PROJECT AKARI

AKARI

知乎 @ 余命数（水无月雫）

中学数学之旅

第二卷·周期·离散·随机

Volume II · PERIOD·DISCRETE·CHANCE

$$\binom{n}{k}+\binom{n}{k+1}=\binom{n+1}{k+1}$$

公开版 · 10 月 5 日

A Project Akari

余命数

知乎@余命数

知乎@余命数

中学数学之旅

知乎@余命数

知乎@余命数

知乎@余命数

知乎@余命数

知乎@余命数

rikka.moe

知乎@余命数

余命数 (Midnight Forever)
知乎 @ 余命数
https://rikka.moe

知乎@余命数

中学数学之旅
An Excursion through Secondary School Mathematics
最后更新：October 7, 2026

知乎@余命数

© 2026 余命数
保留所有权利. 书稿版本，禁止修改、倒卖等任何侵犯作者著作权的行为.

知乎@余命数

知乎@余命数

书稿声明

Draft Notice

本书目前为**书稿状态**，意味着：

- 内容尚未完整，部分章节可能缺失或正在编写中
- 文中可能存在笔误、排版错误或数学推导疏漏
- 章节结构与内容安排可能在后续版本中调整
- 习题答案与索引等附录内容可能尚未完善

若您在阅读过程中发现任何错误或有改进建议，
欢迎通过以下方式联系作者：

个人网站：https://rikka.moe

电子邮箱：me@rikka.moe

书籍主页：https://rikka.moe/books

当前版本：Draft October 7, 2026

知乎@余命数

献词

Dedication

谨以此书

献给星街彗星

愿你如彗星划过星街，把梦唱到银河的尽头

每一个未曾舞蹈的日子
都是对生命的一种辜负

„Und verloren sei uns der Tag,
wo nicht ein Mal getanzt wurde!”

— 弗里德里希·尼采
Friedrich Nietzsche

— 2026 —

知乎@余命数

关于作者

余命数



Project Akari ·Midnight Forever

2025 届广西新高考二卷考生, 政史地组合, 数学 139 分.

这套书整理自我中学阶段学习数学时的笔记与讲义. 我按一条理想的阅读骨架把代数、离散与随机、几何、分析、线性结构依次铺开, 每条结论都给出推导路径, 把概念彼此生长的过程摊开在纸面上. 愿它陪你走完一段中学数学的旅程.

书籍主页

https://rikka.moe/books

更多信息、勘误表、联系方式等均可在此查看

交流与联系

在线平台

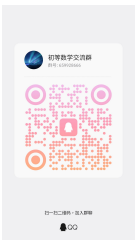
个人主页	`https://rikka.moe`
书籍主页	`https://rikka.moe/books`
知乎	@ 余命数
爱发电	`https://afdian.com/a/writeamathbook`

即时通讯

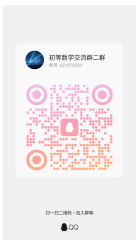
QQ	`2454167821`
交流群	`659928666`
邮箱	`me@rikka.moe`



扫码添加好友



交流群一



交流群二



爱发电赞助

知乎@余命数

欢迎通过以上方式与我交流讨论

知乎@余命数

前言

如果你基于本书制作修改版 PDF, 请保留这里的前言与声明, 谢谢. 想了解更多, 请访问我的主页https://rikka.moe

推动我写这套书的最主要动因是贫穷. 我高中就读于一所普通的省重点, 家里没钱给我补习, 能找到的老师水平也一般.

好在我有一台能联网的手机和一颗热爱学习的心, 于是便用互联网找学习资源, 听网课、看电子书.

这个过程里我攒下大量笔记, 大多记在 Obsidian 上. 高考结束后的 2025 年暑假, 我把这些笔记整理成册, 发到几个群里反响不错, 于是继续写了下去.

我也希望这本书能集百家之长. 市面上的资料和教辅很多, 我想读一本成体系、又贴合自己阅读习惯的书, 于是干脆自己动手写, 权当半个教科书.

这本书花了我不少时间. 如果你觉得有帮助, 欢迎去爱发电https://afdian.com/a/writeamathbook支持一下, 请我吃一碗拼好饭也可以. 书中的疑问或错误, 欢迎随时联系, 邮件 me@rikka.moe、QQ2454167821, 或加入 QQ 群 659928666.

知乎@余命数

吊诡的是, 我觉得教辅太贵, 所以都看电子版.

Obsidian 是一款很好用的笔记软件, 用好了可以打造属于自己的知识库.

关于联系与提问

本人经常在外打工兼职或忙别的事, QQ 私聊不一定看得到, 重要的事情请发知乎私信. 已经写清楚的问题我一般不再重复, 比如分卷本与合订本的区别、各版本之间的区别.

学习方法我也说不出什么有用的. 你观察过学得好的同学就会发现, 无论被谁请上台分享经验, 大多只能说些政治正确的话, 扎实基础、回归课本、多跟老师. 这怪不得他们, 真正掌握一项技能后, 就很难想象不懂它是什么感觉, 一项已经内化的能力, 很难再拆开讲给你听.

即便方法就是“极度的自律、极高的抗挫折能力、成百上千小时的枯燥重复”, 即便有些人同时有智商与基因、家庭托底和超强的专注力与抗挫折力, 学习也没有廉价速成的魔法. 每次有人问我, 我也只能给出和所有人一样的回答, 巩固基础、回归课本、多跟老师、总结错题.

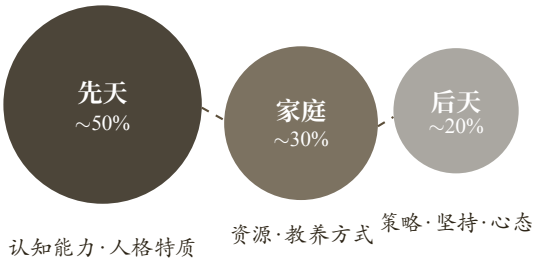
用卖课的人的话说, 就是欢迎多多支持我的项目和电子书. 所以这方面的问题, 就不必再问了.

不过, 考虑到还是有人想知道学习方法. 很好, 下面从比较科学的角度出发, 拒绝政治正确和哔哩哔哩经典学长或学姐分享学习方法的套话, 来说说好的学习方法.

学习成绩的好坏可以认为取决于三大类因素: 先天禀赋、家庭条件和后天努力. 下图用圆的大小表示各自的影响占比, 圆越大代表该因素对成绩差异的解释份额越多.

占比数据综合自 Silventoinen et al. (2020) 在 Scientific Reports 上发表的 28 国双胞胎整合分析(遗传力约 40%–70%, 共享环境约 20%–40%)与 de Zeeuw et al. (2015) 的荷兰双胞胎元分析; 本文取中位数附近的近似值.

知乎@余命数



知乎@余命数

图中三个圆分别代表：

- **先天**，约占一半：认知加工速度、工作记忆容量等人脑硬件参数，以及尽责性、抗压性等人格倾向. 这部分相对稳定，但并非锁死. 早期营养、睡眠和运动会影响其发挥. 有些人的确在某些方面的所谓的感觉更好，又或者说，假如你在婴幼儿时期，甚至还没出生的时候，就接受到了一些对神经系统不太好的物质，这都是实打实的.
- **家庭**，约占三成：家里的藏书量、父母的词汇量和教育期望、能否提供安静的学习环境与课外资源. 家庭条件好在早期尤为关键，随着年级升高其权重逐渐下降. 这里也比较显然，如果家庭是那种书香家庭，父母都是文化人，或者都是中产阶级甚至是富人阶级，请得起清华北大的私教一对一，买学区房、上重高等等，获取好的成绩的概率更大、更轻松. 如果家庭是那种一天到晚都在吵架、家庭暴力、父母没文化而且也不怎么有钱（遗憾的是编者的家庭都是如此，因此我付出了很大的努力），很大概率最后都是成绩一般甚至很差，成为混混或其他.
- **后天**，约占两成：学习方法是否科学、能不能坚持、面对难题时的心态. 虽然占比最小，却是个人最能把控的部分，也是拉开同等条件下人与人差距的主战场. 这也就是唯一可以努力的地方，除非你是家长，否则对于学生，也就是我们而言，只有此路！

虚线表示三类因素之间存在交互，即先天禀赋需要在合适的家庭环境中才能充分发挥，而后天努力的效率又受限于先天的认知基础和家庭提供的资源.

把可控因素提炼出来，成绩的近似模型可以写成：

$$\text{成绩} \approx \underbrace{(\text{认知基础} + \text{尽责性})}_{\text{较稳定}} \times \underbrace{(\text{科学策略} + \text{自我调节})}_{\text{高度可控}} + \underbrace{\text{家庭支持} + \text{教学质量}}_{\text{环境系数}} \tag{0-1}$$

说说，公式里用的是乘法和加法，含义各有侧重.“较稳定”部分与“高度可控”部分之间是乘号，即便先天基础只有六七分，只要方法和自律跟上，乘积依然可观；反过来基础再好，方法为零则乘积也为零. 加在最外面的“环境系数”是加分项，好的家庭和老师能把整体水平往上抬一截，但它们替代不了主体部分的乘积.

Poropat (2009) 的元分析，样本逾七万人，显示尽责性是大五人格中对学业成绩最强的非智力预测因子，其对大学成绩的预测效力与智力相当.

ACE 模型源自行为遗传学经典范式：A 为加性遗传，C 为共享环境，E 为非共享环境.Lee et al. (2018) 在 Nature Genetics 上发表了百万级样本的 GWAS，其多基因评分解释受教育年限约 11%–13% 的变异.

Hart & Risley (1995) 在 Meaningful Differences 中跟踪观测了 42 户家庭三年，发现专业人员家庭幼儿在三岁前累计听到约 3000 万词，领取福利金的家庭约 1000 万词，“三千万词鸿沟”由此得名；后续研究对其取样和方法有争议，但方向一致.

Duckworth & Seligman (2005) 在 Psychological Science 上对八年级生的纵向研究中发现，自控力对期末成绩的方差解释量约为 IQ 的两倍有余.

知乎关注@余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

“科学策略”指的是有实证支撑的学习技术. 合上书本逼自己回忆的主动检索, 以及在快要忘记时复习的间隔重复, 其效果远大于被动重读和划重点.“自我调节”涵盖制定计划、监控进度和在出错时及时纠正方向的能力.

先天和家庭设定了起跑线和跑道状况, 后天的高质量努力决定你能在这条道上跑多远.

所以, 动用你聪明的小脑袋瓜, 评估自身的情况, 决定从何入手, 别去问别人的学习方法. 你首先是你自己的, 如果你自己都不知道自己要做什么, 就不要偷懒问别人了.Like WTF, 如果你是个普通家庭, 你去问一个有钱人出身的竞赛背景的学霸, 问学习方法, 这本身就荒谬.

当然, 还有一个很好的学习习惯, 那就是去爱发电支持作者:), 推出更多的学习资源, 助你一臂之力.

Dunlosky et al. (2013) 在 Psychological Science in the Public Interest 上评定了十种常用学习技术, practice testing 与 distributed practice 获评高效用, rereading 与 highlighting 获评低效用.

知乎@余命数

如何获取无水印版

本书的基础版本， 含水印免费开放， 任何人都可以在我的主页`https://rikka.moe/books`下载. 如果你想获取无水印版, 可以在爱发电上支持本项目:

1. 打开爱发电页面:`https://afdian.com/a/writeamathbook`
2. 选择对应的赞助档位(通常会标注“无水印版”或“完整版”), 完成赞助
3. 赞助后在爱发电的消息里会收到无水印版 PDF 的下载链接; 如果没有收到, 请发邮件或在知乎私信告诉我你的爱发电昵称

无水印版与免费下载版的**内容完全一致**, 差别只在水印. 你不花钱也能读到一模一样的数学内容;付费只是为了去掉水印、支持项目运转.

既然可以免费下到, 为什么要花钱买?

我一直用中世纪或文艺复兴时期王公贵族资助艺术家来比喻这种捐赠的模式, 或者更恰当的, 是 18-19 世纪文人的集资预订, 又或者是数字公共品理论。

许多高中生(包括我)因为高中的时候穷, 没钱, 所以喜欢白漂电子书资料。毕竟在我那时候, 任何没有 DRM(数字版权管理技术)的电子书都能被随意复制粘贴和传播, 盗版网课也有很多。很多时候不是不想支持正版, 而是没钱。与其到处找资源, 不如加成千上万个资料群, 去点那些夸克网盘(天下没有免费的午餐, 夸克网盘对分享资料的人也是有利可图的), 下一堆资料。虽然下了或者存了以后, 可能没有时间看或者基本不看, 但至少满足了收集欲, 以及一种“我在学习”的感觉。

但是, 与其在网盘里存 500G 的名师网课和状元笔记, 或者花重金买各种名师的网课和资料, 不如把钱留给真正对你有用的东西。如果你觉得 “这本书确实对我有用”、写得好, 或者在其他方面给你带来了价值, 不妨去我的爱发电`https://afdian.com/a/writeamathbook`花 10～30 块来支持我, 并获得无水印版。

其实 10 块钱，吃一次拼好饭差不多就花完了。所以我想了一句广告语：用一次拼好饭的钱，来支持下一位创作者。即使实际上你可以白漂（毕竟这是公共品，一个人的占用不会导致另一个人占用的减少），但你的支持，能让互联网上的创作者感到生活更加美好。

更重要的是，你今天怎么对待创作者，就是你希望将来别人怎么对待你。总有一天，你也会在网上创作：可能是一篇文章，一门课。到那时你会明白，被人认可的感觉有多重要，哪怕对方只付了一次拼好饭的钱。

公共品之所以能一直存在，是因为总有一些人，在可以不付钱的时候，仍然愿意为自己认可的东西付一点点。创作者因此能继续创作，受众也因此能继续读到好东西，这才是真正的双赢。而你今天的这一点点，就是在把这个循环往下传。

知乎 知乎@余命数

试用 Akari Web 平台

知乎@余命数

除了纸质书之外，Project Akari 还在做一套在线学习工具，即 Akari Web 平台（`https://kirie.io`）.

Project Akari 是一项旨在结合人工智能教育、教育数据挖掘、学习分析，并基于认知心理学的研究成果，搭建一款旨在帮助学生高效地学习数学和计算机科学方面知识的智能辅导系统，或称为学习引擎.

余命数 知乎@余命数 知乎@余命数

因此，我们的目的就是，尽可能量化学生的知识状态，并结合算法，智能评价和判断学生目前的学习状态，辅助其真正习得和掌握一样东西.

我们的哲学是，让用户用效率最高的方式学到知识，为此我们开发了一整套配合的算法，并在课程等设计上尝试最小化认知负担，采取更互动的方式以激发主动性. 量化学习的一切，并且优化，便是一切的出发点和落脚点.

知乎@余命数

这本书写给谁

知乎@余命数

本书的目标读者有三类：

- **在校高中生**：跟着课堂教学同步使用，课上没听懂的章节回来对着书再看一遍推导. 标题带 $*$ 的选读内容可以先跳过，等基础扎实了再回头啃.
- **复读生或自学备考者**：按目录顺序系统过一遍，重点关注每节的例题和章末习题. 如果你的时间紧张，优先读函数、三角、数列和解三角形四章，即高考分值密集的区域.
- **想重温数学的成年人**：你可能忘了大半高中内容但又不想从初中补起. 挑感兴趣的章节直接读就好，遇到前置知识不清楚的地方顺着交叉引用回去翻.

不管你是哪种读者，建议的做法都一样，先看正文搞懂逻辑链，再做例题（盖上解答自己想），最后做章末习题检验. 光看不练等于白看.

知乎关注@余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

本书的结构

各卷可独立阅读，也可合订通读. 分卷本在文前附有本卷的分册前言与本卷导读；合订本在每卷分隔页之后给出本卷导读，便于按卷回看. 标题中带 * 的章节多为选读或拓展，初读时可略，需要时再补.

关于版本与系列

本书目前仍是书稿，个别地方尚未完整，可能存在笔误或疏漏，若发现错误或有建议，欢迎联系作者. 本版本是 2026 年按学习路径重排的重要更新，调整了五卷章序，把数列章尾的分析内容并入极限与导数，把平面向量提到解三角形之前.

本书属于 Project Akari 系列.Akari 取自日文“あかり”，意为光，寓意是让好的学习材料更容易被找到、更容易被用起来.

Project Akari 的另一部书稿《函数与导数》将函数、极限与导数相关内容单独成书，系统展开. 两本书可以独立阅读，也可以互相参照. 配套的 Akari Web 平台（`https://kirie.io`）提供笔记、记忆卡片和在线练习，与书稿配合使用.

本系列的基础版本免费开放，完整版本以极低价格提供. 知识不应该是奢侈品，这是我做这件事的底线.

从写作、排版到软件开发，Project Akari 的每一步都是我一个人完成的. 服务器和运营成本靠书的销售收入以及我个人的工作收入来覆盖. 这个项目既是学习方法的实践，也是工程能力的验证.

发现错误怎么办

一个人的精力和眼力终究有限，书里难免有笔误、推导疏漏乃至概念性错误. 如果你发现了，请按以下方式反馈，我能更快地定位和修正：

1. 说明在哪一卷、哪一章、哪一节，或直接给页码和小标题
2. 引用书中的原文或公式编号
3. 写出你认为正确的版本或质疑的理由

最快的渠道是知乎私信 `https://www.zhihu.com/people/yushumi-shu`，其次是邮件 `me@rikka.moe`. 有价值纠错的读者名字会列入后续版本的鸣谢名单.

致谢

这本书能走到今天，离不开很多人的帮助.

首先要感谢每一位读过草稿、提出意见的同学和朋友. 你们在群里随手指出的一道错题、一处不通顺的表达，聚沙成塔地把这本书往前推了一步.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

感谢爱发电上的赞助者们. 你们的每一笔支持都实实在在地变成了服务器费用和我碗里的拼好饭, 让我能在打工之余腾出手来继续写.

感谢家人容忍我把无数个周末和假期耗在了电脑屏幕前. 虽然你们不完全理解我在做什么, 但从没拦过我.

最后感谢你, 正在读这段字的陌生人. 一本书的价值在被翻开的那一刻才算兑现.

版权与授权

本系列的基础版本依据**知识共享署名-非商业性使用-禁止演绎 4.0 国际协议**(CC BY-NC-ND 4.0) 开放给个人学习者免费使用. 完整版本以极低价格提供, 销售收入用于覆盖服务器与运营成本. 商业使用、机构采购或翻译授权请联系作者.

防盗版与使用声明

本书为 Project Akari 系列出版物. 本 PDF 仅供合法获取者个人学习使用, 禁止以下行为:

- 将本 PDF 或其任何部分上传至任何公开网络平台、网盘、论坛或即时通讯群组;
- 对本 PDF 进行二次售卖、捆绑销售或用于任何商业用途(已获作者书面授权者除外);
- 将本书内容用作机器学习或人工智能训练语料;
- 移除、涂改或遮盖本书的水印、版权信息、读者指纹或页脚标识;
- 修改 PDF 元数据后以原书名、作者名重新分发.

如发现盗版或侵权行为, 欢迎通过以下渠道举报:

- 书籍主页:https://rikka.moe/books
- 邮箱:me@rikka.moe

版本指纹(用于溯源与正版核验):

读者标识 DRAFT-REVIEW

文档哈希 DRAFT-REVIEW

编译日期 October 7, 2026

关于修改与再分发的声明

本书由作者独自维护并发布, 是本项目唯一的**权威正版来源**. 任何对本 PDF 或其源文件的内容、水印、版权信息、读者指纹、PDF 元数据的删改、涂改或遮盖, 均视为对本声明与著作权法的违反.

修改后的版本**不得**保留以下任一标识并以任何形式分发, 否则视为侵权:

- 书名《中学数学之旅》及 Project Akari 系列名称;
- 作者署名"余命数"及相关社交账号;
- 版权页、前言中任何声明性文字;
- 页脚隐性指纹与页面水印.

欢迎读者基于本书内容做学习笔记与摘录，但请在分发时注明来源与原作者. 如需做系统性改编，如翻译、缩写、教材化使用，请先联系作者取得书面授权.

正版核验与联系方式见版权页，或访问 https://rikka.moe/books.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

分册前言

学完第一卷的函数语言后，我把视线转到三类新的现象：圆周上的周期变奏、正整数上的离散递推、以及数据与随机度量. 第二卷就围绕这三件事展开.

三角函数从弧度与单位圆建立基本量与恒等式，并通向解三角形. 数列章聚焦归纳法、等差等比、递推与通项等**初等**工具；严格的 ε-N 极限、级数与泰勒公式已迁入第四卷分析主干，避免在数列章末预支微积分. 统计、计数与概率则串起描述、可数模型与随机度量.

把 ε-N 留到第四卷是后来才决定的事. 最早一稿把它放在数列章末尾，结果发现数列章就被这半本数学分析压垮了.

本卷主线是：连续的周期、离散的递推、在计数之上给随机事件分配度量. 它把第一卷的函数与不等式接到更广阔的建模场景，并为第三卷的几何主干留下接口.

数列章中偶有涉及 ε-N 极限、级数与泰勒之处，相应位置都注明了出处在第四卷. 基础轨读者可先按初等递推与计数理解，完整轨读者学完第四卷后回看这些旁注.

本书为书稿状态，内容尚不完善，可能存在笔误或疏漏. 如有问题或建议，欢迎联系作者 `https://rikka.moe`，或发送邮件至 `me@rikka.moe`.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

本卷导读

本卷线索

第二卷把周期、离散与随机串在同一条函数与度量的线索上. 三角描述圆周上的连续振动, 数列描述正整数上的递推演化, 计数为随机事件提供可数的模型空间, 统计与概率则分别从数据与公理两端逼近不确定性.

章节地图

- ▷ **三角函数**　从弧度与单位圆建立正弦、余弦等基本量, 展开图像与性质、同角关系、诱导与和差倍角. 单位圆是推导恒等式的支点, 这里的图像与变换同第一卷的函数图像、复合变换一脉相承; 三角函数既是几何量, 也是第三卷解三角形的公共语言.
- ▷ **数列**　以数学归纳法与 $\sum / \prod$ 记号为工具, 研究等差、等比、差比与斐波那契等模型, 进而讨论递推、通项、不动点、单调与周期性. 归纳法是本章论证的基石, 递推模型的每一步分析都建立在它之上.
- ▷ **统计**　从描述数据到抽样设计, 再到用样本推断总体.
- ▷ **计数原理**　加法与乘法原理是计数的起点, 由此建立排列与组合, 再以二项式定理收束.
- ▷ **概率**　在计数提供的可数模型上, 用事件与概率公理刻画不确定性, 并讨论条件概率、独立性与错排等经典问题.

统计、计数、概率三章依次沿描述、计数、度量推进, 合起来构成从处理数据到公理化不确定性的完整链条.

向后展望

第三卷进入几何主干, 平面向量先行, 再解三角形与解析几何, 参数与极坐标紧贴坐标法. 第四卷建立极限与导数的严格理论, 数列的极限行为、级数与泰勒展开以及概率中的渐近结果在那里获得完整处理.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

目录

关于作者 i

前言 iii

分册前言 xi

本卷导读·第二卷 xiii

第一章 三角函数 2

1.1 弧度制 . . . 3
1.2 正弦函数与余弦函数 . . . 5
1.3 正弦与余弦函数的图像 . . . 9
1.4 正弦函数与余弦函数的性质 . . . 11
1.5 正切函数 . . . 14
1.6 余切、正割与余割函数 . . . 18
1.7 同角三角函数的基本关系 . . . 22
1.8 诱导公式 . . . 28
1.9 和差角公式 . . . 35
1.10 倍角公式 . . . 42
1.11 万能公式与齐次式思想 . . . 53
1.12 角的变换 . . . 61
1.13 和差化积与积化和差 . . . 65
1.14 $y = A\sin(\omega x + \varphi) + B$ . . . 73
1.15 ω 的约束与求解 . . . 82
1.16 反三角函数 . . . 102

第二章 数列 110

2.1 数学归纳法 . . . 110
2.2 $\sum$ 和 $\prod$ 符号 . . . 114
2.3 数列的表示 . . . 126
2.4 等差数列 . . . 127
2.5 等比数列 . . . 131
2.6 差比数列 . . . 133
2.7 斐波那契数列 . . . 147
2.8 数列求和的通用策略 . . . 158
2.9 递推关系与通项公式 . . . 191

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.10 一阶递推的基本形式 . . . 196
2.11 迭代与不动点 . . . 207
2.12 数列的单调性 . . . 215
2.13 奇偶数列 * . . . 218
2.14 数列的周期性 . . . 220
2.15 子列与 Bolzano-Weierstrass 定理 * . . . 225

第三章 统计 **230**
3.1 描述数据 . . . 231
3.2 抽样方法 . . . 234
3.3 从样本到总体 . . . 241
3.4 相关与回归 . . . 244
3.5 章末习题 . . . 249
习题 . . . 249

第四章 计数原理 **252**
4.1 加法原理与乘法原理 . . . 253
4.2 组合与排列 . . . 263
4.3 二项式定理 . . . 268
4.4 从三项式定理到多项式定理 * . . . 273
4.5 排列组合综合应用 . . . 280
习题 . . . 310

第五章 概率 **314**
5.1 事件与概率 . . . 314
5.2 经典概率问题 . . . 329
5.3 条件概率 . . . 338
5.4 事件的相互独立性 . . . 343
5.5 全概率公式与贝叶斯公式 . . . 348
5.6 离散型随机变量及其分布列 . . . 355
5.7 期望与方差 . . . 358
5.8 常见分布模型 . . . 362
习题 . . . 367

习题解答 **369**
第 3 章统计 . . . 369

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

第 1 章

三角函数

基础语言

「几何无王者之路。」

——欧几里得

第一章 三角函数

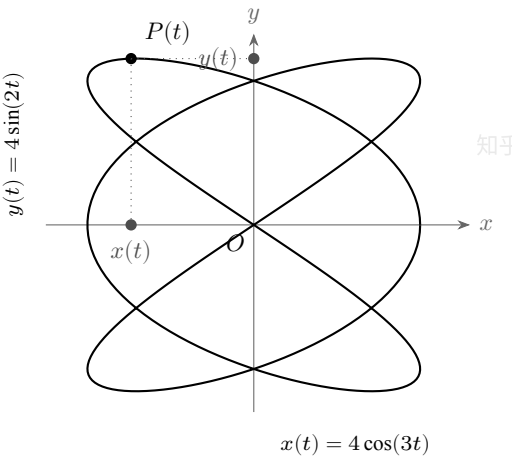


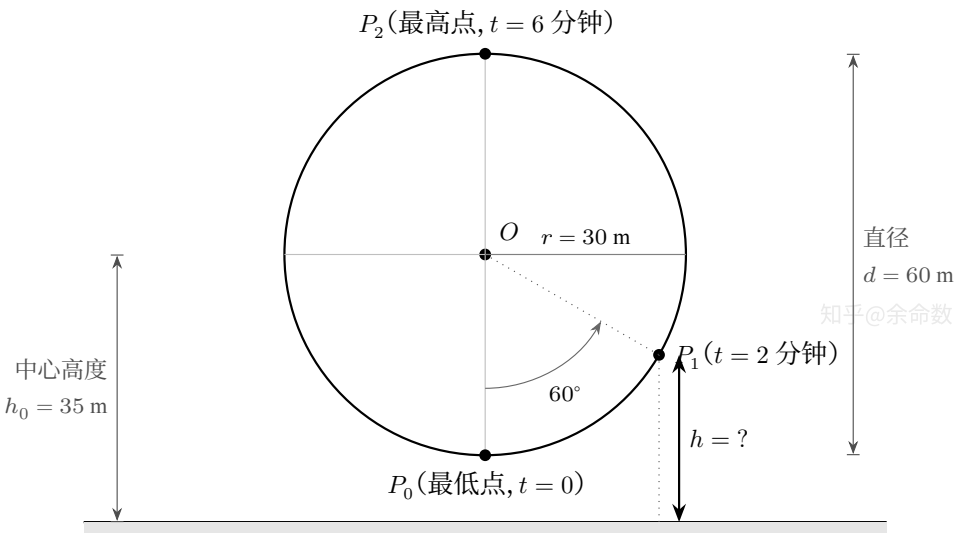
图 1.1 利萨茹曲线:两个独立三角振动的叠加

屏幕上的光点同时参与两种独立的运动. 一个余弦波控制它水平往复振动,另一个正弦波控制它垂直往复振动. 两种运动叠加后, 光点描出稳定、封闭而复杂的图案, 我们称之为利萨茹曲线. 图案的最终形态由频率比与初相位决定. 同频同相得直线段, 同频异相得椭圆(相位差为 π 时退化为直线段), 频率比为简单整数比时得封闭的多环图案. 改变水平或垂直振动的频率或起始时机, 图形随之变形. 这些现象直接体现了三角函数的周期性与相位性质.

探索课题

我们考虑一座摩天轮. 设它的半径为 30 米, 其中心点距离地面 35 米. 你从摩天轮的最低点坐上座舱, 摩天轮以恒定的角速度开始逆时针转动, 并且每 12 分钟转一圈.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵



问题

1. 启动 2 分钟后, 你距离地面的高度是多少?
2. 启动 8 分钟后, 你距离地面的高度又是多少?

我们已知最低点高度为 $35-30=5$ 米, 最高点为 $35+30=65$ 米. 从最低点到最高点需要 $12/2=6$ 分钟. 你可能会想, 既然 6 分钟升高了 60 米, 那么 2 分钟后高度就是 $5+\frac{2}{6}\times 60=25$ 米. 请仔细观察上图. 当你在 P_1 点时, 你上升的垂直高度与沿圆弧走过的路程不成正比. 你的垂直速度在不断变化 (在最低点和最高点最慢, 在与中心等高处最快). 这说明高度随时间的变化不是线性关系, 我们需要一种新的函数模型来描述这种周而复始的非线性变化.

在此之前, 我们遇到的函数, 如一次函数、二次函数、指数函数、对数函数大多描述的是增长、衰减或简单的 V 形、U 形变化. 然而现实世界中充满着**周期性现象**, 这些循环往复的运动, 无法用旧有的函数模型来刻画.

三**角函数**正是为描述周期性现象而建立的数学语言. 本章将超越直角三角形内的边角比, 从单位圆上的坐标这一更广阔的视角重新定义三角函数, 进而系统研究其**图像特征**与**性质**, 并借助解析式 $y=A\sin(\omega x+\varphi)+k$ 分析函数的行为, 解决与周期、频率、相位相关的实际问题.

1.1 弧度制

来龙去脉

我们先用度作角的单位. 把圆周分成 360 等份, 每份对应 1°. 古巴比伦人选择 360, 因为它能被许多整数整除, 便于天文、航海和日常计算.

知乎@余命数

360 出自历史约定, 与圆的半径、周长等几何性质无关. 角度制下许多公式因此带一个转换系数 $\frac{\pi}{180}$, 表达式不够简洁.

更自然的做法是用圆自身的量来度量角, 即取弧长与半径的比值. 这个比值就是**弧度制**.

1.1.1 定义

定义 1.1 (弧度) 在圆中, 我们规定**长度等于半径长的圆弧, 它所对的圆心角的大小为 1 弧度**. 设一段弧长为 l, 它所对的圆心角为 α(单位为弧度), 所在圆的半径为 r. 那么角 α 的大小由弧长与半径的比值唯一确定.

$$\alpha = \frac{l}{r} \tag{1-1}$$

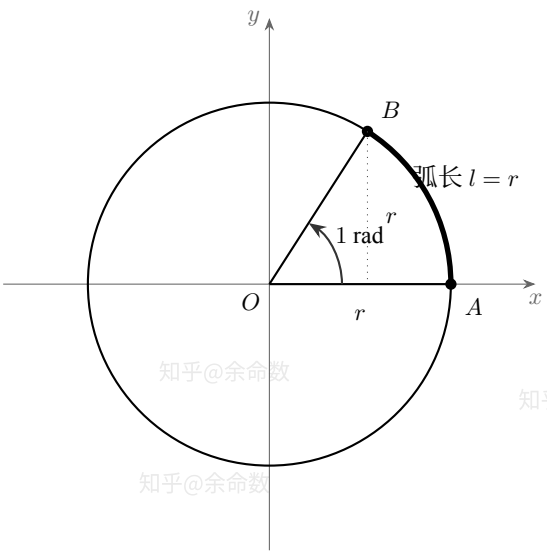


图 1.2 弧度的定义. 当弧长等于半径时, 所对应的圆心角为 1 弧度

弧度制是一个纯粹的**比值**, 没有单位(也可说单位为 1). 角的大小因此与所选圆的半径无关, 成为一个只与形状有关的纯数, 能在数学公式中与其他物理量直接、协调地参与运算.

1.1.2 角度制与弧度制的换算

我们先以周角为锚点进行换算. 一个完整圆周的弧长 $l = 2\pi r$. 由弧度制定义, 它对应的弧度角为

$$\alpha = \frac{l}{r} = \frac{2\pi r}{r} = 2\pi(\text{弧度}).$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

一个周角等于 360°, 所以

$$360^\circ = 2\pi \text{ rad} \quad \Longleftrightarrow \quad 180^\circ = \pi \text{ rad} \tag{1-2}$$

由此得到任意角度的换算公式.

- **角度化弧度**.$1^\circ = \frac{\pi}{180}$ rad, 那么 $n^\circ = n \cdot \frac{\pi}{180}$ rad.
- **弧度化角度**.$1 \text{ rad} = \left(\frac{180}{\pi}\right)^\circ \approx 57.3^\circ$.

常用特殊角的换算

我们把这些特殊角的弧度值列在下表中.

角度制	0°	30°	45°	60°	90°	180°	270°	360°
弧度制	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π

从现在起, 在讨论函数、方程等数学问题时, 凡未指明单位的角度, 一**律默认为弧度制**. 例如, $\sin 2$ 指的是 $\sin(2 \text{ rad})$, 不是 $\sin(2^\circ)$. 这是高中和高等数学的标准约定.

1.2 正弦函数与余弦函数

在初等几何中, 我们最初在**直角三角形**内定义正弦、余弦.

$$\sin A = \frac{\text{对边}}{\text{斜边}}, \quad \cos A = \frac{\text{邻边}}{\text{斜边}}.$$

这个定义只适用于锐角 $A \in (0, \frac{\pi}{2})$. 它能处理直角三角形的边角计算, 却不能说明 $\sin\frac{2\pi}{3}$ 或 $\cos(-\frac{\pi}{6})$ 的意义, 更无法把正弦、余弦看成定义在 $\mathbb{R}$ 上的函数.

要描述任意大小的旋转, 例如摩天轮或行星轨道, 就要突破直角三角形的限制. **单位圆**把角 α 的终边与圆上点的坐标直接联系起来.$\sin\alpha$ 是该点的纵坐标, $\cos\alpha$ 是该点的横坐标. 这样, 定义域从 $(0, \frac{\pi}{2})$ 扩展到整个实数集 $\mathbb{R}$, 周期性也直接来自圆周运动.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.2.1 单位圆上的参数化

定义 1.2 (任意角的三角函数) 在平面直角坐标系中, 我们构造一个以原点 O 为圆心, 半径 $r = 1$ 的圆, 称之为**单位圆**, 其方程为 $x^2 + y^2 = 1$. 设角 α 的顶点与原点重合, 始边与 x 轴的非负半轴重合. 其终边与单位圆交于唯一点 $P(x, y)$. 那么点 P 的坐标与角 α 之间的关系如下.

- 点 P 的**纵坐标** y 称为角 α 的**正弦** (Sine), 记作 $\sin\alpha = y$.
- 点 P 的**横坐标** x 称为角 α 的**余弦** (Cosine), 记作 $\cos\alpha = x$.

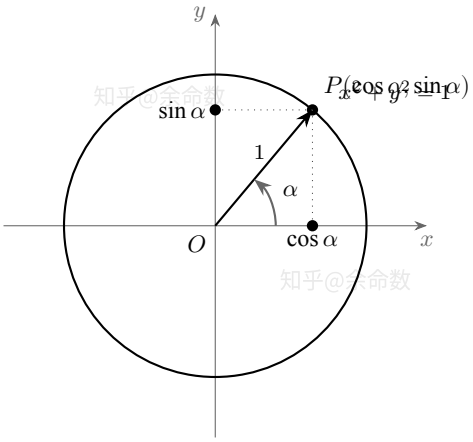


图 1.3 单位圆中三角函数的定义

由单位圆方程 $x^2 + y^2 = 1$ 可得, 对任意角 α 都有

$$\sin^2\alpha + \cos^2\alpha = 1.$$

这个等式就是同角三角函数的平方关系, 它的几何意义是单位圆上任意一点到原点的距离为 1.

欧拉公式预告

注 1.1 (欧拉公式的启示)
在复分析领域, 欧拉公式说明了一个更为精细的联系. 它给出 $e^{i\alpha} = \cos\alpha + i\sin\alpha$. 因此, 复数 $e^{i\alpha}$ 是单位圆上辐角为 α 的点. 正弦与余弦就是这个基本复指数函数在实轴与虚轴上的投影, 它统一了指数函数与三角函数.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.2.2 图像的生成

$\sin x$ 和 $\cos x$ 已经是单位圆上动点的纵、横坐标. 把它们对应的坐标值沿角度 x 展开, 就得到一维的函数图像.

我们让质点 P 从 $(1,0)$ 出发沿单位圆逆时针运动.

- 把 P 的**纵坐标** $y=\sin x$ 投影到水平数轴上, 得到**正弦曲线**;
- 把 P 的**横坐标** $x'=\cos x$ 投影到同一条水平数轴上, 得到**余弦曲线**.

这两条曲线是最基本的波动形态.

知乎@余命数

1.2.3 性质的几何根源

知乎@余命数

正弦与余弦函数的性质都可以从单位圆直接读出.

1.2.3.1 定义域与值域

- **定义域**. 对于任意实数角 $x\in\mathbb{R}$, 其终边总能在单位圆上唯一确定一个点 P. 因此点 P 的横纵坐标 $\cos x$ 和 $\sin x$ 总是有定义的, 故 $y=\sin x$ 和 $y=\cos x$ 的定义域均为 $\mathbb{R}$.
- **值域**. 单位圆上的点, 其横坐标 x 和纵坐标 y 的取值范围都被限制在 $[-1,1]$ 之内, 故 $y=\sin x$ 和 $y=\cos x$ 的值域均为 $[-1,1]$.

1.2.3.2 周期性

当角 x 增加 2π(一个完整的圆周)时, 其终边回到原来的位置, 单位圆上的点 P 也重合. 于是点的坐标值将重复出现.

$$\sin(x+2\pi)=\sin x,\quad \cos(x+2\pi)=\cos x$$

因此, 2π 是它们的一个周期. 我们还可以验证, 这也是它们的最小正周期. 周期性是圆周运动最根本的属性.

知乎@余命数

1.2.3.3 奇偶性与对称性

我们考察角 x 与角 $-x$ 的终边. 它们在单位圆上所对应的点 $P(\cos x,\sin x)$ 与 $P'(\cos(-x),\sin(-x))$ 关于 x 轴对称.

由对称性可知, 点 P 与 P' 的横坐标相同, 纵坐标互为相反数. 据此得到

$$\cos(-x)=\cos x\quad(\text{横坐标不变})$$

$$\sin(-x)=-\sin x\quad(\text{纵坐标反号})$$

因此, $y=\cos x$ 是**偶函数**, 其图像关于 y 轴对称; $y=\sin x$ 是**奇函数**, 其图像关于原点对称.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

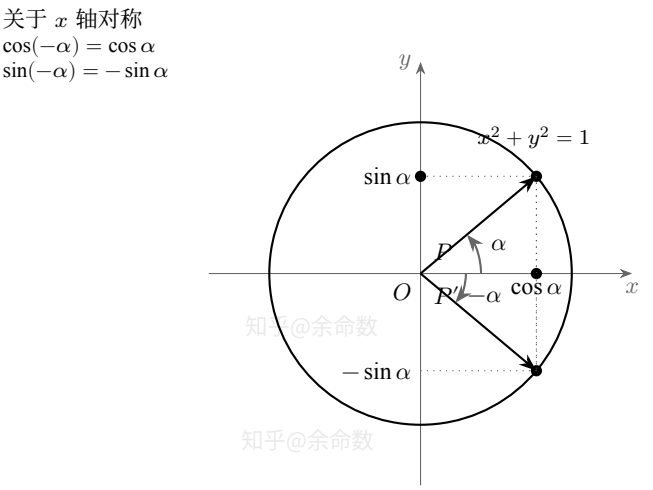
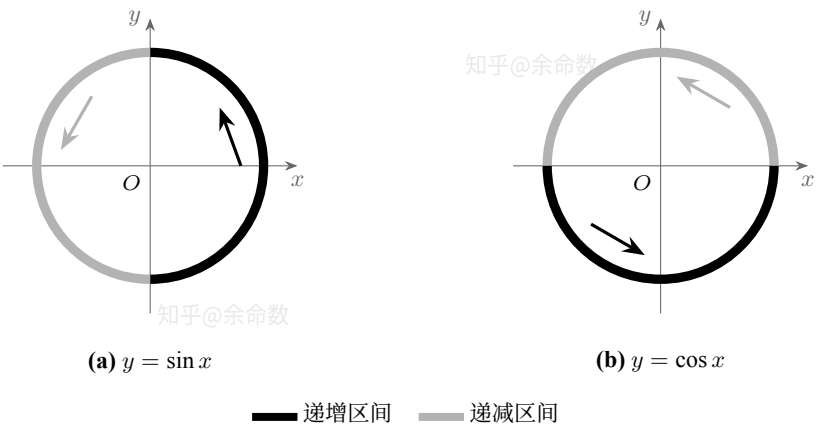


图 1.4 三角函数的奇偶性

1.2.3.4 单调性

函数的单调性完全取决于动点 P 在单位圆上运动时,其横纵坐标的变化趋势.



(a) $y=\sin x$ **(b)** $y=\cos x$

图 1.5 三角函数在单位圆上的单调性

- **对于** $y=\sin x$**(考察纵坐标)**. 当角 x 从 $-\frac{\pi}{2}$ 运动至 $\frac{\pi}{2}$ 时, 动点 P 在单位圆的右半圆上从最低点运动至最高点, 其纵坐标 y 从 -1 单调地**增加**至 1. 当角 x 从 $\frac{\pi}{2}$ 运动至 $\frac{3\pi}{2}$ 时, 动点 P 在单位圆的左半圆上从最高点运动至最低点, 其纵坐标 y 从 1 单调地**减少**至 -1.
- **对于** $y=\cos x$**(考察横坐标)**. 当角 x 从 0 运动至 π 时, 动点 P 在单位圆的上半圆上从最右点运动至最左点, 其横坐标 x' 从 1 单调地**减少**至 -1. 当角 x 从 π 运动至 2π 时, 动点 P 在单位圆的下半圆上从最左点运动至最右点, 其横坐标 x' 从 -1 单调地**增加**至 1.

我们将这些观察与周期性结合, 就能得到它们在整个定义域上的所有单调区间.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

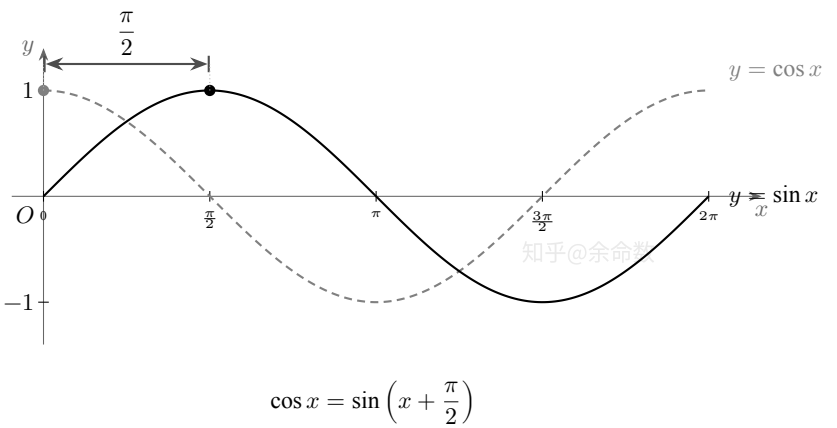
1.2.4 正弦与余弦的内在联系

正弦与余弦的图像形状完全相同，只是位置错开. 下面我们用一个诱导公式说明这种关系.

定理 1.1 (诱导公式) 对于任意角 x, 恒有

$$\cos x = \sin\left(x + \frac{\pi}{2}\right)$$

证明 我们用几何方法证明. 在单位圆上，将角 x 的终边逆时针旋转 $\frac{\pi}{2}$，得到角 $x + \frac{\pi}{2}$ 的终边. 设角 x 终边上的点为 $P(\cos x, \sin x)$，则角 $x + \frac{\pi}{2}$ 终边上的点为 $P'(-\sin x, \cos x)$. 根据定义, $\sin(x + \frac{\pi}{2})$ 就是点 P' 的纵坐标, 即 $\cos x$. □



$$\cos x = \sin\left(x + \frac{\pi}{2}\right)$$

图 1.6 正弦函数与余弦函数的相位关系.$\cos x$ 是 $\sin x$ 向左平移 $\frac{\pi}{2}$ 个单位

余弦函数可以由正弦函数向左平移 $\frac{\pi}{2}$ 得到, 两者描述的是同一类波动, 只是相位不同. 余弦比正弦超前 $\frac{\pi}{2}$, 正弦比余弦滞后 $\frac{\pi}{2}$. 它们都是简谐运动的基础.

1.3 正弦与余弦函数的图像

从单位圆到函数图像

当角 x 从 0 变到 2π 时, 单位圆上的点 $P(\cos x, \sin x)$ 刚好转一周. 在这个过程中, P 的纵坐标 $y = \sin x$ 和横坐标 $x' = \cos x$ 都在周期性变化.

我们把这种圆周运动画成函数图像. 让单位圆沿 x 轴正方向滚动, 同时把 P 的纵坐标投影到对应的 x 轴位置上. 投影点描出的轨迹就是正弦函数的图像.

这条曲线称为**正弦曲线**, 是最基本的波动形态.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

1.3.1 五点作图法

画一条曲线需要很多点. 先标出一个周期内的最大、最小和零点, 就能快速勾出轮廓. 对 $y=\sin x$ 在 $[0,2\pi]$ 上, 这五个关键点是

- **起点(零值点)**.$(0,0)$
- **最高点**.$(\frac{\pi}{2},1)$
- **中间点(零值点)**.$(\pi,0)$
- **最低点**.$(\frac{3\pi}{2},-1)$
- **终点(零值点)**.$(2\pi,0)$

用光滑曲线连接这五点, 就得到一个周期的正弦图像. 由于函数以 2π 为周期, 把这个波形向左右重复延伸, 即得完整图像.

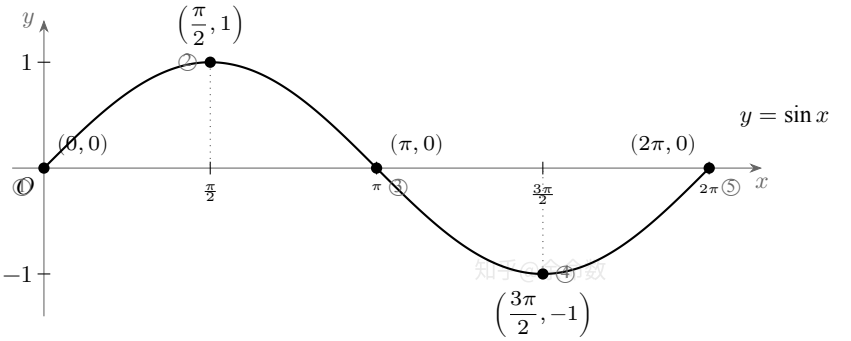


图 1.7 正弦函数 $y=\sin x$ 在 $[0,2\pi]$ 上的图象及五个关键点

1.3.2 余弦函数的图像

同样地, 把单位圆上点 $P(\cos x,\sin x)$ 的**横坐标** $\cos x$ 沿角度展开, 就得到**余弦曲线** $y=\cos x$. 它在 $[0,2\pi]$ 上的五个关键点是 $(0,1),(\frac{\pi}{2},0),(\pi,-1),(\frac{3\pi}{2},0),(2\pi,1)$.

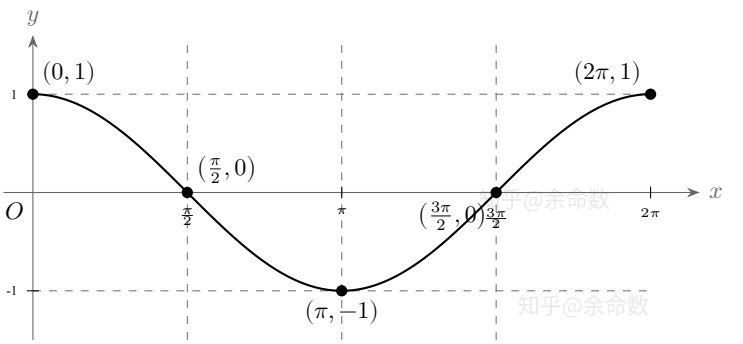


图 1.8 函数 $y=\cos x$ 在 $[0,2\pi]$ 上的五点法作图

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.4 正弦函数与余弦函数的性质

有了正弦、余弦函数的图像，我们便可以研究它们的性质. 这些性质包括定义域、值域、周期性、奇偶性、单调性和对称性. 我们还将用它们来理解 $y = A\sin(\omega x + \varphi) + B$.

知乎@余命数

1.4.1 定义域、值域与周期性

从单位圆的定义看，角 x 旋转一周（2π 弧度），终边上的点 P 回到原位，其坐标 $(\cos x, \sin x)$ 也重复出现. 这正是三角函数周期性的来源. 我们有 $\sin(x+2\pi) = \sin x$，$\cos(x + 2\pi) = \cos x$. 因此 2π 是它们的一个周期，也是最小正周期.

- **定义域**. 从单位圆定义可知，任意实数角 α 都有唯一确定的正弦值和余弦值与之对应. 因此，函数 $y = \sin x$ 和 $y = \cos x$ 的定义域均为实数集 $\mathbb{R}$.
- **值域**. 在单位圆 $x^2 + y^2 = 1$ 上，点的横、纵坐标的取值范围都是 $[-1, 1]$. 因此，函数 $y = \sin x$ 和 $y = \cos x$ 的值域均为 $[-1, 1]$.
  - 当 $x = \frac{\pi}{2} + 2k\pi, k \in \mathbb{Z}$ 时，$y = \sin x$ 取得最大值 1.
  - 当 $x = \frac{3\pi}{2} + 2k\pi, k \in \mathbb{Z}$ 时，$y = \sin x$ 取得最小值 -1.
  - 当 $x = 2k\pi, k \in \mathbb{Z}$ 时，$y = \cos x$ 取得最大值 1.
  - 当 $x = \pi + 2k\pi, k \in \mathbb{Z}$ 时，$y = \cos x$ 取得最小值 -1.
- **周期性**. 正弦函数和余弦函数都是周期函数，它们的最小正周期都是 2π. 我们有 $\sin(x + 2k\pi) = \sin x$，$\cos(x + 2k\pi) = \cos x$，其中 $k \in \mathbb{Z}$.

1.4.2 奇偶性与单调性

- **奇偶性**. 在单位圆上，角 x 与 $-x$ 的终边关于 x 轴对称. 设角 x 终边上点为 $P(\cos x, \sin x)$，则角 $-x$ 终边上点为 $P'(\cos x, -\sin x)$. 根据定义，我们得到

$$\cos(-x) = \cos x$$

$$\sin(-x) = -\sin x$$

  因此

  - $y = \cos x$ 是**偶函数**，其图像关于 y 轴对称.
  - $y = \sin x$ 是**奇函数**，其图像关于原点对称.
- **单调性**. 我们先在一个周期内观察单调性，再由周期性推广到整个定义域.
  - 对于函数 $y = \sin x$：
    - **单调递增区间**为 $\left[-\frac{\pi}{2} + 2k\pi, \frac{\pi}{2} + 2k\pi\right], k \in \mathbb{Z}$.

- **单调递减区间**为 $\left[\frac{\pi}{2}+2k\pi, \frac{3\pi}{2}+2k\pi\right], k\in\mathbb{Z}$.

- 对于函数 $y=\cos x$：
  - **单调递增区间**为 $\left[-\pi+2k\pi, 2k\pi\right], k\in\mathbb{Z}$.
  - **单调递减区间**为 $\left[2k\pi, \pi+2k\pi\right], k\in\mathbb{Z}$.

记单调区间时，我们可以在单位圆上观察点的运动. 对于 $\sin x$，我们看纵坐标 y. 从 -90° $(-\pi/2)$ 转到 90° $(\pi/2)$，点在圆的右半部分，纵坐标 y 从 -1 **增**至 1；从 90° 转到 270° $(3\pi/2)$，点在圆的左半部分，纵坐标 y 从 1 **减**至 -1. 对于 $\cos x$，我们看横坐标 x. 从 180° (π) 转到 360°（2π 或 0），点在圆的下半部分移动到右侧，横坐标 x 从 -1 **增**至 1；从 0° 转到 180°，点在圆的上半部分，横坐标 x 从 1 **减**至 -1.

1.4.3 对称性

我们讨论对称轴和对称中心.

- **函数** $y=\sin x$.
  - **对称轴**. 图像的最高点和最低点所在的竖直线，即直线 $x=k\pi+\frac{\pi}{2}, k\in\mathbb{Z}$.
  - **对称中心**. 图像与 x 轴的交点，即点 $(k\pi, 0), k\in\mathbb{Z}$.
- **函数** $y=\cos x$.
  - **对称轴**. 图像的最高点和最低点所在的竖直线，即直线 $x=k\pi, k\in\mathbb{Z}$.（这包含了作为偶函数对称轴的 y 轴 $x=0$）
  - **对称中心**. 图像与 x 轴的交点，即点 $\left(k\pi+\frac{\pi}{2}, 0\right), k\in\mathbb{Z}$.

对称轴经过函数取到**最值**的位置；对称中心是函数图像与 x 轴的交点，也是图像的拐点. 若函数满足 $f(a+x)=f(a-x)$，则其图像关于直线 $x=a$ 对称；若满足 $f(a+x)=-f(a-x)$（或 $f(x)+f(2a-x)=0$），则图像关于点 $(a,0)$ 中心对称.

1.4.4 性质汇总

我们将正弦函数与余弦函数的性质汇总成表.

知乎@余命数

表 1.1　正弦函数与余弦函数性质对比

性质	$y=\sin x$	$y=\cos x$
定义域	$\mathbb{R}$	$\mathbb{R}$
值域	$[-1,1]$	$[-1,1]$
最小正周期	2π	2π
奇偶性	奇函数	偶函数
单调递增区间	$[2k\pi-\frac{\pi}{2},2k\pi+\frac{\pi}{2}]$	$[(2k-1)\pi,2k\pi]$
单调递减区间	$[2k\pi+\frac{\pi}{2},2k\pi+\frac{3\pi}{2}]$	$[2k\pi,(2k+1)\pi]$
对称轴	直线 $x=k\pi+\frac{\pi}{2}$	直线 $x=k\pi$
对称中心	点 $(k\pi,0)$	点 $(k\pi+\frac{\pi}{2},0)$

表中所有 $k\in\mathbb{Z}$

知乎@余命数

1.4.5 高考真题选讲

例题 1.1 (2012 年福建卷文科数学第 8 题) 函数 $f(x)=\sin\left(x-\frac{\pi}{4}\right)$ 的图象的一条对称轴是

A. $x=\frac{\pi}{4}$　　C. $x=-\frac{\pi}{4}$

B. $x=\frac{\pi}{2}$　　D. $x=-\frac{\pi}{2}$

解 答案:C.

令 $x-\frac{\pi}{4}=k\pi+\frac{\pi}{2},k\in\mathbb{Z}$, 得对称轴方程 $x=k\pi+\frac{3\pi}{4}$.

当 $k=-1$ 时, $x=-\frac{\pi}{4}$, 对应选项 C.

例题 1.2 (2019 年北京卷理科数学第 9 题) 函数 $f(x)=\sin^2 2x$ 的最小正周期是_____.

解 答案: $\frac{\pi}{2}$.

知乎@余命数

用降幂公式 $\sin^2 2x = \dfrac{1-\cos 4x}{2}$, 故

$$f(x) = \frac{1}{2} - \frac{1}{2}\cos 4x,$$

周期 $T = \dfrac{2\pi}{4} = \dfrac{\pi}{2}$.

例题 1.3 (2017 年全国 □ 卷理科数学第 6 题) 设函数 $f(x) = \cos\left(x + \dfrac{\pi}{3}\right)$, 则下列结论错误的是

A. $f(x)$ 的一个周期为 -2π

B. $y = f(x)$ 的图象关于直线 $x = \dfrac{8\pi}{3}$ 对称

C. $f(x+\pi)$ 的一个零点为 $x = \dfrac{\pi}{6}$

D. $f(x)$ 在 $\left(\dfrac{\pi}{2}, \pi\right)$ 单调递减

解 答案:D.

A: cos 的周期为 $2k\pi$, $k = -1$ 时 $T = -2\pi$, 正确.

B: $f\left(\dfrac{8\pi}{3}\right) = \cos\left(\dfrac{8\pi}{3} + \dfrac{\pi}{3}\right) = \cos 3\pi = -1$, 取最小值, 故 $x = \dfrac{8\pi}{3}$ 是对称轴, 正确.

C: $f\left(\dfrac{\pi}{6} + \pi\right) = \cos\left(\dfrac{\pi}{6} + \pi + \dfrac{\pi}{3}\right) = \cos\dfrac{3\pi}{2} = 0$, 正确.

D: 当 $x \in \left(\dfrac{\pi}{2}, \pi\right)$ 时, $x + \dfrac{\pi}{3} \in \left(\dfrac{5\pi}{6}, \dfrac{4\pi}{3}\right)$, cos 在该区间内先减后增(在 $x + \dfrac{\pi}{3} = \pi$ 处取最小值), 并非单调递减, 故 D 错误.

1.5 正切函数

正弦、 余弦对应单位圆上点的坐标 (x, y). 角 α 确定了终边上一点 $P(\cos\alpha, \sin\alpha)$. 这条终边 OP 是一条射线, 除点的坐标外还有一个几何量, 即**斜率**. 根据斜率的定义, $k_{OP} = \dfrac{\Delta y}{\Delta x} = \dfrac{\sin\alpha - 0}{\cos\alpha - 0} = \dfrac{\sin\alpha}{\cos\alpha}$ (当 $\cos\alpha \neq 0$ 时). 这个比值就是角 α 的**正切**.

正切的几何意义可以直接画出来. 过单位圆右侧点 $(1, 0)$ 作 x 轴的垂线, 这条直线与圆**相切**. 延长角 α 的终边, 与这条切线交于点 T. 由相似三角形可知, 点 T 的纵坐标是 $\dfrac{\sin\alpha}{\cos\alpha}$. 所以正切值就是**单位圆切线上的一段有向线段的长度**.

知乎@余命数 知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.5.1 定义与图像

定义 1.3 (正切) 我们定义角 α 的**正切**为

$$\tan\alpha = \frac{\sin\alpha}{\cos\alpha} \tag{1-3}$$

它的定义域是所有使得分母 $\cos\alpha \neq 0$ 的角 α 的集合,也就是 $\alpha \neq k\pi + \frac{\pi}{2}, k \in \mathbb{Z}$.

知乎@余命数

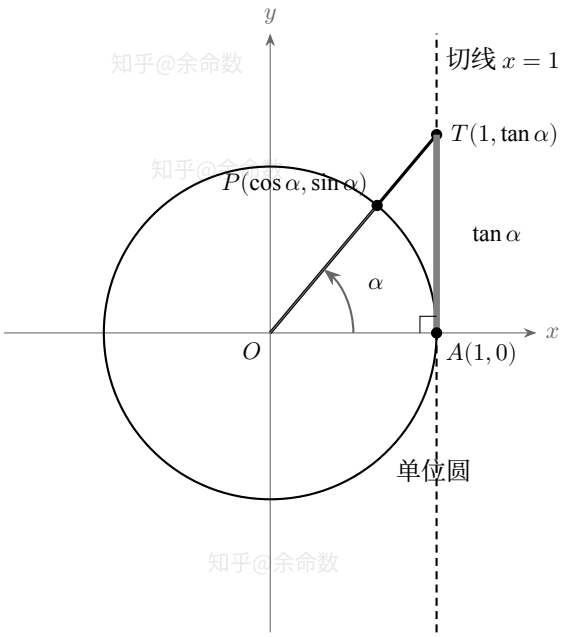


图 1.9 正切函数的几何意义

正切函数的图像与正弦、余弦曲线有明显的不同. 由于定义域中排除了使 $\cos x = 0$ 的点, 函数图像在这些点上是**不连续**的. 这些不连续点 $x = k\pi + \frac{\pi}{2}$ 形成了函数图像的一条条**垂直渐近线**.

在一个周期 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 内,取几个关键点即可描绘其形态. $(-\frac{\pi}{4}, -1)$、$(0,0)$、$(\frac{\pi}{4}, 1)$. 图像从左侧贴近渐近线 $x = -\frac{\pi}{2}$ 的负无穷处开始,穿过这三点,最终向上贴近渐近线 $x = \frac{\pi}{2}$ 的正无穷处.

1.5.2 正切函数的性质

观察正切函数的定义与图像,我们可总结出它的主要性质.

- **定义域**.$\{x \in \mathbb{R} \mid x \neq k\pi + \frac{\pi}{2}, k \in \mathbb{Z}\}$.
- **值域**. 实数集 $\mathbb{R}$. 正切函数可以取到任意实数值.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

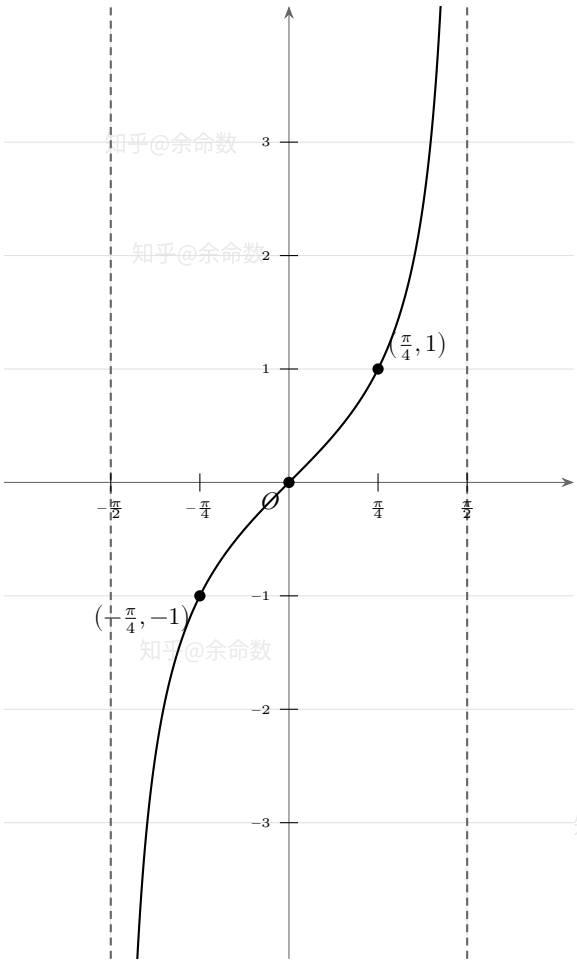


图 1.10 函数 $y = \tan x$ 的图像

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **周期性**. 观察图像, 其形态每隔 π 的长度就重复一次. 可以验证

$$\tan(x+\pi)=\frac{\sin(x+\pi)}{\cos(x+\pi)}=\frac{-\sin x}{-\cos x}=\tan x$$

所以正切函数是周期函数, 其**最小正周期为** π.

- **奇偶性**. 根据定义

$$\tan(-x)=\frac{\sin(-x)}{\cos(-x)}=\frac{-\sin x}{\cos x}=-\tan x$$

所以正切函数是**奇函数**, 其图像关于原点 $(0,0)$ 对称. 更一般地, 图像关于所有零点 $(k\pi,0),k\in\mathbb{Z}$ 中心对称.

- **单调性**. 正切函数在它的每一个连续区间上都是**单调递增**的. 其单调递增区间为

$$\left(k\pi-\frac{\pi}{2},k\pi+\frac{\pi}{2}\right),k\in\mathbb{Z}$$

常见的误解是以为 $y=\tan x$ 在整个定义域上单调递增. 其实它在每个连续区间 $(-\frac{\pi}{2},\frac{\pi}{2})$、$(\frac{\pi}{2},\frac{3\pi}{2})$ 等内部都递增, 但不能跨区间说整体递增. 例如取 $x_1=\frac{\pi}{4}$ 和 $x_2=\frac{3\pi}{4}$, 有 $x_1<x_2$, 而 $\tan x_1=1,\tan x_2=-1$, 所以 $\tan x_1>\tan x_2$, 这与整体单调递增矛盾.

1.5.3 高考真题选讲

例题 1.4 (2020 年全国 □ 卷理科数学第 9 题) 已知 $2\tan\theta-\tan\left(\theta+\frac{\pi}{4}\right)=7$, 则 $\tan\theta=$

A. -2　　C. 1

B. -1　　D. 2

解 答案: D.

用正切和角公式 $\tan\left(\theta+\frac{\pi}{4}\right)=\frac{\tan\theta+1}{1-\tan\theta}$ 代入, 得

$$2\tan\theta-\frac{\tan\theta+1}{1-\tan\theta}=7.$$

两边乘以 $1-\tan\theta$ 并整理, 得

$$2\tan\theta-2\tan^2\theta-\tan\theta-1=7-7\tan\theta,$$

即 $2\tan^2\theta-8\tan\theta+8=0$, 亦即 $(\tan\theta-2)^2=0$, 故 $\tan\theta=2$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 1.5 (2020 年浙江卷第 13 题) 已知 $\tan\theta = 2$, 则 $\cos 2\theta =$_____, $\tan\left(\theta - \frac{\pi}{4}\right) =$_____.

解 答案: $-\frac{3}{5}, \frac{1}{3}$.

用万能公式求 $\cos 2\theta$:

$$\cos 2\theta = \frac{1-\tan^2\theta}{1+\tan^2\theta} = \frac{1-4}{1+4} = -\frac{3}{5}.$$

用正切差角公式求 $\tan\left(\theta - \frac{\pi}{4}\right)$:

$$\tan\left(\theta - \frac{\pi}{4}\right) = \frac{\tan\theta - 1}{1+\tan\theta} = \frac{2-1}{1+2} = \frac{1}{3}.$$

知乎@余命数

1.6 余切、正割与余割函数

我们已经在单位圆上用点 (x, y) 的坐标定义了 $\sin\alpha, \cos\alpha$, 用终边斜率定义了 $\tan\alpha$. 现在我们把这三个函数分别取倒数, 得到另外三个函数.

- **余切 (Cotangent)** $\cot\alpha = \dfrac{1}{\tan\alpha} = \dfrac{x}{y}$.
- **正割 (Secant)** $\sec\alpha = \dfrac{1}{\cos\alpha} = \dfrac{1}{x}$.
- **余割 (Cosecant)** $\csc\alpha = \dfrac{1}{\sin\alpha} = \dfrac{1}{y}$.

这三个倒数函数在历史上曾被用来简化除法运算; 在现代数学中, 它们在积分、微分方程和某些物理问题里仍然出现.

1.6.1 定义与几何意义

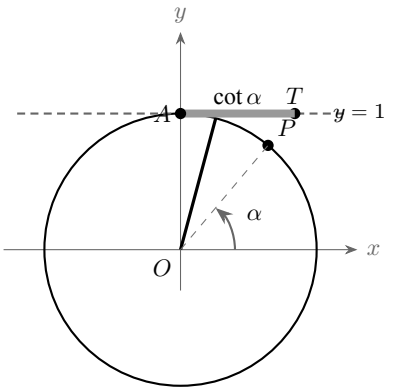
定义 1.4 (余切、正割、余割) 对于任意角 α, 我们定义.

- **余切** $\cot\alpha = \dfrac{\cos\alpha}{\sin\alpha}$, 定义域为 $\{\alpha \in \mathbb{R} \mid \sin\alpha \neq 0\}$.
- **正割** $\sec\alpha = \dfrac{1}{\cos\alpha}$, 定义域为 $\{\alpha \in \mathbb{R} \mid \cos\alpha \neq 0\}$.
- **余割** $\csc\alpha = \dfrac{1}{\sin\alpha}$, 定义域为 $\{\alpha \in \mathbb{R} \mid \sin\alpha \neq 0\}$.

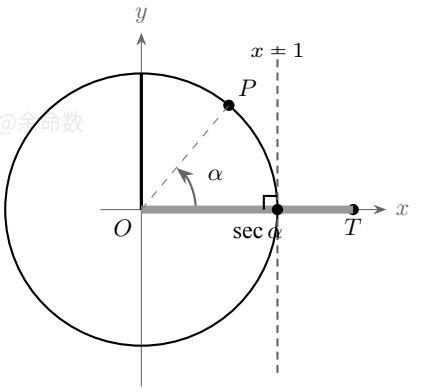
这三个函数同样可以在单位圆上用切线段的长度来解释.

- 余切 $\cot\alpha$ 的几何意义如下. 作单位圆过点 $(0, 1)$ 的切线 $y = 1$. 角 α 的终边或其反向延长线与该切线交于点 T_c. 点 T_c 的横坐标即为 $\cot\alpha$.

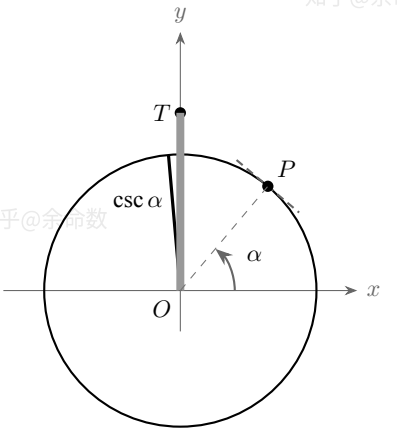
知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵



(a) 余切.$\cot\alpha = \overline{AT}$



(b) 正割.$\sec\alpha = \overline{OT}$



(c) 余割.$\csc\alpha = \overline{OT}$

图 1.11　余切、正割、余割的几何意义

知乎@余命数

- 正割 $\sec\alpha$ 与余割 $\csc\alpha$ 的几何意义如下. 作单位圆过终边上点 $P(\cos\alpha,\sin\alpha)$ 的切线. 该切线与 x 轴的交点为 T_s, 与 y 轴的交点为 T_{cs}. 原点到这两个交点的有向距离 OT_s 和 OT_{cs} 分别为 $\sec\alpha$ 和 $\csc\alpha$.

1.6.2 函数图像与性质

这三个函数的图像与性质由倒数定义直接决定. 凡是使分母为零的点, 都是图像的垂直渐近线.

1.6.2.1 余切函数 $y=\cot x$

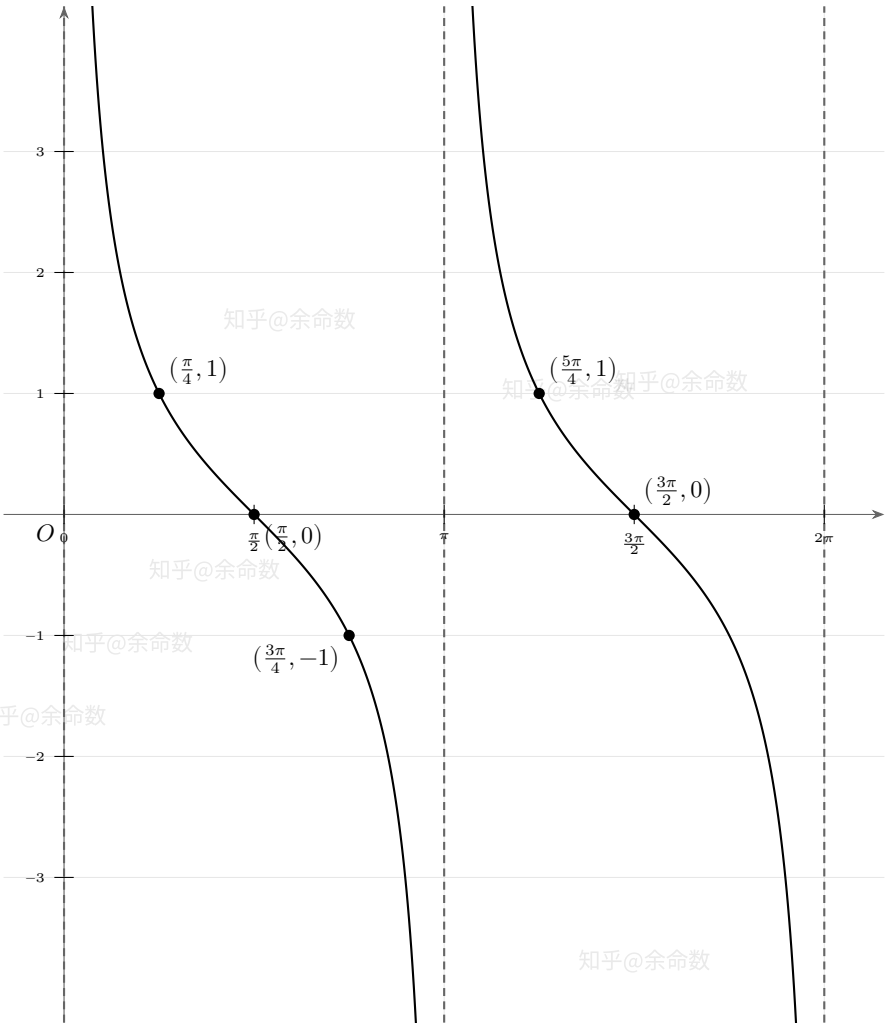


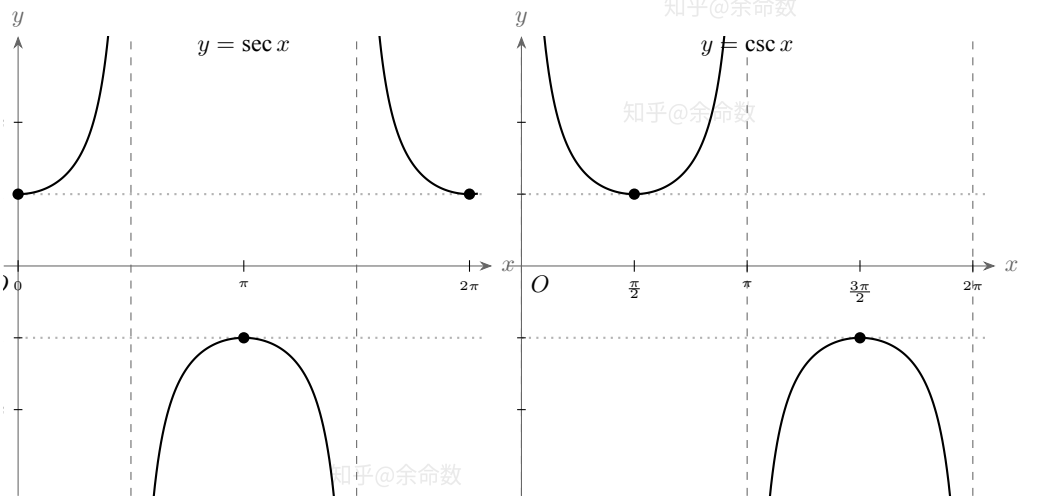
图 1.12 函数 $y=\cot x$ 的图像

- **定义域** $\{x\in\mathbb{R}\mid x\neq k\pi,k\in\mathbb{Z}\}$.
- **值域** $\mathbb{R}$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **周期性** 最小正周期为 π.
- **奇偶性** $\cot(-x) = -\cot x$, 是奇函数, 图像关于原点对称.
- **单调性** 在每个定义区间 $(k\pi, (k+1)\pi)$ 内单调递减.

1.6.2.2 正割函数 $y = \sec x$ 与余割函数 $y = \csc x$



(a) 正割函数　　(b) 余割函数

图 1.13　正割函数和余割函数的图像

- **定义域** $\sec x$ 为 $\{x \mid x \neq k\pi + \frac{\pi}{2}\}$; $\csc x$ 为 $\{x \mid x \neq k\pi\}$.
- **值域**均为 $(-\infty, -1] \cup [1, \infty)$.
- **周期性**最小正周期均为 2π.
- **奇偶性** $y = \sec x$ 是偶函数; $y = \csc x$ 是奇函数.

1.6.3 性质汇总

我们将六个三角函数的性质汇总成表.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

表 1.2 六种三角函数重要性质对比

函数	定义域	值域	最小正周期	奇偶性
$y=\sin x$	$\mathbb{R}$	$[-1,1]$	2π	奇函数
$y=\cos x$	$\mathbb{R}$	$[-1,1]$	2π	偶函数
$y=\tan x$	$\{x \mid x \neq k\pi+\frac{\pi}{2}\}$	$\mathbb{R}$	π	奇函数
$y=\cot x$	$\{x \mid x \neq k\pi\}$	$\mathbb{R}$	π	奇函数
$y=\sec x$	$\{x \mid x \neq k\pi+\frac{\pi}{2}\}$	$(-\infty,-1]\cup[1,\infty)$	2π	偶函数
$y=\csc x$	$\{x \mid x \neq k\pi\}$	$(-\infty,-1]\cup[1,\infty)$	2π	奇函数

注. 表中所有 $k\in\mathbb{Z}$

1.7 同角三角函数的基本关系

我们已经用单位圆上点的坐标定义了 $\sin\alpha,\cos\alpha$, 用终边斜率定义了 $\tan\alpha$. 它们来自同一个角, 因此有固定的联系. 三角计算与恒等变形从这些关系出发.

1.7.1 两个基本关系式

这些关系可以直接从定义得到.

- **第一条, 单位圆方程**.
  点 $P(x,y)$ 在单位圆 $x^2+y^2=1$ 上, 而 $x=\cos\alpha, y=\sin\alpha$. 将坐标代入圆的方程, 得到对任意角 α 都成立的平方关系.
- **第二条, 正切的定义**.
  我们定义了 $\tan\alpha=\dfrac{\sin\alpha}{\cos\alpha}$. 这本身就是一个关系式, 我们称之为商数关系.

定理 1.2 (同角三角函数基本关系) 对于任意角 α, 以下关系式恒成立.

1. **平方关系**

$$\sin^2\alpha+\cos^2\alpha=1 \tag{1-4}$$

2. **商数关系**

$$\tan\alpha=\frac{\sin\alpha}{\cos\alpha}\quad\left(\alpha\neq k\pi+\frac{\pi}{2},k\in\mathbb{Z}\right) \tag{1-5}$$

证明 我们用定义直接证明.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

设角 α 的终边上有一点 $P(x,y)$(非原点),它到原点 O 的距离为 $r=\sqrt{x^2+y^2}$. 则

$$\sin\alpha=\frac{y}{r},\quad \cos\alpha=\frac{x}{r},\quad \tan\alpha=\frac{y}{x}\quad (x\neq 0)$$

关于平方关系,直接将定义代入等式左侧计算,得

$$\sin^2\alpha+\cos^2\alpha=\left(\frac{y}{r}\right)^2+\left(\frac{x}{r}\right)^2=\frac{y^2}{r^2}+\frac{x^2}{r^2}=\frac{x^2+y^2}{r^2}$$

因为 $x^2+y^2=r^2$,所以上式可以继续化简,得

$$\frac{x^2+y^2}{r^2}=\frac{r^2}{r^2}=1$$

这就证明了 $\sin^2\alpha+\cos^2\alpha=1$ 恒成立.

关于商数关系,同样从定义出发. 在定义域要求 $\cos\alpha\neq 0$,即 $x\neq 0$ 的前提下,

$$\frac{\sin\alpha}{\cos\alpha}=\frac{y/r}{x/r}=\frac{y}{x}$$

而 $\frac{y}{x}$ 是 $\tan\alpha$ 的定义. 故商数关系 $\tan\alpha=\frac{\sin\alpha}{\cos\alpha}$ 成立. 该关系式的成立条件 $\cos\alpha\neq 0$,意味着角 α 的终边不能落在 y 轴上,即 $\alpha\neq k\pi+\frac{\pi}{2},k\in\mathbb{Z}$. □

平方关系和商数关系是所有三角恒等变形的代数起点. 更复杂的三角恒等式都可以通过对它们作代数运算得到.

1.7.2 公式的变形与应用

这两个公式形式简单,但变形很多.

- **知一求二**. 由平方关系开方,得

$$\sin\alpha=\pm\sqrt{1-\cos^2\alpha},\qquad \cos\alpha=\pm\sqrt{1-\sin^2\alpha}.$$

  正负号由角 α 所在象限唯一确定. 开方符号是这一步最容易出错的地方.

- **和差与积的转化**. 我们可以用下面的恒等式把 $\sin\alpha\pm\cos\alpha$ 与 $\sin\alpha\cos\alpha$ 互相转化.

$$(\sin\alpha\pm\cos\alpha)^2=\sin^2\alpha+\cos^2\alpha\pm 2\sin\alpha\cos\alpha=1\pm 2\sin\alpha\cos\alpha$$

  知道和或差,就能求出积;知道积,也能求出和或差的平方.

- **切化弦与弦化切**. 商数关系把正切与正、余弦联系在一起.
  - 遇到正切,可以统一化为正余弦计算,即切化弦.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

- 遇到关于 $\sin\alpha$ 和 $\cos\alpha$ 的齐次式(分子分母各项的幂次相同), 可以分子分母同除以 $\cos^k\alpha$, 其中 k 为幂次, 从而转化为关于 $\tan\alpha$ 的式子, 即弦化切.

例题 1.6 (知一求二) 已知角 α 的终边在第三象限, 且 $\cos\alpha=-\frac{4}{5}$, 求 $\sin\alpha$ 和 $\tan\alpha$ 的值.

知乎@余命数

解 我们先用平方关系求出 $\sin\alpha$ 的绝对值, 再由象限确定符号, 最后用商数关系求 $\tan\alpha$.

知乎@余命数

将 $\cos\alpha=-\frac{4}{5}$ 代入平方关系, 得

$$\sin^2\alpha=1-\cos^2\alpha=1-\left(-\frac{4}{5}\right)^2=1-\frac{16}{25}=\frac{9}{25},$$

故 $\sin\alpha=\pm\frac{3}{5}$. 角 α 的终边在第三象限, 其正弦值为负, 因而 $\sin\alpha=-\frac{3}{5}$. 再由商数关系得

知乎@余命数

$$\tan\alpha=\frac{\sin\alpha}{\cos\alpha}=\frac{-3/5}{-4/5}=\frac{3}{4}.$$

开方符号由象限决定. 忽略第三象限这个条件, $\sin\alpha$ 的符号就无法确定, 答案也会出错. 任何涉及开方的三角运算, 都应先检查象限.

例题 1.7 (和差积关系的运用) 已知角 $\alpha\in(\frac{\pi}{2},\pi)$, 且 $\sin\alpha+\cos\alpha=\frac{1}{5}$, 求下列各式的值.

(1) $\sin\alpha\cos\alpha$ (2) $\sin\alpha-\cos\alpha$ (3) $\tan\alpha$

知乎@余命数
知乎@余命数

解 我们可以对已知等式两边平方, 利用完全平方公式 $(a+b)^2=a^2+b^2+2ab$ 把和转化为积.

(1) 将已知等式 $\sin\alpha+\cos\alpha=\frac{1}{5}$ 两边平方, 并以平方关系 $\sin^2\alpha+\cos^2\alpha=1$ 代入, 得

$$\begin{aligned}\frac{1}{25}&=(\sin\alpha+\cos\alpha)^2=\sin^2\alpha+\cos^2\alpha+2\sin\alpha\cos\alpha\\&=1+2\sin\alpha\cos\alpha,\end{aligned}$$

移项解出

$$\sin\alpha\cos\alpha=-\frac{12}{25}.$$

知乎@余命数

(2) 直接求 $\sin\alpha-\cos\alpha$ 不便, 我们改求它的平方. 由上一步结果,

$$\begin{aligned}(\sin\alpha-\cos\alpha)^2&=\sin^2\alpha+\cos^2\alpha-2\sin\alpha\cos\alpha\\&=1-2\left(-\frac{12}{25}\right)=1+\frac{24}{25}=\frac{49}{25},\end{aligned}$$

开方得 $\sin\alpha-\cos\alpha=\pm\frac{7}{5}$. 由 $\alpha\in(\frac{\pi}{2},\pi)$ 知 $\sin\alpha>0$ 且 $\cos\alpha<0$, 故 $\sin\alpha-\cos\alpha$ 为正, 即 $\sin\alpha-\cos\alpha=\frac{7}{5}$.

(3) 联立 $\sin\alpha+\cos\alpha=\frac{1}{5}$ 与 $\sin\alpha-\cos\alpha=\frac{7}{5}$, 两式相加得 $\sin\alpha=\frac{4}{5}$, 相减得 $\cos\alpha=-\frac{3}{5}$, 这与 α 在第二象限的符号一致. 由商数关系,

$$\tan\alpha=\frac{\sin\alpha}{\cos\alpha}=\frac{4/5}{-3/5}=-\frac{4}{3}.$$

平方把和差变成积;开方从积得到和差;象限条件决定开方符号.

例题 1.8 (弦化切齐次式处理) 已知 $\tan\alpha=-3$, 求 $\dfrac{2\sin\alpha-3\cos\alpha}{4\sin\alpha+5\cos\alpha}$.

解 所求表达式的分子、分母都是关于 $\sin\alpha$ 和 $\cos\alpha$ 的一次齐次式(每一项的幂次均为 1), 因此可以分子分母同除以 $\cos\alpha$, 把整个式子转化为关于 $\tan\alpha$ 的表达式.

当 $\tan\alpha=-3$ 有意义时, 必有 $\cos\alpha\neq0$, 因此可在分子分母同除以 $\cos\alpha$.

$$\begin{aligned}\text{原式}&=\frac{\frac{2\sin\alpha-3\cos\alpha}{\cos\alpha}}{\frac{4\sin\alpha+5\cos\alpha}{\cos\alpha}}\\&=\frac{2\frac{\sin\alpha}{\cos\alpha}-3\frac{\cos\alpha}{\cos\alpha}}{4\frac{\sin\alpha}{\cos\alpha}+5\frac{\cos\alpha}{\cos\alpha}}\\&=\frac{2\tan\alpha-3}{4\tan\alpha+5}\end{aligned}$$

将 $\tan\alpha=-3$ 代入上式, 得

$$\text{原式}=\frac{2(-3)-3}{4(-3)+5}=\frac{-6-3}{-12+5}=\frac{-9}{-7}=\frac{9}{7}.$$

弦化切用于关于 $\sin\alpha,\cos\alpha$ 的齐次分式. 通过分子分母同除以 $\cos\alpha$ 的适当幂次, 把两个变量化为一个变量 $\tan\alpha$. 二次齐次式则同除以 $\cos^2\alpha$.

例题 1.9 (切化弦由正切求正余弦) 已知 $\tan\alpha=2$, 求 $\dfrac{1}{\sin^2\alpha-\sin\alpha\cos\alpha}$ 的值.

解 我们用两种方法求解.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

法一数形结合. 由 $\tan\alpha=2>0$ 知角 α 在第一或第三象限. 构造一个直角三角形, 使其一锐角 α' 满足 $\tan\alpha'=2=\frac{2}{1}$, 即对边为 2、邻边为 1, 由勾股定理斜边为 $\sqrt{2^2+1^2}=\sqrt{5}$. 于是 $|\sin\alpha|=\frac{2}{\sqrt{5}}$, $|\cos\alpha|=\frac{1}{\sqrt{5}}$, 再按象限分类.

- 若 α 在第一象限, 则 $\sin\alpha=\frac{2}{\sqrt{5}}$, $\cos\alpha=\frac{1}{\sqrt{5}}$,

$$\text{所求表达式}=\frac{1}{(\frac{2}{\sqrt{5}})^2-(\frac{2}{\sqrt{5}})(\frac{1}{\sqrt{5}})}=\frac{1}{\frac{4}{5}-\frac{2}{5}}=\frac{1}{\frac{2}{5}}=\frac{5}{2}.$$

- 若 α 在第三象限, 则 $\sin\alpha=-\frac{2}{\sqrt{5}}$, $\cos\alpha=-\frac{1}{\sqrt{5}}$,

$$\text{所求表达式}=\frac{1}{(-\frac{2}{\sqrt{5}})^2-(-\frac{2}{\sqrt{5}})(-\frac{1}{\sqrt{5}})}=\frac{1}{\frac{4}{5}-\frac{2}{5}}=\frac{1}{\frac{2}{5}}=\frac{5}{2}.$$

两种情形下结果相同, 故所求表达式的值均为 $\frac{5}{2}$.

法二代数变形. 所求表达式并非齐次式, 我们先把它改造成齐次式. 分子 1 可以写成 $\sin^2\alpha+\cos^2\alpha$; 代入后分子分母皆为二次齐次式, 再同除以 $\cos^2\alpha$, 得

$$\begin{aligned}\text{原式}&=\frac{\sin^2\alpha+\cos^2\alpha}{\sin^2\alpha-\sin\alpha\cos\alpha}\quad\text{(分子以平方关系代换)}\\&=\frac{\tan^2\alpha+1}{\tan^2\alpha-\tan\alpha}\qquad\text{(同除以 }\cos^2\alpha\text{)}.\end{aligned}$$

代入 $\tan\alpha=2$, 得

$$\text{原式}=\frac{2^2+1}{2^2-2}=\frac{5}{2}.$$

构造三角形法先求出绝对值, 最后需要由象限补符号. 代数变形法把 1 写成 $\sin^2\alpha+\cos^2\alpha$, 把非齐次式化为齐次式, 适用范围更广.

例题 1.10 (2023 年 6 月全国甲卷) 设甲为 $\sin^2\alpha+\sin^2\beta=1$, 乙为 $\sin\alpha+\cos\beta=0$, 则(　)

A. 甲是乙的充分条件但不是必要条件

B. 甲是乙的必要条件但不是充分条件

C. 甲是乙的充要条件

D. 甲既不是乙的充分条件也不是乙的必要条件

解 判断甲、乙的充分必要关系, 需分别检验"乙 $\Longrightarrow$ 甲"与"甲 $\Longrightarrow$ 乙"两个方向.

先证"乙 $\Longrightarrow$ 甲". 由条件乙 $\sin\alpha+\cos\beta=0$ 得 $\sin\alpha=-\cos\beta$, 两边平方并以平方关系 $\cos^2\beta=1-\sin^2\beta$ 代入,

$$\sin^2\alpha=\cos^2\beta=1-\sin^2\beta,$$

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

移项即得 $\sin^2\alpha+\sin^2\beta=1$, 此即条件甲. 故“乙 $\Longrightarrow$ 甲”成立, 甲是乙的必要条件.

再考察“甲 $\Longrightarrow$ 乙”. 由条件甲 $\sin^2\alpha+\sin^2\beta=1$ 及平方关系得 $\sin^2\alpha=1-\sin^2\beta=\cos^2\beta$, 开方有 $\sin\alpha=\pm\cos\beta$, 此式含 $\sin\alpha=\cos\beta$ 与 $\sin\alpha=-\cos\beta$ 两种可能, 未必推出条件乙, 取 $\alpha=\beta=\frac{\pi}{4}$ 即可说明. 此时

$$\sin^2\frac{\pi}{4}+\sin^2\frac{\pi}{4}=\frac{1}{2}+\frac{1}{2}=1$$

满足甲, 而 $\sin\frac{\pi}{4}+\cos\frac{\pi}{4}=\sqrt{2}\neq 0$ 不满足乙. 故“甲 $\Longrightarrow$ 乙”不成立, 甲不是乙的充分条件.

由于“乙 $\Longrightarrow$ 甲”成立而“甲 $\Longrightarrow$ 乙”不成立, 甲是乙的必要条件, 但不是充分条件, 答案为 B.

1.7.3 高考真题选讲

知乎@余命数

例题 1.11 (2020 年全国 □ 卷理科数学第 9 题) 已知 $\alpha\in(0,\pi)$, 且 $3\cos 2\alpha-8\cos\alpha=5$, 则 $\sin\alpha=$

A. $\frac{\sqrt{5}}{3}$　　C. $\frac{1}{3}$

B. $\frac{2}{3}$　　D. $\frac{\sqrt{5}}{9}$

解 答案: A.

用二倍角公式 $\cos 2\alpha=2\cos^2\alpha-1$ 代入, 得

$$3(2\cos^2\alpha-1)-8\cos\alpha=5,$$

整理得 $3\cos^2\alpha-4\cos\alpha-4=0$, 解得 $\cos\alpha=2$(舍去)或 $\cos\alpha=-\frac{2}{3}$.

由 $\alpha\in(0,\pi)$ 且 $\cos\alpha<0$ 知 $\alpha\in\left(\frac{\pi}{2},\pi\right)$, 故

$$\sin\alpha=\sqrt{1-\cos^2\alpha}=\sqrt{1-\frac{4}{9}}=\frac{\sqrt{5}}{3}.$$

例题 1.12 (2019 年全国 □ 卷理科数学第 10 题) 已知 $\alpha\in\left(0,\frac{\pi}{2}\right)$, $2\sin 2\alpha=\cos 2\alpha+1$, 则 $\sin\alpha=$

A. $\frac{1}{5}$　　C. $\frac{\sqrt{3}}{3}$

B. $\frac{\sqrt{5}}{5}$　　D. $\frac{2\sqrt{5}}{5}$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

解 答案:B.

用二倍角公式展开:$4\sin\alpha\cos\alpha = 2\cos^2\alpha$. 由 $\alpha \in \left(0, \dfrac{\pi}{2}\right)$ 知 $\cos\alpha > 0$, 两边除以 $2\cos\alpha$ 得 $\cos\alpha = 2\sin\alpha$.

代入 $\sin^2\alpha + \cos^2\alpha = 1$ 得 $5\sin^2\alpha = 1$, 故 $\sin\alpha = \dfrac{\sqrt{5}}{5}$.

1.8 诱导公式

知乎@余命数

我们已经会求 $\frac{\pi}{6}$、$\frac{\pi}{4}$、$\frac{\pi}{3}$ 这些锐角的三角函数值. 但实际旋转的角度是任意的, 如 $\sin\frac{2\pi}{3}$ 或 $\cos(-\frac{\pi}{6})$ 这类值就不能直接读出.

$120°$、$210°$、$-30°$ 等角的终边与 $30°$ 的终边在单位圆上具有对称性, 或关于 y 轴对称, 或关于原点对称, 或关于 x 轴对称.

诱导公式把这些对称性写成代数等式, 从而把任意角的三角函数计算化为锐角的三角函数计算.

我们以 α 为基准, 在单位圆上考察 $2k\pi+\alpha$、$\pi+\alpha$、$-\alpha$、$\pi-\alpha$ 等角的对称点, 建立它们与 α 的三角函数关系. 为看清象限, 先设 α 为锐角;所得的公式对任意角 α 都成立.

设角 α 的终边与单位圆交于点 $P(\cos\alpha, \sin\alpha)$.

知乎@余命数

1.8.0.1 关于原点与 x 轴的对称

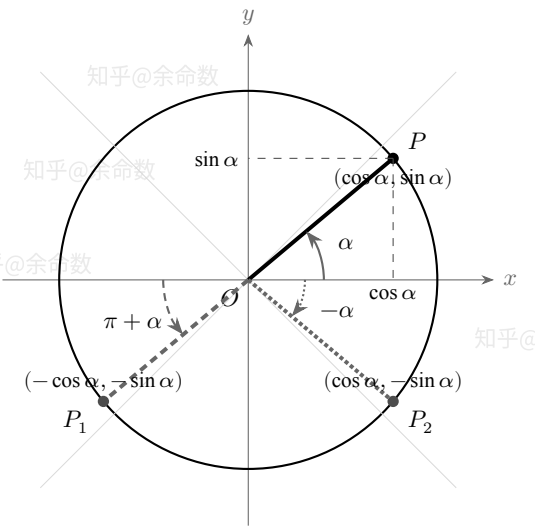


图 1.14 诱导公式的几何意义. 角 α、$\pi+\alpha$ 和 $-\alpha$ 的终边关系

- **周期性(公式一)** 角 $2k\pi+\alpha$ 与角 α 的终边完全相同, 所以它们的三角函数值也完全相同. 这是周期性的体现.

$$\sin(2k\pi+\alpha) = \sin\alpha \quad \cos(2k\pi+\alpha) = \cos\alpha \quad \tan(2k\pi+\alpha) = \tan\alpha$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

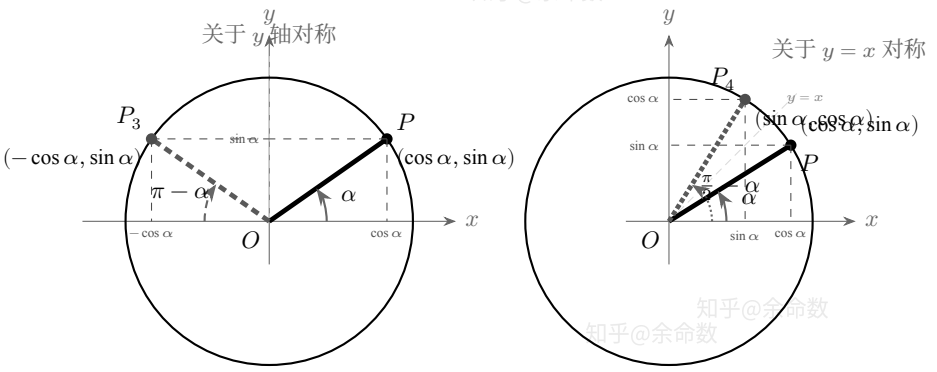
- **关于原点对称(公式二)**角 $\pi+\alpha$ 的终边是角 α 终边的反向延长线, 点 P_1 与点 P 关于原点对称. 因此 P_1 的坐标为 $(-\cos\alpha,-\sin\alpha)$.

$$\sin(\pi+\alpha)=-\sin\alpha\quad\cos(\pi+\alpha)=-\cos\alpha\quad\tan(\pi+\alpha)=\tan\alpha$$

- **关于 x 轴对称(公式三)**角 $-\alpha$ 的终边与角 α 的终边关于 x 轴对称. 点 P_2 与点 P 的横坐标相同, 纵坐标互为相反数. 因此 P_2 的坐标为 $(\cos\alpha,-\sin\alpha)$.

$$\sin(-\alpha)=-\sin\alpha\quad\cos(-\alpha)=\cos\alpha\quad\tan(-\alpha)=-\tan\alpha$$

1.8.0.2 关于 y 轴与直线 $y=x$ 的对称



(a) 角 α 和 $\pi-\alpha$　　**(b)** 角 α 和 $\frac{\pi}{2}-\alpha$

图 1.15　互余角和互补角的终边关系

- **关于 y 轴对称(公式四)**角 $\pi-\alpha$ 的终边与角 α 的终边关于 y 轴对称. 点 P_3 与点 P 的纵坐标相同, 横坐标互为相反数. 因此 P_3 的坐标为 $(-\cos\alpha,\sin\alpha)$.

$$\sin(\pi-\alpha)=\sin\alpha\quad\cos(\pi-\alpha)=-\cos\alpha\quad\tan(\pi-\alpha)=-\tan\alpha$$

- **关于直线 $y=x$ 对称(公式五)**角 $\frac{\pi}{2}-\alpha$ 的终边与角 α 的终边关于直线 $y=x$ 对称. 点 P_4 的坐标是点 P 坐标的横纵互换. 因此 P_4 的坐标为 $(\sin\alpha,\cos\alpha)$.

$$\sin(\frac{\pi}{2}-\alpha)=\cos\alpha\quad\cos(\frac{\pi}{2}-\alpha)=\sin\alpha\quad\tan(\frac{\pi}{2}-\alpha)=\cot\alpha$$

1.8.0.3 名称变换的诱导公式

上面的公式中, 函数名称没有改变. 而涉及 $\frac{\pi}{2}$ 和 $\frac{3\pi}{2}$ 的变换, 则会引起函数名称的改变, 正弦变余弦, 余弦变正弦.

- **角 $\frac{\pi}{2}+\alpha$(公式六)**可由已有公式推导.

$$\sin(\frac{\pi}{2}+\alpha)=\sin\Big(\pi-(\frac{\pi}{2}-\alpha)\Big)=\sin(\frac{\pi}{2}-\alpha)=\cos\alpha$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$\cos(\frac{\pi}{2}+\alpha)=\cos\left(\pi-(\frac{\pi}{2}-\alpha)\right)=-\cos(\frac{\pi}{2}-\alpha)=-\sin\alpha$$

几何上, 角 $\frac{\pi}{2}+\alpha$ 的终边是将角 α 的终边逆时针旋转 90° 得到的, 点 (x,y) 旋转 90° 变为 $(-y,x)$, 同样可得.

$$\sin(\frac{\pi}{2}+\alpha)=\cos\alpha\quad\cos(\frac{\pi}{2}+\alpha)=-\sin\alpha\quad\tan(\frac{\pi}{2}+\alpha)=-\cot\alpha$$

1.8.1 奇变偶不变, 符号看象限

这些公式的内在规律可以概括为一句口诀, **奇变偶不变, 符号看象限**, 适用于所有 $k\cdot\frac{\pi}{2}\pm\alpha\,(k\in\mathbb{Z})$ 形式的诱导公式.

- **奇变偶不变**. 指角 $k\cdot\frac{\pi}{2}\pm\alpha$ 中 k 的奇偶性.
  - 当 k 为偶数时, 如 $\pi\pm\alpha=2\cdot\frac{\pi}{2}\pm\alpha$、$2\pi\pm\alpha=4\cdot\frac{\pi}{2}\pm\alpha$, 函数名称不变 ($\sin\to\sin,\cos\to\cos$).
  - 当 k 为奇数时, 如 $\frac{\pi}{2}\pm\alpha$、$\frac{3\pi}{2}\pm\alpha$, 函数名称改变 ($\sin\leftrightarrow\cos$).
- **符号看象限**. 判断结果正负时, 先设 α 为锐角, 确定原角 $k\cdot\frac{\pi}{2}\pm\alpha$ 所在象限, 再取原函数在该象限的符号.

符号看象限时, 我们看的是原函数在原角所在象限的符号. 例如, 判断 cos 在第四象限的符号, 结果为正. 设 α 为锐角只是为了方便判断象限; 这条口诀对任意角 α 都成立, 依据是三角函数的周期性与对称性.

1.8.2 典例解析

例题 1.13 计算 $\sin(-\frac{13\pi}{6})+\cos(\frac{19\pi}{3})-\tan(-\frac{29\pi}{4})$ 的值.

解 我们先把负角化成正角, 再把角化到 $[0,2\pi)$ 内, 最后用诱导公式化为锐角. 先用公式三去掉负号,

$$\begin{aligned}\text{原式}&=\sin(-\frac{13\pi}{6})+\cos(\frac{19\pi}{3})-\tan(-\frac{29\pi}{4})\\&=-\sin(\frac{13\pi}{6})+\cos(\frac{19\pi}{3})+\tan(\frac{29\pi}{4})\quad\text{(奇偶性)}.\end{aligned}$$

再由周期性把每一项化为 $[0,2\pi)$ 内的熟悉角,

$$\begin{aligned}&\sin(\frac{13\pi}{6})=\sin(2\pi+\frac{\pi}{6})=\sin(\frac{\pi}{6}),\\&\cos(\frac{19\pi}{3})=\cos(6\pi+\frac{\pi}{3})=\cos(\frac{\pi}{3}),\\&\tan(\frac{29\pi}{4})=\tan(7\pi+\frac{\pi}{4})=\tan(\pi+\frac{\pi}{4})=\tan(\frac{\pi}{4})\quad\text{(周期 }\pi\text{)}.\end{aligned}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

代入特殊角值 $\sin(\frac{\pi}{6})=\frac{1}{2}$、$\cos(\frac{\pi}{3})=\frac{1}{2}$、$\tan(\frac{\pi}{4})=1$，得

$$\text{原式}=-\left(\frac{1}{2}\right)+\frac{1}{2}+1=1.$$

奇偶性、周期性与诱导公式的组合，把任意角的三角函数化为锐角的三角函数，再落到一组熟记的特殊值上.

知乎@余命数

注 1.2 (方法总结)
任意角的三角函数求值可以分三步.

- **化正角**对负角用 $\sin(-\alpha)=-\sin\alpha$、$\cos(-\alpha)=\cos\alpha$、$\tan(-\alpha)=-\tan\alpha$.
- **化小角**利用周期性，把角加上或减去 2π（正切是 π）的整数倍，化到 $[0,2\pi)$ 之类的熟悉区间.
- **化锐角**再用诱导公式把大于 $\frac{\pi}{2}$ 的角化为锐角.

例题 1.14 化简分式 $\dfrac{\sin(\pi+\alpha)\cos(2\pi-\alpha)\tan(3\pi-\alpha)}{\cos(\frac{\pi}{2}+\alpha)\sin(-\alpha-\pi)}$.

知乎@余命数

解 我们用口诀把表达式中的每一项化为只含 α 的三角函数. 遇到 $\sin(-\alpha-\pi)$ 这类带多重符号的角，先处理括号内的负号.

知乎@余命数 知乎@余命数

分子的三项依次化简，得

$$\begin{aligned}
&\sin(\pi+\alpha)=-\sin\alpha && \text{（偶不变，III 象限 sin 为负），}\\
&\cos(2\pi-\alpha)=\cos(-\alpha)=\cos\alpha && \text{（周期性，IV 象限 cos 为正），}\\
&\tan(3\pi-\alpha)=\tan(-\alpha)=-\tan\alpha && \text{（周期 }\pi\text{，II 象限 tan 为负）.}
\end{aligned}$$

分母的两项为

$$\begin{aligned}
&\cos(\frac{\pi}{2}+\alpha)=-\sin\alpha && \text{（奇变，II 象限 cos 为负），}\\
&\sin(-\alpha-\pi)=\sin(-(\pi+\alpha))=-\sin(\pi+\alpha)=-(-\sin\alpha)=\sin\alpha && \text{（先提内层负号）.}
\end{aligned}$$

将上述结果代回原分式，

$$\text{原式}=\frac{(-\sin\alpha)(\cos\alpha)(-\tan\alpha)}{(-\sin\alpha)(\sin\alpha)}=\frac{\sin\alpha\cdot\cos\alpha\cdot\tan\alpha}{-\sin^2\alpha},$$

再以商数关系 $\tan\alpha=\frac{\sin\alpha}{\cos\alpha}$ 切化弦，得

$$\text{原式}=\frac{\sin\alpha\cdot\cos\alpha\cdot(\frac{\sin\alpha}{\cos\alpha})}{-\sin^2\alpha}=\frac{\sin^2\alpha}{-\sin^2\alpha}=-1.$$

知乎@余命数

化简结果为 -1.

注 1.3 (经验总结)
处理复杂的三角表达式时, 我们把每一项单独化简; 对 $\sin(-\alpha-\pi)$ 这类复合角, 先提取括号内的负号; 最后若同时出现 sin、cos、tan, 可以把 tan 写成 $\frac{\sin}{\cos}$, 方便约分.

知乎@余命数

例题 1.15 在 $\triangle ABC$ 中, 求证.
(1) $\sin(A+B)=\sin C$ (2) $\cos A=-\cos(B+C)$ (3) $\sin\frac{A+B}{2}=\cos\frac{C}{2}$

知乎@余命数

解 在三角形中, $A+B+C=\pi$. 我们用它把所证式中的角改写为诱导公式的标准形式, 即可得证.

(1) 由 $A+B+C=\pi$ 得 $A+B=\pi-C$, 代入等式左边并用诱导公式 $\sin(\pi-C)=\sin C$,

$$\sin(A+B)=\sin(\pi-C)=\sin C,$$

故 $\sin(A+B)=\sin C$.

(2) 由 $A+B+C=\pi$ 得 $B+C=\pi-A$, 考察等式右边并用诱导公式 $\cos(\pi-A)=-\cos A$,

$$-\cos(B+C)=-\cos(\pi-A)=-(-\cos A)=\cos A,$$

右边即等于左边的 $\cos A$, 故 $\cos A=-\cos(B+C)$.

知乎@余命数

(3) 由 $A+B+C=\pi$, 两边同除以 2 得 $\frac{A+B}{2}+\frac{C}{2}=\frac{\pi}{2}$, 即 $\frac{A+B}{2}=\frac{\pi}{2}-\frac{C}{2}$, 代入等式左边并用诱导公式 $\sin(\frac{\pi}{2}-x)=\cos x$,

知乎@余命数

$$\sin\frac{A+B}{2}=\sin\left(\frac{\pi}{2}-\frac{C}{2}\right)=\cos\frac{C}{2},$$

知乎@余命数

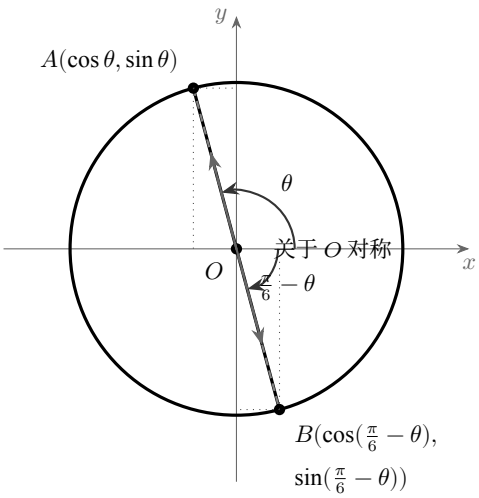
故 $\sin\frac{A+B}{2}=\cos\frac{C}{2}$.

注 1.4 (经验总结)
在三角形中, $A+B+C=\pi$ 可以把两角和写成 π 减第三角, 也可以把半角写成 $\frac{\pi}{2}$ 减第三角的一半. 这些是诱导公式的标准形式.

例题 1.16 (2023 年 12 月广东佛山高三测试)若点 $A(\cos\theta,\sin\theta)$ 关于原点对称点为 $B\left(\cos(\frac{\pi}{6}-\theta),\sin(\frac{\pi}{6}-\theta)\right)$, 写出 θ 的一个取值为 _____.

解 点 A 在单位圆上, 关于原点对称的点 B 对应的角与 A 对应的角相差 π 的奇数倍.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵



点 A 与点 B 关于原点对称, 意味着其终边互为反向延长线, 对应的角相差 π 的奇数倍. 设 $\theta_A=\theta, \theta_B=\dfrac{\pi}{6}-\theta$, 由对称性有

$$\theta_B-\theta_A=(2k+1)\pi,\quad k\in\mathbb{Z}.$$

代入 θ_A, θ_B 并整理,

$$\begin{aligned}\left(\frac{\pi}{6}-\theta\right)-\theta&=(2k+1)\pi,\\ \frac{\pi}{6}-2\theta&=(2k+1)\pi,\\ -2\theta=(2k+1)\pi-\frac{\pi}{6}&=\frac{12k\pi+6\pi-\pi}{6}=\frac{(12k+5)\pi}{6},\end{aligned}$$

解得

$$\theta=-\frac{(12k+5)\pi}{12},\quad k\in\mathbb{Z}.$$

取 $k=0$ 得 $\theta=-\dfrac{5\pi}{12}$, 取 $k=-1$ 得 $\theta=-\dfrac{(-12+5)\pi}{12}=\dfrac{7\pi}{12}$. 因此 θ 的一个可能取值为 $\dfrac{7\pi}{12}$ 或 $-\dfrac{5\pi}{12}$ 等.

注 1.5 (经验总结)
本题把几何对称转化为代数关系. 几何上, 关于原点对称说明两个角的终边在同一直线上; 代数上, 这等价于两个角相差 π 的奇数倍, 即 $\theta_2-\theta_1=(2k+1)\pi$. 建立方程后, 再用代数运算求解. 处理单位圆问题时, 这种几何直观常常能直接给出突破口.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

例题 1.17 （2023 年 11 月广东金太阳联考）设 $\alpha \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right], \beta \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$，且 $\sin\alpha + \cos\alpha = \sqrt{2}\cos\beta$，则（　）

A. $\alpha + \beta = \frac{\pi}{4}$

B. $\alpha - \beta = \frac{\pi}{4}$

C. $\alpha + \beta = \frac{\pi}{2}$

D. $\alpha - \beta = -\frac{\pi}{4}$

知乎@余命数

解 等式左边可用辅助角公式合并，右边是 $\sqrt{2}\cos\beta$，结合两边在给定区间上的值域即可锁定 α 和 β.

对等式左边用辅助角公式变形，

$$\sin\alpha + \cos\alpha = \sqrt{2}\left(\frac{\sqrt{2}}{2}\sin\alpha + \frac{\sqrt{2}}{2}\cos\alpha\right) = \sqrt{2}\sin\left(\alpha + \frac{\pi}{4}\right),$$

代回原等式并约去 $\sqrt{2}$，得

知乎@余命数

$$\sin\left(\alpha + \frac{\pi}{4}\right) = \cos\beta.$$

此方程在给定区间上不能直接求解，我们改先考察两边的值域. 由 $\alpha \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$ 知 $\alpha + \frac{\pi}{4} \in \left[\frac{\pi}{2}, \frac{3\pi}{4}\right]$，在该区间上 sin 的值域为 $\left[\sin\frac{3\pi}{4}, \sin\frac{\pi}{2}\right] = \left[\frac{\sqrt{2}}{2}, 1\right]$；而 $\beta \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$，cos 的值域为 $\left[\cos\frac{\pi}{2}, \cos\frac{\pi}{4}\right] = \left[0, \frac{\sqrt{2}}{2}\right]$. 两边值域的唯一公共值为 $\frac{\sqrt{2}}{2}$，故必须同时满足

知乎@余命数

知乎@余命数

$$\sin\left(\alpha + \frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} \quad \text{且} \quad \cos\beta = \frac{\sqrt{2}}{2}.$$

在 $\alpha + \frac{\pi}{4} \in \left[\frac{\pi}{2}, \frac{3\pi}{4}\right]$ 中 $\sin(\alpha + \frac{\pi}{4}) = \frac{\sqrt{2}}{2}$ 给出 $\alpha + \frac{\pi}{4} = \frac{3\pi}{4}$，即 $\alpha = \frac{\pi}{2}$；在 $\beta \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$ 中 $\cos\beta = \frac{\sqrt{2}}{2}$ 给出 $\beta = \frac{\pi}{4}$. 因此 $\alpha - \beta = \frac{\pi}{2} - \frac{\pi}{4} = \frac{\pi}{4}$，答案为 B.

注 1.6 (经验总结)
当三角方程难以直接求解，且自变量范围已知时，可以分别求出等式两边的值域，取交集后往往能得到唯一可能的函数值，从而确定变量的取值.

1.8.3 高考真题选讲

例题 1.18 (2010 年全国 □ 卷文科数学第 3 题) 已知 $\sin\alpha = \frac{2}{3}$，则 $\cos(\pi - 2\alpha) =$

A. $-\frac{\sqrt{5}}{3}$　　C. $\frac{1}{9}$

B. $-\frac{1}{9}$　　D. $\frac{\sqrt{5}}{3}$

解 答案：B. 先用诱导公式 $\cos(\pi - 2\alpha) = -\cos 2\alpha$，再用倍角公式 $\cos 2\alpha = 1 - 2\sin^2\alpha$，

$$\cos(\pi - 2\alpha) = -(1 - 2\sin^2\alpha) = -\left(1 - 2\cdot\frac{4}{9}\right) = -\frac{1}{9}.$$

知乎@余命数

知乎关注@余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

1.9 和差角公式

我们已经会求 30°、45°、60° 等特殊角的三角函数值. 遇到 75° 或 15° 这类角时, 我们想用已知角表示它们. 例如

$$75^\circ = 45^\circ + 30^\circ, \qquad 15^\circ = 45^\circ - 30^\circ.$$

一个自然的猜测是 $\cos(\alpha-\beta)$ 等于 $\cos\alpha-\cos\beta$. 但这个等式不成立. 取 $\alpha = 60^\circ, \beta = 30^\circ$, 则 $\cos(60^\circ - 30^\circ) = \cos 30^\circ = \frac{\sqrt{3}}{2}$, 而 $\cos 60^\circ - \cos 30^\circ = \frac{1}{2} - \frac{\sqrt{3}}{2}$, 两者不相等.

知乎@余命数

我们关心两个角的和差如何进入三角函数. 下面的推导借助单位圆上的向量点积.

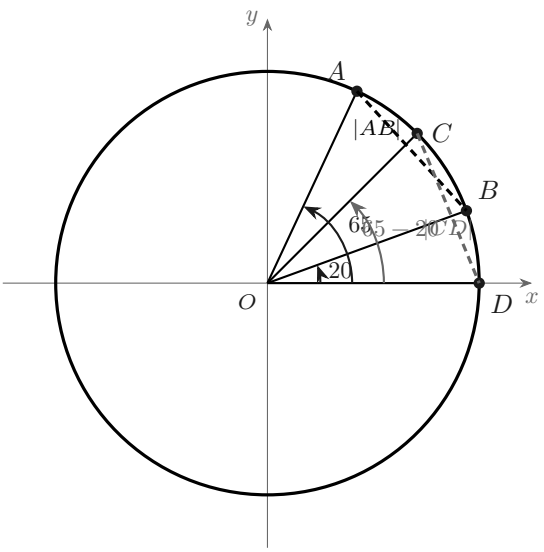
1.9.1 两角差的余弦

定理 1.3 (两角差的余弦) 对于任意角 α, β, 恒有

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta \tag{1-6}$$

证明 我们用单位圆上等弧对等弦的性质来证明. 如图所示, 在单位圆上取四点 A, B, C, D, 对应的角分别为 $\alpha, \beta, \alpha-\beta, 0$.

知乎@余命数



知乎@余命数

证明依赖等弧对等弦. 圆心角 $\angle AOB = \angle COD = \alpha - \beta$, 故弦长相等, 即 $|AB| = |CD|$. 两边平方得 $|AB|^2 = |CD|^2$, 用两点间距离公式把两端分别展开,

$$(\cos\alpha - \cos\beta)^2 + (\sin\alpha - \sin\beta)^2 = (\cos(\alpha-\beta) - 1)^2 + \sin^2(\alpha-\beta).$$

知乎@余命数

左端展开并利用 $\sin^2\theta+\cos^2\theta=1$ 化简,

$$\begin{aligned}&\cos^2\alpha-2\cos\alpha\cos\beta+\cos^2\beta+\sin^2\alpha-2\sin\alpha\sin\beta+\sin^2\beta\\&=2-2(\cos\alpha\cos\beta+\sin\alpha\sin\beta),\end{aligned}$$

右端同理化简,

$$\begin{aligned}&\cos^2(\alpha-\beta)-2\cos(\alpha-\beta)+1+\sin^2(\alpha-\beta)\\&=2-2\cos(\alpha-\beta).\end{aligned}$$

两端相等, 即

$$2-2(\cos\alpha\cos\beta+\sin\alpha\sin\beta)=2-2\cos(\alpha-\beta),$$

整理即得

$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta.$$

□

这个公式是和差角公式的起点, 其余公式都由它通过代数变形得到. 它的结构可记为余余正正, 符号反.

1.9.2 公式体系的构建

我们由 $\cos(\alpha-\beta)$ 推出其余和差角公式.

- **推导** $\cos(\alpha+\beta)$
  将 $\alpha+\beta$ 看作 $\alpha-(-\beta)$.

$$\begin{aligned}\cos(\alpha+\beta)&=\cos(\alpha-(-\beta))\\&=\cos\alpha\cos(-\beta)+\sin\alpha\sin(-\beta)\\&=\cos\alpha\cos\beta-\sin\alpha\sin\beta\end{aligned}$$

- **推导** $\sin(\alpha+\beta)$
  利用诱导公式 $\sin\theta=\cos(\frac{\pi}{2}-\theta)$.

$$\begin{aligned}\sin(\alpha+\beta)&=\cos\left(\frac{\pi}{2}-(\alpha+\beta)\right)=\cos\left(\left(\frac{\pi}{2}-\alpha\right)-\beta\right)\\&=\cos(\frac{\pi}{2}-\alpha)\cos\beta+\sin(\frac{\pi}{2}-\alpha)\sin\beta\\&=\sin\alpha\cos\beta+\cos\alpha\sin\beta\end{aligned}$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

定理 1.4 (和差角公式)

$$\sin(\alpha\pm\beta)=\sin\alpha\cos\beta\pm\cos\alpha\sin\beta \qquad \text{(正余余正, 符号同)}$$

$$\cos(\alpha\pm\beta)=\cos\alpha\cos\beta\mp\sin\alpha\sin\beta \qquad \text{(余余正正, 符号反)}$$

$$\tan(\alpha\pm\beta)=\frac{\tan\alpha\pm\tan\beta}{1\mp\tan\alpha\tan\beta} \qquad \text{(分子符号同, 分母符号反)}$$

其中正切公式要求 $\tan\alpha,\tan\beta,\tan(\alpha\pm\beta)$ 均有意义, 且分母 $1\mp\tan\alpha\tan\beta\neq0$.

证明 我们用 $\cos(\alpha-\beta)$ 的公式、奇偶性与诱导公式直接推出正弦、余弦的和差公式. 例如

$$\cos(\alpha+\beta)=\cos(\alpha-(-\beta))=\cos\alpha\cos\beta-\sin\alpha\sin\beta,$$

而

$$\sin(\alpha+\beta)=\cos\left(\frac{\pi}{2}-(\alpha+\beta)\right)=\sin\alpha\cos\beta+\cos\alpha\sin\beta.$$

将 β 替换为 $-\beta$ 即得相应的差角公式.

当相关正切均有意义时, 由商数关系可得

$$\begin{aligned}\tan(\alpha+\beta)&=\frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)}\\&=\frac{\sin\alpha\cos\beta+\cos\alpha\sin\beta}{\cos\alpha\cos\beta-\sin\alpha\sin\beta}.\end{aligned}$$

在 $\cos\alpha\cos\beta\neq0$ 时, 分子分母同除以 $\cos\alpha\cos\beta$, 得

$$\tan(\alpha+\beta)=\frac{\tan\alpha+\tan\beta}{1-\tan\alpha\tan\beta}.$$

差角正切公式同理, 或将 β 替换为 $-\beta$ 得到. 分母非零正对应 $\cos(\alpha\pm\beta)\neq0$, 即 $\tan(\alpha\pm\beta)$ 有意义. □

1.9.3 辅助角公式

我们经常遇到函数 $f(x)=a\sin x+b\cos x$; 它表示两个同频率但不同相位的简谐波叠加. 和差角公式把这个和式合并成一个单一的正弦或余弦函数, 从而可以直接分析最值、周期和单调性.

我们称这种变形为**辅助角公式**.

$$a\sin x+b\cos x=\sqrt{a^2+b^2}\left(\frac{a}{\sqrt{a^2+b^2}}\sin x+\frac{b}{\sqrt{a^2+b^2}}\cos x\right)$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

关键在于构造一个辅助角. 令 $\cos\varphi=\frac{a}{\sqrt{a^2+b^2}}$, $\sin\varphi=\frac{b}{\sqrt{a^2+b^2}}$, 则 $\tan\varphi=\frac{b}{a}$. 于是

$$\begin{aligned}\text{原式}&=\sqrt{a^2+b^2}(\cos\varphi\sin x+\sin\varphi\cos x)\\&=\sqrt{a^2+b^2}\sin(x+\varphi)\end{aligned}$$

注 1.7 (辅助角变形)
要把 $a\sin x+b\cos x$ 写成 $\sqrt{a^2+b^2}\sin(x+\varphi)$ 的形式, 我们先提取系数 $\sqrt{a^2+b^2}$, 再把括号内配成 $\sin x\cos\varphi+\cos x\sin\varphi$ 或 $\cos x\cos\varphi+\sin x\sin\varphi$ 的形式, 最后由 $\tan\varphi=\dfrac{b}{a}$ 与 a,b 的符号确定 φ 所在象限.

1.9.4 典例解析

知乎@余命数

例题 1.19 (给值求值) 已知 $\sin\alpha=\frac{5}{13}$、$\alpha\in(\frac{\pi}{2},\pi)$、$\cos\beta=-\frac{3}{5}$、$\beta$ 是第三象限角. 求 $\sin(\alpha+\beta)$ 的值.

余命数 知乎@余命数

解 我们先用象限条件补全各角的正弦、余弦, 再代入和差角公式. 由 $\alpha\in(\frac{\pi}{2},\pi)$ 知 $\cos\alpha<0$, 故

知乎@余命数

$$\cos\alpha=-\sqrt{1-\sin^2\alpha}=-\sqrt{1-(\frac{5}{13})^2}=-\sqrt{\frac{144}{169}}=-\frac{12}{13};$$

又 β 是第三象限角, $\sin\beta<0$, 故

知乎@余命数

$$\sin\beta=-\sqrt{1-\cos^2\beta}=-\sqrt{1-(-\frac{3}{5})^2}=-\sqrt{\frac{16}{25}}=-\frac{4}{5}.$$

知乎@余命数

代入正弦和角公式 $\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta$, 得

$$\begin{aligned}\sin(\alpha+\beta)&=\left(\frac{5}{13}\right)\left(-\frac{3}{5}\right)+\left(-\frac{12}{13}\right)\left(-\frac{4}{5}\right)\\&=-\frac{15}{65}+\frac{48}{65}=\frac{33}{65}.\end{aligned}$$

知乎@余命数

例题 1.20 (函数性质分析) 求函数 $f(x)=\sqrt{3}\sin x-\cos x$ 的最小正周期和单调递增区间.

解 我们先用辅助角公式把函数化为 $y=A\sin(\omega x+\varphi)+k$ 的形式, 再读出周期与单调性. 此处 $a=\sqrt{3},b=-1$, 故 $\sqrt{a^2+b^2}=\sqrt{3+1}=2$, 从而

$$f(x)=2\left(\frac{\sqrt{3}}{2}\sin x-\frac{1}{2}\cos x\right)$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$
\begin{aligned}
&= 2\left(\sin x\cos\frac{\pi}{6} - \cos x\sin\frac{\pi}{6}\right)\\
&= 2\sin\left(x-\frac{\pi}{6}\right).
\end{aligned}
$$

由标准形式 $2\sin(x-\frac{\pi}{6})$ 知最小正周期 $T=\frac{2\pi}{|\omega|}=\frac{2\pi}{1}=2\pi$. 求单调递增区间, 须令内层 $u=x-\frac{\pi}{6}$ 落在正弦的标准递增区间 $2k\pi-\frac{\pi}{2}\le u\le 2k\pi+\frac{\pi}{2}$ $(k\in\mathbb{Z})$ 上, 即

$$
\begin{aligned}
&2k\pi-\frac{\pi}{2}\le x-\frac{\pi}{6}\le 2k\pi+\frac{\pi}{2},\\
&2k\pi-\frac{\pi}{3}\le x\le 2k\pi+\frac{2\pi}{3}\quad(\text{两端加 }\frac{\pi}{6}),
\end{aligned}
$$

故单调递增区间为 $\left[2k\pi-\frac{\pi}{3},2k\pi+\frac{2\pi}{3}\right],k\in\mathbb{Z}$.

例题 1.21 (2024 年 1 月金丽衢十二校第一次联考) 已知 α 是第二象限角, $\beta\in(0,\frac{\pi}{2})$, $\tan(\alpha+\frac{\pi}{4})=-\frac{1}{4}$. 现将角 α 的终边逆时针旋转 β 后得到角 γ, 若 $\tan\gamma=\frac{1}{7}$, 则 $\tan\beta=$ _____.

解 我们把逆时针旋转 β 理解为 $\gamma=\alpha+\beta$, 于是 $\beta=\gamma-\alpha$, 再用差角公式求 $\tan\beta$. 这要求我们先求出 $\tan\gamma$ 和 $\tan\alpha$.

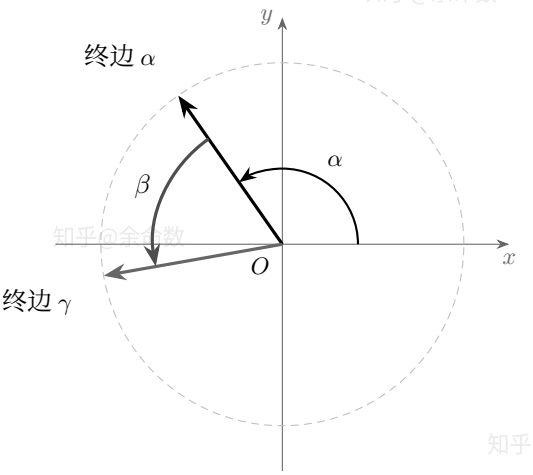


图 1.16　角 α 的终边逆时针旋转角 β 得到角 γ 的终边

先求 $\tan\alpha$. 由正切和角公式

$$
\tan\left(\alpha+\frac{\pi}{4}\right)=\frac{\tan\alpha+\tan\frac{\pi}{4}}{1-\tan\alpha\tan\frac{\pi}{4}}=\frac{\tan\alpha+1}{1-\tan\alpha},
$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

代入已知值 $\tan(\alpha+\frac{\pi}{4})=-\frac{1}{4}$ 得方程 $\frac{1+\tan\alpha}{1-\tan\alpha}=-\frac{1}{4}$, 交叉相乘解得

$$4(1+\tan\alpha)=-(1-\tan\alpha)\implies 4+4\tan\alpha=-1+\tan\alpha\implies 3\tan\alpha=-5,$$

即 $\tan\alpha=-\frac{5}{3}$(α 是第二象限角, $\tan\alpha$ 为负, 与题意相符).

再由 $\tan\gamma=\frac{1}{7}$ 与差角公式求 $\tan\beta=\tan(\gamma-\alpha)$,

$$\begin{aligned}\tan\beta=\tan(\gamma-\alpha)&=\frac{\tan\gamma-\tan\alpha}{1+\tan\gamma\tan\alpha}\\&=\frac{\frac{1}{7}-\left(-\frac{5}{3}\right)}{1+\frac{1}{7}\cdot\left(-\frac{5}{3}\right)}\\&=\frac{\frac{3+35}{21}}{1-\frac{5}{21}}=\frac{\frac{38}{21}}{\frac{16}{21}}=\frac{38}{16}=\frac{19}{8}.\end{aligned}$$

故 $\tan\beta=\dfrac{19}{8}$.

注 1.8 (旋转与差角)
几何描述给出终边旋转关系 $\gamma=\alpha+\beta$, 于是 $\beta=\gamma-\alpha$. 要求 $\tan\beta$, 只需求出 $\tan\gamma$ 和 $\tan\alpha$. 而已知 $\tan(\alpha+\frac{\pi}{4})$ 可通过正切和角公式化为关于 $\tan\alpha$ 的方程.

例题 1.22 (2023 年 12 月江苏省决胜新高考三大联考) 已知实数 m,n 满足 $(m+1)(n+1)=2$, 则 m,n 可能是(　)

A. $m=\tan\frac{\pi}{16},n=\tan\frac{3\pi}{16}$

B. $m=\tan\frac{\pi}{8},n=\tan\frac{3\pi}{8}$

C. $m=\cos\frac{\pi}{16},n=\tan\frac{3\pi}{16}$

D. $m=\cos\frac{\pi}{8},n=\tan\frac{3\pi}{8}$

解 我们先把代数式变形, 看出它与正切和角公式的结构相同, 再用三角函数检验选项. 将 $(m+1)(n+1)=2$ 展开并移项, 得

$$mn+m+n+1=2\implies m+n=1-mn.$$

当 $1-mn\neq 0$ 时, 上式可化为 $\frac{m+n}{1-mn}=1$, 此形式与正切和角公式 $\tan(A+B)=\frac{\tan A+\tan B}{1-\tan A\tan B}$ 一致, 故令 $m=\tan A,n=\tan B$, 条件即转化为 $\tan(A+B)=1$, 即 $A+B=k\pi+\frac{\pi}{4}$ ($k\in\mathbb{Z}$).

逐一检验哪个选项满足 $A+B=k\pi+\frac{\pi}{4}$.

- **A**: $A=\frac{\pi}{16},B=\frac{3\pi}{16}$, $A+B=\frac{4\pi}{16}=\frac{\pi}{4}$, 满足 ($k=0$), 故 A 可能.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **B**: $A=\frac{\pi}{8}, B=\frac{3\pi}{8}, A+B=\frac{\pi}{2}$, 此时 $\tan(A+B)$ 无定义, 对应分母 $1-mn=0$, 即 $mn=1$; 代入原式 $mn+m+n+1=2$ 得 $m+n=0$, 但 $m=\tan\frac{\pi}{8}>0, n=\tan\frac{3\pi}{8}>0$, 和不能为 0, 故 B 错误.
- **C**: $m=\cos\frac{\pi}{16}>\frac{1}{2}$, 且 $n=\tan\frac{3\pi}{16}>\tan\frac{\pi}{8}=\sqrt{2}-1$, 因而 $(m+1)(n+1)>\frac{3}{2}\cdot\sqrt{2}>2$, 不满足条件.
- **D**: $m=\cos\frac{\pi}{8}>0$, 且 $n=\tan\frac{3\pi}{8}>1$, 因而 $(m+1)(n+1)>1\cdot 2=2$, 不满足条件.

正确答案为 A.

@余命数　知乎@余命数

注 1.9 (结构识别)
把 $(m+1)(n+1)=2$ 化简为 $m+n=1-mn$, 于是 $\frac{m+n}{1-mn}=1$. 这个式子与正切和角公式 $\tan(A+B)=\frac{\tan A+\tan B}{1-\tan A\tan B}$ 结构相同, 因此可令 $m=\tan A, n=\tan B$, 问题化为检验 $A+B=k\pi+\frac{\pi}{4}$.

例题 1.23 (2023 年 12 月山东中学联盟高三联考) 已知 $\cos(\beta-\alpha)=\frac{3}{5}$, 且 $\tan\alpha\tan\beta=\frac{1}{2}$, 求 $\cos 2(\alpha+\beta)$ 的值(　)

A. $-\frac{2}{25}$
B. $\frac{23}{25}$
C. $\frac{2}{25}$
D. $-\frac{23}{25}$

解 我们把 $\cos\alpha\cos\beta$ 与 $\sin\alpha\sin\beta$ 看作两个整体未知数, 解方程组求出它们, 再求目标值.

由已知条件得到两个关系式, 即 $\cos(\beta-\alpha)=\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta=\frac{3}{5}$ 与 $\tan\alpha\tan\beta=\frac{\sin\alpha\sin\beta}{\cos\alpha\cos\beta}=\frac{1}{2}$. 令 $X=\cos\alpha\cos\beta, Y=\sin\alpha\sin\beta$, 上述两式化为关于 X,Y 的方程组

$$\begin{cases} X+Y=\frac{3}{5} \\ \frac{Y}{X}=\frac{1}{2} \end{cases}$$

知乎@余命数

由第二式得 $X=2Y$, 代入第一式有 $3Y=\frac{3}{5}$, 即 $Y=\frac{1}{5}, X=\frac{2}{5}$, 故

知乎@余命数

$$\sin\alpha\sin\beta=\frac{1}{5},\quad \cos\alpha\cos\beta=\frac{2}{5}.$$

要计算 $\cos 2(\alpha+\beta)$, 我们先求 $\cos(\alpha+\beta)$, 再用二倍角公式 $\cos 2\theta = 2\cos^2\theta - 1$. 由余弦和角公式,

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta = X - Y = \frac{2}{5} - \frac{1}{5} = \frac{1}{5},$$

代入二倍角公式,

$$\begin{aligned}\cos 2(\alpha+\beta) &= 2\cos^2(\alpha+\beta) - 1 = 2\left(\frac{1}{5}\right)^2 - 1\\ &= \frac{2}{25} - \frac{25}{25} = -\frac{23}{25}.\end{aligned}$$

正确答案为 D.

1.9.5 高考真题选讲

例题 1.24 (2011 年浙江卷理科数学第 6 题) 若 $0 < \alpha < \frac{\pi}{2}$、$-\frac{\pi}{2} < \beta < 0$、$\cos\left(\frac{\pi}{4}+\alpha\right) = \frac{1}{3}$、$\cos\left(\frac{\pi}{4}-\frac{\beta}{2}\right) = \frac{\sqrt{3}}{3}$, 则 $\cos\left(\alpha+\frac{\beta}{2}\right) =$

A. $\frac{\sqrt{3}}{3}$　　C. $\frac{5\sqrt{3}}{9}$

B. $-\frac{\sqrt{3}}{3}$　　D. $-\frac{\sqrt{6}}{9}$

解 答案:C. 注意到 $\alpha+\frac{\beta}{2} = \left(\frac{\pi}{4}+\alpha\right) - \left(\frac{\pi}{4}-\frac{\beta}{2}\right)$, 可直接用余弦差角公式. 先求两个角的正弦值. 由 $0 < \alpha < \frac{\pi}{2}$ 知 $\frac{\pi}{4} < \frac{\pi}{4}+\alpha < \frac{3\pi}{4}$, 故 $\sin\left(\frac{\pi}{4}+\alpha\right) = \sqrt{1-\frac{1}{9}} = \frac{2\sqrt{2}}{3}$; 由 $-\frac{\pi}{2} < \beta < 0$ 知 $\frac{\pi}{4} < \frac{\pi}{4}-\frac{\beta}{2} < \frac{\pi}{2}$, 故 $\sin\left(\frac{\pi}{4}-\frac{\beta}{2}\right) = \sqrt{1-\frac{1}{3}} = \frac{\sqrt{6}}{3}$.

代入余弦差角公式,

$$\cos\left(\alpha+\frac{\beta}{2}\right) = \frac{1}{3}\cdot\frac{\sqrt{3}}{3} + \frac{2\sqrt{2}}{3}\cdot\frac{\sqrt{6}}{3} = \frac{\sqrt{3}}{9} + \frac{4\sqrt{3}}{9} = \frac{5\sqrt{3}}{9}.$$

1.10 倍角公式

来龙去脉

我们在和角公式中令两个角相同, 即设 $\alpha = \beta$, 便得到二**倍角公式**.

二倍角公式把角和它的二倍角联系起来. 由它移项可得**降幂公式**; 把倍角换成半角又可得**半角公式**. 这些公式在化简与求值中都会用到.

知乎@余命数

1.10.1 公式体系

定理 1.5 (倍角、半角公式)　1. **二倍角公式**

$$\begin{aligned}\sin(2\alpha) &= 2\sin\alpha\cos\alpha\\ \cos(2\alpha) &= \cos^2\alpha-\sin^2\alpha\\ &= 2\cos^2\alpha-1\\ &= 1-2\sin^2\alpha\\ \tan(2\alpha) &= \frac{2\tan\alpha}{1-\tan^2\alpha}\quad(\tan\alpha\text{ 有意义且 }\tan\alpha\neq\pm1)\end{aligned}$$

2. **降幂公式**

$$\begin{aligned}\sin^2\alpha &= \frac{1-\cos(2\alpha)}{2}\\ \cos^2\alpha &= \frac{1+\cos(2\alpha)}{2}\end{aligned}$$

3. **半角公式**

$$\begin{aligned}\sin\frac{\alpha}{2} &= \pm\sqrt{\frac{1-\cos\alpha}{2}}\\ \cos\frac{\alpha}{2} &= \pm\sqrt{\frac{1+\cos\alpha}{2}}\\ \tan\frac{\alpha}{2} &= \pm\sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}}\end{aligned}$$

在分母有意义时, 还可写成

$$\tan\frac{\alpha}{2}=\frac{\sin\alpha}{1+\cos\alpha}\quad(\cos\alpha\neq-1),\qquad \tan\frac{\alpha}{2}=\frac{1-\cos\alpha}{\sin\alpha}\quad(\sin\alpha\neq0).$$

半角公式根号前的正负号由角 $\frac{\alpha}{2}$ 所在的象限决定. 正切的后两个无根号形式在化简中更为常用, 但使用前必须检查分母是否为零.

证明　我们用两角和的三角函数公式推导这些公式.

先推二倍角公式. 在两角和公式中令两个角相等, 即设 $\beta=\alpha$. 对正弦函数, 由 $\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta$ 得

$$\sin(2\alpha)=\sin(\alpha+\alpha)=\sin\alpha\cos\alpha+\cos\alpha\sin\alpha=2\sin\alpha\cos\alpha;$$

对余弦函数, 由 $\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta$ 得

$$\cos(2\alpha)=\cos(\alpha+\alpha)=\cos^2\alpha-\sin^2\alpha,$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

再结合平方关系 $\sin^2\alpha+\cos^2\alpha=1$, 得 $\cos(2\alpha)$ 的另外两种常用形式

$$\cos(2\alpha)=\cos^2\alpha-(1-\cos^2\alpha)=2\cos^2\alpha-1,$$

$$\cos(2\alpha)=(1-\sin^2\alpha)-\sin^2\alpha=1-2\sin^2\alpha;$$

对正切函数, 由商数关系及已证二倍角公式,

$$\tan(2\alpha)=\frac{\sin(2\alpha)}{\cos(2\alpha)}=\frac{2\sin\alpha\cos\alpha}{\cos^2\alpha-\sin^2\alpha},$$

为得到用 $\tan\alpha$ 表示的形式, 分子分母同除以 $\cos^2\alpha$(要求 $\cos\alpha\neq0$), 得

$$\tan(2\alpha)=\frac{2\frac{\sin\alpha}{\cos\alpha}}{1-\frac{\sin^2\alpha}{\cos^2\alpha}}=\frac{2\tan\alpha}{1-\tan^2\alpha}.$$

再推降幂公式. 它由余弦二倍角公式移项得到. 由 $\cos(2\alpha)=1-2\sin^2\alpha$ 移项得 $\sin^2\alpha=\frac{1-\cos(2\alpha)}{2}$; 同理由 $\cos(2\alpha)=2\cos^2\alpha-1$ 得 $\cos^2\alpha=\frac{1+\cos(2\alpha)}{2}$.

最后推半角公式. 在降幂公式中令角为 $\frac{\alpha}{2}$ 即可. 在 $\sin^2\theta=\frac{1-\cos(2\theta)}{2}$ 中令 $\theta=\frac{\alpha}{2}$, 则 $2\theta=\alpha$, 得

$$\sin^2\frac{\alpha}{2}=\frac{1-\cos\alpha}{2}\implies\sin\frac{\alpha}{2}=\pm\sqrt{\frac{1-\cos\alpha}{2}};$$

同理在 $\cos^2\theta=\frac{1+\cos(2\theta)}{2}$ 中作相同代换, 得

$$\cos^2\frac{\alpha}{2}=\frac{1+\cos\alpha}{2}\implies\cos\frac{\alpha}{2}=\pm\sqrt{\frac{1+\cos\alpha}{2}}.$$

正切的半角公式还有无根号形式. 利用商数关系及倍角、降幂公式,

$$\tan\frac{\alpha}{2}=\frac{\sin\frac{\alpha}{2}}{\cos\frac{\alpha}{2}}=\frac{2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}}{2\cos^2\frac{\alpha}{2}}=\frac{\sin\alpha}{1+\cos\alpha},$$

或分子分母同乘 $2\sin\frac{\alpha}{2}$, 得另一形式

$$\tan\frac{\alpha}{2}=\frac{\sin\frac{\alpha}{2}}{\cos\frac{\alpha}{2}}=\frac{2\sin^2\frac{\alpha}{2}}{2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}}=\frac{1-\cos\alpha}{\sin\alpha}.$$

这两个有理形式避免了开方和符号判断, 常在化简中使用. □

1.10.2 内在联系与应用策略

- **$\cos(2\alpha)$ 的三种形式**
  $\cos(2\alpha)$ 既能连起倍角与单角, 移项后又给出升幂、降幂公式, 是处理三角幂次的工具.

知乎@余命数

- **角的相对性**
  倍角是相对的.α 是 $\frac{\alpha}{2}$ 的倍角，4α 是 2α 的倍角. 公式可以跨层次使用，例如 $\sin\alpha = 2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}$，$\cos(4x) = 2\cos^2(2x) - 1$.
- **1 的代换与弦切互化**
  当式子中出现 1 时，可写成 $1 = \sin^2\alpha + \cos^2\alpha$ 来齐次化. 例如 $\dfrac{1-\cos\alpha}{\sin\alpha} = \tan\dfrac{\alpha}{2}$.

知乎@余命数

1.10.3 典例解析

例题 1.25　已知 $\tan\alpha = -\frac{1}{7}$，$\sin\beta = \frac{\sqrt{10}}{10}$ 且 β 为锐角，求 $\tan(2\alpha+\beta)$ 的值.

解　我们要求 $\tan(2\alpha+\beta)$，先求 $\tan 2\alpha$ 与 $\tan\beta$，再用正切和角公式.

已知 $\tan\alpha = -\frac{1}{7}$，由正切二倍角公式

$$\tan(2\alpha) = \frac{2\tan\alpha}{1-\tan^2\alpha} = \frac{2(-\frac{1}{7})}{1-(-\frac{1}{7})^2} = \frac{-\frac{2}{7}}{1-\frac{1}{49}} = \frac{-\frac{2}{7}}{\frac{48}{49}} = -\frac{7}{24}.$$

再求 $\tan\beta$. 由 β 为锐角知 $\cos\beta > 0$，故

$$\cos\beta = \sqrt{1-\sin^2\beta} = \sqrt{1-\left(\frac{\sqrt{10}}{10}\right)^2} = \sqrt{\frac{90}{100}} = \frac{3\sqrt{10}}{10},$$

从而 $\tan\beta = \frac{\sin\beta}{\cos\beta} = \frac{\sqrt{10}/10}{3\sqrt{10}/10} = \frac{1}{3}$. 最后由正切和角公式，

$$\begin{aligned}\tan(2\alpha+\beta) &= \frac{\tan(2\alpha)+\tan\beta}{1-\tan(2\alpha)\tan\beta} = \frac{-\dfrac{7}{24}+\dfrac{1}{3}}{1-\left(-\dfrac{7}{24}\right)\left(\dfrac{1}{3}\right)} \\ &= \frac{\dfrac{-7+8}{24}}{1+\dfrac{7}{72}} = \frac{\dfrac{1}{24}}{\dfrac{79}{72}} = \frac{1}{24}\cdot\frac{72}{79} = \frac{3}{79}.\end{aligned}$$

因此 $\tan(2\alpha+\beta) = \frac{3}{79}$.

例题 1.26 (倍角公式的连乘技巧)（2024 年 1 月湖南邵阳高三第一次联考）已知 $\cos\theta\cos 2\theta\cos 4\theta = -\frac{1}{8}$，$\theta \in (\frac{\pi}{7}, \frac{\pi}{4})$，则 $2\cos^2 4\theta - \cos\theta =$ _____.

注 1.10 (变量准备) 本题需要先求 $\tan 2\alpha$ 与 $\tan\beta$，再把 $2\alpha+\beta$ 看作整体和角代入正切和角公式.

知乎@余命数

解 左侧是余弦的倍角连乘. 我们在等式两边同乘一个合适的正弦因子, 连续使用二倍角公式把它折叠成简洁方程. 设 $P=\cos\theta\cos 2\theta\cos 4\theta=-\frac{1}{8}$, 在等式两边同乘 $8\sin\theta$(由 $\theta\in(\frac{\pi}{7},\frac{\pi}{4})$ 知 $\sin\theta\neq 0$), 反复用 $2\sin x\cos x=\sin 2x$ 折叠,

$$\begin{aligned}
8\sin\theta\cos\theta\cos 2\theta\cos 4\theta &= -\sin\theta,\\
4(2\sin\theta\cos\theta)\cos 2\theta\cos 4\theta &= -\sin\theta,\\
4\sin 2\theta\cos 2\theta\cos 4\theta &= -\sin\theta,\\
2(2\sin 2\theta\cos 2\theta)\cos 4\theta &= -\sin\theta,\\
2\sin 4\theta\cos 4\theta &= -\sin\theta,\\
\sin 8\theta &= -\sin\theta \quad \text{(折叠完成)}.
\end{aligned}$$

由 $\sin 8\theta=-\sin\theta=\sin(-\theta)$ 及正弦函数的性质, 得

$$8\theta=-\theta+2k\pi \quad \text{或} \quad 8\theta=\pi-(-\theta)+2k\pi,\quad k\in\mathbb{Z},$$

即 $9\theta=2k\pi \implies \theta=\frac{2k\pi}{9}$, 或 $7\theta=(2k+1)\pi \implies \theta=\frac{(2k+1)\pi}{7}$. 将 $\theta\in(\frac{\pi}{7},\frac{\pi}{4})$ 即 $\theta\in(\frac{4\pi}{28},\frac{7\pi}{28})$ 代入检验: 对 $\theta=\frac{2k\pi}{9}$, 取 $k=1$ 得 $\theta=\frac{2\pi}{9}=\frac{56\pi}{252}$, 而 $\frac{\pi}{7}=\frac{36\pi}{252}$, $\frac{\pi}{4}=\frac{63\pi}{252}$, 由 $36<56<63$ 知 $\theta=\frac{2\pi}{9}$ 是一个解; 对 $\theta=\frac{(2k+1)\pi}{7}$, $k=0$ 给出 $\theta=\frac{\pi}{7}$ 不在区间内, $k\geq 1$ 给出 $\theta\geq\frac{3\pi}{7}>\frac{\pi}{4}$. 故唯一解为 $\theta=\frac{2\pi}{9}$.

所求表达式 $2\cos^2 4\theta-\cos\theta$ 用降幂公式 $2\cos^2 x=\cos(2x)+1$ 化为 $\cos(8\theta)+1-\cos\theta$. 由 $9\theta=2\pi$ 知 $8\theta=2\pi-\theta$, 故 $\cos(8\theta)=\cos(2\pi-\theta)=\cos\theta$, 代入得

$$\cos\theta+1-\cos\theta=1.$$

注 1.11 (经验总结)
形如 $\cos x\cos(2x)\cos(4x)\cdots$ 的连乘式, 常用方法是两边同乘最简角的正弦(这里为 $\sin\theta$), 连续使用 $2\sin\alpha\cos\alpha=\sin(2\alpha)$, 把式子逐步折叠成一个简洁的方程.

例题 1.27 (2024 年广州广雅中学二调) 已知锐角 θ 满足 $2\cos 2\theta=1+\sin 2\theta$, 则 $\tan\theta=$(　)

A. $\frac{1}{3}$

B. $\frac{1}{2}$

C. 2

D. 3

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

解 方程两边都含倍角 2θ, 先把它化为只含单角 θ 的式子. 左边用余弦二倍角公式,

$$2\cos 2\theta = 2(\cos^2\theta - \sin^2\theta) = 2(\cos\theta - \sin\theta)(\cos\theta + \sin\theta);$$

右边把 1 写成 $\sin^2\theta + \cos^2\theta$ 后与 $\sin 2\theta$ 配成完全平方,

$$1 + \sin 2\theta = (\sin^2\theta + \cos^2\theta) + 2\sin\theta\cos\theta = (\sin\theta + \cos\theta)^2.$$

于是原方程化为

$$2(\cos\theta - \sin\theta)(\cos\theta + \sin\theta) = (\sin\theta + \cos\theta)^2.$$

由 θ 是锐角知 $\sin\theta + \cos\theta \neq 0$, 两边可同除以 $\sin\theta + \cos\theta$, 得

$$2(\cos\theta - \sin\theta) = \sin\theta + \cos\theta,$$
$$2\cos\theta - 2\sin\theta = \sin\theta + \cos\theta,$$
$$\cos\theta = 3\sin\theta.$$

再在 $\cos\theta = 3\sin\theta$ 两边同除以 $\cos\theta$(θ 为锐角, $\cos\theta \neq 0$), 得

$$1 = 3\frac{\sin\theta}{\cos\theta} \implies \tan\theta = \frac{1}{3}.$$

答案为 A.

注 1.12 (经验总结)
解这道题, 要活用 $1 \pm \sin 2\theta = (\sin\theta \pm \cos\theta)^2$.

- 常数 1 可写成 $\sin^2\theta + \cos^2\theta$, 与 $\sin 2\theta$ 搭配后得到完全平方.
- 把 2θ 统一化为单角 θ, 便于与另一边比较.
- 两边约去 $\sin\theta + \cos\theta$ 前, 先说明它不为零; 这里锐角条件正好保证这一点.

例题 1.28 (2023 年 11 月湖北省部分高中联考协作体期中考) 若 $\theta \in (0, \frac{\pi}{2})$, $\dfrac{\sin\theta}{2\cos^2\theta} = \dfrac{\sin\theta - \cos\theta}{1 - \sin 2\theta}$, 则 $\tan\theta =$ _____.

解 等式两边的分母都可化为单角 θ 的形式, 先统一角, 再化简方程. 左边分母 $2\cos^2\theta$ 已是单角形式; 右边分母可配方,

$$1 - \sin 2\theta = \sin^2\theta + \cos^2\theta - 2\sin\theta\cos\theta = (\sin\theta - \cos\theta)^2,$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

代入原方程得

$$\frac{\sin\theta}{2\cos^2\theta}=\frac{\sin\theta-\cos\theta}{(\sin\theta-\cos\theta)^2}.$$

化简右式前须考虑分母 $\sin\theta-\cos\theta$ 是否为零，若 $\sin\theta-\cos\theta=0$，则 $\tan\theta=1$，$\theta=\frac{\pi}{4}$，此时 $1-\sin 2\theta=0$，原方程右边分母为零而无意义，故 $\theta=\frac{\pi}{4}$ 不在取值范围内，$\sin\theta-\cos\theta\neq 0$，可约分得

$$\frac{\sin\theta}{2\cos^2\theta}=\frac{1}{\sin\theta-\cos\theta}.$$

交叉相乘，

$$\sin\theta(\sin\theta-\cos\theta)=2\cos^2\theta,$$

$$\sin^2\theta-\sin\theta\cos\theta-2\cos^2\theta=0,$$

这是关于 $\sin\theta,\cos\theta$ 的二次齐次方程. 由 $\theta\in(0,\frac{\pi}{2})$ 知 $\cos\theta\neq 0$，两边同除以 $\cos^2\theta$，把方程化为关于 $\tan\theta$ 的方程，

$$\tan^2\theta-\tan\theta-2=0,$$

因式分解得 $(\tan\theta-2)(\tan\theta+1)=0$，解得 $\tan\theta=2$ 或 $\tan\theta=-1$. 由 $\theta\in(0,\frac{\pi}{2})$ 知 $\tan\theta>0$，舍去 $\tan\theta=-1$.

答案为 2.

1.10.4 三倍角与 n 倍角公式(拓展)

来龙去脉

和角公式与二倍角公式能处理 $\alpha\pm\beta$ 与 2α. 把 3α 拆成 $2\alpha+\alpha$，就能用同样的办法推出三倍角公式.

三倍角公式无需死记硬背，但它的推导展示出如何把未知角拆成已知角的和，再把结果化归到单一三角函数. 这种递推思路还可以继续推广到任意整数倍角 $n\alpha$.

定理 1.6 (三倍角公式)

$$\sin(3\alpha)=3\sin\alpha-4\sin^3\alpha$$

$$\cos(3\alpha)=4\cos^3\alpha-3\cos\alpha$$

证明 把 3α 写成 $2\alpha+\alpha$，再用和角公式与二倍角公式展开. 先推 $\sin(3\alpha)$，

$$\sin(3\alpha)=\sin(2\alpha+\alpha)$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$
\begin{aligned}
&= \sin(2\alpha)\cos\alpha + \cos(2\alpha)\sin\alpha && \text{(应用和角公式)}\\
&= (2\sin\alpha\cos\alpha)\cos\alpha + (1-2\sin^2\alpha)\sin\alpha \\
&\qquad\qquad \text{(展开 } 2\alpha\text{, 注意 } \cos(2\alpha) \text{ 选用含 sin 的形式)}\\
&= 2\sin\alpha\cos^2\alpha + \sin\alpha - 2\sin^3\alpha && \text{(化简)}\\
&= 2\sin\alpha(1-\sin^2\alpha) + \sin\alpha - 2\sin^3\alpha && \text{(将 } \cos^2\alpha \text{ 换掉, 实现“归一”)}\\
&= 2\sin\alpha - 2\sin^3\alpha + \sin\alpha - 2\sin^3\alpha \\
&= 3\sin\alpha - 4\sin^3\alpha.
\end{aligned}
$$

类似地推 $\cos(3\alpha)$, 把结果只保留 $\cos\alpha$,

$$
\begin{aligned}
\cos(3\alpha) &= \cos(2\alpha+\alpha) \\
&= \cos(2\alpha)\cos\alpha - \sin(2\alpha)\sin\alpha && \text{(应用和角公式)}\\
&= (2\cos^2\alpha - 1)\cos\alpha - (2\sin\alpha\cos\alpha)\sin\alpha \\
&\qquad\qquad \text{(展开 } 2\alpha\text{, 注意 } \cos(2\alpha) \text{ 选用含 cos 的形式)}\\
&= 2\cos^3\alpha - \cos\alpha - 2\cos\alpha\sin^2\alpha && \text{(化简)}\\
&= 2\cos^3\alpha - \cos\alpha - 2\cos\alpha(1-\cos^2\alpha) && \text{(将 } \sin^2\alpha \text{ 换掉, 实现“归一”)}\\
&= 2\cos^3\alpha - \cos\alpha - 2\cos\alpha + 2\cos^3\alpha \\
&= 4\cos^3\alpha - 3\cos\alpha.
\end{aligned}
$$

□

1.10.4.1 推广至 n 倍角

三倍角公式启发我们, $\cos(n\alpha)$ 可以表示为关于 $\cos\alpha$ 的多项式, 并且存在递推关系. 考虑和差角公式.

$$\cos(n\alpha+\alpha) = \cos(n\alpha)\cos\alpha - \sin(n\alpha)\sin\alpha$$

$$\cos(n\alpha-\alpha) = \cos(n\alpha)\cos\alpha + \sin(n\alpha)\sin\alpha$$

将以上两式相加, 可得

$$\cos((n+1)\alpha) + \cos((n-1)\alpha) = 2\cos(n\alpha)\cos\alpha$$

整理后得到递推公式.

$$\cos((n+1)\alpha) = 2\cos\alpha\cdot\cos(n\alpha) - \cos((n-1)\alpha) \tag{1-7}$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

这个公式表明，只要知道相邻两项 $\cos((n-1)\alpha)$ 和 $\cos(n\alpha)$，就可以推出下一项 $\cos((n+1)\alpha)$.

验证一下.

- 令 $n=1$，递推式变为 $\cos(2\alpha)=2\cos\alpha\cdot\cos\alpha-\cos(0\alpha)=2\cos^2\alpha-1$. 与二倍角公式完全吻合.
- 令 $n=2$，递推式变为 $\cos(3\alpha)=2\cos\alpha\cdot\cos(2\alpha)-\cos\alpha$. 将 $\cos(2\alpha)=2\cos^2\alpha-1$ 代入，得

$$\begin{aligned}\cos(3\alpha)&=2\cos\alpha(2\cos^2\alpha-1)-\cos\alpha\\&=4\cos^3\alpha-2\cos\alpha-\cos\alpha\\&=4\cos^3\alpha-3\cos\alpha\end{aligned}$$

知乎@余命数

与我们刚刚推导的三倍角公式完全吻合.

例题 1.29 求值 $\dfrac{3\sin 15^\circ-4\sin^3 15^\circ}{4\cos^3 15^\circ-3\cos 15^\circ}$.

解 分子、分母的结构分别对应正弦、余弦三倍角公式，直接套用即可. 由正弦三倍角公式 $\sin(3\alpha)=3\sin\alpha-4\sin^3\alpha$ 取 $\alpha=15^\circ$，得

$$3\sin 15^\circ-4\sin^3 15^\circ=\sin(3\times 15^\circ)=\sin 45^\circ=\frac{\sqrt{2}}{2};$$

由余弦三倍角公式 $\cos(3\alpha)=4\cos^3\alpha-3\cos\alpha$ 取 $\alpha=15^\circ$，得

$$4\cos^3 15^\circ-3\cos 15^\circ=\cos(3\times 15^\circ)=\cos 45^\circ=\frac{\sqrt{2}}{2}.$$

将分子、分母代回原分式，

$$\text{原式}=\frac{\sin 45^\circ}{\cos 45^\circ}=\tan 45^\circ=1.$$

定理 1.7 (正切的三倍角公式)

$$\tan(3\alpha)=\frac{3\tan\alpha-\tan^3\alpha}{1-3\tan^2\alpha}$$

该公式要求 $\tan\alpha$ 与 $\tan(3\alpha)$ 均有意义，且 $1-3\tan^2\alpha\neq 0$.

证明 推导正切的三倍角公式有两种思路.

知乎关注@余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

其一, 利用商数关系, 得

$$\tan(3\alpha)=\frac{\sin(3\alpha)}{\cos(3\alpha)}=\frac{3\sin\alpha-4\sin^3\alpha}{4\cos^3\alpha-3\cos\alpha},$$

分子分母同除以 $\cos^3\alpha$(要求 $\cos\alpha\neq 0$),

$$\begin{aligned}\tan(3\alpha)&=\frac{3\dfrac{\sin\alpha}{\cos\alpha}\dfrac{1}{\cos^2\alpha}-4\dfrac{\sin^3\alpha}{\cos^3\alpha}}{4-3\dfrac{1}{\cos^2\alpha}}\\&=\frac{3\tan\alpha(1+\tan^2\alpha)-4\tan^3\alpha}{4-3(1+\tan^2\alpha)}\\&=\frac{3\tan\alpha-\tan^3\alpha}{1-3\tan^2\alpha},\end{aligned}$$

上式分母 $1-3\tan^2\alpha$ 非零, 正对应 $\cos(3\alpha)\neq 0$, 即 $\tan(3\alpha)$ 有意义.

其二, 利用和角公式,

$$\begin{aligned}\tan(3\alpha)&=\tan(2\alpha+\alpha)\\&=\frac{\tan(2\alpha)+\tan\alpha}{1-\tan(2\alpha)\tan\alpha}&&\text{(应用和角公式)}\\&=\frac{\frac{2\tan\alpha}{1-\tan^2\alpha}+\tan\alpha}{1-\frac{2\tan\alpha}{1-\tan^2\alpha}\cdot\tan\alpha}&&\text{(展开 }\tan(2\alpha)\text{)}\\&=\frac{\frac{2\tan\alpha+\tan\alpha(1-\tan^2\alpha)}{1-\tan^2\alpha}}{\frac{(1-\tan^2\alpha)-2\tan^2\alpha}{1-\tan^2\alpha}}&&\text{(通分)}\\&=\frac{2\tan\alpha+\tan\alpha-\tan^3\alpha}{1-3\tan^2\alpha}\\&=\frac{3\tan\alpha-\tan^3\alpha}{1-3\tan^2\alpha}.\end{aligned}$$

□

例题 1.30 (2024 年 2 月青桐鸣联考) 已知 $\tan\alpha=\dfrac{\sqrt{2}}{2}$, 求 $\tan 3\alpha$ 的值, $\tan 3\alpha=$ _____.

解 可直接代入正切三倍角公式, 也可分步用 $3\alpha=2\alpha+\alpha$ 计算. 现用分步法.

法一采用分步计算. 先由正切二倍角公式,

$$\tan(2\alpha)=\frac{2\tan\alpha}{1-\tan^2\alpha}=\frac{2\cdot\frac{\sqrt{2}}{2}}{1-\left(\frac{\sqrt{2}}{2}\right)^2}=\frac{\sqrt{2}}{1-\frac{2}{4}}=\frac{\sqrt{2}}{\frac{1}{2}}=2\sqrt{2};$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

再由正切和角公式求 $\tan(3\alpha)=\tan(2\alpha+\alpha)$,

$$\tan(3\alpha)=\frac{\tan(2\alpha)+\tan\alpha}{1-\tan(2\alpha)\tan\alpha}=\frac{2\sqrt{2}+\frac{\sqrt{2}}{2}}{1-(2\sqrt{2})\left(\frac{\sqrt{2}}{2}\right)}=\frac{\frac{5\sqrt{2}}{2}}{1-2}=-\frac{5\sqrt{2}}{2}.$$

法二直接代入三倍角公式. 将 $\tan\alpha=\frac{\sqrt{2}}{2}$ 代入 $\tan(3\alpha)=\dfrac{3\tan\alpha-\tan^3\alpha}{1-3\tan^2\alpha}$,

$$\tan(3\alpha)=\frac{3\left(\frac{\sqrt{2}}{2}\right)-\left(\frac{\sqrt{2}}{2}\right)^3}{1-3\left(\frac{\sqrt{2}}{2}\right)^2}=\frac{\frac{3\sqrt{2}}{2}-\frac{\sqrt{2}}{4}}{1-\frac{3}{2}}=\frac{\frac{5\sqrt{2}}{4}}{-\frac{1}{2}}=-\frac{5\sqrt{2}}{2}.$$

两种方法结果一致, 答案为 $-\dfrac{5\sqrt{2}}{2}$.

1.10.5 高考真题选讲

以下三题均用倍角公式将 $\cos 2x$ 或 $\sin^2 x$ 降幂, 再以换元 $t=\cos x$ 或 $t=\sin x$ 转化为闭区间上二次函数的最值问题.

例题 1.31 (2021 年北京卷第 7 题) 函数 $f(x)=\cos x-\cos 2x$ 的奇偶性及最大值为

A. 奇函数, 最大值为 2　　C. 奇函数, 最大值为 $\frac{9}{8}$

B. 偶函数, 最大值为 2　　D. 偶函数, 最大值为 $\frac{9}{8}$

解 答案: D.

由 $f(-x)=\cos(-x)-\cos(-2x)=\cos x-\cos 2x=f(x)$ 知 $f(x)$ 为偶函数.

用二倍角公式 $\cos 2x=2\cos^2 x-1$ 降幂,

$$f(x)=\cos x-(2\cos^2 x-1)=-2\cos^2 x+\cos x+1.$$

令 $t=\cos x\in[-1,1]$, 则

$$g(t)=-2t^2+t+1=-2\left(t-\frac{1}{4}\right)^2+\frac{9}{8}.$$

对称轴 $t=\dfrac{1}{4}\in[-1,1]$, 故 $f(x)$ 的最大值为 $\dfrac{9}{8}$.

例题 1.32 (2017 年全国 □ 卷理科数学第 14 题) 函数 $f(x)=\sin^2 x+\sqrt{3}\cos x-\frac{3}{4}\left(x\in\left[0,\frac{\pi}{2}\right]\right)$ 的最大值是_____.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 答案：1.

用 $\sin^2 x = 1 - \cos^2 x$ 降幂，

$$f(x) = 1 - \cos^2 x + \sqrt{3}\cos x - \frac{3}{4} = -\cos^2 x + \sqrt{3}\cos x + \frac{1}{4}.$$

令 $t = \cos x$. 由 $x \in \left[0, \frac{\pi}{2}\right]$ 得 $t \in [0, 1]$.

$$g(t) = -t^2 + \sqrt{3}t + \frac{1}{4} = -\left(t - \frac{\sqrt{3}}{2}\right)^2 + 1.$$

对称轴 $t = \frac{\sqrt{3}}{2} \in [0, 1]$，故最大值为 1.

例题 1.33 (2016 年全国 □ 卷文科数学第 11 题) 函数 $f(x) = \cos 2x + 6\cos\left(\frac{\pi}{2} - x\right)$ 的最大值为

A. 4　　C. 6

B. 5　　D. 7

解 答案：B.

由诱导公式 $\cos\left(\frac{\pi}{2} - x\right) = \sin x$ 及二倍角公式 $\cos 2x = 1 - 2\sin^2 x$，

$$f(x) = 1 - 2\sin^2 x + 6\sin x.$$

令 $t = \sin x \in [-1, 1]$，则

$$g(t) = -2t^2 + 6t + 1 = -2\left(t - \frac{3}{2}\right)^2 + \frac{11}{2}.$$

对称轴 $t = \frac{3}{2} \notin [-1, 1]$，抛物线开口向下，故 $g(t)$ 在 $[-1, 1]$ 上单调递增，最大值在 $t = 1$ 处取得：$g(1) = -2 + 6 + 1 = 5$.

1.11 万能公式与齐次式思想

来龙去脉

在处理三角函数问题时，同一个方程或表达式里常同时出现 $\sin\alpha, \cos\alpha, \tan\alpha$ 等不同三角函数. 化简这类问题的首要步骤，是把它们统一成同类函数.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

已用弦化切处理过**齐次式**. 把这个思路再推进一步. 若能用一个单一的代数变量把所有三角函数都表示出来, 那么任何三角方程都能化为纯代数方程, 从而避开三角函数特有的变形.

万能公式说明这样的表示是可行的. 它以二倍角公式和 $1=\sin^2\frac{\alpha}{2}+\cos^2\frac{\alpha}{2}$ 为根据, 把 $\sin\alpha,\cos\alpha,\tan\alpha$ 都写成半角正切 $t=\tan\frac{\alpha}{2}$ 的有理分式, 完成了三角函数名之间的统一.

知乎@余命数

之所以冠以“万能”之名, 在于它的适用范围不限于某一类三角方程. 只要一个表达式由 $\sin\alpha,\cos\alpha,\tan\alpha$ 经四则运算复合而成, 代入万能公式后所得的必然是关于 t 的有理分式. 有理分式的化简与求值只依赖代数运算, 不再涉及任何三角恒等变形. 一个代换把整类问题归约为代数问题, “万能”二字正由此而来.

这一代换最早服务于积分学. 对于形如 $\int R(\sin x,\cos x)\,dx$ 的有理三角积分, 令 $t=\tan\frac{x}{2}$ 可将被积函数连同 dx 一并化为 t 的有理函数, 从而把超越的积分问题转化为有理函数的积分. 这一技巧以德国数学家魏尔斯特拉斯的名字命名, 称为**魏尔斯特拉斯代换**, 其积分应用见 第四卷第 17.3 节.

知乎@余命数

1.11.1 公式体系

知乎@余命数

定理 1.8 (万能公式) 当 $\alpha\neq(2k+1)\pi$ $(k\in\mathbb{Z})$ 时, 令 $t=\tan\frac{\alpha}{2}$, 则有

$$\sin\alpha=\frac{2t}{1+t^2} \tag{1-8}$$

$$\cos\alpha=\frac{1-t^2}{1+t^2} \tag{1-9}$$

$$\tan\alpha=\frac{2t}{1-t^2}\quad(\text{当}t\neq\pm1\text{ 时}) \tag{1-10}$$

知乎@余命数

1.11.2 齐次化构造

证明 (推导) 先推 $\sin\alpha$. 从二倍角公式出发, 利用 $1=\sin^2\frac{\alpha}{2}+\cos^2\frac{\alpha}{2}$ 把分子改造成齐次分式,

$$\begin{aligned}\sin\alpha&=\sin\left(2\cdot\frac{\alpha}{2}\right)=2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}\\&=\frac{2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}}{1}&&\text{(除以 1)}\\&=\frac{2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}}{\sin^2\frac{\alpha}{2}+\cos^2\frac{\alpha}{2}}.&&\text{(1 的代换, 构造齐次分式)}\end{aligned}$$

知乎@余命数

知乎@余命数

此时分子分母都是关于 $\sin\frac{\alpha}{2},\cos\frac{\alpha}{2}$ 的二次齐次式, 同除以 $\cos^2\frac{\alpha}{2}$(设 $\cos\frac{\alpha}{2}\neq 0$),

$$\sin\alpha=\frac{\frac{2\sin(\alpha/2)\cos(\alpha/2)}{\cos^2(\alpha/2)}}{\frac{\sin^2(\alpha/2)+\cos^2(\alpha/2)}{\cos^2(\alpha/2)}}=\frac{2\frac{\sin(\alpha/2)}{\cos(\alpha/2)}}{\frac{\sin^2(\alpha/2)}{\cos^2(\alpha/2)}+1}=\frac{2\tan\frac{\alpha}{2}}{1+\tan^2\frac{\alpha}{2}}=\frac{2t}{1+t^2}.$$

再推 $\cos\alpha$, 采用相同策略. 由二倍角公式及 1 的代换,

$$\begin{aligned}\cos\alpha&=\cos\left(2\cdot\frac{\alpha}{2}\right)=\cos^2\frac{\alpha}{2}-\sin^2\frac{\alpha}{2}\\&=\frac{\cos^2\frac{\alpha}{2}-\sin^2\frac{\alpha}{2}}{1}=\frac{\cos^2\frac{\alpha}{2}-\sin^2\frac{\alpha}{2}}{\sin^2\frac{\alpha}{2}+\cos^2\frac{\alpha}{2}},\end{aligned}$$

分子分母同除以 $\cos^2\frac{\alpha}{2}$, 得

$$\cos\alpha=\frac{1-\frac{\sin^2(\alpha/2)}{\cos^2(\alpha/2)}}{1+\frac{\sin^2(\alpha/2)}{\cos^2(\alpha/2)}}=\frac{1-\tan^2\frac{\alpha}{2}}{1+\tan^2\frac{\alpha}{2}}=\frac{1-t^2}{1+t^2}.$$

最后推 $\tan\alpha$. 当 $\cos\alpha\neq 0$ 时, 由商数关系及刚得的两式,

$$\tan\alpha=\frac{\sin\alpha}{\cos\alpha}=\frac{\frac{2t}{1+t^2}}{\frac{1-t^2}{1+t^2}}=\frac{2t}{1-t^2},$$

此处要求 $1-t^2\neq 0$, 这正对应 $\cos\alpha\neq 0$. □

万能公式提供了一条纯代数的解题路径. 它的计算量通常较大, 且要注意 $\alpha=(2k+1)\pi$ 时 $\tan\frac{\alpha}{2}$ 无定义. 因此它在其他方法, 如辅助角公式不易见效时, 是一种可靠但繁琐的备选.

1.11.3 典例解析

例题 1.34 (解三角方程) 求解方程 $\sin\theta+\cos\theta=1$.

解 此题可用多种方法求解, 下用万能公式法演示. 令 $t=\tan\frac{\theta}{2}$, 将万能公式代入原方程,

$$\frac{2t}{1+t^2}+\frac{1-t^2}{1+t^2}=1.$$

因分母 $1+t^2>0$ 恒成立, 可直接去分母,

$$2t+1-t^2=1+t^2\implies 2t^2-2t=0\implies 2t(t-1)=0,$$

解得 $t=0$ 或 $t=1$. 由 $t=\tan\frac{\theta}{2}$ 反解: $t=0$ 给出 $\frac{\theta}{2}=k\pi$, 即 $\theta=2k\pi$; $t=1$ 给出 $\frac{\theta}{2}=k\pi+\frac{\pi}{4}$, 即 $\theta=2k\pi+\frac{\pi}{2}$ $(k\in\mathbb{Z})$.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

万能公式的使用前提是 $\cos\frac{\theta}{2}\neq 0$,即 $\theta\neq(2k+1)\pi$,故须单独检验 $\theta=(2k+1)\pi$,当 $\theta=\pi$ 时 $\sin\pi+\cos\pi=0-1=-1\neq 1$,这些点不是解.

原方程的解集为 $\{\theta\mid\theta=2k\pi$ 或$\theta=2k\pi+\frac{\pi}{2},k\in\mathbb{Z}\}$.

例题 1.35 (2024 年 1 月浙江省宁波九校联考) 已知 $\tan A=\dfrac{1}{2}$,求

$$\frac{\cos 4A-4\cos 2A+3}{\cos 4A+4\cos 2A+3}$$

的值(　)

A. $\dfrac{1}{16}$

B. $\dfrac{1}{8}$

C. $\dfrac{1}{4}$

D. $\dfrac{1}{2}$

知乎@余命数

解 题目给出单角的正切值,要求一个含其二倍角和四倍角的复杂分式的值,宜用由切求弦的思路,把所有三角函数都用 $\tan A$ 表示.

由万能公式的余弦表达式直接求 $\cos 2A$,

$$\cos 2A=\frac{1-\tan^2 A}{1+\tan^2 A}=\frac{1-(\frac{1}{2})^2}{1+(\frac{1}{2})^2}=\frac{\frac{3}{4}}{\frac{5}{4}}=\frac{3}{5}.$$

再把 $2A$ 视为整体,用余弦二倍角公式 $\cos(2\theta)=2\cos^2\theta-1$ 求 $\cos 4A$,

$$\cos 4A=\cos(2\cdot 2A)=2\cos^2(2A)-1=2\left(\frac{3}{5}\right)^2-1=2\left(\frac{9}{25}\right)-1=\frac{18}{25}-1=-\frac{7}{25}.$$

将 $\cos 2A=\frac{3}{5}$ 与 $\cos 4A=-\frac{7}{25}$ 代入原分式,

$$\begin{aligned}
\text{原式}&=\frac{(-\frac{7}{25})-4(\frac{3}{5})+3}{(-\frac{7}{25})+4(\frac{3}{5})+3}=\frac{-\frac{7}{25}-\frac{12}{5}+3}{-\frac{7}{25}+\frac{12}{5}+3},\\
\text{分子}&=\frac{-7-60+75}{25}=\frac{8}{25},\\
\text{分母}&=\frac{-7+60+75}{25}=\frac{128}{25},\\
\text{原式}&=\frac{8/25}{128/25}=\frac{8}{128}=\frac{1}{16}.
\end{aligned}$$

知乎@余命数

另法：令 $x=\cos 2A$，将 $\cos 4A=2x^2-1$ 代入，分子为 $2x^2-1-4x+3=2(x-1)^2$，分母为 $2x^2+4x+2=2(x+1)^2$，故原式 $=\frac{2(\cos 2A-1)^2}{2(\cos 2A+1)^2}=\left(\frac{\cos 2A-1}{\cos 2A+1}\right)^2$；代入 $\cos 2A=3/5$ 得 $\left(\frac{-2/5}{8/5}\right)^2=\frac{1}{16}$.

两种方法都得 $\frac{1}{16}$，故选 A.

例题 1.36 (2023 年 12 月福建百校 & 河北金科联考) 若 $\tan\left(\alpha+\dfrac{\pi}{4}\right)=-3$，$\tan\beta=3$，求

$$\frac{\cos(\alpha-\beta)}{\sin(\alpha+\beta)}$$

的值(　)

A. -1

B. $\dfrac{7}{5}$

C. $\dfrac{3}{5}$

D. $\dfrac{4}{5}$

知乎@余命数

解 题目给出两个角的正切相关值，要求解一个复杂的三角分式. 直接展开会很繁琐，该分式的结构正适合弦化切.

先求 $\tan\alpha$. 由 $\tan(\alpha+\frac{\pi}{4})=-3$ 及和角公式，

$$\frac{\tan\alpha+1}{1-\tan\alpha}=-3,$$

解得 $\tan\alpha+1=-3(1-\tan\alpha)=-3+3\tan\alpha$，即 $2\tan\alpha=4$，故 $\tan\alpha=2$.

再对目标分式弦化切. 将其分子分母分别用和差角公式展开，

$$\text{原式}=\frac{\cos\alpha\cos\beta+\sin\alpha\sin\beta}{\sin\alpha\cos\beta+\cos\alpha\sin\beta},$$

这是关于 $\sin,\cos$ 的伪齐次式，分子分母同除以 $\cos\alpha\cos\beta$（设其不为 0），

$$\begin{aligned}\text{原式}&=\frac{\frac{\cos\alpha\cos\beta+\sin\alpha\sin\beta}{\cos\alpha\cos\beta}}{\frac{\sin\alpha\cos\beta+\cos\alpha\sin\beta}{\cos\alpha\cos\beta}}\\&=\frac{1+\frac{\sin\alpha}{\cos\alpha}\frac{\sin\beta}{\cos\beta}}{\frac{\sin\alpha}{\cos\alpha}+\frac{\sin\beta}{\cos\beta}}=\frac{1+\tan\alpha\tan\beta}{\tan\alpha+\tan\beta}.\end{aligned}$$

代入 $\tan\alpha=2$ 与 $\tan\beta=3$，

$$\text{原式}=\frac{1+(2)(3)}{2+3}=\frac{1+6}{5}=\frac{7}{5}.$$

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

正确答案为 B.

例题 1.37 (2023 年 11 月江淮十校第二次联考) 已知角 θ 为第二象限角, 且满足 $\sin(\theta-\frac{\pi}{3})\sin(\pi+\theta)=\cos 2\theta$, 则 $\tan\theta=$(　　)

A. $\dfrac{\sqrt{3}-\sqrt{11}}{2}$

B. $\dfrac{\sqrt{3}-\sqrt{7}}{2}$

C. $\dfrac{-\sqrt{3}-\sqrt{11}}{2}$

D. $\dfrac{-\sqrt{3}-\sqrt{7}}{2}$

解 这个恒等式把所有三角函数都化为单角 θ 的正弦和余弦后, 即可整理成齐次方程, 再弦化切求解.

用和差角公式、诱导公式与倍角公式把方程各项展开,

$$\sin\left(\theta-\frac{\pi}{3}\right)=\sin\theta\cos\frac{\pi}{3}-\cos\theta\sin\frac{\pi}{3}=\frac{1}{2}\sin\theta-\frac{\sqrt{3}}{2}\cos\theta,$$

$$\sin(\pi+\theta)=-\sin\theta,$$

$$\cos(2\theta)=\cos^2\theta-\sin^2\theta,$$

代入原方程,

$$\left(\frac{1}{2}\sin\theta-\frac{\sqrt{3}}{2}\cos\theta\right)(-\sin\theta)=\cos^2\theta-\sin^2\theta.$$

展开并化简, 两边同乘 2 后移项至同一侧, 得到关于 $\sin\theta,\cos\theta$ 的二次齐次方程,

$$-\frac{1}{2}\sin^2\theta+\frac{\sqrt{3}}{2}\sin\theta\cos\theta=\cos^2\theta-\sin^2\theta,$$

$$-\sin^2\theta+\sqrt{3}\sin\theta\cos\theta=2\cos^2\theta-2\sin^2\theta\quad(\text{两边同乘 }2),$$

$$\sin^2\theta+\sqrt{3}\sin\theta\cos\theta-2\cos^2\theta=0\quad(\text{移项合并}).$$

因为 θ 是第二象限角, $\cos\theta\neq 0$, 两边同除以 $\cos^2\theta$ 弦化切, 化为关于 $\tan\theta$ 的一元二次方程, 再用求根公式,

$$\tan^2\theta+\sqrt{3}\tan\theta-2=0,$$

$$\tan\theta=\frac{-\sqrt{3}\pm\sqrt{(\sqrt{3})^2-4(1)(-2)}}{2(1)}=\frac{-\sqrt{3}\pm\sqrt{3+8}}{2}=\frac{-\sqrt{3}\pm\sqrt{11}}{2}.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

θ 为第二象限角, 其正切必为负: $\frac{-\sqrt{3}+\sqrt{11}}{2}$ 因 $\sqrt{11}>\sqrt{3}$ 而为正, 舍去; $\frac{-\sqrt{3}-\sqrt{11}}{2}$ 为负, 符合题意.

故正确答案为 C.

例题 1.38 (2024 届长郡中学月考) 已知 $\alpha\in(0,\frac{\pi}{2})$, 且 $3\sqrt{2}\sin(\alpha-\frac{\pi}{4})=\cos\alpha$, 则 $\sin\alpha+2\cos\alpha=$(　)

A. -2

B. $\dfrac{25}{16}$

C. $-\dfrac{25}{16}$

D. 2

解 先由方程解出 α 的某个三角函数值,再代入目标式. 展开 $3\sqrt{2}\sin(\alpha-\frac{\pi}{4})=\cos\alpha$:

$$\begin{aligned}3\sqrt{2}\Big(\sin\alpha\cdot\frac{\sqrt{2}}{2}-\cos\alpha\cdot\frac{\sqrt{2}}{2}\Big)&=\cos\alpha,\\ 3(\sin\alpha-\cos\alpha)&=\cos\alpha,\\ 3\sin\alpha&=4\cos\alpha.\end{aligned}$$

因 $\alpha\in(0,\frac{\pi}{2})$, $\cos\alpha\neq0$, 故 $\tan\alpha=\frac{4}{3}$.

法一采用求弦代入. 由 $\tan\alpha=\frac{4}{3}$ 与 α 锐角, 构造边长 $3,4,5$ 的直角三角形, 得 $\sin\alpha=\frac{4}{5}$, $\cos\alpha=\frac{3}{5}$, 于是

$$\sin\alpha+2\cos\alpha=\frac{4}{5}+\frac{6}{5}=2.$$

法二采用 tan 表达弦. 在 $\alpha\in(0,\frac{\pi}{2})$ 下,

$$\sin\alpha=\frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}}=\frac{4/3}{5/3}=\frac{4}{5},\qquad\cos\alpha=\frac{1}{\sqrt{1+\tan^2\alpha}}=\frac{3}{5},$$

正确答案为 D.

1.11.4 基本思路

上面的例题反复用到以下四条思路.

- **思想一**(配凑与拆解)
  已知角与目标角不一致时, 需要构造它们之间的联系.
  - **拼凑角**, 如已知 $\tan(\alpha-\beta)$ 与 $\tan\beta$ 求 $\tan\alpha$, 可用 $\alpha=(\alpha-\beta)+\beta$.
  - **拆分角**, 如推导三倍角公式时, 把 3α 写成 $2\alpha+\alpha$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **互补与互余**, 在三角形中, 利用 $A+B+C=\pi$ 构造 $\pi-C$ 或 $\frac{\pi}{2}-\frac{C}{2}$ 等形式.

角的变换最终都是把未知角用已知角表示, 从而便于套用公式.

- **思想**二(识别公式结构)
观察代数式的结构, 看它接近哪个公式.
  - **公式识别**, 由 $(m+1)(n+1)=2$ 得 $\frac{m+n}{1-mn}=1$, 可联想到 $\tan(A+B)$ 的结构; $1\pm\sin 2\theta$ 可联想到 $(\sin\theta\pm\cos\theta)^2$.
  - **齐次式构造**, 分式可考虑弦化切; $\sin\alpha+2\cos\alpha$ 等非齐次式也可除以 $1=\sin^2\alpha+\cos^2\alpha$ 化为齐次式.

这种思路把复杂式子变形成熟悉的公式模型.

- **思想**三(整体代换)
当局部变量, 如 α,β 难以直接求解时, 可把某些组合视为整体.
  - **整体代换**, 例如已知 $\cos(\beta-\alpha)$ 与 $\tan\alpha\tan\beta$, 可把 $\cos\alpha\cos\beta$ 与 $\sin\alpha\sin\beta$ 设为整体未知数 X,Y 处理.

关注关系本身, 能简化复杂问题.

- **思想四**(归一)
式子里出现多种三角函数时, 常把它们化为同一类函数, 例如正切.
  - **齐次方程化正切方程**, 二次齐次方程 $\sin^2\theta+A\sin\theta\cos\theta+B\cos^2\theta=0$ 可化为关于 $\tan\theta$ 的代数方程.
  - **万能公式化有理分式**, 万能公式把各三角函数写成 $t=\tan(\frac{\alpha}{2})$ 的有理分式.

变量归一后, 代数运算和化简更容易进行.

注 1.13 (三角恒等变换的做法)

1. **分析条件**
拿到题目后, 先分析已知条件和求解目标, 重点关注两个方面.
   - **分析角**, 已知角与目标角之间是否存在和、差、倍、半、互补、互余等关系?
   - **分析结构**, 表达式是齐次式吗? 是否可以因式分解? 是否与某个公式的结构相似? 函数名是否统一? 幂次是否一致?
2. **变形与转化**
根据对条件与结构的分析, 选择最合适的思想方法变形.
   - 若**角不统一**, 优先使用**角的变换**思想, 进行拆角或凑角.
   - 若**名不统一**, 优先使用**统一思想**, 进行弦化切或应用同角关系.
   - 若**幂不统一**, 优先使用**降幂公式**或**升幂公式**(二倍角余弦).
   - 若**结构特殊**, 应用**结构变形**思想, 直接套用模型或进行因式分解.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 若问题**极其复杂**, 考虑**整体思想**或启用**万能公式**作为最后手段.

3. **求解与检验**
经过变形后, 问题通常会化简为一个可以直接求解的代数问题或简单的三角函数求值问题.
   - **求解**. 进行最后的代数运算, 得出结果.
   - **检验**. 将结果带回思考, 是否符合题目给出的角的范围, 如象限、锐角等? 在变形过程中, 是否有除以一个可能为零的式子? 逻辑上是否存在漏洞?

1.11.5 高考真题选讲

例题 1.39 (2021 年全国 □ 卷数学第 6 题) 若 $\tan\theta=-2$, 则 $\dfrac{\sin\theta(1+\sin 2\theta)}{\sin\theta+\cos\theta}=$

A. $-\frac{6}{5}$　　C. $\frac{2}{5}$

B. $-\frac{2}{5}$　　D. $\frac{6}{5}$

解 答案:C. 将 $1+\sin 2\theta=\sin^2\theta+\cos^2\theta+2\sin\theta\cos\theta=(\sin\theta+\cos\theta)^2$ 代入,

$$\frac{\sin\theta(1+\sin 2\theta)}{\sin\theta+\cos\theta}=\frac{\sin\theta(\sin\theta+\cos\theta)^2}{\sin\theta+\cos\theta}=\sin\theta(\sin\theta+\cos\theta)=\sin^2\theta+\sin\theta\cos\theta.$$

再除以 $\sin^2\theta+\cos^2\theta=1$ 弦化切,

$$\sin^2\theta+\sin\theta\cos\theta=\frac{\tan^2\theta+\tan\theta}{1+\tan^2\theta}=\frac{4-2}{1+4}=\frac{2}{5}.$$

1.12 角的变换

来龙去脉

我们在解三角变换题时,常把已知角与目标角写成不同形式. 例如已知 $\sin(\theta-\frac{\pi}{3})$, 却要求 $\cos(2\theta+\frac{\pi}{3})$, 它们乍看之下没有明显联系.

我们的目标是把目标角用已知角表示出来. 通过加减、倍半等代数运算, 把它向已知角的方向改写, 就能直接套用和、差、倍、半角公式.

换元法把这种关系写得更清楚. 令 $x=\theta-\frac{\pi}{3}$, 则复杂的角表达式改写为关于 x 的简单式子, 倍角、半角的关系也变得更明显.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.12.1 典例剖析

例题 1.40 (2024 年 1 月浙江省嘉兴市高三上期末) 已知 $\sin(\theta-\frac{\pi}{3})=\frac{\sqrt{5}}{5}$, 则 $\cos(2\theta+\frac{\pi}{3})=$ _____.

解 我们先识别已知角与目标角. 已知角是 $\theta-\frac{\pi}{3}$, 目标角是 $2\theta+\frac{\pi}{3}$. 令 $x=\theta-\frac{\pi}{3}$, 就能把两者关系显化.

换元简化令 $x=\theta-\frac{\pi}{3}$, 则已知条件变为 $\sin x=\frac{\sqrt{5}}{5}$. 再把目标角用 x 表示. 由 $x=\theta-\frac{\pi}{3}$ 得 $\theta=x+\frac{\pi}{3}$.

$$\begin{aligned}2\theta+\frac{\pi}{3}&=2\left(x+\frac{\pi}{3}\right)+\frac{\pi}{3}\\&=2x+\frac{2\pi}{3}+\frac{\pi}{3}\\&=2x+\pi\end{aligned}$$

转化问题换元后, 原问题等价于 已知 $\sin x=\frac{\sqrt{5}}{5}$, 求 $\cos(2x+\pi)$ 的值.

现在求解新问题. 先用诱导公式, 再用二倍角公式.

$$\cos(2x+\pi)=-\cos(2x)$$

已知 $\sin x$, 我们选用只含 $\sin x$ 的余弦二倍角公式.

$$-\cos(2x)=-(1-2\sin^2 x)=2\sin^2 x-1$$

代入已知值 $\sin x=\frac{\sqrt{5}}{5}$:

$$2\left(\frac{\sqrt{5}}{5}\right)^2-1=2\left(\frac{5}{25}\right)-1=\frac{10}{25}-1=\frac{2}{5}-1=-\frac{3}{5}$$

故最终答案为 $-\frac{3}{5}$.

例题 1.41 (2023 年 11 月 THUSSAT 诊断性测试) 若 $\frac{9\pi}{16}<\alpha<\frac{17\pi}{16}$, 且满足 $\cos\left(2\alpha-\frac{\pi}{8}\right)=-\frac{1}{4}$, 则 $\sin\left(\frac{9\pi}{16}-\alpha\right)=$(　　)

A. $-\frac{\sqrt{6}}{4}$

B. $\frac{\sqrt{6}}{4}$

C. $-\frac{\sqrt{10}}{4}$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

D. $\dfrac{\sqrt{10}}{4}$

解 这道题的角关系更隐蔽. 我们仍用换元法揭示它, 再根据角的范围判断三角函数值的符号.

换元简化令 $x=2\alpha-\frac{\pi}{8}$. 已知条件变为 $\cos x=-\frac{1}{4}$. 再把目标角 $\frac{9\pi}{16}-\alpha$ 用 x 表示. 由 $x=2\alpha-\frac{\pi}{8}$, 可得 $2\alpha=x+\frac{\pi}{8}\implies\alpha=\frac{x}{2}+\frac{\pi}{16}$. 把它代入目标角, 得

$$\frac{9\pi}{16}-\alpha=\frac{9\pi}{16}-\left(\frac{x}{2}+\frac{\pi}{16}\right)=\frac{8\pi}{16}-\frac{x}{2}=\frac{\pi}{2}-\frac{x}{2}$$

转化问题于是原问题等价于下面这个问题. 已知 $\cos x=-\frac{1}{4}$, 求 $\sin(\frac{\pi}{2}-\frac{x}{2})$ 的值.

下面用诱导公式与半角关系求解. 根据诱导公式, $\sin(\frac{\pi}{2}-\frac{x}{2})=\cos(\frac{x}{2})$. 问题进一步化为已知 $\cos x$ 求 $\cos\frac{x}{2}$, 即半角问题. 由二倍角公式 $\cos x=2\cos^2\frac{x}{2}-1$, 解得 $\cos^2\frac{x}{2}$.

$$2\cos^2\frac{x}{2}=1+\cos x=1+\left(-\frac{1}{4}\right)=\frac{3}{4}$$

$$\cos^2\frac{x}{2}=\frac{3}{8}\implies\cos\frac{x}{2}=\pm\sqrt{\frac{3}{8}}=\pm\frac{\sqrt{6}}{4}$$

接下来由范围确定符号. 符号由角 $\frac{x}{2}$ 所在象限决定. 先根据 α 的范围求出 x 的范围, 再得到 $\frac{x}{2}$ 的范围. 已知 $\frac{9\pi}{16}<\alpha<\frac{17\pi}{16}$. 两边同乘以 2, 得 $\frac{9\pi}{8}<2\alpha<\frac{17\pi}{8}$. 再同减去 $\frac{\pi}{8}$, 得 $\frac{9\pi}{8}-\frac{\pi}{8}<2\alpha-\frac{\pi}{8}<\frac{17\pi}{8}-\frac{\pi}{8}$. 即 $\pi<x<2\pi$. 两边同除以 2, 得 $\frac{\pi}{2}<\frac{x}{2}<\pi$. 所以, 角 $\frac{x}{2}$ 在第二象限. 在第二象限, 余弦值为负, 因此 $\cos\frac{x}{2}$ 取负值.

$$\cos\frac{x}{2}=-\frac{\sqrt{6}}{4}$$

所以 $\sin(\frac{9\pi}{16}-\alpha)=-\frac{\sqrt{6}}{4}$. 正确答案为 A.

1.12.2 四大思想

这四条思路在上面的例题中反复出现.

- **思想一配凑与拆解**
  当已知角与目标角不一致时, 我们要构造它们之间的联系.
  - **拼凑角**. 已知 $\tan(\alpha-\beta)$ 与 $\tan\beta$ 求 $\tan\alpha$, 可用 $\alpha=(\alpha-\beta)+\beta$.
  - **拆分角**. 推导三倍角公式时, 把 3α 写成 $2\alpha+\alpha$.
  - **换元凑角**. 令 $x=\theta-\frac{\pi}{3}$, 则目标角 $2\theta+\frac{\pi}{3}=2x+\pi$.

  角的变换最终都是把未知角用已知角表示, 从而直接应用公式.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **思想二识别公式结构**
  观察代数式的结构, 看它接近哪个公式.
  - **公式识别**. 由 $(m+1)(n+1)=2$ 得 $\frac{m+n}{1-mn}=1$, 可联想到 $\tan(A+B)$ 的结构; $1\pm\sin 2\theta$ 可联想到 $(\sin\theta\pm\cos\theta)^2$.
  - **齐次式构造**. 分式可考虑弦化切; $\sin\alpha+2\cos\alpha$ 等非齐次式也可除以 $1=\sin^2\alpha+\cos^2\alpha$ 化为齐次式.

  它把复杂式子变形成常见公式模型.
- **思想三整体代换**
  当局部变量(如 α,β)难以直接求解时, 可把某些组合视为整体.
  - **整体代换**. 例如已知 $\cos(\beta-\alpha)$ 与 $\tan\alpha\tan\beta$, 可把 $\cos\alpha\cos\beta$ 与 $\sin\alpha\sin\beta$ 设为整体未知数 X,Y 处理.

  这样注意力放在关系本身, 从而简化复杂问题.
- **思想四归一**
  式子里出现多种三角函数时, 常把它们化为同一类函数, 例如正切.
  - **齐次方程化正切方程**. 二次齐次方程 $\sin^2\theta+A\sin\theta\cos\theta+B\cos^2\theta=0$ 可化为关于 $\tan\theta$ 的代数方程.
  - **万能公式化有理分式**. 万能公式把各三角函数写成 $t=\tan(\frac{\alpha}{2})$ 的有理分式.

  变量归一后, 代数运算和化简变得更直接.

1.12.3 高考真题选讲

例题 1.42 (2012 年江苏卷第 11 题) 设 α 为锐角, 若 $\cos\left(\alpha+\frac{\pi}{6}\right)=\frac{4}{5}$, 则 $\sin\left(2\alpha+\frac{\pi}{12}\right)=$ _____.

解 令 $\beta=\alpha+\frac{\pi}{6}$, 则 $2\alpha+\frac{\pi}{12}=2\left(\alpha+\frac{\pi}{6}\right)-\frac{\pi}{4}=2\beta-\frac{\pi}{4}$, 于是

$$\sin\left(2\alpha+\frac{\pi}{12}\right)=\sin\left(2\beta-\frac{\pi}{4}\right)=\frac{\sqrt{2}}{2}(\sin 2\beta-\cos 2\beta).$$

由 α 为锐角知 $\beta=\alpha+\frac{\pi}{6}\in\left(\frac{\pi}{6},\frac{2\pi}{3}\right)$, 又 $\cos\beta=\frac{4}{5}$, 故 $\sin\beta=\frac{3}{5}$. 由此

$$\sin 2\beta=2\cdot\frac{3}{5}\cdot\frac{4}{5}=\frac{24}{25},\quad \cos 2\beta=2\cdot\frac{16}{25}-1=\frac{7}{25}.$$

代回原式,

$$\sin\left(2\alpha+\frac{\pi}{12}\right)=\frac{\sqrt{2}}{2}\left(\frac{24}{25}-\frac{7}{25}\right)=\frac{\sqrt{2}}{2}\cdot\frac{17}{25}=\frac{17\sqrt{2}}{50}.$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.13 和差化积与积化和差

至此,我们已会处理角的和差,例如 $\sin(\alpha+\beta)$. 更复杂的问题中还会出现三角函数的和差,例如 $\sin\alpha+\sin\beta$. 这两者形式不同,但可以通过公式互相转化.

积化和差与和差化积是和差角公式的变形与组合. 它们互为逆运算.

- **积化和差**. 将两个三角函数的乘积化为和差. 它常用于降幂和化简,把乘积项变成可以逐项处理的和差项.
- **和差化积**. 将两个三角函数的和差化为乘积. 它常用于分式约分和解三角方程,把 $f(x)+g(x)=0$ 化为 $A(x)B(x)=0$ 的形式.

这两套公式是三角恒等变换中处理复杂变形的常用工具.

1.13.1 公式

1.13.1.1 积化和差

我们回顾正弦与余弦的和差角公式.

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta \quad \cdots ①$$

$$\sin(\alpha-\beta)=\sin\alpha\cos\beta-\cos\alpha\sin\beta \quad \cdots ②$$

$$\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta \quad \cdots ③$$

$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta \quad \cdots ④$$

通过将以上四式两两相加或相减,即可得到积化和差公式. 例如,$(①+②)\div 2$ 可得

$$\sin\alpha\cos\beta=\frac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)]$$

定理 1.9 (积化和差公式)

$$\sin\alpha\cos\beta=\frac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)]$$

$$\cos\alpha\sin\beta=\frac{1}{2}[\sin(\alpha+\beta)-\sin(\alpha-\beta)]$$

$$\cos\alpha\cos\beta=\frac{1}{2}[\cos(\alpha+\beta)+\cos(\alpha-\beta)]$$

$$\sin\alpha\sin\beta=-\frac{1}{2}[\cos(\alpha+\beta)-\cos(\alpha-\beta)]$$

知乎@余命数

证明 由和差角公式

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta,$$
$$\sin(\alpha-\beta)=\sin\alpha\cos\beta-\cos\alpha\sin\beta,$$
$$\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta,$$
$$\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta.$$

将前两式相加、相减, 分别得到前两个积化和差公式; 将后两式相加、相减, 分别得到后两个公式. □

知乎@余命数

1.13.1.2 和差化积公式

在积化和差公式的基础上, 我们进行一次巧妙的换元. 令 $\theta=\alpha+\beta, \phi=\alpha-\beta$, 则 $\alpha=\frac{\theta+\phi}{2}$, $\beta=\frac{\theta-\phi}{2}$. 将此代入第一个积化和差公式 $\sin\alpha\cos\beta=\frac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)]$ 中, 反解可得

$$\sin\theta+\sin\phi=2\sin\frac{\theta+\phi}{2}\cos\frac{\theta-\phi}{2}$$

定理 1.10 (和差化积公式)

$$\sin\theta+\sin\phi=2\sin\frac{\theta+\phi}{2}\cos\frac{\theta-\phi}{2}$$
$$\sin\theta-\sin\phi=2\cos\frac{\theta+\phi}{2}\sin\frac{\theta-\phi}{2}$$
$$\cos\theta+\cos\phi=2\cos\frac{\theta+\phi}{2}\cos\frac{\theta-\phi}{2}$$
$$\cos\theta-\cos\phi=-2\sin\frac{\theta+\phi}{2}\sin\frac{\theta-\phi}{2}$$

证明 在积化和差公式中令

$$\alpha=\frac{\theta+\phi}{2},\qquad \beta=\frac{\theta-\phi}{2}.$$

则 $\alpha+\beta=\theta, \alpha-\beta=\phi$. 将这一换元分别代入四个积化和差公式并整理, 即得四个和差化积公式. □

这四个公式的结构可统一如下. 两角正弦之和, 右端正弦在前; 两角正弦之差, 右端余弦在前; 两角余弦之和, 右端余弦在前; 两角余弦之差, 右端正弦在前且带负号. 各式右端的角度永远是和的一半、差的一半.

知乎@余命数

1.13.2 典例剖析

例题 1.43 (和差化积的应用) 化简分式 $\dfrac{\sin 5x+\sin 3x}{\cos 5x+\cos 3x}$.

解 本题的分子、分母都是三角函数的和. 我们先用和差化积公式化简, 再约分.

对分子应用和差化积

$$\sin 5x+\sin 3x=2\sin\frac{5x+3x}{2}\cos\frac{5x-3x}{2}=2\sin 4x\cos x$$

知乎@余命数

对分母应用和差化积

$$\cos 5x+\cos 3x=2\cos\frac{5x+3x}{2}\cos\frac{5x-3x}{2}=2\cos 4x\cos x$$

知乎@余命数

整合与约分将化简后的结果代回原分式.

$$\text{原式}=\frac{2\sin 4x\cos x}{2\cos 4x\cos x}$$

假设 $\cos 4x\neq 0$ 且 $\cos x\neq 0$, 可约去公因式 $2\cos x$.

知乎@余命数

$$\text{原式}=\frac{\sin 4x}{\cos 4x}=\tan 4x$$

例题 1.44 (积化和差的应用) 求函数 $f(x)=\sin(x+\frac{\pi}{6})\cos(x-\frac{\pi}{3})$ 的最大值.

解 本题是两个三角函数的乘积, 直接求最值较繁琐. 我们把它用积化和差化为标准正弦函数.

知乎@余命数

应用积化和差公式令 $\alpha=x+\frac{\pi}{6}$, $\beta=x-\frac{\pi}{3}$. 应用公式 $\sin\alpha\cos\beta=\frac{1}{2}[\sin(\alpha+\beta)+\sin(\alpha-\beta)]$. 先计算 $\alpha+\beta$ 和 $\alpha-\beta$.

$$\alpha+\beta=\left(x+\frac{\pi}{6}\right)+\left(x-\frac{\pi}{3}\right)=2x-\frac{\pi}{6}$$

$$\alpha-\beta=\left(x+\frac{\pi}{6}\right)-\left(x-\frac{\pi}{3}\right)=\frac{\pi}{6}+\frac{\pi}{3}=\frac{\pi}{2}$$

代入公式后.

$$\begin{aligned}f(x)&=\frac{1}{2}\left[\sin\left(2x-\frac{\pi}{6}\right)+\sin\left(\frac{\pi}{2}\right)\right]\\&=\frac{1}{2}\left[\sin\left(2x-\frac{\pi}{6}\right)+1\right]\\&=\frac{1}{2}\sin\left(2x-\frac{\pi}{6}\right)+\frac{1}{2}\end{aligned}$$

知乎@余命数

分析最值经过积化和差, 我们把 $f(x)$ 化为 $y=A\sin(\omega x+\varphi)+k$ 的形式. 已知 $\sin(2x-\frac{\pi}{6})$ 的取值范围是 $[-1,1]$. 因此, $f(x)$ 的最大值在 $\sin(2x-\frac{\pi}{6})=1$ 时取得.

$$f(x)_{\max}=\frac{1}{2}(1)+\frac{1}{2}=1$$

函数的最大值为 1.

例题 1.45 (2022 年 6 月南京大学强基测试复试) 已知 $\alpha,\beta\in(0,\pi)$, 且 $\cos\alpha+\cos\beta-\cos(\alpha+\beta)=\frac{3}{2}$, 试求 α,β 的值.

解 这个方程混合了单角与和角, 直接求解 α,β 困难. 我们先把各项化为相同的角, 再统一处理.

统一视角并应用公式观察到方程中有 $\cos\alpha+\cos\beta$ 和 $\cos(\alpha+\beta)$. 我们先把它们都与 $\frac{\alpha+\beta}{2}$ 和 $\frac{\alpha-\beta}{2}$ 联系起来. 对 $\cos\alpha+\cos\beta$ 应用**和差化积**.

$$\cos\alpha+\cos\beta=2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

对 $\cos(\alpha+\beta)$ 应用二**倍角公式**, 将其视为 $\cos(2\cdot\frac{\alpha+\beta}{2})$.

$$\cos(\alpha+\beta)=2\cos^2\frac{\alpha+\beta}{2}-1$$

换元与代数化将上述两式代入原方程, 得

$$2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}-\left(2\cos^2\frac{\alpha+\beta}{2}-1\right)=\frac{3}{2}$$

我们再做换元以简化结构. 令 $a=\cos\frac{\alpha+\beta}{2}$, $b=\cos\frac{\alpha-\beta}{2}$. 则方程变为

$$2ab-(2a^2-1)=\frac{3}{2}$$

整理得到一个关于 a,b 的代数方程

$$2ab-2a^2+1=\frac{3}{2}\implies 4ab-4a^2+2=3\implies 4a^2-4ab+1=0$$

配方与求解接下来对关于 a,b 的式子配方.

$$(4a^2-4ab+b^2)-b^2+1=0$$

$$(2a-b)^2+(1-b^2)=0$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

由范围定值. 因为 $\alpha,\beta\in(0,\pi)$, 所以 $\alpha-\beta\in(-\pi,\pi)$, 则 $\frac{\alpha-\beta}{2}\in(-\frac{\pi}{2},\frac{\pi}{2})$. 在此区间内, $b=\cos\frac{\alpha-\beta}{2}>0$. 同时 $b\leq 1$. 因此 $b^2\leq 1\implies 1-b^2\geq 0$. 接着, 方程 $(2a-b)^2+(1-b^2)=0$ 是两个非负数之和等于零, 那么唯一的可能是两个非负数同时为零.

$$\begin{cases}(2a-b)^2=0\implies 2a=b\\ 1-b^2=0\end{cases}$$

由 $1-b^2=0$ 及 $b>0$ 可得 $b=1$. 代入 $2a=b$ 可得 $a=\frac{1}{2}$.

反解定角将 a,b 的值代回去.

$$b=\cos\frac{\alpha-\beta}{2}=1\implies\frac{\alpha-\beta}{2}=0\implies\alpha=\beta$$

$$a=\cos\frac{\alpha+\beta}{2}=\frac{1}{2}$$

因为 $\alpha=\beta$, 所以 $\cos\frac{2\alpha}{2}=\cos\alpha=\frac{1}{2}$. 又因为 $\alpha\in(0,\pi)$, 所以 $\alpha=\frac{\pi}{3}$.

因此, $\alpha=\beta=\dfrac{\pi}{3}$.

例题 1.46 (2010 年清华大学自主招生测试) 求 $\sin^4 10^\circ+\sin^4 50^\circ+\sin^4 70^\circ$ 的值.

解 本题角度特殊、幂次较高. 直接计算较繁琐, 我们先用降幂公式降次, 再结合和差化简化简求和.

降幂处理处理高次幂通常先用降幂公式降次.

$$\sin^2 x=\frac{1-\cos 2x}{2}\implies\sin^4 x=\left(\frac{1-\cos 2x}{2}\right)^2=\frac{1-2\cos 2x+\cos^2 2x}{4}$$

将 $10^\circ,50^\circ,70^\circ$ 分别代入, 得

$$\begin{aligned}\text{原式}&=\frac{1-2\cos 20^\circ+\cos^2 20^\circ}{4}+\frac{1-2\cos 100^\circ+\cos^2 100^\circ}{4}\\&\quad+\frac{1-2\cos 140^\circ+\cos^2 140^\circ}{4}\\&=\frac{1}{4}\Big[3-2(\cos 20^\circ+\cos 100^\circ+\cos 140^\circ)\\&\quad+(\cos^2 20^\circ+\cos^2 100^\circ+\cos^2 140^\circ)\Big]\end{aligned}$$

问题转化为计算两个括号内的和.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

计算一次项之和令 $S_1 = \cos 20^\circ + \cos 100^\circ + \cos 140^\circ$. 先对后两项使用和差化积.

$$\begin{aligned}\cos 100^\circ + \cos 140^\circ &= 2\cos\frac{100^\circ + 140^\circ}{2}\cos\frac{100^\circ - 140^\circ}{2} \\ &= 2\cos 120^\circ \cos(-20^\circ) = 2\left(-\frac{1}{2}\right)\cos 20^\circ = -\cos 20^\circ\end{aligned}$$

所以, $S_1 = \cos 20^\circ + (-\cos 20^\circ) = 0$.

计算二次项之和令 $S_2 = \cos^2 20^\circ + \cos^2 100^\circ + \cos^2 140^\circ$. 再次进行降幂处理.

$$\begin{aligned}S_2 &= \frac{1+\cos 40^\circ}{2} + \frac{1+\cos 200^\circ}{2} + \frac{1+\cos 280^\circ}{2} \\ &= \frac{3}{2} + \frac{1}{2}(\cos 40^\circ + \cos 200^\circ + \cos 280^\circ)\end{aligned}$$

计算括号内的和 $S_3 = \cos 40^\circ + \cos 200^\circ + \cos 280^\circ$. 根据诱导公式, $\cos 200^\circ = \cos(180^\circ + 20^\circ) = -\cos 20^\circ$, $\cos 280^\circ = \cos(360^\circ - 80^\circ) = \cos 80^\circ$.

$$S_3 = \cos 40^\circ - \cos 20^\circ + \cos 80^\circ$$

对 $\cos 40^\circ + \cos 80^\circ$ 应用和差化积.

$$\begin{aligned}\cos 40^\circ + \cos 80^\circ &= 2\cos\frac{40^\circ + 80^\circ}{2}\cos\frac{40^\circ - 80^\circ}{2} \\ &= 2\cos 60^\circ \cos(-20^\circ) = 2\left(\frac{1}{2}\right)\cos 20^\circ = \cos 20^\circ\end{aligned}$$

所以, $S_3 = \cos 20^\circ - \cos 20^\circ = 0$. 因此, $S_2 = \frac{3}{2} + \frac{1}{2}(0) = \frac{3}{2}$.

整合求值将 $S_1 = 0$ 和 $S_2 = \frac{3}{2}$ 代回原式.

$$\text{原式} = \frac{1}{4}\left[3 - 2(0) + \frac{3}{2}\right] = \frac{1}{4}\left(3 + \frac{3}{2}\right) = \frac{1}{4}\left(\frac{9}{2}\right) = \frac{9}{8}$$

最终结果为 $\dfrac{9}{8}$.

来龙去脉

我们已研究了最基本的正弦函数 $y = \sin x$; 它的振幅为 1, 周期为 2π, 从原点开始振动. 现实中的周期现象更复杂.

- 声音有高低**音量**之分;
- 琴弦有不同**音调**之别;
- 振动有不同**起始时刻**之差;
- 潮汐有不同**基准海平面**之高低.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

描述这些现象需要为基础波形 $y=\sin x$ 引入参数，写成 $y=A\sin(\omega x+\varphi)+B$ $(A>0,\omega>0)$. 参数 A,ω,φ,B 分别对应振幅、周期、相位和垂直平移四种几何变换. 我们在本节中研究每个参数对函数图像的单独作用，以及它们共同作用时的效果.

$y=A\sin(\omega x+\varphi)+B$ 的图像可看成由 $y=\sin x$ 经过一系列几何变换得到. 参数 A,B,ω,φ 分别控制一种特定形变. 我们逐一考察每个参数的效果.

1.13.2.1 参数 B 与垂直平移

参数 B 控制图像的**整体上下平移**. 它不改变图像的形状(振幅、周期)，只改变其垂直位置.

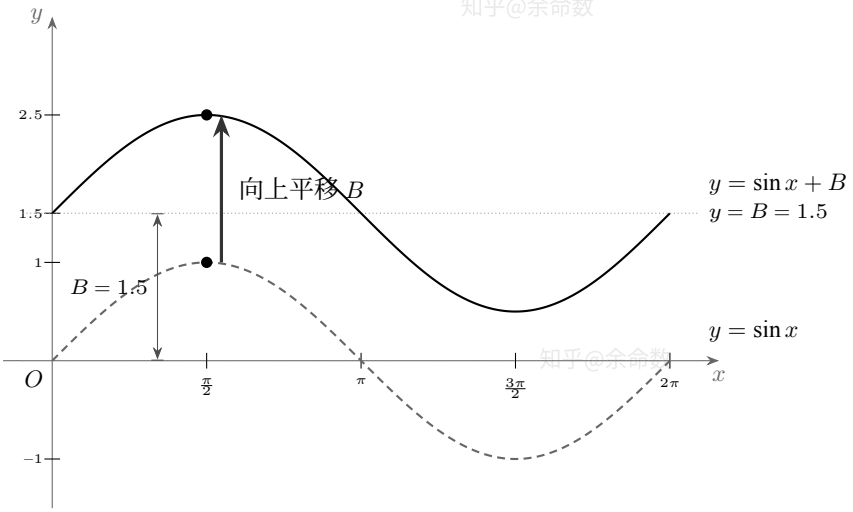


图 1.17　参数 B 导致图像垂直平移，得到 $y=\sin x+B$

- **变换规则**. 将 $y=f(x)$ 的图像向上 $(B>0)$ 或向下 $(B<0)$ 平移 $|B|$ 个单位长度，得到 $y=f(x)+B$ 的图像.
- **性质影响**.
  - **对称中心线**. 图像的平衡位置由 x 轴 $(y=0)$ 平移至直线 $y=B$.
  - **最值**. 最大值由 A 变为 $A+B$，最小值由 $-A$ 变为 $-A+B$.

1.13.2.2 参数 A 与垂直伸缩(振幅变换)

参数 A $(A>0)$ 控制图像在垂直方向上的**拉伸或压缩**，即改变图像的**振幅**.

- **变换规则**. 将 $y=f(x)$ 图像上所有点的纵坐标变为原来的 A 倍(横坐标不变)，得到 $y=Af(x)$ 的图像.
- **性质影响**.
  - **值域**. 值域由 $[-1,1]$ 变为 $[-A,A]$.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

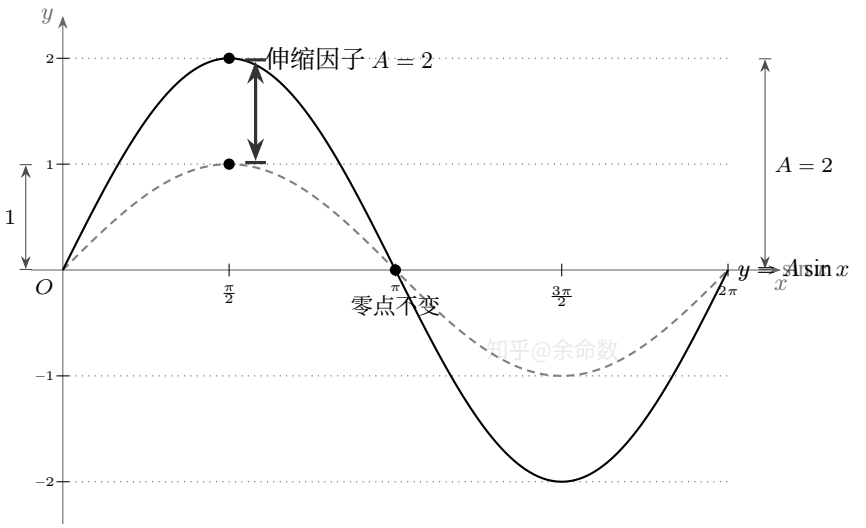


图 1.18 参数 A 使图像在垂直方向伸缩 A 倍(振幅变为 A)

- **最值**. 最大值为 A, 最小值为 $-A$.
- **不变性**. 图像的周期、对称轴位置和零点位置均**不改变**.

1.13.2.3 参数 ω 与水平伸缩(周期变换)

参数 ω ($\omega > 0$) 控制图像在水平方向上的**压缩或拉伸**, 直接改变图像的**周期**.

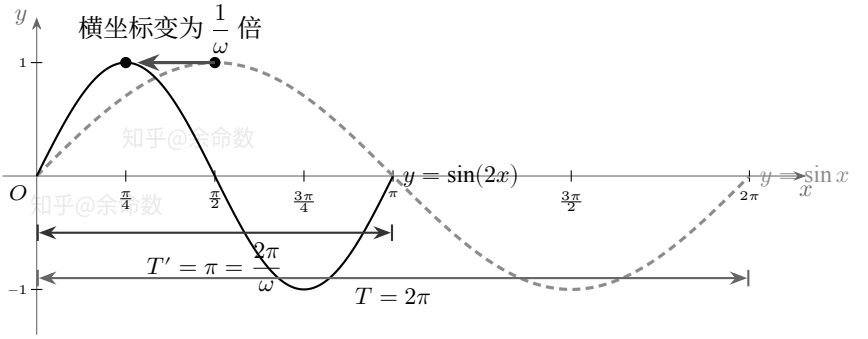


图 1.19 参数 ω 使图像在水平方向压缩 ω 倍(周期变为 $\dfrac{2\pi}{\omega}$)

- **变换规则**. 将 $y = f(x)$ 图像上所有点的横坐标变为原来的 $\frac{1}{\omega}$ 倍(纵坐标不变), 得到 $y = f(\omega x)$ 的图像.
  - 当 $\omega > 1$ 时, 图像被**压缩**.
  - 当 $0 < \omega < 1$ 时, 图像被**拉伸**.
- **性质影响**.
  - **周期**. 最小正周期由 T 变为 $T' = \dfrac{T}{\omega}$. 对于正弦函数, 由 2π 变为 $\dfrac{2\pi}{\omega}$.
  - **不变性**. 图像的振幅、值域和平衡位置均**不改变**.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.13.2.4 参数 φ 与水平平移(相位变换)

参数 φ 控制图像的**整体左右平移**. 它不改变图像的形状, 只改变其水平位置, 即**相位**.

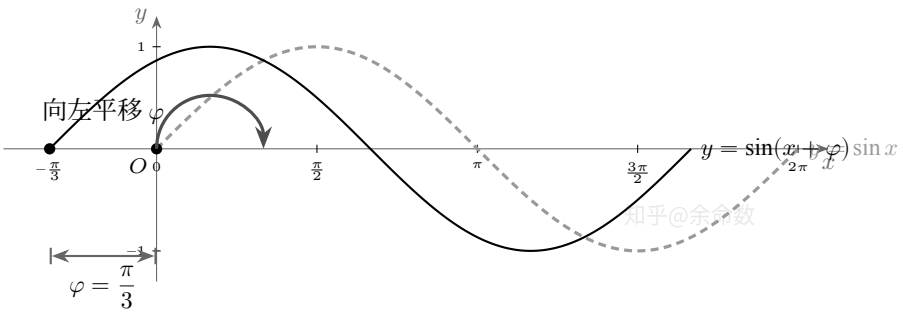


图 1.20　参数 φ 使图像沿 x 轴平移; $\varphi > 0$ 时向左平移, $\varphi < 0$ 时向右平移

- **变换规则**. 将 $y = f(x)$ 的图像进行平移.
  - **向左**平移 φ 个单位 ($\varphi > 0$), 得到 $y = f(x + \varphi)$ 的图像.
  - **向右**平移 φ 个单位 ($\varphi > 0$), 得到 $y = f(x - \varphi)$ 的图像.

  即向左平移对应函数中的加号, 向右平移对应函数中的减号.
- **性质影响**.
  - **位置**. 图像上所有点的位置发生水平移动, 因此对称轴、对称中心、最值点、零点的位置都发生改变.
  - **不变性**. 图像的形状完全不变, 即振幅、周期和值域都**不改变**.

1.14 $y = A\sin(\omega x + \varphi) + B$

1.14.1 由 $y = \sin x$ 到 $y = A\sin(\omega x + \varphi) + B$ 的图像变换

我们想把 $y = \sin x$ 的图像变成 $y = A\sin(\omega x + \varphi) + B$. 伸缩与平移的顺序不同, 得到的结果也会不同. 我们先伸缩、后平移, 这样最不容易混淆.

路径一

$$
\begin{aligned}
y = \sin x &\xrightarrow{\text{周期变换}} y = \sin(\omega x) \\
&\xrightarrow{\text{相位变换}} y = \sin(\omega x + \varphi) \\
&\xrightarrow{\text{振幅变换}} y = A\sin(\omega x + \varphi) \\
&\xrightarrow{\text{垂直平移}} y = A\sin(\omega x + \varphi) + B
\end{aligned}
$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

我们一步一步地看.

1. **周期变换**. 将 $y=\sin x$ 的图像上所有点的横坐标变为原来的 $\frac{1}{\omega}$ 倍(纵坐标不变),得到 $y=\sin(\omega x)$.
2. **相位变换**. 将 $y=\sin(\omega x)$ 的图像**向左**平移 $\frac{\varphi}{\omega}$ 个单位长度,得到 $y=\sin\left(\omega\left(x+\frac{\varphi}{\omega}\right)\right)=\sin(\omega x+\varphi)$.
3. **振幅变换**. 将 $y=\sin(\omega x+\varphi)$ 的图像上所有点的纵坐标变为原来的 A 倍(横坐标不变),得到 $y=A\sin(\omega x+\varphi)$.
4. **垂直平移**. 将 $y=A\sin(\omega x+\varphi)$ 的图像向上 $(B>0)$ 或向下 $(B<0)$ 平移 $|B|$ 个单位长度,得到最终图像.

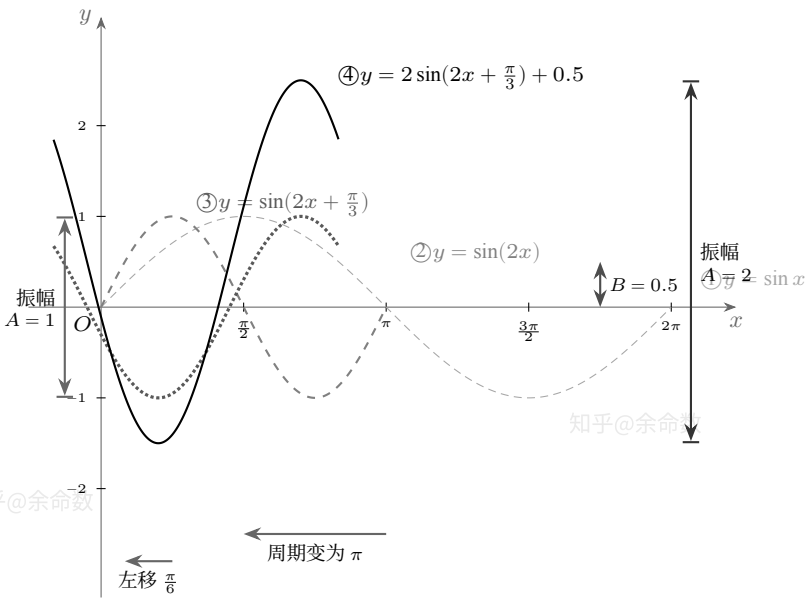


图 1.21 由 $y=\sin x$ 到 $y=2\sin(2x+\frac{\pi}{3})+0.5$ 的图像变换过程

注 1.14 (平移量的确定)
我们进行水平平移时,先把 x 的系数 ω 提出,得到 $y=\sin\left(\omega\left(x+\frac{\varphi}{\omega}\right)\right)$.

- 从 $y=\sin(\omega x)$ 到 $y=\sin(\omega x+\varphi)$,图像向左平移了 $\frac{\varphi}{\omega}$ 个单位.
- 另一条路径先把 $y=\sin x$ 向左平移 φ 个单位,得到 $y=\sin(x+\varphi)$,再把横坐标缩为原来的 $\frac{1}{\omega}$ 倍,于是平移量也缩为 $\frac{\varphi}{\omega}$,最终结果仍是 $y=\sin(\omega x+\varphi)$.

因此,我们采用先伸缩、后平移的路径;平移量是 $\frac{\varphi}{\omega}$.

知乎@余命数 知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1.14.2 整体换元思想

我们分析函数 $f(x) = A\sin(\omega x + \varphi) + B$ 的性质时，把内层函数 $u = \omega x + \varphi$ 看成一个整体. 先求出基础函数 $y = A\sin u + B$ 的对应性质区间，再反解出 x 的取值范围.

- **值域与最值**. 当 $\sin(\omega x + \varphi) = 1$ 时，$y_{\max} = |A| + B$. 当 $\sin(\omega x + \varphi) = -1$ 时，$y_{\min} = -|A| + B$.
- **单调区间**.（以 $A > 0, \omega > 0$ 为例）要求单调递增区间，即求满足 $2k\pi - \frac{\pi}{2} \leq \omega x + \varphi \leq 2k\pi + \frac{\pi}{2}$ 的 x 的范围. 要求单调递减区间，即求满足 $2k\pi + \frac{\pi}{2} \leq \omega x + \varphi \leq 2k\pi + \frac{3\pi}{2}$ 的 x 的范围.
- **对称性**.（以 $A > 0, \omega > 0$ 为例）对称轴由 $\omega x + \varphi = k\pi + \frac{\pi}{2}$ 解出. 对称中心由 $\omega x + \varphi = k\pi$ 解出其横坐标.

1.14.3 典例剖析

例题 1.47 (由图求参) 已知函数 $f(x) = A\sin(\omega x + \varphi) + B$ $(A > 0, \omega > 0, |\varphi| < \frac{\pi}{2})$ 的部分图像如图所示，求其解析式.

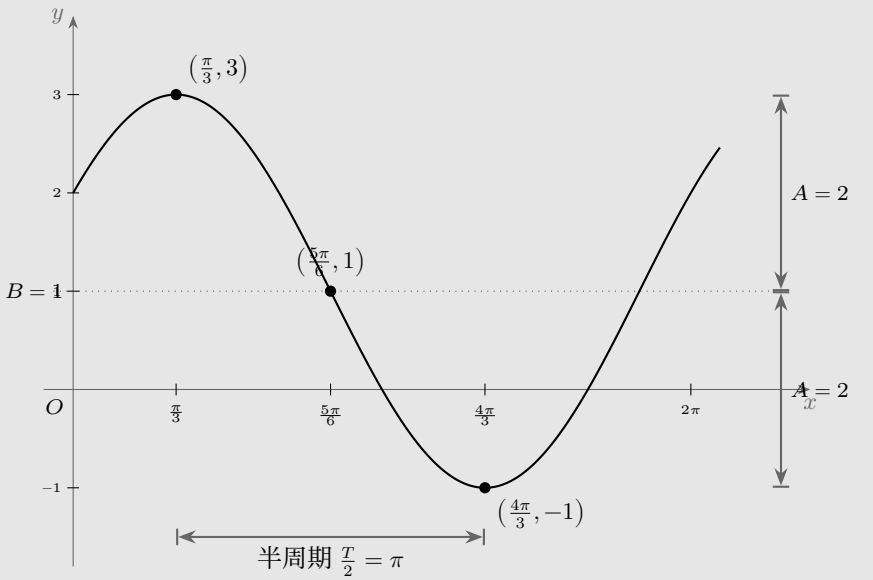


图 1.22　函数 $f(x) = A\sin(\omega x + \varphi) + B$ 的部分图像

解 我们按 **定** B、**定** A、**定** ω、**定** φ 的顺序，从图像中逐一求出参数.

先确定平衡位置 B

知乎@余命数

由图可知, 函数的最大值 $y_{\max}=3$, 最小值为 $y_{\min}=-1$.

$$B=\frac{y_{\max}+y_{\min}}{2}=\frac{3+(-1)}{2}=1$$

再确定振幅 A

$$A=\frac{y_{\max}-y_{\min}}{2}=\frac{3-(-1)}{2}=2$$

接着确定角频率 ω

图中给出了一个最高点 $\left(\frac{\pi}{3},3\right)$ 和一个相邻的最低点 $\left(\frac{4\pi}{3},-1\right)$. 它们之间的水平距离恰好是半个周期.

$$\frac{T}{2}=\frac{4\pi}{3}-\frac{\pi}{3}=\pi \implies T=2\pi$$

由周期公式 $T=\frac{2\pi}{\omega}$ 可得

$$\omega=\frac{2\pi}{T}=\frac{2\pi}{2\pi}=1$$

最后确定初相 φ

将 A,B,ω 的值代入, 函数解析式为 $f(x)=2\sin(x+\varphi)+1$.

我们取最高点 $\left(\frac{\pi}{3},3\right)$ 代入来求 φ.

$$3=2\sin\left(\frac{\pi}{3}+\varphi\right)+1 \implies 2=2\sin\left(\frac{\pi}{3}+\varphi\right) \implies \sin\left(\frac{\pi}{3}+\varphi\right)=1$$

于是内层函数满足 $\frac{\pi}{3}+\varphi=\frac{\pi}{2}+2k\pi, k\in\mathbb{Z}$.

$$\frac{\pi}{3}+\varphi=\frac{\pi}{2}+2k\pi \implies \varphi=\frac{\pi}{6}+2k\pi$$

由条件 $|\varphi|<\frac{\pi}{2}$, 取 $k=0$, 得 $\varphi=\frac{\pi}{6}$.

答案. 所求函数的解析式为 $f(x)=2\sin\left(x+\frac{\pi}{6}\right)+1$

注 1.15 (五点法反求 φ)

我们确定 φ 时, 把五点法中的特殊点代入, 这些点对应最高点、最低点或与平衡位置的交点. 这样得到的方程容易求解.

- **代入最高点**. 此时 $\sin(\omega x+\varphi)=1$, 即 $\omega x+\varphi=\frac{\pi}{2}+2k\pi$.
- **代入最低点**. 此时 $\sin(\omega x+\varphi)=-1$, 即 $\omega x+\varphi=-\frac{\pi}{2}+2k\pi$.
- **代入中心线上升点**. 此时 $\sin(\omega x+\varphi)=0$ 且函数递增, 即 $\omega x+\varphi=2k\pi$.

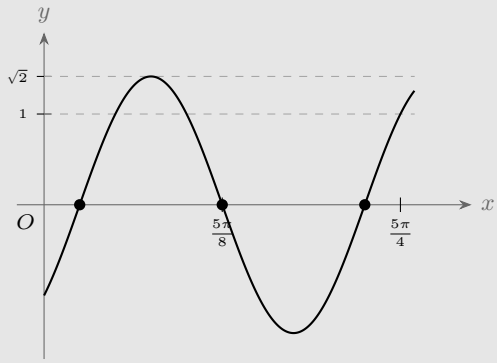
知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

- **代入中心线下降点**. 此时 $\sin(\omega x + \varphi) = 0$ 且函数递减, 即 $\omega x + \varphi = \pi + 2k\pi$.

这样得到的方法称为“五点法求 φ”.

综合题中的导数判别

例题 1.48 (多选, 2023 山东师大附中 10 月月考) 函数 $f(x) = \sqrt{2}\sin(\omega x + \varphi)$ $(\omega > 0, |\varphi| < \frac{\pi}{2})$ 的部分图象如图所示, 则下列说法中正确的有(　)



A. $f(x)$ 的最小正周期 T 为 π

B. $f(x)$ 向右平移 $\dfrac{3\pi}{8}$ 个单位后得到的新函数是偶函数

C. 若方程 $f(x) = 1$ 在 $(0, m)$ 上共有 4 个根, 则这 4 个根的和为 $\dfrac{7\pi}{2}$

D. $f(x)(x \in \left[0, \dfrac{5\pi}{4}\right])$ 图象上的动点 M 到直线 $2x - y + 4 = 0$ 的距离最小时, M 的横坐标为 $\dfrac{\pi}{4}$

解　我们先用图像确定函数解析式, 再逐一判断四个选项. (D 选项用到切线斜率, 见 第四卷第 16 章.)

假设选项 A 正确, 即 $T = \pi$. 则 $\omega = \frac{2\pi}{T} = 2$. 函数为 $f(x) = \sqrt{2}\sin(2x + \varphi)$. 将图中的零点 $(\frac{5\pi}{8}, 0)$ 代入, 得

$$\sqrt{2}\sin\left(2 \cdot \frac{5\pi}{8} + \varphi\right) = 0 \implies \sin\left(\frac{5\pi}{4} + \varphi\right) = 0$$

即 $\frac{5\pi}{4} + \varphi = k\pi, k \in \mathbb{Z}$.

$$\varphi = k\pi - \frac{5\pi}{4}$$

由条件 $|\varphi| < \frac{\pi}{2}$, 取 $k = 1$, 得 $\varphi = \pi - \frac{5\pi}{4} = -\frac{\pi}{4}$. 于是函数解析式为 $f(x) = \sqrt{2}\sin(2x - \frac{\pi}{4})$. 代入 $x = \frac{\pi}{8}$ 和 $x = \frac{5\pi}{8}$ 均得 $f(x) = 0$, 与图象吻合.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

逐一判断选项

- **对于 A**. 它基于 $T=\pi$ 推导出的解析式, 故 A 正确.
- **对于 B**. 将 $f(x)$ 向右平移 $\frac{3\pi}{8}$ 个单位得到 $g(x)=f(x-\frac{3\pi}{8})$.

$$\begin{aligned} g(x) &= \sqrt{2}\sin\left(2\left(x-\frac{3\pi}{8}\right)-\frac{\pi}{4}\right) \\ &= \sqrt{2}\sin\left(2x-\frac{3\pi}{4}-\frac{\pi}{4}\right)=\sqrt{2}\sin(2x-\pi)=-\sqrt{2}\sin(2x) \end{aligned}$$

  该函数是奇函数, 故 B 错误.

- **对于 C**. 解方程 $f(x)=1 \implies \sqrt{2}\sin(2x-\frac{\pi}{4})=1 \implies \sin(2x-\frac{\pi}{4})=\frac{\sqrt{2}}{2}$.

$$2x-\frac{\pi}{4}=2k\pi+\frac{\pi}{4} \quad \text{或} \quad 2x-\frac{\pi}{4}=2k\pi+\frac{3\pi}{4}$$

$$2x=2k\pi+\frac{\pi}{2} \quad \text{或} \quad 2x=2k\pi+\pi$$

$$x=k\pi+\frac{\pi}{4} \quad \text{或} \quad x=k\pi+\frac{\pi}{2}$$

  在 $(0,m)$ 内的 4 个正根, 从小到大依次为. $x_1=\frac{\pi}{4}$, $x_2=\frac{\pi}{2}$（当 $k=0$）$x_3=\pi+\frac{\pi}{4}=\frac{5\pi}{4}$, $x_4=\pi+\frac{\pi}{2}=\frac{3\pi}{2}$, 它们对应 $k=1$. 这 4 个根的和为 $\frac{\pi}{4}+\frac{2\pi}{4}+\frac{5\pi}{4}+\frac{6\pi}{4}=\frac{14\pi}{4}=\frac{7\pi}{2}$. 故 C 正确.

- **对于 D**. 求动点 $M(x,f(x))$ 到直线 $2x-y+4=0$ 的距离最小值, 即求

$$d(x)=\frac{|2x-f(x)+4|}{\sqrt{5}}$$

  在 $x\in[0,\frac{5\pi}{4}]$ 上的最小值. 由于 $2x-f(x)+4\geq 4-\sqrt{2}>0$, 可直接比较 $h(x)=2x-f(x)+4$. 先找内部驻点.

$$h'(x)=2-f'(x)=2-2\sqrt{2}\cos\left(2x-\frac{\pi}{4}\right)$$

  令 $h'(x)=0$, 得 $\cos(2x-\frac{\pi}{4})=\frac{\sqrt{2}}{2}$, 所以区间内的候选点为 $x=0,\frac{\pi}{4},\pi,\frac{5\pi}{4}$. 再比较端点与候选点.

$$h(0)=5, \quad h\left(\frac{\pi}{4}\right)=\frac{\pi}{2}+3, \quad h(\pi)=2\pi+5, \quad h\left(\frac{5\pi}{4}\right)=\frac{5\pi}{2}+3.$$

  其中 $\frac{\pi}{2}+3<5<2\pi+5$ 且 $\frac{\pi}{2}+3<\frac{5\pi}{2}+3$, 故距离在 $x=\frac{\pi}{4}$ 处最小.D 正确.

正确的选项为 ACD.

切线条件用导数

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 1.49 (多选，2023 湖南长郡中学月考) 已知 $x_0=\dfrac{2\pi}{3}$ 是函数 $f(x)=\sin(2x+\varphi)(0<\varphi<\pi)$ 的一个零点，则(　)

A. $f(x)$ 在区间 $\left[0,\dfrac{5\pi}{12}\right]$ 上单调递减

B. $f(x)$ 在区间 $\left(-\dfrac{\pi}{12},\dfrac{11\pi}{12}\right)$ 只有一个极值点

C. 直线 $x=\dfrac{7\pi}{6}$ 是曲线 $y=f(x)$ 的对称轴

D. 直线 $y=\dfrac{\sqrt{3}}{2}-x$ 是曲线 $y=f(x)$ 的切线

知乎@余命数

解 我们先用给定的零点求出参数，再判断各选项.

我们先确定函数解析式. 将零点 $x_0=\frac{2\pi}{3}$ 代入 $f(x)=\sin(2x+\varphi)$.

$$\sin\left(2\cdot\frac{2\pi}{3}+\varphi\right)=\sin\left(\frac{4\pi}{3}+\varphi\right)=0$$

即 $\frac{4\pi}{3}+\varphi=k\pi,k\in\mathbb{Z}$.

$$\varphi=k\pi-\frac{4\pi}{3}$$

由条件 $0<\varphi<\pi$，取 $k=2$，得 $\varphi=2\pi-\frac{4\pi}{3}=\frac{2\pi}{3}$. 因此 $f(x)=\sin(2x+\frac{2\pi}{3})$.

逐一判断选项

- **对于 A**. 求单调递减区间. 令 $u=2x+\frac{2\pi}{3}$，$\sin u$ 的递减区间为 $[\frac{\pi}{2}+2k\pi,\frac{3\pi}{2}+2k\pi]$.

$$\frac{\pi}{2}+2k\pi\le 2x+\frac{2\pi}{3}\le\frac{3\pi}{2}+2k\pi$$

$$-\frac{\pi}{6}+2k\pi\le 2x\le\frac{5\pi}{6}+2k\pi\implies-\frac{\pi}{12}+k\pi\le x\le\frac{5\pi}{12}+k\pi$$

  当 $k=0$ 时，一个单调递减区间为 $[-\frac{\pi}{12},\frac{5\pi}{12}]$. 题目给出的区间 $[0,\frac{5\pi}{12}]$ 是上述区间的子集，因此 $f(x)$ 在此区间上单调递减. 故 A 正确.

- **对于 B**. 求极值点. 极值点满足 $u=2x+\frac{2\pi}{3}=\frac{\pi}{2}+k\pi$.

$$2x=\frac{\pi}{2}+k\pi-\frac{2\pi}{3}=-\frac{\pi}{6}+k\pi\implies x=\frac{k\pi}{2}-\frac{\pi}{12}$$

  在区间 $(-\frac{\pi}{12},\frac{11\pi}{12})$ 内寻找满足条件的 x：若 $k=1,x=\frac{\pi}{2}-\frac{\pi}{12}=\frac{5\pi}{12}$，在区间内. 若 $k=2,x=\pi-\frac{\pi}{12}=\frac{11\pi}{12}$，不在区间内. 若 $k=0,x=-\frac{\pi}{12}$，不在区间内. 因此该区间内只有一个极值点. 故 B 正确.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

- **对于 C**. 判断对称轴. 对称轴方程为 $x=\frac{k\pi}{2}-\frac{\pi}{12}$. 将 $x=\frac{7\pi}{6}$ 代入该式看能否取得.

$$\frac{7\pi}{6}=\frac{k\pi}{2}-\frac{\pi}{12}\implies\frac{14\pi}{12}+\frac{\pi}{12}=\frac{k\pi}{2}$$
$$\implies\frac{15\pi}{12}=\frac{k\pi}{2}\implies\frac{5\pi}{4}=\frac{k\pi}{2}\implies k=\frac{5}{2}$$

  k 不是整数, $x=\frac{7\pi}{6}$ 不是对称轴. 故 C 错误.

- **对于 D**. 判断切线. 需找一个点 x_t, 同时满足 $f'(x_t)=-1$ 和 $f(x_t)=\frac{\sqrt{3}}{2}-x_t$. $f'(x)=2\cos(2x+\frac{2\pi}{3})$. 令 $f'(x_t)=-1\implies\cos(2x_t+\frac{2\pi}{3})=-\frac{1}{2}$. 一个简单的解是 $2x_t+\frac{2\pi}{3}=\frac{2\pi}{3}\implies x_t=0$. 验证 $x_t=0$. 切点坐标为 $x=0$, $y=f(0)=\sin(\frac{2\pi}{3})=\frac{\sqrt{3}}{2}$. 切点为 $(0,\frac{\sqrt{3}}{2})$. 将该点代入切线方程 $y=\frac{\sqrt{3}}{2}-x$, 左边为 $\frac{\sqrt{3}}{2}$, 右边为 $\frac{\sqrt{3}}{2}-0=\frac{\sqrt{3}}{2}$. 左右相等, 说明点在直线上. 斜率和点均满足, 该直线是曲线在 $x=0$ 处的切线. 故 D 正确.

正确的选项为 ABD.

1.14.4 高考真题选讲

以下选编近年高考中与正弦型函数 $A\sin(\omega x+\varphi)+b$ 相关的试题, 涵盖化简、性质与图像变换.

例题 1.50 (2016 年浙江卷理科数学第 10 题) 已知 $2\cos^2 x+\sin 2x=A\sin(\omega x+\varphi)+b\,(A>0)$, 则 $A=$_____, $b=$_____.

解 用降幂公式 $\cos^2 x=\frac{1+\cos 2x}{2}$ 把 $2\cos^2 x$ 降为一次,

$$2\cos^2 x+\sin 2x=(1+\cos 2x)+\sin 2x=\sin 2x+\cos 2x+1.$$

再用辅助角公式合并 $\sin 2x+\cos 2x$,

$$\sin 2x+\cos 2x=\sqrt{2}\sin\left(2x+\frac{\pi}{4}\right),$$

所以 $2\cos^2 x+\sin 2x=\sqrt{2}\sin(2x+\frac{\pi}{4})+1$, 故 $A=\sqrt{2}, b=1$.

例题 1.51 (2021 年全国乙卷文科数学第 4 题) 函数 $f(x)=\sin\frac{x}{3}+\cos\frac{x}{3}$ 的最小正周期和最大值分别是

A. 3π 和 $\sqrt{2}$　　C. 6π 和 $\sqrt{2}$

B. 3π 和 2　　D. 6π 和 2

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 答案:C. 用辅助角公式合并,

$$f(x)=\sin\frac{x}{3}+\cos\frac{x}{3}=\sqrt{2}\sin\left(\frac{x}{3}+\frac{\pi}{4}\right),$$

最小正周期 $T=\frac{2\pi}{1/3}=6\pi$, 最大值为 $\sqrt{2}$.

例题 1.52 (2021 年全国 □ 卷数学第 4 题) 下列区间中, 函数 $f(x)=7\sin\left(x-\frac{\pi}{6}\right)$ 单调递增的区间是

A. $\left(0,\frac{\pi}{2}\right)$　　C. $\left(\pi,\frac{3\pi}{2}\right)$

B. $\left(\frac{\pi}{2},\pi\right)$　　D. $\left(\frac{3\pi}{2},2\pi\right)$

知乎@余命数

解 答案:A. 令 $-\frac{\pi}{2}+2k\pi\leqslant x-\frac{\pi}{6}\leqslant\frac{\pi}{2}+2k\pi\,(k\in\mathbb{Z})$, 解得 $-\frac{\pi}{3}+2k\pi\leqslant x\leqslant\frac{2\pi}{3}+2k\pi$. 取 $k=0$, 单调递增区间为 $[-\frac{\pi}{3},\frac{2\pi}{3}]$, 而 $(0,\frac{\pi}{2})\subset[-\frac{\pi}{3},\frac{2\pi}{3}]$.

例题 1.53 (2017 年全国 □ 卷理科数学第 9 题) 已知曲线 $C_1:y=\cos x$, $C_2:y=\sin\left(2x+\frac{2\pi}{3}\right)$, 则

A. 把 C_1 上各点的横坐标伸长到原来的 2 倍, 纵坐标不变, 再把得到的曲线向右平移 $\frac{\pi}{6}$ 个单位长度, 得到曲线 C_2

B. 把 C_1 上各点的横坐标伸长到原来的 2 倍, 纵坐标不变, 再把得到的曲线向左平移 $\frac{\pi}{12}$ 个单位长度, 得到曲线 C_2

C. 把 C_1 上各点的横坐标缩短到原来的 $\frac{1}{2}$ 倍, 纵坐标不变, 再把得到的曲线向右平移 $\frac{\pi}{6}$ 个单位长度, 得到曲线 C_2

D. 把 C_1 上各点的横坐标缩短到原来的 $\frac{1}{2}$ 倍, 纵坐标不变, 再把得到的曲线向左平移 $\frac{\pi}{12}$ 个单位长度, 得到曲线 C_2

解 答案:D. 先统一函数名, 把 C_2 化为余弦形式,

$$C_2:y=\sin\left(2x+\frac{2\pi}{3}\right)=\cos\left(2x+\frac{2\pi}{3}-\frac{\pi}{2}\right)=\cos\left(2x+\frac{\pi}{6}\right).$$

知乎@余命数

从 $C_1:y=\cos x$ 出发, 横坐标缩短到原来的 $\frac{1}{2}$ 倍, 即 ω 从 1 变为 2, 得到 $y=\cos 2x$; 再向左平移 $\frac{\pi}{12}$ 个单位, 得到 $y=\cos\left[2\left(x+\frac{\pi}{12}\right)\right]=\cos\left(2x+\frac{\pi}{6}\right)$, 即 C_2.

例题 1.54 (2016 年全国 □ 卷理科数学第 14 题) 函数 $y=\sin x-\sqrt{3}\cos x$ 的图象可由函数 $y=\sin x+\sqrt{3}\cos x$ 的图象至少向右平移 _____ 个单位长度得到.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 分别化简两个函数,

$$y=\sin x+\sqrt{3}\cos x=2\sin\left(x+\frac{\pi}{3}\right),\quad y=\sin x-\sqrt{3}\cos x=2\sin\left(x-\frac{\pi}{3}\right).$$

设向右平移 φ ($\varphi>0$) 个单位,则 $2\sin\left(x+\frac{\pi}{3}-\varphi\right)=2\sin\left(x-\frac{\pi}{3}\right)$,需要

$$\frac{\pi}{3}-\varphi=-\frac{\pi}{3}+2k\pi\quad(k\in\mathbb{Z}),$$

解得 $\varphi=\frac{2\pi}{3}-2k\pi$. 取 $k=0$ 得最小正值 $\varphi=\frac{2\pi}{3}$.

知乎@余命数

1.15 ω 的约束与求解

注 1.16
个别变式使用导数表示切线斜率,见第四卷第 16 章.

我们常需要根据函数在区间上的性质,反向推断解析式中的参数,尤其是决定周期的参数 ω. 这些性质包括单调性、最值点的数量、零点的个数等. 我们把这类问题统称为参数 ω 的约束与求解.

这类问题可归结为建立周期与区间长度之间的不等关系. 函数在给定区间内能容纳多少个周期,决定了该区间内最值点和零点的数量. 下面给出两条常用的分析路径.

- **整体换元法**. 将 $u=\omega x+\varphi$ 视为一个整体,分析当 x 在给定区间内变化时 u 所在的区间. 然后考察标准正弦函数 $y=\sin u$ 在这个有效区间内必须包含哪些特征点(最高点、最低点、零点),以及必须排除哪些特征点,从而建立关于区间端点的不等式来求解 ω.
- **特征点定位法**,亦称 ω 卡根法. 直接写出函数 $f(x)$ 的所有特征点横坐标的通式(该通式含有 ω). 然后根据题意,确定第 n 个特征点必须落在给定区间内,而第 $n+1$ 个特征点必须落在区间外或恰好在端点上. 通过这种卡位的方式,建立关于 ω 的不等式组.

这两条路径最终都归结为建立并求解不等式.

用导数找极值/周期约束

例题 1.55 (2022 年 6 月全国甲卷 (理) 第 11 题) 设函数 $f(x)=\sin(\omega x+\frac{\pi}{3})$ 在区间 $(0,\pi)$ 恰有三个极值点,两个零点,则 ω 的取值范围是(　)

A. $\left(\frac{5}{3},\frac{13}{6}\right]$　　C. $\left(\frac{13}{6},\frac{8}{3}\right]$

B. $\left(\frac{5}{3},\frac{19}{6}\right)$　　D. $\left(\frac{13}{6},\frac{19}{6}\right)$

解 我们从整体换元与特征点定位两个角度处理.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

法一采用整体换元法. 令 $t=\omega x+\frac{\pi}{3}$, 则 $f(x)$ 的性质等价于 $y=\sin t$ 的性质. 当 $x\in(0,\pi)$ 时,

$$t\in\left(\frac{\pi}{3},\omega\pi+\frac{\pi}{3}\right).$$

要求 f 在 $(0,\pi)$ 内恰有三个极值点、两个零点, 等价于 $\sin t$ 在上述区间内恰有三个极值点、两个零点.

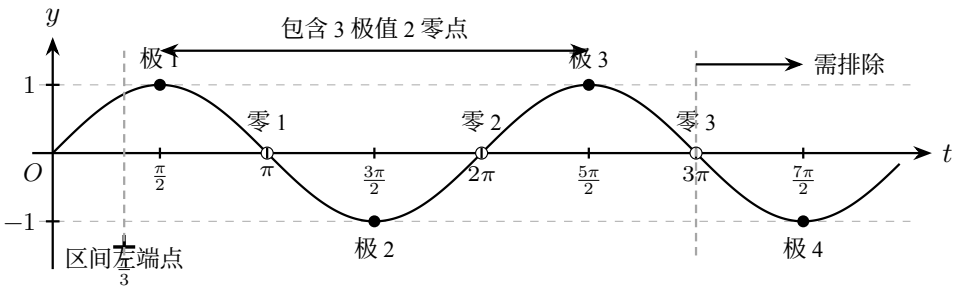
知乎@余命数

$t>0$ 时 $\sin t$ 的极值点与零点分别如下.

- 极值点 ($\sin t=\pm1$).$\frac{\pi}{2},\frac{3\pi}{2},\frac{5\pi}{2},\frac{7\pi}{2},\ldots$
- 零点 ($\sin t=0$).$\pi,2\pi,3\pi,4\pi,\ldots$

知乎@余命数

区间左端为 $t=\frac{\pi}{3}$. 恰含三个极值点时, 这三点只能是 $\frac{\pi}{2},\frac{3\pi}{2},\frac{5\pi}{2}$; 恰含两个零点时, 这两点只能是 $\pi,2\pi$.



区间 $\left(\frac{\pi}{3},\omega\pi+\frac{\pi}{3}\right)$ 必须包含的最后一个点是第三个极值点 $\frac{5\pi}{2}$. 同时, 它必须排除第三个零点 3π 和第四个极值点 $\frac{7\pi}{2}$. 其中, 3π 是更小的需要被排除的点. 因此, 区间的右端点 $\omega\pi+\frac{\pi}{3}$ 必须满足.

$$\frac{5\pi}{2}<\omega\pi+\frac{\pi}{3}\le3\pi$$

由于 $x\in(0,\pi)$, 即使 $\omega\pi+\frac{\pi}{3}=3\pi$ 成立, 对应的 x 值也不在原区间内, 因此右边界可以取等号. 解此不等式组.

$$\frac{5\pi}{2}-\frac{\pi}{3}<\omega\pi\le3\pi-\frac{\pi}{3}$$

$$\frac{13\pi}{6}<\omega\pi\le\frac{8\pi}{3}$$

$$\frac{13}{6}<\omega\le\frac{8}{3}$$

知乎@余命数

法二采用特征点定位法(ω 卡根法). 直接写出极值点与零点的横坐标通式. 由 $f(x)=0$ 得 $\omega x+\frac{\pi}{3}=k\pi$, 零点 $x_k=\frac{(3k-1)\pi}{3\omega}$, $k\in\mathbb{Z}$. 由 $f'(x)=0$ 得 $\omega x+\frac{\pi}{3}=k\pi+\frac{\pi}{2}$, 极值点 $x'_k=\frac{(6k+1)\pi}{6\omega}$, $k\in\mathbb{Z}$.

需在 $(0,\pi)$ 内恰有三个正极值点、两个正零点.

- 零点 $(x_k>0).k=1,2,3,\cdots \implies x_1=\frac{2\pi}{3\omega},x_2=\frac{5\pi}{3\omega},x_3=\frac{8\pi}{3\omega},\ldots$
- 极值点 $(x'_k>0).k=0,1,2,3,\cdots \implies x'_0=\frac{\pi}{6\omega},x'_1=\frac{7\pi}{6\omega},x'_2=\frac{13\pi}{6\omega},x'_3=\frac{19\pi}{6\omega},\ldots$

于是第三个极值点 x'_2 与第二个零点 x_2 须在 $(0,\pi)$ 内, 而第四个极值点 x'_3 与第三个零点 x_3 须在区间外, 即 $\geq\pi$. 等价地说, 区间内诸点中最大者小于 π, 区间外诸点中最小者 $\geq\pi$.

比较得 $x'_2=\frac{13\pi}{6\omega}$ 大于 $x_2=\frac{5\pi}{3\omega}$, 故区间内最大点为 x'_2; $x_3=\frac{8\pi}{3\omega}$ 小于 $x'_3=\frac{19\pi}{6\omega}$, 故区间外最小点为 x_3. 于是

$$\begin{cases}x'_2=\dfrac{13\pi}{6\omega}<\pi\\ x_3=\dfrac{8\pi}{3\omega}\geq\pi\end{cases}$$

分别解得 $\omega>\frac{13}{6}$ 与 $\omega\leq\frac{8}{3}$. 合并解得

$$\frac{13}{6}<\omega\leq\frac{8}{3}$$

两种方法结果一致, 故选 C.

求解此类关于 ω 的范围问题, 最终都是建立不等式.

- **整体换元法**侧重于图像理解. 它将问题转化为在标准正弦曲线上框选一个区间, 使得该区间恰好包含指定数量的特征点. 关键在于确定区间的左右端点必须卡在哪两个特征点之间.
- **特征点定位法**侧重于代数计算. 它通过求解通式, 直接定位最后一个应在区间内的点和第一个应在区间外的点, 然后利用区间端点将它们卡住. 这种方法需要细心处理多个点的通式和大小比较.

无论哪种方法, 处理不等式的边界条件 (等号能否取到) 都决定结果是否正确. 这通常取决于原区间的开闭性以及函数在端点处的取值情况.

我们根据函数在固定长度的区间内展现出的性质, 如零点、极值点的数量, 反向约束决定周期性伸缩的参数 ω.

问题等价于判断函数图像在区间 (a,b) 内能容纳多少个波形. ω 越大, 周期 $T=\frac{2\pi}{\omega}$ 越小, 图像被压缩得越厉害, 同样长度的区间内就能容纳更多的零点和极值点; ω 越小则反之.

我们的任务就是将“恰有 N 个极值点, M 个零点”这一性质描述, 翻译成一个关于 ω 的数学不等式.

于是, 卡位思想是解决此类问题的关键. 它指的是, 我们必须找到最后一个应该在区间内的特征点, 以及第一个不应该在区间内的特征点, 然后用区间的端点将它们卡住.

具体地, 假设区间为 (a,b), 得

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **内卡条件**. 最后一个需要被包含的特征点(设其横坐标为 $x_{\text{尾内}}$), 必须严格落在区间 (a,b) 内部. 这给出关于 ω 的一个下界约束.

$$x_{\text{尾内}} < b$$

- **外卡条件**. 第一个不应被包含的特征点(设其横坐标为 $x_{\text{首外}}$), 必须落在区间右侧或恰好在右端点上. 这给出关于 ω 的一个上界约束.

$$x_{\text{首外}} \geq b$$

联立这两个不等式, 即可解出 ω 的范围.

下面给出两条路径, 都用 $x_{\text{尾内}}$ 和 $x_{\text{首外}}$ 来建立不等式, 二者等价, 操作不同.

1.15.0.1 整体换元法(图像视角)

固定标准图像, 移动考察区间.

操作步骤.

1. **换元**. 令 $t = \omega x + \varphi$, 将问题转化为分析标准函数 $y = \sin t$.
2. **定区间**. 根据 $x \in (a,b)$, 确定新变量 t 的范围 $(\omega a + \varphi, \omega b + \varphi)$.
3. **数点**. 在 $y = \sin t$ 的标准图像上, 从 t 的左端点开始, 数出题目要求的 N 个极值点和 M 个零点.
4. **卡位**. 确定 t 的右端点 $\omega b + \varphi$ 必须满足.

$$t_{\text{尾内}} < \omega b + \varphi \leq t_{\text{首外}}$$

其中 $t_{\text{尾内}}$ 是最后一个应包含点的横坐标, $t_{\text{首外}}$ 是第一个应排除点的横坐标.

5. **求解**. 解此不等式, 得到 ω 的范围.

图像视角的思路直观, 与函数图像的联系紧密, 能清晰地看到特征点被框选的过程.

1.15.0.2 特征点定位法(代数视角)

固定考察区间, 计算伸缩后的特征点.

操作步骤.

1. **求通式**. 直接写出原函数 $f(x)$ 的所有极值点横坐标通式 x'_k 和零点横坐标通式 x_k(均含有 ω).
2. **定序号**. 根据 $x > a$ 的条件, 确定满足题意的第 1、2、……、N 个极值点和第 1、2、……、M 个零点的具体表达式.
3. **找关键点**. 在所有应包含的点中, 找出横坐标最大的那个 $x_{\text{尾内}}$. 在所有不应包含的点中, 找出横坐标最小的那个 $x_{\text{首外}}$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

4. **卡位**. 根据卡位思想, 建立不等式组.

$$\begin{cases} x_{\text{尾内}} < b \\ x_{\text{首外}} \geq b \end{cases}$$

5. **求解**. 解此不等式组, 得到 ω 的范围.

代数视角的方法程序化, 在特征点类型单一, 如只问零点或初相 φ 简单时, 计算可能更直接.

在建立不等式时, 等号能否取到, 决定最终结果是否正确. 判断依据如下.

- **看原区间开闭**. 若原区间是开区间 (a,b), 特征点恰好落在端点 b 上时, 它并不属于区间内. 因此, 外卡条件中的 $x_{\text{首外}}$ 可以取到 b, 即 $x_{\text{首外}} \geq b$.
- **看函数在端点取值**. 如果题目对函数在端点处的行为有特殊要求, 例如, 在闭区间 $[a,b]$ 内恰有 N 个极值点, 则需要额外考虑端点是否可能成为极值点, 这会影响不等式的取舍.

处理边界时, 必须仔细审题, 弄清题中恰好一词的确切含义.

例题 1.56 (单调性约束) 已知函数 $f(x)=\sin(\omega x+\frac{\pi}{4})$ 在区间 $(0,\frac{\pi}{2})$ 上单调递增, 求正实数 ω 的取值范围.

解 这里的条件是对函数在特定区间上单调性的要求. 要使函数在指定区间上单调, 该区间必须是函数某个标准单调区间的子集.

对于正弦函数 $y=\sin t$, 单调递增区间为 $[2k\pi-\frac{\pi}{2},2k\pi+\frac{\pi}{2}],k\in\mathbb{Z}$.

令 $t=\omega x+\frac{\pi}{4}$. 当 $x\in(0,\frac{\pi}{2})$ 时, t 的取值范围为

$$t\in\left(\omega\cdot 0+\frac{\pi}{4},\omega\cdot\frac{\pi}{2}+\frac{\pi}{4}\right)=\left(\frac{\pi}{4},\frac{\omega\pi}{2}+\frac{\pi}{4}\right)$$

要使 $f(x)$ 在 $(0,\frac{\pi}{2})$ 上单调递增, 就必须使区间 $(\frac{\pi}{4},\frac{\omega\pi}{2}+\frac{\pi}{4})$ 是 $y=\sin t$ 某个标准单调递增区间的子集.

由于该区间的左端点为 $\frac{\pi}{4}$, 它自然落在 $k=0$ 时所对应的标准递增区间 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 内部. 因此, 只需保证该区间的右端点不超过标准递增区间的右端点即可.

$$\frac{\omega\pi}{2}+\frac{\pi}{4}\leq\frac{\pi}{2}$$

解此不等式.

$$\frac{\omega\pi}{2}\leq\frac{\pi}{2}-\frac{\pi}{4}=\frac{\pi}{4}$$

$$\omega\leq\frac{1}{2}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

又因为题目要求 ω 为正实数, 即 $\omega > 0$.

ω 的取值范围是 $(0, \frac{1}{2}]$.

例题 1.57 (零点个数与闭区间约束) 已知函数 $f(x) = \cos(\omega x + \frac{\pi}{6})$ 在闭区间 $[0, \pi]$ 上恰有三个零点, 求 ω 的取值范围.

解 这里把考察对象换成零点, 区间也换成闭区间. 端点处若恰有零点, 同样计入总数, 因此边界条件的判断更为严格.

令 $t = \omega x + \frac{\pi}{6}$. 当 $x \in [0, \pi]$ 时, t 的取值范围为

$$t \in \left[\omega \cdot 0 + \frac{\pi}{6}, \omega\pi + \frac{\pi}{6}\right] = \left[\frac{\pi}{6}, \omega\pi + \frac{\pi}{6}\right]$$

问题转化为, 标准余弦函数 $y = \cos t$ 在闭区间 $[\frac{\pi}{6}, \omega\pi + \frac{\pi}{6}]$ 上恰有三个零点.

标准余弦函数 $y = \cos t$ 的零点位于 $t = k\pi + \frac{\pi}{2}, k \in \mathbb{Z}$. 其正值零点依次如下.$\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \frac{7\pi}{2}, \ldots$ 考察区间从 $t = \frac{\pi}{6}$ 开始. 要使其恰好包含三个零点, 这三个零点必须是 $\frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}$.

根据卡位思想, 区间的右端点 $\omega\pi + \frac{\pi}{6}$ 必须大于等于第三个零点 $\frac{5\pi}{2}$, 同时必须严格小于第四个零点 $\frac{7\pi}{2}$.

$$\frac{5\pi}{2} \leq \omega\pi + \frac{\pi}{6} < \frac{7\pi}{2}$$

- 左边界. 当 $\omega\pi + \frac{\pi}{6} = \frac{5\pi}{2}$ 时, 第三个零点恰好落在区间的右端点上. 由于原区间是闭区间, 该点被计入, 满足恰有三个零点的要求, 故可以取等号.
- 右边界. 当 $\omega\pi + \frac{\pi}{6} = \frac{7\pi}{2}$ 时, 第四个零点被包含, 变为四个零点, 不满足题意, 故不能取等号.

解此不等式组.

$$\frac{5\pi}{2} - \frac{\pi}{6} \leq \omega\pi < \frac{7\pi}{2} - \frac{\pi}{6}$$

$$\frac{14\pi}{6} \leq \omega\pi < \frac{20\pi}{6}$$

$$\frac{7\pi}{3} \leq \omega\pi < \frac{10\pi}{3}$$

$$\frac{7}{3} \leq \omega < \frac{10}{3}$$

ω 的取值范围是 $[\frac{7}{3}, \frac{10}{3})$.

知乎@余命数

例题 1.58 (对称性与极值唯一性约束) 已知函数 $f(x)=\sin(\omega x+\frac{\pi}{4})$ $(\omega>0)$ 的图像关于直线 $x=\frac{\pi}{2}$ 对称, 且在区间 $(0,\pi)$ 上有唯一最大值, 求 ω 的值.

解 这里的条件是对称性与极值唯一性的组合, 需要从两个性质出发, 建立关于 ω 的方程和不等式, 并求出它们的交集.

函数图像关于直线 $x=a$ 对称, 意味着 $x=a$ 必为函数的一条对称轴, 即函数在该点取得极值. 因此, $x=\frac{\pi}{2}$ 必须是函数的一个极值点. 令 $t=\omega x+\frac{\pi}{4}$, 则当 $x=\frac{\pi}{2}$ 时, t 必须满足 $t=k\pi+\frac{\pi}{2}, k\in\mathbb{Z}$.

$$\omega\cdot\frac{\pi}{2}+\frac{\pi}{4}=k\pi+\frac{\pi}{2}$$

$$\frac{\omega\pi}{2}=k\pi+\frac{\pi}{4}$$

$$\omega=2k+\frac{1}{2},\quad k\in\mathbb{Z}$$

再利用唯一最大值条件筛选. 当 $x\in(0,\pi)$ 时, 相位

$$t=\omega x+\frac{\pi}{4}$$

的取值区间为

$$\left(\frac{\pi}{4},\omega\pi+\frac{\pi}{4}\right).$$

若 $k\geq 1$, 则 $\omega=2k+\frac{1}{2}\geq\frac{5}{2}$, 于是相位区间至少为

$$\left(\frac{\pi}{4},\frac{11\pi}{4}\right),$$

其中同时包含正弦函数的两个最大值相位 $\frac{\pi}{2}$ 与 $\frac{5\pi}{2}$, 这会使 $f(x)$ 在 $(0,\pi)$ 上至少有两个最大值点, 与题意矛盾. 因此只能有 $k=0$.

结合 $\omega>0, k<0$ 不合题意; 而 $k=0$ 时 $\omega=\frac{1}{2}$, 此时相位区间为 $(\frac{\pi}{4},\frac{3\pi}{4})$, 其中恰含一个最大值相位 $\frac{\pi}{2}$, 符合题意. 因此, 唯一符合所有条件的 ω 值是 $\frac{1}{2}$.

ω 的值为 $\dfrac{1}{2}$.

上面三道变式的切入点不同, 但都遵循同一逻辑. **将函数的性质语言, 翻译为关于参数 ω 的代数约束语言**.

1.15.0.3 思想一. 单调性条件的区间包含

当题目给出的条件是单调性这类贯穿整个区间的性质时, 我们的策略从数点转变为区间包含.

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

1. **确定标准区间**. 写出基础函数, 如 $y=\sin t$ 具备该性质的标准区间通式. 例如, 单调递增区间为 $[2k\pi-\frac{\pi}{2},2k\pi+\frac{\pi}{2}]$.
2. **建立包含关系**. 利用整体换元法, 得到 t 的实际考察区间 $(\omega a+\varphi,\omega b+\varphi)$. 要使原函数在 (a,b) 上具备该性质, 就必须保证 t 的实际区间是某个标准性质区间的**子集**.
3. **转化为不等式**. 将子集关系转化为关于区间端点的不等式. 例如, 要使函数在 $(0,\frac{\pi}{2})$ 上单调递增, 就必须满足.

$$\left(\frac{\pi}{4},\frac{\omega\pi}{2}+\frac{\pi}{4}\right)\subseteq\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$$

这直接引出了关于 ω 的决定性不等式.

1.15.0.4 思想二. 开闭区间对等号的影响

原问题定义域的开闭性, 决定卡位不等式中等号能否取到.
在建立卡位不等式 $x_{\text{尾内}}\le b<x_{\text{首外}}$ 时, 必须仔细审视边界.

- **开区间** (a,b). 特征点若恰好落在端点 b 上, 则不计入区间内. 因此, 第一个区间外的点可以等于 b, 即 $x_{\text{首外}}\ge b$.
- **闭区间** $[a,b]$. 特征点若恰好落在端点 b 上, 则计入区间内. 例如变式二中, 第三个零点可以恰好等于区间的右端点, 从而不等式变为 $\frac{5\pi}{2}\le\omega\pi+\frac{\pi}{6}$.

对边界条件的精确判断, 依赖严格推理.

1.15.0.5 思想三. 多约束条件的分解与交集

当题目给出多个性质条件时, 应将它们分解, 逐一翻译成独立的数学约束, 最后求解这些约束的**交集**.

例如变式三中的问题可分解为两个独立任务.

1. **处理对称性**. 对称性通常对应函数在某点取得极值, 这是一个很强的条件, 往往能导出一个关于 ω 的**方程**, 将 ω 的取值范围从连续区间量化为离散点集, 如 $\omega=4m+\frac{1}{2}$.
2. **处理唯一性**. 极值点的唯一性, 是一个关于函数周期与区间长度关系的几何约束, 它通常导出一个关于 ω 的**不等式**, 为 ω 划定一个连续的取值范围, 如 $\omega\le 2$.
3. **求解交集**. 最后, 寻找在不等式划定的范围内, 满足方程结构的所有解. 这通常意味着在一个有限的离散点集中进行筛选, 从而得到唯一确定的答案.

化整为零, 分别约束, 再求交集, 是处理多条件问题的常用策略.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 1.59 (2022 年 6 月全国乙卷 (理) 第 15 题) 记函数 $f(x)=\cos(\omega x+\varphi)$ $(\omega>0,0<\varphi<\pi)$ 的最小正周期为 T. 若 $f(T)=\frac{\sqrt{3}}{2}$, $x=\frac{\pi}{9}$ 为 $f(x)$ 的零点, 则 ω 的最小值为 _____.

解 我们利用函数的两个特定点的函数值, 反向求解参数 φ 和 ω. 先用周期性确定 φ, 再用零点条件建立关于 ω 的方程, 最后根据 ω 的正性求其最小值.

根据最小正周期的定义, 有 $T=\frac{2\pi}{\omega}$, 故 $\omega T=2\pi$. 将 $x=T$ 代入第一个已知条件 $f(T)=\frac{\sqrt{3}}{2}$ 中.

$$f(T)=\cos(\omega T+\varphi)=\cos(2\pi+\varphi)=\frac{\sqrt{3}}{2}$$

利用余弦函数的周期性, 上式简化为.

$$\cos\varphi=\frac{\sqrt{3}}{2}$$

结合题目给出的范围 $0<\varphi<\pi$, 可唯一确定 φ 的值.

$$\varphi=\frac{\pi}{6}$$

已知 $x=\frac{\pi}{9}$ 是函数 $f(x)$ 的一个零点, 即 $f(\frac{\pi}{9})=0$.

$$f\left(\frac{\pi}{9}\right)=\cos\left(\omega\cdot\frac{\pi}{9}+\varphi\right)=0$$

余弦函数值为零的点, 其相位是 $\frac{\pi}{2}$ 的奇数倍. 因此,

$$\omega\frac{\pi}{9}+\varphi=k\pi+\frac{\pi}{2},\quad k\in\mathbb{Z}$$

将已求得的 $\varphi=\frac{\pi}{6}$ 代入, 得

$$\omega\frac{\pi}{9}+\frac{\pi}{6}=k\pi+\frac{\pi}{2}$$

从上述方程中解出 ω.

$$\omega\frac{\pi}{9}=k\pi+\frac{\pi}{2}-\frac{\pi}{6}=k\pi+\frac{\pi}{3}$$

两边同乘以 $\frac{9}{\pi}$, 得到 ω 关于整数 k 的表达式.

$$\omega=9k+3$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

题目要求 $\omega>0$, 因此.

$$9k+3>0 \implies 9k>-3 \implies k>-\frac{1}{3}$$

由于 k 必须是整数, 满足此条件的最小整数 k 为 0. 将 $k=0$ 代入, 即可得到 ω 的最小正值.

$$\omega_{\min}=9(0)+3=3$$

因此, ω 的最小值为 3.

例题 1.60 (2024 年 1 月浙江省宁波市宁波九校第 15 题) 将函数 $f(x)=\sin(\omega x+\frac{\pi}{6})$ $(\omega>0)$ 的图像向左平移 $\frac{\pi}{3}$ 个单位长度后得到函数 $g(x)$ 的图像, 且 $g(x)$ 在区间 $[\frac{\pi}{6},\frac{2\pi}{3}]$ 上单调递增, 则 ω 的取值范围是 _____.

知乎@余命数

解 我们先进行图像平移, 得到 $g(x)$ 的解析式, 再利用单调性条件建立不等式.

将函数 $f(x)$ 的图像向左平移 $\frac{\pi}{3}$ 个单位, 意味着用 $x+\frac{\pi}{3}$ 替换原解析式中的 x.

$$g(x)=f\left(x+\frac{\pi}{3}\right)=\sin\left[\omega\left(x+\frac{\pi}{3}\right)+\frac{\pi}{6}\right]=\sin\left(\omega x+\frac{\omega\pi}{3}+\frac{\pi}{6}\right)$$

问题转化为, 函数 $g(x)=\sin\left(\omega x+\frac{\omega\pi}{3}+\frac{\pi}{6}\right)$ 在区间 $[\frac{\pi}{6},\frac{2\pi}{3}]$ 上单调递增. 令 $t=\omega x+\frac{\omega\pi}{3}+\frac{\pi}{6}$. 要使 $g(x)$ 单调递增, 其内层函数 t 的取值区间必须是标准正弦函数 $y=\sin t$ 某个单调递增区间的子集.

先确定当 $x\in[\frac{\pi}{6},\frac{2\pi}{3}]$ 时, t 的取值范围.

- 当 $x=\frac{\pi}{6}$ 时(左端点), $t_{\min}=\omega\left(\frac{\pi}{6}\right)+\frac{\omega\pi}{3}+\frac{\pi}{6}=\frac{\omega\pi}{2}+\frac{\pi}{6}$.
- 当 $x=\frac{2\pi}{3}$ 时(右端点), $t_{\max}=\omega\left(\frac{2\pi}{3}\right)+\frac{\omega\pi}{3}+\frac{\pi}{6}=\omega\pi+\frac{\pi}{6}$.

因此, t 的取值区间为 $[\frac{\omega\pi}{2}+\frac{\pi}{6},\omega\pi+\frac{\pi}{6}]$.

标准正弦函数 $y=\sin t$ 的单调递增区间为 $[2k\pi-\frac{\pi}{2},2k\pi+\frac{\pi}{2}]$, $k\in\mathbb{Z}$. 由于 $\omega>0$, t 的区间下限 $\frac{\omega\pi}{2}+\frac{\pi}{6}$ 恒为正. 因此, 只需考虑 $k\geq 0$ 的情况.

当 $k=0$ 时, 标准递增区间为 $[-\frac{\pi}{2},\frac{\pi}{2}]$. t 的区间必须是该标准区间的子集, 得.

知乎@余命数

$$\begin{cases}\frac{\omega\pi}{2}+\frac{\pi}{6}\geq-\frac{\pi}{2} & (\text{此式对}\omega>0\text{ 恒成立})\\ \omega\pi+\frac{\pi}{6}\leq\frac{\pi}{2}\end{cases}$$

解第二个不等式.

$$\omega\pi\leq\frac{\pi}{2}-\frac{\pi}{6}=\frac{\pi}{3} \implies \omega\leq\frac{1}{3}$$

还需检验是否存在 $k \geq 1$ 的解. 当 $k=1$ 时, 标准递增区间为 $[\frac{3\pi}{2}, \frac{5\pi}{2}]$. 此时需满足.

$$\begin{cases} \frac{\omega\pi}{2} + \frac{\pi}{6} \geq \frac{3\pi}{2} \implies \frac{\omega\pi}{2} \geq \frac{4\pi}{3} \implies \omega \geq \frac{8}{3} \\ \omega\pi + \frac{\pi}{6} \leq \frac{5\pi}{2} \implies \omega\pi \leq \frac{7\pi}{3} \implies \omega \leq \frac{7}{3} \end{cases}$$

$\omega \geq \frac{8}{3}$ 与 $\omega \leq \frac{7}{3}$ 矛盾, 故无解. 对于更大的 k 值, 同样无解.

结合 $\omega > 0$ 的前提条件, 最终的取值范围为 $0 < \omega \leq \frac{1}{3}$.

因此, ω 的取值范围是 $\left(0, \frac{1}{3}\right]$.

例题 1.61 (变式四) 已知函数 $f(x) = \cos(\omega x + \frac{\pi}{3})$ $(\omega > 0)$ 在区间 $[0, \frac{\pi}{2}]$ 上的值域为 $[-\frac{1}{2}, \frac{1}{2}]$, 求 ω 的取值范围.

解 这里的约束条件是函数在特定区间上的值域. 我们先分析内层函数 $t = \omega x + \frac{\pi}{3}$ 的取值区间, 再考察标准余弦函数 $y = \cos t$ 在该区间上的最值行为.

令 $t = \omega x + \frac{\pi}{3}$. 当 $x \in [0, \frac{\pi}{2}]$ 时, t 的取值范围为

$$t \in \left[\omega \cdot 0 + \frac{\pi}{3}, \omega \cdot \frac{\pi}{2} + \frac{\pi}{3}\right] = \left[\frac{\pi}{3}, \frac{\omega\pi}{2} + \frac{\pi}{3}\right]$$

问题转化为, 函数 $y = \cos t$ 在区间 $[\frac{\pi}{3}, \frac{\omega\pi}{2} + \frac{\pi}{3}]$ 上的值域为 $[-\frac{1}{2}, \frac{1}{2}]$.

由于 $\omega > 0$, 该区间的右端点大于左端点. 在 $[\frac{\pi}{3}, \pi]$ 上, $\cos t$ 单调递减. 左端点处函数值已经等于 $\frac{1}{2}$, 因而值域的最大值必为 $\frac{1}{2}$. 要使值域的最小值恰好为 $-\frac{1}{2}$, 区间右端点必须恰好到达第一个满足 $\cos t = -\frac{1}{2}$ 的点 $t = \frac{2\pi}{3}$; 若未到达, 最小值大于 $-\frac{1}{2}$, 若越过, 最小值小于 $-\frac{1}{2}$. 因此

$$\frac{\omega\pi}{2} + \frac{\pi}{3} = \frac{2\pi}{3}.$$

解得

$$\omega = \frac{2}{3}.$$

ω 的取值范围为 $\{\frac{2}{3}\}$.

例题 1.62 (变式五) 已知函数 $f(x) = \sin(\omega x)$ $(\omega > 0)$ 在区间 $[0, \pi]$ 上至少有一个最大值点, 但不存在最小值点, 求 ω 的取值范围.

解 这里的约束条件是至少有一个最大值点、同时不存在最小值点. 我们仍然用卡位思想处理.

对于标准正弦函数 $y = \sin t$,

- 其正的最大值点位于 $t = 2k\pi + \frac{\pi}{2}, k \in \mathbb{Z}$. 第一个正的最大值点是 $\frac{\pi}{2}$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 其正的最小值点位于 $t = 2k\pi + \frac{3\pi}{2}, k \in \mathbb{Z}$. 第一个正的最小值点是 $\frac{3\pi}{2}$.

令 $t = \omega x$. 当 $x \in [0, \pi]$ 时, t 的取值范围是 $[0, \omega\pi]$. 问题转化为, 区间 $[0, \omega\pi]$ 需满足.

1. **至少包含一个最大值点**. 即区间的右端点必须大于等于第一个最大值点的位置.

$$\omega\pi \geq \frac{\pi}{2}$$

2. **不包含任何最小值点**. 即区间的右端点必须严格小于第一个最小值点的位置.

$$\omega\pi < \frac{3\pi}{2}$$

解第一个不等式.

$$\omega \geq \frac{1}{2}$$

解第二个不等式.

$$\omega < \frac{3}{2}$$

将两个条件合并, ω 的取值范围是 $[\frac{1}{2}, \frac{3}{2})$.

切线斜率为负的导数表述

例题 1.63 (变式六) 已知函数 $f(x) = \sin(\omega x - \frac{\pi}{3})$ $(\omega > 0)$ 满足对任意 $x \in (0, \pi)$, 都有 $f(x) + f(\pi - x) = 0$. 若 $f(x)$ 在点 $(\frac{\pi}{2}, f(\frac{\pi}{2}))$ 处的切线斜率为负, 求 ω 的最小值.

解 第一个条件给出隐式的对称性, 第二个条件约束函数在该对称中心处的变化趋势. 我们先把几何语言翻译为代数条件, 再结合导数信息求解.

条件 $f(x) + f(\pi - x) = 0$ 对任意 $x \in (0, \pi)$ 成立, 描述的是函数图像关于点 $(\frac{\pi}{2}, 0)$ 中心对称. 要使点 $(\frac{\pi}{2}, 0)$ 成为 $y = \sin t$ 型函数的对称中心, 该点对应的相位 t 必须是 π 的整数倍. 令 $t = \omega x - \frac{\pi}{3}$. 将 $x = \frac{\pi}{2}$ 代入, 得

$$\omega \cdot \frac{\pi}{2} - \frac{\pi}{3} = k\pi, \quad k \in \mathbb{Z}$$

解出 ω 关于 k 的表达式.

$$\frac{\omega\pi}{2} = k\pi + \frac{\pi}{3} \implies \omega = 2k + \frac{2}{3}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

函数在点 $x=\frac{\pi}{2}$ 处的切线斜率为负, 即 $f'(\frac{\pi}{2})<0$. 首先求导数.

$$f'(x)=\omega\cos\left(\omega x-\frac{\pi}{3}\right)$$

将 $x=\frac{\pi}{2}$ 代入, 得

$$f'\left(\frac{\pi}{2}\right)=\omega\cos\left(\omega\frac{\pi}{2}-\frac{\pi}{3}\right)<0$$

由前面对对称中心的分析, $\omega\frac{\pi}{2}-\frac{\pi}{3}=k\pi$. 代入上式, 得

$$\omega\cos(k\pi)<0$$

因为已知 $\omega>0$, 所以必须满足 $\cos(k\pi)<0$. 这当且仅当 k 为**奇数**时成立.

现在有两个关于 ω 的条件.

1. $\omega=2k+\frac{2}{3}$
2. k 是奇数
3. $\omega>0$

要使 $\omega>0$, 则 $2k+\frac{2}{3}>0\Longrightarrow k>-\frac{1}{3}$. 结合 k 为奇数, 满足条件的 k 的最小值为 1. 当 $k=1$ 时, 得 ω 的最小值.

$$\omega_{\min}=2(1)+\frac{2}{3}=\frac{8}{3}$$

因此, ω 的最小值为 $\frac{8}{3}$.

处理此类参数约束题时, 应将题设条件逐一转化为相位方程或不等式, 并充分利用所有约束信息.

例题 1.64 (2023 年 12 月重庆南开中学校高三第四次期中质检) 已知函数 $f(x)=\sin(\omega x+\varphi)$ $(\omega>0, 0<\varphi<\frac{\pi}{2})$, 曲线 $y=f(x)$ 的一个对称中心为 $(\frac{\pi}{12},0)$, 一条对称轴为 $x=\frac{\pi}{4}$, 则 ω 的最小值为 _____.

解 我们用一个对称中心与一条对称轴约束 ω 与 φ, 可以从代数方程与几何关系两条路推进.

法一采用代数法, 把几何性质写成相位方程后消元. 对称中心 $(\frac{\pi}{12},0)$ 要求相位为 π 的整数倍.

$$\omega\cdot\frac{\pi}{12}+\varphi=k_1\pi,\quad k_1\in\mathbb{Z}. \tag{1-11}$$

对称轴 $x=\frac{\pi}{4}$ 要求相位为 $\frac{\pi}{2}$ 的奇数倍:

$$\omega\cdot\frac{\pi}{4}+\varphi=k_2\pi+\frac{\pi}{2},\quad k_2\in\mathbb{Z}. \tag{1-12}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

(1-12) 减去 (1-11) 消去 φ.

$$\omega\left(\frac{\pi}{4}-\frac{\pi}{12}\right)=(k_2-k_1)\pi+\frac{\pi}{2}\implies\omega=6(k_2-k_1)+3.$$

再消 ω 以求 φ: 将 (1-11) 乘以 3 得 $\omega\frac{\pi}{4}+3\varphi=3k_1\pi$, 减去 (1-12),

$$2\varphi=\left(3k_1-k_2-\frac{1}{2}\right)\pi\implies\varphi=\frac{3k_1-k_2-1/2}{2}\pi.$$

由 $0<\varphi<\frac{\pi}{2}$ 得 $\frac{1}{2}<3k_1-k_2<\frac{3}{2}$. k_1,k_2 为整数, 故 $3k_1-k_2$ 为整数, 唯一可能是 1, 即 $k_2=3k_1-1$. 代回 ω 的式子: $\omega=12k_1-3$. 由 $\omega>0$ 得 $k_1>\frac{1}{4}$, 最小整数 $k_1=1$, 此时

$$\omega_{\min}=9.$$

法二采用几何法. 对称中心与对称轴的水平距离必为四分之一周期的奇数倍. 取 $d=\left|\frac{\pi}{4}-\frac{\pi}{12}\right|=\frac{\pi}{6}$, 周期 $T=\frac{2\pi}{\omega}$, 则

$$\frac{\pi}{6}=\frac{2n+1}{4}T=\frac{(2n+1)\pi}{2\omega},\quad n\in\{0,1,2,\dots\},$$

解得 $\omega=3(2n+1)$, 候选值 $\omega\in\{3,9,15,\dots\}$.

对每个候选值, 由对称中心条件 $\varphi=k\pi-\omega\frac{\pi}{12}$ 检验是否存在 $\varphi\in(0,\frac{\pi}{2})$. 当 $\omega=3$ 时 $\varphi=k\pi-\frac{\pi}{4}$, 对一切整数 k 均不在 $(0,\frac{\pi}{2})$ 内, 舍去. 当 $\omega=9$ 时, $k=1$ 给出 $\varphi=\frac{\pi}{4}\in(0,\frac{\pi}{2})$; 再验对称轴:

$$9\cdot\frac{\pi}{4}+\frac{\pi}{4}=\frac{5\pi}{2}$$

恰为 $\frac{\pi}{2}$ 的奇数倍, 故 $\omega=9$ 有效. 候选值随 n 递增, 第一个有效值即最小值.

两种方法均得 ω 的最小值为 9.

例题 1.65 (变式七) 已知函数 $f(x)=\cos(\omega x+\varphi)$ $(\omega>0,0<\varphi<\frac{\pi}{2})$ 的图像关于直线 $x=\frac{\pi}{3}$ 和 $x=\frac{5\pi}{6}$ 均对称, 求 ω 的最小值.

解 这里给出两条对称轴. 由对称轴可得到函数在特定点的相位信息; 两条对称轴之间的距离又与周期有确定的几何关系. 函数图像的两条对称轴之间的距离, 必定是半周期 $T/2$ 的整数倍. 两条对称轴的距离为.

$$d=\left|\frac{5\pi}{6}-\frac{\pi}{3}\right|=\frac{\pi}{2}$$

周期 $T=\frac{2\pi}{\omega}$. 因此有.

$$d=n\cdot\frac{T}{2},\quad n\in\{1,2,3,\dots\}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$\frac{\pi}{2} = n \cdot \frac{1}{2} \cdot \frac{2\pi}{\omega} = \frac{n\pi}{\omega}$$

解出 ω 关于正整数 n 的表达式.

$$\omega = 2n, \quad n \in \{1, 2, 3, \dots\}$$

这给出了 ω 的一个离散候选值集 $\{2, 4, 6, 8, \dots\}$. 还需要利用 φ 的范围来筛选出符合条件的最小 n 值.

因为 $x = \frac{\pi}{3}$ 是一条对称轴, 所以函数在该点的相位必须是 π 的整数倍 (对于余弦函数).

知乎@余命数

$$\omega \cdot \frac{\pi}{3} + \varphi = k\pi, \quad k \in \mathbb{Z}$$

$$\varphi = k\pi - \omega\frac{\pi}{3}$$

将 $\omega = 2n$ 代入, 得

$$\varphi = k\pi - \frac{2n\pi}{3} = \left(k - \frac{2n}{3}\right)\pi$$

知乎@余命数

根据题设 $0 < \varphi < \frac{\pi}{2}$, 有

$$0 < \left(k - \frac{2n}{3}\right)\pi < \frac{\pi}{2} \implies 0 < k - \frac{2n}{3} < \frac{1}{2}$$

需寻找最小的正整数 n, 使得存在整数 k 满足此不等式.

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

- 当 $n = 1$ 时, $0 < k - \frac{2}{3} < \frac{1}{2} \implies \frac{2}{3} < k < \frac{7}{6}$. 此时整数 $k = 1$ 满足条件.

因此, 满足条件的最小正整数 n 是 1. 对应的 ω 最小值为 $\omega_{\min} = 2(1) = 2$.

ω 的最小值为 2.

例题 1.66 (变式八) 已知函数 $f(x) = \cos(\omega x + \varphi)$ $(\omega > 0, 0 < \varphi < \pi)$. 若将 $y = f(x)$ 的图像向右平移 $\frac{\pi}{12}$ 个单位后得到的函数 $g(x)$ 为偶函数, 且 $f(\frac{\pi}{6}) = 0$, 求 ω 的最小值.

解 这里把对称性条件施加在一个变换后的函数上. 我们先进行图像变换的代数运算, 再将函数的性质转化为关于参数的方程组. 函数 $g(x) = f(x - \frac{\pi}{12}) = \cos\left(\omega(x - \frac{\pi}{12}) + \varphi\right) = \cos\left(\omega x - \frac{\omega\pi}{12} + \varphi\right)$. $g(x)$ 为偶函数, 意味着其图像关于 y 轴对称. 对于余弦型函数, 这要求其相位中的常数项(不含 x 的部分)为 π 的整数倍.

$$-\frac{\omega\pi}{12} + \varphi = k\pi, \quad k \in \mathbb{Z} \tag{1-13}$$

知乎@余命数 知乎@余命数

$f(\frac{\pi}{6})=0$ 意味着函数在 $x=\frac{\pi}{6}$ 处的相位是 $\frac{\pi}{2}$ 的奇数倍.

$$\omega\frac{\pi}{6}+\varphi=m\pi+\frac{\pi}{2},\quad m\in\mathbb{Z} \tag{1-14}$$

由方程 (1-13) 可得 $\varphi=k\pi+\frac{\omega\pi}{12}$. 将其代入方程 (1-14):

$$\omega\frac{\pi}{6}+\left(k\pi+\frac{\omega\pi}{12}\right)=m\pi+\frac{\pi}{2}$$

$$\omega\left(\frac{\pi}{6}+\frac{\pi}{12}\right)=(m-k)\pi+\frac{\pi}{2}$$

$$\omega\frac{\pi}{4}=(m-k)\pi+\frac{\pi}{2}$$

令 $n=m-k$, n 仍为整数.

$$\omega=4n+2,\quad n\in\mathbb{Z}$$

将 $\omega=4n+2$ 代回 φ 的表达式.

$$\varphi=k\pi+\frac{(4n+2)\pi}{12}=k\pi+\frac{(2n+1)\pi}{6}$$

根据题设 $0<\varphi<\pi$, 有, 得

$$0<k\pi+\frac{(2n+1)\pi}{6}<\pi\implies 0<k+\frac{2n+1}{6}<1$$

$$-\frac{2n+1}{6}<k<1-\frac{2n+1}{6}=\frac{5-2n}{6}$$

需寻找最小的整数 n(使得 $\omega>0$), 使得上述关于 k 的不等式区间内存在整数解.

- 因为 $\omega=4n+2>0$, 所以 $n>-1/2$. 故 n 的最小整数取值为 0.
- 当 $n=0$ 时, 不等式变为 $-\frac{1}{6}<k<\frac{5}{6}$. 此时存在整数解 $k=0$.

因此, 使得 ω 取最小值的整数 n 为 0.

$$\omega_{\min}=4(0)+2=2$$

ω 的最小值为 2.

例题 1.67 (变式九) 已知函数 $f(x)=\sin(\omega x+\varphi)$ $(\omega>0,0<\varphi<\frac{\pi}{2})$ 满足对任意 $x\in\mathbb{R}$, 都有 $f(\frac{\pi}{3}+x)=f(\frac{\pi}{3}-x)$, 且函数在区间 $(0,\frac{\pi}{2})$ 内恰有两个零点, 求 ω 的取值范围.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

解 这里将一个明确的对称轴条件与一个零点数量的卡位条件相结合, 需要建立一个关于 ω 和整数参数 k 的不等式系统来求解.

$f(\frac{\pi}{3}+x)=f(\frac{\pi}{3}-x)$ 表明, 直线 $x=\frac{\pi}{3}$ 是函数的一条对称轴. 因此, 函数在 $x=\frac{\pi}{3}$ 处的相位必须是 $\frac{\pi}{2}$ 的奇数倍.

$$\omega\frac{\pi}{3}+\varphi=k\pi+\frac{\pi}{2},\quad k\in\mathbb{Z} \tag{1-15}$$

函数在 $(0,\frac{\pi}{2})$ 内恰有两个零点. 令 $t=\omega x+\varphi$. 当 $x\in(0,\frac{\pi}{2})$ 时, t 的取值范围是 $(\varphi,\frac{\omega\pi}{2}+\varphi)$. 函数的零点对应 $t=m\pi, m\in\mathbb{Z}$. 由于 $0<\varphi<\frac{\pi}{2}$, t 的区间左端点 φ 在 $(0,\frac{\pi}{2})$ 内. 因此, 区间内可能包含的第一个零点是 $t=\pi$. 要恰好包含两个零点, 这两个零点必须是 π 和 2π. 根据卡位思想, 区间 $(\varphi,\frac{\omega\pi}{2}+\varphi)$ 必须包含 2π, 但不包含 3π.

$$2\pi<\frac{\omega\pi}{2}+\varphi\le 3\pi\quad\cdots(*)$$

现在有一个方程和一个不等式, 它们共同约束着 ω 和 φ. 从方程 (1-15) 中解出 φ:

$$\varphi=k\pi+\frac{\pi}{2}-\frac{\omega\pi}{3}$$

利用 φ 的范围 $0<\varphi<\frac{\pi}{2}$:

$$0<k\pi+\frac{\pi}{2}-\frac{\omega\pi}{3}<\frac{\pi}{2}\implies -\frac{\pi}{2}<k\pi-\frac{\omega\pi}{3}<0\implies -\frac{1}{2}<k-\frac{\omega}{3}<0$$

$$k<\frac{\omega}{3}<k+\frac{1}{2} \tag{1-16}$$

这个关系表明, 对于一个给定的 ω, 整数 k 的取值是唯一确定的.

接着, 将 φ 的表达式代入不等式 $(*)$:

$$2\pi<\frac{\omega\pi}{2}+\left(k\pi+\frac{\pi}{2}-\frac{\omega\pi}{3}\right)\le 3\pi$$

$$2\pi<\frac{\omega\pi}{6}+k\pi+\frac{\pi}{2}\le 3\pi$$

$$\frac{3\pi}{2}<\frac{\omega\pi}{6}+k\pi\le\frac{5\pi}{2}$$

$$\frac{3}{2}-k<\frac{\omega}{6}\le\frac{5}{2}-k\implies 9-6k<\omega\le 15-6k$$

现在需要在 $\omega>0$ 的前提下, 寻找存在整数 k 使得不等式 $9-6k<\omega\le 15-6k$ 和不等式 (1-16) $(3k<\omega<3k+1.5)$ 同时成立.

为使两个关于 ω 的区间有交集, 区间的端点必须满足.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- $3k < 15 - 6k \implies 9k < 15 \implies k < \frac{5}{3}$
- $9 - 6k < 3k + 1.5 \implies 7.5 < 9k \implies k > \frac{5}{6}$

结合 k 为整数, 唯一可能的取值是 $k = 1$.

当 $k = 1$ 时, 将两个区间具体化.

- 由 (1-16) 得: $3(1) < \omega < 3(1) + 1.5 \implies 3 < \omega < 4.5$.
- 由主不等式得: $9 - 6(1) < \omega \leq 15 - 6(1) \implies 3 < \omega \leq 9$.

取这两个区间的交集, 得到 ω 的最终取值范围.

因此, ω 的取值范围是 $(3, \frac{9}{2})$.

通过对例题及其变式题的剖析, 我们总结出处理复杂 ω 约束问题的几点原则.

1.15.0.6 原则一. 对称性的两种处理

对称性是三角函数最重要的性质之一, 也是 ω 约束问题的高频考点.

知乎@余命数

1.15.0.6.1 1. 代数法 将几何性质直接转化为代数方程.

- **对称轴** $x = x_0$ 当且仅当相位 $\omega x_0 + \varphi = k\pi + \frac{\pi}{2}$ 对 sin 成立, 或 $k\pi$ 对 cos 成立.
- **对称中心** $(x_0, 0)$ 当且仅当相位 $\omega x_0 + \varphi = k\pi$(对 sin)或 $k\pi + \frac{\pi}{2}$ (对 cos.
- **偶函数**当且仅当初相 $\varphi = k\pi$(对 cos)或 $k\pi + \frac{\pi}{2}$ (对 sin.
- **隐式对称** $f(a + x) = f(a - x)$ 当且仅当对称轴为 $x = a$.
- **隐式对称** $f(a + x) = -f(a - x)$ 当且仅当对称中心为 $(a, 0)$.

1.15.0.6.1.1 说明 代数翻译是最直接的方法. 它将几何性质转化为关于 ω, φ 和整数 k, m 的不定方程组. 当涉及多个整数参数时, 需要细致地讨论整数解的存在性.

知乎@余命数

1.15.0.6.2 2. 几何法 利用对称点之间的几何距离关系.

- 两条**同类**对称点(轴与轴, 中心与中心)的距离是**半周期**的整数倍.
- 两条**异类**对称点(轴与中心)的距离是**四分之一周期**的奇数倍.

1.15.0.6.2.1 说明 几何法在题目给出两个对称点时, 能迅速建立一个仅关于 ω 和一个整数参数 n 的方程, 简化计算. 但候选集必须代回原条件, 结合 φ 的范围进行检验, 以排除不符合初相条件的伪解.

1.15.0.7 原则二. 多约束条件的分解

当题目给出多个不同类型的约束条件时, 如例 20 及其变式, 分而治之是有效的策略.

1. **分解**. 将复杂的题目描述分解为若干个独立的性质陈述. 例如, 平移后为偶函数且某点为零点.
2. **翻译**. 对每一个性质陈述, 选择最恰当的方法(代数或几何), 将其翻译为关于参数的方程或不等式约束.
3. **整合**. 联立求解所有约束. 这个过程通常演变为.
   - 方程与方程联立, 通过消元法求解.
   - 方程与不等式联立, 将由方程解出的参数表达式代入不等式, 从而约束整数参数的取值.
   - 不等式与不等式联立, 求解不等式组的交集.

知乎@余命数

1.15.0.7.0.1 说明 这种分治-整合的策略将复杂问题拆解为若干子问题, 降低出错风险. 关键在于准确地翻译每一个条件, 并在最后进行严谨的逻辑整合.

知乎@余命数

1.15.0.8 原则三. 存在性与唯一性

从至少或至多有一个, 到恰好有一个, 问题的逻辑复杂度在提升.

- **存在性**(至少/至多)通常对应一个单边的不等式. 例如, 至少一个最大值点意味着考察区间的右端点必须越过第一个最大值点, 即 $\omega\pi \geq \frac{\pi}{2}$.
- **唯一性**(恰好一个)通常对应一个双边不等式, 即卡位思想. 例如, 唯一最大值意味着区间必须包含第一个最大值点, 同时排除第二个最大值点和第一个最小值点.

知乎@余命数

1.15.0.8.0.1 说明 处理这类逻辑词汇时, 要把它精确地转化为区间端点的开闭与大小关系. 这既考验计算, 也考验对数学语言精确性的把握.

- 对于**单一性质**(单调性、值域、零点个数), **整体换元法**通常是最稳健的选择.
- 对于**两个或以上明确的对称点**, **几何距离法**是首选的捷径, 但务必进行后续检验.
- 对于**复合的、混合的性质约束**, **建立方程组(代数法)**, 并采用分治-整合的策略, 是最可靠的方法.

知乎@余命数

ω 约束问题的每种性质都有明确的几何意义, 需要在代数表示与几何意义之间切换.

1.15.1 高考真题选讲

例题 1.68 (2008 年辽宁卷理科数学第 16 题) 已知 $f(x)=\sin\left(\omega x+\dfrac{\pi}{3}\right)$ $(\omega>0)$, $f\left(\dfrac{\pi}{6}\right)=f\left(\dfrac{\pi}{3}\right)$, 且 $f(x)$ 在区间 $\left(\dfrac{\pi}{6},\dfrac{\pi}{3}\right)$ 上有最小值无最大值, 则 $\omega=$_____.

解 答案: $\dfrac{14}{3}$.

由 $f\left(\dfrac{\pi}{6}\right)=f\left(\dfrac{\pi}{3}\right)$ 知, $x=\dfrac{\pi}{4}$, 即 $\dfrac{\pi}{6}$ 与 $\dfrac{\pi}{3}$ 的中点是一条对称轴. 又因 $f(x)$ 在 $\left(\dfrac{\pi}{6},\dfrac{\pi}{3}\right)$ 上有最小值无最大值, 该最小值在 $x=\dfrac{\pi}{4}$ 处取得, 故

$$\omega\cdot\frac{\pi}{4}+\frac{\pi}{3}=2k\pi-\frac{\pi}{2},\quad k\in\mathbb{Z},$$

解得 $\omega=\dfrac{24k-10}{3}$.

由 $\omega>0$ 知 $k\geqslant 1$. 当 $k=1$ 时, $\omega=\dfrac{14}{3}$, 半周期 $\dfrac{\pi}{\omega}=\dfrac{3\pi}{14}$ 大于区间长度 $\dfrac{\pi}{6}$, 故区间内恰含一个最小值点而不含最大值点, 符合题意. 当 $k=2$ 时, $\omega=\dfrac{38}{3}$, 半周期 $\dfrac{3\pi}{38}$ 小于区间长度 $\dfrac{\pi}{6}$, 区间内已出现最大值, 不合题意.

故 $\omega=\dfrac{14}{3}$.

例题 1.69 (2018 年北京卷理科数学第 11 题) 设函数 $f(x)=\cos\left(\omega x-\dfrac{\pi}{6}\right)$ $(\omega>0)$, 若 $f(x)\leqslant f\left(\dfrac{\pi}{4}\right)$ 对任意实数 x 都成立, 则 ω 的最小值为_____.

解 答案: $\dfrac{2}{3}$.

$f(x)\leqslant f\left(\dfrac{\pi}{4}\right)$ 对任意实数 x 成立, 说明 $x=\dfrac{\pi}{4}$ 是最大值点. 对余弦型函数, 最大值点处的相位等于 $2k\pi$:

$$\omega\cdot\frac{\pi}{4}-\frac{\pi}{6}=2k\pi,\quad k\in\mathbb{Z},$$

解得 $\omega=8k+\dfrac{2}{3}$.

由 $\omega>0$ 知 $k\geqslant 0$. 当 $k=0$ 时, $\omega=\dfrac{2}{3}$, 即为所求最小值.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

例题 1.70 (2018 年江苏卷第 7 题) 已知函数 $y=\sin(2x+\varphi)\left(-\dfrac{\pi}{2}<\varphi<\dfrac{\pi}{2}\right)$ 的图象关于直线 $x=\dfrac{\pi}{3}$ 对称, 则 φ 的值为_____.

解 答案: $-\dfrac{\pi}{6}$.

正弦型函数的对称轴处, 相位等于 $k\pi+\dfrac{\pi}{2}$:

$$2\cdot\frac{\pi}{3}+\varphi=k\pi+\frac{\pi}{2},\quad k\in\mathbb{Z},$$

解得 $\varphi=k\pi-\dfrac{\pi}{6}$.

由 $-\dfrac{\pi}{2}<\varphi<\dfrac{\pi}{2}$ 知 $k=0$, 故 $\varphi=-\dfrac{\pi}{6}$.

例题 1.71 (2013 年山东卷理科数学第 5 题) 函数 $y=\sin(2x+\varphi)$ 的图象沿 x 轴向左平移 $\dfrac{\pi}{8}$ 个单位后, 得到一个偶函数的图象, 则 φ 的一个可能的值为

A. $\dfrac{3\pi}{4}$　　C. 0

B. $\dfrac{\pi}{4}$　　D. $-\dfrac{\pi}{4}$

解 答案: B.

向左平移 $\dfrac{\pi}{8}$ 个单位后, 函数变为

$$y=\sin\left[2\left(x+\frac{\pi}{8}\right)+\varphi\right]=\sin\left(2x+\frac{\pi}{4}+\varphi\right).$$

该函数为偶函数, 故初相 $\dfrac{\pi}{4}+\varphi=k\pi+\dfrac{\pi}{2}$, $k\in\mathbb{Z}$, 即 $\varphi=k\pi+\dfrac{\pi}{4}$.

当 $k=0$ 时, $\varphi=\dfrac{\pi}{4}$, 对应选项 B.

1.16 反三角函数

函数 $y=f(x)$ 存在反函数 $x=f^{-1}(y)$ 的充要条件是 f 为一一映射. 三角函数由于周期性都是多对一, 例如 $\frac{\pi}{6},\frac{5\pi}{6},\frac{13\pi}{6},\ldots$ 都对应同一个正弦值 $\frac{1}{2}$.

因此不能直接把整个三角函数反转, 否则一个输入值会对应无穷多个输出角, 不符合函数定义.

解决方法是限制定义域, 截取一段单调且值域完整的分支. 这段分支需满足两个条件.

1. **完备性**. 函数值恰好覆盖原函数的全部值域. 例如 $y=\sin x$ 在选定区间上的值域必须是 $[-1,1]$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2. **单调性**. 函数在该区间上严格单调,从而是一一映射,保证反函数存在且唯一.

满足上述条件的区间称为**主值区间**. 在其上构造的反函数称为**反三角函数**. 它要回答的问题是, 在主值区间内哪个角的三角函数值等于给定值.

1.16.1 反正弦函数

1.16.1.1 定义与主值区间的选取

考察函数 $y=\sin x$. 为了定义其反函数, 需选取一个单调区间, 使得该区间上的函数值域覆盖 $[-1,1]$.

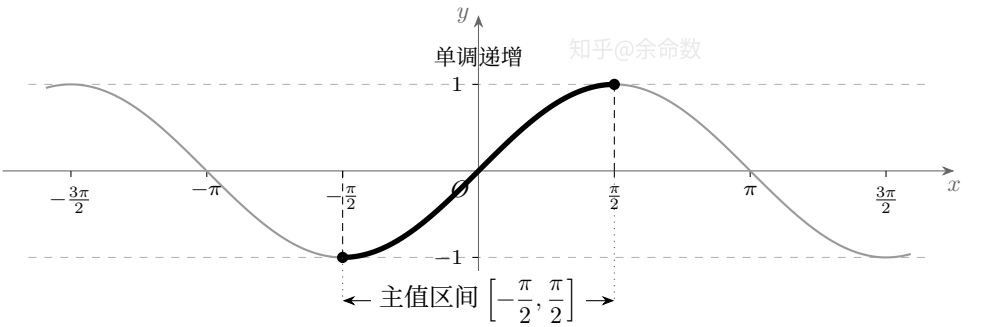


图 1.23　正弦函数主值区间 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 的选取

观察图像可知, 区间 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 是一个理想的选择. 它同时满足满射性(值域恰好覆盖 $[-1,1]$)、严格单调递增, 以及关于原点对称并包含锐角范围的对称性, 因而便于计算与记忆. 虽然 $[\frac{\pi}{2},\frac{3\pi}{2}]$ 等区间也满足单调性和值域条件, 但数学界统一规定选取 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 作为反正弦函数的主值区间.

定义 1.5 (反正弦) 函数 $y=\sin x$ 在其主值区间 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 上的反函数, 称为**反正弦函数**, 记作 $y=\arcsin x$.

- **定义域**: $[-1,1]$
- **值域**: $[-\frac{\pi}{2},\frac{\pi}{2}]$

符号 $\arcsin x$ 表达的意义是. 一个在 $[-\frac{\pi}{2},\frac{\pi}{2}]$ 内的角, 其正弦值为 x.

1.16.1.2 性质与图像

- **单调性**: 正弦函数在主值区间的单调递增性, 决定了反正弦函数在其整个定义域 $[-1,1]$ 上也是**单调递增**的.
- **奇偶性**: 正弦函数主值支的图像关于原点对称, 这一几何特性被反函数继承. 故 $y=\arcsin x$ 是**奇函数**, 满足 $\arcsin(-x)=-\arcsin x$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

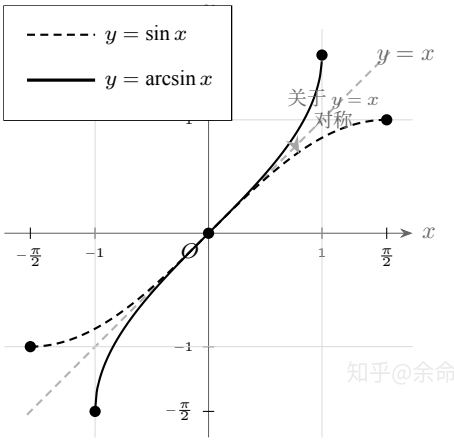
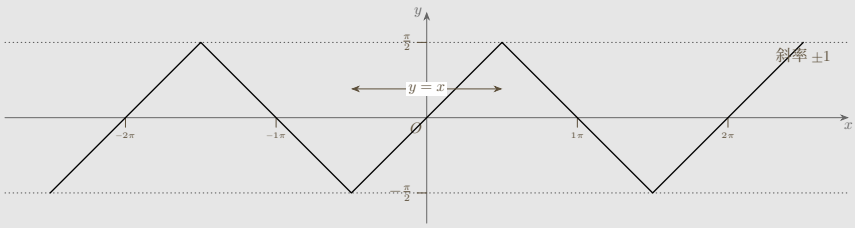


图 1.24 反正弦函数与正弦函数主值支的对称关系

注 1.17 ($\sin(\arcsin x)$ 与 $\arcsin(\sin x)$)
根据反函数的定义, $\sin(\arcsin x) = x$ 与 $\arcsin(\sin x) = x$ 两个式子看似相同, 但成立范围不同.
此关系式对于 $\arcsin x$ 的整个定义域 $x \in [-1, 1]$ 是**恒成立**的.
此关系式**仅当** $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$(主值区间)时成立.

- 当 x 超出主值区间时, 运算结果会将 x **折叠**回主值区间.
- 函数 $y = \arcsin(\sin x)$ 的图像是周期为 2π 的锯齿波.



1.16.2 反余弦函数

1.16.2.1 定义与主值区间的选取

考察 $y = \cos x$. 区间 $[-\frac{\pi}{2}, \frac{\pi}{2}]$ 上 $\cos x$ 先增后减, 不是单射, 不能作为反余弦的主值区间.

为了定义反函数, 需选取一个既能覆盖值域 $[-1, 1]$ 又严格单调的区间. 数学界选定 $[0, \pi]$ 作为标准.

- 在 $[0, \pi]$ 上, $\cos x$ 从 1 严格单调递减至 -1.
- 该区间是非负的, 且包含了 0 点, 形式最简.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

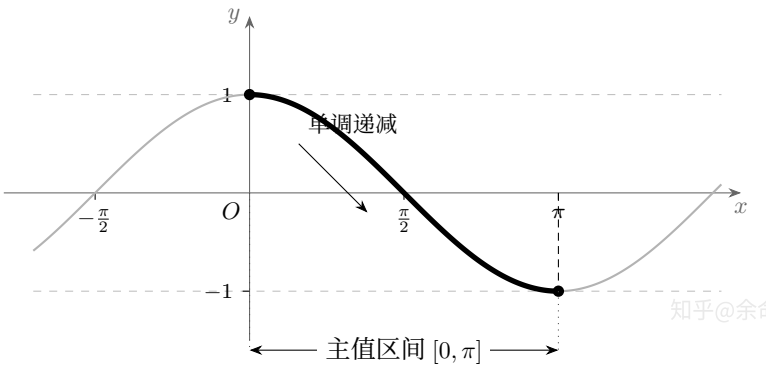


图 1.25　余弦函数主值区间 $[0,\pi]$ 的选取示意图

定义 1.6 (反余弦) 函数 $y=\cos x$ 在其主值区间 $[0,\pi]$ 上的反函数, 称为**反余弦函数**, 记作 $y=\arccos x$.

- **定义域**: $[-1,1]$
- **值域**: $[0,\pi]$

符号 $\arccos x$ 表达的意义是. 一个在 $[0,\pi]$ 内的角, 其余弦值为 x.

1.16.2.2 性质与图像

- **单调性**: 在其整个定义域 $[-1,1]$ 上**单调递减**.
- **奇偶性**: 既非奇函数也非偶函数. 但它具有一个重要的对称关系.

$$\arccos(-x)=\pi-\arccos x.$$

证明 (关于 $\arccos(-x)=\pi-\arccos x$) 令 $\alpha=\arccos x$. 由定义知 $\cos\alpha=x$ 且 $\alpha\in[0,\pi]$. 现证 $\pi-\alpha=\arccos(-x)$. 根据反余弦函数的定义, 只需验证两点.

1. $\cos(\pi-\alpha)=-x$.
2. $\pi-\alpha$ 落在反余弦的值域 $[0,\pi]$ 内.

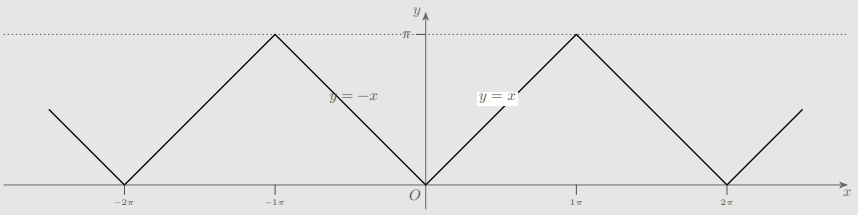
第一点, 由诱导公式可得 $\cos(\pi-\alpha)=-\cos\alpha=-x$, 成立. 第二点, 因为 $0\le\alpha\le\pi$, 所以 $-\pi\le-\alpha\le0$, 故 $0\le\pi-\alpha\le\pi$, 也成立. 证毕. □

注 1.18 ($\cos(\arccos x)$ 与 $\arccos(\cos x)$)
1. 恒等式 $\cos(\arccos x)=x$
此关系式对于 $\arccos x$ 的定义域 $x\in[-1,1]$ 是**恒成立**的.
2. 关系式 $\arccos(\cos x)=x$ **的成立范围**
此关系式**仅当** $x\in[0,\pi]$(主值区间)时成立.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 当 x 超出此范围时, $\arccos(\cos x)$ 的结果是**返回到主值区间** $[0,\pi]$ **内**.
- 几何意义. 结果是与 x 终边位置具有相同余弦值, 且位于 0 到 π 之间的那个角.
- 例如.$\arccos(\cos(2\pi))=0$; $\arccos(\cos(-\frac{\pi}{3}))=\frac{\pi}{3}$.

函数 $y=\arccos(\cos x)$ 的图像是一条在 x 轴上方、周期为 2π 的三角波.



1.16.3 重要的恒等关系

定理 1.11 (余角关系) 对于任意 $x\in[-1,1]$, 恒有.

$$\arcsin x+\arccos x=\frac{\pi}{2} \tag{1-17}$$

证明 令 $\alpha=\arcsin x$. 由定义知 $\sin\alpha=x$ 且 $\alpha\in[-\frac{\pi}{2},\frac{\pi}{2}]$. 考察角 $\frac{\pi}{2}-\alpha$. 首先,其值域为 $0\le\frac{\pi}{2}-\alpha\le\pi$,这恰好是反余弦函数的值域. 其次,其值为 $\cos(\frac{\pi}{2}-\alpha)=\sin\alpha=x$. 根据反余弦函数的唯一定义,一个值域在 $[0,\pi]$ 且余弦值为 x 的角,必定是 $\arccos x$. 故 $\arccos x=\frac{\pi}{2}-\alpha=\frac{\pi}{2}-\arcsin x$. 移项即得证. □

定理 1.12 (反正切的加法公式) 对于任意实数 a,b,

$$\arctan a+\arctan b=\begin{cases}\arctan\left(\frac{a+b}{1-ab}\right), & \text{if } ab<1\\ \frac{\pi}{2}, & \text{if } ab=1 \text{ and } a>0\\ -\frac{\pi}{2}, & \text{if } ab=1 \text{ and } a<0\\ \pi+\arctan\left(\frac{a+b}{1-ab}\right), & \text{if } ab>1 \text{ and } a>0\\ -\pi+\arctan\left(\frac{a+b}{1-ab}\right), & \text{if } ab>1 \text{ and } a<0\end{cases} \tag{1-18}$$

证明 令 $A=\arctan a, B=\arctan b$. 则 $\tan A=a,\tan B=b$, 且 $A,B\in(-\frac{\pi}{2},\frac{\pi}{2})$. 考察其和 $A+B$ 的正切值.

$$\tan(A+B)=\frac{\tan A+\tan B}{1-\tan A\tan B}=\frac{a+b}{1-ab}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这说明 $A+B$ 与 $\arctan\left(\frac{a+b}{1-ab}\right)$ 的正切值相同,它们之间相差 π 的整数倍. 即 $A+B=k\pi+\arctan\left(\frac{a+b}{1-ab}\right)$. 要点在于确定整数 k 的值, 这取决于 $A+B$ 所在的区间.

情形一. $ab<1$ 因为 $\cos A>0$ 且 $\cos B>0$, 有

$$\cos(A+B)=\cos A\cos B(1-\tan A\tan B)=\cos A\cos B(1-ab)>0.$$

又 $A+B\in(-\pi,\pi)$, 所以 $A+B\in(-\frac{\pi}{2},\frac{\pi}{2})$. 这正是 arctan 的主值区间, 因而 $k=0$.

情形二. $ab=1$ 此时 a 与 b 同号. 若 $a>0$, 则 $A,B\in(0,\frac{\pi}{2})$ 且 $\tan A\tan B=1$, 故 $A+B=\frac{\pi}{2}$. 若 $a<0$, 则 $A,B\in(-\frac{\pi}{2},0)$, 同理 $A+B=-\frac{\pi}{2}$.

情形三. $ab>1$ **且** $a>0$ 此时必有 $b>0$.$A,B\in(0,\frac{\pi}{2})$, 故 $A+B\in(0,\pi)$. $ab>1\implies 1-ab<0\implies\frac{a+b}{1-ab}<0\implies\tan(A+B)<0$. 这说明 $A+B$ 位于第二或第四象限. 结合 $A+B\in(0,\pi)$, 可知 $A+B\in(\frac{\pi}{2},\pi)$. 而 $\arctan(\frac{a+b}{1-ab})$ 位于 $(-\frac{\pi}{2},0)$. 要使等式成立, 必须有 $k=1$.

情形四. $ab>1$ **且** $a<0$ 此时必有 $b<0$.$A,B\in(-\frac{\pi}{2},0)$, 故 $A+B\in(-\pi,0)$. $ab>1\implies 1-ab<0\implies\frac{a+b}{1-ab}>0\implies\tan(A+B)>0$. 这说明 $A+B$ 位于第一或第三象限. 结合 $A+B\in(-\pi,0)$, 可知 $A+B\in(-\pi,-\frac{\pi}{2})$. 而 $\arctan(\frac{a+b}{1-ab})$ 位于 $(0,\frac{\pi}{2})$. 要使等式成立, 必须有 $k=-1$. □

1.16.4 典例剖析

例题 1.72 (复合函数求值) 求 $\cos\left(\arcsin\left(-\frac{4}{5}\right)\right)$ 的值.

解 令 $\alpha=\arcsin\left(-\frac{4}{5}\right)$. 现求 $\cos\alpha$. 根据反正弦函数的定义, 可知.

$$\sin\alpha=-\frac{4}{5}\quad\text{且}\quad\alpha\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$$

利用同角三角函数平方关系 $\sin^2\alpha+\cos^2\alpha=1$,

$$\cos^2\alpha=1-\sin^2\alpha=1-\left(-\frac{4}{5}\right)^2=1-\frac{16}{25}=\frac{9}{25}$$

所以 $\cos\alpha=\pm\frac{3}{5}$.

接着, 利用角的范围确定符号. 因为 $\sin\alpha=-\frac{4}{5}$ 是负值, 且 $\alpha\in[-\frac{\pi}{2},\frac{\pi}{2}]$, 所以角 α 必定落在第四象限, 或其边界, 即 $\alpha\in[-\frac{\pi}{2},0)$. 在第四象限, 余弦值为正.

$$\therefore\cos\alpha=\frac{3}{5}$$

例题 1.73 (反正切的和) 求值.$\arctan 2+\arctan 3$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 直接应用反正切的加法公式. 令 $a=2, b=3$. 注意到 $a>0, b>0$ 且 $ab=2\times 3=6>1$. 因此, 应使用第二种情形的公式.

$$\begin{aligned}\arctan 2+\arctan 3 &= \pi+\arctan\left(\frac{2+3}{1-2\cdot 3}\right)\\ &= \pi+\arctan\left(\frac{5}{-5}\right)\\ &= \pi+\arctan(-1)\end{aligned}$$

根据反正切函数的定义, $\arctan(-1)$ 是在 $(-\frac{\pi}{2}, \frac{\pi}{2})$ 内正切值为 -1 的角, 即 $-\frac{\pi}{4}$.

$$\therefore \arctan 2+\arctan 3=\pi+\left(-\frac{\pi}{4}\right)=\frac{3\pi}{4}$$

例题 1.74 (透彻理解恒等式) 求值.$\arccos\left(\cos\left(\frac{8\pi}{3}\right)\right)$.

解 角 $\frac{8\pi}{3}$ 不在反余弦函数的主值区间 $[0,\pi]$ 内, 所以结果不是它本身. 先需计算内层函数的值.

$$\cos\left(\frac{8\pi}{3}\right)=\cos\left(2\pi+\frac{2\pi}{3}\right)=\cos\left(\frac{2\pi}{3}\right)=-\frac{1}{2}$$

接着, 原问题转化为求 $\arccos(-\frac{1}{2})$. 需寻找在 $[0,\pi]$ 区间内, 余弦值为 $-\frac{1}{2}$ 的角. 该角为 $\frac{2\pi}{3}$.

$$\therefore \arccos\left(\cos\left(\frac{8\pi}{3}\right)\right)=\frac{2\pi}{3}$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

第 2 章

数列

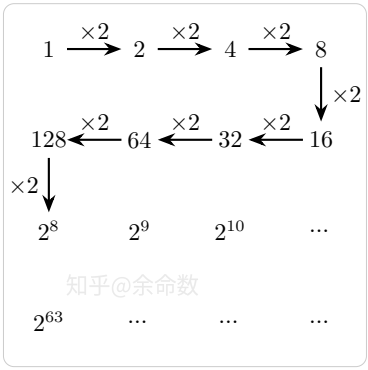


基础语言

「数学的本质在于其自由。」

——康托尔

第二章 数列



知乎@余命数

棋盘上的米粒

古老的传说中, 国际象棋的发明者向国王提出一个请求. 他在棋盘第 1 格放 1 粒米, 第 2 格放 2 粒, 第 3 格放 4 粒, 之后每一格的米粒数都是前一格的两倍, 直到填满全部 64 格. 国王欣然应允, 却不知自己许下了一个无法兑现的承诺.

这个序列 $1, 2, 4, 8, \ldots$ 增长速度惊人. 到第 20 格, 米粒数已超过一百万; 到第 41 格, 已超过当时全世界的米粒总产量. 我们要算清 64 格总共有多少米粒, 就要刻画这串数字的生成规律, 并找到不必逐一相加便能求出总数的办法.

这个故事引出一个基本数学对象, 即**数列**, 特别是**等比数列**. 复利增长、细胞分裂等利滚利现象, 都可以用等比数列描述. 本章我们建立描述数列每一项的记法, 并学习高效计算其和的方法.

2.1 数学归纳法

在数学的探索中, 我们常常从观察具体的例子开始.

$$1 = 1^2$$

$$1 + 3 = 4 = 2^2$$

$$1 + 3 + 5 = 9 = 3^2$$

$$1 + 3 + 5 + 7 = 16 = 4^2$$

观察这一组等式. 前 n 个正奇数的和似乎总是等于 n^2. 要确定这个规律对所有正整数 n 成立, 无论是 $n = 100$ 还是 $n = 10^{10}$, 验证再多例子也无法构成严格的**证明**. 需要一种从有限验证出发、得到关于无限情况的确定性结论的方法.

这个逻辑工具就是**数学归纳法**, 它把归纳猜想与演绎证明联系起来, 是离散数学[1]的重要证明方法.

[1] 离散数学是数学的几个分支的总称, 研究基于离散空间的数学结构. 与连续变化的实数不同, 离散数学的研究对象, 例如整数、图和数学逻辑中的命题, 拥有不等、分立的值. 因此离散数学不包含微积分和分析等连续数学的内容.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

公理 2.1 (数学归纳法) 设 $P(n)$ 是一个关于正整数 n 的命题, 且起始值 n_0 为正整数. 如果以下两个条件都被满足,

- **奠基步骤:** 命题 $P(n)$ 对起始值 $n = n_0$ 成立(通常 $n_0 = 1$).
- **归纳步骤:** **假设**命题对任意整数 $k \geq n_0$ 成立, 即 $P(k)$ 为真, 能够**推导出**命题对下一个整数 $k+1$ 也成立, 即 $P(k+1)$ 为真.

那么, 可断定, 命题 $P(n)$ 对所有大于或等于 n_0 的整数 n 都成立.

数学归纳法的原理可以用一排无穷无尽的多米诺骨牌来比喻. 要让所有的骨牌都倒下, 需要满足两个条件,

知乎@余命数

1. 首先要推倒第一块骨牌, 这就是**奠基步骤** $P(n_0)$. 第一块不倒, 连锁反应就无从谈起.
2. 骨牌的排列必须保证, 任意一块(第 k 块)倒下时, 都必定会撞倒第 $k+1$ 块, 这就是**归纳步骤** $P(k) \implies P(k+1)$. 它保证了倒下这个性质可以无休止地传递下去.

只要这两个条件同时满足, 整排骨牌就将依次倒下.

例题 2.1 使用数学归纳法证明,对所有正整数 n,等式 $1+3+5+\cdots+(2n-1) = n^2$ 成立.

证明 设命题 $P(n)$ 为 $1+3+5+\cdots+(2n-1) = n^2$.

当 $n = 1$ 时, 等式左边为 1, 右边为 $1^2 = 1$, 两端相等, 故 $P(1)$ 成立, 奠基步骤完成.

假设当 $n = k$($k \geq 1$ 为任意整数)时命题成立, 即

$$1+3+5+\cdots+(2k-1) = k^2.$$

由此只需证明当 $n = k+1$ 时命题也成立, 即

$$1+3+5+\cdots+(2(k+1)-1) = (k+1)^2.$$

对目标等式的左边作恒等变形, 将其分离为前 k 个奇数之和与新增的末项, 得

$$\begin{aligned} \text{左边} &= 1+3+5+\cdots+(2k-1)+(2(k+1)-1) \\ &= \underbrace{[1+3+5+\cdots+(2k-1)]}_{\text{归纳假设部分}}+(2k+2-1) \\ &= k^2+(2k+1) \quad \text{(代入归纳假设)} \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= k^2 + 2k + 1$$

$$= (k+1)^2$$

$$= \text{右边}.$$

上式表明, 若第 k 块骨牌倒下, 它必然会推倒第 $k+1$ 块, 归纳步骤成立.

由数学归纳法原理可知, 命题 $P(n)$ 对所有正整数 n 均成立. □

归纳假设不是循环论证. 我们在归纳步骤中未直接断言结论. 我们证明的是一个**条件命题**, "**如果** $P(k)$ 成立, **那么** $P(k+1)$ 成立". 我们把 $P(k)$ 作为一个前提条件, 来证明这个传递的过程是有效的. 两步缺一不可. 只有奠基步骤, 我们只能证明命题对起始值成立, 无法推广; 只有归纳步骤, 我们只证明了"如果一块倒下, 下一块也会倒", 但如果第一块就没倒, 那么什么都不会发生. 数学归纳法同样能处理整除与不等式问题.

知乎@余命数

例题 2.2 证明当实数 $x > -1$ 时, 对所有整数 $n \geq 1$, 不等式 $(1+x)^n \geq 1+nx$ 成立.

解 设命题 $P(n)$ 为 $(1+x)^n \geq 1+nx$.

当 $n=1$ 时, 不等式左边为 $(1+x)^1 = 1+x$, 右边为 $1+1\cdot x = 1+x$, 两端相等, 故 $P(1)$ 成立.

假设当 $n=k$ $(k \geq 1)$ 时命题成立, 即

$$(1+x)^k \geq 1+kx.$$

由此只需证明当 $n=k+1$ 时命题也成立, 即 $(1+x)^{k+1} \geq 1+(k+1)x$.

由条件 $x > -1$ 知 $1+x > 0$, 故不等式两边可同乘正数 $1+x$ 而不改变方向. 将目标不等式左边写为 $(1+x)^{k+1} = (1+x)^k \cdot (1+x)$, 代入归纳假设得

$$(1+x)^{k+1} \geq (1+kx)(1+x) \quad \text{(归纳假设, 两边同乘 } 1+x>0\text{)}$$

$$= 1 + x + kx + kx^2$$

$$= 1 + (k+1)x + kx^2.$$

由于 $k \geq 1$ 且 $x^2 \geq 0$, 故 $kx^2 \geq 0$, 从而

$$1 + (k+1)x + kx^2 \geq 1 + (k+1)x.$$

由不等式的传递性将两式连接得

$$(1+x)^{k+1} \geq 1 + (k+1)x + kx^2 \geq 1 + (k+1)x,$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

即 $(1+x)^{k+1} \geq 1+(k+1)x$, 归纳步骤成立.

由数学归纳法原理, 命题对所有整数 $n \geq 1$ 成立.

此即**伯努利不等式**.

例题 2.3　证明对所有正整数 n, 表达式 7^n-1 能被 6 整除.

解　设命题 $P(n)$ 为"7^n-1 能被 6 整除".

当 $n=1$ 时, $7^1-1=6$ 是 6 的倍数, 故 $P(1)$ 成立.

假设当 $n=k\,(k \geq 1)$ 时命题成立, 即 7^k-1 是 6 的倍数, 可记作 $7^k-1=6m$, 其中 m 为整数, 于是 $7^k=6m+1$. 由此只需证明当 $n=k+1$ 时 $7^{k+1}-1$ 能被 6 整除.

对目标表达式变形以引入归纳假设中的 7^k, 得

$$\begin{aligned} 7^{k+1}-1 &= 7\cdot 7^k-1 \\ &= 7(6m+1)-1 \quad (\text{代入 } 7^k=6m+1) \\ &= 42m+6 \\ &= 6(7m+1). \end{aligned}$$

因 m 为整数, 故 $7m+1$ 亦为整数, 从而 $7^{k+1}-1$ 等于 6 与一整数之积, 符合能被 6 整除的定义, 归纳步骤成立.

由数学归纳法原理, 对所有正整数 n, 7^n-1 都能被 6 整除.

例题 2.4 (缺一格的棋盘)　设 n 为正整数, 从 $2^n\times 2^n$ 的棋盘中挖去任意一格, 证明剩下的部分都能用 L 形骨牌(2 × 2 的方块去掉一格)恰好铺满.

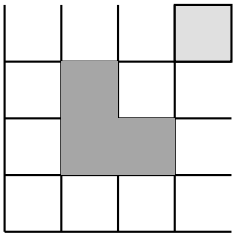


图 2.1　4×4 棋盘挖去一格(右上角的浅色格). 先在中心放一块 L 形骨牌, 使不含挖去格的三个象限各少一格, 四个 2×2 小棋盘于是都成为"挖去一格"的形状.

证明　对 n 归纳.$n=1$ 时, 2×2 的棋盘挖去一格, 剩下一格恰好是一块 L 形骨牌, 结论成立.

设结论对 n 成立, 考虑 $2^{n+1}\times 2^{n+1}$ 的棋盘挖去一格. 把它等分成四个 $2^n\times 2^n$ 的象限, 挖去的格落在其中一个象限里. 在棋盘正中心放一块 L 形骨牌, 让它的三

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

格分别落在另外三个象限里紧贴中心的那一格. 这样一来, 四个象限各自恰好少了一格, 含挖去格的象限少的是原来那格, 另外三个象限少的是被骨牌占去的那格. 四个象限都成了“$2^n \times 2^n$ 挖去一格”的形状, 由归纳假设各自都能铺满, 合起来就铺满了整个棋盘.

归纳假设能用在四个象限上, 是因为每个象限的形状与假设中的形状完全相同. 换一个尺寸或换一种骨牌, 这一步就未必走得通. □

知乎@余命数

例题 2.5 (直线分平面的区域数) 平面上 n 条直线两两不平行, 且任意三条不共点, 证明这 n 条直线把平面分成的区域数为 $\dfrac{n(n+1)}{2}+1$.

证明 记区域数为 R_n.$n=1$ 时一条直线把平面分成 2 个区域, 而 $\dfrac{1\times 2}{2}+1=2$, 公式成立.

设 n 条直线已把平面分成 R_n 个区域, 添入第 $n+1$ 条直线. 它与原有的 n 条直线各有一个交点, 由题设这些交点互不重合, 于是这条直线上共有 n 个交点, 它们把直线分成 $n+1$ 段. 每一段都落在某个已有的区域里, 并把该区域一分为二, 所以区域数增加 $n+1$, 即

$$R_{n+1}=R_n+(n+1).$$

知乎@余命数

由归纳假设 $R_n=\dfrac{n(n+1)}{2}+1$, 得

$$R_{n+1}=\frac{n(n+1)}{2}+1+(n+1)=\frac{(n+1)(n+2)}{2}+1,$$

公式对 $n+1$ 也成立. □

涉及新定义时还会再用到数学归纳法. 现先认识两个简化运算的符号.

2.2 $\sum$ 和 $\prod$ 符号

进入数列前, 有一类操作会反复出现, 即计算数列中许多项的总和或总乘积. 例如, 要表达前 100 个正整数的和, 写出 $1+2+3+\cdots+100$ 尚可接受, 但若要表达一个更抽象或更长的和, 比如 $a_1+a_2+\cdots+a_n$, 这种写法就显得冗长且不够精确. 为简洁、清晰、无歧义地表达这种运算, 数学家引入了两个实用符号, 即求和符号 ($\sum$) 和连乘符号 ($\prod$).

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定义 2.1 (求和符号) 一个数列 $\{a_n\}$ 从第 m 项到第 n 项的和 $a_m + a_{m+1} + \cdots + a_n$, 其中 $m \le n$ 可以用求和符号 $\sum$ (Sigma) 紧凑地写成

$$\sum_{k=m}^{n} a_k$$

其中各部分的名称如下.

- k 被称为**求和指标**. 它是一个哑变量, 仅起占位作用, 可用任何其他不引起混淆的字母代替, 如 i, j 等.
- m 是求和的**下界**, 即求和指标的起始值.
- n 是求和的**上界**, 即求和指标的终止值.
- a_k 是被求和的**通项**.

例题 2.6
- 前 100 个正整数的和可写成 $1 + 2 + \cdots + 100 = \sum_{k=1}^{100} k$.
- 前 n 个正奇数的和可写成 $1 + 3 + 5 + \cdots + (2n-1) = \sum_{k=1}^{n} (2k-1)$.
- 等差数列前 n 项和可写成 $S_n = a_1 + \cdots + a_n = \sum_{k=1}^{n} a_k$.

知乎@余命数

定义 2.2 (连乘符号) 与求和符号类似, 一个数列 $\{a_n\}$ 从第 m 项到第 n 项的积 $a_m \cdot a_{m+1} \cdot \cdots \cdot a_n$ 可以用连乘符号 $\prod$ (Pi) 表示如下.

$$\prod_{k=m}^{n} a_k$$

其各部分的名称与求和符号类似, 分别为**连乘指标**、**下界**和**上界**.

例题 2.7
- 阶乘可写成 $n! = 1 \cdot 2 \cdot 3 \cdot \cdots \cdot n = \prod_{k=1}^{n} k$.
- 等比数列前 n 项的积可写成 $P_n = a_1 \cdot a_2 \cdot \cdots \cdot a_n = \prod_{k=1}^{n} a_k = \prod_{k=1}^{n} (a_1 q^{k-1})$.

知乎@余命数

它们继承了加法和乘法的基本定律, 如分配律、结合律和交换律.

设 $\{a_k\}, \{b_k\}$ 为数列, c 为常数.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **常数倍数法则**.

$$\sum_{k=m}^{n} ca_k = c\sum_{k=m}^{n} a_k \qquad \text{而} \qquad \prod_{k=m}^{n}(ca_k) = c^{n-m+1}\prod_{k=m}^{n} a_k$$

注意连乘时, 常数 c 被乘了 $n-m+1$ 次.

- **加法/乘法法则**.

$$\sum_{k=m}^{n}(a_k \pm b_k) = \sum_{k=m}^{n} a_k \pm \sum_{k=m}^{n} b_k$$

$$\prod_{k=m}^{n}(a_k b_k) = \Big(\prod_{k=m}^{n} a_k\Big)\Big(\prod_{k=m}^{n} b_k\Big)$$

- **拆分法则**. 对于任意整数 p 满足 $m \le p < n$,

$$\sum_{k=m}^{n} a_k = \sum_{k=m}^{p} a_k + \sum_{k=p+1}^{n} a_k$$

$$\prod_{k=m}^{n} a_k = \left(\prod_{k=m}^{p} a_k\right)\left(\prod_{k=p+1}^{n} a_k\right)$$

- **指标平移**.

$$\sum_{k=m}^{n} a_k = \sum_{j=m+p}^{n+p} a_{j-p}$$

这相当于做了一个变量代换 $j = k+p$.

逐条说明这些法则.

2.2.1 运算法则

2.2.1.1 线性性质

这条性质允许把复杂的求和分解为若干个简单的部分. 该性质仅适用于**求和符号**.

$$\sum_{k=m}^{n}(c\cdot a_k \pm d\cdot b_k) = c\sum_{k=m}^{n} a_k \pm d\sum_{k=m}^{n} b_k$$

其中 c, d 是与求和指标 k 无关的常数. 这个性质可以看作两条规则的结合. 常数可提出求和号外, 各项可分别求和.

例题 2.8 利用等差数列求和公式计算 $\sum_{k=1}^{n}(4k-3)$.

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

解 直接将通项 $4k-3$ 看作首项为 1、末项为 $4n-3$ 的等差数列求和固然可行，此处则借助线性性质将其分解为更基本的求和.

$$\begin{aligned}\sum_{k=1}^{n}(4k-3) &= \sum_{k=1}^{n}(4k)-\sum_{k=1}^{n}3 && \text{（加减法拆分）}\\ &= 4\sum_{k=1}^{n}k-3\sum_{k=1}^{n}1 && \text{（常数因子提出）}\end{aligned}$$

代入两个基本求和公式 $\sum_{k=1}^{n}k=\frac{n(n+1)}{2}$ 与 $\sum_{k=1}^{n}1=n$，得

$$\begin{aligned}\sum_{k=1}^{n}(4k-3) &= 4\cdot\frac{n(n+1)}{2}-3\cdot n\\ &= 2n(n+1)-3n\\ &= 2n^2-n.\end{aligned}$$

这与直接用等差数列求和公式 $S_n=\frac{n(a_1+a_n)}{2}=\frac{n(1+4n-3)}{2}=n(2n-1)=2n^2-n$ 所得结果一致.

注 2.1 (线性性质不适用于乘积)
求和的线性性质**不适用于**乘积. 一般地，

$$\sum_{k=1}^{n}(a_kb_k)\neq\left(\sum_{k=1}^{n}a_k\right)\left(\sum_{k=1}^{n}b_k\right)$$

例如，$\sum_{k=1}^{2}k^2=1^2+2^2=5$，但 $(\sum_{k=1}^{2}k)^2=(1+2)^2=9$.
连乘符号没有线性性质. 一般地，

$$\prod_{k=1}^{n}(a_k+b_k)\neq\prod_{k=1}^{n}a_k+\prod_{k=1}^{n}b_k$$

2.2.1.2 指标平移

这项技巧通过改变求和的起点和终点，来简化求和式内部的表达式，使其变为我们熟悉的形式. 这好比解方程时换元.

具体做法是，令新的指标 j 与旧的指标 k 满足一个简单线性关系，如 $j=k+c$ 或 $j=k-c$.

$$\sum_{k=m}^{n}a_k=\sum_{j=m+c}^{n+c}a_{j-c}$$

在替换时，要同时更新**下界**、**上界**和**通项**中的变量.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

例题 2.9 计算和式 $S=\sum_{k=3}^{n+2}(k-2)^2$.

解 这个和式的通项是 $(k-2)^2$, 上下界也比较奇特. 把它变成 $\sum j^2$ 的标准形式即可直接套用平方和公式.

作换元, 令新指标 $j=k-2$, 并同时更新求和的三要素, 得

- **下界**. 当旧指标 $k=3$ 时, 新指标 $j=3-2=1$.
- **上界**. 当旧指标 $k=n+2$ 时, 新指标 $j=(n+2)-2=n$.
- **通项**.$a_k=(k-2)^2$ 变为 $a_{j+2}=j^2$.

于是原来的和式转化为

$$S=\sum_{j=1}^{n}j^2.$$

套用平方和公式(将在后续章节证明)

$$\sum_{j=1}^{n}j^2=\frac{n(n+1)(2n+1)}{6},$$

即得 $S=\frac{n(n+1)(2n+1)}{6}$.

2.2.1.3 嵌套求和

当求和的通项本身又是一个求和时, 就构成了嵌套求和.

$$\sum_{i=m}^{n}\left(\sum_{j=p}^{q}a_{i,j}\right)$$

当内层求和的上下界 (p,q) 与外层指标 (i) 无关时, 可自由地**交换求和次序**.

$$\sum_{i=m}^{n}\sum_{j=p}^{q}a_{i,j}=\sum_{j=p}^{q}\sum_{i=m}^{n}a_{i,j}$$

[2] 矩阵即按行列排成的数表.

理解这个性质的最佳方式是想象一个 $m\times n$ 的数字矩阵[2]. 上述等式无非是说, 先按行求和再把各行之和加起来, 与先按列求和再把各列之和加起来, 结果是相等的.

例题 2.10 证明 $\left(\sum_{i=1}^{m}x_i\right)\left(\sum_{j=1}^{n}y_j\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}x_iy_j$.

知乎@余命数 知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

证明 从左边入手, 将第一个和式视作一个整体 $C=\sum_{i=1}^{m} x_i$, 则

$$\left(\sum_{i=1}^{m} x_i\right)\left(\sum_{j=1}^{n} y_j\right)=C\cdot\sum_{j=1}^{n} y_j.$$

由求和的线性性质, 常数 C 可乘入第二个和式内部, 得 $\sum_{j=1}^{n}(C\cdot y_j)$; 再把 $C=\sum_{i=1}^{m} x_i$ 代回, 有

$$\sum_{j=1}^{n}\left(\left(\sum_{i=1}^{m} x_i\right)y_j\right).$$

对于内层的和 $\sum_{i=1}^{m} x_i, y_j$ 关于指标 i 是常数, 故同样可乘入, 得到

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m} x_i y_j\right).$$

有限求和可交换次序, 因此

$$\sum_{j=1}^{n}\sum_{i=1}^{m} x_i y_j=\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j,$$

这正是右侧的嵌套求和形式. 此恒等式在概率统计中计算期望与方差时常被引用. □

2.2.1.4 裂项求和与求积

我们后续会用专门的章节来分析裂项求和与求积. 这里的要点是用 ∑ 和 ∏ 符号写出它的结构.

- **裂项求和**. 如果通项可以表示为 $a_k=f(k)-f(k+1)$, 那么

$$\begin{aligned}\sum_{k=m}^{n} a_k &= \sum_{k=m}^{n}(f(k)-f(k+1))\\ &=(f(m)-f(m+1))+(f(m+1)-f(m+2))\\ &\quad+\cdots+(f(n)-f(n+1))\\ &=f(m)-f(n+1)\end{aligned}$$

  中间的项 $-f(m+1)$ 与 $+f(m+1)$ 等前后抵消, 如同收起的望远镜.

- **裂项求积**. 如果通项可以表示为 $a_k=\frac{f(k)}{f(k+1)}$, 那么

$$\prod_{k=m}^{n} a_k=\prod_{k=m}^{n}\frac{f(k)}{f(k+1)}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= \frac{f(m)}{f(m+1)} \cdot \frac{f(m+1)}{f(m+2)} \cdot \cdots \cdot \frac{f(n)}{f(n+1)}$$
$$= \frac{f(m)}{f(n+1)}$$

2.2.2 有趣的应用 *

例题 2.11 计算和式 $S = \sum_{k=0}^{n-1}(k+1)^3$.

解 这个和式的通项是 $(k+1)^3$,求和指标从 0 到 $n-1$,与熟知的立方和公式 $\sum_{j=1}^{n} j^3$ 形式接近, 故作指标平移化为标准形式.

令新指标 $j = k+1$, 求和的三要素随之更新, 得

- **下界**. 当旧指标 $k=0$ 时, 新指标 $j = 0+1 = 1$.
- **上界**. 当旧指标 $k = n-1$ 时, 新指标 $j = (n-1)+1 = n$.
- **通项**.$(k+1)^3$ 变为 j^3.

原和式转化为 $S = \sum_{j=1}^{n} j^3$, 套用立方和公式(将在后续章节证明)即得

$$S = \left(\frac{n(n+1)}{2}\right)^2 = \frac{n^2(n+1)^2}{4}.$$

例题 2.12 计算和式 $S_n = \sum_{k=1}^{n} \frac{1}{k(k+2)}$.

解 通项 $\frac{1}{k(k+2)}$ 的形式暗示了裂项求和的可能, 以待定系数法将其分解为部分分式, 即

$$\frac{1}{k(k+2)} = \frac{A}{k} + \frac{B}{k+2}.$$

通分得 $1 = A(k+2) + Bk$. 令 $k=0$ 得 $A = 1/2$, 令 $k=-2$ 得 $B = -1/2$, 故

$$\frac{1}{k(k+2)} = \frac{1}{2}\left(\frac{1}{k} - \frac{1}{k+2}\right).$$

代回原和式,

$$S_n = \sum_{k=1}^{n} \frac{1}{2}\left(\frac{1}{k} - \frac{1}{k+2}\right) = \frac{1}{2}\sum_{k=1}^{n}\left(\frac{1}{k} - \frac{1}{k+2}\right).$$

这是间隔两项的裂项求和. 写出前后各项以观察抵消规律.

$$2S_n = \left(\frac{1}{1} - \frac{1}{3}\right) + \left(\frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{5}\right) + \left(\frac{1}{4} - \frac{1}{6}\right) + \cdots$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

$$+\left(\frac{1}{n-1}-\frac{1}{n+1}\right)+\left(\frac{1}{n}-\frac{1}{n+2}\right).$$

从第三项起, $-\frac{1}{3}$ 与 $+\frac{1}{3}$ 抵消, $-\frac{1}{4}$ 与 $+\frac{1}{4}$ 抵消, 以此类推, 最终只留下开头的两项正数与结尾的两项负数.

$$2S_n=\frac{1}{1}+\frac{1}{2}-\frac{1}{n+1}-\frac{1}{n+2}=\frac{3}{2}-\frac{n+2+n+1}{(n+1)(n+2)}=\frac{3}{2}-\frac{2n+3}{(n+1)(n+2)}.$$

故

$$S_n=\frac{3}{4}-\frac{2n+3}{2(n+1)(n+2)}=\frac{n(3n+5)}{4(n+1)(n+2)}.$$

例题 2.13　计算二重和 $S=\sum_{i=1}^{n}\sum_{j=i}^{n}\frac{i}{j(j+1)}$.

解　内层 $\sum_{j=i}^{n}$ 的求和范围依赖外层指标 i, 直接计算内层和有困难, 故交换求和次序.

指标 (i,j) 需满足 $1\le i\le j\le n$. 当前的求和次序是固定 i、令 j 从 i 扫到 n(按行求和); 改为固定 j 再看 i 的取值范围, 由条件 $1\le i\le j\le n$ 可知 j 从 1 到 n, 固定 j 时 i 从 1 到 j. 于是得到等价形式

$$S=\sum_{j=1}^{n}\sum_{i=1}^{j}\frac{i}{j(j+1)}.$$

在内层和 $\sum_{i=1}^{j}$ 中, $\frac{1}{j(j+1)}$ 与指标 i 无关, 可提出.

$$S=\sum_{j=1}^{n}\left(\frac{1}{j(j+1)}\sum_{i=1}^{j}i\right).$$

内层为标准等差数列求和 $\sum_{i=1}^{j}i=\frac{j(j+1)}{2}$, 代回得

$$S=\sum_{j=1}^{n}\left(\frac{1}{j(j+1)}\cdot\frac{j(j+1)}{2}\right)=\sum_{j=1}^{n}\frac{1}{2}=n\cdot\frac{1}{2}=\frac{n}{2}.$$

交换求和次序后, 复杂的二重和归约为平凡的 n 个 $\frac{1}{2}$ 相加.

例题 2.14　计算连乘积 $P_n=\prod_{k=2}^{n}\frac{k^3-1}{k^3+1}$, 其中 $n\ge 2$.

解　连乘积的目标通常是裂项相消, 要点在于因式分解通项. 用立方差与立方和公式

$$a^3-b^3=(a-b)(a^2+ab+b^2),\quad a^3+b^3=(a+b)(a^2-ab+b^2)$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

分解分子分母, 得

$$k^3-1=(k-1)(k^2+k+1),\quad k^3+1=(k+1)(k^2-k+1),$$

故每一项为

$$\frac{k^3-1}{k^3+1}=\frac{(k-1)(k^2+k+1)}{(k+1)(k^2-k+1)}.$$

此式直接看不出相消关系, 将其拆成两个部分的乘积.

$$P_n=\prod_{k=2}^{n}\left(\frac{k-1}{k+1}\right)\cdot\prod_{k=2}^{n}\left(\frac{k^2+k+1}{k^2-k+1}\right).$$

知乎@余命数

第一个乘积是标准的裂项求积.

$$\prod_{k=2}^{n}\frac{k-1}{k+1}=\frac{1}{3}\cdot\frac{2}{4}\cdot\frac{3}{5}\cdot\cdots\cdot\frac{n-1}{n+1}=\frac{1\cdot 2}{n(n+1)}=\frac{2}{n(n+1)}.$$

对于第二个乘积, 令 $f(k)=k^2-k+1$, 则 $f(k+1)=(k+1)^2-(k+1)+1=k^2+k+1$, 于是通项的第二部分可写为 $\frac{f(k+1)}{f(k)}$.

$$\prod_{k=2}^{n}\frac{f(k+1)}{f(k)}=\frac{f(3)}{f(2)}\cdot\frac{f(4)}{f(3)}\cdots\cdots\frac{f(n+1)}{f(n)}=\frac{f(n+1)}{f(2)}=\frac{(n+1)^2-(n+1)+1}{2^2-2+1}=\frac{n^2+n+1}{3}.$$

两部分相乘即得

$$P_n=\frac{2}{n(n+1)}\cdot\frac{n^2+n+1}{3}=\frac{2(n^2+n+1)}{3n(n+1)}.$$

例题 2.15 计算和式 $S_n=\sum_{k=1}^{n}\frac{k}{k^4+k^2+1}$.

解 通项分母 k^4+k^2+1 形式特殊, 是突破口. 通过配方法和平方差公式对其因式分解, 即所谓苏菲·热尔曼恒等式[3].

$$k^4+k^2+1=(k^4+2k^2+1)-k^2=(k^2+1)^2-k^2=(k^2-k+1)(k^2+k+1).$$

注意到分母两因子之差 $(k^2+k+1)-(k^2-k+1)=2k$ 恰为分子的常数倍, 据此作裂项, 得

$$\begin{aligned}\frac{k}{k^4+k^2+1}&=\frac{k}{(k^2-k+1)(k^2+k+1)}\\&=\frac{1}{2}\cdot\frac{2k}{(k^2-k+1)(k^2+k+1)}\end{aligned}$$

[3] 索菲·热尔曼恒等式是一个常用的因式分解技巧, 其经典形式为 $a^4+4b^4=(a^2+2b^2-2ab)(a^2+2b^2+2ab)$. 其做法是通过增项再减项配成一个平方差, 例如将 a^4+4b^4 变为 $(a^4+4a^2b^2+4b^4)-4a^2b^2$. 本例中对 k^4+k^2+1 的分解正是该做法的直接应用, $(k^4+2k^2+1)-k^2$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= \frac{1}{2} \cdot \frac{(k^2+k+1)-(k^2-k+1)}{(k^2-k+1)(k^2+k+1)}$$
$$= \frac{1}{2}\left(\frac{1}{k^2-k+1} - \frac{1}{k^2+k+1}\right).$$

令 $f(k) = \frac{1}{k^2-k+1}$. 因 $k^2+k+1 = (k+1)^2-(k+1)+1$, 故 $\frac{1}{k^2+k+1} = f(k+1)$, 原和式化为

$$S_n = \sum_{k=1}^{n} \frac{1}{2}\left(f(k) - f(k+1)\right) = \frac{1}{2}\sum_{k=1}^{n}(f(k) - f(k+1))$$
$$= \frac{1}{2}\left[(f(1)-f(2)) + (f(2)-f(3)) + \cdots + (f(n)-f(n+1))\right]$$
$$= \frac{1}{2}(f(1) - f(n+1)).$$

知乎@余命数　知乎@余命数

计算首末两项, $f(1) = \frac{1}{1^2-1+1} = 1$, $f(n+1) = \frac{1}{(n+1)^2-(n+1)+1} = \frac{1}{n^2+n+1}$. 代入得

$$S_n = \frac{1}{2}\left(1 - \frac{1}{n^2+n+1}\right) = \frac{1}{2} \cdot \frac{n^2+n}{n^2+n+1} = \frac{n(n+1)}{2(n^2+n+1)}.$$

例题 2.16　计算连乘积 $P_n = \prod_{k=2}^{n}\left(1 - \frac{1}{k^2}\right)$.

解　直接逐项相乘会非常复杂. 连乘计算的关键第一步通常是**因式分解**通项, 以期发现可裂项相消的结构. 通项为

$$a_k = 1 - \frac{1}{k^2} = \frac{k^2-1}{k^2} = \frac{(k-1)(k+1)}{k \cdot k},$$

知乎@余命数

代入连乘积得

$$P_n = \prod_{k=2}^{n} \frac{(k-1)(k+1)}{k \cdot k}.$$

展开乘积以观察相消规律.

$$P_n = \frac{(2-1)(2+1)}{2 \cdot 2} \cdot \frac{(3-1)(3+1)}{3 \cdot 3} \cdot \frac{(4-1)(4+1)}{4 \cdot 4} \cdot \cdots \cdot \frac{(n-1)(n+1)}{n \cdot n}$$
$$= \frac{1 \cdot 3}{2 \cdot 2} \cdot \frac{2 \cdot 4}{3 \cdot 3} \cdot \frac{3 \cdot 5}{4 \cdot 4} \cdot \cdots \cdot \frac{(n-1)(n+1)}{n \cdot n}.$$

知乎@余命数

将分子与分母分别归类.

$$P_n = \frac{1 \cdot 2 \cdot 3 \cdot \cdots \cdot (n-1)}{2 \cdot 3 \cdot 4 \cdot \cdots \cdot n} \cdot \frac{3 \cdot 4 \cdot 5 \cdot \cdots \cdot (n+1)}{2 \cdot 3 \cdot 4 \cdot \cdots \cdot n}.$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

两式都呈大量对消. 第一式

$$\frac{1\cdot 2\cdot 3\cdot\cdots\cdot(n-1)}{2\cdot 3\cdot\cdots\cdot(n-1)\cdot n}=\frac{1}{n},$$

第二式

$$\frac{3\cdot 4\cdot 5\cdot\cdots\cdot n}{2\cdot 3\cdot 4\cdot\cdots\cdot n}\cdot(n+1)=\frac{n+1}{2}.$$

故

$$P_n=\frac{1}{n}\cdot\frac{n+1}{2}=\frac{n+1}{2n}.$$

例题 2.17 计算 $P=\prod_{k=2}^{63}\log_k(k+1)$.

解 通项是对数, 底数与真数都在变化, 无法直接计算. 处理不同底数对数的乘积, 标准思路是用**对数换底公式** $\log_a b=\frac{\ln b}{\ln a}$ 把各项统一到同一底(自然对数 ln). 于是 $\log_k(k+1)=\frac{\ln(k+1)}{\ln k}$, 原乘积变为

$$P=\prod_{k=2}^{63}\frac{\ln(k+1)}{\ln k}.$$

展开后得裂项求积形式.

$$P=\frac{\ln(3)}{\ln(2)}\cdot\frac{\ln(4)}{\ln(3)}\cdot\frac{\ln(5)}{\ln(4)}\cdot\cdots\cdot\frac{\ln(63)}{\ln(62)}\cdot\frac{\ln(64)}{\ln(63)}.$$

前一项的分子恰为后一项的分母, 依次抵消.

$$P=\frac{\cancel{\ln(3)}}{\ln(2)}\cdot\frac{\cancel{\ln(4)}}{\cancel{\ln(3)}}\cdot\frac{\cancel{\ln(5)}}{\cancel{\ln(4)}}\cdot\cdots\cdot\frac{\cancel{\ln(63)}}{\cancel{\ln(62)}}\cdot\frac{\ln(64)}{\cancel{\ln(63)}}=\frac{\ln(64)}{\ln(2)}.$$

反向使用换底公式, 或直接计算得

$$P=\log_2(64)=\log_2(2^6)=6.$$

例题 2.18 计算 $P_n=\prod_{k=1}^{n}\cos\left(\frac{x}{2^k}\right)$, 其中 $\sin(x/2^n)\neq 0$.

解 这个连乘积涉及三角函数, 常规因式分解不再适用, 需寻找三角恒等式来创造相消结构. 核心恒等式是正弦二倍角公式 $\sin(2\theta)=2\sin\theta\cos\theta$, 由此得 $\cos\theta=\frac{\sin(2\theta)}{2\sin\theta}$. 此式把一个 cos 项与两个不同角度的 sin 项联系起来, 结构上正适合裂项. 乘积中并无 sin 项, 故引入一个**辅助乘子**来启动连锁反应, 最合适的乘子是与乘积中角度最小的项 $\cos(x/2^n)$ 配套的 $\sin(x/2^n)$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

令 $P_n = \cos\left(\frac{x}{2}\right)\cos\left(\frac{x}{4}\right)\ldots\cos\left(\frac{x}{2^n}\right)$, 两边同乘 $2^n\sin(x/2^n)$, 得

$$\begin{aligned}2^n\sin\left(\frac{x}{2^n}\right)P_n &= 2^n\cos\left(\frac{x}{2}\right)\ldots\cos\left(\frac{x}{2^{n-1}}\right)\left[\sin\left(\frac{x}{2^n}\right)\cos\left(\frac{x}{2^n}\right)\right]\\ &= 2^{n-1}\cdot 2\cos\left(\frac{x}{2}\right)\ldots\left[\frac{1}{2}\sin\left(2\cdot\frac{x}{2^n}\right)\right]\\ &= 2^{n-1}\cos\left(\frac{x}{2}\right)\ldots\cos\left(\frac{x}{2^{n-1}}\right)\sin\left(\frac{x}{2^{n-1}}\right).\end{aligned}$$

最小角度的 sin, cos 项合并为角度大一倍的 sin 项, 此过程持续下去.

$$\begin{aligned}\ldots &= 2^{n-2}\cos\left(\frac{x}{2}\right)\ldots\sin\left(\frac{x}{2^{n-2}}\right)\\ &\vdots\\ &= 2\cos\left(\frac{x}{2}\right)\sin\left(\frac{x}{2}\right)\\ &= \sin(x).\end{aligned}$$

故 $2^n\sin\left(\frac{x}{2^n}\right)P_n = \sin(x)$. 由 $\sin(x/2^n)\neq 0$, 解出

$$P_n = \frac{\sin x}{2^n\sin(x/2^n)}.$$

例题 2.19　计算 $P_n = \prod_{k=2}^{n}\left(1-\frac{2}{k(k+1)}\right)$, 其中 $n\geq 2$.

解　先对通项作代数变形, 得

$$a_k = 1-\frac{2}{k(k+1)} = \frac{k(k+1)-2}{k(k+1)} = \frac{k^2+k-2}{k(k+1)},$$

分解分子得 $a_k = \frac{(k+2)(k-1)}{k(k+1)}$. 将此分式**分组**, 使每一组都构成简单的裂项结构.

$$a_k = \frac{k-1}{k}\cdot\frac{k+2}{k+1}.$$

原连乘积随之拆分为两个连乘积之积.

$$P_n = \left(\prod_{k=2}^{n}\frac{k-1}{k}\right)\cdot\left(\prod_{k=2}^{n}\frac{k+2}{k+1}\right).$$

分别计算. 第一个

$$\prod_{k=2}^{n}\frac{k-1}{k} = \frac{1}{2}\cdot\frac{2}{3}\cdot\frac{3}{4}\cdot\ldots\cdot\frac{n-1}{n} = \frac{1}{n},$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

第二个

$$\prod_{k=2}^{n}\frac{k+2}{k+1}=\frac{4}{3}\cdot\frac{5}{4}\cdot\frac{6}{5}\cdot\cdots\cdot\frac{n+2}{n+1}=\frac{n+2}{3}.$$

两部分相乘即得

$$P_n=\frac{1}{n}\cdot\frac{n+2}{3}=\frac{n+2}{3n}.$$

这些例题显示, 复杂的和式与积式经过适当变形, 往往会暴露出隐藏的结构. 其中用到的技巧按作用方式可归为几类.

最基础的操作是**变形与化归**, 它把陌生或复杂的形式通过等价变换变回熟悉的形式. **指标平移**就是一例, 令 $j=k+1$, 可把 $\sum(k+1)^3$ 写成标准立方和 $\sum j^3$. 多数例题的第一步是对通项做代数变形, 例如把 $\prod(1-1/k^2)$ 改写成 $\frac{(k-1)(k+1)}{k^2}$. 对数换底公式、三角二倍角公式、立方和/差公式等恒等式, 都能把通项改写成便于下一步处理的形式.

裂项是和式与积式求值中最常用的技巧. 它的做法是把 a_k 拆成 $f(k)-f(k+1)$ 或 $f(k)/f(k+1)$, 让中间项在累加或连乘中相互抵消.

构造裂项需要观察通项结构. 有理分式常用**部分分式分解**, 如 $\sum\frac{1}{k(k+2)}$; 复杂代数式则要利用恒等式, 例如用**索菲·热尔曼恒等式**分解 k^4+k^2+1. 有时把通项拆成多个因子的乘积, 每组各自构成简单裂项结构, 这就是**分组裂项**, 如 $\prod(1-\frac{2}{k(k+1)})$ 中的处理.

对二重和或多重和, 交换求和次序能化简内层计算. 例如 $\sum_{i=1}^{n}\sum_{j=i}^{n}$ 若按 i 先、j 后求和, 内层依赖外层; 改为先固定 j、再对 i 求和, 内层变量可能变成常数, 计算随之简化.

主动引入辅助因子可创造裂项或相约条件. 计算 $\prod\cos(x/2^k)$ 时, 乘上 $\sin(x/2^n)$ 后用二倍角公式 $\sin(2\theta)=2\sin\theta\cos\theta$ 逐级合并, 便得到简洁结果.

2.3 数列的表示

来龙去脉

在数学中, 当我们想研究一串按照特定规律排列的数时, 需一套精确的记法来描述它们. 数列, 就是这套记法的基础.

定义 2.3 (数列) 按一定次序排列的一列数称为**数列**. 数列中的每一个数都称为这个数列的**项**.
排在第一位的数称为第一项(或首项), 排在第二位的数称为第二项, ……, 排在第 n 位的数称为第 n 项.

知乎@余命数 知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

一个数列通常记作 $\{a_n\}$. 例如，奇数数列可以写成 $1,3,5,7,\ldots$. 其中，首项 $a_1=1$, 第二项 $a_2=3$, 等等.

要完全确定一个数列, 有两种主要的方式.

定义 2.4 (通项公式) 如果数列 $\{a_n\}$ 的第 n 项 a_n 与项数 n 之间的关系可以用一个公式 $a_n=f(n)$ 来表示, 那么这个公式就叫做这个数列的**通项公式**.

通项公式的优点是直截了当, 给定 n 就可直接算出对应的项.

定义 2.5 (递推公式) 如果已知数列 $\{a_n\}$ 的第一项, 或前几项, 且从第二项, 或某一项开始的任意一项 a_n 都可由它前面的一项, 或几项来确定, 那么这个关系式就叫做这个数列的**递推公式**.

递推公式描述相邻项之间的关系, 即从已知项推出下一项, 层层传递.

例题 2.20 对于奇数数列 $1,3,5,7,\ldots$

- **通项公式**是 $a_n=2n-1$. 例如, 第 100 项是 $a_{100}=2(100)-1=199$.
- **递推公式**是 $a_1=1, a_{n+1}=a_n+2\ (n\geq 1)$.

知乎@余命数

2.4 等差数列

来龙去脉

在所有数列中, 最简单的一种规律是每次固定增加或减少一个量. 比如, 楼房的层数、自然数的序列、储蓄罐里每天固定存入 10 元后的总额, 都是等差数列.

定义 2.6 (等差数列) 如果一个数列从第二项起, 每一项与它的前一项的差等于同一个常数, 那么这个数列就叫做**等差数列**. 这个常数叫做等差数列的**公差**, 通常用字母 d 表示.

根据定义, 公差 $d=a_{n+1}-a_n$.

- 当 $d>0$ 时, 后一项总比前一项大 $(a_{n+1}>a_n)$.
- 当 $d<0$ 时, 后一项总比前一项小 $(a_{n+1}<a_n)$.
- 当 $d=0$ 时, 数列是一个常数列, 所有项都相等 $(a_{n+1}=a_n)$.

知乎@余命数

2.4.1 通项公式与前 n 项和

2.4.1.1 通项公式

需建立一个从项数 n 到项的值 a_n 的直接联系, 从而不必一步步地从 a_1 开始计算.

我们从定义 $a_{k+1}=a_k+d$ 出发.

从首项 a_1 到达第 n 项 a_n 要走 $n-1$ 步, 每步相差一个公差 d.

$$
\begin{aligned}
a_2 &= a_1+d \\
a_3 &= a_2+d=(a_1+d)+d=a_1+2d \\
a_4 &= a_3+d=(a_1+2d)+d=a_1+3d \\
&\vdots
\end{aligned}
$$

注意到, 第 n 项 a_n 是在 a_1 的基础上, 累加了 $n-1$ 个公差 d 得到的.

一种更严谨的推导方法是**累加法**. 将定义式改写为 $a_{k+1}-a_k=d$. 然后, 将从 $k=1$ 到 $k=n-1$ 的所有等式写出来.

$$
\begin{aligned}
a_2-a_1 &= d \\
a_3-a_2 &= d \\
a_4-a_3 &= d \\
&\vdots \\
a_n-a_{n-1} &= d
\end{aligned}
$$

将这 $n-1$ 个等式全部相加, 左侧的中间项会两两抵消, 例如 $-a_2$ 与 $+a_2$, 这正是**裂项相消**.

$$
(a_2-a_1)+(a_3-a_2)+\cdots+(a_n-a_{n-1})=\underbrace{d+d+\cdots+d}_{n-1\text{ 个}}
$$

化简后, 左边只剩下首末两项的差, 右边是 $n-1$ 个 d 的和.

$$
a_n-a_1=(n-1)d
$$

移项后便得到等差数列的通项公式.

$$
a_n=a_1+(n-1)d \tag{2-1}
$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

直观的归纳和累加法给出了公式的图像. 要严格证明公式 $a_n = a_1 + (n-1)d$ 对**所有**正整数 n 成立, 可以使用**数学归纳法**. 先验证 $n = 1$ 时命题成立, 再证明若 $n = k$ 时成立则 $n = k + 1$ 也成立.

设命题 $P(n)$ 为 $a_n = a_1 + (n-1)d$.

当 $n = 1$ 时, 公式左边为 a_1, 右边为 $a_1 + (1-1)d = a_1$, 两端相等, 故 $P(1)$ 成立, 奠基步骤完成.

假设当 $n = k(k \geq 1$ 且 k 为整数)时命题成立, 即 $a_k = a_1 + (k-1)d$. 由此只需证明当 $n = k + 1$ 时 $a_{k+1} = a_1 + ((k+1)-1)d = a_1 + kd$ 也成立.

由等差数列的定义 $a_{k+1} = a_k + d$ 出发, 代入归纳假设得

$$\begin{aligned} a_{k+1} &= (a_1 + (k-1)d) + d \quad \text{（代入归纳假设）} \\ &= a_1 + kd - d + d \\ &= a_1 + kd, \end{aligned}$$

与归纳目标 $a_{k+1} = a_1 + kd$ 一致, 归纳步骤成立. 故若命题对 $n = k$ 成立, 则对 $n = k + 1$ 亦成立.

由数学归纳法原理, 命题 $P(n)$ 对所有正整数 n 均成立, 通项公式 $a_n = a_1 + (n-1)d$ 得证.

2.4.1.2 前 n 项和公式的推导与视角

求 $S_n = a_1 + a_2 + \cdots + a_n$ 有多种推导途径, 下面给出三种方法.

法一采用倒序相加.

倒序相加相传由年幼的高斯发现, 要点是创造对称性.

将和 S_n 写两遍, 第二遍顺序颠倒.

$$\begin{aligned} S_n &= a_1 + a_2 + \cdots + a_{n-1} + a_n \\ S_n &= a_n + a_{n-1} + \cdots + a_2 + a_1 \end{aligned}$$

将这两式按位相加, 得到.

$$2S_n = (a_1 + a_n) + (a_2 + a_{n-1}) + \cdots + (a_{n-1} + a_2) + (a_n + a_1)$$

考察每一对的和. 对于第 k 对 $(a_k + a_{n-k+1})$:

$$\begin{aligned} a_k + a_{n-k+1} &= [a_1 + (k-1)d] + [a_1 + (n-k)d] \\ &= 2a_1 + (k - 1 + n - k)d \\ &= 2a_1 + (n-1)d \end{aligned}$$

知乎@余命数

这个结果是一个与 k 无关的常数, 恰好等于首项与末项之和 a_1+a_n. 由于一共有 n 对这样的和, 得

$$2S_n = n(a_1+a_n)$$

于是得第一个求和公式, 它连接了和、项数、首项与末项.

$$S_n = \frac{n(a_1+a_n)}{2} \tag{2-2}$$

法二采用代入与分组.

此法说明求和公式中两个组成部分的来源. 将通项公式 $a_k = a_1+(k-1)d$ 直接代入和式 S_n 的每一项, 得

$$\begin{aligned} S_n &= a_1+a_2+a_3+\cdots+a_n \\ &= a_1+[a_1+d]+[a_1+2d]+\cdots+[a_1+(n-1)d] \end{aligned}$$

接着, 将式子中的 a_1 和包含 d 的部分分开重新组合, 得

$$S_n = \underbrace{(a_1+a_1+\cdots+a_1)}_{n\text{个}} + \underbrace{[0\cdot d+1\cdot d+2\cdot d+\cdots+(n-1)d]}_{\text{公差部分}}$$

第一部分显然是 na_1. 第二部分可以提出公因子 d:

$$d[1+2+\cdots+(n-1)]$$

括号内是前 $n-1$ 个正整数的和, 这是一个我们熟知的简单等差数列, 其和为 $\frac{(n-1)n}{2}$. 将两部分加起来, 我们便得到求和公式的第二种形式.

$$S_n = na_1+\frac{n(n-1)}{2}d \tag{2-3}$$

法三从 S_n 的函数结构入手.

我们在提示框中提到, S_n 是关于 n 的一个没有常数项的二次函数. 可利用这一点, 反过来推导出公式本身. 设 $S_n = An^2+Bn$. 我们的任务是确定系数 A 和 B. 我们利用数列的一个基本关系. 当 $n\geq 2$ 时, $a_n = S_n - S_{n-1}$.

$$\begin{aligned} a_n &= (An^2+Bn)-[A(n-1)^2+B(n-1)] \\ &= (An^2+Bn)-[A(n^2-2n+1)+B(n-1)] \\ &= An^2+Bn-An^2+2An-A-Bn+B \\ &= (2A)n+(B-A) \end{aligned}$$

知乎@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这个结果表明，a_n 是关于 n 的一次式，这正是等差数列的特征. 比较通项公式 $a_n = dn + (a_1 - d)$，得系数的对应关系.

$$公差 d = 2A \implies A = \frac{d}{2}$$

$$a_1 = (2A)\cdot 1 + (B - A) = A + B \implies B = a_1 - A = a_1 - \frac{d}{2}$$

将求出的 A 和 B 代回 $S_n = An^2 + Bn$：

$$S_n = \frac{d}{2}n^2 + \left(a_1 - \frac{d}{2}\right)n$$

整理后即得

$$S_n = na_1 + \frac{n(n-1)}{2}d$$

知乎@余命数

注 2.2 (等差数列的函数视角)

- **通项公式的本质**.$a_n = dn + (a_1 - d)$. 故 a_n 是关于项数 n 的**一次函数**.
- **求和公式的本质**.$S_n = \frac{d}{2}n^2 + (a_1 - \frac{d}{2})n$. 故 S_n 是关于项数 n 的**二次函数**，且其**常数项为零**.
- 这一视角在判断题、选择题中能提供更宏观的判断依据.

2.5 等比数列

知乎@余命数

来龙去脉

等差数列刻画求和的规律，等比数列刻画求积的规律. 现实中翻倍、按比例增长或衰减的模型，在米粒与棋盘问题里已有体现.

知乎@余命数

定义 2.7 (等比数列) 如果一个数列从第二项起，每一项与它的前一项的比等于同一个常数，那么这个数列就叫做**等比数列**. 这个常数叫做等比数列的**公比**，通常用字母 q 表示 $(q \neq 0)$.

根据定义，公比 $q = \dfrac{a_{n+1}}{a_n}$.

2.5.1 通项公式与前 n 项和

2.5.1.1 通项公式

等差数列累加公差得到通项, 等比数列则连乘公比. 目标同样是建立项数 n 到项值 a_n 的直接关系.

法一采用注意力法.

这个方法很直观. 从首项 a_1 到第 n 项 a_n 需要连乘 $n-1$ 次公比 q.

$$
\begin{aligned}
a_2 &= a_1 \cdot q = a_1 q^1 \\
a_3 &= a_2 \cdot q = (a_1 q^1) \cdot q = a_1 q^2 \\
a_4 &= a_3 \cdot q = (a_1 q^2) \cdot q = a_1 q^3 \\
&\vdots
\end{aligned}
$$

规律是, 第 n 项 a_n 中, 公比 q 的指数比项数 n 小 1. 由此, 我们归纳出等比数列的通项公式.

$$
a_n = a_1 q^{n-1} \tag{2-4}
$$

法二采用累乘法.

这种方法与等差数列的累加法完全对称. 等比数列的定义式可写作 $\dfrac{a_{k+1}}{a_k} = q$. 将从 $k=1$ 到 $k=n-1$ 的所有等式写出来.

$$
\begin{aligned}
\frac{a_2}{a_1} &= q \\
\frac{a_3}{a_2} &= q \\
\frac{a_4}{a_3} &= q \\
&\vdots \\
\frac{a_n}{a_{n-1}} &= q
\end{aligned}
$$

将这 $n-1$ 个等式全部相乘, 左边的分子和分母依次约去, 这正是乘法中的**裂项相消**.

$$
\frac{a_2}{a_1} \cdot \frac{a_3}{a_2} \cdot \frac{a_4}{a_3} \cdot \cdots \cdot \frac{a_n}{a_{n-1}} = \underbrace{q \cdot q \cdot \cdots \cdot q}_{n-1\text{个}}
$$

化简后, 左边只剩下首末两项的比, 右边是 $n-1$ 个 q 的乘积.

$$
\frac{a_n}{a_1} = q^{n-1}
$$

两边同乘 a_1, 再次得到通项公式.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

法三采用数学归纳法.

要严格证明该公式对所有正整数 n 成立, 可以使用数学归纳法.

设命题 $P(n)$ 为 $a_n = a_1 q^{n-1}$.

当 $n = 1$ 时, 公式左边为 a_1, 右边为 $a_1 q^{1-1} = a_1 q^0 = a_1$ (约定 $q \neq 0$), 两端相等, 故 $P(1)$ 成立, 奠基步骤完成.

归纳假设 $a_k = a_1 q^{k-1}$ ($k \geq 1$ 为整数), 归纳目标是证明 $a_{k+1} = a_1 q^{(k+1)-1} = a_1 q^k$.

由等比数列定义 $a_{k+1} = a_k \cdot q$ 出发, 代入归纳假设得

$$\begin{aligned} a_{k+1} &= (a_1 q^{k-1}) \cdot q \quad \text{（代入归纳假设）} \\ &= a_1 q^{(k-1)+1} \\ &= a_1 q^k, \end{aligned}$$

与归纳目标一致, 归纳步骤成立. 故若命题对 $n = k$ 成立, 则对 $n = k + 1$ 亦成立.

由数学归纳法原理, 命题 $P(n)$ 对所有正整数 n 均成立.

2.6 差比数列

我们已经分别讨论了等差数列与等比数列. 现在把两者对应项相乘, 得到一类新的数列.

若一个数列 $\{c_n\}$ 的每一项都是某个等差数列的对应项与某个等比数列的对应项的乘积, 我们就称它为**差比数列**. 其通项同时带有等差与等比的特征, 而求和通常使用错位相减法.

定义 2.8 (差比数列) 一个数列 $\{c_n\}$ 若能表示为 $c_n = a_n \cdot b_n$ 的形式, 其中 $\{a_n\}$ 是一个首项为 a_1、公差为 d 的等差数列, $\{b_n\}$ 是一个首项为 b_1、公比为 q 的等比数列, 则称 $\{c_n\}$ 为**差比数列**. 其通项公式为

$$c_n = (a_1 + (n-1)d) \cdot b_1 q^{n-1}$$

例如, 我们在前面求和过的数列 $\{n \cdot 2^n\}$, 就是由等差数列 $\{n\}$, 其中 $a_1 = 1, d = 1$ 和等比数列 $\{2^n\}$, 其中 $b_1 = 2, q = 2$ 相乘得到的.

2.6.1 差比求和

我们来求差比数列 $\{(a_1 + (n-1)d) \cdot b_1 q^{n-1}\}$ 的前 n 项和. 标准工具是错位相减法. 其做法是**用等比数列部分的公比 q 去乘 S_n, 再与原式相减, 使相减后的新数列化简为一个等比数列的和**.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

我们先推导一般情形. 为简化书写, 设差比数列为 $c_n=(a+(n-1)d)r^{n-1}$. 令 $S_n=\sum_{k=1}^{n}(a+(k-1)d)r^{k-1}$.

$$S_n=a+(a+d)r+(a+2d)r^2+\cdots+(a+(n-1)d)r^{n-1}$$

两边同乘以公比 r $(r\neq 1)$, 得

$$rS_n=\qquad ar+(a+d)r^2+\cdots+(a+(n-2)d)r^{n-1}+(a+(n-1)d)r^n$$

两式相减, 得

$$\begin{aligned}(1-r)S_n&=a+[(a+d)r-ar]+[(a+2d)r^2-(a+d)r^2]+\ldots\\&\quad+[(a+(n-1)d)r^{n-1}-(a+(n-2)d)r^{n-1}]-(a+(n-1)d)r^n\\(1-r)S_n&=a+[dr+dr^2+\cdots+dr^{n-1}]-(a+(n-1)d)r^n\end{aligned}$$

观察上式, 中括号内的部分是一个首项为 dr, 公比为 r 的等比数列, 共有 $n-1$ 项.

$$[dr+dr^2+\cdots+dr^{n-1}]=\frac{dr(1-r^{n-1})}{1-r}$$

代入后得到关于 S_n 的方程, 解出 S_n 即可. 这个一般公式较复杂, 我们重点理解推导过程.

注 2.3 (差比数列求和要领)
求差比数列的和时, 我们先把通项拆成等差部分与等比部分, 再用公比 q 乘和式并作**错位相减**. 相减后的结果通常呈 **首项 + 等比数列 + 尾项**结构, 从而化归成熟悉的形式.

例题 2.21 求数列 $\{(2n+1)\cdot 3^n\}$ 的前 n 项和 S_n.

解 该数列是等差数列 $\{2n+1\}$(首项 3, 公差 2)与等比数列 $\{3^n\}$ 的乘积, 后者首项为 3, 公比为 3.

写出和式 S_n.

$$S_n=3\cdot 3^1+5\cdot 3^2+7\cdot 3^3+\cdots+(2n+1)3^n$$

两边同乘以公比 3, 得

$$3S_n=\qquad 3\cdot 3^2+5\cdot 3^3+\cdots+(2n-1)3^n+(2n+1)3^{n+1}$$

上式减下式, 得

$$S_n-3S_n=3\cdot 3^1+(5-3)3^2+(7-5)3^3+\ldots$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$+((2n+1)-(2n-1))3^n-(2n+1)3^{n+1}$$

$$-2S_n=9+2\cdot 3^2+2\cdot 3^3+\cdots+2\cdot 3^n$$

$$-(2n+1)3^{n+1}$$

$$-2S_n=9+2(3^2+3^3+\cdots+3^n)-(2n+1)3^{n+1}$$

知乎@余命数

括号内是一个首项为 $3^2=9$, 公比为 3, 项数为 $n-1$ 的等比数列.

知乎@余命数

$$3^2+3^3+\cdots+3^n=\frac{9(3^{n-1}-1)}{3-1}=\frac{9}{2}(3^{n-1}-1)=\frac{1}{2}(3^{n+1}-9)$$

知乎@余命数

代回原式, 得

$$-2S_n=9+2\cdot\frac{1}{2}(3^{n+1}-9)-(2n+1)3^{n+1}$$

$$-2S_n=9+3^{n+1}-9-(2n+1)3^{n+1}$$

$$-2S_n=(1-(2n+1))3^{n+1}$$

$$-2S_n=-2n\cdot 3^{n+1}$$

两边同除以 -2, 最终得到

$$S_n=n\cdot 3^{n+1}$$

知乎@余命数　知乎@余命数

2.6.2 苹果公式

知乎@余命数

错位相减法需要构造辅助等式. 另一种思路是用待定系数法把差比数列的通项 c_k 拆成 $F(k)-F(k-1)$, 通过裂项相消求和.

知乎@余命数

把这种做法推广, 可以得到适用于一般差比数列的求和公式. 该公式直接给出待定系数法的结果, 代入参数即可计算.

定理 2.1 (差比数列求和公式) 对于一个通项为 $c_k=(ak+b)q^{k-1}$ 的差比数列, 其中 a,b,q 为常数, $a\neq 0, q\neq 1$, 其前 n 项和为

$$S_n=\sum_{k=1}^{n}(ak+b)q^{k-1}=(An+B)q^n-B$$

其中系数 A 与 B 由下式确定

$$A=\frac{a}{q-1},\quad B=\frac{b-A}{q-1}=\frac{b(q-1)-a}{(q-1)^2}=\frac{bq-b-a}{(q-1)^2}$$

知乎@余命数

证明 我们用待定系数裂项法来证明. 寻找一个函数 $F(k)$, 使差比数列的通项 $c_k=(ak+b)q^{k-1}$ 等于其差分 $F(k)-F(k-1)$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

通项 c_k 是线性多项式 $(ak+b)$ 与指数项 q^{k-1} 的乘积, 因此反差分函数 $F(k)$ 也应为多项式乘以指数项的形式. 设

$$F(k)=(Pk+Q)q^k,$$

其中 P,Q 为待定常数. 计算其差分并令其与 c_k 相等, 得

$$\begin{aligned}F(k)-F(k-1)&=(Pk+Q)q^k-[P(k-1)+Q]q^{k-1}\\&=q(Pk+Q)q^{k-1}-[P(k-1)+Q]q^{k-1}\\&=[q(Pk+Q)-(Pk-P+Q)]q^{k-1}\\&=[(Pq-P)k+(P+Qq-Q)]q^{k-1}\\&=[P(q-1)k+P+Q(q-1)]q^{k-1}.\end{aligned}$$

要求此式恒等于 $c_k=(ak+b)q^{k-1}$.

$$[P(q-1)k+P+Q(q-1)]q^{k-1}=(ak+b)q^{k-1}.$$

比较两端 q^{k-1} 的系数多项式中 k 的同次幂系数, 得到关于 P,Q 的方程组.

- k 的一次项系数为 $P(q-1)=a \implies P=\dfrac{a}{q-1}$.
- 常数项为 $P+Q(q-1)=b \implies Q(q-1)=b-P \implies Q=\dfrac{b-P}{q-1}$.

这正是定理中系数 A 和 B 的定义, 即 $A=P,B=Q$.

通项 c_k 已分解为 $F(k)-F(k-1)$ 的形式, 求和随之成为标准的裂项相消.

$$\begin{aligned}S_n=\sum_{k=1}^{n}c_k&=\sum_{k=1}^{n}[F(k)-F(k-1)]\\&=(F(1)-F(0))+(F(2)-F(1))+\cdots+(F(n)-F(n-1))\\&=F(n)-F(0).\end{aligned}$$

将 $F(k)=(Ak+B)q^k$ 代入首末两项. 于是 $F(n)=(An+B)q^n$,$F(0)=(A\cdot 0+B)q^0=B$. 故

$$S_n=(An+B)q^n-B.$$

证毕. □

证明 (Abel 变换证法) 考虑和式 $S_n=\sum_{k=1}^{n}(ak+b)q^{k-1}$. 阿贝尔变换把求和项 c_k 写成 u_kv_k, 让其中一部分易于求和, 另一部分易于求差.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

阿贝尔变换公式为

$$\sum_{k=1}^{n} u_k v_k = U_n v_n - \sum_{k=1}^{n-1} U_k (v_{k+1} - v_k), \quad \text{其中} U_k = \sum_{i=1}^{k} u_i$$

对通项 $c_k = (ak+b)q^{k-1}$, 可取

- 令 $v_k = ak + b$, 这是等差数列项, 其差分 $v_{k+1} - v_k$ 为常数, 能简化新的求和.
- 令 $u_k = q^{k-1}$, 这是等比数列项, 其前 k 项和 U_k 有简洁的封闭形式.

接着, 计算阿贝尔变换所需的各个构件, 得

- 差分项 $v_{k+1} - v_k = [a(k+1)+b] - [ak+b] = a$.
- 部分和 $U_k = \sum_{i=1}^{k} u_i = \sum_{i=1}^{k} q^{i-1} = \frac{q^k-1}{q-1}$.
- 最终项 $v_n = an + b$ 以及 $U_n = \frac{q^n-1}{q-1}$.

将这些构件代入阿贝尔变换公式, 得

$$\begin{aligned} S_n &= U_n v_n - \sum_{k=1}^{n-1} U_k (v_{k+1} - v_k) \\ &= \left(\frac{q^n-1}{q-1}\right)(an+b) - \sum_{k=1}^{n-1} \left(\frac{q^k-1}{q-1}\right) a \\ &= \frac{(an+b)(q^n-1)}{q-1} - \frac{a}{q-1} \sum_{k=1}^{n-1} (q^k - 1) \end{aligned}$$

接下来计算新的求和部分, 得

$$\begin{aligned} \sum_{k=1}^{n-1} (q^k - 1) &= \left(\sum_{k=1}^{n-1} q^k\right) - \left(\sum_{k=1}^{n-1} 1\right) \\ &= \frac{q(q^{n-1}-1)}{q-1} - (n-1) \end{aligned}$$

将此结果代回 S_n 的表达式并整理, 目标是分离出含 q^n 的项与常数项.

$$\begin{aligned} S_n &= \frac{(an+b)q^n - (an+b)}{q-1} - \frac{a}{q-1}\left[\frac{q^n-q}{q-1} - (n-1)\right] \\ &= \frac{(an+b)q^n}{q-1} - \frac{an+b}{q-1} - \frac{a(q^n-q)}{(q-1)^2} + \frac{a(n-1)}{q-1} \end{aligned}$$

首先, 合并所有包含 q^n 的项, 得

$$\begin{aligned} q^n \left[\frac{an+b}{q-1} - \frac{a}{(q-1)^2}\right] &= q^n \left[\frac{(an+b)(q-1)-a}{(q-1)^2}\right] \\ &= q^n \left[\frac{an(q-1)+b(q-1)-a}{(q-1)^2}\right] \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= q^n \left[\frac{a(q-1)}{(q-1)^2} n + \frac{bq-b-a}{(q-1)^2} \right]$$
$$= \left(\frac{a}{q-1} n + \frac{bq-b-a}{(q-1)^2} \right) q^n$$

这恰好是 $(An+B)q^n$ 的形式, 其中 $A = \dfrac{a}{q-1}$ 且 $B = \dfrac{bq-b-a}{(q-1)^2}$.

接着, 合并所有不含 q^n 的项, 即常数部分.

$$\begin{aligned}
&-\frac{an+b}{q-1} - \frac{-aq}{(q-1)^2} + \frac{a(n-1)}{q-1} \\
&= \frac{-(an+b)+a(n-1)}{q-1} + \frac{aq}{(q-1)^2} \\
&= \frac{-(an+b)(q-1)+a(n-1)(q-1)+aq}{(q-1)^2} \\
&= \frac{-an(q-1)-b(q-1)+an(q-1)-a(q-1)+aq}{(q-1)^2} \\
&= \frac{-b(q-1)-a(q-1)+aq}{(q-1)^2} \\
&= \frac{-bq+b-aq+a+aq}{(q-1)^2} = \frac{-bq+b+a}{(q-1)^2} \\
&= -\left(\frac{bq-b-a}{(q-1)^2} \right) = -B
\end{aligned}$$

将两部分结果组合起来, 最终得到

$$S_n = (An+B)q^n - B$$

□

证明 (Gosper 裂项法证法) Gosper 算法系统地寻找给定通项 c_k 的反差分函数 $F(k)$, 使得 $c_k = F(k) - F(k-1)$. 一旦找到 $F(k)$, 求和便可通过裂项相消 $S_n = F(n) - F(0)$ 直接得到.

通项 $c_k = (ak+b)q^{k-1}$ 是典型的超几何项(其相邻项之比为有理函数).Gosper 算法断言, 其反差分函数 $F(k)$ 若存在, 则必具相似结构, 即一个多项式乘以一个指数项.

求和公式的结果含 $(An+B)q^n$, 因此我们猜测反差分函数 $F(k)$ 具 $F(k) = (Pk+Q)q^k$ 的形式, 其中 P, Q 为待定常数. 再求解差分方程定出 P 与 Q.

设 $F(k) = (Pk+Q)q^k$ 并计算其差分, 得

$$\begin{aligned}
F(k) - F(k-1) &= (Pk+Q)q^k - [P(k-1)+Q]q^{k-1} \\
&= q \cdot (Pk+Q)q^{k-1} - [Pk-P+Q]q^{k-1}
\end{aligned}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= [q(Pk+Q) - (Pk - P + Q)]q^{k-1}$$
$$= [(Pq - P)k + (P + Qq - Q)]q^{k-1}$$
$$= [P(q-1)k + P + Q(q-1)]q^{k-1}$$

要求这个差分结果恒等于原数列的通项 $c_k = (ak+b)q^{k-1}$.

$$[P(q-1)k + P + Q(q-1)]q^{k-1} = (ak+b)q^{k-1}$$

这是一个关于 k 的多项式恒等式. 通过比较等式两边 k 的同次幂的系数, 可建立关于 P, Q 的线性方程组.

- 比较 k^1 的系数, 得

$$P(q-1) = a \implies P = \frac{a}{q-1}$$

- 比较 k^0(常数项)的系数, 得

$$P + Q(q-1) = b \implies Q(q-1) = b - P \implies Q = \frac{b-P}{q-1}$$

将求得的 P 代入 Q 的表达式中, 得

$$Q = \frac{b - \frac{a}{q-1}}{q-1} = \frac{\frac{b(q-1)-a}{q-1}}{q-1} = \frac{b(q-1)-a}{(q-1)^2} = \frac{bq-b-a}{(q-1)^2}$$

因此, 用这个系统的方法求出的常数 P, Q 与定理中给出的系数 A, B 完全一致, 即 $A = P, B = Q$.

我们已找到反差分函数 $F(k) = (Ak+B)q^k$. 根据离散微积分基本定理, 和式 S_n 等于 $F(n)$ 与 $F(0)$ 之差.

$$S_n = \sum_{k=1}^{n} c_k = F(n) - F(0)$$

计算 $F(n)$ 和 $F(0)$, 得

- $F(n) = (An+B)q^n$
- $F(0) = (A \cdot 0 + B)q^0 = B$

因此, 得

$$S_n = (An+B)q^n - B$$

□

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

例题 2.22 求数列 $\{(2k-1)\cdot 3^{k-1}\}$ 的前 n 项和 S_n.

解 先识别数列的类型. 其通项 $c_k=(2k-1)\cdot 3^{k-1}$ 是等差数列项 $(2k-1)$ 与等比数列项 3^{k-1} 的乘积, 因此是一个标准的差比数列.

我们可以直接套用差比数列求和公式. 先将通项与标准形式 $(ak+b)q^{k-1}$ 比对, 确定参数 a,b,q.

- 线性部分 $2k-1$ 对应 $ak+b$, 可得 $a=2,b=-1$.
- 指数部分 3^{k-1} 对应 q^{k-1}, 可得 $q=3$.

接着, 计算公式所需的中间系数 A 和 B.

$$A=\frac{a}{q-1}=\frac{2}{3-1}=1$$

$$B=\frac{bq-b-a}{(q-1)^2}=\frac{(-1)(3)-(-1)-2}{(3-1)^2}=\frac{-3+1-2}{4}=\frac{-4}{4}=-1$$

最后, 将求得的 A,B 以及参数 q 代入求和公式 $S_n=(An+B)q^n-B$, 得

$$\begin{aligned}S_n&=(1\cdot n+(-1))\cdot 3^n-(-1)\\&=(n-1)3^n+1\end{aligned}$$

因此, 该数列的前 n 项和为 $S_n=(n-1)3^n+1$.

验证. 为检验结果, 我们计算前两项的和. $S_2=c_1+c_2=(2\cdot 1-1)3^0+(2\cdot 2-1)3^1=1+3\cdot 3=10$. 将 $n=2$ 代入推导的公式, 得 $S_2=(2-1)3^2+1=1\cdot 9+1=10$. 结果吻合.

例题 2.23 求数列 $\{k\cdot 2^k\}$ 的前 n 项和 S_n.

解 该数列的通项 $c_k=k\cdot 2^k$ 也是差比数列, 但指数 k 与标准公式 $(ak+b)q^{k-1}$ 不一致. 套用公式前要先做代数变形.

目标是把 2^k 化为含 q^{k-1} 的形式.

$$2^k=2\cdot 2^{k-1}$$

将此代回通项表达式, 得

$$c_k=k\cdot(2\cdot 2^{k-1})=(2k)\cdot 2^{k-1}$$

经过变形, 通项现在完全符合 $(ak+b)q^{k-1}$ 的结构. 进行参数比对.

- 线性部分 $2k$ 对应 $ak+b$, 可得 $a=2,b=0$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 指数部分 2^{k-1} 对应 q^{k-1}, 可得 $q=2$.

  接着, 计算系数 A, B.

$$A=\frac{a}{q-1}=\frac{2}{2-1}=2$$

$$B=\frac{bq-b-a}{(q-1)^2}=\frac{(0)(2)-0-2}{(2-1)^2}=\frac{-2}{1}=-2$$

  将 A, B, q 代入求和公式 $S_n=(An+B)q^n-B$, 得

$$\begin{aligned}S_n&=(2n-2)\cdot 2^n-(-2)\\&=2(n-1)\cdot 2^n+2\\&=(n-1)\cdot 2^{n+1}+2\end{aligned}$$

  因此, 该数列的前 n 项和为 $S_n=(n-1)2^{n+1}+2$.

例题 2.24　求和式 $S_n=\sum_{k=1}^{n}\frac{k+1}{3^k}$ 的值.

解　该和式的通项 $c_k=\dfrac{k+1}{3^k}$ 要先化成差比数列形式. 利用求和的线性性质, 把通项拆成熟悉的部分之和.

将通项 c_k 拆分.

$$c_k=\frac{k}{3^k}+\frac{1}{3^k}$$

因此, 原求和式可分解为两个独立和式之和.

$$S_n=\sum_{k=1}^{n}\frac{k}{3^k}+\sum_{k=1}^{n}\frac{1}{3^k}$$

分别计算这两个和式.

令 $T_n=\sum_{k=1}^{n}\frac{k}{3^k}$. 其通项为 $d_k=k\cdot\left(\frac{1}{3}\right)^k$. 这是一个差比数列. 为了应用公式, 将其变形为标准形式 $(ak+b)q^{k-1}$.

$$d_k=k\cdot\frac{1}{3}\cdot\left(\frac{1}{3}\right)^{k-1}=\left(\frac{1}{3}k\right)\cdot\left(\frac{1}{3}\right)^{k-1}$$

与标准形式比对, 得参数 $a=\frac{1}{3}, b=0, q=\frac{1}{3}$. 计算系数 A, B, 得

$$A=\frac{a}{q-1}=\frac{1/3}{1/3-1}=\frac{1/3}{-2/3}=-\frac{1}{2}$$

知乎@余命数　知乎@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$B=\frac{bq-b-a}{(q-1)^2}=\frac{0-0-1/3}{(1/3-1)^2}=\frac{-1/3}{(-2/3)^2}=\frac{-1/3}{4/9}=-\frac{3}{4}$$

代入公式 $T_n=(An+B)q^n-B$, 得

$$T_n=\left(-\frac{1}{2}n-\frac{3}{4}\right)\left(\frac{1}{3}\right)^n-\left(-\frac{3}{4}\right)=\frac{3}{4}-\left(\frac{n}{2}+\frac{3}{4}\right)\frac{1}{3^n}$$

令 $G_n=\sum_{k=1}^{n}\frac{1}{3^k}=\sum_{k=1}^{n}\left(\frac{1}{3}\right)^k$. 这是一个首项为 $1/3$, 公比为 $1/3$ 的等比数列.

$$G_n=\frac{\frac{1}{3}\left(1-(\frac{1}{3})^n\right)}{1-\frac{1}{3}}=\frac{\frac{1}{3}\left(1-\frac{1}{3^n}\right)}{2/3}=\frac{1}{2}\left(1-\frac{1}{3^n}\right)=\frac{1}{2}-\frac{1}{2\cdot 3^n}$$

将两部分相加.

$$\begin{aligned}S_n&=T_n+G_n\\&=\left[\frac{3}{4}-\left(\frac{n}{2}+\frac{3}{4}\right)\frac{1}{3^n}\right]+\left[\frac{1}{2}-\frac{1}{2\cdot 3^n}\right]\\&=\frac{3}{4}+\frac{1}{2}-\left(\frac{n}{2}+\frac{3}{4}+\frac{1}{2}\right)\frac{1}{3^n}\\&=\frac{5}{4}-\left(\frac{n}{2}+\frac{5}{4}\right)\frac{1}{3^n}\\&=\frac{5}{4}-\frac{2n+5}{4\cdot 3^n}\end{aligned}$$

例题 2.25 已知数列 $\{c_n\}$ 的前 n 项和 $S_n=(n-1)2^n+1$.

1. 求数列 $\{c_n\}$ 的通项公式.
2. 求数列 $\{(-1)^{k+1}c_k\}$ 的前 n 项和 T_n.

解 (1) 这是一个由前 n 项和 S_n 反求通项 c_n 的问题. 利用基本关系式 $c_n=S_n-S_{n-1}$, 当 $n\geq 2$ 时.

$$S_{n-1}=((n-1)-1)2^{n-1}+1=(n-2)2^{n-1}+1$$

对于 $n\geq 2$:

$$\begin{aligned}c_n&=S_n-S_{n-1}\\&=[(n-1)2^n+1]-[(n-2)2^{n-1}+1]\\&=(n-1)\cdot 2\cdot 2^{n-1}-(n-2)2^{n-1}\\&=[2(n-1)-(n-2)]2^{n-1}\\&=(2n-2-n+2)2^{n-1}\end{aligned}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= n \cdot 2^{n-1}$$

需检验此公式在 $n=1$ 时是否成立. 当 $n=1$ 时, $c_1 = S_1 = (1-1)2^1 + 1 = 1$. 将 $n=1$ 代入推导的公式, $c_1 = 1 \cdot 2^{1-1} = 1 \cdot 2^0 = 1$. 公式对 $n=1$ 也成立. 因此, 数列 $\{c_n\}$ 的通项公式为 $c_n = n \cdot 2^{n-1}$.

(2) 需求和的数列是 $\{d_k\}$, 其通项为 $d_k = (-1)^{k+1}c_k = (-1)^{k+1}k \cdot 2^{k-1}$. 这个通项看似复杂, 仍是差比数列. 重组后即可匹配标准形式.

$$\begin{aligned} d_k &= k \cdot [(-1)^{k+1}2^{k-1}] \\ &= k \cdot [(-1)^2(-1)^{k-1}2^{k-1}] \\ &= k \cdot [(-2)^{k-1}] \end{aligned}$$

知乎@余命数

现在通项 $d_k = k \cdot (-2)^{k-1}$ 是一个标准形式的差比数列. 与 $(ak+b)q^{k-1}$ 比对, 得参数 $a=1, b=0, q=-2$.

计算系数 A, B, 得

$$A = \frac{a}{q-1} = \frac{1}{-2-1} = -\frac{1}{3}$$

$$B = \frac{bq-b-a}{(q-1)^2} = \frac{0-0-1}{(-2-1)^2} = -\frac{1}{9}$$

代入求和公式 $T_n = (An+B)q^n - B$:

$$\begin{aligned} T_n &= \left(-\frac{1}{3}n - \frac{1}{9}\right)(-2)^n - \left(-\frac{1}{9}\right) \\ &= -\frac{3n+1}{9}(-2)^n + \frac{1}{9} \\ &= \frac{1-(3n+1)(-2)^n}{9} \end{aligned}$$

此即为所求的前 n 项和.

例题 2.26 (当公比为 1 时) 一个数列的通项为 $c_k = (4k-3) \cdot 1^k$. 求其前 n 项和 S_n.

解 初看此题, 通项 $c_k = (4k-3) \cdot 1^k$ 像是线性部分 $(4k-3)$ 与指数部分 1^k 的乘积, 容易让人套用差比数列求和公式.

作参数比对, 先将通项变形为标准形式 $(ak+b)q^{k-1}$. 于是

$$c_k = (4k-3) \cdot 1^k = (4k-3) \cdot 1 \cdot 1^{k-1},$$

可得 $a=4, b=-3, q=1$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

随之计算系数 A, 得

$$A=\frac{a}{q-1}=\frac{4}{1-1}=\frac{4}{0},$$

此时出现除以零的错误. 差比数列求和公式在此失效, 因为公式推导(无论是错位相减法还是待定系数法)都隐含 $q\neq 1$ 的前提. 当 $q=1$ 时, 分母 $(q-1)$ 为零, 公式不再成立.

须回到数列本身分析. 当公比 $q=1$ 时, 指数部分 $q^{k-1}=1^{k-1}=1$ 对所有 k 恒成立, 等比部分退化为常数数列 $\{1,1,1,\dots\}$, 因此原数列的通项 c_k 发生**退化**.

知乎@余命数

$$c_k=(4k-3)\cdot 1^{k-1}=4k-3,$$

知乎@余命数

实质上是一个早已熟悉的**等差数列**.

问题随之化归为求解首项 $c_1=4(1)-3=1$、公差 $d=4$ 的等差数列前 n 项和. 直接使用等差数列求和公式

$$S_n=\frac{n(c_1+c_n)}{2},$$

其中末项 $c_n=4n-3$, 得

$$S_n=\frac{n(1+(4n-3))}{2}=\frac{n(4n-2)}{2}=n(2n-1)=2n^2-n.$$

注 2.4 (启示)
公式都有适用范围. 差比数列求和公式要求 $q\neq 1$. 当 $q=1$ 时, 差比数列退化为等差数列. 解题时应先确定数列类型, 再选择对应公式.

例题 2.27 (综合应用) 已知数列 $\{c_n\}$ 满足关系式 $\sum_{k=1}^{n}\frac{c_k}{3^k}=2n^2+n$.

1. 求数列 $\{c_n\}$ 的通项公式.
2. 设 $d_n=\dfrac{(-1)^n c_n}{3^{n-1}}$, 求数列 $\{d_n\}$ 的前 n 项和 S_n.

知乎@余命数

解 (1) 题目给出变换后的数列 $u_n=c_n/3^n$ 的前 n 项和, 要求反求原数列 c_n 的通项.

令新数列 $\{u_n\}$ 的通项为 $u_n=\dfrac{c_n}{3^n}$. 那么, 题目给出的条件即为数列 $\{u_n\}$ 的前 n 项和 $T_n=\sum_{k=1}^{n}u_k=2n^2+n$.

知乎@余命数

可利用基本关系式 $u_n=T_n-T_{n-1}$, 当 $n\geq 2$ 时来求解 u_n.

$$T_{n-1}=2(n-1)^2+(n-1)=2(n^2-2n+1)+n-1=2n^2-3n+1$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

对于 $n \geq 2$:

$$\begin{aligned} u_n &= T_n - T_{n-1} \\ &= (2n^2 + n) - (2n^2 - 3n + 1) \\ &= 4n - 1 \end{aligned}$$

知乎@余命数

接下来,检验此公式在 $n = 1$ 时是否成立. 当 $n = 1$ 时,$u_1 = T_1 = 2(1)^2 + 1 = 3$. 将 $n = 1$ 代入已得的公式, $u_1 = 4(1) - 1 = 3$. 公式对 $n = 1$ 也成立. 因此, 数列 $\{u_n\}$ 的通项公式为 $u_n = 4n - 1$.

最后, 从 $u_n = \dfrac{c_n}{3^n}$ 反解出 c_n. 于是

$$c_n = u_n \cdot 3^n = (4n - 1)3^n$$

(2) 需求和的数列是 $\{d_n\}$, 其通项 $d_k = \dfrac{(-1)^k c_k}{3^{k-1}}$. 首先, 将第一问求得的 $c_k = (4k - 1)3^k$ 代入, 得

$$d_k = \frac{(-1)^k (4k - 1)3^k}{3^{k-1}} = (-1)^k (4k - 1) \cdot 3$$

这个通项实质上是差比数列, 但形式需变形后才与标准公式 $(ak + b)q^{k-1}$ 匹配. 关键是处理交错项 $(-1)^k$.

$$d_k = 3 \cdot (-1) \cdot (-1)^{k-1} \cdot (4k - 1) = (-3)(4k - 1) \cdot (-1)^{k-1}$$

$$d_k = (-12k + 3) \cdot (-1)^{k-1}$$

现在通项完全符合标准形式. 进行参数比对.

- 线性部分 $-12k + 3$ 对应 $ak + b$, 可得 $a = -12, b = 3$.
- 指数部分 $(-1)^{k-1}$ 对应 q^{k-1}, 可得 $q = -1$.

接着, 计算系数 A, B, 得

$$A = \frac{a}{q - 1} = \frac{-12}{-1 - 1} = \frac{-12}{-2} = 6$$

$$B = \frac{bq - b - a}{(q - 1)^2} = \frac{(3)(-1) - 3 - (-12)}{(-1 - 1)^2} = \frac{-3 - 3 + 12}{(-2)^2} = \frac{6}{4} = \frac{3}{2}$$

将 A, B, q 代入求和公式 $S_n = (An + B)q^n - B$, 得

$$\begin{aligned} S_n &= \left(6n + \frac{3}{2}\right)(-1)^n - \frac{3}{2} \\ &= \frac{12n + 3}{2}(-1)^n - \frac{3}{2} \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= \frac{(12n+3)(-1)^n - 3}{2}$$

注 2.5 (启示)
本题综合了数列的多个知识点.

1. 第一问引入辅助数列 $\{u_n\}$, 将复杂关系式**化归**为由和求通项的问题.
2. 第二问不能直接把 d_k 的原始形式代入公式, 而要先做代数恒等变形, 严格化为 $(ak+b)q^{k-1}$ 的标准结构. 处理交错项 $(-1)^k = -(-1)^{k-1}$, 是识别公比 $q=-1$ 的关键.
3. 推导完成后, 一旦问题化为标准形式, 代入公式即可得到结果, 比错位相减法更直接.

知乎@余命数

2.6.3 裂项相消法

错位相减法通过构造把差比数列求和化归为等比数列求和. 另一种思路是把通项 c_n 直接拆成 $F(n)-F(n-1)$, 像处理 $\frac{1}{n(n+1)}$ 那样裂项相消.

这需要从观察规律转向主动构造. 此法计算上可能更繁琐, 但它说明求和与差分互为逆运算, 也是差分方程的基础[4].

[4] 后面讨论概率相关题目时, 还会回到这里.

2.6.3.1 待定系数法

我们寻找一个函数 $F(n)$, 使得 $c_n = F(n) - F(n-1)$. 差比数列的通项 $c_n = (a_1+(n-1)d)b_1q^{n-1}$ 是 n 的一次多项式与指数函数的乘积. 由此猜测, 函数 $F(n)$ 也应为多项式乘以指数函数的形式.

设差比数列通项为 $c_n = (\alpha n+\beta)q^n$, 当 $q \neq 0$ 时, 常见的差比数列通项都可以化为此形式. 设待定函数 $F(n)$ 具有 $F(n) = (Pn+Q)q^n$ 的形式, 其中 P 和 Q 是待定常数.

接着, 计算 $F(n)-F(n-1)$, 得

$$\begin{aligned} F(n)-F(n-1) &= (Pn+Q)q^n - [P(n-1)+Q]q^{n-1} \\ &= q(Pn+Q)q^{n-1} - [Pn-P+Q]q^{n-1} \\ &= [q(Pn+Q) - (Pn-P+Q)]q^{n-1} \\ &= [Pqn+Qq-Pn+P-Q]q^{n-1} \\ &= [n(Pq-P) + (P+Qq-Q)]q^{n-1} \end{aligned}$$

要使 $c_n = F(n)-F(n-1)$, 只需将上式与 $c_n = (\alpha n+\beta)q\cdot q^{n-1} = [(\alpha q)n+\beta q]q^{n-1}$ 作比较, 令对应项系数相等, 即可解出 P 和 Q.

例题 2.28 再次求解数列 $\{(2n+1)\cdot 3^n\}$ 的前 n 项和 S_n.

解

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

设 $F(n) = (Pn + Q)3^n$.

希望 $c_n = F(n) - F(n-1)$.

知乎@余命数

$$\begin{aligned} F(n) - F(n-1) &= (Pn + Q)3^n - (P(n-1) + Q)3^{n-1} \\ &= 3(Pn + Q)3^{n-1} - (Pn - P + Q)3^{n-1} \\ &= [3Pn + 3Q - Pn + P - Q]3^{n-1} \\ &= [2Pn + (P + 2Q)]3^{n-1} \end{aligned}$$

为了与 $c_n = (2n + 1)3^n$ 比较, 将 c_n 也变形.

$$c_n = (2n + 1)3^n = 3(2n + 1)3^{n-1} = (6n + 3)3^{n-1}$$

知乎@余命数

比较 $[2Pn + (P + 2Q)]3^{n-1}$ 与 $(6n + 3)3^{n-1}$ 的系数, 得

- n 的系数为 $2P = 6 \implies P = 3$
- 常数项系数为 $P + 2Q = 3 \implies 3 + 2Q = 3 \implies Q = 0$

解得 $Q = 0$, 所以待定函数是 $F(n) = 3n \cdot 3^n = n \cdot 3^{n+1}$.

$$S_n = \sum_{k=1}^{n} c_k = F(n) - F(0)$$

计算 $F(n)$ 和 $F(0)$, 得

知乎@余命数

知乎@余命数

$$F(n) = n \cdot 3^{n+1}$$

$$F(0) = 0 \cdot 3^{0+1} = 0$$

因此,

$$S_n = n \cdot 3^{n+1} - 0 = n \cdot 3^{n+1}$$

我们得到与错位相减法相同的结果, 但路径不同. 这一方法把依赖技巧的问题变成按固定步骤做代数运算的问题.

2.7 斐波那契数列

在 1202 年, 意大利数学家斐波那契在他的著作《算盘书》中提出了一个看似简单的问题.

假设一对兔子每个月可以生一对新的兔子, 而新生的兔子又需要一个月的时间才能成熟并开始生育. 如果从一对新生的兔子开始, 那么一年后会有多少对兔子?

这个问题的答案, $1, 1, 2, 3, 5, 8, 13, \ldots$, 催生了数学中最著名、最迷人的数列.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定义 2.9 (斐波那契数列) 斐波那契数列 $\{F_n\}$ 由以下线性递推关系定义

$$F_n = F_{n-1} + F_{n-2} \quad (n \geq 2)$$

其初始项通常定义为 $F_0 = 0, F_1 = 1$. 数列的前几项为

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

知乎@余命数

2.7.1 比内公式

知乎@余命数

递推公式计算效率低. 要从 F_0, F_1 一路递推到 F_{100} 很繁琐. 与等差、等比数列一样, 目标是建立直接从 n 计算 F_n 的**通项公式**.

这是一种二阶线性递推关系, 可以使用**特征方程法**求解. 该方法先猜测解的形式. 我们大胆假设数列的通项具有等比数列的形式, 即 $F_n = x^n$, 其中 $x \neq 0$. 将此形式代入递推关系 $F_n - F_{n-1} - F_{n-2} = 0$ 中, 得

知乎@余命数

$$x^n - x^{n-1} - x^{n-2} = 0$$

两边同除以 x^{n-2}, 得到这个数列的**特征方程**

$$x^2 - x - 1 = 0$$

知乎@余命数 知乎@余命数 知乎@余命数

利用二次方程求根公式, 解得两个根

知乎@余命数 知乎@余命数

$$\varphi = \frac{1+\sqrt{5}}{2}, \quad \psi = \frac{1-\sqrt{5}}{2}$$

φ 是著名的**黄金分割比**. 由于 $F_n = \varphi^n$ 和 $F_n = \psi^n$ 都是满足递推关系的解, 那么它们的任意线性组合 $F_n = A\varphi^n + B\psi^n$ 也同样满足. 接着, 利用初始条件 $F_0 = 0, F_1 = 1$ 来确定常数 A 和 B.

$$n = 0: \quad F_0 = A\varphi^0 + B\psi^0 = A + B = 0 \implies B = -A$$

$$n = 1: \quad F_1 = A\varphi^1 + B\psi^1 = A\varphi - A\psi = A(\varphi - \psi) = 1$$

我们计算 $\varphi - \psi = \frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2} = \sqrt{5}$. 于是 $A\sqrt{5} = 1$, 解得 $A = \frac{1}{\sqrt{5}}$, 从而 $B = -\frac{1}{\sqrt{5}}$. 将 A, B 代回, 得斐波那契数列的通项公式, 即**比内公式**.

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right] \tag{2-5}$$

知乎@余命数

这个充满了无理数 $\sqrt{5}$ 的公式, 对于任意非负整数 n, 其计算结果永远是一个整数. 在代数运算中, 所有的 $\sqrt{5}$ 项都相互抵消了.

2.7.2 前 n 项和

斐波那契数列的求和公式也很简洁. 现推导 $S_n = \sum_{k=1}^{n} F_k$ 的封闭表达式.

由定义式 $F_{k+2} = F_{k+1} + F_k$ 移项, 得到裂项所需的形式

$$F_k = F_{k+2} - F_{k+1}$$

它表明, 每一个斐波那契数都可以表示为后两个斐波那契数之差. 接着将和式 S_n 中的每一项都用这种形式替换:

$$\begin{aligned} S_n &= F_1 + F_2 + F_3 + \cdots + F_n \\ &= (F_3 - F_2) + (F_4 - F_3) + (F_5 - F_4) + \cdots + (F_{n+2} - F_{n+1}) \end{aligned}$$

这是**裂项相消**求和. 中间项 $-F_3$ 与 $+F_3$, $-F_4$ 与 $+F_4$ 等依次抵消, 最后只剩下 $-F_2$ 和 $+F_{n+2}$.

$$S_n = F_{n+2} - F_2$$

根据我们的定义 $F_2 = 1$, 得最终的求和公式

$$\sum_{k=1}^{n} F_k = F_{n+2} - 1 \tag{2-6}$$

这个结果就是斐波那契恒等式. 前 n 个斐波那契数的和等于 $F_{n+2} - 1$.

例题 2.29　我们验证 $n = 5$.

$$S_5 = F_1 + F_2 + F_3 + F_4 + F_5 = 1 + 1 + 2 + 3 + 5 = 12$$

根据公式, 结果应为

$$F_{5+2} - 1 = F_7 - 1 = 13 - 1 = 12$$

2.7.3 平方和

求出前 n 项和后, 自然可以研究**前 n 项的平方和** $\sum_{k=1}^{n} F_k^2$. 这个和也有简洁的封闭形式.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定理 2.2 对于斐波那契数列 $\{F_n\}$, 其前 n 项平方和为

$$\sum_{k=1}^{n} F_k^2 = F_n F_{n+1}$$

证明 需找一个关于 F_k^2 的恒等式来实现裂项. 从恒等式出发, 对任意 $k \geq 1$:

$$F_{k+1} - F_{k-1} = (F_k + F_{k-1}) - F_{k-1} = F_k.$$

据此把 F_k^2 中的一个 F_k 替换掉.

$$F_k^2 = F_k \cdot (F_{k+1} - F_{k-1}) = F_k F_{k+1} - F_k F_{k-1}.$$

令 $A_k = F_k F_{k+1}$, 则 $F_k F_{k-1}$ 即 A_{k-1}, 上式写成裂项形式

$$F_k^2 = A_k - A_{k-1}.$$

对整个数列求和.

$$\begin{aligned}\sum_{k=1}^{n} F_k^2 &= \sum_{k=1}^{n} (A_k - A_{k-1}) \\ &= (A_1 - A_0) + (A_2 - A_1) + (A_3 - A_2) + \cdots + (A_n - A_{n-1}).\end{aligned}$$

裂项求和时, 中间项 $-A_1$ 与 $+A_1$, $-A_2$ 与 $+A_2$ 等依次抵消, 最后只剩下 $-A_0$ 和 $+A_n$.

$$\sum_{k=1}^{n} F_k^2 = A_n - A_0.$$

计算 $A_n = F_n F_{n+1}$, $A_0 = F_0 F_{0+1} = F_0 F_1 = 0 \cdot 1 = 0$, 故

$$\sum_{k=1}^{n} F_k^2 = F_n F_{n+1} - 0 = F_n F_{n+1}.$$

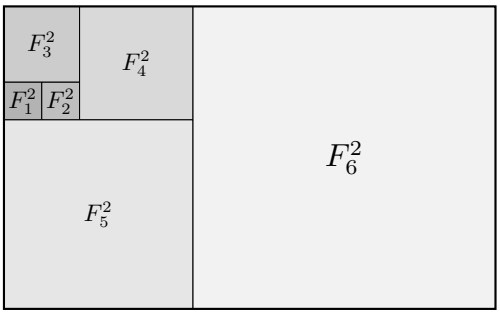
证毕. □

这个恒等式还有一个几何证明. 用边长为 F_k 的正方形面积表示 F_k^2, 再把它们拼成一个矩形.

观察这个拼接过程.

- 从 $F_1^2 (1 \times 1)$ 开始, 这是一个 $F_1 \times F_2$ 的矩形.
- 加上 $F_2^2 (1 \times 1)$, 得一个 2×1 的矩形, 其面积为 $F_1^2 + F_2^2 = F_2 F_3$.
- 加上 $F_3^2 (2 \times 2)$, 得一个 2×3 的矩形, 其面积为 $\sum_{k=1}^{3} F_k^2 = F_3 F_4$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵



- 加上 $F_4^2(3\times 3)$, 得一个 5×3 的矩形, 其面积为 $\sum_{k=1}^{4} F_k^2 = F_4F_5$.

每一步, 把边长为 F_n 的正方形拼到长宽为 F_{n-1} 和 F_n 的矩形上, 得到长为 F_n、宽为 $F_{n-1}+F_n = F_{n+1}$ 的新矩形. 这个过程不断进行, 说明所有正方形面积之和等于最后两个斐波那契数所构成矩形的面积.

知乎@余命数

2.7.4 相邻项乘积之和

在分析了平方和之后, 自然要问相邻两项乘积之和是否也有简洁闭式.

$$\sum_{k=1}^{n} F_kF_{k+1}.$$

答案是肯定的, 现给出按奇偶情形区分的精确公式.

定理 2.3 设 $S_n = \sum_{k=1}^{n} F_kF_{k+1}$. 则

$$S_n = \begin{cases} F_{n+1}^2 & \text{如果}n\text{是奇数} \\ F_nF_{n+2} & \text{如果}n\text{是偶数} \end{cases}$$

证明 这个结论按奇偶区分, 可用交错数学归纳法证明. 我们同时证明对任意正整数 m, 命题在 $n=2m-1$ 和 $n=2m$ 时都成立.

设 $S_n = \sum_{k=1}^{n} F_kF_{k+1}$. 于是有两个递推关系.

$$S_{2m} = S_{2m-1} + F_{2m}F_{2m+1} \tag{1}$$

$$S_{2m+1} = S_{2m} + F_{2m+1}F_{2m+2} \tag{2}$$

对奇数 $n=1, S_1 = F_1F_2 = 1\cdot 1 = 1$, 公式结果为 $F_{1+1}^2 = F_2^2 = 1^2 = 1$, 命题成立. 对偶数 $n=2, S_2 = S_1 + F_2F_3 = 1+1\cdot 2 = 3$, 公式结果为 $F_2F_{2+2} = F_2F_4 = 1\cdot 3 = 3$, 命题成立.

假设当 $m=k$ 时, 即对于 $n=2k-1$ 和 $n=2k$, 命题均成立.

$$S_{2k-1} = F_{2k}^2 \qquad \text{(归纳假设 1)}$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

$$S_{2k} = F_{2k}F_{2k+2} \qquad \text{(归纳假设 2)}$$

需证当 $m = k+1$ 时, 即对于 $n = 2k+1$ 和 $n = 2k+2$, 命题也成立.

先证 $S_{2k+1} = F_{(2k+1)+1}^2 = F_{2k+2}^2$. 由关系式 (2) 与归纳假设 2 出发, 得

$$\begin{aligned} S_{2k+1} &= S_{2k} + F_{2k+1}F_{2k+2} \\ &= (F_{2k}F_{2k+2}) + F_{2k+1}F_{2k+2} \\ &= (F_{2k} + F_{2k+1})F_{2k+2} \\ &= F_{2k+2} \cdot F_{2k+2} \quad (\text{斐波那契定义 } F_{2k} + F_{2k+1} = F_{2k+2}) \\ &= F_{2k+2}^2, \end{aligned}$$

即奇数 $n = 2k+1$ 的目标.

再证 $S_{2k+2} = F_{2k+2}F_{(2k+2)+2} = F_{2k+2}F_{2k+4}$. 由关系式 (1)(将 m 换为 $k+1$) 与刚证得的 $S_{2k+1} = F_{2k+2}^2$ 出发, 得

$$\begin{aligned} S_{2k+2} &= S_{2k+1} + F_{2k+2}F_{2k+3} \\ &= F_{2k+2}^2 + F_{2k+2}F_{2k+3} \\ &= F_{2k+2}(F_{2k+2} + F_{2k+3}) \\ &= F_{2k+2}F_{2k+4} \quad (\text{斐波那契定义 } F_{2k+2} + F_{2k+3} = F_{2k+4}). \end{aligned}$$

即偶数 $n = 2k+2$ 的目标.

奠基步骤成立, 归纳步骤证明了从 k 到 $k+1$ 的传递性, 故该定理对所有正整数 n 均成立. □

2.7.5 对称项之和

斐波那契数列中还有许多对称恒等式. 下列关系式说明任意一项与它前后等距两项之间的联系.

定理 2.4 对于斐波那契数列 $\{F_n\}$, 任意一项的三倍, 等于它前面第二项与后面第二项之和.

$$3F_n = F_{n+2} + F_{n-2} \quad (n \geq 2)$$

证明 从右侧入手, 反复用斐波那契数列的定义化简 $F_{n+2} + F_{n-2}$. 将 F_{n+2} 向下展开, 用 F_n 及其相邻项表示.

$$F_{n+2} = F_{n+1} + F_n$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= (F_n + F_{n-1}) + F_n$$
$$= 2F_n + F_{n-1}.$$

对 F_{n-2}, 由定义式 $F_n = F_{n-1} + F_{n-2}$ 移项得

$$F_{n-2} = F_n - F_{n-1},$$

故 F_{n-2} 也以 F_n 和 F_{n-1} 表示. 将两式代回原表达式, 得

$$\begin{aligned} F_{n+2} + F_{n-2} &= (2F_n + F_{n-1}) + (F_n - F_{n-1}) \\ &= 2F_n + F_n + F_{n-1} - F_{n-1} \\ &= 3F_n. \end{aligned}$$

知乎@余命数

等式右侧化简为左侧, 证毕. □

黄金分割极限

注 2.6
数值上 F_{n+1}/F_n 很快靠近 $\varphi \approx 1.618$; 现用比内公式作严格极限论证, 依赖 $|q| < 1$ 时 $q^n \to 0$ 等结论, 见 第四卷第 16 章.

例题 2.30 (黄金分割极限) 证明斐波那契数列相邻两项的比值, 当 n 趋于无穷时, 收敛于黄金分割比 φ.

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \frac{1+\sqrt{5}}{2}$$

解 这是数列极限问题, 描述斐波那契数列的渐近行为. 递推公式无法直接处理, 需用通项公式, 即比内公式. 记 $\varphi = \frac{1+\sqrt{5}}{2}$ 和 $\psi = \frac{1-\sqrt{5}}{2}$, 比内公式为 $F_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}$. 构造比值

$$\frac{F_{n+1}}{F_n} = \frac{\frac{\varphi^{n+1} - \psi^{n+1}}{\sqrt{5}}}{\frac{\varphi^n - \psi^n}{\sqrt{5}}} = \frac{\varphi^{n+1} - \psi^{n+1}}{\varphi^n - \psi^n}.$$

求极限需识别主导项. 注意到 $|\varphi| = \frac{1+\sqrt{5}}{2} \approx 1.618 > 1$, $|\psi| = \left|\frac{1-\sqrt{5}}{2}\right| \approx 0.618 < 1$, 当 $n \to \infty$ 时 $\psi^n \to 0$ 而 $\varphi^n \to \infty$, 故 φ^n 为主导项. 将分子分母同除以 φ^n:

$$\frac{F_{n+1}}{F_n} = \frac{\varphi - \psi \cdot (\frac{\psi}{\varphi})^n}{1 - (\frac{\psi}{\varphi})^n}.$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

因 $|\frac{\psi}{\varphi}| = |\frac{1-\sqrt{5}}{1+\sqrt{5}}| < 1$, 故 $\lim_{n\to\infty}\left(\frac{\psi}{\varphi}\right)^n = 0$, 从而

$$\lim_{n\to\infty}\frac{F_{n+1}}{F_n} = \frac{\varphi - \psi\cdot 0}{1-0} = \varphi.$$

斐波那契数列的相邻项之比确收敛于黄金分割比 $\varphi = \frac{1+\sqrt{5}}{2}$.

例题 2.31 计算前 10 项斐波那契数的平方和 $\sum_{k=1}^{10} F_k^2$.

解 逐项计算平方再相加很繁琐. 利用已证的平方和公式, 问题可化简.

已证明的平方和公式为

$$\sum_{k=1}^{n} F_k^2 = F_n F_{n+1}$$

在本题中, $n = 10$. 因此, 只需要计算出 F_{10} 和 F_{11} 即可. 回忆斐波那契数列.

$$1, 1, 2, 3, 5, 8, 13, 21, 34, \underbrace{55}_{F_{10}}, \underbrace{89}_{F_{11}}, \ldots$$

所以 $F_{10} = 55$, $F_{11} = 89$.

代入公式, 得

$$\sum_{k=1}^{10} F_k^2 = F_{10}F_{11} = 55 \times 89$$

计算可得

$$55 \times 89 = 55 \times (90 - 1) = 4950 - 55 = 4895$$

因此, 前 10 项斐波那契数的平方和为 4895.

例题 2.32 求 $\sum_{k=1}^{8}(F_k + F_{k+1})^2$ 的值.

解 通项是完全平方式, 直接计算复杂, 先展开再拆成已知形式, 得

$$(F_k + F_{k+1})^2 = F_k^2 + 2F_kF_{k+1} + F_{k+1}^2,$$

故原求和式拆分为三部分.

$$\sum_{k=1}^{8}(F_k + F_{k+1})^2 = \sum_{k=1}^{8} F_k^2 + 2\sum_{k=1}^{8} F_kF_{k+1} + \sum_{k=1}^{8} F_{k+1}^2.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

逐个处理. 部分一为平方和, 由公式 $\sum_{k=1}^{n} F_k^2 = F_n F_{n+1}$ 得

$$\sum_{k=1}^{8} F_k^2 = F_8 F_9 = 21 \times 34 = 714.$$

部分二为相邻项乘积之和. 因 $n = 8$ 是偶数, 用公式 $\sum_{k=1}^{n} F_k F_{k+1} = F_n F_{n+2}$, 得

$$2\sum_{k=1}^{8} F_k F_{k+1} = 2(F_8 F_{10}) = 2 \times (21 \times 55) = 2310.$$

知乎@余命数

部分三 $\sum_{k=1}^{8} F_{k+1}^2$ 是变形的平方和. 换元 $j = k + 1$ 得

$$\sum_{k=1}^{8} F_{k+1}^2 = F_2^2 + F_3^2 + \cdots + F_9^2 = \left(\sum_{k=1}^{9} F_k^2\right) - F_1^2,$$

由平方和公式

$$\left(\sum_{k=1}^{9} F_k^2\right) - F_1^2 = (F_9 F_{10}) - 1^2 = (34 \times 55) - 1 = 1869.$$

三部分相加得

$$714 + 2310 + 1869 = 4893.$$

最终结果为 4893.

例题 2.33　证明 $\sum_{k=1}^{n} F_{2k-1} = F_{2n}$.

证明　这个和式只对奇数项求和, 没有现成公式可用, 但通项形式适合裂项, 需把 F_{2k-1} 写成两项之差. 由斐波那契数列的定义, 偶数项与其前两项的关系为

$$F_{2k} = F_{2k-1} + F_{2k-2},$$

移项得

$$F_{2k-1} = F_{2k} - F_{2k-2},$$

此式把一个奇数项表示成两个相邻偶数项之差. 对整个和式裂项.

知乎@余命数

$$\begin{aligned}\sum_{k=1}^{n} F_{2k-1} &= \sum_{k=1}^{n} (F_{2k} - F_{2k-2}) \\ &= (F_2 - F_0) + (F_4 - F_2) + (F_6 - F_4) + \cdots + (F_{2n} - F_{2n-2}).\end{aligned}$$

裂项求和时,中间项 $-F_2$ 与 $+F_2$,$-F_4$ 与 $+F_4$ 等全部抵消,最终只剩第一项的 $-F_0$ 和最后一项的 $+F_{2n}$.

$$\sum_{k=1}^{n} F_{2k-1} = F_{2n} - F_0.$$

知乎@余命数

由定义 $F_0 = 0$,得

$$\sum_{k=1}^{n} F_{2k-1} = F_{2n}.$$

知乎@余命数 知乎@余命数

证毕. □

知乎@余命数

例题 2.34 (23-24 河南漯河期末) 分析并判断以下关于斐波那契数列 $\{F_n\}$ (定义为 $F_0 = 0, F_1 = 1, \ldots$)的四个命题是否正确,并给出严格的证明或反例.
命题 A 数列 $\{F_n\}$ 是一个递增数列.
命题 B 89 是数列 $\{F_n\}$ 中的一项.
命题 C 对所有 $n \geq 1$,恒等式 $\sum_{k=1}^{n} F_k^2 = F_n F_{n+1}$ 成立.(原题是要求从前 10 项注意到,稍微有些不严谨.)
命题 D 比值 $\dfrac{F_n}{F_{n+1}}$ 收敛,且其极限为 $\dfrac{\sqrt{5}-1}{2}$.

解 逐一判断四个命题的真伪.

命题 A 错误. 严格递增数列要求对任意 $n \geq 0$ 都有 $F_{n+1} > F_n$. 但在本数列中,有 $F_1 = 1$ 且 $F_2 = 1$,不满足 $F_2 > F_1$. 因此,斐波那契数列是一个**非减数列**,但不是严格递增数列.

知乎@余命数 知乎@余命数

命题 B 正确. 通过逐项计算,得数列的前几项,得

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ \underline{89},\ \ldots$$

可见 89 是数列的第 11 项 ($F_{11} = 89$).

命题 C 正确. 这正是在本节平方和部分,通过代数裂项和几何拼接两种方法证明过的恒等式. 它对所有正整数 n 都成立.

命题 D 正确. 在前面的例题中,已证明斐波那契数列相邻两项的比值收敛于黄金分割比 φ.

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \varphi = \frac{1+\sqrt{5}}{2}$$

知乎@余命数

本题所求的极限是其倒数. 在已接受 $\lim F_{n+1}/F_n = \varphi$ 的前提下,由极限运算法则 (第四卷第 16 章),得

知乎@余命数

$$\lim_{n\to\infty} \frac{F_n}{F_{n+1}} = \frac{1}{\lim_{n\to\infty} \frac{F_{n+1}}{F_n}} = \frac{1}{\varphi}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

对 $\frac{1}{\varphi}$ 作分子有理化.

$$\frac{1}{\varphi}=\frac{2}{1+\sqrt{5}}=\frac{2(\sqrt{5}-1)}{(\sqrt{5}+1)(\sqrt{5}-1)}=\frac{2(\sqrt{5}-1)}{5-1}=\frac{2(\sqrt{5}-1)}{4}=\frac{\sqrt{5}-1}{2}$$

因此, 该命题正确. 这个值有时被称为黄金分割比的共轭, 记作 $\frac{1}{\varphi}$ 或 Φ.

命题 A 错误, 命题 B、C、D 均正确.

例题 2.35 (2024 年新高考一卷第 8 题) 已知定义在 $\mathbb{R}$ 上的函数 $f(x)$ 满足 $f(x)>f(x-1)+f(x-2)$, 且当 $x<3$ 时, $f(x)=x$. 判断下列结论中哪一个必定正确?

(A) $f(10)>100$

(B) $f(20)>1000$

(C) $f(10)<1000$

(D) $f(20)<10000$

解 问题要求理解不等式 $f(x)>f(x-1)+f(x-2)$ 的含义. 该函数的增长速度**超过**斐波那契数列, 因此可用一个标准斐波那契数列作为 $f(x)$ 的**下界**, 再据此估值.

不等式只给出下限, $f(x)$ 的值可以比下限大很多, 例如 $f(3)$ 可以是 3.1, 也可以是 10^9. 因此, 任何给出 $f(x)$ **上界**的结论, 如 (C) 和 (D), 都无法从已知条件保证, 不可能是必定正确的. 只需关心与下界相关的 (A) 和 (B).

构造一个基准数列 $\{g_n\}$, 它满足斐波那契的**等式**关系.

$$g_n=g_{n-1}+g_{n-2}\quad(n\geq 3).$$

用 $f(x)$ 的已知值设定初始项. 对于整数 n, 得

$$f(1)=1\quad(\text{因为}1<3)$$

$$f(2)=2\quad(\text{因为}2<3),$$

故设 $g_1=1, g_2=2$. 计算 $\{g_n\}$ 的前几项, 得

$$g_1=1,\quad g_2=2,\quad g_3=g_2+g_1=3,$$

$$g_4=g_3+g_2=5,\quad g_5=g_4+g_3=8.$$

数列 $\{g_n\}$ 正是斐波那契数列 $\{F_n\}$ 的平移. 于是 $g_n=F_{n+1}$.

用数学归纳法证明 $f(n)>g_n=F_{n+1}$ 对所有整数 $n\geq 3$ 成立.

当 $n=3$ 时, 由题设 $f(3)>f(2)+f(1)=2+1=3$, 而 $g_3=3$, 故 $f(3)>g_3$ 成立, 奠基步骤完成.

知乎@余命数

假设当 $3 \le k < n$ 时不等式 $f(k) > g_k$ 成立, 需证 $f(n) > g_n$. 由题设

$$f(n) > f(n-1) + f(n-2).$$

根据归纳假设, $f(n-1) > g_{n-1}$ 且 $f(n-2) > g_{n-2}$(只要 $n-2 \ge 3$). 当 $n-2 = 2$ 即 $n = 4$ 时, $f(2) = g_2$, 需单独考虑.

$$\text{若} n = 4, \quad f(4) > f(3) + f(2) > g_3 + g_2 = g_4;$$

$$\text{若} n \ge 5, \quad f(n) > f(n-1) + f(n-2) > g_{n-1} + g_{n-2} = g_n.$$

综合 $n = 4$ 与 $n \ge 5$ 的讨论, 对所有整数 $n \ge 3$ 都有 $f(n) > g_n = F_{n+1}$.

用此下界检验选项 (A) 和 (B).

选项 (A) $f(10) > 100$. 由结论 $f(10) > g_{10} = F_{11}$. 计算斐波那契数列 $F_1 = 1, \dots, F_8 = 21, F_9 = 34, F_{10} = 55, F_{11} = 89$, 只能确定 $f(10) > 89$, 并不能保证 $f(10)$ 一定大于 100. 可构造满足条件但 $f(10) < 100$ 的函数. 令 $x < 3$ 时 $f(x) = x$, $x \ge 3$ 时递推定义 $f(x) = f(x-1) + f(x-2) + 0.1$. 它满足题设不等式, 而整数点上依次有 $f(3) = 3.1, f(4) = 5.2, \dots, f(10) = 94.4 < 100$. 故 (A) 不一定正确.

选项 (B) $f(20) > 1000$. 由结论 $f(20) > g_{20} = F_{21}$. 继续计算斐波那契数列, 得

$$F_{11} = 89, \quad F_{12} = 144, \quad F_{13} = 233, \quad F_{14} = 377,$$

$$F_{15} = 610, \quad F_{16} = 987, \quad F_{17} = 1597, \quad F_{18} = 2584,$$

$$F_{19} = 4181, \quad F_{20} = 6765, \quad F_{21} = 10946.$$

得 $f(20) > F_{21} = 10946$. 因 $10946 > 1000$, 故 $f(20) > 1000$ 一定成立, (B) 正确.

本题把函数与数列结合在一起. 条件 $f(x) > f(x-1) + f(x-2)$ 与斐波那契递推 $F_n = F_{n-1} + F_{n-2}$ 只差一点, 即不等号. 这说明 f 的增长速度不低于斐波那契数列.

处理这类不等式时, 可先构造一个满足等式的标准斐波那契数列 g_n 作为下界. 令 $g_1 = f(1) = 1$, $g_2 = f(2) = 2$, 则 $g_n = g_{n-1} + g_{n-2}$, 也就是 $g_n = F_{n+1}$. 再用数学归纳法证明 $f(n) > g_n$ 对所有 $n \ge 3$ 成立, 下界就落到实处.

2.8 数列求和的通用策略

我们已经讨论了等差数列与等比数列的求和公式. 实际遇到的数列常常比这两种模型更复杂, 既非等差, 也非等比. 我们要计算这些数列的和, 就需要一套通用的代数技巧.

这些技巧的目标一致. 我们总是把陌生、复杂的求和问题**化归**为熟悉或简单的形式.

知乎@余命数

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.8.1 基本公式法

当数列的通项 a_n 是关于 n 的多项式时，我们利用求和符号 $\sum$ 的线性性质，把原求和式拆成若干基本幂次和，再用已知公式计算.

我们先写下几个基本的自然数幂次和公式.

- 常数的和是 $\sum_{k=1}^{n} c = cn$.
- 一次方的和是 $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.
- 二次方的和是 $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$.

下面补充平方和公式的推导.

知乎@余命数

2.8.1.1 扰动法

我们从寻找与目标和式 $\sum k^2$ 相关的裂项求和开始.$(k+1)^3 - k^3$ 是关于 k 的二次多项式，正好把 k^2 与裂项联系起来.

考虑恒等式

$$(k+1)^3 - k^3 = k^3 + 3k^2 + 3k + 1 - k^3 = 3k^2 + 3k + 1$$

这个恒等式对所有 k 均成立. 现在，对这个等式的两边同时从 $k=1$ 到 n 求和，得

$$\sum_{k=1}^{n} \left[(k+1)^3 - k^3\right] = \sum_{k=1}^{n} (3k^2 + 3k + 1)$$

分别计算等式左右两边.

左侧 (LHS) 是一个典型的裂项求和，计算如下.

$$\begin{aligned}\sum_{k=1}^{n} \left[(k+1)^3 - k^3\right] &= (2^3 - 1^3) + (3^3 - 2^3) + (4^3 - 3^3) \\ &\quad + \cdots + ((n+1)^3 - n^3) \\ &= -1^3 + (2^3 - 2^3) + (3^3 - 3^3) + \ldots \\ &\quad + (n^3 - n^3) + (n+1)^3 \\ &= (n+1)^3 - 1\end{aligned}$$

右侧 (RHS) 利用求和的线性性质拆分，得

$$\sum_{k=1}^{n} (3k^2 + 3k + 1) = 3\sum_{k=1}^{n} k^2 + 3\sum_{k=1}^{n} k + \sum_{k=1}^{n} 1$$

令 $S_2 = \sum_{k=1}^{n} k^2$，并代入已知的一次和与常数和公式，得

$$\text{RHS} = 3S_2 + 3 \cdot \frac{n(n+1)}{2} + n$$

联立左右两侧的结果, 有

$$(n+1)^3-1=3S_2+\frac{3n(n+1)}{2}+n$$

任务现转化为解出 S_2.

$$\begin{aligned}3S_2&=(n+1)^3-1-\frac{3n(n+1)}{2}-n\\&=(n+1)^3-(n+1)-\frac{3n(n+1)}{2}\\&=(n+1)\left[(n+1)^2-1-\frac{3n}{2}\right]\\&=(n+1)\left[n^2+2n+1-1-\frac{3n}{2}\right]\\&=(n+1)\left[n^2+\frac{n}{2}\right]\\&=(n+1)\cdot n\left(n+\frac{1}{2}\right)\\&=\frac{n(n+1)(2n+1)}{2}\end{aligned}$$

两边同除以 3, 即得平方和公式

$$S_2=\sum_{k=1}^{n}k^2=\frac{n(n+1)(2n+1)}{6}$$

2.8.1.2 数学归纳法

已通过扰动法得到公式, 现用归纳法严格证明它.

设命题 $P(n)$ 为 $\sum_{k=1}^{n}k^2=\frac{n(n+1)(2n+1)}{6}$.

当 $n=1$ 时, 左侧 $=\sum_{k=1}^{1}k^2=1$, 右侧 $=\frac{1(1+1)(2\cdot1+1)}{6}=\frac{1\cdot2\cdot3}{6}=1$, 两端相等, 故 $P(1)$ 成立, 奠基步骤完成.

假设当 $n=m(m\geq1$ 为任意整数)时命题成立, 即

$$\sum_{k=1}^{m}k^2=\frac{m(m+1)(2m+1)}{6}.$$

由此只需证明当 $n=m+1$ 时

$$\sum_{k=1}^{m+1}k^2=\frac{(m+1)((m+1)+1)(2(m+1)+1)}{6}=\frac{(m+1)(m+2)(2m+3)}{6}$$

也成立. 我们从左侧开始推演.

$$\text{左侧}=\sum_{k=1}^{m+1}k^2=\left(\sum_{k=1}^{m}k^2\right)+(m+1)^2$$

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

$$
\begin{aligned}
&= \frac{m(m+1)(2m+1)}{6} + (m+1)^2 \quad \text{（代入归纳假设）} \\
&= (m+1)\left[\frac{m(2m+1)}{6} + (m+1)\right] \quad \text{（提取公因子 } (m+1)\text{）} \\
&= (m+1)\left[\frac{2m^2+m+6(m+1)}{6}\right] \\
&= \frac{(m+1)(2m^2+7m+6)}{6} \\
&= \frac{(m+1)(m+2)(2m+3)}{6} \quad \text{（因式分解 } 2m^2+7m+6\text{）} \\
&= \text{右侧}.
\end{aligned}
$$

故若命题对 $n=m$ 成立, 则对 $n=m+1$ 亦成立, 归纳步骤完成.

由数学归纳法原理, 命题 $P(n)$ 对所有正整数 n 均成立.

知乎@余命数

2.8.1.3 待定系数法

我们注意到, d 次多项式的求和结果是一个 $d+1$ 次多项式. 由于 k^2 是关于 k 的二次多项式, 其前 n 项和 $S_n=\sum_{k=1}^{n} k^2$ 必然是关于 n 的三次多项式.

设 $S_n = An^3+Bn^2+Cn+D$. 我们的任务是确定系数 A, B, C, D. 我们利用关系 $a_n = S_n - S_{n-1}$, 得到

$$
\begin{aligned}
n^2 &= S_n - S_{n-1} \\
&= (An^3+Bn^2+Cn+D) - [A(n-1)^3+B(n-1)^2+C(n-1)+D] \\
&= A(n^3-(n^3-3n^2+3n-1)) + B(n^2-(n^2-2n+1)) + C(n-(n-1)) \\
&= A(3n^2-3n+1) + B(2n-1) + C \\
&= 3An^2 + (2B-3A)n + (A-B+C)
\end{aligned}
$$

比较等式两边 n 的同次幂系数, 得到

- n^2 的系数: $3A=1 \implies A=\frac{1}{3}$
- n 的系数: $2B-3A=0 \implies 2B=3(\frac{1}{3})=1 \implies B=\frac{1}{2}$
- 常数项: $A-B+C=0 \implies C=B-A=\frac{1}{2}-\frac{1}{3}=\frac{1}{6}$

因此, $S_n=\frac{1}{3}n^3+\frac{1}{2}n^2+\frac{1}{6}n+D$. 为了确定 D, 我们利用初始条件 $S_1=1$.

$$
S_1 = \frac{1}{3}(1)^3+\frac{1}{2}(1)^2+\frac{1}{6}(1)+D = \frac{1}{3}+\frac{1}{2}+\frac{1}{6}+D = 1+D
$$

由于 $S_1=1$, 故 $1+D=1 \implies D=0$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

将系数代回, 得到

$$\begin{aligned} S_n &= \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n = \frac{2n^3+3n^2+n}{6} \\ &= \frac{n(2n^2+3n+1)}{6} = \frac{n(n+1)(2n+1)}{6}. \end{aligned}$$

例题 2.36 求和 $S_n = \sum_{k=1}^{n}(k^2 - 3k + 2)$.

解 我们利用求和符号的线性性质将和式分解, 得

$$S_n = \sum_{k=1}^{n} k^2 - \sum_{k=1}^{n} 3k + \sum_{k=1}^{n} 2$$

再将常数因子提出, 得

$$S_n = \sum_{k=1}^{n} k^2 - 3\sum_{k=1}^{n} k + 2\sum_{k=1}^{n} 1$$

最后代入基本幂次和公式, 得

$$\begin{aligned} S_n &= \frac{n(n+1)(2n+1)}{6} - 3\cdot\frac{n(n+1)}{2} + 2n \\ &= \frac{n}{6}[(n+1)(2n+1) - 9(n+1) + 12] \\ &= \frac{n}{6}[2n^2+3n+1-9n-9+12] \\ &= \frac{n}{6}[2n^2-6n+4] \\ &= \frac{n}{3}(n^2-3n+2) \\ &= \frac{n(n-1)(n-2)}{3} \end{aligned}$$

2.8.2 裂项相消法

若数列的通项 a_k 能分解为两项之差, 即 $a_k = f(k) - f(k+c)$ 或 $a_k = f(k+c) - f(k)$, 我们就可以把求和化简.

把数列的每一项都拆成两部分, 求和时中间的项前后两两抵消, 最终只剩下有限的几项(通常是首尾几项), 从而大幅简化计算.

这一方法的难点在于构造裂项. 以下给出常见的几类代数结构.

- **部分分式型 (I) - 基本型.**

$$\frac{1}{k(k+d)} = \frac{1}{d}\left(\frac{1}{k} - \frac{1}{k+d}\right)$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这是最基本的分式裂项. 其构造的要点是在分子上创造出与分母因子相匹配的差, 即 $(k+d)-k=d$. 通过在原式分子分母同乘以 d, 再进行拆分即可得到.

- **部分分式型 (II) - 奇偶项.**

$$\frac{1}{(2k-1)(2k+1)}=\frac{1}{2}\left(\frac{1}{2k-1}-\frac{1}{2k+1}\right)$$

这是上述公式在 $d=2$ 时的特例, 适用于分母是由公差为 2 的等差数列相邻项构成的情况.

- **根式有理化型.**

$$\frac{1}{\sqrt{k+d}+\sqrt{k}}=\frac{1}{d}(\sqrt{k+d}-\sqrt{k})$$

构造时对分母进行有理化. 分子分母同乘以其共轭表达式 $\sqrt{k+d}-\sqrt{k}$, 利用平方差公式 $(\sqrt{a}+\sqrt{b})(\sqrt{a}-\sqrt{b})=a-b$ 消去分母的根号, 使其变为常数 d.

知乎@余命数

- **平方差裂项型.**

$$\frac{2k+1}{k^2(k+1)^2}=\frac{1}{k^2}-\frac{1}{(k+1)^2}$$

从右侧反向推导更为直观. 通分 $\frac{1}{k^2}-\frac{1}{(k+1)^2}$, 其分子为 $(k+1)^2-k^2=(k^2+2k+1)-k^2=2k+1$, 恰好匹配左侧. 此形式在处理分母为相邻项的完全平方之积时极为有效.

- **指数型.**

$$\frac{2^k}{(2^k-1)(2^{k+1}-1)}=\frac{1}{2^k-1}-\frac{1}{2^{k+1}-1}$$

同样从右侧通分验证. 其分子为 $(2^{k+1}-1)-(2^k-1)=2^{k+1}-2^k=2\cdot 2^k-2^k=2^k$, 与左侧分子完全吻合. 本质是利用了指数函数 $f(k)=2^k-1$ 的差分性质.

- **混合型(指数与多项式).**

$$\frac{2^k(k-1)}{k(k+1)}=\frac{2^{k+1}}{k+1}-\frac{2^k}{k}$$

发现此类混合裂项的通常途径, 是定义函数 $f(k)=\frac{2^k}{k}$ 并考察其差分 $f(k+1)-f(k)$. 计算 $f(k+1)-f(k)=\frac{2^{k+1}}{k+1}-\frac{2^k}{k}$ 并通分, 即可得到左侧的表达式.

- **交错分式型.**

$$\frac{(-1)^k(k+1)}{(2k+1)(2k+3)}=\frac{1}{4}\left(\frac{(-1)^k}{2k+1}-\frac{(-1)^{k+1}}{2k+3}\right)$$

从右侧方括号内通分, 分子为 $(-1)^k(2k+3)-(-1)^{k+1}(2k+1)$. 利用 $(-1)^{k+1}=-(-1)^k$, 分子化为 $(-1)^k[(2k+3)+(2k+1)]=(-1)^k(4k+4)=4(-1)^k(k+1)$, 再与系数 $\frac{1}{4}$ 相乘, 恰好得到左侧.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **交错根式型.**

$$\frac{(-1)^k}{\sqrt{k}+\sqrt{k-1}} = (-1)^k\sqrt{k} + (-1)^{k-1}\sqrt{k-1}$$

- **高阶部分分式型.**

$$\frac{1}{k(k+1)(k+2)} = \frac{1}{2}\left(\frac{1}{k(k+1)} - \frac{1}{(k+1)(k+2)}\right)$$

这是我们在待定系数法一节中推导过的例子, 以 $f(k) = \frac{A}{k(k+1)}$ 为差分原函数进行构造, 是基本部分分式思想的延伸.

- **阶乘添项型 (I).**

$$k \cdot k! = (k+1)! - k!$$

这是阶乘裂项中最经典的形式. 关键技巧在于将因子 k 改写为 $(k+1)-1$, 然后利用乘法分配律展开, 即

$$k \cdot k! = ((k+1)-1)k! = (k+1)k! - k!$$

- **阶乘添项型 (II).**

$$\frac{k}{(k+1)!} = \frac{1}{k!} - \frac{1}{(k+1)!}$$

与上一条技巧完全相同, 我们把分子 k 改写为 $(k+1)-1$.

$$\frac{k}{(k+1)!} = \frac{(k+1)-1}{(k+1)!} = \frac{k+1}{(k+1)!} - \frac{1}{(k+1)!} = \frac{1}{k!} - \frac{1}{(k+1)!}$$

例题 2.37 求和 $S_n = \sum_{k=1}^{n} \frac{1}{k(k+2)}$.

解 通项 $a_k = \frac{1}{k(k+2)}$ 符合部分分式分解的结构, 因此

$$a_k = \frac{1}{2}\left(\frac{1}{k} - \frac{1}{k+2}\right).$$

这是一个间隔为 2 的裂项. 我们写出和式以观察相消规律.

$$\begin{aligned} 2S_n &= \sum_{k=1}^{n}\left(\frac{1}{k} - \frac{1}{k+2}\right) \\ &= \left(1 - \frac{1}{3}\right) + \left(\frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{5}\right) + \cdots \\ &\quad + \left(\frac{1}{n-1} - \frac{1}{n+1}\right) + \left(\frac{1}{n} - \frac{1}{n+2}\right) \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

抵消发生在前一项的负项与后一项的正项之间, 例如 $-\frac{1}{3}$ 与 $+\frac{1}{3}$. 最终, 只有开头的两项正数 1 和 $\frac{1}{2}$, 以及结尾的两项负数 $-\frac{1}{n+1}$ 和 $-\frac{1}{n+2}$ 被保留下来.

$$2S_n = 1 + \frac{1}{2} - \frac{1}{n+1} - \frac{1}{n+2} = \frac{3}{2} - \frac{2n+3}{(n+1)(n+2)}$$

因此,

$$S_n = \frac{3}{4} - \frac{2n+3}{2(n+1)(n+2)} = \frac{n(3n+5)}{4(n+1)(n+2)}$$

例题 2.38　已知正项数列 $\{a_n\}$ 中, $a_1 = 1, a_{n+1}^2 - a_n^2 = 1$, 则数列 $\left\{\frac{1}{a_{n+1}+a_n}\right\}$ 的前 99 项和为(　)

(A) 4950

(B) 10

(C) 9

(D) $\frac{1}{4950}$

解　已知的是 $\{a_n\}$ 的递推, 所求却是 $b_n = \frac{1}{a_{n+1}+a_n}$ 的前 99 项和. 沟通已知与目标的关键, 在于把递推式等价变形为 b_n 的简单表达式.

递推左侧 $a_{n+1}^2 - a_n^2 = 1$ 是平方差, 因式分解得

$$(a_{n+1} - a_n)(a_{n+1} + a_n) = 1.$$

因子 $a_{n+1} + a_n$ 恰为 b_n 的分母. 因 $\{a_n\}$ 为正项数列, $a_{n+1} + a_n \neq 0$, 两边同除得

$$\frac{1}{a_{n+1} + a_n} = a_{n+1} - a_n.$$

分式被化归为差分, 原求和化为裂项相消:

$$\begin{aligned} S_{99} &= \sum_{n=1}^{99} \frac{1}{a_{n+1} + a_n} = \sum_{n=1}^{99} (a_{n+1} - a_n) \\ &= (a_2 - a_1) + (a_3 - a_2) + \cdots + (a_{100} - a_{99}) = a_{100} - a_1. \end{aligned}$$

余下只需求 a_{100}. 令 $c_n = a_n^2$, 则 $c_{n+1} - c_n = 1$, 故 $\{a_n^2\}$ 是首项 $c_1 = a_1^2 = 1$、公差 1 的等差数列, 通项 $c_n = n$, 即 $a_n^2 = n$. 由正项知 $a_n = \sqrt{n}$, 从而 $a_{100} = 10$.

$$S_{99} = a_{100} - a_1 = 10 - 1 = 9.$$

故选 (C).

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 2.39 数列 $\{a_n\}$ 的通项公式为 $a_n=\dfrac{2n+1}{n^2(n+1)^2}(n\in\mathbb{N}^*)$, 该数列的前 8 项和为 _____.

解 通项较复杂, 宜先裂项. 分母 $n^2(n+1)^2$ 恰为 $\frac{1}{n^2}$ 与 $\frac{1}{(n+1)^2}$ 通分后的分母, 提示 a_n 可能是二者之差. 直接验证:

$$\frac{1}{n^2}-\frac{1}{(n+1)^2}=\frac{(n+1)^2-n^2}{n^2(n+1)^2}=\frac{2n+1}{n^2(n+1)^2}=a_n.$$

故 $a_n=\frac{1}{n^2}-\frac{1}{(n+1)^2}$, 属标准裂项. 前 8 项和

$$S_8=\sum_{n=1}^{8}\Big(\frac{1}{n^2}-\frac{1}{(n+1)^2}\Big)=1-\frac{1}{9^2}=\frac{80}{81}.$$

例题 2.40 (2013 全国课标 I 卷) 已知等差数列 $\{a_n\}$ 的前 n 项和为 S_n, 满足 $S_3=0, S_5=-5$.

1. 求 $\{a_n\}$ 的通项公式;
2. 求数列 $\left\{\dfrac{1}{a_{2n-1}a_{2n+1}}\right\}$ 的前 n 项和.

解 设首项 a_1、公差 d. 由 $S_n=na_1+\frac{n(n-1)}{2}d$ 与已知条件,

$$\begin{cases}S_3=3a_1+3d=0\\S_5=5a_1+10d=-5\end{cases}\Longrightarrow\begin{cases}a_1+d=0\\a_1+2d=-1\end{cases},$$

解得 $d=-1, a_1=1$, 通项 $a_n=2-n$.

第二问令 $b_n=\frac{1}{a_{2n-1}a_{2n+1}}$. 由通项 $a_{2n-1}=3-2n, a_{2n+1}=1-2n$, 故

$$b_n=\frac{1}{(3-2n)(1-2n)}.$$

分母两因子之差为 2, 裂项得

$$b_n=\frac{1}{2}\Big(\frac{1}{1-2n}-\frac{1}{3-2n}\Big).$$

令 $f(k)=\frac{1}{3-2k}$, 则 $f(k+1)=\frac{1}{1-2k}$, 故 $b_k=\frac{1}{2}(f(k+1)-f(k))$.

令所求前 n 项和为 T_n, 则

$$T_n=\sum_{k=1}^{n}b_k=\frac{1}{2}\sum_{k=1}^{n}(f(k+1)-f(k))$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= \frac{1}{2}\left[(f(2)-f(1))+(f(3)-f(2))+\cdots+(f(n+1)-f(n))\right]$$

这是一个首尾相连的裂项求和, 中间项相互抵消, 最终结果为

$$\begin{aligned} T_n &= \frac{1}{2}(f(n+1)-f(1)) \\ &= \frac{1}{2}\left(\frac{1}{3-2(n+1)}-\frac{1}{3-2(1)}\right) \\ &= \frac{1}{2}\left(\frac{1}{1-2n}-1\right) \\ &= \frac{1}{2}\left(\frac{1-(1-2n)}{1-2n}\right) \\ &= \frac{n}{1-2n} \end{aligned}$$

例题 2.41 (2014 大纲卷) 已知等差数列 $\{a_n\}$ 的前 n 项和为 S_n、$a_1=10$、a_2 为整数, 且对任意 $n\in\mathbb{N}^*$ 都有 $S_n\le S_4$.

1. 求 $\{a_n\}$ 的通项公式;
2. 设 $b_n=\dfrac{1}{a_n a_{n+1}}$, 求数列 $\{b_n\}$ 的前 n 项和 T_n.

解 条件 $S_n\le S_4$ 对所有正整数 n 成立, 意味着数列 $\{S_n\}$ 在 $n=4$ 处取得最大值. 对于等差数列, 其前 n 项和 S_n 是关于 n 的二次函数, $S_n=\frac{d}{2}n^2+(a_1-\frac{d}{2})n$. 欲使 S_n 存在最大值, 其二次项系数必须为负, 即公差 $d<0$.

S_n 在 $n=4$ 取得最大值, 这也意味着从第 5 项开始, 数列的项 a_n 开始对总和产生负贡献, 或零贡献. 这等价于 $S_4\ge S_3$ 且 $S_4\ge S_5$. 由 $S_n-S_{n-1}=a_n$, 上述条件转化为:

$$\begin{cases} a_4\ge 0 \\ a_5\le 0 \end{cases}$$

利用通项公式 $a_n=a_1+(n-1)d$ 及已知 $a_1=10$, 得关于公差 d 的不等式组:

$$\begin{cases} 10+3d\ge 0 \\ 10+4d\le 0 \end{cases} \Longrightarrow \begin{cases} d\ge -\frac{10}{3} \\ d\le -\frac{10}{4} \end{cases}$$

即 $-\frac{10}{3}\le d\le -\frac{5}{2}$.

又因为 $a_2=a_1+d=10+d$ 是整数, 且 $a_1=10$ 是整数, 所以公差 d 必须是整数. 在区间 $[-\frac{10}{3},-\frac{5}{2}]$, 即 $[-3.33\ldots,-2.5]$ 内唯一的整数是 $d=-3$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

由此唯一确定该等差数列. 其通项公式为:

知乎@余命数

$$a_n = 10 + (n-1)(-3) = 13 - 3n$$

第二问求数列 $\{b_n\}$ 的和, 其通项为 $b_n = \dfrac{1}{a_n a_{n+1}}$. 代入 $a_n = 13 - 3n$, 有 $a_{n+1} = 13 - 3(n+1) = 10 - 3n$.

$$b_n = \frac{1}{(13-3n)(10-3n)}$$

分母两因子之差为 $(13-3n)-(10-3n)=3$. 可进行裂项分解:

知乎@余命数

$$b_n = \frac{1}{3}\left(\frac{1}{10-3n} - \frac{1}{13-3n}\right)$$

知乎@余命数

令 $f(k) = \dfrac{1}{13-3k}$, 那么 $f(k+1) = \dfrac{1}{13-3(k+1)} = \dfrac{1}{10-3k}$. 于是, $b_k = \dfrac{1}{3}(f(k+1) - f(k))$.

所求前 n 项和 T_n 为:

知乎@余命数

$$\begin{aligned} T_n &= \sum_{k=1}^{n} b_k = \frac{1}{3}\sum_{k=1}^{n}(f(k+1) - f(k)) \\ &= \frac{1}{3}\left[(f(2)-f(1)) + (f(3)-f(2)) + \cdots + (f(n+1)-f(n))\right] \end{aligned}$$

此为裂项求和, 结果为:

知乎@余命数

知乎@余命数

$$\begin{aligned} T_n &= \frac{1}{3}(f(n+1) - f(1)) \\ &= \frac{1}{3}\left(\frac{1}{13-3(n+1)} - \frac{1}{13-3(1)}\right) \\ &= \frac{1}{3}\left(\frac{1}{10-3n} - \frac{1}{10}\right) \\ &= \frac{1}{3}\left(\frac{10-(10-3n)}{10(10-3n)}\right) \\ &= \frac{n}{10(10-3n)} \end{aligned}$$

例题 2.42 (2017 全国 III 卷) 设数列 $\{a_n\}$ 满足 $a_1 + 3a_2 + \cdots + (2n-1)a_n = 2n$.

1. 求 $\{a_n\}$ 的通项公式;
2. 求数列 $\left\{\dfrac{a_n}{2n+1}\right\}$ 的前 n 项和.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 (1) 设 $T_n = a_1 + 3a_2 + \cdots + (2n-1)a_n = 2n$. 这个结构提示使用作差法来分离出通项 a_n. 当 $n \geq 2$ 时, 有 $T_{n-1} = a_1 + 3a_2 + \cdots + (2(n-1)-1)a_{n-1} = 2(n-1)$.

考察 T_n 与 T_{n-1} 的差:

$$\begin{aligned} T_n - T_{n-1} &= (a_1 + \cdots + (2n-1)a_n) - (a_1 + \cdots + (2n-3)a_{n-1}) \\ &= (2n-1)a_n \end{aligned}$$

另一方面,

$$T_n - T_{n-1} = 2n - 2(n-1) = 2$$

因此, 对于 $n \geq 2$, 有 $(2n-1)a_n = 2$, 即 $a_n = \dfrac{2}{2n-1}$.

接着, 需检验 $n = 1$ 的情况. 在原条件中令 $n = 1$, 可得 $1 \cdot a_1 = 2(1)$, 解得 $a_1 = 2$. 将 $n = 1$ 代入已得的公式, $a_1 = \dfrac{2}{2(1)-1} = 2$. 公式对 $n = 1$ 也成立. 故数列 $\{a_n\}$ 的通项公式为 $a_n = \dfrac{2}{2n-1}$.

(2) 设新数列为 $\{c_n\}$, 其通项为 $c_n = \dfrac{a_n}{2n+1}$. 将 a_n 的表达式代入, 得

$$c_n = \frac{2/(2n-1)}{2n+1} = \frac{2}{(2n-1)(2n+1)}$$

分母两因子之差为 $(2n+1) - (2n-1) = 2$, 这是一个典型的裂项求和结构.

$$c_n = \frac{1}{2n-1} - \frac{1}{2n+1}$$

令所求前 n 项和为 S_n, 得

$$\begin{aligned} S_n &= \sum_{k=1}^{n} \left(\frac{1}{2k-1} - \frac{1}{2k+1} \right) \\ &= \left(\frac{1}{1} - \frac{1}{3} \right) + \left(\frac{1}{3} - \frac{1}{5} \right) + \cdots + \left(\frac{1}{2n-1} - \frac{1}{2n+1} \right) \end{aligned}$$

中间项全部抵消, 得到:

$$S_n = 1 - \frac{1}{2n+1} = \frac{2n}{2n+1}$$

例题 2.43 (2011 大纲卷) 设数列 $\{a_n\}$ 满足 $a_1 = 0$, 且 $\dfrac{1}{1-a_{n+1}} - \dfrac{1}{1-a_n} = 1$.

1. 求 $\{a_n\}$ 的通项公式;
2. 设 $b_n = \dfrac{1-\sqrt{a_{n+1}}}{\sqrt{n}}$, 记 $S_n = \sum_{k=1}^{n} b_k$, 求证: $S_n < 1$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 (1) 观察给定的递推关系, 其形式暗示了换元法. 令新数列 $\{c_n\}$ 的通项为 $c_n = \dfrac{1}{1-a_n}$. 则原递推关系化为 $c_{n+1} - c_n = 1$. 故数列 $\{c_n\}$ 是一个公差为 1 的等差数列.

其首项为 $c_1 = \dfrac{1}{1-a_1} = \dfrac{1}{1-0} = 1$. 根据等差数列通项公式, $c_n = c_1 + (n-1)d = 1 + (n-1)\cdot 1 = n$.

由 $c_n = n$ 反解 a_n:

$$\frac{1}{1-a_n} = n \implies 1 - a_n = \frac{1}{n} \implies a_n = 1 - \frac{1}{n} = \frac{n-1}{n}$$

该公式即为 $\{a_n\}$ 的通项公式.

(2) 为了对 $\{b_n\}$ 求和, 先需化简其通项. 将 $a_{n+1} = \dfrac{(n+1)-1}{n+1} = \dfrac{n}{n+1}$ 代入 b_n 的表达式, 得

$$\sqrt{a_{n+1}} = \sqrt{\frac{n}{n+1}} = \frac{\sqrt{n}}{\sqrt{n+1}}$$

于是,

$$\begin{aligned} b_n &= \frac{1 - \frac{\sqrt{n}}{\sqrt{n+1}}}{\sqrt{n}} \\ &= \frac{(\sqrt{n+1} - \sqrt{n})/\sqrt{n+1}}{\sqrt{n}} \\ &= \frac{\sqrt{n+1} - \sqrt{n}}{\sqrt{n}\sqrt{n+1}} \\ &= \frac{\sqrt{n+1}}{\sqrt{n}\sqrt{n+1}} - \frac{\sqrt{n}}{\sqrt{n}\sqrt{n+1}} \\ &= \frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n+1}} \end{aligned}$$

通项 b_n 就此分解为可以裂项相消的形式.

接着, 计算前 n 项和 S_n, 得

$$\begin{aligned} S_n &= \sum_{k=1}^{n} b_k = \sum_{k=1}^{n} \left(\frac{1}{\sqrt{k}} - \frac{1}{\sqrt{k+1}} \right) \\ &= \left(\frac{1}{\sqrt{1}} - \frac{1}{\sqrt{2}} \right) + \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{3}} \right) + \cdots + \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n+1}} \right) \end{aligned}$$

中间项全部抵消, 得到:

$$S_n = \frac{1}{\sqrt{1}} - \frac{1}{\sqrt{n+1}} = 1 - \frac{1}{\sqrt{n+1}}$$

知乎@余命数 知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

因为 n 为正整数, $n+1\geq 2$, 所以 $\sqrt{n+1}>0$, 从而 $\dfrac{1}{\sqrt{n+1}}>0$. 因此, $S_n=1-\dfrac{1}{\sqrt{n+1}}<1$. 证毕.

例题 2.44 (2006 全国 I 卷理) 设数列 $\{a_n\}$ 的前 n 项和为 S_n, 且满足 $S_n=\dfrac{4}{3}a_n-\dfrac{1}{3}\cdot 2^{n+1}+\dfrac{2}{3}$ $(n\in\mathbb{N}^*)$.

1. 求首项 a_1 与通项公式 a_n;
2. 设 $T_n=\dfrac{2^n}{S_n}$, 求证: $\sum_{i=1}^{n}T_i<\dfrac{3}{2}$.

解 (1) 首先求首项 a_1. 当 $n=1$ 时, 数列的前 1 项和即为其本身, 故有 $S_1=a_1$. 将此条件代入已知关系式, 得

$$a_1=\frac{4}{3}a_1-\frac{1}{3}\cdot 2^{1+1}+\frac{2}{3}$$

$$a_1=\frac{4}{3}a_1-\frac{4}{3}+\frac{2}{3}\Longrightarrow -\frac{1}{3}a_1=-\frac{2}{3}$$

解得 $a_1=2$.

接下来求通项公式. 已知 S_n 与 a_n 的关系, 需建立一个只含数列项的递推关系. 通项与前 n 项和的最基本关系式 $a_n=S_n-S_{n-1}$, 当 $n\geq 2$ 时是实现这一目标的关键, 因为它能把关于前 n 项和的条件转化为相邻两项之间的递推关系.

对于 $n\geq 2$, 有

$$S_n=\frac{4}{3}a_n-\frac{1}{3}\cdot 2^{n+1}+\frac{2}{3}$$

$$S_{n-1}=\frac{4}{3}a_{n-1}-\frac{1}{3}\cdot 2^{n}+\frac{2}{3}$$

两式相减, 左侧为 $S_n-S_{n-1}=a_n$.

$$a_n=\left(\frac{4}{3}a_n-\frac{2}{3}\cdot 2^n\right)-\left(\frac{4}{3}a_{n-1}-\frac{1}{3}\cdot 2^n\right)+\frac{2}{3}-\frac{2}{3}$$

$$a_n=\frac{4}{3}a_n-\frac{4}{3}a_{n-1}-\frac{1}{3}(2^{n+1}-2^n)=\frac{4}{3}a_n-\frac{4}{3}a_{n-1}-\frac{1}{3}\cdot 2^n$$

整理上式以得到 a_n 与 a_{n-1} 的关系:

$$-\frac{1}{3}a_n=-\frac{4}{3}a_{n-1}-\frac{1}{3}\cdot 2^n\implies a_n=4a_{n-1}+2^n\quad(n\geq 2)$$

知乎@余命数

这是一个一阶线性递推关系. 为求解此式, 可两边同除以 4^n, 将其转化为一个更简单的形式.

$$\frac{a_n}{4^n}=\frac{4a_{n-1}}{4^n}+\frac{2^n}{4^n}\implies\frac{a_n}{4^n}=\frac{a_{n-1}}{4^{n-1}}+\left(\frac{1}{2}\right)^n$$

令 $c_n=\dfrac{a_n}{4^n}$, 则 $c_1=\dfrac{a_1}{4}=\dfrac{1}{2}$, 且 $c_n-c_{n-1}=(\frac{1}{2})^n$. 通过累加法,

$$c_n=c_1+\sum_{k=2}^{n}\left(\frac{1}{2}\right)^k=\frac{1}{2}+\frac{(\frac{1}{2})^2(1-(\frac{1}{2})^{n-1})}{1-\frac{1}{2}}=\frac{1}{2}+\frac{1}{2}\left(1-\frac{1}{2^{n-1}}\right)=1-\frac{1}{2^n}$$

因此 $a_n=4^nc_n=4^n\left(1-\frac{1}{2^n}\right)=4^n-2^n$. 检验 $n=1$ 时, $a_1=4^1-2^1=2$, 与所求一致. 故通项公式为 $a_n=4^n-2^n$.

(2) 先需得到 S_n 的显式表达式.

$$S_n=\sum_{k=1}^{n}a_k=\sum_{k=1}^{n}(4^k-2^k)=\sum_{k=1}^{n}4^k-\sum_{k=1}^{n}2^k$$

利用等比数列求和公式:

$$S_n=\frac{4(4^n-1)}{4-1}-\frac{2(2^n-1)}{2-1}=\frac{4^{n+1}-4}{3}-(2^{n+1}-2)=\frac{4^{n+1}-3\cdot2^{n+1}+2}{3}$$

S_n 的表达式看似复杂, 但若将其中的指数项看作一个整体, 它的结构便清晰起来. 将 2^{n+1} 视为整体变量后, 分子可直接因式分解, 从而为后续裂项提供结构. 令 $x=2^{n+1}$, 则 $4^{n+1}=(2^2)^{n+1}=(2^{n+1})^2=x^2$. 于是分子变为 $x^2-3x+2=(x-1)(x-2)$.

$$S_n=\frac{(2^{n+1}-1)(2^{n+1}-2)}{3}=\frac{2(2^{n+1}-1)(2^n-1)}{3}$$

现在可计算 T_n 的通项, 得

$$T_n=\frac{2^n}{S_n}=\frac{3\cdot2^n}{2(2^{n+1}-1)(2^n-1)}=\frac{3\cdot2^{n-1}}{(2^{n+1}-1)(2^n-1)}$$

分母是结构相似的项的乘积, 这强烈暗示了裂项相消的可能性. 尝试将其分解为 $\dfrac{A}{2^n-1}-\dfrac{B}{2^{n+1}-1}$ 的形式.

$$\frac{A(2^{n+1}-1)-B(2^n-1)}{(2^n-1)(2^{n+1}-1)}=\frac{(2A-B)2^n-(A-B)}{(2^n-1)(2^{n+1}-1)}$$

比较分子 $3\cdot2^{n-1}=\frac{3}{2}\cdot2^n$, 得到 $2A-B=\frac{3}{2}$ 且 $A-B=0$, 解得 $A=B=\frac{3}{2}$. 故

$$T_n=\frac{3}{2}\left(\frac{1}{2^n-1}-\frac{1}{2^{n+1}-1}\right)$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这是一个标准的裂项形式. 求其前 n 项和:

$$\begin{aligned}\sum_{i=1}^{n} T_i &= \frac{3}{2}\sum_{i=1}^{n}\left(\frac{1}{2^i-1}-\frac{1}{2^{i+1}-1}\right)\\ &= \frac{3}{2}\left[\left(\frac{1}{2^1-1}-\frac{1}{2^2-1}\right)+\cdots+\left(\frac{1}{2^n-1}-\frac{1}{2^{n+1}-1}\right)\right]\\ &= \frac{3}{2}\left(1-\frac{1}{2^{n+1}-1}\right)\end{aligned}$$

由于 $n\geq 1$, 则 $n+1\geq 2$, $2^{n+1}-1\geq 3>0$, 所以 $\dfrac{1}{2^{n+1}-1}>0$. 因此,

$$\sum_{i=1}^{n} T_i = \frac{3}{2}-\frac{3}{2(2^{n+1}-1)}<\frac{3}{2}$$

证毕.

例题 2.45 (2023 年 4 月杭州、长沙、郑州、西安部分名校联考) 已知正项数列 $\{a_n\}$ 的前 n 项和为 S_n, 且 $a_1=1$, $a_n=\sqrt{S_n}+\sqrt{S_{n-1}}(n\in\mathbb{N}^*$ 且 $n\geq 2)$.

1. 求数列 $\{a_n\}$ 的通项公式;
2. 设数列 $\left\{\dfrac{a_{n+2}}{2^n a_n a_{n+1}}\right\}$ 的前 n 项和为 T_n, 求证:$T_n<1$.

解 (1) 给定的递推关系式 $a_n=\sqrt{S_n}+\sqrt{S_{n-1}}$ 结构特殊, 直接处理较为困难. 观察此式的结构, 它与平方差公式 $(\sqrt{x}-\sqrt{y})(\sqrt{x}+\sqrt{y})=x-y$ 有着紧密的联系. 已知, 数列通项与其前 n 项和之间存在基本关系 $a_n=S_n-S_{n-1}$, 当 $n\geq 2$ 时. 这启发将 a_n 替换为 $(\sqrt{S_n}-\sqrt{S_{n-1}})(\sqrt{S_n}+\sqrt{S_{n-1}})$.

对于 $n\geq 2$, 有

$$a_n=S_n-S_{n-1}=(\sqrt{S_n}-\sqrt{S_{n-1}})(\sqrt{S_n}+\sqrt{S_{n-1}})$$

将已知条件 $a_n=\sqrt{S_n}+\sqrt{S_{n-1}}$ 代入上式, 得到

$$a_n=(\sqrt{S_n}-\sqrt{S_{n-1}})a_n$$

由于 $\{a_n\}$ 为正项数列, 故 $a_n>0$, 等式两边可以约去 a_n.

$$\sqrt{S_n}-\sqrt{S_{n-1}}=1\quad(n\geq 2)$$

故数列 $\{\sqrt{S_n}\}$ 从第二项起是一个公差为 1 的等差数列. 这一步把含根号的递推关系转化为等差数列问题, 是求出 S_n 与 a_n 的关键转折.

计算其首项:$\sqrt{S_1}=\sqrt{a_1}=\sqrt{1}=1$. 因此, 数列 $\{\sqrt{S_n}\}$ 是一个首项为 1, 公差为 1 的等差数列. 其通项为 $\sqrt{S_n}=1+(n-1)\cdot 1=n$. 故 $S_n=n^2$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

当 $n \geq 2$ 时,

$$a_n = S_n - S_{n-1} = n^2 - (n-1)^2 = 2n - 1$$

当 $n = 1$ 时, $a_1 = S_1 = 1^2 = 1$. 将 $n = 1$ 代入 $2n-1$ 得 1, 公式同样成立. 所以, $\{a_n\}$ 的通项公式为 $a_n = 2n - 1$.

(2) 设所求数列为 $\{b_n\}$, 其通项为 $b_n = \dfrac{a_{n+2}}{2^n a_n a_{n+1}}$. 将 $a_n = 2n - 1$ 代入, 得

$$a_{n+1} = 2(n+1) - 1 = 2n + 1$$

$$a_{n+2} = 2(n+2) - 1 = 2n + 3$$

知乎@余命数 知乎@余命数 知乎@余命数

于是,

$$b_n = \frac{2n+3}{2^n(2n-1)(2n+1)}$$

此通项结构复杂, 直接求和十分困难. 现将其分解为可以裂项相消的形式. 经观察, 有恒等式

$$b_n = \frac{1}{2^{n-1}(2n-1)} - \frac{1}{2^n(2n+1)}$$

(请自行通分验证此等式为何成立.) 这个形式是 $f(k) - f(k+1)$ 的理想裂项结构, 其中 $f(k) = \dfrac{1}{2^{k-1}(2k-1)}$.

计算前 n 项和 T_n, 得

$$\begin{aligned} T_n = \sum_{k=1}^{n} b_k &= \sum_{k=1}^{n} \left(\frac{1}{2^{k-1}(2k-1)} - \frac{1}{2^k(2k+1)} \right) \\ &= \left(\frac{1}{2^0 \cdot 1} - \frac{1}{2^1 \cdot 3} \right) + \left(\frac{1}{2^1 \cdot 3} - \frac{1}{2^2 \cdot 5} \right) \\ &\quad + \cdots + \left(\frac{1}{2^{n-1}(2n-1)} - \frac{1}{2^n(2n+1)} \right) \end{aligned}$$

中间项逐一抵消, 最终只剩下首项与末项:

$$T_n = \frac{1}{1} - \frac{1}{2^n(2n+1)} = 1 - \frac{1}{2^n(2n+1)}$$

因为 n 是正整数, 所以 $2^n > 0$ 且 $2n + 1 > 0$, 即 $\dfrac{1}{2^n(2n+1)}$ 是一个正数. 因此, $T_n = 1 - (\text{一个正数}) < 1$. 证毕.

构造裂项还有更系统的做法.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.8.2.1 待定系数法

这个方法根据通项 a_k 的结构**猜测**出差分原函数 $f(k)$ 的代数形式, 再用待定系数法精确求解.

目标是把 a_k 表示为 $f(k)-f(k+1)$ 或其变体. 考虑一个比 $\frac{1}{k(k+1)}$ 更复杂的例子.

例题 2.46　求和 $S_n=\sum_{k=1}^{n}\frac{1}{k(k+1)(k+2)}$.

知乎@余命数

解 通项 $a_k=\frac{1}{k(k+1)(k+2)}$, 需找一个函数 $f(k)$ 使得 $a_k=f(k)-f(k+1)$. a_k 的分母是三个连续项的乘积, 故猜测原函数 $f(k)$ 的分母少一项, 即只含两个连续因子.

设 $f(k)=\frac{A}{k(k+1)}$, 其中 A 是待定常数.

计算 $f(k)-f(k+1)$ 并令其等于 a_k, 得

$$\begin{aligned}f(k)-f(k+1)&=\frac{A}{k(k+1)}-\frac{A}{(k+1)(k+2)}\\&=A\left(\frac{1}{k(k+1)}-\frac{1}{(k+1)(k+2)}\right)\\&=A\cdot\frac{(k+2)-k}{k(k+1)(k+2)}\\&=\frac{2A}{k(k+1)(k+2)}\end{aligned}$$

欲使该结果等于 $a_k=\frac{1}{k(k+1)(k+2)}$. 通过比较分子, 得 $2A=1$, 解得 $A=\frac{1}{2}$.

由此得到通项分解

$$a_k=\frac{1}{2}\left(\frac{1}{k(k+1)}-\frac{1}{(k+1)(k+2)}\right)$$

这是一个标准的裂项形式, 其中 $f(k)=\frac{1}{2k(k+1)}$.

现在进行求和.

$$\begin{aligned}S_n&=\sum_{k=1}^{n}\frac{1}{2}\left(\frac{1}{k(k+1)}-\frac{1}{(k+1)(k+2)}\right)\\&=\frac{1}{2}\left[\left(\frac{1}{1\cdot2}-\frac{1}{2\cdot3}\right)+\left(\frac{1}{2\cdot3}-\frac{1}{3\cdot4}\right)\right.\\&\qquad\left.+\cdots+\left(\frac{1}{n(n+1)}-\frac{1}{(n+1)(n+2)}\right)\right]\\&=\frac{1}{2}\left(\frac{1}{2}-\frac{1}{(n+1)(n+2)}\right)\\&=\frac{1}{4}-\frac{1}{2(n+1)(n+2)}=\frac{n(n+3)}{4(n+1)(n+2)}\end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.8.2.2 差分与和分 *

裂项相消法的背后是离散化的微积分. 如果微分与积分互为逆运算, 那么数列中也有对应的运算, 即差分与和分.

定义 2.10 (差分算子) 设 $f(k)$ 是任意函数. 我们定义其(前向)**差分** $\Delta f(k)$ 为

$$\Delta f(k) = f(k+1) - f(k)$$

知乎@余命数

有了差分算子, 裂项求和就纳入了差分与和分的框架. 求和 $\sum_{k=m}^{n}(f(k+1)-f(k))$ 正是 $\sum_{k=m}^{n}\Delta f(k)$.

由此, **任何求和问题 $\sum_{k=1}^{n} a_k$ 都等价于寻找一个函数 $F(k)$, 称为 a_k 的反差分, 使得 $\Delta F(k) = F(k+1) - F(k) = a_k$.** 一旦找到反差分 $F(k)$, 求和只需计算端点, 得

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} \Delta F(k) = F(n+1) - F(1)$$

寻找反差分 $F(k)$ 的过程, 也就是构造裂项的过程, 这一思路可以推广到更广的通项.

寻找 $F(k)$ 时, 对多项式型的通项可再次运用待定系数法.

例题 2.47 运用反差分的思想, 推导平方和公式 $\sum_{k=1}^{n} k^2$.

解 现求 $S_n = \sum_{k=1}^{n} k^2$. 通项 $a_k = k^2$. 专注于寻找一个函数 $F(k)$ 使得 $\Delta F(k) = F(k+1) - F(k) = k^2$.

差分运算会使多项式的次数降低 1. 因此, 一个二次多项式 k^2 的反差分 $F(k)$ 必然是一个三次多项式. 设 $F(k) = Ak^3 + Bk^2 + Ck$.(常数项在差分中会消失, 故可省略).

计算其差分 $\Delta F(k)$, 得

$$\begin{aligned}
\Delta F(k) &= [A(k+1)^3 + B(k+1)^2 + C(k+1)] - [Ak^3 + Bk^2 + Ck] \\
&= A((k+1)^3 - k^3) + B((k+1)^2 - k^2) + C((k+1) - k) \\
&= A(3k^2 + 3k + 1) + B(2k+1) + C \\
&= (3A)k^2 + (3A + 2B)k + (A + B + C)
\end{aligned}$$

要求 $\Delta F(k) = k^2 = 1 \cdot k^2 + 0 \cdot k + 0$. 通过比较等式两边 k 的同次幂系数, 得到一个线性方程组.

$$k^2\text{的系数}: 3A = 1$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$k\text{的系数:}3A+2B=0$$

$$\text{常数项:}A+B+C=0$$

解这个方程组. 由方程一得 $A=\frac{1}{3}$. 代入方程二得 $3(\frac{1}{3})+2B=0 \implies 1+2B=0 \implies B=-\frac{1}{2}$. 代入方程三得 $\frac{1}{3}-\frac{1}{2}+C=0 \implies C=\frac{1}{6}$.

已找到反差分函数.

$$F(k)=\frac{1}{3}k^3-\frac{1}{2}k^2+\frac{1}{6}k$$

根据离散微积分基本定理, 平方和为

$$S_n=\sum_{k=1}^{n}k^2=F(n+1)-F(1)$$

计算 $F(1)=\frac{1}{3}-\frac{1}{2}+\frac{1}{6}=\frac{2-3+1}{6}=0$. 计算 $F(n+1)$, 得

$$\begin{aligned}F(n+1)&=\frac{1}{3}(n+1)^3-\frac{1}{2}(n+1)^2+\frac{1}{6}(n+1)\\&=\frac{n+1}{6}[2(n+1)^2-3(n+1)+1]\\&=\frac{n+1}{6}[2(n^2+2n+1)-3n-3+1]\\&=\frac{n+1}{6}[2n^2+4n+2-3n-2]\\&=\frac{n+1}{6}[2n^2+n]\\&=\frac{n(n+1)(2n+1)}{6}\end{aligned}$$

因此, $S_n=F(n+1)-0=\frac{n(n+1)(2n+1)}{6}$.

2.8.3 Gosper 裂项法 *

裂项相消法在特定结构的数列求和问题上成效显著. 无论部分分式分解还是阶乘恒等式, 都是把通项 a_k 改写为 $F(k)-F(k-1)$ 的形式. 然而, 这些技巧在很大程度上依赖于观察、经验与代数变形.

我们希望有一种系统性的方法, 也就是一种**算法**, 能够机械地判断一个给定的通项 a_k 是否存在一个简单的反差分 $F(k)$, 并在存在时直接把它算出来.

对于数学中最为常见和重要的一类序列, **超几何序列**, 美国数学家 R. William Gosper 在 20 世纪 70 年代给出了一个完整的解决方案.Gosper 算法将寻找裂项公式这一创造性活动, 转化为了一套可以由计算机执行的确定程序.

如今, 在 Mathematica 或 Maple 等计算机代数系统中输入复杂求和指令时, 其背后往往就是 Gosper 算法或其后继算法在运行.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.8.3.1 超几何项

Gosper 算法的适用范围, 是被称为超几何项的序列.

知乎@余命数

定义 2.11 (超几何项) 一个序列 $\{a_k\}$ 如果其相邻项之比 $\frac{a_{k+1}}{a_k}$ 是一个关于 k 的**有理函数**, 即两个关于 k 的多项式之比, 则称 $\{a_k\}$ 是一个**超几何项**或**超几何序列**.

这个定义涵盖了我们常遇到的大部分基本构件.

- 常数与幂函数.$a_k = c$ 时 $\frac{a_{k+1}}{a_k} = 1$; $a_k = k^m$ 时 $\frac{a_{k+1}}{a_k} = \left(\frac{k+1}{k}\right)^m$.
- 指数函数.$a_k = q^k$ 时 $\frac{a_{k+1}}{a_k} = q$.
- 阶乘.$a_k = k!$ 时 $\frac{a_{k+1}}{a_k} = k+1$.
- 二项式系数.$a_k = \binom{n}{k}$ 时 $\frac{a_{k+1}}{a_k} = \frac{n-k}{k+1}$.

更重要的是, 超几何项的集合在乘法、除法、平移等运算下是封闭的. 即, 由以上基本构件通过乘除组合而成的复杂项, 例如 $a_k = \frac{(k^2+1)k!}{2^k\binom{2k}{k}}$, 仍然是超几何项.

2.8.3.2 Gosper 的洞察与方法

知乎@余命数

面对一个超几何求和 $\sum a_k$, 现寻找其反差分 $F(k)$, 使得 $a_k = F(k) - F(k-1)$. Gosper 的第一个关键洞察如下.

定理 2.5 (Gosper 判别方程) 设非零超几何项 a_k 的相邻项比值为 $r(k) = \frac{a_k}{a_{k-1}}$. 若存在有理函数 $y(k)$ 满足

$$r(k)y(k) - y(k-1) = r(k),$$

则 $F(k) = y(k)a_k$ 是 a_k 的一个超几何反差分, 即 $F(k) - F(k-1) = a_k$.

证明 由 $F(k) = y(k)a_k$ 得

$$F(k) - F(k-1) = y(k)a_k - y(k-1)a_{k-1} = \left(y(k) - \frac{y(k-1)}{r(k)}\right)a_k.$$

判别方程 $r(k)y(k) - y(k-1) = r(k)$ 等价于 $y(k) - \frac{y(k-1)}{r(k)} = 1$, 因而 $F(k) - F(k-1) = a_k$.

知乎@余命数

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

又因为 a_k 是超几何项, $r(k)$ 为有理函数; $y(k)$ 为有理函数时,

知乎@余命数

$$\frac{F(k)}{F(k-1)}=\frac{y(k)}{y(k-1)}\,r(k)$$

也是关于 k 的有理函数(在相应表达式有意义处). 因此 $F(k)$ 也是超几何项. □

Gosper 完整算法进一步说明, 对超几何项求超几何反差分时, 搜索这样的有理函数 $y(k)$ 不会漏掉可求解的情形. 因而这个判别方程大幅缩小了对未知函数 $F(k)$ 的搜寻范围.

将 $F(k)=y(k)a_k$ 代入裂项关系式 $a_k=F(k)-F(k-1)$, 得

$$a_k=y(k)a_k-y(k-1)a_{k-1}$$

知乎@余命数

在 $a_k\neq 0$ 的情况下, 两边同除以 a_k, 得

$$1=y(k)-y(k-1)\frac{a_{k-1}}{a_k}$$

令 $r(k)=\dfrac{a_k}{a_{k-1}}$ 为相邻项的比值(这是一个已知的有理函数), 则上式变为

$$r(k)=r(k)y(k)-y(k-1)$$

这是一个关于未知有理函数 $y(k)$ 的函数方程.Gosper 的完整算法给出了求解这个方程的系统步骤, 要点是证明 $y(k)$ 必须具有一种特殊的多项式结构, 从而将问题转化为一个求解多项式系数的线性方程组问题.

用一个例子体会这种待定系数求解的思想.

知乎@余命数

例题 2.48 求和 $S_n=\sum_{k=1}^{n}(k^2+1)k!$.

解 通项 $a_k=(k^2+1)k!$ 是一个超几何项. 遵循 Gosper 的思想, 尝试寻找其反差分 $F(k)$, 并猜测 $F(k)$ 的结构与 a_k 相似, 设 $F(k)=P(k)\cdot k!$, 其中 $P(k)$ 是待定多项式.

计算 $F(k)-F(k-1)$, 得

$$\begin{aligned}F(k)-F(k-1)&=P(k)k!-P(k-1)(k-1)!\\&=P(k)\cdot k\cdot(k-1)!-P(k-1)(k-1)!\\&=[kP(k)-P(k-1)](k-1)!\end{aligned}$$

欲使该结果等于 $a_k=(k^2+1)k!=k(k^2+1)(k-1)!$.

由此得到关于未知多项式 $P(k)$ 的方程

$$kP(k)-P(k-1)=k(k^2+1)=k^3+k$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

接下来，比较等式两边多项式的次数以确定 $P(k)$ 的次数. 设 $\deg(P) = d$，则 $P(k) = c_d k^d + \ldots$ ($c_d \neq 0$). 方程左侧的最高次项来自 $kP(k)$, 其次数为 $d+1$. 方程右侧的最高次项是 k^3, 次数为 3. 因此, 必然有 $d+1=3$, 解得 $d=2$. 故 $P(k)$ 是一个二次多项式.

设 $P(k) = Ak^2 + Bk + C$. 代入方程 $kP(k) - P(k-1) = k^3 + k$:

$$\begin{aligned}\text{LHS} &= k(Ak^2+Bk+C) - [A(k-1)^2 + B(k-1) + C] \\ &= (Ak^3 + Bk^2 + Ck) - [A(k^2-2k+1) + B(k-1) + C] \\ &= Ak^3 + (B-A)k^2 + (C+2A-B)k + (-A+B-C)\end{aligned}$$

RHS $= 1 \cdot k^3 + 0 \cdot k^2 + 1 \cdot k + 0$.

比较等式两边各项系数, 得到一个线性方程组.

k^3系数: $A = 1$

k^2系数: $B - A = 0 \implies B = A = 1$

k^1系数: $C + 2A - B = 1 \implies C + 2(1) - 1 = 1 \implies C = 0$

k^0系数: $-A + B - C = 0 \implies -1 + 1 - 0 = 0$.(验证成立)

解得 $P(k) = 1 \cdot k^2 + 1 \cdot k + 0 = k^2 + k = k(k+1)$.

因此, 找到的反差分为

$$F(k) = P(k)k! = k(k+1)k! = k \cdot (k+1)!$$

于是求和只剩端点值的计算.

$$\begin{aligned}S_n = \sum_{k=1}^{n} a_k &= \sum_{k=1}^{n} [F(k) - F(k-1)] \\ &= F(n) - F(0) \\ &= n(n+1)! - 0 \cdot (0+1)! \\ &= n(n+1)!\end{aligned}$$

例题 2.49 求和 $S_n = \sum_{k=1}^{n} k \cdot k!$.

解 通项 $a_k = k \cdot k!$ 是超几何项. 寻找其反差分 $F(k)$, 使得 $a_k = F(k) - F(k-1)$.

一个自然的猜测是 $F(k)$ 与 $k!$ 具有相同的结构, 形如 $F(k) = P(k) \cdot k!$, 其中 $P(k)$ 是一个待定的多项式. 计算差分, 得

$$F(k) - F(k-1) = P(k)k! - P(k-1)(k-1)!$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= [kP(k) - P(k-1)](k-1)!$$

欲使该结果等于 $a_k = k \cdot k! = k^2(k-1)!$.

由此得到关于 $P(k)$ 的方程

$$kP(k) - P(k-1) = k^2$$

比较等式两边多项式的次数. 若 $\deg(P) = d$, 则左侧最高次项来自 $kP(k)$, 次数为 $d+1$. 右侧次数为 2. 故 $d+1=2 \implies d=1$.$P(k)$ 是一个一次多项式.

设 $P(k) = Ak + B$. 代入方程, 得

$$k(Ak+B) - [A(k-1)+B] = k^2$$

$$Ak^2 + Bk - Ak + A - B = k^2$$

$$Ak^2 + (B-A)k + (A-B) = 1 \cdot k^2 + 0 \cdot k + 0$$

比较系数. $A=1$. $B-A=0 \implies B=1$. $A-B=0 \implies 1-1=0$, 验证成立.

解得 $P(k) = k+1$. 因此, 反差分是 $F(k) = (k+1)k! = (k+1)!$.

求和.

$$S_n = \sum_{k=1}^{n} a_k = F(n) - F(0)$$

$$= (n+1)! - (0+1)!$$

$$= (n+1)! - 1$$

例题 2.50　求和 $S_n = \sum_{k=1}^{n} k \cdot 2^k$.

解　通项 $a_k = k \cdot 2^k$ 是超几何项. 猜测其反差分具有形式 $F(k) = P(k) \cdot 2^k$. 计算差分, 得

$$F(k) - F(k-1) = P(k)2^k - P(k-1)2^{k-1}$$

$$= [2P(k) - P(k-1)]2^{k-1}$$

欲使该结果等于 $a_k = k \cdot 2^k = 2k \cdot 2^{k-1}$.

由此得到关于 $P(k)$ 的方程

$$2P(k) - P(k-1) = 2k$$

比较次数可知 $\deg(P) = 1$. 设 $P(k) = Ak + B$. 代入方程, 得

$$2(Ak+B) - [A(k-1)+B] = 2k$$

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

$$2Ak+2B-Ak+A-B=2k$$

$$Ak+(A+B)=2k+0$$

比较系数. $A=2$. $A+B=0 \implies 2+B=0 \implies B=-2$.

解得 $P(k)=2k-2$. 因此, 反差分是 $F(k)=(2k-2)2^k=(k-1)2^{k+1}$.

求和.

$$\begin{aligned}S_n=\sum_{k=1}^{n}a_k&=F(n)-F(0)\\&=(n-1)2^{n+1}-(0-1)2^{0+1}\\&=(n-1)2^{n+1}-(-1)\cdot 2\\&=(n-1)2^{n+1}+2\end{aligned}$$

这个结果与使用错位相减法得到的结果完全一致.

例题 2.51 求和 $S_n=\sum_{k=1}^{n}k(k+1)(k+2)$.

解 通项 $a_k=k(k+1)(k+2)$ 是一个三次多项式. 寻找其反差分 $F(k)$ 使得 $F(k)-F(k-1)=a_k$. 差分运算会使多项式的次数减一, 故 $F(k)$ 必然是一个四次多项式.

这种上升阶乘幂的形式 $k^{\overline{m}}=k(k+1)\ldots(k+m-1)$ 具有简单的差分性质, 故猜测反差分 $F(k)$ 具有更高一阶的上升阶乘幂形式, 即 $F(k)=C\cdot k(k+1)(k+2)(k+3)$.

计算其差分, 得

$$\begin{aligned}F(k)-F(k-1)&=C\cdot k(k+1)(k+2)(k+3)-C\cdot(k-1)k(k+1)(k+2)\\&=C\cdot k(k+1)(k+2)\left[(k+3)-(k-1)\right]\quad\text{(提取公因子 } C\cdot k(k+1)(k+2)\text{)}\\&=C\cdot k(k+1)(k+2)\cdot 4.\end{aligned}$$

令此式等于 $a_k=k(k+1)(k+2)$, 比较可知 $4C=1 \implies C=\dfrac{1}{4}$.

因此, 反差分是 $F(k)=\dfrac{1}{4}k(k+1)(k+2)(k+3)$.

求和.

$$\begin{aligned}S_n=\sum_{k=1}^{n}a_k&=F(n)-F(0)\\&=\frac{1}{4}n(n+1)(n+2)(n+3)-\frac{1}{4}\cdot 0\cdot(1)\cdot(2)\cdot(3)\\&=\frac{n(n+1)(n+2)(n+3)}{4}\end{aligned}$$

这种方法比将 a_k 展开为 k^3+3k^2+2k 再利用幂和公式计算要简洁得多.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 2.52　求和 $S_n = \sum_{k=1}^{n} k^2$.

解 通项 $a_k = k^2$ 是一个二次多项式. 寻找其反差分 $F(k)$ 使得 $a_k = F(k) - F(k-1)$. 由于差分运算会使多项式的次数降低 1, 一个二次多项式的反差分必然是三次多项式. 设 $F(k) = Ak^3 + Bk^2 + Ck$(常数项在差分中会消失, 故可省略).

计算差分 $F(k) - F(k-1)$, 得

$$\begin{aligned} F(k) - F(k-1) &= (Ak^3 + Bk^2 + Ck) - [A(k-1)^3 + B(k-1)^2 + C(k-1)] \\ &= A(3k^2 - 3k + 1) + B(2k - 1) + C \\ &= 3Ak^2 + (2B - 3A)k + (A - B + C) \end{aligned}$$

欲使该结果等于 $a_k = 1 \cdot k^2 + 0 \cdot k + 0$.

比较系数可得

$$3A = 1 \implies A = 1/3$$

$$2B - 3A = 0 \implies 2B = 3(1/3) = 1 \implies B = 1/2$$

$$A - B + C = 0 \implies C = B - A = 1/2 - 1/3 = 1/6$$

因此, 反差分是 $F(k) = \dfrac{1}{3}k^3 + \dfrac{1}{2}k^2 + \dfrac{1}{6}k = \dfrac{k(k+1)(2k+1)}{6}$.

求和.

$$\begin{aligned} S_n = \sum_{k=1}^{n} a_k &= F(n) - F(0) \\ &= \frac{n(n+1)(2n+1)}{6} - 0 \\ &= \frac{n(n+1)(2n+1)}{6} \end{aligned}$$

例题 2.53　求和 $S_n = \sum_{k=1}^{n} \dfrac{k}{(k+1)!}$.

解 设 $F(k) = \dfrac{C}{(k+1)!}$. 计算差分, 得

$$\begin{aligned} F(k) - F(k-1) &= \frac{C}{(k+1)!} - \frac{C}{k!} \\ &= \frac{C - C(k+1)}{(k+1)!} \\ &= \frac{-Ck}{(k+1)!} \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

欲使该结果等于 $a_k = \dfrac{k}{(k+1)!}$. 比较可知 $-C = 1 \implies C = -1$.

因此, 反差分是 $F(k) = -\dfrac{1}{(k+1)!}$.

求和.

$$\begin{aligned} S_n = \sum_{k=1}^{n} a_k &= F(n) - F(0) \\ &= -\frac{1}{(n+1)!} - \left(-\frac{1}{(0+1)!}\right) \\ &= 1 - \frac{1}{(n+1)!} \end{aligned}$$

知乎@余命数

例题 2.54 求和 $S_n = \sum_{k=1}^{n} \frac{1}{(2k-1)(2k+1)}$.

知乎@余命数 知乎@余命数 知乎@余命数

解 通项 $a_k = \dfrac{1}{(2k-1)(2k+1)}$, 寻找其反差分 $F(k)$.

观察通项分母的结构, 猜测反差分 $F(k)$ 的分母可能只包含其中一个因子, 设 $F(k) = \dfrac{C}{2k+1}$.

计算差分 $F(k) - F(k-1)$, 得

$$\begin{aligned} F(k) - F(k-1) &= \frac{C}{2k+1} - \frac{C}{2(k-1)+1} \\ &= \frac{C}{2k+1} - \frac{C}{2k-1} \\ &= C\left(\frac{(2k-1)-(2k+1)}{(2k+1)(2k-1)}\right) \\ &= \frac{-2C}{(2k-1)(2k+1)} \end{aligned}$$

知乎@余命数

欲使该结果等于 $a_k = \dfrac{1}{(2k-1)(2k+1)}$. 比较分子可知 $-2C = 1 \implies C = -1/2$.

因此, 反差分是 $F(k) = -\dfrac{1}{2(2k+1)}$.

知乎@余命数

求和.

$$\begin{aligned} S_n = \sum_{k=1}^{n} a_k &= F(n) - F(0) \\ &= -\frac{1}{2(2n+1)} - \left(-\frac{1}{2(2\cdot 0+1)}\right) \\ &= -\frac{1}{2(2n+1)} + \frac{1}{2} \\ &= \frac{1}{2}\left(1 - \frac{1}{2n+1}\right) \end{aligned}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= \frac{1}{2} \cdot \frac{2n}{2n+1} = \frac{n}{2n+1}$$

此方法提供了一种绕过部分分式分解, 直接获得裂项公式的系统途径.

待定系数裂项法可按下列次序推进. 先观察通项 a_k 的代数结构, 猜测其反差分 $F(k)$ 可能的形式. 关键原则是, $F(k)$ 中非多项式的部分应与 a_k 保持一致, 而多项式部分的次数或系数待定.

设 $P(k), Q(k)$ 为关于 k 的多项式. 常见猜想结构如下.

- 若 $a_k = P(k)$(多项式), 则猜想 $F(k) = Q(k)$, 其中 $\deg(Q) = \deg(P) + 1$.
- 若 $a_k = P(k)q^k$(差比数列), 则猜想 $F(k) = Q(k)q^k$.
- 若 $a_k = P(k)k!$(阶乘相关), 则猜想 $F(k) = Q(k)k!$ 或 $Q(k)(k-1)!$.
- 若 a_k 为分式, 则猜想 $F(k)$ 为一个结构更简单的分式.

写出核心关系式 $F(k) - F(k-1) = a_k$, 并将带有待定系数的猜想形式代入此方程. 随后进入纯代数计算.

- 对代入后的方程化简, 通常需要通分或提取公因子, 如 q^{k-1} 或 $(k-1)!$.
- 化简后, 方程两边会呈现出关于 k 的多项式相等的形式.
- 通过比较等式两边多项式的最高次项, 确定待定多项式 $Q(k)$ 的次数.
- 设出 $Q(k)$ 的一般形式, 如 $Ak^2 + Bk + C$, 然后通过比较等式两边 k 的所有同次幂系数, 建立一个关于待定系数的线性方程组.
- 解这个方程组, 求出所有待定系数的值.

将已求出的系数代回, 得到确定的反差分函数 $F(k)$. 利用裂项求和的最终公式计算总和. 如果求和从 $k = 1$ 开始, 则

$$S_n = \sum_{k=1}^{n} a_k = F(n) - F(0)$$

注意仔细计算 $F(n)$ 和 $F(0)$ 的值.

2.8.4 错位相减法

我们用错位相减法处理**差比数列**求和. 差比数列的通项形如 $c_n = a_n b_n$, 其中 $\{a_n\}$ 是等差数列, $\{b_n\}$ 是等比数列.

求和时模仿等比数列求和公式的推导过程. 设等比部分的公比为 q.

1. 写出和式 $S_n = \sum_{k=1}^{n} a_k b_k$.
2. 将整个等式两边同乘以公比 q, 得到 qS_n.
3. 将 S_n 与 qS_n 的表达式对齐项错位相减, 得到 $(1-q)S_n$.
4. 相减后的结果, 通常会变成一个简单的等比数列之和, 外加首末两个独立项.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这个过程通过乘公比再作差, 把差比数列化简为一个纯粹的等比数列.

例题 2.55 求和 $S_n = \sum_{k=1}^{n} k \cdot x^k \ (x \neq 1)$.

解 这是由等差数列 $\{k\}$ 与等比数列 $\{x^k\}$ 构成的差比数列. 先写出和式

$$S_n = 1 \cdot x + 2 \cdot x^2 + 3 \cdot x^3 + \cdots + n \cdot x^n$$

知乎@余命数

两边同乘以公比 x, 得

$$xS_n = \quad 1 \cdot x^2 + 2 \cdot x^3 + \cdots + (n-1) \cdot x^n + n \cdot x^{n+1}$$

知乎@余命数

两式相减.

$$\begin{aligned}(1-x)S_n &= x + (2-1)x^2 + (3-2)x^3 + \cdots + (n-(n-1))x^n - n \cdot x^{n+1} \\ &= (x + x^2 + x^3 + \cdots + x^n) - n \cdot x^{n+1}\end{aligned}$$

括号内是一个首项为 x, 公比为 x 的等比数列之和.

$$(1-x)S_n = \frac{x(1-x^n)}{1-x} - nx^{n+1}$$

两边同除以 $(1-x)$, 得

$$S_n = \frac{x(1-x^n)}{(1-x)^2} - \frac{nx^{n+1}}{1-x}$$

知乎@余命数

2.8.5 倒序相加法

当数列的和式具有某种对称性 (典型特征是首末项之和、第二项与倒数第二项之和……遵循统一的规律) 时, 可把和式 S_n 的各项顺序颠倒, 得到另一个表达式, 再将两个表达式按位相加. 如果每一对应项的和都是常数或简单形式, 问题就得以简化. 高斯求和等差数列是此方法的经典应用.

阿贝尔变换与离散分部

注 2.7
裂项求和与分部积分在形式上对应. 前者对差分求和, 后者对导数积分.

知乎@余命数

2.8.6 阿贝尔 (Abel) 变换 *

在数列求和中常会遇到形如 $\sum_{k=1}^{n} a_k b_k$ 的乘积和. 错位相减法处理等差乘以等比这一特殊情况. 更一般的乘积求和可用 Abel 变换, 也称分部求和法.

这个方法把求和项 $a_k b_k$ 的求和难度在 $\{a_k\}$ 与 $\{b_k\}$ 之间转换. 可能 $\{a_k\}$ 易于求和, 例如等比数列, 其前 k 项和 A_k 有简洁公式, 而 $\{b_k\}$ 的**相邻项差分** $\Delta b_k = b_{k+1} - b_k$ 很简单, 例如多项式, 差分后降次.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

分部求和法把原求和式中对 a_k 的逐项累加，转化为对 A_k 与 Δb_k 乘积的求和，其中 A_k 是 a_k 的和，Δb_k 是 b_k 的差.

这样用一个易于求和的部分换取一个易于求差的部分，把原问题转化为阶次更低、结构更简单的求和问题.

熟悉微积分的场合中，这与分部积分法 $\int u\,dv = uv - \int v\,du$ 完全对偶.

定理 2.6 (阿贝尔变换) 设 $\{a_k\}$ 与 $\{b_k\}$ 为两个数列，并记 $A_n = \sum_{i=1}^{n} a_i$ 为 $\{a_k\}$ 的前 n 项和(规定 $A_0 = 0$). 则

$$\sum_{k=1}^{n} a_k b_k = A_n b_n - \sum_{k=1}^{n-1} A_k (b_{k+1} - b_k)$$

知乎@余命数

证明 现将求和式中的 a_k 替换掉，引入其部分和 A_k. 数列的通项 a_k 与其部分和 A_k 的最基本关系是差分关系. 当 $k \geq 1$ 时，a_k 等于其第 k 项部分和与第 $k-1$ 项部分和之差.

$$a_k = A_k - A_{k-1}$$

这是整个推导的逻辑起点.

将上述关系代入原求和式，并利用求和的线性性质将其拆分为两个部分，以便进行后续的代数变形.

知乎@余命数

$$\begin{aligned}\sum_{k=1}^{n} a_k b_k &= \sum_{k=1}^{n} (A_k - A_{k-1}) b_k \\ &= \sum_{k=1}^{n} A_k b_k - \sum_{k=1}^{n} A_{k-1} b_k\end{aligned}$$

为了将两个求和式合并，需让其中求和指标 k 所对应的项具有相同的结构，特别是让 A_k 的下标对齐. 对第二个求和式 $\sum_{k=1}^{n} A_{k-1} b_k$ 进行指标平移.

令新指标 $j = k - 1$. 当旧指标 k 从 1 变化到 n 时，新指标 j 从 0 变化到 $n-1$. 同时，旧通项中的 k 必须用 $j+1$ 替换. 于是

$$\sum_{k=1}^{n} A_{k-1} b_k = \sum_{j=0}^{n-1} A_j b_{j+1}$$

由于求和指标只是一个哑变量，可将 j 重新写为 k:

知乎@余命数

$$\sum_{k=0}^{n-1} A_k b_{k+1}$$

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

现在的表达式变为 $\sum_{k=1}^{n} A_k b_k - \sum_{k=0}^{n-1} A_k b_{k+1}$. 这两个和式的求和范围不完全重合. 为了合并它们, 将各自范围中不重合的项分离出来.

知乎@余命数

$$\begin{aligned}\sum_{k=1}^{n} a_k b_k &= \left(\sum_{k=1}^{n-1} A_k b_k + A_n b_n\right) - \left(A_0 b_1 + \sum_{k=1}^{n-1} A_k b_{k+1}\right) \\ &= A_n b_n - A_0 b_1 + \sum_{k=1}^{n-1} A_k b_k - \sum_{k=1}^{n-1} A_k b_{k+1}\end{aligned}$$

根据规定 $A_0 = 0$, 所以 $A_0 b_1 = 0$. 接着合并两个求和范围完全相同的和式, 得

$$\begin{aligned}\sum_{k=1}^{n} a_k b_k &= A_n b_n + \sum_{k=1}^{n-1} (A_k b_k - A_k b_{k+1}) \\ &= A_n b_n - \sum_{k=1}^{n-1} A_k (b_{k+1} - b_k)\end{aligned}$$

证毕. □

例题 2.56 再次使用阿贝尔变换求解差比数列和 $S_n = \sum_{k=1}^{n} k \cdot x^k$ $(x \neq 1)$.

解 需将通项 $k \cdot x^k$ 分解为 $a_k b_k$ 两部分. 阿贝尔变换的作用在于, 如果 $b_{k+1} - b_k$ 比 b_k 简单, 且 $A_k = \sum a_i$ 易于计算, 那么变换后的新求和式会变得更简单.

- 考虑 k: 它是一个一次多项式. 其差分 $(k+1) - k = 1$ 是一个常数, 形式急剧简化. 这使其成为 b_k 的理想选择.
- 考虑 x^k: 它是一个等比数列. 其部分和 $A_k = \sum_{i=1}^{k} x^i$ 有一个简洁的求和公式. 这使其成为 a_k 的理想选择.

因此令 $a_k = x^k$ 且 $b_k = k$.

根据所选, 计算阿贝尔变换公式所需的各个部分, 得

- $b_k = k \implies b_{k+1} - b_k = 1$.
- $A_k = \sum_{i=1}^{k} x^i = \frac{x(1-x^k)}{1-x}$.
- $A_n = \frac{x(1-x^n)}{1-x}$.
- $A_n b_n = \frac{nx(1-x^n)}{1-x}$.

将上述构件代入阿贝尔变换公式, 得

$$\sum_{k=1}^{n} k x^k = A_n b_n - \sum_{k=1}^{n-1} A_k (b_{k+1} - b_k)$$

知乎@余命数

$$S_n = \frac{nx(1-x^n)}{1-x} - \sum_{k=1}^{n-1} \frac{x(1-x^k)}{1-x} \cdot 1$$

知乎@余命数

问题转化为计算新的、但更简单的求和式 $\sum_{k=1}^{n-1} \frac{x(1-x^k)}{1-x}$.

$$\begin{aligned}\sum_{k=1}^{n-1} \frac{x(1-x^k)}{1-x} &= \frac{x}{1-x} \sum_{k=1}^{n-1} (1-x^k) \\ &= \frac{x}{1-x} \left(\sum_{k=1}^{n-1} 1 - \sum_{k=1}^{n-1} x^k \right) \\ &= \frac{x}{1-x} \left((n-1) - \frac{x(1-x^{n-1})}{1-x} \right)\end{aligned}$$

知乎@余命数　知乎@余命数

代回, 得

$$S_n = \frac{nx(1-x^n)}{1-x} - \frac{x}{1-x} \left(n-1 - \frac{x(1-x^{n-1})}{1-x} \right)$$

继续通分整理, 得

$$\begin{aligned}S_n &= \frac{nx(1-x^n)(1-x) - x(n-1)(1-x) + x^2(1-x^{n-1})}{(1-x)^2} \\ &= \frac{x-(n+1)x^{n+1}+nx^{n+2}}{(1-x)^2}.\end{aligned}$$

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

这与错位相减法得到的结果一致. 阿贝尔变换将一个需要巧妙构造的问题, 转化为了一个遵循固定范式的程序化计算.

2.8.6.1 部分分式型

我们来推导 $\frac{1}{k(k+d)}$ 的裂项形式.

希望找到 $f(k)$ 使得 $f(k) - f(k+1)$ 与 $\frac{1}{k(k+d)}$ 成比例. 差分运算通常会使分式的分母变得更复杂 (通分导致阶次升高) . 因此, 一个合理的猜想是, 反差分函数 $f(k)$ 的分母应该比原通项**更简单**. 原通项分母是两个因子的乘积, 我们猜想 $f(k)$ 的分母只包含其中一个因子.

我们不妨猜想 $f(k) = \frac{1}{k}$.

计算这个猜想的差分, 得

$$f(k) - f(k+1) = \frac{1}{k} - \frac{1}{k+1} = \frac{(k+1)-k}{k(k+1)} = \frac{1}{k(k+1)}$$

知乎@余命数　知乎@余命数

这个结果对于 $d=1$ 的情况直接成功. 对于更一般的情况, 我们猜想 $f(k)=\frac{1}{k}$.

$$f(k)-f(k+d)=\frac{1}{k}-\frac{1}{k+d}=\frac{(k+d)-k}{k(k+d)}=\frac{d}{k(k+d)}$$

这个结果与待求的 $\frac{1}{k(k+d)}$ 只差一个常数因子 d. 因此, 两边同除以 d 即得

$$\frac{1}{k(k+d)}=\frac{1}{d}\left(\frac{1}{k}-\frac{1}{k+d}\right)$$

这个裂项公式便由系统的猜想与验证推得.

2.8.6.2 阶乘添项型

我们来推导 $k\cdot k!$ 的裂项形式.

我们寻找一个函数 $f(k)$, 其差分 $f(k+1)-f(k)$ 等于 $k\cdot k!$. 阶乘世界的基本运算单元是 $k!$ 本身. 一个自然的猜想是, 反差分函数 $f(k)$ 应该与 $k!$ 结构相似.

最大胆的猜想是 $f(k)=k!$.

计算该猜想的差分, 得

$$f(k+1)-f(k)=(k+1)!-k!$$

利用阶乘的性质 $(k+1)!=(k+1)\cdot k!$:

$$(k+1)k!-k!=((k+1)-1)k!=k\cdot k!$$

猜想一次成功. 我们系统地推导出了恒等式 $k\cdot k!=(k+1)!-k!$.

2.8.6.3 平方差裂项型

我们寻找 $f(k)$ 使得 $f(k)-f(k+1)$ 等于目标通项. 观察分母 $k^2(k+1)^2$, 它提示我们反差分函数 $f(k)$ 的分母可能更简单, 例如只包含 k^2.

我们猜想 $f(k)=\dfrac{1}{k^2}$.

计算这个猜想的差分, 得

$$f(k)-f(k+1)=\frac{1}{k^2}-\frac{1}{(k+1)^2}=\frac{(k+1)^2-k^2}{k^2(k+1)^2}$$

计算分子, 得

$$(k+1)^2-k^2=(k^2+2k+1)-k^2=2k+1$$

因此,

$$f(k)-f(k+1)=\frac{2k+1}{k^2(k+1)^2}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.9 递推关系与通项公式

我们希望通过数列的局部规则得到它的整体表达式. 递推关系描述相邻项之间的联系, 通项公式则给出每一项的具体值. 本节建立由递推关系推导通项公式的代数框架.

2.9.1 和与项的互化关系

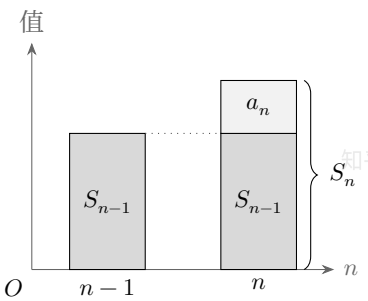
在数列研究中, 前 n 项和 S_n 与通项 a_n 包含的信息是等价的.S_n 描述了数列的累积性质, 而 a_n 描述了数列的局部特征. 两者通过差分运算相互转化.

2.9.1.1 基本原理

定义 2.12　设数列 $\{a_n\}$ 的前 n 项和为 $S_n=\sum_{k=1}^{n}a_k$. 则通项 a_n 与和 S_n 的关系如下.

$$a_n=\begin{cases}S_1 & n=1\\ S_n-S_{n-1} & n\geq 2\end{cases}$$

几何上, 这对应于从总长度中减去前一部分的长度, 剩余即为当前项.



2.9.1.2 由 S_n 的解析式求 a_n

这是最直接的应用. 计算时需严格区分 $n=1$ 与 $n\geq 2$ 的情况, 最后检验是否可以合并.

例题 2.57　已知数列 $\{a_n\}$ 的前 n 项和为 $S_n=n^2+1$. 求通项公式.

解　先计算首项, 得

$$n=1:\quad a_1=S_1=1^2+1=2$$

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

再计算一般项 $(n \geq 2)$, 得

$$\begin{aligned} a_n &= S_n - S_{n-1} \\ &= (n^2+1) - [(n-1)^2+1] \\ &= n^2 - (n^2 - 2n + 1) \\ &= 2n - 1 \end{aligned}$$

最后检验一致性. 令 $n = 1$ 代入 $2n-1$, 得 $2(1)-1 = 1$. 由于 $1 \neq a_1$ (实际 $a_1 = 2$), 故公式不能合并.

故通项公式为

$$a_n = \begin{cases} 2 & n = 1 \\ 2n-1 & n \geq 2 \end{cases}$$

知乎@余命数

注 2.8

若 $S_n = An^2 + Bn + C$, 则 $\{a_n\}$ 为等差数列的充要条件是常数项 $C = 0$. 上例中 $C = 1 \neq 0$, 故 a_1 游离于等差规则之外.

知乎@余命数 知乎@余

2.9.1.3 由含 S_n 的递推关系求 a_n

当已知条件给出 S_n 与 a_n 的混合关系时, 我们用移位作差消去 S_n, 把它化为 a_n 与 a_{n-1} 的纯递推关系.

例题 2.58 已知数列 $\{a_n\}$ 满足 $a_1 = 1$, 且 $2S_n = 3a_n - 1$. 求通项公式.

解 构造相邻项的等式即可消去 S_n.

当 $n \geq 2$ 时, 写出原式与降阶后的式子.

知乎@余命数

$$2S_n = 3a_n - 1 \tag{2-7}$$

$$2S_{n-1} = 3a_{n-1} - 1 \tag{2-8}$$

两式相减 (2-7) − (2-8), 得

$$2(S_n - S_{n-1}) = 3a_n - 3a_{n-1}$$

代入 $S_n - S_{n-1} = a_n$, 得

$$2a_n = 3a_n - 3a_{n-1}$$

知乎@余命数

整理得

$$a_n = 3a_{n-1}$$

故 $\{a_n\}$ 从第 2 项开始构成公比 $q=3$ 的等比数列.

下面检验首项. 由已知 $a_1=1$. 当 $n=2$ 时, 由递推式得 $a_2=3a_1=3$.

若通项为 $a_n=1\cdot 3^{n-1}$, 则 $a_1=3^0=1$, 符合.

因此, 通项公式为 $a_n=3^{n-1}$.

2.9.1.4 连乘积与项的转化

类似于和与项的关系, 若定义前 n 项积 $T_n=\prod_{k=1}^{n}a_k(a_k\neq 0)$, 则有乘法形式的互化关系.

定义 2.13

$$a_n=\begin{cases}T_1 & n=1\\ \frac{T_n}{T_{n-1}} & n\geq 2\end{cases}$$

例题 2.59　已知数列 $\{a_n\}$ 的前 n 项积 $T_n=n^2$. 求通项公式.

解　当 $n=1$ 时, $a_1=T_1=1^2=1$.

当 $n\geq 2$ 时,

$$a_n=\frac{T_n}{T_{n-1}}=\frac{n^2}{(n-1)^2}=\left(\frac{n}{n-1}\right)^2$$

检验 $n=1$: $\left(\frac{1}{0}\right)^2$ 无意义, 或趋于无穷, 且原式计算 $n=1$ 时代入公式分母为零. 故公式不可合并, 通项公式为

$$a_n=\begin{cases}1 & n=1\\ \left(\frac{n}{n-1}\right)^2 & n\geq 2\end{cases}$$

常见类型可归纳如下.

2.9.1.5 S_n 为显式函数

当 S_n 以 n 的函数形式给出时, 直接利用定义求解. 注意常数项往往导致首项不符合后续规律.

例题 2.60 (2022 全国卷)　已知数列 $\{a_n\}$ 的前 n 项和为 $S_n=n^2+4n-3$. 求通项公式.

解　先考察 $n=1$ 的情形.

$$a_1=S_1=1^2+4(1)-3=2$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

再考察 $n \geq 2$ 的情形.

$$\begin{aligned} a_n &= S_n - S_{n-1} \\ &= (n^2 + 4n - 3) - [(n-1)^2 + 4(n-1) - 3] \\ &= n^2 + 4n - 3 - (n^2 - 2n + 1 + 4n - 4 - 3) \\ &= n^2 + 4n - 3 - (n^2 + 2n - 6) \\ &= 2n + 3 \end{aligned}$$

再检验一致性. 令 $n = 1$, 得 $2(1) + 3 = 5$. 由于 $5 \neq a_1$(实际 $a_1 = 2$), 故通项公式不统一, 即

$$a_n = \begin{cases} 2 & n = 1 \\ 2n + 3 & n \geq 2 \end{cases}$$

知乎@余命数

2.9.1.6 线性混合递推

当条件给出 a_n 与 S_n 的线性关系时, 利用移位作差法消去 S_n, 转化为 a_n 与 a_{n-1} 的纯递推关系.

例题 2.61 (2013 全国卷) 若数列 $\{a_n\}$ 的前 n 项和为 $S_n = \frac{2}{3}a_n + \frac{1}{3}$, 求数列 $\{a_n\}$ 的通项公式.

知乎@余命数

解 先确定首项 a_1. 令 $n = 1$, 得

$$S_1 = \frac{2}{3}a_1 + \frac{1}{3} \implies a_1 = \frac{2}{3}a_1 + \frac{1}{3} \implies \frac{1}{3}a_1 = \frac{1}{3} \implies a_1 = 1$$

对于 $n \geq 2$, 利用移位作差法.

$$\begin{aligned} S_n &= \frac{2}{3}a_n + \frac{1}{3} \\ S_{n-1} &= \frac{2}{3}a_{n-1} + \frac{1}{3} \end{aligned}$$

知乎@余命数 知乎@余命数

两式相减得 $a_n = \frac{2}{3}a_n - \frac{2}{3}a_{n-1}$. 移项整理, 得

$$\frac{1}{3}a_n = -\frac{2}{3}a_{n-1} \implies a_n = -2a_{n-1}$$

由此可知, 数列 $\{a_n\}$ 从第 2 项开始满足公比 $q = -2$ 的等比关系. 下面验证首项. 若公式 $a_n = a_1 \cdot q^{n-1}$ 对 $n = 1$ 成立, 则需 $1 \cdot (-2)^0 = 1$, 这与实际 a_1 一致.

因此, 通项公式为 $a_n = (-2)^{n-1}$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.9.1.7 隐式与二次关系

当关系式中包含 a_n 的高次项时, 通常需结合代数变形, 如配方, 挖掘其内在的结构.

例题 2.62 (2015 全国 I 卷) 已知 $a_n>0$, 且 $a_n^2+2a_n=4S_n+3$. 求 $\{a_n\}$ 的通项公式.

解 先求 a_1. 令 $n=1$, 得

$$a_1^2+2a_1=4a_1+3 \implies a_1^2-2a_1-3=0 \implies (a_1-3)(a_1+1)=0$$

因 $a_n>0$, 解得 $a_1=3$.

对于 $n\geq 2$, 观察原式结构, 等号左边配方为 $(a_n+1)^2-1$.

$$(a_n+1)^2=4S_n+4$$

同理可写出

$$(a_{n-1}+1)^2=4S_{n-1}+4$$

两式相减, 得

$$(a_n+1)^2-(a_{n-1}+1)^2=4(S_n-S_{n-1})=4a_n$$

利用平方差公式展开左边, 或者直接展开计算, 得

$$(a_n^2+2a_n+1)-(a_{n-1}+1)^2=4a_n$$

$$a_n^2-2a_n+1=(a_{n-1}+1)^2$$

$$(a_n-1)^2=(a_{n-1}+1)^2$$

两边开方. 由于 $a_n>0$, 只能取正号 $a_n-1=a_{n-1}+1$(若取负号则导致 $a_n+a_{n-1}=0$, 矛盾).

$$a_n-a_{n-1}=2$$

故 $\{a_n\}$ 是首项为 3, 公差为 2 的等差数列.

通项公式为 $a_n=3+(n-1)2=2n+1$.

2.9.1.8 结构化消元

当递推式呈现特定的代数结构, 如平方差形式、根式形式时, 应优先考虑整体代换.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 2.63 (2020 辽宁) 已知数列 $\{a_n\}$ 各项均为正数, 其前 n 项和为 S_n. 若 $a_1=1$, 且 $\sqrt{S_{n+1}}+\sqrt{S_n}=a_{n+1}$. 求通项公式.

解 利用 $a_{n+1}=S_{n+1}-S_n$, 将原式变形为

$$\sqrt{S_{n+1}}+\sqrt{S_n}=S_{n+1}-S_n$$

利用平方差公式分解右侧, 得

$$\sqrt{S_{n+1}}+\sqrt{S_n}=(\sqrt{S_{n+1}}-\sqrt{S_n})(\sqrt{S_{n+1}}+\sqrt{S_n})$$

由于各项为正, $\sqrt{S_{n+1}}+\sqrt{S_n}\neq 0$, 两边约去该公因子, 得

$$1=\sqrt{S_{n+1}}-\sqrt{S_n}$$

故数列 $\{\sqrt{S_n}\}$ 是一个等差数列, 公差 $d=1$.

首项为 $\sqrt{S_1}=\sqrt{a_1}=1$. 所以 $\sqrt{S_n}=1+(n-1)\cdot 1=n$. 两边平方得 $S_n=n^2$.

最后还原 a_n: 当 $n=1$ 时, $a_1=1$. 当 $n\geq 2$ 时, $a_n=S_n-S_{n-1}=n^2-(n-1)^2=2n-1$. 经检验, $n=1$ 时 $2(1)-1=1$, 符合.

故通项公式为 $a_n=2n-1$.

2.10 一阶递推的基本形式

当递推关系仅涉及相邻两项的差或比时, 数列的通项可以通过遍历所有项的积累效应(求和或求积)直接还原.

2.10.1 累加法

形如 $a_{n+1}-a_n=f(n)$ 的递推关系. 该式表明, 数列的增长率是关于 n 的函数.

定理 2.7 若 $a_{n+1}-a_n=f(n)$, 则

$$a_n=a_1+\sum_{k=1}^{n-1}f(k)\quad(n\geq 2)$$

证明 对递推式从 $k=1$ 到 $k=n-1$ 求和, 得

$$\sum_{k=1}^{n-1}(a_{k+1}-a_k)=\sum_{k=1}^{n-1}f(k).$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

左侧裂项相消为 $a_n - a_1$, 故

$$a_n = a_1 + \sum_{k=1}^{n-1} f(k).$$

□

知乎@余命数

裂项相消求解的要点在于计算 $\sum f(k)$. 若 $f(k)$ 能被拆解为 $g(k+1) - g(k)$ 的形式, 求和过程将极度简化.

例题 2.64 (根式裂项) 已知数列 $\{a_n\}$ 满足 $a_1 = 1$, 且 $a_{n+1} = a_n + \frac{1}{\sqrt{n}+\sqrt{n+1}}$. 求 a_n.

解 将递推项变形, 用分母有理化构造差分形式, 得

$$f(n) = \frac{1}{\sqrt{n}+\sqrt{n+1}} = \frac{\sqrt{n+1}-\sqrt{n}}{(n+1)-n} = \sqrt{n+1}-\sqrt{n}$$

原递推式化为 $a_{n+1} - a_n = \sqrt{n+1} - \sqrt{n}$.
对 $n \geq 2$ 累加, 得

$$\begin{aligned} a_n &= a_1 + \sum_{k=1}^{n-1} (\sqrt{k+1} - \sqrt{k}) \\ &= 1 + [(\sqrt{2}-\sqrt{1}) + (\sqrt{3}-\sqrt{2}) + \cdots + (\sqrt{n}-\sqrt{n-1})] \end{aligned}$$

中间项全部抵消, 只留首尾.

$$a_n = 1 + (\sqrt{n} - \sqrt{1}) = \sqrt{n}$$

检验 $n = 1$, $\sqrt{1} = 1$, 符合. 故 $a_n = \sqrt{n}$.

例题 2.65 (分式裂项) 已知 $a_1 = 2$, 且 $a_{n+1} = a_n + \frac{1}{n^2+2n}$. 求 a_n.

解 对 $f(n) = \frac{1}{n(n+2)}$ 作部分分式分解, 得

$$\frac{1}{n(n+2)} = \frac{1}{2}\left(\frac{1}{n} - \frac{1}{n+2}\right)$$

累加求和(注意跨度为 2, 需保留两项), 得

$$\begin{aligned} a_n - a_1 &= \frac{1}{2}\sum_{k=1}^{n-1}\left(\frac{1}{k} - \frac{1}{k+2}\right) \\ &= \frac{1}{2}\left[\left(1 - \frac{1}{3}\right) + \left(\frac{1}{2} - \frac{1}{4}\right) + \left(\frac{1}{3} - \frac{1}{5}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n+1}\right)\right] \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

剩余项为起始的正项 $1,\frac{1}{2}$ 与末尾的负项 $-\frac{1}{n},-\frac{1}{n+1}$：

$$a_n-2=\frac{1}{2}\left(1+\frac{1}{2}-\frac{1}{n}-\frac{1}{n+1}\right)$$

知乎@余命数

整理得

$$a_n=\frac{11}{4}-\frac{2n+1}{2n(n+1)}$$

知乎@余命数

例题 2.66 (2020 内蒙古) 在数列 $\{a_n\}$ 中，$a_1=3,a_{n+1}=a_n+\frac{3}{n(n+1)}$. 求通项公式 a_n.

解 差分项为 $f(n)=\frac{3}{n(n+1)}$. 作部分分式分解，把 $f(n)$ 裂项.

$$f(n)=3\left(\frac{1}{n}-\frac{1}{n+1}\right)$$

当 $n\geq 2$ 时累加，得

$$\begin{aligned}a_n&=a_1+\sum_{k=1}^{n-1}f(k)\\&=3+3\sum_{k=1}^{n-1}\left(\frac{1}{k}-\frac{1}{k+1}\right)\\&=3+3\left[\left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{n-1}-\frac{1}{n}\right)\right]\end{aligned}$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

中间项相互抵消，只留首尾.

$$a_n=3+3\left(1-\frac{1}{n}\right)=6-\frac{3}{n}$$

检验 $n=1$：$6-3=3$，与已知一致. 故 $a_n=6-\frac{3}{n}$.

2.10.1.1 构造性累加

当递推关系不直接为差分形式，但两边结构相似时，可通过构造新数列转化为累加问题.

例题 2.67 (2021 全国乙卷) 在数列 $\{a_n\}$ 中，$a_1=2$，且 $\frac{a_{n+1}}{n+1}=\frac{a_n}{n}+\ln(1+\frac{1}{n})$. 求 a_n.

解 等式两边形式均为项除以项数. 令 $b_n=\frac{a_n}{n}$，则原式化为

$$b_{n+1}=b_n+\ln\left(\frac{n+1}{n}\right)$$

知乎关注@余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

即 $b_{n+1}-b_n=\ln(n+1)-\ln n$.

这是一个标准的累加形式. 对于 $n\geq 2$:

$$\begin{aligned} b_n &= b_1+\sum_{k=1}^{n-1}[\ln(k+1)-\ln k] \\ &= \frac{a_1}{1}+(\ln n-\ln 1) \\ &= 2+\ln n \end{aligned}$$

所以 $b_n=2+\ln n$.

还原 a_n:

$$a_n=nb_n=n(2+\ln n)$$

经检验 $n=1$ 时成立.

2.10.2 累乘法

形如 $a_{n+1}=f(n)\cdot a_n$ 的递推关系. 该式表明, 数列的增长比例是关于 n 的函数.

定理 2.8 若数列各项非零, 且 $\frac{a_{n+1}}{a_n}=f(n)$、$f(n)\neq 0$, 则

$$a_n=a_1\cdot\prod_{k=1}^{n-1}f(k)\quad(n\geq 2)$$

证明 对等式 $\frac{a_{k+1}}{a_k}=f(k)$ 从 $k=1$ 到 $k=n-1$ 连乘, 得

$$\prod_{k=1}^{n-1}\frac{a_{k+1}}{a_k}=\prod_{k=1}^{n-1}f(k).$$

左侧相邻分子分母逐项约去, 只剩 $\frac{a_n}{a_1}$. 由于 $a_1\neq 0$, 两边同乘 a_1, 即得

$$a_n=a_1\prod_{k=1}^{n-1}f(k).$$

□

因子缩并求解要点在于将 $f(k)$ 写成分子与分母错位对应的形式, 使连乘积发生大规模约分.

知乎@余命数

例题 2.68 (Wallis 型约分) 已知数列 $\{a_n\}$ 满足 $a_1=1$, $a_2=1$, 且 $a_{n+1}=\frac{n^2}{n^2-1}a_n$ $(n\geq 2)$. 求 a_n.

解 注意 $n^2-1=(n-1)(n+1)$. 递推式从 $n=2$ 开始, 故需由 a_2 起累乘.

将因子结构化.

$$\frac{a_{k+1}}{a_k}=\frac{k\cdot k}{(k-1)(k+1)}$$

累乘展开, 得

$$a_n=a_2\cdot\prod_{k=2}^{n-1}\frac{k^2}{k^2-1}=a_2\cdot\left(\frac{2\cdot 2}{1\cdot 3}\right)\cdot\left(\frac{3\cdot 3}{2\cdot 4}\right)\cdot\left(\frac{4\cdot 4}{3\cdot 5}\right)\cdots\left(\frac{(n-1)(n-1)}{(n-2)n}\right)$$

观察交叉约分规律.

$$a_n=a_2\cdot\frac{2}{1}\cdot\underbrace{\left(\frac{2}{3}\cdot\frac{3}{2}\right)}_{1}\cdot\underbrace{\left(\frac{3}{4}\cdot\frac{4}{3}\right)}_{1}\cdots\cdot\frac{n-1}{n}$$

剩余项为

$$a_n=a_2\cdot 2\cdot\frac{n-1}{n}=\frac{2(n-1)}{n}a_2$$

由 $a_2=1$, 得

$$a_1=1,\qquad a_n=\frac{2(n-1)}{n}\quad(n\geq 2).$$

例题 2.69 (指数型累乘) 已知 $a_1=1$, 且 $a_{n+1}=3^n a_n$. 求 a_n.

解 由递推式可知

$$\frac{a_{n+1}}{a_n}=3^n$$

累乘得

$$a_n=a_1\cdot\prod_{k=1}^{n-1}3^k=1\cdot 3^1\cdot 3^2\cdot 3^3\cdots 3^{n-1}$$

于是转化为指数求和(等差数列).

$$a_n=3^{1+2+\cdots+(n-1)}=3^{\frac{(n-1)n}{2}}$$

知乎@余命数

例题 2.70 (2023 全国乙卷) 数列 $\{a_n\}$ 满足 $a_1=\frac{1}{2}$, 且 $S_n=n^2a_n$. 求 $\{a_n\}$ 的通项公式.

解 我们先寻找相邻项的比值关系. 当 $n\geq 2$ 时, 由 $a_n=S_n-S_{n-1}$ 出发,

$$a_n=n^2a_n-(n-1)^2a_{n-1}$$

移项整理, 得

$$(n-1)^2a_{n-1}=(n^2-1)a_n$$

$$(n-1)^2a_{n-1}=(n-1)(n+1)a_n$$

因 $n\geq 2$, 可约去 $n-1$:

$$(n-1)a_{n-1}=(n+1)a_n \implies \frac{a_n}{a_{n-1}}=\frac{n-1}{n+1}$$

下面用累乘法求解 a_n $(n\geq 2)$.

$$\begin{aligned} a_n&=a_1\cdot\prod_{k=2}^{n}\frac{k-1}{k+1}\\ &=\frac{1}{2}\cdot\left(\frac{1}{3}\cdot\frac{2}{4}\cdot\frac{3}{5}\cdot\frac{4}{6}\cdots\frac{n-2}{n}\cdot\frac{n-1}{n+1}\right)\end{aligned}$$

观察分子分母的对消模式(跨度为 2). 分子保留 $1,2$ 两项. 分母保留 $n,n+1$ 两项.

$$a_n=\frac{1}{2}\cdot\frac{1\cdot 2}{n(n+1)}=\frac{1}{n(n+1)}$$

检验 $n=1$: $\frac{1}{1(2)}=\frac{1}{2}$, 与 a_1 一致. 故 $a_n=\frac{1}{n(n+1)}$.

2.10.2.1 加权和的剥离

形如 $\sum c_ka_k=f(n)$ 的关系式, 实质上是给出了一个新数列 $\{c_na_n\}$ 的前 n 项和. 利用差分即可求出 c_na_n, 进而求出 a_n.

例题 2.71 (2019 江苏) 已知 $\sum_{k=1}^{n}2^{k-1}a_k=9-6n$. 求数列 $\{a_n\}$ 的通项公式.

解 令 $T_n=\sum_{k=1}^{n}2^{k-1}a_k=9-6n$. 则数列 $\{2^{n-1}a_n\}$ 的第 n 项可表示为 T_n-T_{n-1}.

当 $n=1$ 时,

$$2^0a_1=T_1=9-6(1)=3 \implies a_1=3$$

知乎@余命数
知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

当 $n \geq 2$ 时,

$$\begin{aligned} 2^{n-1}a_n &= T_n - T_{n-1} \\ &= (9-6n) - [9-6(n-1)] \\ &= 9-6n-(9-6n+6) \\ &= -6 \end{aligned}$$

解出 a_n:

$$a_n = \frac{-6}{2^{n-1}} = -3 \cdot \frac{2}{2^{n-1}} = -3 \cdot \left(\frac{1}{2}\right)^{n-2}$$

检验 $n=1$: 代入公式得 $-3 \cdot (\frac{1}{2})^{-1} = -6$, 而实际 $a_1 = 3$. 故公式不统一, 通项须分段给出.

$$a_n = \begin{cases} 3 & n = 1 \\ -3 \cdot (\frac{1}{2})^{n-2} & n \geq 2 \end{cases}$$

知乎@余命数

2.10.3 结构匹配与变形

有时递推关系并不直接呈现为差或比的形式, 但通过移项或同除, 可以构造出具备上述特征的新数列.

同型构造若递推式形如 $g(n+1)a_{n+1} = g(n)a_n$, 则可定义 $b_n = g(n)a_n$, 此时 $b_{n+1} = b_n$, 即 $\{b_n\}$ 为常数列.

例题 2.72 已知 $a_1 = 1$, 且 $na_{n+1} = (n+1)a_n + 2n(n+1)$. 求 a_n.

解 观察系数 n 与 $n+1$, 将等式两边同除以 $n(n+1)$, 得

$$\frac{a_{n+1}}{n+1} = \frac{a_n}{n} + 2$$

令 $b_n = \frac{a_n}{n}$. 则 $b_{n+1} - b_n = 2$. 数列 $\{b_n\}$ 是公差为 2 的等差数列.
首项为 $b_1 = \frac{a_1}{1} = 1$.

$$b_n = 1 + (n-1)2 = 2n-1$$

还原 a_n:

$$a_n = nb_n = n(2n-1) = 2n^2 - n$$

知乎@余命数

2.10.4 线性递推与代数构造

处理 $a_{n+1}=pa_n+f(n)$ $(p\neq 1)$ 形式的递推关系, 一般通过待定系数法或同除变形, 构造新的等比数列或等差数列.

2.10.4.1 常数项构造

当 $f(n)=q$ 为常数时, 即 $a_{n+1}=pa_n+q$. 设 $a_{n+1}+\lambda=p(a_n+\lambda)$, 展开比对得 $\lambda(p-1)=q$, 即 $\lambda=\frac{q}{p-1}$. 此时数列 $\{a_n+\lambda\}$ 为公比为 p 的等比数列.

2.10.4.2 同除构造

当 $f(n)$ 为指数函数 $c\cdot p^n$ 或多项式时, 常在等式两边同除以 p^{n+1}, 转化为累加形式.

例题 2.73 已知 $a_1=1$, 且 $a_{n+1}=2a_n+3^n$. 求 a_n.

知乎@余

证明 (解析) 观察非齐次项 3^n, 等式两边同除以 3^{n+1}, 得

$$\frac{a_{n+1}}{3^{n+1}}=\frac{2a_n}{3\cdot 3^n}+\frac{3^n}{3^{n+1}}$$

整理得

$$\frac{a_{n+1}}{3^{n+1}}=\frac{2}{3}\cdot\frac{a_n}{3^n}+\frac{1}{3}$$

令 $b_n=\frac{a_n}{3^n}$, 则 $b_{n+1}=\frac{2}{3}b_n+\frac{1}{3}$. 设 $b_{n+1}+\lambda=\frac{2}{3}(b_n+\lambda)$, 解得 $\lambda=-1$. 数列 $\{b_n-1\}$ 是首项为 $b_1-1=\frac{1}{3}-1=-\frac{2}{3}$, 公比为 $\frac{2}{3}$ 的等比数列.

$$b_n-1=\left(-\frac{2}{3}\right)\cdot\left(\frac{2}{3}\right)^{n-1}=-\left(\frac{2}{3}\right)^n$$

$$b_n=1-\left(\frac{2}{3}\right)^n$$

知乎@余命数

还原 a_n:

$$a_n=3^n\cdot b_n=3^n\left[1-\left(\frac{2}{3}\right)^n\right]=3^n-2^n$$

□

2.10.5 非线性递推的变换技巧

知乎@余命数

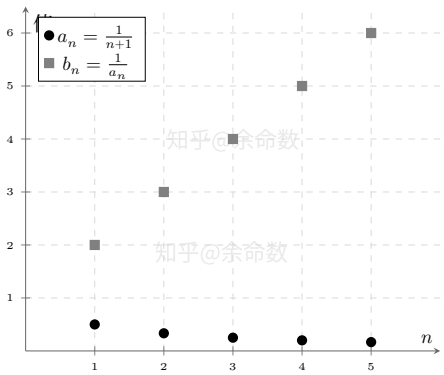
通过代数变换, 可将部分非线性递推转化为线性递推.

2.10.5.1 倒数变换

针对分式线性形式 $a_{n+1}=\frac{Aa_n}{Ba_n+C}$. 若 $C\neq 0$, 取倒数得

$$\frac{1}{a_{n+1}}=\frac{Ba_n+C}{Aa_n}=\frac{C}{A}\cdot\frac{1}{a_n}+\frac{B}{A}$$

令 $b_n=\frac{1}{a_n}$, 即转化为 $b_{n+1}=pb_n+q$ 形式.



倒数变换将双曲线型衰减序列转化为线性增长序列.

2.10.5.2 对数变换

针对幂指形式 $a_{n+1}=Ca_n^k$ $(a_n>0)$. 两边取对数, 得

$$\ln a_{n+1}=k\ln a_n+\ln C$$

令 $b_n=\ln a_n$, 即转化为线性递推.

2.10.6 二阶常系数线性递推

形如 $a_{n+2}=pa_{n+1}+qa_n$ 的递推关系, 其通项结构由特征方程 $x^2-px-q=0$ 的根决定.

定理 2.9 设 $q\neq 0$, 特征方程 $x^2-px-q=0$ 的两根为 λ_1,λ_2.

1. 若 $\lambda_1\neq\lambda_2$, 则 $a_n=c_1\lambda_1^n+c_2\lambda_2^n$.
2. 若 $\lambda_1=\lambda_2=\lambda$, 则 $a_n=(c_1+c_2n)\lambda^n$.

其中 c_1,c_2 由初始值 a_1,a_2 确定.

证明 若 x 是特征方程的根, 则序列 $u_n=x^n$ 满足

$$u_{n+2}=x^{n+2}=px^{n+1}+qx^n=pu_{n+1}+qu_n.$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

当 $\lambda_1 \neq \lambda_2$ 时, λ_1^n 与 λ_2^n 都是递推方程的解, 因而其线性组合 $c_1\lambda_1^n + c_2\lambda_2^n$ 也是解. 因 $q \neq 0$, 两根均非零, 由初值方程

$$c_1\lambda_1 + c_2\lambda_2 = a_1, \qquad c_1\lambda_1^2 + c_2\lambda_2^2 = a_2$$

的系数行列式为 $\lambda_1\lambda_2(\lambda_2 - \lambda_1) \neq 0$, 可唯一确定 c_1, c_2.

当 $\lambda_1 = \lambda_2 = \lambda$ 时, 有 $p = 2\lambda, q = -\lambda^2$, 且 $\lambda \neq 0$. 除 λ^n 外, 序列 $n\lambda^n$ 也满足递推式:

$$(n+2)\lambda^{n+2} = 2\lambda(n+1)\lambda^{n+1} - \lambda^2 n\lambda^n.$$

因而 $(c_1 + c_2 n)\lambda^n$ 是解. 由 a_1, a_2 得到的两元一次方程组行列式为 $\lambda^3 \neq 0$, 故 c_1, c_2 也唯一确定.

最后, 任意两个满足同一二阶递推且前两项相同的数列, 其差数列前两项为零, 由递推式可归纳推出以后各项也全为零. 因此上述形式给出了满足初值的唯一通项. □

例题 2.74 求斐波那契数列变体 $a_1 = 1, a_2 = 5$, 且 $a_{n+2} = 5a_{n+1} - 6a_n$ 的通项.

证明 (解析) 先写出特征方程.

$$x^2 - 5x + 6 = 0 \implies (x-2)(x-3) = 0$$

特征根为 $\lambda_1 = 2, \lambda_2 = 3$. 设通项为 $a_n = c_1 \cdot 2^n + c_2 \cdot 3^n$. 再代入初值.

$$\begin{cases} 2c_1 + 3c_2 = 1 \\ 4c_1 + 9c_2 = 5 \end{cases}$$

解得 $c_1 = -1, c_2 = 1$. 故 $a_n = 3^n - 2^n$. □

2.10.7 特征方程法的由来

二阶常系数线性递推关系形如 $a_{n+2} = pa_{n+1} + qa_n$. 特征方程 $x^2 - px - q = 0$ 是基于**试探解**的自然产物, 其背后的数学逻辑与微分方程求解一脉相承.

回到一阶递推 $a_{n+1} = pa_n$. 一阶递推 $a_{n+1} = pa_n$ 的通项公式为等比数列形式

$$a_n = a_1 \cdot p^{n-1} = C \cdot p^n$$

这提示我们, 指数形式 x^n 是线性递推关系最自然的解的结构. 因此, 对于二阶关系, 我们假设通项公式也具有 $a_n = x^n$ 的形式, 其中 $x \neq 0$ 为待定常数.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

将 $a_n = x^n$ 代入原递推关系式,

$$a_{n+2} - pa_{n+1} - qa_n = 0$$

得到

$$x^{n+2} - px^{n+1} - qx^n = 0$$

由于 $x \neq 0$, 等式两边可以同时除以 x^n.

$$x^2 - px - q = 0$$

这便得到了该递推关系的**特征方程**.

这个方程表明, 只有当 x 取特征方程的根 (λ_1, λ_2) 时, 函数 x^n 才是原递推关系的解.

- **基本解.** λ_1^n 和 λ_2^n 都是满足条件的特解.
- **线性叠加.** 由于递推关系是线性的(不含 a_n^2 或 $a_n a_{n+1}$ 等项), 特解的线性组合依然是解.

由此构造出通解, 得

$$a_n = c_1\lambda_1^n + c_2\lambda_2^n$$

特征方程这一名称借用于线性代数, 见 第五卷第 22 章. 将递推关系改写为矩阵形式, 结构会更清楚.

定义向量 $\mathbf{v}_n = \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix}$. 递推关系可以表示为线性变换.

$$\begin{pmatrix} a_{n+2} \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} p & q \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix}$$

即 $\mathbf{v}_{n+1} = A\mathbf{v}_n$. 由此推导 $\mathbf{v}_n = A^n\mathbf{v}_0$.

求解 A^n 的要点在于求矩阵 A 的特征值. 计算 A 的特征多项式 $|\lambda I - A| = 0$:

$$\det \begin{pmatrix} \lambda - p & -q \\ -1 & \lambda \end{pmatrix} = (\lambda - p)\lambda - (-q)(-1) = \lambda^2 - p\lambda - q = 0$$

这与代数法推导出的方程完全一致. 数列的通项公式, 实质上就是矩阵 A 在特征向量基下的幂运算结果.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.10.8 隔项与周期

部分数列的规律需依据项数的奇偶性或周期性分别讨论.

2.10.8.1 奇偶项讨论

若递推关系跨度为 2(如 a_{n+2} 与 a_n 相关), 或系数含 $(-1)^n$, 通常将 $\{a_{2k}\}$ 与 $\{a_{2k-1}\}$ 视为两个子数列分别求解.

2.10.8.2 周期性

若计算发现 $a_{n+k}=a_n$ 对某定值 k 成立, 则数列为周期数列. 利用归纳法或模运算确定通项. 常见于分式迭代, 如 $a_{n+1}=1-\frac{1}{a_n}$(周期为 3).

例题 2.75 已知 $a_1=2$, 且 $a_{n+1}=-a_n+2$. 求 a_n.

证明 (解析) 计算前几项, 得 $a_1=2$. $a_2=-2+2=0$. $a_3=-0+2=2$. 观察知数列呈 $2,0,2,0,\ldots$ 交替.

下面分奇偶讨论.

$$a_n=\begin{cases}2 & n\text{ 为奇数}\\ 0 & n\text{ 为偶数}\end{cases}$$

也可合写为 $a_n=1+(-1)^{n+1}$. □

2.11 迭代与不动点

在前文中, 我们讨论了递推公式, 即由前一项或几项计算后一项的规则. 若把这一规则看成一个函数 f, 则数列的生成过程就是该函数的反复作用. 我们称这种重复应用函数的过程为**迭代**.

除了生成数列, 我们还用迭代求解方程和优化模型. 从古巴比伦人计算平方根到现代计算机训练神经网络, 都建立在迭代之上.

注 2.9
压缩映射、导数判别与牛顿迭代依赖极限与导数, 见 第四卷第 16 章.

2.11.1 不动点与方程求解

我们研究的不动点, 是既在几何上有直观意义, 又能把方程求解与迭代过程联系起来的一个对象.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定义 2.14 (不动点) 设 $f : D \to D$ 是一个函数. 如果存在 $x^* \in D$ 使得

$$f(x^*) = x^*$$

则称 x^* 为函数 f 的**不动点**.

从几何上看, 函数 $y = f(x)$ 的图像与直线 $y = x$ 的交点的横坐标, 就是该函数的不动点.

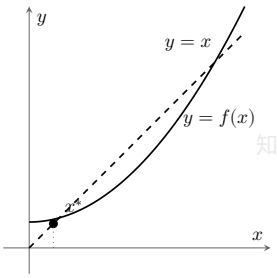


图 2.2 不动点的几何意义是 $y = f(x)$ 与 $y = x$ 的交点 (x^*, x^*),其横坐标 x^* 满足 $f(x^*) = x^*$.

求解代数方程 $g(x) = 0$ 是数学中古老的问题. 对于五次及以上的多项式方程或超越方程, 如 $x = \cos x$, 通常不存在代数求根公式. 此时可以把求根问题改写为迭代问题.

考虑方程 $g(x) = 0$. 可经代数变形改写为等价形式 $x = f(x)$. 例如

- 简单移项给出 $x = x + \lambda g(x)$, 其中 $\lambda \neq 0$.
- 牛顿法构造给出 $x = x - \frac{g(x)}{g'(x)}$.

此时, 方程 $g(x) = 0$ 的**根**, 就是辅助函数 $f(x)$ 的**不动点**. 一旦将方程转化为 $x = f(x)$ 的形式, 我们便可以构造迭代序列, 得

$$x_{n+1} = f(x_n), \quad n = 0, 1, 2, \ldots \tag{2-9}$$

给定初始猜测 x_0, 得序列 $x_0, x_1, x_2, \ldots$. 若该序列收敛于 x^*, 且 f 是连续函数, 则

$$x^* = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f(x^*)$$

即极限值 x^* 必为 f 的不动点, 也就是原方程 $g(x) = 0$ 的根.

2.11.2 不动点法求通项的原理

不动点法的工作方式是坐标变换. 它把原点平移到不动点, 或者以不动点为基准建立参照系, 从而把复杂的非齐次或非线性递推关系化为简单的等比或等差数列.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.11.2.1 线性递推中的平移变换

考虑一阶线性递推关系

知乎@余命数

$$a_{n+1} = pa_n + q \quad (p \neq 1, q \neq 0)$$

这是一个非齐次线性方程, 含有常数项 q. 需消去常数项 q, 将其转化为齐次形式, 即等比数列.

设该迭代函数 $f(x) = px + q$ 的不动点为 c. 则 c 满足

$$c = pc + q$$

将原递推式与不动点方程相减, 得

知乎@余命数

$$\begin{aligned} a_{n+1} &= pa_n + q \\ c &= pc + q \\ \Longrightarrow\ a_{n+1} - c &= p(a_n - c) \end{aligned}$$

令 $b_n = a_n - c$. 上式即为 $b_{n+1} = pb_n$. 故, 虽然数列 $\{a_n\}$ 本身不是等比数列, 但它**相对于不动点 c 的距离** $\{a_n - c\}$ 构成了一个公比为 p 的等比数列.

原理. 找不动点就是找一个参照位置. 通过变换 $b_n = a_n - c$(坐标平移), 非齐次的仿射变换 $x \mapsto px + q$ 化为齐次的线性变换 $y \mapsto py$.

知乎@余命数 知乎@余命数

2.11.2.2 分式线性递推中的比值变换

考虑分式线性递推关系(莫比乌斯变换)

$$a_{n+1} = \frac{pa_n + q}{ra_n + s} \quad (r \neq 0, ps - qr \neq 0)$$

设其迭代函数 $f(x) = \frac{px+q}{rx+s}$. 若直接迭代, 通项公式很复杂. 不动点法通过构造交比把它线性化.

2.11.2.3 有两个相异不动点 α, β

不动点满足 $f(\alpha) = \alpha, f(\beta) = \beta$. 我们考察 a_{n+1} 与不动点的距离之比.

$$\begin{aligned} a_{n+1} - \alpha &= \frac{pa_n + q}{ra_n + s} - \frac{p\alpha + q}{r\alpha + s} \\ &= \frac{(pa_n + q)(r\alpha + s) - (p\alpha + q)(ra_n + s)}{(ra_n + s)(r\alpha + s)} \\ &= \frac{(ps - qr)(a_n - \alpha)}{(ra_n + s)(r\alpha + s)} \end{aligned}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

同理可得

知乎@余命数

$$a_{n+1}-\beta=\frac{(ps-qr)(a_n-\beta)}{(ra_n+s)(r\beta+s)}$$

两式相除, 消去复杂的公因子, 得

$$\frac{a_{n+1}-\alpha}{a_{n+1}-\beta}=\left(\frac{r\beta+s}{r\alpha+s}\right)\cdot\frac{a_n-\alpha}{a_n-\beta}$$

令 $b_n=\frac{a_n-\alpha}{a_n-\beta}$. 上式即为 $b_{n+1}=kb_n$, 其中 $k=\frac{r\beta+s}{r\alpha+s}$ 为常数.

原理. 这种变换是射影几何中的坐标变换. 把两个不动点 α,β 映射为 0 和 ∞, 非线性迭代就化为简单的乘法迭代(等比数列).

2.11.2.4 有唯一重根不动点 α

知乎@余命数

此时 $\alpha=\beta$, 上述比值变换失效 (分母为零) . 此时不动点 α 满足 $r\alpha^2+(s-p)\alpha-q=0$, 且判别式为零, 即 $\alpha=\frac{p-s}{2r}$. 我们考察倒数变换

$$\frac{1}{a_{n+1}-\alpha}=\frac{ra_n+s}{(pa_n+q)-\alpha(ra_n+s)}$$

利用 α 的代数性质化简分母, 最终可得

$$\frac{1}{a_{n+1}-\alpha}=\frac{1}{a_n-\alpha}+\frac{r}{p-r\alpha}$$

知乎@余命数

知乎@余命数

令 $b_n=\frac{1}{a_n-\alpha}$. 上式即为 $b_{n+1}=b_n+d$.

原理. 当两个不动点重合时, 几何上的旋转/缩放退化为平移. 倒数变换把原数列化为等差数列. 这对应矩阵的若尔当标准型理论. 特征值相等时无法对角化, 只能化为若尔当块, 对应线性增长(等差数列).

2.11.2.5 线性化的一般框架

利用不动点求通项公式, 是动力系统线性化的一个例子.

1. **寻找平衡位置.** 解方程 $f(x)=x$ 找到系统的不动点.
2. **建立新坐标系.** 以不动点为参照, 定义新数列 b_n, 如 a_n-c 或 $\frac{a_n-\alpha}{a_n-\beta}$.
3. **线性化迭代.** 在新坐标系下, 复杂的迭代关系退化为最基本的线性关系 ($b_{n+1}=kb_n$ 或 $b_{n+1}=b_n+d$).
4. **还原.** 利用等比/等差数列公式求出 b_n, 再逆变换回 a_n.

压缩映射原理

知乎@余命数

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.11.3 压缩映射原理

有些迭代形式 $x_{n+1}=f(x_n)$ 会发散. 迭代是否收敛取决于函数 f 是否压缩空间. 波兰数学家巴拿赫 (Stefan Banach) 将其形式化为**压缩映射原理**(巴拿赫不动点定理), 是数值分析中迭代法收敛性的基础.

定义 2.15 (压缩映射) 设 f 定义在区间 $[a,b]$ 上, 且其值域仍在 $[a,b]$ 内. 若存在常数 $0\le k<1$, 使得对于区间内任意 x,y, 都有

$$|f(x)-f(y)|\le k|x-y|$$

则称 f 为该区间上的一个**压缩映射**, 常数 k 称为压缩系数.

直观上, 压缩映射使得任意两点在经过函数作用后, 其距离变得更近.

定理 2.10 (压缩映射原理) 若 f 是闭区间 $[a,b]$ 上的压缩映射, 则

1. f 在 $[a,b]$ 内存在且仅存在一个不动点 x^*.
2. 对于任意初始值 $x_0\in[a,b]$, 迭代序列 $x_{n+1}=f(x_n)$ 均收敛于 x^*.
3. 误差满足估计式 $|x_n-x^*|\le\frac{k^n}{1-k}|x_1-x_0|$.

证明 任取 $x_0\in[a,b]$, 定义迭代 $x_{n+1}=f(x_n)$. 因为 $f([a,b])\subseteq[a,b]$, 所以所有 x_n 都属于 $[a,b]$.

由压缩条件,

$$|x_{n+1}-x_n|=|f(x_n)-f(x_{n-1})|\le k|x_n-x_{n-1}|\le k^n|x_1-x_0|.$$

因而当 $m>n$ 时,

$$|x_m-x_n|\le\sum_{j=n}^{m-1}|x_{j+1}-x_j|\le\sum_{j=n}^{m-1}k^j|x_1-x_0|\le\frac{k^n}{1-k}|x_1-x_0|.$$

这说明 $\{x_n\}$ 是柯西数列. 闭区间 $[a,b]$ 完备, 故存在 $x^*\in[a,b]$ 使 $x_n\to x^*$.

压缩映射满足 Lipschitz 条件, 因而连续. 对 $x_{n+1}=f(x_n)$ 两边取极限, 得 $x^*=f(x^*)$, 所以不动点存在.

若 u,v 都是不动点, 则

$$|u-v|=|f(u)-f(v)|\le k|u-v|.$$

由于 $0\le k<1$, 只能有 $|u-v|=0$, 即 $u=v$. 不动点唯一.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

在上面的估计中令 $m \to \infty$, 即得

$$|x_n - x^*| \leq \frac{k^n}{1-k}|x_1 - x_0|.$$

因此任意初始值产生的迭代序列都收敛到唯一不动点, 且满足所给误差估计. □

若函数 f 可导, 根据拉格朗日中值定理, $|f(x) - f(y)| = |f'(\xi)||x - y|$. 因此, f 在区间内恒有 $|f'(x)| \leq k < 1$ 是它成为压缩映射的一个充分条件.

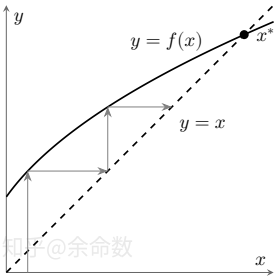


图 2.3 迭代过程的几何表示(蛛网图). 数列 $\{x_n\}$ 在 $y = x$ 与 $y = f(x)$ 之间折返, 逼近交点, 即不动点.

反之, 若 $f(x)$ 的图像与 $y = x$ 不相交, 则不动点不存在, 迭代序列无从收敛. 取 $f(x) = x + \frac{1}{2}$, 其图像是 $y = x$ 的一条平行线, 二者永不相交; 迭代 $x_{n+1} = x_n + \frac{1}{2}$ 的蛛网呈阶梯状持续上移, 远离任何可能的极限.

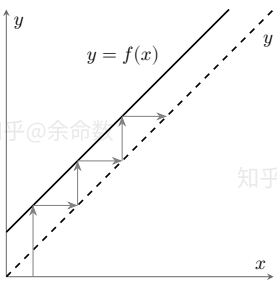
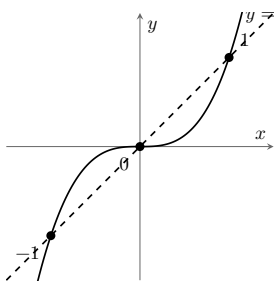


图 2.4 不动点不存在的情形, 如 $f(x) = x + \frac{1}{2}$ 与 $y = x$ 平行, 无交点, 迭代序列发散.

不动点还可以不唯一. 当 $f(x)$ 的图像与 $y = x$ 多次相交时, 每个交点都是不动点. 取 $f(x) = x^3$, 它与 $y = x$ 在 $x = -1, 0, 1$ 三处相交, 故有三个不动点. 此时迭代序列收敛到哪一个, 取决于初始值 x_0 的选取, 从 $x_0 \in (0, 1)$ 出发迭代趋于 0, 从 $x_0 > 1$ 出发迭代趋于 $+\infty$, 不动点 $x^* = 1$ 虽存在却非吸引点.



知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

图 2.5　不动点不唯一的情形, 如 $f(x)=x^3$ 与 $y=x$ 有三个交点 $x^*=-1,0,1$, 迭代收敛与否取决于初始值.

即便不动点存在且迭代收敛, 收敛速度也未必快. 收敛的快慢取决于 f 在不动点处的斜率.$|f'(x^*)|$ 越接近 1, 每步缩减的误差越小, 所需迭代步数越多. 取 $f(x)=-0.9x+1$, 不动点 $x^*=\frac{10}{19}\approx 0.526$, 此处 $|f'(x^*)|=0.9$. 从 $x_0=0$ 出发, 蛛网在 $y=x$ 与 $y=f(x)$ 之间反复折返, 锯齿幅度每步仅缩减一成, 数十步后方才逼近交点.

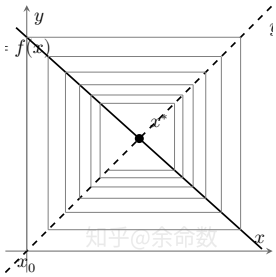


图 2.6　收敛缓慢的情形, 如 $f(x)=-0.9x+1$, $|f'(x^*)|=0.9$ 接近 1, 蛛网锯齿幅度每步仅缩减一成, 需数十步方能逼近不动点.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

2.11.4 巴比伦算法

在计算器出现之前, 古巴比伦人就用一种基于迭代的算法计算 $\sqrt{S}$ 的值.

考虑方程 $x^2=S$ $(S>0)$. 将其改写为 $x=\frac{S}{x}$. 若 x_n 小于 $\sqrt{S}$, 则 $\frac{S}{x_n}$ 必然大于 $\sqrt{S}$; 反之亦然. $\sqrt{S}$ 位于 x_n 与 $\frac{S}{x_n}$ 之间. 取二者的算术平均值作为下一次的近似, 应当比 x_n 更接近真实值.

由此得到**巴比伦迭代公式**.

$$x_{n+1}=\frac{1}{2}\left(x_n+\frac{S}{x_n}\right)$$

知乎@余命数

巴比伦算法收敛的导数讨论

例题 2.76　利用巴比伦算法计算 $\sqrt{2}$ 的近似值, 取初始值 $x_0=1$.

解　代入公式 $S=2$:

$$x_1=\frac{1}{2}\left(1+\frac{2}{1}\right)=1.5$$

$$x_2=\frac{1}{2}\left(1.5+\frac{2}{1.5}\right)=\frac{1}{2}\left(\frac{3}{2}+\frac{4}{3}\right)=\frac{17}{12}\approx 1.41666\ldots$$

$$x_3=\frac{1}{2}\left(\frac{17}{12}+\frac{2}{17/12}\right)=\frac{1}{2}\left(\frac{17}{12}+\frac{24}{17}\right)=\frac{577}{408}\approx 1.4142156\ldots$$

知乎@余命数

已知 $\sqrt{2}\approx 1.41421356$. 仅迭代三次, 误差已小于 10^{-5}.

考察该迭代的收敛性. 设函数 $f(x)=\frac{1}{2}(x+\frac{S}{x})$. 不动点方程为 $x=\frac{1}{2}(x+\frac{S}{x}) \implies 2x=x+\frac{S}{x} \implies x=\frac{S}{x} \implies x^2=S$. 不动点确为 $\sqrt{S}$.

计算导数, 得

$$f'(x)=\frac{1}{2}\left(1-\frac{S}{x^2}\right)$$

在不动点 $x^*=\sqrt{S}$ 处,

$$f'(\sqrt{S})=\frac{1}{2}\left(1-\frac{S}{(\sqrt{S})^2}\right)=0$$

由于 $|f'(\sqrt{S})|=0<1$, 根据收敛原理, 该迭代在 $\sqrt{S}$ 附近收敛很快(二阶收敛).

牛顿迭代法

知乎@余命数

2.11.5 牛顿迭代法

函数一章已经介绍过这个迭代. 我们指出, 巴比伦算法是**牛顿迭代法**的一个特例.

计算机求解方程 $f(x)=0$ 时通常使用迭代法.

牛顿法的几何直观是用切线逼近曲线. 在点 $(x_n,f(x_n))$ 处作曲线 $y=f(x)$ 的切线. 切线方程为

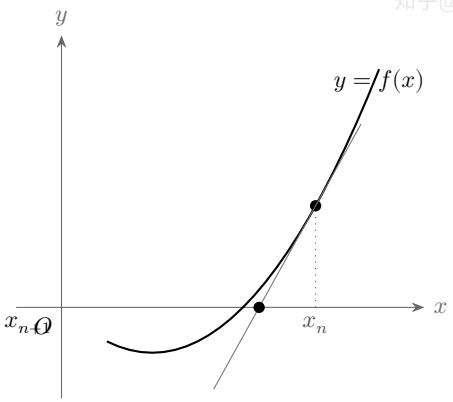
$$y-f(x_n)=f'(x_n)(x-x_n)$$

令 $y=0$, 求出切线与 x 轴的交点, 作为下一次的近似值 x_{n+1}:

$$-f(x_n)=f'(x_n)(x_{n+1}-x_n)$$

整理得牛顿迭代公式.

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)} \tag{2-10}$$



知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

图 2.7　牛顿迭代法的几何意义(利用切线的根作为曲线根的近似).

例题 2.77　利用牛顿法推导求 k 次方根 $\sqrt[k]{S}$ 的迭代公式.

解　求 $\sqrt[k]{S}$ 等价于求解方程 $x^k - S = 0$. 设 $f(x) = x^k - S$, 则 $f'(x) = kx^{k-1}$. 代入牛顿迭代公式,

$$\begin{aligned} x_{n+1} &= x_n - \frac{x_n^k - S}{kx_n^{k-1}} \\ &= x_n - \frac{1}{k}\left(\frac{x_n^k}{x_n^{k-1}} - \frac{S}{x_n^{k-1}}\right) \\ &= x_n - \frac{1}{k}x_n + \frac{1}{k}\frac{S}{x_n^{k-1}} \\ &= \frac{k-1}{k}x_n + \frac{1}{k}\frac{S}{x_n^{k-1}} \end{aligned}$$

当 $k = 2$ 时, 公式变为 $x_{n+1} = \frac{1}{2}x_n + \frac{1}{2}\frac{S}{x_n}$, 这正是巴比伦算法.

2.12 数列的单调性

观察一个数列, 除了关心各项的数值, 我们还要看它的整体趋势. 随着 n 增大, 数列有的持续增大, 有的持续减小, 有的上下波动, 描述这种增减趋势的数学语言就是**单调性**.

这个概念与函数的单调性定义相同, 都刻画取值随变量的增减变化. 函数的单调性描述因变量如何随自变量连续变化, 数列的单调性则描述数列的项如何随离散的整数变量 n 变化. 判断并证明数列的单调性, 是分析其收敛性、估计其界限、解决相关不等式问题的前提.

注 2.10
用连续函数导数判别数列单调性属分析工具, 见 第四卷 第 16 章.

2.12.1 定义

定义 2.16 (单调数列) 设有一个数列 $\{a_n\}$.

- 如果对于任意正整数 n, 恒有 $a_{n+1} > a_n$, 那么称数列 $\{a_n\}$ 为**递增数列**.
- 如果对于任意正整数 n, 恒有 $a_{n+1} < a_n$, 那么称数列 $\{a_n\}$ 为**递减数列**.
- 如果对于任意正整数 n, 恒有 $a_{n+1} \geq a_n$, 那么称数列 $\{a_n\}$ 为**非减数列**.
- 如果对于任意正整数 n, 恒有 $a_{n+1} \leq a_n$, 那么称数列 $\{a_n\}$ 为**非增数列**.

递增数列和递减数列统称为**严格单调数列**. 非减数列和非增数列统称为**单调数列**.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.12.2 判断方法

判断一个数列的单调性,实质上是比较相邻两项 a_{n+1} 与 a_n 的大小关系. 根据通项公式的结构特征,常用以下几种方法.

2.12.2.1 作差法

作差法直接套用定义. 比较相邻两项的差 $d_n = a_{n+1} - a_n$,判断差值的符号.

- 若对于一切 $n \in \mathbb{N}^*$,恒有 $a_{n+1} - a_n > 0$,则 $\{a_n\}$ 是递增数列.
- 若对于一切 $n \in \mathbb{N}^*$,恒有 $a_{n+1} - a_n < 0$,则 $\{a_n\}$ 是递减数列.

知乎@余命数

对于非严格单调的情况,只需将不等号换为 $\geq$ 或 $\leq$.

知乎@余命数 知乎@余命数

例题 2.78 判断数列 $\left\{\dfrac{n}{2n+1}\right\}$ 的单调性.

解 解析: 设 $a_n = \dfrac{n}{2n+1}$. 考察相邻两项之差.

$$\begin{aligned} a_{n+1} - a_n &= \frac{n+1}{2(n+1)+1} - \frac{n}{2n+1} \\ &= \frac{n+1}{2n+3} - \frac{n}{2n+1} \\ &= \frac{(n+1)(2n+1) - n(2n+3)}{(2n+3)(2n+1)} \\ &= \frac{(2n^2+3n+1) - (2n^2+3n)}{(2n+3)(2n+1)} \\ &= \frac{1}{(2n+3)(2n+1)} \end{aligned}$$

知乎@余命数 知乎@余命数

由于对于任意正整数 n,分母 $(2n+3)(2n+1)$ 恒为正数. 因此,$a_{n+1} - a_n = \dfrac{1}{(2n+3)(2n+1)} > 0$ 恒成立. 故数列 $\left\{\dfrac{n}{2n+1}\right\}$ 是一个递增数列.

2.12.2.2 作商法

知乎@余命数

当数列的各项均为正数时,比较 a_{n+1} 与 a_n 的大小等价于比较它们的比值 $\dfrac{a_{n+1}}{a_n}$ 与 1 的大小. 这种方法尤其适用于通项中含有指数、阶乘或连乘积形式的数列. 设数列 $\{a_n\}$ 满足 $a_n > 0$.

- 若对于一切 $n \in \mathbb{N}^*$,恒有 $\dfrac{a_{n+1}}{a_n} > 1$,则 $\{a_n\}$ 是递增数列.
- 若对于一切 $n \in \mathbb{N}^*$,恒有 $\dfrac{a_{n+1}}{a_n} < 1$,则 $\{a_n\}$ 是递减数列.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 2.79　判断数列 $\left\{\dfrac{3^n}{n!}\right\}$ 的单调性.

解　解析：设 $a_n=\dfrac{3^n}{n!}$. 显然对于所有 $n\in\mathbb{N}^*$, $a_n>0$, 故可以使用作商法. 考察相邻两项之比.

$$\begin{aligned}\frac{a_{n+1}}{a_n}&=\frac{3^{n+1}/(n+1)!}{3^n/n!}\\&=\frac{3^{n+1}}{(n+1)!}\cdot\frac{n!}{3^n}\\&=\frac{3^{n+1}}{3^n}\cdot\frac{n!}{(n+1)n!}\\&=3\cdot\frac{1}{n+1}\\&=\frac{3}{n+1}\end{aligned}$$

需比较这个比值与 1 的大小关系. 当 $n=1$ 时, $\dfrac{a_2}{a_1}=\dfrac{3}{2}>1$. 当 $n=2$ 时, $\dfrac{a_3}{a_2}=\dfrac{3}{3}=1$. 当 $n\geq 3$ 时, $n+1\geq 4$, 此时 $\dfrac{3}{n+1}<1$.

因此, $a_2>a_1$ 且 $a_3=a_2$, 而对于 $n\geq 3$, 有 $a_{n+1}<a_n$. 该数列不是一个单调数列. 它先增后不变, 然后开始递减. 具体地, $a_1=3, a_2=4.5, a_3=4.5, a_4=3.375, \ldots$.

2.12.2.3　函数法

如果数列的通项 a_n 可以看作是某个函数 $f(x)$ 在正整数点上的取值, 即 $a_n=f(n)$, 那么函数 $f(x)$ 的单调性便可以为我们提供关于数列 $\{a_n\}$ 单调性的相关信息. 若函数 $f(x)$ 在区间 $[1,+\infty)$ 上是单调递增, 或递减的, 那么数列 $\{a_n\}$ 也必然是单调递增, 或递减的.

函数法/导数判别数列单调

在实践中, 我们通常利用导数来判断函数 $f(x)$ 的单调性.

注 2.11 (注意)
我们只关心函数在整数点上的行为. 因此, 即使函数 $f(x)$ 不是在整个 $[1,+\infty)$ 区间上单调, 只要它在某个区间 $[N,+\infty)$(N 为正整数)上单调, 我们就可以断定数列 $\{a_n\}$ 从第 N 项开始是单调的.

例题 2.80　判断数列 $\left\{\dfrac{\ln n}{n}\right\}$ 的单调性.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 解析: 设 $a_n = \dfrac{\ln n}{n}$. 考虑构造函数 $f(x) = \dfrac{\ln x}{x}$ for $x \geq 1$. 利用导数来研究其单调性.

$$f'(x) = \frac{(\ln x)' \cdot x - \ln x \cdot (x)'}{x^2} = \frac{\frac{1}{x} \cdot x - \ln x}{x^2} = \frac{1 - \ln x}{x^2}$$

令 $f'(x) = 0$, 得 $1 - \ln x = 0$, 解得 $x = e$. 当 $1 \leq x < e$ 时, $\ln x < 1$, 故 $1 - \ln x > 0$, $f'(x) > 0$. 当 $x > e$ 时, $\ln x > 1$, 故 $1 - \ln x < 0$, $f'(x) < 0$.

故函数 $f(x)$ 在 $(0, e)$ 上单调递增, 在 $(e, +\infty)$ 上单调递减.

由于 $e \approx 2.718$, 需考察整数点 $n = 1, 2, 3$ 附近的行为. $a_1 = \dfrac{\ln 1}{1} = 0$. $a_2 = \dfrac{\ln 2}{2} \approx \dfrac{0.693}{2} = 0.3465$. $a_3 = \dfrac{\ln 3}{3} \approx \dfrac{1.098}{3} = 0.366$. $a_4 = \dfrac{\ln 4}{4} = \dfrac{2\ln 2}{4} = \dfrac{\ln 2}{2} \approx 0.3465$.

有 $a_1 < a_2 < a_3$, 而 $a_3 > a_4$. 由于函数 $f(x)$ 在 $x > e$ 后单调递减, 那么对于所有满足 $n \geq 3$ 的整数, 必然有 $a_{n+1} = f(n+1) < f(n) = a_n$. 因此, 该数列从第 3 项开始是严格递减的, 但它不是一个整体上的单调数列.

注 2.12 (单调性判断方法总结)

- **作差法:** 通用性最强, 是定义的直接体现. 适用于大多数多项式或分式形式的通项.
- **作商法:** 适用于各项为正, 且通项包含指数、阶乘等乘法结构的情况.
- **函数法:** 通项可延拓为连续函数时可用; **求导判别** 见 第四卷第 16 章.

选择哪种方法, 取决于通项 a_n 的结构.

2.13 奇偶数列 *

整数的奇偶性是其最基本的数论属性. 当将研究对象从单个整数扩展到按规律排列的整数序列时, 一个自然的问题是, 数列中各项的奇偶性是否也遵循某种规律.

研究数列的奇偶性, 相当于在模 2 的算术体系下看数列. 这种视角转换常能把复杂的递推关系化简为简单、甚至周期性的新关系. 奇偶性分析除了用于判断项的基本性质, 还用于解决数论问题, 例如证明某些项不可能是完全平方数. 数列在二元结构中常表现出周期性.

2.13.1 定义与判断

对于一个各项均为整数的数列 $\{a_n\}$, 可根据其项的奇偶性对其进行分类, 并引出其伴随的奇偶性序列.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定义 2.17 (奇数列与偶数列) 设 $\{a_n\}$ 是一个整数数列.

- 如果对于任意正整数 n, 数列的项 a_n 均为奇数, 则称 $\{a_n\}$ 为**奇数列**.
- 如果对于任意正整数 n, 数列的项 a_n 均为偶数, 则称 $\{a_n\}$ 为**偶数列**.

知乎@余命数

判断一个数列的奇偶性, 主要有两种途径. 一种是从通项公式入手, 另一种是从递推关系入手.

2.13.1.1 基于通项公式的分析

如果数列的通项公式 $a_n = f(n)$ 已知, 可直接分析函数值 $f(n)$ 在 n 取遍正整数时的奇偶性. 这通常需要讨论 n 的奇偶性.

知乎@余命数

例题 2.81 判断数列 $a_n = n^2 + 3n + 5$ 的奇偶性.

解 解析: 将通项进行因式分解和重组, 以说明其结构:

$$a_n = n(n+3) + 5$$

对项数 n 的奇偶性讨论.

- **若 n 为偶数**, 设 $n = 2k$ ($k \in \mathbb{N}^*$). 则 n 是偶数, 故 $n(n+3)$ 是偶数. $a_n =$ (偶数) $+ 5 =$ (偶数) $+$ (奇数) $=$ (奇数).
- **若 n 为奇数**, 设 $n = 2k - 1$ ($k \in \mathbb{N}^*$). 则 $n + 3$ 是偶数, 故 $n(n+3)$ 是偶数. $a_n =$ (偶数) $+ 5 =$ (偶数) $+$ (奇数) $=$ (奇数).

知乎@余命数　知乎@余命数

综合两种情况, 对于所有正整数 n, a_n 均为奇数. 因此, $\{a_n\}$ 是一个奇数列.

2.13.1.2 基于递推关系的分析

当数列由递推关系定义时, 直接读出每一项的性质并不容易. 但可以分析奇偶性如何通过递推关系**遗传**下去. 这需要模 2 的同余算术. 基本运算法则如下:

- 奇 $\pm$ 奇 = 偶, $1 + 1 \equiv 0 \pmod 2$
- 奇 $\pm$ 偶 = 奇, $1 + 0 \equiv 1 \pmod 2$
- 偶 $\pm$ 偶 = 偶, $0 + 0 \equiv 0 \pmod 2$
- 奇 $\times$ 奇 = 奇, $1 \times 1 \equiv 1 \pmod 2$
- 奇 $\times$ 偶 = 偶, $1 \times 0 \equiv 0 \pmod 2$
- 偶 $\times$ 偶 = 偶, $0 \times 0 \equiv 0 \pmod 2$

知乎@余命数

通过将原递推关系中的每一项都替换为其模 2 的余数(奇数 $\to 1$, 偶数 $\to 0$), 可得到一个关于奇偶性的新递推关系.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 2.82 分析斐波那契数列 $\{F_n\}$ $(F_1=1, F_2=1)$ 的各项奇偶性规律.

解 解析: 斐波那契数列的递推关系为 $F_{n+2}=F_{n+1}+F_n$. 考察其各项模 2 的余数, 设 $b_n \equiv F_n \pmod 2$. 则 $\{b_n\}$ 满足递推关系 $b_{n+2} \equiv b_{n+1}+b_n \pmod 2$.

从初始项开始计算这个新的奇偶性序列, 得 $F_1=1 \implies b_1=1$, 为奇 $F_2=1 \implies b_2=1$, 为奇 $b_3 \equiv b_2+b_1=1+1 \equiv 0$, 为偶 $b_4 \equiv b_3+b_2=0+1 \equiv 1$, 为奇 $b_5 \equiv b_4+b_3=1+0 \equiv 1$, 为奇 $b_6 \equiv b_5+b_4=1+1 \equiv 0$, 为偶

序列 $\{b_n\}$ 的前几项为 $1,1,0,1,1,0,\ldots$. 观察到, 序列以奇、奇、偶的模式循环出现, 其周期为 3. 因此, 斐波那契数列中 F_n 是偶数, 当且仅当项数 n 是 3 的倍数.

例题 2.83 已知数列 $\{a_n\}$ 满足 $a_1=1, a_{n+1}=5a_n+8$. 证明 $\{a_n\}$ 是一个奇数列.

解 证明: 使用数学归纳法, 并结合奇偶性分析. 设命题 $P(n)$ 为"a_n 是一个奇数".

奠基步骤: 当 $n=1$ 时, $a_1=1$, 是一个奇数. 故 $P(1)$ 成立.

归纳步骤: 假设当 $n=k$ $(k \geq 1)$ 时命题成立, 即 a_k 是一个奇数. 需证明当 $n=k+1$ 时, a_{k+1} 也是一个奇数.

考察递推关系 $a_{k+1}=5a_k+8$. 根据归纳假设, a_k 是奇数. 因此:

$$\begin{aligned} a_{k+1} &= (\text{奇数}) \times a_k + (\text{偶数}) \\ &= (\text{奇数}) \times (\text{奇数}) + (\text{偶数}) \\ &= (\text{奇数}) + (\text{偶数}) \\ &= (\text{奇数}) \end{aligned}$$

这证明了 $P(k+1)$ 成立.

或者, 可使用更为形式化的同余语言. $a_{n+1} \equiv 5a_n+8 \pmod 2$. 由于 $5 \equiv 1 \pmod 2$ 且 $8 \equiv 0 \pmod 2$, 上式化为:

$$a_{n+1} \equiv 1 \cdot a_n + 0 \pmod 2 \implies a_{n+1} \equiv a_n \pmod 2$$

故后一项的奇偶性与前一项完全相同. 由于首项 $a_1=1$ 是奇数, 即 $a_1 \equiv 1 \pmod 2$. 因此, 对于所有正整数 n, 都有 $a_n \equiv a_1 \equiv 1 \pmod 2$. 即 $\{a_n\}$ 的每一项都是奇数.

根据数学归纳法原理, 命题得证.

2.14 数列的周期性

知乎@余命数

来龙去脉

知乎@余命数

数列的项在经过固定步长后开始重复取值, 这种现象称为数列的周期性.

研究周期性直接通向数列的代数结构. 当递推关系涉及模运算或有限集合上的运算时, 周期性几乎必然出现. 判断周期性的方法常用于解决数论与组合问题, 也用于处理项数过大、无法逐项计算的数列问题.

2.14.1 定义与性质

先对数列的周期性给出严格的数学定义.

知乎@余命数

定义 2.18 (周期数列) 对于数列 $\{a_n\}$, 如果存在一个正整数 T, 使得对于所有大于或等于某个正整数 n_0 的 n, 恒有等式 $a_{n+T}=a_n$ 成立, 则称数列 $\{a_n\}$ 为**周期数列**.

周期数列的定义中满足条件的正整数 T 不止一个, 需要区分周期与最小正周期.

知乎@余命数

定义 2.19 (周期与最小正周期) 在上述定义中, 满足条件的任意一个正整数 T 都被称为数列 $\{a_n\}$ 的一个**周期**.
在所有的周期中, 最小的那个正整数被称为该数列的**最小正周期** (或主周期). 没有特别说明时, 数列的周期指的就是最小正周期.

一个直接的推论是, 若 T 是数列的最小正周期, 则任何形如 kT ($k\in\mathbb{N}^*$) 的整数都是该数列的周期. 周期数列最重要的性质是, 其所有项的取值落在一个有限集合内; 长度为 T 的一段取值重复出现, 即可确定整个数列.

知乎@余命数

反过来, 任一周期都是最小正周期的整数倍.

命题 2.1 (最小正周期的整除性) 若 T 是数列 $\{a_n\}$ 的最小正周期, T' 是它的任一周期, 则 $T\mid T'$.

证明 由带余除法, $T'=qT+r$, 其中 $0\le r<T$. 因 T 是周期, qT 也是周期, 故 $a_{n+qT}=a_n$; 又 T' 是周期, $a_{n+T'}=a_n$. 于是

$$a_{n+r}=a_{n+T'-qT}=a_{n+T'}=a_n,$$

即 r 也是周期. 但 $0\le r<T$ 且 T 是最小正周期, 故 $r=0$, 即 $T\mid T'$. □

周期数列的取值被一个长度为 T 的循环块完全决定, 由此得到两条常用性质.

知乎@余命数

命题 2.2 (周期求和) 若数列 $\{a_n\}$ 以 T 为周期, 则任意连续 T 项之和为定值, 即对任意正整数 n,

$$\sum_{k=0}^{T-1} a_{n+k} = \sum_{k=1}^{T} a_k.$$

证明 记 $S_n = \sum_{k=0}^{T-1} a_{n+k}$. 把求和起点从 n 平移到 $n+1$, 得

$$S_{n+1} = \sum_{k=0}^{T-1} a_{n+1+k} = \sum_{k=1}^{T} a_{n+k} = S_n - a_n + a_{n+T}.$$

由 $a_{n+T} = a_n$ 知 $S_{n+1} = S_n$, 故 S_n 与 n 无关. □

知乎@余命数

命题 2.3 (和与积的周期) 若 $\{a_n\}$ 以 T_1 为周期, $\{b_n\}$ 以 T_2 为周期, 则 $\{a_n+b_n\}$ 与 $\{a_n b_n\}$ 均以 $L = \operatorname{lcm}(T_1, T_2)$ 为周期; 两者的最小正周期都整除 L.

知乎@余命数

证明 L 同时是 T_1 与 T_2 的倍数, 故

$$a_{n+L} + b_{n+L} = a_n + b_n, \qquad a_{n+L}\, b_{n+L} = a_n\, b_n.$$

知乎@余命数 知乎@余命数

即 L 是和数列与积数列的周期. 由 命题 2.1, 最小正周期整除任一周期, 故整除 L. □

还有一种更广义的最终周期性, 高中阶段较少涉及.

知乎@余命数

定义 2.20 (最终周期数列) 如果一个数列从某一项之后才开始呈现周期性, 即存在正整数 N 和 T, 使得对于所有 $n > N$, 恒有 $a_{n+T} = a_n$ 成立, 则称该数列为**最终周期数列**.
数列的前 N 项 $\{a_1, a_2, \dots, a_N\}$ 构成了数列的**预周期**部分.

2.14.2 周期性的判定

判定由递推关系定义的数列是否具有周期性, 关键在于考察系统状态的演化. 对于一个 k 阶递推关系 $a_{n+k} = f(a_{n+k-1}, \dots, a_n)$, 第 n 步的状态可以用 k 维向量 $\mathbf{v}_n = (a_n, a_{n+1}, \dots, a_{n+k-1})$ 完全描述. 若某一步的状态 $\mathbf{v}_m$ 与之前某一步的状态 $\mathbf{v}_l$ $(l < m)$ 完全相同, 由递推关系的确定性可知, 此后各项将重复之前的演化路径, 数列进入周期.

知乎@余命数

这一观察与同余理论结合后尤为有效. 如果一个整数数列 $\{a_n\}$ 在模 m 的意义下被考察, 那么每一项 $a_n \pmod m$ 的取值只有 m 种可能. 相应地, k 维状态向量 $\mathbf{v}_n$ 的每个分量也只有 m 种可能, 总的状态数不超过 m^k 种. 根据**鸽巢原理**, 在至多 m^k+1 次状态转移后, 必然有两个状态完全相同. 我们由此得到一个有效的结论.

定理 2.11　设整数数列 $\{a_n\}$ 由 k 阶整系数线性递推关系

$$a_{n+k} = c_{k-1}a_{n+k-1} + \cdots + c_1 a_{n+1} + c_0 a_n$$

定义. 对任意正整数 m, 其模 m 的余数数列必然是最终周期数列; 等价地, 存在正整数 l 与 T, 使得从第 l 个状态开始按周期 T 重复.

证明　在模 m 的意义下, 状态向量

$$\mathbf{v}_n = (a_n, a_{n+1}, \dots, a_{n+k-1}) \pmod m$$

的每个分量只有 m 种可能, 因此状态总数不超过 m^k 个. 递推关系的系数为整数, 所以在模 m 下, 下一状态

$$\mathbf{v}_{n+1} = (a_{n+1}, \dots, a_{n+k})$$

由当前状态 $\mathbf{v}_n$ 唯一确定.

考察 m^k+1 个状态 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m^k+1}$. 根据鸽巢原理, 存在 $1 \le l < r \le m^k+1$, 使得 $\mathbf{v}_l = \mathbf{v}_r$. 由于后续演化由状态唯一决定, 归纳可得对所有 $j \ge 0$, 都有 $\mathbf{v}_{l+j} = \mathbf{v}_{r+j}$. 因此从第 l 个状态开始, 状态序列以 $T = r-l$ 为周期, 余数数列 $\{a_n \pmod m\}$ 也从某一项开始以 T 为周期. 它是最终周期数列. □

定理保证了最终周期性的存在. 具体求周期时, 逐项计算状态 $\mathbf{v}_n$ 直至首次出现重复 $\mathbf{v}_r = \mathbf{v}_l$ $(l < r)$, 则周期为 $r-l$; 对模 m 的 k 阶递推, 这样的重复在 m^k+1 个状态内必然出现. 最小正周期通常远小于上界 m^k, 须由计算确认.

例题 2.84　已知数列 $\{a_n\}$ 满足 $a_1 = 1$, $a_{n+1} = 2a_n \pmod 7$. 求 $\{a_n\}$ 的最小正周期, 并计算 $a_1 + a_2 + \cdots + a_{2024}$.

解　解析: 这是一阶递推, 状态即 a_n 本身, 取值只有 7 种. 从 $a_1 = 1$ 起逐项计算:

$$a_1 = 1, \quad a_2 = 2, \quad a_3 = 4, \quad a_4 = 8 \equiv 1 \pmod 7.$$

$a_4 = a_1$, 此后序列重复, 故最小正周期为 3, 循环块为 $(1, 2, 4)$.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

一个周期内三项之和为 $1+2+4=7$. 由 命题 2.2, 任意连续 3 项之和都等于 7. 把 2024 项按每 3 项一组划分:

$$2024 = 3 \times 674 + 2.$$

前 2022 项恰为 674 个完整周期, 和为 $674 \times 7 = 4718$; 余下两项为 $a_1 = 1, a_2 = 2$. 故

$$a_1 + a_2 + \cdots + a_{2024} = 4718 + 1 + 2 = 4721.$$

例题 2.85 已知数列 $\{a_n\}$ 满足 $a_1 = 1, a_2 = 1$ 且 $a_{n+2} = a_{n+1} - a_n$. 分析其周期性.

解 解析: 从初始项逐项计算, 观察规律.

知乎@余命数

$$a_1 = 1$$

$$a_2 = 1$$

$$a_3 = a_2 - a_1 = 1 - 1 = 0$$

$$a_4 = a_3 - a_2 = 0 - 1 = -1$$

$$a_5 = a_4 - a_3 = -1 - 0 = -1$$

$$a_6 = a_5 - a_4 = -1 - (-1) = 0$$

$$a_7 = a_6 - a_5 = 0 - (-1) = 1$$

$$a_8 = a_7 - a_6 = 1 - 0 = 1$$

可见 $a_7 = a_1$ 且 $a_8 = a_2$. 由于这是二阶递推关系, 当连续两项重复时, 后续所有项也将随之重复.

该数列为 $1, 1, 0, -1, -1, 0, 1, 1, \ldots$. 其重复的周期块为 $(1, 1, 0, -1, -1, 0)$. 因此, 该数列是一个周期数列, 其最小正周期为 6.

注 2.13 (拓展旁注) 复单位根/复指数写法见 第五卷第 20 章.

注 2.14 (延伸思考)
该递推关系的特征方程为 $x^2 - x + 1 = 0$, 其根为 $x = \frac{1 \pm \sqrt{-3}}{2} = e^{\pm i\pi/3}$, 这是单位圆上的 6 次单位根. 这揭示了该数列周期为 6 的代数原因.

例题 2.86 设 $\{F_n\}$ 为斐波那契数列 ($F_1 = 1, F_2 = 1$). 求 F_{2024} 的个位数字.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 解析: 整数的个位数字等于它模 10 的余数. 考察斐波那契数列模 10 后的余数数列 $\{a_n\}$, 其中 $a_n \equiv F_n \pmod{10}$.

该数列满足递推关系 $a_{n+2} \equiv a_{n+1} + a_n \pmod{10}$, 初始项为 $a_1 = 1, a_2 = 1$. 状态由相邻两项构成的序对 (a_n, a_{n+1}) 决定, 这样的序对最多有 $10 \times 10 = 100$ 种. 由鸽巢原理, 该数列必在有限步后进入周期. 现通过计算确定周期.

数列 $\{a_n\}$ 的项为:

$$1,1,2,3,5,8,3,1,4,5,9,4,3,7,0,7,7,4,1,5,$$

$$6,1,7,8,5,3,8,1,9,0,9,9,8,7,5,2,7,9,6,5,$$

$$1,6,7,3,0,3,3,6,9,5,4,9,3,2,5,7,2,9,1,0,\ldots$$

当计算到第 61 项和第 62 项时, 得 $a_{61} \equiv a_{60} + a_{59} \equiv 0 + 1 = 1$ $a_{62} \equiv a_{61} + a_{60} \equiv 1 + 0 = 1$

知乎@余命数

此时状态序对 $(a_{61}, a_{62}) = (1, 1)$, 与初始状态 (a_1, a_2) 相同. 因此, 该数列的最小正周期为 60.

为了求 F_{2024} 的个位数字, 计算 2024 模 60 的余数.

$$2024 = 60 \times 33 + 44$$

因此, $a_{2024} = a_{44}$.

查阅已计算的序列, 第 44 项是 3. 故 F_{2024} 的个位数字是 3.

注 2.15 (周期性的分析路径)

- **存在性**. 整系数线性递推在模 m 下必为最终周期数列, 周期不超过 m^k (k 为阶数), 依据是鸽巢原理.
- **求周期**. 逐项计算状态直至首次重复 $\mathbf{v}_r = \mathbf{v}_l$, 周期即为 $r - l$; 最小正周期须实际算出, 通常远小于上界.
- **用性质**. 周期求和把长求和化为一个周期内的求和; 和与积的周期取 lcm, 用于由已知周期合成新数列的周期.

知乎@余命数

2.15 子列与 Bolzano-Weierstrass 定理 *

2.15.1 子列

知乎@余命数

从数列 $\{a_n\}$ 中按原有顺序选取一部分项, 得到的新数列称为子列.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

定义 2.21 (子列) 设 $\{a_n\}$ 是一个数列. 若 $n_1 < n_2 < n_3 < \cdots$ 是一列严格递增的正整数, 则

$$a_{n_1}, a_{n_2}, a_{n_3}, \ldots$$

称为 $\{a_n\}$ 的一个**子列**, 记作 $\{a_{n_k}\}$.

下标 n_k 严格递增, 保证子列中各项的先后顺序与原数列一致. 取 $n_k = 2k-1$, 得到奇数项子列 $a_1, a_3, a_5, \ldots$; 取 $n_k = 2k$, 得到偶数项子列 $a_2, a_4, a_6, \ldots$.

子列继承了原数列的收敛性.

定理 2.12 (子列的极限) 若数列 $\{a_n\}$ 收敛于 L, 则它的任意子列 $\{a_{n_k}\}$ 也收敛于 L.

知乎@余命

知乎@余命数

证明 设 $\lim_{n\to\infty} a_n = L$. 对任意 $\epsilon > 0$, 存在正整数 N, 使得当 $n > N$ 时, $|a_n - L| < \epsilon$.

严格递增的正整数序列满足 $n_k \geq k$. 取 $K = N$, 当 $k > K$ 时, $n_k \geq k > N$, 故 $|a_{n_k} - L| < \epsilon$.

因此 $\lim_{k\to\infty} a_{n_k} = L$. □

这个定理的逆否命题提供了一个判断数列发散的工具.

知乎

定理 2.13 (发散判据) 若数列 $\{a_n\}$ 有两个子列分别收敛于不同的极限, 则 $\{a_n\}$ 发散.

证明 设子列 $\{a_{n_k}\}$ 收敛于 L_1, $\{a_{m_k}\}$ 收敛于 L_2, 且 $L_1 \neq L_2$.

用反证法. 假设 $\{a_n\}$ 收敛于某个 L. 根据子列的极限定理, $\{a_{n_k}\}$ 和 $\{a_{m_k}\}$ 都应收敛于 L. 但它们分别收敛于 L_1 和 L_2, 故 $L = L_1$ 且 $L = L_2$, 与 $L_1 \neq L_2$ 矛盾.

因此 $\{a_n\}$ 发散. □

数列 $a_n = (-1)^n$ 是最直接的例子. 它的奇数项子列 $a_1, a_3, a_5, \cdots = -1, -1, -1, \ldots$ 收敛于 -1; 偶数项子列 $a_2, a_4, a_6, \cdots = 1, 1, 1, \ldots$ 收敛于 1. 两个子列极限不同, 故 $\{a_n\}$ 发散.

知乎@余命数

这个判据同样适用于函数极限. 在 第四卷第 16.2 节中, 函数 $g(x) = \sin(1/x)$ 在 $x \to 0$ 时没有极限, 证明方法正是构造两条趋近于 0 的路径 $x_n = \dfrac{1}{2n\pi + \pi/2}$ 和 $y_n = \dfrac{1}{n\pi}$, 对应的函数值子列分别趋于 1 和 0. 两条路径给出不同的极限, 函数极限便不存在. 数列的子列判据与函数的路径判据在逻辑上完全一致.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

2.15.2 Bolzano-Weierstrass 定理

收敛数列的子列仍收敛, 但发散数列的子列未必发散. 一个有界的发散数列, 虽然整体没有极限, 却可能存在收敛的子列.

定理 2.14 (Bolzano–Weierstrass 定理) 任何有界数列都存在收敛的子列.

定理的证明采用二分法. 反复对分区间, 每次保留包含无穷多项的那一半, 由此构造一列嵌套收缩的区间, 最终锁定一个收敛子列的极限点.

证明 设 $\{a_n\}$ 是有界数列. 存在实数 a, b, 使得对所有 n, $a \le a_n \le b$.

将区间 $[a, b]$ 等分为两个子区间. 由于数列有无穷多项, 至少有一个子区间包含无穷多项. 记该子区间为 $I_1 = [a_1, b_1]$.

对 I_1 重复此操作, 等分并选取包含无穷多项的那一半, 记为 $I_2 = [a_2, b_2]$.

如此继续, 得到一列嵌套区间

$$[a, b] \supseteq I_1 \supseteq I_2 \supseteq \cdots$$

每个区间的长度是前一个的一半:

$$|I_k| = \frac{b-a}{2^k}$$

当 $k \to \infty$ 时, $|I_k| \to 0$.

嵌套区间的左端点序列 $\{a_k\}$ 非减且有上界 b, 右端点序列 $\{b_k\}$ 非增且有下界 a. 根据 第四卷第 16.1 节中的完备性公理, $\{a_k\}$ 有上确界, $\{b_k\}$ 有下确界. 令

$$L = \sup\{a_k : k \ge 1\} = \inf\{b_k : k \ge 1\}$$

则 $a_k \le L \le b_k$ 对所有 k 成立. 由于 $|I_k| \to 0$, L 是嵌套区间的唯一公共点.

从每个区间 I_k 中选取一项 a_{n_k}, 要求下标严格递增. 这是可行的, 因为每个 I_k 包含无穷多项, 总能在其中找到下标大于 n_{k-1} 的项.

由于 $a_{n_k} \in I_k = [a_k, b_k]$, 有 $a_k \le a_{n_k} \le b_k$. 当 $k \to \infty$ 时, $a_k \to L$ 且 $b_k \to L$, 由夹逼准则得 $a_{n_k} \to L$.

因此 $\{a_{n_k}\}$ 是 $\{a_n\}$ 的一个收敛子列. □

Bolzano-Weierstrass 定理与完备性公理 (第四卷第 16.1 节) 等价. 在有理数系 $\mathbb{Q}$ 中, 有界数列未必有收敛子列, 例如 $\sqrt{2}$ 的十进制逼近序列 $1, 1.4, 1.41, \ldots$ 在 $\mathbb{Q}$ 中没有极限. 定理成立的前提正是实数轴没有空隙.

单调收敛定理要求数列单调且有界, Bolzano-Weierstrass 定理去掉了单调性条件, 代价是仅保证子列收敛, 整体未必收敛. 对于有界但振荡的数列 (如 $a_n = (-1)^n$), 整体极限不存在, 但奇数项子列和偶数项子列分别收敛.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

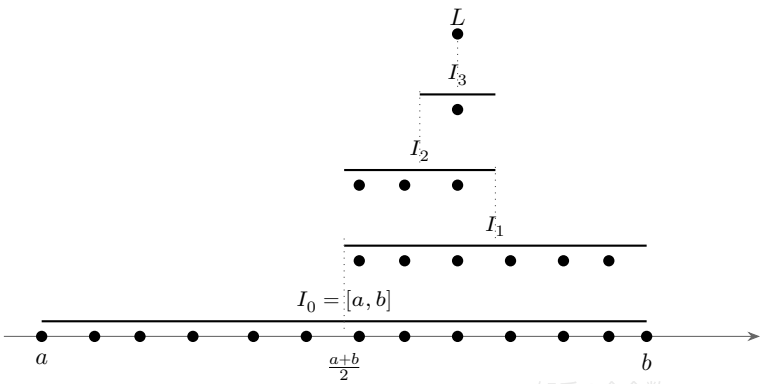


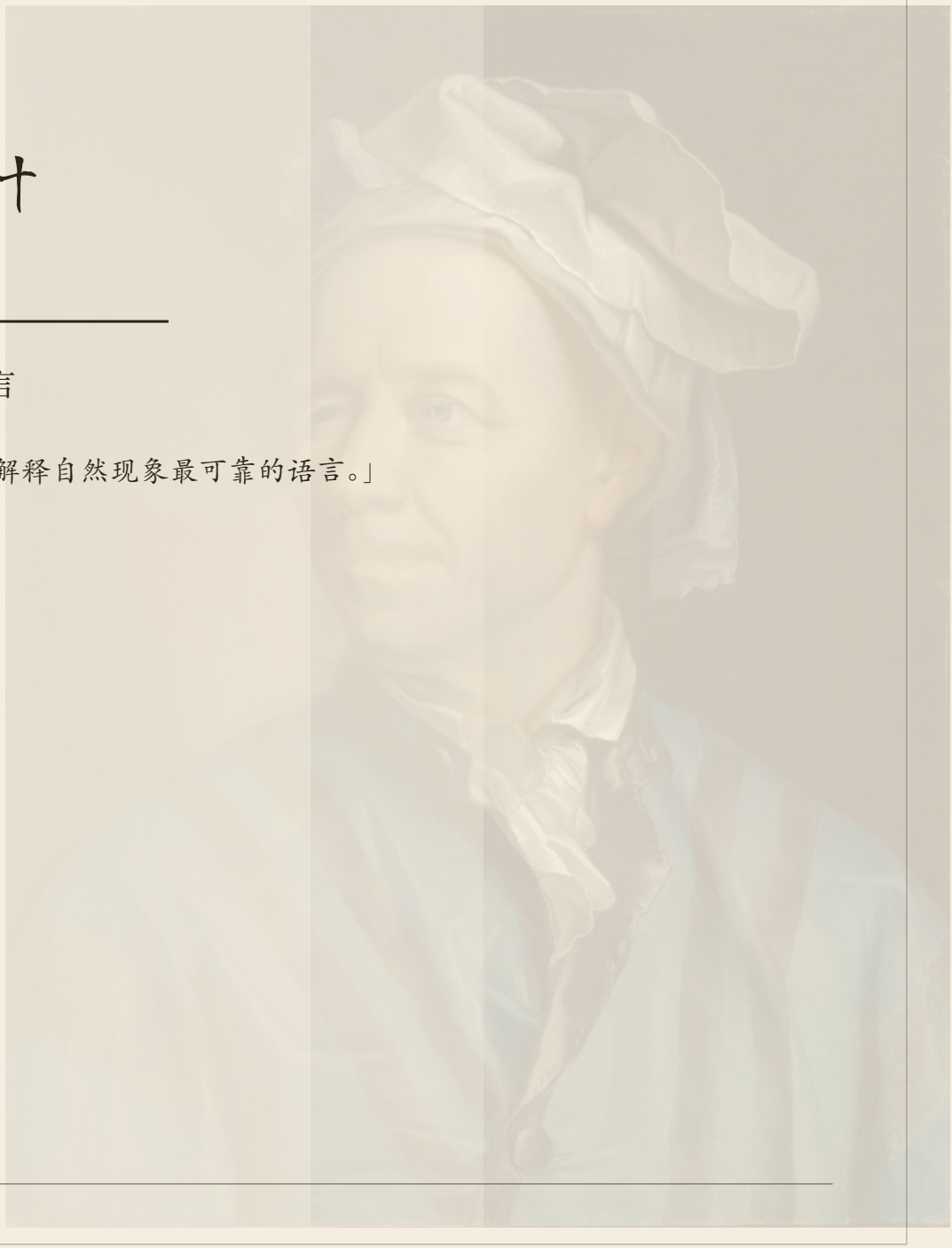
图 2.8 二分法构造收敛子列, 每一步保留包含无穷多项的半区间, 区间长度趋于零, 嵌套区间收缩到极限点 L.

注 2.16 (分析续篇的位置)
本章止于初等数列工具. 数列极限的 ε-N 语言、无穷级数, 以及泰勒公式与泰勒级数, 已迁入第四卷第 16 章(在极限定义与高阶导数之后). 等比公比 $|q| < 1$ 时各项越来越接近零、以及等比无穷和的图像, 可先由有限和的极限来理解. 严格证明见 第四卷第 16 章.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

第 3 章

统计



基础语言

「数学是解释自然现象最可靠的语言。」

——欧拉

第三章 统计

知乎@余命数

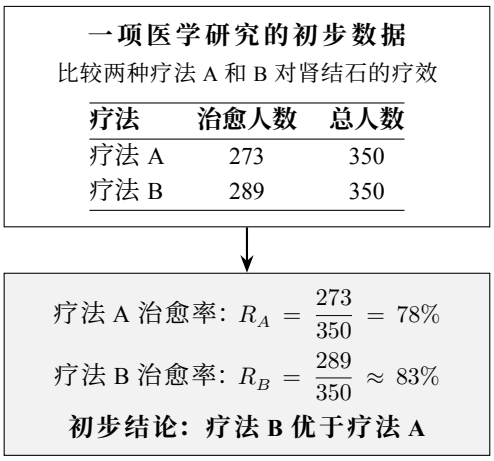


图 3.1 辛普森悖论示例(一):总体数据与初步结论

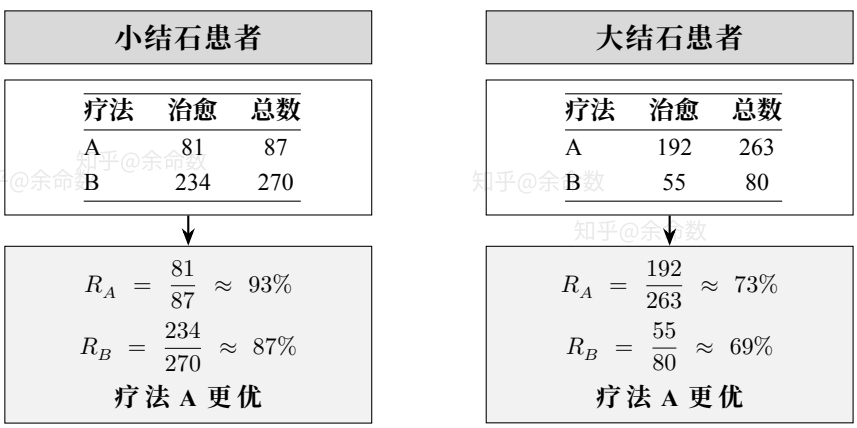


图 3.2 辛普森悖论示例(二):分组数据展现真相

辛普森悖论

我们来看一个思想实验. 某医疗中心汇总了疗法 A 与疗法 B 对肾结石的临床数据. 从整体来看, 疗法 B 的治愈率为 83%, 疗法 A 的治愈率为 78%. 只看这两个数字, 我们大概会推广疗法 B.

但是, 如果把病人按结石大小分成小结石与大结石两组, 结果就变了. 在小结石患者群中, 疗法 A 的治愈率高于疗法 B; 在大结石患者群中, 疗法 A 的治愈率也高于疗法 B. 结石大小是一个重要的潜在变量, 它改变了我们从汇总数据得出的结论.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

一个疗法在每个分组都表现更好，合并后反而显得更差，这种现象称为**辛普森悖论**. 它说明，简单汇总可能掩盖数据内部的结构，从而扭曲事实. 数据本身不会说谎，但如果我们呈现和解读数据的方式不当，就会得到错误的结论.

为了避免这类陷阱，我们不能只满足于表面的数字. 我们需要一套系统的方法，来科学地收集数据，严谨地分析数据内部的结构，并最终做出可靠的推断. 这套方法就是**统计学**.

3.1 描述数据

我们来讨论描述性统计. 它的起点是一组原始数据，例如一个班级所有同学的身高记录. 这些数据在整理之前只是一些无序的数字. 我们的任务是把它们整理、提炼，使其蕴含的规律显现出来.

3.1.1 总体与样本

我们先界定研究对象的范围.

定义 3.1 (总体与样本)
- **总体：**我们所研究的对象的全体构成的集合. 例如，中国所有高三学生的身高.
- **个体：**构成总体的每一个对象. 例如，每一位高三学生的身高数据.
- **样本：**从总体中实际抽取并观测的一部分个体的集合. 例如，随机抽取的 1000 名高三学生的身高数据.
- **样本容量：**样本中所包含的个体数量，通常记为 n.

在大多数情况下，我们受成本、时间或可行性的限制，无法观测整个总体. 于是我们研究一个有代表性的**样本**，再由它推断**总体**的特征. 样本是我们手中可测的数据，总体是我们要推断的对象.

3.1.2 频率分布

拿到样本数据后，我们先了解数据的整体分布形态.

定义 3.2 (频率分布) 我们将样本数据分组，统计每组内数据出现的次数，称为**频数**，并计算频数占样本总量的比例，称为**频率**，然后用表格或图形把这种分布情况呈现出来，这就得到**频率分布**.

常用的图形有条形图（适用于离散或分类数据），以及适用于连续数据的**频率分布直方图**.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

例题 3.1 某校随机抽取了 50 名学生, 测量其身高(单位 cm), 然后制作其频率分布表和直方图.

168, 175, 159, 171, 166, 182, 163, 170, 176, 169, 161, 173, 178, 165, 172, 157, 180, 167, 174, 169, 164, 171, 177, 162, 168, 170, 175, 166, 179, 158, 163, 172, 168, 174, 160, 173, 167, 171, 165, 181, 169, 176, 164, 170, 166, 172, 161, 175, 168, 173.

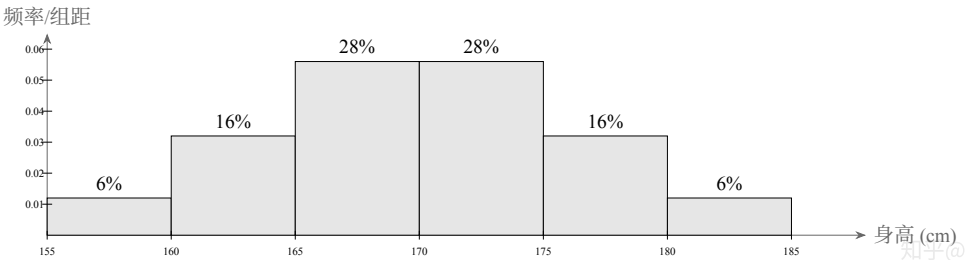


图 3.3 50 名学生身高频率分布直方图

解 这些数据的最大值为 182, 最小值为 157. 取组距为 5, 把数据分成 6 组, 得到频率分布表.

分组	频数	频率	频率/组距
$[155, 160)$	3	0.06	0.012
$[160, 165)$	8	0.16	0.032
$[165, 170)$	14	0.28	0.056
$[170, 175)$	14	0.28	0.056
$[175, 180)$	8	0.16	0.032
$[180, 185)$	3	0.06	0.012
合计	50	1.00	—

在直方图中, 每个矩形的高度正比于该组的频率, 纵轴通常标为频率/组距. 这样所有矩形的**面积之和**恰好为 1.

3.1.3 数据的数字特征

图形给出直观印象, 但若要精确比较和分析, 还需用具体数字概括样本的主要特征.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

3.1.3.1 衡量中心趋势

定义 3.3 (样本均值) 对于一个容量为 n 的样本 $\{x_1, x_2, \dots, x_n\}$, 其**样本均值**, 记为 $\bar{x}$, 定义为.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

样本均值即数据集中所有数值的算术平均, 它描述数据围绕的中心位置.

3.1.3.2 衡量离散程度

但仅知道中心位置还不够. 两组数据可以有相同的均值, 但围绕中心的分散程度可能不同. 为此我们需要一个度量离散程度的指标. 一个直接的办法是计算每个数据点到均值的平均距离 $\frac{1}{n}\sum|x_i - \bar{x}|$, 称为平均绝对离差. 绝对值在数学处理上不够便利, 因此统计学改用离差平方和来度量变异.

知乎@余命数

定义 3.4 (样本方差与样本标准差 (高中课本定义)) 对于样本 $\{x_1, \dots, x_n\}$ 及其均值 $\bar{x}$, 其**样本方差**, 记为 s^2, 定义为.

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

样本方差的算术平方根 $s = \sqrt{s^2}$ 称为**样本标准差**.

高中课本以 n 为分母定义样本方差, 统计学则采用 $n-1$ 为分母的定义.

定义 3.5 (样本方差与样本标准差) 对于样本 $\{x_1, \dots, x_n\}$ 及其均值 $\bar{x}$, 其**样本方差**, 记为 s^2, 定义为.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

样本方差的算术平方根 $s = \sqrt{s^2}$ 称为**样本标准差**.

样本方差的分母是 $n-1$, 这常令人困惑. 既然样本均值是对 n 个数据求平均, 计算平均离差平方时却要除以 $n-1$, 这个调整称为**贝塞尔校正**.

我们计算样本方差 s^2, 是为了**估计**未知的**总体方差** σ^2. 总体方差由数据围绕**总体真实均值** μ 的离散程度决定, 但 μ 往往未知, 只能用**样本均值** $\bar{x}$ 替代, 这就会引入偏差.

知乎@余命数

样本均值 $\bar{x}$ 是由这 n 个数据点决定的中心. 算术平均值有一个最优性质, 它使该组数据的离差平方和 $\sum(x_i-c)^2$ 达到**最小值**. 因此用 $\bar{x}$ 计算的离差平方和 $\sum(x_i-\bar{x})^2$, 比用任何其他中心点(包括未知的 μ)计算的都小.

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 \le \sum_{i=1}^{n}(x_i-\mu)^2$$

于是, 若用 $\bar{x}$ 并除以 n 计算方差, 所得值会系统性地**偏小**. 在统计学中, 这种系统性估计偏差称为**偏误**.

为了修正这一偏差, 我们需把偏小的结果适当放大. 数学推导表明, 分母取 $n-1$ 恰可校正偏差, 使得 s^2 在平均意义上等于所估的 σ^2. 这样的估计量称为**无偏估计量**.

$n-1$ 在统计学中又称**自由度**.n 个独立样本点算出样本均值 $\bar{x}$ 后, 满足约束 $\sum(x_i-\bar{x})=0$. 在这一约束下, 知道其中 $n-1$ 个离差, 第 n 个就被完全确定. 因此估计总体离散程度时, 实际可独立变化的信息只有 $n-1$ 个.

标准差 s 的单位与原始数据相同, 它度量数据点偏离其均值的典型距离.s 越大, 数据越分散; s 越小, 数据越集中.

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

3.2 抽样方法

3.2.1 简单随机抽样

我们先讨论简单随机抽样. 它是所有抽样方法中最基础的一种, 其目的在于消除抽样过程中的主观偏见与系统性误差, 并确保样本具有**代表性**. 要理解它的定义, 我们回到一个根本问题, 即如何保证公平性.

假设我们调查全校学生的平均身高, 但只测量了校篮球队的成员. 这样得到的样本均值会系统性地高于真实值, 样本不具备代表性. 这种因抽样方法缺陷导致的系统性偏差称为**抽样偏误**. 为了克服偏误, 我们需要一种机制, 使总体中每个个体都有平等的机会被选中.

这一思想源自抽签. 公平的抽签要求每一支签被抽中的可能性完全相同. 我们把抽签的思想数学化, 便得到简单随机抽样的定义.

定义 3.6 (简单随机抽样) 设一个总体含有 N 个个体, 从中逐个不放回地抽取 n 个个体作为样本 ($n \le N$). 如果在整个抽样过程中, **每一次抽取时, 总体中余下的每一个个体被抽到的机会都相等**, 则称这种抽样为**简单随机抽样**.

知乎@余命数

这个定义刻画了一个动态的、程序化的**过程**, 并为公平性提供可操作的保证.

- **逐个不放回**, 明确抽样的基本机制. 一个个体一旦被抽中, 便移出总体, 不再参与后续抽取. 这是社会调查与质量检验的标准做法.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **每一次抽取时机会都相等**, 这一定义的关键, 也是公平性的数学形式. 它要求抽第一步时每个个体被抽中的概率为 $1/N$; 抽出一个后, 剩下的 $N-1$ 个个体每一个在第二步被抽中的概率为 $1/(N-1)$, 以此类推. 该条件确保任何阶段抽样过程对剩余个体均无偏好.

这种程序公平性的来源是**随机化**, 它也是现代统计学的基础. 引入纯粹随机机制, 可打乱总体中可能存在的内在结构, 例如学生按身高、成绩或班级排序. 随机化下某一次抽得的样本未必恰好代表总体, 但**抽样方法**本身无偏, 长期来看由此方法产生的样本能公正反映总体.

上述过程化的定义可导出一个在组合数学上极为重要的等价性质, 即在简单随机抽样下, 总体中每一个容量为 n 的可能子集被抽为最终样本的概率完全相等.

从 N 个个体中抽取 n 个, 总共存在 $\binom{N}{n}$ 个不同的样本组合. 考虑其中任意一个特定的样本组合 $\{a_1, a_2, \dots, a_n\}$, 按照定义的抽样过程, 以特定顺序, 例如 $a_1, a_2, \dots, a_n$ 被抽出的概率为

知乎@余命数

$$\frac{1}{N} \times \frac{1}{N-1} \times \frac{1}{N-2} \times \cdots \times \frac{1}{N-n+1} = \frac{1}{P_N^n}.$$

该样本中的 n 个个体可按 $n!$ 种不同顺序被抽出, 且每种顺序的发生概率相同, 故此特定样本组合被抽中的总概率为

$$\begin{aligned} n! \times \frac{1}{P_N^n} &= n! \times \frac{(N-n)!}{N!} \qquad (\text{由 } P_N^n = N!/(N-n)!) \\ &= \frac{1}{N!/(n!(N-n)!)} = \frac{1}{\binom{N}{n}}. \end{aligned}$$

因此, 任何一个容量为 n 的样本组合, 其被选中的概率都是 $1/\binom{N}{n}$. 这一性质是所有基于古典概型的统计推断计算的理论基础, 它把一个动态的抽样过程与一个静态的、所有可能结果等概率的样本空间联系起来.

例题 3.2 设总体容量为 N, 从中抽取一个容量为 n 的简单随机样本. 证明总体中任意一个个体 a 被抽入样本的概率为 $\frac{n}{N}$.

证明 我们用等价性质来证明. 由简单随机抽样的等价性质, 任何容量为 n 的子集被抽为样本的概率均等, 故所有可能样本组合总数为 $\binom{N}{n}$. 若样本须包含特定个体 a, 只需从其余 $N-1$ 个个体中再抽取 $n-1$ 个与之组合, 故包含 a 的样本数为 $\binom{N-1}{n-1}$. 于是 a 被抽入样本的概率为

$$P(a \text{ 于样本里}) = \frac{\text{包含 } a \text{ 的样本数}}{\text{总样本数}} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}}.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

将组合数按定义展开并约分, 得

$$\begin{aligned}\frac{\binom{N-1}{n-1}}{\binom{N}{n}} &= \frac{(N-1)!}{(n-1)!(N-n)!}\cdot\frac{n!(N-n)!}{N!} && \text{(展开组合数)}\\ &= \frac{(N-1)!}{(n-1)!}\cdot\frac{n(n-1)!}{N(N-1)!} && \text{(消去 } (N-n)!\text{)}\\ &= \frac{n}{N}.\end{aligned}$$

同一结论亦可由顺序视角导出. 将事件“个体 a 被抽入样本”分解为 n 个互斥事件 E_k(“a 在第 k 次抽取时被抽中”, $k=1,2,\dots,n$)的并集, 则 $P(a\text{ 于样本里})=\sum_{k=1}^{n}P(E_k)$. 显然 $P(E_1)=P(a\text{ 在第 1 次被抽中})=\frac{1}{N}$. 对于 E_2, 须第 1 次未抽中 a 且第 2 次抽中, 由乘法公式

$$P(E_2)=\frac{N-1}{N}\times\frac{1}{N-1}=\frac{1}{N}.$$

一般地, E_k 要求前 $k-1$ 次均未抽中 a 且第 k 次抽中, 故

$$\begin{aligned}P(E_k) &= \frac{N-1}{N}\times\frac{N-2}{N-1}\times\cdots\times\frac{N-k+1}{N-k+2}\times\frac{1}{N-k+1}\\ &= \frac{1}{N} && \text{(中间项依次约分)}.\end{aligned}$$

即在简单随机抽样过程中, 任一个体在任一次被抽中的概率都相同. 因此 a 被抽入容量为 n 的样本的总概率为

$$P(a\text{ 于样本里})=\sum_{k=1}^{n}P(E_k)=\sum_{k=1}^{n}\frac{1}{N}=\frac{n}{N}.$$

□

这个结论是简单随机抽样公平性的另一种体现. 它与每个样本组合被抽中的概率相等这一性质等价, 在理论上极为重要. 它保证了使用样本均值 $\bar{x}$ 估计总体均值 μ 时, 这种估计在方法论上无偏, 因为总体中每一个成员都有完全相同的机会对这个估计值做出贡献.

例题 3.3 在第二次世界大战期间, 盟军分析缴获的德军坦克的序列号, 以估计其总产量. 设坦克的序列号从 1 到 N 连续编号, 其中 N 是未知的总产量. 盟军缴获了 n 辆坦克, 其序列号构成从总体 $\{1,2,\dots,N\}$ 中抽取的容量为 n 的简单随机样本. 若观测到的最大序列号为 m, 试给出 N 的一个合理估计.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 令观测到的 n 个序列号为 $x_1, x_2, \dots, x_n$, 样本最大值 $m = \max\{x_1, \dots, x_n\}$. 由于 $N \geq m$, 而 m 几乎必然是低估(恰好抽到最大序列号坦克的概率极小), 需在 m 之上追加一个修正项.

这 n 个被抽中的序列号将从 1 到 N 的整个序列分割开来. 在真正的随机样本中, 成员在总体中应大致均匀散布, 故可用样本内部的平均间距估计总体中未观测到的末尾间距.n 个样本点大致将区间 $[1, m]$ 分割为 n 个小段, 平均每段长度约为 m/n, 以此估计从 m 到 N 这段末尾间距的长度, 便得到初步模型 $\hat{N} \approx m + \frac{m}{n}$.

此估计尚有改进空间. 注意 n 个样本点将总体 N 个数字分成 $n+1$ 个间隔(从 0 到最小样本、样本间的间隔、以及从最大样本到 $N+1$), 在随机抽样下这些间隔的期望长度应相等, 由此消除系统性偏差, 得到更精确的估计量

$$\hat{N} = m + \frac{m}{n} - 1.$$

例如, 缴获 $n = 5$ 辆坦克且最大序列号 $m = 120$ 时, 总产量的估计为

$$\hat{N} = 120 + \frac{120}{5} - 1 = 120 + 24 - 1 = 143.$$

德国坦克问题是统计推断中参数估计的一个经典范例. 它展示了统计学如何基于一个不完整的随机样本信息, 对一个未知的总体参数(这里是总体最大值 N)做出推断. 这里给出的 $\hat{N} = \frac{n+1}{n} m - 1$ 是基于样本最大值的常用无偏估计量.

3.2.2 分层抽样

定义 3.7 (分层抽样) 当总体由若干个有显著差异、互不重叠的子群体, 称为**层**构成时, 我们将总体划分为若干层, 然后**在每一层内部分别进行简单随机抽样**, 最后把各层抽取的样本合并起来. 这种方法称为**分层抽样**.

辛普森悖论的例子告诉我们, 当总体存在显著的内部结构时, 对这种结构视而不见可能会导致错误的结论. 简单随机抽样凭纯粹的随机化可保证长期公平性, 但无法保证**每一次**抽样都能准确反映总体的内部结构. 例如, 在一个男女比例为 $1:4$ 的总体中进行容量为 10 的简单随机抽样, 可能因随机性的波动抽到一个男女比例为 $1:1$ 的样本, 从而对依赖于性别的某些特征做出有偏的估计.

分层抽样是对纯粹随机性的一种约束和优化. 它承认并利用了总体的内在异质性, 做法是先分层后抽样. 这样, 一个异质性高的复杂总体被分解为若干个同质性高的子总体(层), 我们在每个子总体内部实施随机化.

这种方法有几条优势. 一条是**提升估计精度**. 确保所有子群体都按其在总体中的重要性(通常是规模)被恰当地代表, 分层抽样可降低抽样误差, 得到比同等规模简单随机抽样更精确的总体参数估计值, 即方差更小的估计量. 另一条是**保证子群体代表性**. 它确保样本中包含来自所有层的足够数量的个体, 我们既能对

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

总体进行推断, 又能对各子群体自身的特征进行分析和比较. 第三条是实施方便. 在组织层面, 把大规模的抽样任务分解到各个层级, 往往更易于管理.

在实际操作中, 最常见的分层抽样方法是**按比例分配**, 即每个层所抽取的样本容量占总样本容量的比例, 等于该层所含个体数占总体个体数的比例. 设总体容量为 N, 被划分为 L 个层, 每层的容量分别为 $N_1, N_2, \dots, N_L$. 若要抽取容量为 n 的样本, 则在第 h 层 ($h = 1, \dots, L$) 中抽取的样本容量 n_h 为

$$n_h = n \cdot \frac{N_h}{N}$$

例题 3.4 某大学共有学生 $N = 5000$ 人, 其中低年级学生 $N_1 = 3000$ 人, 高年级学生 $N_2 = 2000$ 人. 学校希望估计全体学生平均每周的自习时间. 根据以往经验, 低年级学生自习时间的波动性较大, 而高年级学生则相对稳定. 现采用分层抽样方法, 抽取一个容量为 $n = 100$ 的样本.

1. 采用按比例分配的原则, 应在两个年级各抽取多少名学生?
2. 假设抽样结果为低年级样本的平均自习时间 $\bar{x}_1 = 15$ 小时, 高年级样本的平均自习时间 $\bar{x}_2 = 25$ 小时. 试估计全校学生平均每周的自习时间.
3. 试讨论为何在这种情况下, 分层抽样的估计结果比简单随机抽样可能更为可靠.

解 总体根据年级这一关键变量自然地分为两层, 这是分层抽样的典型应用场景.

(1) 按比例分配, 第 h 层抽取样本量 $n_h = n \cdot \frac{N_h}{N}$. 代入数据, 低年级层应抽取 $n_1 = 100 \cdot \frac{3000}{5000} = 60$ 人, 高年级层应抽取 $n_2 = 100 \cdot \frac{2000}{5000} = 40$ 人.

(2) 由于样本结构与总体结构成比例, 以各层在总体中的比例 $W_h = N_h/N$ 为权重对各层样本均值加权平均, 即得总体均值 μ 的估计

$$\begin{aligned}\bar{x}_{st} &= W_1\bar{x}_1 + W_2\bar{x}_2 = \frac{N_1}{N}\bar{x}_1 + \frac{N_2}{N}\bar{x}_2 \\ &= \frac{3000}{5000} \cdot 15 + \frac{2000}{5000} \cdot 25 = 0.6 \cdot 15 + 0.4 \cdot 25 = 9 + 10 = 19 \text{ 小时}.\end{aligned}$$

故全校学生平均每周自习时间估计为 19 小时.

(3) 简单随机抽样依赖纯粹的随机机会获得代表性. 本例中容量为 100 的简单随机样本, 其年级比例可能因随机性偏离真实的 $3:2$, 若偶然抽到 40 名低年级与 60 名高年级学生, 且样本均值仍为 15 与 25, 则整体均值为 $0.4 \cdot 15 + 0.6 \cdot 25 = 21$ 小时, 高估真实总体均值. 分层抽样强制样本结构与总体结构一致, 从根本上消除了样本中各子群体比例不当引入的抽样误差, 将随机性限制在各同质层内部, 使最终的加权估计量更为稳定和精确.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

此例导出了分层抽样下总体均值的估计方法, 即 $\bar{x}_{st}=\sum_{h=1}^{L}W_h\bar{x}_h$. 这是一个统计性质优良的无偏估计量.

例题 3.5　某市教育部门计划对一项新的教学改革方案进行民意调查, 以了解全体高中生的支持率. 该市共有高中生 $N=10000$ 人, 其中高一年级 $N_1=4000$ 人, 高二年级 $N_2=3500$ 人, 高三年级 $N_3=2500$ 人. 由于不同年级的学生面临的学业压力和课程内容不同, 他们对改革方案的态度可能存在显著差异. 因此, 决定采用分层抽样的方法, 抽取一个容量为 $n=400$ 的样本.

1. 采用按比例分配的原则, 计算应从每个年级分别抽取多少名学生?
2. 调查结束后, 样本数据显示高一、高二、高三学生的支持率分别为 85%、70%、60%. 请估计该市全体高中生对该方案的支持率.

解　各年级是天然的层, 层内学生同质性较高而层间可能存在显著差异, 这是分层抽样的典型应用场景.

1. 按比例分配要求样本中各层的比例与总体中各层的比例保持一致. 各年级学生占总体的比例分别为 $W_1=\frac{N_1}{N}=\frac{4000}{10000}=0.4$、$W_2=\frac{N_2}{N}=\frac{3500}{10000}=0.35$、$W_3=\frac{N_3}{N}=\frac{2500}{10000}=0.25$. 据此分配样本容量 $n=400$, 高一抽取 $n_1=n\times W_1=400\times0.4=160$ 人, 高二抽取 $n_2=n\times W_2=400\times0.35=140$ 人, 高三抽取 $n_3=n\times W_3=400\times0.25=100$ 人.
2. 已知各层内样本支持率 $\hat{p}_1=0.85,\hat{p}_2=0.70,\hat{p}_3=0.60$. 各层在总体中的权重不同, 不能将三者取算术平均, 须以各层比例 W_h 为权重加权平均, 得

$$\begin{aligned}\hat{p}_{st}&=W_1\hat{p}_1+W_2\hat{p}_2+W_3\hat{p}_3\\&=0.4\times0.85+0.35\times0.70+0.25\times0.60\\&=0.34+0.245+0.15=0.735.\end{aligned}$$

故该市全体高中生对该方案的支持率约为 73.5%.

本例说明, 当层间均值差异显著时, 简单随机抽样的总体方差会同时包含层内波动与层间差异; 按比例分层抽样先在各层内抽样再加权汇总, 可消除由层间结构混杂带来的额外方差. 因此在相同样本量下, 分层抽样通常能给出更稳定的总体均值估计.

例题 3.6　一家连锁企业拥有两种类型的门店, 即旗舰店和社区店. 共有旗舰店 $N_1=50$ 家, 社区店 $N_2=450$ 家. 根据历史数据, 旗舰店的月均销售额为 $\mu_1=200$ 万元, 销售额的标准差为 $\sigma_1=50$ 万元; 社区店的月均销售额为

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$\mu_2 = 80$ 万元, 标准差为 $\sigma_2 = 15$ 万元. 现计划抽取 $n = 50$ 家门店来估计整个连锁企业的平均月销售额. 试比较简单随机抽样和按比例分层抽样两种方法的效率.

解 抽样方法的效率由其估计量的方差衡量, 方差越小则估计波动性越小、精确度越高, 效率也越高. 以下比较两种抽样方法下样本均值 $\bar{x}$ 的方差.

先计算总体相关参数. 总体规模 $N = N_1 + N_2 = 500$, 各层权重 $W_1 = \frac{N_1}{N} = 0.1$, $W_2 = \frac{N_2}{N} = 0.9$, 总体均值

$$\mu = W_1\mu_1 + W_2\mu_2 = 0.1 \times 200 + 0.9 \times 80 = 20 + 72 = 92 \text{ 万元}.$$

总体方差 σ^2 由层内方差的加权平均与层间方差两部分构成, 代入数据得

$$\begin{aligned}\sigma^2 &= \sum_{h=1}^{2} W_h\sigma_h^2 + \sum_{h=1}^{2} W_h(\mu_h - \mu)^2 \\ &= (0.1 \times 50^2 + 0.9 \times 15^2) + [0.1(200-92)^2 + 0.9(80-92)^2] \\ &= (250 + 202.5) + (1166.4 + 129.6) = 452.5 + 1296 = 1748.5.\end{aligned}$$

1. 简单随机抽样. 在此方法下样本均值 $\bar{x}_{srs}$ 的方差为 (忽略有限总体校正系数)

$$\mathrm{Var}(\bar{x}_{srs}) = \frac{\sigma^2}{n} = \frac{1748.5}{50} = 34.97.$$

2. 按比例分层抽样按比例分配样本量 $n_1 = nW_1 = 50 \times 0.1 = 5$ 家, $n_2 = nW_2 = 50 \times 0.9 = 45$ 家. 分层抽样下总体均值估计量 $\bar{x}_{st} = W_1\bar{x}_1 + W_2\bar{x}_2$ 的方差为

$$\begin{aligned}\mathrm{Var}(\bar{x}_{st}) &= \sum_{h=1}^{2} W_h^2 \frac{\sigma_h^2}{n_h} = \frac{W_1^2\sigma_1^2}{nW_1} + \frac{W_2^2\sigma_2^2}{nW_2} = \frac{1}{n}\sum_{h=1}^{2} W_h\sigma_h^2 \\ &= \frac{1}{50}(0.1 \times 50^2 + 0.9 \times 15^2) = \frac{1}{50}(250 + 202.5) = \frac{452.5}{50} = 9.05.\end{aligned}$$

比较二者, $\mathrm{Var}(\bar{x}_{st}) = 9.05 < \mathrm{Var}(\bar{x}_{srs}) = 34.97$, 分层抽样的估计量方差远小于简单随机抽样. 故在相同样本量下, 分层抽样提供更为精确的估计.

本例用到分层抽样的两步, 即按比例分配样本量和对结果进行加权汇总. 按比例分配确保样本在结构上是总体的缩影, 从而保证了样本的代表性. 加权平均是从样本数据推断总体参数的正确逻辑, 它承认并利用了总体的内在分层结构, 避免了因某一层样本比例过高或过低而导致的估计偏差. 如果研究者错误地对 $85\%, 70\%, 60\%$ 求简单算术平均, 将得到 $(0.85 + 0.7 + 0.6)/3 \approx 71.7\%$, 这会低估总体支持率, 因为占比较大的高一年级的较高支持率没有得到足够的体现.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

3.2.3 系统抽样

知乎@余命数

定义 3.8(系统抽样) 将总体的 N 个个体按序编号. 确定抽样间隔 $k=\lfloor N/n \rfloor$. 在第 1 到第 k 个个体中随机抽取一个作为起始个体 i. 之后, 依次抽取编号为 $i+k, i+2k, \ldots$ 的个体. 这种方法称为**系统抽样**.

3.3 从样本到总体

第3.2节 解决了怎么抽的问题. 样本到手之后, 真正的问题才开始, 那就是手里的数字离总体的真值有多远. 答案可以算得很准, 不过得从一段历史讲起.

1936 年美国总统选举,《文学文摘》(Literary Digest) 杂志预测共和党候选人兰登将以压倒性优势击败寻求连任的罗斯福, 依据是约 240 万份问卷回复, 规模空前, 而且这家杂志从 1916 年起届届测准. 开票结果是罗斯福以 62% 对 38% 大胜, 预测误差高达十九个百分点, 杂志不久便倒闭了. 同一年, 成立不久的盖洛普用 3000 人的样本正确预测了罗斯福获胜, 甚至提前一个月预测出了《文学文摘》会给出什么数字, 误差不到一个百分点.240 万人输给 3000 人, 原因不在人数, 在样本之外.

《文学文摘》的名单取自电话簿和俱乐部会员名册. 大萧条尚未过去, 四户人家才有一部电话, 穷人几乎进不了名单, 而穷人恰恰支持罗斯福.1000 万份寄出的问卷只收回 240 万份, 愿意回信的人又与不肯回信的人不同. 两道筛子朝着同一个方向偏, 样本再大, 也只是在更大的规模上重复同一个错误. 这种系统性的偏移叫**抽样偏差**, 它与抽签式的随机波动是两种不同性质的误差. 偏差要靠抽样设计消除, 随机误差则可以靠数学度量, 接着我们就来看看数学怎么把随机误差算清楚.

3.3.1 盒子模型与标准误

把总体想成一只盒子. 要估计总体中具有某种特征的个体占比 p, 就为每个个体放一张票, 具有该特征的记 1, 其余记 0, 于是盒中 1 的比例恰为 p. 从总体中做样本量为 n 的简单随机抽样, 相当于从盒中无放回地抽 n 张票. 第5.8节 讨论过, 当总体远大于样本时, 无放回与有放回几乎无异, 样本中 1 的张数 X 近似服从二项分布 $B(n,p)$, 期望为 np, 方差为 $np(1-p)$.

实际报告里用的是比例 $\hat{p}=X/n$, 即样本中 1 的占比. 由期望的线性性与方差的性质 (第5.7节),

$$E(\hat{p})=p, \qquad \sqrt{D(\hat{p})}=\sqrt{\frac{p(1-p)}{n}}. \tag{3-1}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

期望恰好等于真值, 说明简单随机抽样没有系统偏差, $\hat{p}$ 的波动大小由方差的平方根刻画. 这个平方根称为**标准误**, 它度量机会误差的典型大小, 样本比例偏离真值的幅度通常在一两倍标准误之内.

例题 3.7 (一千人够不够) 某健康调查的 6672 名受访者中男性占 46%. 一位研究者从中简单随机抽取 100 人复核, 样本中男性占比的波动有多大? 若改抽 400 人呢?

解 把男性记 1、女性记 0, 则 $p = 0.46$, 样本占比 $\hat{p}$ 的标准误为

$$\sqrt{\frac{0.46 \times 0.54}{100}} = \sqrt{0.002484} \approx 0.050,$$

知乎@余命数

约五个百分点. 换成 400 人, 标准误为

知乎@余命数

$$\sqrt{\frac{0.46 \times 0.54}{400}} \approx 0.025,$$

恰好减半. 样本量翻两番, 精度只翻一番, 这就是标准误的**平方根律**, 它由 式 (3-1) 中分母的 n 直接决定. 精度提高一倍要付出四倍的调查成本, 统计的代价在这里明码标价.

知乎@余命数

100 人样本的标准误约 5%, 400 人约 2.5%, 精度随 $\sqrt{n}$ 增长.

式 (3-1) 里没有总体规模的位置. 只要总体远大于样本, 从千万人口的国家和从十万人口的城市各抽一千人, 对占比估计的精度几乎相同. 这与直觉相悖, 却正是抽样的经济性所在. 新闻里民调常标注“误差范围 ± 3 个百分点”, 那正是约一千人的样本、按最保守的 $p = 0.5$ 算出的两倍标准误

知乎@余命数

$$2\sqrt{\frac{0.5 \times 0.5}{1000}} \approx 3.2\%.$$

3.3.2 正态近似

标准误给出波动的尺度, 要算出概率还需要分布形状. 第5.8节 介绍的正态分布在此接手. 当样本量不太小、p 不太靠近 0 或 1 时, $\hat{p}$ 的分布近似正态, 中心在 p, 标准差就是标准误. 样本比例落在真值左右一个标准误内的概率约为 68%, 落在两个标准误内约为 95%.

例题 3.8 (电话公司的市场调查) 某电话公司有 10 万用户, 其中 20% 年收入超过五万元. 公司简单随机抽取 400 人做市场研究, 抽到的用户中高收入者占比介于 18% 与 22% 之间的概率约是多少?

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 高收入人数 $X \sim B(400,\ 0.2)$ 近似，样本占比 $\hat{p}$ 的标准误为

$$\sqrt{\frac{0.2 \times 0.8}{400}} = \sqrt{0.0004} = 0.02.$$

以 20% 为中心、2% 为标准差，区间 $[18\%, 22\%]$ 恰好是两个标准差之内的对称区间，由正态近似，概率约为 95%.

约为 95%.

3.3.3 置信区间

实际调查的方向与上面的例子相反. 真值 p 未知，手里只有一个样本占比 $\hat{p}$，要反过来给 p 圈定范围. 思路是把 $\hat{p}$ 的波动折算回去. 标准误 式 (3-1) 里的 p 用观测值 $\hat{p}$ 代替，得到标准误的估计

$$\widehat{\mathrm{SE}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

然后以 $\hat{p}$ 为中心向两边各取两倍标准误，得到区间

$$\hat{p} \pm 2\,\widehat{\mathrm{SE}} = \left[\hat{p} - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]. \tag{3-2}$$

定义 3.9 (置信区间) 按 式 (3-2) 构造的区间称为总体比例 p 的**置信区间**，置信水平约为 95%. 把两端的标准误倍数换成 1 倍或 3 倍，置信水平相应约为 68% 或 99.7%.

例题 3.9 (选前民调) 某镇有 25000 名选民，随机抽取 1600 人，其中 917 人支持某候选人. 给出全体选民支持率的 95% 置信区间.

解 样本支持率

$$\hat{p} = \frac{917}{1600} \approx 0.573,$$

标准误的估计

$$\widehat{\mathrm{SE}} = \sqrt{\frac{0.573 \times 0.427}{1600}} \approx \sqrt{0.000153} \approx 0.012,$$

置信区间为 $57.3\% \pm 2.4\%$，即 $[54.9\%, 59.8\%]$. 以 95% 的把握，该候选人在全体选民中的支持率落在这个区间内，过半在望，但优势大小仍有不确定的成分.

95% 置信区间约为 $[54.9\%, 59.8\%]$.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

“95% 的把握”的确切含义值得多说一句. 总体支持率是一个固定的数, 不随机, 随机的是区间, 每抽一次样就得到一个新区间.95% 说的是方法的长期能力, 反复抽样反复构造区间, 约有 95% 的区间套住真值, 5% 的错过. 单次调查得到的区间要么套住了, 要么没套住, 无法自知, 但采用这个方法的赌注是划算的.

置信区间也有边界.p 极端靠近 0 或 1 而样本又不大时, 二项分布明显偏斜, 正态近似失真, 式 (3-2) 便不再可靠. 判断的办法是想象总体与样本同构成, 此时预期 1 的张数 $n\hat{p}$ 与 0 的张数 $n(1-\hat{p})$ 都该不少于五左右, 正态近似才稳妥.

现在可以完整地复盘 1936 年了.《文学文摘》的名单偏向富人, 回信又自我筛选, 抽样偏差把 19 个百分点系统地搬走了, 任何标准误都救不回来. 盖洛普从同样的名单里随机抽 3000 人, 偏差消失, 标准误只有约一个百分点, 所以他连别人的错误都能提前测准. 好的推断是两件事的合成, 无偏的设计挡住系统误差, 标准误诚实地报告剩下的随机误差, 缺一不可.

练习 3.1 某市为估计成年居民中经常锻炼者的比例, 简单随机调查了 900 人, 其中 432 人经常锻炼. 求样本比例、标准误的估计, 并给出全市经常锻炼者比例的 95% 置信区间.

练习 3.2 若要把上题的标准误减半, 样本量需增大到多少? 由此说明误差范围 ± 1.5 个百分点大致需要多大的样本.

练习 3.3 某网络平台发起在线问卷, 收到 50 万份回复, 其中 83% 反对某项收费调整, 平台据此宣布“绝大多数用户反对”. 用抽样偏差与标准误的语言评论这一结论.

3.4 相关与回归

前两节的统计对象一次只有一个变量. 更多的问题里, 每个个体带着两个数字出场, 父亲的身高与儿子的身高, 学习年限与收入, 温度与饮料销量. 单看两组平均数会丢掉最重要的信息, 两个变量如何相互联动, 这要从画图开始.

3.4.1 散点图与相关系数

把每一对数据 (x_i, y_i) 画成一个点, 得到的图叫**散点图**. 下图是计算机模拟的两组父子身高数据, 左图散点紧贴一条上升趋势的直线, 右图散点松散, 直线趋势

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

微弱. 两个图的平均数与标准差几乎相同, 区别全在点围绕直线的紧密程度, 需要一个数来度量它.

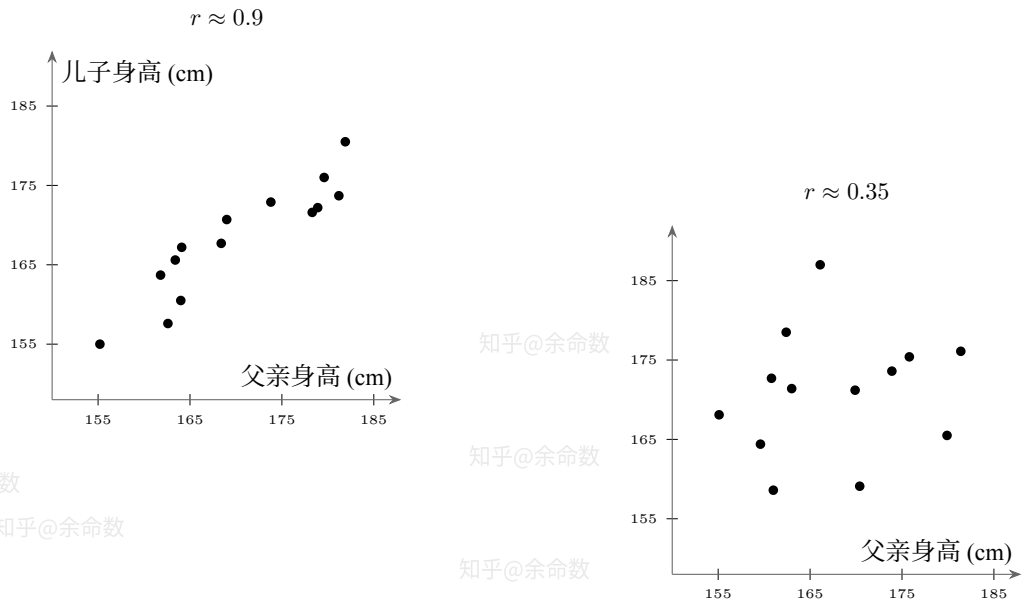


图 3.4　模拟的两组父子身高数据, 平均数与标准差几乎相同, 相关强弱迥异.

度量的想法是把两个变量都标准化, 再看它们同起同落还是各行其是. 记 $\bar{x}, \bar{y}$ 为两组平均数, s_x, s_y 为两组标准差 (第3.1节), 偏差 $d_{x,i} = x_i - \bar{x}$, $d_{y,i} = y_i - \bar{y}$. 衡量一对偏差是否同步, 最自然的乘积 $d_{x,i}\, d_{y,i}$ 同号时增大、异号时缩小, 于是把全部乘积加起来除以恰当的尺度, 得到**相关系数**

$$r = \frac{\sum_i d_{x,i}\, d_{y,i}}{\sqrt{\left(\sum_i d_{x,i}^2\right)\left(\sum_i d_{y,i}^2\right)}}. \tag{3-3}$$

写成标准化的样子即 $r = \dfrac{1}{n}\sum_i \dfrac{d_{x,i}}{s_x} \cdot \dfrac{d_{y,i}}{s_y}$, 分子是标准化变量的平均乘积, 分母的折算把 r 的取值压进 $[-1, 1]$.

注 3.1 ($|r| \le 1$ 的两行证明)
对任意实数 t, $\sum_i (d_{x,i} - t\, d_{y,i})^2 \ge 0$. 展开得 $t^2 \sum d_y^2 - 2t \sum d_x d_y + \sum d_x^2 \ge 0$ 对一切 t 成立, 故关于 t 的二次函数判别式非正, $\left(\sum d_x d_y\right)^2 \le \sum d_x^2 \sum d_y^2$, 代入 式 (3-3) 即 $|r| \le 1$. 等号成立当且仅当存在 t 使所有 $d_{x,i} = t\, d_{y,i}$, 即全部点恰好落在一条直线上, 此时 $r = \pm 1$, 称完全相关.

r 的符号读作趋势方向.$r > 0$ 表示一个变量增大时另一个倾向于增大, 散点整体向上爬; $r < 0$ 表示一增一减, 散点下坡. 绝对值大小读作围绕直线的紧密度, $|r|$ 越接近 1 点带越窄, 越接近 0 云团越散. 有两点常被误读.$r = 0.8$ 并非表示 80% 的点在直线上, r 也不衡量两倍于 $r = 0.4$ 的线性强度, 它只是单调的紧密度指标. 另

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

外 r 只度量**线性**关联, 点落在一条弯曲的光滑曲线上时 r 可以等于 0, 后文的习题会给出这样的例子.

例题 3.10 (手算相关系数) 五名学生的数学成绩 x 与物理成绩 y 为 $(1,3),(2,1),(3,4),(4,2),(5,5)$. 计算 r.

解 $\bar{x}=3,\bar{y}=3$, 偏差

$$d_x=(-2,-1,0,1,2),\qquad d_y=(0,-2,1,-1,2).$$

逐对相乘求和

$$\sum d_x d_y=0+2+0-1+4=5,\qquad \sum d_x^2=\sum d_y^2=10.$$

代入 式 (3-3),

$$r=\frac{5}{\sqrt{10\times 10}}=\frac{1}{2}.$$

$r=0.5$, 中等程度的正相关.

3.4.2 回归直线与最小二乘法

散点图上点带既已成形, 自然想画一条直线概括它, 并用它由一个变量预测另一个. 设直线为 $\hat{y}=a+bx$, 每个点离直线的竖直距离 y_i-a-bx_i 称为**残差**. 让直线贴住全部点, 标准做法是让残差平方和

$$Q(a,b)=\sum_i (y_i-a-bx_i)^2$$

尽可能小, 这就是**最小二乘法**, 相应的直线称为**回归直线**. 最小值可以用配方法干净地求出, 不需要任何新工具.

定理 3.1 (回归系数公式) 使 $Q(a,b)$ 最小的直线过平均点 $(\bar{x},\bar{y})$, 斜率与截距为

$$b=\frac{\sum_i d_{x,i}\,d_{y,i}}{\sum_i d_{x,i}^2}=r\,\frac{s_y}{s_x},\qquad a=\bar{y}-b\,\bar{x}.$$

证明 改写残差. 记 $c=a-\bar{y}+b\bar{x}$, 则 $y_i-a-bx_i=d_{y,i}-b\,d_{x,i}-c$, 于是

$$Q=\sum_i \big(d_{y,i}-b\,d_{x,i}-c\big)^2=\sum_i \big(d_{y,i}-b\,d_{x,i}\big)^2-2c\sum_i\big(d_{y,i}-b\,d_{x,i}\big)+n\,c^2.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

由偏差和为零, $\sum d_y = \sum d_x = 0$, 中项消失, 剩下

$$Q = \sum_i \left(d_{y,i} - b\, d_{x,i}\right)^2 + n\, c^2 \geq \sum_i \left(d_{y,i} - b\, d_{x,i}\right)^2,$$

等号当且仅当 $c = 0$, 即 $a = \bar{y} - b\bar{x}$, 直线必过平均点. 对剩下的二次式再展开,

$$\sum d_y^2 - 2b \sum d_x d_y + b^2 \sum d_x^2,$$

这是关于 b 的开口向上的二次函数, 顶点在

$$b = \frac{\sum_i d_{x,i}\, d_{y,i}}{\sum_i d_{x,i}^2}.$$

最后, 由 $s_x^2 = \frac{1}{n}\sum d_x^2$ 与 $s_y^2 = \frac{1}{n}\sum d_y^2$,

$$r\,\frac{s_y}{s_x} = \frac{\sum d_x d_y}{n\, s_x s_y} \cdot \frac{s_y}{s_x} = \frac{\sum d_x d_y}{\sum d_x^2} = b.$$

□

两处细节值得停一停. 回归线必过平均点, 且最优时残差和 $\sum_i (y_i - \hat{y}_i) = \sum (d_{y,i} - b d_{x,i}) = 0$, 高估与低偿恰好抵消. 斜率 $b = r\, s_y / s_x$ 把两种信息合在一起, r 管方向与紧密程度, 标准差之比管量纲换算, 父亲身高每增加一个标准差, 儿子身高平均增加 r 个标准差.

例题 3.11 (手算回归线) 求上例五名学生成绩数据的回归直线, 并据此预测 $x = 6$ 时的 y.

解 斜率

$$b = \frac{\sum d_x d_y}{\sum d_x^2} = \frac{5}{10} = \frac{1}{2}, \qquad a = \bar{y} - b\,\bar{x} = 3 - \frac{1}{2} \times 3 = \frac{3}{2},$$

回归直线 $\hat{y} = \dfrac{1}{2}x + \dfrac{3}{2}$. 预测 $\hat{y}(6) = \dfrac{15}{4}$. 五个点的残差为 $(1, -\frac{3}{2}, 1, -\frac{3}{2}, 1)$, 残差和恰为零, 残差平方和为 7.5, 已是所有直线能达到的最小值.

$\hat{y} = \frac{1}{2}x + \frac{3}{2}$, 预测值 $\frac{15}{4}$.

例题 3.12 (教育与收入) 美国 2005 年一项抽样调查记录了 562 名 25～29 岁男性的受教育年限与年收入, 摘要统计为平均受教育 12.5 年、标准差 3 年, 平均收入约 30000 美元、标准差约 24000 美元, 相关系数约 0.25. 求由受教育年限预测收入的回归直线, 并解读斜率.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 斜率

$$b = r\frac{s_y}{s_x} \approx 0.25 \times \frac{24000}{3} = 2000, \qquad a \approx 30000 - 2000 \times 12.5 = 5000,$$

回归直线为 $\hat{y} \approx 2000x + 5000$(美元),其中 x 为教育年限,$\hat{y}$ 为预测年收入. 斜率读作在该调查人群中,教育年限每多一年,平均收入约高 2000 美元. 截距 5000 美元对应零年教育的预测,已经处在数据边缘,外推到远离数据的地方并不可靠,同一调查中女性样本的回归截距甚至是负数,这并非说不上学反而倒欠钱,只是直线在数据范围外的延伸没有意义.

斜率约 2000 美元/年,教育每多一年平均收入约高 2000 美元.

3.4.3 均值回归

把斜率换算成标准差语言,会看到一件有趣的事. 预测值的偏差为 $b\,d_x$,除以 s_y 后等于 r 乘以 d_x/s_x,用回归线预测时,一个高出平均 2 个标准差的 x,对应的 y 平均只高出 $2r$ 个标准差. 只要 $|r| < 1$,预测就比观测值更靠近平均.

高尔顿在十九世纪末研究父子身高时正遇到了这个现象. 极高的父亲,儿子平均略矮于父亲;极矮的父亲,儿子平均略高于父亲,身高像被一股力量往平均处拉. 他把这条直线称为“回归线”,“回归”一词从此进入统计学,本意正是**向平均回归**. 数值上,设父亲身高平均 170 cm、标准差 6 cm,儿子身高平均 172 cm、标准差 6 cm,$r = 0.5$,回归线为 $\hat{y} = 0.5x + 87$. 父亲 182 cm(高出平均两个标准差),儿子预测 178 cm,只高出平均一个标准差,高个家庭的后代平均上向中心退了半步,但仍高于人群平均.

3.4.4 相关不等于因果

回归线的斜率常被读成干预效果,教育每多一年收入高 2000 美元,让所有人都多读一年书,收入并不会因此普遍上涨. 上述数据来自观察性调查,研究者只是记录了自然形成的差异,受教育多的人与受教育少的人在能力、家境、机遇上都不同,这些因素与教育纠缠在一起,统计学称为**混杂**. 斜率诚实地度量了人群中的关联,却分不清关联里有多少来自教育本身. 要回答因果问题,靠的是第3.2节 里随机化分组的对照实验,把其他因素的分布搅平,斜率才能读成干预效果.

这正是章首辛普森悖论的回声. 潜在变量既能让汇总数据颠倒分组结论,也能让两个本无因果的变量呈现高度相关. 夏季气温升高同时推高冰淇淋销量与溺水人数,两者的相关系数不低,冰淇淋却救不了溺水者. 遇到相关先问三个问题,数据如何产生,有无潜在的第三变量,换成对照实验结论还成立吗. 相关系数与回归线是描述数据的利器,因果的门槛它们自己跨不过去.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

练习 3.4　六次观测的数据为 $(1,2),(2,1),(3,4),(4,3),(5,5),(6,4)$. 计算相关系数, 求回归直线, 并预测 $x=8$ 时的 y.

练习 3.5　五个点 $(1,8),(2,3),(3,0),(4,3),(5,8)$ 全部落在抛物线 $y=(x-3)^2$ 上. 验证 $r=0$, 并说明 y 由 x 完全决定.

练习 3.6　某地统计发现家庭藏书量与孩子考试成绩的相关系数约为 0.5. 有人提议“多买书就能提高成绩”. 给出至少两条理由说明这个推断不成立.

3.5 章末习题

习题

3.1. (线性变换)一组数据 $x_1,\dots,x_n$ 的平均数为 $\bar{x}$, 标准差为 s. 数据 $y_i=3x_i+2$ 的平均数与标准差各是多少? 提示: 平均数线性, $\bar{y}=3\bar{x}+2$; 方差按 $D(aX+b)=a^2D(X)$ 缩放, 标准差乘 $|3|$.

3.2. (样本均值的标准误) 某地区成年男性身高的标准差约为 7 cm. 从该总体简单随机抽取 49 人, 样本平均身高偏离总体平均的典型幅度是多少? 提示: 样本均值的方差为 $\dfrac{\sigma^2}{n}$, 标准误为 $\dfrac{7}{\sqrt{49}}=1$ cm.

3.3. (过半的概率) 某修正案在选民中的真实支持率为 52%. 随机访问 2500 人, 样本支持率超过一半的概率约是多少? 提示: 标准误 $\sqrt{0.52\times0.48/2500}\approx1\%$, 50% 在真值下方两个标准误处, 用正态近似.

3.4. (样本量设计) 要使支持率的 95% 置信区间半宽不超过 2 个百分点, 按最保守的 $p=0.5$ 估计, 至少需要多大样本? 提示: 解 $2\sqrt{0.25/n}\le0.02$.

3.5. (置信区间的读法) 某次调查得到支持率的 95% 置信区间为 $[51\%,55\%]$, 于是有人断言“真值有 95% 的概率落在这个区间里”. 这句话错在哪里? 提示: 真值是固定常数, 不随机; 随机的是区间, 95% 说的是反复抽样时约 95% 的区间套住真值.

3.6. (父子身高回归) 设父亲身高平均 172 cm、标准差 7 cm, 儿子身高平均 174 cm、标准差 7 cm, 相关系数 0.6. 求由父亲身高预测儿子身高的回归直线, 并预测身高 190 cm 的父亲之子的身高. 提示: $b=r\,s_y/s_x=0.6$, $a=174-0.6\times172$, 代入 $x=190$.

3.7. (向平均回归) 某体能训练营按首日测试把学员分组, 教练发现首日成绩最差的一组次日平均进步明显, 首日成绩最好的一组次日平均反而退步, 于是认为“分组训练法

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

见效”. 给出另一种解释. 提示: 首日极端成绩含有极端的随机波动, 次日波动回归常态, 与训练法无关, 这正是高尔顿的均值回归.

3.8. (相关的符号)判断下列相关系数的符号并简述理由, 二手车的车龄与价格, 身高与体重, 海拔与水的沸点温度. 提示: 依次为负、正、负, 分别来自折旧、体格的同向增长与气压降低沸点下降.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

第 4 章

计数原理

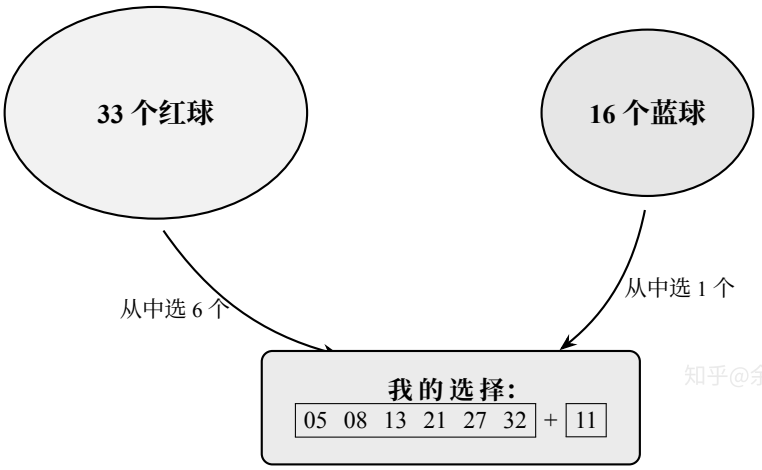


基础语言

「人追求完美，却往往只能得到近似。」

——柯西

第四章 计数原理



中头奖究竟有多难?

图 4.1 彩票抽奖示意图: 从 33 个红球中选 6 个, 从 16 个蓝球中选 1 个

中头奖究竟有多难

知乎上有这么一个有趣的问题“给你选的话, 你想选几百万资产, 还是上清华北大?”[1] 我相信每个人的答案不同, 但是许多人也为这百万现金有过追求, 正如每周开奖的彩票吸引着无数人, 每个人都梦想着能选中那组独一无二的幸运号码. 但这种幸运究竟有多么渺茫, 要精确计算中得头奖的概率, 必须先回答一个根本问题, 即从所有可能的号码组合中到底有多少种不同的组合.

[1] 免责声明. 给一些没有幽默感的人说句话, 如果你不知道的话, 这是一个笑话, 可能并不好笑, 如果你不经常上知乎这个软件平台, 自然是不知道的.

以一种常见的彩票游戏为例. 需从 33 个红色球中选出 6 个, 再从 16 个蓝色球中选出 1 个. 这项任务可以分为两步. 第一步挑选红球, 第二步挑选蓝球. 挑选蓝球很简单, 有 16 种选择. 真正的挑战是第一步.

从 33 个红球中不重复地选出 6 个, 一共有多少种选法?

不妨试着列举, $(1,2,3,4,5,6),(1,2,3,4,5,7),\ldots$ 很快我们就会发现, 这是一个不可能完成的任务, 其总数将是一个天文数字. 更重要的是, 我们只关心选出了**哪些**号码, 而不在乎它们被摇出的**先后顺序**. 选出 ‘05、08、13、21、27、32‘ 和选出 ‘32、21、13、08、27、05‘ 是完全一样的结果.

这种只选不排的问题, 正是我们要研究的对象. 我们需要一套方法, 用来系统地计算在各种限制下有多少种方法. 这些方法既包括需要考虑顺序的**排列**, 也包括不需要考虑顺序的**组合**. 这就是本章将要探索的计数原理.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

4.1 加法原理与乘法原理

我们要处理形形色色的计数问题. 无论问题表面多么不同, 归根到底都用到两条原理, 即**加法原理**与**乘法原理**. 它们本身很朴素, 组合起来却能产生各种计数方法.

4.1.1 加法原理

从北京到上海, 可以乘坐高铁(有 3 个不同的车次), 也可以乘坐飞机(有 5 个不同的航班). 我们来看看一共有多少种选择.

乘坐高铁与乘坐飞机是两类**互斥**的选择. 一旦选定高铁的某个车次, 就不可能再选任何一班飞机. 因此总选择数是两类选择数之和, 这是加法原理.

定理 4.1 (加法原理) 如果完成一件事有 k 类方法, 这些类别两两互斥, 且任何一类方法中的任何一种方法都能独立地完成任务. 在第一类方法中有 m_1 种不同的途径, 在第二类方法中有 m_2 种不同的途径, …, 在第 k 类方法中有 m_k 种不同的途径, 那么完成这件事共有

$$N = m_1 + m_2 + \cdots + m_k$$

种不同的方法.

证明 设第 i 类方法构成有限集合 A_i, 则 $|A_i| = m_i$. 类别两两互斥意味着当 $i \neq j$ 时 $A_i \cap A_j = \emptyset$. 完成这件事的全部方法是并集

$$A_1 \cup A_2 \cup \cdots \cup A_k.$$

因为这些集合两两不交, 对有限集合反复使用两个集合的并集计数公式 $|A \cup B| = |A| + |B|$, 可得

$$|A_1 \cup \cdots \cup A_k| = |A_1| + \cdots + |A_k| = m_1 + \cdots + m_k.$$

因此完成任务的方法总数为 $N = m_1 + \cdots + m_k$. □

加法原理的要点在于分类恰当. 当一个问题可以拆成几个互不相干、彼此独立的情形时, 总数就是各情形之和. 在上面的例子中, 乘坐高铁是一类, 乘坐飞机是另一类; 总选择数就是 $3 + 5 = 8$ 种.

从集合论的观点看, 如果 A_i 代表第 i 类方法的集合, 那么这些集合两两不交, 即 $A_i \cap A_j = \emptyset$ for $i \neq j$. 完成任务的总方法数等于这些集合并集的大小.

$$|A_1 \cup A_2 \cup \cdots \cup A_k| = |A_1| + |A_2| + \cdots + |A_k|$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

下面看几个例子.

例题 4.1 某大学的课程目录中, 数学系开设了 5 门不同的选修课, 计算机系开设了 3 门不同的选修课. 一名学生希望从中选择一门课来学习, 他有多少种不同的选择?

解 这是加法原理最直接的应用, 学生的目标是选择一门课. 可选课程分为互斥的两类. 选数学系课程有 5 种, 选计算机系课程有 3 种. 因一名学生只能选其中一门课, 二者互斥, 由加法原理总的选择数为

$$N = 5 + 3 = 8,$$

即该学生共有 8 种不同的选择.

例题 4.2 使用字母 $\{A, B, C\}$ 和数字 $\{1, 2\}$, 可以构成多少个长度为 1 或长度为 2 的密码? 密码中的字符不能重复.

解 构成一个密码这件事, 可按密码长度分为两种互斥情况.

当密码长度为 1 时, 从 5 个可用字符 $\{A, B, C, 1, 2\}$ 中任选一个, 有 5 种方法.

当密码长度为 2 且字符不重复时, 这是一个分步过程. 先为第一位选一个字符, 有 5 种选择; 因不能重复, 第二位只剩 4 种选择. 由乘法原理, 长度为 2 的密码共有 $5 \times 4 = 20$ 个.(分步相乘的依据, 是下一小节将严格表述的**乘法原理**.)

一个密码的长度不可能既是 1 又是 2, 二者互斥, 由加法原理, 总的密码个数为

$$N = 5 + 20 = 25,$$

即总共可构成 25 个这样的密码.

例题 4.3 (斐波那契数列) 要登上一个有 10 级台阶的楼梯, 如果每次只能上 1 级或 2 级, 总共有多少种不同的登法?

解 直接列举所有情况会非常困难. 以 $f(n)$ 记登上 n 级台阶的方法数, 所求是 $f(10)$.

考虑踏上第 n 级台阶的最后一步, 它只有两种互斥可能. 要么从第 $n-1$ 级一步跨上, 要么从第 $n-2$ 级一步跨两阶上. 前者须先到达第 $n-1$ 级, 有 $f(n-1)$ 种方法; 后者须先到达第 $n-2$ 级, 有 $f(n-2)$ 种方法. 由于最后一步的出发点唯一, 二者互斥, 由加法原理得到递推关系

$$f(n) = f(n-1) + f(n-2).$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

为使用此关系, 需先确定初始条件. 对 1 级台阶只有一种登法, $f(1)=1$; 对 2 级台阶有连续两步 1 级或一步 2 级两种登法, $f(2)=2$.

由此依次推算

$$f(3)=f(2)+f(1)=2+1=3,$$

$$f(4)=f(3)+f(2)=3+2=5,$$

$$f(5)=5+3=8,$$

$$f(6)=8+5=13,$$

$$f(7)=13+8=21,$$

$$f(8)=21+13=34,$$

$$f(9)=34+21=55,$$

$$f(10)=55+34=89,$$

故总共有 89 种不同的登法.

这个例子说明, 加法原理可以用来支撑递归构造. 把一个大问题分解为几个互斥的、结构相同的子问题, 再把它们的解相加, 就得到原问题的解. 序列 $1,2,3,5,8,\ldots$ 是斐波那契数列的一个变体.

例题 4.4 在一个平面直角坐标系中, 一个质点从原点 $(0,0)$ 出发, 每步只能使 y 坐标或 x 坐标加 1, 即向上或向右移动一个单位. 问该质点到达点 $(3,2)$ 有多少条不同的路径?

解 令 $P(m,n)$ 表示从 $(0,0)$ 到达点 (m,n) 的路径数, 所求是 $P(3,2)$. 与登台阶一例同理, 考虑质点是如何到达终点 (m,n) 的. 最后一步必然来自左邻点 $(m-1,n)$ 或下邻点 $(m,n-1)$, 前者对应向右一步到达, 后者对应向上一步到达, 二者互斥. 前者须先到达 $(m-1,n)$, 有 $P(m-1,n)$ 种方法; 后者须先到达 $(m,n-1)$, 有 $P(m,n-1)$ 种方法. 由加法原理得到递推关系

$$P(m,n)=P(m-1,n)+P(m,n-1).$$

边界条件是明显的. 从 $(0,0)$ 到达 x 轴上任意点 $(m,0)$ 只有一条路径(一直向右), 即 $P(m,0)=1$; 同理 $P(0,n)=1$.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

据此由近及远依次推算

$$
\begin{aligned}
&P(1,0)=1, && P(0,1)=1,\\
&P(1,1)=P(0,1)+P(1,0)=2,\\
&P(2,0)=1, && P(0,2)=1,\\
&P(2,1)=P(1,1)+P(2,0)=3,\\
&P(1,2)=P(0,2)+P(1,1)=3,\\
&P(2,2)=P(1,2)+P(2,1)=6,\\
&P(3,1)=P(2,1)+P(3,0)=4,\\
&P(3,2)=P(2,2)+P(3,1)=10,
\end{aligned}
$$

知乎@余命数

故共有 10 条不同的路径.

例题 4.5 正整数 $N=180$ 的所有正因数中, 有多少个是 6 的倍数或者是 10 的倍数?

解 对 N 作素数分解, $N=180=2^2\cdot 3^2\cdot 5^1$, 故其任一正因数 d 都可表示为 $d=2^a\cdot 3^b\cdot 5^c$, 其中 $0\le a\le 2$、$0\le b\le 2$、$0\le c\le 1$.

此题看似须用容斥原理, 因为一个数可能既是 6 的倍数又是 10 的倍数; 然而通过更精细的分类, 即可纯粹地套用加法原理. 将目标因数分为两类.

第一类是 6 的倍数. 因数是 $6=2\cdot 3$ 的倍数, 意味着 $a\ge 1$ 且 $b\ge 1$. 此时 a 可取 $\{1,2\}$ 共 2 种, b 可取 $\{1,2\}$ 共 2 种, c 可取 $\{0,1\}$ 共 2 种, 由乘法原理这类因数共有 $2\times 2\times 2=8$ 个.

第二类是 10 的倍数但不是 6 的倍数. 因数是 $10=2\cdot 5$ 的倍数, 意味着 $a\ge 1$ 且 $c\ge 1$; 同时它不是 6 的倍数, 即条件 $a\ge 1$ 且 $b\ge 1$ 不成立. 既然已要求 $a\ge 1$, 欲使其非 6 的倍数, 必须 $b<1$, 即 $b=0$. 于是 a 仍有 2 种选择, b 仅 1 种, c 仅 1 种, 这类因数共有 $2\times 1\times 1=2$ 个.

这两类因数的定义互斥 (一类是 6 的倍数, 另一类不是 6 的倍数), 由加法原理, 满足条件的因数总数为

$$N=8+2=10,$$

即共有 10 个这样的因数.

精确定义分类标准 (A 和 $B\setminus A$), 把一个看似需要容斥原理的问题转化为纯粹的加法原理问题. 这显示了恰当分类在计数中的作用.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

4.1.2 乘法原理

现在考虑一个不同的场景. 假设你正在为一次晚宴搭配服装, 有 3 件不同的衬衫和 4 条不同的裤子. 搭配方式的总数按下面的步骤算出.

完成搭配一套服装这个任务, 须**依次**完成两个步骤. 先选一件衬衫, 再选一条裤子. 第一步的选择, 例如选白色衬衫并不影响第二步的选择数量(仍有 4 条裤子可选). 对每一件衬衫都有 4 条裤子与之搭配, 有 3 件衬衫, 故总的搭配方式就是把每一步的选择数相乘. 这就是乘法原理.

定理 4.2 (乘法原理) 如果完成一件事需要依次分成 k 个步骤, 完成第一个步骤有 m_1 种不同的方法, 并且在任意已经完成前 $i-1$ 个步骤的选择之后, 第 i 个步骤总有 m_i 种不同的方法 $(i=2,\dots,k)$, 那么完成这件事共有

$$N = m_1 \times m_2 \times \cdots \times m_k$$

种不同的方法.

证明 按分步过程构造选择树. 第 1 步有 m_1 个分支; 对每一条已成前缀, 第 2 步都有 m_2 个分支, 故前两步共 $m_1 m_2$ 种前缀. 一般地, 若前 $i-1$ 步共有 $m_1 m_2 \cdots m_{i-1}$ 种前缀, 由题设每条前缀均可接上 m_i 种第 i 步选择, 故前 i 步共有

$$m_1 m_2 \cdots m_{i-1} m_i$$

种前缀. 对 $i = 1, 2, \dots, k$ 归纳, 完成全部 k 步的方法数为 $m_1 m_2 \cdots m_k$. □

从集合论的观点看, 乘法原理是**笛卡尔积**基数的直接体现. 为此先引入有序对与笛卡尔积.

一个**有序对** (a, b) 是由两个元素组成的集合, 但与普通集合 $\{a, b\}$ 不同, 它规定了元素的次序. 除非 $a = b$, 否则 $(a, b) \neq (b, a)$. 这是分步思想的数学抽象. 第一步的选择是第一个元素, 第二步的选择是第二个元素, 顺序是二者的决定性特征.

知乎@余命数

定义 4.1 (笛卡尔积) 设 A 和 B 为两个集合, 它们的笛卡尔积, 记作 $A \times B$(读作“A cross B”), 是所有可能的有序对 (a,b) 的集合, 其中 $a \in A$ 且 $b \in B$. 用集合符号表示为

$$A \times B = \{(a,b) \mid a \in A, b \in B\}$$

这个概念可以推广到任意 k 个集合. 设 $A_1, A_2, \ldots, A_k$ 为 k 个集合, 它们的笛卡尔积是所有可能的有序 k-元组 $(a_1, a_2, \ldots, a_k)$ 的集合, 其中对任意 $i = 1, 2, \ldots, k$, 都有 $a_i \in A_i$.

$$A_1 \times A_2 \times \cdots \times A_k = \{(a_1, a_2, \ldots, a_k) \mid a_i \in A_i \text{ 对于} i = 1, \ldots, k\}$$

知乎@余命数

笛卡尔积的大小, 即其元素的个数, 恰好就是乘法原理给出的结果. 若这些集合都是有限集, 则

$$|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \times |A_2| \times \cdots \times |A_k| = \prod_{i=1}^{k} |A_i|$$

这个定义把分步完成任务这一直观操作数学化. 每一个步骤对应于从一个集合中选择一次, 而整个任务的所有可能结果构成一个笛卡尔积.

回到服装搭配的问题. 如果衬衫的集合是 $S = \{s_1, s_2, s_3\}$, 裤子的集合是 $P = \{p_1, p_2, p_3, p_4\}$, 那么一套搭配就是一个有序对 (s,p), 其中 $s \in S, p \in P$. 所有可能的搭配构成了集合 $S \times P$, 其大小为

$$|S \times P| = |S| \times |P| = 3 \times 4 = 12$$

下面用笛卡尔积写几个搭配计数的例子.

例题 4.6 一家餐厅的午市套餐允许顾客从 2 种汤中选择一种, 并从 3 种主菜中选择一种, 汤的集合为 $T = \{$罗宋汤, 奶油蘑菇汤$\}$, 主菜的集合为 $M = \{$牛排, 意面, 披萨$\}$. 请问共有多少种不同的套餐组合? 并写出这个组合集合.

知乎@余命数

解 每一种套餐组合都可以被视为一个有序对 (汤, 主菜). 所有可能的套餐组合构成的集合, 是汤的集合 T 与主菜的集合 M 的笛卡尔积 $T \times M$.

其大小为

$$|T \times M| = |T| \times |M| = 2 \times 3 = 6$$

这个集合的具体内容是

$$T \times M = \{(\text{罗宋汤}, \text{牛排}), (\text{罗宋汤}, \text{意面}), (\text{罗宋汤}, \text{披萨}),$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

(奶油蘑菇汤, 牛排), (奶油蘑菇汤, 意面), (奶油蘑菇汤, 披萨)}

共有 6 种不同的套餐组合.

例题 4.7　组装一台电脑, CPU 有 2 种选择 ($C = \{\text{Intel}, \text{AMD}\}$), 内存有 3 种选择 ($R = \{8\text{GB}, 16\text{GB}, 32\text{GB}\}$), 硬盘有 2 种选择 ($H = \{1\text{TB SSD}, 2\text{TB SSD}\}$), 并且显卡也有 3 种选择 ($G = \{\text{GPU-A}, \text{GPU-B}, \text{GPU-C}\}$). 一种完整的配置方案由 CPU、内存、硬盘、显卡各选一种构成. 请问有多少种不同的配置方案?

解　每一种配置方案都可以被视为一个有序**四元组** (cpu, ram, hdd, gpu). 所有可能的配置方案集合是这四个选择集合的笛卡尔积 $C \times R \times H \times G$.

根据笛卡尔积的基数公式, 总的方案数是

$$|C \times R \times H \times G| = |C| \times |R| \times |H| \times |G| = 2 \times 3 \times 2 \times 3 = 36$$

所以, 共有 36 种不同的电脑配置方案.

4.1.3 原理的辨析与应用

区分加法原理和乘法原理是计数推理的常见起点. 一个简单的判断方法如下.

- **加法原理**关注分类. 完成任务的各种方式是平行的, 选择其中**任何一种**方式都能完成任务, 对应关键词**或者**.
- **乘法原理**关注分步. 完成任务需要经历**所有**步骤, 一步都不能少, 对应关键词**并且**.

在更复杂的问题中, 这两个原理往往需要结合使用.

例题 4.8　求在所有三位正整数中, 至少含有一个数字 7 的数有多少个?

解　直接计算至少含有一个 7 的数字个数会比较复杂. 须分类讨论恰好含有一个 7、恰好含有两个 7、恰好含有三个 7 三种情形, 思路可行却繁琐易错. 换一个角度往往能化繁为简.

不妨先计算它的**反面**, 即一个 7 都没有的三位正整数有多少个.

先看三位正整数的总数. 这是一个分步过程. 依次确定百位、十位、个位. 百位不能为 0, 有 9 种选择 (1 到 9); 十位和个位可取 0 到 9 中任意数字, 各有 10 种选择. 由乘法原理, 共有

$$9 \times 10 \times 10 = 900$$

个三位正整数.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

表 4.1 加法原理与乘法原理的比较

比较维度	加法原理	乘法原理
要领	**分类讨论**. 完成任务有若干类彼此独立的方法.	**分步完成**. 完成任务需要经过若干个连续进行的步骤.
关键词	或者、任选其一、情况一、情况二……、分为几类	并且、依次、接着、每一个……都……
任务关系	各类方法**相互排斥**. 选择任何一类中的一种方法, 任务即告完成.	各个步骤**相互依存**. 必须完成所有步骤, 任务才算完成.
判断准则	完成任务的方式是**择一即可**.	完成任务的步骤是**缺一不可**.
集合模型	**不交并集** $\lvert A_1 \cup \cdots \cup A_k\rvert = \sum_{i=1}^{k} \lvert A_i\rvert$, 其中 $A_i \cap A_j = \emptyset$ 对所有 $i \neq j$.	**笛卡尔积** $\lvert A_1 \times \cdots \times A_k\rvert = \prod_{i=1}^{k} \lvert A_i\rvert$.
经典比喻	从北京到上海, 可以坐火车**或**飞机.	搭配一套服装, 需要选上衣**和**裤子.

再算**不含数字 7** 的三位数. 同样分步. 百位不能是 0 且不能是 7, 有 8 种选择 (1,2,3,4,5,6,8,9); 十位和个位都不能是 7, 各有 9 种选择 (0,1,2,3,4,5,6,8,9). 由乘法原理, 不含数字 7 的三位数共有

$$8 \times 9 \times 9 = 648$$

个.

所有三位数可分成互斥的两类, 即至少含有一个 7 的和一个 7 也没有的, 二者之和即全部三位数. 由加法原理的变形, 也称**补集原理**, 是**容斥原理**的最简形式得

$$(\text{至少含有一个 7 的个数}) = (\text{总个数}) - (\text{一个 7 也没有的个数}),$$

$$N = 900 - 648 = 252,$$

即至少含有一个数字 7 的三位正整数共有 252 个.

从问题的反面入手,可以避开复杂的分类讨论. 上面的例子把一次加法与乘法的混合应用, 转化为两次乘法应用和一次减法. 这种策略在组合计数中相当实用.

接着看几个关于乘法原理的应用.

例题 4.9 令集合 $A = \{a_1, a_2, a_3\}$ 且 $B = \{1, 2, 3, 4, 5\}$. 从集合 A 到集合 B 可以定义多少个不同的函数 $f : A \to B$?

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 定义一个函数 $f: A \to B$, 就是为 A 中的**每一个**元素在 B 中指定**唯一一个**与之对应的元素, 称为它的像.A 有三个元素 a_1, a_2, a_3, 因此这项任务可分解为三个连续的决策. 先定 $f(a_1)$, 再定 $f(a_2)$, 最后定 $f(a_3)$.

这三步彼此独立.$f(a_1)$ 可取 B 中任意一个元素, 有 5 种选择; 同样 $f(a_2)$ 与 $f(a_3)$ 也各有 5 种选择. 由乘法原理, 方法数是各步选择数之积

$$N = 5 \times 5 \times 5 = 5^3 = 125,$$

故总共可定义 125 个不同的函数.

此例说明了一个普遍结论. 从一个包含 m 个元素的集合到一个包含 n 个元素的集合, 可以定义 n^m 个不同的函数.

例题 4.10 在一场田径比赛的决赛中, 共有 8 名选手. 假设没有并列名次, 那么总共有多少种可能的最终排名?

解 确定最终排名, 等价于为第 1 名到第 8 名这 8 个位置依次安排一位选手, 这是一个典型的分步过程.

第 1 名可从 8 位选手中产生, 有 8 种选择; 第 1 名一经确定, 第 2 名只能从剩下 7 位中产生, 有 7 种选择; 以此类推, 第 i 名有 $9-i$ 种选择, 直至第 8 名只剩最后一位选手, 仅 1 种选择. 由乘法原理, 所有可能的排名总数为

$$N = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40320.$$

这个连乘积用专门的符号**阶乘**表示, 记作 8!.

将 n 个不同的物体进行全排列, 其方法数恰好是 $n! = n \times (n-1) \times \cdots \times 1$. 阶乘是乘法原理最直接的推论之一, 也是组合数与排列数计算中反复出现的工具.

例题 4.11 求正整数 $N = 720$ 的所有正因数的个数.

解 逐一列举 720 的因数既繁琐又易遗漏, 宜从其素数分解的结构入手. 由算术基本定理先作素数分解

$$720 = 72 \times 10 = (8 \times 9) \times (2 \times 5) = (2^3 \times 3^2) \times (2 \times 5) = 2^4 \cdot 3^2 \cdot 5^1,$$

故 720 的任一正因数 d 必具 $2^a \cdot 3^b \cdot 5^c$ 的形式, 其中 $0 \le a \le 4$、$0 \le b \le 2$、$0 \le c \le 1$.

构造一个因数, 等价于依次确定三个指数 a, b, c 的取值.a 可取 $0, 1, 2, 3, 4$ 中任一值, 共 $4+1=5$ 种选择;b 可取 $0, 1, 2$ 中任一值, 共 $2+1=3$ 种选择;c 可取 $0, 1$ 中任一值, 共 $1+1=2$ 种选择. 三次决策相互独立, 由乘法原理得正因数的总个数为

$$\tau(720) = (4+1) \times (2+1) \times (1+1) = 5 \times 3 \times 2 = 30,$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

即 720 共有 30 个正因数.

这个方法具有普适性. 对于任意正整数 $N = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$, 其正因数的个数为 $\tau(N) = (a_1+1)(a_2+1)\dots(a_k+1)$, 这也是乘法原理在数论中的一项应用.

例题 4.12 (路径着色) 有 4 个排成一排的方格, 现在用 5 种不同的颜色对其进行染色, 要求相邻的方格颜色不能相同. 问有多少种不同的染色方法?

解 将 4 个方格从左到右依次编号为 1, 2、3、4, 染色是一个分步决策过程, 每一步只需避开左邻一格的颜色.

第 1 格无任何限制, 可从 5 种颜色中任选一种, 有 5 种方法. 此后第 i 格 ($i = 2,3,4$) 只需与第 $i-1$ 格不同色, 无论第 $i-1$ 格具体选了哪种颜色, 都恰好排除一种可能, 故各有 $5-1=4$ 种选择. 由乘法原理, 总的染色方法数为

$$N = 5 \times 4 \times 4 \times 4 = 5 \times 4^3 = 320,$$

即共有 320 种不同的染色方法.

这个问题的要点在于, 每一步选择的**数量**是固定的, 不受前一步**具体选择**的影响, 例如, 无论第 1 格是红色还是蓝色, 第 2 格都剩下 4 种选择. 如果问题的限制条件更为复杂, 例如首尾方格颜色不同, 则可能需要更精细的分类讨论, 即加法与乘法原理的结合.

例题 4.13 (集合的子集个数) 一个包含 n 个不同元素的有限集合 S, 共有多少个不同的子集?(子集包括空集 $\emptyset$ 和集合 S 本身)

解 此题有一个著名答案, 其证法之一恰巧运用乘法原理, 要点在于转变思考角度.

传统想法是挑选元素组成子集. 挑选 0 个元素的子集, 挑选 1 个元素的子集, ..., 再用加法原理求和. 这条路可行但更繁, 第4.2节 会用组合数给出另一种算法.

换一个角度. 对集合 S 中的**每一个**元素, 逐一决定它的归属. 设 $S = \{s_1, s_2, \dots, s_n\}$, 为构造一个子集 $A \subseteq S$, 依次对每个元素决策其是否属于 A. 对 s_1 有属于或不属于两种选择, 对 s_2 亦然, 且这一决策与对 s_1 的决策无关; 如此继续, 直至对 s_n 仍有两种选择. 当 n 个决策全部完成, 一个确定的子集就被唯一地构造出来. 若对所有元素都判定不属于, 得到空集 $\emptyset$; 若都判定属于, 则得到 S 本身.

这 n 个决策相互独立, 各有 2 种选择, 由乘法原理得总方法数

$$N = \underbrace{2 \times 2 \times \cdots \times 2}_{n\text{次}} = 2^n,$$

故一个包含 n 个元素的集合, 其子集的个数为 2^n.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这个证明展示了一种常用的组合技巧. 把从整体中挑选一部分的观点, 改为对每个个体进行独立决策, 从而把一个分类求和问题转化为一个连续决策问题. 这种视角转换在解决高级计数问题时相当常用.

4.2 组合与排列

在上一节, 我们用加法原理和乘法原理处理了计数问题的两类基本情形. 本节用这两条原理定义并计算两个对象, 即**组合**与**排列**.

回到本章开头提出的彩票问题.

从 33 个红球中不重复地选出 6 个, 一共有多少种选法?

这里我们只关心**选出了哪些**号码,不在乎它们的**先后顺序**. 选出集合 $\{5,8,13,21,27,32\}$ 与选出集合 $\{32,27,21,13,8,5\}$ 是完全等价的. 这种只选不排的问题, 正是我们要研究的对象.

4.2.1 组合

定义 4.2 (组合) 从 n 个不同的元素中, **不计顺序**地选取 k 个元素组成一个子集, 称为一个从 n 个元素中取出 k 个元素的**组合**. 所有不同组合的个数, 称为组合数, 记作 $C(n,k)$ 或更常用的 $\binom{n}{k}$(读作"从 n 中选 k 个").

关于组合数的记号

组合数在不同的教材和领域中有多种常见的记法, 例如 $C(n,k)$, C_n^k, 甚至 C_k^n. 它们都表示同一个数学概念.

在本书中, 我们主要采用国际通用的**二项式系数**记法

$$\binom{n}{k}$$

这个符号读作"从 n 中选 k 个", 它与常见记法满足如下恒等关系

$$\binom{n}{k} \equiv C(n,k) \equiv C_n^k$$

直接计数 $\binom{n}{k}$ 并不容易, 但可以借助乘法原理, 用一个思想实验推导出来. 以彩票问题为例, 计算 $\binom{33}{6}$.

思想实验从顺序重要的情形入手. 设想从袋子里一个接一个地、**按顺序**取出 6 个球. 根据乘法原理逐步计数.

- 第一个球有 33 种选择;

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 第二个球有 32 种选择;
- …
- 第六个球有 $33-6+1=28$ 种选择.

所以, 有序的选法总数为 $33\times32\times31\times30\times29\times28$.

现在回到顺序不重要的现实. 考虑一个具体的组合, 例如集合 $\{1,2,3,4,5,6\}$. 在刚才的有序计数中, 这个组合对应于这 6 个数字的全部全排列. $(1,2,3,4,5,6)$、$(1,2,3,4,6,5)$、…、$(6,5,4,3,2,1)$. 这 6 个数字的全排列数, 根据乘法原理, 是 $6\times5\times4\times3\times2\times1=6!$(6 的阶乘).

算出的有序选法总数, 把每一个真正想要的无序组合都重复计数了 $6!$ 次. 因此, 组合数等于有序选法总数除以这个重复倍数.

$$\binom{33}{6}=\frac{33\times32\times31\times30\times29\times28}{6\times5\times4\times3\times2\times1}$$

知乎@余命数

这个逻辑可以推广到一般情况. 从 n 个元素中按顺序选出 k 个, 有 $n(n-1)\cdots(n-k+1)$ 种方法. 而这 k 个被选出的元素自身的全排列数为 $k!$. 每一个组合都对应着 $k!$ 个有序的选法. 因此, 得组合数的计算公式.

知乎@余命数

定理 4.3 (组合数公式) 从 n 个不同元素中取出 k 个元素的组合数 $(0\le k\le n)$ 为

$$\binom{n}{k}=\frac{n(n-1)\cdots(n-k+1)}{k!}=\frac{n!}{k!(n-k)!}$$

其中 $n!$ 表示 n 的阶乘, $n!=n\times(n-1)\times\cdots\times1$. 特别地, 我们定义 $0!=1$.

知乎@余命数

证明 用双重计数. 对同一计数问题以两种方式计算, 所得必等, 由此解出未知量.

所计对象是从含 n 个不同元素的集合 S 中选出 k 个元素并按顺序排列的方法总数, 暂记为 $A(n,k)$, 即下一小节将定义的**排列数**.

法一直接用乘法原理. 构造一个 k-排列即填充 k 个有序位置. 第一位有 n 种选法, 第二位有 $n-1$ 种, …, 第 k 位有 $n-k+1$ 种. 故

$$A(n,k)=n(n-1)\cdots(n-k+1).$$

分子分母同乘 $(n-k)!$, 写成阶乘形式

$$A(n,k)=\frac{n!}{(n-k)!}.$$

知乎@余命数

知乎@余命数

法二经组合数间接计算. 构造 k-排列可分两步. 先从 n 个元素中不计顺序选出 k 个, 有 $\binom{n}{k}$ 种; 再对这 k 个作全排列, 有 $k!$ 种. 由乘法原理,

$$A(n,k)=\binom{n}{k}\times k!.$$

两式相等, 得

$$\binom{n}{k}\times k!=\frac{n!}{(n-k)!}.$$

两边同除以 $k!$($k\geq 0$ 时 $k!\geq 1$), 即

$$\binom{n}{k}=\frac{n!}{k!(n-k)!}.$$

若改用 $A(n,k)=n(n-1)\cdots(n-k+1)$, 则得常用形式

$$\binom{n}{k}=\frac{n(n-1)\cdots(n-k+1)}{k!}.$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

特别地, $k=0$ 时选出空集仅一种方法, 公式给出 $\binom{n}{0}=\frac{n!}{0!\,n!}=1$, 与事实相符, 亦说明定义 $0!=1$ 的合理性. □

现在可以回答彩票问题了.

$$\binom{33}{6}=\frac{33\times 32\times 31\times 30\times 29\times 28}{6\times 5\times 4\times 3\times 2\times 1}=1,107,568$$

挑选红球的方法数超过了一百万种, 而这只是中头奖的第一步. 完成整张彩票, 还要从 16 个蓝球中选出 1 个, 有 $\binom{16}{1}=16$ 种选法. 根据乘法原理, 全部可能的号码组合共有

知乎@余命数

$$\binom{33}{6}\times\binom{16}{1}=1,107,568\times 16=17,721,088$$

知乎@余命数

种. 全部组合超过一千七百万种, 而其中只有一种是当期的头奖号码. 至于中头奖的概率应当如何严格定义与计算, 我们还将在 第5章 中回到这个问题.

在结束组合的讨论之前, 再补充组合数的一个重要性质, 它能简化计算.

定理 4.4 (组合数的对称性) 对于所有满足 $0\leq k\leq n$ 的整数 n,k, 有

$$\binom{n}{k}=\binom{n}{n-k}$$

知乎@

证明 代数上, 由组合数公式直接可得

$$\binom{n}{n-k} = \frac{n!}{(n-k)!\,(n-(n-k))!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{k}$$

这一等式的组合意义同样清楚. 从 n 个元素中选出 k 个带走, 与选出 $n-k$ 个留下, 是同一个决定的两种说法, 两种计数必然相等. □

这个性质的实际价值在于化大为小. 当 k 接近 n 时, 直接计算 $\binom{n}{k}$ 的分子分母都很长, 转而计算 $\binom{n}{n-k}$ 就轻快得多. 例如

$$\binom{100}{98} = \binom{100}{2} = \frac{100\times 99}{2\times 1} = 4950$$

4.2.2 排列

理解了组合之后, 排列便很自然. 排列关心选出了哪些元素, 并关心这些元素被排成的**顺序**.

定义 4.3 (排列) 从 n 个不同的元素中, 选取 k 个元素并按照**一定的顺序**排成一列, 称为一个从 n 个元素中取出 k 个元素的**排列**. 所有不同排列的个数, 称为排列数. 排列数的常见记号有 $A(n,k)$、A_n^k、$P(n,k)$、P_n^k 等; 与组合数的情形类似, 本书统一使用 $A(n,k)$.

计算排列数 $A(n,k)$, 把完成一次排列看作两个步骤.

1. 先从 n 个元素中, 选出要参与排列的 k 个元素. 这一步有 $\binom{n}{k}$ 种方法.
2. 将选出的 k 个元素进行全排列. 这一步有 $k!$ 种方法.

必须完成**所有**步骤, 任务才算完成, 由乘法原理得到排列数与组合数之间的基本关系

$$A(n,k) = \binom{n}{k}\times k!$$

这个关系说明了排列与组合的联系. 一个排列, 实质上是一个组合再加上对所选元素的排序.

将组合数的公式代入, 我们就能得到排列数的计算公式.

定理 4.5 (排列数公式) 从 n 个不同元素中取出 k 个元素的排列数 $(0\le k\le n)$ 为

$$A(n,k) = \frac{n!}{k!(n-k)!}\times k! = \frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1)$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

证明 目标是计算从 n 个不同元素中，按顺序取出 k 个元素的所有方法的总数，即 $A(n,k)$. 把构造这样一个排列的过程视为一个分步完成的任务，直接应用乘法原理.

设想有 k 个空置的位置，需从一个含 n 个元素的集合 S 中选元素填充，且每个元素最多用一次.

填充位置 1 时，可从集合 S 中任选一个元素，有 n 种选择. 填充位置 2 时，集合 S 中只剩下 $n-1$ 个元素可用，故有 $n-1$ 种选择. 填充位置 3 时，已用去两个元素，剩下 $n-2$ 个元素可用，故有 $n-2$ 种选择. 依此类推，填充第 k 个（也是最后一个）位置时，已用去 $k-1$ 个元素，剩下 $n-(k-1)=n-k+1$ 个元素可用，故有 $n-k+1$ 种选择.

根据乘法原理，完成这 k 个步骤的总方法数是每一步方法数的乘积. 因此，

知乎@余命数

$$A(n,k)=n\times(n-1)\times(n-2)\times\cdots\times(n-k+1)$$

知乎@余命数

这证明了公式的第一个等式.

为得到阶乘形式的表达式，对上式做代数变换，在分子和分母上同时乘以 $(n-k)!$.

$$A(n,k)=n(n-1)\cdots(n-k+1)\times\frac{(n-k)(n-k-1)\cdots1}{(n-k)(n-k-1)\cdots1}$$

分子部分现在是一个从 n 开始到 1 的完整连乘积，即 $n!$. 分母部分则是 $(n-k)!$. 所以，

$$A(n,k)=\frac{n!}{(n-k)!}$$

这证明了公式的第二个等式.

最后，由上一节推导出的排列与组合的关系 $A(n,k)=\binom{n}{k}\times k!$ 验证一致性.

$$\binom{n}{k}\times k!=\frac{n!}{k!(n-k)!}\times k!=\frac{n!}{(n-k)!}=A(n,k)$$

知乎@余命数

这既验证了公式，也再次说明一个 k-排列可以看作先进行一次 k-组合，再对选出的元素进行一次全排列. □

知乎@余命数

例题 4.14 一个班级有 10 名学生，需要从中选出 3 人，分别担任班长、学习委员和体育委员. 有多少种不同的选举结果?

解 此题是典型的排列问题，因为职位不同，人选的顺序至关重要. 例如，张三任班长、李四任学委，与李四任班长、张三任学委，是两种不同的结果.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

直接套用排列数公式, 即从 10 人中选出 3 人排列, 得

$$A(10,3)=\frac{10!}{(10-3)!}=\frac{10!}{7!}=10\times 9\times 8=720.$$

也可走先组合、后排序的路径. 先从 10 人中不计顺序地选出 3 人, 有 $\binom{10}{3}=\dfrac{10\times 9\times 8}{3\times 2\times 1}=120$ 种方法; 再为这 3 人分配 3 个不同职位, 有 $3!=6$ 种方法. 由乘法原理, 总的选举结果为

$$N=\binom{10}{3}\times 3!=120\times 6=720,$$

与直接使用排列数公式的结果一致, 正说明排列与组合的联系.

4.3 二项式定理

我们在上一节定义了组合数 $\binom{n}{k}$, 即 n 个不同元素中选取 k 个的方案数. 这个符号也被称为**二项式系数**, 与形如 $(x+y)$ 的二项式有内在联系. 本节说明这一联系, 它集中体现在二项式定理上.

从简单的例子开始, 观察展开式的系数规律.

$$\begin{aligned}(x+y)^0&=1\\(x+y)^1&=1x+1y\\(x+y)^2&=1x^2+2xy+1y^2\\(x+y)^3&=1x^3+3x^2y+3xy^2+1y^3\\(x+y)^4&=1x^4+4x^3y+6x^2y^2+4xy^3+1y^4\end{aligned}$$

观察这些展开式, 系数 $1,4,6,4,1$ 恰好是 $\binom{4}{0},\binom{4}{1},\binom{4}{2},\binom{4}{3},\binom{4}{4}$ 的值. 这并非巧合.

定理 4.6 (二项式定理) 对于任意可交换的变量或数 x,y 和任意非负整数 n, 下述展开式成立.

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$$

展开来看, 即

$$(x+y)^n=\binom{n}{0}x^n+\binom{n}{1}x^{n-1}y+\binom{n}{2}x^{n-2}y^2+\cdots+\binom{n}{n-1}xy^{n-1}+\binom{n}{n}y^n$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

证明 考虑乘积的完全展开形式

$$(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n\text{个因子}}$$

根据乘法分配律, 展开后的每一项都是通过从 n 个 $(x+y)$ 因子中, 各自选择 x 或 y 相乘得到的. 例如, 要得到 x^n 这一项, 须从每一个因子中都选择 x.

现在考虑一个更普遍的项, 形如 $x^{n-k}y^k$ 是如何产生的. 为构成这一项, 必须从 n 个 $(x+y)$ 因子中.

- 恰好从 k 个因子中选择 y.
- 从其余的 $n-k$ 个因子中选择 x.

每一种这样的选择方案都贡献一个 $x^{n-k}y^k$ 项. 因此, 展开式中 $x^{n-k}y^k$ 的系数, 等于构成该项的不同选择方案的总数.

这个选择方案的总数是一个熟悉的问题. 它等价于从 n 个不同的因子中, 选出 k 个来提供 y 的方案数.

根据组合的定义, 这个方案数恰好是 $\binom{n}{k}$.

因此, $x^{n-k}y^k$ 项的系数必然是 $\binom{n}{k}$.

k 可从 0(一个 y 都不选, 得到 x^n)取到 n, 全部选 y 时则得到 y^n, 把所有这些项加起来, 就得到二项式定理的完整形式

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}y^k$$

□

这个证明解释了组合数为何会作为系数出现在二项式的展开式中.

4.3.1 杨辉三角

将二项式系数 $\binom{n}{k}$ 排列成一个三角形的阵列, 我们就得到了著名的杨辉三角.

杨辉三角最引人注目的性质是, **每个数都等于它肩上两个数之和**. 例如, 在 $n=4$ 行, 6 是由它上一行 $(n=3)$ 中的 3 和 3 相加得到的. 这个性质在数学上称为**帕斯卡恒等式**.

定理 4.7 (帕斯卡恒等式) 对于所有满足 $n \geq 1$ 且 $1 \leq k \leq n-1$ 的整数 n, k, 有

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

代数代入组合数公式再通分即可验证这个恒等式. 下面给出一个组合学证明.

证明 考虑一个组合问题. 从 n 个人中选出 k 人组成班委会, 方法数是 $\binom{n}{k}$.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

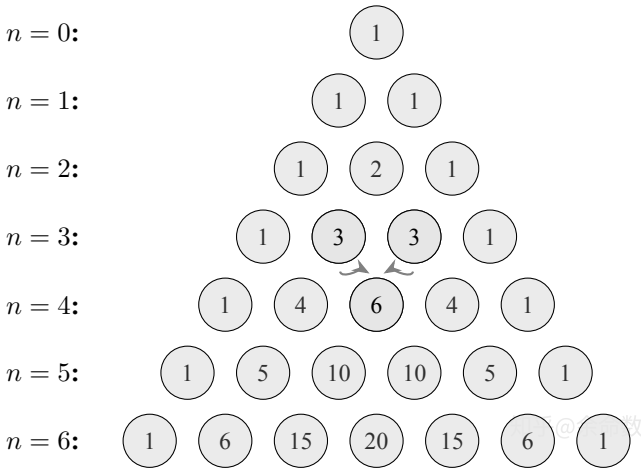


图 4.2 注意恒等式 $\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}$ 的一个例子, $\binom{4}{2}=\binom{3}{1}+\binom{3}{2}$

换一种方式计数. 指定一个特殊的人李华, 最终选出的委员会只有两种互斥情况.

- 李华在班委中. 此时只需从剩下 $n-1$ 个人中再选出 $k-1$ 个人填补剩余名额, 方法数是 $\binom{n-1}{k-1}$.
- 李华不在班委中. 此时必须从剩下 $n-1$ 个人中选出全部 k 个班委成员, 方法数是 $\binom{n-1}{k}$.

由加法原理, 选班委的总方法数等于上述两种互斥情况的方法数之和. 故

$$\binom{n}{k}=\binom{n-1}{k-1}+\binom{n-1}{k}$$

□

4.3.2 二项式系数的性质

再来观察杨辉三角, 可从中猜测出二项式系数的一系列规律. 每一行从左读到右与从右读到左完全相同; 每一行的数先增大、后减小, 在正中间达到最大; 每一行所有数的和恰好是上一行的两倍. 这些从数值表中归纳出的猜想, 都可以借助组合数公式和二项式定理严格证明.

首先是**对称性**, $\binom{n}{k}=\binom{n}{n-k}$, 这是杨辉三角每一行左右镜像的原因. 我们已在第4.2节 中证明这一性质; 在二项式定理中交换 x 与 y 的位置, 展开式也保持不变.

接下来看每一行的和所对应的规律. 证明它的关键技巧在于, 二项式定理是一个关于 x,y 的**恒等式**, 对其中的变量赋予特殊的值, 就能得到各种各样的组合恒等式. 这种方法称为**赋值法**.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定理 4.8 (二项式系数之和) 对于任意非负整数 n, 有

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n$$

知乎@余命数

证明 在二项式定理 $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$ 中令 $x=y=1$,左边变为 $(1+1)^n=2^n$, 右边恰为全部二项式系数之和. □

这个等式其实是已知事实的另一种表述. 在 第4.1节 中, 我们曾用乘法原理证明, 一个 n 元集合的子集共有 2^n 个. 另一方面, 按子集的大小分类计数, 大小为 k 的子集有 $\binom{n}{k}$ 个, 对 k 从 0 到 n 求和同样数出了全部子集. 两种方法数的是同一个对象, 结果自然相等. 这是推导组合数公式时用过的双重计数方法.

知乎@余命数

定理 4.9 (二项式系数的交错和) 对于任意正整数 n, 有

$$\binom{n}{0}-\binom{n}{1}+\binom{n}{2}-\cdots+(-1)^n\binom{n}{n}=0$$

从而奇数项二项式系数之和等于偶数项二项式系数之和, 二者都等于 2^{n-1}.

证明 在二项式定理中令 $x=1$, $y=-1$. 由于 $n\geq 1$, 左边为 $(1+(-1))^n=0$, 右边恰为交错和 $\sum_{k=0}^{n}(-1)^k\binom{n}{k}$, 故交错和为 0. 把所有偶数项系数留在一边、奇数项系数移到另一边, 即得奇数项之和等于偶数项之和; 又由二项式系数之和为 2^n, 所以每一方都等于 $2^n/2=2^{n-1}$. □

知乎@余命数　知乎@余命数

最后是先增后减、中间最大的规律, 这在数学上称为**单峰性**.

定理 4.10 (二项式系数的单峰性) 对于固定的正整数 n, 二项式系数序列

$$\binom{n}{0},\binom{n}{1},\ldots,\binom{n}{n}$$

先严格递增, 再严格递减. 当 n 为偶数时, 最大值在正中间一项 $\binom{n}{n/2}$ 处取得; 当 n 为奇数时, 中间两项 $\binom{n}{(n-1)/2}$ 与 $\binom{n}{(n+1)/2}$ 相等且同为最大值.

证明 考察相邻两项的比值. 对于 $1\leq k\leq n$,

知乎@余命数

$$\frac{\binom{n}{k}}{\binom{n}{k-1}}=\frac{n!}{k!(n-k)!}\cdot\frac{(k-1)!\,(n-k+1)!}{n!}=\frac{n-k+1}{k}$$

知乎@余命数

这个比值大于 1 当且仅当 $n-k+1>k$, 即 $k<\dfrac{n+1}{2}$. 因此

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 当 $k < \dfrac{n+1}{2}$ 时, $\binom{n}{k} > \binom{n}{k-1}$, 序列严格递增;
- 当 $k > \dfrac{n+1}{2}$ 时, $\binom{n}{k} < \binom{n}{k-1}$, 序列严格递减.

若 n 为偶数, 则 $\dfrac{n+1}{2}$ 不是整数, 序列在 $k = \dfrac{n}{2}$ 处取得唯一的最大值 $\binom{n}{n/2}$. 若 n 为奇数, 则当 $k = \dfrac{n+1}{2}$ 时比值恰为 1, 即中间两项相等, 它们同时成为最大值 $\binom{n}{(n-1)/2} = \binom{n}{(n+1)/2}$. 这也再次印证了对称性. □

单峰性说明, $(x+y)^n$ 展开式中最中间的项系数最大. 例如 $(x+y)^{10}$ 的展开式中, 系数最大的项是 $\binom{10}{5}x^5y^5 = 252x^5y^5$. 这一性质在二项分布等后续内容中还会反复出现.

4.3.3 杨辉三角的奇偶图案 *(选读)

把杨辉三角中每个二项式系数按奇偶涂色, 奇数涂黑, 偶数留白, 得到的图形会向自己相似, 每往下一层, 图案都由三个上一层的图案拼成, 中间留出一块空白. 本节数清每一行里究竟有多少个奇数.

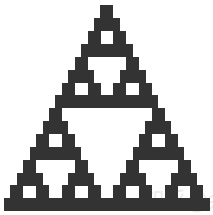


图 4.3 杨辉三角第 0 到第 15 行的奇偶图案, 黑格表示奇数. 图案自相似, 每一层由三个上一层图案拼成, 中间留空.

先看指数是 2 的幂的那些行. 在模 2 的意义下计算 $(1+x)^{2^k}$, 用平方代替乘法.

$$(1+x)^{2^k} = \left((1+x)^{2^{k-1}}\right)^2 \equiv \left(1+x^{2^{k-1}}\right)^2 = 1 + 2x^{2^{k-1}} + x^{2^k} \equiv 1 + x^{2^k} \pmod 2.$$

最右边只有两项, 中间各项的系数都是偶数. 反复使用这一步, 从 $k=0$ 推起, 可知第 2^k 行里只有两端的两个 1 是奇数.

再看一般的行. 记 n 的二进制写成 $n = 2^{k_1} + 2^{k_2} + \cdots + 2^{k_s}$, 其中 $k_1 > k_2 > \cdots > k_s \geq 0$, s 是 n 的二进制中 1 的个数. 由上一段的结论,

$$(1+x)^n = \prod_{i=1}^{s} (1+x)^{2^{k_i}} \equiv \prod_{i=1}^{s} \left(1 + x^{2^{k_i}}\right) \pmod 2.$$

右边每个括号贡献一次选择, 要么取 1, 要么取 $x^{2^{k_i}}$. 由于各个 2^{k_i} 互不相同, 不同的选择得到的指数互不相同, 展开后不会有两项合并. 于是右边展开后恰好有 2^s 项系数为 1, 其余项系数为 0.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定理 4.11 (一行的奇数个数) 设 n 的二进制表示中 1 的个数为 s, 则杨辉三角第 n 行中奇数的个数等于 2^s. 特别地, 第 2^k-1 行的 2^k 个数全是奇数.

第 2^k-1 行的二进制的 k 位全是 1, 故 $s=k$, 奇数个数为 2^k, 恰好等于该行的项数, 整行都是奇数. 第 2^k 行则恰好转过一轮, 只剩两端两个奇数, 中间全为偶数.

例题 4.15 求杨辉三角第 100 行中奇数的个数.

解 把 100 写成二进制

$$100=64+32+4=1100100_{(2)},$$

其中 1 的个数 $s=3$, 故第 100 行中奇数有 $2^3=8$ 个.

奇数的占比还可以整体估算. 在第 0 到第 2^k-1 行中, 二进制含 j 个 1 的行共有 $\binom{k}{j}$ 行, 每行的奇数个数为 2^j, 于是这些行里的奇数总数为

$$\sum_{j=0}^{k}\binom{k}{j}2^j=3^k.$$

而这 2^k 行的项数共为 $1+2+\cdots+2^k=2^k\left(2^k+1\right)/2$. 两者相除得

$$\frac{3^k}{2^k\left(2^k+1\right)/2}=\frac{2\cdot 3^k}{2^k\left(2^k+1\right)},$$

分母中的 2^{2k} 增长更快, 比值的极限为 0. 行数越多, 奇数所占的比例越接近于零, 杨辉三角里几乎所有的数都是偶数.

4.4 从三项式定理到多项式定理 *

一旦得出二项式定理, 自然会考虑是否存在类似的三项式定理、四项式定理 …. 沿用证明二项式定理时的组合思想来分析 $(x+y+z)^n$ 的展开式.

$$(x+y+z)^n=\underbrace{(x+y+z)(x+y+z)\cdots(x+y+z)}_{n\text{个因子}}$$

展开后的任意一项必然具有 $x^{k_1}y^{k_2}z^{k_3}$ 的形式. 由于从 n 个因子中各取一个变量相乘, 这些变量的指数总和必须等于 n, 即

$$k_1+k_2+k_3=n\quad(\text{其中}k_1,k_2,k_3\text{为非负整数})$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

接下来确定项 $x^{k_1}y^{k_2}z^{k_3}$ 的系数. 这个系数等于从 n 个因子中, 选出 k_1 个提供 x、选出 k_2 个提供 y、再选出 k_3 个提供 z 的总方案数. 可以分步计数来解决.

1. 为 x 选择因子. 从 n 个可用的因子中, 选出 k_1 个. 这有 $\binom{n}{k_1}$ 种方法.
2. 为 y 选择因子. 从剩下的 $n-k_1$ 个因子中, 选出 k_2 个. 这有 $\binom{n-k_1}{k_2}$ 种方法.
3. 为 z 选择因子. 最后剩下的 $n-k_1-k_2=k_3$ 个因子, 全部用于提供 z. 这只有 $\binom{k_3}{k_3}=1$ 种方法.

根据乘法原理, 总方案数为

$$\begin{aligned}\binom{n}{k_1}\binom{n-k_1}{k_2} &= \frac{n!}{k_1!(n-k_1)!}\cdot\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\\ &= \frac{n!}{k_1!(n-k_1)!}\cdot\frac{(n-k_1)!}{k_2!k_3!}\\ &= \frac{n!}{k_1!k_2!k_3!}\end{aligned}$$

知乎@余命数

这个结果引出了三项式系数的定义.

知乎@余命数

定义 4.4 (三项式系数) 对于满足 $k_1+k_2+k_3=n$ 的非负整数 k_1,k_2,k_3, 三项式系数被定义为

$$\binom{n}{k_1,k_2,k_3}=\frac{n!}{k_1!k_2!k_3!}$$

分组问题

三项式系数 $\binom{n}{k_1,k_2,k_3}$ 有一个非常直观的组合解释. 它表示将 n 个不同的物体分成三组, 各组物体的数量分别为 k_1,k_2,k_3 的方法数. 这与上面的推导过程(将 n 个因子分成三组)完全一致. 它也等价于对包含 k_1 个 x、k_2 个 y 和 k_3 个 z 的多重集进行全排列的数目.

4.4.1 多项式定理

有了三项式系数的定义, 可将二项式定理正式推广到三项式.

知乎@余命数

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定理 4.12 (三项式定理) 对于任意可交换的变量或数 x, y, z 和任意非负整数 n, 下述展开式成立.

$$(x+y+z)^n = \sum_{k_1+k_2+k_3=n} \binom{n}{k_1, k_2, k_3} x^{k_1} y^{k_2} z^{k_3}$$

其中求和遍历所有满足条件的非负整数解 (k_1, k_2, k_3).

在证明之前, 先弄清三项式定理中求和符号 $\sum_{k_1+k_2+k_3=n}$ 的含义.

这个记号表示, 需要对所有满足以下两个条件的有序整数三元组 (k_1, k_2, k_3) 进行求和.

1. k_1, k_2, k_3 均为**非负整数**, 即 $k_1 \geq 0, k_2 \geq 0, k_3 \geq 0$.
2. 它们的和恰好等于 n, 即 $k_1 + k_2 + k_3 = n$.

这个过程也被称为"遍历 n 的所有整数分拆到三个部分".

以具体的例子 $n = 2$ 来阐明. 需要找出所有满足 $k_1 + k_2 + k_3 = 2$ 的非负整数解 (k_1, k_2, k_3), 系统地列出它们.

- 如果 $k_1 = 2$, 那么 $k_2 = 0, k_3 = 0$. 得到 $(2, 0, 0)$.
- 如果 $k_1 = 1$, 那么 $k_2 + k_3 = 1$. 这有两种可能, 即 $(1, 1, 0)$ 和 $(1, 0, 1)$.
- 如果 $k_1 = 0$, 那么 $k_2 + k_3 = 2$. 这有三种可能, 即 $(0, 2, 0), (0, 1, 1), (0, 0, 2)$.

所以, 对于 $n = 2$, 求和符号意味着将对应于以下 6 个三元组的项全部加起来.

$$(2,0,0), (0,2,0), (0,0,2), (1,1,0), (1,0,1), (0,1,1)$$

因此, $(x+y+z)^2$ 的展开式实际上是

$$\begin{aligned}
\sum_{k_1+k_2+k_3=2} \binom{2}{k_1, k_2, k_3} x^{k_1} y^{k_2} z^{k_3} = \\
&\binom{2}{2,0,0} x^2 y^0 z^0 + \binom{2}{0,2,0} x^0 y^2 z^0 + \binom{2}{0,0,2} x^0 y^0 z^2 \\
&+ \binom{2}{1,1,0} x^1 y^1 z^0 + \binom{2}{1,0,1} x^1 y^0 z^1 + \binom{2}{0,1,1} x^0 y^1 z^1 \\
&= 1x^2 + 1y^2 + 1z^2 + 2xy + 2xz + 2yz
\end{aligned}$$

这与直接用分配律展开得到的结果一致.

证明 用组合论证确定展开式中任意一个通项 $x^{k_1} y^{k_2} z^{k_3}$ 的系数, 其中 k_1, k_2, k_3 是满足 $k_1 + k_2 + k_3 = n$ 的非负整数.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

考虑乘积的展开形式

$$(x+y+z)^n = \underbrace{(x+y+z)(x+y+z)\cdots(x+y+z)}_{n\text{个因子}}$$

根据乘法分配律, 展开后的每一项都是从这 n 个 $(x+y+z)$ 因子中的每一个里选择一个变量(x, y 或 z)相乘得到的.

为构成一个具体的项 $x^{k_1}y^{k_2}z^{k_3}$, 必须

- 恰好从 n 个因子中的 k_1 个因子里选择变量 x.
- 恰好从剩下的 $n-k_1$ 个因子中的 k_2 个因子里选择变量 y.
- 恰好从最后剩下的 $n-k_1-k_2=k_3$ 个因子里选择变量 z.

因此, 项 $x^{k_1}y^{k_2}z^{k_3}$ 在最终展开式中出现的次数, 即它的系数, 等于完成上述三步选择的总方案数.

用乘法原理计算这个方案数, 得

1. **为 x 选择因子.** 从 n 个不同的因子中, 选出 k_1 个来提供变量 x. 这是一个组合问题, 方法数是 $\binom{n}{k_1}$.
2. **为 y 选择因子.** 为 x 选好因子后, 还剩下 $n-k_1$ 个因子. 从这些剩余因子中, 选出 k_2 个来提供变量 y. 方法数是 $\binom{n-k_1}{k_2}$.
3. **为 z 选择因子.** 最后剩下 $n-k_1-k_2=k_3$ 个因子. 从这 k_3 个因子中选出 k_3 个来提供变量 z. 方法数是 $\binom{k_3}{k_3}=1$.

根据乘法原理, 总方案数是以上三步方法数的乘积.

$$\begin{aligned}
\text{系数} &= \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{k_3}{k_3} \\
&= \frac{n!}{k_1!(n-k_1)!}\cdot\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdot\frac{k_3!}{k_3!0!} \\
&= \frac{n!}{k_1!\cancel{(n-k_1)!}}\cdot\frac{\cancel{(n-k_1)!}}{k_2!(k_3)!}\cdot 1 \quad (\text{因为} n-k_1-k_2=k_3) \\
&= \frac{n!}{k_1!k_2!k_3!} \\
&= \binom{n}{k_1,k_2,k_3}
\end{aligned}$$

已经证明, 对于任意满足 $k_1+k_2+k_3=n$ 的非负整数组 (k_1,k_2,k_3), 其对应项 $x^{k_1}y^{k_2}z^{k_3}$ 的系数恰好是三项式系数 $\binom{n}{k_1,k_2,k_3}$.

把所有这些可能的项相加, 就构成 $(x+y+z)^n$ 的完整展开式. 定理得证. □

例题 4.16 求在 $(2x-y+3z^2)^5$ 的展开式中, x^2yz^4 项的系数.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 将表达式视为 $(A+B+C)^5$，其中 $A=2x, B=-y, C=3z^2$，所求项 x^2yz^4 应具 $A^{k_1}B^{k_2}C^{k_3}$ 的形式，$k_1+k_2+k_3=5$. 由 $(2x)^{k_1}$ 贡献 x^2 得 $k_1=2$；由 $(-y)^{k_2}$ 贡献 y^1 得 $k_2=1$；由 $(3z^2)^{k_3}$ 贡献 z^4，即 $z^{2k_3}=z^4$，得 $k_3=2$. 指数和 $k_1+k_2+k_3=2+1+2=5$，与 $n=5$ 相符.

据三项式定理，含 $A^2B^1C^2$ 的完整项为 $\binom{5}{2,1,2}A^2B^1C^2$，其中三项式系数

$$\binom{5}{2,1,2}=\frac{5!}{2!\,1!\,2!}=\frac{120}{2\cdot 1\cdot 2}=30.$$

代回 A,B,C 并整理得

$$\begin{aligned}30\cdot(2x)^2\cdot(-y)\cdot(3z^2)^2&=30\cdot(4x^2)\cdot(-y)\cdot(9z^4)\\&=(30\cdot 4\cdot(-1)\cdot 9)\,x^2yz^4\\&=-1080\,x^2yz^4,\end{aligned}$$

故 x^2yz^4 项的系数是 -1080.

这个思想可以自然地推广到任意数量的项，得到更一般的**多项式定理**.

考虑展开式 $(x_1+x_2+\cdots+x_m)^n$. 其通项必然具有 $x_1^{k_1}x_2^{k_2}\cdots x_m^{k_m}$ 的形式，其中指数和为 n，即 $k_1+k_2+\cdots+k_m=n$.

遵循之前的组合论证，该项的系数等于将 n 个不同的因子分成 m 组，并指定第 i 组提供变量 x_i 的方案数，其中第 i 组的大小恰好为 k_i. 这个数称为**多项式系数**. 其组合意义是

将 n 个不同的物体分成 m 个有区别的组，使得第 i 组恰好有 k_i 个物体的总方法数.

与二项式系数的关系

二项式系数是多项式系数的一个特例. 从 n 个元素中选出 k 个，等价于将 n 个元素分成两组. 一组是被选中的 k 个元素，另一组是未被选中的 $n-k$ 个元素. 因此，

$$\binom{n}{k}=\binom{n}{k,n-k}=\frac{n!}{k!(n-k)!}$$

这个视角统一了这些计数工具，它们数的都是把 n 个元素分成若干组的方法数.

有了对这个系数组合意义的理解，我们正式陈述多项式定理. 该定理给出展开式的完整形式，同时给出计算多项式系数的代数公式.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

定理 4.13 (多项式定理) 对于任意非负整数 n、整数 $m \geq 2$ 以及两两可交换的变量或数 $x_1, x_2, \ldots, x_m$, 有

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n} \binom{n}{k_1, k_2, \ldots, k_m} x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}$$

其中**多项式系数**定义为

$$\binom{n}{k_1, k_2, \ldots, k_m} = \frac{n!}{k_1! k_2! \cdots k_m!}$$

知乎@余命数

证明 我们用组合论证. 基本思路是分析代数展开的组合结构, 从而确定各项的系数.

首先, 将 $(x_1 + x_2 + \cdots + x_m)^n$ 写作 n 个相同因子的乘积

$$\underbrace{(x_1 + x_2 + \cdots + x_m)(x_1 + x_2 + \cdots + x_m) \cdots (x_1 + x_2 + \cdots + x_m)}_{n\text{个因子}}$$

根据乘法分配律, 展开后的每一项都是从这 n 个因子中的**每一个**里选择一个变量 $(x_1, x_2, \ldots, \text{或} x_m)$相乘得到的.

考虑展开式中一个**通项**的形式, 它必然是 $x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}$, 其中 $k_1, k_2, \ldots, k_m$ 是非负整数. 由于从 n 个因子中进行选择, 指数的总和必须等于 n, 即

$$k_1 + k_2 + \cdots + k_m = n$$

现在的任务是计算这个通项的系数. 这个系数等于能够产生该项的不同选择方案的总数, 即从 n 个因子中, 选出 k_1 个来提供 x_1, 选出 k_2 个来提供 x_2, …, 并选出 k_m 个来提供 x_m 的总方法数.

这实质上是将 n 个不同的因子(不妨将它们编号为 $1, 2, \ldots, n$)分成 m 个不同组的问题, 第 i 组的大小为 k_i. 用乘法原理, 按变量次序依次选取因子.

1. 从 n 个可用因子中为 x_1 选 k_1 个, 有 $\binom{n}{k_1}$ 种方法;
2. 再从剩下的 $n - k_1$ 个中为 x_2 选 k_2 个, 有 $\binom{n-k_1}{k_2}$ 种方法;
3. 再从剩下的 $n - k_1 - k_2$ 个中为 x_3 选 k_3 个, 有 $\binom{n-k_1-k_2}{k_3}$ 种方法;
4. 依此类推, 直至从最后剩下的 $n - k_1 - \cdots - k_{m-1} = k_m$ 个因子中为 x_m 选 k_m 个, 有 $\binom{k_m}{k_m} = 1$ 种方法.

由**乘法原理**, 总方案数是以上各步方法数的乘积.

$$\text{系数} = \binom{n}{k_1} \binom{n-k_1}{k_2} \binom{n-k_1-k_2}{k_3} \cdots \binom{k_m}{k_m}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

现在将这些二项式系数用阶乘形式展开, 观察化简过程.

$$\begin{aligned}\text{系数} &= \frac{n!}{k_1!(n-k_1)!}\cdot\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\cdot\frac{(n-k_1-k_2)!}{k_3!(n-k_1-k_2-k_3)!}\cdots\frac{k_m!}{k_m!0!}\\ &= \frac{n!}{k_1!\cancel{(n-k_1)!}}\cdot\frac{\cancel{(n-k_1)!}}{k_2!\cancel{(n-k_1-k_2)!}}\cdot\frac{\cancel{(n-k_1-k_2)!}}{k_3!\cdots}\cdots 1\\ &= \frac{n!}{k_1!k_2!k_3!\cdots k_m!}\end{aligned}$$

这是多项式系数 $\binom{n}{k_1,k_2,\cdots,k_m}$ 的定义.

已经证明, 对于任意满足 $k_1+\cdots+k_m=n$ 的非负整数组 $(k_1,\cdots,k_m)$, 其对应项 $x_1^{k_1}\cdots x_m^{k_m}$ 的系数恰好是多项式系数 $\binom{n}{k_1,\cdots,k_m}$.

把所有这些可能的项 (遍历所有满足条件的指数集) 加总, 便构成 $(x_1+\cdots+x_m)^n$ 的完整展开式. 定理得证. □

4.4.2 骰子与多项式 *(选读)

多项式乘法还能给计数问题提供另一套语言. 这一节用它解决一个关于骰子的经典问题, 答案由 Sicherman 给出.

例题 4.17 (点数重新标注的两枚骰子) 普通骰子的六个面分别标着 $1,2,3,4,5,6$. 能否把两枚骰子的点数重新标注成正整数, 允许同一枚骰子上出现相同的点数, 使得两枚骰子都不是普通的骰子, 而掷出的点数之和的各种情况仍与两枚普通骰子逐一相同?

解 把一枚骰子翻译成一个多项式. 若六个面的点数是 $a_1,a_2,\cdots,a_6$, 就记

$$A(x)=x^{a_1}+x^{a_2}+\cdots+x^{a_6}.$$

这个多项式的系数和 $A(1)$ 等于面数 6. 掷两枚骰子得到和为 s 的方案数, 恰好是乘积 $A(x)B(x)$ 中 x^s 的系数, 因为每一项 $x^{a_i}x^{b_j}=x^{a_i+b_j}$ 的指数是两个点数之和.

普通骰子对应

$$D(x)=x+x^2+x^3+x^4+x^5+x^6=x(1+x)(1+x+x^2)(1-x+x^2),$$

末一个等号把右端逐项乘开即可验证. 普通的两枚骰子对应 $D(x)^2$, 于是问题变成, 找出两个系数为 0 或 1、最低次项为 x、系数和都等于 6 的多项式 $A(x)$ 与 $B(x)$, 使

$$A(x)B(x)=D(x)^2.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这是一次多项式的分解. 由上面的分解式

$$D(x)^2 = x^2(1+x)^2\left(1+x+x^2\right)^2\left(1-x+x^2\right)^2,$$

把右端的八个因子分给 A 与 B 两方. 令 $x=1$ 看系数和, 得 $A(1)=B(1)=6$. 因子 $1+x$ 在 $x=1$ 处取 2, 而 6 只含一个因子 2, 所以两枚骰子各得一个 $1+x$. 因子 $1+x+x^2$ 在 $x=1$ 处取 3, 同理两枚骰子各得一个. 又因为两枚骰子的点数都是正整数, 各自的多项式都要被 x 整除, 于是 x^2 也是一方一个.

剩下的因子 $(1-x+x^2)^2$ 有两种分法. 一方一个, 得到 $A=B=x(1+x)(1+x+x^2)(1-x+x^2)=D$, 两枚都是普通骰子, 不合要求. 把两个都分给同一方, 则

$$A(x)=x(1+x)(1+x+x^2)=x+2x^2+2x^3+x^4,$$

$$B(x)=x(1+x)(1+x+x^2)(1-x+x^2)^2=x+x^3+x^4+x^5+x^6+x^8.$$

两个多项式各有六个系数为 1 的项(第二式中 x^2 与 x^7 的系数为零), 面数都是 6, 点数都是正整数. 对应的两枚骰子分别是

$$1,\,2,\,2,\,3,\,3,\,4 \quad 与 \quad 1,\,3,\,4,\,5,\,6,\,8.$$

由构造过程 $A(x)B(x)=D(x)^2$, 两枚骰子掷出的和从 2 到 12, 各种和的方案数依次是 $1,2,3,4,5,6,5,4,3,2,1$, 与两枚普通骰子完全一致.

这样的一对骰子称为 Sicherman 骰子. 全过程没有枚举任何一种掷法, 只做了多项式分解, 原因是乘法把点数的和与指数的和对应了起来. 这类充当计数载体的多项式称为**生成函数**, 第一卷第 3.20 节处理牛顿和时用过同一个工具.

4.5 排列组合综合应用

排列与组合给出了计数的代数工具. 要把复杂的问题对应到这些基本模型上, 我们先对问题进行剖析与转化.

4.5.1 相邻与不相邻问题

在排列问题中, 元素之间的相对位置关系是最常出现的约束条件之一. 相邻与不相邻构成一对典型的对立约束, 分别对应**捆绑法**与**插空法**两种经典的处理技巧.

4.5.1.1 相邻问题与捆绑法

若若干特定元素必须排在一起, 即相邻, 我们就把它们在逻辑上视为一个不可分割的整体, 即一个复合元素, 再与其他元素一起排列.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

捆绑法

解决元素相邻的排列问题, 通常遵循以下步骤.

1. **捆绑**. 将所有必须相邻的 m 个元素视为一个单一的、不可分割的大元素.
2. **外排**. 将这个大元素与其余的 $n-m$ 个元素进行全排列. 如果共有 k 个元素(包括大元素), 则此步的方案数为 $k!$.
3. **内排**. 考虑被捆绑的 m 个元素内部自身的排列方式. 此步的方案数通常为 $m!$.
4. **相乘**. 根据乘法原理, 将外排和内排的方案数相乘, 得到最终结果.

例题 4.18　有 5 名男生和 3 名女生站成一排, 如果要求 3 名女生必须站在一起, 问有多少种不同的排法?

解　这是典型的相邻问题, 我们用捆绑法处理. 令 3 名女生组成的整体为 G, 则待排列的对象变为 5 名男生 $B_1,\dots,B_5$ 与 G, 共 $5+1=6$ 个元素. 对这 6 个元素作全排列, 由 $A(6,6)=6!=720$ 种方法.

在整体 G 内部, 3 名女生自身尚可重排, 有 $A(3,3)=3!=6$ 种方法. 外排与内排两步相互独立, 由乘法原理得

知乎@余命数　知乎@余命数

$$N=6!\times 3!=720\times 6=4320,$$

即共有 4320 种不同的排法.

例题 4.19　用数字 0、1、2、3、4、5 组成没有重复数字的六位数, 要求偶数数字必须相邻, 奇数数字也必须相邻, 问能组成多少个这样的六位数?

解　先按奇偶性分集合. 令偶数集合 $E=\{0,2,4\}$, 奇数集合 $O=\{1,3,5\}$, 各有 3 个元素. 题目要求偶数相邻且奇数也相邻, 故需两次捆绑. 将 3 个偶数捆绑为整体 B_E, 3 个奇数捆绑为整体 B_O. 外层只需排列 B_E,B_O 这两个大元素, 全排列数为 $2!$. 但所成者为六位数, 首位不能为 0, 因此按哪个大元素居首分类讨论.

情况一.B_O 居首. 此时排列为 (B_O,B_E), 首位必为奇数 (1, 3 或 5), 不可能是 0, 所有排列皆有效.B_O 内部全排列 $3!=6$ 种, B_E 内部全排列 $3!=6$ 种, 此情况共 $3!\times 3!=36$ 种.

知乎@余命数

情况二.B_E 居首. 此时排列为 (B_E,B_O), 首位须取 B_E 中一数字且不能为 0, 故 B_E 内部排列的首位只能选 2 或 4(2 种), 余下 2 个偶数, 含 0 排在后两位有 $2!$ 种, B_E 内部有效排列数为 $2\times 2!=4$ 种; B_O 内部仍有 $3!=6$ 种. 此情况共 $4\times 6=24$ 种.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

两情况互斥, 由加法原理得

知乎@余命数

$$N = 36 + 24 = 60,$$

即共能组成 60 个这样的六位数.

4.5.1.2 不相邻问题与插空法

若若干特定元素必须互不相邻, 我们就先排列那些没有位置限制的元素. 这些元素会自然形成一些空隙, 再将要求不相邻的元素插入这些空隙.

插空法

解决元素互不相邻的排列问题, 通常遵循以下步骤.

1. **排不受限元素**. 首先排列那些没有不相邻限制的元素. 假设有 n 个这样的元素, 它们的排列数为 $n!$.
2. **造空**. 这 n 个元素排好后, 将产生 $n+1$ 个可供插入的空隙(包括队伍的最前端和最后端).
3. **插入**. 将 m 个要求不相邻的元素, 插入到这 $n+1$ 个空隙中. 由于每个空隙最多只能放一个元素, 这相当于从 $n+1$ 个空隙中选出 m 个进行排列. 方案数为 $A(n+1, m)$.
4. **相乘**. 根据乘法原理, 将上述步骤的方案数相乘, 得到最终结果.

知乎@余命数

例题 4.20 将 3 本不同的小说和 4 本不同的诗集排成一排, 要求 3 本小说互不相邻, 问有多少种不同的排法?

解 这是典型的互不相邻问题, 我们用插空法处理. 小说是受限元素, 诗集是不受限元素. 先排列 4 本不同的诗集, 有 $A(4,4) = 4! = 24$ 种方法; 4 本诗集排好后形成 5 个可供插入的空隙

$$_P_1_P_2_P_3_P_4_,$$

将 3 本不同的小说插入这 5 个空隙, 每空至多一本, 等价于从 5 个空隙中选出 3 个并考虑小说顺序, 有 $A(5,3) = 5 \times 4 \times 3 = 60$ 种方法. 由乘法原理, 总的排法数为

知乎@余命数

$$N = 4! \times A(5,3) = 24 \times 60 = 1440,$$

即共有 1440 种不同的排法.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

不相邻与不全相邻的辨析

在处理计数问题时，我们必须先弄清约束条件的含义. 互不相邻与不全相邻是两个容易混淆的概念.

- **互不相邻**. 指任意两个指定的元素都不能相邻. 这是**插空法**的适用范围.
- **不全相邻**. 指指定的元素不全部连在一起即可，允许其中一部分相邻.

对于不全相邻问题，直接计数通常很困难，因为它包含多种复杂情况. 更简洁的策略是使用**补集思想**（正难则反），即用总的排列数减去其对立面，也就是全部相邻的排列数.

例题 4.21　有 5 名男生和 3 名女生站成一排，回答下列问题.

1. 3 名女生互不相邻的排法有多少种?
2. 3 名女生不全相邻的排法有多少种?

解　本题意在澄清互不相邻与不全相邻的区别.

(1) 3 名女生互不相邻，适用插空法. 先排 5 名男生，有 $5!$ 种方法；5 名男生形成 6 个空隙；将 3 名女生插入这 6 个空隙，有 $A(6,3)$ 种方法. 由乘法原理得

$$N_1 = 5! \times A(6,3) = 120 \times (6 \times 5 \times 4) = 120 \times 120 = 14400.$$

(2) 3 名女生不全相邻，宜用补集思想，其对立面是 3 名女生全部相邻. 先算 8 人全排列总数 $8! = 40320$；对立事件即 3 名女生全部相邻，是前例用捆绑法算过的情形，将 3 名女生捆绑与 5 名男生共 6 个元素全排列，再乘以女生内部全排列，得 $N_{\text{全相邻}} = 6! \times 3! = 720 \times 6 = 4320$. 由总数减去对立事件数

$$N_2 = 8! - (6! \times 3!) = 40320 - 4320 = 36000.$$

比较 $N_1 < N_2$，不全相邻的情况包含了恰有两名女生相邻以及三名女生互不相邻两种情况，而 N_1 只是后者.

4.5.2 捆绑法与插空法的原理

上一节引入了处理相邻与不相邻问题的两种流程，即捆绑法与插空法. 本节分析它们为何有效，即把它们还原为乘法原理的构造性解释.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

4.5.2.1 捆绑法

相邻问题的约束是某些元素必须保持相对位置不变, 这破坏了全排列中每个元素独立性的前提. 捆绑法的做法, 是把一个带有复杂**约束**的计数问题, 通过逻辑上的等价变换, 转化为一个无约束的、分步的**构造**问题.

原理

现在证明一个事实. 一个满足 m 个元素相邻的排列, 可以由以下两个独立的步骤唯一地构造出来.

1. **内部构造**. 将被指定的 m 个元素进行全排列, 形成一个有序的、不可分割的复合元素.
2. **外部构造**. 将此复合元素与其余 $n-m$ 个普通元素进行全排列.

我们来分析这个构造过程的完备性与唯一性.

- **完备性**. 任何一个满足条件的最终排列, 都能通过这个两步构造法得到. 在任意一个满足条件的排列中, 那 m 个相邻的元素必然以某种顺序排列, 对应内部构造, 并且这个由它们组成的元素块必然在整个序列中占据一个位置 (对应外部构造). 因此, 不存在任何一种合法的排列会被构造过程遗漏.
- **唯一性**. 不存在两个不同的构造步骤序列会产生同一个最终排列. 改变内部构造, 例如, 将捆绑的 $\{A,B\}$ 顺序从 (A,B) 改为 (B,A), 最终排列中这部分元素的顺序也必然改变. 同样, 改变外部构造, 例如, 移动复合元素的位置, 最终排列也必然不同. 因此, 每一个独特的内部构造与外部构造组合都唯一对应一个最终的合法排列.

由于构造步骤与合法排列之间建立了一一对应, 合法排列的总数就精确等于构造步骤的总数. 根据乘法原理, 构造步骤的总数是每一步方法数的乘积.

- 内部构造的方法数是对 m 个元素的全排列, 即 $m!$.
- 外部构造的方法数是将 $n-m+1$ 个元素($n-m$ 个普通元素和 1 个复合元素)进行全排列, 即 $(n-m+1)!$.

因此, 总方法数为 $(n-m+1)!\times m!$. 这从根本上证明了捆绑法计算逻辑的正确性. 它把一个静态的满足约束的排列集合有多大的问题, 转变为一个动态的生成一个这样的排列有多少种方法的问题.

4.5.2.2 插空法

不相邻问题的约束是某些元素之间必须被其他元素隔开. 直接处理这种排斥性约束是困难的. 插空法通过逆向思维, 把一个复杂的**定位**问题转化为一个有序的**填充**问题.

插空法同样是一个构造性过程, 其正确性依赖于以下分步构造的唯一性与完备性.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

1. **构建框架**. 首先排列那些没有位置限制的 n 个元素. 这一步的目的是创建一个稳定的骨架或框架.
2. **填充框架**. 将要求不相邻的 m 个元素, 放置到由上述框架产生的空隙中.

我们审视这个过程的数学基础.

- **空隙的定义**.n 个已经排好的元素 $U_1, U_2, \ldots, U_n$, 会在它们之间以及两端, 严格地定义出 $n+1$ 个不同的、可识别的空隙位置.

$$_U_1_U_2_\cdots_U_n_$$

这些空隙是放置不相邻元素的理想容器, 因为任何被放入不同空隙的两个元素, 都必然被至少一个框架元素分隔, 从而天然满足不相邻的约束.

- **填充即排列**. 将 m 个**不同**的受限元素放入 $n+1$ 个空隙中, 并且每个空隙至多放一个元素时, 这个过程本身是一个排列问题. 它等价于下面两步.
  1. 从 $n+1$ 个空隙中**选出** m 个空隙.
  2. 将 m 个不同的元素**排入**这 m 个选定的空隙中.

  它是排列 $A(n+1, m)$ 的组合意义.

与捆绑法类似, 这个两步构造过程(构建框架 $\rightarrow$ 填充框架)与所有满足互不相邻条件的最终排列之间也构成一一对应关系.

- **完备性**. 任何一个合法的排列, 其不受限元素必然构成一个有序子序列, 对应框架构建, 而受限元素必然分布在由这些不受限元素隔开的位置上 (对应填充框架).
- **唯一性**. 如果框架的排列不同, 或者填充空隙的方案不同, 最终得到的总排列也必然不同.

因此, 满足条件的总排列数等于构建框架的方法数与填充框架的方法数的乘积.

- 构建框架的方法数是对 n 个不受限元素的全排列, 即 $n!$.
- 填充框架的方法数是从 $n+1$ 个空隙中选 m 个进行排列, 即 $A(n+1, m)$.

最终结果为 $n! \times A(n+1, m)$. 这就证明了插空法是乘法原理的一个构造性应用.

例题 4.22 (2020.4 学军中学高二下期中) 一条街道上有 10 盏路灯, 将路灯依次排列并编号 1-10. 有关部门要求晚上这 10 盏路灯中相邻的两盏灯不能全开, 且这 10 盏路灯中至少打开两盏路灯. 则符合要求的开法总数是 _____.

解 此问题的约束是开着的灯互不相邻; 又因只关心哪些位置的灯是开的, 灯本身无差别, 所以这是一个组合问题. 直接挑选 k 个不相邻的位置并不方便, 我们反过来思考. 利用关着的灯为开着的灯制造空隙, 这就是插空思想.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

按打开的灯数 $k(k\geq 2)$ 分类. 设打开 k 盏灯, 则有 $10-k$ 盏灯关闭. 先将这 $10-k$ 盏关着的灯视作相同元素排成一排, 形成 $(10-k)+1=11-k$ 个空隙, 含两端; 为保证 k 盏灯彼此不相邻, 将它们放入这 $11-k$ 个空隙, 每空至多一盏. 因灯无差别, 此即从 $11-k$ 个空隙中选 k 个, 方法数为 $\binom{11-k}{k}$. 由 $k\geq 2$ 及 $11-k\geq k$, 即 $2k\leq 11$ 得 $k\in\{2,3,4,5\}$. 各情况相加

$$\begin{aligned}N&=\binom{9}{2}+\binom{8}{3}+\binom{7}{4}+\binom{6}{5}\\&=\tfrac{9\cdot 8}{2}+\tfrac{8\cdot 7\cdot 6}{3!}+\binom{7}{3}+6\\&=36+56+35+6=133,\end{aligned}$$

故符合要求的开法总数是 133.

同一结果亦可由变换法导出. 问题即从 10 个位置中选出若干个互不相邻的位置, 是纯组合问题. 设打开 k 盏灯, 等价于从 $\{1,\dots,10\}$ 中取 k 个数且任意两数不相邻. 令选出的数为 $1\leq x_1<x_2<\cdots<x_k\leq 10$, 约束可写为 $x_{i+1}-x_i\geq 2$.

为处理此不等式, 构造双射 $y_i=x_i-(i-1)$. 则 $y_1=x_1\geq 1, y_k=x_k-(k-1)\leq 10-(k-1)=11-k$, 且

$$y_{i+1}-y_i=(x_{i+1}-i)-(x_i-(i-1))=(x_{i+1}-x_i)-1\geq 1,$$

即 $y_1<y_2<\cdots<y_k$. 于是从 $\{1,\dots,10\}$ 中选 k 个互不相邻的数 x_i, 等价于从 $\{1,\dots,11-k\}$ 中选 k 个不同的数 y_i, 方法数同为 $\binom{11-k}{k}$.k 的取值范围仍由 $k\geq 2$ 与 $11-k\geq k$ 决定, 得 $k\in\{2,3,4,5\}$, 相加得

$$N=\binom{9}{2}+\binom{8}{3}+\binom{7}{4}+\binom{6}{5}=36+56+35+6=133.$$

例题 4.23 (2014 重庆理) 某次联欢会要安排 3 个歌舞类节目、2 个小品类节目和 1 个相声类节目的演出顺序, 则同类节目不相邻的排法种数是(　).

A. 72　　B. 120　　C. 144　　D. 168

解 本题涉及对多个元素集合的不相邻约束, 正面处理很复杂, 使用**容斥原理**(正难则反)正合适.

首先假定所有 6 个节目都是独一无二的. 总的排列数(没有任何限制)为

$$N_{总}=6!=720$$

需要排除所有不符合同类节目不相邻要求的排列. 同类节目不相邻的对立面是至少有一类同类节目相邻. 具体地, 就是 2 个小品相邻或者 3 个歌舞中有节目相邻.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

设 U 为所有排列的全集,$|U| = 720$. 令 A 为 2 个小品节目相邻的排列构成的集合. 令 B 为至少有 2 个歌舞节目相邻的排列构成的集合. 要求解的是 $|U| - |A \cup B|$. 根据容斥原理, $|A \cup B| = |A| + |B| - |A \cap B|$.

将 2 个小品节目捆绑成一个复合元素. 现在排列 {歌舞 1, 歌舞 2, 歌舞 3, 相声, (小品块)} 这 5 个对象, 有 5! 种方法. 小品块内部有 2! 种排列.

$$|A| = 5! \times 2! = 120 \times 2 = 240$$

直接计算至少有 2 个歌舞相邻比较复杂, 计算其补集, 即 3 个歌舞节目互不相邻(记作 B^c). 为此使用插空法. 先排列非歌舞节目 {小品 1, 小品 2, 相声}, 有 3! 种方法. 这 3 个节目形成了 4 个空隙. 将 3 个不同的歌舞节目插入这 4 个空隙中, 有 $A(4,3)$ 种方法.

$$|B^c| = 3! \times A(4,3) = 6 \times 24 = 144$$

因此, $|B| = |U| - |B^c| = 720 - 144 = 576$.

最后计算 $|A \cap B|$, 这个集合表示小品相邻 **且**歌舞至少有 2 个相邻. 同样可利用补集思想计算这个交集的大小, 即 $|A \cap B| = |A| - |A \cap B^c|$. $|A \cap B^c|$ 表示小品相邻 **且**歌舞互不相邻. 再次使用插空法. 将被捆绑的小品块和相声节目作为框架. 排列 {(小品块), 相声} 有 2! 种方法. 小品块内部有 2! 种排列. 故框架的排列数为 $2! \times 2! = 4$. 这个框架形成了 3 个空隙. 将 3 个不同的歌舞节目插入这 3 个空隙中, 有 $A(3,3) = 3!$ 种方法.

$$|A \cap B^c| = (2! \times 2!) \times 3! = 4 \times 6 = 24$$

所以, $|A \cap B| = |A| - |A \cap B^c| = 240 - 24 = 216$.

不符合要求的排列总数为

$$|A \cup B| = |A| + |B| - |A \cap B| = 240 + 576 - 216 = 600$$

因此, 符合要求的排列种数为

$$N = |U| - |A \cup B| = 720 - 600 = 120$$

故选 B.

4.5.3 定序问题与倍缩策略

在排列问题中, 除了相邻与不相邻的约束外, 另一类常见的限制是**定序问题**, 即要求某几个元素必须保持特定的相对顺序. 例如, 在排列 A、B、C、D 时, 要求 A 必须始终在 B 的左边.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

面对这类问题, 一种有效的思想是**倍缩策略**, 做法是先放后缩. 先忽略定序的约束, 对所有相关元素进行全排列, 得到一个总数. 然后分析在这个总数中, 所关注的、被定了序的元素有多少种排列方式. 由于题目只允许其中的一种特定顺序, 原来的计数夸大或重复了, 其倍数恰好是被定序元素的全排列数. 因此, 用总排列数除以这个重复倍数, 即可得到最终答案.

这个思想来自组合数公式, 而组合数公式本身就是倍缩策略最经典的体现.

$$\binom{n}{k} = \frac{A(n,k)}{k!}$$

此公式的组合意义是, 要从 n 个元素中选出 k 个而不计顺序, 可以先选出 k 个并排列它们, 共 $A(n,k)$ 种方法. 但我们不关心这 k 个元素的顺序, 而它们自身有 $k!$ 种排列, 因此将 $A(n,k)$ 除以 $k!$ 以消除顺序的影响.

倍缩策略是用总排列数除以被定序元素的全排列数. 这种除法是乘法原理的逆向应用, 背后是集合划分与一一对应的思想. 我们来说明它为何有效.

设集合 S 含 n 个元素, 我们要计算 S 的所有排列中, 满足特定 m 个元素 $M = \{e_1, e_2, \dots, e_m\}$ 保持固定相对顺序, 例如 e_1 在 e_2 之前, e_2 在 e_3 之前, 以此类推的排列总数. 令这个符合条件的排列集合为 $\mathcal{A}$, 其大小为 $N = |\mathcal{A}|$.

直接计算 N 较困难, 因此引入一个更大的集合作为参照. 令 $\mathcal{U}$ 为 S 中所有 n 个元素的全排列集合, 则 $|\mathcal{U}| = n!$.

关键在于建立集合 $\mathcal{U}$ 与 $\mathcal{A}$ 之间的关系. 我们用构造性论证与乘法原理表达 $|\mathcal{U}|$.

考虑如何构造出 $\mathcal{U}$ 中的任意一个排列. 每一个排列都可由一个两步过程唯一地生成.

1. **选择一个骨架排列.**
   从符合定序条件的集合 $\mathcal{A}$ 中任意选择一个排列. 在这个排列中, 那 m 个特殊元素的相对顺序是正确的. 这一步有 N 种选择.
2. **对特殊元素进行内部重排.**
   在上一步选出的骨架排列中, 那 m 个特殊元素占据了 m 个确定的位置. 保持其他 $n-m$ 个元素不动, 只将这 m 个特殊元素在这 m 个位置上进行任意的全排列. 这一步有 $m!$ 种方法.

根据**乘法原理**, 通过这两个步骤, 总共可以构造出 $N \times m!$ 个不同的排列.

现在证明这个构造过程与集合 $\mathcal{U}$ 等价, 即它不重不漏地生成了 $\mathcal{U}$ 中的所有 $n!$ 个排列.

- **无遗漏:** $\mathcal{U}$ 中的任何一个排列都能被上述方法构造出来. 任取一个 $\mathcal{U}$ 中的排列 p, 其中 m 个特殊元素的位置和内部顺序是任意的. 可以通过一个标准化操作, 将这 m 个元素按照题目要求的固定顺序重新排列, 从而得到一个唯一对应的骨架排列(它必然属于集合 $\mathcal{A}$). 这证明了任何排列 p 都有其构造来源.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- **无重复：**两个不同的构造步骤组合，不会产生同一个最终排列. 假设通过 $(\text{骨架}_1, \text{重排}_1)$ 和 $(\text{骨架}_2, \text{重排}_2)$ 得到了同一个最终排列 p. 一个排列中，所有非特殊元素的位置，以及所有特殊元素所占据的位置集合，都是唯一的. 因此，骨架_1 和 骨架_2 必须具有相同的结构，即相同的非特殊元素排列和相同的特殊元素占位. 又因为在骨架排列中特殊元素的相对顺序固定，所以 $\text{骨架}_1 = \text{骨架}_2$. 既然骨架相同，且最终结果 p 相同，那么施加在骨架上的内部重排操作也必须相同，即 $\text{重排}_1 = \text{重排}_2$. 因此每一种构造组合都唯一对应一个最终排列.

上述论证建立了从构造步骤对到全排列集合 $\mathcal{U}$ 的一一对应关系. 因此它们的大小必然相等.

$$|\mathcal{U}| = N \times m!$$

已知 $|\mathcal{U}| = n!$，所以

$$n! = N \times m!$$

由此解出最初的目标 N.

$$N = \frac{n!}{m!}$$

这个推导把倍缩策略归结为对乘法原理方程 $n! = N \times m!$ 的求解.

从另一个角度看，这个过程等价于将 $n!$ 个全排列进行**分组**或**划分**. 我们将所有排列按照除 m 个特殊元素的内部顺序外、其他一切相同的规则分组. 例如，排列 (e_2, x, e_1, y) 和 (e_1, x, e_2, y) 就属于同一组，因为它们的骨架（非 e 元素的位置和 e 元素所占的位置）是相同的. 每一组内的排列个数，恰好等于这 m 个特殊元素自身的全排列数，即 $m!$ 个. 而题目要求的定序排列，恰好是每一组中那**唯一一个**满足特定顺序的排列. 因此，符合条件的排列总数 N 等于这个分组的数量. 根据除法原理，分组的数量为

$$N = \frac{\text{总元素数}}{\text{每组的元素数}} = \frac{n!}{m!}$$

这两种视角都证明了倍缩策略的正确性. 它把一个带有复杂相对顺序约束的问题，转化为关于等价类的计数问题.

倍缩策略

若在一组共 n 个元素的排列中，有 m 个特定元素被要求保持固定的相对顺序，则符合条件的排列数可按以下步骤计算.

1. 计算这 n 个元素的全排列数，即 $n!$.
2. 这 m 个被定序的元素在上述全排列中，自身有 $m!$ 种不同的排列方式.
3. 由于约束只允许这 $m!$ 种排列中的一种，将总数除以 $m!$.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

最终的排列数为 $\dfrac{n!}{m!}$. 如果有多组元素被定序, 则需除以各组元素内部全排列数的乘积.

例题 4.24 (2024 年 1 月浙江省宁波九校联考) 体育课上, 罗老师让 8 名身高各不相同的同学排队, 要求排成前后两排, 每排 4 人, 且每排同学从左到右身高依次递增, 则不同排法的种数为(　).

A. 60　　B. 70　　C. 80　　D. 90

解 此问题的突破口在于, 一旦确定了哪些同学在前排、哪些同学在后排, 那么由于身高依次递增的严格约束, 每一排的站位顺序就**唯一确定**了. 因此, 问题从一个排列问题, 退化为一个分组或组合问题.

任务可以简化为把 8 名同学分成两组, 一组是前排组, 另一组是后排组, 每组 4 人.

确定前排组的人选即可完成这个任务. 从 8 名同学中选出 4 名到前排, 这是一个组合问题, 其方法数为

$$\binom{8}{4}=\frac{8\times7\times6\times5}{4\times3\times2\times1}=70$$

一旦这 4 名同学被选定, 他们只能以身高从低到高的一种方式站立.

剩下的 $8-4=4$ 名同学自动成为后排组, 同样, 他们的站立方式也只有一种.

由乘法原理, 总的排法数就是分组的方法数, 即 $70\times1\times1=70$.

也可以从**倍缩策略**的角度理解这个问题, 这种视角更能说明重复计数来自哪里.

首先不考虑任何身高顺序的限制, 将 8 名同学安排到 8 个具体的位置上(4 个前排位置, 4 个后排位置). 这相当于 8 名同学的全排列, 总方法数为 $8!$.

现在引入约束条件. 对于最终被安排在前排的 4 名同学, 无论他们是谁, 在 $8!$ 种总排列中, 他们在这 4 个前排位置上可以有 $4!$ 种不同的排列方式. 然而, 题目要求他们必须按身高递增排列, 这只是 $4!$ 种排列中的**一种**. 因此, 对前排的排列多算了 $4!$ 倍.

同理, 对于最终被安排在后排的 4 名同学, 他们的顺序也被唯一确定, 所以对后排的排列也多算了 $4!$ 倍.

由倍缩策略, 将总排列数除以这两个重复的倍数, 得到符合条件的排法数.

$$N=\frac{8!}{4!\times4!}=\frac{40320}{24\times24}=\frac{40320}{576}=70$$

这与从组合角度得到的结果一致. 故选 B.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 4.25 (2008 安徽理) 12 名同学合影, 站成前排 4 人后排 8 人, 现摄影师要从后排 8 人中抽 2 人调整到前排, 若其他人的相对顺序不变, 则不同调整方式的种数是(　).

A. $C_8^2A_3^2$　　C. $C_8^2A_6^2$

B. $C_8^2A_6^6$　　D. $C_8^2A_5^2$

解 此问题分两步. 先**选人**, 从后排 8 名同学中选 2 人调到前排, 只关心选哪两人、不涉及顺序, 是组合问题, 方法数为

$$\binom{8}{2}$$

知乎@余命数

再**排位**, 此时前排 $4+2=6$ 名同学, 后排剩 $8-2=6$ 名.

问题的约束是其他人的相对顺序不变, 即

- 原前排 4 人的相对顺序保持不变.
- 原后排剩下的 6 人的相对顺序也保持不变.

由于后排 6 人的位置和相对顺序都已固定, 我们只需考虑前排这 6 人的新排法.

在前排的 6 个位置上, 需要安排 4 名旧同学和 2 名新同学. 其中, 4 名旧同学的相对顺序是固定的. 这是一个典型的**定序问题**.

我们用插空法解决排位问题. 把前排的 6 个位置看作 6 个空位, 为新选出的 2 名同学选择位置. 为第一位新同学选择位置, 有 6 种选择. 为第二位新同学选择位置, 有 5 种选择. 这两位新同学是不同的个体, 顺序是重要的. 因此为这两位新同学安排位置的方法数是一个排列问题, 共有

$$A(6,2)=6\times5=30$$

一旦这两位新同学的位置确定, 剩下的 4 个位置将由 4 名旧同学按原有相对顺序填入, 方法只有 1 种.

或者, 也可以运用**倍缩策略**思考排位问题. 如果不考虑任何顺序限制, 将这 6 名同学(4 旧 2 新)排在前排 6 个位置上, 共有 6! 种方法. 但题目要求 4 名旧同学的相对顺序不变. 在这 6! 种排列中, 这 4 名旧同学自身有 4! 种排列方式, 而只接受其中的 1 种. 因此, 符合条件的排法数为

$$\frac{6!}{4!}=\frac{720}{24}=30=A(6,2)$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

最后由乘法原理, 将选人和排位两个步骤的方法数相乘, 得到总的调整方式种数.

$$N = \binom{8}{2} \times A(6,2)$$

在题目选项的记法中, 即为 $C_8^2 A_6^2$. 故选 C.

例题 4.26 (重复元素的排列) 计算由单词 MISSISSIPPI 中所有字母构成的不同排列(字符串)的数量.

解 这个问题的新复杂性在于元素重复. 单词 MISSISSIPPI 共含 11 个字母, 其中 1 个 M, 4 个 I, 4 个 S, 2 个 P.

直接应用排列公式行不通, 因为它要求所有元素可区分. 但借助倍缩策略与一个思想实验, 此问题即可解决.

首先假想所有字母都可区分. 给相同字母加上下标来区分它们.

$$M_1, I_1, S_1, S_2, I_2, S_3, S_4, I_3, P_1, P_2, I_4$$

现在我们有 11 个完全不同的对象, 将它们全排列, 总方法数为 11!.

接下来去除下标, 回到原问题. 考虑任意一个具体的排列, 例如 $S_1 I_1 P_1 M_1 I_2 S_2 S_3 I_3 P_2 S_4 I_4$. 现在关注其中的 4 个 I ($I_1, I_2, I_3, I_4$). 在这个排列的特定位置上, 这 4 个 I 自身可以有 4! 种排列方式, 例如 I_2 和 I_1 交换位置. 在假想的可区分世界里, 这些都是不同的排列. 但在现实世界中, 所有 I 都相同, 因此这 4! 种排列都对应同一个字符串 SIPMISSIPSI.

最初计算的 11! 对每一个最终字符串, 都因 I 的存在而重复计数了 4! 倍. 因此必须用总数除以 4! 来消除由 I 引起的重复.

同理, 对于 4 个 S, 也重复计数了 4! 倍. 对于 2 个 P, 重复计数了 2! 倍.

由倍缩策略, 将总的假想排列数除以各重复元素内部排列所产生的重复倍数.

$$N = \frac{11!}{4! \times 4! \times 2! \times 1!} = \frac{39916800}{(24) \times (24) \times (2) \times (1)} = \frac{39916800}{1152} = 34650$$

因此共有 34650 个不同的排列.

此例导出一个普遍结论. 对于包含 n 个对象的多重集, 其中有 k_1 个第一类相同对象,k_2 个第二类相同对象,...,k_m 个第 m 类相同对象,其中 $k_1+k_2+\cdots+k_m = n$, 其不同的排列数为

$$\binom{n}{k_1, k_2, \ldots, k_m} = \frac{n!}{k_1! k_2! \cdots k_m!}$$

这是我们之前遇到的**多项式系数**. 倍缩策略为其提供了一个组合解释.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 4.27 (圆排列) 将 n 位不同的宾客安排在一个圆形的餐桌旁就座, 有多少种不同的就座方式?(如果两种就座方式中, 每位宾客的左邻和右邻都完全相同, 则认为这两种方式是相同的.)

解 这是典型的圆排列问题. 圆桌没有起点或终点的概念, 判断标准从绝对位置变为相对位置.

我们再次用倍缩策略, 从一个更熟悉的线性排列开始.

首先暂时忽略圆桌的特性, 设想将这 n 位宾客排成一条直线. 那么总的线性排列数为 $n!$.

现在将任意一个线性排列, 例如 $(P_1, P_2, \dots, P_n)$, 弯曲成一个圆圈, 使其首尾相连. 旋转这个线性排列, 依次得到下列各排列.

$$
\begin{aligned}
&(P_1, P_2, P_3, \dots, P_n)\\
&(P_n, P_1, P_2, \dots, P_{n-1})\\
&(P_{n-1}, P_n, P_1, \dots, P_{n-2})\\
&\vdots\\
&(P_2, P_3, P_4, \dots, P_1)
\end{aligned}
$$

通过 n 次旋转, 得到 n 个不同的线性排列. 然而, 当将它们都布置到圆桌上时, 它们描述的是**同一种**就座方式, 因为在每一种情况中, P_1 的左边总是 P_n、右边总是 P_2; P_2 的左边总是 P_1、右边总是 P_3, 以此类推. 相对位置关系完全一样.

这说明了一个关键事实. 最初计算的 $n!$ 种线性排列, 可以将它们 n 个一组进行划分. 每一组中的 n 个线性排列都对应着同一个圆排列. 因此初始计数 $n!$ 将每一种真正的圆排列都重复计算了 n 倍.

由倍缩策略, 将线性排列总数除以这个重复的倍数 n, 即得到圆排列的数目.

$$
N = \frac{n!}{n} = (n-1)!
$$

因此有 $(n-1)!$ 种不同的就座方式.

圆排列的一个自然延伸是**项链问题**. 将 n 颗不同的珠子串成一条项链, 求不同的串法数. 项链既能旋转, 也能从空间中拿起来**翻转**. 对于 $n > 2$ 且珠子两两不同的情况, 每一次翻转都会将一个圆排列变成另一个镜像圆排列, 但二者对应同一条项链. 因此之前的计数 $(n-1)!$ 将每条项链重复计算了 2 倍. 项链的不同串法数为 $\dfrac{(n-1)!}{2}$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 4.28 (棋盘问题 *) 在一个 $m \times n$ 的棋盘格上, 一个棋子从左下角 $(0,0)$ 出发, 要到达右上角 (m,n). 如果每次只能向上或向右移动一个单位, 总共有多少条不同的最短路径?

解 这个经典的格路计数问题解法众多, 倍缩策略为它提供了一个新视角.

首先分析路径的构成. 任何一条从 $(0,0)$ 到 (m,n) 的最短路径, 都必须包含 m 次向右移动与 n 次向上移动, 分别记作 R 与 U. 因此, 每一条不同的路径都唯一对应一个由 m 个 R 和 n 个 U 构成的序列, 该序列总长度为 $m+n$.

例如, 在 2×1 的棋盘上, 从 $(0,0)$ 到 $(2,1)$ 的路径如下.

- 路径 $(0,0) \to (1,0) \to (2,0) \to (2,1)$ 对应序列 RRU.
- 路径 $(0,0) \to (1,0) \to (1,1) \to (2,1)$ 对应序列 RUR.
- 路径 $(0,0) \to (0,1) \to (1,1) \to (2,1)$ 对应序列 URR.

于是原问题转化为计算由 m 个 R 和 n 个 U 构成的不同序列的数量.

这是一个带有重复元素的排列问题, 是倍缩策略的用武之地.

设想 $m+n$ 次移动, 它们分别是 $R_1, R_2, \dots, R_m$ 和 $U_1, U_2, \dots, U_n$. 将这 $m+n$ 个不同的操作进行全排列, 总方法数为 $(m+n)!$.

在现实问题中, 所有的向右移动 R 没有区别, 所有的向上移动 U 也没有区别. 在假想的 $(m+n)!$ 种排列中, 那 m 个不同的向右操作 $R_1, \dots, R_m$ 自身可以有 $m!$ 种排列方式. 但这些都对应同一个由 m 个 R 组成的移动模式. 因此计数因 R 的存在而重复了 $m!$ 倍.

同理, 计数也因 U 的存在而重复了 $n!$ 倍.

将总的假想排列数除以由重复元素产生的倍数, 得到最终的路径数.

$$N = \frac{(m+n)!}{m!n!}$$

这个结果是组合数 $\binom{m+n}{m}$, 也等于 $\binom{m+n}{n}$. 这为组合数提供了一个新的组合解释.

$$\binom{m+n}{m} = \text{从 (0,0) 到 (m,n) 的格路数}$$

也可以直接用组合的定义解决. 一条路径总共需要 $m+n$ 步. 只需在这 $m+n$ 步中决定哪几步向右移动 (R), 剩下的自动是向上移动 (U). 从 $m+n$ 个总步数中选出 m 步来执行向右移动, 方法数为

$$\binom{m+n}{m}$$

这与倍缩策略得到的结果一致. 这种将排列问题转化为序列编码、再用倍缩策略求解的方法, 是组合数学中的常用技巧.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

例题 4.29 (2024 年 1 月湖南省邵阳市高三第一次联考) 城步苗族自治县六月六山歌节是湖南省四大节庆品牌之一, 至今已举办 25 届. 假设在即将举办的第 26 届六月六山歌节中, 组委会要在原定排好的 10 个本土歌舞节目中增加 2 个歌王对唱节目. 若保持原来 10 个节目的相对顺序不变, 则不同的排法种数为(　).

A. 110　　B. 144　　C. 132　　D. 156

解　问题的约束是保持原来 10 个节目的相对顺序不变. 记原来的 10 个节目为 $O_1, O_2, \ldots, O_{10}$, 则在最终的节目单中 O_1 必须出现在 O_2 之前, O_2 必须出现在 O_3 之前, 以此类推. 这是一个典型定序问题, 适合用倍缩策略.

增加 2 个新的歌王对唱节目后(视这两个新节目为不同个体 N_1, N_2), 总节目数量为 $10+2=12$ 个.

方法一采用倍缩策略.

首先忽略相对顺序不变的约束, 考虑将这 12 个不同节目全排列. 总的排列方法数为

$$12!$$

现在引入约束条件. 在上述 12! 种排列中, 那 10 个原来的节目自身可以有 10! 种不同的内部排列方式. 然而, 题目要求只接受它们按原有顺序排列的**那一种**.

对于任何一种满足条件的最终排法, 最初的 12! 计数都将其重复计算了 10! 次(因为这 10! 次内部重排都被算作不同情况). 因此必须用总排列数除以这个重复的倍数.

符合条件的排法种数为

$$N=\frac{12!}{10!}=\frac{12\times 11\times (10!)}{10!}=12\times 11=132$$

方法二采用选位法(插空思想的变体).

也可从构造的角度思考. 总共有 12 个节目位置需要安排, 问题可转化为先为 2 个新节目选择位置, 剩下的位置自然由 10 个旧节目按其固有顺序填充. 从 12 个时间槽中为 2 个不同的新节目选定位置, 等价于从 12 个位置中选出 2 个进行排列, 方法数为 $A(12,2)=12\times 11=132$. 新节目位置一旦确定, 只剩 $12-2=10$ 个空位, 而 10 个旧节目的相对顺序固定, 填充这 10 个空位的方式**唯一确定**, 方法数为 1. 由乘法原理, 总的排法种数为 $A(12,2)\times 1=132$.

两种方法得到相同的结果. 故选 C.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 4.30 (待插入元素相同) 沿用上题背景, 假设增加的 2 个歌王对唱节目是**完全相同**的, 例如, 都是同一首歌的两个不同时段重播, 若仍要保持原来 10 个节目的相对顺序不变, 则不同的排法种数为(　).

A. 55　　B. 66　　C. 78　　D. 132

解 这个问题与原题的区别在于新增的 2 个节目从可区分变为不可区分, 这直接影响计数方式.

方法一采用选位法(组合思想).

总共有 12 个节目位置. 任务是为 2 个**相同**的新节目选择位置. 由于新节目相同, 选择第 3 位和第 8 位与选择第 8 位和第 3 位最终产生的节目单相同. 因此这是一个纯粹的组合问题.

从 12 个可用位置中选出 2 个位置安放这两个相同的新节目. 方法数为

$$\binom{12}{2} = \frac{12 \times 11}{2 \times 1} = 66$$

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

一旦新节目的位置被确定, 剩下的 10 个位置将由 10 个旧节目按其固有顺序唯一填充. 由乘法原理, 总排法种数为 $\binom{12}{2} \times 1 = 66$.

方法二采用倍缩策略的叠加.

从 12 个节目全排列出发, 总数为 12!.

首先应用第一个约束, 即 10 个旧节目的相对顺序不变. 因此将总数除以 10! 来消除这 10 个节目内部顺序的重复计数.

$$N_1 = \frac{12!}{10!}$$

知乎@余命数

接着应用第二个约束.2 个新节目是完全相同的. 即在上述结果中, 仍然将这两个新节目视为可区分的, 从而对它们自身的 2! 种排列进行了重复计数. 因此, 必须**再次**应用倍缩策略, 除以 2!.

最终的排法种数为

$$N = \frac{12!}{10! \times 2!} = \frac{A(12,2)}{2!} = \frac{132}{2} = 66$$

这个结果是组合数 $\binom{12}{2}$ 的定义. 它把组合数 $\binom{n}{k}$ 解释为在 n 个对象的排列中同时对选出的 k 个元素和未选的 $n-k$ 个元素施加定序, 或视其为相同约束的结果. 故选 B.

知乎@余命数 知乎@余命数

例题 4.31 (增加不相邻约束) 沿用原题背景, 即增加的 2 个歌王对唱节目是**不同**的. 若既要保持原来 10 个节目的相对顺序不变, 又要求这两个新增的歌王对唱节目**不能相邻**, 则不同的排法种数为(　).

A. 110　　B. 120　　C. 121　　D. 132

解 这个问题在原题的基础上增加了一个不相邻约束. 这种定序与不相邻的复合约束, 是插空法与倍缩思想结合的典型场景.

方法一采用插空法.

处理不相邻问题最直接的方法是用插空法. 在这里, 10 个旧节目是骨架, 2 个新节目是等待插入的元素.

先排列骨架. 先安排 10 个旧节目. 由于它们的相对顺序是固定的, 排列它们的方法只有一**种**.

知乎@余命数

$$O_1, O_2, O_3, O_4, O_5, O_6, O_7, O_8, O_9, O_{10}$$

再创造空隙. 这 10 个排好的节目形成了 11 个可供插入的空隙(包括最前和最后).

$$_O_1_O_2_\cdots_O_{10}_$$

接着插入元素. 将 2 个**不同**的新节目插入到这 11 个空隙中, 每个空隙至多放一个节目, 以保证它们不相邻. 这相当于从 11 个空隙中选出 2 个, 并将 2 个不同的节目排列进去. 这是一个排列问题. 方法数为

$$A(11,2) = 11 \times 10 = 110$$

由乘法原理, 总的排法种数为 $1 \times A(11,2) = 110$.

方法二采用补集思想(正难则反).

也可以从原问题的总数中减去不满足不相邻约束的情况, 即减去 2 个新节目必须相邻的情况.

原问题(只要求旧节目定序)的总排法数, 前面已计算过是 132.

现在计算旧节目定序且新节目相邻的排法数. 使用**捆绑法**. 将 2 个不同的新节目捆绑成一个大节目块. 这个大节目块内部, 2 个新节目可以有 $2! = 2$ 种排列方式.

现在的问题转化为将这个大节目块和 10 个旧节目进行排列, 同时保持 10 个旧节目的相对顺序不变. 这等价于将 1 个大节目块插入到 10 个旧节目形成的 11 个空隙(包括首尾)中. 选择插入位置的方法数有 $\binom{11}{1} = 11$ 种.

因此新节目相邻的排法数为

$$N_{\text{相邻}} = (\text{内部排列}) \times (\text{外部插入}) = 2! \times \binom{11}{1} = 2 \times 11 = 22$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

最后用总数减去相邻数.

$$N_{\text{不相邻}} = N_{\text{总}} - N_{\text{相邻}} = 132 - 22 = 110$$

两种方法再次得到一致的结果. 故选 A.

4.5.4 分组问题

在组合计数中, 经常需要将一个包含 n 个不同元素的集合, 划分成若干个互不相交的子集, 即组. 解决这类问题的要点在于精确辨析一个要素, 即这些最终形成的组本身是**可区分的**还是**不可区分的**.

这两种模型的根本差异可以从一个场景看出. 设有 6 本完全不同的书, 需要将它们分成三份, 每份 2 本.

- **情景一.** 将这些书分给甲、乙、丙三个人, 每人 2 本.
- **情景二.** 将这些书打包成三份, 每份 2 本, 堆在桌子上.

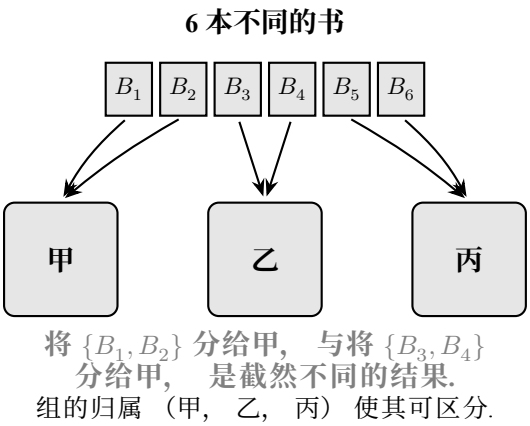


图 4.4 组可区分的有序分组

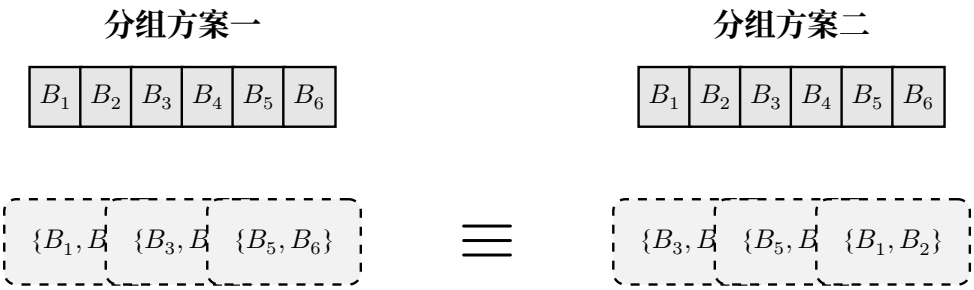
4.5.4.1 有序分组

在情景一中, 最终分到书的三个人甲、乙、丙是明确可区分的. 甲拿到 $\{B_1, B_2\}$ 和乙拿到 $\{B_1, B_2\}$ 是两种完全不同的分配结果. 因此, 这些组因为其归属(甲、乙、丙)而被赋予了身份, 是可区分的. 这称为**有序分组**或**定序分组**.

计算这类问题的过程是一个连续的组合决策.

- 首先, 从 6 本书中为甲选出 2 本, 方法数为 $\binom{6}{2}$.
- 接着, 从剩下的 4 本书中为乙选出 2 本, 方法数为 $\binom{4}{2}$.
- 最后, 剩下的 2 本书自动归丙所有, 方法数为 $\binom{2}{2}$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵



知乎@余命数

图 4.5　组不可区分的无序分组

由乘法原理, 总的分配方法数为

$$N_1 = \binom{6}{2}\binom{4}{2}\binom{2}{2} = \frac{6\times 5}{2} \times \frac{4\times 3}{2} \times 1 = 15 \times 6 \times 1 = 90$$

知乎@余命数

将这个结果用阶乘形式展开, 可以得到

$$\frac{6!}{2!4!} \times \frac{4!}{2!2!} \times \frac{2!}{2!0!} = \frac{6!}{2!2!2!}$$

这是将 6 个不同元素分成大小为 2, 2, 2 的三个**有序**组的方法数, 其形式与多项式系数一致.

知乎@余命数

有序分组模型

将 n 个不同元素分成 m 个**可区分的**组, 例如, 分配给 m 个不同的人, 或放入 m 个有标签的盒子, 各组的元素个数分别为 $k_1, k_2, \ldots, k_m$, 其中 $\sum k_i = n$, 其总方法数为

$$N = \binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{k_m}{k_m} = \frac{n!}{k_1!k_2!\cdots k_m!}$$

证明　我们用构造性论证证明此公式. 思路是将完成一次有序分组这一任务, 分解为一系列连续且独立的组合选择, 然后应用乘法原理.

设原始集合为 S, 包含 n 个不同元素. 现将其划分为 m 个有序子集 $G_1, G_2, \ldots, G_m$, 使得对每个 $i = 1, \ldots, m$, 子集 G_i 的大小为 $|G_i| = k_i$.

可按顺序为这 m 个可区分的组选择它们的成员.

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

首先为第一个组 G_1 选择成员. 需从 S 中的 n 个元素里不计顺序地选出 k_1 个. 根据组合数的定义, 完成这一步的方法数为

$$\binom{n}{k_1}$$

在确定 G_1 的成员后, 集合 S 中还剩下 $n-k_1$ 个元素可供分配. 接下来为第二个组 G_2 选择成员. 需从这 $n-k_1$ 个剩余元素中选出 k_2 个. 方法数为

$$\binom{n-k_1}{k_2}$$

这个过程依次进行. 当为第 i 个组 G_i 选择成员时, 已有 $k_1+k_2+\cdots+k_{i-1}$ 个元素被分配. 因此, 需从剩下的 $n-\sum_{j=1}^{i-1}k_j$ 个元素中, 选出 k_i 个.

直至最后一个组 G_m, 此时剩下的元素数量为 $n-\sum_{j=1}^{m-1}k_j=k_m$. 需从这 k_m 个元素中选出 k_m 个, 方法数为 $\binom{k_m}{k_m}=1$.

由于整个分组任务必须完成所有这些步骤, 根据乘法原理总的方法数 N 是每一步方法数的乘积.

$$N=\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-\sum_{j=1}^{m-2}k_j}{k_{m-1}}\binom{k_m}{k_m}$$

这便证明了公式的第一个等式.

为了推导其更简洁的阶乘形式, 将每一个组合数用其阶乘定义展开.

$$\begin{aligned}N&=\frac{n!}{k_1!(n-k_1)!}\times\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}\times\cdots\times\frac{(k_m)!}{k_m!(k_m-k_m)!}\\&=\frac{n!}{k_1!\cancel{(n-k_1)!}}\times\frac{\cancel{(n-k_1)!}}{k_2!\cancel{(n-k_1-k_2)!}}\times\cdots\times 1\end{aligned}$$

观察到前一项分母中的阶乘项, 恰好被后一项的分子抵消. 这种级联约分会一直持续到最后.

最终所有中间的阶乘项都被消去, 只剩下最初的分子 $n!$ 和每一项分母中代表各组大小的阶乘 $k_i!$.

$$N=\frac{n!}{k_1!k_2!\cdots k_m!}$$

这与多项式系数的形式完全吻合. □

例题 4.32 某学校的志愿者社团共有 9 名成员, 需要将他们分配到 A、B、C 三个不同的社区服务中心.A 中心需要 2 名成员, B 中心需要 3 名成员, C 中心需要 4 名成员. 问共有多少种不同的分配方案?

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 这个问题要求将 9 名不同成员分配到三个可区分的中心, 且每个中心所需人数已经确定, 因此是一个有序分组问题. 把分配过程视为一个连续的决策步骤.

首先为 A 中心从 9 名成员中选派 2 人. 这是一个组合问题, 因为成员在此步骤中仅被选中, 其顺序无关紧要. 方法数为

$$\binom{9}{2}$$

接着从余下的 $9-2=7$ 名成员中为 B 中心选派 3 人. 方法数为

$$\binom{7}{3}$$

最后剩下的 $7-3=4$ 名成员将全部被派往 C 中心. 方法数为

$$\binom{4}{4}$$

根据乘法原理, 完成整个分配任务的总方案数是上述步骤方法数的乘积.

$$N=\binom{9}{2}\binom{7}{3}\binom{4}{4}=\frac{9\times 8}{2\times 1}\times\frac{7\times 6\times 5}{3\times 2\times 1}\times 1=36\times 35\times 1=1260$$

此结果也可直接通过有序分组的通用公式得到. 将 $n=9$ 个元素分成大小为 $k_1=2,k_2=3,k_3=4$ 的可区分三组, 其方法数由多项式系数给出.

$$\binom{9}{2,3,4}=\frac{9!}{2!3!4!}=\frac{362880}{2\times 6\times 24}=\frac{362880}{288}=1260$$

两种计算路径的结果一致. 故共有 1260 种不同的分配方案.

例题 4.33 将 5 名新入职的员工分配到三个不同的项目中, 其中项目 A 需要 1 名员工, 项目 B 需要 2 名员工, 项目 C 需要 2 名员工. 问有多少种不同的分配方案?

解 此任务的要点是将 5 名可区分的员工, 分配到三个可区分的项目 (A、B、C) 中. 这是一个有序分组问题, 即使其中两个项目所需的人数相同, 都是 2 人, B 与 C 这两个项目本身的差异性使分组是**有序**的.

我们采用连续组合选择的方式来解决. 首先, 为项目 A 从 5 名员工中选定 1 人. 方法数为

$$\binom{5}{1}=5$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

接着, 为项目 B 从剩下的 4 名员工中选定 2 人. 方法数为

$$\binom{4}{2} = \frac{4 \times 3}{2} = 6$$

最后剩下的 2 名员工自动被分配到项目 C. 方法数为

$$\binom{2}{2} = 1$$

依据乘法原理, 总的分配方案数为

$$N = \binom{5}{1}\binom{4}{2}\binom{2}{2} = 5 \times 6 \times 1 = 30$$

也可直接应用有序分组公式验证. 将 $n = 5$ 个元素分成大小为 $k_1 = 1, k_2 = 2, k_3 = 2$ 的三组, 其方法数为

$$\binom{5}{1,2,2} = \frac{5!}{1!2!2!} = \frac{120}{1 \times 2 \times 2} = 30$$

两种路径再次得到相同的结果. 故共有 30 种不同的分配方案.

例题 4.34 在一副标准的 52 张扑克牌 (所有牌均不相同) 中, 将所有牌平均发给东、南、西、北四位玩家, 每位玩家获得 13 张牌. 问总共有多少种不同的发牌结果?

解 这个问题要求将 52 张不同的牌分配给四个可区分的玩家. 这是将 $n = 52$ 个不同元素分成四组、每组大小均为 $k = 13$ 的有序分组问题.

我们模拟发牌的连续过程来构建解决方案. 首先为东这位玩家从 52 张牌中发出 13 张. 方法数为

$$\binom{52}{13}$$

接着为南这位玩家从剩下的 $52 - 13 = 39$ 张牌中发出 13 张. 方法数为

$$\binom{39}{13}$$

然后为西这位玩家从剩下的 $39 - 13 = 26$ 张牌中发出 13 张. 方法数为

$$\binom{26}{13}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

最后剩下的 13 张牌全部发给北这位玩家. 方法数为

$$\binom{13}{13}$$

根据乘法原理, 总的发牌结果数为

$$N=\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}$$

现在利用有序分组的通用公式从更宏观的视角验证这个结果. 该问题是把 $n=52$ 个元素分成大小分别为 $k_1=13,k_2=13,k_3=13,k_4=13$ 的四个可区分的组. 其总方法数由以下公式直接给出.

$$N=\frac{52!}{13!13!13!13!}=\frac{52!}{(13!)^4}$$

我们验证这两种表达方式的等价性, 以弄清公式的来龙去脉.

$$\begin{aligned}
&\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}\\
&=\frac{52!}{13!(52-13)!}\cdot\frac{39!}{13!(39-13)!}\cdot\frac{26!}{13!(26-13)!}\cdot\frac{13!}{13!(13-13)!}\\
&=\frac{52!}{13!\cdot 39!}\cdot\frac{39!}{13!\cdot 26!}\cdot\frac{26!}{13!\cdot 13!}\cdot\frac{13!}{13!\cdot 0!}\\
&=\frac{52!}{13!\cdot \cancel{39!}}\cdot\frac{\cancel{39!}}{13!\cdot \cancel{26!}}\cdot\frac{\cancel{26!}}{13!\cdot 13!}\cdot 1 \quad (\text{注意}0!=1)\\
&=\frac{52!}{13!\cdot 13!\cdot 13!\cdot 13!}
\end{aligned}$$

两种方法的结果完全吻合. 最终结果十分庞大, 通常无需计算其具体值, 以阶乘或组合数形式表示即可.

4.5.4.2 无序分组

现在转向情景二. 在这里只是将书打包成三份, 这些份或堆之间没有任何身份标签, 因此不可区分. 例如, 分组结果 $\{\{B_1,B_2\},\{B_3,B_4\},\{B_5,B_6\}\}$ 与 $\{\{B_3,B_4\},\{B_1,B_2\},\{B_5,B_6\}\}$ 是完全**相同**的一种分组, 因为它们都只是把这六本书分成了同样的三堆.

直接计算无序分组较困难, 但可借助有序分组的结果和倍缩策略. 我们先把这些组视为有区别的, 比如给它们贴上 A、B、C 的标签. 正如有序分组计算所得, 这样做有 90 种方法.

现在分析这种贴标签的操作导致了多少倍的重复计数. 考虑一个具体的无序分组结果, 例如 $G=\{\{B_1,B_2\},\{B_3,B_4\},\{B_5,B_6\}\}$. 在我们的有序分组计数中, 这个单一结果对应于将 A、B、C 三个标签分配给这三个小组的全部分配方式.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- (A: $\{B_1, B_2\}$, B: $\{B_3, B_4\}$, C: $\{B_5, B_6\}$)
- (A: $\{B_1, B_2\}$, C: $\{B_3, B_4\}$, B: $\{B_5, B_6\}$)
- ……

将 A、B、C 三个标签全排列, 共有 $3! = 6$ 种方式. 即每一个我们真正想要的无序分组, 在有序分组的计算中都被重复计数了 3! 次.

因此根据倍缩策略, 我们必须用有序分组的结果除以这个重复的倍数.

$$N_2 = \frac{N_1}{3!} = \frac{\binom{6}{2}\binom{4}{2}\binom{2}{2}}{3!} = \frac{90}{6} = 15$$

因此将 6 本不同的书分成无法区分的三堆, 每堆 2 本, 只有 15 种方法.

这个思想可以推广. 如果将 n 个元素分组, 其中有 m 个组的大小完全相同, 那么在计算有序分组后, 需要除以 $m!$ 来消除这些相同大小的组之间的顺序带来的重复.

知乎@余命数

无序分组模型

将 n 个不同元素分组.

1. **各组大小均不相同**:若各组大小为 $k_1, k_2, \ldots, k_m$ 且彼此不等, 则无序分组与有序分组的结果相同, 为 $\dfrac{n!}{k_1!k_2!\cdots k_m!}$.(因为组的大小本身就成为了它们的天然标签)
2. **部分组大小相同**:若其中有 m_1 个组的大小为 k_1, m_2 个组的大小为 k_2, …, 则总方法数为

$$N = \frac{\binom{n}{\cdots}}{m_1!m_2!\cdots} = \frac{n!}{(k_1!)^{m_1}(k_2!)^{m_2}\cdots(m_1!)(m_2!)\cdots}$$

先按有序分组计算, 然后对每一类大小相同的组, 除以该类组数的阶乘.

例题 4.35 将 6 个颜色互不相同的球全部放入三个完全相同的盒子里, 每个盒子里至少放一个球且数目不同, 则不同的放球方法有(　)种.

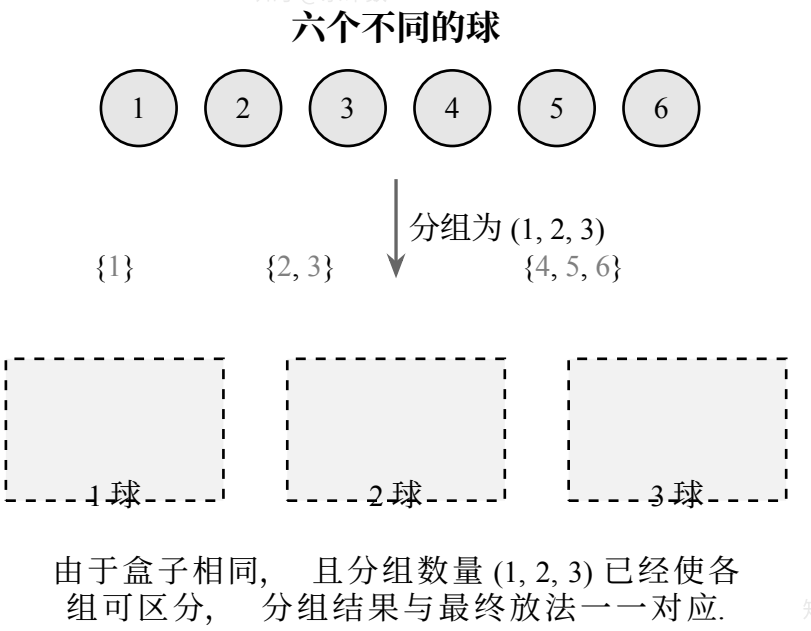
A. 10　　B. 20　　C. 360　　D. 60

解 相异球放入相同盒子, 且附加每盒至少一个、各盒数目互异的约束. 先确定各组球数, 再计算分配方案.

设三盒球数为 k_1, k_2, k_3, 需 $k_1 + k_2 + k_3 = 6$, $k_i \geq 1$ 且两两不同. 即把 6 分拆为三个不同正整数之和. 设 $1 \leq k_1 < k_2 < k_3$. 若 $k_1 = 1$, 则 $k_2 + k_3 = 5$, 唯一解

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数



知乎@余命数
知乎@余命数

图 4.6 分组过程图解

$k_2=2, k_3=3$, 得方案 $(1,2,3)$. 若 $k_1\geq 2$, 最小和为 $2+3+4=9>6$, 不可能. 故唯一数量方案是分成 $1,2,3$ 三组.

再把 6 个不同球按大小 $1,2,3$ 分组, 先选 1 个有 $\binom{6}{1}$ 种, 再选 2 个有 $\binom{5}{2}$ 种, 余下 3 个自成一组, 故

$$N=\binom{6}{1}\binom{5}{2}\binom{3}{3}=6\times 10=60.$$

盒子相同, 分组本应无序; 但三组大小分别为 $1,2,3$, 大小本身即可区分各组, 无需再除以 3!. 因而放法总数即 60. 故选 D.

例题 4.36 将 10 名实习生分配到三个不同的部门 A、B、C 进行实习, 其中 A 部门 3 人, B 部门 3 人, C 部门 4 人. 问有多少种分配方案?

解 这是一个典型有序分组问题, 因为部门 A、B、C 明确可区分.

分步选择. 首先为 A 部门从 10 人中选出 3 人, 有 $\binom{10}{3}$ 种方法. 然后为 B 部门从剩下的 7 人中选出 3 人, 有 $\binom{7}{3}$ 种方法. 最后剩下的 4 人全部到 C 部门, 有 $\binom{4}{4}$ 种方法.

根据乘法原理, 总方案数为

$$N=\binom{10}{3}\binom{7}{3}\binom{4}{4}=\frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1}\times\frac{7\cdot 6\cdot 5}{3\cdot 2\cdot 1}\times 1=120\times 35=4200$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

因此共有 4200 种不同的分配方案.

例题 4.37 将 10 名实习生分成三组, 两组 3 人, 一组 4 人, 再将这三组派往三个不同的城市进行社会调研. 问有多少种派遣方案?

解 问题包含分组与分配两阶段.

方法一采用分步处理, 先分组, 后分配. 先将 10 人分成两组 3 人、一组 4 人, 此时组本身**没有身份**, 只关心成员构成; 两个 3 人组大小相同, 不可区分. 若先按有序分组计算, 方法数为 $\binom{10}{3}\binom{7}{3}\binom{4}{4}=4200$; 因两个 3 人组不可区分, 需再除以 2!, 得

$$N_{\text{分组}}=\frac{\binom{10}{3}\binom{7}{3}\binom{4}{4}}{2!}=2100.$$

再将这三组(人员构成已定, 因而可区分)派往三个不同城市, 有 $3!=6$ 种分配. 总方案数 $N=2100\times 6=12600$.

方法二采用直接分配. 将 10 人直接分到三个城市岗位, 其中两城各 3 人、一城 4 人, 按哪个城市得 4 人分类. 城市 1 得 4 人、其余各 3 人, 方法数为 $\binom{10}{4}\binom{6}{3}\binom{3}{3}=4200$; 城市 2 或城市 3 得 4 人时同样各为 4200. 由加法原理, $N=3\times 4200=12600$.

两种方法结果一致, 并体现计数路径的等价关系

$$\left(\frac{\binom{10}{3}\binom{7}{3}\binom{4}{4}}{2!}\right)\times 3!=3\times\binom{10}{4}\binom{6}{3}.$$

4.5.5 相同元素的分配与隔板法

到目前为止, 我们处理的元素都是**可区分**的, 例如不同的人、不同的书、不同的节目. 但在另一大类问题中, 待分配的元素是**完全相同**的, 例如把 10 个完全相同的名额分配给 3 个班级. 此时哪几个元素分在一起不再有意义, 唯一重要的是每个接收者**分到几个**. 处理这类问题的标准工具是**隔板法**.

我们从一个具体问题开始分析.

例题 4.38 把 10 个完全相同的小球放入 3 个不同的盒子中, 每个盒子至少放 1 个球, 有多少种不同的放法?

解 元素完全相同意味着无法像处理分组问题那样选出某些元素组成一组. 但换一个视角, 把 10 个小球在桌上排成一行.

• • • • • • • • • •

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

只要在这一行小球中插入 2 块隔板, 小球就被分成三段; 把三段依次放入三个盒子, 就得到一种符合要求的放法. 例如

$$\bullet\bullet \mid \bullet\bullet\bullet \mid \bullet\bullet\bullet\bullet\bullet$$

对应“盒子一放 2 个、盒子二放 3 个、盒子三放 5 个”.

10 个小球排成一行, 球与球之间共有 $10-1=9$ 个内部空隙. 要保证每盒至少 1 球, 隔板不能放在两端, 同一个空隙里也不能放两块隔板, 因此只需从这 9 个空隙中选出 2 个放置隔板. 这是一个组合问题, 方法数为

$$\binom{9}{2}=\frac{9\times 8}{2\times 1}=36$$

因此共有 36 种不同的放法.

这个构造有效, 是因为放隔板的方案与放球的方案之间存在一一对应. 每一种放球方案唯一决定 2 块隔板的位置, 反过来, 每一种隔板放法也唯一决定一种放球方案. 于是我们得到一般性结论.

定理 4.14 (隔板法) 把 n 个完全相同的元素放入 k 个不同的盒子 ($n\geq k\geq 1$), 每个盒子至少放 1 个, 则不同的放法数为

$$\binom{n-1}{k-1}$$

等价地, 方程 $x_1+x_2+\cdots+x_k=n$ 的**正整数解** $(x_1,x_2,\ldots,x_k)$ 的个数也是 $\binom{n-1}{k-1}$.

证明 将 n 个相同元素排成一行, 它们之间形成 $n-1$ 个内部空隙. 从中选出 $k-1$ 个空隙各插入一块隔板, 元素就被分成 k 段, 且每段非空; 按顺序将各段放入 k 个盒子, 即得一种合法放法.

这个映射是双射. 任意一种合法放法, 把各盒中的元素依次排开, 在盒与盒的交界处放回隔板, 就唯一还原出一种隔板放法; 反之, 任意一种隔板放法显然唯一决定一种放球方案. 因此放法总数等于隔板放法数, 即从 $n-1$ 个空隙中选出 $k-1$ 个的组合数 $\binom{n-1}{k-1}$.

令 x_i 表示第 i 个盒子中元素的个数, 则每个盒子至少 1 个等价于 $x_i\geq 1$, 放法与方程 $x_1+\cdots+x_k=n$ 的正整数解一一对应, 故正整数解的个数同为 $\binom{n-1}{k-1}$. □

如果允许盒子为空, 一个借球技巧可以把它化归为已经解决的情形.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定理 4.15 (允许空盒的情形) 把 n 个完全相同的元素放入 k 个不同的盒子, 允许某些盒子为空, 则不同的放法数为

$$\binom{n+k-1}{k-1}$$

等价地, 方程 $x_1+x_2+\cdots+x_k=n$ 的**非负整数解**的个数也是 $\binom{n+k-1}{k-1}$.

证明 向每个盒子中借入 1 个相同元素, 于是盒中共有 $n+k$ 个元素. 对每一种允许为空的原始放法, 加上借入的元素后每个盒子都至少有 1 个元素, 对应一种把 $n+k$ 个元素放入 k 个盒子且每盒非空的放法; 反过来, 对 $n+k$ 个元素的任意一种每盒非空的放法, 从每个盒子中各拿走 1 个元素, 就唯一还原出一种允许为空的原始放法.

二者一一对应, 故所求放法数等于"$n+k$ 个相同元素放入 k 个不同盒子、每盒非空"的放法数, 由隔板法即 $\binom{n+k-1}{k-1}$. 非负整数解的情形同样由 $x_i\geq 0$ 与放法的一一对应得到. □

隔板法使用要点

- **适用条件**. 待分配的元素**完全相同**, 接收的盒子**彼此不同**, 且只关心每个盒子分到几个.
- **每盒非空**.n 个元素、k 个盒子, 放法数为 $\binom{n-1}{k-1}$, 对应**正整数解**计数.
- **允许空盒**. 放法数为 $\binom{n+k-1}{k-1}$, 对应**非负整数解**计数; 可通过每盒先借 1 个化为非空情形.
- **每盒至少 m 个**. 先给每盒各预放 m 个, 或等价地换元, 剩余元素再按非空情形分配.

例题 4.39 某校把 12 个三好学生名额分配给高三年级的 4 个班, 每班至少 1 个名额, 有多少种分配方案? 若允许某些班没有名额, 又有多少种?

解 名额是完全相同的元素, 班级是彼此不同的盒子, 这是隔板法的标准场景.

第一问. 每班至少 1 个名额. 由隔板法, 分配方案数为

$$\binom{12-1}{4-1}=\binom{11}{3}=\frac{11\times 10\times 9}{3\times 2\times 1}=165$$

第二问. 允许有班级没有名额. 由允许空盒的公式, 分配方案数为

$$\binom{12+4-1}{4-1}=\binom{15}{3}=\frac{15\times 14\times 13}{3\times 2\times 1}=455$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 4.40　求方程 $x_1+x_2+x_3+x_4=20$ 满足 $x_1\geq 1$、$x_2\geq 2$、$x_3\geq 3$、$x_4\geq 4$ 的整数解的个数.

解　各未知数的下界不同, 先通过预付把它们统一化为正整数解问题. 这等价于先往第 2、3、4 个盒子里分别预放 $1,2,3$ 个元素, 再把剩下的元素按每盒至少 1 个分配.

令 $y_1=x_1$、$y_2=x_2-1$、$y_3=x_3-2$、$y_4=x_4-3$, 则原约束等价于 $y_i\geq 1$, 且

$$y_1+y_2+y_3+y_4=20-(0+1+2+3)=14$$

由隔板法, 这个方程的正整数解个数为

$$\binom{14-1}{4-1}=\binom{13}{3}=\frac{13\times 12\times 11}{3\times 2\times 1}=286$$

所以原方程满足条件的整数解共有 286 个.

4.5.6 涂色问题 *

在加法原理与乘法原理一节中, 我们解决过 4 个方格排成一行、相邻不同色的问题. 从左到右逐格染色, 每格只需避开左边一格的颜色, 乘法原理直接给出答案. 但如果这些方格首尾相接围成一个**环**, 或者染色的对象换成地图、多面体的顶点, 单纯的分步就不再够用, 因为后面某一步的选择数会依赖于前面某两步的颜色是否相同. 这时必须把加法原理与乘法原理结合起来分类讨论. 下面用两个经典的例子说明其中的思想.

例题 4.41 (环形方格)　把 4 个方格围成一圈, 即第 1 格与第 4 格也相邻, 用 5 种不同的颜色染色, 要求相邻方格颜色不同, 有多少种染色方法?

解　与链状情形的差别在于, 染第 4 格时既要避开第 3 格的颜色, 又要避开第 1 格的颜色, 而第 1 格和第 3 格是否同色直接影响第 4 格的选择数. 因此按第 1 格与第 3 格是否同色分类.

先染第 1 格, 有 5 种选择; 第 2 格与第 1 格不同色, 有 4 种选择.

情况一. 第 3 格与第 1 格同色第 3 格只需避开第 2 格的颜色, 而第 1 格的颜色恰好满足条件, 因此第 3 格的颜色被唯一确定(1 种选择). 此时第 4 格只需避开第 3 格, 亦即第 1 格这一种颜色, 有 4 种选择. 此类共有 $5\times 4\times 1\times 4=80$ 种.

情况二. 第 3 格与第 1 格不同色第 3 格既要避开第 2 格、又要避开第 1 格的颜色, 有 $5-2=3$ 种选择. 此时第 4 格要避开第 3 格与第 1 格这两种不同的颜色, 有 $5-2=3$ 种选择. 此类共有 $5\times 4\times 3\times 3=180$ 种.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

两类互斥, 由加法原理, 总的染色方法数为

$$N = 80 + 180 = 260$$

这个按首尾是否呼应分类的思想可以推广到任意个方格的环, 并得到一般公式. 用 k 种颜色染 n 个环形方格 ($n \geq 2$), 相邻不同色的方法数为

$$a_n = (k-1)^n + (-1)^n(k-1)$$

知乎@余命数

它的推导需要一点递推技巧, 我们把它留作本章习题. 可以验证 $n = 4$, $k = 5$ 时 $a_4 = 4^4 + 4 = 260$, 与上面的结果一致.

例题 4.42 (2003 全国理) 用 5 种不同的颜色给一个四棱锥 $S-ABCD$ 的 5 个顶点染色, 要求同一条棱的两个端点颜色不同, 有多少种染色方法?

解 顶点 S 与底面四个顶点都相邻, 底面 $ABCD$ 本身又是一个四元环. 仿照上例, 按 A 与 C 是否同色分类处理底面.

先染 S, 有 5 种选择; 再染 A, 与 S 不同, 有 4 种选择; 染 B, 与 S, A 都不同, 有 3 种选择.

情况一. C 与 A 同色 C 只需与 S, B 不同色, 而 A 的颜色恰满足此条件, 因此 C 的颜色被唯一确定(1 种选择). 此时 D 与 S、A、C 相邻, 而 A, C 同色, 所以 D 只需避开 2 种颜色, 有 3 种选择. 此类共有 $5 \times 4 \times 3 \times 1 \times 3 = 180$ 种.

情况二. C 与 A 不同色 C 既要与 S, B 不同色, 又不与 A 同色, 故 C 有 $5-3=2$ 种选择. 此时 D 要避开 S、A、C 这 3 种互不相同的颜色, 有 2 种选择. 此类共有 $5 \times 4 \times 3 \times 2 \times 2 = 240$ 种.

由加法原理, 总的染色方法数为

$$N = 180 + 240 = 420$$

涂色问题的难点在于找到影响后续选择数的关键位置, 如本节的第 1 格与第 3 格、顶点 A 与 C, 并围绕它展开分类.

习题

4.1. (加法原理与乘法原理) 用数字 $0, 1, 2, \ldots, 9$ 组成没有重复数字的三位偶数, 共能组成多少个? 按个位是否为 0 分类, 注意百位不能为 0. 答案 328.

4.2. (组合数的对称性) 计算 $\binom{15}{12} + \binom{20}{18}$. 先利用 $\binom{n}{k} = \binom{n}{n-k}$ 化大为小. 答案 645.

4.3. (组合与补集) 从 5 名男生和 4 名女生中选出 3 人参加活动, 要求至少有 1 名女生, 有多少种不同的选法? 用总数减去全是男生的情形. 答案 74.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

4.4. (捆绑法)7 名同学站成一排, 其中甲、乙两人必须相邻, 有多少种不同的排法? 把甲、乙视为一个复合元素, 注意内部排列. 答案 1440.

4.5. (插空法)7 名同学站成一排, 其中甲、乙两人互不相邻, 有多少种不同的排法? 先排其余 5 人, 再在空隙中安排甲、乙. 答案 3600.

4.6. (倍缩策略)7 名同学站成一排, 要求甲、乙、丙三人的相对顺序保持不变(甲在乙前, 乙在丙前, 三人不必相邻), 有多少种不同的排法? 在全排列中, 三人的 3! 种内部顺序只有 1 种被接受. 答案 840.

4.7. (无序分组)将 6 名不同的同学分成两组, 每组 3 人, 有多少种不同的分法? 若改为分给甲、乙两个小组, 每组 3 人, 又有多少种? 前一问中两个组不可区分, 需除以 2!; 后一问中组有归属, 是有序分组. 答案 10 与 20.

4.8. (隔板法)把 12 个相同的小球放入 4 个不同的盒子, 每个盒子至少 1 个, 有多少种放法? 若允许盒子为空, 又有多少种? 分别对应正整数解与非负整数解的计数. 答案 165 与 455.

4.9. (二项式定理)求 $\left(2x-\dfrac{1}{x}\right)^6$ 展开式中的常数项. 通项为 $\binom{6}{k}(2x)^{6-k}\left(-\frac{1}{x}\right)^k$, 令 x 的指数为 0. 答案 -160.

4.10. (赋值法)求 $(x+2y)^7$ 展开式中所有项的系数之和. 令 $x=y=1$. 答案 2187.

4.11. (交错和)求 $\binom{10}{0}+\binom{10}{2}+\binom{10}{4}+\cdots+\binom{10}{10}$ 的值. 偶数项二项式系数之和等于 2^{n-1}. 答案 512.

4.12. (环形涂色)用 k 种颜色给 n 个围成一圈的方格染色 ($n\geq 2$), 相邻方格颜色不同, 记方法数为 a_n. 证明:

$$a_n=(k-1)^n+(-1)^n(k-1)$$

先染第 1 格, 再沿环依次染色, 可得递推关系 $a_n+a_{n-1}=k(k-1)^{n-1}$ ($n\geq 3$); 剪开环后, 首尾同色与首尾不同色的情形分别对应 a_{n-1} 与 a_n. 再由递推式用数学归纳法或构造等比数列求出通项.

知乎@余命数

知乎@余命数

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

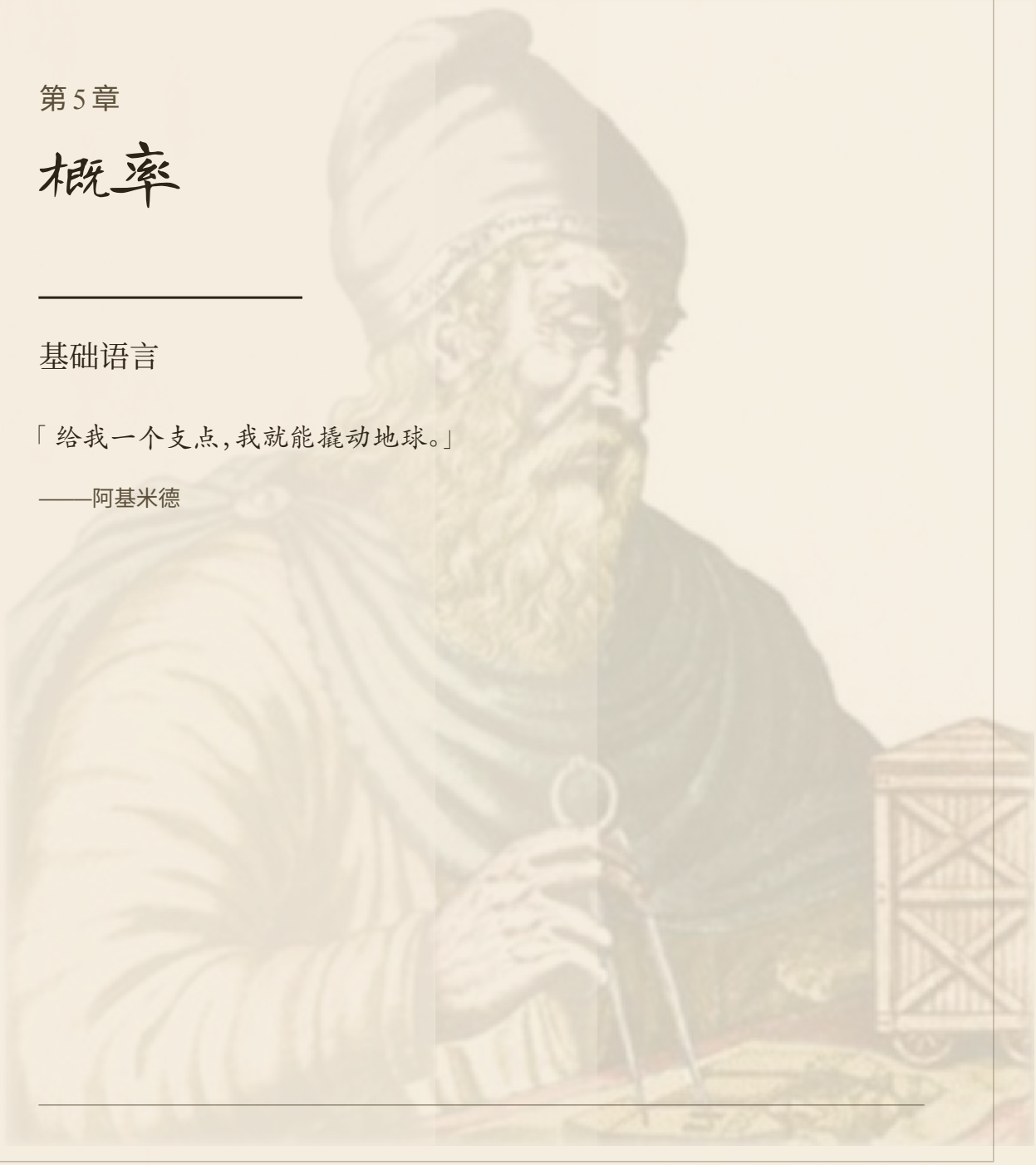
第 5 章

概率

基础语言

「给我一个支点，我就能撬动地球。」

——阿基米德



知乎@余命数

第五章 概率

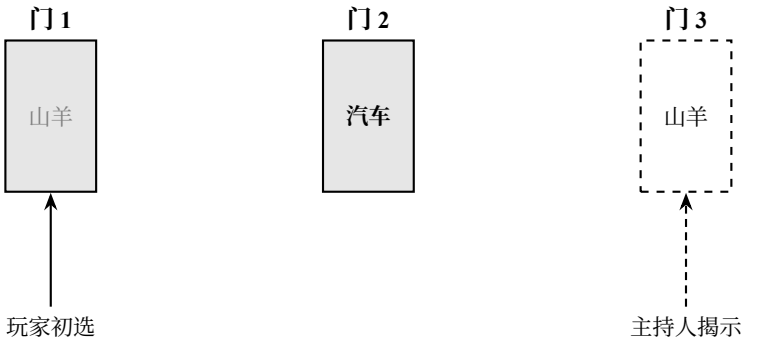


图 5.1 蒙提霍尔问题:玩家初选门 1 后, 主持人揭示门 3 为山羊

蒙提霍尔问题

我们先来看一个源自美国电视游戏的著名谜题. 舞台上有三扇关闭的门, 其中一扇门后面是一辆汽车, 另外两扇门后面各是一只山羊. 我们选择一扇门, 例如门 1, 但暂时不能打开. 此时, 知道门后真相的主持人会打开剩下两扇门中的一扇, 我们便会看到其后面是一只山羊, 例如门 3.

接着他问,“你是坚持最初选择的门 1, 还是换成剩下那扇未打开的门 2?”第一感觉往往是, 既然门 3 已经被排除, 剩下门 1 和门 2, 汽车在其中一扇门后的机会似乎均等, 各有一半. 那么换与不换, 结果应该没有区别.

但这个直觉是错误的; **交换选择会使赢得汽车的机会翻倍**. 要解释这种看似合理的直觉为何失效, 我们需要一套能够精确量化可能性大小的数学工具, 即**概率论**. 用严谨的逻辑分析不确定性, 许多直觉性的判断便会现出原形.

待到 第5.5节 建立条件概率与贝叶斯公式之后, 我们再来回看蒙提霍尔问题, 给出严格解答.

5.1 事件与概率

抛掷一枚硬币会出现正面还是反面, 投掷一颗骰子会得到几点, 我们每天都面对诸如此类的不确定性. 概率论提供了一套描述和量化这种不确定性的数学语言. 本章从最基本的概念(事件)出发, 搭建起初步的概率论体系.

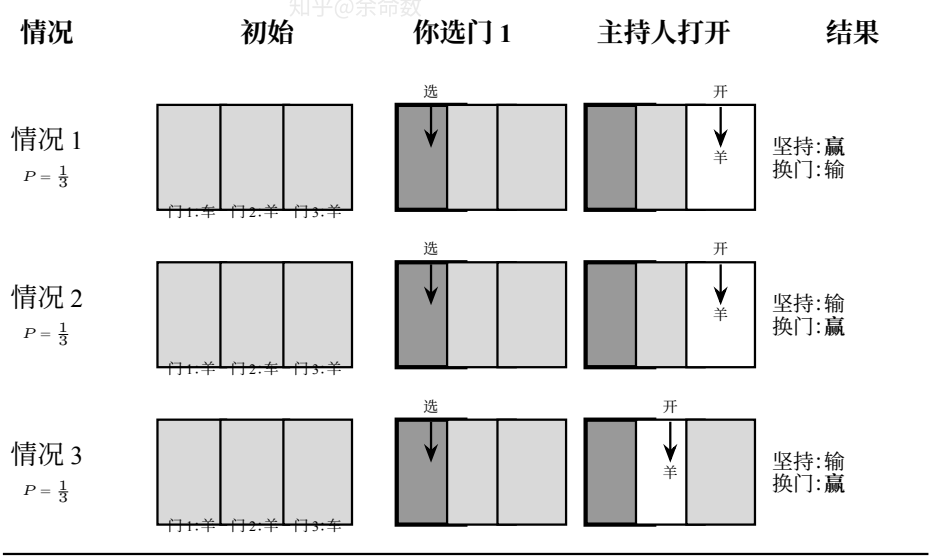
5.1.1 随机试验与事件

要讨论可能性的大小, 我们先要精确界定所讨论的对象. 概率论的起点是**随机试验**.

一个过程或试验, 若它满足以下三个基本特征, 我们便称之为随机试验.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数



在 3 种等可能情况中，　坚持策略仅在情况 1 获胜 ($P = \frac{1}{3}$), 换门策略在情况 2 和 3 获胜 ($P = \frac{2}{3}$)

图 5.2　蒙提霍尔问题:穷举所有可能情况(假设玩家初选门 1)

1. **可重复性**试验可以在相同的条件下重复进行.
2. **可知性**试验的所有可能结果是明确已知的.
3. **不确定性**每次试验前无法预知具体结果.

例题 5.1 (随机试验的例子)
- **抛掷一枚硬币**可以在相同条件下重复抛掷, 所有可能结果是正面或反面, 但每次抛掷前不知道具体是哪一面朝上.
- **掷一颗骰子**所有可能结果是点数 $\{1,2,3,4,5,6\}$.
- **记录某路口一小时内通过的汽车数量**所有可能结果是 $\{0,1,2,\dots\}$.
- **从区间 $[0,1]$ 上随机取一个数**所有可能结果是该区间内的任何一个实数.

对于随机试验, 我们把它所有可能的基本结果构成的集合称为**样本空间**, 通常用大写希腊字母 Ω 表示. 样本空间中的每一个元素, 即每一个不能再分解的基本结果, 称为一个**样本点**.

定义 5.1 (样本空间与样本点) 随机试验 E 的所有可能结果构成的集合称为 E 的**样本空间**, 记为 Ω. 样本空间 Ω 中的元素, 即试验的每一个基本结果, 称为**样本点**.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

例题 5.2 (样本空间的确定)

- 抛掷一枚硬币一次, 记正面为 H, 反面为 T. 样本空间为 $\Omega = \{H, T\}$.
- 掷一颗骰子一次, 样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- 抛掷一枚硬币两次, 记录每次的结果. 样本空间为

$$\Omega = \{(H, H),\ (H, T),\ (T, H),\ (T, T)\}.$$

- 从装有 2 个红球和 3 个白球的袋中, 不放回地抽取 2 个球. 若关心球的颜色组合, 则样本空间可设为

$$\Omega = \{\{红, 红\},\ \{红, 白\},\ \{白, 白\}\}.$$

若对每个球加以区分, 例如 R_1, R_2, W_1, W_2, W_3, 则样本空间会复杂得多. **样本空间的定义方式取决于我们所关心的问题**.

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

上面最后一个例子表明, 定义样本空间时, 我们既要罗列结果, 也要围绕研究目标建模. 样本空间是为解决特定问题而构建的数学模型.

5.1.1.1 模型的选择

5.1.1.2 等可能性

在应用古典概型时, 一个根本性的要求是样本空间中的所有样本点必须**等可能**. 然而, 等可能性往往只是建模时做出的一种**理想化假设**. 通过恰当地定义样本点, 我们可以*构造*出满足等可能性的样本空间.

例题 5.3 (区分与不区分的抉择) 一个袋中装有 2 个红球和 2 个白球, 除颜色外完全相同. 从中不放回地随机抽取 2 个球, 求抽到 2 个红球的概率是多少?

解 对此试验, 我们至少可以构想出两种不同的样本空间.

第一种只关心最终抽出的颜色组合, 不区分同色球. 可能的结果有三种, 即两红、一红一白、两白. 样本空间取为

$$\Omega_1 = \{两红, 一红一白, 两白\}$$

如果草率地认定这三个结果等可能, 我们就会得出 $P(两红) = \frac{1}{3}$ 的错误结论. 这个模型虽然直观, 但样本点并不等可能; 它忽略了不同球之间的差异, 压缩了信息.

知乎@余命数

第二种改从抽取过程入手, 区分同色球以构造等可能的样本点. 即便这些球物理上无法分辨, 数学上仍可给它们贴上标签 R_1, R_2, W_1, W_2. 于是抽取两球等

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

同于从这 4 个可区分的物体中取出 2 个, 不考虑顺序. 所有可能的组合构成新样本空间

$$\Omega_2 = \{\{R_1, R_2\}, \{W_1, W_2\}, \{R_1, W_1\}, \{R_1, W_2\}, \{R_2, W_1\}, \{R_2, W_2\}\}$$

这个样本空间共有 $\binom{4}{2} = 6$ 个样本点. 由于每个球被抽到的机会均等, 我们可以合理地假设这 6 个组合是**等可能的**.

现在考察事件 A, 它表示抽到 2 个红球. 在 Ω_2 中, 这个事件对应于子集 $A = \{\{R_1, R_2\}\}$. 该事件只包含了 1 个样本点. 根据古典概型, 正确的概率为

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega_2)} = \frac{1}{6}$$

这个对比清楚地展示了建模思路. 在数学上将物理上不可区分的物体视为可区分的, 便能构造出满足等可能性的样本空间, 从而套用古典概型. **好的样本空间能把一个复杂的、非等可能的问题化为一个简单的、等可能的组合计数问题**.

样本空间的构建还依赖于所关心问题的**细节层次**. 一个过于粗糙的样本空间可能丢失解决问题所需的信息, 而一个过于精细的样本空间则可能徒增不必要的计算复杂性. 例如, 在连续抛掷三次硬币的试验中, 若关心恰好出现两次正面的概率, 那么记录每次抛掷结果的**有序**样本空间 $\Omega = \{HHH, HHT, \ldots, TTT\}$ 是必要的. 若只关心正面的总数, 则样本空间可取为 $\Omega = \{0, 1, 2, 3\}$, 它虽然合法, 但样本点不等可能, 反而增加了计算难度.

5.1.1.3 事件的定义与表示

定义了样本空间之后, 我们便可定义事件. 在日常语言中, 事件指发生的事情.

定义 5.2 (事件) 样本空间 Ω 的任意一个**子集** A ($A \subseteq \Omega$), 称为一个**随机事件**, 简称**事件**. 当且仅当一次试验的结果, 即样本点属于子集 A 时, 我们称**事件 A 发生**.

这个定义是现代概率论的基础. 它把随机、不确定的概念(事件的发生)与确定、严谨的数学对象(集合)联系起来.

例题 5.4 (事件的集合表示) 考虑掷一颗骰子的试验, 其样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$. 看几个事件的例子.

- 事件 A 表示掷出的点数为偶数. 用集合表示, $A = \{2, 4, 6\}$.
- 事件 B 表示掷出的点数大于 4. 用集合表示, $B = \{5, 6\}$.
- 事件 C 表示掷出的点数为 7. 这是一个不可能发生的情况, 对应的集合是空集 $C = \emptyset$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

- 事件 D 表示掷出的点数小于 10. 这是一个必然发生的情况, 对应的集合是整个样本空间 $D=\Omega$.

从这个例子中可见两种特殊的事件.

- **必然事件**每次试验中必定发生的事件. 它对应于整个样本空间 Ω.
- **不可能事件**每次试验中都绝不会发生的事件. 它对应于空集 $\emptyset$.

5.1.2 事件的运算与关系

既然事件就是集合, 集合论中的各种关系和运算便对应事件之间的逻辑关系, 这套语言足以精确描述复杂的概率问题. 我们以掷一颗骰子的试验为例, 令样本空间 $\Omega=\{1,2,3,4,5,6\}$.

知乎@余命数

定义 5.3 (包含关系) 若事件 A 的发生必然导致事件 B 的发生, 则称事件 B **包含**事件 A, 记作 $A\subseteq B$ 或 $B\supseteq A$.

例如, 把点数为 2 或 4 的事件记作 $C=\{2,4\}$, 把点数为偶数的事件记作 $B=\{2,4,6\}$. 那么 $C\subseteq B$, 因为掷出 2 点或 4 点必然意味着掷出了偶数点.

知乎@余命数

定义 5.4 (和事件) 事件"A 与 B **至少有一个**发生"称为 A 与 B 的**和事件**, 记作 $A\cup B$.

例如, 把点数不大于 4 的事件记作 $A=\{1,2,3,4\}$, 把点数为偶数的事件记作 $B=\{2,4,6\}$. 那么和事件 $A\cup B=\{1,2,3,4,6\}$, 表示点数不大于 4 或者点数为偶数.

定义 5.5 (积事件) 事件"A 与 B **同时**发生"称为 A 与 B 的**积事件**, 记作 $A\cap B$ 或 AB.

知乎@余命数

例如, 对于上述事件 A,B, 积事件 $A\cap B=\{2,4\}$, 表示点数既不大于 4 又是偶数.

知乎@余命数

定义 5.6 (互斥事件) 若事件 A 与 B **不能同时**发生, 则称它们为**互斥事件**或**互不相容事件**. 在集合上表现为它们的交集是空集, 即 $A\cap B=\emptyset$.

例如, 事件 $B=\{2,4,6\}$ 与事件 $D=\{1,3,5\}$ 互斥(前者为偶数, 后者为奇数), 因为 $B\cap D=\emptyset$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定义 5.7 (差事件) 事件“A 发生而事件 B **不发生**”称为 A 与 B 的**差事件**, 记作 $A-B$. 它等价于 $A\cap\bar{B}$.

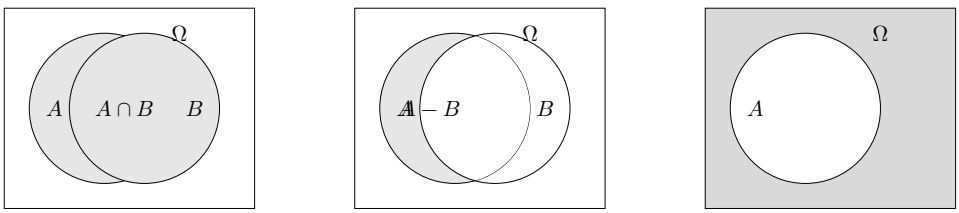
例如, 令 $A=\{1,2,3,4\}, B=\{2,4,6\}$. 那么差事件 $A-B=\{1,3\}$, 表示点数不大于 4 却不是偶数.

定义 5.8 (对立事件) 事件“A **不发生**”称为 A 的**对立事件**或**逆事件**, 记作 $\bar{A}$. 它由样本空间 Ω 中所有不属于 A 的元素构成.

事件 A 与其对立事件 $\bar{A}$ 满足 $A\cup\bar{A}=\Omega$ 且 $A\cap\bar{A}=\emptyset$.A 和 $\bar{A}$ 既互斥又完备, 它们构成对样本空间的一个划分.

例如, 若事件 $A=\{1,2,3,4\}$ 表示点数不大于 4, 则其对立事件为 $\bar{A}=\{5,6\}$ (点数大于 4).

这些事件间的关系和运算, 可用韦恩图直观地表示.



$A\cup B$(整个涂色区域)　　$A-B=A\cap\bar{B}$　　对立事件 $\bar{A}$

图 5.3　事件运算的韦恩图表示

事件的运算满足与集合运算完全相同的定律, 例如交换律、结合律、分配律以及德·摩根定律.

定理 5.1 (德·摩根定律) 对于任意两个事件 A 和 B:

1. $\overline{A\cup B}=\bar{A}\cap\bar{B}$　(A 或 B 都不发生等价于 A 不发生且 B 不发生)
2. $\overline{A\cap B}=\bar{A}\cup\bar{B}$　即 A 和 B 不都发生等价于 A 不发生或 B 不发生

证明 对任意样本点 $\omega\in\Omega$, 由补集与并集的定义依次推得

$$\omega\in\overline{A\cup B}\iff\omega\notin A\cup B\iff(\omega\notin A)\text{ 且}(\omega\notin B)\iff\omega\in\bar{A}\cap\bar{B}.$$

该等价对一切 ω 成立, 故 $\overline{A\cup B}=\bar{A}\cap\bar{B}$.

交换 $\cup$ 与 $\cap$ 的角色, 同理有

$$\omega\in\overline{A\cap B}\iff\omega\notin A\cap B\iff(\omega\notin A)\text{ 或}(\omega\notin B)\iff\omega\in\bar{A}\cup\bar{B},$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

从而 $\overline{A \cap B} = \bar{A} \cup \bar{B}$. □

例题 5.5 高中生李华决定报名参加学校的社团活动, 有三种选择, 即 ACG 社、辩论社与虚拟联合国, 依次记作 A、B、C. 设 A, B, C 分别表示李华报名参加了 ACG 社、报名参加了辩论社、报名参加了虚拟联合国这三个事件. 请使用 A, B, C 的运算关系表示下列事件. (1) 李华至少报名参加了一个社团. (2) 李华同时报名参加了全部三个社团. (3) 李华报名参加了 ACG 社, 但没有报名参加辩论社. (4) 李华三个社团都没有报名参加. (5) 李华恰好只报名参加了一个社团.

解 本题要点在于理解或、且、非等逻辑词与事件运算的对应关系.

(1) 至少报名参加一个社团意味着参加了 A 或参加了 B 或参加了 C. 这是三个事件的和事件.

$$A \cup B \cup C$$

(2) 同时报名参加了全部三个社团意味着参加了 A 且参加了 B 且参加了 C. 这是三个事件的积事件.

$$A \cap B \cap C \quad \text{或简记为} \quad ABC$$

(3) “报名参加了 ACG 社, 但没有报名参加辩论社”意味着事件 A 发生且事件 B 不发生, 即 $\bar{B}$ 发生.

$$A \cap \bar{B} \quad \text{或记为} \quad A - B$$

(4) 三个社团都没有报名参加是至少报名参加一个社团的对立事件.

$$\overline{A \cup B \cup C}$$

根据德·摩根定律, 这也可以表示为没有参加 A 且没有参加 B 且没有参加 C.

$$\bar{A} \cap \bar{B} \cap \bar{C}$$

(5) 恰好只报名参加一个社团是一个更复杂的逻辑, 它由三个互斥的部分组成, 即只参加了 A、只参加了 B、只参加了 C. 只参加了 A 意味着“参加了 A, 且没参加 B, 且没参加 C”, 即 $A\bar{B}\bar{C}$. 因此, 该事件为这三个互斥事件的和事件.

$$(A\bar{B}\bar{C}) \cup (\bar{A}B\bar{C}) \cup (\bar{A}\bar{B}C)$$

这个练习把日常语言的逻辑结构分解, 再映射到集合运算.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

例题 5.6　某电子设备由两个部件构成. 令事件 A 表示部件 1 在一年内失效, 事件 B 表示部件 2 在一年内失效. 已知 $P(A)=0.1$、$P(B)=0.08$、$P(AB)=0.02$. 求该设备在一年内能正常工作的概率.

解　设备能正常工作, 当且仅当部件 1 没有失效且部件 2 没有失效, 即事件 $\bar{A}\cap\bar{B}$.

直接计算 $P(\bar{A}\cap\bar{B})$ 较为困难, 转而利用其与和事件的关系. 由德·摩根定律

$$\bar{A}\cap\bar{B}=\overline{A\cup B},$$

其直观意义是两个部件都没坏恰好为至少有一个部件坏了的对立事件, 于是

$$P(\bar{A}\cap\bar{B})=P(\overline{A\cup B})=1-P(A\cup B).$$

再由广义加法公式 $P(A\cup B)=P(A)+P(B)-P(AB)$ 代入已知数据,

$$P(A\cup B)=0.1+0.08-0.02=0.16,$$

故设备正常工作的概率为

$$P(\text{正常工作})=1-P(A\cup B)=1-0.16=0.84.$$

例题 5.7　某班级有 50 名学生, 参加了数学、物理、化学三科的兴趣小组. 已知参加数学的有 30 人, 物理 25 人, 化学 20 人. 同时参加数学和物理的有 15 人, 数学和化学 10 人, 物理和化学 8 人. 三科都参加的有 5 人. 问题如下. (1) 至少参加一科的有多少人? (2) 恰好只参加一科的有多少人?

解　设 M,P,C 分别表示参加数学、物理、化学兴趣小组的学生构成的集合. 这是一个典型的容斥原理问题.

(1) 至少参加一科的人数即为 $\text{card}(M\cup P\cup C)$. 根据三事件的容斥原理公式, 我们先把单个事件的基数相加, 再减去两两相交事件的基数, 最后加上三者相交事件的基数.

$$\begin{aligned}\text{card}(M\cup P\cup C)=&(\text{card}(M)+\text{card}(P)+\text{card}(C))\\&-(\text{card}(MP)+\text{card}(MC)+\text{card}(PC))\\&+\text{card}(MPC)\end{aligned}$$

代入数据.

$$\begin{aligned}\text{card}(M\cup P\cup C)&=(30+25+20)-(15+10+8)+5\\&=75-33+5=47\end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

所以至少参加一科的有 47 人;另外 3 人什么都没参加.

(2) 恰好只参加一科的人数需要从总数中剔除参加两科及三科的人,利用韦恩图辅助思考会非常清晰. 设 N_1, N_2, N_3 分别为恰好参加一科、两科、三科的人数.

三科都参加的人数 $N_3 = \text{card}(MPC) = 5$ 人. 在此基础上求恰好参加两科的人数. 恰好参加数学和物理两科,即参加这两科但不参加化学的人数为

知乎@余命数

$$\text{card}(MP) - \text{card}(MPC) = 15 - 5 = 10$$

同理,恰好参加物理和化学两科的人数为 $8 - 5 = 3$,恰好参加数学和化学两科的人数为 $10 - 5 = 5$,故恰好参加两科的总人数 $N_2 = 10 + 3 + 5 = 18$ 人.

再由 N_3 与 N_2 反求恰好只参加一科的人数. 恰好只参加数学一科的人数,等于参加数学的总人数减去所有与数学有交集的恰好两科和三科的人数.

$$\begin{aligned} &30 - (\text{恰好数物}) - (\text{恰好数化}) - (\text{三科都参加}) \\ &= 30 - 10 - 5 - 5 = 10 \end{aligned}$$

同理,恰好只参加物理一科的人数为 $25 - 10 - 3 - 5 = 7$,恰好只参加化学一科的人数为 $20 - 5 - 3 - 5 = 7$,故恰好只参加一科的总人数 $N_1 = 10 + 7 + 7 = 24$ 人.

检验.$N_1 + N_2 + N_3 = 24 + 18 + 5 = 47$,与 (1) 的结果吻合.

例题 5.8 证明对于任意两个事件 A, B,以下等式恒成立.

$$(A \cup B) \cap (A \cup \bar{B}) = A$$

解 我们用分配律来证明.

把事件的并集与交集分别类比为数的加法与乘法,分配律 $x(y+z) = xy + xz$ 在此对应于 $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$,因而公共因子 A 可提取. 整个化简只需三步.

$$\begin{aligned} (A \cup B) \cap (A \cup \bar{B}) &= A \cup (B \cap \bar{B}) && \text{(分配律)} \\ &= A \cup \emptyset && \text{(对立事件} B \cap \bar{B} = \emptyset\text{)} \\ &= A && \text{(空集与任意事件的并集为自身).} \end{aligned}$$

知乎@余命数

我们用逻辑分析来证明.

分析事件 $(A \cup B) \cap (A \cup \bar{B})$ 发生的条件. 该事件发生意味着 $(A \cup B)$ 与 $(A \cup \bar{B})$ 同时发生. 分两种情况.

- **若事件 A 发生**,则 $A \cup B$ 与 $A \cup \bar{B}$ 均必然发生,整个积事件发生.
- **若事件 A 不发生,即 $\bar{A}$ 发生**,则要使 $A \cup B$ 发生须 B 发生,要使 $A \cup \bar{B}$ 发生须 $\bar{B}$ 发生,二者不能同时成立.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

该复合事件发生的充要条件即事件 A 发生, 故它等价于事件 A.

熟练运用事件的运算法则, 可以像代数一样简洁地化简和证明复杂的事件关系.

例题 5.9 从一副标准的 52 张扑克牌中随机抽取一张. 定义下列事件. 设 A 表示抽到红桃. 设 B 表示抽到黑桃. 设 C 表示抽到红色牌, 即红桃或方片. 设 D 表示抽到 J、Q、K 中的一张(人头牌).
请判断下列说法的正误. (1) 事件 A 与 B 是互斥的吗? (2) 事件 A 与 C 是互斥的吗? (3) 事件 A 与 B 是对立事件吗? (4) 事件 C 与 $\bar{C}$ 是对立事件吗? (5) 事件 A 与 D 是互斥的吗?

解 此题旨在厘清互斥与对立这两个容易混淆的概念的区别.

(1) **正确**. 一张牌不能既是红桃又是黑桃, 所以它们的交集为空, 即 $A \cap B = \emptyset$.

(2) **错误**. 红桃本身就是红色的牌. 事件 A 的发生必然导致事件 C 的发生, 即 $A \subseteq C$. 因此它们的交集 $A \cap C = A \neq \emptyset$.

(3) **错误**. 对立事件必须满足两个条件. 第一, 它们互斥 ($A \cap B = \emptyset$); 第二, 它们的并集是整个样本空间 ($A \cup B = \Omega$). 从 (1) 可知, 事件 A 和 B 确实满足互斥的条件. 但是, 它们的并集 $A \cup B$ 代表抽到红桃或黑桃, 这并未包含所有可能性, 例如方片和梅花. 因此 $A \cup B \neq \Omega$. 由于第二个条件不满足, 它们不是对立事件.

(4) **正确**. 我们验证对立事件的两个条件. 事件 C 是抽到红色牌, 其对立事件 $\bar{C}$ 就是没有抽到红色牌, 即抽到黑色牌. (1) 一张牌不能既红又黑, 所以 $C \cap \bar{C} = \emptyset$, 即两事件互斥. (2) 任何一张牌非红即黑, 所以 $C \cup \bar{C} = \Omega$(并集为全集). 两个条件都满足, 因此它们是对立事件.

(5) **错误**. 存在一张牌既是红桃又是人头牌, 即红桃 J、红桃 Q、红桃 K. 因此 $A \cap D \neq \emptyset$.

互斥只要求事件不能同时发生 ($A \cap B = \emptyset$). 对立则更为严格, 它在互斥的基础上, 还要求两事件的并集穷尽所有可能性 ($A \cup B = \Omega$). 任何对立事件必然互斥, 但互斥事件不一定对立.

至此, 事件的形式化描述已经建立. 我们把原本不确定的概念**事件**转化为严谨的数学对象**集合**. 接下来为每一个事件 A 赋予一个数值 $P(A)$, 用以量化它发生的可能性. 这就是概率.

5.1.3 概率的定义

知乎@余命数

在详细分析概率的严格数学体系之前, 我们先回到一个更符合直觉的问题. 抛掷均匀硬币、正面朝上的概率是 $\frac{1}{2}$, 这句话究竟意味着什么, 需要仔细说明.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

5.1.3.1 概率的统计定义

最朴素、也最符合科学实验精神的一种想法是, 概率是一个事件在大量重复试验中发生的**频率的稳定值**.

具体地, 在相同条件下对一个随机试验重复进行 n 次, 其中事件 A 发生了 n_A 次. 我们称比值 $f_n(A)=\frac{n_A}{n}$ 为事件 A 在这 n 次试验中发生的**频率**. 大量的实践经验表明, 当试验次数 n 越来越大时, 这个频率 $f_n(A)$ 会表现出稳定性, 它会围绕着某个固定的数值摆动, 且摆动的幅度越来越小. 这个现象被称为**大数定律**. 于是, 一个自然的定义便产生了.

> **定义 5.9 (概率的统计定义)** 在不变的条件下, 重复进行 n 次试验, 事件 A 发生的频率 $f_n(A)$ 随着 n 的增大稳定于某一常数 p. 我们就称这个常数 p 为事件 A 的概率, 记作 $P(A)=p$.

这个定义将概率与可观测的物理世界联系起来, 非常直观. 但要充当数学定义, 它存在一些基本的困难.

首先, 它依赖于频率会稳定于一个常数这个事实; 但这个事实本身需要用概率论的理论 (大数定律) 来证明. 用一个需要被证明的结论来作定义, 这在逻辑上不够严谨.

其次, 我们永远无法做无穷多次试验以求那个极限值. 任何实际的频率都只是概率的一个**近似值**.

最后, 也是对于实践最不利的, 对于很多现实问题, 试验根本不具备重复进行的可能性. 例如, 明天会下雨的概率或某支球队赢得下届世界杯的概率, 这些事件都是一次性的, 无法重复一万个明天或一万个下届世界杯来统计频率.

由于这些缺陷, 统计定义不能作为概率论的逻辑基础. 于是数学家们尝试从另一个角度出发.

5.1.3.2 古典概型及其不足

历史上最早取得成功的数学定义是**古典概型**, 它避开了重复试验, 转而分析试验结果的内在对称性. 它基于一个非常直观的思想. 如果一个试验有 n 个**等可能**的基本结果, 而事件 A 包含了其中的 m 个结果, 那么事件 A 发生的概率就是 $P(A)=\frac{m}{n}$. 这种定义简洁、易于理解, 并且在我们处理像掷骰子、抽扑克牌这类理想化问题时非常有效.

然而, 古典概型的应用范围同样有限. 它要求样本空间是有限的, 并且依赖于等可能性这个无法被定义和验证的前提. 对于一枚质地不均匀的硬币, 或一个包含无穷结果的试验, 如从区间 $[0,1]$ 内随机取一个数, 古典概型便无能为力.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

5.1.3.3 概率的公理化定义

为了克服上述定义的局限，我们需要一个更具普适性的概率定义. 这个定义从一些最基本、不证自明的公理出发，用逻辑演绎的方式构建起整个概率论的体系. 这就是苏联数学家安德雷·柯尔莫戈洛夫在 1933 年提出的**概率公理化定义**.

它不关心概率的数值从何而来(可以来自频率统计，也可以来自对称性假设)，只关心这些数值必须遵守哪些基本的规则.

定义 5.10 (概率的公理化定义) 设 E 是一个随机试验，Ω 是其样本空间. 对于 Ω 中的每一个事件 A，赋予一个实数，记为 $P(A)$，称为事件 A 的**概率**. 如果这个赋值函数 $P(\cdot)$ 满足以下三条公理，则称之为一个**概率测度**.

1. **非负性公理**对于每一个事件 A，有 $P(A) \geq 0$.
2. **规范性公理**必然事件的概率为 1，即 $P(\Omega) = 1$.
3. **可加性公理**若 $A_1, A_2, \ldots$ 是一列**两两互斥**的事件，即对于任意 $i \neq j$，都有 $A_i \cap A_j = \emptyset$，则有

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots \tag{5-1}$$

对于有限个互斥事件，此公理同样成立；$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$.

柯尔莫戈洛夫的定义规定了任何一种合理的概率度量**必须遵守的根本法则**.

首先，“对于…每一个事件 A，赋予一个实数…记为 $P(A)$”这句话，实质上定义了一个**函数**. 我们在高中阶段所熟悉的函数，其定义域和值域通常都是数的集合，例如 $f(x) = x^2$. 而这里的概率 $P(\cdot)$ 是一个更广义的函数. 它的**定义域**是样本空间 Ω 的所有事件构成的集合，函数的自变量是一个**集合**；其**值域**则是实数集 $\mathbb{R}$(根据公理，实际上是区间 $[0, 1]$). 因此，$P(A)$ 的写法与 $f(x)$ 完全对应，它表示将事件 A 这个输入，映射到它的概率值这个输出.

而概率测度这一术语，意为测量. 就像长度测量一维线段、面积测量二维图形一样，概率测度就是一套用来**测量**一个事件发生可能性大小的规则. 任何满足这三条公理的函数 $P(\cdot)$，都被认为是一种合法的、自洽的测量方式.

5.1.3.4 三条公理的内涵

5.1.3.4.1 非负性 这是最基本的要求. 一个事件发生的可能性，最低对应概率为 0，即事件不发生. 此公理为概率设定了下界.

5.1.3.4.2 规范性 这是为测量设定一个**单位尺度**. 规范性的意思就是标准化. 将必然发生这件事的确定性程度规定为 1(或者说 100%). 于是，所有其他事件的概

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

率都成为了相对于这个完全确定的基准的一个比例,自然地落在了 0 和 1 之间. 就像将 1 米定义为一个标准长度一样,这里将完全的确定性标定为 1.

5.1.3.4.3 可加性 对于一系列两两互斥的事件 A_i, 其并集的概率等于它们各自概率的和. 这是三条公理中最实用的一条, 它将整体的概率与局部的概率联系起来.

概率的公理化定义, 实质上是为可能性这个模糊概念建立了一个数学模型. 它声称,任何一个函数,只要它取非负值,把总的可能性校准为 1,并且允许将互斥可能性的概率进行相加,也就是满足三条公理,那么它就是一个有效的概率模型.

这三条公理形式简单, 却是概率论的基础; 所有我们熟知的概率性质都可以从它们推导出来.

性质 不可能事件的概率为零.

$$P(\emptyset) = 0$$

知乎@余命数

证明 考虑必然事件 Ω 与不可能事件 $\emptyset$. 二者互斥 $(\Omega \cap \emptyset = \emptyset)$,且并集为 $\Omega \cup \emptyset = \Omega$. 由公理 3(可加性),

$$P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset).$$

代入 $\Omega \cup \emptyset = \Omega$,上式化为 $P(\Omega) = P(\Omega) + P(\emptyset)$;再由公理 2, $P(\Omega) = 1$,得

$$1 = 1 + P(\emptyset),$$

故 $P(\emptyset) = 0$. □

知乎@余命数

反之则不一定成立. 概率为 0 的事件不一定是不可能事件. 例如, 在从区间 $[0,1]$ 上随机取一个数的试验中,取到数字 0.5 的概率是 0,但它显然是一个可能发生的结果.

性质 对于任意事件 A,其对立事件 $\bar{A}$ 的概率为

$$P(\bar{A}) = 1 - P(A)$$

这个性质十分常用,因为直接计算事件 A 的概率往往比较复杂,而计算其对立事件 $\bar{A}$ 的概率则比较简单.

证明 对任意事件 A 与其对立事件 $\bar{A}$, 二者互斥 $(A \cap \bar{A} = \emptyset)$ 且完备 $(A \cup \bar{A} = \Omega)$. 由公理 3(可加性),

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

又 $A \cup \bar{A} = \Omega$, 由公理 2, $P(A \cup \bar{A}) = P(\Omega) = 1$, 故

$$1 = P(A) + P(\bar{A}),$$

移项即得 $P(\bar{A}) = 1 - P(A)$. □

性质 对于任意事件 A, 其概率介于 0 和 1 之间.

$$0 \leq P(A) \leq 1$$

证明 由公理 1(非负性)直接得 $P(A) \geq 0$.

为证 $P(A) \leq 1$, 套用刚证得的对立事件概率性质

$$P(A) = 1 - P(\bar{A}).$$

公理 1 保证 $P(\bar{A}) \geq 0$, 故

$$P(A) = 1 - P(\bar{A}) \leq 1 - 0 = 1,$$

即 $0 \leq P(A) \leq 1$. □

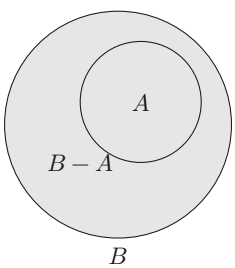
性质 若事件 A 包含于事件 B ($A \subseteq B$), 则事件 A 的概率不大于事件 B 的概率.

$$P(A) \leq P(B)$$

并且, 事件差的概率为

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

证明 当 $A \subseteq B$ 时, 事件 B 可分解为两个互斥部分的并集. 这两个部分分别是事件 A 与事件 $B - A$, 即属于 B 但不属于 A 的部分.



知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

$B = A \cup (B - A)$, 且 $A \cap (B - A) = \emptyset$. 由公理 3 的可加性对此不交并分解,

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

移项得事件差的概率公式 $P(B - A) = P(B) - P(A)$. 又由公理 1（非负性）有 $P(B - A) \geq 0$, 代入上式即 $P(B) - P(A) \geq 0$, 故

$$P(A) \leq P(B).$$

□

性质 对于任意两个事件 A 和 B, 它们并集的概率为

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

证明 为套用公理 3, 将事件 $A \cup B$ 分解为互斥事件的并. 由韦恩图可见 $A \cup B$ 可写成三个互斥区域 $A - B$、$B - A$、$A \cap B$ 的并集; 更简洁的分解是

$$A \cup B = A \cup (B - A),$$

其中 A 与 $B - A$ 互斥. 由公理 3,

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

另一方面, $B = (A \cap B) \cup (B - A)$ 且 $A \cap B$ 与 $B - A$ 互斥, 故 $P(B) = P(A \cap B) + P(B - A)$, 即 $P(B - A) = P(B) - P(A \cap B)$. 代回上式即得

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

□

这个公式表明, 当简单地将两个事件的概率相加时, 我们把它们**共同发生**的部分 $(A \cap B)$ 计算了两遍, 因此必须减去一次. 这个思想可以推广到任意多个事件, 并引出著名的**容斥原理**. 例如, 对于三个事件 A, B, C, 有

$$P(A \cup B \cup C) = \sum P(A) - \sum P(AB) + P(ABC)$$

其中 $\sum P(A) = P(A) + P(B) + P(C)$, $\sum P(AB) = P(AB) + P(AC) + P(BC)$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

5.2 经典概率问题

我们这一节来看两个著名的古典概率问题, 即生日问题与错排问题. 为求解它们, 我们会用到上一节建立的全部工具, 包括样本空间的建模、对立事件与容斥原理.

注 5.1 (集合符号提示)
处理多个事件的关系时, 逐一写出 $A_1 \cup A_2 \cup \cdots \cup A_n$ 会非常繁琐. 为此我们引入一套简洁的符号, 用来表示多个事件的并集与交集.
已知 $A \cup B$ 表示 A 或 B 发生, $A \cap B$ 表示 A 与 B 同时发生. 我们把它推广到 n 个事件 $A_1, A_2, \ldots, A_n$, 于是得到以下记号.

- 大并集符号 $\bigcup$

$$\bigcup_{i=1}^{n} A_i \quad \text{读作从 i 等于 1 到 n, } A_i \text{ 的并集}$$

它等价于 $A_1 \cup A_2 \cup \cdots \cup A_n$, 所代表的事件是 $A_1, A_2, \ldots, A_n$ 中**至少有一个**发生.

- 大交集符号 $\bigcap$

$$\bigcap_{i=1}^{n} A_i \quad \text{读作从 i 等于 1 到 n, } A_i \text{ 的交集}$$

它等价于 $A_1 \cap A_2 \cap \cdots \cap A_n$, 所代表的事件是 $A_1, A_2, \ldots, A_n$ **全部都**发生.

5.2.1 生日问题 *

我们先看一个经典的问题.

例题 5.10 (生日问题) 在一个有 n 个人的房间里 ($n \leq 365$), 假设每个人的生日等可能地分布在一年 365 天中的任何一天, 且不考虑闰年. 求至少有两个人生日相同的概率是多少? 计算当 $n = 23$ 时, 这个概率大约是多少.

解 直接计算至少有两人生日相同的概率颇为困难, 因为它混合了恰好两人相同、恰好三人相同、两对不同的人生日相同等多种情形. 于是我们转求它的对立事件 A, 即"所有人的生日都不同".

样本空间 Ω 是 n 个人所有可能的生日组合, 每人都有 365 种可能, 故 $\text{card}(\Omega) = 365^n$. 事件 A 的样本点数相当于从 365 天中选出 n 个不同生日并分配给 n 个人, 即

$$\text{card}(A) = P_{365}^{n} = 365 \times 364 \times \cdots \times (365 - n + 1),$$

因而

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{P_{365}^{n}}{365^n} = \frac{365}{365} \times \frac{364}{365} \times \cdots \times \frac{365 - n + 1}{365}.$$

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

所求的至少有两人生日相同的概率为

$$P(\bar{A}) = 1 - P(A) = 1 - \prod_{k=0}^{n-1} \left(1 - \frac{k}{365}\right).$$

代入 $n = 23$, 由 $P(A) = \frac{P_{365}^{23}}{365^{23}} \approx 0.4927$ 得

$$P(\bar{A}) = 1 - P(A) \approx 1 - 0.4927 = 0.5073.$$

仅 23 人的房间里, 至少两人生日相同的概率已超过 50%. 这是一个反直觉的结果, 其根源在于配对数量随 n 二次增长.

我们常感到意外, 因为直觉倾向于**线性思考**. 它容易把问题理解为我与另外 22 个人生日相同的概率, 或者简单地比较人数 23 与天数 365, 从而认为概率应当很小.

问题的真正要点是**任意两个人**生日相同的概率. 任意二字把组合学引入问题. 对 n 个人, 总共有 $\binom{n}{2} = \frac{n(n-1)}{2}$ 个可能的配对. 当 $n = 23$ 时, 总共有

$$\binom{23}{2} = \frac{23 \times 22}{2} = 253 \text{ 个配对.}$$

于是, 一个只有 23 人的房间里, 存在 **253** 个可能出现生日相同的配对. 单个配对生日相同的概率很低, 但配对数如此之多, 至少出现一对相同的概率便相当高. 直觉没有捕捉到由人数增加导致的**配对数量的二次方增长**. 从上面的解法可见, 直接处理至少一词非常复杂; 转向其对立事件(所有人生日都不同), 则使问题简化.

事件 A(所有人生日都不同)的概率可以写成一个连乘积.

$$P(A) = 1 \times \left(1 - \frac{1}{365}\right) \times \left(1 - \frac{2}{365}\right) \times \ldots$$

每个乘上的因子都小于 1. 多个小于 1 的数相乘, 结果会**迅速减小**. 这种衰减是指数性的, 比直觉中的线性减少快得多. 因此 $P(A)$ 迅速从 1 降到 0.5 以下, 其对立事件 $P(\bar{A})$ 便迅速攀升超过 0.5.

生日问题的模型在计算机科学中有一个重要应用, 即**哈希碰撞**.

哈希函数把任意长度的数据, 如一个文件、一张图片映射为固定长度的、看似随机的字符串, 我们称这个字符串为**哈希值**, 它相当于数据的数字指纹. **碰撞**则是指两个**不同**的原始数据经过哈希函数计算后得到**完全相同**的哈希值.

哈希值的可能空间 (对应生日问题中的 365 天) 通常非常巨大, 例如 2^{128} 或 2^{256}. 然而生日问题表明, 我们并不需要 2^{128} 个不同文件才会看到明显的碰撞风险. 按生日问题的近似, 当文件数量达到 $\sqrt{2^{128}} = 2^{64}$ 量级时, 碰撞概率已经不可忽视.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这一原理是著名的**生日攻击**的基础. 攻击者可以利用这一概率原理, 以远低于暴力破解的计算量伪造数字签名或寻找加密算法的弱点. 因此, 设计安全的密码系统时必须考虑生日问题所揭示的碰撞风险.

5.2.2 错排问题

例题 5.11 (德·蒙莫尔问题) n 位客人参加晚宴, 他们都把自己的帽子交给了衣帽间的服务员. 晚宴结束后, 服务员由于匆忙, 把 n 顶帽子随机地还给了每位客人. 求没有任何一位客人拿到自己帽子的概率 p_n. 当 $n \to \infty$ 时, 这个概率收敛到何值?

这是一个经典的概率与组合问题. 我们从古典概型出发求解它.

先确定样本空间的总大小. 服务员把 n 顶不同的帽子随机还给 n 位客人, 总的排列方式有 $n!$ 种.

知乎@余命数

$$\text{card}(\Omega) = n!$$

接下来, 我们需要计算事件 A 所包含的样本点数, 其中事件 A 表示没有任何一位客人拿到自己的帽子. 这是一个复杂的计数问题, 因此我们先用严格的方式定义这一类排列.

定义 5.11 (错排与错排数) 设一个 n 个元素的排列中没有任何元素出现在它原来的位置上, 我们就称它为一个 n 元**错排**. 这类排列的总数称为**错排数**, 记为 D_n 或 $!n$.

例如, 当 $n = 3$ 时, 集合 $\{1, 2, 3\}$ 的全排列共有 $3! = 6$ 种, 其中错排是 $(2, 3, 1)$ 与 $(3, 1, 2)$, 因此 $D_3 = 2$.

知乎@

根据上面的定义, 我们要求的样本点数就是 D_n. 于是原概率问题化为求 D_n 的通项公式.

$$p_n = P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{D_n}{n!}$$

知乎@余命数

我们用两种方法求 D_n.

解

直接计算所有人都错位的排列数 D_n 很困难. 我们先算其对立面, 即至少有一人拿对自己帽子的排列数, 再用总排列数 $n!$ 减去它.

令 A_i 为“第 i 位客人拿到自己帽子”的排列所构成的集合 $(i = 1, \dots, n)$, 我们需要计算 $|A_1 \cup A_2 \cup \cdots \cup A_n|$. 由容斥原理,

$$|A_1 \cup \cdots \cup A_n| = \sum |A_i| - \sum_{i<j} |A_i \cap A_j| + \cdots + (-1)^{n-1} |A_1 \cap \cdots \cap A_n|.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

其中 $|A_i|$ 表示第 i 位客人拿对帽子而其余 $n-1$ 人任意排，有 $(n-1)!$ 种，这样的集合有 $\binom{n}{1}$ 个；$|A_i \cap A_j|$ 表示第 i,j 位客人拿对帽子而其余 $n-2$ 人任意排，有 $(n-2)!$ 种，这样的集合有 $\binom{n}{2}$ 个. 以此类推，k 位指定客人拿对帽子的排列数为 $(n-k)!$，共有 $\binom{n}{k}$ 组.

由此至少一人拿对的总数为

$$\binom{n}{1}(n-1)! - \binom{n}{2}(n-2)! + \cdots + (-1)^{n-1}\binom{n}{n}(n-n)!$$

$$= \frac{n!}{1!(n-1)!}(n-1)! - \frac{n!}{2!(n-2)!}(n-2)! + \cdots = \frac{n!}{1!} - \frac{n!}{2!} + \frac{n!}{3!} - \cdots + \frac{(-1)^{n-1}n!}{n!},$$

故错排数 D_n 为

$$\begin{aligned} D_n &= n! - \left(\frac{n!}{1!} - \frac{n!}{2!} + \cdots + \frac{(-1)^{n-1}n!}{n!}\right) \\ &= \frac{n!}{0!} - \frac{n!}{1!} + \frac{n!}{2!} - \cdots + \frac{(-1)^n n!}{n!} = n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}. \end{aligned}$$

分析第 n 个元素的位置还可建立 D_n 的递推关系. 考虑元素 n，在一个错排中它不能在第 n 位，只能在其余 $n-1$ 个位置之一，设其在第 k 位 $(k \neq n)$，这有 $n-1$ 种选择. 对此选择再就元素 k 的位置分两种情形. 若元素 k 恰好放在第 n 位，相当于 k 与 n 互换位置，剩下 $n-2$ 个元素需完全错排，方法数为 D_{n-2}；若元素 k 不在第 n 位，则 n 已占第 k 位，对剩下 $n-1$ 个元素与 $n-1$ 个位置，其中元素不含 n、位置不含 k，问题化为将元素 $j \in \{1,\ldots,n-1\}, j \neq k$ 放在非 j 号位且将元素 k 放在非 n 号位，等价于一个 $n-1$ 元错排，方法数为 D_{n-1}. 由乘法与加法原理得二阶递推关系

$$D_n = (n-1)(D_{n-1} + D_{n-2}) \quad (n \geq 3),$$

初始值 $D_1 = 0, D_2 = 1$. 由此递推亦可推出通项公式.

回到最初的概率问题，代入 D_n 的通项公式，

$$p_n = \frac{D_n}{n!} = \frac{n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}}{n!} = \sum_{k=0}^{n}\frac{(-1)^k}{k!}.$$

注 5.2

由 $p_n = \sum_{k=0}^{n}\frac{(-1)^k}{k!}$ 可直接算出前几项，$p_2 = 0.5$、$p_3 \approx 0.333$、$p_4 = 0.375$、$p_5 = 0.3666\ldots$，数值上很快接近 0.37. 通项公式与递推关系解决了有限 n 的错排计数；极限常数见下.

知乎关注@余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

错排概率极限 $1/e$

当 $n\to\infty$ 时, 部分和 $\sum_{k=0}^{n}\frac{(-1)^k}{k!}$ 收敛到 e^x 在 $x=-1$ 处的泰勒级数.

$$\lim_{n\to\infty} p_n=\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}=e^{-1}=\frac{1}{e}\approx 0.36788.$$

因而大规模宴会中无人拿对帽子的概率趋于常数 $1/e$. 级数与余项的严格语言见第四卷第 16 章.

5.2.2.1 错排数的性质与公式

上述例题给出了关于错排数的一些重要结论, 我们把它汇总如下.

定理 5.2 (错排数的通项公式) n 元错排数 D_n 的显式通项公式为

$$D_n=n!\sum_{k=0}^{n}\frac{(-1)^k}{k!}=\frac{n!}{0!}-\frac{n!}{1!}+\frac{n!}{2!}-\cdots+(-1)^n\frac{n!}{n!} \tag{5-2}$$

证明 设全集为 n 个元素的全排列集合, 其大小为 $n!$. 对每个 $i=1,2,\dots,n$, 令 A_i 表示“第 i 个元素仍在第 i 个位置”的排列集合. 错排就是不属于任何 A_i 的排列, 因而

$$D_n=n!-|A_1\cup A_2\cup\cdots\cup A_n|.$$

若指定 r 个元素固定在原位, 其余 $n-r$ 个元素可任意排列, 因此对应排列数为 $(n-r)!$. 这样的 r 个固定元素集合有 $\binom{n}{r}$ 种. 由容斥原理,

$$|A_1\cup\cdots\cup A_n|=\sum_{r=1}^{n}(-1)^{r-1}\binom{n}{r}(n-r)!.$$

所以

$$D_n=n!+\sum_{r=1}^{n}(-1)^r\binom{n}{r}(n-r)!=\sum_{r=0}^{n}(-1)^r\frac{n!}{r!}=n!\sum_{r=0}^{n}\frac{(-1)^r}{r!}.$$

□

定理 5.3 (错排数的递推关系) 错排数 D_n 满足以下两个等价的递推关系, 初始值为 $D_1=0, D_2=1$.

1. **(二阶递推)** $D_n=(n-1)(D_{n-1}+D_{n-2})$, 　对于 $n\ge 3$.
2. **(一阶递推)** $D_n=nD_{n-1}+(-1)^n$, 　对于 $n\ge 2$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

证明 对二阶递推, 在一个 n 元错排中元素 n 不能在第 n 位, 设它被放到第 k 位 $(k \neq n)$, 这有 $n-1$ 种选择. 若元素 k 放在第 n 位, 则 k 与 n 形成互换, 剩下 $n-2$ 个元素必须错排, 有 D_{n-2} 种方法; 若元素 k 不放在第 n 位, 则删去第 k 位和元素 n 后, 剩余问题等价于一个 $n-1$ 元错排, 有 D_{n-1} 种方法. 因此

$$D_n = (n-1)(D_{n-1} + D_{n-2}) \qquad (n \geq 3).$$

一阶递推由通项公式直接导出:

$$D_n = n! \sum_{r=0}^{n} \frac{(-1)^r}{r!} = n\,(n-1)! \sum_{r=0}^{n-1} \frac{(-1)^r}{r!} + (-1)^n = nD_{n-1} + (-1)^n.$$

初始值 $D_1 = 0, D_2 = 1$ 可直接列举得到. □

当 n 较小时, 使用递推关系通常比展开通项公式更快捷. 例如 $D_3 = 2(D_2 + D_1) = 2(1+0) = 2, D_4 = 3(D_3 + D_2) = 3(2+1) = 9$.

极限概率 $D_n/n! \to 1/e$

性质 (极限概率) n 个元素的全排列中, 是一个错排的概率 $p_n = D_n/n!$ 收敛于 $1/e$.

$$\lim_{n\to\infty} \frac{D_n}{n!} = \frac{1}{e} \approx 0.36788 \tag{5-3}$$

证明 由错排数通项公式,

$$\frac{D_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

右端是指数函数泰勒级数 $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ 在 $x = -1$ 处的前 n 项部分和. 因此当 $n \to \infty$ 时,

$$\frac{D_n}{n!} \to \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} = \frac{1}{e}.$$

□

最佳整数近似 $D_n = \lfloor n!/e + 1/2 \rfloor$

注 5.3

小 n 时用递推或通项有限和; 现用 e^{-1} 的交错级数余项说明四舍五入公式.

知乎@余命数 知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

性质 (最佳整数近似) 对于所有正整数 $n \geq 1$, D_n 是最接近 $\frac{n!}{e}$ 的整数. 因此 D_n 可以通过四舍五入计算得到.

$$D_n = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor.$$

这个性质为快速估算甚至精确计算 D_n 提供了一个极为便捷的方法.

证明 由 e^{-1} 的交错级数展开与错排数通项公式,

$$\frac{n!}{e} - D_n = n! \sum_{k=n+1}^{\infty} \frac{(-1)^k}{k!}.$$

交错级数余项的绝对值小于首个被舍去项, 所以

$$\left| \frac{n!}{e} - D_n \right| < \frac{n!}{(n+1)!} = \frac{1}{n+1} \leq \frac{1}{2}.$$

因此整数 D_n 与 $\frac{n!}{e}$ 的距离小于 $1/2$, 它就是距离 $\frac{n!}{e}$ 最近的整数, 等价地

$$D_n = \left\lfloor \frac{n!}{e} + \frac{1}{2} \right\rfloor.$$

□

现在看一些错排理论在具体问题中的应用.

例题 5.12 (信件与信封问题) 将 4 封写好的不同信件, 随机装入写好对应地址的 4 个信封中, 每个信封恰好装一封信. 求 4 封信全部装错信封的装法共有多少种?

解 这是最直接的错排问题. 元素是 4 封信, 位置是 4 个信封, 所求即 4 个元素的错排数 D_4. 代入错排数通项公式, 当 $n = 4$ 时

$$\begin{aligned} D_4 &= 4! \left(\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} \right) \\ &= 24 \left(1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} \right) \\ &= 24 \left(\frac{1}{2} - \frac{1}{6} + \frac{1}{24} \right) \\ &= 12 - 4 + 1 = 9. \end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

或由递推公式, 已知 $D_2=1, D_3=2$,

$$D_4=(4-1)(D_3+D_2)=3\times(2+1)=9.$$

故共有 9 种全部装错的方法.

公式给出答案, 而公式背后正是容斥原理的一次直接应用. 为看清来路, 我们不妨手算 D_4.

总排列数,即样本空间大小为 $4!=24$. 记 A_i 为第 i 封信装对信封 $(i=1,2,3,4)$, 由容斥原理

$$D_4=24-|A_1\cup A_2\cup A_3\cup A_4|,$$

右端 $|A_1\cup\cdots\cup A_4|$ 即至少一封装对的情况数, 按容斥原理逐项展开, 得

- 任选 1 封装对, 其余 3 封任意排, 共 $\binom{4}{1}\times 3!=4\times 6=24$ 种.
- 指定 2 封装对时在上一项中被重复计入, 减去 $\binom{4}{2}\times 2!=6\times 2=12$ 种.
- 指定 3 封装对时前两层分别被加、减各 3 次而净计入 0 次, 补回 $\binom{4}{3}\times 1!=4\times 1=4$ 种.
- 四封全装对累计被计入 $4-6+4=2$ 次, 应保留 1 次, 再减 $\binom{4}{4}\times 0!=1$ 种.

于是至少一封装对的情况数为

$$\binom{4}{1}3!-\binom{4}{2}2!+\binom{4}{3}1!-\binom{4}{4}0!=24-12+4-1=15,$$

从而 $D_4=24-15=9$, 与公式计算吻合.

例题 5.13 (枚举法) 三位同学甲、乙、丙, 分别坐在 1 号、2 号、3 号座位上. 现在让他们重新就座, 要求每个人都不能坐在自己原来的位置上. 请问有多少种不同的坐法?

解 此为 $n=3$ 的错排问题, 总数很少, 用枚举法直接列出即可. 对甲、乙、丙三人作全排列, 括号内按顺序表示坐在 1、2、3 号座位上的人.

1. (甲, 乙, 丙)$\to$ 所有人都坐对了.
2. (甲, 丙, 乙)$\to$ 甲坐对了.
3. (乙, 甲, 丙)$\to$ 丙坐对了.
4. (乙, 丙, 甲)$\to$ **没有任何人坐对, 这是一个错排**.
5. (丙, 甲, 乙)$\to$ **没有任何人坐对, 这是另一个错排**.
6. (丙, 乙, 甲)$\to$ 乙坐对了.

逐项核对, 仅第 4、第 5 种坐法满足“每个人都不能坐在自己原来的位置上”, 故共有 2 种不同的坐法, 验证了 $D_3=2$.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

例题 5.14 (四球四盒问题) 有 4 个写着号码 1, 2、3, 4 的球, 和 4 个对应号码的盒子. 现将 4 个球随机放入 4 个盒子中, 每盒一个. 求至少有一个球放进与自己号码相同的盒子的情况有多少种?

解 所求为至少一个球放对的情况数. 直接分类恰好一个对、恰好两个对等会很繁琐, 转而用容斥原理直接计算并事件. 令 A_i 为"第 i 号球放进第 i 号盒"的事件, 则所求即 $|A_1 \cup A_2 \cup A_3 \cup A_4|$. 由容斥原理,

$$\begin{aligned}|A_1 \cup A_2 \cup A_3 \cup A_4| &= \binom{4}{1}3! - \binom{4}{2}2! + \binom{4}{3}1! - \binom{4}{4}0! \\ &= 4 \times 6 - 6 \times 2 + 4 \times 1 - 1 \times 1 \\ &= 24 - 12 + 4 - 1 = 15,\end{aligned}$$

其中每一项 $\binom{4}{k}(4-k)!$ 表示选定 k 个球放对、其余 $4-k$ 个球任意放的取法数. 故至少一个球放对有 15 种.

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

所有球都放错的情况数 D_4 则是 $24 - 15 = 9$ 种.

例题 5.15 一位粗心的药剂师手边有为 4 位病人准备的 4 种不同药品. 他随机地将这 4 种药品分发给了这 4 位病人, 每人一种. 请问, 恰好只有两位病人拿到了正确的药品的情况有多少种?

解 此为部分错排问题, 可分解为两个独立步骤, 再用乘法原理相乘.

先从 4 位病人中选出拿到正确药品的 2 位, 这是组合问题, 选法数为 $\binom{4}{2} = \frac{4\times3}{2\times1} = 6$ 种. 剩下 2 位病人则需完全错排. 设他们为甲、乙, 对应药品为药 A、药 B, 要让两人都拿错, 只能甲拿药 B 且乙拿药 A, 故 $D_2 = 1$ 种.

知乎@余命数

由乘法原理将两步方法数相乘,

$$\begin{aligned}\text{总数} &= (\text{选出 2 个正确病人的方法数}) \times (\text{剩下 2 个病人错排的方法数}) \\ &= \binom{4}{2} \times D_2 = 6 \times 1 = 6.\end{aligned}$$

故恰好只有两位病人拿到正确药品的情况有 6 种.

例题 5.16 书架上有《红楼梦》《西游记》《三国演义》《水浒传》四本书, 按从左到右的顺序排好. 现将它们取下后随机放回, 求恰好有一本书放回原位的概率.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

解 此题结合古典概型与部分错排. 将 4 本不同的书随机放回, 样本空间总数为

$$\text{card}(\Omega) = 4! = 24.$$

事件恰好有一本书放回原位可分解为两步. 先从 4 本书中选出 1 本放回原位, 有 $\binom{4}{1} = 4$ 种选法, 再让其余 3 本完全错排($D_3 = 2$ 种). 由乘法原理,

$$\text{card}(A) = \binom{4}{1} \times D_3 = 4 \times 2 = 8.$$

代入古典概型公式 $P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$,

$$P(\text{恰好一本对}) = \frac{8}{24} = \frac{1}{3}.$$

故恰好有一本书放回原位的概率是 $\frac{1}{3}$.

知乎@余命数

5.3 条件概率

知乎@余命数

到目前为止, 我们计算概率时总是站在对整个试验结果一无所知的立场上. 现实中, 我们常常已经掌握关于试验结果的部分信息. 例如掷完骰子后有人瞥见点数为偶数, 体检后化验单上写着某项指标异常. 新的信息会改变我们对事件可能性的判断. 这一节研究的问题是, 在已知事件 A 发生的前提下, 事件 B 发生的概率该如何计算.

知乎@余命数

5.3.1 条件概率的定义

我们先看一个引例. 掷一颗骰子, 样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$. 记事件 B 为点数大于 3, 显然

知乎@余命数

$$P(B) = \frac{\text{card}(B)}{\text{card}(\Omega)} = \frac{3}{6} = \frac{1}{2}.$$

现在增加一条信息. 有人提前表明点数为偶数. 记这个事件为 $A = \{2, 4, 6\}$. 在这条信息之下, 点数大于 3 的概率需要重新计算.

信息改变了问题的格局. 既然 A 已经发生, 试验结果只可能落在 $\{2, 4, 6\}$ 之中, 原来的样本空间 Ω 实际上已经**收缩**为 A. 在这三个等可能的结果里, 满足点数大于 3 的只有 4 和 6 两个. 所以我们直观地认为, 所求概率应该是 $\frac{2}{3}$, 它大于原来的 $\frac{1}{2}$.

知乎@余命数

知乎@余命数

把这个过程用符号重写一遍. 在古典概型下, “在 A 发生的条件下 B 发生”的样本点, 恰好构成交集 $A \cap B$; 而新的样本空间就是 A 本身. 于是

$$\text{所求概率} = \frac{\text{card}(A \cap B)}{\text{card}(A)} = \frac{\text{card}(A \cap B)/\text{card}(\Omega)}{\text{card}(A)/\text{card}(\Omega)} = \frac{P(A \cap B)}{P(A)}.$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

最右端的表达式只用到了概率, 不再依赖古典概型的计数结构. 这提示我们把这个比值直接当作条件概率的定义.

定义 5.12 (条件概率) 设 A, B 是样本空间 Ω 中的两个事件, 且 $P(A) > 0$. 称

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

为**在事件 A 发生的条件下事件 B 发生的条件概率**.

定义中要求 $P(A) > 0$ 是自然的. 如果 A 根本不可能发生, 或者它以概率零发生, 那么已知 A 发生这个前提便失去意义. 读法上, $P(B \mid A)$ 读作在 A 发生的条件下 B 的概率, 竖线之后的部分是条件.

我们可以从两个等价的角度理解条件概率, 它们随时切换会带来便利.

- **收缩的样本空间**. 已知 A 发生后, 样本空间从 Ω 缩小为 A, 各样本点的相对比例保持不变, 再重新归一化.
- **概率的比值**. 直接用公式 $P(B \mid A) = P(AB)/P(A)$ 计算, 其中 $P(AB)$ 与 $P(A)$ 都在原样本空间中求得.

例题 5.17 掷一颗骰子, 记 A 为点数为偶数, B 为点数大于 3. 用定义验证引例中的直观结果.

解 有 $A = \{2, 4, 6\}$, $B = \{4, 5, 6\}$, 于是 $A \cap B = \{4, 6\}$.

$$P(A) = \frac{3}{6} = \frac{1}{2}, \qquad P(A \cap B) = \frac{2}{6} = \frac{1}{3}.$$

由条件概率的定义,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/3}{1/2} = \frac{2}{3},$$

与引例中收缩样本空间的算法一致.

例题 5.18 (两个孩子的家庭) 某家庭有两个孩子. 假设男孩与女孩的出生机会均等, 四个结果(依长幼顺序记为男男、男女、女男、女女)等可能. (1) 已知这个家庭至少有一个男孩, 求两个孩子都是男孩的概率. (2) 已知老大是男孩, 求两个孩子都是男孩的概率.

解 记 $B = \{(男, 男)\}$ 为两个孩子都是男孩这一事件.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

(1) 记 A_1 为至少有一个男孩, 则

$$A_1 = \{(男, 男), (男, 女), (女, 男)\}, \qquad P(A_1) = \frac{3}{4}.$$

由于 $B \subseteq A_1$, 有 $B \cap A_1 = B$, 故 $P(B \cap A_1) = \frac{1}{4}$. 所以

$$P(B \mid A_1) = \frac{P(B \cap A_1)}{P(A_1)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

(2) 记 A_2 为老大是男孩, 则 $A_2 = \{(男, 男), (男, 女)\}$, $P(A_2) = \frac{1}{2}$. 所以

$$P(B \mid A_2) = \frac{P(B \cap A_2)}{P(A_2)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

两问的答案不同, 因为两条信息的内容不同. 至少有一个男孩只排除了女女这一种结果, 而老大是男孩还额外排除了女男. 信息越多, 样本空间收缩得越厉害, 概率的改变也越大. 使用条件概率时必须把已知的信息精确地写成事件; 含糊其词会得出错误的答案.

条件概率 $P(B \mid A)$ 中的 A 一旦固定, 它作为 B 的函数, 自身就是一个合法的概率测度.

性质 (条件概率是概率测度) 设 $P(A) > 0$. 令 $Q(B) = P(B \mid A)$, 则 $Q(\cdot)$ 满足概率的三条公理.

1. 对任意事件 B, $Q(B) \geq 0$;
2. $Q(\Omega) = 1$;
3. 若 $B_1, B_2, \ldots$ 两两互斥, 则 $Q(B_1 \cup B_2 \cup \ldots) = Q(B_1) + Q(B_2) + \ldots$.

证明 (1) 由 $P(A \cap B) \geq 0$ 且 $P(A) > 0$, 得

$$Q(B) = \frac{P(A \cap B)}{P(A)} \geq 0.$$

(2) 由于 $\Omega \cap A = A$,

$$Q(\Omega) = \frac{P(\Omega \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$

(3) 若 $B_1, B_2, \ldots$ 两两互斥, 则 $B_1 \cap A, B_2 \cap A, \ldots$ 也两两互斥. 对 P 使用可加性公理,

$$Q\Big(\bigcup_{i=1}^{\infty} B_i\Big) = \frac{P((\bigcup_{i=1}^{\infty} B_i) \cap A)}{P(A)} = \frac{P(\bigcup_{i=1}^{\infty}(B_i \cap A))}{P(A)}$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$= \frac{\sum_{i=1}^{\infty} P(B_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} Q(B_i).$$

□

这条性质带来一批推论. 在条件 A 之下, 对立事件公式、加法公式等全部照旧成立. 例如

$$P(\bar{B} \mid A) = 1 - P(B \mid A),$$

$$P(B \cup C \mid A) = P(B \mid A) + P(C \mid A) - P(B \cap C \mid A).$$

它们与普通概率的用法相同, 只是所有概率都挂在同一个条件 A 上.

5.3.2 乘法公式

把条件概率的定义式交叉相乘, 立即得到一个计算积事件概率的公式.

定理 5.4 (乘法公式) 设 $P(A) > 0$, 则

$$P(A \cap B) = P(A)\, P(B \mid A).$$

对称地, 若 $P(B) > 0$, 则 $P(A \cap B) = P(B)\, P(A \mid B)$.

证明 由条件概率的定义 $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$, 两边乘以 $P(A)$ 即得 $P(A \cap B) = P(A)\, P(B \mid A)$. 对称地交换 A, B 得另一形式. □

要求 A 与 B 同时发生的概率, 可以先让 A 发生(概率 $P(A)$), 再在 A 已经发生的前提下让 B 发生, 后者概率为 $P(B \mid A)$. 这与计数中的乘法原理一脉相承. 一个分两步完成的过程, 总方法数等于各步方法数的乘积; 概率的乘法公式正是这个思想在概率上的投影.

例题 5.19 (不放回抽球) 袋中装有 3 个红球和 2 个白球, 除颜色外完全相同. 不放回地依次取出 2 个球, 求两次都取到红球的概率.

解 记 A_1 为第一次取到红球, A_2 为第二次取到红球.

第一次取球时, 袋中有 3 个红球共 5 个球, 故

$$P(A_1) = \frac{3}{5}.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

在 A_1 已经发生的条件下, 袋中剩下 2 个红球共 4 个球, 故

$$P(A_2 \mid A_1) = \frac{2}{4} = \frac{1}{2}.$$

由乘法公式,

$$P(A_1 \cap A_2) = P(A_1)\,P(A_2 \mid A_1) = \frac{3}{5} \times \frac{1}{2} = \frac{3}{10}.$$

用古典概型验证. 从 5 个球中取 2 个, 共有 $\binom{5}{2} = 10$ 种等可能的组合, 其中全是红球的有 $\binom{3}{2} = 3$ 种, 概率为 $\frac{3}{10}$. 两种方法结果一致.

乘法公式可以推广到任意多个事件的积. 要求 n 个事件依次全部发生的概率, 只需沿着时间顺序, 每一步都在此前各步均已发生的条件下计算下一步的概率, 再全部乘起来.

定理 5.5 (乘法公式的链式形式) 设 $A_1, A_2, \ldots, A_n$ 是 n 个事件 ($n \geq 2$), 且 $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, 则

$$\begin{aligned} &P(A_1 \cap A_2 \cap \cdots \cap A_n) \\ &\quad = P(A_1)\,P(A_2 \mid A_1)\,P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$

证明 对 n 作数学归纳法.

$n = 2$ 时, 结论就是乘法公式本身.

假设结论对 $n-1$ 个事件成立 ($n \geq 3$). 把 $A_1 \cap \cdots \cap A_{n-1}$ 视为一个事件 C. 由归纳假设成立的前提 $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, 即 $P(C) > 0$, 可以对 C 与 A_n 使用乘法公式.

$$P(C \cap A_n) = P(C)\,P(A_n \mid C).$$

再由归纳假设展开 $P(C) = P(A_1 \cap \cdots \cap A_{n-1})$, 即得

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) = {} & P(A_1)\,P(A_2 \mid A_1) \cdots P(A_{n-1} \mid A_1 \cap \cdots \cap A_{n-2}) \\ & \times P(A_n \mid A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$

归纳成立, 结论对一切 $n \geq 2$ 成立.

条件 $P(A_1 \cap \cdots \cap A_{n-1}) > 0$ 的用意在于保证链上每一项都有意义. 由于 $A_1 \cap \cdots \cap A_{n-1} \subseteq A_1 \cap \cdots \cap A_{k-1}$ 对任意 $k \leq n$ 成立, 由概率的单调性, $P(A_1 \cap \cdots \cap A_{n-1}) > 0$ 蕴含 $P(A_1 \cap \cdots \cap A_{k-1}) > 0$, 故每个条件概率 $P(A_k \mid A_1 \cap \cdots \cap A_{k-1})$ 都有定义. □

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

例题 5.20　一批产品共 10 件, 其中有 3 件次品. 从中不放回地依次抽取 3 件, 求“前两次抽到正品、第三次抽到次品”的概率.

解　记 A_i 为第 i 次抽到正品 $(i=1,2)$, A_3 为第三次抽到次品. 要求 $P(A_1\cap A_2\cap A_3)$.

沿着抽取顺序逐个写出条件概率.

$$P(A_1)=\frac{7}{10},\qquad P(A_2\mid A_1)=\frac{6}{9},\qquad P(A_3\mid A_1\cap A_2)=\frac{3}{8}.$$

第一个等号是因为 10 件产品中有 7 件正品; 第二个等号是因为抽走 1 件正品后, 剩下 9 件中有 6 件正品; 第三个等号是因为抽走 2 件正品后, 剩下 8 件中有 3 件次品.

由链式乘法公式,

$$P(A_1\cap A_2\cap A_3)=\frac{7}{10}\times\frac{6}{9}\times\frac{3}{8}=\frac{126}{720}=\frac{7}{40}.$$

链式乘法公式把一个整体上难以直接计数的事件, 拆解为一系列每一步都容易判断的条件概率. 抽样问题中带先后顺序的事件, 用它处理格外顺手.

回到本节的出发点. 条件概率刻画了信息对判断的影响. 一般情况下 $P(B\mid A)\neq P(B)$, 知道 A 发生会改变对 B 的估计. 但也存在一类特殊的事件对, A 的发生与否完全不影响 B 的概率, 即 $P(B\mid A)=P(B)$. 这种互不影响的关系在概率论中极为重要, 下一节专门研究它.

5.4 事件的相互独立性

上一节末尾留下一个问题, 即是否存在 $P(B\mid A)=P(B)$ 的情形, 也就是事件 A 的发生完全不改变事件 B 发生的概率. 抛两枚硬币时, 第一枚的结果不会影响第二枚; 一台设备的两个独立部件, 一个的寿命也不会影响另一个. 这种互不影响的关系就是本节要研究的**相互独立**.

5.4.1 两个事件的独立性

从条件概率出发, $P(B\mid A)=P(B)$ 等价于

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

最右端的等式有很多好处. 它对 A,B 完全对称; 不要求 $P(A)>0$; 只涉及乘积, 验证起来十分方便. 我们就用它作为独立性的定义.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

定义 5.13 (两个事件的相互独立) 设 A, B 是样本空间 Ω 中的两个事件. 若

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

则称事件 A 与事件 B **相互独立**, 简称**独立**; 否则称它们**不独立**或**相依**.

当 $P(A) > 0$ 时, 独立性等价于 $P(B \mid A) = P(B)$; 当 $P(B) > 0$ 时, 又等价于 $P(A \mid B) = P(A)$. **一个事件的发生, 不改变另一个事件发生的概率**.

例题 5.21 抛两枚均匀的硬币. 记 A 为第一枚正面朝上, B 为第二枚正面朝上. 判断 A 与 B 是否相互独立.

知乎@余命数

解 样本空间为 $\{(H,H),(H,T),(T,H),(T,T)\}$, 四个样本点等可能.

知乎@余命数

$$P(A) = \frac{2}{4} = \frac{1}{2}, \qquad P(B) = \frac{2}{4} = \frac{1}{2}, \qquad P(A \cap B) = P(\{(H,H)\}) = \frac{1}{4}.$$

由于 $P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A)\,P(B)$, 故事件 A 与 B 相互独立.

性质 (独立性与对立事件) 若事件 A 与 B 相互独立, 则 A 与 $\bar{B}$, $\bar{A}$ 与 B, $\bar{A}$ 与 $\bar{B}$ 也都相互独立.

知乎@余命数

证明 只证 A 与 $\bar{B}$ 独立, 其余两组的证法完全相同.

由差事件概率公式, $P(A \cap \bar{B}) = P(A - B) = P(A) - P(A \cap B)$. 代入独立性条件 $P(A \cap B) = P(A)P(B)$:

$$P(A \cap \bar{B}) = P(A) - P(A)\,P(B) = P(A)\,(1 - P(B)) = P(A)\,P(\bar{B}).$$

这正是 A 与 $\bar{B}$ 独立的定义式. □

这条性质符合直觉. 如果 A 的发生影响不了 B, 那么它也影响不了 B 不发生. 在计算中它非常实用, 只要四个事件中有一对独立, 其余三对的乘积公式全部免费成立.

知乎@

注 5.4 (独立与互斥)
互斥与独立描述的是两个完全不同的维度, 初学者极易混淆.

- **互斥**是事件层面的关系, 它意味着 $A \cap B = \emptyset$, 两个事件不能同时发生. 一个发生了, 另一个就**绝不发生**, 这其实是极强的相互影响.
- **独立**是概率层面的关系, 它要求 $P(A \cap B) = P(A)P(B)$, 一个事件的发生不改变另一个事件的概率.

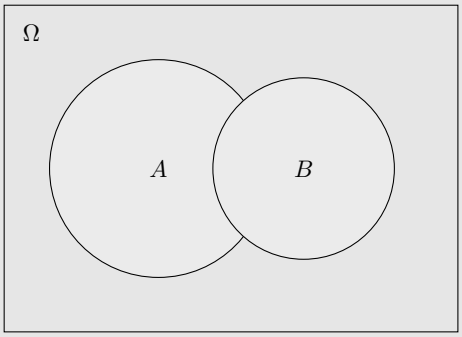
知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

二者在正概率下甚至互相排斥. 若 $P(A)>0, P(B)>0$ 且 A, B 互斥, 则

$$P(A\cap B)=P(\emptyset)=0<P(A)\,P(B),$$

故事件 A, B 必不独立. 反过来, 若 A, B 独立且概率都为正, 则 $P(A\cap B)=P(A)P(B)>0$, 两个事件必定相容.
先看两个极端情形. 不可能事件 $\emptyset$ 与任何事件独立, 因为 $P(\emptyset\cap B)=0=P(\emptyset)P(B)$ 恒成立; 必然事件 Ω 也与任何事件独立, 因为 $P(\Omega\cap B)=P(B)=P(\Omega)P(B)$ 恒成立.

例题 5.22 (人教 A 版必修二 226 页复习参考题 10) 下图是一个古典概型的样本空间 Ω 和事件 A, B 的 Venn 图. 已知样本空间的总样本点数 $\text{card}(\Omega)=24$, 事件 A 包含的样本点数 $\text{card}(A)=12$, 事件 B 包含的样本点数 $\text{card}(B)=8$, 事件 $A\cup B$ 包含的样本点数 $\text{card}(A\cup B)=16$.



(1) 求 $\text{card}(AB)$、$P(AB)$、$P(A\cup B)$ 以及 $P(\bar{A}\bar{B})$.
(2) 判断事件 A 与 B 是否互斥? 是否相互独立?

解 本题把事件的关系与概率的基本性质放进同一个模型中考查.

(1) 计算基本事件的样本点数与概率. 由集合的容斥原理计算事件 A 和 B 交集(积事件)的样本点数,

$$\text{card}(A\cup B)=\text{card}(A)+\text{card}(B)-\text{card}(A\cap B).$$

代入已知数值 $16=12+8-\text{card}(A\cap B)$, 解得积事件 AB, 即 $A\cap B$ 的样本点数 $\text{card}(AB)=4$. 由古典概型定义算出相关概率:

知乎@余命数

$$P(AB)=\frac{\text{card}(AB)}{\text{card}(\Omega)}=\frac{4}{24}=\frac{1}{6},$$

$$P(A\cup B)=\frac{\text{card}(A\cup B)}{\text{card}(\Omega)}=\frac{16}{24}=\frac{2}{3}.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

至于 $P(\bar{A}\bar{B})$, 由德·摩根定律 $\bar{A}\cap\bar{B}=\overline{A\cup B}$, 它即事件 $A\cup B$ 的对立事件概率,

$$P(\bar{A}\bar{B})=P(\overline{A\cup B})=1-P(A\cup B)=1-\frac{2}{3}=\frac{1}{3}.$$

(2) 判断事件的关系.

互斥性. 互斥要求交集为空, 即 $\text{card}(A\cap B)=0$; 而 (1) 已得 $\text{card}(AB)=4\neq 0$, 故事件 A 与 B **不互斥**.

独立性. 相互独立要求 $P(AB)=P(A)P(B)$.(1) 已给左边 $P(AB)=\frac{1}{6}$; 由古典概型算右边

$$P(A)=\frac{\text{card}(A)}{\text{card}(\Omega)}=\frac{12}{24}=\frac{1}{2},\qquad P(B)=\frac{\text{card}(B)}{\text{card}(\Omega)}=\frac{8}{24}=\frac{1}{3},$$

故 $P(A)P(B)=\frac{1}{2}\times\frac{1}{3}=\frac{1}{6}=P(AB)$ 成立, 事件 A 与 B **相互独立**.

这道题还透露出一个信息. 独立与互斥的判定途径完全不同. 互斥与否看集合是否相交, 独立与否看概率的乘积关系, 两者各算各的账.

知乎@余命数

独立性在工程中的一个直接应用是评估系统的可靠性.

例题 5.23 (串联与并联系统) 某设备的两个部件各自独立工作, 在一个使用周期内, 部件 1 正常工作的概率为 0.9, 部件 2 正常工作的概率为 0.8. (1) 若两个部件**串联**, 即任何一个失效都会导致系统失效, 求系统正常工作的概率. (2) 若两个部件**并联**, 即只要还有一个正常系统就能运行, 求系统正常工作的概率.

解 记 A_i 为部件 i 正常工作 $(i=1,2)$, 则 $P(A_1)=0.9$, $P(A_2)=0.8$, 且 A_1 与 A_2 相互独立.

(1) 串联系统正常, 当且仅当两个部件都正常, 即事件 $A_1\cap A_2$. 由独立性,

$$P(\text{系统正常})=P(A_1\cap A_2)=P(A_1)\,P(A_2)=0.9\times 0.8=0.72.$$

(2) 并联系统失效, 当且仅当两个部件都失效, 即事件 $\bar{A}_1\cap\bar{A}_2$. 由上面的性质, $\bar{A}_1$ 与 $\bar{A}_2$ 也相互独立, 故

$$P(\text{系统失效})=P(\bar{A}_1)\,P(\bar{A}_2)=(1-0.9)(1-0.8)=0.1\times 0.2=0.02.$$

知乎@余命数

系统正常的概率为

知乎@余命数

$$P(\text{系统正常})=1-0.02=0.98.$$

对比两个结果. 单个部件的可靠性最高不过 0.9, 并联之后系统可靠性提升到 0.98. 用多个部件互为备份来对冲单点失效的风险, 这就是工程上**冗余设计**的概率原理.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

5.4.2 多个事件的独立性

三个或更多事件的独立性, 不能简单地用两两独立来概括.

定义 5.14 (n 个事件的相互独立) 设 $A_1, A_2, \dots, A_n$ 是 n 个事件. 若对任意满足 $2 \le k \le n$ 的整数 k, 以及任意 k 个不同的下标 $1 \le i_1 < i_2 < \cdots < i_k \le n$, 都有

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

则称这 n 个事件**相互独立**.

以 $n = 3$ 为例, 相互独立要求以下四条同时成立.

$$P(AB) = P(A)P(B), \quad P(AC) = P(A)P(C),$$

$$P(BC) = P(B)P(C), \quad P(ABC) = P(A)P(B)P(C).$$

前三条只涉及两个事件, 合起来称为**两两独立**. 两两独立推不出相互独立, 最后一条三者联合的等式可能单独失效.

例题 5.24 (两两独立但不相互独立) 抛两枚均匀的硬币. 记 A 为第一枚正面朝上, B 为第二枚正面朝上, C 为两枚结果相同. 验证 A, B, C 两两独立, 但不相互独立.

解 样本空间仍是 $\{(H,H),(H,T),(T,H),(T,T)\}$. 逐一计算, 得

$$P(A) = P(B) = P(C) = \frac{2}{4} = \frac{1}{2}.$$

先算两两积事件.

$$P(A \cap B) = P(\{(H,H)\}) = \frac{1}{4} = P(A)P(B),$$

$$P(A \cap C) = P(\{(H,H)\}) = \frac{1}{4} = P(A)P(C),$$

$$P(B \cap C) = P(\{(H,H)\}) = \frac{1}{4} = P(B)P(C).$$

三个事件两两独立. 再算三者同时发生.

$$P(A \cap B \cap C) = P(\{(H,H)\}) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C).$$

所以 A, B, C 不相互独立.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

直观的解释如下. 单独看 A 或 B, 都改变不了 C 的概率; 但 A 与 B **合起来**完全决定了 C, 两枚的结果都知道了, 是否相同也就知道了. 两两独立只排除了单个事件的影响, 管不住多个事件的联合影响, 这正是定义中要求任意 k 个的原因.

多个事件的相互独立最常见的出场方式, 是**独立重复试验**. 同一个试验在相同条件下重复进行, 各次的结果互不影响.

例题 5.25 (重复独立试验) 连续抛一枚均匀的硬币 5 次, 求恰好出现 3 次正面的概率.

解 把 5 次抛掷的结果按顺序记录下来, 每一种完整的结果序列都是一个样本点, 例如

$$(\text{正}, \text{正}, \text{反}, \text{正}, \text{反}).$$

各次抛掷相互独立, 由乘法公式, 这个特定序列的概率为

$$\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\left(\frac{1}{2}\right)^5=\frac{1}{32}.$$

同理, 任何一个指定了每次正反的序列, 概率都是 $\frac{1}{32}$.

事件恰好 3 次正面包含的序列, 相当于在 5 个位置中选出 3 个放正, 共有 $\binom{5}{3}=10$ 个. 这些序列两两互斥, 由可加性公理,

$$P(\text{恰好 3 次正面})=\binom{5}{3}\times\frac{1}{32}=10\times\frac{1}{32}=\frac{5}{16}.$$

每种指定序列的概率用独立性算, 序列的条数用组合数算, 这个两步走是处理独立重复试验的标准动作, 将来研究二项分布时还会反复用到它.

本节从互不影响的直觉出发, 把独立性写成了乘积公式, 并用它拆解了系统可靠性、重复试验等问题. 下一节换一个方向. 当一个事件可以由若干种原因引发时, 我们如何把各种原因的概率拼出该事件的总概率; 反过来, 看到结果之后, 又如何反推每种原因的可能性.

5.5 全概率公式与贝叶斯公式

很多事件的概率难以直接求得, 但它总是由若干种可以一一枚举的原因引发的. 每种原因之下, 事件的条件概率往往好算得多. 把各种原因的情形拼起来, 就得到了事件的总概率, 这就是**全概率公式**. 反过来, 试验结束后看到结果, 回头追问哪种原因的可能性最大, 这就是**贝叶斯公式**. 这两个公式是条件概率的重要应用, 本章开头悬而未决的蒙提霍尔问题, 也将在本节得到正式的解答.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

5.5.1 全概率公式

先引入一个术语. 把样本空间切成若干互不重叠的块, 每块代表一种原因或情形.

定义 5.15 (样本空间的划分) 设 $B_1, B_2, \dots, B_n$ 是样本空间 Ω 中的一组事件. 若它们满足

1. **两两互斥**. 对任意 $i \neq j$, $B_i \cap B_j = \emptyset$;
2. **完备**. $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$;
3. **正概率**. $P(B_i) > 0$ $(i = 1, 2, \dots, n)$,

则称 $B_1, B_2, \dots, B_n$ 为样本空间 Ω 的一个**划分**, 或称**完备事件组**.

最熟悉的划分莫过于 $\{A, \bar{A}\}$. 任何事件与它的对立事件, 恰好把样本空间一分为二.

定理 5.6 (全概率公式) 设 $B_1, B_2, \dots, B_n$ 是样本空间 Ω 的一个划分, 则对任意事件 A,

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

证明 由于 $B_1, \dots, B_n$ 的并集是整个 Ω,

$$A = A \cap \Omega = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

由 B_i 两两互斥, 可知 $A \cap B_i$ 也两两互斥. 应用可加性公理,

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

再由 $P(B_i) > 0$, 对每一项使用乘法公式 $P(A \cap B_i) = P(B_i)\, P(A \mid B_i)$, 即得

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

□

全概率公式按互斥的原因划分事件. 直接求 $P(A)$ 困难时, 找一组把问题切成若干简单情形的划分 $\{B_i\}$, 如图 5.4, 在每种情形下分别计算条件概率, 再按各情形自身的概率加权求和.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

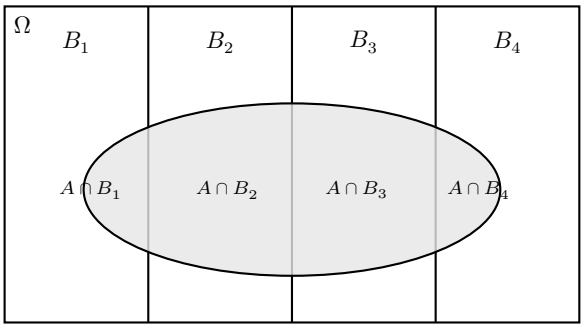


图 5.4 全概率公式的几何图像. 划分 B_1, B_2, B_3, B_4 把事件 A 切成互斥的四块, $P(A)$ 是各块概率按 $P(B_i)$ 加权之和

例题 5.26 (抽签的公平性) 5 张签中有 1 张有奖. 甲、乙两人依次不放回地各抽一张签. 求乙抽到奖签的概率, 并与甲比较.

解 乙能否中奖, 取决于甲抽走了什么. 记 B 为甲抽到奖签, 则 $\{B, \bar{B}\}$ 构成一个划分, 且

$$P(B) = \frac{1}{5}, \qquad P(\bar{B}) = \frac{4}{5}.$$

记 A 为乙抽到奖签. 我们分两种情形写出条件概率.

- 若甲已经抽走奖签, 则乙绝不可能中奖, 即 $P(A \mid B) = 0$.
- 若甲没中奖, 则剩下 4 张签中有 1 张奖签, 即 $P(A \mid \bar{B}) = \frac{1}{4}$.

由全概率公式,

$$P(A) = P(B)\,P(A \mid B) + P(\bar{B})\,P(A \mid \bar{B}) = \frac{1}{5} \times 0 + \frac{4}{5} \times \frac{1}{4} = \frac{1}{5}.$$

整个推理过程可以画成一棵概率树 (图 5.5). 第一阶段按甲的结果分枝, 第二阶段按乙的结果分枝; 每条边上标注相应的 (条件) 概率, 沿路径相乘得到各结局的联合概率, 再把属于乙中奖的叶子加起来, 就是全概率公式在树上读出的结果.

乙中奖的概率与甲完全相同, 都是 $\frac{1}{5}$. 同样的计算可以推广到更多人. 无论先抽后抽, 每个人中奖的概率都相等. 抽签作为一种分配手段是公平的, 先到者并不占便宜, 直觉里后抽的人机会被先抽者抢走的担心, 被全概率公式证伪了.

例题 5.27 (两个供应商) 某工厂从甲、乙两家供应商处采购同一种零件, 甲厂供货占 60%, 次品率为 2%; 乙厂供货占 40%, 次品率为 5%. 从仓库中随机取出一个零件, 求它是次品的概率.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

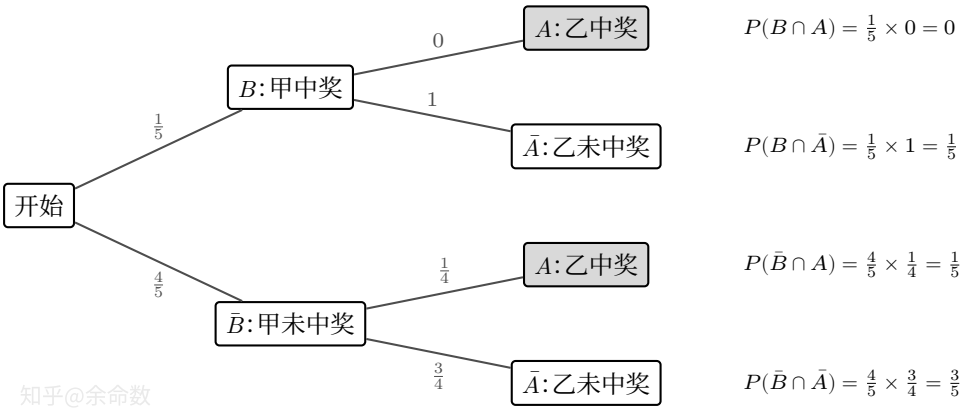


图 5.5　抽签问题的概率树. 沿路径相乘、按叶求和. 两个灰色叶子对应乙中奖, 概率之和为 $0+\frac{1}{5}=\frac{1}{5}$

解　取出的零件来自哪家供应商, 构成一个划分. 记 B_1 为零件来自甲厂, B_2 为零件来自乙厂, A 为零件是次品. 已知

$$P(B_1)=0.6,\quad P(B_2)=0.4,\quad P(A\mid B_1)=0.02,\quad P(A\mid B_2)=0.05.$$

由全概率公式,

$$\begin{aligned}P(A)&=P(B_1)\,P(A\mid B_1)+P(B_2)\,P(A\mid B_2)\\&=0.6\times 0.02+0.4\times 0.05=0.012+0.020=0.032.\end{aligned}$$

零件为次品的概率是 3.2%.

5.5.2　贝叶斯公式

上一个例题还可以追问一步. 如果抽到的零件**确实是次品**, 它更可能来自哪家供应商, 这样的问题颠倒了条件与结果的顺序. 全概率公式由原因推出结果, 现在要由结果反推原因.

定理 5.7 (贝叶斯公式)　设 $B_1,B_2,\dots,B_n$ 是样本空间 Ω 的一个划分, A 是满足 $P(A)>0$ 的事件. 则对每个 $i=1,2,\dots,n$,

$$P(B_i\mid A)=\frac{P(B_i)\,P(A\mid B_i)}{\sum_{j=1}^{n}P(B_j)\,P(A\mid B_j)}.$$

证明　由条件概率的定义,

$$P(B_i\mid A)=\frac{P(B_i\cap A)}{P(A)}.$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

对分子使用乘法公式 $P(B_i \cap A) = P(B_i)\,P(A \mid B_i)$, 对分母使用全概率公式, 即得所证等式. □

公式中各项有约定俗成的名字.$P(B_i)$ 称为**先验概率**, 它表示在看到结果 A 之前, 我们对各种原因可能性的原有判断; $P(B_i \mid A)$ 称为**后验概率**, 它表示结果 A 发生后, 我们对各种原因可能性的修正判断. 贝叶斯公式回答的正是如何用证据更新信念. 后验概率与先验概率以及该原因产生证据的能力 $P(A \mid B_i)$ 成正比.

例题 5.28 (次品溯源) 继续上面的供应商例题. 已知取出的零件是次品, 求它来自甲厂的概率.

解 由贝叶斯公式,

$$P(B_1 \mid A) = \frac{P(B_1)\,P(A \mid B_1)}{P(B_1)\,P(A \mid B_1) + P(B_2)\,P(A \mid B_2)} = \frac{0.012}{0.032} = \frac{3}{8} = 0.375.$$

甲厂供货占 60%, 但次品中只有 37.5% 来自甲厂. 两个数字不矛盾. 甲厂的次品率低, 摊到每件次品上的嫌疑自然小于它的供货份额. 后验概率把供货份额与质量两方面一并计入.

例题 5.29 (医学筛查) 某种疾病在人群中的患病率为 0.1%. 现有一种检测方法. 患病者检测呈阳性的概率为 99%(灵敏度), 健康者检测呈阳性的概率为 1%, 即假阳性率. 某人检测结果为阳性, 求他真正患病的概率.

解 记 B 为此人患病, A 为检测呈阳性. 已知

$$P(B) = 0.001, \quad P(\bar{B}) = 0.999, \quad P(A \mid B) = 0.99, \quad P(A \mid \bar{B}) = 0.01.$$

由贝叶斯公式,

$$P(B \mid A) = \frac{P(B)\,P(A \mid B)}{P(B)\,P(A \mid B) + P(\bar{B})\,P(A \mid \bar{B})} = \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.01}.$$

计算分子分母, 得

$$P(B \mid A) = \frac{0.00099}{0.00099 + 0.00999} = \frac{0.00099}{0.01098} \approx 0.0902.$$

检测呈阳性的人, 真正患病的概率只有约 9%.

直觉的猜测往往是“检测有 99% 的准确率, 阳性结果应该有九成九的把握”, 计算给出的数字却不到一成. 反差来自基础概率. 人群中的健康者数量是患病者的近千倍, 哪怕只有 1% 的健康者被误判, 假阳性的人数也会淹没真阳性的人数.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

这正是先验概率的作用, 面对罕见事件, 单一证据很难彻底翻转极低的先验, 需要追加独立的复查来逐步更新判断.

这个结论用自然频率数一遍人会看得更清楚 (图 5.6).

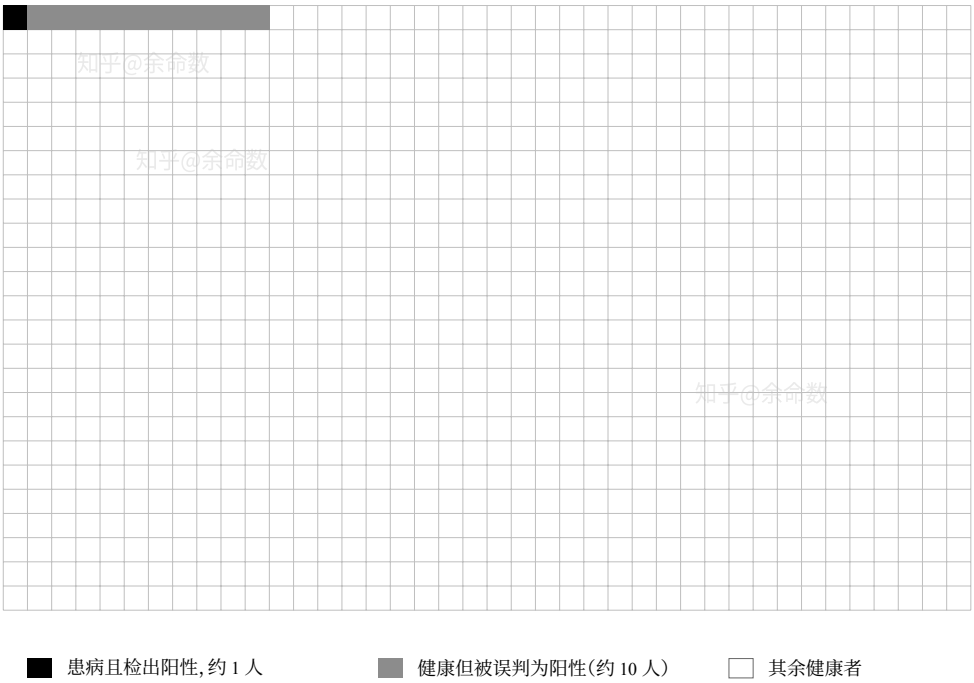


图 5.6　医学筛查的自然频率图像. 每格代表 1 人, 共 1000 人. 阳性者约 11 人(1 黑格加 10 灰格), 其中真正患病的只有约 1 人, 占 $\frac{1}{11} \approx 9\%$

5.5.3 回到蒙提霍尔问题

现在补齐本章开头的欠账. 回顾蒙提霍尔问题的规则. 三扇门后有一辆汽车和两只山羊, 车的位置等可能, 见图 5.1; 玩家初选门 1; 主持人知道车的位置, 并且遵守固定的规则, 即他总是打开剩下一扇有山羊的门, 如果两扇可选的门后都是山羊, 即车恰好在门 1 后, 他随机打开其中一扇. 主持人打开门 3 之后, 需要比较坚持门 1 与换到门 2 哪个胜算更大.

记 C_i 为车在门 i 后 ($i = 1, 2, 3$), H_3 为主持人打开门 3. 由对称性,

$$P(C_1) = P(C_2) = P(C_3) = \frac{1}{3}.$$

按照主持人的规则, 我们逐个写出条件概率.

- 若车在门 1 后, 门 2、门 3 后都是山羊, 主持人随机开一扇, 于是 $P(H_3 \mid C_1) = \frac{1}{2}$.
- 若车在门 2 后, 主持人只能打开门 3, 于是 $P(H_3 \mid C_2) = 1$.
- 若车在门 3 后, 主持人不能打开门 3, 于是 $P(H_3 \mid C_3) = 0$.

知乎关注 @ 余命数谢谢喵, 有帮助的话去爱发电支持一下谢谢喵

由全概率公式,

$$P(H_3)=\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times 1+\frac{1}{3}\times 0=\frac{1}{2}.$$

再由贝叶斯公式, 分别计算两种策略的胜率, 得

$$P(C_1\mid H_3)=\frac{P(C_1)\,P(H_3\mid C_1)}{P(H_3)}=\frac{\frac{1}{3}\times\frac{1}{2}}{\frac{1}{2}}=\frac{1}{3},$$

$$P(C_2\mid H_3)=\frac{P(C_2)\,P(H_3\mid C_2)}{P(H_3)}=\frac{\frac{1}{3}\times 1}{\frac{1}{2}}=\frac{2}{3}.$$

坚持门 1 的胜率仍是 $\frac{1}{3}$, 换到门 2 的胜率是 $\frac{2}{3}$. 换门确实使胜率翻倍, 章首宣称的结论得到严格的证明.

其实还有一个更干脆的论证, 一条式子都不用写. 注意主持人永远不会打开有车的门, 因此有下面两条.

- 若玩家初选正确(概率 $\frac{1}{3}$), 换门必然扑空;
- 若玩家初选错误(概率 $\frac{2}{3}$), 车在两扇未选的门之一后面, 主持人被迫打开了另一扇山羊门, 剩下那扇未开的门后必为汽车, 换门必胜.

换门获胜, 当且仅当初选错误, 故换门胜率为 $P(\text{初选错误})=\frac{2}{3}$. 这层关系用策略树直接表示 (图 5.7).

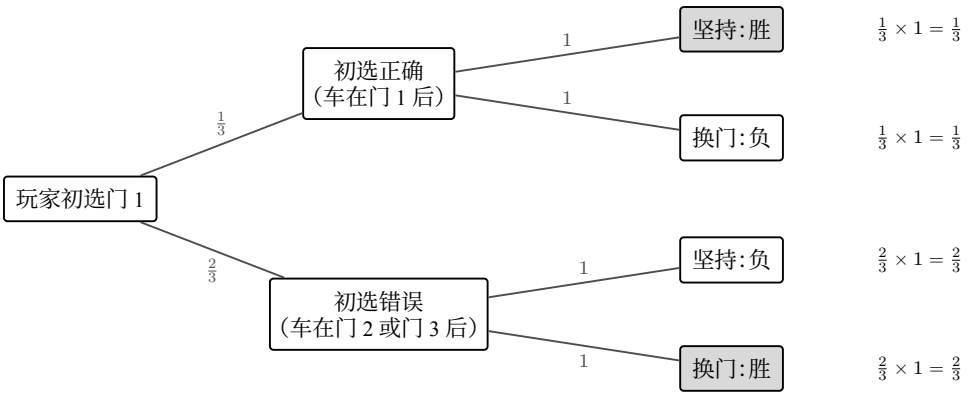


图 5.7 蒙提霍尔问题的策略树. 主持人总会打开一扇山羊门, 胜负在初选时就已注定. 坚持胜与换门胜各对应一个灰色叶子, 概率分别为 $\frac{1}{3}$ 与 $\frac{2}{3}$

直觉在这里失灵,是因为直觉把主持人开门当成了一次随机的排除. 三扇门随便排除一扇, 剩下两扇自然五五开. 但主持人是**带着知识**开门的, 他刻意避开汽车. 这个有偏向的动作, 把初选错误的那 $\frac{2}{3}$ 概率全部集中到了剩下那扇未开的门上. 章首的三情况穷举图 (图 5.2) 数出了同样的答案, 而条件概率与贝叶斯公式把这条信息的来龙去脉记成了明账. 先验是三个 $\frac{1}{3}$, 主持人开门这条证据按 $P(H_3\mid C_i)$ 加权, 后验就变成 $\frac{1}{3}$ 与 $\frac{2}{3}$ 的分化.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

不过到目前为止，我们回答的问题都是某事件发生的概率是多少. 更多的场合里，我们关心的其实是试验产生的数值. 掷骰子的点数、抽检中的次品数、一名射手的命中环数. 把这些数值作为研究对象，需要一个新的概念，即随机变量.

知乎@余命数

5.6 离散型随机变量及其分布列

前面几节里，我们关心的都是某事件发生的概率. 但一次试验的结果本身也可以变成数. 掷一颗骰子时，我们自然关心点数；抽检一批产品时，我们关心次品的个数；连抛硬币时，我们也可把正面记为 1、反面记为 0. 给每个样本点配上一个数，这个数就成了新的研究对象.

5.6.1 随机变量的概念

定义 5.16 (随机变量) 设随机试验的样本空间为 Ω. 若对每个样本点 $\omega \in \Omega$，都有唯一确定的实数 $X(\omega)$ 与之对应，则称这个对应关系 X 为一个**随机变量**，通常用大写字母 X, Y, Z 等表示.

随机变量并不神秘. 用 第一卷第 2 章中已经熟悉的话来说，它就是定义在样本空间 Ω 上、取值于实数集 $\mathbb{R}$ 的一个**函数**. 它与普通函数的唯一区别，是定义域里的每个样本点都带着一个概率；于是 X 取某个值这件事也有了概率.

例题 5.30 (随机变量的例子)

- 掷一颗骰子，令 X 为掷出的点数. 样本空间 $\Omega = \{1, 2, \dots, 6\}$，$X(\omega) = \omega$.
- 抛两枚硬币，令 Y 为出现正面的次数. 样本空间为

$$\{(H,H), (H,T), (T,H), (T,T)\},$$

  Y 在四个样本点处的取值分别为 $2, 1, 1, 0$.
- 抛一枚硬币，令

$$Z = \begin{cases} 1, & \text{正面朝上}, \\ 0, & \text{反面朝上}. \end{cases}$$

  试验结果本身不是数，人为赋值之后，Z 也是一个随机变量.

若一个随机变量的所有可能取值可以一一列举出来（有限个，或像 $0, 1, 2, \dots$ 那样排成一列），就称它为**离散型随机变量**. 上面三个例子都是离散型的；本章只讨论离散型. 取值可数无限时，分布列与期望中的求和改为相应的级数，其收敛性

知乎关注 @ 余命数谢谢喵，有帮助的话去爱发电支持一下谢谢喵

需要单独判断. 像灯泡寿命这种取值充满一个区间的随机变量, 称为连续型; 严格处理需要积分, 本书把它放在 第四卷第 17 章.

引入随机变量之后, 事件获得了新的表达方式.$\{X=2\}$ 表示 X 取值为 2 这一事件, 它对应样本点的集合 $\{\omega\in\Omega : X(\omega)=2\}$; 同样, $\{X\le 3\}$、$\{1<X\le 5\}$ 也都是事件. 例如抛两枚硬币时, $\{Y=1\}=\{(H,T),(T,H)\}$, 所以

$$P(Y=1)=\frac{2}{4}=\frac{1}{2}.$$

随机变量把事件翻译成了关于数的等式与不等式, 许多实际问题因此变得易于书写.

5.6.2 分布列

光知道随机变量能取哪些值还不够, 还要知道它取每个值的概率有多大. 把这两样信息列成一张表, 就得到了分布列.

定义 5.17 (分布列) 设离散型随机变量 X 的所有可能取值为 $x_1,x_2,\dots,x_n$, 取每个值的概率为

$$P(X=x_i)=p_i\quad(i=1,2,\dots,n).$$

则称该表为随机变量 X 的**概率分布列**, 简称**分布列**.

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

性质 (分布列的性质) 离散型随机变量 X 的分布列满足下面两条.

1. $p_i\ge 0\ (i=1,2,\dots,n)$;
2. $p_1+p_2+\cdots+p_n=1$.

证明 (1) 每个 p_i 都是事件 $\{X=x_i\}$ 的概率, 由非负性公理即得.

(2) 事件 $\{X=x_1\},\{X=x_2\},\dots,\{X=x_n\}$ 两两互斥 (X 一次只能取一个值), 且它们的并集是整个样本空间, 因为 X 总要取某个值. 由可加性公理与规范性公理,

$$\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}P(X=x_i)=P\Big(\bigcup_{i=1}^{n}\{X=x_i\}\Big)=P(\Omega)=1.$$

□

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

这两条性质是检验一张分布列是否合法的标准, 也常用来反求分布列中待定的参数.

例题 5.31 (两颗骰子点数之和的分布列) 掷两颗均匀的骰子, 记 X 为两颗骰子的点数之和. 求 X 的分布列.

解 X 的可能取值为 $2,3,\dots,12$. 两颗骰子的结果共有 36 个等可能样本点, 和为 k 的样本点数等于方程 $x_1+x_2=k$ 满足 $1\le x_1,x_2\le 6$ 的整数解个数, 例如和为 4 的有 $(1,3),(2,2),(3,1)$ 共 3 个. 逐一枚举, 得下表. 为书写紧凑, 第二行记录的是样本点个数, 概率为样本点数除以 36.

X	2	3	4	5	6	7	8	9	10	11	12
样本点数	1	2	3	4	5	6	5	4	3	2	1

故 X 的分布列为

知乎@余命数　知乎@余命数　知乎@余命数　知乎@余命数

$$P(X=k)=\frac{6-|k-7|}{36},\quad k=2,3,\dots,12.$$

直接检验, $1+2+3+4+5+6+5+4+3+2+1=36$, 各概率之和 $\frac{36}{36}=1$, 符合分布列的两条性质.

例题 5.32 (抽检中的次品数) 一批产品共 5 件, 其中有 2 件次品. 从中不放回地取出 3 件, 记 X 为取到的次品件数. 求 X 的分布列.

解 X 的可能取值为 $0,1,2$. 从 5 件中取 3 件, 共有 $\binom{5}{3}=10$ 种等可能取法. 取到 k 件次品等价于从 2 件次品中选 k 件、从 3 件正品中选 $3-k$ 件, 故

知乎@余命数　知乎@余命数

$$P(X=k)=\frac{\binom{2}{k}\binom{3}{3-k}}{\binom{5}{3}},\quad k=0,1,2.$$

分别代入得

$$P(X=0)=\frac{\binom{2}{0}\binom{3}{3}}{\binom{5}{3}}=\frac{1}{10},\quad P(X=1)=\frac{\binom{2}{1}\binom{3}{2}}{\binom{5}{3}}=\frac{6}{10},\quad P(X=2)=\frac{\binom{2}{2}\binom{3}{1}}{\binom{5}{3}}=\frac{3}{10}.$$

故分布列为

知乎@余命数　知乎@余命数

X	0	1	2
P	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$

直接检验, $\frac{1}{10}+\frac{6}{10}+\frac{3}{10}=1$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

求分布列, 即列出 X 的所有可能取值, 逐一计算 $P(X=x_i)$, 并用两条性质检验. 计算每个取值的概率, 是把概率问题归结为计数问题, 这一步依赖 第4章 与前面各节建立的工具.

分布列完整地刻画了随机变量, 但一张表往往信息过载. 比较两个随机变量时, 我们常希望各用一个数字概括它们. 一个数字代表平均水平, 另一个代表波动程度. 这就是下一节要定义的期望与方差.

5.7 期望与方差

分布列给出了离散型随机变量的全部信息, 但一整张表读起来并不轻松. 比较两名射手的水平、估计一家保险公司的盈亏, 我们想要的往往是几个有代表性的数字. 本节引入最常用的两个数字, **期望**刻画平均水平, **方差**刻画波动程度.

5.7.1 数学期望

知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数 知乎@余命数

先从一个思想实验说起. 掷一颗骰子 n 次, 把每次的点数都记下来, 计算这 n 个数的算术平均值. 当 n 很大时, 根据频率的稳定性, 点数 k 出现的频率会稳定在 $\frac{1}{6}$ 附近, 于是平均值稳定在

$$1\times\frac{1}{6}+2\times\frac{1}{6}+\cdots+6\times\frac{1}{6}=3.5$$

附近. 这个按概率加权的平均值, 就是期望的雏形.

定义 5.18 (数学期望) 设离散型随机变量 X 的分布列为 $P(X=x_i)=p_i$ $(i=1,2,\dots,n)$. 称

$$E(X)=\sum_{i=1}^{n}x_ip_i$$

为随机变量 X 的**数学期望**, 简称**期望**, 也称**均值**.

期望是以概率为权重的加权平均. 取值越容易出现, 权重越大, 对平均水平的贡献也越大. 期望未必是 X 能取到的值, 骰子的期望 3.5 就不在 $\{1,2,\dots,6\}$ 之中, 它描述的是长期平均意义下的中心位置.

知乎@余命数 知乎@余命数

例题 5.33 求上一节抽检次品数一例中, 次品数 X 的期望.

解 由分布列

X	0	1	2
P	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{3}{10}$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

按期望定义直接计算,

$$E(X)=0\times\frac{1}{10}+1\times\frac{6}{10}+2\times\frac{3}{10}=\frac{12}{10}=1.2.$$

平均每次抽检会抽到 1.2 件次品.

期望有一组用起来极为顺手的运算性质.

性质 (期望的线性性质) 设 X,Y 是离散型随机变量, a,b 为常数, 则

1. $E(aX+b)=aE(X)+b$;
2. $E(X+Y)=E(X)+E(Y)$.

证明 (1) 设 X 的分布列为 $P(X=x_i)=p_i$. 随机变量 $aX+b$ 取值 ax_i+b 的概率也是 p_i, 故

$$E(aX+b)=\sum_{i=1}^{n}(ax_i+b)\,p_i=a\sum_{i=1}^{n}x_ip_i+b\sum_{i=1}^{n}p_i=aE(X)+b.$$

(2) 在有限样本空间 Ω 上, 期望可以改写成对样本点求和的形式. 把 $\sum_{\omega}X(\omega)\,P(\{\omega\})$ 中取值相同的项合并, 恰好得到 $\sum_i x_iP(X=x_i)$. 于是

$$\begin{aligned}E(X+Y)&=\sum_{\omega\in\Omega}(X(\omega)+Y(\omega))\,P(\{\omega\})\\&=\sum_{\omega\in\Omega}X(\omega)\,P(\{\omega\})+\sum_{\omega\in\Omega}Y(\omega)\,P(\{\omega\})=E(X)+E(Y).\end{aligned}$$

□

第二条性质最省力的地方在于, 它不要求 X 与 Y 独立. 无论两个随机变量如何纠缠, 和的期望永远等于期望的和.

例题 5.34 (两颗骰子点数之和的期望) 掷两颗均匀的骰子, 求点数之和 X 的期望.

解 法一采用分布列. 由上一节求得的分布列,

$$\begin{aligned}E(X)=\frac{1}{36}(&2\times1+3\times2+4\times3+5\times4+6\times5+7\times6\\&+8\times5+9\times4+10\times3+11\times2+12\times1)=\frac{252}{36}=7.\end{aligned}$$

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

法二采用线性性质. 设 X_1, X_2 分别为两颗骰子的点数, 则 $X = X_1 + X_2$. 单颗骰子的期望为

$$E(X_1) = E(X_2) = \frac{1+2+3+4+5+6}{6} = 3.5,$$

故

$$E(X) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7.$$

两种方法结果一致. 把一个复杂的随机变量拆成若干简单随机变量之和, 再用线性性质求期望, 是概率论中反复使用的技巧.

5.7.2 方差

期望相同, 水平未必一样. 看两名射手.

例题 5.35 (两名射手的成绩) 甲、乙两名射手射击一次, 命中环数 X_1, X_2 的分布列分别为

X_1	8	9	10
P	0.2	0.6	0.2

X_2	8	9	10
P	0.4	0.2	0.4

比较两人的射击水平.

解 按期望定义计算两人平均环数,

$$E(X_1) = 8 \times 0.2 + 9 \times 0.6 + 10 \times 0.2 = 9,$$

$$E(X_2) = 8 \times 0.4 + 9 \times 0.2 + 10 \times 0.4 = 9.$$

两人平均环数相同, 但直观上甲的成绩集中在 9 环附近, 乙则经常打出 8 环或 10 环, 甲更稳定. 期望看不出来这个差别, 还需要一个专门度量偏离平均水平程度的数字.

度量偏离, 自然的想法是看各取值与期望之差 $x_i - E(X)$. 直接加权平均会被正负抵消(结果恒为 0), 把差平方再平均就解决了这个问题.

定义 5.19 (方差与标准差) 设离散型随机变量 X 的分布列为 $P(X = x_i) = p_i$ $(i = 1, 2, \dots, n)$. 称

$$D(X) = \sum_{i=1}^{n} \big(x_i - E(X)\big)^2 p_i$$

为随机变量 X 的**方差**;称 $\sqrt{D(X)}$ 为 X 的**标准差**.

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

方差是与期望偏差的平方的期望, 可以记作 $D(X)=E[(X-E(X))^2]$. 方差越小, 取值越集中于期望附近, 随机变量越稳定; 方差越大, 波动越剧烈. 标准差与 X 同量纲, 解释起来更直观.

按定义算方差, 每算一个都要先减期望再平方, 相当繁琐. 下列公式把计算化为两个期望之差.

性质 (方差的计算公式)

$$D(X)=E(X^2)-\big[E(X)\big]^2.$$

知乎@余命数

证明 记 $\mu=E(X)$. 展开平方并套用期望的线性性质,

知乎@余命数

$$\begin{aligned}D(X)&=\sum_{i=1}^{n}(x_i-\mu)^2p_i=\sum_{i=1}^{n}(x_i^2-2\mu x_i+\mu^2)\,p_i\\&=\sum_{i=1}^{n}x_i^2p_i-2\mu\sum_{i=1}^{n}x_ip_i+\mu^2\sum_{i=1}^{n}p_i\\&=E(X^2)-2\mu\cdot\mu+\mu^2=E(X^2)-\mu^2.\end{aligned}$$

知乎@余命数

□

性质 设 a,b 为常数, 则

$$D(aX+b)=a^2D(X).$$

证明 由期望的线性性质, $E(aX+b)=aE(X)+b$, 于是

$$(aX+b)-E(aX+b)=aX+b-aE(X)-b=a\big(X-E(X)\big).$$

两边平方取期望即

知乎@余命数

$$D(aX+b)=E\big[a^2(X-E(X))^2\big]=a^2E\big[(X-E(X))^2\big]=a^2D(X).$$

常数平移 b 不影响方差, 因为整批数据一起移动, 离散程度不变; 缩放 a 会把偏差放大 a 倍, 方差放大 a^2 倍. □

例题 5.36 完成射手比较. 计算 X_1,X_2 的方差, 再求骰子点数的方差.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

解 套用计算公式 $D(X)=E(X^2)-[E(X)]^2$. 对甲射手, $E(X_1^2)=64\times0.2+81\times0.6+100\times0.2=81.4$, 故

$$D(X_1)=81.4-9^2=0.4.$$

对乙射手, $E(X_2^2)=64\times0.4+81\times0.2+100\times0.4=81.8$, 故

$$D(X_2)=81.8-9^2=0.8.$$

$D(X_1)<D(X_2)$, 甲的成绩更稳定, 与直观判断一致, 期望比平均水平, 方差比稳定程度, 两个数字合起来才构成完整的评价.

骰子点数 X 同法处理, $E(X^2)=\frac{1^2+2^2+\cdots+6^2}{6}=\frac{91}{6}$, 故

$$D(X)=\frac{91}{6}-\left(\frac{7}{2}\right)^2=\frac{91}{6}-\frac{49}{4}=\frac{182-147}{12}=\frac{35}{12}\approx2.92.$$

期望与方差是分布列的摘要. 前者回答平均是多少, 后者回答围绕平均晃多远. 下一节要认识的几个分布模型, 都配有现成的期望与方差公式, 届时可以直接套用.

5.8 常见分布模型

现实中有几类随机试验反复出现. 一次成败的尝试、重复多次的成败尝试、不放回抽检、受大量微小因素叠加影响的测量, 都对应着现成的分布公式与数字特征. 认出模型, 就可以直接套用. 本节认识四个出场率最高的分布模型.

5.8.1 两点分布

最简单的随机试验只有两种结果, 成功与失败、正面与反面、合格与不合格.

定义 5.20 (两点分布) 设随机变量 X 只取 0 和 1 两个值, 分布列为

X	0	1
P	$1-p$	p

其中 $0<p<1$. 则称 X 服从**两点分布**, 也称**伯努利分布**.

按定义直接算其数字特征,

$$E(X)=0\times(1-p)+1\times p=p,$$

知乎关注@余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

$$E(X^2) = 0^2 \times (1-p) + 1^2 \times p = p,$$

$$D(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p).$$

5.8.2 二项分布

把同一个两点分布的试验在相同条件下独立地重复 n 次, 记录成功的总次数, 这就是第四节讨论过的独立重复试验, 也称 n **重伯努利试验**.

知乎@余命数

定义 5.21 (二项分布) 在 n 重伯努利试验中, 每次试验成功的概率为 p $(0 < p < 1)$, 记 X 为成功的总次数. 则

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

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

知乎@余命数

这个公式的来路, 第四节已经走过一遍. 任何一个指定的、恰好含 k 次成功的结果序列, 例如成成败成…, 由独立性其概率为 $p^k(1-p)^{n-k}$; 这样的序列相当于在 n 个位置中选出 k 个放成功, 共 $\binom{n}{k}$ 个, 且两两互斥, 相加即得.

各概率之和为 1 可由我们已在 第4.3节 中了解过的二项式定理直接验证,

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \big(p + (1-p)\big)^n = 1.$$

二项分布的名字正来自这个二项展开.

定理 5.8 (二项分布的期望) 若 $X \sim B(n, p)$, 则

$$E(X) = np.$$

证明 由期望定义,

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

知乎@余命数

$k = 0$ 的项为 0. 对 $k \geq 1$, 利用恒等式 $k\binom{n}{k} = n\binom{n-1}{k-1}$ (两边都是“从 n 人中选 k 人并指定一名组长”的计数), 得

$$\begin{aligned} E(X) &= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\ &= np\big(p + (1-p)\big)^{n-1} = np. \end{aligned}$$

□

直觉上这个结论理所当然. 每次试验平均贡献 p 次成功, n 次独立重复自然平均成功 np 次. 用期望的线性性质也能看出这一点. 记 X_i 为第 i 次试验成功的两点分布变量, 则 $X = X_1 + \cdots + X_n$, 故 $E(X) = nE(X_i) = np$.

定理 5.9 (二项分布的方差) 若 $X \sim B(n,p)$, 则

$$D(X) = np(1-p).$$

证明 先算 $E[X(X-1)]$. 对 $k \geq 2$, 利用 $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$:

$$\begin{aligned} E[X(X-1)] &= \sum_{k=2}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{(n-2)-(k-2)} \\ &= n(n-1)p^2 (p+(1-p))^{n-2} = n(n-1)p^2. \end{aligned}$$

于是

$$E(X^2) = E[X(X-1)] + E(X) = n(n-1)p^2 + np,$$

$$D(X) = E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p).$$

□

例题 5.37 (投篮问题) 某篮球运动员的罚球命中率为 0.6. 他独立罚球 10 次, 记 X 为命中次数. (1) 求恰好命中 6 次的概率; (2) 求 X 的期望与方差.

解 每次罚球是一次伯努利试验, $X \sim B(10, 0.6)$.

(1) 由分布公式,

$$P(X=6) = \binom{10}{6} \times 0.6^6 \times 0.4^4 = 210 \times 0.046656 \times 0.0256 \approx 0.2508.$$

(2) 直接套公式,

$$E(X) = np = 10 \times 0.6 = 6, \qquad D(X) = np(1-p) = 10 \times 0.6 \times 0.4 = 2.4.$$

他平均命中 6 次, 命中次数围绕 6 上下波动, 标准差为 $\sqrt{2.4} \approx 1.55$ 次.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

5.8.3 超几何分布

二项分布要求各次试验相互独立, 放回抽样满足这个要求; 不放回抽样时, 每次抽取都改变了剩下的成分, 独立性被破坏, 分布也随之改变.

定义 5.22 (超几何分布) 一批产品共 N 件, 其中次品 M 件. 从中不放回地取出 n 件 ($n \le N$), 记 X 为取到的次品件数. 则

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=0,1,2,\dots,\min(M,n),$$

称 X 服从**超几何分布**.

这个公式我们在第六节的抽检次品数一例中已经用过. 分母是全部取法数,分子是"k 件次品从 M 件次品中选、$n-k$ 件正品从 $N-M$ 件正品中选"的取法数.

性质 (超几何分布的期望) 若 X 服从参数为 N, M, n 的超几何分布, 则

$$E(X)=\frac{nM}{N}.$$

证明 记 $C=\binom{N}{n}$. 由期望定义与恒等式 $k\binom{M}{k}=M\binom{M-1}{k-1}$ ($k\ge 1$),

$$E(X)=\frac{1}{C}\sum_{k=1}^{n}k\binom{M}{k}\binom{N-M}{n-k}=\frac{M}{C}\sum_{k=1}^{n}\binom{M-1}{k-1}\binom{N-M}{n-k}.$$

右端求和是范德蒙恒等式. 从 $M-1$ 个一类元素与 $N-M$ 个二类元素中共取出 $n-1$ 个, 按一类元素取 $k-1$ 个分类求和, 总数为 $\binom{N-1}{n-1}$. 故

$$E(X)=\frac{M\binom{N-1}{n-1}}{\binom{N}{n}}=M\cdot\frac{n}{N}=\frac{nM}{N}.$$

□

结果的读法很直白. 次品率为 $\frac{M}{N}$, 抽 n 件平均抽到 $\frac{nM}{N}$ 件次品. 第六节的例子中 $N=5$、$M=2$、$n=3$, 公式给出 $E(X)=\frac{3\times 2}{5}=1.2$, 与当时按分布列硬算的结果一致.

注 5.5 (放回与不放回)
同样的次品率 $p=\frac{M}{N}$, 放回抽检 n 次, 次品数服从 $B(n,p)$; 不放回抽检 n 件, 次品数服从超几何分布. 二者的期望都是 np, 但方差不同, 因为不放回时各次抽取互相牵制, 波动更小. 当总体很大而抽样量相对很小, 例如从十万件中抽十件时, 抽走几件几乎不改变总体成分, 不放回近似于放回, 超几何分布可用二项分布近似计算, 这是统计抽检中常用的简化.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

正态分布与 3σ 直观

注 5.6
离散分布(两点、二项、超几何)是本章主干. 正态分布给出钟形曲线与 3σ 规则的图像直觉;密度曲线面积的严格含义见 第四卷第 17 章.

5.8.4 正态分布

最后一个模型属于连续型随机变量, 严格处理需要积分; 它的应用极广, 本节给出直观认识.

测量误差、身高体重、考试成绩、工厂零件的尺寸偏差, 即由无数微小因素叠加而成的随机量, 其取值分布往往呈现同一种形态. 中间高、两边低、左右对称的钟形曲线. 这背后有一条严格的数学定理, 称为中心极限定理. 大量独立微小随机因素的叠加, 会趋向同一类分布, 即**正态分布**.

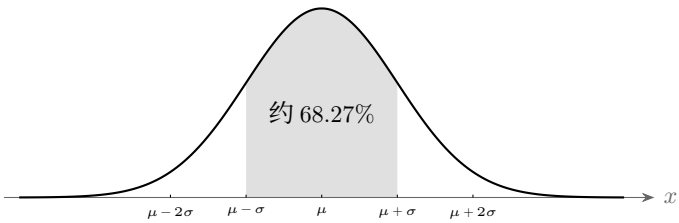


图 5.8 正态分布的密度曲线. 以 μ 为中心、σ 为尺度的钟形曲线,阴影部分面积约为 68.27%

正态分布由两个参数完全确定. 期望 μ 决定曲线的中心位置, 标准差 σ 决定曲线的胖瘦(σ 越小曲线越瘦高, 取值越集中). 记作 $X\sim N(\mu,\sigma^2)$. 它的概率用曲线下方的面积来度量.$P(a\le X\le b)$ 等于曲线在区间 $[a,b]$ 上与 x 轴围成的面积, 总面积为 1.

性质 (3σ 规则) 若 $X\sim N(\mu,\sigma^2)$, 则

$$P(\mu-\sigma\le X\le\mu+\sigma)\approx 68.27\%,$$

$$P(\mu-2\sigma\le X\le\mu+2\sigma)\approx 95.45\%,$$

$$P(\mu-3\sigma\le X\le\mu+3\sigma)\approx 99.73\%.$$

由曲线的对称性, 中心两侧的概率可以互相换算, 例如

$$P(X\ge\mu+a)=P(X\le\mu-a)=\frac{1-P(\mu-a\le X\le\mu+a)}{2}.$$

几乎全部的取值都落在期望的 3σ 之内, 这是正态分布最实用的画像.

例题 5.38 某年级的一次数学考试成绩 X 近似服从 $N(80,10^2)$. (1) 估计成绩在 70 分到 90 分之间的学生比例;(2) 估计成绩不低于 100 分的学生比例.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数

解 (1) $70 = 80 - 10, 90 = 80 + 10$, 即区间 $[\mu - \sigma, \mu + \sigma]$. 由 3σ 规则,

$$P(70 \le X \le 90) \approx 68.27\%.$$

约有六成八的学生成绩落在 70 到 90 分之间.

(2) $100 = 80 + 2 \times 10 = \mu + 2\sigma$. 由对称性,

$$P(X \ge 100) = \frac{1 - P(60 \le X \le 100)}{2} \approx \frac{1 - 95.45\%}{2} \approx 2.28\%.$$

百分以上的学生约占 2.3%.

本章以样本空间与概率公理为起点, 用条件概率、独立性、全概率公式与贝叶斯公式刻画信息, 并用随机变量、分布列、期望与方差描述分布, 最后给出几个常用分布模型. 正态分布钟形曲线与 3σ 规则的严格推导以积分为基础 (第四卷第 17 章).

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

习题

5.1. (事件运算)设 A, B 为两个事件. 化简: $(A \cup B) \cap (\bar{A} \cup B)$. 提示: 对 B 用分配律提取公因子.

5.2. (古典概型)从 $1, 2, \ldots, 100$ 中任取一个整数, 求它能被 6 或 8 整除的概率. 提示: 用容斥原理, 注意 6 与 8 的最小公倍数是 24.

5.3. (错排)5 位同学每人准备一份礼物, 随机交换后每人收到一份. 求没有任何一人拿到自己礼物的方案数. 提示: 即错排数 D_5, 用递推 $D_n = (n-1)(D_{n-1} + D_{n-2})$ 及 $D_3 = 2$, $D_4 = 9$.

5.4. (条件概率) 掷两颗均匀的骰子, 已知两颗的点数不同, 求至少有一颗为 6 点的概率. 提示: 在收缩后的样本空间中直接计数.

5.5. (独立性)甲、乙、丙三人各自独立破译同一份密码, 能破译的概率分别为 $\frac{1}{5}$、$\frac{1}{4}$、$\frac{1}{3}$. 求密码被破译的概率. 提示: 考虑对立事件三人都未能破译.

5.6. (全概率公式)三个箱子中分别装有 2 红 1 白、1 红 2 白、3 红共 3 个球. 随机选一个箱子, 从中取出一个球, 求取到红球的概率. 提示: 以选中哪个箱子作样本空间的划分.

5.7. (贝叶斯公式)承上题. 已知取到的是红球, 求它来自装有 3 个红球的箱子的概率. 提示: 分母用第 6 题的结果.

5.8. (期望与方差) 设随机变量 X 的分布列为 $P(X = 0) = 0.2$、$P(X = 1) = 0.5$、$P(X = 2) = 0.3$. 求 $E(X)$ 与 $D(X)$. 提示: 方差用计算公式 $D(X) = E(X^2) - [E(X)]^2$.

5.9. (二项分布) 某选手投篮命中率为 0.7, 独立投篮 4 次. 求恰好命中 3 次的概率. 提示: 命中次数服从 $B(4, 0.7)$.

5.10. (正态分布)设 $X \sim N(2, 9)$, 用 3σ 规则估计 $P(X > 5)$. 提示: $\sigma = 3$, 而 $5 = \mu + \sigma$.

知乎关注 @ 余命数谢谢喵,有帮助的话去爱发电支持一下谢谢喵

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

习题解答

第 3 章统计

3.1. **解** $\hat{p}=432/900=0.48$, $\widehat{\mathrm{SE}}=\sqrt{0.48\times0.52/900}\approx0.0167$, 置信区间 $48\%\pm3.3\%$, 即 $[44.7\%,51.3\%]$.

3.2. **解** 标准误与 $\sqrt{n}$ 成反比, 减半需 n 扩大四倍, 即 3600 人. 一般地按最保守的 $p=0.5$ 估计, ±1.5 个百分点为两倍标准误, $2\sqrt{0.25/n}=0.015$ 给出 $n\approx4400$.

3.3. **解** 50 万份回复的随机误差确实极小, 但问卷的样本是自愿点击进入的活跃用户, 不回复者与回复者对收费调整的态度通常恰恰不同, 抽样偏差可能远大于随机误差.《文学文摘》的教训正是样本巨大而偏差更大, 结论是否可信取决于样本如何产生, 与样本量的大小关系不大.

3.4. **解** $\bar{x}=\dfrac{21}{6}=3.5$, $\bar{y}=\dfrac{19}{6}$, 于是

$$d_x=\left(-\tfrac{5}{2},-\tfrac{3}{2},-\tfrac{1}{2},\tfrac{1}{2},\tfrac{3}{2},\tfrac{5}{2}\right),\qquad d_y=\tfrac{1}{6}\left(-7,-13,5,-1,11,5\right).$$

逐项计算得 $\sum d_xd_y=\dfrac{63}{6}=10.5$、$\sum d_x^2=\dfrac{35}{2}=17.5$、$\sum d_y^2=\dfrac{65}{6}$. 代入得

$$r=\frac{10.5}{\sqrt{17.5\times\frac{65}{6}}}\approx0.76,\qquad b=\frac{10.5}{17.5}=\frac{3}{5},\qquad a=\frac{19}{6}-\frac{3}{5}\times\frac{7}{2}=\frac{16}{15}.$$

回归直线为 $\hat{y}=\dfrac{3}{5}x+\dfrac{16}{15}$, 预测 $\hat{y}(8)=\dfrac{88}{15}\approx5.87$.

3.5. **解** $\bar{x}=3$、$d_x=(-2,-1,0,1,2)$、$d_y=(5,0,-3,0,5)$, 于是 $\sum d_xd_y=-10+10=0$, $r=0$. 但每个 x 唯一决定 $y=(x-3)^2$, 关系是完全确定的非线性关系. 相关为零只排除直线关系, 不排除其他形态的依赖.

3.6. **解** 其一, 重视教育的家庭既多买书又多陪伴, 家庭条件与教育态度是混杂变量, 藏书量只是它们的外显标志. 其二, 观察性数据里的斜率只描述人群关联, 因果结论需要随机化对照实验, 而随机指定家庭买书并不现实. 相关为预测提供依据, 不为干预背书.

知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数
知乎@余命数

PROJECT AKARI

AKARI

内容简介

《中学数学之旅》把初等数学当作一个完整的体系，从初中运算的起点一路讲到大学数学的门口. 全书五卷按数学自身的生长顺序编排，从代数语言与函数出发，经周期、离散与随机，走过几何主干，拧成极限、导数与积分的分析主干，最后扩到复数、空间与线性结构. 每条结论尽量给出推导；高考与竞赛的主干题材全部覆盖，取舍只看一条，它在数学结构里是否承重.

无论你是准备升学的学生、带学生的老师，还是想把数学重新学一遍的人，这本书都用得上.

本系列分卷

第 1 卷：预备知识、集合与映射、函数、函数方程、不等式
第 2 卷：三角函数、数列、统计、计数原理、概率
第 3 卷：平面向量、解三角形、解析几何、极坐标与参数方程、射影几何
第 4 卷：极限与导数、简单的积分、微分方程初步、数学思想浅谈
第 5 卷：复数、立体几何、矩阵与行列式、连分数、常数 e、π 与 γ、无穷乘积

公开版 • 10 月 7 日

Project Akari