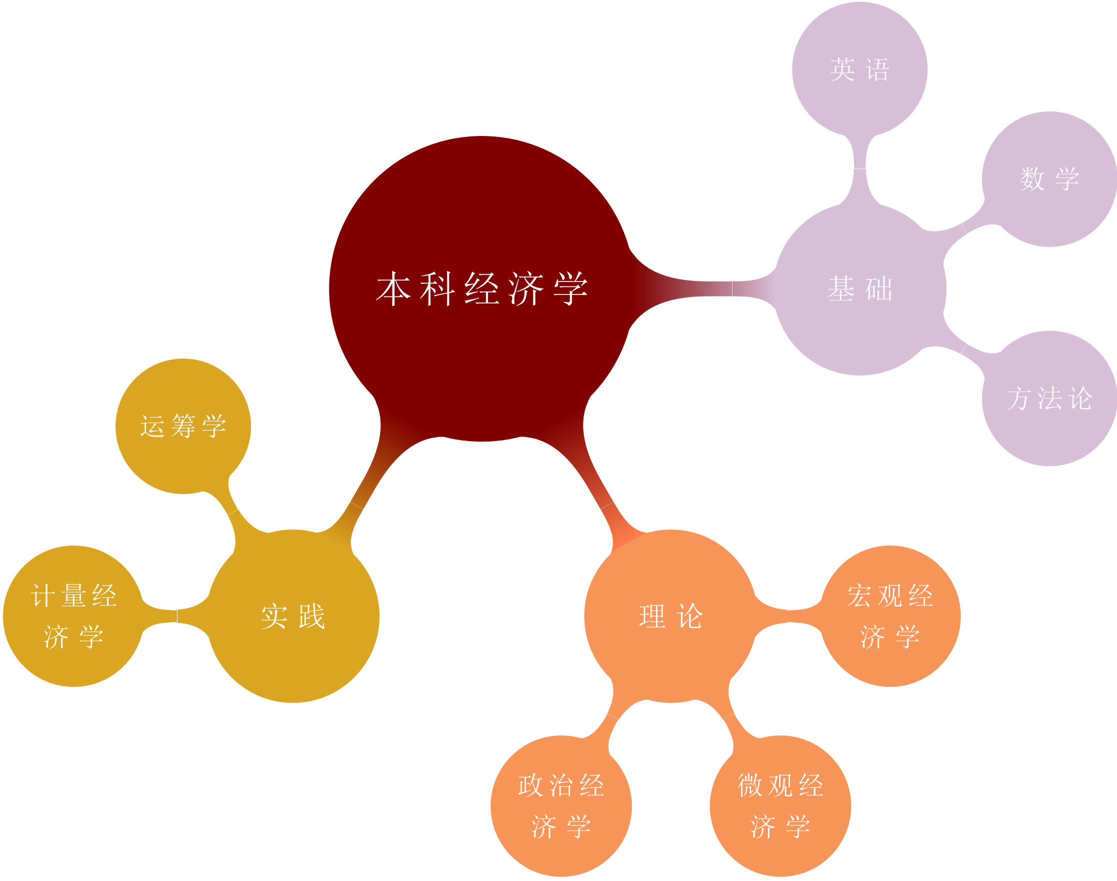

回顾本科经济学

Review of Undergraduate Economics

凌伟 苏念悠 王秋力 编著

吉林大学商学与管理学院
SCHOOL OF BUSINESS AND MANAGEMENT, JILIN UNIVERSITY

版权声明

如您有任何疑问与建议，请联系邮箱：lingwei3418@mails.jlu.edu.cn。

前言

在吉林大学商学与管理学院（原商学院）数量经济系的本科学习生活即将结束，回顾本科阶段的学习收获以形成完整的知识体系既对接下来的研究生学习阶段大有裨益，又能够将学习经验传递给正在学习相关知识的师弟师妹，让他们能够在本书的积累基础上更上一层楼。

写作本书的想法发端于我在保研中的夏令营准备过程，我常常思考吉大数量经济学的学生相对于其他学校学生的优势是什么、所学过的课程的联系在哪里、怎么认识经济学的科学性等等问题。这促使我在保研结束后的空窗期开始着手整理本科期间所学过程的知识。借助大三期间 R 语言书籍写作的经验积累和对 LaTeX 的掌握，我很快开始了本书的写作过程，在大四寒假之前逐步整理了中级微观经济学、中级宏观经济学、计量经济学等课程的知识点。到了今年 5 月，在完成毕业论文以后，我邀请同专业的苏念悠和王秋力同学参与本书的写作，逐步完成了余下内容的写作工作。

本书的主要内容包括方法论、知识体系和阅读书单三部分。方法论包括西方经济学的哲学原理和笔者自创的管理学习方法，知识体系包括英语、数学、微观经济学、宏观经济学、计量经济学、运筹学和政治经济学等课程的核心知识点，阅读书单包括笔者在本科期间度过的有价值的书籍，供读者参考。

本书的排版是在 ElegantLaTeX 项目的 ElegantBook 模板的基础上完成的，ElegantLaTeX 项目的官网是https://elegantlatex.org/，Github 地址是https://github.com/ElegantLaTeX/ 。

读者如何使用本书：对于大一大二的读者，建议阅读本书的前三章，学习实践管理学习学习方法和英语学习建议；对于大三大四的读者，建议将第 4-9 章作为专业课的辅助复习资料，可以尝试构建自己的知识体系。建议读者在学习经济学的过程中，对同一理论或问题参考多个文献，注意培养研究与写作的能力。

本书第 1、2、4、5、6、7、8、10 章由我撰写，第 3 章由苏念悠撰写，第 9 章由王秋力撰写。感谢二位合作者的辛勤付出。笔者在学习和写作过程中受到吉林大学诸多老师和同学的帮助，特别是商学与管理学院数量经济系的老师，在此表示由衷的感谢。特别感谢方毅老师、刘洋老师和刘达禹老师在学业和科研方面对我的指导和帮助。限于作者水平有限，时间有限，不免有错漏之处，欢迎读者批评指正。

凌伟
June 30, 2022

目录

第一部分　方法论 **1**

第 1 章　哲学基础与经济学原理 **2**

第 2 章　管理学四要素学习方法 **4**

第二部分　知识体系 **8**

第 3 章　英语 **9**
- 3.1 语言习得理论 9
- 3.2 英语实践 10
- 3.3 四六级备考建议 12
- 3.4 总结 13

第 4 章　数学 **14**
- 4.1 微积分 14
- 4.2 线性代数 36
- 4.3 概率论与数理统计 49

第 5 章　微观经济学 **67**
- 5.1 引言 67
- 5.2 消费者理论 67
- 5.3 生产者理论 72
- 5.4 价格歧视和博弈 77
- 5.5 交换和福利 82
- 5.6 公共选择理论 84

第 6 章　宏观经济学 **86**
- 6.1 引言 86
- 6.2 古典理论：长期中的经济 88
- 6.3 增长理论：超长期中的经济 92
- 6.4 经济周期理论 94
- 6.5 宏观经济理论与政策 107

第 7 章　计量经济学 **112**

7.1 引言 112
7.2 经典计量经济学 112
7.3 微观经济计量学 123
7.4 时间序列分析 143

第 8 章 运筹学 **156**
8.1 线性规划与单纯形法 156
8.2 对偶问题 161
8.3 运输问题 167
8.4 整数规划 175
8.5 0-1 规划 176
8.6 网络优化 184

第 9 章 政治经济学 **190**
9.1 生产力、生产关系和生产方式 190
9.2 劳动、商品和价值 192
9.3 货币和货币流通量 195
9.4 资本 196
9.5 社会总资本再生产和市场实现 199
9.6 信用制度与虚拟资本 200
9.7 竞争和垄断 203

第三部分 阅读书单 **205**

第 10 章 笔者已读书单 **206**
10.1 经济理论与实践 206
10.2 研究与写作 207
10.3 哲学与文化 207
10.4 程序设计 208
10.5 其他 209

第一部分

方法论

第 1 章 哲学基础与经济学原理

哲学是所有知识的终极指向，西方经济学和马克思主义政治经济学发端于不同的哲学理论，掌握一定的哲学知识可以帮助更好地理解和应用经济学原理。可惜的是，目前笔者没有接触到讲解经济学的哲学原理的课程，最多是经济思想史这样的课程。但是从哲学层次把握经济学的原理和应用对于学习经济学又比较重要，至少可以破除经济学无用论，坚定读者对经济学原理的科学性的信心。因此，笔者尝试结合《西方经济学理论的经验论哲学基础》、《理解现代经济学》等文献，和读者分享笔者对经济学理论的认识。

西方经济学的哲学基础之一是经验论，洛克主张"知识源于经验，人的心灵在经验印入之前是一张白板，经验由感官印象而来，任何观念最终都可还原为感觉印象，或者说是对感觉的反省。"经验论是功利主义、主观相对主义和方法论个人主义的哲学基础。功利主义的创立者是边沁，有三个方面的内容，

1. 从个体角度，功利主义是快乐主义；
2. 从社会角度，功利主义追求最大多数人的最大利益；
3. 从分析工具角度，功利主义表现为心理联想主义，道德观念是在苦乐感觉的联想基础上形成的。

效用理论就建立在功利主义基础上，个人追求效用最大化。群体效应最大化的理论便是 GDP，产出品越多，全体效用越高。

主观相对主义认为"由于人们的感官所感知的只是感觉和观念而不是外部世界，因此，我们没有任何根据从这种感觉和观念推断外部事物的存在。相反，从感觉只能推断存在一个感知它的感知者。因为观念只能存在于一个主动的感知它的实体中，这个感知者或实体就是心灵、精神、灵魂或自我。"主观相对主义在经济学中的体现是

1. 资源是稀缺的
2. 边际效用价值理论

边际效用价值理论认为物品的价值取决于效用和稀缺性，效用是消费一种物品或服务得到的主观享受或有用性，稀缺性是指只有当物品相对于人的欲望来说是稀缺的，它才对人有价值。价值在经济学中的实现就是价格，供给曲线和需求曲线既体现了消费者对物品的效用评价，又体现了由生产限制导致的物品的稀缺性。在运筹学中，如果物品的生产限制是松弛的，那么它的价格为零。

方法论个人主义强调知识来自个人的心灵所接受的感觉和经验，而非普遍理性，因而个人是对自己的幸福的最好判断者，个人最有资格关心自己的福利。在西方主流经济学中，人们的经济行动不是整个国家和社会的经济行动，而总是个人的行动，以利用某种物品的一定量为目标。从认识论上，经济现象的解释要从个人的经济决定和行为出发，上升到整个社会的经济现象，也就是说宏观现象必定要有微观基础。

理解西方经济学哲学基础的过程是在经济学的学习过程中潜移默化完成的，对于初学者来说，掌握经济学的原理、培养经济学直觉更为重要。钱颖一将经济学原理总结为两个出发

点、一个落脚点和三个基本原理，如下

- 出发点（假定）：资源稀缺性、个人理性；
- 落脚点（标准）：效率；
- 基本原理：人们对激励做出反应、市场值资源配置的有效方式、创新是推动经济持续增长的最终力量。

经济学原理只在初级经济学课程中讲授，在中高级课程中被当做公认的原理。当笔者重新阅读曼昆的《经济学原理》教材时，才体会到这一点，经济学原理是经济学与其他社会科学的根本区别，也是经济学大厦的理论基础。这里简单谈一谈效率，在经济学中，效率是把事情做得更好，对于个人来说，效率不只是在单位时间内获取和消化了多少知识、完成了多少工作，在广义层面上，效率是个人是否做到了可以做到的最好，比如在学习中，时间和精力的投入、学习方法、学习规划等决定是否能使个人达到潜在的学习水平，而对内卷、挂科等的担心是无效率的。

第 2 章 管理学四要素学习方法

管理学四要素学习方法（以下简称管理学习方法）是笔者在大一学习蔡玉程老师讲授的《管理学》课程时总结出的学习方法，其核心是将采用管理学中的四要素规划学习过程。笔者三年的学习实践证明此方法适用于经管学科的本科学习，故分享给读者。

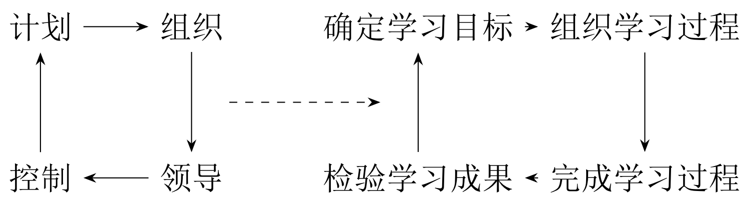

图 2.1: 管理学四要素与学习过程

管理学四要素指的是*计划、组织、领导、控制*，这分别对应学习的四个阶段，*确定学习目标、组织学习过程、完成学习过程、检验学习成果*。将工作过程转化为清晰的、可执行的步骤的思维方式是结构化思维。在学习中运用结构化思维的好处是使学习者对学习过程有清晰的宏观把握，清楚知道目前在学习的哪一个阶段，是否因跳过了某个阶段导致学习中的困难，并使得学习者对当期的学习工作充满自信，对阶段性的学习成果有合理预期。下面以以笔者《中级宏观经济学》期末复习过程为例，具体介绍每个学习阶段的工作和实施准则。

图2.2是中级宏观经济学复习过程的规划，包含学习目标、过程规划以及阶段性的学习成果预期，这是实施管理学习方法的主要依据。首先，在*确定学习目标*阶段，根据实现时间的长短可将学习目标分为两类，长期学习目标和短期学习目标，前者如培养经济学研究能力，后者如通过期末考核。长期和短期学习目标的关键差别在于清晰度、可量化程度和可执行程度，短期学习目标更容易转化为清晰的、可量化的、可执行的目标，比如期末考核的目标可以是成绩达到 90 分以上。但是短期目标的问题在于缺乏可持续性，并且可能导致短视行为，比如拒绝合作，这就需要用长期目标来指导和调和短期目标，达到长期学习的一致性和持续性。作为经管专业的学生，专业知识、通识、计算机能力、英语交流能力、研究与写作能力、思维能力等等都可以作为长期的学习目标。在制定短期的学习目标时，应该和长期学习目标相契合，高屋建瓴。比如图2.2中的复习目标，就和专业知识与研究与写作能力的长期学习目标相契合。读者可能会期待这里有“期末考核达到 A”这样的短期目标，这里没有写的原因是在整体的复习规划下，这个短期目标是附带的结果，其重要性不及长期目标的短期实现。笔者在设定学习目标时的原则是长期目标的短期实现、清晰和可执行，但是建议读者在初次应用管理学习方法设定学习目标时在前三条原则基础上增加可量化原则，建立短期的反馈机制，更容易养成学习习惯。

其次，在*组织学习过程*阶段，可以从两个维度来组织学习过程第一，根据学习的投入要素来分类组织，包括学习材料、地点和时间。学习材料的组织既包括书籍、期刊、文具等具体材料的准备，确保在指定的学习时点有相应的材料，不再花时间临时搜寻，又包括学习材料的学习内容和学习方式。地点的组织要求学习环境安静、相对固定和方便学习，目的是营造

中级宏观经济学复习规划

凌伟 2020.11.28

一、考试时间

第 17 周或 18 周，闭卷考试，题型包括选择、名词解释、简答、计算

二、复习目标

中级宏观经济学的特色为经济学理论结合数学模型分析宏观经济学问题，最重要的是分析经济学理论对经济系统运行的描述，并从理论过渡到模型，进一步分析经济现象。故复习目标：掌握中级宏观经济学的教学脉络，掌握中级宏观所关注的经济现象及相关的理论解释，掌握中级宏观经济学模型方法。

三、复习规划

1. 第一阶段——回顾教材

时间：11.28-12.3，8 学时

要求：浏览教材，记录教学重点并形成教学思维导图和重要知识点清单。

产出：教学思维导图，重点清单。

2. 第二阶段——全线出击

时间：12.3-12.17，24 学时

要求：细致复习教材和课后习题，关注重点概念，经济现象及相应的理论解释和经济模型，并学会运用理论模型分析经济现象。

产出：重点概念清单，经济现象及理论汇总清单，主要经济模型的假设、分析、表达、结论清单。

3. 第三阶段——重回教材

时间：12.18-12.20，6 学时

要求：重新细致阅读教材，回顾主要的教学内容，了解相对冷门的知识点，并运用经济理论和模型分析新的经济问题。

产出：教学内容回顾笔记，冷门知识点清单，新的经济现象分析笔记。

图 2.2: 中级宏观经济学复习规划示例

学习的氛围。笔者一般选择图书馆和教学楼。时间的组织要求在充足而不富余的时间内有效完成学习任务，包括时间点和持续时间。学习材料、地点和时间的组织是在学习目标的指导下相互协调统一的，比如学习微积分时，需要进行大量推导，就要求环境安静，时间相对充裕，而且要避免被打扰。第二，将学习目标转换为阶段性目标，如图2.2中的三个阶段，笔者的思路是初步宏观把握、具体的微观把握和宏观把握，比如图2.2中复习规划的第一个阶段是制作思维导图和知识点清单，目的是初步把握全局内容，第二个阶段是细致复习教材和习题，目的是把握具体的知识点，第三个阶段是回顾全局内容，目的是建立具体知识点之间的连接，完成从点到面的升华。学习过程和组织应该把阶段性的学习内容组织和材料、地点、时间的组织结合起来，一步一步规划学习过程。

再次，在*完成学习过程*阶段，学习者在“组织学习过程”的要求下，具体完成学习的任务，既包括知识的输入，又包括知识的输出。学习过程是学习内容的输入、理解和输出的过程。学习材料决定学习内容的输入内容，学习的时间和地点决定学习内容的输入是否顺畅。在精力充沛的时间和安静整洁的环境中，学习内容的输入最为高效，同时大脑也更为专注，利于理解。学习内容的输入是理解和输出的前提，理解是输入和输出之间的桥梁，理解的目的是将新的内容增添到学习者原有的知识体系中，输出和检验和保存传播理解程度的过程，也可以倒逼学习者主动吸收输入的学习内容。因此，学习过程最好有输出的要求，比如图2.2中复习规划的各阶段中的产出。完成学习过程阶段是学习过程中投入时间和精力最多的阶段，也是学习收获的所有来源，只规划不完成不会带来任何收获。完成学习阶段要求学习者专注和坚持。

最后，在*检验学习成果*阶段，根据完成学习阶段的输出和额外检测分析是否达到学习目标，反思整个学习过程的问题与收获。完成学习阶段的输出是初步检验学习成果的依据，如果输出的数量和质量达到和规划中所预期的水平，那么认为达到了学习的目标，如果无法达到预期水平，那么需要判断核心的学习任务是否完成，比如图2.2中期末复习的核心任务是完成第二阶段的学习，如果核心学习任务未完成，那就需要反思问题所在并及时处理。如果学习目标过高，那就需要降低对自己的预期，如果学习规划的执行程度低，那就需要增加外部激励、改良学习习惯。在诸如期末考试等量化考试的学习准备过程中，额外检测是检验学习成果的有效方法，但是检测不能替代学习过程，检测的目的是判断学习的整体效果。

根据管理学习方法，学习过程的四个阶段的执行要点如下：

1. 确定学习目标
    - 长期目标的短期实现
    - 清晰、可量化、可执行
2. 组织学习过程
    - 组织学习材料、地点、时间
    - 划分学习的具体阶段，宏观微观宏观
3. 完成学习过程
    - 输入、理解、输出
    - 专注、坚持

4. 检验学习成果
    - 是否达到学习目标
    - 反思学习过程

经管学科的知识体系的特点是体量大、逻辑严密，核心是经济学原理。因此，读者在学习中应以经济学原理为基础，将新的学科知识架构在经济学原理基础之上，高等数学等方法类课程也应为经济学原理服务，要避免追求高深的数学形式和高超的编程技巧。另外，考试仅仅是一种检测学习成果的方式，在学习中应以长期的成长性目标为重，把绩点当做附加的奖励。建议养成规划和学习的习惯，自主学习，自由成长。

最后，时间是最宝贵的资源，笔者非常受用《管理学》给出的时间管理建议，在此与读者共勉：

1. Make and keep a list of all your current, upcoming, and routine goals.
2. Rank your goals according to importance.
3. List the activities/tasks necessary to achieve your goals.
4. Divide these activities/tasks into categories using an A, B, and C classification.
5. Schedule your activities/tasks according to the priorities you’ve set.
6. Plan your to-do list each day so that it includes a mixture of A, B, and C activities/tasks.
7. Realize that priorities may change as your day or week proceeds.
8. Remember that your goal is to manage getting your work done as efficiently and effectively as you can.

第二部分

知识体系

第 3 章 英语

在讲述具体理论之前，请大家思考一个问题：为什么要学英语？

有些人只想在英语考试中取得高分；有些人面临着像考试、升学、工作等压力，不得不学习英语；有些人觉得周围人英语好，自己也不能落下；有些人觉得说一口流利的英语让自己看起来很厉害；对于有些人来说，英语是工具——辅助他们更好地了解其他知识的工具；或者，也有些人觉得英语（以及其他语言）是一门艺术，值得细细推敲。对待英语的态度其实一部分决定了你的英语水平上限，所以如果真的想学好英语的话，反问一下自己，为什么要学好英语？究竟是在以哪种态度看待英语呢？

3.1 语言习得理论

3.1.1 语言学习本质

史蒂芬克拉申教授的语言习得理论中，学习语言的过程是：

> We acquire language in one way, and only one way, when we get comprehensible input in a low anxiety environment.

也就是说，在一个不那么焦虑的环境中，通过语言（你想学的那门语言）获得可以被我们理解的信息的过程，是最好的语言习得过程。

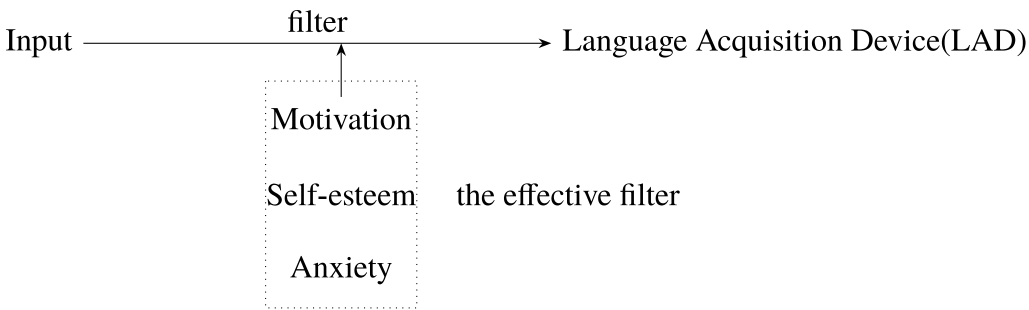

图 3.1: 语言习得过程

首先，语言其实就是我们传递信息的载体，要使用好语言，需要用这门语言去获得“可理解的信息输入”。什么是可理解的信息输入呢？如果对一个完全不懂英语的人说“apple”，无论你说的次数再多，说的声音有多大，或者你把这个词写下来给他看，他都不知道你要干什么。但是如果你拿着一个苹果，一边指示一边对他说“apple”，效果会好许多，而这时，“apple”就是可理解的信息输入。当“apple”成为能够被理解的信息时，语言习得的过程就发生了。因此，在学英语的过程中，一遍遍的重复等机械方法效果其实并不好，而能够通过英语来获取信息的过程才是你在与这门语言磨合的过程。

其次，有了可理解的信息输入之后，这些信息能不能以新语言的方式被大脑消化掉，则取决于“动力”、“自尊”、“焦虑程度”三者对大脑消化的阻碍程度。动力越强、越是自尊自信、焦虑越少的人，阻碍越少，往往会在英语学习中提高更快。相反，如果很排斥语言学习，如果在学习过程中不自信，总是怀疑“我能学好吗？”，或者觉得语言学习的过程暴露了自己的弱势所在，那么即使你在获得可理解信息输入，却阻碍了大脑的消化过程，语言习得的效果将大大降低。总结一下，在心理阻碍较少的情境下获得可理解的信息输入，语言习得就发生了。

3.1.2 减少思维弯路

对于第二语言的学习，局限于传统课堂的学习环境有时会让我们习惯于在思维上“绕弯路”。还是以“apple”为例：

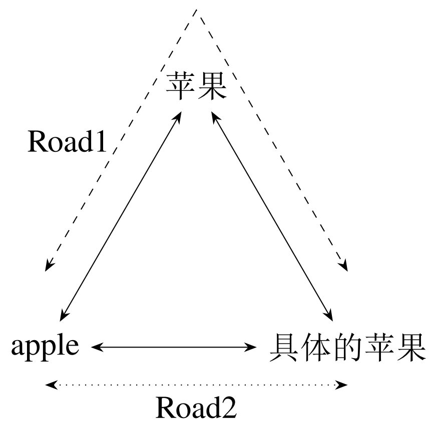

图 3.2: 语言理解路径

当遇到“apple”时，若先将“apple”翻译成中文“苹果”，然后再理解出那个很多人爱吃的水果的话，那其实采取了路径 1，绕弯路了。最佳的途径是由“apple”直接就能联想到苹果这个水果，即路径 2。这两条路径反过来走也是一样的，看到一个水果，你能直接脱口而出它的名字吗？还是想到中文，再把它翻译成英语呢？

在英语的学习过程中，我们需要做的就是减少走路径 1，从习惯于走路径 1 转变为习惯于走路径 2。当听到或读到一句英语，如果一开始需要把一些词换成中文才能理解，但是后来做到能立马理解语句所传达的信息，可以肯定，英语水平已经得到了很大提高。

3.2 英语实践

本节将介绍在实践中，有哪些方法可以通过可理解的信息输入减少思维弯路，提升英语水平。

3.2.1 长期阅读

长期阅读十分有助于培养英语思维。关于阅读有两方面需要注意：阅读材料的选取、长期坚持。

英语阅读材料种类很多，可以是英文原著、杂志、报纸、公众号的推送等等。阅读材料的选取上，第一点要尽量选取比自己现有的英语水平略高一些的阅读材料。阅读材料难度要适度，即做到：让阅读的过程成为获取“可理解的信息输入”的过程。如果硬读太难的文章，而文章里面几乎没有能看懂的句子，那么此时通过英语阅读获取的信息会非常少，这个过程会很费劲、也很难让人提起兴致。此外，也一定要保证阅读材料要比自己现有水平高一点，不要读过于简单的资料。毕竟要给大脑一些挑战，才能有提高。如果在阅读的过程中，虽然会碰到不认识的词汇，或者有一些句子看不太懂，但是能够理解文章的大意，此时阅读材料难度就是较为合适的。

第二点，阅读材料的选取要选择自己感兴趣的材料读。如果你喜欢经济学，那就找经济学的材料读；如果想读童话故事，那就找故事书读。阅读自己感兴趣的内容可以减少大脑消化新语言的阻碍。正如前文所说，动力越强、越是自尊自信、焦虑越少的人，阻碍越少，往往会在英语学习中提高更快。选择自己喜欢的内容，会提高英语阅读的动力，对所获取的信息更自信。而当英语阅读变成主动接收知识的过程时，对于英语学习的焦虑也变少了。

当你选择了合适的材料，并且读到非常精彩的内容时，有时会暂时忘记你所使用的语言是英语，此时重要的是你获取的信息，而不是英语本身——路径 2 发生了，你的语言水平也正在迅速提升。

除了会挑选材料，长期坚持同样重要。虽然有很多方法可以让大家短期快速提升考试分数，但从本质上提升自己的英语能力则需要长期积淀。若只有一两次通过路径 2 使用英语，只会让大脑接收到短期刺激，时间一长，大脑也就忘记怎么走路径 2 了。但是长期通过路径 2 使用语言，那么大脑就会将路径 2 视为习惯，英语的运用也越来越娴熟。

3.2.2 磨耳朵

在听力方面，听有声英语可以提升英语听力水平。与长期阅读类似，在听英语时，要选取难度适中、感兴趣的听力材料（通过耳朵获取可理解的信息输入）。判断标准同样是：可以有听不懂的单词或语句，能整体听懂大意；同样，也需要长期坚持，让大脑习惯于通过听到英语理解信息，“磨一磨”耳朵与大脑的配合。

3.2.3 英语交流

英语交流是传统教学中较为缺失的环节。若说阅读与听力是信息的输入，那么交流则同时涉及信息的输入与输出。同样，交流过程重在信息的传递，而不是英语本身。有条件的话可以找外国人交流，也可以找小伙伴一起练习。英语交流方面尤其需要注意“动力”、“自尊”、“焦虑程度”三个因素。阻碍很多人提升口语水平的因素并不是口语有多难，而是不好

意思开口、觉得自己说的不好、不自信这些阻碍因素。英语交流，关键在于交流信息，而不是口音有多纯正。学有余力可以再去练习好听的语调口音，但是准确传情达意其实更加重要。练习的过程中，可以跟小伙伴（或者外教）一起选择共同感兴趣的话题，提前做一下词汇、语句表达方面的准备，再一起交流。

3.3 四六级备考建议

如果考试迫在眉睫，或者只是想有一个漂亮的英语分数的话，短期也有一些提升方法。以四六级考试为例，这里将介绍一些备考技巧，希望有所帮助。

3.3.1 听力

听力要注意审题。四六级听力没有题干，在每部分听力开始之前要用最快的速度阅读选项。阅读选项的过程中，注意圈画一下重点词，着重注意选项中的主语（是 they，she，he，it，还是其他）以及重要的名词。四六级听力的四个选项往往都十分接近，要注意四个选项之间有区别的地方。这样的勾画也有助于大家在听力过程中定位，大体了解听力播放到了哪一个题目。

介绍一个方法——“听写法”：当做完整篇听力后，看看是哪一个 Section 错的最多或者是最听不懂，把最薄弱的那一篇拿出来听写一遍。听写的过程中，放一句然后按暂停键，把这一句写下来，再听下一句。听写第一遍肯定会有很多地方没写上，没关系，我们再听第二遍，在第一遍听写稿的基础上修改。还有听不出来的地方再听第三遍。直到自己觉得再也不能听到更多内容的时候，可以拿出听力原文对照修改听写稿了，这一次修改要用红笔（或者其他不同颜色的笔）。改完以后再比照着自己的听写稿听一遍，着重听有红笔修改的部分，思考一下为什么听了好几遍这个地方还是没听到，是词汇不会，还是出现了连读等原因呢？

这个方法需要大量的时间，但是亲测对听力提高很有帮助。如果大家实在是没有时间，可以先着重听写错题中选项的听力原文内容。

3.3.2 阅读

Section A 选词填空，最重要的就是词汇的积累，要背单词。此外也可以通过语法来选词，比如，空出来的地方应该填名词，形容词，动词还是副词？动词的话前面有没有第三人称？是过去时，一般现在时还是正在进行的 ing？被动还是主动？利用语法可以帮助大家缩小选词的范围，但是理解单词的意思仍是最重要的。

Section B 这一部分文章的篇幅很长，要求将题目信息与段落进行匹配。其实后面的题目就是将文章中某些句子进行改述，因此“定位”很重要。我会先通篇阅读原文，在阅读的过程中把人名、地名以及其他专有词（比如大学的名字，书籍名称等）标出来帮助定位。阅读全篇后看题目，再读题干并把题目对应到某几段甚至某一段，再快速浏览这一部分，能找到相近的表述的话，一道题就完成了。注意！这样的做题方法可能会费时间，如果阅读速度不是

很快的话，听说先读题目再阅读文章的方法也可以做到快速准确。很多方法都是因人而异的，希望大家能找到适合自己的方法。

Section C 有点像高中的阅读题，这部分的题目分值很高，需要尽量少出错。大部分的题目和选项也是能对应到原文的。也有选项需要对全文的把握。对于这一部分，个人偏向于先阅读原文再看题，本身文章篇幅不长，而且可以避免后面某些选项给我们造成先验印象（比如某个选项有可能会把我们的思路带偏，在读的过程中不自觉地往这个方向想）。阅读过程中同样要勾画重点，标注人名、地名以及其他专有词。

短期提高的话，就是要刷题，培养做题的感觉。刷题注意控制时间，提高做题的速度。

3.3.3 翻译写作

四六级翻译和写作重点是对词汇和语法的考察。这一部分真正的提高需要长期积累。短期备考的话，首先需要自己动手写一份答案，再对照答案，看看自己的词汇和句式可以怎样提高。作文部分还可以把多篇范文拿出来一起看，自己总结一下句型词汇，整理一份“私人订制”的模板。另外，注意及时应用好的句型词汇，若上一篇范文中的优秀语句，要有意识地在自己写作或翻译过程中套用一下。只有会用了知识才是自己的。

3.4 总结

回到最开始的问题，你为什么要学英语？不同的态度下，在英语上付出的精力也不同。这些态度没有好坏之分，重要的是，让自己的努力程度与想得到的结果匹配。以上内容是笔者十分受用的英语学习理念以及学习方法，供大家选择。提高英语的方法有很多，希望大家根据自己目的，选择合适的，整理出一套专属自己的学习方法。

第 4 章 数学

4.1 微积分

集合论是现代数学的基础，函数研究的是集合与集合之间的映射关系，微积分研究函数的微分和积分。微积分从集合与函数开始，逐步分析导数、微分、积分及其在多元上的拓展，最后到微分方程与差分方程。微积分在本科阶段的经济学中的应用主要是导数、积分与函数优化。由于数学概念非常多，因此本文在首次使用某个数学概念时会强调而不是单独进行定义，如*函数*。

4.1.1 函数与极限

*集合*是具有某种确定性质的对象的全体，组成集合的每一个对象称为该集合的*元素*。约定集合采用大写拉丁字母表示，如集合 A，而元素采用小写拉丁字母表示，如元素 a。元素 a 要么属于集合 A，记为 $a \in A$，要么不属于集合 A，记为 $a \notin A$。

集合按元素数量可分为三类：有限集、无限集和空集，空集用 $\emptyset$ 表示。集合的表示法分为列举法和描述法，如 $A = \{1, -1\}$ 和 $A = \{x|x^2 - 1 = 0\}$。

如果 $\forall a \in A, a \in B$，则集合 A 是集合 B 的*子集*，记为 $A \subset B$ 或 $B \supset A$。如果 $A \subset B$ 和 $B \subset A$ 同时成立，则集合 A, B *相等*，记为 $A = B$。

微积分研究的是*数集*，即元素均为数的集合。重要的数集如下

- 自然数集 $N = \{0, 1, \cdots\}$
- 整数集 $Z = \{\cdots, -1, 0, 1 \cdots\}$
- 有理数集 Q
- 实数集 R
- 复数集 C

特别的，正整数集表示为 Z^+，负整数集表示为 Z^-，其他类似。

集合有三种运算，并交差

- 并集 $A \cup B = \{x|x \in A \text{ or } x \in B\}$
- 交集 $A \cap B = \{x|x \in A \text{ and } x \in B\}$
- 差集 $A \backslash B = \{x|x \in A \text{ but } x \notin B\}$

特别的，对于特定问题涉及元素的集合称为*全集*，记为 Ω，将 Ω 中属于集合 A 的元素删去，得到集合 A 相对于全集 Ω 的*补集*，记为 $\bar{A} = \Omega \backslash A$。

集合的运算服从以下规律，这些规律对于分析集合之间的关系非常有用。

1. 交换律：$A \cup B = B \cup A, A \cap B = B \cap A$
2. 结合律：$(A \cup B) \cup C = A \cup (B \cup C), (A \cap B) \cap C = A \cap (B \cap C)$
3. 分配律：$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

4. 自反律：$A \cup A = A, A \cap A = A, A \cup \emptyset = A, A \cap \emptyset = \emptyset$
5. 并一律：如果 $\forall A_i \subset B, i = 1, 2, \cdots$，则 $\cup_{i=1}^{\infty} A_i \subset B$；如果 $\forall A_i \supset B, i = 1, 2, \cdots$，则 $\cap_{i=1}^{\infty} A_i \supset B$
6. 对偶律：$\overline{\cup_{i=1}^{\infty} A_i} = \cap_{i=1}^{\infty} \bar{A}_i, \overline{\cap_{i=1}^{\infty} A_i} = \cup_{i=1}^{\infty} \bar{A}_i$

区间是一类特殊的数集，是由连续的实数构成的集合，包括四类：闭区间 $[a, b] = \{x|a \leq x \leq b\}$、开区间 $(a, b) = \{x|a < x < b\}$、半开半闭区间，如 $(a, b] = \{x|a < x \leq b\}$ 和无限区间，如 $(a, +\infty) = \{x|a < x\}$。前三类是有限区间，$a, b$ 分别是区间的左端点和右端点，$b - a$ 为区间长度，∞ 为无穷大，加 $+$ 为正无穷大，加 $-$ 为负无穷大。区间用 I 或 X 表示。

邻域是由到某点一定距离以内的实数构成的集合。设 $a, \delta \in R, \delta > 0$，则数集 $U(a, \delta) = \{x||x - a| < \delta\} = (a - \delta, a + \delta)$ 为点 a 的 δ 邻域，点 a 称为邻域的中心，δ 称为邻域的半径。去掉点 a 以后，领域 $U(a, \delta)$ 变为去心 δ 邻域，记为 $\dot{U}(a, \delta) = x|0 < |x - \delta| < \delta = (a - \delta) \cup (a + \delta)$，$(a - \delta)$ 为左 δ 邻域，$(a + \delta)$ 为右 δ 邻域。

对于非空集合 X, Y，如果 $\forall x \in X$，按照对应法则 f 都有唯一的 $y \in Y$ 与之对应，则 f 为从 X 到 Y 的映射。y 为在映射 f 下的 x 的像，$y = f(x)$，x 为在映射 f 下 y 的原像。集合 X 为映射 f 的定义域，记为 $D_f = X$，X 所有元素的像的组合为映射 f 的值域，记为 $R_f = \{y|y = f(x), y \in Y, x \in X\}$。

如果 $\forall x_1 \neq x_2, x_1, x_2 \in X$，都有对应的像 $y_1 \neq y_2$，则映射 f 为单射。如果 $R_f = Y$，则映射 f 为满射。如果 f 既是单射又是满射，则 f 为一一映射。对于一一映射，对于任意的 $y \in Y$，都存在唯一的 $x \in X$ 与之对应，这样的对应关系为映射 f 的逆映射，记为 $f^{-1} : Y \to X$。

函数是从非空数集 $X \subset R$ 到 $Y = R$ 的映射，记为 $f : X \to Y$。X 为函数 f 的定义域，记为 D_f，$R_f = \{f(x)|x \in D_f\}$ 为函数 f 的值域。函数可记为

$$y = f(x), x \in D_f \tag{4.1}$$

x 为自变量，y 为因变量。一般地，函数的定义域默认为使公式有意义的所有数值。

在平面直角坐标系中，点集 $\{(x, y)|y = f(x), x \in D_f\}$ 称为函数 $y = f(x)$ 的图像，通常是一条曲线。与映射不同，在函数中，可能存在一个自变量对应多个因变量的情形，此时的函数为多值函数，否则为单值函数。

对于两个函数，$y = f(u), u = \phi(x)$，修正定义域以后，对于 $\forall x \in D$，都有 $u \in R_D, y \in R_f$ 与之对应，则函数 $g(x) = f[\phi(x)]$ 为复合函数，u 为中间变量。

初步把握一个函数，可以从以下方面入手：

- 有界性：如果存在正数 M 使得 $\forall x \in X$，都有 $|f(x)| \leq M$，则 $f(x)$ 在 X 上有界。
- 单调性：$\forall x_1 < x_2 \in D_f$，都有 $f(x_1) < (>) f(x_2)$，则称 $f(x)$ 为在 D_f 上单调增加（减少）函数。将 $>, <$ 变为 $\geq, \leq$，相应的 $f(x)$ 为单调不减（不增）函数。
- 奇偶性：如果函数 $f(x)$ 关于坐标原点对称，则为奇函数，即 $f(-x) = -f(x)$；如果函数 $f(x)$ 关于 $x = 0$ 对称，则为偶函数，即 $f(x) = f(-x)$。口诀：函数的和，奇偶不变；函数的积，奇奇得偶，奇偶得奇，偶偶得偶；函数复合，内奇同外，内偶则偶。
- 周期性：存在 $T > 0$，$\forall x \in D_f$，都有 $x + T \in D_f, f(x) = f(x + T)$。$T$ 为 $f(x)$ 的一个周期。

在本阶段，微积分的研究对象都是由 6 个基本初等函数经过有限次四则运算和复合运算构成的初等函数。基本初等函数如下：

- 常函数：$y = C, C$ 为常数
- 幂函数：$y = x^{\mu}, \mu$ 为实常数
- 指数函数：$y = a^x, a > 0, a \neq 1, a$ 为常数
- 对数函数：$y = log_a x, x > 0, a \neq 1, a$ 为常数
- 三角函数：$y = sin(x), cos(x), tan(x), sec(x), csc(x), cot(x)$
- 反三角函数：$y = Arcsin(x), Arccos(x), Arctan(x)$，默认取主值分支

给出函数的基本定义以后，下面分析函数的极限和连续性。首先从数列极限开始。

数列是按照正整数的顺序排列起来的无穷个数：$x_1, x_2, \cdots$，记为 $\{x_n\}$，每个数为数列的项，第 n 个数为数列的第 n 项或通项。类似函数，数列也有单调性和有界性。

对于数列 $\{x_n\}$ 和常数 a，如果对于任意给定的正数 ε，$\exists N \in N^+$，当 $n > N$ 时，都有 $|x_n - a| < \varepsilon$。则称 a 是数列 $\{x_n\}$ 的极限，记为 $\lim\limits_{n\to\infty} x_n = a$ 或 $x_n \to a(n \to \infty)$。当数列有极限时，它是收敛的，否则是发散的。

数列极限有三个性质：

1. 唯一性：若数列收敛，则极限唯一。
2. 有界性：收敛数列必有界。
3. 保号性：若 $\lim\limits_{n\to\infty} x_n = a, \lim\limits_{n\to\infty} y_n = b$，且 $a > b$，则存在正整数 N 使得当 $n > N$ 时有 $x_n > y_n$。特别的，$y_n = b$ 或 $y_n = 0$ 时这一性质也成立。如果 $\exists N \in N^+$，当 $n > N$ 时，$x_n \geq (\leq) 0$，则 $a \geq (\leq) 0$。

设函数 $f(x)$ 在点 x_0 的某去心邻域有定义，A 为常数。如果对于任意给定的整数 ε，总存在正数 δ，使得当 $0 < |x - x_0| < \delta$ 时，恒有不等式 $|f(x) - A| < \varepsilon$ 成立，则函数 $f(x)$ 当 x 趋向 x_0 时有极限 A，或 $f(x)$ 在 x_0 处有极限 A，记为 $\lim\limits_{x\to x_0} f(x) = A$ 或 $f(x) \to A(x \to x_0)$。当 x 在 x_0 的左邻域（右邻域）里变化时，相应的极限是函数在 x_0 处的左极限（右极限），分别记为 $\lim\limits_{x\to x_0^-} f(x) = A$ 或 $\lim\limits_{x\to x_0^+} f(x) = A$。左极限和右极限都是单侧极限。极限 $\lim\limits_{x\to x_0} f(x) = A$ 等价于 $\lim\limits_{x\to x_0^-} f(x) = \lim\limits_{x\to x_0^+} f(x) = A$。

当 $x \to \infty, +\infty, -\infty$ 时，函数的极限可类比上述去心邻域的极限定义。同样的，极限 $\lim\limits_{x\to\infty} f(x) = A$ 等价于 $\lim\limits_{x\to-\infty} f(x) = \lim\limits_{x\to+\infty} f(x) = A$。

类比数列极限的性质，函数极限也有唯一性、局部有界性和局部保号性。复杂函数的极限可以根据函数极限的运算法则计算。函数的四则运算法则，设 $\lim f(x) = A, \lim g(x) = B$，则

- $\lim[f(x) \pm g(x)] = A \pm B$
- $\lim[f(x)g(x)] = AB$
- $B \neq 0, \lim \frac{f(x)}{g(x)} = \frac{A}{B}$

特别的，$\lim[cf(x)] = c \lim f(x), c$ 为常数；$\lim[f(x)]^k = [\lim f(x)]^k, k$ 为正整数。复合函数的运算法则为：$\lim\limits_{x\to x_0} f[\varphi(x)] = f[\lim\limits_{x\to x_0} \varphi(x)]$。

以上两个运算法则的前提是知道 $f(x), g(x)$ 的极限，这里给出极限存在准则，可以帮助确定极限。

夹逼准则：设函数 $f(x), g(x), h(x)$ 在点 x_0 的某去心邻域有定义，且

1. $g(x) \leq f(x) \leq h(x)$
2. $\lim\limits_{x \to x_0} g(x) = \lim\limits_{x \to x_0} h(x) = A$

则极限 $\lim\limits_{x \to x_0} f(x)$ 存在且等于 A。

单调有界准则：单调有界数列必有极限。

根据以上两个准则，可以得出两个重要极限：

- $\lim\limits_{x \to 0} \frac{sin(x)}{x} = 1$
- $\lim\limits_{x \to \infty} (1 + \frac{1}{x})^x = e$

特别的，如果在 x 的某个变化过程中，$\lim f(x) = 0$，则 $f(x)$ 为 x 在该变化过程中的无穷小量。无穷小的以下性质可用于快速确定某个函数是否为无穷小：

- 有限个无穷小的代数和为无穷小
- 有界变量与无穷小的乘积为无穷小

如果某个变化过程中函数可以拆分为无穷小和常数，则极限存在，即 $\lim f(x) = A$ 等价于 $f(x) = A + \alpha(x), \alpha(x)$ 为无穷小。

在函数极限分析中，我们还关心极限收敛的快慢，这可以通过无穷小的比较分析。设 $\alpha = \alpha(x), \beta = \beta(x)$ 是自变量在同一变化过程中的无穷小，

1. 如果 $\lim \frac{\beta}{\alpha} = c \neq 0$，则 β 与 α 是同阶无穷小
2. 如果 $\lim \frac{\beta}{\alpha} = 1$，则 β 与 α 是等价无穷小，记为 $\beta \sim \alpha$
3. 如果 $\lim \frac{\beta}{\alpha} = 0$，则 β 是 α 的高阶无穷小
4. 如果 $\lim \frac{\beta}{\alpha^k} = c \neq 0$，则 β 是 α 是 k 阶无穷小

特别的，$\alpha \sim \beta$ 的充要条件的 $\alpha - \beta$ 是 α 或 β 的高阶无穷小。如果 $\alpha, \beta, \alpha', \beta'$ 是自变量同一变化过程中的无穷小，且 $\alpha \sim \alpha', \beta \sim \beta', \lim \frac{\beta'}{\alpha'}$ 存在，则 $\lim \frac{\beta}{\alpha} = \lim \frac{\beta'}{\alpha'}$。

常用的等价无穷小如下：$x \to 0, x \sim sin(x) \sim tan(x) \sim arcsin(x) \sim arctan(x) \sim ln(1 + x) \sim e^x - 1; 1 - cos(x) \sim \frac{x^2}{2}; (1+x)^a - 1 \sim ax(a \neq 0); a^x - 1 \sim xln(a), a > 0, a \neq 1$。

与无穷小相对的概念是无穷大。设函数 $f(x)$ 在 x_0 的某个去心邻域内有定义，如果对于任意给定的任意大证书 M，都存在正数 δ，当 $0 < |x - x_0| < \delta$ 时，都有 $|f(x)| > M$。则称函数 $f(x)$ 为当 $x \to x_0$ 时的无穷大量，记为 $\lim\limits_{x \to x_0} = \infty$ 或 $f(x) \to \infty(x \to x_0)$。类似的，当 $x \to \infty$ 时也可定义无穷大。

极限是为函数连续性分析做铺垫的。设函数 $y = f(x)$ 在点 x_0 的某个去心邻域内有定义，如果当 $x \to x_0$ 时 $f(x)$ 的极限存在，且等于 $f(x_0)$，即 $\lim\limits_{x \to x_0} f(x) = f(x_0)$。则称函数 $f(x)$ 在点 x_0 处连续，且称 x_0 为函数 $f(x)$ 的连续点。如果只取左邻域或右邻域，即 $\lim\limits_{x \to x_0^-} f(x) = f(x_0)$ 或 $\lim\limits_{x \to x_0^+} f(x) = f(x_0)$，则函数 $f(x)$ 在点 x_0 处左连续或右连续。

$\lim\limits_{x \to x_0} f(x) = f(x_0)$ 等价于 $\lim\limits_{x \to x_0^-} f(x) = f(x_0)$ 和 $\lim\limits_{x \to x_0^+} f(x) = f(x_0)$ 同时成立。

如果函数在区间 I 内每一点都连续，则称函数是 I 上的连续函数。特别的，区间 I 包含端点时，函数 $f(x)$ 只需在左端点处右连续，在右端点处左连续。

如果函数 $f(x)$ 在点 x_0 处不连续，则称点 x_0 是函数 $f(x)$ 的间断点，共有两种类型：

1. 第一类间断点：$f(x)$ 在 x_0 处左极限和右极限都存在。
   - 可去间断点：$\lim\limits_{x\to x_0^-} f(x) = \lim\limits_{x\to x_0^+} f(x) \neq f(x_0)$
   - 跳跃间断点：$\lim\limits_{x\to x_0^-} f(x) \neq \lim\limits_{x\to x_0^+} f(x)$
2. 第二类间断点：除第一类间断点的其他情形，即左极限或右极限不存在。如无穷间断点、振荡间断点。

根据极限的四则运算法则，连续函数的四则运算得到的函数仍然是连续函数。而且，连续函数经过复合后也是连续函数。基本初等函数都是连续的，根据以上两条法则，所有初等函数都是连续函数。函数的连续性是运用微积分分析函数性质的重要前提。

连续函数有两个重要性质，经常会用到：

- 最值定理：如果函数 $f(x)$ 在闭区间 $[a,b]$ 上连续，则必定存在最大值和最小值。
- 介值定理：如果函数 $f(x)$ 在闭区间 $[a,b]$ 上连续，且 $f(a)\neq f(b)$，值域为 R_f，则 $\forall \mu \in R_f$，都至少存在一点 ε 使得 $f(\varepsilon)=\mu$。

特别的，在介值定理中，如果 $f(a), f(b)$ 异号，则必定存在 $\varepsilon \in (a,b)$ 使得 $f(\varepsilon)=0$，这就是根的存在性定理，也称零点定理。使用零点定理比介值定理更为方便。

4.1.2 导数与微分

导数是分析函数变化过程的有力工具，导数是函数变化过程最精确的刻画。

设函数 $y=f(x)$ 在点 x_0 的某邻域内 U 有定义，当自变量 x 在点 x_0 处取得增量 $\Delta x, x+\Delta \in U$ 时，相应的函数增量为 $\Delta y = f(x+\Delta x) - f(x)$，如果极限 $\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}$ 存在，则称函数 $y=f(x)$ 在点 x_0 处可导，并且上述极限为函数 $y=f(x)$ 在点 x_0 处的导数，记为 $f'(x_0)=\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}$。如果上述极限不存在，则函数 $f(x)$ 在点 x_0 处不可导。如果上述极限为无穷大，则函数 $f(x)$ 在点 x_0 处的导数为无穷大，记为 $f'(x_0)=\infty$。

相应的，将上述 Δx 变化过程调整为 $\Delta x \to 0^-$，上述导数定义变为左导数，记为 $f'_-(x_0)$。同样可以定义右导数 $f'_+(x_0)$。根据极限的性质，函数 $f(x)$ 在点 x_0 处可导等价于左导数和右导数存在且相等，即 $f'(x_0) \Leftrightarrow f'_-(x_0) = f'_+(x_0)$。

如果函数 $f(x)$ 在区间 I 内每一点都可导，自变量和导数的映射构成新的函数，称为导函数，简称导数，记为 $f'(x), x\in I$。在函数的图像中，导数 $f'(x_0)$ 就是函数的曲线在点 $(x_0, f(x_0))$ 处的切线斜率，即切线方程为 $y-f(x_0)=f'(x_0)(x-x_0)$。同样的，法线方程为 $y-f(x_0) = -\frac{1}{f'(x_0)}(x-x_0)$。

函数的可导性和连续性有如下关系：函数 $y=f(x)$ 在点 x_0 处可导 $\to$ 函数 $y=f(x)$ 在点 x_0 处连续，注意反过来不成立。

对于基本初等函数的函数，可以通过求导法则得到，而初等函数的函数可以由导数的四则运算法则和链式法则得到。

导数的四则运算法则如下：

- $[u(x) \pm v(x)]' = u'(x) \pm v(x)'$
- $[u(x)v(x)]' = u'(x)v(x) + u(x)v'(x)$
- $[\frac{u(x)}{v(x)}]' = \frac{u'(x)v(x)-u(x)v'(x)}{v^2(x)}$

特别的，$[Cu(x)]' = Cu(x)', [\frac{1}{v(x)}]' = -\frac{v'(x)}{v^2(x)}$，

$(u_1 u_2 \cdots u_n)' = u_1' u_2 \cdots u_n + u_1 u_2' \cdots u_n + \cdots + u_1 u_2 \cdots u_n'$。

导数的链式法则：$\{f[\varphi(x)]\}' = [f(u)|_{u=\varphi(x)}]' = f'(u)|_{u=\varphi(x)}\varphi'(x)$。

反函数的求导法则：如果函数 $x = g(u)$ 在 I_y 上单调可导，值域为 I_x，$g'(y) \neq 0$，则其反函数 $y = g^{-1}(x) = f(x)$ 在 I_x 上可导，且：$f'(x) = \frac{1}{g'(y)}$ 或 $\frac{dx}{dy} = \frac{1}{dy/dx}$。

基本初等函数的导数如下：

- $C' = 0, C$ 为常数
- $x\mu' = \mu x^{\mu-1}$
- $sin(x)' = cos(x)$
- $cos(x)' = sin(x)$
- $tan(x)' = sec^2(x)$
- $cot(x)' = -csc^2(x)$
- $sec(x)' = sec(x)tan(x)$
- $csc(x)' = -csc(x)cot(x)$
- $a^{x\prime} = a^x ln(a), a > 0, a \neq 1$
- $e^x = e^x$
- $log_a(x)' = \frac{1}{xln(a)}$
- $ln(x)' = \frac{1}{x}$
- $arcsin(x)' = \frac{1}{\sqrt{1-x^2}}, |x| < 1$
- $arccos(x)' = -\frac{1}{\sqrt{1-x^2}}, |x| < 1$
- $arctan(x)' = \frac{1}{1+x^2}$
- $arccot(x)' = -\frac{1}{1+x^2}$

如果导数可导，那么导函数的导数就是函数的二阶导数，依次可以定义高阶导数。记二阶导数为 $y'', f''(x), \frac{d^2y}{dx^2}$，$n$ 导数为 $y^{(n)}, f^{(n)}, \frac{d^ny}{dx^n}$。

高阶导数的求导法则为：

- $(au + bv)^{(n)} = av^{(n)} + bv^{(n)}$
- Leibniz 公式：$(uv)^{(n)} = \sum_{k=0}^{n} C_n^k u^{(n-k)v^{(k)}}$

有些函数无法用解析式表达，求导数并不方便。但是如果可以用公式表达变量之间的依赖关系，也可以通过解方程的形式求出导数，这样的函数为*隐函数*，即设有方程 $F(x, y) = 0$，如果存在定义在区间 I 上的函数 $f(x)$ 使得 $F(x, f(x))$ 恒等于 0，则 $y = f(x)$ 是由 $F(x, y) = 0$ 确定的隐函数。隐函数的求法是对方程 $F(x, y)$ 两边取导数，解出 y' 的表达式。

如果两个变量都可以用第三个变量表达，也可以确定这两个变量间的对应关系，即由参数方程 $x = u(t), y = v(t)$ 确定的函数。去掉中间变量以后 $y = v(u^{-1}(x))$，现在求导：$\frac{dy}{dx} =$

$\frac{dy/dt}{dx/dt} = \frac{u'(t)}{v'(t)}$，可以继续求高阶导数。

现在来到微分。设函数 $y = f(x)$ 在区间 I 上有定义，$x_0, x_0 + \Delta x \in I$，如果函数的增量 Δy 可以表示为 $\Delta y = A\Delta x + o(\Delta x)$，这里 A 是与 Δx 无关的常数，$o(\Delta)$ 是 $\Delta \to 0$ 时 Δx 的高阶无穷小。此时称 $y = f(x)$ 在点 x_0 处可微，$A\Delta x$ 为函数 $y = f(x)$ 在点 x_0 处的微分，记为 $dy|_{x=x_0}$。函数 $y = f(x)$ 在点 x_0 处可微的充要条件是它在 x_0 处可导，$dy|_{x=x_0} = f'(x_0)dx$，因而 $\frac{dy}{dx} = f'(x)$。

基本初等函数的微分、微分的四则运算法则和导数类似，这里不再列出，记住 $dy|_{x=x_0} = f'(x_0)dx$。需要注意的是一阶微分形式不变形：$dy = f'(u)du$。

函数与导数的联系可以用三个微分中值定理初步分析，这三个定理是 L'Hospital 法则的依据。

1. Rolle 中值定理：函数 $f(x)$ 在闭区间 $[a, b]$ 上连续，在开区间 (a, b) 上可导，且 $f(a) = f(b)$，则至少存在一点 $\varepsilon \in (a, b)$ 使得 $f'(\varepsilon) = 0$。
2. Lagrange 中值定理：函数 $f(x)$ 在闭区间 $[a, b]$ 上连续，在开区间 (a, b) 上可导，则至少存在一点 $\varepsilon \in (a, b)$ 使得 $f'(\varepsilon) = \frac{f(b)-f(a)}{b-a}$。
3. Cauchy 中值定理：函数 $f(x), g(x)$ 在闭区间 $[a, b]$ 上连续，在开区间 (a, b) 上可导，且 $g'(x) \neq 0$，则至少存在一点 $\varepsilon \in (a, b)$ 使得 $\frac{f'(\varepsilon)}{g'(\varepsilon)} = \frac{f(b)-f(a)}{g(b)-g(a)}$。

Lagrange 中值定理是 Cauchy 中值定理的特例，Rolle 中值定理是 Lagrange 中值定理的特例。特别的，如果 $\forall x \in (a, b), f'(x) = 0$，则在 (a, b) 内 $f(x) = c, c$ 为常数。如果 $\forall x \in (a, b), f'(x) = g'(x)$，则在 (a, b) 内 $f(x) - g(x) = c, c$ 为常数。

L'Hospital 法则用于求 $\frac{0}{0}, \frac{\infty}{\infty}$ 两种类型的公式值。设函数 $f(x), g(x)$ 在 x 的某个变化过程中有定义，且满足:(1)$\lim f(x) = \lim g(x) = 0$；（2）$f'(x), g'(x)$ 都存在且 $g'(x) \neq 0$；（3）$\lim \frac{f'(x)}{g'(x)} = A$ 或 ∞，则 $\lim \frac{f(x)}{g(x)} = \lim \frac{f'(x)}{g'(x)}$。将条件（1）中的 0 替换为 ∞ 即为 $\frac{\infty}{\infty}$ 型算式。

以下 5 种未定式都可以转换为 $\frac{0}{0}$ 或 $\frac{\infty}{\infty}$ 两种未定式。

1. $0 \times \infty, \infty - \infty$ 直接转换为 $\frac{0}{0}, \frac{\infty}{\infty}$
2. $1^\infty, 0^0, \infty^0$ 将 a^b 形式转化为 $e^{bln(a)}$ 形式，$bln(a)$ 为 $0 \times \infty$ 形式，再转化为 $\frac{0}{0}, \frac{\infty}{\infty}$

根据 Cauchy 定理，可以给出任何高阶可导函数的 Taylor 展开。设函数 $f(x)$ 在区间 (a, b) 内有 n 阶导数，且 $f^{(n)}(x)$ 在 (a, b) 内连续，$f(x)$ 在点 $x_0(x_0 \in (a, b))$ 关于 $(x - x_0)$ 的 n 次 Taylor 多项式为

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n \tag{4.2}$$

函数 $f(x)$ 与 Taylor 多项式的差函数为 $R_n(x) = f(x) - P_n(x)$。那么函数 $f(x)$ 在点 x_0 处的 n 阶 *Taylor 展开式*为

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + R_n(x) \tag{4.3}$$

余项 $R_n(x)$ 有两种形式：

- Peano 型余项：$R_n(x) = o((x - x_0)^n), x \to x_0$
- Lagrange 型余项：$R_n(x) = \frac{f^{(n+1)}(\varepsilon)}{(n+1)!}(x - x_0)^{n+1}, \varepsilon \in (min\{x, x_0\}, max\{x, x_0\})$

特别的，当 $x_0 = 0$，Taylor 公式又被称为 Maclaurin 公式。常用基本初等函数的 Taylor 展

开式为

- $e^x \approx 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n!}$
- $sin(x) \approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^{(m-1)}\frac{x^{2m-1}}{(2m-1!)}$
- $cos(x) \approx 1 - \frac{x^2}{x!} + \frac{x^4}{4!} - \cdots + (-1)^m\frac{x^{2m}}{(2m)!}$
- $ln(1+x) \approx x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + (-1)^{(n-1)}\frac{x^n}{n}$
- $(1+x)^\alpha \approx 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)\cdots(\alpha-n+1)}{n!}x^n$

函数的单调性和导数的关系非常清楚：对于函数 $f(x)$ 和定义域 I，$f'(x) > 0 \Leftrightarrow f(x)$ 在 I 上单调增，可类比单调减的等价条件。函数的极值的解法就是在众多的局部极值中取出最值。在点 x_0 的某个邻域中，如果 $\forall x \neq x_0$ 都有 $f(x) < f(x_0)$，则 $f(x_0)$ 为函数 $f(x)$ 的一个极大值，点 x_0 为函数的一个极大值点。相应的定义极小值。如果函数 $f(x)$ 在点 x_0 处的导数为 0，则 x_0 为函数 $f(x)$ 的极值点（反过来不成立，x_0 可以不连续但为极值点）。

导数为 0 的点为驻点，驻点和不可导的点是可能的极值点。确定函数极值的步骤为

1. 在定义域内求出所有的驻点和不可导的点
2. 确定 $f'(x)$ 在上述点两侧附近的符号
3. 如果 $f'_-(x_0) < 0, f'_+(x_0) > 0$，则点 x_0 为极小值点，如果 $f'_-(x_0) > 0, f'_+(x_0) < 0$，则点 x_0 为极大值点。如果两种情况都不成立，则点 x_0 不是极值点。最后比较极大值（极小值）点的函数值得到极值。（如果函数存在二阶导数，$f'(x_0) = 0, f''(x_0) > 0$ 时，点 x_0 是函数的极小值点，$f'(x_0) = 0, f''(x_0) < 0$ 时，点 x_0 是函数的极大值点）

一阶导数决定函数的单调性，二阶导数决定函数的凸性。函数 $f(x)$ 在区间 (a,b) 上连续，如果对任意的两点 $x_1, x_2 \in (a,b)$，都有

$$f(\frac{x_1+x_2}{2}) < \frac{1}{2}[f(x_1) + f(x_2)] \tag{4.4}$$

则函数 $f(x)$ 在 (a,b) 内是下凸函数，曲线 $y = f(x)$ 为下凸曲线；（$>$）反之则 $f(x)$ 为上凸函数，曲线 $y = f(x)$ 为上凸曲线。二阶导数和凸性的对应关系为：

- $f''(x) > 0 \Leftrightarrow$ 下凸函数
- $f''(x) < 0 \Leftrightarrow$ 上凸函数

连续曲线的上凸部分和下凸部分的分界点为拐点，必要条件为二阶导数为 0，如果二阶导不存在，则可以分析该点两侧的二阶导数的符号是否相反。

4.1.3 不定积分与定积分

不定积分是微分的逆运算。原函数是与导数密切相关的概念。设 $f(x)$ 在区间 I 上有定义，$\forall x \in I$，都有 $F'(x) = f(x)$ 或者 $dF(x) = f(x)dx$，则称 $F(x)$ 为 $f(x)$ 在区间 I 上的一个原函数。连续函数一定有原函数。如果 $f(x)$ 有原函数 $F(x)$，则 $G(x) = F(x) + C, C$ 为常数，也为 $f(x)$ 的原函数。

设函数 $F(x)$ 是 $f(x)$ 的一个原函数，C 为任意的常数，称 $F(x) + C$ 为 $f(x)$ 的不定积分，即 $\int f(x)dx = F(x) + C$，$\int$ 为积分号，$f(x)$ 为被积函数，x 为被积变量，C 为积分常数。不定积分的性质如下：

- 微分与积分互逆：$d\int f(x)dx = f(x), \int dF(x) = F(x) + C$
- 函数可加性：$\int[f(x) \pm g(x)]dx = \int f(x)dx \pm \int g(x)dx$
- 倍乘可分性：$\int kf(x)dx = k\int f(x)dx, k \neq 0$

基本的积分公式如下，这是进一步计算复杂积分的基础。

- $\int kdx = kx + C, k$ 为常数
- $\int x^{\mu}dx = \frac{1}{\mu+1}x^{\mu+1} + C, \mu \neq 1$
- $\int \frac{1}{x}dx = ln|x| + C$
- $\int e^x dx = e^x + C$
- $\int a^x dx = \frac{a^x}{lna} + C$
- $\int cos(x)dx = sin(x) + C$
- $\int sin(x)dx = -cos(x) + C$
- $\int \frac{1}{cos^2(x)}dx = \int sec^2(x)dx = tan(x) + C$
- $\int \frac{1}{sin^2(x)}dx = \int sec^2(x)dx = -cot(x) + C$
- $\int sec(x)tan(x)dx = sec(x) + C$
- $\int csc(x)cot(x)dx = -csc(x) + C$
- $\int \frac{1}{1+x^2}dx = arctan(x) + C$
- $\int \frac{1}{\sqrt{1-x^2}}dx = arcsin(x) + C$

不定积分的基本方法是换元积分法和分部积分法，前者的依据是一阶微分形式不变形，后者的依据是导数的运算法则。换元积分法如下：

- 凑微分法：设 $f(u)$ 有原函数，且 $u(x) = \phi(x)$ 具有连续的导数，则 $f[\phi(x)]\phi'(x)$ 有原函数，且 $\int f[\phi(x)]\phi'(x)dx = [\int f(u)du]_{u=\phi(x)}$
- 复合函数法：设 $f(x)$ 连续，$x = \phi(t)$ 具有连续的导数，且 $\phi'(t) \neq 0$，则 $t = \phi^{-1}(x)$ 为 $\phi(x)$ 的反函数。则 $\int f(x)dx = [\int f[\phi(t)]\phi'(x)dt]_{t=\phi^{-1}(x)}$

基于换元积分法得出的基本积分公式为

- $\int tan(x)dx = -ln|cos(x)| + C$
- $\int cot(x)dx = ln|sin(x)| + C$
- $\int sec(x)dx = ln|sec(x) + tan(x)| + C$
- $\int csc(x)dx = ln|csc(x) - cot(x)| + C$
- $\int \frac{1}{x^2+a^2}dx = \frac{1}{a}arctan\frac{x}{a} + C$
- $\int \frac{1}{x^2-a^2}dx = \frac{1}{2a}ln|\frac{x-a}{x+a}| + C$
- $\int \frac{1}{\sqrt{a^2-x^2}}dx = arcsin\frac{x}{a} + C$
- $\int \frac{1}{\sqrt{x^2\pm a^2}}dx = ln|\sqrt{x^2 \pm a^2} + x| + C$

分部积分法：$udv = uv - vdu$。

有理函数的积分是将有理函数化成最简分式的和。三角函数的积分是采用万能变换 $tan(\frac{x}{2}) = u$ 化成有理函数，即 $sin(x) = \frac{2u}{1+u^2}, cos(x) = \frac{1-u^2}{1+u^2}, dx = \frac{2}{1+u^2}du$。因此三角函数和有理函数的积分总是可以计算。

不定积分难以在经济学中直接应用，而定积分是经济分析的有力工具，二者的区别在于

积分限。

设函数 $f(x)$ 在区间 $[a,b]$ 上有界，在 $[a,b]$ 内任意插入 $n-1$ 和分点，$a=x_0<x_1<\cdots<x_{n-1}<x_n=b$，将 $[a,b]$ 分成 n 个小区间 $[x_{i-1},x_i](i=1,2,\cdots,n)$。记 $\Delta x_i=x_i-x_{i-1}$，在第 i 个小区间上任取一点 $\varepsilon_i\in[x_{i-1},x_i]$，用点 ε_i 的函数值 $f(\varepsilon_i)$ 乘上小区间的长度 Δx_i 即 $f(\varepsilon_i)\Delta x_i, i=1,2,\cdots,n$，并加和 $\sum_{i=1}^{n}f(\varepsilon_i)\Delta x_i$。记 $\lambda=max\{\Delta x_1,\Delta x_2,\cdots,\Delta x_n\}$，如果不论对 $[a,b]$ 怎样分割，也不论 ε_i 怎样取，当 $\lambda\to 0$ 时，式子 $\sum_{i=1}^{n}f(\varepsilon_i)\Delta x_i$ 总是趋向于同一极限，则称该极限为函数 $f(x)$ 在区间 $[a,b]$ 上的定积分，记为 $\int_a^b f(x)dx$，即

$$\int_a^b f(x)dx=\lim_{\lambda\to 0}\sum_{i=1}^{n}f(\varepsilon_i)\Delta x_i \tag{4.5}$$

其中 $\int$ 为积分号，$f(x)$ 为被积函数，$f(x)dx$ 为被积表达式，x 为积分变量，a 为积分下限，b 为积分上限，$[a,b]$ 为积分区间，$sum_{i=1}^{n}f(\varepsilon_i)\Delta x_i$ 为积分和，当积分和的极限存在是，称 $f(x)$ 在 $[a,b]$ 上可积。

规定，如果 $a=b, \int_a^b f(x)dx=0$，如果 $a>b, \int_a^b f(x)dx=-\int_b^a f(x)dx$。

$f(x)$ 在区间 $[a,b]$ 上积分存在的充分条件有（其一即可）：

- 连续
- 有界，只有有限个第一类间断点
- 单调有界

定积分的性质如下：

- 可分离性：$\int_a^b kf(x)dx=k\int_a^b f(x)dx, k$ 为常数
- 函数可加性：$\int_a^b[f(x)\pm g(x)]dx=\int_a^b f(x)dx\pm\int_a^b f(x)dx$
- 区间可加性：$\int_a^b f(x)dx=\int_a^c f(x)dx+\int_c^a f(x)dx$
- 常数积分：$f(x)=C, x\in[a,b]$，则 $\int_a^b f(x)dx=(b-a)C$
- 保号性：$\forall x\in[a,b], f(x)\geq 0$，则 $\int_a^b f(x)dx\geq 0\Rightarrow|\int_a^b f(x)dx|\leq\int_a^b|f(x)|dx$
- 估值性：$M=max\{f(x),x\in[a,b]\}, m=min\{f(x),x\in[a,b]\}$，则 $m(b-a)\leq\int_a^b f(x)dx\leq M(b-a)$
- 定积分中值定理：$f(x)$ 在区间 $[a,b]$ 上连续，则 $\exists\varepsilon\in[a,b]$，使得 $\int_a^b f(x)dx=f(\varepsilon)(b-a)$

具体微分的计算依赖微积分基本定理，首先了解积分上限函数的含义。如果函数 $f(x)$ 在区间 $[a,b]$ 上连续，则积分上限的函数 $\Phi(x)=\int_a^x f(t)dt$ 在 $[a,b]$ 上可导，且 $\Phi'(x)=\frac{d}{dx}\int_a^x f(t)dt=f(x), a\leq x\leq b$。如果函数 $f(x)$ 在区间 $[a,b]$ 上连续，则积分上限函数就是它的一个原函数。

微积分基本定理如下：设函数 $f(x)$ 在区间 $[a,b]$ 上连续，函数 $F(x)$ 是 $f(x)$ 在 $[a,b]$ 上的一个原函数，则 $\int_a^b f(x)dx=F(b)-F(a)$（Newton-Leibniz 公式）。

根据不定积分的换元积分法和分部积分法，可以得出定积分的换元积分法和分部积分法。

- 换元积分法：函数 $f(x)$ 在 $[a,b]$ 上连续，函数 $x=\phi(t)$ 满足（1）在 $[\alpha,\beta]$ 上单调，$\phi(t)$ 值域为 $[a,b]$，且 $\phi(\alpha)=a,\phi(\beta)=b$（或反过来）;（2）$[\alpha,\beta]$ 上有连续导数。则 $\int_a^b f(x)dx=\int_\alpha^\beta f[\phi(t)]\phi'(t)dt$
- 分部积分法：函数 $u(x),v(x)$ 在 $[a,b]$ 上可导，且导数 $u'(x),v'(x)$ 连续，则 $\int_a^b udv=[uv]_a^b\int_a^b cdu$

前面讨论的是有限区间、有界函数的积分，但积分区间可能无限，被积函数也可能有无穷间断点，这两种情况为广义积分，广义积分的计算只需在定积分的基础上增加一个极限条件，这里不再介绍。一个重要的广义积分是 Γ 函数：$\Gamma(s)=\int_0^{+\infty}x^{s-1}e^{-x}dx$。其性质为

- 递推公式：$\Gamma(s+1)=s\Gamma(s), s>0$
- 极限 0 不存在：$s\to 0^+, \Gamma(s)\to+\infty$
- 余元公式：$\Gamma(s)\Gamma(1-s)=\frac{\pi}{sin(\pi s)}, 0<s<1$
- 常用公式：$\int_0^{+\infty}e^{-u^2}u^t du=\frac{1}{2}\Gamma(\frac{1+t}{2})$，如 $\int_0^{+\infty}e^{-u^2}du=\frac{\sqrt{\pi}}{2}$

平面图形的面积和立体的体积可以用元素法转化为定积分求解。

4.1.4 向量代数与空间解析几何

空间解析几何将几何问题转化为代数问题，也是多元函数微积分的知识基础。描述空间几何体的基本概念是坐标系、坐标和向量。

在空间中取一点 O，以 O 为原点作三条长度单位相同、两两垂直的数轴，依次记为 x 轴、y 轴和 z 轴，称为坐标轴。最常用的坐标系为 Oxyz 直角坐标系，坐标轴的方向符合右手规则，点 O 称为坐标系的原点。

由两条坐标轴可以确定一个平面，分别是 Oxy 平面、Oxz 平面和 Oyz 平面。对空间中的任意一点 M，都可以过该点做三个平面分别垂直于 x 轴、y 轴和 z 轴，交点为对应的坐标，即 $M(x,y,z)$，x,y,z 分别为点 M 的横坐标、纵坐标和竖坐标。

在空间中两点的距离公式为：

$$d(M_1,M_2)=|M_1M_2|=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2} \tag{4.6}$$

特别的，点 $M(x,y,z)$ 到原点的距离为 $\sqrt{x^2+y^2+z^2}$。

向量是既有大小又有方向的量，用黑体字母 $\boldsymbol{a}$、加箭头的字母 $\vec{a}$ 或者有向线段 $\overrightarrow{M_1M_2}$。向量的大小叫做向量的模，记为 $|\vec{a}|$，模为 0 的向量为零向量，记为 $\vec{0}$，它的方向是任意的。模为 1 的向量为单位向量。

向量相等 $\Leftrightarrow$ 模相等和方向相同，记为 $\vec{a}=\vec{b}$。这里不要求起点一致，因为讨论的是自由向量。向量平行 $\Leftrightarrow$ 方向相同或相反，记为 $\vec{a}//\vec{b}$，零向量与任意向量平行。在直角坐标系中，向量 $\overrightarrow{OM}$ 称为点 M 关于点 O 的向径，记为 $\vec{r}=\overrightarrow{OM}$，点和向径一一对应。

向量的线性运算法则如下：

- 加法：（平行四边形法则）$\overrightarrow{M_1M_2}+\overrightarrow{M_2M_3}=\overrightarrow{M_1M_3}$。适用交换律和结合律。
- 数乘：大小，$|\lambda\vec{a}|=|\lambda||\vec{a}|$，方向，$\lambda>0$，与 $\vec{a}$ 同向，$\lambda<0$，与 $\vec{a}$ 反向，$\lambda=0$，$\lambda\vec{a}=\vec{0}$。适用分配律和结合律。

特别的，$\vec{a}-\vec{b}=\vec{a}+(-1\vec{b})$, $-\vec{b}$ 为负向量。任意的非零向量都可以单位化：$\vec{a}=|a|\vec{e_a}, \vec{e_a}=\frac{\vec{a}}{|\vec{a}|}$。$\forall\vec{a},\vec{b},\vec{a}\neq\vec{0},\vec{a}//\vec{b}\Leftrightarrow\exists\lambda,\vec{b}=\lambda\vec{a}$，即数乘等价于平行。

空间中的向量用基本单位向量（x,y,z 轴方向的单位向量）表示：

$$\vec{a}=(x_2-x_1)\vec{i}+(y_2-y_1)\vec{j}+(z_2-z_1)\vec{k}=a_x\vec{i}+a_y\vec{j}+a_z\vec{k} \tag{4.7}$$

有序数组 (a_x,a_y,a_z) 为向量 $\vec{a}$ 的坐标，记为 $\vec{a}=(a_x,a_y,a_z)$。特别的，向径 $\overrightarrow{OM}=(x,y,z)$。

空间中向量 $\vec{a}$ 的模为 $|\vec{a}| = \sqrt{a_x^2 + a_y^2 + a_z^2}$，方向用向量 $\vec{a}$ 与 x,y,z 轴的方向角表示，即 $cos(\alpha) = \frac{a_x}{\vec{a}}, cos(\beta) = \frac{a_y}{\vec{a}}, cos(\gamma) = \frac{a_z}{\vec{a}}$。方向余弦满足 $cos(\alpha)^2 + cos(\beta)^2 + cos(\beta)^2 = 1$。与 $\vec{a}$ 同方向的单位向量的坐标表达式为 $\vec{e_a} = \frac{\vec{a}}{|\vec{a}|} = \frac{(a_x,a_y,a_z)}{|\vec{a}|} = (cos(\alpha), cos(\beta), cos(\gamma))$。

向量的线性运算和平行用坐标表示为

- $\vec{a} \pm \vec{b} = (a_x \pm b_x, a_y \pm b_y, a_z \pm b_z)$
- $\lambda\vec{a} = (\lambda a_x, \lambda a_y, \lambda a_z)$
- $\vec{a}//\vec{b}, \vec{a} \neq \vec{0} \Leftrightarrow b_x = \lambda a_x, b_y = \lambda a_y, b_z = \lambda a_z$

设向量 $\vec{a}, \vec{b}$ 的夹角为 $\theta = \widehat{(\vec{a}, \vec{b})}$，二者的数量积（内积、点积）为 $\vec{a}\cdot\vec{b}$，值为 $\vec{a}\cdot\vec{b} = |\vec{a}||\vec{b}|cos(\theta)$。$|\vec{b}|cos(\theta)$ 为向量 $\vec{b}$ 在向量 $\vec{a}$ 上的投影，记为 $Prj_{\vec{a}}\vec{b} = |\vec{b}|cos(\theta)$。

内积满足 $\vec{a} \cdot \vec{b} = |\vec{a}|^2$，$\vec{a} \perp \vec{b} \Leftrightarrow \vec{a} \cdot \vec{b} = \vec{0}$。

内积的运算律如下

- 交换律：$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$
- 分配律：$\vec{a} \cdot (\vec{b} \pm \vec{c}) = \vec{a} \cdot \vec{b} \pm \vec{a} \cdot \vec{c}$
- 结合律：$(\lambda\vec{a}) \cdot \vec{b} = \vec{a}(\lambda\vec{b}) = \lambda(\vec{a} \cdot \vec{b}), \lambda$ 为实数

数量积的坐标表达式为 $\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y + a_z b_z$。$\vec{a} \perp \vec{b} \Leftrightarrow a_x b_x + a_y b_y + a_z b_z = 0$。夹角公式 $\cos(\theta) = \frac{\vec{a}\cdot\vec{b}}{|\vec{a}||\vec{b}|} = \frac{a_x b_x + a_y b_y + a_z b_z}{\sqrt{a_x^2+a_y^2+a_z^2}\sqrt{b_x^2+b_y^2+b_z^2}}$。

向量的向量积（叉积、外积）是一个向量，模为 $|\vec{a} \times \vec{b}| = |\vec{a}||\vec{b}|sin(\theta), \theta = \widehat{(\vec{a}, \vec{b})}$，方向同时垂直于 $\vec{a}, \vec{b}$，符合右手规则。向量积的性质如下

- $\vec{a} \times \vec{a} = \vec{a}$
- $\vec{a}//\vec{b} \Leftrightarrow \vec{a}\vec{b} = \vec{0}$

向量积的运算规律如下

- 反交换律：$\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$
- 分配律：$\vec{a} \times (\vec{b} + (\vec{c})) = \vec{a} \times \vec{b} + \vec{a} \times \vec{b}$
- 结合律：$(\lambda\vec{a}) \times \vec{b} = \vec{a} \times (\lambda\vec{b}) = \lambda(\vec{a} \times \vec{b}), \lambda$ 为实数

向量积的坐标表达式需要借助矩阵代数来表达

$$\vec{a} \times \vec{b} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_x & a_y & a_z \\ b_x & b_y & b_z \end{vmatrix} \tag{4.8}$$

根据坐标可以确定点的位置，根据坐标的方程可以确定曲面，必须一一对应。曲面的方程和方程的图形：曲线上的点$M(x, y, z)$ 和 $F(x, y, z) = 0$ 的解一一对应。

平面的方程有如下表示方法：

- 点法式方程：$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$，法向量 $\vec{n} = (A, B, C)$
- 三点式方程：$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0$
- 一般方程：$Ax + By + Cz = 0$

- 截距式方程：$\frac{x}{a}+\frac{y}{b}+\frac{z}{c}=1$

两个平面的夹角用法向量来表示：

$$cos(\theta)=\frac{|\vec{n_1}\cdot\vec{n_2}|}{|\vec{n_1}||\vec{n_2}|}=\frac{A_1A_2+B_1B_2C_1C_2}{\sqrt{A_1^2+B_1^2+C_1^2}\sqrt{A_2^2+B_2^2+C_2^2}} \tag{4.9}$$

因此，$\pi_1\perp\pi_2\Leftrightarrow A_1A_2+B_1B_2+C_1C_2=0,\pi_1//\pi_2\Leftrightarrow\frac{A_1}{A_2}=\frac{B_1}{B_2}=\frac{C_1}{C_2}$，当 $\frac{A_1}{A_2}=\frac{B_1}{B_2}=\frac{C_1}{C_2}=\frac{D_1}{D_2}$ 时，两平面重合。

点到直线的距离公式

$$d=\frac{|Ax_0+By_0+Cz_0+D|}{\sqrt{A^2+B^2+C^2}} \tag{4.10}$$

直线的方程为

- 对称式方程：$\frac{x-x_0}{m}=\frac{y-y_0}{n}=\frac{z-z_0}{p}$，方向向量 $\vec{s}=(m,n,p)$
- 参数方程：$\begin{cases}x=x_0+mt,\\ y=y_0+nt,\\ z=z_0+pt\end{cases}$
- 一般方程：$\begin{cases}A_1x+B_1y+C_1z+D_1=0,\\ A_2x+B_2y+C_2z+D_2=0\end{cases}$

直线的夹角余弦：$cos(\theta)=\frac{|\vec{s_1}\cdot\vec{s_2}|}{|\vec{s_1}||\vec{s_2}|}=\frac{|m_1m_2+n_1n_2+p_1p_2|}{\sqrt{m_1^2+n_1^2+p_1^2}\sqrt{m_2^2+n_2^2+p_2^2}}$。因而，直线平行或垂直有代数表示：$l_1\perp l_2\Leftrightarrow m_1m_2+n_1n_2+p_1p_2=0,l_1//l_2\Leftrightarrow\frac{m_1}{m_2}=\frac{n_1}{n_2}=\frac{p_1}{p_2}$。

直线和平面的夹角正弦为 $sin(\phi)=\frac{|\vec{n}\cdot\vec{s}|}{|\vec{n}||\vec{s}|}=\frac{Am+Bn_Cz}{\sqrt{A^2+B^2+C^2}\sqrt{m^2+n^2+p^2}}$。

同样，$l//\pi\Leftrightarrow Am+Bn+Cz=0,l\perp\pi\Leftrightarrow\frac{A}{m}=\frac{B}{n}=\frac{C}{p}$。

柱面和旋转面是立体几何中规则的几何体的要素。平行于定直线的直线沿定曲线移动所形成的曲面为*柱面*，其方程是只含有两个变量的方程，如 $F(x,y)=0$。平面上的曲线绕该平面上的一条定直线旋转所形成的曲面为*旋转曲面*，该曲线为旋转曲面的母线，该直线为旋转曲面的轴，其方程也只有两个变量，如 $F(\pm r,z)=0,r=\sqrt{x^2+y^2}$。常用曲面如下：

- 旋转椭球面：$\frac{x^2}{a^2}+\frac{y^2+z^2}{b^2}=1$
- 旋转抛物面：$x^2+y^2=a^2z$
- 圆锥面：$x^2+y^2=a^2z^2$
- 二次曲面
  - 椭球面：$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1,a,b,c>0$
  - 椭圆锥面：$z^2=\frac{x^2}{a^2}+\frac{y^2}{b^2},a,b>0$
  - 双曲面：$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1,\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1,a,b,c>0$
  - 抛物面：$\frac{x^2}{a^2}+\frac{y^2}{b^2}=z,\frac{x^2}{a^2}-\frac{y^2}{b^2}=z,a,b>0$

曲线的一般方程为

$$\begin{cases}F(x,y,z)=0\\ G(x,y,z)=0\end{cases} \tag{4.11}$$

曲线的参数方程为

$$\begin{cases} x = x(t) \\ y = y(t) \\ z = z(t) \end{cases} \tag{4.12}$$

曲线在坐标面上的投影：

$$\begin{cases} F(x,y,z) = 0 \\ G(x,y,z) = 0 \end{cases} \Rightarrow H(x,y) = 0 \text{投影柱面）}$$

令 $z = 0$ 得到在 Oxy 面上的投影曲线。

4.1.5 多元函数微分与重积分

多元函数的微分和积分是多元优化中最常用的工具，多元函数微积分是一元函数微积分的拓展，二者的概念相通。

由二元有序实数组 (x,y) 的全体所构成的集合称为二维空间，记为 R^2，即 $R^2 = \{(x,y)|x,y \in R\}$。$R^2$ 中的点集为平面点集。设 $P_0(x_0,y_0) \in R^2, \delta$ 为一正数，称 R^2 中与点 P_0 的距离小于 δ 的点的集合为点 P_0 的 δ 领域，记为 $U(P_0,\delta) = \{(x,y)|\sqrt{(x-x_0)^2 + (y-y_0)^2} < \delta\}$。$U(P_0,\delta)$ 中去掉点 P_0 的部分为去心 δ 领域。

点 P 与点集 E 的关系有三种：内点、外点和边界点。如果 E 中所有点均为内点，则 E 为开集。如果 E 为开集，若 $\forall P_1, P_2 \in E$，在 E 内都存在一条连接 P_1, P_2 的折线，则 E 为区域（连通的开集）。区域与区域的边界的集合为闭区域。

设 D 为 R^n 的非空子集，从 D 到实数 R 的映射 f 称为定义在 D 上的 n 元函数，记为 $f : D \in R^n \to R$ 或 $y = f(x_1, \cdots, x_n)$，其中 $x_1, \cdots, x_n$ 为自变量，y 为因变量，D 为函数 f 的定义域，集合 $f(D)$ 为函数的值域。

设函数 $f(x,y)$ 在点 $P\_(x_0,y_0)$ 的某去心领域 $\dot{U}(P_0)$ 内有定义，A 为常数，如果当 $P(x,y)$ 无限趋于 $P_0(x_0,y_0)$ 时，$f(x,y)$ 无限趋于数 A，则称 A 是 $f(x,y)$ 当 $P(x,y) \to P_0(x_0,y_0)$ 时的极限，记为 $\lim\limits_{P \to P_0} f(P) = A$ 或 $f(P) \to A(P \to P_0)$。二元函数的极限称为二重极限。

设函数 $z = f(x,y)$ 在点 $P_0(x_0,y_0)$ 的某个领域 $U(P_0)$ 内有定义，如果 $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y) = f(x_0,y_0)$，则称函数 $f(x,y)$ 在点 P_0 处连续，否则称 $f(x,y)$ 在点 P_0 处间断或不连续。如果函数 $f(x,y)$ 在区域 D 上的每一点都连续，则称函数 $f(x,y)$ 在区域 D 上连续。

多元函数的极限、连续性等性质与一元函数一致，如有界性定理、最大最小值定理、介值定理等。

设函数 $z = f(x,y)$ 在点 P_0 的某个领域 $U(P_0)$ 内有定义，令 $y = y_0$，自变量 x 自 x_0 取得增量 Δx，相应的函数的增量为 $\Delta_x z = f(x_0 + \Delta x, y_0) - f(x_0, y_0)$，如果极限 $\lim\limits_{\Delta x \to 0} \frac{\Delta_x z}{\Delta x} = \lim\limits_{\Delta x \to 0} \frac{f(x_0+\Delta x,y_0) - f(x_0,y_0)}{\Delta x}$ 存在，则称此极限为函数 $z = f(x,y)$ 在点 P_0 处关于自变量 x 的偏导数，记为 $\frac{\partial f}{\partial x}|_{(x_0,y_0)}$ 或 $f'_x(x_0,y_0)$。

当函数 $z = f(x,y)$ 在点 P_0 处关于 x 和 y 的偏导数都存在时，称 $f(x,y)$ 在点 P_0 处可偏

导，相应的可以定义偏导数。由于偏导数只有两个方向，因此可偏导和连续无任何关系。如果偏导数仍可偏导，则可定义高阶偏导数。特别的，如果函数 $f(x,y)$ 的两个二阶混合偏导数 $f''_{xy}(x,y), f''_{yx}$ 都在 (x,y) 处连续，则 $f''_{xy}(x,y) = f''_{yx}(x,y)$。

设函数 $z = f(x,y)$ 在点 $P_0(x_0,y_0)$ 的某领域 $U(P_0)$ 内有定义，如果函数在点 P_0 处取得全增量 $\Delta x = f(x_0+\Delta x, y_0+\Delta y) - f(x_0,y_0)$ 可以表示为 $\Delta z = A\Delta x + B\Delta y + o(\rho)$，其中 A, B 是不依赖于 $\Delta x, \Delta y$ 的常数，$\rho = \sqrt{\Delta x^2 + \Delta y^2}$。则称函数 $z = f(x,y)$ 在点 (x_0,y_0) 处可微，称 $A\Delta x + \Delta y$ 为 $z = f(x,y)$ 在点 (x_0,y_0) 的全微分，记为 $dz = A\Delta x + B\Delta y$。

当函数 $z = f(x,y)$ 在区域 D 内的每一点都可微时，称 $z = f(x,y)$ 为 D 内的可微函数。

函数 $z = f(x,y)$ 在点 (x_0,y_0) 处可微 $\Rightarrow f(x,y)$ 在点 (x_0,y_0) 处连续，可偏导，且 $A = f'_x(x_0,y_0), B = f'_y(x_0,y_0)$。全微分公式：$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$。

函数 $z = f(x,y)$ 在点 (x_0,y_0) 的某领域 $U(P_0)$ 内可偏导，若偏导数 $f'_x(x,y), f'_y(x,y)$ 都在点 (x_0,y_0) 处连续，则 $f(x,y)$ 在点 (x_0,y_0) 处可微。

链锁规则：设函数 $z = f(x,y)$ 及 $x = \phi(t), y = \varphi(t)$ 可以构成复合函数 $z = f[\phi(t), \varphi(t)]$，$x = \phi(t), y = \varphi(t)$ 都在点 t 处可导，$z = f(x,y)$ 在对应点 (x,y) 处可微，则符合函数 $z = f[\phi(t), \varphi(t)]$ 在点 t 处可导，且

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} \tag{4.13}$$

链锁规则对其他形式的复合函数也适用。

一阶全微分形式不变性：若函数 $z = f(u,v)$ 在点 (u,v) 处可微，而 $u = u(x,y), v = v(x,y)$ 在对应的点 (x,y) 处也可微，则无论 u, v 作为 $z = f(u,v)$ 的自变量，还是作为复合函数 $z = f[u(x,y), v(x,y)]$ 的中间变量，都有

$$dz = \frac{\partial z}{\partial u}du + \frac{\partial z}{\partial v}dv \tag{4.14}$$

微分法则：

- $d(u \pm v) = du \pm dv$
- $d(uv) = vdu + udv$
- $d(\frac{u}{v}) = \frac{vdu - udv}{v^2}(v \neq 0)$

隐函数也可以求微分。隐函数存在定理：设函数 $F(x,y)$ 在包含点 $P_0(x_0,y_0)$ 的某区域 D 内具有连续偏导数，且 $F(x_0,y_0) = 0, F'_y(x_0,y_0) \neq 0$，则存在唯一的定义在点 x_0 的某领域 $U(x_0)$ 内的函数 $y = f(x)$，满足 $f(x_0) = y_0$ 及恒等式 $F(x, f(x)) = 0$，在 $U(x_0)$ 内具有连续导数，且

$$\frac{dy}{dx} = -\frac{F'_x}{F'_y} \tag{4.15}$$

推导法：对 $F(x,y) = 0$ 两边同时求 x 的偏导，得 $F'_x + F'_y\frac{dy}{dx} = 0 \to \frac{dy}{dx} = -\frac{F'_x}{F'_y}$。

微分法：对 $F(x,y)$ 取微分，得 $F'_x dx + F'_y dy = 0 \to \frac{dy}{dx} = -\frac{F'_x}{F'_y}$

运用多元函数的偏导数可以求多元含的极值。设二元函数 $f(x,y)$ 在点 $P_0(x_0,y_0)$ 的某个领域 $U(P_0)$ 内有定义，如果对任何 $(x,y) \in \dot{U}(P_0)$，都有 $f(x,y) < (>) f(x_0,y_0)$，则称 $f(x_0,y_0)$ 是 $f(x,y)$ 的一个极大值（极小值），点 (x_0,y_0) 称为 $f(x,y)$ 的极大值点（极小值点）。极大值

和极小值统称为极值，极大值点和极小值点统称为极值点。

函数取极值的必要条件：设函数 $z = f(x, y)$ 在点 (x_0, y_0) 处可偏导，且在点 (x_0, y_0) 处取得极值，则 $f'_x(x_0, y_0) = 0, f'_y(x_0, y_0) = 0$。

使的 $f'_x(x_0, y_0) = 0, f'_y(x_0, y_0) = 0$ 同时成立的点为驻点，驻点和不可偏导的点统称为可能极值点。

函数取极值的充分条件：设函数 $z = f(x, y)$ 在点 (x_0, y_0) 的某个领域内具有连续二阶偏导数，且 $f'_x(x_0, y_0) = f'_y(x_0, y_0) = 0$，令 $f''_{xx}(x_0, y_0) = A, f''_{xy}(x_0, y_0) = B, f''_{yy}(x_0, y_0) = C$，则

1. $B^2 - AC < 0$，$f(x, y)$ 在点 (x_0, y_0) 处取极值，$A < 0$ 时取极大值，$A > 0$ 时取极小值
2. $B^2 - AC > 0$，$f(x, y)$ 在点 (x_0, y_0) 处无极值
3. $B^2 - AC = 0$，无法判断

依据以上两个定理，具有连续二阶偏导数的函数的极值求法如下：

1. 解方程 $\begin{cases} f'_x(x, y) = 0 \\ f'_y(x, y) = 0 \end{cases}$ 求出驻点
2. 对每一个驻点求出 A, B, C
3. 用 $B^2 - AC$ 的符号定是否为极值，然后用 A 的符合定极大值还是极小值

求多元函数的最值，只需先求出极值和不可偏导点的函数值，然后一一比较，特别注意边界点是不可偏导的，需要分开求。

在约束条件下求多元函数极值的通用方法是 Lagrange 乘数法，即

$$z = f(x, y) \tag{4.16}$$

$$s.t. \phi(x, y) = 0 \tag{4.17}$$

解法如下：

1. 引入 Lagrange 函数：$L(x, y, \lambda) = f(x, y) - \lambda\phi(x, y)$
2. 解驻点方程：$\begin{cases} f'_x(x, y) + \lambda\phi'_x(x, y) = 0 \\ f'_y(x, y) + \lambda\phi'_y(x, y) = 0 \\ \phi(x, y) = 0 \end{cases}$ 得到驻点
3. 根据驻点和极值的充分条件进一步求解

接下来介绍重积分。设函数 $f(x, y)$ 是有界闭区域 D 上的有界函数，将闭区域 D 划分为任意的 n 和小闭区域 $\Delta\sigma_1, \cdots, \Delta\sigma_n$，其中 $\Delta\sigma_k$ 是第 k 个小闭区域的面积。在每个 $\Delta\sigma_k$ 上任意取一点 (ε_k, η_k)，作和数 $\sum_{k=1}^{n} f(\varepsilon_k, \eta_k)\Delta\sigma_k$。如果当各小闭区域的直径的最大值 λ 趋于 0 时，和数的极限存在且与对区域的分法和点的取法无关，则称函数 $f(x, y)$ 在闭区域 D 上可积，此极限为函数 $f(x, y)$ 在闭区域 D 上的二重积分，记为

$$\iint_D f(x, y)d\sigma = \iint_D f(x, y)dxdy = \lim_{\lambda \to 0} \sum_{k=1}^{n} f(\varepsilon_k, \eta_k)\Delta\sigma_k \tag{4.18}$$

二重积分是由曲面 $f(x, y)$ 和闭区域 D 构成的曲顶柱体的体积。

可积的充分条件：函数 $f(x, y)$ 在有界闭区域 D 上连续 $\Rightarrow f(x, y)$ 在 D 上可积。

二重积分的性质与定积分性质类似：

- $f(x,y)\equiv 1 \Rightarrow \iint\limits_D d\sigma=\sigma(D),\sigma(D)$ 为 D 的面积
- 线性性：$\iint\limits_D(\alpha f(x,y)+\beta g(x,y))d\sigma=\alpha\iint\limits_D f(x,y)d\sigma+\beta\iint\limits_D g(x,y)d\sigma$
- 区域可加性：$\iint\limits_D f(x,y)d\sigma=\iint\limits_{D_1} f(x,y)d\sigma+\iint\limits_{D_2} f(x,y)d\sigma, D_1\cup D_2=D, D_1\cap D_2=\emptyset$
- 保号性：$f(x,y)\le g(x,y),(x,y)\in D\Rightarrow \iint\limits_D f(x,y)d\sigma\le\iint\limits_D g(x,y)d\sigma$
- 绝对值不等式：$|\iint\limits_D f(x,y)d\sigma|\le\iint\limits_D |f(x,y)|d\sigma$
- 有界性：$min\{f(x,y)\}\sigma(D)\le\iint\limits_D f(x,y)d\sigma\le max\{f(x,y)\}$
- 二重积分中值定理：$f(x,y)$ 在有界闭区域 D 上连续，则 $\exists(\varepsilon,\eta)\in D$ 使 $\iint\limits_D f(x,y)d\sigma=f(\varepsilon,\eta)\sigma(D)$

二重积分可在直角坐标系和极坐标系中计算。在直角坐标系中，需要将区域 D 划分为 x 型域和 y 型域运用区间可加性计算。x 型域是形如 $D=\{(x,y)|\phi_1(x)\le y\le\phi_2(x),a\le x\le b\}$ 的区域，同理，y 型域是形如 $D=\{(x,y)|\varphi_1(y)\le x\le varphi_2(y),c\le y\le d\}$ 的区域。x 型域和 y 型域将闭区域 D 划分为规则的小闭区域。运用元素法可得 x 型域的二重积分为

$$\iint\limits_D f(x,y)d\sigma=\int_a^b[\iint_{\phi_1(x)}^{\phi_2(x)} f(x,y)dy]dx=\int_a^b dx\int_{\phi_1(x)}^{\phi_2(x)} f(x,y)dy \tag{4.19}$$

同样，y 型域的二重积分为

$$\iint\limits_D f(x,y)d\sigma=\int_c^d dy\int_{\varphi_1(y)}^{\varphi_2(y)} f(x,y)dx \tag{4.20}$$

直角坐标系和极坐标系的对应关系：

- $\begin{cases} x=rcos(\theta)\\ y=rsin(\theta)\end{cases}$
- $d\sigma=rd\theta dr=rdrd\theta$
- $\iint\limits_D f(x,y)d\sigma=\iint\limits_D f(rsin(\theta),rcos(\theta))rdrd\theta$

极坐标系下的二重积分：

$$\iint\limits_D f(rcos(\theta),rsin(\theta))rdrd\theta=\int_\alpha^\beta d\theta\int_{r_1(\theta)}^{r_2(\theta)} f(rcos(\theta),rsin(\theta))rdr \tag{4.21}$$

二重积分也有反常积分，这里不再讨论。二重积分可以拓展到三重积分，甚至更高阶的积分。运用元素法求三种积分，定义域的元素为体积元素：

$$\iiint\limits_\Omega f(x,y,z)dV=\iiint\limits_\Omega f(x,y,z)dxdydz=\lim_{\lambda\to 0}\sum_{k=1}^n f(\varepsilon_k,\eta_k,\varsigma_k)\Delta V_k \tag{4.22}$$

直角坐标系中计算三重积分有两种方法：

- 先一后二法：$\iiint\limits_\Omega f(x,y,z)dxdydz=\iint\limits_{D_{xy}} dxdy\int_{z_1(x,y)}^{z_2(x,y)} f(x,y,z)dz$（先对 z 积分，后对 z,y 积分，可换序）
- 先二后一法：$\iiint \liminf_\Omega f(x,y,z)dxdydz=\int_p^q dz\iint\limits_{D_z} f(x,y,z)dxdy$（先对 xy 积分，后对 z 积分，可换序）

在柱坐标系中计算三重积分：

- $\begin{cases} x = rcos(\theta) \\ y = rsin(\theta) \\ z = z \end{cases}$
- $\iiint\limits_{\Omega} f(x,y,z)dxdydz = \int_{\alpha}^{\beta} d\theta \int_{r_1(\theta)}^{r_2(\theta)} rdr \int_{z_1(r,\theta)}^{z_2(r,\theta)} f(rcos(\theta), rsin(\theta), z)dz$

在球面坐标系中计算三重积分：

- $\begin{cases} x = rsin(\phi)cos(\theta) \\ y = rsin(\phi)sin(\theta) \\ z = rcos(\phi) \end{cases}$
- $\iiint\limits_{\Omega} f(x,y,z)dxdydz = \iiint\limits_{\Omega} f(rsin(\phi)cos(\theta), rsin(\phi)sni(\theta), rcos(\phi))r^2sin(\phi)drd\theta d\phi$

4.1.6 无穷级数

无穷级数是研究函数和数值计算的重要数学工具。设给定数列 $\{u_n\}$，将表达式 $u_1 + u_2 + \cdots + u_n + \cdots$ 称为常数项无穷级数，简称级数，记为 $\sum_{n=1}^{\infty} u_n = u_1 + u_2 + \cdots + u_n + \cdots$，这里 u_n 为第 n 项或通项。

如果级数的部分和数列 $\{S_n\}$ 有极限 S，即 $\lim\limits_{n\to\infty} S_n = S$，则级数收敛，$S$ 为级数的和，$S_n = u_1 + u_2 + \cdots + u_n, S = \sum_{n=1}^{\infty} u_n$。否则级数发散，没有和。当级数收敛时，余项 $r_n = S - S_n = u_{n+1} + u_{n+1} + cdots$ 是部分和的近似误差。

常用级数：

- 等比/几何级数：$\sum_{n=0}^{\infty} aq^n, a \neq 0$，$|q| < 1$，级数收敛，和为 $\frac{a}{1-q}$；$|q| \geq 1$，级数发散
- 调和级数：$\sum_{n=1}^{\infty} \frac{1}{n}$ 发散

级数的性质：

- 可加性：$\sum_{n=1}^{\infty} u_n, \sum_{n=1}^{\infty} v_n$ 均收敛，和为 S_1, S_2，则级数 $\sum_{n=1}^{\infty}$ 收敛，和为 $S_1 + S_2$
- 线性：$\sum_{n=1}^{\infty} u_n$ 和 $\sum_{n=1}^{\infty} ku_n, k \neq 0$ 同敛散，k 为常数，且收敛时二者的和成 k 倍
- 有限项无关性：增加、减少或改变级数的有限项后不改变技术的敛散性
- 分块加和不变性：收敛的级数任意加括号后所得的新级数仍然收敛，且和不变
- 级数收敛的必要条件：余项趋于 0，$\lim\limits_{n\to\infty} u_n = 0$

常数项级数的收敛性判别从两种基础形式入手，正项级数（通项非负）和交错项级数（逐项换号）。

正项级数收敛的充要条件是部分和数列 $\{S_n\}$ 有界（其实是单调增有上上界，则极限收敛）。正项级数的判别方法：

- 比较判别法：$\forall n, u_n \leq v_n$，则 $\sum_{n=1}^{\infty} v_n$ 收敛 $\Rightarrow \sum_{n=1}^{\infty} u_n$ 收敛；$\sum_{n=1}^{\infty} v_n$ 发散 $\Rightarrow \sum_{n=1}^{\infty} v_n$ 发散。比较判别法的极限形式为 $\lim\limits_{n\to\infty} \frac{u_n}{v_n} = l$，若 $l = 0$，则 $\sum_{n=1}^{\infty} v_n$ 收敛 $\Rightarrow \sum_{n=1}^{\infty} u_n$ 收敛；若 $l = +\infty$，则 $\sum_{n=1}^{\infty} v_n$ 发散 $\Rightarrow \sum_{n=1}^{\infty} u_n$ 发散；若 $0 < l < \infty$，二者同敛散。

- 比值判别法：如果 $\lim\limits_{n\to\infty}\frac{u_{n+1}}{u_n}=\rho$，若 $\rho<1$，级数收敛；若 $\rho>1$ 或 $\rho=+\infty$，级数发散；若 $\rho=1$，无法判断。
- 根值判别法：如果 $\lim\limits_{n\to\infty}\sqrt[n]{u_n}=\rho$，当 $\rho<1$，级数收敛；当 $\rho>1$，级数发散；当 $\rho=1$，无法判断。

P-级数：$1+\frac{1}{2^p}+\cdots+\frac{1}{n^p}+\cdots$，当 $p\le 1$，级数发散，当 $p>1$，级数收敛。

交错级数收敛的 Leibniz 判别法：设交错级数 $\sum_{n=1}^{\infty}(-1)^{n-1}u_n$，如果满足（1）$u_n\ge u_{n+1},n=1,2,\cdots$；（2）$\lim\limits_{n\to\infty}u_n=0$，则级数收敛，且和 $S\le u_1$，余项 $|r_n|\le u_{n+1}$。

通项的符号没有规定的级数是任意项级数，其敛散性从由每项绝对值组成的正项级数开始分析。如果级数 $\sum_{n=1}^{\infty}|u_n|$ 收敛，则级数 $\sum_{n=1}^{\infty}u_n$ 收敛。此时 $\sum_{n=1}^{\infty}u_n$ 绝对收敛，如果 $\sum_{n=1}^{\infty}u_n$ 收敛而 $\sum_{n=1}^{\infty}|u_n|$ 发散，则 $\sum_{n=1}^{\infty}u_n$ 条件收敛。

设函数列 $u_1(x),u_2(x),\cdots,u_n(x),\cdots$ 中的每一项都在 I 上有定义，则称 $\sum_{n=1}\infty u_n(x)$ 为定义在区间 I 上的**函数项级数**。如果 $x_o\in I$，使得 $\sum_{n=1}^{\infty}u_n(x_0)$ 收敛，则称点 x_0 为函数项级数的收敛点，否则为发散点。由所有的收敛点构成的集合为收敛域，记为 D。在收敛域上，函数项级数的和函数为 $S(x)=\sum_{n=1}^{\infty}u_n(x),x\in D$。

幂级数是特殊的函数项级数：$\sum_{n=0}^{\infty}a_n(x-x_0)^n$，其中 $a_0,a_1,\cdots$ 为幂级数的系数。（Abel 定理）当 $x=x_1\ne 0$ 时幂级数收敛，则 $\forall x,|x|<|x_1|$，幂级数都绝对收敛。当 $x=x_2$ 时幂级数发散，则 $\forall x,|x|>|x_2|$，幂级数发散。可以导出：如果幂级数不仅在 $x=0$ 处收敛，也不是在所有定义域上收敛，则必然存在唯一的正数 R，使得 $\forall x,|x|<R$，级数绝对收敛，$\forall x,|x|>R$，级数发散，$|x|=R$，无法判断。此时 R 为收敛半径，$(-R,R)$ 为收敛区间。

收敛半径的计算：设幂级数 $\sum_{n=0}^{\infty}a_nx^n$，且 $\lim\limits_{n\to\infty}|\frac{a_{n+1}}{a_n}|=\rho$，则 $0<\rho<+\infty$ 时，$R=1/\rho$，$\rho=0$ 时，$R=+\infty$，$\rho=+\infty$ 时，$R=0$。

幂级数的运算法则，$\sum_{n=0}^{\infty}a_nx^n,\sum_{n=0}^{\infty}b_nx^n$ 的收敛半径分别为 R_1,R_2，如果 $R=min\{R_1,R_2\}$，则当 $x\in -R,R$

- $\sum_{n=0}^{\infty}a_nx^n\pm\sum_{n=0}^{\infty}b_nx^n=\sum_{n=0}^{\infty}(a_n\pm b_n)x^n$
- $(\sum_{n=0}^{\infty}a_nx^n)(\sum_{n=0}^{\infty}b_nx^n)=\sum_{n=0}^{\infty}c_nx^n,c_n=\sum_{k=0}^{n}a_kb_{n-k}$

和函数的性质，设 $\sum_{n=0}^{\infty}a_nx^n$ 的收敛半径为 R，和函数为 $S(x)$，则

1. 连续性：$S(x)$ 在 $(-R,R)$ 上连续，当幂级数在 $x=R(-R)$ 处收敛时，$S(x)$ 在 $x=R$ 处左连续（$x=-R$ 处右连续）
2. 逐项微分：$S(x)$ 在 $(-R,R)$ 内可导，并可逐项求导，即 $S'(x)=(\sum_{n=0}^{\infty}a_nx^n)'=\sum_{n=0}^{\infty}(a_nx^n)'=\sum_{n=0}^{\infty}na_nx^{n-1},x\in(-R,R)$
3. 逐项积分：$S(x)$ 在 $(-R,R)$ 内可积，且可逐项积分，即 $\forall x\in(-R,R)$，$\int_0^x S(x)dx=\int_0^x(\sum_{n=0}^{\infty}a_nx^n)dx=\sum_{n=0}^{\infty}\int_0^x a_nx^ndx=\sum_{n=0}^{\infty}\frac{1}{n+1}a_nx^{n+1}$

设函数 $f(x)$ 在 $x=x_0$ 点的某邻域 $U(x_0)$ 内有任意阶导数，且 $f(x)$ 在 x_0 点的 Taylor 公式为

$$f(x)=\sum_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k+R_n(x) \tag{4.23}$$

则 $f(x)$ 在该领域内可以展开为 Taylor 级数的充分必要条件为 $\forall x\in U(x_0),\lim\limits_{n\to\infty}R_n(x)=0$。

从而，函数在 x_0 点处展开为幂级数的一般步骤如下：

1. 求 $f(x_0), f'(x_0), \cdots, f^{(n)}(x_0), \cdots$
2. 得到幂级数 $\sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$，并求出收敛半径 R
3. $\forall x \in (-R, R)$，证明 $\lim\limits_{n\to\infty} R_n(x) = 0$

4.1.7 微分方程与差分方程

变量之间的显式关系通常要从微分方程和差分方程的解中获得，解方程也许是问题分析的第一步。微分方程是包含未知函数、未知函数的导数或微分、自变量之间关系的方程，导数的最高阶为微分方程的阶。

n 阶微分方程的一般形式为

$$F(x, y, y', \cdots, y^{(n)}) = 0 \tag{4.24}$$

有的微分方程的解中函数任意的常数，这样的解为通解。通解加上定解条件就得到特解。

一阶微分方程的基础形式如下：

- 可分离变量的微分方程：$f_1(x)dx = f_2(y)dx \Rightarrow F_2(y) = F_1(x) + C$
- 齐次方程：$\frac{dy}{dx} = \phi(\frac{y}{x}) \Rightarrow u = \frac{y}{x}, \frac{dy}{dx} = u + x\frac{du}{dx} \Rightarrow x\frac{du}{dx} = \phi(u) - u, \frac{du}{\phi(u)-u} = \frac{1}{x}dx \Rightarrow$ 化归为可分离变量形式
- 一阶齐次线性微分方程 $\frac{dy}{dx} + P(x)y = 0$ 的通解为 $y = Ce^{-\int P(x)dx}$。一阶线性微分方程：$\frac{dy}{dx} + P(x)y = Q(x)$ 的通解为 $e^{-\int P(x)dx}\int Q(x)e^{\int P(x)dx}dx + Ce^{-\int P(x)dx}$
- Bernoulli 方程：$y' + P(x)y = Q(x)y^n \Rightarrow z = y^{1-n} \Rightarrow$ 一阶线性微分方程

二阶及二阶以上的微分方程为高阶微分方程。几种特殊类型的高阶微分方程可以转化为低阶微分方程：

- $y^{(n)} = f(x) \Rightarrow$ 两边逐阶积分
- $y'' = f(x, y') \Rightarrow p = y' = \frac{dy}{dx}, p' = y'' \Rightarrow \frac{dp}{dx} = f(x, p)$（一阶微分方程）
- $y'' = f(y, y') \Rightarrow p = y', y'' = p\frac{dp}{dy} \Rightarrow p\frac{dp}{dy} = f(y, p)$（一阶微分方程）

二阶齐次线性微分方程：

$$y'' + p(x)y' + q(x)y = 0 \tag{4.25}$$

设 y_1, y_2 是二阶齐次线性微分方程的解，则其线性组合 $y = C_1y_1 + C_2y_2$ 也是该方程的解。

函数的线性相关性：定义在 I 上的函数 $y_1, \cdots, y_n$，如果 $\exists k_1, k_2, \cdots, k_n \neq 0$，使得 $k_1y_1 + k_2y_2 + \cdots + k_ny_n \equiv 0, x \in I$，则 $y_1, y_2, \cdots, y_n$ 在 I 上线性相关，否则线性无关。

齐次线性方程通解结构定理：如果 y_1, y_2 为二阶齐次线性微分方程的线性无关的解，则 $y = C_1y_1 + C_2y_2$ 为该方程的通解，C_1, C_2 为任意的常数。

非齐次线性微分方程通解结构定理：设 y^* 是二阶线性非齐次微分方程的一个特解，Y 是相应的齐次方程的通解，则 $y = Y + y^*$ 为非齐次线性微分方程的通解。$\Rightarrow$ 叠加原理：设二阶非齐次线性微分方程 $y'' + p(x)y' + q(x)y = \sum_{k=1}^{n} f_k(x)$，如果 y^* 为非齐次线性微分方程 $y'' + p(x)y' + q(x)y = f_k(x), k = 1, 2, \cdots, n$ 的一个特解，则 $y^* = \sum_{k=1}^{n} y_k^*$ 为该方程的一个特

解。

二阶常系数齐次线性微分方程 $y''+py'+qy=0, p, q$ 为常数，特征方程为 $r^2+pr+q=0$，特征根为 r_1, r_2

- $p^2-4q>0, r_1\neq r_2 \Rightarrow y=C_1e^{r_1x}+C_2e^{r_2x}$
- $p^2-4q=0, r_1=r_2 \Rightarrow y=(C_1+C_2x)e^{r_1x}$
- $p^2-4q<0, r_{1,2}=\alpha\pm\beta i \Rightarrow e^{\alpha x}(C_1cos(\beta x)+C_2sin(\beta x))$

有两类二阶常系数非齐次线性微分方程的特解容易求：$y''+py'+qy=f(x)$，特征方程根为 r_1, r_2

- $f(x)=P_n(x)e^{\lambda x} \Rightarrow y^*=x^kQ_n(x)e^{\lambda x}, Q_n(x), P_n(x)$ 同阶，采用待定系数法求 $Q_n(x), k$ 的取值如下：$\begin{cases}\lambda\notin\{r_1,r_2\}, k=0\\ \lambda\in\{r_1,r_2\}, r_1\neq r_2, k=1\\ \lambda=r_1=r_2, k=2\end{cases}$
- $f(x)=e^{\lambda x}(P_l(x)cos(\omega x)+P_nsin(\omega x)) \Rightarrow y^*=x^ke^{\lambda x}[Q_1(x)cos(\omega x)+Q_2(x)sin(\omega x)], Q_1, Q_2$ 同阶且阶数为 $min\{l,n\}$，仍由待定系数法求，k 取值如下：$\begin{cases}\lambda\pm\omega i\notin\{r_1,r_2\}, k=0\\ \lambda\pm\omega i\in\{r_1,r_2\}, k=1\end{cases}$

n 阶 Euler 方程可以通过变量代换转变为常系数线性微分方程：

$$x^ny^{(n)}+a_1x^{n-1}y^{n-1}+\cdots+a_ny=f(x), x=e^t \Rightarrow \text{常系数非齐次线性微分方程} \tag{4.26}$$

下面介绍差分方程。设函数 $y=f(x)$，当自变量 x 依次取遍非负整数时，相应的函数值可以排成数列 $f(0), f(1), \cdots, f(x), \cdots$，记为 $y_0, y_1, \cdots, y_n, \cdots$。当自变量从 x 变化到 $x+1$ 时，相应的函数变化量 $y_{x+1}-y_x$ 为函数 y 在 x 点的步长为 1 的差分（一阶差分），记为 $\Delta y_x=y_{x+1}-y_x, x=0,1,\cdots$。

一阶差分的四则运算法则：

- $\Delta(Cy_x)=C\Delta y_x$
- $\Delta(y_x+z_x)=\Delta y_z+\Delta z_x$
- $\Delta(y_xz_x)=y_{x+1}\Delta z_x+z_x\Delta y_x=y_x\Delta z_x+z_{x+1}\Delta y_x$
- $\Delta(\frac{y_x}{z_x})=\frac{z_x\Delta y_x-y_x\Delta z_x}{z_xz_{x+1}}=\frac{z_{x+1}\Delta y_x-y_{x+1}\Delta z_x}{z_xz_{x+1}}$

二阶差分：$\Delta^2y_x=\Delta(\Delta y_x)=\Delta(y_{x+1}-y_x)=y_{x+2}-2y_{x+1}+y_x$，三阶差分：$\Delta^3y_x=y_{x+3}-3y_{x+2}+3y_{x+1}-y_x$，$n$ 阶差分：$\Delta^ny_x=\Delta(\Delta^{n-1}y_x), n=2,3,\cdots$。

含有未知函数差分的方程为*差分方程*：

$$F(x, y_x, \Delta y_x, \cdots, \Delta^n y_x)=0 \tag{4.27}$$

等价于

$$F(x, y_x, y_{x+1}, \cdots, y_{x+n})=0 \tag{4.28}$$

含有差分的最高阶数为差分方程的阶，和微分方程一样，差分方程也有解、通解、特解、初始条件等概念。

n 阶常系数线性差分方程的一般形式为

$$y_{x+n} + a_1 y_{x+n-1} + \cdots + a_{n-1} y_{x+1} + a_n y_x = f(x) \tag{4.29}$$

当 $f(x) \equiv 0$ 时，该方程为 n 阶常系数齐次线性差分方程，否则该方程为 n 阶常系数非齐次线性差分方程，有如下结论：

- 如果 $y_i(x), i = 1, \cdots, n$ 都是常系数齐次线性差分方程的解，则对任意的常数 $c_i, i = 1, \cdots, n$ 有 $y(x) = c_1 y_1(x) + \cdots + c_n y_n(x)$ 也是该方程的解。
- 齐次线性方程通解结构：如果 $y_i(x), i = 1, \cdots, n$ 是常系数齐次线性差分方程的 n 个线性无关的解，则对任意的常数 $c_i, i = 1, \cdots, n$，都有 $y(x) = c_1 y_1(x) + \cdots + c_n y_n(x)$ 是该方程的通解。
- 非齐次线性微分方程通解结构定理：如果 y^* 是非齐次线性微分方程的一个特解，Y_x 为相应的齐次差分方程的通解，则 $y(x) = y^* + Y$ 是差分方程的通解。
- 叠加原理：设 y_i^* 分别为方程 $y_{x+n} + a_1 y_{x+n-1} + \cdots + a_{n-1} y_{x+1} + a_n y_x = f_i(x), i = 1, \cdots, k$ 的特解，则 $y(x) = y_1^* + \cdots + y_k^*$ 为差分方程 $y_{x+n} + a_1 y_{x+n-1} + \cdots + a_{n-1} y_{x+1} + a_n y_x = f_1(x) + \cdots + f_k(x)$ 的一个特解。

一阶常系数齐次线性差分方程：$y_{x+1} - a y_x = 0, a \neq 0$ 的通解为

$$y_x = C a^x \tag{4.30}$$

其中 C 为任意常数，求解方法为迭代法或特征根法。一阶常系数线性非齐次差分方程的解：$y_{x+1} - a y_x = f(x), a \neq 0$

- $f(x) = P_n(x)$（多项式）$\Rightarrow \begin{cases} y_x^* = Q_n(x), a \neq 1 \\ y_x^* = x Q_n(x), a = 1 \end{cases}$ （待定系数法）
- $f(x) = u^x P_n(x), u \neq 0, 1 \Rightarrow y_x = u^z z_x \Rightarrow z_{x+1} - \frac{a}{u} z_x = \frac{1}{u} P_n(x)$（化归第一种形式）
- $f(x) = b_1 cos(\omega x) + b_2 sin(\omega x) \Rightarrow D = (cos(\omega) - a)^2 + sin^2(\omega)$

  $$\begin{cases} D \neq 0, y_x = C a^x + B_1 cos(\omega x) + B_2 sin(\omega x) \\ D = 0, \begin{cases} y_x = C + x(b_1 cos(2k\pi) + b_2 sin(2k\pi)), a = 1 \\ y_x = C(-1)^x - x[b_1 cos((2k+1)\pi x) + b_2 sin((2k+1)\pi x)], a = -1 \end{cases} \end{cases}$$

二阶常系数线性差分方程：

$$y_{x+2} + a y_{x+1} + b y_x = f(x) \tag{4.31}$$

二阶常系数齐次线性差分方程：$f(x) = 0$，特征方程 $\lambda^2 + a\lambda + b = 0$ 的根为 λ_1, λ_2

- $\lambda_1 \neq \lambda_2 \Rightarrow y_x = C_1 \lambda_1^x + C_2 \lambda_2^x, C_1, C_2$ 为任意常数
- $\lambda_1 = \lambda_2 = \lambda \Rightarrow y_x = (C_1 + C_2 x)\lambda^x$
- $\lambda_{1,2} = \alpha \pm \beta i \Rightarrow y_x = C_1 r^x cos(\theta x) + C_2 r^x sin(\theta x), r = \sqrt{\alpha^2 + \beta^2}, \theta = arctan(\frac{\beta}{\alpha})$

二阶常系数非齐次线性差分方程的解法，特征方程 $\lambda^2 + a\lambda + b = 0$ 的根为 r_1, r_2

- $f(x) = P_n(x) \Rightarrow y_x^* = x^k Q_n(x) \begin{cases} k = 0, 1 \notin r_1, r_2 \\ k = 1, 1 = r_1 \neq r_2 \\ k = 2, 1 = r_1 = r_2 \end{cases}$ （待定系数法）

- $f(x) = u^x P_n(x), u \neq 0, 1 \Rightarrow y_x = u^x z_x \Rightarrow z_{x+2} + \frac{a}{u} z_{x+1} + \frac{b}{u^2} z_x = \frac{1}{u^2} P_n(x)$（化归为第一种形式）

4.2 线性代数

线性代数是数据表示、数值特征分析和优化的工具之一，是计量经济学的主要数学工具。同时线性代数所揭示的变量特征也是微积分所需要的，如二次型和海塞矩阵分析多元函数的优化问题。线性代数的基本概念也在计量经济学中使用，比如线性相关、特征值。本科阶段用到的线性代数包括矩阵运算、线性相关性、解方程组、特征值、二次型等。

4.2.1 行列式与矩阵

由 n 个自然数 $1, 2, \cdots, n$ 按照任何次序排列得到的有序数组 $j_1, j_2, \cdots, j_n$ 为一个n *级排列*。规定从小到大为自然数间的标准顺序，在一个排列中，当两个元素的顺序与标准顺序不同时，就产生了*逆序*，一个 n 级排列的所有逆序的总和为*逆序数*，记为 $\tau(j_1 j_2 \cdots j_n)$，计算方法如下：

$$\tau(j_1 j_2 \cdots j_n) = k_1 + k_2 \cdots + k_n \tag{4.32}$$

其中 k_i 是元素 j_i 右边比 j_i 大的元素的数量。逆序数为奇数的排列为*奇排列*，逆序数为偶数的排列为*偶排列*。将排列中两个元素的位置互换为对换。一次对换必定改变排列的奇偶性。

设有 n^2 个数，排成 n 行 n 列的数表

$$\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{matrix}$$

表中不同行列的 n 个数的乘积共有 $n!$ 项，每项的符号为 $(-1)^\tau$，其代数和称为n *阶行列式*，记为

$$det(a_{ij}) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{j_1 j_2 \cdots j_n} (-1)^{\tau(j_1 j_2 \cdots j_n)} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \tag{4.33}$$

其中 $a_{1j_1} a_{2j_2} \cdots a_{nj_n}$ 称为 n 阶行列式的均匀分布项（均布项），$(-1)^{\tau(j_1 j_2 \cdots j_n)}$ 为均布项的符号因子，其正负取决于排列 $j_1 j_2 \cdots j_n$ 的奇偶性。

以下是几种特殊的行列式，值只与对角线元素有关。

- 主对角行列式：$D_1 = \begin{vmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{vmatrix} = \lambda_1\lambda_2\cdots\lambda_n$
- 次对角行列式：$D_2 = \begin{vmatrix} & & & \lambda_1 \\ & & \lambda_2 & \\ & \cdots & & \\ \lambda_n & & & \end{vmatrix} = (-1)^{\frac{n(n-1)}{2}}\lambda_1\lambda_2\cdots\lambda_n$
- 下三角行列式：$D = \begin{vmatrix} a_{11} & & & \\ a_{21} & a_{22} & & \\ \cdots & \cdots & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}$
- 上三角行列式：$D = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \cdots & \cdots \\ & & & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}$
- 反下三角行列式：

相应的，还有反下三角行列式和反上三角行列式，值只与对角线元素有关，但是符号为 $(-1)^{\frac{n(n-1)}{2}}$。运用分块矩阵将三角行列式推广：

- 分块对角行列式：$D = \begin{vmatrix} D_1 & \\ & D_2 \end{vmatrix} = |D_1||D_2|$
- 分块下/上三角行列式：$D = \begin{vmatrix} D_1 & \\ C & D_2 \end{vmatrix} = \begin{vmatrix} D_1 & C \\ & D_2 \end{vmatrix} = |D_1||D_2|$
- 分块反下/上三角行列式：$D = \begin{vmatrix} & D_1 \\ D_1 & C \end{vmatrix} = \begin{vmatrix} C & D_1 \\ D_2 & \end{vmatrix} = (-1)^{mn}|D_1||D_2|$（$m, n$ 为 D_1, D_2 的阶）

行列式的性质如下

- 行列式等于转置行列式：$D = D^T$，即 $det(a_{ij}) = det(a_{ji})$
- 换法性质：互换行列式两行/列仅改变行列式的符号。
- 倍法性质：行列式某行/列乘以 k 等于 k 乘以行列式。
- 行列式具有分行/列可加性。
- 消法性质：将行列式的某行/列的各元素乘以 k 加到另一行/列，行列式不变。

运用以上性质可以将行列式转换为三角行列式求值。对于一般的行列式，则需要运用行列式展开定理降阶。

在 n 阶行列式 $D = det(a_{ij})$ 中，把元素 a_{ij} 所在行列划去后，留下的 $n-1$ 阶行列式称为元素 a_{ij} 的余子式，记为 M_{ij}，则 $A_{ij} = (-1)^{i+j}M_{ij}$ 为元素 a_{ij} 的代数余子式。

n 阶行列式 D 等于它的任意一行/列的各元素与其对应的代数余子式乘积之和：

$$D=\sum_{k=1}^{n}a_{ik}A_{ik}=\sum_{k=1}^{n}a_{kj}A_{kj},i,j=1,2,\cdots,n \tag{4.34}$$

行列式某一行/列的元素与另一行/列对应元素的代数余子式的乘积之和为 0：

$$\sum_{k=1}^{n}a_{ik}A_{jk}=0,\sum_{k=1}^{n}a_{ki}A_{kj}=0,i\neq j \tag{4.35}$$

将以上结果推广得到 Laplace 定理，即设 D 为 n 阶行列式，在 D 中取定某 $k,1\leq k\leq n-1$ 行，则含于此 k 行中的所有 k 阶子式与其代数余子式的乘积之和为 D。

行列式是解线性方程组的利器。依据 Cramer 法则，对于线性方程组

$$\begin{cases}a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1,\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2,\\ \cdots a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=b_n,\end{cases} \tag{4.36}$$

其系数行列式为

$$D=\begin{vmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\ a_{21}&a_{22}&\cdots&a_{2n}\\ &&\cdots&\\ a_{n1}&a_{n2}&\cdots&a_{nn}\end{vmatrix} \tag{4.37}$$

替换行列式为

$$D_j=\begin{vmatrix}a_{11}&\cdots&a_{1,j-1}&b_1&a_{1,j+1}&\cdots&a_{1n}\\ a_{21}&\cdots&a_{2,j-1}&b_2&a_{2,j+1}&\cdots&a_{2n}\\ &&\cdots&\cdots&\cdots&&\\ a_{n1}&\cdots&a_{n,j-1}&b_n&a_{n,j+1}&\cdots&a_{nn}\end{vmatrix} \tag{4.38}$$

如果 $D\neq 0$，则方程组有唯一解：

$$x_j=\frac{D_j}{D},j=1,2,\cdots,n \tag{4.39}$$

反过来，如果该方程组无解或无穷解，则 $D=0$。当 $b_1=b_2=\cdots=0$ 时，该方程组为齐次线性方程组，$x_1=x_2=\cdots=0$ 为它的一个解，称为零解。如果齐次线性方程组的系数行列式 $D=0$，则该方程组只有零解，没有非零解。反过来，如果齐次线性方程组有非零解，则 $D\neq 0$。

由 $m\times n$ 个数 $a_{ij},i=1,2,\cdots,m,j=1,2,\cdots,n$ 按序排成一个 m 行 n 列的数表，称为矩阵，记为

$$A_{m\times n}=(a_{ij})_{m\times n}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\ a_{21}&a_{22}&\cdots&a_{2n}\\ &&\cdots&\\ a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix} \tag{4.40}$$

矩阵用大写字母表示。元素为实数的矩阵为实矩阵，元素为复数的矩阵为复矩阵。如果 A,B 的行列数都相同，则为同型矩阵。如果 $A=(a_{ij})_{m\times n},B=(b_{ij})_{m\times n}$ 有 $a_{ij}=b_{ij},i=$

$1,2,\cdots,m,j=1,2,\cdots,n$，则 A 与 B 相等，记为 $A=B$。

元素均为 0 的矩阵为零矩阵，记为 $O_{m\times n}$。当 $m=1$，$A_{m\times n}$ 为行矩阵，当 $n=1$，$A_{m\times n}$ 为列矩阵，当 $m=n$，$A_{m\times n}$ 为方阵，记为 A_n。

几种基础的矩阵形式如下：

- 对角矩阵：$\Lambda = diag(\lambda_1,\lambda_2,\cdots,\lambda_n) = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \cdots & \\ & & & \lambda_n \end{bmatrix}$
- 标量矩阵（数量矩阵）：$A = \begin{bmatrix} a & & & \\ & a & & \\ & & \cdots & \\ & & & a \end{bmatrix}$
- 单位矩阵：$E_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \cdots & \\ & & & 1 \end{bmatrix}$
- 上三角矩阵：$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \cdots & \cdots \\ & & & a_{nn} \end{bmatrix}$；下三角矩阵：$B = \begin{bmatrix} b_{11} & & & \\ b_{21} & b_{22} & & \\ \cdots & \cdots & \cdots & \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$
- 对称矩阵：$A=(a_{ij}),a_{ij}=a_{ji},i,j=1,2,\cdots,n$；反对称矩阵：$A=(a_{ij}),a_{ij}=-a_{ij},i,j=1,2,\cdots,n \Rightarrow a_{ii}=0,i=1,2,\cdots,n$

矩阵的运算包括加法、数乘、乘法、转置、可逆和初等变换。矩阵加法要求矩阵同型，然后对应元素相加，即 $A=(a_{ij})_{m\times n},B=(b_{ij})_{m\times n}$，

$$A+B=(a_{ij}+b_{ij})_{m\times n} \tag{4.41}$$

矩阵加法满足交换律、结合律，特别的，$A+(-A)=O$，当 $B=O$ 时，$A+O=A$。矩阵的数乘即为每个元素进行数乘，

$$\lambda A=(\lambda a_{ij})_{m\times n} \tag{4.42}$$

矩阵数乘满足如下规律：

- $\lambda A = A\lambda$
- $\lambda(\mu A) = (\lambda\mu)A$
- $(\lambda+\mu)A = \lambda A + \mu A$
- $\lambda(A+B) = \lambda A + \lambda B$
- $(-1)A = -A$

矩阵的乘法满足对乘加法则，即要求：$A=(a_{ij})_{m\times n}, B=(b_{ij})_{n\times p}$，

$$C=AB=(c_{ij})_{m\times p}, c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj}, i=1,2,\cdots,m, j=1,2,\cdots,p \tag{4.43}$$

对于单位矩阵 E，总有 $EA=AE=A$。矩阵乘法不满足交换律，但满足以下定律：

- $ABC=A(BC)$
- $\lambda(AB)=(\lambda A)B=A(\lambda B)$
- $A(B+C)=AB+AC, (B+C)A=BA+CA$

如果 $AB=BA$，则同型矩阵 A,B 是可换的。矩阵的幂就是矩阵的连乘，$A^k=A^{k-1}A$，矩阵的幂满足 $A^kA^l=A^{k+l}, (A^k)^l=A^{kl}, k,l\in N^+$，特别的 $A^0=E$。

矩阵的转置即行列互换：

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \Rightarrow A^T=\begin{bmatrix} a_{11} & a_{22} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ & & \cdots & \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \tag{4.44}$$

矩阵的转置满足以下规律：

- $(A^T)^T=A$
- $(A+B)^T=A^T+B^T$
- $(\lambda A)^T=\lambda A^T$
- $(AB)^T=B^TA^T$

因而 A 为对称矩阵 $\Leftrightarrow A^T=A$；A 为反对称矩阵 $\Leftrightarrow A^T=-A$。

由 n 阶方阵 A 的元素所构成的行列式为方阵 A 的行列式，记为 $|A|$ 或 $detA$，满足以下运算定律

- $|A^T|=A$
- $|\lambda A|=\lambda^n|A|$
- $|AB|=|A||B|$

设 A 为 n 阶方阵，如果存在方阵 B 使得 $AB=BA=E$，则方阵 A 是可逆的，称 B 为 A 的逆矩阵，简称 A 的逆，逆是唯一的，记为 A^{-1}。

设 n 阶方阵 $A=(a_{ij})$，元素 a_{ij} 在 $|A|$ 中的代数余子式为 $A_{ij}, i,j=1,2,\cdots,n$，则矩阵

$$A^*=\begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ & & \cdots & \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix} \tag{4.45}$$

为 A 的伴随矩阵。矩阵 A 可逆的充要条件是 $|A|\neq 0$，此时

$$A^{-1}=\frac{1}{|A|}A^* \tag{4.46}$$

矩阵 A 可逆的充要条件是存在 n 阶方阵 B 使得 $AB=E$ 或 $BA=E$，当 A 可逆时，$B=A^{-1}$。同时 $A=B^{-2}$。可逆矩阵 A 的性质如下：

- A^{-1} 可逆，且 $(A^{-1})^{-1} = A$
- $\lambda A, \lambda \neq 0$ 可逆，$(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$
- 若 B 也可逆，则 AB 可逆，且 $(AB)^{-1} = B^{-1}A^{-1}$
- A^T 可逆，$(A^T)^{-1} = (A^{-1})^T$

在分析高阶矩阵时，可以采用分块矩阵运算的办法。分块矩阵的运算规律和矩阵一致。特别注意的分块矩阵的转置中需要对每一字块都转置。

矩阵的初等变换：设 $A = (a_{ij})_{m\times n}$，则以下三种变换为矩阵 A 的初等行/列变换，统称为*初等变换*：

- 换法性质：交换 A 的两行/列，记为 $r_i \leftrightarrow r_j$
- 倍法性质：用 $k, k \neq 0$ 乘以 A 的某行/列，记为 $r_i \times k$
- 消法性质：用一个数乘以 A 的某行/列的各元素，再加到 A 的另一行/列对应的元素上去，记为 $kr_j + r_i$

初等变换可逆。如果矩阵 A 经过有限次初等变换后变成 B，则矩阵 A 与 B 等价，记为 $A \cong B$，性质如下：

- 反身性：$A \cong A$
- 对称性：$A \cong B \Rightarrow B \cong A$
- 传递性：$A \cong B, B \cong C \Rightarrow A \cong C$

通过初等变换，可以将任意的矩阵转换为等价的行阶梯矩阵，满足：

- 所有元素为 0 的行都在矩阵底部
- 每行左起第一个非零元素下方的元素均为 0

即 $A \cong A_2 \cong A_3$，A_2 为行阶梯矩阵，也称行最简矩阵。对 A_2 继续变换可以得到 A_3，标准型矩阵，其左上角是一个单位矩阵，其余元素为 0。

对单位矩阵 E 实施一次初等变换得到的矩阵为*初等矩阵*，有三类初等矩阵：

- 初等换法矩阵：$E, r_i \leftrightarrow r_j \Rightarrow P(i,j)$
- 初等倍法矩阵：$E, r_i \times k \Rightarrow P(i[k])$
- 初等消法矩阵：$E, kr_j + r_i \Rightarrow P(i, j[k])$

初等矩阵的转置矩阵仍为初等矩阵，初等矩阵为可逆矩阵，逆矩阵仍为同类型的初等矩阵。

用初等矩阵左乘 A，相当于对 A 实施相应的初等行变换，用初等矩阵右乘 A，相当于对 A 实施相应的初等列变换。如果方阵 A 可逆，则可以经过有限次初等变换化为单位矩阵，即 $A \cong E$，因而可逆矩阵可以表示为多个初等矩阵的乘积。运用初等矩阵可以得出矩阵的逆，这称为初等变换法：

- 初等行变换：$(A, E) \cong (E, A^{-1})$
- 初等列变换：$\begin{bmatrix} A \\ E \end{bmatrix} \cong \begin{bmatrix} E \\ A^{-1} \end{bmatrix}$

秩是矩阵的数字特征。在矩阵 $A_{m\times n}$ 中，任取 $k, 1 \leq k \leq min\{m, n\}$ 列，位于这些行列交叉处的 k^2 个元素，不改变它们的次序得到的 k 阶行列式为 A 的一个*k 阶子式*，共 $C_m^k C_n^k$ 个。

如果矩阵 A 中有一个 r 阶子式 $D_r \neq 0$，而所有 $r+1$ 阶子式的值都等于 0，则称 D_r 为矩阵 A 的一个最高阶非零子式，其阶数 r 称为矩阵 A 的秩，记为 $R(A)$。性质如下：

- $R(A^T) = R(A)$
- $R(A) \leq min\{m, n\}$

由于初等变换不改变矩阵的秩，因而可以用初等变换求矩阵的秩。行阶梯矩阵的秩为非零行的行数，列阶梯矩阵的秩为非零列的列数。若 $A_{m\times n}$ 的秩为 $R(A) = m$，则 A 为行满秩矩阵，若 $R(A) = n$，则 A 为列满秩矩阵。若 $R(A_n) = n$，则方阵 A 为满秩矩阵。从而方阵矩阵满秩的充要条件是 $|A| \neq 0$，这也是方阵 A 可逆的充要条件，因此

$$R(A_n) = n \Leftrightarrow |A| \neq 0 \Leftrightarrow \exists A^{-1} \tag{4.47}$$

而且，设 $A_{m\times n}$，齐次线性方程组 $Ax = 0$ 有非零解的充要条件是 $R(A) < n$。

4.2.2 线性相关性

设 n 个有序数 $a_1, a_2, \cdots, a_n$ 所组成的数组称为 n 维向量，这 n 个数称为该向量的 n 个分量，第 i 个数 a_i 称为第 i 个分量。向量用小写希腊字母表示，分量用带下表的拉丁字母表示，向量分列向量 α 和行向量 α^T，如下：

$$\alpha = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{bmatrix}, \alpha^T = \begin{bmatrix} a_1, a_2, \cdots, a_n \end{bmatrix} \tag{4.48}$$

所有分量均为 0 的向量为零向量。所有分量乘以 -1 得到原向量的负向量。如果 n 维向量 α, β 的所有对应分量相等，则向量相等，记为 $\alpha = \beta$。向量加法，$\alpha = (a_1, a_2, \cdots, a_n), \beta = (b_1, b_2, \cdots, b_n)$

$$\alpha + \beta = \begin{bmatrix} a_1 + b_1, a_2 + b_2, \cdots, a_n + b_n \end{bmatrix} \tag{4.49}$$

可以定义负向量 $-\alpha$。向量数乘，设 $\alpha = (a_1, a_2, \cdots, a_n)$，$\lambda$ 是数，则 $\lambda\alpha = \begin{bmatrix} \lambda a_1, \lambda a_2, \cdots, \lambda a_n \end{bmatrix}$。

向量加法和数乘称为向量的线性运算，满足以下运算法则：

- $\alpha + \beta = \beta + \alpha$
- $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$
- $\alpha + \vec{0} = \alpha$
- $\alpha + (-\alpha) = \vec{0}$
- $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$
- $(\lambda + \mu)\alpha = \lambda\alpha + \mu\alpha$
- $(\lambda\mu)\alpha = \lambda(\mu\alpha)$
- $1\alpha = \alpha$

若干同维的列向量/行向量所组成的集合为向量组。矩阵 $A=(a_{ij})_{m\times n}$ 可以看做列向量组：

$$A=\begin{bmatrix}\alpha_1,\alpha_2,\cdots,\alpha_n\end{bmatrix},\alpha_j=\begin{bmatrix}a_{1j}\\a_{2j}\\\cdots\\a_{mj}\end{bmatrix},j=1,2,\cdots,n \tag{4.50}$$

或行向量组：

$$A=\begin{bmatrix}\beta_1\\\beta_2\\\cdots\\\beta_n\end{bmatrix},\beta_i=\begin{bmatrix}a_{i1},a_{i2},\cdots,a_{in}\end{bmatrix},i=1,2,\cdots,m \tag{4.51}$$

设 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 为 n 维向量组，$\lambda_1,\lambda_2,\cdots,\lambda_s$ 为一组数，称 $\lambda_1\alpha_1+\lambda_2\alpha_2+\cdots+\lambda_s\alpha_s$ 为向量组 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 的一个线性组合，称 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 为线性组合的系数。如果存在一组数 $k_1,k_2,\cdots,k_s$ 使

$$k_1\alpha_1+k_2\alpha_2+\cdots+k_s\alpha_s=\alpha \tag{4.52}$$

成立，则称 α 能由向量组 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性表示。如果存在一组不全为 0 的数 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 使得

$$\lambda_1\alpha_1+\lambda_2\alpha_2+\cdots+\lambda_s\alpha_s=\vec{0} \tag{4.53}$$

则称向量组 $\alpha_1,\alpha_2,\cdots,\alpha_s$ 线性无关，否则称其线性相关。有如下结论：

- 一个向量 α 线性相关 $\Leftrightarrow\alpha=\vec{0}$；一个向量 α 线性无关 $\Leftrightarrow\alpha\neq\vec{0}$
- 两个向量线性相关 $\Leftrightarrow$ 对应分量成比例
- 线性相关向量组的扩大组必定线性相关
- 线性无关向量族组的非空部分向量组仍然线性无关

向量组 $\alpha_1,\alpha_2,\cdots,\alpha_m,m\geq 2$ 线性相关的充要条件是至少有一个向量可以由其余 $m-1$ 个向量线性表示。如果向量组 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 线性无关，而 $\alpha_1,\alpha_2,\cdots,\alpha_m,\beta$ 线性相关，则 β 能由 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 线性表示，且表达式唯一。

线性相关性与线性方程组的解、秩联系紧密。线性方程组可以写成

$$Ax=b,x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n=b \tag{4.54}$$

因而 b 能由向量组 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性表示 $\Leftrightarrow$ 线性方程组 $x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n=b$ 有解。向量组 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性相关（无关）$\Leftrightarrow$ 齐次线性方程组 $x_1\alpha_1+x_2\alpha_2+\cdots+x_n\alpha_n=\vec{0}$ 有（无）非零解 $\Leftrightarrow$ 当 $A=A_n,|A|=0(\neq 0)$。当 $A_{m\times n},m>n$，m 个 n 维向量一定线性相关。

设两个向量组

$$A:\alpha_j=\begin{bmatrix}a_{1j}\\a_{2j}\\\cdots\\a_{rj}\end{bmatrix},B:\beta=\begin{bmatrix}a_{1j}\\a_{2j}\\\cdots\\a_{rj}\\a_{(r+1)j},j=1,2,\cdots,m\end{bmatrix} \tag{4.55}$$

即 β_j 在 α_j 上增加了一个分量，如果 A 线性无关，则 B 线性无关。拓展一下，r 维向量组增加 $n-r$ 个分量得到 n 维向量组，如果 r 维向量组线性无关，则 n 维向量组线性无关，反过来，如果 n 维向量组线性相关，则 r 维向量组线性相关。

如果向量组 $A:\alpha_1,\alpha_2,\cdots,\alpha_s$ 中每个向量都能由向量组 $B\beta_1,\beta_2,\cdots,\beta_t$ 线性表示，则称向量组 A 能由向量组 B 线性表示，如果二者能够相互线性表示，则称它们等价。向量组等价关系具有反身性、对称性和传递性。

如果向量组 $A:\alpha_1,\alpha_2,\cdots,\alpha_s$ 能由向量组 $B:\beta_1,\beta_2,\cdots,\beta_t$ 线性表示，且 $s>t$，则 A 必定线性相关。因而，如果 A 线性无关，且能由 B 线性表示，一定有 $s\le t$。如果 A,B 等价，且都线性无关，则 $s=t$。

对于向量组 $A:\alpha_1,\alpha_2,\cdots,\alpha_m$，如果存在 A 的部分向量组 $A_0:\alpha_{j1},\alpha_{j2},\cdots,\alpha_{jr}$ 满足以下条件，则称 A_0 是 A 的一个吉大线性无关向量组，简称极大无关组：（1）向量组 A_0 线性无关；（2）向量组 A 中任一向量可用 A_0 线性表示。极大无关组所含向量的个数 r 称为向量组的秩，记为 $R(\alpha_1,\alpha_2,\cdots,\alpha_m)$。规定只含零向量的向量组秩为 0。

向量组线性无关 $\Leftrightarrow R(A)=m$，向量组线性相关 $\Leftrightarrow R(A)<m$。A_0 为 A 的极大无关组 $\Leftrightarrow A$ 中任意 $r+1$ 个向量都线性相关。如果 $R(A)=r$，则 A 中任意 r 个线性无关的向量都是 A 的一个极大无关组。

向量组等价于它的任意一个极大无关组。$\Rightarrow$ 向量组的任意两个极大无关组等价，向量组的秩唯一。

如果向量组 A 能由向量组 B 线性表示，则向量组 A 的秩不大于向量组 B 的秩。等价向量组的秩相等。矩阵的秩等于行向量组的秩，也等于列向量组的秩。

设 A,B 均为 $m\times n$ 矩阵，则

$$R(A+B)\le R(A)+R(B) \tag{4.56}$$

设 $C=AB$，则 $R(C)\le min\{R(A),R(B)\}$。设 $A_{m\times n},P,Q$ 分别是 m,n 阶的可逆矩阵，则

$$R(A)=R(PA)=R(AQ)=R(PAQ) \tag{4.57}$$

由 n 维向量的全体所构成的集合

$$R^n=\{x=(x_1,x_2,\cdots,x_n)|x_1,x_2,\cdots,x_n\in R\} \tag{4.58}$$

称为 n 维向量空间。定义在 R^n 上的向量线性运算是封闭的，并满足相应的运算律。设 W 是向量空间 R^n 的子集合，如果 W 关于 R^n 中的两条运算也封闭且满足 8 条运算律，则称 W 为 R^n 的子空间，在判断时只需判断是否两种线性运算封闭即可。一般地，设向量组

$A: \alpha_1, \alpha_2, \cdots, \alpha_r \in R^n$，称

$$V = \{x = \lambda_1\alpha_1 + \lambda_2\alpha_2 + \cdots + \lambda_r\alpha_r | \lambda_1, \lambda_2, \cdots, \lambda_r \in R\} \tag{4.59}$$

为有向量组 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 生成的 R^n 的子空间。如果向量组 A 满足：(1) 线性无关；(2) V 中任意一个向量都可以由 A 线性表示，则 A 为向量空间 V 的一个基，r 称为向量空间 V 的维数，称 V 为一个 r 维向量空间。

4.2.3 线性方程组

齐次线性方程组：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases} \tag{4.60}$$

或写成

$$Ax = 0, A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \cdots x_n \end{bmatrix} \tag{4.61}$$

或

$$x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = 0 \tag{4.62}$$

如果向量 ε 使得上述方程组 $A\varepsilon = 0$ 成立则称 ε 为其解向量。$Ax = 0$ 总有零解。若 $x = \varepsilon \neq 0$，则 $x = \varepsilon$ 为非零解，条件是 $R(A) < n$。

齐次线性方程组解的性质：

- 加法性质：$A\varepsilon_1 = 0, A\varepsilon_2 = 0 \Rightarrow A(\varepsilon_1 + \varepsilon_2) = 0$
- 数乘性质：$A\varepsilon = 0 \Rightarrow A(k\varepsilon) = 0$

由齐次线性方程组的所有解构成的向量空间为解空间。n 元齐次线性方程组，当 $R(A) = r < n$ 时，其解空间为维数为 $n - r$，通解为

$$x = k_1\varepsilon_1 + k_2\varepsilon_2 + \cdots + k_{n-r}\varepsilon_{n-r} \tag{4.63}$$

其中 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{n-r}$ 为 $Ax = 0$ 的基础解系，$k_1, k_2, \cdots, k_{n-r}$ 为任意常数。

非齐次线性方程组：

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases} \tag{4.64}$$

或写成

$$Ax = b \tag{4.65}$$

或

$$x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n = b \tag{4.66}$$

当非齐次线性方程组有解时，称其相容。该方程组有解 $\Leftrightarrow b$ 能用 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 线性表示 $\Leftrightarrow R(B) = R(A), B = [A, b]$，$A$ 为系数矩阵，B 为增广矩阵。非齐次线性方程组解的性质：

- 减法性质：$A\eta_1 = b, A\beta_2 = b \Rightarrow A(\eta_1 - A\eta_2) = 0$
- 加法性质：$A\varepsilon = 0, A\eta^* = b \Rightarrow A(\varepsilon + \eta^*) = b$

从而，设 η^* 是 $Ax = b, b \neq 0$ 的一个解，$\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{n-r}$ 是 $Ax = 0$ 的基础解系，则方程 $Ax = b$ 的通解为

$$x = k_1\varepsilon_1 + k_2\varepsilon_2 + \cdots + k_{n-r}\varepsilon_{n-r} + \eta^* \tag{4.67}$$

其中 $k_1, k_2, \cdots, k_{n-r}$ 为任意常数。总结非齐次线性方程组的解如下：

$$\begin{cases} R([A,b]) \neq R(A) \Rightarrow \text{无解} \\ R([A,b]) = R(A) \begin{cases} = n, \text{唯一解} \\ < n, \text{无穷多解} \end{cases} \end{cases}$$

4.2.4 矩阵特征值、特征向量与对角化

设 n 维实向量 x, y 称 (x, y) 为 x, y 的内积。

$$(x,y) = x_1y_1 + x_2y_2 + \cdots + x_ny_n, x = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} \tag{4.68}$$

当 x, y 均为列向量，$(x, y) = x^T y$，当 x, y 均为行向量，$(x, y) = xy^T$。内积满足以下运算律：

- 交换律：$(x, y) = (y, x)$
- 数乘：$(\lambda x, y) = \lambda(x, y)$
- 分配律：$(x + y, z) = (x, z) + (y, z)$

设 n 维列向量 $x = [x_1, x_2, \cdots, x_n]^T$，令

$$||x|| = \sqrt{(x,x)} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x} \tag{4.69}$$

称 $||x||$ 为 x 的范数（长度），具有如下性质：

- 非负性：$||x|| \geq 0$，当且仅当 $x = \vec{0}$ 时，$||x|| = 0$
- 齐次性：$||\lambda x|| = |\lambda| ||x||$
- 三角不等式：$||x + y|| \leq ||x|| + ||y||$

长度为 1 的向量为单位向量。$\forall \aleph \neq 0, e = \frac{\alpha}{||\aleph||}$ 都是单位向量。向量的内积满足 Chauchy-Schwarz 不等式

$$(x,y)^2 \leq (x,x)(y,y) \tag{4.70}$$

当 $||x|| \neq 0, ||y|| \neq 0$ 时，称 $\theta = arccos\frac{(x,y)}{||x||||y||}$ 为 x, y 的夹角。

如果 n 维实向量 $\aleph, \beta$ 内积为 0，即 $(\aleph, \beta) = 0$，则 α, β 正交，记为 $\alpha \perp \beta$。如果非零向量组 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 满足

$$(\alpha_i, \alpha_j) = \begin{cases} 0, i \neq j k_i > 0, i = j \end{cases}, i, j = 1, 2, \cdots, n, k \in R \tag{4.71}$$

则称该向量组为正交向量组。正交向量组必定线性无关。如果正交向量组中每一个向量都是单位向量，则该向量组为正交规范向量组，简称正交规范组。如果空间向量的基为正交规范组，则称为正交规范基。将任意一组线性无关的向量转换为正交规范向量组的方法为 Schmidt 正交化方法：

$$\alpha_1, \alpha_2, \cdots, \alpha_r \Rightarrow \begin{cases} \beta_1 = \alpha_1, e_1 = \frac{\beta_1}{||\beta_1||} \\ \beta_2 = \alpha_2 - (\alpha_2, e_1)e_1, e_2 = \frac{\beta_2}{||\beta_2||} \\ \beta_3 = \alpha_3 - (\alpha_3, e_1)e_1 - (\alpha_3, e_2)e_2, e_3 = \frac{\beta_3}{||\beta_3||} \\ \cdots \\ \beta_r = \alpha_r - (\alpha_r, e_1)e_1 - \cdots - (\alpha_r, e_{r-1})e_{r-1}, e_r = \frac{\beta_r}{||\beta r||} \end{cases} \tag{4.72}$$

如果 n 阶方阵 A 满足 $A^T A = E$，则称其为正交矩阵。性质如下：

- $A^-1 = A^T$
- $|A| = \pm 1$，可逆
- A^T 为正交阵
- 若 B 为正交阵，则 AB 也为正交阵

正交阵的行/列向量组均为正交规范向量组。若 P 为正交阵，则称线性变换 $y = Px$ 为正交变换，且 $||y|| = ||x||$。

设 A 为 n 阶方阵，如果对于数 λ，存在非零列向量 $\alpha \in R^n$ 使得 $A\alpha = \lambda\alpha$，则称 λ 为 A 的一个特征值，α 为 A 对应于特征值 λ 的特征向量。根据上式解出 λ, α 的条件是 $(A - \lambda E)\alpha = 0$ 有非零解，即 $|A - \lambda E| = 0$（特征方程），因而特征值也称为特征根。

设 A 为 n 阶方阵，则 A, A^T 具有相同的特征值。设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为方阵 $A = (a_{ij})$ 的 n 个特征值，则

- $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii} = trA$（迹）
- $\prod_{i=1}^n \lambda_i = |A|$

特别的，当 $0 \in \{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ 时，$|A| = 0$，方阵 A 不可逆。设 $p_1, p_2, \cdots, p_n$ 为方阵 A 对应于特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 的特征向量，则 $p_1, p_2, \cdots, p_n$ 线性无关。

设 A, B 是 n 阶方阵，如果存在 n 阶可逆矩阵 P，使得 $P^{-1}AP = B$，则称矩阵 A 相似于矩阵 B，记为 $A \sim B$，具有如下性质：

- 反身性：$A \sim A$

- 对称性：$A \sim B \Rightarrow B \sim A$
- 传递性：$A \sim B, B \sim C \Rightarrow A \sim C$

以可逆矩阵 P 对 A 进行运算称为相似变换，P 称为相似因子。相似矩阵性质如下：

- $A \sim B \Rightarrow R(A) = R(B)$
- 相似矩阵具有相同的特征值。对角矩阵的特征值为对角线元素。

n 阶方阵 A 与对角矩阵 Ω 相似 $\Leftrightarrow$ 是 A 有 n 个线性无关的特征向量。如果 A 有 n 个线性无关的特征值，则 n 阶方阵可对角化。

实对称矩阵的特征值一定为实数，而且不同的特征值对应的特征向量是正交的。设 A 为 n 阶实对称矩阵，λ_r 是 A 的 r 重特征值，则 $R(\lambda_r E - A) = n - r$，从而特征值 λ_r 恰好对应 r 个线性无关的特征向量。必定存在 n 阶正交矩阵 P 使得 $P^{-1}AP = \Omega$，Ω 是以 A 的 n 个特征值为对角线元素的对角矩阵。

4.2.5 二次型

含有 n 个变量 $x_1, x_2, \cdots, x_n$ 的二次多项式

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) &= a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n \\ &+ a_{22}x_2^2 + \cdots + 2a_{2n}x_2x_n \\ &\cdots \\ &+ a_{nn}x_n^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j, a_{ij} = a_{ji} \end{aligned} \tag{4.73}$$

称为 n 元二次型。用矩阵表示

$$f = x^T Ax, A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \tag{4.74}$$

A 为实对称矩阵，称为二次型 f 的矩阵，f 称为矩阵 A 的二次型，A 的秩称为二次型 f 的秩。

称只含有二次项的二次型为二次型的标准形，即 $f = y^T\Omega y$。可以将二次型转化为标准形。

对于 n 阶方阵 A, B，如果存在 k 阶可逆矩阵 C 使得 $C^TAC = B$，则称 A 合同于 B，记为 $A \simeq B$，满足

- 自反性、对称性、传递性
- 保秩：$A \simeq B \Rightarrow R(A) = R(B)$
- 保持对称性：$A \simeq B, A = A^T \Rightarrow B^T = T$
- 合同与相似无关。当 $PP^T = E$ 时，合同变换与相似变换一致。

二次型一定可以转化为标准形，$f = x^TAx \Rightarrow y^T\Omega y$ 有三种转换方法：

- 正交变换：$x = Py, f = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$，$\lambda_1, \lambda_2, \cdots, \lambda_n$ 为矩阵 A 的特征值。P 的求法：解 $|\lambda E - A = 0|$ 得到 A 的特征值，然后求特征向量，并单位化、正交化。P 是规范正交的单位向量组。
- 配方法。任何 n 元实二次型都可以经过可逆线性变换化成标准形，而且是合同变换。
- 合同变换（初等变换）：$(A, E) \cong (\Lambda, C), x = Cy$

惯性定律：设实二次型 $f = x^T Ax$ 的秩为 r，并有两个实可逆变换 $x = Cy, x = Pz$ 分别把二次型化为标准形 $f = k_1 y_1^2 + k_2 y_2^2 + \cdots + k_r y_r^2, f = \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_r z^2$，则 $k_1, k_2, \cdots, k_r$ 和 $\lambda_1, \lambda_2, \cdots, \lambda_r$ 中的正数（负数）个数相等，分别称为正惯性指数（负惯性指数）。

设实二次型 $f = x^T Ax$，如果对于任意的非零向量 x，都有

- $f = x^T Ax > 0(< 0)$，则 f 为正（负）定二次型，A 为正（负）定矩阵
- $f = x^T Ax \geq 0(\leq 0)$ 则 f 为半正（负）定二次型，A 为半正（负）定矩阵
- 无法判断 f 的符号，则 f 为不定二次型，A 为不定矩阵

对二次型作可逆线性变换不改变其正定性。n 元实二次型正定 $\Leftrightarrow$ 标准形中平方项的系数均大于 0。从而 $f = x^T Ax$ 正定 $\Leftrightarrow$ A 的特征值均大于 0 $\Leftrightarrow$ 正惯性指数为 n。$Rightarrow$ 正定矩阵行列式大于 0。

可以直接从二次型判断正定性：实对称矩阵 A 正定的充要条件是 A 的各阶顺序主子式均为正。实对称矩阵 A 正定的充要条件是 A 的奇数阶顺序主子式为负而偶数阶顺序主子式为正。

4.3 概率论与数理统计

概率论与数理统计是理解随机现象的工具，是从确定现象走向不确定现象，理解现实世界的理论。概率统计是计量经济学的数理基础，是经济学实证分析必不可少的工具，并且在经济理论中也占有重要地位，比如分位数、显著性。

4.3.1 随机事件与随机变量

随机现象是在一定条件下可能发生也可能不发生的现象。随机现象的重复观测过程就是试验。随机试验满足以下条件：可重复性、可观测性和随机性。试验用 E 表示，试验的基本结果用 ω 表示，称为样本点，样本点的集合为样本空间（基本空间），记为 $\Omega = \{\omega\}$。

随机事件是若干样本点的集合，是样本空间的子集，用大写字母表示。由一个样本点组成的事件为基本事件，由 Ω 组成的事件为必然事件，由 $\emptyset$ 组成的事件为不可能事件。如果样本空间的样本点有限，则称为有限样本空间。如果样本空间的样本点无限，则称为无限样本空间。

随机事件之间的关系和运算如下：

- 事件包含：A 包含于 B（B 包含 A），$A \subset B$，即若 $\omega \in A$，则 $\omega \in B$
- 事件相等：$A = B \Leftrightarrow A \subset B, B \subset A$

- 事件的并（和）：$A\cup B=\{\omega|\omega\in A或\omega\in B\}$
- 事件的交（积）：$AB=A\cap B=\{\omega|\omega\in A且\omega\in B\}$
- 事件的差：$A-B=\{\omega|\omega\in A但\omega\notin B\}$
- 互不相容事件：$A\cap B=\emptyset$
- 对立事件（互逆事件）：$\bar{A}=B\Leftrightarrow A\cap B=\emptyset, A\cup B=\Omega$
- 有限个或无穷个事件的并与交：$\cup_{i=1}^{n}A_i, \cap_{i=1}^{n}A_i, \cup_{i=1}^{\infty}A_i, \cap_{i=1}^{\infty}A_i$
- 完备事件组（（有限）分割）：事件 $A_1, A_2, \cdots, A_n, \cdots$ 满足（1）$A_iA_j=\emptyset, i\neq j, i,j=1,2,\cdots$；（2）$\cup_i A_i=\Omega$
- 事件的关系与运算用文氏图表示

事件的运算满足以下法则（对比集合）：

- 交换律：$A\cup B=B\cup A, A\cap B=B\cap A$
- 结合律：$(A\cup B)\cup C=A\cup(B\cup C), (AB)C=A(BC)$
- 分配律：$(A\cup B)\cap C=(A\cap C)\cup(B\cap C), (A\cap B)\cup C=(A\cup C)\cap(B\cup C)$
- De Morgan 定律（对偶律）：$\overline{\cup_{i=1}^{n}A_i}=\cap_{i=1}^{n}\bar{A}_i, \overline{\cap_{i=1}^{n}A_i}=\cup_{i=1}^{n}\bar{A}_i$

设在相同条件下进行的 n 次试验中，事件 A 发生了 n_A 次，则称 n_A 为事件 A 的频数，称 $\frac{n_A}{n}$ 为事件 A 发生的频率，记为 $f_n(A)$，即 $f_n(A)=\frac{n_A}{n}$。频率具有以下性质：

- 有界性：$0\leq f_n(A)\leq 1$
- 规范性：$f_n(\Omega)=1$
- 有限可加性：若 $A_1, A_2, \cdots, A_m$ 是两两互不相容的事件，则 $f_n(A_1\cup A_2\cup\cdots\cup A_m)=f_n(A_1)+f_n(A_2)+\cdots+f_n(A_m)$

在相同条件下重复进行 n 次试验，如果当 n 增大时，事件 A 的频率 $\frac{n_A}{n}$ 稳定地在某一常数 p 附近摆动，则称常数 p 为事件 A 的概率，记为 $P(A)=p$。概率的严格定理如下：设随机试验 E 的样本空间为 Ω，如果对于 Ω 中的每一个事件 A，有惟一的实数 $P(A)$ 与之对应，且这一事件的函数 $P(A)$ 满足：

- 非负性：$P(A)\geq 0$
- 规范性：$P(\Omega)=1$
- 可列可加性：对于两两不相容的事件 $A_1, A_2, \cdots$ 有 $P(\cup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)$

概率有如下性质：

- 对于不可能事件：$P(\emptyset)=0$
- 有限可加性：对于两两互不相容事件 $A_1, A_2, \cdots, A_n$，有 $P(\cup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)$
- 对于任意事件都有，$P(\bar{A})=1-P(A)$
- 如果 $A\subset B$，则 $P(B-A)=P(B)-P(A), P(A)\leq P(B)$
- 对于任意事件都有，$P(A)\leq 1$
- 减法公式：$P(B-A)=P(B)-P(AB)$
- 加法公式：$P(A\cup B)=P(A)+P(B)-P(AB), P(A\cup B)\leq P(A)+P(B)$
- 一般加法公式：$P(\cup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leq i<j\leq n}P(A_iA_j)+\sum_{1\leq i<j<k\leq n}P(A_iA_jA_k)+\cdots+(-1)^{(n-1)}P(A_1A_2\cdots A_n)$

如果随机事件：（1）只有有限个可能结果；（2）每种结果发生的可能性相同，则这种试验为**古典概型**。设试验 E 为古典概型，样本空间为 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_n\}$，则基本事件是完备事件组，且 $P(\omega_i)=\frac{1}{n}, i=1,2,\cdots,n$，如果事件 A 中包含了 r 个基本事件，则 $P(A)=\frac{r}{n}$。古典概型有三种应用：摸球问题、分房问题和随机取数问题等。

古典概型是离散事件，与之相对的连续时间的试验为**几何概型**，对于任意有度量的子区域 $A\subset\Omega$，定义事件 A 为任取一点落在区域 A 中，其概率为 $P(A)=\frac{A\text{的度量}}{\Omega\text{的度量}}$，其应用包括会面问题、Buffon 投针问题等。

设 A,B 是试验 E 中的两个事件，且 $P(A)>0$，称 $P(AB)/P(A)$ 为在事件 A 已经发生的情况下事件 B 发生的**条件概率**，记为 $P(B|A)$，即 $P(B|A)=\frac{P(AB)}{P(A)}$，因而 $P(AB)=P(A)P(B|A)$（乘法公式），可以推广为

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{4.75}$$

条件概率具有如下性质：

- 非负性：$P(B|A)\geq 0$
- 规范性：$P(\Omega|A)=1$
- 可列可加性：对于两两不相容的事件 $B_1,B_2,\cdots,B_n,\cdots$ 有 $P(\cup_{i=1}^{\infty}B_i|A)=\sum_{i=1}^{\infty}P(B_i|A)$

如果事件 $A_1,A_2,\cdots,A_n$ 为样本空间 Ω 的一个（有限）完备事件组或分割，且 $P(A_i)>0, i=1,2,\cdots,n$，则对于任意事件 B，

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i) \tag{4.76}$$

这就是**全概率公式**。全概率公式的应用借助概率树。根据全概率公式和条件概率公式可推导出**贝叶斯公式**，

$$P(A_j|B)=\frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{n}P(A_i)P(B|A_i)}, j=1,2,\cdots,n \tag{4.77}$$

设 A,B 是同一试验 E 的链各个事件，如果

$$P(AB)=P(A)P(B) \tag{4.78}$$

则事件 A,B 是**相互独立**的，可以推导 $P(AB)=P(A)P(B)\Leftrightarrow P(A|B)=P(A), P(B)>0\Leftrightarrow P(A\bar{B})=P(A)P(\bar{B})$。对于同一试验 E 的三个事件 A,B,C，如果

$$P(AB)=P(A)P(B), P(BC)=P(B)P(C), P(AC)=P(A)P(C) \tag{4.79}$$

则称三个事件 A,B,C **两两独立**。如果 $P(ABC)=P(A)P(B)P(C)$ 也成立，则事件 A,B,C **相互独立**。

一般的，设 $A_1,A_2,\cdots,A_n$ 是同一试验 E 中的 n 个事件，如果对于任意正整数 k 及这 n 个事件中的任意 $k, 2\leq k\leq n$ 个事件 $A_{i1},A_{i2},\cdots,A_{ik}$ 都有 $P(A_{i1}A_{i2}\cdots A_{ik})=P(A_{i1})P(A_{i2})\cdots P(A_{ik})$ 成立，则事件 $A_1,A_2,\cdots,A_n$ 是**相互独立**的。

如果事件 E 只有两种结果 $A,\bar{A}$，且 $P(A)=p, P(\bar{A})=1-p, 0<p<1$，则将试验 E 独立重复进行 n 次所构成的一串试验称为 n 重 Bernoulli 试验，称为伯努利概型。在 n 次独立重

复试验中，事件 A 恰好发生 $k, 0 \leq k \leq n$ 次的概率为

$$P_n(k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \cdots, n \tag{4.80}$$

实际上

$$[p + (1-p)]^n = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \tag{4.81}$$

因而 $P_n(k)$ 公式又称为二项概率公式。

4.3.2 多维随机变量与概率分布

设随机试验 E 的样本空间为 $\Omega = \{\omega\}$，如果对于每一个 $\omega \in \Omega$，都有一个实数 $X(\omega)$ 与之对应，则称 $X = X(\omega), \omega \in \Omega$ 为随机变量，记为 X。随机变量用大写拉丁字母或小写希腊字母表示。

设 X 是一个随机变量，对于任意实数 x，函数

$$F(x) = P(X \leq x), -\infty < x < +\infty \tag{4.82}$$

称为 X 的分布函数。分布函数性质如下：

- 有界性：$0 \leq F(x) \leq 1, \forall x \in R$
- 单调性：$\forall x_1 < x_2, F(x_1) \leq F(x_2)$
- 有极限：$F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0, F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$
- 连续性：$\lim\limits_{x \to x^+} F(x) = F(x_0), x_0 \in R$（右连续）

可以用分布函数表示任意事件的概率。如果随机变量 X 的可能取值是有限个或可数无穷个，并且以确定的概率取不同的值，则 X 为离散型随机变量。设所有可能取值为 $x_k, k = 1, 2, \cdots$，事件 $\{X = x_k\}$ 的概率为

$$P\{X = x_k\} = p_k, k = 1, 2, \cdots \tag{4.83}$$

并且 p_k 满足：$p_k \geq 0, k = 1, 2, \cdots, \sum_{k=1}^{\infty} p_k = 1$。则称上式为离散型随机变量 X 的概率分布，也称为分布律。概率分布有三种表示方法：

- 列表法：

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

- 解析式法：$\{X \leq x\} = \cup_{x_k \leq x} \{X = x_k\}$
- 矩阵法：$\begin{bmatrix} X & x_1 & x_2 & \cdots & x_n & \cdots \\ P & p_1 & p_2 & \cdots & p_n & \cdots \end{bmatrix}$

离散型随机变量的分布函数为

$$F(x) = P\{X \leq x\} = \sum_{x_k \leq x} P\{X = x_k\} = \sum_{x_k \leq x} p_k \tag{4.84}$$

常用离散型随机变量如下：

- 离散型均匀分布：$P\{X = x_k\} = \frac{1}{n}, k = 1, 2, \cdots, n, x_i \neq x_j, i \neq j$
- 0-1 分布（两点分布）：$P\{X = k\} = p^k (1-p)^{1-k}, 0 < p < 1, k = 0, 1$
- 二项分布：$P\{X = k\} = C_n^k p^k (1-p)^{n-k}, k = 1, 2, \cdots, n, 0 < p < 1$

- 几何分布：$X \sim G(p), P\{X=k\}=(1-p)^{k-1}p \Rightarrow P\{X>n+m|X>n\}=P\{X>m\}$（记忆性）
- Poisson 分布：$X \sim \pi(\lambda)(P(\lambda)), P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}, k=0,1,2,\cdots,\lambda>0$

可以用泊松分布近似二项分布：（泊松定理）设 $\lambda>0$ 为常数，n 为任意正整数，$np_n=\lambda$，则对任意固定非负整数 k，有

$$\lim_{n\to\infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!} \tag{4.85}$$

设随机变量 X 的分布函数 $F(x)$，若存在非负可积函数 $f(x)$，使得对任意的实数 x 有

$$F(x)=\int_{-\infty}^{x} f(t)dt \tag{4.86}$$

则称 X 为连续型随机变量，其中 $f(x)$ 称为概率密度函数，简称概率密度，记为 $X\sim f(x)$。性质如下：

- $f(x)\geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx=1$
- $\forall a<b, P\{a<x\leq b\}=F(b)-F(a)=\int_a^b f(x)dx$
- X 的分布函数 $F(x)$ 处处连续
- $\forall a\in R, P\{X=a\}=0$
- 可导性：$F'(x)=f(x)$

常用分布如下：

- 均匀分布：$X\sim U(a,b), f(x)=\begin{cases}\frac{1}{b-a}, 0<x<b \\ 0, otherwise\end{cases}, F(x)=\begin{cases}0, x\leq a \\ \frac{x-a}{b-a}, a<x<b \\ 1, x\geq b\end{cases}$
- 指数分布：$X\sim E(\theta), \theta>0, f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, x\geq 0 \\ 0, x<0\end{cases}, F(x)=\begin{cases}1-e^{-\frac{x}{\theta}}, x\geq 0 \\ 0, x<0\end{cases}$

正态分布是最重要的分布。设随机变量 X 具有概率密度

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<+\infty, \mu,\sigma>0 \tag{4.87}$$

则称 X 服从参数为 μ,σ^2 的正态分布，记为 $X\sim N(\mu,\sigma^2)$。其分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt, -\infty<x<+\infty \tag{4.88}$$

设 $X\sim N(\mu,\sigma^2)$，如果 $\mu=0,\sigma=1$，则 X 服从标准正态分布，记为 $X\sim N(0,1)$，其概率密度函数和分布函数分别为

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt, -\infty<x<+\infty \tag{4.89}$$

根据正态分布的对称性，可得

- $\Phi(-x)+\Phi(x)=1$
- $P\{|X|>a\}=2[1-\Phi(a)]$
- $P\{|X|<a\}=2\Phi(a)-1$

任意正态分布的分布函数可用 $\Phi(x)$ 表示，即 $X \sim N(\mu,\sigma^2), F(X)=\phi(\frac{x-\mu}{\sigma})$。设 $X \sim N(0,1)$，对于给定的 $\alpha, 0<\alpha<1$，如果 u_α 满足

$$P\{X \geq u_\alpha\} = \frac{1}{\sqrt{2\pi}} \int_{u\alpha}^{+\infty} e^{-\frac{x^2}{2}} dx = \alpha \tag{4.90}$$

则称点 u_α 为标准正态分布的上 α 分位点，且 $\Phi(u_\alpha)=1-\alpha, u_{1-\alpha}=-u_\alpha$。

基础分布不能完全满足实践要求，这时需要求随机变量的函数的分布。对于离散型随机变量的函数，设 X 的概率密度为 $P\{X=x_k\}=p_k, k=1,2,\cdots$，$y=g(x)$ 是连续函数，则对于 X 的函数 $Y=g(X)$，有 $P\{Y=g(x_k)\}=p_k, k=1,2,\cdots$，然后需要加总相同的 Y 的概率密度，得到 $Y=g(X)$ 的概率分布。

设 X,Y 为连续型随机变量，$Y=g(X)$，则有两种方法求 Y 的概率密度

1. 分布函数法：$F_Y(y)=P\{Y\leq y\}=P\{g(Y)\leq y\}=P\{X\in S\}, S=\{x|g(x)\leq y\}$，然后对 y 求导，$F_Y(y)=\begin{cases} \frac{dF_Y(y)}{dy}, \text{可导处} \\ 0, \text{不可导处} \end{cases}$
2. 公式法：设 X 为连续性随机变量，概率密度为 $f_X(x)$，函数 $y=g(x)$ 处处可导严格单调，其反函数 $h(y)$ 有连续导数，则 $Y=g(X)$ 的概率密度为 $f_Y(y)=\begin{cases} f_X[h(y)]|h'(y)|, \alpha<y<\beta \\ 0, otherwise \end{cases}$，其中 $\alpha=min\{g(-\infty),g(+\infty)\}, \beta=max\{g(-\infty),g(+\infty)\}$。

接下来介绍多维随机变量及其概率分布。设随机试验 E 的样本空间为 Ω，X,Y 是定义在 Ω 上的两个随机变量，由他们构成的向量 (X,Y) 为二维随机变量。如果对于任意实数 x,y，记事件 $\{X\leq x\},\{Y\leq y\}$ 的交为 $\{X\leq x, Y\leq y\}$，称二元函数

$$F(x,y)=P\{X\leq x, Y\leq y\} \tag{4.91}$$

为二维随机变量 (X,Y) 的分布函数，或称为联合分布函数。分布函数性质如下：

- 有界性、有极限：$0\leq F(x,y)\leq 1, F(x,-\infty)=\lim\limits_{y\to-\infty}F(x,y)=0, F(-infty,y)=\lim\limits_{x\to-\infty}F(x,y)=0, F(-\infty,-\infty)=0, F(+\infty,+\infty)=0$
- 单调不减性：$\forall x_1<x_2, F(x_1,y)\leq F(x_2,y), \forall y_1<y_2, F(x,y_1)\leq F(x,y_2)$
- 右连续性：$F(x,y)=F(x+0,y), F(x,y)=F(x,y+0)$
- 封闭区域：$P\{x_1<X\leq x_2, y_1<y\leq y_2\}=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\geq 0$

如果 (X,Y) 的取值是有限对或无穷可列对，则 (X,Y) 是二维离散型随机变量。设 (X,Y) 所有可能取值为 $(x_i,y_j), i,j=1,2,\cdots$，记事件 $\{X=x_i, Y=y_j\}$，如果 $P\{X=x_i,Y=y_j\}=p_{ij}, i,j=1,2,\cdots$，则由概率的定义：$p_{ij}\geq 0, i,j=1,2,\cdots; \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}p_{ij}=1$，满足以上条件的 p_{ij} 为 (X,Y) 的概率分布或分布律，也叫做联合概率分布或联合分布律。其分布函数为

$$F(x,y)=P\{X\leq x, Y\leq y\}=\sum_{x_i\leq x}\sum_{y_j\leq y}p_{ij} \tag{4.92}$$

如果二维随机变量 (X,Y) 的分布函数为 $F(x,y)$，如果非负二元函数 $f(x,y)$ 使得 $\forall x,y$

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv \tag{4.93}$$

则称 (X,Y) 为二维连续型随机变量，$f(x,y)$ 为其概率密度，或称为联合概率密度。其具

有如下性质：

- $f(x,y) \geq 0$
- $\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy = 1$
- 如果 $f(x,y)$ 在 (x,y) 处连续，则 $f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$
- 设 G 是 xOy 平面的一个区域，则 $P\{(X,Y) \in G\} = \iint\limits_{G} f(x,y)dxdy$

均匀分布和正态分布是常用的二维分布。设 D 是 xOy 平面的有界区域，其面积为 A，如果二维随机变量 (X,Y) 具有概率密度

$$f(x,y) = \begin{cases} 1/A, (x,y) \in D \\ 0, otherwise \end{cases} \tag{4.94}$$

则称 (X,Y) 在 D 上服从均匀分布。设随机变量 (X,Y) 的概率密度为

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}]\} \tag{4.95}$$

其中 $-\infty < x, y < +\infty, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 都是常数，且 $\sigma_1, \sigma_2 > 0, -1 < \rho < 1$，则称 (X,Y) 服从参数为 $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$ 的**二维正态分布**，记为 $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$。

已知二维随机变量 (X,Y) 的分布函数 $F(x,y)$，则有

$$F_X(x) = P\{X \leq x\} = P\{X \leq x, Y \leq +\infty\} = F_X(x) = F(x, +\infty) = \lim_{y \to +\infty} F(x,y) \tag{4.96}$$

称 $F_X(x)$ 为二维随机变量 (X,Y) 关于 X 的**边缘分布函数**，同样可以定义 $F_Y(y)$。随机变量的边缘概率密度（边缘分布律）求法如下：

- 离散型随机变量：$P\{X = x_i\} = \sum_{j=1}^{\infty} P\{X = x_i, Y = y_j\} = \sum_{j=1}^{\infty} p_{ij} = p_{i\cdot}, i = 1, 2, \cdots$
- 连续型随机变量：$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy$

给定 (X,Y) 的联合分布函数和边缘分布函数，如果

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\} \Leftrightarrow F(x,y) = F_X(x)F_Y(y) \tag{4.97}$$

则称随机变量 X, Y **相互独立**，同时

$$F(x,y) = F_X(x)F_Y(y) \Leftrightarrow \begin{cases} p_{ij} = p_{i\cdot}p_{\cdot j}, i, j = 1, 2, \cdots \\ f(x,y) = f_X(x)f_Y(y) \end{cases} \tag{4.98}$$

二维正态分布的边缘分布是正态分布，其中两个随机变量独立的充要条件是 $\rho = 0$。

对于固定的 $j, p_{\cdot j} > 0$，则在事件 $\{Y = y_j\}$ 已经发生的条件下，事件 $\{X = x_i\}$ 发生的条件概率为

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_{\cdot j}}, i = 1, 2, \cdots \tag{4.99}$$

称其为在 $Y = y_j$ 条件下随机变量 X 的**条件概率分布**或**条件分布律**，此时 $P\{X = x_i | Y = y_j\} > 0, \sum_{i=1}^{\infty} P\{X = x_i | Y = y_j\} = 1$。设二维连续随机变量 (X,Y) 的分布函数为 $F(x,y)$，概率密度为 $f(x,y)$，边缘概率密度为 $f_X(x), f_Y(y)$，则在 $Y = y$ 的条件下 X 的**条件概率密度**为

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} \tag{4.100}$$

如果一个随机变量是两个随机变量的函数，如 $Z = f(X, Y)$，则其概率分布求解方法如下。首先是离散型随机变量，(X, Y) 的概率分布为 $P\{X = x_i, Y = y_j\} = p_{ij}, i, j = 1, 2, \cdots$，则 $Z = g(X, Y)$ 也是离散型随机变量，概率分布为

$$P\{Z = z_k\} = P\{g(X, Y) = z_k\} = \sum_{g(x_i, y_j) = z_k} p_{ij}, k = 1, 2, \cdots \tag{4.101}$$

对于连续性随机变量，

$$F_Z(z) = P\{Z \leq z\} = P\{g(X, Y) \leq z\} = \iint_{g(x,y) \leq z} f(x, y)dxdy, f_Z(z) = F_Z'(z) \tag{4.102}$$

对于以下情况，$F_Z(z)$ 的表达式非常简洁：

- $Z = X + Y, f_Z(z) = f_X * f_Y = \int_{-\infty}^{+\infty} f_X(x)f_Y(z - x)dx = \int_{-\infty}^{+\infty} f_X(z - y)f_Y(y)dy$（卷积公式）
- X, Y 相互独立，$M = max\{X, Y\}, F_M(z) = F_X(z)F_Y(z)$
- X, Y 相互独立，$N = min\{X, Y\} = 1 - [1 - F_X(z)][1 - F_Y(z)]$

4.3.3 随机变量数字特征、大数定理与中心极限定理

随机变量数字特征包括期望、方差、协方差、相关系数、矩。设离散随机变量 X 的概率分布为 $P\{X = x_k\} = p_k, k = 1, 2, \cdots$，如果无穷级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛，则称无穷级数 $\sum_{k=1}^{\infty}$ 的和为离散型随机变量 X 的数学期望或均值，记为

$$E(X) = \sum_{k=1}^{\infty} x_k p_k \tag{4.103}$$

设连续型随机变量 X 的概率密度为 $f(x)$，如果广义积分 $\int_{-\infty}^{+\infty} xf(x)dx$ 绝对收敛，则称广义积分 $\int_{-\infty}^{+\infty} xf(x)dx$ 的值为连续型随机变量 X 的数学期望或均值，即

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \tag{4.104}$$

设随机变量 Y 是随机变量 X 的函数，$Y = g(X)$，其中 g 是一元连续函数，如果 X 是离散型随机变量，概率分布为 $P\{X = x_k\} = p_k, k = 1, 2, \cdots$，如果无穷级数 $\sum_{k=1}^{\infty} g(x_k)p_k$ 绝对收敛，则随机变量 Y 的数学期望为

$$E(Y) = E[g(X)] = \sum_{k=1}^{\infty} g(x_k)p_k \tag{4.105}$$

设 X 是连续型随机变量，其概率密度为 $f(x)$，如果广义积分 $\int_{-\infty}^{+\infty} g(x)f(x)dx$ 绝对收敛，则随机变量 Y 的数学期望为

$$E(Y) = E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx \tag{4.106}$$

上述结论可以推广到多个随机变量的函数的概率分布。数学期望具有如下性质：

- 常函数期望：$E(C) = C$
- 倍法性：$E(CX) = CE(X)$
- 线性性：$E(X + Y) = E(X) + E(Y)$

- 独立可分离性：X,Y 相互独立，则 $E(XY)=E(X)E(Y)$
- Cauchy-Schwarz 不等式：$[E(XY)]^2\le E(X^2)E(Y^2)$

设 X 是一个随机变量，如果 $E\{[X-E(X)]^2\}$ 存在，则称之为随机变量 X 的方差，记为 $D(X)$ 或 $var(X)$，

$$D(X)=E\{[X-E(X)]^2\} \tag{4.107}$$

称 $\sqrt{D(X)}$ 为随机变量 X 的标准差或均方差，记为

$$\sigma(X)=\sqrt{D(X)} \tag{4.108}$$

方差度量随机变量与数学期望之间的偏离程度。方差和数学期望的关系为

$$D(X)=E(X^2)-[E(X)]^2 \tag{4.109}$$

方差性质如下：

- $D(C)=0$，$D(X)=0\Leftrightarrow P\{X=C\}=1$
- $D(CX)=C^2D(X)$
- $D(C+X)=D(X)$
- X,Y 相互独立，则 $D(X+Y)=D(X)+D(Y)$

常用分布的均值和方差如下：

	$E(X)$	$D(X)$
两点分布：$X\sim(0-1)$	p	$p(1-p)$
二项分布：$X\sim B(n,p)$	np	$np(1-p)$
几何分布：$X\sim G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
均匀分布：$X\sim U(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布：$X\sim E(\theta)$	θ	θ^2
泊松分布：$X\sim\pi(\lambda)$	λ	λ
正态分布：$X\sim N(\mu,\sigma^2)$	μ	σ^2

随机变量的标准化是将随机变量化为均值为 0，方差为 1 的随机变量，

$$X^*=\frac{X-\mu}{\sigma},E(X^*)=0,D(X^*)=1 \tag{4.110}$$

设随机变量 X,Y 的数学期望为 $E(X),E(Y)$ 存在，如果 $E\{[X-E(X)][Y-E(Y)]\}$ 存在，则称之为随机变量 X,Y 的*协方差*，记为

$$cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}=E(XY)-E(X)E(Y) \tag{4.111}$$

协方差性质如下：

- 对称性：$cov(X,Y)=cov(Y,X)$
- 线性：$Cov(aX,bY)=abCov(X,Y),a,b$ 为常数
- 分配性：$cov(X+Y,Z)=cov(X,Z)+cov(Y,Z)$

设随机变量 X,Y 的方差都存在且不等于零，协方差 $cov(X,Y)$ 存在，则称 $\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$

为随机变量 X, Y 的相关系数，记为

$$\rho_{xy} = \frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{4.112}$$

当 $\rho_{xy} = 0$，称 X, Y 不相关。相关系数性质如下：

- 有界性：$|\rho_{xy}| \leq 1$
- 独立不相关：如果 X, Y 相互独立，则 $\rho = 0$
- 完全相关性：$|\rho_{xy}| = 1 \Leftrightarrow \exists a, b P\{Y = a + bX\} = 1$

随机变量 X, Y 不相关有如下等价形式:X, Y 不相关 $\Leftrightarrow cov(X,Y) \Leftrightarrow E(XY) = E(X)E(Y) \Leftrightarrow D(X+Y) = D(X)+D(Y)$。特别的，对于正态分布，$X, Y$ 相互独立 $\Leftrightarrow \rho = 0 \Leftrightarrow X, Y$ 不相关。

设 X, Y 是随机变量，如果

$$E(X^k), k = 1, 2, \cdots \tag{4.113}$$

存在，则称之为随机变量 X 的k 阶原点矩。如果

$$E\{[X - E(X)]^k\}, k = 1, 2\cdots \tag{4.114}$$

存在，则称之为随机变量 X 的k 阶中心矩。如果

$$E(X^kY^l), k, l = 1, 2, \cdots \tag{4.115}$$

则称之为随机变量 X, Y 的 $k + l$ 阶混合原点矩。如果

$$E\{[X - E(X)]^k[Y - E(Y)]^l\}, k, l = 1, 2, \cdots \tag{4.116}$$

则称之为随机变量 X, Y 的 $k + l$ 阶混合中心矩。

设二维随机变量 (X_1, X_2) 关于 X_1, X_2 的二阶中心矩和二阶混合中心矩

$$C_{ij} = E\{[X_i - E(X_i)][Y_j - E(Y_j)]\}, i, j = 1, 2 \tag{4.117}$$

都存在，则称矩阵

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{4.118}$$

为二维随机变量 (X_1, X_2) 的协方差矩阵，它是对称矩阵。同样，可以定义 n 维随机变量的协方差矩阵。

概率论中用来阐述大量随机现象的平均结果的稳定性的理论称为大数定律。设随机变量 X 具有数学期望 $E(X) = \mu$ 和方差 $D(X) = \sigma^2$，则对于任意给定的正数 ε，Chebyshev 不等式成立：

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2} \tag{4.119}$$

等价形式为

$$P\{|X - \mu| \leq \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2} \tag{4.120}$$

设 $X_1, X_2, \cdots, X_n, \cdots$ 是一个随机变量序列，a 是一个常数，如果对于任意给定的正数 ε，有

$$\lim_{n\to\infty} P\{|X_n - a| \leq \varepsilon\} = 1 \tag{4.121}$$

则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 依概率收敛于 a，记为 $X_n \xrightarrow{P} a$。设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，函数 $g(x,y)$ 在点 (a,b) 处连续，则

$$g(X_n, Y_n) \xrightarrow{P} g(a,b) \tag{4.122}$$

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，分布具有期望 $E(X_1), E(X_2), \cdots, E(X_n), \cdots$ 和方差 $D(X_1), D(X_2), \cdots, D(X_n), \cdots$，并且方差一致有上界，即存在正数 M 使得 $D(X_n) \leq M, n = 1, 2, \cdots$，则对于任意给定的正数 ε，恒有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)| \leq \varepsilon\} = 1 \tag{4.123}$$

这就是切比雪夫定理。如果上述随机变量相互独立，同时均值都等于 μ，方差都等于 σ^2，则当 $n \to \infty$ 时算术平均值 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 依概率收敛于数学期望 μ。

（Bernoulli 定理）设 n_A 是在 n 次独立重复试验中的事件 A 发生的次数，p 是事件 A 在每次试验中发生的概率，则对于任意给定的正数 ε，有

$$\lim_{n\to\infty} P\{|\frac{n_A}{n} - p| \leq \varepsilon\} = 1 \tag{4.124}$$

（Khintchine 定理）设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k) = \mu, k = 1, 2, \cdots$，则对于任意给定的正数 ε，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu|\varepsilon\} = 1 \tag{4.125}$$

Bernoulli 定理是 Khintchine 定理的特殊情况。最后介绍中心极限定理。

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 的分布函数依次为 $F_1(x), F_2(x), \cdots, F_n(X), \cdots$，如果对于 $F(x)$ 的每一个连续点 x，都有 $\lim\limits_{n\to\infty} F_n(x) = F(x)$，则称随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 依分布收敛于 X，记为 $X_n \xrightarrow{L} X$。

（Levy-Lindberg 定理，也称独立同分布的中心极限定理）设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k) = \mu$ 和方差 $D(X_k) = \sigma^2 \neq 0, k = 1, 2, \cdots$，随机变量 $Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$ 的分布函数为 $F_n(x)$，则

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{4.126}$$

即 $Y_n \xrightarrow{L} u, u \sim N(0,1)$。特别的，（DeMoivre-Laplace）对于二项分布 $Y_n \sim B(n,p), n = 1, 2, \cdots$，对于任意实数 x，有

$$\lim_{n\to\infty} P\{\frac{Y_n - np}{\sqrt{np(1-p)} \leq x}\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{4.127}$$

即正态分布是二项分布的极限分布。（Liapunov 定理）设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有数学期望和方差 $E(X_k) = \mu_k, D(X_k) = \sigma_k^2 \neq 0, k = 1, 2, \cdots$，记 $B_n^2 = \sum_{k=1}^{n} \sigma_k^2$，设随机变量 $Z_n = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$，其分布函数为 $F_n(x)$，如果存在正数 δ，使得当 $n \to \infty$ 时，有

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^{n} E\{|X_k - \mu_k|^{2+\delta}\} \to 0 \tag{4.128}$$

则对任意实数 x，恒有

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \tag{4.129}$$

4.3.4 数理统计基础

数理统计是以概率论为基础，根据试验或观测到的数据，运用有效方法对数据进行整理、分析和判断，从而对研究对象的性质和统计规律作出合理和科学的判断。

研究对象的全体称为总体，总体中的每个元素称为个体，个体总数为总体容量，根据总体容量是否有限，可以将总体分为有限总体和无限总体。从总体中抽取的待测个体组成的集合称为样本，样本包含的个体数目称为样本容量，样本记为 $X_1, X_2, \cdots, X_n$，这里每个 X_i 都是随机变量。样本的确切值 $x_1, x_2, \cdots, x_n$ 称为样本观测值，也称样本值。样本 $X_1, X_2, \cdots, X_n$ 的所有可能取值的全体称为样本空间，观测值 $x_1, x_2, \cdots, x_n$ 是样本空间中的一个点。

抽样的方式会影响样本的特征，最基本的样本为简单随机样本（默认），满足以下两个条件：

1. 独立性：$X_1, X_2, \cdots, X_n$ 相互独立
2. 随机性：样本 $X_1, X_2, \cdots, X_n$ 与总体具有相同的分布

如果总体 X 的分布函数为 $F(x)$，则样本 $X_1, X_2, \cdots, X_n$ 的联合分布函数为

$$F(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} F(x_i) \tag{4.130}$$

如果总体为离散型随机变量，概率分布为 $P\{X = x\} = p(x)$，则样本 $X_1, X_2, \cdots, X_n$ 的联合概率分布为

$$P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\} = \prod_{i=1}^{n} p(x_i) \tag{4.131}$$

如果总体为连续型随机变量，概率密度为 $f(x)$，则样本 $X_1, X_2, \cdots, X_n$ 的联合概率分布为

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i) \tag{4.132}$$

设总体的分布函数为 $F(x)$，从总体中抽取容量为 n 个样本，观测值为 $x_1, x_2, \cdots, x_n$，假如其中有 k 个不同的值，依次排序为 $x_{(1)<x_{(2)}} < \cdots < x_{(k)}$，每个 $x_{(i)}$ 出现的频率为 $f_i = n_i/n$，则 $\sum_{i=1}^{k} f_i = 1$，设函数

$$F_n(x) = \begin{cases} 0, x < x_{(1)} \\ \sum_{j=1}^{i} f_j, x_{(i)} \leq x < x_{(i+1)}, i = 1, 2, \cdots, k-1 \\ 1x \geq x_{(k)} \end{cases} \tag{4.133}$$

称 $F_n(x)$ 为样本分布函数或经验分布函数，具有如下性质：

- 规范性：$0 \leq F_n(x) \leq 1$
- 非减性

- 有极限：$F_n(-\infty)=0, F_n(+\infty)=1$
- 连续性：$F_n(x)$ 在每个 $x_{(i)}$ 处右连续，点 $x_{(i)}$ 是其跳跃间断点

（W.Glivenko 定理）当 $n\to\infty$，样本分布函数 $F_n(x)$ 依概率 1 关于 x 均匀收敛于总体分布函数 $F(x)$，即

$$P\{\lim_{n\to\infty}\sup_{-\infty<x<+\infty}|F_n(x)-F(x)|=0\}=1 \tag{4.134}$$

这里 sup 是上确界。该定理是用样本推断总体的理论依据。

设总体 X 的样本为 $X_1,X_2,\cdots,X_n$，观测值为 $x_1,x_2,\cdots,x_n$，如果 $g(t_1,t_2,\cdots,t_n)$ 为已知的 n 元函数，则称 $g(X_1,X_2,\cdots,X_n)$ 为样本函数，它是一个随机变量，称 $g(x_1,x_2,\cdots,x_n)$ 为样本函数的观测值。如果 $g(X_1,X_2,\cdots,X_n)$ 中不含有未知数，则称该样本函数为统计量。

常用统计量如下：

	统计量	观测值
样本均值	$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$	$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$
样本方差	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$	$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$
样本标准差	$S=\sqrt{S^2}$	$s=\sqrt{s^2}$
样本 k 阶原点矩	$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k, k=1,2,\cdots$	$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k, k=1,2,\cdots$
样本 k 阶中心矩	$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k, k=1,2,\cdots$	$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k, k=1,2,\cdots$
样本偏度		$\gamma_1=b_3/b_2^{\frac{3}{2}}$
样本峰度		$\gamma_2=b_4/b_2^2-3$
样本最大值	$X_{(n)}=max\{X_1,X_2,\cdots\}$	$x_{(n)}=max\{x_1,x_2,\cdots,x_n\}$
样本最小值	$X_{(1)}=min\{X_1,X_2,\cdots\}$	$x_{(1)}=min\{x_1,x_2,\cdots,x_n\}$
样本协方差	$S_{XY}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$	$s_{xy}^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$
样本相关系数	$\rho_{XY}=\frac{S_{XY}^2}{S_XS_Y}$	$\rho_{xy}=\frac{s_{xy}^2}{s_xs_y}$

最值统计量的分布函数为 $F_{max}(x)=[F(x)]^n, F_{min}(x)=1-[1-F(x)]^n$。样本二阶中心矩和样本方差的关系为 $B_2=\frac{n-1}{n}S^2$。

正态分布总体的常用统计量的分布。设 $X_1,X_2,\cdots,X_n$ 是来自标准正态分布 $N(0,1)$ 的样本，则统计量

$$\chi^2=X_1^2+X_2^2+\cdots+X_n^2 \tag{4.135}$$

服从自由度为 n 的 χ^2 分布，即 $\chi^2\sim\chi^2(n)$，其概率密度为

$$f(x)=\begin{cases}\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}, x>0\\ 0, x\le 0\end{cases} \tag{4.136}$$

其中 $\Gamma(t)=\int_0^{+\infty}x^{t-1}e^{-x}dx, t>0$ 为 Γ 函数。χ^2 分布具有如下性质：

- 统计量：$\chi^2\sim\chi^2(n), E(\chi^2)=n, D(\chi^2)=2n$
- 线性性：$X\sim\chi^2(n_1), Y\sim\chi^2(n_2), X,Y$ 相互独立，则 $X+Y\sim\chi^2(n_1+n_2)$
- 中心极限定理：$\chi^2\sim\chi^2(n), \forall x, \lim_{n\to\infty}P\{\frac{\chi^2-n}{2n}\le x\}=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$

类似正态分布，也可以定义 χ^2 分布的上 α 分位点，记为 $\chi^2_\alpha(n)$（其他分布同理）。设 $X_1, X_2, \cdot, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本，则随机变量

$$\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n) \tag{4.137}$$

同时，样本均值和方差为 $\bar{X}, S^2$，可以证明

- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
- $\bar{X}, S^2$ 相互独立

设 $X \sim N(0,1), Y \sim \chi^2(n), X, Y$ 相互独立，则称统计量

$$t = \frac{X}{\sqrt{Y/n}} \tag{4.138}$$

服从自由度为 n 的 t 分布，记为 $t \sim t(n)$，其概率密度为

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, -\infty < x < +\infty \tag{4.139}$$

t 分布关于 $x=0$ 对称，且 $\lim\limits_{n\to\infty} f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$，其统计量为 $t \sim t(n), E(t)=0, D(t)=\frac{n}{n-1}$。$t$ 分布的上 α 分位点有如下性质：$t_{1-\alpha}(n) = -t_\alpha(n)$。

正态分布样本的均值和方差可以构成 t 统计量：

$$t = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1) \tag{4.140}$$

设从两个正态总体 $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$ 中分别独立抽取样本，样本容量为 n_1, n_2，样本均值为 $\bar{X}, \bar{Y}$，样本方差为 S_1^2, S_2^2，记 $S_\omega = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$，则随机变量

$$t = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2) \tag{4.141}$$

设 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$，且 X, Y 相互独立，则称统计量

$$F = \frac{X/n_1}{Y/n_2} \tag{4.142}$$

服从第一自由度为 n_1，第二自由度为 n_2 的 F 分布，记为 $F \sim F(n_1, n_2)$，其概率密度为

$$f(x) = \begin{cases} \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}(\frac{n_1}{n_2})^{\frac{n_1}{2}}x^{\frac{n_1}{2}-1}(1+\frac{n_1}{n_2}x)^{-frac(n_1+n_2)2}, x>0 \\ 0, x \leq 0 \end{cases} \tag{4.143}$$

如果 $F \sim F(n_1, n_2)$，则 $\frac{1}{F} \sim F(n_2, n_1)$。$F$ 分布的上 α 分位点有如下性质：

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)} \tag{4.144}$$

设从两个正态总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 中分别独立抽样 $X_1, X_2, \cdots, X_n$ 和 $Y_1, Y_2, \cdots, Y_n$，则有 F 统计量

$$F = \frac{n_2}{n_1}\frac{\sigma_2^2}{\sigma_1^2}\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2} \sim F(n_1, n_2) \tag{4.145}$$

其均值和方差也可以构成 F 统计量

$$F = \frac{\sigma_2^2}{\sigma_1^2}\frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1) \tag{4.146}$$

4.3.5 参数估计与假设检验

前面的所有工作都是本节的基础，获取和处理数据的最终目的是获取有价值的信息，即关于参数的信息和假设是否成立。

参数估计包括点估计和区间估计，点估计的常用方法是矩估计和极大似然估计，区间估计需要构造已知分布的统计量。

设总体的分布类型已知，但部分参数未知，设 $X_1, X_2, \cdots, X_n$ 是总体 X 的样本，如果构造统计量 $th\hat{e}ta(X_1, X_2, \cdots, X_n)$，其观测 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 可以作为参数 θ 的估计值，则称 $th\hat{e}ta(X_1, X_2, \cdots, X_n)$ 为 θ 的点估计量，$th\hat{e}ta(x_1, x_2, \cdots, x_n)$ 为参数 θ 的点估计值。如果总体中有 r 个未知参数，则需要构造 r 个统计量作为估计量。

矩估计法的思想是，根据 Khintchine 大数定理，当 $n \to \infty$，样本的 k 阶原点矩 A_k 依概率收敛于总体 X 的 k 阶原点矩 μ_k，即可以用样本原点矩作为总体原点矩的估计量，再解方程。

如果总体 X 有 r 个未知参数，$1, 2, \cdots, r$ 阶原点矩都存在，则 $A_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_r) = E(X^k), k = 1, 2, \cdots, r$

$$\begin{cases} \mu_1(\theta_1, \theta_2, \cdots, \theta_r) = A_1 \\ \mu_2(\theta_1, \theta_2, \cdots, \theta_r) = A_2 \\ \cdots \\ \mu_r(\theta_1, \theta_2, \cdots, \theta_r) = A_r \end{cases} \Rightarrow \begin{cases} \hat{\theta}_1 = \theta_1(A_1, A_2, \cdots, A_r) \\ \hat{\theta}_2 = \theta_2(A_1, A_2, \cdots, A_r) \\ \cdots \\ \hat{\theta}_r = \theta_r(A_1, A_2, \cdots, A_r) \end{cases} \Rightarrow \begin{cases} \hat{\theta}_1 = \theta_1(a_1, a_2, \cdots, a_r) \\ \hat{\theta}_2 = \theta_2(a_1, a_2, \cdots, a_r) \\ \cdots \\ \hat{\theta}_r = \theta_r(a_1, a_2, \cdots, a_r) \end{cases} \tag{4.147}$$

这就是 θ 的矩估计量和矩估计值。矩估计量不具有唯一性。

极大似然法的思想是已知总体的概率分布和样本，则概率最大的事件在一次试验中最可能出现。定义似然函数

$$L(\theta) = \begin{cases} \prod_{i=1}^{n} p(x_i; \theta), \text{离散型随机变量} \\ \prod_{i=1}^{n} f(x_i; \theta), \text{连续型随机变量} \end{cases}, \theta \in \Theta \tag{4.148}$$

则使得 $L(\theta)$ 取得最大值的 $\hat{\theta}$ 为参数 θ 的极大似然估计值，相应的，$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 为 θ 的极大似然估计量。从而，参数估计就转化为极值问题。当 $L(\theta)$ 可导时，$L(\theta)$ 取极值的一阶条件为

$$\frac{dL(\theta)}{d\theta} = 0 \tag{4.149}$$

改方程称为似然方程，这等价于对数似然方程，如下

$$\frac{dlnL(\theta)}{d\theta} = 0 \tag{4.150}$$

在其他情况下，需要根据定义求极值。极大似然估计量是唯一的，但不一定和矩估计量一致。

未知参数的估计量很多，要选择一个最佳的估计量，有如下三条标准，设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 是未知参数 θ 的估计量，则

1. 无偏性：如果 $E(\hat{\theta})$ 存在，则 $\forall\theta \in \Theta, E(\hat{\theta}) = \theta$

2. 有效性：$\hat{\theta}_1, \hat{\theta}_2$ 都是 θ 的无偏估计量，如果 $\forall\theta \in \Theta, D(\hat{\theta}_1) \le D(\hat{\theta}_2)$ 都成立，且 $\exists\theta$ 使得上述不等号成立，则 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效
3. 一致性（相合性）：$\forall\theta \in \Theta$，当 $n \to \infty$，$\hat{\theta}_n$ 依概率收敛于 θ，即 $\forall\varepsilon > 0, \lim\limits_{n\to\infty} P\{|\hat{\theta}-\theta| < \varepsilon\} = 1$

设总体 X 的分布中含有一个未知参数 θ，$X_1, X_2, \cdots, X_n$ 是来自总体的样本，如果对于给定概率 $1-\alpha, 0 < \alpha < 1$，存在两个统计量 θ_1, θ_2 使得

$$P\{\theta_1 < \theta < \theta_2\} = 1 - \alpha \tag{4.151}$$

则把 $1-\alpha$ 称为*置信度*或*置信水平*，随机区间 (θ_1, θ_2) 称为未知参数 θ 的置信水平为 $1-\alpha$ 的*置信区间*，θ_1, θ_2 分别称为置信度为 $1-\alpha$ 的双侧置信区间的*置信下限*和*置信上限*。这种方法称为区间估计。一般情况下，总是寻找长度最短的置信区间。

求置信区间的一般步骤如下：

1. 构造已知分布的统计量：$T = T(X_1, X_2, \cdots, X_n; \theta)$
2. 根据分布和置信度构造置信区间：$(a, b) \Rightarrow P\{a < T < b\} = 1 - \alpha$
3. 反解出随机事件 $P\{a < T < b\}$ 的等价事件 $P\{\theta_1(X_1, X_2, \cdots, X_n) < \theta < \theta_2(X_1, X_2, \cdots, X_n)\} = 1-\alpha$，这里 θ_1, θ_2 不含有未知参数，(θ_1, θ_2) 为 θ 的置信度为 $1-\alpha$ 的置信区间

与正态分布相关的区间估计结果整理如表4.1.

置信区间也可以是单侧置信区间，即满足 $P\{\theta > \theta_1\} = 1-\alpha$ 或 $P\{\theta < \theta_2\} = 1-\alpha$ 的随机区间 $(\theta_1, +\infty)$ 或 $(-\infty, \theta_2)$ 是单侧置信区间，θ_1, θ_2 分别称为 θ 的置信水平为 $1-\alpha$ 的单侧置信下限或单侧置信上限。只要正确构造出统计量，单侧置信区间也容易求出。在样本容量充分大时，可以用渐进分布来构造置信区间，如两点分布 $B(1, P)$ 的样本均值 $\bar{X}$ 的渐进分布为 $N(P, \frac{P(1-P)}{n})$。

现在介绍假设检验的内容。对总体分布的类型或参数提出明确的假设，称为*原假设*或*零假设*，与原假设对立的假设为*备择假设*或*对立假设*。为了推断原假设是否成立，我们先假设原假设成立，如果推导出了小概率事件，我们就拒绝原假设，否则没有理由拒绝原假设。这个小概率记为 α，称为*显著性水平*，通常取 0.01、0.05 和 0.1。

为了检验原假设是否成立，需要根据样本构造已知分布的统计量，称之为*检验统计量*，当其取某个区域 W 中的值时，我们就拒绝原假设，则称区域 W 为*拒绝域*。在假设检验中，可能存在错判的可能：

		决策	
		接受 H_0	拒绝 H_0
真实情况	H_0 为真		弃真错误
	H_0 为假	取伪错误	

在实践中，我们一般只能控制第一类错误（弃真）的概率不超过显著性水平 α，这类检验称为*显著性检验*。

只检验总体分布中的参数的假设检验称为*参数检验*。如果原假设为 $H_0: \theta = \theta_0$，备择假设为 $H_1: \theta \neq \theta_0$，这类检验为双边检验，如果备择假设为 $H_1: \theta < \theta_0$，则称为左边检验，如果备择假设为 $H_1: \theta > \theta_0$，则称为右边检验。一般的，假设检验的步骤如下：

表 4.1: 正态分布的区间估计结果

$X \sim N(\mu, \sigma^2)$	统计量	置信区间
已知 σ^2，估计 μ	$u = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$(\bar{X} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$
未知 σ^2，估计 μ	$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$	$(\bar{X} - t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}})$
已知 μ，估计 σ^2	$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$	$(\frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{10\frac{\alpha}{2}}(n)})$
未知 μ，估计 σ^2	$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)})$
$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$	统计量	置信区间
已知 σ_1^2, σ_2^2，估计 $\mu_1 - \mu_2$	$u = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$	$(\bar{X} - \bar{Y} - L, \bar{X} - \bar{Y} + L)$ $L = u_{\frac{\alpha}{2}} \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$
未知 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 估计 $\mu_1 - \mu_2$	$t = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S_\omega \sqrt{1/n_1+1/n_2}} \sim t(n_1 + n_2 - 2)$	$(\bar{X} - \bar{Y} - L, \bar{X} - \bar{Y} + L)$ $L = t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_\omega \sqrt{1/n_1 + 1/n_2}$
未知 σ_1^2, σ_2^2，但 $n_1 = n_2 = n$ 估计 $\mu_1 - \mu_2$	$Z = X - Y, t = \frac{\bar{Z}-(\mu_1-\mu_2)}{S_Z/\sqrt{n}} \sim t(n-1)$ $S_\omega = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$	$(\bar{X} - \bar{Y} - L, \bar{X} - \bar{Y} + L)$ $L = t_{\frac{\alpha}{2}}(n-1) \frac{S_Z}{\sqrt{n}}$
已知 μ_1, μ_2，估计 $\frac{\sigma_1^2}{\sigma_2^2}$	$F = \frac{n_2}{n_1} \frac{\sigma_2^2}{\sigma_1^2} \frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2} \sim F(n_1, n_2)$	$(A \frac{1}{F_{\frac{\alpha}{2}}(n_1,n_2)}, A F_{\frac{\alpha}{2}}(n_2, n_1))$ $A = \frac{n_2 \sum_{i=1}^{n_1}(X_i-\mu_1)^2}{n_1 \sum_{j=1}^{n_2}(Y_j-\mu_2)^2}$
未知 μ_1, μ_2，估计 $\frac{\sigma_1^2}{\sigma_2^2}$	$F = \frac{\sigma_2^2}{\sigma_1^2} \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1)$	$(\frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)}, \frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}}(n_2-1,n_1-1))$

1. 提出原假设 H_0 和备择假设 H_1
2. 选择适当的统计量，在原假设 H_0 下确定统计量的分布
3. 根据显著性水平 α 和统计量分布，确定拒绝域 W
4. 根据样本值计算统计量的观测值，记为 s，如果 $s\in W$，则拒绝原假设 H_0，否则接受原假设

正态分布的参数检验和统计量整理如下：

H_0	条件	检验统计量	统计量分布
$\mu=\mu_0$	已知 σ^2	$u=\frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$	$N(0,1)$
	未知 σ^2	$t=\frac{\bar{X}-\mu_0}{S}\sqrt{n}$	$t(n-1)$
$\sigma^2=\sigma_0^2$	已知 μ	$\chi^2=\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i-\mu)^2$	$\chi^2(n)$
	未知 μ	$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2(n-1)$
$\mu_1-\mu_2=\delta$	已知 σ_1^2,σ_2^2	$u=\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$	$N(0,1)$
	未知 $\sigma_1^2=\sigma_2^2=\sigma^2$	$t=\frac{\bar{X}-\bar{Y}-\delta}{S_\omega\sqrt{1/n_1+1/n_2}}$	$t(n_1+n_2-2)$
$\sigma_1^2=\sigma_2^2$	已知 μ_1,μ_2	$F=\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}\frac{n_2}{n_1}$	$F(n_1,n_2)$
	未知 μ_1,μ_2	$F=\frac{S_1^2}{S_2^2}$	$F(n_1-1,n_2-1)$

第 5 章 微观经济学

5.1 引言

根据范里安《微观经济学：现代观点》，本文将微观经济学（以下简称微经）的研究内容分为 5 个部分，分别是消费者理论、生产者理论、价格歧视和博弈、交换和福利和公共选择理论。微经的研究内容是**稀缺性资源的最优配置**，也就是约束最优化。微经的中心就是权衡（trade off），选择的成本就是机会成本（opportunity cost），即由于选择而放弃的可能收入、商品等。微经的最基本模型（workhorse model）是供求模型。如冰淇淋市场中，需求曲线为 $D = 1 - P$，供给曲线为 $S = P$，则市场在供求相等（$D = S$）时达到均衡，即 $1 - P = P$，得到 $P = 0.5D = S = 0.5$。故市场均衡价格为 0.5，均衡数量为 0.5（市场出清，供给等于需求，厂家的产量等于消费者的需求量，等于市场交易量）。

理解了微经的研究内容和分析方法，下文将详细分析 5 个研究部分，重点介绍每个部分的核心内容和逻辑体系，并尝试将微经与经济学的其他课程联系起来，每个部分都会附上延申内容和思考题，请跟随我的思维路径，体会经济学的奇妙之处！

5.2 消费者理论

5.2.1 导言

消费者理论研究**消费者选择能够负担的最佳物品**，它涉及两方面内容，一是消费者的优化目标，二是消费选择的约束。古典理论认为消费者追求的是幸福（happiness），消费者根据**偏好**（preferences）在不同的消费束中进行选择。经济学用效用刻画幸福的程度，消费者比较消费束的效用进行选择。预算约束有两个方面，一个是市场价格，另一个是收入，调节预算约束的措施包括配给、税收和补贴。

5.2.2 偏好和效用

偏好描述消费者如何在消费束中进行选择，包括三种情况：严格偏好（$>$）、无差异（$-$）、弱偏好（$\geq$）。偏好具有三个特征：

1. 完备性：消费者对不同的消费束有明确的偏好，如要么 $A > B$，要么 $B \geq A$。
2. 传递性：如果 $A \geq B$，并且 $B \geq C$，则 $A \geq C$，严格偏好同理。
3. 不满足性：消费者追求越多越好，如果数量上 A 大于等于 B，则 A 弱偏好于 B，即 $A \geq B$。

无差异的消费束组成的曲线就是无差异曲线，如图，无差异曲线具有四个特征：

1. 不满足性，离原点越远效用越高
2. 边际效用递减，向下倾斜

3. 传递性，两条无差异曲线不交叉
4. 完备性，穿过一个消费束的无差异曲线只有一条

效用函数是效用的数学刻画，如 $U=\sqrt{x_1*x_2}$。无差异曲线是 $U=\hat{U}$ 的所有可能解的集合，如 $3=\sqrt{x_1*x_2}$。边际效用（MU）是增加一单位商品带来的效用增量，一般地有**边际效用递减规律**。效用函数的最重要特征是边际效用替代率（$MRS_{1,2}$ 代表替代一单位商品 1 所需的商品 2 的数量）

$$MRS_{1,2}=\frac{\Delta x_2}{\Delta x_1}=-\frac{MU_1}{MU_2} \tag{5.1}$$

5.2.3 预算约束

消费者选择时的首要预算约束是支出小于等于收入，即：$p_1*x_1+p_2*x_2\leq m$（以两种商品的选择为例），满足该条件的集合为预算集。如果只考虑一期，根据偏好的不满足性，消费者总是在预算集的边界（预算线）上进行选择（瓦尔拉斯法则），这里持有剩余的收入不会增加效用。预算线的斜率为：

$$MRT=\frac{\Delta x_2}{\Delta x_1}=-\frac{p_1}{p_2} \tag{5.2}$$

预算线分别与 x_1 和 x_2 轴交于 m/p_1 和 m/p_2。由此决定预算线的因素现在有两个：价格和收入。价格上升，预算线向原点转动；收入上升，预算线远离远点平移。其他的预算约束包括：

1. 配给
2. 补贴
3. 税收

记住，预算约束中唯一重要的是预算线，现在考虑以上几种预算约束如何影响预算线。首先是配给，即消费者可以免费得到一定量的某种商品，如以下的预算约束：

$$s.t.\begin{cases}p_2*x_2\leq m & x_1\leq x_0,\\ p_1*(x_1-x_0)+p_2*x_2=m & x_1>x_0\end{cases} \tag{5.3}$$

其中 x_0 是商品 1 的配给。

补贴和税收是一个硬币的两面，对预算约束是相反的影响，下面统一起来有两种情况，一是从量补贴/从量税，二是总价补贴/从价税。如果政府对商品 1 收取 s 的税收（s 小于 0 则是补贴，也可以是按比例收税），那么消费者的预算约束为：

$$(p_1+s)*x_1+p_2*x_2\leq m \tag{5.4}$$

从价税将使预算线向原点转动。而所得税导致预算线的平移：

$$p_1*x_1+p_2*x_2\leq m-tax \tag{5.5}$$

如果政府要求的税收一致，那么 $s*x_1=tax$。可以证明两种税收方式中从价税下单一消费者的福利更大。

5.2.4 消费者选择

在良态偏好下（单调、严格凸），消费者的最优选择满足三种等价形式：

$$MRS = MRT \tag{5.6}$$

$$\frac{MU_1}{MU_2} = \frac{p_1}{p_2} \tag{5.7}$$

$$\frac{MU_1}{p_1} = \frac{MU_2}{p_2} \tag{5.8}$$

最后一个公式表明，消费者的最后一单位货币花在两种商品上带来的效用一致，因此消费者没有改变现状的倾向。理由是如果 $\frac{MU_1}{p_1} < \frac{MU_2}{p_2}$，消费者可以降低 1 的消费增加 2 的消费提高效用。在良态偏好的条件下，唯一的均衡点就是最优解。因此典型的消费者选择问题可以归纳为以下的约束优化问题：

$$\begin{cases} max : u(x_1, x_2) = x_1^{\alpha} * x_2^{1-\alpha} \\ s.t. p_1 * x_1 + p_2 * x_2 = m \end{cases} \tag{5.9}$$

注意这里是等号而非不等号。

5.2.5 价格变化的效应

价格变化的效应包括**替代效应**（由于两种商品之间的交换比例发生变化引起的需求变化）和**收入效应**（由于收入变化引起的需求变化）。

$$\Delta x_1 = x_1(p_1', m) - x_1(p_1, m) \tag{5.10}$$

$$= x_1(p_1', m') - x_1(p_1, m) + x_1(p_1', m) - x_1(p_1', m') \tag{5.11}$$

$$= \Delta x_1^s + \Delta x_1^n \tag{5.12}$$

$$= \Delta x_1^s - \Delta x_1^m \tag{5.13}$$

变化率：

$$\frac{\Delta x_1}{\Delta p_1} = \frac{\Delta x_1^s}{\Delta p_1} - \frac{\Delta x_1^m}{\Delta p_1} \tag{5.14}$$

$$= \frac{\Delta x_1^s}{\Delta p_1} - \frac{\Delta x_1^m}{\Delta m} x_1 \tag{5.15}$$

斯拉茨基分解：转动 + 移动，转动时消费者的购买力不变，仍能负担起原来的最优选择，补偿需求 $\Delta m = \Delta p_1 * x_1$。希克斯分解：转动时消费者的效用不变。以下是两种分解方式示意图。

需求法则: 如果一种商品的需求随着收入增加而增加，则这种商品的需求一定随价格上升而下降。

商品按需求分类：正常品（必需品和奢侈品）随收入上升需求上升，次级品（包括吉芬品）随收入上升需求下降。

5.2.6 重要知识点和解答

1. $p_1 = 1, p_2 = 2, m = 12, u = x_1 * x_2$，消费者 A 获取的商品 1 的配给券为 δ 单位，请问在离散和连续两种情况下消费者 A 出售配给券的保留价格和交易量是多少？

2. 分析两种税收的消费者福利。

3. 分析以下效用函数下的消费者需求函数：

- $u = a * x_1 + b * x_2$
- $u = min\{a * x_1, b * x_2\}$
- $u = x_1^{\alpha} * x_2^{1-\alpha}$
- $u = (x_1^{\rho} + x_2^{\rho})(1/\rho, \rho > 0$
- $u = \sqrt{x_1} + x_2$

4. 分析完全互补、完全替代和拟线性偏好三种情况下的分解效应的符号，并计算 $u = x_1^{1/3} * x_2^{2/3}$ 情况下，$p_1 = p_2 = 1, m = 12, p_1' = 2$ 的斯拉茨基和希克斯分解结果。

5. 绘制必需品、奢侈品、次级品和吉芬品的价格变化的分解效应。

解答如下：

1. 解：首先根据效应函数求出需求函数和收入提供线。由 $u = x_1 * x_2$ 为柯布道格拉斯型效应函数，变换得 $u = x_1^{0.5} * x_2^{0.5}$，则需求函数为：

$$x_i = \frac{m}{2 * p_i}, i = 1, 2 \tag{5.16}$$

现在分析出售配给券对预算约束的影响。

首先，消费者的最优选择要么在斜预算线上，要么在角点上。出售配给券使水平的预算线上移，使斜预算线左移。因此新的斜预算线在旧的斜预算线左边，如果原最优选择在斜预算线上，则原来的选择优于移动后的任一消费束。即满足如下条件，消费者不出售配给券：

$$x_1(p_1, p_2, m + p_1 * \delta) \geq \delta \tag{5.17}$$

带入得 $\delta \leq 12$。现在考虑 $\delta > 12$，即出售配给券的情形。现在有两个待决定的变量，出售价格 p_c 和出售量 q_c。根据以上的分析，如果消费者的最优选择在斜预算线上，会停止出售配给券，理由很简单，如果消费者需要购买更多的商品 1，使用配给券的成本为 0，而出售配给券并用现金购买的成本为 $p_1 - p_c$，由于 $p_1 - p_c > 0$（否则不会有人购买配给券）因此新的最优选择总是在角点处取得。角点处的坐标为（$\delta - q_c, (m + p_c * q_c)/2$），此时的效用为:

$$u = [\delta - q_c] * [m + p_c * q_c)/2] \tag{5.18}$$

新的最优选择的效(5.18)用应该高于最初的最优选择，即 $q_c = p_c = 0$ 的(5.18)。因此

$$[\delta - q_c] * [m + p_c * q_c)/2] > \delta * m/2 \tag{5.19}$$

$$\delta * p_c * q_c > q_c * m + p_c * q_c^2 \tag{5.20}$$

$$p_c > \frac{m}{\delta - q_c} \tag{5.21}$$

由于可选的 $0 \leq q_c \leq \delta$，则 p_c 的最小值为 $\frac{m}{\delta}$，即 $\frac{12}{\delta}$。最优的交易量满足(5.18)的一阶最优

化条件，即 $q_c = min\frac{\delta * p_c - m}{2 * p_c}, 0$。综上，消费者出售配给券的决策如下：

$$q_c = \begin{cases} 0 & \text{if } \delta \leq 12 \text{ or } p_c < \frac{m}{\delta}, \\ \frac{\delta * p_c - m}{2 * p_c} & \text{if } \delta > 12 \text{ and } p_c > \frac{m}{\delta} \end{cases} \tag{5.22}$$

2. 解：征收从价税后消费者的预算线为 $(p_1+s)*x_1+p_2*x_2=m$，最优消费束为 (x_1^*, x_2^*)，征收征收所得税后消费者的预算线为 $p_1*x_1+p_2*x_2=m-tax$，最优消费束为 (x_1^{**}, x_2^{**})。如果政府征税力度一致，即 $s*x_1=tax$，则

$$(p_1+s)*x_1^*+p_2*x_2^*=mp_1*x_1^*+p_2*x_2^*=m-s*x_1^*p_1*x_1^*+p_2*x_2^*=m-tax \tag{5.23}$$

说明 (x_1^*, x_2^*) 处于征收所得税的预算线上，因此 (x_1^{**}, x_2^{**}) 弱偏好于 (x_1^*, x_2^*)，即征收所得税时消费者福利的损失更小。

3. 解：根据效用函数求需求函数。

- $u=a*x_1+b*x_2$

  角点解，$MRT=-\frac{p_1}{p_2}, MRS=-\frac{a}{b}$。消费者会选择相对便宜的商品。最优选择如下

$$x_1, x_2 = \begin{cases} x_1=\frac{m}{p_1}, x_2=0 & \text{if } \frac{a}{b} > \frac{p_1}{p_2}, \\ x_1*p_1+x_2*p_2=m & \text{if } \frac{a}{b} = \frac{p_1}{p_2}, \\ x_1=0, x_2=\frac{m}{p_2} & \frac{a}{b} < \frac{p_1}{p_2} \end{cases} \tag{5.24}$$

- $u=min\{a*x_1, b*x_2\}$ 角点解。当 $a*x_1=b*x_2$ 时取最优解，即有

$$x_1=\frac{b*m}{b*p_1+a*p_2}, x_2=\frac{a*m}{b*p_1+a*p_2} \tag{5.25}$$

- $u=x_1^{\alpha}*x_2^{1-\alpha}$ 内点解。α 为商品 1 消费占收入的份额。即 $x_1=\frac{\alpha*m}{p_1}, x_2=\frac{(1-\alpha)*m}{p_2}$
- $u=(x_1^{\rho}+x_2^{\rho})(1/\rho, \rho>0$ 内点解，首先单调变换 $u=x_1^{\rho}+x_2^{\rho}$，然后求 $MRS=-(x_1/x_2)^{1/(\rho-1)}$，由 $MRT=MRS$ 和预算约束，求出：

$$x_1=\frac{m*p_1^{\frac{1}{\rho-1}}}{p_1^{\frac{\rho}{\rho-1}}+p_2^{\frac{\rho}{\rho-1}}}, x_2=\frac{m*p_2^{\frac{1}{\rho-1}}}{p_1^{\frac{\rho}{\rho-1}}+p_2^{\frac{\rho}{\rho-1}}} \tag{5.26}$$

- $u=\sqrt{x_1}+x_2$ 内点解，首先求出 $MRS=-\frac{1}{2*\sqrt{x_1}}$，由 $MRT=MRS$ 和预算约束，求出：

$$x_1=\frac{p)1}{2*p_2}^2 \tag{5.27}$$

  则所需的收入为 $m_0=x_1*p_1=\frac{p_2^2}{4*p1}$，由于 m_0 可能小于 m，因此实际需求为：

$$x_1, x_2 = \begin{cases} x_1=\frac{m}{p_1}, x_2=0 & \text{if } \frac{p_2^2}{4*p1} \leq m, \\ x_1=\frac{p)1}{2*p_2}^2, x_2=\frac{m}{p_2}-\frac{p_2}{4*p_1} & \text{if } \frac{p_2^2}{4*p1} > m \end{cases} \tag{5.28}$$

  特别要注意这里不满足预算约束的最优解的情况！

4. 解：在斯拉茨基分解中，几种效用函数情况下的替代效应和收入效应符号：

	替代效应	收入效应
完全互补	0	$\pm$
完全替代	$\pm$	0
拟线性	$\pm$	0

现在计算斯拉茨基分解和希克斯分解效应。首先由需求函数计算总效应，

$x_1(p_1, m) = \frac{m}{3*p_1}, x_2(p_2, m) = \frac{2*m}{3*p_2}$，即 $\Delta x_1 = x_1(2, 12) - x_1(1, 12) = -2$。补偿收入，$\Delta m = x_1(1, 12) * \Delta p_1 = 6$

即 $m' = m + \Delta m = 18$，替代效应 $\Delta x_1^s = x_1(2, 18) - x_1(1, 12) = -1$，收入效应 $\Delta x_1^n = \Delta x_1 - \Delta x_1^s = -1$。

现在计算希克斯的分解效应。初始效用 $u_1 = x_1^{1/3} * x_2^{2/3} = 4$，保持效用不变，补偿收入 Δm^h，即 $p_1' * x_1 + p_2 * x_2 = m + \Delta m^h = m'^h$，从而 $u_1 = \frac{m'^h}{3*p_1'}^{1/3} * \frac{2*m'^h}{3*p_2'}^{2/3}$，带入得到 $m'^h = 12 * 2^{1/3}$。因此替代效应为 $\Delta x_1^s = x_1(2, 12 * 2^{1/3}) - x_1(1, 12) = 2 * 21/3 - 4$，收入效应 $\Delta x_1^n = \Delta x_1 - \Delta x_1^s = 2 - 2 * 2^{1/3}$。

记住：由于斯拉茨基分解后的消费者效用大于初始效用，而希克斯分解保持效用不变，因此希克斯替代效应的绝对值小于斯拉茨基的替代效应的绝对值。

5. 解：通过分解效用来观察商品的需求特征，首先的，替代效应非正，然后分解主要观察收入效应，可以用收入提供线作为辅助。结果如下：

5.2.7 评价

消费者理论源于古典经济学，效用最大化可以追溯到哥森第一、第二定律，分析方法是边际分析法，更为准确的是马歇尔创立的局部均衡分析，也叫比较静态分析。数理方面可以学习蒋中一《数理经济学的基本方法》第三篇。消费理论是微观经济学的基础，它决定市场的需求，因此必须掌握，核心就是效用约束最优化。

5.3 生产者理论

厂商如何运作？简单地说，厂商从事两类活动：生成和销售。厂商面对产品市场，决定产品的供给，厂商面对生产要素市场，决定如何配置生产要素。前者由产品市场的需求和其他厂商的供给决定，目标是利润最大化，后者由生产要素的价格决定，目标是成本最小化。并且，厂商的供给决策决定了产量，因此决定生产要素的配置规模和结构。同时，厂商在短期和长期面临不同的环境约束，短期中一些生产要素的使用量是不变的（fixed），如厂房，长期中全部生产要素的使用量都是可变（variable）的。

5.3.1 利润最大化

假设产品市场的价格为 p，厂商的供给量为 q，总成本函数为 $C(q)$，则厂商的利润最大化目标和约束条件为：

$$\text{Max } \Sigma = pq - C(q) \tag{5.29}$$

$$\text{s.t. } p = p(q) \tag{5.30}$$

一阶最优化条件为：

$$\text{F.O.C } p'q + p - C'(q) = 0 \tag{5.31}$$

在完全竞争的产品市场，$p = \bar{p}$，并假设不变的边际生产成本 $C(q) = cq$，则利润为 $\Sigma = pq - cq$。当 $p = c$ 时厂商利润为 0。从生产和销售统一的角度，厂商的利润是投入品价值和产品价值的差额，即生产函数为 $q = f(x_1, x_2)$，成本函数为 $C(q) = p_1 x_1 + p_2 x_2$，则厂商的最优化目标为：

$$\text{Max } \Sigma = pf(x_1, x_2) - p_1 x_1 - p_2 x_2 \tag{5.32}$$

这里假设 $p = \bar{p}$，则一阶最优化条件为 $\bar{p}\frac{\partial f}{\partial x_i} - p_i = 0, i = 1, 2$，即 $MP_i = \frac{p_i}{\bar{p}}$，$MP_i$ 为生产要素 i 的边际产品，该条件的经济含义为*边际产品等于生产该产品的投入品的相对价值*。这里，未知的因素有两个，一是生产函数的形式（技术），二是生产要素的价格。由于生产要素市场完全竞争，这里假设生产要素的价格已知。下面着重分析**技术**（即投入品到产品的生产过程）的特征。

常见的生产函数形式：

1. 完全替代：$f(x_1, x_2) = ax_1 + bx_2$
2. 完全互补：$f(x_1, x_2) = \min\{ax_1, bx_2\}$
3. 柯布道格拉斯：$f(x_1, x_2) = Ax_1^a x_2^b$

生产函数的特征：

1. 单调性：增加一种生产要素的投入，至少能生产出和原来一样多的产品。$f(x_1, x_2 + \delta x_1) \geq f(x_1, x_2)$，即 $\frac{\partial f}{x_i} > 0, i = 1, 2$。
2. 凸性：等产量线上两种投入组合的加权平均能比两种投入组合生产出更多的产品。$f(x_1^{(1)}, x_2^{(1)}) = (x_1^{(2)}, x_2^{(2)}) = q_0, f(tx_1^{(1)} + (1-t)x_1^{(2)}, tx_2^{(1)}) + (1-t)x_2^{(2)}$，即 $\frac{\partial^2 f}{\partial x_i^2} > 0, i = 1, 2$。

边际产品递减规模和边际技术替代率递减规律：

- 边际产品：$MP_i = \frac{\partial f}{\partial x_i}, i = 1, 2$，$MP_i$ 随 x_i 增加而减少。
- 边际技术替代率：$MRTS_{1,2} = \frac{\delta x_2}{\delta x_1} = -\frac{MP_1}{MP_2}$，随 x_1 增加或者 x_2 减少，$MRTS_{1,2}$ 下降。

规模报酬：投入品成倍增加引起产量如何增加。设 $q = f(x), t^k = f(tx)$

- 规模报酬不变: $k = 1$。如 $f(x_1, x_2) = Ax_1^a x_2^b, a + b = 1$。
- 规模报酬递增: $k > 1$。如 $f(x_1, x_2) = Ax_1^a x_2^b, a + b > 1$。
- 规模报酬递减: $k < 1$。如 $f(x_1, x_2) = Ax_1^a x_2^b, a + b < 1$。

回到式子(5.32)，换个角度：

$$\text{Max } \Sigma \tag{5.33}$$

$$s.t. \begin{cases} q = \frac{\Sigma}{p} + \frac{p_1}{p}x_1 + \frac{p_2}{p}x_2 \\ q \leq f(x_1, x_2) \end{cases} \tag{5.34}$$

这里最优化条件为 $MP_i = \frac{p_i}{p}, i = 1, 2$。请自行绘图分析！这样可以直观地将产品市场销售决策和生产要素市场的分配决策统一起来。请根据该最优化函数的图形表达分析 p_1, p_2, p 的变化对 x_1, x_2, q 的影响（比较静态分析）。但是，进行比较静态分析的前提是有均衡解，这依赖于规模报酬递减，至少是局部规模报酬递减。在规模报酬递增时，利润随着生产规模扩大而增加，因为相对的，生产成本是规模不变的。在规模报酬不变时，生产规模对利润没有影响。

短期和长期的决策不同之处在于生产要素投入的选择集。假设短期 $x_2 = \bar{x_2}$，即短期利润最大化优化形式：

$$\text{Max } \Sigma \tag{5.35}$$

$$s.t. \begin{cases} q = \frac{\Sigma}{p} + \frac{p_1}{p}x_1 + \frac{p_2}{p}\bar{x_2} \\ q \leq f(x_1, \bar{x_2}) \end{cases} \tag{5.36}$$

5.3.2 成本最小化

考虑利润最大化并没有告诉我们厂商如何配置生产要素，你可以想象厂商包括两个部门——生产部门和销售部门，销售部门负责市场定价，并告诉生产部门产量要求。厂商如何控制产品供给的内容涉及市场形式和生产者博弈的复杂内容，这里我们考虑给定产量要求，厂商如何决定生产要素的配置。厂商的唯一目标是成本最小化：

$$\text{Min } C = w_1x_1 + w_2x_2 \tag{5.37}$$

$$\text{s.t. } f(x_1, x_2) = q \tag{5.38}$$

这个最优化形式的对偶形式就是效用最大化的约束优化。优化目标改写为等成本线：

$$x_2 = \frac{C}{w_2} - \frac{w_1}{w_2}x_1 \tag{5.39}$$

移动等成本线直到在与等产量线（$f(x_1, x_2) = qx_2 = Isoquant(x_1, q)$）相交的前提下截距最小。得到连续、凸性的等产量线假设下的最优化充分条件（相切）：边际技术替代率等于要素的价格比率。

$$MRTS_{1,2} = -\frac{MP_1}{MP_2} = -\frac{w_1}{w_2} \tag{5.40}$$

5.3.3 成本曲线

在给定产量下，厂商总是选择最低成本的生产要素配置方式，因此生产成本是产量的唯一单调映射，即 $C(q)$，这里，可以把销售成本等也纳入 $C(q)$。通常来说，厂商的成本根据是否与产

量相关可以划分为两类，一是不变成本，二是可变成本，即 $C = VC+FC$ OR $C(q) = C_v(q)+F$，前者如短期内的固定资产投资，后者如生产要素投入。平均成本函数是单位产品的成本，这是我们主要关心的曲线，它的积分就是对应的成本函数，因此在决策和福利分析时非常有利。包括以下曲线：

1. 平均总成本：$AC(q) = \frac{C(q)}{q}$
2. 平均可变成本：$AVC(q) = \frac{C_v(q)}{q}$
3. 平均固定成本：$AFC(q) = \frac{F}{q}$
4. 边际成本：$MC(q) = C'(q) = C'_v(q)$

边际成本曲线一定穿过平均总成本和平均可变成本的最底点。这四条曲线的基本形态如图5.1.

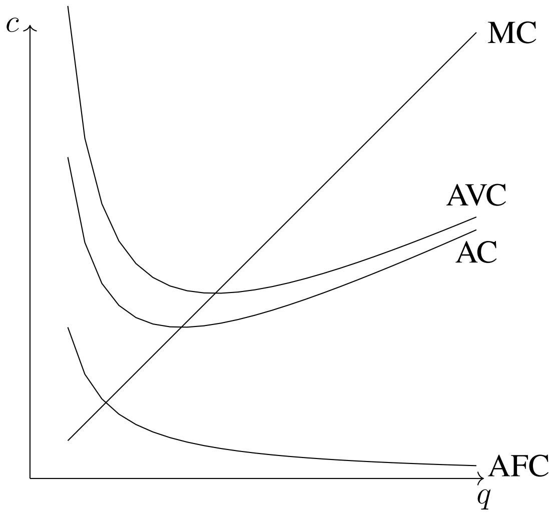

图 5.1: 成本曲线图

并且：$C_v(q) = MC(0) + MC(1) + ... + MC(q-1)$。

长期成本曲线如何取得？短期中部分生产要素不变，假设短期成本函数 $C_s(q, k^*)$，那么长期成本函数为 $C(q, k(q))$。即长期中厂商调整生产要素 k 的使用达到最小化的长期生产成本。因此：

$$C(q) \le C_s(q, k^*) \tag{5.41}$$

并且存在 $q = q^*$ 使得 $k = k^*$，从而 $C(q) = C_s(q, k^*)$，即短期成本等于长期成本。因此，短期平均成本曲线总是在长期平均成本曲线之上，长期平均成本曲线总是与短期平均成本曲线相切，即长期平均成本曲线是短期平均成本曲线的包络线。

5.3.4 厂商供给

首先分析典型的成本曲线如何决定供给曲线，然后分析短期和长期的厂商供给，并且分析是从完全竞争市场的基准情形入手，之后分析完全垄断市场的市场供给选择。下一章将分析寡头垄断的市场供给如何决定以及厂商之间的博弈。

5.3.4.1 完全竞争市场

完全竞争市场中，市场价格与每一个厂商的产量无关，厂商是价格接受者。成立的条件：

1. 厂商数量：厂商数量足够多，每个厂商的产量占市场份额非常小
2. 产品质量：同质产品，即不同厂商的产品完全替代
3. 准入限制：自由进入和退出

对于单个厂商来说，其经营目标仍然是利润最大化，即：

$$\text{Max } \bar{p}q - C(q) \tag{5.42}$$

即其最优产量决定于 $\bar{p} - \frac{\delta c}{\delta q} = 0$，或者写为 $\bar{p}\delta q - \delta c = 0$，即最后一单位产品的收入等于生产成本时利润最大（$P = MR = MC$）。因此，厂商的供给曲线就是其边际成本曲线（在给定的价格水平下，厂商选择利润最大化的产量，此时 $MC = \bar{p}$），但是只包含边际成本曲线的一部分。不包含的部分如下

- 下降的边际成本曲线部分（随产量上升边际成本下降）：此时增加产量，利润上升。因此该商品不是吉芬商品的供给曲线。WHY?
- 价格低于平均变动成本的部分：此时由 $\bar{p}q - C(q) = q(\bar{p} - \frac{C_v(q)}{q}) - F < -F$ 知，利润低于不进行生产的利润（不变资本损失），厂商短期选择不生产，长期退出市场。值得注意的是，当价格高于平均变动成本而低于平均总成本时，厂商短期选择生产，此时的利润虽然为负，但是高于不生产的利润，在长期，厂商退出市场。

值得注意的是，边际成本曲线穿过平均可变成本曲线的最低点，因此，对于典型的成本曲线，厂商供给曲线是该点以右的边际成本曲线。

在短期，厂商不能自由进入和退出市场，因此可能出现盈利或损失的情况，但是在长期，厂商盈利会吸引竞争者进入，从而扩大市场供给，价格下降，厂商损失会导致供给者退出，从而减少市场供给，价格上涨，最终完全竞争市场的厂商没有利润（经济利润），价格等于平均总成本曲线最低点对应的价格。

5.3.4.2 完全垄断市场

完全垄断市场上，厂商的供给就是市场供给，厂商的产量决定了其面临的价格，因此决定其利润。其优化目标为：

$$\text{Max } \pi = D(q)q - C(q) \tag{5.43}$$

其中 $D(q)$ 是反市场需求曲线。边际成本为 $MC = C'(q)$，边际收入为 $MR = D'(q)q + D(q)$，同样，当 $MC = MR$ 时，利润取最值。但是此时，由 $D'(q) < 0$，$P = D(q) > MR$，即价格高于边际收入，也就高于边际成本。因此，厂商依靠垄断地位获取了利润。

对于不完全竞争市场，长期内厂商可能获得利润。此时利润可用经济租金衡量：支付给生产要素的报酬超出为获得该要素而必须支付的最低报酬部分，即 $\pi_{rent} = p^*q^* - C_v(q^*)$。租金可以理解为厂商为获取垄断地位必须付出的成本。这里唯一的未知参数是 q^*，它是如何决定的呢？显然，垄断厂商在利润最大化的目标下，生产太多或太少都是不利的，在最优水平

上，一定有 $MR = MC$。即：

$$\frac{\Delta r}{\Delta q} = \frac{\Delta c}{\Delta q} \tag{5.44}$$

$$\Delta r = q\Delta p + p\Delta q \tag{5.45}$$

$$\frac{\Delta r}{\Delta q} = p + \frac{\Delta p}{\Delta q}p \tag{5.46}$$

$$\frac{\Delta r}{\Delta q} = p(1 + \frac{1}{\epsilon(q)}) = MC(q) \tag{5.47}$$

$$\frac{\Delta r}{\Delta q} = p(1 - \frac{1}{|\epsilon(q)|}) = MC(q) \tag{5.48}$$

$$\tag{5.49}$$

厂商所面临的需求状况（也就是其垄断地位）决定了市场价格。当市场完全有弹性，即 $|\epsilon| = +\infty$，市场变成完全竞争市场，$p = MC$；当市场需求缺乏弹性，$|\epsilon| < 1$，降低产量能够提高利润，因此厂商不会进行生产；当市场需求具有弹性，$|\epsilon| \geq 1$，$p \geq MC$，厂商可以获取垄断利润。以线性需求曲线和线性成本曲线为例。$D(q) = a - bq, C(q) = h + rq$，则 $r(q) = aq - bq^2$，$MR = a - 2bq$，$MC = MR$，即 $a - 2bq = r$，设解为 q^*，此时 $p^* = a - bq^* > a - 2bq^*$，厂商的垄断利润为 bp^*q^*。

5.4 价格歧视和博弈

5.4.1 价格歧视

拥有垄断地位的厂商，可以通过分散的交易价格来获取消费者剩余，**价格歧视手段**（对相同商品向不同的消费者索取不同价格的行为）。价格歧视有三类广义形式：

- 一级价格歧视：向每个顾客都收取其保留价格的行为。也称完全价格歧视，可以摄取所有的消费者剩余。
- 二级价格歧视：对同一商品或服务的不同购买量索取不同价格的行为。如数量折扣和分段定价。
- 三级价格歧视，根据不同的需求曲线将消费者分成两个或者更多的群体并对每一个群体索取不同价格的行为。相对价格由需求价格弹性决定：$MR_1 = MR_2 = MC, \frac{p_1}{1-|\epsilon_1|} = \frac{p_2}{1-|\epsilon_2|}$。但是当某一消费者群体很小并且边际成本上升很快的时候，为该市场服务的额外成本就会高于收益，因此索取单一价格并只向较大的消费者群体服务是有利的。

实践中的价格歧视有如下的几类：

- 跨期价格歧视：利用不同的需求函数将消费者分为不同的组别，通过在不同时点对消费者索取不同价格的行为。
- 高峰负荷定价：当负荷能力限制造成边际成本很高时，在高峰时期索取更高价格的行为。
- 两部收费制：消费者需要同时支付入场费和使用费的一种定价形式。
- 捆绑销售：将两件或者更多商品捆绑在一起销售的行为。

- 搭售：要求某一产品的购买者同时购买同一企业的另一种产品的行为。

两部收费制定价机制是其中重要的内容。现在进行详细分析，然后将进一步分析广告投入的决策条件。

5.4.1.1 两部收费制

消费者要消费特定商品，需要先支付入场费 T^*，然后再为消费行为支付使用费 P^*。如果只有一个消费者或者只有一个需求曲线，那么厂商为尽可能摄取消费者剩余应该将使用费设定为边际成本，而进入费则是全部的消费者剩余。如果有两类消费者，厂商只能设定一个使用费和入场费，此时如果使用费定在边际成本水平，入场费只能定在较低需求的消费者的消费者剩余的水平，而厂商将使用费适当提高并将入场费仍设定在上述水平可以获取更高的利润：$\pi = 2T^* + (P^* - MC)(Q_1 + Q_2)$。如 $D_1(q) = 10 - p, D_2(q) = 4 - 0.25p, MC = 0$，两条需求曲线代表的消费者数量一致，现在分别计算消费者剩余随价格变化的情况。

$$surplus_1 = \begin{cases} \frac{(10-p)^2}{2}, & p \le 10 \\ 0, & p > 10 \end{cases} \tag{5.50}$$

$$surplus_2 = \begin{cases} \frac{(16-p)^2}{8}, & p \le 16 \\ 0, & p > 16 \end{cases} \tag{5.51}$$

比较两个消费者剩余：

$$\begin{cases} surplus_1 > surplus_2, p < 4 \\ surplus_1 = surplus_2, p = 4 \\ surplus_1 < surplus_2, p > 4 \end{cases} \tag{5.52}$$

如果使用费小于 4，入场费应设定为 $surplus_2$，这样可以留住消费者 1，如果使用费大于 4，入场费应设定为 $surplus_1$，这样可以留住消费者 2，但是，留住消费者带来的利益可能会低于完全摄取另一类消费者的利益，因此还需要分类计算。如果只服务一类消费者，那么使用费应该设为边际成本，进入费设为消费者剩余，即：

$$\pi = \begin{cases} 50, & \text{for } D_1 \\ 32, & \text{for } D_2 \end{cases} \tag{5.53}$$

如果服务两个消费者，则：

$$\pi = \begin{cases} 2\frac{(16-p)^2}{8} + p(4 - 0.25p + 10 - p) = -(p-3)^2 + 73 \le 73, & p < 4 \\ 2\frac{(10-p)^2}{2} + p(4 - 0.25p + 10 - p) = -\frac{p^2}{4} - 6p + 100 \le 72, & p \ge 4 \end{cases} \tag{5.54}$$

因此，当使用费等于 3，进入费等于 $surplus_2 = 169/8$ 时，利润最大为 73。

5.4.1.2 广告

厂商进行广告宣传的目的是提高对产品的需求量，设厂商的利润为：

$$\pi = pq(p, a) - c(q) - a \tag{5.55}$$

其中 a 为广告费用。利润最大化的一阶条件为 F.O.C: $\frac{\partial \pi}{\partial a} = p\frac{\partial q}{\partial a} - c'(q)\frac{\partial q}{\partial a} - 1 = 0$，即 $(p - MC)\partial\pi\partial a = 1$，由 $p(1 - \frac{1}{|\epsilon(q)|}) = MC$ 得 $p - MC = \frac{p}{|\epsilon(q)|}$，而广告的需求弹性可以设为 $\epsilon(a) = \frac{\Delta q/q}{\Delta a/a}$，即得到：

$$\frac{a}{pq} = \frac{\epsilon(a)}{|\epsilon(q)|} \tag{5.56}$$

即厂商的广告销售比应该等于广告需求弹性和需求价格弹性的比率（绝对值）。

5.4.2 寡头垄断

在市场上存在两个及以上的垄断厂商，厂商在进行价格和产量决策时必须考虑其他厂商的决策，厂商件可以合作也可以不合作，决策可能是同时做出也可能有先有后。这里，寡头垄断竞争模型有四类：

- 斯塔克尔伯格模型：产量领导，非合作博弈
- 古诺模型：同时决定产量，非合作博弈
- 伯特兰模型：同时决定价格，非合作博弈
- 串谋：同时决定产量，合作博弈

这里，价格领导和产量领导的分析方式一致，不再进行详细分析。其中最重要的模型是古诺模型。下面以例子的形式分析四个博弈形成的解。在进行分析前有必要先理解反应函数，即在假定的其它厂商的产量决策下，厂商的最优产量。假设市场反需求函数为 $D(q) = a - bq$，边际成本不变，即 $C(q) = \delta q$，则对于只有两个垄断厂商的市场，$q = q_1 + q_2$，给定 q_1，则对于厂商 2，利润为

$$\pi_2 = (a - b(q_1 + q_2))q_2 - \delta q_2 = (a - bq_1 - \delta - bq_2)q_2 \tag{5.57}$$

利润最大化产量决定于 $F.O.C : q_2 = \frac{a-\delta-bq_1}{2b}$。

同时，分析串谋产量可以为寡头竞争模型提供参照点。串谋情况下，厂商的共同利润最大化：

$$\pi = (a - bq)q - \delta q = (a - \delta - bq)q \tag{5.58}$$

F.O.C：$q_gong = \frac{a-\delta}{2b}$。

5.4.2.1 斯塔克尔伯格模型

假设市场上有两个厂商，厂商 A 是产量领导者，厂商 B 是产量追随者，厂商 A 宣布产量决策，然后厂商 B 再决定产量，两个厂商的目标都是利润最大化。那么给定厂商 A 的产量 q_1，厂商 B 的产量是 $q_2(q_1) = \frac{a-\delta-bq_1}{2b}$，厂商 A 当然了解厂商 B 的反应函数，它会将其考虑在产量

决策中：

$$\pi_1 = D(q_1+q_2)q_1 - C(q_1) = (a-b(q_1+q_2)-\delta)q_1 = \frac{a-\delta-bq_1}{q_1} \tag{5.59}$$

F.O.C：$q_1 = \frac{a-\delta}{2b}$。带入 B 的反应函数为 $q_2 = \frac{a-\delta}{4b}$。现在市场供给为 $\frac{3(a-\delta)}{4b}$，高于共谋的市场供给 $q_gong = \frac{a-\delta}{2b}$。

5.4.2.2 古诺均衡

厂商可能不了解其他厂商的产量决策，因此其生产是按照预期的产量决策来决定的，即市场信息是不透明的。经过向均衡点的调整，厂商的最终产量决策决定于：

$$\begin{cases} q_1 = q_1(q_2^e) \\ q_2 = q_2(q_1^e) \end{cases} \tag{5.60}$$

在线性需求函数和不变边际成本的情况下，$q_1 = \frac{a-\delta-bq_2}{2b}, q_2 = \frac{a-\delta-bq_1}{2b}$，带入得到 $q_1 = q_2 = \frac{a-\delta}{3b}$，总供给为 $\frac{2(a-\delta)}{3b}$，低于产量领导的博弈结果，高于共谋的博弈结果。

5.4.2.3 伯特兰模型

厂商同时决定价格，让市场决定产量。厂商必须预测其他厂商的价格决策，并最大化自己的利润，这是一种竞争性均衡，每个厂商的价格都等于边际成本，即 $p = MC = MR$，在线性需求函数和不变边际成本的情况下，$q_i = \frac{a-\delta}{b}, i = 1, 2$，市场供给远远大于共谋和其他情况下的寡头垄断模型。

5.4.2.4 卡特尔

在理解寡头垄断的基础上，有必要了解卡特尔（厂商们公开在定价和产量上合作），卡特尔成功的条件包括：

1. 对产品的总需求具有较小的价格弹性
2. 要么卡特尔几乎控制所有的世界供给，要么非卡特尔生产商的供给绝对不能是价格弹性很大的

可以自行分析欧佩克石油卡特尔和西佩克铜卡特尔以印证以上两个条件。

5.4.3 博弈论

博弈论：关于策略相互策略和相互作用的理论。从经济角度看，博弈论研究经济主体行为方式之间相互依存、相互影响、相互作用及其产生的各种相应的结果。根据博弈双方对对方信息的了解和博弈过程是静态还是动态将博弈论分为四种类型：

- 完全信息静态博弈：纳什均衡
- 完全信息动态博弈：子博弈精炼纳什均衡

- 不完全信息静态博弈：贝叶斯纳什均衡
- 不完全信息动态博弈：精炼贝叶斯纳什均衡

博弈论的基本要素包括：

1. 参与者：决策人，目标是效用最大化
2. 行动：游戏者 i 所采取的行动或者步骤，记为 a_i，所有可能行动的集合为行动集，记为 $A_i = a_i$，一次博弈中游戏者各采取一个行动的集合称为行动组合，记为 $a = a_i$，行动组合是博弈的一个可行解
3. 博弈顺序：指博弈决策同时进行还是序列进行
4. 信息集：参与者在某一特定时点上关于不同变量取值的全部知识集合
5. 策略：参与者的行事规则，策略空间和策略组合分别是所有策略和集合和一个可行的策略解
6. 支付：选择的行动给参与者带来的效用或损失
7. 结果：博弈的结果

均衡：博弈中各个参与者都采取其最优策略而产生的策略组合，记为 $s* = (s_1^*, ..., s_N^*)$。**占优策略**：无论竞争者如何行动，该策略都是最好的策略。**占优策略均衡**：无论竞争对手的策略如何，每个厂商的行为总是最优的博弈结果。**纳什均衡**：指任何人单独改变策略都不会获得好处的策略均衡。**极大化极小策略**：选择所有最小收益中的最大值的策略。

5.4.3.1 囚徒困境

参与者 A 和 B 的支付分别用支付矩阵表示为：

$$(A, B) = \begin{pmatrix} (-5,-5) & (-1,-10) \\ (-10,-1) & (-2,-2) \end{pmatrix} \tag{5.61}$$

对于囚犯来说，双方都坦白的结果是获刑 5 年，双方都不坦白的结果是获刑 2 年，如果一方坦白另一方不坦白，则坦白的一方获刑 1 年，不坦白的一方获刑 10 年。因此，不管对方是否坦白，坦白对于一方参与者来说都是占优策略。此时占优策略均衡是双方都坦白，同时该策略也是纳什均衡。求解纳什均衡通常采用划线法。

5.4.3.2 零和博弈

所有结果下博弈双方的支付之和为常数。如：

$$(A, B) = \begin{pmatrix} (5,5) & (10,0) \\ (8,2) & (12,-2) \end{pmatrix} \tag{5.62}$$

对于风险规避者来说，最大化最小策略是合适的。即：

$$max_i\ min_j\ a_ij \tag{5.63}$$

$$max_i\ min_j\ b_ij \tag{5.64}$$

结果是 $a_{21} = 8, b_{21} = 2$，因此 A 选择 2，B 选择 1 是稳定解。同时，由于一方的支付矩阵包含另一方支付矩阵的全部信息，可以通过

$$max_i\ min_j\ a_i j \tag{5.65}$$

$$min_i\ max_j\ a_i j \tag{5.66}$$

解是否相同来判断是否存在均衡解。稍微改变一下支付矩阵的值，如下：

$$(A, B) = \begin{pmatrix} (4,1) & (1,4) \\ (2,3) & (3,2) \end{pmatrix} \tag{5.67}$$

此时 $max_i\ min_j\ a_i j \neq min_i\ max_j\ a_i j$，即不存在纯策略的纳什均衡解。但是存在混合策略纳什均衡解。假设 A 以 p_1 的概率选 1，则 B 选择 1，2 时 A 的预期收益分别为 $2+2p_1, 3-2p_1$，根据最大化最小策略：

$$MIN_{ER} = \begin{cases} 2+2p_1, & p_1 < 0.25 \\ 2.5, & p_1 = 0.25 \\ 3-2p_1, & 0.25 < p_1 < 1 \end{cases} \tag{5.68}$$

即当 $p_1 = 0.25$ 时，A 的预期收益取得最大化最小值。同理，B 的最优策略是 $p_2 = 0.5$。因此，该博弈的混合策略纳什均衡是 A 以 0.25 的概率选择行动 1 而 B 以 0.5 的概率选择行动 1。特别需要记住的是: 如果存在混合策略纳什均衡，则当对方的行动具有发生的正概率时，对方的行动选择对于混合策略决策者一方都是无差异的。

完全信息的动态博弈一般采用反向归纳的方法求解。关于博弈论更深入的探讨可以参考平新乔微观经济学十八讲和范里安微观经济学

5.5 交换和福利

交换经济学考察多个产品市场上的需求和供给条件如何相互影响，从而决定多种商品的价格，这是一般均衡分析的方法。这里我们将着重介绍埃奇沃斯盒心图和福利经济学。首先分析的是纯交换经济，经济学工具是埃奇沃斯盒心图。

5.5.1 埃奇沃斯盒心图

设参与交换的行为人有 A 和 B,所拥有的初始禀赋商品 1 和 2 分别是 (w_A^1, w_A^2) 和 (w_B^1, w_B^2)，假设经过交换，行为人 A 和 B 分别选择了消费束 $X_A = (x_A^1, x_A^2), X_B = (x_B^1, x_B^2)$，则市场商品的总量守恒，有：

$$w_A^1 + w_B^1 = x_A^1 + x_B^1 \tag{5.69}$$

$$w_A^2 + w_B^2 = x_A^2 + x_B^2 \tag{5.70}$$

行为人 A 和 B 进行交换的目的是最大化个人效用，因此按照消费者理论的思路，可以根据预算集和效用函数确定最优消费束。但是在交换经济中价格是可以商量的，交换数量也不

是任意的。因此分析的方法稍有不同。这里首先介绍**帕累托改进**的概念：在不损害其他方福利的同时改进一方的福利。无法进行帕累托改进的资源配置方式就是**帕累托最优配置**。在如下的盒心图中：

绘图!!!

由通过初始禀赋点的无差异曲线围成的透镜状区域是潜在的帕累托改进目标，如果双方均没有议价能力，那么该改进是可以进行的。在两条无差异曲线相交的点上，不存在帕累托改进的可能，因此是帕累托最优配置，所有的曲线切点构成的曲线成为**契约曲线**，在该曲线上始终有 $MRS_{12}^A = MRS_{12}^B$。问题在于，到达契约曲线的均衡一定能够达到吗？或者说交易如何达成。如果存在中间拍卖商，对商品 1 和 2 的交换比例定价，那么分别考察 A 和 B 的最优选择会分别得到最优的消费束，设定为 $X_A = (x_A^1, x_A^2), X_B = (x_B^1, x_B^2)$，此时消费者 A 和 B 对商品 1 和 2 的超额需求（净需求）分别是 $e_i^j = x_i^j - w_i^j, i = A, B, j = 1, 2$，并且均衡时市场的超额需求为 0，即

$$\begin{cases} z^1 = & e_A^1(p_1^*, p_2^*) + e_B^1(p_1^*, p_2^*) = 0 \\ z^2 = & e_A^2(p_1^*, p_2^*) + e_B^2(p_1^*, p_2^*) = 0 \end{cases} \tag{5.71}$$

上述均衡如何能达到呢？瓦尔拉斯法则表明，总的超额需求的值恒等于 0，即：

$$p_1 z^1(p_1, p_2) + p_2 z^2(p_1, p_2) = 0 \tag{5.72}$$

这个等式由预算禀赋和预算线很容易得出。并且,(5.72)表明如果存在 (p_1^*, p_2^*) 使得 $z^1(p_1^*, p_2^*) = 0$ 并且 $p_1^* > 0$，则一定有 $z^2(p_1^*, p_2^*) = 0$。即如果一个市场的供求相等，那么另一个市场的供求也相等。**价格**是决定从初始禀赋达到契约曲线哪一点的关键因素。由此引出福利经济学的基本定理：

- 福利经济学第一定理：在一组竞争市场中均衡时帕累托有效率的。
- 福利经济学第二定理：如果所有交易者的偏好呈凸性，总会存在这样的价格，使得每一组帕累托有效率配置是在适当的商品禀赋条件下的市场均衡。

5.5.2 生产的一般均衡

商品消费的背面是生产，对于生产的一般均衡分析来说，生产的契约线也是两条等产量曲线切点的集合，并且生产 X 和 Y 的边际技术替代率相等 $MRTS_{LK}^X = MRTS_{LK}^Y$。**生产可能性曲线**（PPC）是在技术水平不变的情况下既定资源能够生产的两种产品的最大产量的集合。生产可能性曲线的斜率就是两种产品的边际替代率 $MRT_{1,2}$

在两种投入、两种产出和两个消费者的特殊情况下，市场的一般均衡由生产可能性边界和契约曲线的交点取得，此时生产成本最低，消费者的交换达到帕累托有效率的水平，并且有：

$$MRS_{1,2} = MRT_{1,2} \tag{5.73}$$

5.6 公共选择理论

5.6.1 外部性

外部性：指不经过市场机制的消费和生产活动对其他消费者或者生产者的直接影响。外部性可以分为正的外部性和负的外部性，我们将经济活动所产生的全部收益称为社会收益，全部成本称为社会成本，由决策者自身获得的收益和成本称为私人收益和私人成本，社会成本/收益和私人成本/收益不一致时就产生了外部成本/收益。即：

$$C_{society} = C_{indivdual} + C_{externalty} \tag{5.74}$$

$$R_{society} = R_{indivdual} + R_{externalty} \tag{5.75}$$

决策者按照私人成本和私人收益决策，价格不能反映生产和消费活动的全部成本和收益的准确信息，这将导致相对于社会最优的生产或者消费的过多或者过少，导致低效率。解决外部性主要有两种传统办法：

1. 征收庇古税：对经济行为征收或者补贴，使得私人决策的最优结果与社会最优结果一致，征收和补贴的量等于外部成本/收益。
2. 配额或者限额：限制生产或者消费的量，释放负的外部性对厂商或者消费者直接下令。

科斯定理：在当事人的偏好都为准线性的情况下，如果经济中出现了外部性，则讨价还价过程会产生一个有效的结果（交易成本为零），并且该结果与产权如何配置无关。并且注意三个问题：

- 不同的产权界定会导致不同的收入分配
- 不同的产权界定会导致不同的帕累托有效率点
- 前提是交易成本为零

产权界定不清导致社会无效率的典型例子是**公地悲剧**: 指那些没有明确所有者，人人都可以免费使用的资源，如海洋、湖泊、草场等资源通常会遭到过度使用。

5.6.2 公共物品

公共物品是相对于私人物品而言的，私人物品是指在消费上既具有竞争性又具有排他性的物品。**竞争性**指一个人使用该物品会减少其他人对该物品的使用的特性；**排他性**指可以阻止人们使用这种物品的特性。**公共物品**指在消费上具有非竞争性和非排他性和物品。根据物品是否具有竞争性和排他性，可以分为四类：

- 竞争性 + 排他性：私人物品，如苹果
- 竞争性 + 非排他性：公共资源，如拥挤但不收费的道路
- 非竞争性 + 排他性：俱乐部物品，有线电视
- 非竞争性 + 非排他性：公共物品，国防

对于公共物品，免费搭车者 (free rider) 的存在使得市场很难或者不可能有效地提供商品。既生产私人物品又生产公共物品的经济的帕累托最优条件如下：

1. 私人物品中任两种物品的边际替代率相等
2. 在生产所有物品的生产要素中，任两种生产要素的边际技术替代率相等
3. 在存在公共物品的情况下，私人物品于公共物品的边际替代率之和等于它们之间的边际转换率

阿罗不可能定理：任何社会排序都应该服从以下六个无可争议的公理，

- 集体理性：偏好是可以完备的、排序的
- 无限制性：投票机制不排斥任何形式的个人偏好
- 帕累托最优：如果每个人都相对 B 偏好 A，则社会也相对 B 偏好 A
- 偏好独立性：社会偏好对 A 与 B 的排名与人们对其他不相关的备选方案的排名无关
- 非独裁性：不存在某一个人的偏好成为社会偏好而其他人的偏好无效的可能

阿罗证明了以上五个条件是无法调和的。

第 6 章 宏观经济学

6.1 引言

6.1.1 前言

宏观经济学是研究影响整体经济的力量的经济学领域。本文主要参考曼昆《宏观经济学》（第十版），并辅以罗默高宏的增长模型，意在梳理宏观经济学的整体框架，厘清其中的重点内容。整体而言，宏观经济理论主要包括：古典理论、增长理论、经济周期理论和宏观经济理论与政策四个部分，其中经济周期理论比较接近目前的经济研究实践，也是比较困难的部分。同时，在学习宏观经济学的同时，要特别关注宏观经济变量所反映的微观经济行为。

经济学家常常采用模型来理解世界，用数学术语说明经济变量之间的关系，其中的变量可分为内生变量（模型需要解释）和外生变量（已经给定），在时间序列模型中常常还有前定变量（连接两个时期的经济关系）。模型最重要的部分在于它的假设，通常模型的全部内容就是假设，模型的假设需要阐明对经济现象的思考，但是又不能过于复杂，对所关心的问题作出合适的假设是决定模型是否成功的关键。以供求模型为例，商品的需求决定于商品价格和总收入，即 $Q_D = D(p, y)$，商品的攻击决定于商品价格和原材料的价格，即 $Q_S = S(p, p_m)$，商品的均衡价格决定于 $Q_S = Q_D$，此时双方都没有动力去改变价格和产量。该模型的外生变量是 y, p_m，内生变量是 p, Q。总收入和原材料价格的变动会导致曲线的移动，进而改变均衡价格和产量。该模型可以解释总收入和原材料价格如何影响商品市场，但是不能解释为什么完全竞争市场的价格低于垄断市场的价格。模型是有用的，但是有局限。

市场出清假设指市场处于均衡状态，任何产品与服务的价格都位于供求曲线的交点。该假设要求工资和价格对变化的调整是迅速的，即有弹性而不是黏性。价格弹性在长期经济中是合理的假设。**微观经济学**研究家庭和企业如何做出决策，以及这些决策在市场上如何相互影响。宏观经济学需要微观理论的理由包括：（1）经济整体是经济个体之和（2）总量是个别决策的变量之和。

6.1.2 宏观经济学数据

国内生产总值是描述经济表现状况最重要的指标。计算 GDP 可以采用支出法和收入法两种等价的方法，一般采用支出法。考虑经济变量衡量一单位还是一个时间点的数量也至关重要，存量（stock）衡量给定时点的梳理，流量（flow）衡量每一单位时间内的数量。

国内生产总值（GDP）：给定时期的经济内生产的所有最终产品与服务的市场价值。GDP 的计算规则：

- 所有产品和服务价值的加总：$GDP = \sum_i p_i q_i$
- 二手货：不是当期生产的，不计入 GDP

- 存货：（1）存货变质，产品价值没有实现，不计入 GDP；（2）以后销售，计入当期 GDP 的存货投资
- 中间产品：不是最终的产品，不计入 GDP；每个阶段的增加值的加总也可以计算 GDP
- 估算价值：一些产品和服务不在市场上销售，如果 GDP 要包括这部分产品和服务，则需要估算其价值

由于计算 GDP 的估算是近似的，并且许多产品和服务的价值没有完全纳入 GDP，因此 GDP 不是一个完美的衡量指标。

名义 GDP 是用现期价格衡量的产品与服务的价值，实际 GDP 是用基期不变价格衡量的 GDP。如以 2000 年为基期，$GDP_{norminal} = \sum p_{2021}q_{2021}, GDP_{real} = \sum p_{2000}q_{2021}$。衡量经济总体价格水平的变化可以计算 **GDP 平减指数**：

$$GDP_{deflator} = \frac{GDP_{norminal}}{GDP_{real}} \tag{6.1}$$

也就是名义 GDP 可以分解为两部分：产量和价格

$$GDP_{norminal} = GDP_{real} \times GDP_{deflator} \tag{6.2}$$

理解 GDP 以后，要如何计算 GDP 呢？支出法指出了 GDP 在不同用途之间的配置，即国民收入恒等式：

$$Y = C + I + G + NX \tag{6.3}$$

其中 Y 是 GDP；C 是消费，由家庭在产品和服务的指出构成；I 是投资，由为未来使用而购买的产品构成；G 是政府购买，由联邦政府、州、地方政府购买的产品和服务构成；NX 是净出口，由出口的产品与服务的价值和进口的产品与服务的价值的差额构成。衡量国民收入，也有其他指标，如国民生产总值 GNP，衡量国民所赚取的收入，国民生产净值 NNP，衡量去掉折旧的国民收入。

相比于 GDP 平减指数，消费者价格指数 CPI 是更为常用的价格指标，它与消费者感受到的价格息息相关。**CPI** 是一篮子产品产品与服务的价格相对于同一篮子产品与服务的在基年价格的比值。

$$CPI = \frac{\sum p_{now}q_{pack}}{\sum p_{base}q_{pack}} \tag{6.4}$$

将篮子里的产品和服务换成厂商购买的产品和服务，CPI 就变成了生产者价格指数 PPI。从这个篮子中去掉食物和能源产品，CPI 就变成了核心通货膨胀指数。

GDP 平减指数和 CPI 有三个重要的区别：

1. 研究对象：GDP 平减指数衡量所有产品与服务的价格，CPI 衡量消费者购买的产品与服务的价格
2. 研究范围：GDP 平减指数衡量国内生产的产品和服务而 CPI 衡量国内和国外的消费品
3. 研究方法：CPI 给不同产品和价格分配固定的权重而 GDP 平减指数分配变动的权重

CPI 并不是完美的价格指数，它通常会夸大通货膨胀：

1. 替代偏差：相对价格变动后消费者会替代使用相对便宜的产品和服务
2. 新产品的出现：新产品给了消费者更多的选择，购买力提高，但是不影响 CPI

3. 无法衡量的质量变化：产品质量的提高很难衡量并表现在 CPI 中

失业率衡量了经济利用劳动力的程度，也是关系居民收入的重要变量。全体成年人口 (A) 可以分为劳动力 (L) 和非劳动力，劳动力再分为就业者 (E) 和失业者 (U)。

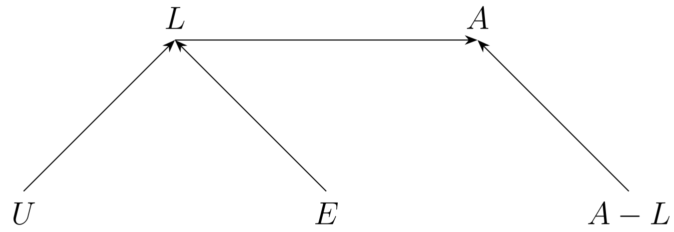

图 6.1: 成年人口构成

- 就业者：包括作为有报酬的雇员工作、在自有企业中工作或者在家庭成员的企业中从事无报酬工作、当时没工作但实际上有工作只是由于天气疾病等原因暂时缺勤的人
- 失业者：愿意工作但没有工作并在此前四个星期努力寻找工作的人，以及被解雇等待召回的人
- 不属于劳动力：想工作但放弃寻找工作——丧失信心的工人

失业率表示为：

$$\text{unemployment rate} = \frac{U}{L} \tag{6.5}$$

劳动力参与率表示为：

$$\text{labor-force participation rate} = \frac{L}{A} \tag{6.6}$$

6.2 古典理论：长期中的经济

6.2.1 国民收入

国民收入取决于两方面，供给和需求，均衡解是国民经济得以维持的解。这里分别考察总供给和总需求如何决定。

6.2.1.1 总供给

总供给取决于生产要素和技术水平。**生产要素**是用于生产产品与服务的投入，**生产函数**决定给定生产要素 (主要是资本和劳动投入) 能够生产多少产出。即

$$Y = F(K, L) \tag{6.7}$$

许多的生产函数具有规模报酬不变 (CRS) 的性质：$zY = F(zK, zL)$。假设一定时期内的资本和劳动存量不变，技术没有进步，则产出也不改变 $Y = F(\bar{K}, \bar{L}) = \bar{Y}$。既然有产出，就需要考虑产出如何在各个生产要素之间进行分配。要素价格是支付给每单位生产要素的报酬数量。对于竞争性企业来说，利润表示为 $\pi = PY - WL - RK = PF(K, L) - WL - RK$，其

中 W 是工资，R 是租金率。简单回想一下微观经济学中企业的要素分配即可了解当要素的边际贡献等于边际成本时企业利润最大。即有：

$$MPK = \frac{R}{P}, MPL = \frac{W}{P} \tag{6.8}$$

并且，这依赖一个假设：边际产量递减。追求利润最大化的竞争性企业遵循一个简答的原则：企业需要每一种生产要素直到该生产要素的边际产量减少到等于其实际要素价格为止。

经济利润是企业支付了生产要素报酬以后留下来的收入：

$$economic\,profit = Y - MPK \times R - MPL \times W \tag{6.9}$$

如果生产函数具有规模报酬不变的性质，那么经济利润等于 0（欧拉定理），即 $F(K, L) = MPK \times R + MPL \times W$。而由于企业所有者往往是资本所有者，企业核算的会计利润往往没有考虑资本报酬，即 $accounting\,profit = economic\,profit + MPK \times K$。

一种常用的生产函数是**柯布-道格拉斯生产函数**：$F(K, L) = AK^{\alpha}L^(1-\alpha)$。它是规模报酬不变的生产函数，并且 $MPK = \frac{\alpha Y}{K}, MPL = \frac{(1-\alpha)Y}{L}$，这里 $\frac{Y}{K}$ 为平均资本生产率，$\frac{Y}{L}$ 为平均劳动生产率，α 为国民收入中分配给资本的比例而 $1-\alpha$ 为国民收入中分配给劳动的比例。

6.2.1.2 总需求

根据国民收入恒等式：$Y = C + I + G + NX$，不考虑净出口，总需求取决于消费、投资和政府购买。消费取决于可支配收入的数量：$C = C(Y - T)$，这里 T 是税收。边际消费倾向 MPC 是收入增加一美元时消费增加的量。投资取决于利率，**利率**衡量为了投资而融资的资金成本，利率分为名义利率和实际利率，前者是报道的利率，后者是剔除通货膨胀，投资者实际感受到的利率。利率水平和投资需求成反比。政府购买是政府支出，其收入源于税收减去转移支付，政府预算有平衡的预算、预算赤字和预算盈余三种情况。

理解了总供给和总需求，现在的问题是经济如何达到均衡？可以将经济分为两个市场分别考察这个问题：

- 产品与服务市场：需求端

$$Y = C + I + G \tag{6.10}$$

$$C = C(Y - T) \tag{6.11}$$

$$I = I(r) \tag{6.12}$$

$$G = \bar{G} \tag{6.13}$$

$$T = \bar{T} \tag{6.14}$$

$$So : Y = C(Y - \bar{T}) + I(r) + \bar{G} \tag{6.15}$$

  供给端：$\bar{Y} = F(\bar{K}, \bar{L})$ 均衡：$\bar{Y} = C(\bar{Y} - \bar{T}) + I(r) + \bar{G}$ 其中调整经济达到均衡的力量是利率。

- 金融市场：需求端是投资需求 $I(r)$，供给端是国民储蓄 $S = S_{public} + S_{private} = (T - G) + (Y - T - C) = I$。金融市场流入的流量一定与金融市场流出的流量相等，即 $\bar{S} =$

$\bar{Y} - C(\bar{Y} - \bar{T}) - \bar{G} = I(r)$。利率将进行调整以达到企业想要投资的量等于家庭想要储蓄的量。

现在基本的模型已经搭建完成，可以自行分析政府购买增加、税收下降、投资需求增加对利率和国民收入的影响。

6.2.2 货币系统

宏观经济政策包括财政政策（政府关于支出和税收的决策）和货币政策（关于一国硬币、通货和银行体系的决策）。**货币**是用于交易的货币存量。货币有三大职能：

1. 价值储藏手段：将现在的购买力转移到未来
2. 计价单位：作为标记价格和记录债务的单位
3. 交换媒介：用于购买产品和服务

货币的类型包括：法定货币（没有内在价值，政府或者法令规定为货币）、商品货币（具有内在价值的货币）。当黄金作为货币时，经济是实行金本位制。

一个经济中可用的货币量称为**货币供给**，政府对货币供给的政策称为**货币政策**，货币政策一般由中央银行执行。货币量有分级的衡量方式：

1. 通货 (C)：未清偿的纸币和硬币总和
2. 活期存款 (D)：人们在银行账户上的存款

货币供给等于通货和活期存款之和：$M = C + D$。银行通过准备金制度创造货币供给。准备金是银行收到但没有贷出去的存款。一个货币供给模型包括以下内容：

- 基础货币：公众以通货持有的货币和银行以准备金形式持有的货币 $B = C + R$
- 存款准备金率：银行持有的准备金占存款的比例 rr，通常受货币政策影响
- 通货存款比：公众持有的通货量对活期存款的比例 cr，表示人民持有货币形式的偏好

$$M = C + D \tag{6.16}$$

$$B = C + R \tag{6.17}$$

$$\frac{M}{B} = \frac{C + D}{C + R} \tag{6.18}$$

$$M = \frac{cr + 1}{cr + rr} \times B \tag{6.19}$$

$$m = \frac{cr + 1}{cr + rr} \tag{6.20}$$

这里 m 是货币乘数。基于货币供给模型，央行影响货币供给的方式分为两类：

1. 影响基础货币：公开市场操作，即央行买卖政府债券；将准备金借给银行；调整贴现率
2. 印象存款准备金率：调整法定准备金率；调整准备金利息

由于央行无法控制银行的经营决策和家庭的财务决策，货币供给常常以央行意想不到的方式变动。

6.2.3 通货膨胀

通货膨胀 (inflation) 指价格水平的总体上升。理解通货膨胀的经典方法是货币数量论，它实际是从货币需求的角度来探讨货币和价格。数量方程表示交易和货币之间的关系：

$$M \times V = P \times T \tag{6.21}$$

其中 M 代表货币，V 是货币的交易流通速度，衡量一定时期内货币换手的次数，P 代表一次典型交易的价格，T 是某一时期的交易总数。在研究货币在经济中的作用时，常常用收入替代交易，数量方程变成：

$$M \times V = P \times Y \tag{6.22}$$

这里 M 代表货币，V 是货币的收入流通速度，P 是 GDP 平减指数，Y 是实际 GDP，故 PY 是实际 GDP。将货币量表示为能够购买的产品和服务的数量，即 $\frac{M}{P}$，它表示实际货币余额，衡量货币存量的购买力。因此货币需求函数表示为

$$(\frac{M}{P})^d = kY, k = \frac{1}{V} \tag{6.23}$$

增加货币流通速度不变的假设，数量方程就变为理解货币的重要理论：货币数量论，$M \times \bar{V} = P \times Y$。因此，

1. 生产要素和生产函数决定经济总产出 Y
2. 央行货币供给 M 决定产出的名义价值 PY
3. 价格水平 P 是名义价值 PY 与产出 Y 的比率

货币数量论意味着价格水平与货币供给成比例。

$$\%\Delta M + \%\Delta V = \%\Delta P\%\Delta Y \tag{6.24}$$

由于央行控制货币供给 ($\%\Delta M$)，货币流通速度的变动反应需求的移动 ($\%\Delta V$)，$\%\Delta P$ 是通货膨胀率，产出变动 $\%\Delta Y$ 取决于生产要素和技术变化，因此货币供给的增长决定通货膨胀率。货币铸造税是通过货币发行筹集到的收入。

费雪方程：

$$i = r + \pi \tag{6.25}$$

其中 i 是实际利率，r 是名义利率，π 是通货膨胀率。准确的说，其中的 π 应该写成预期通货膨胀率 $E\pi$。

利率是持有货币的机会成本，因而也会影响货币需求函数，即 $(\frac{M}{P})^d = L(i, Y) = L(r + E\pi, Y)$。价格水平不仅取决于今天的货币供给，也取决于预期的货币供给。

通货膨胀的社会影响：

1. 预期的通胀的成本：鞋底成本、菜单成本、价格扭曲的成本、税收扭曲的成本、价格变动的不便成本
2. 未预期的通胀的成本：财富的任意再分配、固定养老金实际收入的下降、多变的通胀
3. 好处：使劳动力市场更好地运行

古典二分法：名义变量（用货币表示的变量）和实际变量（数量和相对价格）在理论上

的分离。货币在实际变量决定中的无用性称为**货币中性**。

6.2.4 失业

自然失业率是经济围绕其波动的平均失业率。给定所有阻碍工人立刻找到工作的劳动市场的不完美性，自然失业率是长期中经济趋近的失业率。什么因素决定失业率呢？假设：（1）劳动力是固定的（2）s 代表离职率，f 代表入职率。即稳定转态条件：

$$fU = sE \tag{6.26}$$

得到 $\frac{U}{L} = \frac{1}{1+f/s}$。

工人失业的原因包括两类：摩擦性失业和结构性失业。

- 摩擦性失业：工人由于找工作需要花时间造成的失业。包括部门转移（需求构成在不同行业和地区之间变动）和公共政策（就业机构信息、就业指导、失业保障）
- 结构性失业：工资刚性和工作配给引起的失业。引起工资刚性的原因有最低工资法、工会的垄断力量、效率工资

6.3 增长理论：超长期中的经济

6.3.1 索洛模型

索洛模型：储蓄、人口增长和和技术进步如何影响一个经济的产出水平及其随着时间增长。首先产出取决于资本存量和劳动力，并假设生产函数具有规模报酬不变的特征，即：

$$Y = F(K, L) \tag{6.27}$$

$$\frac{Y}{L} = F(\frac{K}{L}, 1) \tag{6.28}$$

$$Set, y = \frac{Y}{L} \; and \; k = \frac{K}{L} \tag{6.29}$$

$$So, y = f(k) \tag{6.30}$$

$$\tag{6.31}$$

在索洛模型中，产品的需求来自于消费和购买，假设人们储蓄 s 比例的收入，即：

$$c = (1-s)y, i = sy \tag{6.32}$$

人均资本是目前的索洛模型中的唯一变量，影响人均资本的两种力量是投资和折旧，前者提高人均资本而后者降低人均资本。假设资本存量以 δ 的比例磨损，则：

$$\Delta k = i - \delta k = sf(k) - k \tag{6.33}$$

从而，人均资本存量是自激发的时间序列，通过自我调整，人均资本存量在稻田条件（原点处斜率无限大，凹性）下可以达到均衡。如：

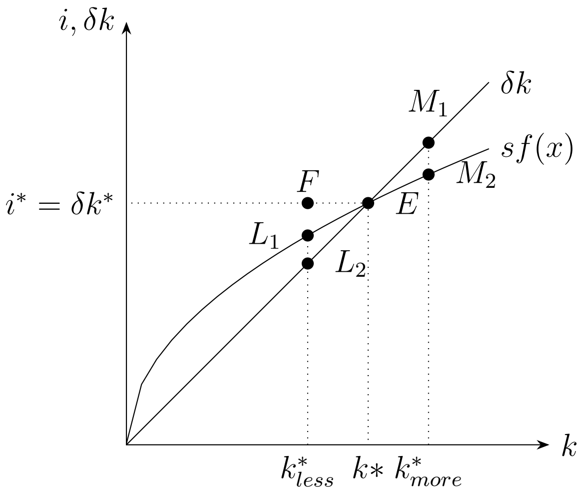

图 6.2: 投资、折旧和向均衡的调整

当 $k < k^*$，$\delta k > 0$，k 增加，并且 $\delta k < |k - k^*|$，即资本存量不足时资本增加但是不会跃迁到资本过剩的状态，这个条件在 $\delta < 1$ 时一定成立（$tan(FL_2E) > 1, So, L_1L_2 < EH$）。

现在考虑消费如何随着经济增长而增长，毕竟消费者的当期福利水平取决于当期的消费，而储蓄是未来的消费，不考虑消费者的跨期消费分配，当人均资本达到什么水平时消费最高呢？即最大化：

$$c = (1 - s)y = f(k) - \delta k \tag{6.34}$$

显然，一阶最优化条件为：$f'(k^*) = \delta$，即 $sf(x)$ 在 k^* 处的切线斜率等于 δ，并称 k^* 为人均资本的黄金率水平，即：

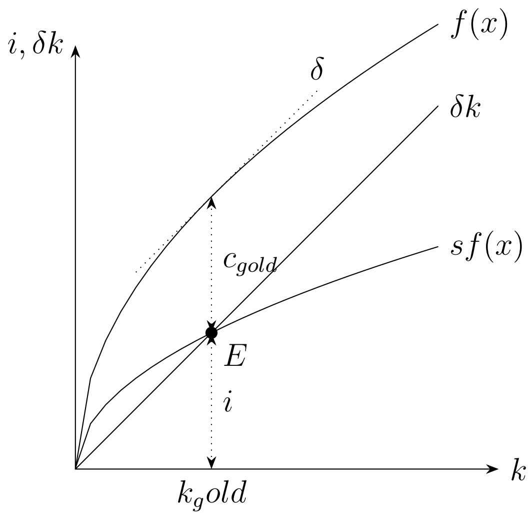

图 6.3: 消费和黄金率水平

6.3.2 平衡的增长

（1）分析平衡增长时资本、收入、劳动力、人均资本、人均收入的变化（2）分析劳动增长率、折旧和储蓄对增长的作用（3）分析储蓄调整的效应

6.4 经济周期理论

6.4.1 经济波动

产出和就业的短期波动称为**经济周期**，当经济经历产出下降和失业上升的时期，称经济处于**衰退**。

6.4.1.1 关于经济周期的事实

在经济中投资比消费的波动大得多。失业与 GDP 之间的负相关关系称为**奥肯定律**。对于美国经济，奥肯定律的表达式为：

$$\%\Delta Y_{real} = 3\% - 2\Delta Rate_{unemployment} \tag{6.35}$$

该式意味着如果失业率保持不变，实际 GDP 增长 3%，这一增长来自劳动力的增长、资本积累和技术进步。经济学家常常会预测 GDP 和失业这样的数据，一种方法就是观察领先指标，即在整体经济运动之前先发生波动的指标。

在分析经济波动时，我们着眼于短期经济。长短期的关键差别在于价格行为，在长期，价格是灵活调整的，能够对需求和供给的变化做出反应，在短期，价格具有黏性，固定在前定水平上，并不会对货币政策做出反应。

6.4.1.2 AD-AS 模型

分析短期经济波动的良好的起点是总需求-总供给（AD-AS）模型。总需求是产出需求量和价格水平之间的关系，总供给是产品和服务的供给量和价格水平之间的关系。回想一下货币数量论的内容，对产出和服务的需求由货币供给决定，当然和价格水平相关，因此总需求函数可以写成：

$$(\frac{M}{P})^d = kY \tag{6.36}$$

当 M 和 V 固定时，AD 曲线是一条向右下方倾斜的曲线。这意味着什么？产出越高，人们进行的交易越多，所需要的实际货币余额越多，对于一个固定的货币供给，实际货币余额越高意味着价格水平越低。反之亦然。总需求曲线的位置受货币供给 M 和货币的收入流通速度 V 影响，M 增加或者 V 增加，总需求增加，AD 曲线右移，M 降低或者 V 降低，总需求降低，AD 曲线左移。

总供给曲线分为长期总供给曲线（LRAS）和短期总供给曲线（SRAS）。长期总供给曲线是垂直的曲线，$\bar{Y} = F(\bar{K}, \bar{L})$（古典二分法，产出达到充分就业的水平），短期总供给曲线是水平的曲线，$P = \bar{P}$，短期中所有价格均固定在前定水平上，短期供给可以快速调整以满足短期的产出和服务需求。总需求的提高在短期会引起产出增加（但价格不变），在长期产出回到自然水平，价格上升。长期均衡位于总需求曲线、短期总供给曲线和长期总供给曲线的交点。如果央行降低货币供给，AD 曲线左移，短期产出下降，价格不变，长期产出回到自然率

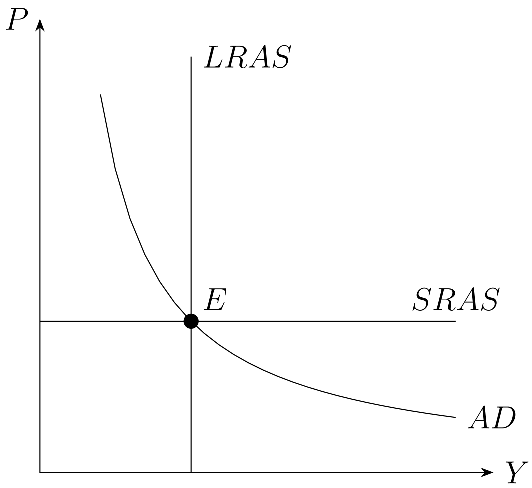

图 6.4: 总需求、总供给曲线和长期均衡

水平，价格下降，经济将经历从短期到长期的复苏调整过程。

6.4.1.3 稳定化政策

冲击（shock）是使需求曲线和供给曲线移动的外生事件，使需求曲线移动的冲击是需求冲击，使供给曲线移动的冲击是供给冲击。旨在减少短期经济波动严重性的政策称为**稳定化政策**。

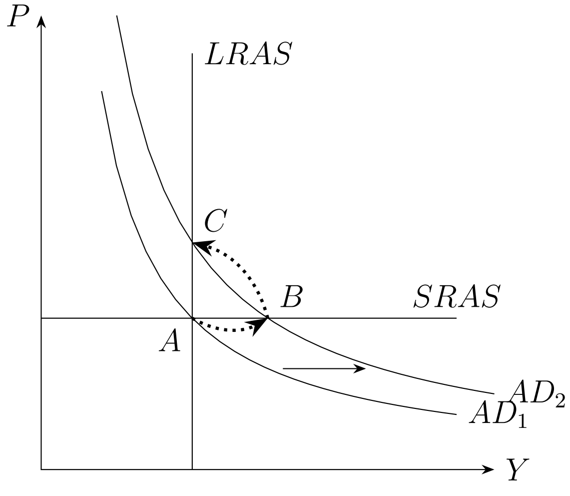

图 6.5: 总需求增加

请自行分析 V 增加对总需求和冲击以及央行的政策选择；分析不利的总供给冲击对经济的影响（滞胀）以及央行的政策后果。

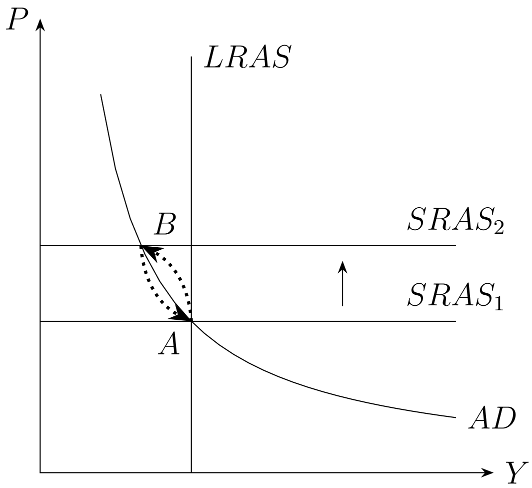

图 6.6: 不利的供给冲击

6.4.2 IS-LM 模型

6.4.2.1 建立 IS-LM 模型

IS-LM 模型可以更为详细地考察总需求的决定因素，IS 曲线分析产品和服务市场上利率和收入水平的关系，LM 曲线分析货币市场上利率和收入水平的关系。IS 曲线从凯恩斯交叉开始。假设（1）简单封闭经济（2）收入取决于支出。实际支出是家庭、企业和政府花在产品和服务上的数额，计划支出是家庭、企业和政府想花在产品和服务商的数额，计划支出和实际支出的差异在于企业销售不及时有额外的存货投资。计划支出决定于消费、投资和政府购买：

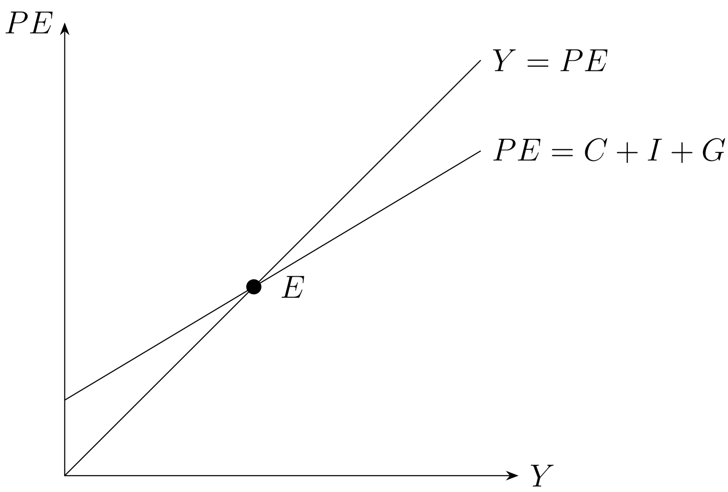

图 6.7: 凯恩斯交叉

$$
\begin{aligned}
PE =& C + I + G && (6.37)\\
PE =& C(Y - \bar{T}) + \bar{I} + \bar{G} && (6.38)\\
\frac{dPE}{dY} =& MPC && (6.39)
\end{aligned}
$$

PE 曲线的斜率是边际消费倾向，一般地，$MPC < 1$。当人们的计划都实现时，他们没有动力去改变当前的行为（均衡），因此：

$$PE = Y \tag{6.40}$$

调整经济达到均衡的力量是存货。政府购买乘数是政府购买增加引起的产出增加的倍数：

$$\frac{\Delta G}{\Delta Y} = \frac{1}{1 - MPC} \tag{6.41}$$

税收乘数是税收增加引起的产出变动的倍数：

$$\frac{\Delta T}{\Delta Y} = -\frac{MPC}{1 - MPC} \tag{6.42}$$

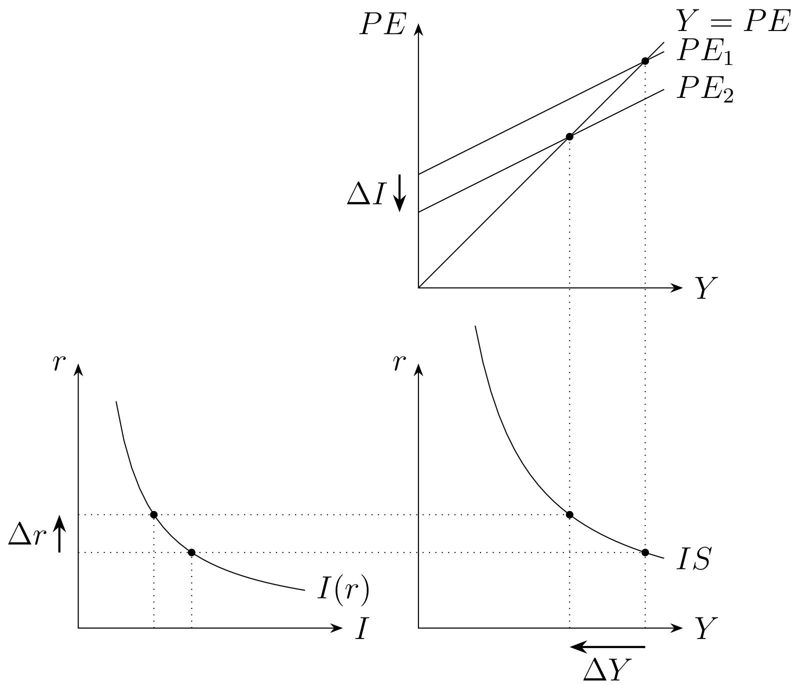

图 6.8: IS 曲线推导

将凯恩斯交叉中的固定投资修改为取决于利率的投资，IS 曲线表明收入如何随着利率的变化而变化。对于 IS 曲线，财政政策影响计划支出，因此传递到 IS 曲线，如政府购买增加，PE 曲线上移，最终 IS 曲线右移，即总需求增加。

LM 曲线的理论基础是流动性偏好理论，利率调整使得经济中最具流动性的资产-货币-的供求均衡。假设存在固定的实际货币余额供给：

$$(\frac{M}{P})^d = \frac{\bar{M}}{\bar{P}} \tag{6.43}$$

考虑实际货币余额的需求，利率是人们持有货币的机会成本，因此利率上升时，持有货

币的机会成本提高，对货币的需求降低，即：

$$(\frac{M}{P})^d = L(r) \tag{6.44}$$

利率的调整使货币市场达到均衡。货币供给的减少引起利率上升，货币供给增长引起利率下降。同时，输入会影响货币需求，收入提高时货币需求也会提高，货币需求函数更正为：

$$(\frac{M}{P})^d = L(r, Y) \tag{6.45}$$

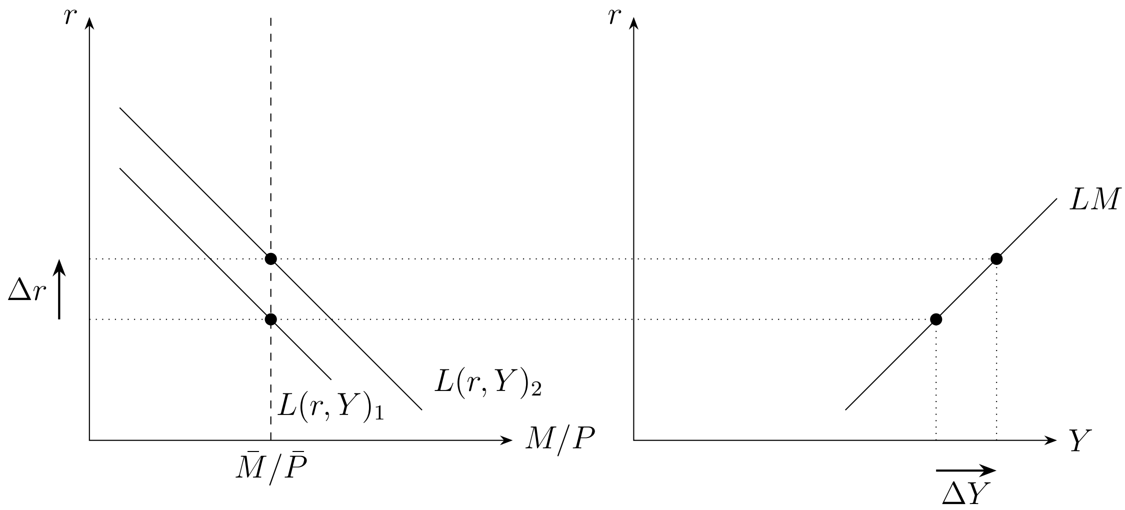

图 6.9: LM 曲线推导

当收入提高时，货币需求曲线上移，而实际货币余额不变，因此利率提高。货币政策如何影响 LM 曲线？实际货币余额是影响 LM 曲线的唯一因素，其减少导致 LM 曲线上移，其增加导致 LM 曲线下移。

IS-LM 模型的全部组成部分：

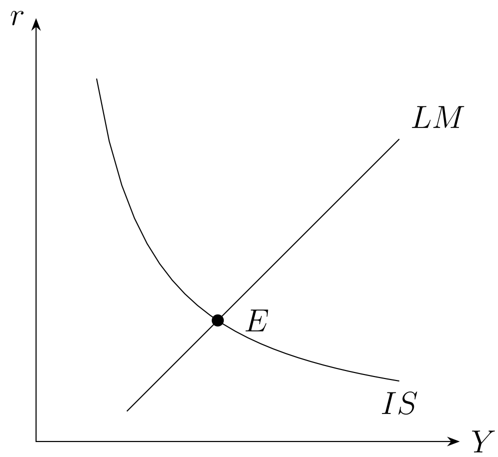

图 6.10: $IS-LM$ 模型

$$\begin{cases} Y = C(Y-T) + I(r) + G, & IS \\ M/P = L(r,Y), & LM \end{cases} \tag{6.46}$$

在 LS 曲线和 LM 曲线的交点处，实际支出等于计划支出，实际货币余额的供给等于需求，经济处于短期均衡。

6.4.2.2 应用 IS-LM 模型

IS-LM 模型决定在价格水平固定的短期中利率和国民收入。可以用于分析三个问题：

1. 国民收入波动的潜在原因: 财政政策、货币政策、二者的相互作用、外生冲击（需求端和供给端）
2. IS-LM 模型如何与 AD-AS 模型相容：推导 AD 曲线，分析短期和长期的 AD 曲线
3. 大萧条

推导 AD 曲线：

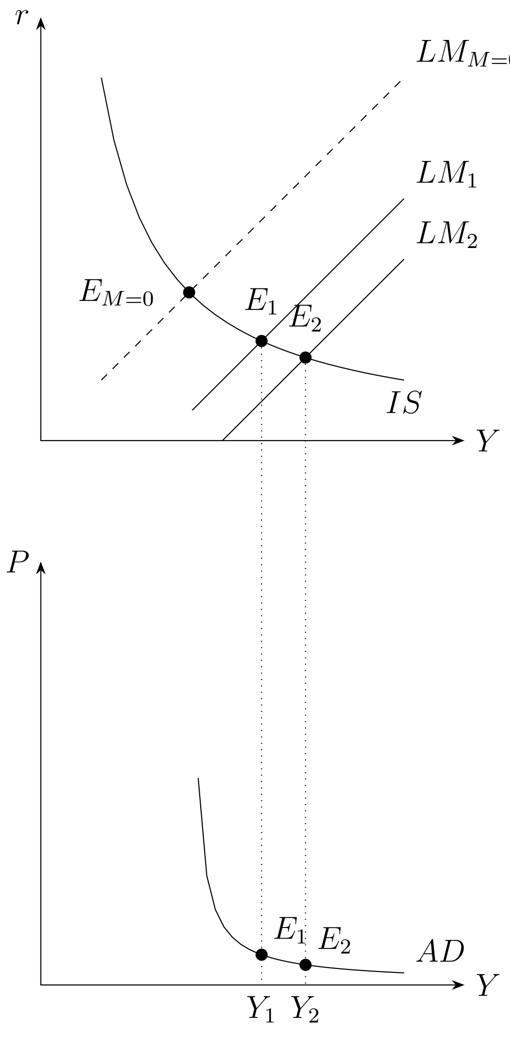

图 6.11: 由 $IS-LM$ 到 AD 曲线: 待增加两条垂线

对大萧条的解释：

- 支出假说：对 IS 曲线的冲击，产品和服务的支出下降。

- 货币假说：对 LM 曲线的冲击，货币的紧缩。价格下降（通货紧缩）的效应：（1）稳定效应，LM 的扩张性移动，收入提高；庇古效应，财富幻觉导致支出更多，IS 曲线扩张性移动。（2）不稳定效应，债务-通缩理论和预期到的通缩效应，IS 曲线的紧缩性移动。
- 流动性陷阱（liquidity trap）：根据 IS-LM 模型，扩张性货币政策可以通过降低利率和刺激投资支出来发生作用，但是如果利率已经降到几乎为零，也许货币政策就不再有效。

6.4.3 蒙代尔-弗莱明模型

6.4.3.1 建立蒙代尔-弗莱明模型

前面的分析均在封闭经济下进行，研究开放经济下的货币政策和财政政策的主导政策范式是**蒙代尔-弗莱明模型**。关键假设：资本完全流动的小型开放经济。即：

$$r = r^* \tag{6.47}$$

该假设的合理性在于经济体相对于世界经济的规模而言非常小，使得它可以在世界进入市场上借入和借出任意量的资金而不影响世界利率。资本完全流动背后代表的过程是：如果国内货币市场的利率高于世界利率，资本会自发流入本国，使利率降低到世界利率水平，如果国内货币市场利率低于世界利率，资本会自发流出本国，使利率提高到世界利率水平。可以发行，套利是维持本国利率不变的动力。与 IS-LM 模型类似，产品市场的用如下方程代表：

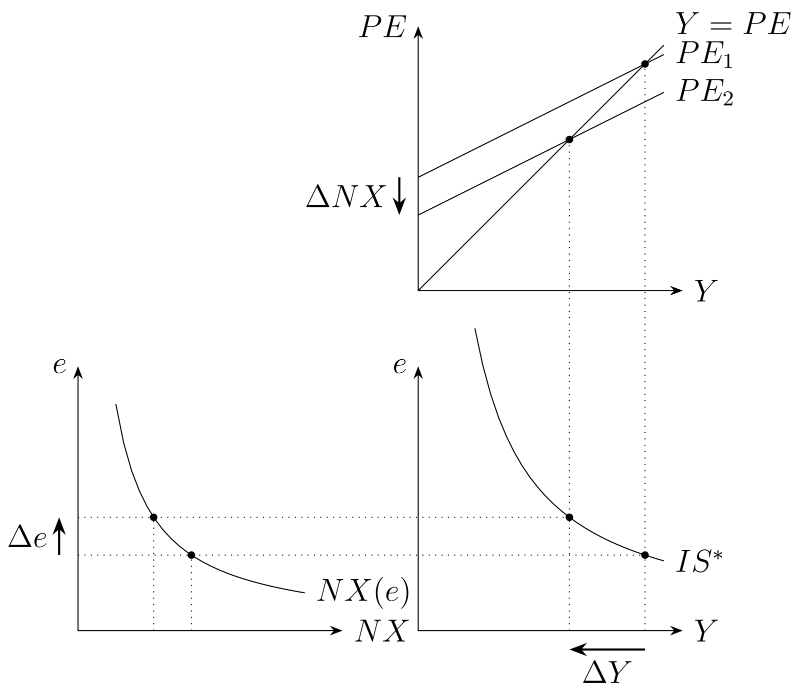

图 6.12: IS^* 曲线推导

$$IS^* : Y = C(Y - T) + I(r^*) + G + NX(e) \tag{6.48}$$

这里将利率固定在固定利率，并增加了净出口，净出口是名义汇率的负相关函数（这里假设国内外价格水平固定不变，因此名义汇率和实际汇率成比例 $\epsilon = e\frac{P_{monther}}{P_{foreign}}$）。$IS^*$ 曲线是向右下倾斜的曲线。

货币市场的方程如下：

$$LM^* : M/P = L(r^*, Y) \tag{6.49}$$

汇率 e 并没有进入该曲线，因此 LM^* 曲线是一条垂线。总的来看，资本完全流动的小型开放经济可以描述为：

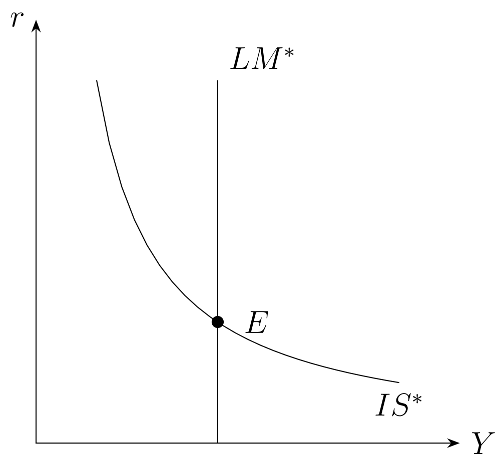

图 6.13: 蒙代尔-弗莱明模型

$$\begin{cases} Y = C(Y-T) + I(r^*) + G + NX(e), & IS^* \\ M/P = L(r^*, Y), & LM^* \end{cases} \tag{6.50}$$

其中的外生变量是财政政策 G 和 T、货币政策 M、价格水平 P 和世界利率 r^*，内生变量是收入 Y 和汇率 e。在两条曲线的交点，产品市场和货币市场都处于均衡状态。下面分析浮动汇率和固定汇率下的小型开放经济在财政政策、货币政策和贸易政策影响下的变化。

6.4.3.2 浮动汇率

浮动汇率：汇率由市场力量决定，可以随着经济状况的变动而波动。财政政策、货币政策、贸易政策对浮动汇率的影响如图6.14、6.15。

6.4.3.3 固定汇率

固定汇率：央行宣布一个法定汇率，并且随时准备买入和卖出本币来维持该汇率。固定汇率支配央行政策的过程如图6.16。

财政政策、货币政策、贸易政策对固定汇率的影响如图6.17、6.18。

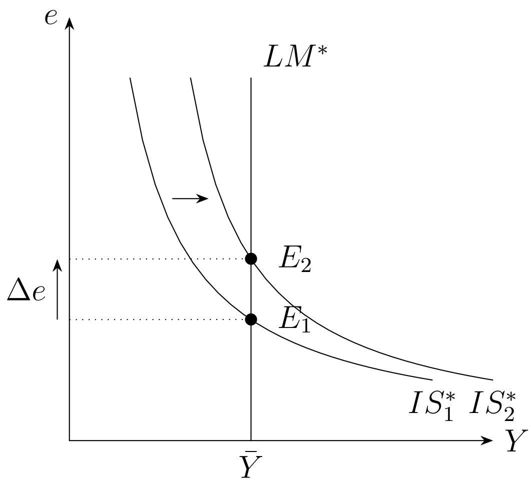

图 6.14: 扩张性财政政策、贸易政策-浮动汇率：$IS^* - LM^*$

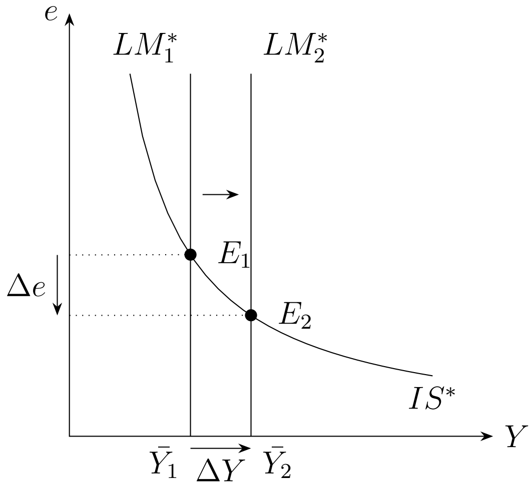

图 6.15: 扩张性货币政策-浮动汇率：$IS^* - LM^*$

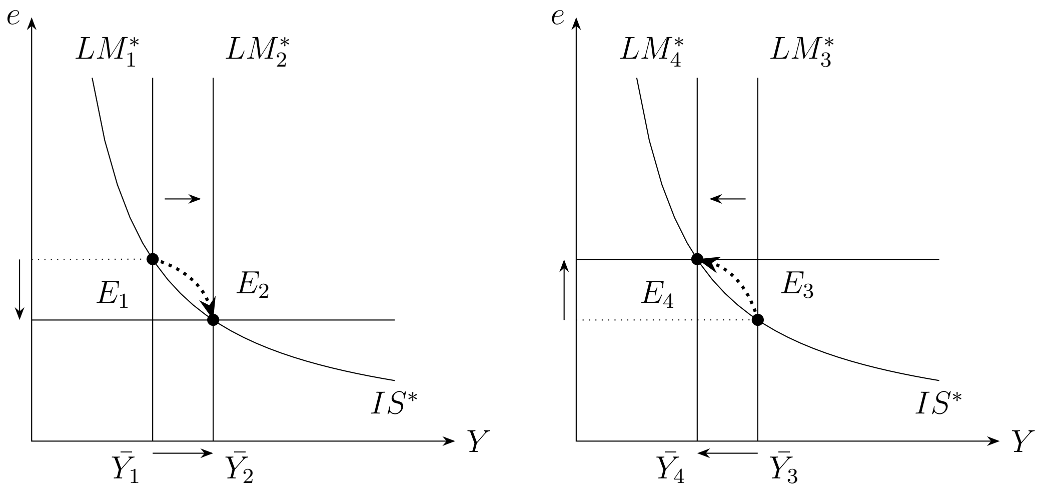

图 6.16: 固定汇率如何支配中央银行政策

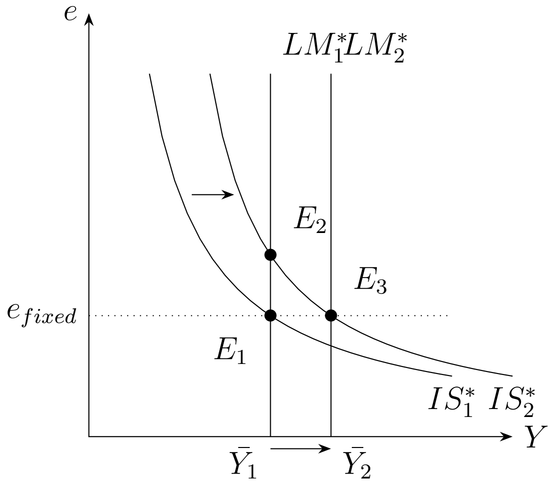

图 6.17: 扩张性财政政策、贸易政策-固定汇率：$IS^* - LM^*$

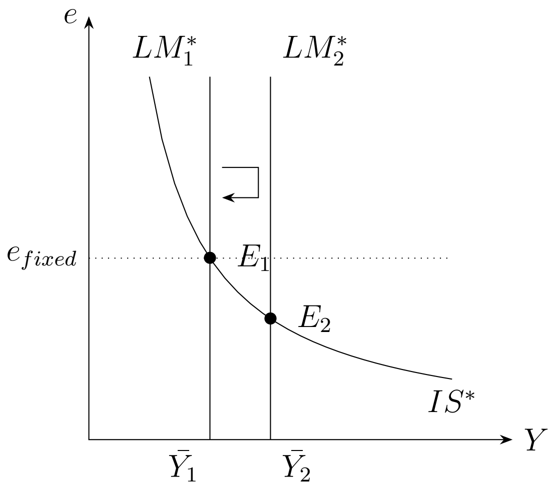

图 6.18: 扩张性货币政策-固定汇率：$IS^* - LM^*$

6.4.3.4 蒙代尔不可能三角

一国不可能同时拥有自由的资本流动、固定汇率和独立的货币政策。

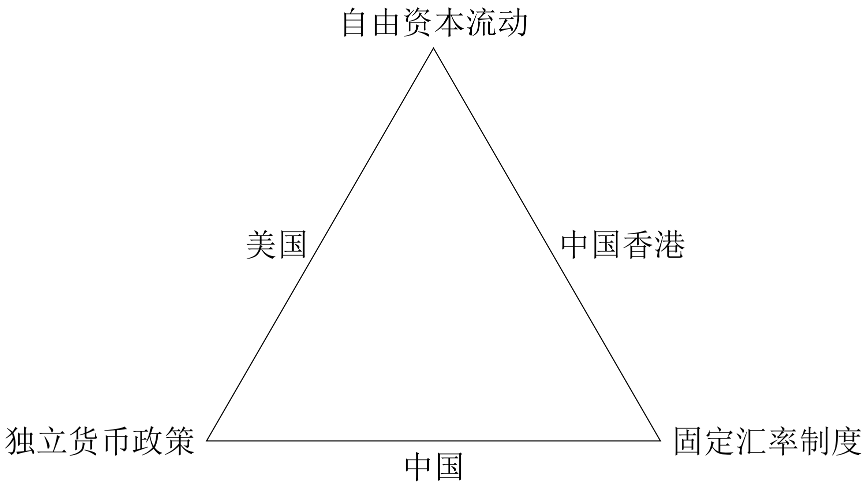

图 6.19: 不可能三角形

6.4.3.5 AD 曲线

在长期，价格水平变动时，资本完全流动的小型开放经济：

$$\begin{cases} Y = C(Y-T) + I(r^*) + G + NX(\epsilon), & IS^* \\ M/P = L(r^*, Y), & LM^* \end{cases} \tag{6.51}$$

此处净出口取决于实际汇率。当价格水平下降，实际货币余额上升，LM^* 曲线右移，收入上升，即 AD 曲线向右下方倾斜。

6.4.4 总供给、通胀和失业的短期权衡

IS-LM 模型是分析总需求的有力工具，那么总供给由什么决定呢？在短期，通胀和失业是描述经济最重要的指标，前者描述价格，后者决定产出，将价格和产出联系在一起的理论是菲利普斯曲线。首先分析短期总供给如何决定。两个经典的总供给模型的结论一致，市场的不完美性使产出偏离其自然水平，短期总供给曲线向右上方倾斜，方程表示如下：

$$Y = \bar{Y} + \alpha(P - EP), \alpha > 0 \tag{6.52}$$

其中 Y 为产出，$\bar{Y}$ 为自然产出水平，P 为价格水平，EP 为预期价格水平。当价格偏离预期价格时，产出就会偏离自然产出，偏离程度由 α 决定，$1/\alpha$ 是总供给曲线的斜率。下面分别介绍以下两个总供给模型。

6.4.4.1 黏性价格模型

黏性价格模型（sticky-price model）强调企业不能根据需求变动立刻调整他们索取的价格。当考虑企业设定价格时，我们假设企业具有某种程度的市场势力。企业面临的定价决策取决于价格总水平 P 和总收入水平 Y，前者决定生产成本，后者决定企业产品的需求，需求越大，边际成本越高，企业的合意价格也越高。企业的合意价格为：

$$p = P + \alpha(Y - \bar{Y}) \tag{6.53}$$

假设市场上一共两种企业，一类企业的价格有弹性，根据式(6.53)定价，另一些企业的价格有黏性，根据预期的经济状况事先宣布自己的价格，定价方式为：

$$p = EP + \alpha(EY - E\bar{\bar{Y}}) \tag{6.54}$$

假设预期产出等于自然产出，则式(6.54)为：

$$p = EP \tag{6.55}$$

如果黏性价格的企业占比为 s，弹性价格的企业占比为 $1 - s$，则价格水平为：

$$P = sEP + (1 - s)(P + \alpha(Y - \bar{Y})) \tag{6.56}$$

$$So, P = EP + \frac{(1 - s)\alpha}{s}(Y - \bar{Y}) \tag{6.57}$$

当企业预期高价格水平时，它们也预期高成本，因此产品定价也提高。当产出高时，对产品的需求也高，价格有弹性的企业就会设定高价格。因此价格水平取决于预期价格和产出。上式可以写成：$Y = \bar{Y} + \alpha(Y - \bar{Y})$。

6.4.4.2 不完备信息模型

假设市场出清，经济中的供给者只生产一种产品和多种消费品，由于信息不完备，他们有时候混淆了价格水平的变动和相对价格的变动。当实际价格超过预期价格时，供给者提高他们的产出。这意味着：

$$Y = \bar{Y} + \alpha(P - EP) \tag{6.58}$$

6.4.4.3 AS 曲线

短期总供给曲线总是想上倾斜：$Y = \bar{Y} + \alpha(P - EP)$。将总需求曲线和总供给曲线放在一起，未预期的总需求增加在短期使价格水平上涨，由于预期停留在原来的位置，产出超过自然产出水平，出现了短期的繁荣。长期的预期价格提高，短期总供给左移，产出回到自然的产水平，但是价格永久的上升。

6.4.4.4 菲利普斯曲线

菲利普斯曲线表明通胀取决于三种力量：预期的通胀率、周期性失业和供给冲击，即：

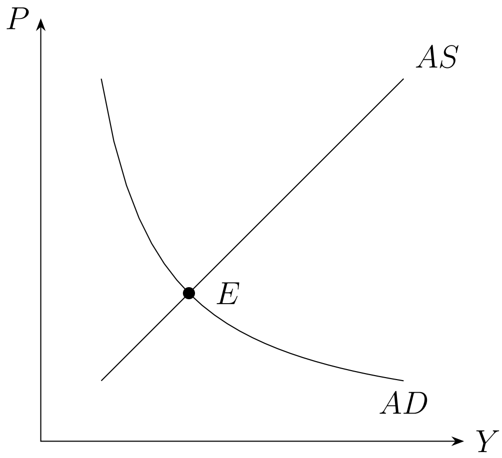

图 6.20: $AD - AS$ 模型

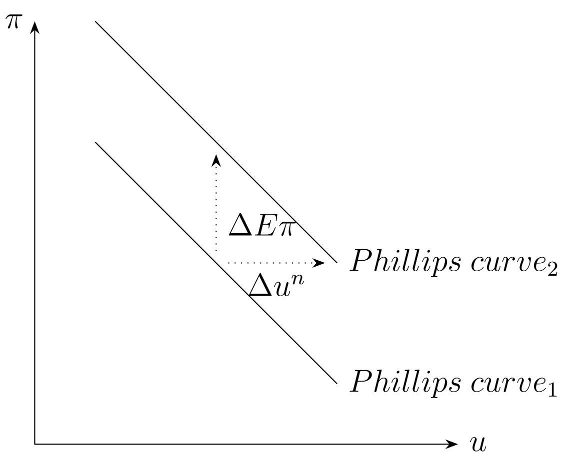

图 6.21: 菲利普斯曲线

$$\pi = E\pi - \beta(u - u^n) + v \tag{6.59}$$

菲利普斯曲线可以由总供给方程推导出来：

$$P = EP + (1/\alpha)(Y - \bar{Y}) + v \tag{6.60}$$
$$P - P_{-1} = EP - P_{-1} + (1/\alpha)(Y - \bar{Y}) + v \tag{6.61}$$
$$\pi = E\pi + (1/\alpha)(Y - \bar{Y}) + v \tag{6.62}$$
$$and\ Okun\ law : (1/\alpha)(Y - \bar{Y}) = -\beta(u - u^n) \tag{6.63}$$
$$So, \pi = E\pi - \beta(u - u^n) + v \tag{6.64}$$

菲利普斯曲线表明，失业与未预期到的通胀变化成负相关关系，总供给曲线和菲利普斯曲线是一块硬币的两面，前者适合研究价格和产出，后者适合研究通胀和失业。

适应性预期：$E\pi = \pi_{-1}$。此时 $\pi = \pi_{-1} - \beta(u - u^n) + v$。影响通胀的两种力量：（1）需求拉动的通胀，失业（2）成本拉动的通胀，供给冲击。短期菲利普斯曲线是向右下方倾斜的曲线，其位置取决于预期的通货膨胀，政策制定者面临通胀和失业之间的权衡。在长期，失业率回到自然率水平，通胀和失业之间不存在权衡。**牺牲率**是为使通货膨胀降低一个百分比而必须放弃的一年实际 GDP 的百分比。**理性预期**指人们可以充分利用可以获得的信息来预测未来。**自然率假说**：总需求的波动仅仅在短时间内影响产出和就业，在长期，经济回到古典模型所描述的产出、就业和失业水平。

6.5 宏观经济理论与政策

6.5.1 DAD-DAS 模型

6.5.1.1 推导 DAD-DAS 模型

DAD-DAS 模型的五个方程：

1. 产品与服务需求：$Y_t = \bar{Y}_t - \alpha(r_t - \rho) + \epsilon_t$。这里 ρ 是自然利率，ϵ_t 是需求的外生移动，政策制定者通过 r_t 影响需求，财政政策通过 ϵ_t 影响需求。
2. 实际利率：$r_t = i_t - E_t\pi_{t+1}$（费雪方程）
3. 通货膨胀：$\pi_t = E_{t-1}\pi_t + \phi(Y_t - \bar{Y}_t) + v_t$
4. 适应性预期：$E_t\pi_{t+1} = \pi_t$
5. 货币政策规则：$i_t = \pi_t + \rho + \theta_\pi(\pi_t - \pi_t^*) + \theta_Y(Y_t - \bar{Y}_t)$，$\theta_\pi, \theta_Y$ 的值决定了央行的货币政策偏向。

在长期均衡下，没有外生冲击 $\epsilon_t = v_t = 0$，并且通货膨胀稳定 $\pi_t = \pi_{t-1}$。得到五个内生

变量的长期值：

$$Y_t = \bar{Y}_t \tag{6.65}$$

$$r_t = \rho \tag{6.66}$$

$$\pi_t = \pi_t^* \tag{6.67}$$

$$E_t\pi_{t+1} = \pi_t^* \tag{6.68}$$

$$i_t = \rho + \pi_t^* \tag{6.69}$$

现在分别推导 DAS,DAD 曲线。将适应性预期带入菲利普斯曲线得到：

$$DAS : \pi_t = \pi_{t-1} + \phi(Y_t - \bar{Y}_t) + v_t \tag{6.70}$$

DAD 曲线则由产品与服务需求方程出发得到：

$$Y_t = \bar{Y}_t - \alpha(r_t - \rho) + \epsilon_t \tag{6.71}$$

$$Y_t = \bar{Y}_t - \alpha(i_t - E_t\pi_{t+1} - \rho) + \epsilon_t \tag{6.72}$$

$$Y_t = \bar{Y}_t - \alpha[\pi_t + \rho + \theta_\pi(\pi_t - \pi_t^*) + \theta_Y(Y_t - \bar{Y}_t) - \pi_t - \rho] + \epsilon_t \tag{6.73}$$

$$(1 + \alpha\theta_Y)(Y_t - \bar{Y}_t) = -\alpha\theta_\pi(\pi_t - \pi_t^*) + \epsilon_t \tag{6.74}$$

$$So, Y_t = \bar{Y}_t - \frac{\alpha\theta_\pi}{1 + \alpha\theta_Y}(\pi_t - \pi_t^*) + \frac{1}{1 + \alpha\theta_Y}\epsilon_t \tag{6.75}$$

$$\tag{6.76}$$

从而 DAD-DAS 模型:

$$\begin{cases} Y_t = \bar{Y}_t - \frac{\alpha\theta_\pi}{1+\alpha\theta_Y}(\pi_t - \pi_t^*) + \frac{1}{1+\alpha\theta_Y}\epsilon_t, & DAD \\ \pi_t = \pi_{t-1} + \phi(Y_t - \bar{Y}_t) + v_t, & DAS \end{cases} \tag{6.77}$$

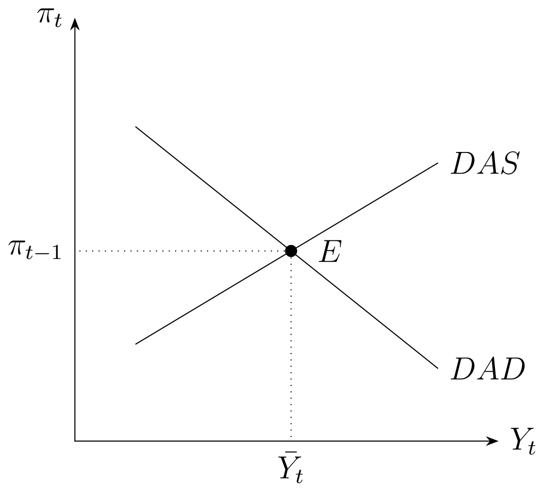

图 6.22: $DAD - DAS$ 模型

6.5.1.2 DAS-DAS 模型动态分析

经济从短期均衡向长期均衡移动，此时 DAD 曲线不包含前定变量，位置不移动，移动的是经济的通胀预期和 DAS 曲线（总供给）。即：

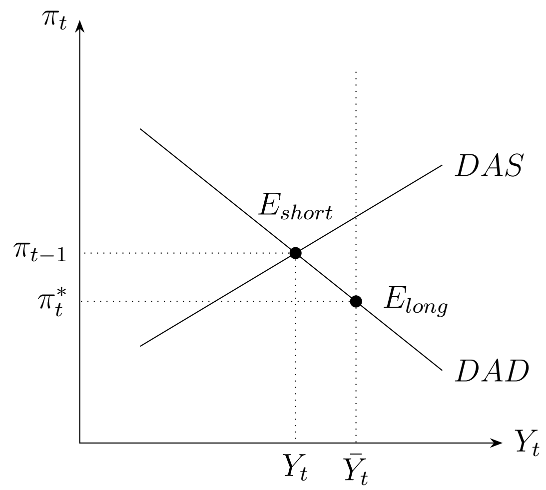

图 6.23: $DAD - DAS$ 模型：短期均衡向长期均衡移动

四种情境：（1）经济增长（2）供给冲击（3）需求冲击（4）货币政策转向下的 DAD-DAS 模型动态过程如图。

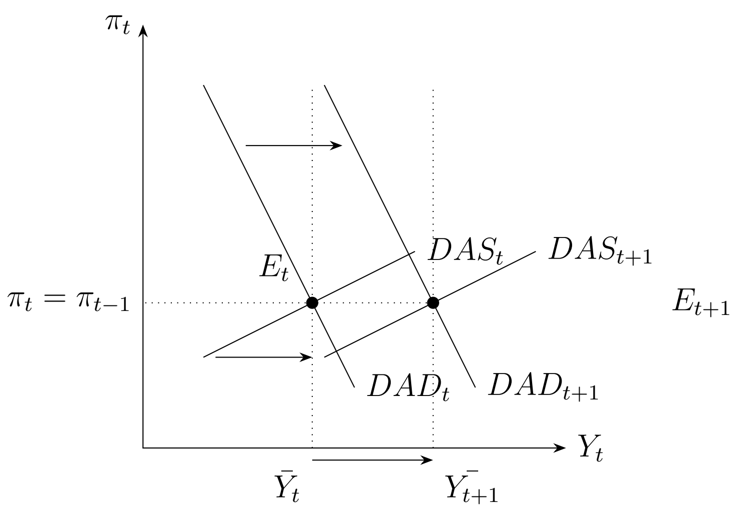

图 6.24: $DAD - DAS$ 模型：经济增长

6.5.2 稳定化政策

内在时滞：经济冲击与应对该冲击的政策行动之间的时间。**外在时滞**：政策行动与其对经济发生影响之间的时间。**自动稳定器**：不用采取任何有意的政策变动就可以在必要时刺激经济或者抑制经济的政策，如所得税制度、失业保障制度。**卢卡斯批判**：传统的政策评估方法没有充分考虑到政策对预期的影响。政策应该按照规则行事还是斟酌处置？对政策制定者和政治过程的不信任：

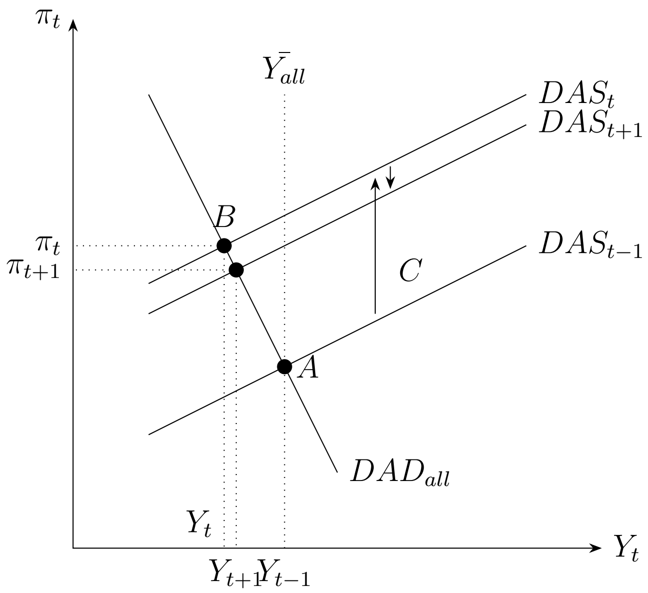

图 6.25: $DAD - DAS$ 模型：总供给冲击

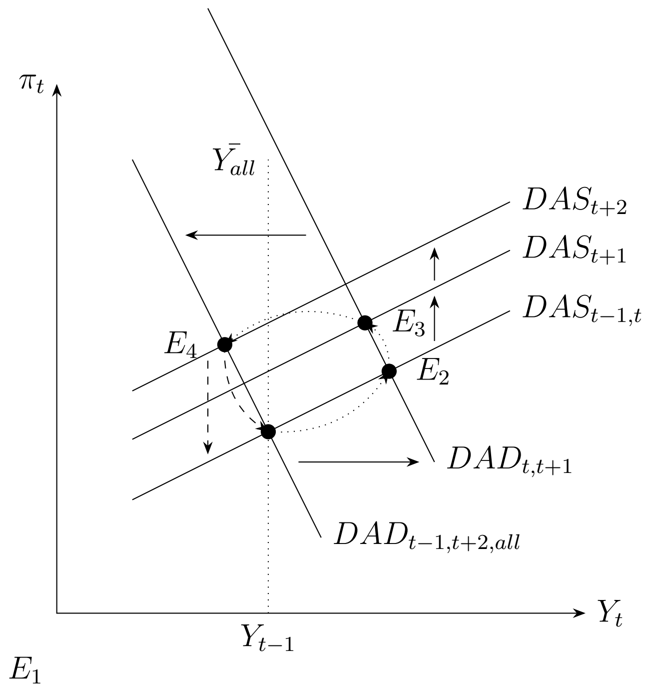

图 6.26: $DAD - DAS$ 模型：总需求冲击

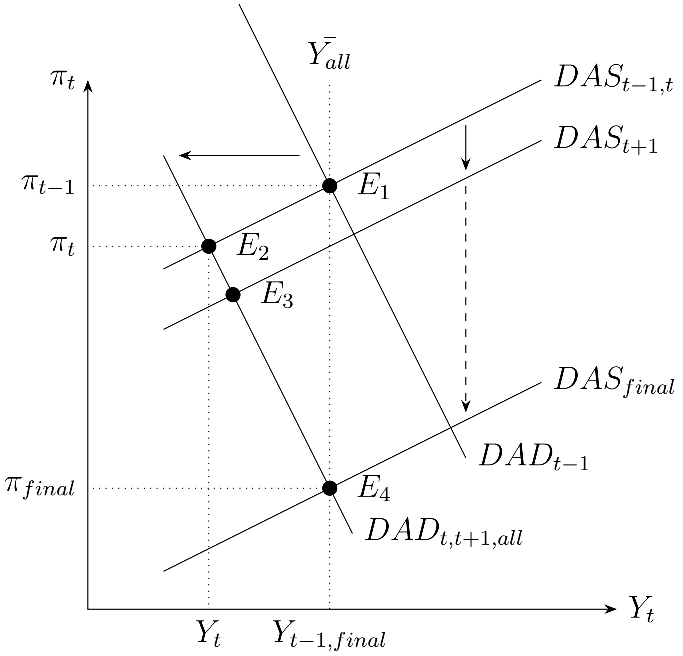

图 6.27: $DAD - DAS$ 模型：货币政策转向（降低通胀目标）

1. 政治家缺少宏观经济学知识来做出有效的判断
2. 政策制定者的目标和公众福利相冲突
3. 政治经济周期
4. 时间不一致性

货币政策规则：

1. 货币主义政策：按稳定的比率增加货币供给
2. 名义 GDP 目标制，产出和价格的高稳定性
3. 通货膨胀目标制，通胀的高稳定性

第 7 章 计量经济学

7.1 引言

计量经济学运用统计方法研究经济理论在现实世界中的适用性，为理论提供证据。计量经济学是理论经济学、数学和统计学的结合，在学习中要注重理论、模型和数据的有机结合，讲理论和做模型是统一的，首先用理论建立模型，然后用模型和数据所反映的经济关系修正理论，解决所研究的问题。本文借鉴伍德里奇《计量经济学》、李子奈《计量经济学》、陈强《计量经济学及 stata 应用》、汉密尔顿《时间序列分析》和郭凤鸣老师在《微观计量经济学》中所教授的内容。计量经济学研究的一般步骤为：

1. 确定研究问题
2. 设计回归模型
3. 收集整理数据
4. 估计模型参数
5. 分析估计结果

常用的数据类型包括：

1. 横截面数据：不同被观测对象在同一时间内的观测值
2. 时间序列数据：相同对象以时间为顺序的观测值
3. 混合横截面数据：不同观测时间不同观测对象的观测值
4. 面板数据：不同观测时间内相同被观测对象的观测值

7.2 经典计量经济学

7.2.1 线性回归模型

回归研究因变量对多个自变量的依赖关系,关心的是**均值**。总体回归方程（PRF）:$E(Y|X_i) = f(X_i)$，总体线性回归方程:$E(Y|X_i) = \beta_0 + \beta_1 * X_i$，个体回归方程（SRF）:$Y_i = \beta_0 + \beta_1 * X_i) + u_i$，$\beta_0$，截距项，$\beta_1$ 斜率系数，衡量 x_i 变化一单位时 y_i 的变化,u_i 误差项。线性指对参数是线性的。

随机扰动项产生的原因：

1. 未知的影响因素（模糊的经济理论）
2. 无法获得的数据（数据局限）
3. 众多细小影响因素（简洁性原则）
4. 变量观测误差（观测局限）
5. 模型设定误差（先验错误）
6. 变量的内在随机性（随机经济行为）

样本回归函数（SRF）

$$\hat{Y} = f(X) = \hat{\beta}_0 + \hat{\beta}_1 * X \tag{7.1}$$

随机形式：

$$Y = \hat{Y} + \hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 * X + e \tag{7.2}$$

!!! 加图：理解 PRF 和 SRF

通常可得的数据是总体的一部分样本，我们希望样本能够尽量代表总体。计量经济学就是通过样本回归函数来推断总体回归函数，在这里就是用 $\hat{\beta}_0$，$\hat{\beta}_1$ 来推断 β_0，β_1。

假设依据经济理论建立以上的回归模型，首要的任务就是估计其中的参数。假设数据是完整的（没有缺失值），并且也没有极端值。在高斯马尔可夫假设下，OLS 估计值是无偏一致最有效的估计量（BLUE）。这五个假设如下，请关注这些假设在系数的估计和特征推理中的重要作用！

1. 线性于参数（模型设定证券）
2. 不存在完全共线性（满秩）
3. 零条件均值 $E(u_i|X) = 0$
4. 同方差 $Var(u_i|X) = \sigma^2, i = 1\ N$
5. 序列不相关 $Cov(u_i, u_j|X) = 0, i \neq j$

7.2.1.1 OLS 参数估计

运用样本信息估计样本回归函数，可以采用最小化离差平方和的方法，即普通最小二乘法估计（OLS），数学表达如下：

$$min\ Q = \sum_1^N e_i^2 = \sum_1^N (Y_i - \hat{Y}_i)^2 \tag{7.3}$$

首先探讨一元线性回归模型的估计和参数分析。式(7.3)变为：

$$min\ Q = \sum_1^N (Y_i - \hat{\beta}_0 - \hat{\beta}_1 * X_i)^2 \tag{7.4}$$

式(7.4)的一阶最优化条件为：

$$\begin{cases} \frac{\partial Q}{\partial \hat{\beta}_0} = -\sum_1^N e_i = 0 \\ \frac{\partial Q}{\partial \hat{\beta}_1} = -2 * \sum_1^N (X_i * e_i) = 0 \end{cases} \tag{7.5}$$

这里 $e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1$。式(7.5)表明，OLS 估计是一种正交估计，残差和向量 $1 = [1, ..., 1]'X = [X_1, ..., X_N]'$ 均正交（也和 $\hat{Y}$ 正交）！并且解出该式：

$$\begin{cases} \hat{\beta}_0 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{N \sum X_i^2 - \sum X_i{}^2} \\ \hat{\beta}_1 = \frac{N \sum X_i Y_i - \sum X_i \sum X_i}{N \sum X_i^2 - \sum X_i{}^2} \end{cases} \tag{7.6}$$

式(7.6)也称为正规方程式，稍微变化其表达方式为：

$$\begin{cases} \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \\ \hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{(\sum (X_i - \bar{X})^2} \end{cases} \tag{7.7}$$

并且其离差形式为 $\hat{x}_i = \hat{X}_i - \bar{X}, \hat{y}_i = \hat{Y}_i - \bar{Y}$，

$$\begin{cases} \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \\ \hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} \end{cases} \tag{7.8}$$

并且有 $\bar{Y} = \bar{\hat{Y}}_i$，

$$\hat{y}_i = \hat{Y}_i - \bar{Y} = (\hat{\beta}_0 + \hat{\beta}_1 X_i) - (\hat{\beta}_0 + \hat{\beta}_1 \bar{X} + \bar{e}) = \hat{\beta}_1 x_i \tag{7.9}$$

式(7.9)是样本回归函数的离差形式。

7.2.1.2 拟合优度

线性回归模型首先是在用数据找模型，判断模型拟合程度的指标就是**拟合优度**，即模型所能够解释的总体的比例，用下式计算：

$$R^2 = SSE/SST = 1 - SSR/SST \tag{7.10}$$

其中，总体平方和 $SST = \sum_1^N y_i^2 = \sum_1^N (Y_i - \bar{Y})^2$，解释平方和 $SSE = \sum_1^N \hat{y_i}^2 = \sum_1^N (\hat{Y}_i - \bar{Y})^2$，残差平方和 $SSE = \sum_1^N e_i^2 = \sum_1^N (Y_i - \hat{Y})^2$。其可以证明 $SET = SSE + SSR$。R^2 越接近 1，模型对数据的拟合程度越好。同时，更重要的是 $\hat{\beta}_i, i = 0, 1$ 和总体回归函数的 $\beta_i, i = 0, 1$ 相差多远？下面说明高斯马尔可夫假设下 ols 估计的小样本性质（线性性、无偏性、有效性）和大样本性质（一致性）。

7.2.1.3 OLS 估计量性质

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i (Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum x_i Y_i}{\sum x_i^2} + \frac{\bar{Y} \sum x_i}{\sum x_i^2} = \sum k_i Y_i \tag{7.11}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \sum (1/n - \bar{X} k_i) Y_i = \sum w_i Y_i \tag{7.12}$$

其中 $k_i = \frac{x_i}{\sum x_i^2}, w_i = 1/n - \bar{X} k_i$。并且 $\sum k_i = 0, X_i = x_i + \bar{X}, \sum k_i X_i = 1, \sum w_i = 1, \sum w_i X_i = 0$，接着将其分解，

$$\hat{\beta}_1 = \sum k_i Y_i = \sum k_i(\beta_0 + \beta_1 X_i + u_i) = \beta_0 \sum k_i + \beta_1 \sum k_i X_i + \sum k_i u_i = \beta_1 + \sum k_i u_i \tag{7.13}$$

$$E(\hat{\beta}_1|X) = \beta_1 + \sum k_i E(u_i|X) = \beta_1 \tag{7.14}$$

$$\hat{\beta}_0 = \sum w_i Y_i = \sum w_i(\beta_0 + \beta_1 X_i + u_i) = \beta_0 + \sum w_i u_i \tag{7.15}$$

$$E(\hat{\beta}_0|X) = \beta_0 + \sum w_i E(u_i|X) = \beta_0 \tag{7.16}$$

上式证明了 OLS 估计量的无偏性，但是证明无偏性时只使用了零条件均值这一条件，但是隐含着线性于参数这一条件，这两个条件意味着：总体是由 PRF 生成的，并且抽样是随机的（误差严格外生），样本可以代表总体，抽样误差用 u_i 表示。

现在计算 OLS 估计量的方差，这在统计推断时十分有用。需要考虑的是，估计量依据什么而变动？如果总体根据 PRF 生产，那么误差 u 就是 0，用抽样样本和 SRF 构造的残差 e 也是 0，因此 OLS 估计量唯一。因此估计量的方差来自于总体误差，现在根据同方差假设计算 OLS 估计量的方差。依据 $Var(u_i|X) = \sigma^2, i = 1\ N$：

$$Var(\hat{\beta}_1|X) = Var(\beta_1 + \sum k_i u_i|X) = \sum k_i^2 \sigma^2 = \sigma^2 \frac{\sum x_i^2}{(\sum x_i^2)^2} = \frac{\sigma^2}{SST_X} \tag{7.17}$$

$$Var(\hat{\beta}_0|X) = Var(\beta_0 + \sum w_i u_i|X) = \sum w_i^2 \sigma^2 = \frac{\sigma^2 \sum X_i^2}{N \sum x_i^2} \tag{7.18}$$

这里 $SST_X = \sum x_i^2 = \sum (X_i - \bar{X})^2$，$\sum w_i^2 = \sum (1/N - \bar{X} k_i)^2 = \sum (1/N^2 - 2\bar{X} k_i + \bar{X}^2 k_i^2) = 1/N - 2\bar{X} \sum k_i + \bar{X}^2 \sum k_i^2 = 1/N + \bar{X}^2 / \sum x_i^2 = \sum X_i^2 / (N \sum x_i^2)$。可以证明，该估计量方差在所有的线性估计量中是最小的。

高斯-马尔可夫定理：普通最小二乘估计量具有线性性、无偏性和有效性等性质，是最优线性无偏估计量（BLUE）。

除了无偏性和有效性，我们还希望当样本量足够大时估计量可以无限逼近真实值，即具有一致性。大样本性质在大数据时代尤其有用。仍然从 $\beta_i, i = 0, 1$ 的形式上入手：

$$Plim(\hat{\beta}_1) = Plim(\beta_1 + \sum k_i u_i) = \beta_1 + Plim \frac{\sum x_i u_i}{\sum x_i^2} = \beta_1 + \frac{Plim(\sum x_i u_i / N)}{Plim(\sum x_i^2 / N)} \tag{7.19}$$

$$Plim(\hat{\beta}_0) = \beta_0 + Plim \frac{\sum w_i u_i}{\sum x_i^2} = \beta_0 + Plim(\sum u_i / N) - \bar{X} Plim(\sum k_i u_i) \tag{7.20}$$

上式中 $Plim(\sum x_i u_i / N) = Cov(X_i, u_i), Plim(\sum x_i^2 / N) = Var(X_i)$。假设 $Cov(X_i, u_i) = 0, \bar{u} = 0$，即 X_i, u_i 没有同期线性相关且后者均值为 0，有 $Plim(\hat{\beta}_1) = \beta_1, Plim(\hat{\beta}_0) = \beta_0$。注意同期线性无关的假设比零条件均值的假设弱得多。

总的来说，OLS 估计量具有正交性、线性性、无偏性、有效性和一致性。但是存在一个问题，

那就是 σ 是多少呢？总体是未知的，可获得的只有样本和样本回归函数的残差项 $e_i, i = 1\ N$。

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{7.21}$$

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \tag{7.22}$$

$$e_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) X_i \tag{7.23}$$

$$\bar{e} = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1) \bar{X} \tag{7.24}$$

$$e_i - \bar{e} = u_i - \bar{u} - (\hat{\beta}_1 - \beta_1)(X_i - \bar{X}) \tag{7.25}$$

$$\sum e_i^2 = \sum (u_i - \bar{u})^2 + \sum (\hat{\beta}_1 - \beta_1)^2 (X_i - \bar{X})^2 - 2 \sum (u_i - \bar{u})(\hat{\beta}_1 - \beta_1)(X_i - \bar{X}) \tag{7.26}$$

$$E(\sum e_i^2) = E(\sum (u_i - \bar{u})^2) + E((\sum k_i u_i)^2 \sum x_i^2) - 2E(\sum k_i u_i \sum x_i u_i) \tag{7.27}$$

$$E(\sum e_i^2) = (N-1)\sigma^2 + (Var(\sum k_i u_i) - (E(\sum k_i u_i)) SST_X - 2\sigma^2 E(\sum k_i x_i) \tag{7.28}$$

$$E(\sum e_i^2) = (N-1)\sigma^2 + \sigma^2 - 2\sigma^2 = (N-2)\sigma^2 \tag{7.29}$$

$$E(SSR/(N-2)) = \sigma^2 \tag{7.30}$$

这里根据估计量的正交性，$\bar{e} = 0$，并且需要 $Cov(u_i, u_j) = 0, i \neq j$。根据 k_i 的表达式，$\sum k_i x_i = 1$。现在，可以开始统计推断的工作。

7.2.1.4 统计推断

假设误差服从正态分布，即 $u_i\ N(0, \sigma^2)$。由于 $\hat{\beta}_i, i = 0, 1$ 是 u_i 的线性函数，因此也服从正态分布，即 $\hat{\beta}_1\ N(\beta_1, \frac{\sigma^2}{SST_X}), \hat{\beta}_0\ N(\beta_0, \frac{\sigma^2 \sum X_i^2}{N \sum x_i^2})$。如果 $sigma^2$ 已知，可以标准化后做统计推断。但是往往未知 σ^2，此时可以用 σ^2 的无偏估计 $SSR/(N-2)$ 替代 σ^2，可以证明:

$$t = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\hat{\sigma}^2 / \sum x_i^2}}\ t(N-2) \tag{7.31}$$

$$t = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \sum X_i^2 / (N \sum x_i^2)}}\ t(N-2) \tag{7.32}$$

在正态分布小样本或者大样本情况下统计量(7.31)和(7.32)都成立。通常报告的显著性是针对 $H_0: \beta_i, i = 0, 1$ 的检验，判断显著性的方式有两个，一是比较估计值和阈值，二是计算 P 值。如在 5% 的显著性水平下的双边检验，当 $|t| > t_{0.025}(N-2)$，系数在 5% 显著性水平下显著的。通常直接利用软件输出的 P 值会比较方便。而参数估计值的置信区间如下：

$$\hat{\beta}_i - t_{\alpha/2} S_{\hat{\beta}_i} < \beta_i < \hat{\beta}_i + t_{\alpha/2} S_{\hat{\beta}_i}, i = 0, 1 \tag{7.33}$$

式(7.33)中 $S_{\hat{\beta}_i}$ 是标准误，即用系数标准差的估计量。

7.2.1.5 预测

线性回归模型运用已有的样本信息预测被接受变量，信息包括：用于估计 SRF 的样本和预测样本的解释变量。预测包括点预测和区间预测，变量预测包括均值预测和个别值的预测。

首先进行均值预测：

$$E(Y_i) = \beta_0 + \beta_1 X_i + E(u_i) = \beta_0 + \beta_1 X_i \tag{7.34}$$

$$E(\hat{Y}_i) = E(\hat{\beta}_0) + E(\hat{\beta}_1 X_i) = \beta_0 + \beta_1 X_i \tag{7.35}$$

$$E(\hat{Y}_i) = E(Y_i) \tag{7.36}$$

$$\tag{7.37}$$

即 $\hat{Y}_i$ 是 Y_i 的无偏的均值预测。下面进行均值区间预测，

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\sum k_i u_i, \sum w_i u_i) = \sigma^2 \sum k_i w_i \tag{7.38}$$

$$\sum k_i w_i = \sum (1/N - \bar{X} k_i) k_i = (1/N \sum k_i - \bar{X} \sum k_i^2) = -\bar{X} / \sum x_i^2 \tag{7.39}$$

$$Var(\hat{Y}_i) = Var(\hat{\beta}_0) + Var(\hat{\beta}_1 X_i) + 2Cov(\hat{\beta}_0, \hat{\beta}_1 X_i) = \frac{\sigma^2 \sum X_i^2}{N \sum x_i^2} + \frac{\sigma^2 X_i^2}{\sum x_i^2} - 2\frac{\sigma^2 X_i \bar{X}}{\sum x_i^2} \tag{7.40}$$

$$Var(\hat{Y}_i) = \sigma^2 (\frac{1}{N} + \frac{(X_i - \bar{X})^2}{\sum x_i^2}) \tag{7.41}$$

从而 $\hat{Y}_i\ N(Y_i, \sigma^2(\frac{1}{N} + \frac{(X_i - \bar{X})^2}{\sum x_i^2}))$。构造统计量：

$$t = \frac{\hat{Y}_i - Y_i}{S_{\hat{Y}_i}}\ t(N-2) \tag{7.42}$$

式(7.42)中，$S_{\hat{Y}_i}$ 是用 $\hat{\sigma}^2$ 替代 σ^2 的标准误。因此总体条件均值的预测区间为：

$$\hat{Y}_i - t_{\alpha/2} S_{\hat{Y}_i} < E(Y|X_i) < \hat{Y}_i + t_{\alpha/2} S_{\hat{Y}_i} \tag{7.43}$$

这里 α 为显著性水平。当 N 无限增大时，该区间的半径无限缩小到下限 $t_{\alpha/2}\hat{\sigma}^2 \frac{(X_i - \bar{X})^2}{\sum x_i^2}$，在 $\bar{X}$ 附近，预测区间半径最小。接下来进行总体个别值的区间预测，

$$Y_i\ N(\beta_0 + \beta_1 X_i, \sigma^2) \tag{7.44}$$

$$\hat{Y}_i\ N(\beta_0 + \beta_1 X_i, \sigma^2 (\frac{1}{N} + \frac{(X_i - \bar{X})^2}{\sum x_i^2})) \tag{7.45}$$

$$e_i = Y_i - \hat{Y}_i\ N(0, \sigma^2 (1 + \frac{1}{N} + \frac{(X_i - \bar{X})^2}{\sum x_i^2})) \tag{7.46}$$

故 $t = \frac{e_i - 0}{S_{e_i}}\ t(N-2)$，得到 Y_i 的置信区间为，

$$\hat{Y}_i - t_{\alpha/2} S_{e_i} < E(Y|X_i) < \hat{Y}_i + t_{\alpha/2} S_{e_i} \tag{7.47}$$

显然个别值的预测区间比均值的预测区间更大，因为它包含了数据生成过程中的误差。

7.2.1.6 多元线性回归模型

考虑多元线性回归模型，以上的结论不变，但是各表达式有所变化。多元线性回归模型的估计量、统计推断和预测的结果分别如下：

总体回归函数：

$$Y_i = X_i \beta + u_i, i = 1\ N \tag{7.48}$$

其中，$X_i = [1, X_{i,1}, ..., X_{i,k}], \beta = [\beta_0, \beta_1, ..., \beta_k]$。

样本回归函数：

$$Y_i = X_i\hat{\beta} + e_i, i = 1\ N \tag{7.49}$$

其中，$X_i = [1, X_{i,1}, ..., X_{i,k}], \hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k]$。

矩阵表达式分别为:$Y = X\beta + u, Y = X\hat{\beta} + e$。并假设 $Plim X'X/N \to Q, E(u|X) = 0, Var(u|X) = \sigma^2 I$ 和误差正态分布 $u|X\ N(0, \sigma^2 I)$。

OLS 估计量：

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{7.50}$$

同样满足正交性、线性性、无偏性、有效性、一致性，$Ie = 0, X'e = 0, I = [1, ..., 1]$，且 $\hat{\sigma}^2 = \frac{e'e}{N-k-1}$。拟合优度和调整的拟合优度为，

$$R^2 = 1 - \frac{SSR}{SST} \tag{7.51}$$

$$\bar{R}^2 = 1 - \frac{SSR/(N-k-1)}{SST/(N-1)} \tag{7.52}$$

此时，$E(\hat{\beta}) = \beta, Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$。令 $C = (X'X)^{-1} = c_{i,j}$，则 $Var(\beta_i) = \sigma^2 c_{ii}, i = 1\ k$，单个系数的显著性检验与一元线性回归模型一致。多个系数的联合检验常常使用 F 检验，即检验 m 个约束条件是否同时成立。设约束回归模型为 r，无约束回归模型为 ur，则

$$F = \frac{(SSR_r - SSR_{ur})/m}{SSR/(N-k)}\ F(m, N-k) \tag{7.53}$$

$$\text{OR } F = \frac{(R_{ur} - R_r)/m}{1 - R_{ur}/(N-k)}\ F(m, N-k) \tag{7.54}$$

进行模型整体显著性检验时，$H_0 : \beta = 0$，即 $m = k, R_r = 0$。在总体预测时，需要的方差如下：

$$Var(\hat{Y}_i) = \sigma^2 X_i(X'X)^{-1}X_i' \tag{7.55}$$

$$Var(e_i) = \sigma^2(1 + X_i(X'X)^{-1}X_i') \tag{7.56}$$

在这里，有必要重新总结高斯-马尔可夫假定。

1. 线性于参数
2. 严格外生性 $E(u_i|X) = E(u_i|x_1, ..., x_N) = 0, i = 1\ N$
3. 不存在完全共线性 $Rank(X) = k + 1$
4. 球形扰动项 $Var(u|X) = \sigma^2 I$

记住，小样本推断时还需要正态分布假定！大样本推断则不需要。违法第一条假设，说明数据生成过程判断失误，必须采用非线性模型；违法第二条假设，说明存在内生性问题，这是论文最关注的问题，因为它代表经济理论设定错误，整个微观计量都是在解决内生性问题；违反第三条假设，则存在多重共线性问题，这只会导致系数标准误变大，统计检验失误；违法第四条假设，说明存在自相关和异方差问题，这在时间序列模型中非常常见。这里，本章只交代多重共线性、异方差和自相关问题如何检验和解决，内生性问题在微观计量中会详细分析。

7.2.2 多重共线性

多重共线性（multicollinearity）指多个解释变量之间出现了相关性。对于总体回归函数 $Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_k X_{ik}$，如果存在不全为 0 的系数 $c_i, i = 1\ k$，有 $c_1 X_{i1} + ... + c_k X_{ik} = 0$，则存在完全多重共线性。如果存在随机误差项使 $c_1 X_{i1} + ... + c_k X_{ik} + v_i = 0$ 成立，则存在近似多重共线性。产生多重共线性的原因包括：

1. 经济变量共同趋势
2. 模型设定错误：包含滞后项、重复度量等
3. 截面数据相依性
4. 样本选择问题

多重共线性对模型估计和统计推断有什么影响呢？请自行用二元线性回归模型进行分析。

1. 完全多重共线性：参数估计量不存在

   由于 $R(X) < k+1$，$X'X$ 不满秩，因此 $(X'X)^{-1}$ 不存在，OLS 估计量 $\hat{\beta} = (X'X)^{-1}X'Y$ 无法计算。
2. 近似多重共线性：参数估计量方差变大

   此时 $|X'X|! = 0$（约等于），$(X'X)^{-1}$ 主对角线元素较大，使得参数估计量方差变大，影响统计推断的准确性。
3. 参数经济意义模糊

   参数表征具有共线性的变量对被解释变量的共同影响，无法分离各自的影响，因此符号和大小与不可信。
4. 模型显著性检验和预测无效

   由于系数估计量方差变大，构造的 t 统计量变小，导致错误接受系数不显著的结果。同时预测的区间会变大，有效性降低。此时，拟合优度非常高，模型显著性检验的 F 检验的 P 值很大，但是单个系数却并不显著。

多重共线性如何检验？从定义和影响两方面检验。

1. 简单相关系数检验：报告相关系数表，如果 $r_{i,j}$ 大于 0.8，可初步判断存在多重共线性。
2. 方差膨胀因子 (VIF) 检验:将 X_j 对其他变量做辅助回归,得到 R_j^2,则 $Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_i^2}\frac{1}{1-R_j^2} = \frac{\sigma^2}{\sum x_i^2}VIF_j$。当 $VIF_j \geq 10$ 时，变量 X_j 与其他解释变量存在严重多重共线性。
3. 直观判断：
   - 调整解释变量或观测值导致参数估计值发生很大变化
   - 理论上重要的解释变量没有通过显著性检验
   - 参数符号与经济理论相违背
   - R^2 和 F 检验显著但是重要解释变量的 t 检验不显著
4. 逐步回归：将变量逐个引入模型，如果拟合优度无显著变化，则该变量不是独立解释变量，如果拟合优度显著提高，则该变量是独立解释变量。

多重共线性的补救措施：

1. 剔除引起多重共线性的变量

2. 增大样本容量
3. 改变模型的形式，如差分、比值
4. 引入非样本先验信息
5. 横截面数据和时间序列数据并用
6. 变量变换：相对指标、实际指标、大类指标、对数指标
7. 逐步回归

7.2.3 异方差

异方差：$Var(u_i|X) = \sigma_i^2 = f(X_i) \neq \sigma^2$。此时 OLS 估计量的无偏性和一致性并不受影响，但是有效性和统计推断不再成立。异方差产生的原因：

1. 模型设定误差，省略相关的解释变量
2. 测量误差的变化
3. 截面数据中总体各单位的差异

异方差的后果：（由于无法获得估计量的方差导致）

1. 参数估计量非有效
2. 变量显著性检验无效
3. 模型预测失效

异方差的检验：

1. 观察残差图的形态：e-y,e-x
2. B-P 检验 (Breusch-Pagan)：
   假设异方差的生成过程为 $u_i^2 = \delta_0 + \delta_1 X_{i1} + ... + \delta_k X_{ik} + \epsilon_i$，则原假设为 $H_0 : \delta_1 = ... = \delta_k = 0$。由于 u_i 未知，可以用 e_i 替代。即回归：$e_i^2 = \delta_0 + \delta_1 X_{i1} + ... + \delta_k X_{ik} + error_i$，构造检验统计量进行检验：
   $$LM = NR_e^2\ \chi^2(k-1) \text{ OR } F = \frac{R_e^2/k}{(1-R_e^2)/(N-k-1)}\ F(k, N-k-1) \tag{7.57}$$
3. White 检验：
   在 B-P 检验的基础上加入高次项，如二元回归模型：$e_i^2 = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \delta_3 X_{i1}^2 + \delta_4 X_{i2}^2 + \delta_5 X_{i1} X_{i2} + error_i$，此时 $H_0 : \delta_1 = ... = \delta_5 = 0$。由于这样会损失较多的自由度，可以用 Y_i 的高次幂替代交叉项。

异方差的补救措施：

1. OLS+ 稳健标准误
   怀特提出用 e_i^2 替代 σ_i^2。即在一元回归模型中：$Var(\beta_1|X) = \frac{\sum x_i e_i}{(\sum x_i^2)^2}$。
2. WLS
   如果已知 $\sigma_i^2 = \sigma^2 f(X_i)$,则可调整总体回归模型:$Y_i/\sqrt{f(X_i)} = \beta_0/\sqrt{f(X_i)} + \beta_1 X_{i1}/\sqrt{f(X_i)} + ... + \beta_k X_{ik}\sqrt{f(X_i)} + u_i/\sqrt{f(X_i)}$，此时 $V(u_i/\sqrt{f(X_i)}|X) = \sigma^2$。
3. FWLS
   但是 $f(X_i)$ 的形式往往未知，可以用残差估计。在 B-P 检验中的辅助回归的基础上取对

数保证方差为正：$lne_i^2 = \delta_0 + \delta_1 X_{i1} + ... + \delta_k X_{ik} + error_i$。得到预测值 $ln\hat{\sigma_i^2}$，则拟合值 $\hat{\sigma_i^2} = exp(ln\hat{\sigma_i^2})$。故权重为 $1/\hat{\sigma_i^2}$，然后进行 WLS 估计。

7.2.4 自相关

自相关（autocorrelation）：存在 $i \neq j$ 使得 $Cov(u_i, u_j) = E(u_i u_j) \neq 0$。这使得 $Var(u_i|X)$ 非对角线元素不为 0，$Var(\hat{\beta}|X) \neq \sigma^2(X'X)^{-1}$，统计检验和预测失效，系数的估计量方差和误差方差将被低估，双重低估导致 t 检验特别显著。并且 R^2 和 F 检验 P 值也非常高。

自相关产生的原因：

1. 经济系统惯性
2. 经济活动滞后效应
3. 数据处理的后果
4. 蛛网现象
5. 模型设定偏误

自相关的检验方法：

1. 图示检验：$e_t - e_{t-1}$ 和 $e_t - t$
2. B-G 检验：考虑自相关的生成过程为 $u_t = \rho_1 u_{t-1} + ... + \rho_p u_{t-p} + v_t$，$v_t$ 满足古典假定。$H_0 : \rho_1 = ... = \rho_p = 0$。检验过程：
   - OLS 估计原模型，得到 e_t
   - 辅助回归：$u_t = \rho_1 u_{t-1} + ... + \rho_p u_{t-p} + \delta X_{i1} + ... + \delta X_{ik} + v_t, t = p+1, ..., N$
   - 原假设 $H_0 : \rho_1 = ... = \rho_p = 0$，LM 统计量 $LM = (N-p)R^2\ \chi^2(p)$
3. D-W 检验：只能检验一阶自相关，要求解释变量严格外生。$DW = \frac{\sum_{t=2}^N (e_t - e_{t-1})^2}{\sum_{t=1}^N e_i^2} = 2(1-\hat{\rho}_1)$。D-W 通过阈值比较进行检验，检验方法在经典教科书上都有。
4. Q 检验：$H_0 : \rho_1 = ... = \rho_p = 0$。$\hat{\rho} = \frac{\sum_{t=j+1}^N e_t e_{t-j}}{\sum_{t=1}^N e_t^2}$
   - Box-Pierce Q 统计量:$Q_{BP} = N \sum_{j=1}^p \hat{\rho_j}^2\ \chi^2(p)$
   - Ljung-Box Q 统计量:$Q_{LB} = N(N+2) \sum_{j=1}^p \frac{\hat{\rho_j}^2}{N-j}\ \chi^2(p)$

自相关的补救办法:

1. OLS+ 异方差自相关稳健标准误 HAC
2. 准差分：已知自相关生成原因为 $u_t = \rho u_t - 1 + v_t$，则做如下变换 $Y_t - \rho Y_{t-1} = (1-\rho)\beta_0 + \beta_1(X_{t,1} - \rho X_{t-1,1} + ... + \beta_k(X_{t,k} - \rho X_{t-1,k}) + v_t$。但是这里的 ρ 未知，只能估计，常用方法包括残差一阶自相关系数，D-W 统计量、CO 迭代和 PW 迭代法，具体见庞皓《计量经济学》。
3. 修改模型设定

7.2.5 模型设定与数据问题

模型就是所猜测的数据生成过程的刻画，模型设定错误意味着理论出了问题，但是，在操作层面，我们可以从模型导致的问题出发修补理论，最终得到良好的解释。同样，样本是信息来源，如果样本存在抽样、度量的问题，也会影响计量分析的结果。现在一一分析常见的模型设定偏误和数据问题。

7.2.5.1 遗漏变量

假设真实的模型为 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$，但是估计的模型为 $Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$，即 $\epsilon_i = \beta_2 X_{i2} + u_i$。如果原模型满足古典假定，则

1. 如果 X_1, X_2 不相关，则 X_1, ϵ 不相关，估计模型满足古典假定，但是估计的 σ^2 增大。
2. 如果 X_1, X_2 相关，则 X_1, ϵ 相关，违背零条件均值假设，导致内生性问题，估计量不再无偏一致有效。

解决遗漏变量：

1. 加入尽可能多的控制变量
2. 随机实验与自然实验
3. 工具变量
4. 采用面试数据

7.2.5.2 无关变量

假设真实的模型为 $Y_i = \beta_0 + \beta_1 X_{i1} + u_i$，但是估计的模型为 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$，即 $\epsilon_i = u_i - \beta_2 X_{i2}$。由于 X_2 是无关变量，理论上 $\beta_2 = 0$，因此估计模型并不违反古典假定，只是 $\hat{\beta}$ 的方差可能变大，可能影响统计推断。

7.2.5.3 建模策略

1. 从大到小：列出所有可能的解释变量，逐一筛选。
2. 从小到大：从最简单的模型开始，逐一增加解释变量。

实际操作时往往折衷处理。在模型之间进行权衡时，可以采用信息准则：

1. 拟合优度 R^2
2. AIC 准则：Min AIC = $ln(\frac{SSR}{N}) + \frac{2}{N}k$
3. BIC 准则：Min BIC = $ln(\frac{SSR}{N}) + \frac{lnN}{N}k$

7.2.5.4 模型形式

检验是否遗漏高次项：RESET 检验。

- 估计模型：$Y_i = \beta_0 + \beta_1 X_{i1} + u_i$，拟合值 $\hat{Y}_i$

- 辅助回归：$Y_i = \beta_0 + \beta_1 X_{i1} + \delta_2 Y_i^2 + \epsilon_i$
- 检验：$H0 : \delta_2 = 0$（F 检验或 LM 检验）

7.2.5.5 极端数据

极端值对 OLS 估计系数有较大影响，可以通过散点图发现极端值，但是采用统计方法更有效，设观测数据 i 对回归系数的影响力为：

$$lev_i = X_i'(X'X)^{-1}X_i \tag{7.58}$$

这里有 $\hat{\beta} - \hat{\beta}^i = \frac{1}{1-lev_i}(X'X^{-1})X_i e_i$，即 lev_i 越大，该数据对回归系数的影响越大，因此越异常。对于异常值，可以从三个方面解决：检查数据输入和操作是否正常、分析个体异质性、同时汇报全样本和去掉异常值的子样本的回归结果。

7.2.5.6 虚拟变量

在使用定性数据时，通常需要引入虚拟变量，这时其系数表征经济变量的不同类别的影响差异。对于 k 类对象，只能引入 $k-1$ 个虚拟变量，剩下一个是基准变量，否则引起完全共线性。引入方式有两种：

1. 单独引入虚拟变量：改变截距项，表达不同组别单独对被接受变量的作用。
2. 引入交叉项：改变解释变量系数，表达不同组别内解释变量的影响差异。

7.3 微观经济计量学

微观经济计量学研究微观经济领域的计量问题，微观数据具有独特的特征，与经典计量模型的要求并不一致。这里主要介绍微观数据特征、二元选择模型、Tobit 模型、样本选择模型、内生性与工具变量、面板数据模型、分层数据模型和处理评估方法等内容。主要参考《微观经济计量学：方法与应用》（卡梅伦）。

7.3.1 微观经济计量学基础

定义 7.1

微观经济数据：个体或者厂商经济行为方面的个体水平数据，包括横截面数据和面板数据。

微观经济数据具有如下特征：

- 异质性：经济主体各自在许多方面具有不同的特征。
- 多样性：经济主体的特征类别和具体表现是多样的。
- 不可观测变量：无法被观测到的个体特征，并且对研究目标有显著的影响。

个体特征和抽样误差是导致微观经济数据违背高斯马尔可夫定理的主要原因。抽样的误差主要是选择偏差和数据缺失。选择偏差意味着样本不是随机从总体中抽样得到，即样本不能代表总体，消除选择偏差的努力集中于如何修正样本以代表总体，数据缺失同样是在非随机数据缺失的情况下会影响样本数据代表总体数据特征的可靠性。简单地说，由于微观经济主体十分庞大，在收集数据时只能采用随机抽样方法，实际操作中有现实的约束无法得到基于总体的随机样本，从而在微观计量中需要处理抽样带来的误差。除了数据方面的调整以外，微观计量还需要基于经济理论找到合适的计量方法，其中特别需要处理的就是内生性。解决内生性是贯穿微观计量的主线。

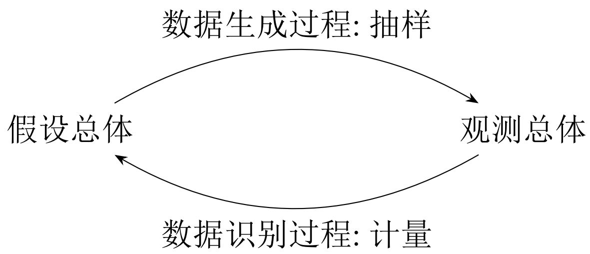

图 7.1: 抽样和识别

微观经济数据的来源一般为观测数据和实验数据，实验包含社会实验和自然实验。观测数据是直接对现实的经济主体进行抽样和记录得到的，抽样方法和抽样对象是核心，观测数据的质量上存在如下的问题：

- 调查无响应：出于非随机原因导致样本缺失或者部分缺失
- 数据缺失或误测：观测数据不真实
- 样板损耗：调查和响应行为前后不一致，不完全参与调查

微观数据的局限性在于数据生成过程（DGP）的模糊性，经济行为不可控，只能根据理论去猜测微观主体的行为方式。

相对于观测数据，社会实验数据可以明确控制经济主体行为，常常用于估计实际的或潜在的社会政策效应。随机实验的优点是可以消除选择偏差，并且处理变量外生，经济现象的原因和结果十分清楚。局限性首先是成本高昂，其次是由于实验对象也是抽样样本，具有内部有效性和外部有效性两方面的问题。内部有效性是针对实验样本而言，处理组和对照组的划分非随机，而且受试者在实验中的表现和现实中的表现可能不一致。外部有效性是实验结果无法适用于假设总体，这可能是由于实验样本和总体在特征和规模上的差异对所关心的经济行为有显著影响。但是总体上说，社会实验数据比观测数据更容易进行计量分析，它的原因和效应都是可控的。

自然实验是指某些社会和政治过程会导致的近似于真正实验的场景。自然实验是社会实验的某种近似，自然实验对于经济学家来说是零成本的，并且是在现实环境下有明确的处理组合对照组。自然实验的局限性在于：（1）干预的外生性值得怀疑（2）干预的影响是否足够大（3）处理组和对照组是否明确。同样，自然实验也存在实验数据的局限：内部有效性和外部有效性均需要考察。自然实验的核心是随机分配，这是保障数据代表性的依据，可以从理论逻辑上和数据上来证明自然实验的随机性。不管是自然实验还是社会实验，我们都希望向

随机控制实验靠拢，即向三个指标靠拢：可比较、随机分配和人为干预。观测数据满足可比较的标准，社会实验可能满足三个标准，自然实验满足第一个标准，近似满足第三个标准。

7.3.2 二元选择模型、Tobit 模型和样本选择模型

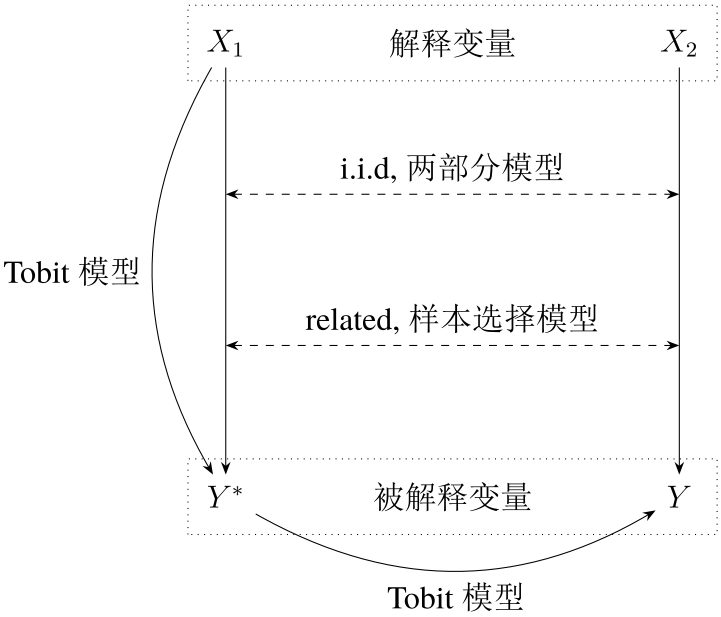

图 7.2: 参与和结果方程

对应经济主体来说，一项观测包含两层信息：是否被观测和观测的结果，这在计量中对应两个分析方程，参与方程和结果方程，一些因素影响是否参与实验（或者说参与某种经济活动），另一些因素影响参与的结果。在实际观测中，我们会遇到截尾和删失数据，截尾数据是部分观测值（大于或小于阈值）无法取得，并不在数据集中体现，删失数据是部分观测值（大于或小于阈值）的因变量无法取得，自变量可以取得，此时设置因变量为某一确定值。截尾和删失也可以看做是一个选择过程，可以采用参与方程来判断是否截尾或删失。下面首先介绍二元选择模型，即考察参与方程。

7.3.2.1 二元选择模型

二元选择模型是被解释变量为 0-1 变量的模型。如果用经典线性回归模型来拟合二元被解释变量，这个模型就是线性概率模型：

$$Y_i = \beta_0 + \beta_1 X_i + u_t \tag{7.59}$$

这里的概率来自于 $E(Y|X) = E(Y = 1|X)$，即拟合值的均值是该点取 1 的概率。线性概率模型容易估计和预测，并且可以提示变量的符号，但是不满足同方差假定，而且预测值可能在 0-1 范围之外，从经济变量考虑，模型显示 $\frac{\Delta Y}{\Delta X_1}|X_1$ 是不变的，这并不合理。我们来尝试改进这个模型，我们想要预测值总是在 0-1 之间，并且是单调的，这样有利于判断符号，累积

分布函数正好符合这个要求，得到非线性概率模型：

$$p_i = P(Y_i = 1|X) = F(X_i'\beta) \tag{7.60}$$

设定 F 为标准正态分布累计分布函数，得到 probit 模型，设定 F 为 Logistic 分布，得到 logit 模型。在线性模型中我们首先关心的是系数，在这里就是边际效应，然后再看看如何估计这个模型。

边际效应定义为：

$$\frac{\partial P(Y_i = 1|X_i)}{\partial x_{i,j}} = F'(X_i'\beta)\beta_j \tag{7.61}$$

样本平均边际效应为 $N^{-1}\sum_i F'(X_i'\hat{\beta})\hat{\beta}_j$，回归元样本均值处的边际效应为 $F'(\bar{X}_i{}'\hat{\beta})\hat{\beta}_j$，Y 样本均值处的边际效应为 $F(X'\beta) = \bar{Y}, F'(X'\beta) = F'(F^{-1}(\bar{Y}))$。带入模型的具体形式，得到边际效应：

$$Probit : p_i = \Phi(X_i'\beta), \frac{\partial p_i}{\partial X_{i,j}} = \phi(X_i'\beta)\beta_j \tag{7.62}$$

$$Logit : p_i = \frac{exp(X_i'\beta)}{1 + exp(X_i'\beta)}, \frac{\partial p_i}{\partial X_{i,j}} = \frac{exp(X_i'\beta)}{(1 + exp(X_i'\beta))^2}\beta_j \tag{7.63}$$

特别的，logit 模型在 Y 均值处对变量 j 的边际效应为 $\bar{Y}(1-\bar{Y})\hat{\beta}_j$。

估计二元选择模型一般采用极大似然法 ML，二值结果变量服从伯努利分布，即：

$$f(Y_i|X_i) = p_i^{Y_i}(1-p_i)^{1-Y_i} \tag{7.64}$$

$$lnf(Y_i|X_i) = Y_i lnp_i + (1-Y_i)ln(1-p_i) \tag{7.65}$$

$$L_N(\beta) = \sum_i \{Y_i lnp_i + (1-Y_i)ln(1-p_i)\} \tag{7.66}$$

$$set, F_i = F(X_i'\beta), F_i' = F(X_i'\beta) \tag{7.67}$$

$$\frac{dL_N}{d\beta} = \sum_i \{\frac{Y_i}{F_i}F_i'X_i - \frac{1-Y_i}{1-F_i}F_i'X_i\} = 0 \tag{7.68}$$

$$\sum_i \frac{Y_i - F_i}{F_i(1-F_i)F_i'X_i} = 0 \tag{7.69}$$

只需解出上式即可（牛顿-拉夫森迭代）。对于 logit 模型来说，ML 一阶条件为 $\sum_i (Y_i - L(X_i'\beta))X_i = 0$,$L$ 为 Logit 分布 $(\frac{1}{1+e^{-x}})$，并且对数优势比 $ln\frac{p}{1-p} = X_i'\beta$ 为线性。相对来说，probit 模型没有这么简单的表达形式。在 logit 和 porbit 模型中进行选择时通常在前者的简单形式和后者容易推广中选择，幸运的是二者在预测值等方面的结果差异不大。有一点可能让你感到很迷惑，非线性概率模型用 0-1 变量拟合，预测值是 0-1 之间的连续变量，这应该怎么判断呢？可用的方法是尝试，随机选择一个阈值，画出 ROC 曲线加以判断。为了理解选择的过程，可以设一个潜变量（不完全观测的变量）。将潜变量和二值变量联系在一起的函数是指

示函数，形式如下：

$$y_i^* = X_i'\beta + u_i \tag{7.70}$$

$$y = \begin{cases} 1, y^* > 0 \\ 0, y^* \le 0 \end{cases} \tag{7.71}$$

$$P(Y_i = 1|X_i) = P(Y_i > 0) = P(X_i'\beta + u > 0) = P(-u < X_i'\beta) = F(X_i'\beta) \tag{7.72}$$

这里 F 为 $-u$ 的累计分布函数。现在可以看到二元选择模型和潜变量的关联，在二元选择模型中我们假设了一个潜变量，并设定当潜变量大于 0 时取 1，否则取 0。由于这里用到的是 $-u$，如果 u 服从对称分布将是容易处理的，这也是 probit 模型比较常用的原因之一。

7.3.2.2 Tobit 模型

Tobit 模型考虑的是结果完全依赖参与方程的情形，即一个潜变量同时决定是否参与和参与结果。Tobit 模型是非常简单的样本选择形式，但是可以从中理解删失数据、截尾数据和它们的期望。当然，这里的删失（censored）和截尾（truncated）机制只是我们假设的数据生成过程而已，它有助于理解现实但不能还原现实。如图：

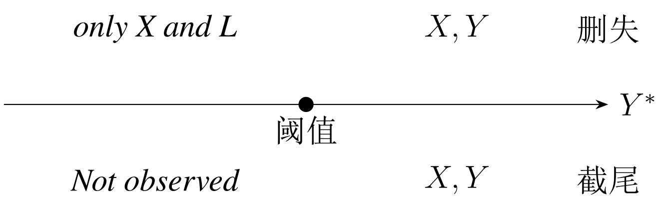

图 7.3: 删失和截尾机制 (左)

删失和截尾（左）机制用函数表达如下：

$$Truncated : Y = \begin{cases} Y^*, Y^* > L \\ L, Y^* \le L \end{cases} \tag{7.73}$$

$$Censored : Y = Y^*, Y^* > L \tag{7.74}$$

现在求均值（假设阈值为 0），这非常重要。

$$Truncated : E(Y) = E(Y^*|Y^* > 0) = E(X'\beta + \epsilon|X'\beta + \epsilon > 0) \tag{7.75}$$

$$= X'\beta + E(\epsilon|X'\beta + \epsilon > 0) = X'\beta + E(\epsilon|\epsilon > -X'\beta) \tag{7.76}$$

$$Censored : E(Y) = P(d = 0)E(Y|d = 0) + P(d = 1)E(Y|d = 1) \tag{7.77}$$

$$= 0 + P(Y^* > 0)E(Y^*|Y^* > 0) = P(Y^* > 0)E(Y^*|Y^* > 0) < E(Y^*|Y^* > 0) \tag{7.78}$$

左删失均值小于左截尾均值。对删失和截尾数据应用线性回归模型，不能采用 ols 估计，

此时的条件均值非线性，一般采用 ML 方法。其密度函数和极大似然函数设定为：

$$Censored: f(Y|X) = \begin{cases} f^*(Y|X), Y > L \\ F^*(L|X), Y = L \end{cases} \tag{7.79}$$

$$set, d = \begin{cases} 1, Y > L \\ 0, Y \leq L \end{cases} \tag{7.80}$$

$$f(Y|X) = f^*(Y|X)^d F^*(L|X)^{1-d} \tag{7.81}$$

$$So, lnL_N = \sum_i d_i ln(f^*(Y|X,\theta) + (1-d_i)lnF^*(L_i|X_i, 0) \tag{7.82}$$

$$Truncated: f(Y|X) = f^*(Y|X), Y > L \tag{7.83}$$

$$= f^*(y)/P(Y|Y > L) \tag{7.84}$$

$$= f^*(y)/(1 - F*(L)) \tag{7.85}$$

$$So, lnL_N = \sum_i ln(f^*(Y_i|X,\theta) + ln[1 - (f^*(Y_i|X,\theta)] \tag{7.86}$$

现在可以介绍 Tobit 模型（删失正态回归模型）是从 0 点删失的模型，潜变量关于回归元是线性的，可加误差是正态分布同方差的。

$$Y^* = X'\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2); \tag{7.87}$$

因此 $Y^* \sim N(X'\beta, \sigma^2)$，而结果变量为：

$$Y = \begin{cases} Y^*, Y^* > 0 \\ \emptyset, Y^* \leq 0 \end{cases} \tag{7.88}$$

对于 Tobit 模型，也可以应用截尾机制，只需要修改上述结果变量的表达式。为了推导出 Tobit 模型的表达式，还需要计算一下正态分布的截尾矩：

$$E(z|z > c) = \frac{\phi(c)}{1 - \Phi(c)}, E(z|z > -c) = \frac{\phi(c)}{\Phi(c)} \tag{7.89}$$

$$E(z^2|z > c) = 1 + \frac{c\phi(c)}{1 - \Phi(c)} \tag{7.90}$$

$$V(z|z > c) = 1 + \frac{c\phi(c)}{1 - \Phi(c)} - \frac{\phi(c)^2}{[1 - \Phi(c)]^2} \tag{7.91}$$

因此 Tobit 模型的误差项为:

$$E(\varepsilon|\varepsilon > -X'\beta) = \sigma E(\frac{\varepsilon}{\sigma}|\frac{\varepsilon}{\sigma} > -\frac{X'\beta}{\sigma}) = \sigma \frac{-\frac{X'\beta}{\sigma}}{1 - \Phi(-\frac{X'\beta}{\sigma})} = \sigma \frac{\frac{X'\beta}{\sigma}}{\Phi(\frac{X'\beta}{\sigma})} = \sigma\lambda(\frac{X'\beta}{\sigma}) \tag{7.92}$$

其中 $\lambda(z) = \frac{\phi(z)}{\Phi(z)}$ 称为逆米尔斯比（Inverse Mill’s Ratio）。可以发现，这一表达式是截尾数

据均值的调整项，也可以调整删失数据均值。这样一来，相应的条件均值变为：

$$潜变量E(Y^*|X) = X'\beta \tag{7.93}$$

$$左截尾E(Y|X) = E(Y^*|X, Y^* > 0) = X'\beta + \sigma\lambda(\frac{X'\beta}{\sigma}) \tag{7.94}$$

$$左删失E(Y|X) = \Phi(\frac{X'\beta}{\sigma})X'\beta + \sigma\phi(\frac{X'\beta}{\sigma}) \tag{7.95}$$

相应的边际效应分别为：

$$潜变量\partial E(Y^*|X)/\partial X = \beta \tag{7.96}$$

$$左截尾\partial E(Y|X, Y > 0)/\partial X = 1 - w\lambda(w) - \lambda(w)^2\beta左删失\partial E(Y|X)/\partial X = \Phi(w)/\beta \tag{7.97}$$

$$here, w = \frac{X'\beta}{\sigma}, \Phi'(z) = \phi(z), \phi'(z) = -z\phi(z) \tag{7.98}$$

要估计出 Tobit 模型中的具体参数，可以采用 MLE（依赖正态同方差加上）和 Heckman 两步估计方法。到这里，你也许感到疑惑，到底 Tobit 模型的具体形式是什么?! 实际上 Tobit 模型的形式决定于数据的删失机制（这里截尾也称为删失的方式之一），即潜变量如何对应结果变量，比如对于左截尾的 Tobit 模型，我们首先依据 $Y^* = X'\beta + u$ 估计参数，然后用 $E(Y|X) = X'\beta + \sigma\lambda(\frac{X'\beta}{\sigma})$ 进行调整预测值。（可能有误！再回来看看这段）

7.3.2.3 两部分模型

Tobit 模型假设删失机制和结果变量直接关联，但是一般情况下删失机制（是否参与）和结果变量可以分离。如果参与方程和结果方程无关（误差项独立），这就是两部分模型。即样本根据是否参与划分为两类（d=0,1），两部分模型表示为：

$$f(Y|X) = \begin{cases} P(d = 0|X), Y = 0 \\ P(d = 1|X)f(Y|d = 1, X), Y > 0 \end{cases} \tag{7.99}$$

此时可以设定两个方程，参与方程 $E(D) = F(X_1'\beta)$ 和结果方程 $E(Y|X_2) = X_2'\beta$，这里的删失机制已经和结果分离（独立），因此结果方程可以采用 ols 估计，参与方程采用 ML 估计。

7.3.2.4 样本选择模型

两部分模型假设参与方程和结果方程独立，这有时并不成立。我们在抽样是常常是外生抽样，样本有意或无意建立在凭借因变量取值的基础上，这种样本是选择样本。选择样本有两类，自选择（结果由个体是否选择参与到活动中决定）和样本选择（参与到活动中的个体被过度抽样）。二变量选择模型的参与方程和结果方程分别是：

$$Y_1 = \begin{cases} 1, Y_1^* > 0 \\ 0, Y_1^* \leq 0 \end{cases} \tag{7.100}$$

$$Y_2 = \begin{cases} Y_2^* > 0, Y_1^* > 0 \\ \emptyset, Y_1^* \leq 0 \end{cases} \tag{7.101}$$

潜变量的形式为：

$$Y_1^* = X_1'\beta_1 + \varepsilon_1 \tag{7.102}$$

$$Y_2^* = X_2'\beta_2 + \varepsilon_2 \tag{7.103}$$

这里设定 ε_1 和 ε_2 相关，从而两个方程必须一起估计。假设相关误差服从联合正态分布并且同方差，则:

$$\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix} \sim N\left[\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}\right] \tag{7.104}$$

这样可以采用 ML 估计。似然函数为：

$$L_N = \prod_{i=1}^{n} P(Y_{1i})^* < 0)^{1-Y_{1i}} f(Y_{2i}|Y_{1i}^* > 0) \times P(Y_{1i}^* > 0) \tag{7.105}$$

极大似然法严重依赖分布假设，这在某些情景下并不合理。这里我们有一种更为常用的方法来估计样本选择模型——Heckman 两部估计量 (Heckit)，这个估计框架如下：

假设线性回归模型为

$$Y_i = X_i'\beta + \varepsilon_i \tag{7.106}$$

Y_i（结果）能否被观测取决于变量 Z_i（参与），决定 Z_i 的是潜变量 Z_i^*，

$$Z_i = \begin{cases}1, Z_i^* > 0\\ 0, Z_i^* \leq 0\end{cases} \quad Z_i^* = w_i'\gamma + u_i \tag{7.107}$$

这里 u_i 服从正态分布，则 Z_i 为 probit 模型，故 $P(Z_i = 1|w_i) = \Phi(w_i'\gamma)$。那么观测样本的条件期望为：

$$E(Y_i|Z_i^* > 0) = E(X_i'\beta + \varepsilon_i|w_i'\gamma + u_i > 0) \tag{7.108}$$

$$= E(X_i'\beta + \varepsilon_i|u_i > -w_i'\gamma) \tag{7.109}$$

$$= X_i'\beta + E(\varepsilon_i|u_i > -w_i'\gamma) \tag{7.110}$$

$$= X_i'\beta + \rho\sigma_\varepsilon\lambda(-w_i'\gamma) \tag{7.111}$$

$$here, E(\varepsilon_i) = E(u_i) = 0, adjust \sigma = 1 \tag{7.112}$$

现在可以介绍 heckit：

- Step1, 用 Probit 模型估计参与方程，得到参数估计值 $\hat{\gamma}$ 并计算逆米尔斯比 $\lambda(-\hat{w_i'\gamma})$
- Step2, 用 OLS 回归 $Y_i, X_i, \hat{\lambda}_i$，得到估计值 $\hat{\beta}, \hat{\rho}, \hat{\sigma}_\varepsilon$

其中第二部的估计方程为：

$$Y_i = X_i'\beta + \sigma_{12}\lambda(-w_i'\hat{\gamma}) + v_i \tag{7.113}$$

注意这里我们只关心调整项前面的系数，因此将 $\rho\sigma_\varepsilon$ 合并在一起。相比于 ML 估计，Heckit 实施简单，适用性更强并且对分布假设不敏感。但是，特别要注意的是，如果 X, w 是完全相同的回归元，正态分布误差模型接近不可识别，因此至少需要一个在参与方程中的因素不出现在结果方程中 $w - X \neq \emptyset$。在实际应用中选择两部分模型还是样本选择模型完全取决于参与方程和结果方程是否相关，如果是依赖可观测因素的选择，二者不相关，如果是依赖不可

观测因素的选择，二者相关。

在这里我们设定的是截尾机制，如果未参与的个体也可以观测到结果，这个模型就变成罗伊模型，即：

$$Y_1 = \begin{cases} 1, Y_1^* > 0 \\ 0, Y_1^* \leq 0 \end{cases} \tag{7.114}$$

$$Y_2 = \begin{cases} Y_2^*, Y_1^* > 0 \\ Y_3^*, Y_2^* < 0 \end{cases} \tag{7.115}$$

对应的潜变量模型也是线性，并且误差服从联合整体分布（三个方程），估计罗伊模型的方法也是 heckit，这里的调整项对参与者和非参与者有相反的符号。

$$E(Y|X, Y_1^* > 0) = X_2'\beta_2 + \sigma_{12}\lambda(X_1'\beta_1) \tag{7.116}$$

$$E(Y|X, Y_1^* \leq 0) = X_3'\beta_3 - \sigma_{13}\lambda(-X_1'\beta_1) \tag{7.117}$$

$$here, \lambda(z) = \frac{\phi(z)}{\Phi(z)} \tag{7.118}$$

7.3.3 工具变量

工具变量是专门用于解决内生性问题的工具，内生性是指误差项与解释变量之间具有相关性，内生性将造成线性模型的 OLS 估计量偏差。相对的外生性表示为:

$$E(X'(Y - X\beta)) = 0, \beta = E(X'X)^{-1}E(X'Y) \tag{7.119}$$

内生性的来源可以分类如下：

- 遗漏变量偏差
- 变量误差偏倚
- 样本选择偏差
- 联立因果偏差

解决内生性问题的方式有代理变量、工具变量、面板固定效应或一阶差分等方程。对于一个存在内生性问题的模型，有效的工具变量需要符合两个条件：

- 相关性：$Corr(X_i, Z_i) \neq 0$
- 外生性：$Corr(Z_i, u_i) = 0$

虽然寻找这样的工具变量比较困难，但是不妨先假设这个完美的工具变量存在，继续分析如何解决内生性。首先了解一下工具变量如何作用：

简单来说，存在内生性的模型，有两条影响 Y 的路径：$u - X - Y$ 和 $X - Y$，工具变量是将前一条路径分离，具体方式是用 $Z - X$ 和 $\hat{X} - Y$ 两条路径代替存在内生性问题的模型，当然这里假设新的回归模型没有内生性。我们从简单的工具变量方法开始，推导单变量单工具的情形。

第一种方式，两步最小二乘估计。第一步，用 Z 对 X 进行回归，得到 X 中与 u 不相关

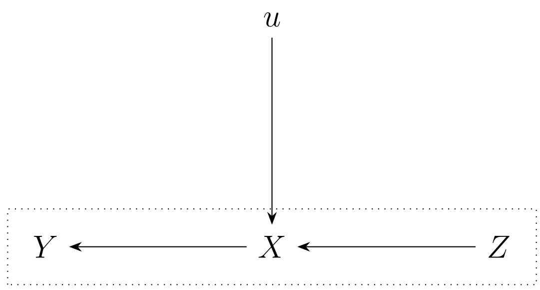

图 7.4: 工具变量作用路径

的部分，然后用拟合值 $\hat{X}$ 做解释变量加入 Y 的解释模型，即：

$$Step1, X = \alpha_0 + \alpha_1 Z + V \tag{7.120}$$

$$get, \hat{X} \tag{7.121}$$

$$Step2, Y = \beta_0 + \beta_1 \hat{X} + u \tag{7.122}$$

这里 $\hat{X}$ 与 u 不相关。

第二种方式，直接代数变换。

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{7.123}$$

$$Cov(Y_i, Z_i) = Cov(\beta_0 + \beta_1 X_i + u_i, Z_i) \tag{7.124}$$

$$= Cov(\beta_1 X_i, Z_i) \tag{7.125}$$

$$= \beta_1 Cov(X_i, Z_i) \tag{7.126}$$

$$So, \beta_1 = \frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)} \tag{7.127}$$

这里可以采用样本协方差估计总体协方差，即 $\hat{\beta}_1 = \frac{S_{YZ}}{S_{XZ}}$。

第三种方式，基于简化形式的推导，这种方式是联立方程：

$$X_i = \alpha_0 + \alpha_1 Z_i + v_i \tag{7.128}$$

$$Y_i = \gamma_0 + \gamma_1 Z_i + w_i \tag{7.129}$$

$$So, \beta_1 = \frac{\gamma}{\alpha} \tag{7.130}$$

这里 β_1 是 X 的外生变化对 Y 的影响，β_1 是用 X_i 表达 Y_i 的系数（解方程得出）。以上三种方法，如果采用 OLS 估计，必须知道第二个方程的标准差有偏误（没有考虑第一个方程），幸运的是常用计量软件可以直接给出稳健的标准差。

考虑一般形式的工具变量，即对多个内生变量应用多个工具变量，这个情形不是单工具变量的简单加和，我们需要考虑群体因素。为了表示方便，我们统一符号：

$$内生变量 X_1, \ldots, X_k \tag{7.131}$$

$$外生变量 W_1, \ldots, W_r \tag{7.132}$$

$$工具变量 Z_1, \ldots, Z_m \tag{7.133}$$

并引入一个概念**识别**，指能否根据模型和变量假设计算出唯一的变量系数。如果信息过多，则可能出现过度识别，如果缺少必要信息，则是识别不足。在 IV 回归中，如果工具变量

个数少于内生变量个数 $(m < k)$，则无法识别，如果 $m = k$，恰好识别，如果 $m > k$，过度识别。如果回归模型中只有一个内生变量，通常也可以用多个工具变量进行回归，处理方法和单变量单工具一致，只是要注意第一步回归中要包含所有的外生变量。虽然这样做可能引起过度识别，但是为一个内生变量找恰好的工具变量总是不容易。在理想状态下，工具变量可以帮助识别因果效应，只要工具变量相关性成立、外生性成立（随机抽样）、影响路径唯一（通过内生变量）。

下面进入工具变量的检验部分：包括不可识别检验、弱工具变量检验、过度识别检验和内生性检验。

- 不可识别检验：有用的工具变量个数小于内生变量个数
- 弱工具变量检验：相关性不成立或者关系微弱，此时估计量的渐进方差很大，估计不准确。检验方法是看第一步回归中的系数 γ 是否显著。如果通过弱工具变量检验，则需要换工具变量或采取 LIM 等方法。
- 过度识别检验：需要找出 IV 所有其他可能影响的渠道并一一排除，才能说明外生性成立，一般是理论分析。在过度识别的情况下（$m > k$），可以进行检验，方式是用内生变量和工具变量对第一步回归的误差回归，如果工具变量系数均为 0，则外生性成立。
- 内生性检验：所有的解释变量均为外生变量。检验方式是比较 IV 和 OLS 估计量的一致性，不具有内生性是两者都一致，具有内生性时 IV 一致而 OLS 不一致。

7.3.4 面板数据模型和分层线性模型

前面的讨论都集中在截面数据，这里我们进入面板数据，你可以将它当做截面数据来用，但是可能出现序列相关性和个体异质性等问题，实际上面板数据提供了相当多的个体行为和群体行为信息，对于经济分析非常有帮助。**面板数据**可以定义为相同截面上的个体在不同的时点反复观测得到的数据序列。根据数据完整性，面板数据可以分为平衡面板（每个观测个体在所有时期都有观测）和非平衡面板（个体在相同事情内缺失若干观测值）。这里我们不介绍怎么处理数据缺失问题，直接假设所用到的数据均为平衡面板。面板数据表示为 $Y_{i,t}, i = 1, \ldots, N; t = 1, \ldots, T$。沿袭前面的思路，这里先介绍建模过程，再介绍估计方法。面板数据模型包括：混合回归模型、固定效应回归模型和随机效应回归模型。

混合模型是将面板数据看做截面数据使用，模型设定为：

$$Y_{i,t} = \alpha + X_{i,t}\beta + \varepsilon_{i,t}, i = 1, \ldots, N; t = 1, \ldots, T \tag{7.134}$$

混合模型要求解释变量和误差项协方差无关，即 $Cov(X_{i,t}, \varepsilon_{i,t}) = 0$，则参数估计量是一致估计量。通常使用混合模型需要处理自相关问题（个体异质性）。

固定效应模型包括三个，个体固定效应模型、时点固定效应模型和个体时点双固定效应

模型。固定效应认为对每个个体或者时期都有各自的特征，需要特别强调。模型表示如下：

$$\text{个体固定效应模型} Y_{i,t} = \alpha_i + X'_{i,t}\beta + \varepsilon_{i,t}, = 1, \ldots, N; t = 1, \ldots, T \tag{7.135}$$

$$Note, \alpha_i unobservable; \textit{strong assumption,} E(\varepsilon_{i,t}|\alpha_i, X_{i,t}) = 0, i = 1, \ldots, N; t = 1, \ldots, T \tag{7.136}$$

$$\text{时点固定效应模型} Y_{i,t} = \gamma_t + X'_{i,t}\beta + \varepsilon_{i,t}, = 1, \ldots, N; t = 1, \ldots, T \tag{7.137}$$

$$here, X_{i,t} \textit{contains } Z_t, \textit{a variable set related to t} \tag{7.138}$$

$$\text{个体时点固定效应模型} Y_{i,t} = \alpha_i + \gamma_t + X'_{i,t}\beta + \varepsilon_{i,t}, = 1, \ldots, N; t = 1, \ldots, T \tag{7.139}$$

$$\tag{7.140}$$

时点固定效应模型和个体时点固定效应模型通常满足假定 $Cov(\varepsilon_{i,t}|X_{i,t}, \alpha_i, \gamma_t) = 0$，并且这里三个模型中固定效应（$\alpha, \gamma$）与变量相关。

随机效应模型的形式与个体固定效应模型一致，但是假设有差异，这是由于二者的思想差异，随机效应认为截距项是随机给定的，与个体的特征（解释变量无关），而固定效应则相当于将一些无法表达的解释力量放在了截距项中。具体形式如下：

$$Y_{i,t} = \alpha_i + X'_{i,t}\beta + \varepsilon_{i,t}, = 1, \ldots, N; t = 1, \ldots, T \tag{7.141}$$

$$assumption, \alpha_i \sim i.i.d.(\alpha, \sigma_\alpha^2), \varepsilon_{i,t} \sim i.i.d.(0, \sigma_\varepsilon^2) \tag{7.142}$$

注意这里并没有假设出误差项的具体分布。特别的，个体随机效应模型又被称为等相关模型，随机效应和误差项都是不可观测的，不妨并在一起组成新的误差项。

$$Y_{i,t} = X'_{i,t}\beta + (\alpha_i + \varepsilon_{i,t}) = X'_{i,t}\beta + u_{i,t} \tag{7.143}$$

$$if \alpha_i \sim (\alpha, \sigma_\alpha^2), \varepsilon_{i,t} \sim (0, \sigma_\varepsilon^2), \textit{and } Cov(\alpha_i, \varepsilon_{i,t} = 0) \tag{7.144}$$

$$So, Cov(u_{i,t}, u_{i,s}) = Cov(\alpha_i + \varepsilon_{i,t}, \alpha_i + \varepsilon_{i,s}) = \begin{cases} \sigma_\alpha^2, t \neq s \\ \sigma_\alpha^2 + \sigma_\varepsilon^2, t = s \end{cases} \tag{7.145}$$

$$\textit{and } Corr(u_{i,t}, u_{i,s}) = \begin{cases} \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}, t \neq s \\ 1, t = s \end{cases} \tag{7.146}$$

上面的相关系数表达式表明 $Corr(u_{i,t}, u_{i,s})$ 与间隔时期长短无关，因此不同时期的相关性是一致的，故称为等相关模型。下面介绍如何估计这些模型，估计方法包括混合最小二乘估计、组间 OLS 估计、组内 OLS 估计、一阶差分估计和随机效应估计。

混合最小二乘估计是将二维数据混合成一维的数据，然后采用 OLS 估计，如果模型设定正确，$Cov(X_{i,t}, \sigma_{i,t}) = 0$，则参数估计量具有一致性，混合最小二乘估计使用与混合模型，但是估计时由于误差项序列相关，信息含量降低，可能导致标准差低估。如果对于固定效应模型仍然采用混合最小二乘估计，则需要调整形式：

$$Y_{i,t} = \alpha + X'_{i,t}\beta + (\alpha_i - \alpha + u_{i,t}) \tag{7.147}$$

由于 α_i 和 $X_{i,t}$ 相关，因此模型具有内生性，估计量不一致，相似的，对于时点固定效应模型和个体时点双固定也适用。

组间估计、组内估计和一阶差分估计方法都是将未知效应删去，然后回到经典线性模型

的估计方法上，详细地说就是：

组间 OLS 估计，对个体或者时间分组，然后求每组的平均值，用平均值来估计就不会存在内生性和自相关问题。以个体随机效应模型的估计为例：

$$Y_{i,t} = \alpha_i + X'_{i,t}\beta + \varepsilon_{i,t} \tag{7.148}$$

$$\bar{Y_i} = \alpha_i + \bar{X_i}'\beta + \bar{\varepsilon_i}, i = 1, \ldots, N \tag{7.149}$$

$$\bar{Y_i} = \alpha + \bar{X_i}'\beta + (\alpha_i - \alpha + \bar{\varepsilon_i}), i = 1, \ldots, N \tag{7.150}$$

如果 $\bar{X_i}$ 和 $\alpha_i - \alpha + \bar{\varepsilon_{i,t}}$ 相互独立，则 α, β 的平均估计量是一致的，这适用于混合模型和个体随机效应模型，但是对于个体固定效应模型上式并不成立，估计量是非一致的。

组内 OLS 估计，每组数据都减去各组的均值，以消除各组的固定效应，这适用于固定效应模型，如个体固定效应模型：

$$Y_{i,t} = \alpha_i + X'_{i,t}\beta + \varepsilon_{i,t} \tag{7.151}$$

$$\bar{Y_i} = \alpha_i + \bar{X_i}'\beta + \bar{\varepsilon_i}, i = 1, \ldots, N \tag{7.152}$$

$$Y_{i,t} - \bar{Y_i} = (X_{i,t} - \bar{X_i})'\beta + (\varepsilon_{i,t} - \bar{\varepsilon_i}), i = 1, \ldots, N; t = 1, \ldots, T and, \hat{\alpha_i} = \bar{Y_i} - \bar{X_i}'\hat{\beta} \tag{7.153}$$

组内 OLS 估计无法估计非时变的变量构成的面板数据模型，那将导致离差为 0。

一阶差分 OLS 估计是将个体数据做时序一阶差分，以去掉未知的个体固定效应，然后进行 OLS 估计。即：

$$Y_{i,t} = \alpha_i + X'_{i,t}\beta + \varepsilon_{i,t} \tag{7.154}$$

$$Y_{i,t-1} = \alpha_i + X'_{i,t-1}\beta + \varepsilon_{i,t-1} \tag{7.155}$$

$$Y_{i,t} - Y_{i,t-1} = (X_{i,t} - X_{i,t-1})'\beta + (\varepsilon_{i,t} - \varepsilon_{i,t-1}) \tag{7.156}$$

$$\tag{7.157}$$

相比于组内估计，一阶差分数据每组损失一个观测值，因此估计的有效性将下降。

随机效应估计（可行 GLS 估计）方法：

$$Y_{i,t} = \alpha_i + X'_{i,t}\beta + \varepsilon_{i,t}, \alpha_i, \varepsilon \sim i.i.d. \tag{7.158}$$

$$Y_{i,t} - \hat{\lambda}\bar{Y_{i,t}} = (X_{i,t} - \hat{\lambda}\bar{X_i})'\lambda + v_{i,t}, v_{i,t} = (1 - \hat{\lambda}\alpha_i) + (\varepsilon_{i,t} - \hat{\lambda}\hat{\varepsilon_i}) \tag{7.159}$$

$$and, \lambda = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_\alpha^2}} \tag{7.160}$$

当 $\hat{\lambda} = 0$，GLS 等同于混合 OLS 估计，当 $\hat{\lambda} = 1$，GLS 等同于组内 OLS 估计。对于随机效应模型，可行 GLS 估计是一致估计量，对于个体固定效应模型，可行 GLS 不是一致估计量。总的来说，各种模型和估计方法总结如下：

前面我们是在模型设定正确的假设下进行分析，现在回到我们的假设，其中的重点是是否存在固定效应，我们重点关注如何区别个体固定效应和个体随机效应。二者的模型形式一致，并且 $E(Y_{i,t}|\alpha_i, \beta) = \alpha_i + X'_{i,t}\beta$，但是固定效应假设 α_i 和 $X_{i,t}$ 相关，而随机效应假设其无关。检验的方法是判断个体截距 α_i 是否不变，即 $H_0 : \alpha_i = \alpha$，第一种方法是 F 检验，统计量

	混合模型	随机效应	固定效应
混合 OLS	一致	一致	不一致
组间估计	一致	一致	不一致
组内估计	一致	一致	一致
一阶差分	一致	一致	一致
可行 GLS	一致	一致	不一致

表 7.1: 面板数据模型和估计方法一致性

为：

$$F = \frac{SSE_r - SSE_{ur}/(N-1)}{SSE_u/(NT-N-k)} \tag{7.161}$$

第二种更具有针对性的检验方法是 Hausman 检验，其思想是如果是随机效应模型，则组内估计和可行 GLS 都是一致的，如果是固定效应模型，组内估计是一致的但可行 GLS 是不一致的。因此 Hausman 假设的原假设是 $H_0 : \hat{\beta_{RE}} - \hat{\beta_w} = 0$（随机效应模型），对应的统计量为：

$$H = \frac{(\hat{\beta_w} - \hat{\beta_{RE}})^2}{s(\hat{\beta_w})^2 - s(\beta_{RE})^2} \sim \chi^2(1) \tag{7.162}$$

在面板回归模型中，如果包含滞后变量，模型就变成动态面板模型，此时的滞后项是不可度量的个体效应的结果。但是模型可能存在内生性问题，即前一期的被解释变量和当期的误差项相关。

7.3.4.1 分层线性模型

分层线性模型考虑的是个体的集聚特征，即考察群体或者组织的特征。分层数据来源于社会结构（如地域、收入分级）和分层抽样方法。这里我们主要考察如何在模型中体现不同层级的影响。这里只介绍两水平分层线性模型，更多层次的线性模型只是外推就可以得到。

$$Y_{i,j} = \beta_{0,j} + \beta_{1,j} X_{i,j} + \varepsilon_{i,j} \tag{7.163}$$

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + u_{0,j} \tag{7.164}$$

$$\beta_{2,j} = \gamma_{1,0} + \gamma_{1,1} R_{1,j} + u_{1,j} \tag{7.165}$$

上述模型的建立过程如下：（太麻烦了，想想怎么简化一下好懂！）

$$\text{地区内个体相关性模型 } \mathit{M0} \tag{7.166}$$

$$Y_{i,j} = \beta_{0,j} + \varepsilon_{i,j}, \varepsilon_{i,j} \sim N(0, \sigma^2) \tag{7.167}$$

$$\beta_{0,j} = \gamma_{0,0} + u_{0,j}, u_{0,j} \sim N(0, \sigma_{u0}^2) \tag{7.168}$$

$$Y_{i,j} = \gamma_{0,0} + (u_{0,j} + \varepsilon_{i,j}), Cov(\varepsilon_{i,j}, u_{0,j}) = 0 \tag{7.169}$$

$$and, \text{组内相关系数} ICC = \frac{\sigma_{u0}^2}{\sigma^2 + \sigma_{u0}^2} \tag{7.170}$$

$$\text{带有区域经济特征解释变量主效应的随机截距模型 } \mathit{M1} \tag{7.171}$$

$$Y_{i,j} = \beta_{0,j} + \varepsilon_{i,j}, \varepsilon_{i,j} \sim N(0, \sigma^2) \tag{7.172}$$

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + u_{0,j}, u_{0,j} \sim N(0, \sigma_{u0}^2) \tag{7.173}$$

$$Y_{i,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + (u_{0,j} + \varepsilon_{i,j}), Cov(\varepsilon_{i,j}, u_{0,j}) = 0 \tag{7.174}$$

$$\text{加入个体层面解释变量解释地区内个体差异 } \mathit{M2} \tag{7.175}$$

$$Y_{i,j} = \beta_{0,j} + \alpha_1 X_{1,i,j} + \varepsilon_{i,j}, \varepsilon_{i,j} \sim N(0, \sigma^2) \tag{7.176}$$

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + u_{0,j}, u_{0,j} \sim N(0, \sigma_{u0}^2) \tag{7.177}$$

$$Y_{i,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + \alpha_1 X_{1,i,j} + (u_{0,j} + \varepsilon_{i,j}), Cov(\varepsilon_{i,j}, u_{0,j}) = 0 \tag{7.178}$$

$$\text{检验个体层面的随机性 } \mathit{M3} \tag{7.179}$$

$$Y_{i,j} = \beta_{0,j} + \alpha_1 X_{1,i,j}^0 + \alpha_1 X_{1,i,j}^1 + \varepsilon_{i,j}, \varepsilon_{i,j} \sim N(0, \sigma^2) \tag{7.180}$$

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + u_{0,j}, u_{0,j} \sim N(0, \sigma_{u0}^2) \tag{7.181}$$

$$\beta_{1,j} = \gamma_{0,j} + u_{1,j} \tag{7.182}$$

$$Y_{i,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + \alpha_1 X_{1,i,j}^0 + \gamma_{1,0} X_{1,i,j}^1 + (u_{0,j} + \varepsilon_{i,j} + X_{1,i,j}^1 u_{1,j}), \tag{7.183}$$

$$where, \begin{bmatrix} u_{0,j} \\ u_{1,j} \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u0}^2 & \sigma_{u01}^2 \\ \sigma_{u01}^2 & \sigma_{u1}^2 \end{bmatrix}\right] \tag{7.184}$$

$$\text{地区层面变量对具有随机效应的个体层面变量系数的解释作用 } \mathit{M4} \tag{7.185}$$

$$Y_{i,j} = \beta_{0,j} + \alpha_1 X_{1,i,j}^0 + \beta_{1,j} X_{1,i,j}^1 + \varepsilon_{i,j} \tag{7.186}$$

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + u_{0,j} \tag{7.187}$$

$$\beta_{1,j} = \gamma_{1,0} + \gamma_{1,1} R_{1,j} + u_{1,j} \tag{7.188}$$

$$Y_{i,j} = \gamma_{0,0} + \gamma_{0,1} R_{1,j} + \alpha_1 X_{1,i,j}^0 + \gamma_{1,0} X_{1,i,j}^1 + \gamma_{1,1} R_{1,j} X_{1,i,j}^1 \tag{7.189}$$

$$+ (u_{0,j} + X_{1,i,j}^1 u_{1,j} + \varepsilon_{i,j}) \tag{7.190}$$

7.3.5 处理评估方法

终于到了政策评价这一章节，这一节是微观计量最有价值的一部分，前面的内容都可以看做是它的铺垫。简单地说，处理评估就是问一个问题：政策实施到底有多大的效果！并且我们尝试得到的是关于总体的处理效应。即对于总体 Y，处理后的结果是 Y_1，处理前的结果

是 Y_0，每个个体的处理效应是 $Y_{1i}-Y_{0i}, i=1,\ldots,N$。可惜的是，我们有两点局限，一是可得数据是样本而不是总体，此时我们需要用随机抽样等方法确信样本的代表性，二是反事实观测问题，我们将样本分为控制组和对照组，只能观察到对照组未被处理的结果并且时间上样本有可能发生系统性变化，此时我们需要一些假设来弥补这个硬约束，如假设对照组和处理组具有平行趋势。现在我们正式进入政策评估问题，我们希望掌握的数据如下：

	处理组 $(D=1)$	对照组 $(D=0)$
处理前	$Y_{0i}\|D=1$	$Y_{0i}\|D=0$
处理后	$Y_{1i}\|D=1$	$Y_{1i}\|D=0$

表 7.2: 政策评价理论数据

我们实际掌握的数据如下：

	处理组 $(D=1)$	对照组 $(D=0)$
处理前	$Y_{0i}\|D_i=1$	$Y_{0i}\|D_i=0$
处理后	$Y_{1i}\|D_i=1$	$\emptyset$

表 7.3: 政策评价观测数据

遗憾的是 $Y_{1i}|D=0$ 不可得，但要完成政策评估，这个值又特别重要，如何处理这个缺失值就是不同的评估方法的核心差异。在进一步了解之前，回到处理效应的定义可以明确我们的目的。

$$\text{平均处理效应}ATE = E(Y_{1i}-Y_{0i}) \tag{7.191}$$

$$\text{处理组平均处理效应}ATT = E(Y_{1i}-Y_{0i}|D_i=1) \tag{7.192}$$

$$\text{非处理组平均处理效应}ATU = E(Y_{1i}-Y_{0i}|D_i=0) \tag{7.193}$$

$$\tag{7.194}$$

我们关心的是平均处理效应，并且注意这里是同一个个体 i，而在观测值中 i 要么是处理组要么是对照组。观测结果 Y_i 可以写成如下的形式：

$$Y_i = \begin{cases} Y_{1i}, D_i=1 \\ Y_{0i}, D_i=0 \end{cases} \tag{7.195}$$

$$= Y_{0i}+(Y_{1i}-Y_{0i})D_i \tag{7.196}$$

将处理的总效应分解为处理组的处理效应和选择偏差如下：

$$E(Y_i|D_i=1)-E[Y_{0i}|D_i=0] \tag{7.197}$$

$$=(E(Y_{1i}|D_i=1)-E[Y_{0i}|D_i=1])+(E[Y_{0i}|D_i=1]-E[Y_{0i}|D_i=0]) \tag{7.198}$$

$$i.e., ATE = ATT + \textit{Selection Bias} \tag{7.199}$$

上式将平均处理效应分解成处理组的平均处理效应和选择偏差，这里的平均处理效应是清楚的，我们需要处理的是选择偏差。第一种方式是随机分组，使得 D_i 独立于 Y_{0i}, Y_{1i}。更为宽松的要求是 D_i 均值独立于 Y_{0i}, Y_{1i}，即 $E[Y_{0i}|D_i=1]=E[Y_{0i}|D_i=0]$，此时 $ATE=ATT$，即平均处理效应等于处理组的处理效应。

下面我们探讨如何取得均值独立条件，这里我们重点关注 D_i 的取值受什么因素影响。

7.3.5.1 依可观测变量的选择

如果个体 i 的 D_i 的选择取决于可观测变量 X_1，则在给定 X_1 的情况下潜在的结果 (Y_{0i}, Y_{1i}) 独立于 D_i。

定义 7.2

可忽略性假定：给定 X_{1i}，则 (Y_{0i}, Y_{1i}) 独立于 D_i，即 $(Y_{0i}, Y_{1i}) \perp D_i | X_{1i}$ ♣

可忽略性假定意味着给定 X_{1i}，(Y_{0i}, Y_{1i}) 在对照组和控制组的分布完全一致。

定义 7.3

均值可忽略性假定：$E(Y_{0i}|X_{1i}, D_i) = E(Y_{0i}|X_{1i}), E(Y_{1i}|X_{1i}, D_i) = E(Y_{1i}|X_{1i})$ ♣

如果可忽略假定成立，则只需在回归如下方程即可解决遗漏变量问题：

$$Y_i = X_{1i}'\beta + \gamma D_i + \varepsilon_i \tag{7.200}$$

推导过程如下。如果假设处理效应在不同个体之间是相同的，即 $Y_{1i} - Y_{0i} = \rho, i = 1, \ldots, N$，则 $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i = \alpha + \beta D_i + u_i, \alpha = E(Y_{0i}), \rho = Y_{1i} - Y_{0i}, u_i = Y_{0i} - E(Y_{0i})$。则平均处理效应为：

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) \tag{7.201}$$

$$= \rho + E(u_i|D_i = 1) - E(u_i|D_i = 0) \tag{7.202}$$

即误差项和处理状态之间的相关性导致了选择偏差，此时只需要控制误差项中相关的变量 X_1 即可消除选择偏差，因此模型变为式(7.200)。

现在我们理解了如何克服选择偏差，重点就落在求平均处理效应 ATE 上。此时我们的分析都是在反事实框架下进行的，都是用观测数据替代不可能观测到的数据。假设处理发生在 k 期，$t' < k < t$

- 前后估计量：假设 $E(Y_{0t} - Y_{0t'}|D_i = 1) = 0$，即处理前的结果是处理后的非处理潜在结果的好的替代。前后估计量为 $(\bar{Y_{1t}} - \bar{Y_{0t}})_1$。
- 横截面估计量：假设 $E(Y_{0t}|D_i = 1) = E_{Y_{0t}|D_i=0}$，即非处理组的非处理结果是处理组非处理的潜在结果的替代变量，横截面估计量为 $\bar{Y_{1t1}} - \bar{Y_{0t0}}$。

以上两种匹配方法都严重依赖假设，这常常并不成立。下面介绍一种常用的匹配估计方法——断点估计 RDD。RDD 要求处理变量完全由连续变量 X_i 是否超过某断电所决定，此时 D_i 是 X_i 的确定性函数，因此满足可忽略性假定 $(Y_{0i}, Y_{1i} \perp D_i | X_i)$。现在我们来估计处理效应。RDD 分为截断 RD 和模糊 RD 两类，这里先介绍截断 RD，即处理状态决定于一个确定的非连续函数：

$$D_i = \begin{cases} 1, X_i \geq X_0 \\ 0, X_i < X_0 \end{cases} \tag{7.203}$$

其中 X_0 是已知的先验参数。实施断点回归，除了需要知道上述函数之外，还需要假设在断点附近的样本除了处理状态，其他方面的特征都一致，即在断点附近样本是被随机分配成处理和非处理的。

$$E(Y_i|X_0-\Delta<X_i<X_0)\approx E(Y_{0i}|X_i=X_0) \tag{7.204}$$

$$E(Y_i|X_0<X_i<X_0+\Delta)\approx E(Y_{1i}|X_i=X_0) \tag{7.205}$$

$$\lim_{\Delta\to 0} E(Y_i|X_0<X_i<X_0+\Delta)-E(Y_i|X_0-\Delta<X_i<X_0)=E(Y_{1i}-Y_{0i}|X_i=X_0) \tag{7.206}$$

假设在实验前结果变量与 X_i 之间存在线性关系：$Y_i=\alpha+\beta X_i+\varepsilon_i$，则设置断点 $X_i=c$，所估计的模型如下：

$$Y_i=\alpha+\beta(X_i-c)+\delta D_i+\gamma(X_i-c)D_i+\varepsilon_i, i=1,\dots,N \tag{7.207}$$

当 $D_i=0,1$ 时上式的截距和斜率不同，而且这里是中心化的模型，截距差异就是处理效应而斜率差异是处理组和对照组中 Y_i 对 X_i 的敏感性差异来源，即:$ATE=\delta$。

除了 RDD，匹配也是政策评估的常用方法，匹配的目的是在控制组中寻找处理组个体非处理结果的良好替代，即 $E(Y_{0i}|X,D_i=1)=E(Y_{0i}|X,D_i=0)$，匹配如何才能实现？匹配有两种方式，一是精确匹配，即可观测特征在处理组和对照组中都一致的个体进行匹配，这要求变量是离散的并且包含大量的样本，精确匹配要求对照组和控制组的全部变量都有部分的重叠，这被称为重叠假。二是倾向分匹配，即利用个体受处理的倾向程度进行匹配，此时我们只需要匹配处理概率即可，而得到处理概率的方法非常简单——二元选择模型 (Logit 或 Probit)，根据倾向分如何进行匹配则包括最近邻匹配、半径匹配、核匹配等方法。倾向分匹配的局限在于样本容量需求大，要求处理组合对照组有较大的共同取值范围并且只能控制可观测变量的影响，无法处理基于不可观测变量的选择偏差。

7.3.5.2 依不可观测因素的选择

依不可观测变量的选择情景下的处理评估方法包括双重差分估计 DID、横截面估计量（IV 和处理效应模型）、双重差分匹配估计量。首先介绍双重差分估计量 DID，DID 要求数据为处理组合对照组的面板数据或者重复截面数据。假设处理发生在 k 期，$t'<k<t$，则 DID 假设处理组合对照组具有平行趋势：

$$E(Y_{0t}-Y_{0t'}|D=1)=E(Y_{0t}-Y_{0t'}|D=0) \tag{7.208}$$

那么 ATE 的计算：

$$E(Y_{1t}-Y0t|D=1) \tag{7.209}$$

$$=E(Y_{1t}-Y0t'|D=1)-E(Y_{0t}-Y0t'|D=1) \tag{7.210}$$

$$=E(Y_{1t}-Y0t'|D=1)-E(Y_{0t}-Y0t'|D=0) \tag{7.211}$$

即 DID 估计量为 $(\bar{Y_{1t}}-\bar{Y_{0t'}})_1-(\bar{Y_{1t}}-\bar{Y_{0t'}})_0$。DID 估计量考虑到了时间趋势并且可以运用重复截面数据，但是平行趋势假设的验证比较困难。估计 DID 估计量的计量模型为：

$$Y=\beta_1 D+\beta_2 T+\beta_3 D*T+\varepsilon \tag{7.212}$$

这里 $D=0,1$ 表示状态变量，为 1 为处理组，为 0 为非处理组，$T=0,1$ 为时间变量，为 1 是处理发生后，为 0 是处理发生前。DID 估计量为 $\hat{\beta}_3$。当然在回归中需要加上控制因素，得到回归方程：

$$Y=\beta_1 D+\beta_2 T+\beta_3 D*T+X'\Gamma+\varepsilon \tag{7.213}$$

第二种政策评估方法是模糊断点回归，与精确断点回归的差异在于在断点两侧个体是否被处理是以概率发发生的，左侧为 a，右侧为 b，$0<a<b<1$，此时 D 的取值不完全由 X 决定因此不能直接采用 OLS 估计。如果给定 X，(Y_1-Y_0) 独立于 D，则 $(Y_{1i}-Y_{0i})\perp D_i|X_i$，则

$$Y=Y_0+(Y_1-Y_0)D \tag{7.214}$$

$$E(Y|X)=E(Y_0|X)+E[(Y_1-Y_0)D|X] \tag{7.215}$$

$$=E(Y_0|X)+E[(Y_1-Y_0)|X]E(D|X) \tag{7.216}$$

这里 $E[(Y_1-Y_0)|X]$ 是平均处理效应，$E(D|X)$ 是倾向得分。对该式取极限并做差为：

$$\lim_{x\to c^-}E(Y|X)=\lim_{x\to c^-}E(Y_0|X)+\lim_{x\to c^-}E(D|X)\lim_{x\to c^-}E[(Y_1-Y_0)|X] \tag{7.217}$$

$$\lim_{x\to c^+}E(Y|X)=\lim_{x\to c^+}E(Y_0|X)+\lim_{x\to c^+}E(D|X)\lim_{x\to c^+}E[(Y_1-Y_0)|X] \tag{7.218}$$

$$and,\lim_{x\to c^-}E(Y_0|X)=lim_{x\to c^+}E(Y_0|X) \tag{7.219}$$

$$\lim_{x\to c^-}E[(Y_1-Y_0)|X]=\lim_{x\to c^+}E[(Y_1-Y_0)|X]=E[(Y_1-Y_0)|X=c] \tag{7.220}$$

$$So,\lim_{x\to c^+}E(Y|X)-\lim_{x\to c^-}E(Y|X)=[\lim_{x\to c^+}E(D|X)-\lim_{x\to c^-}E(D|X)]E[(Y_1-Y_0)|X=c] \tag{7.221}$$

$$because,\lim_{x\to c^+}E(D|X)-\lim_{x\to c^-}E(D|X)=b-a\neq 0 \tag{7.222}$$

$$So,ATE=E[(Y_1-Y_0)|X=c]=\frac{\lim_{x\to c^+}E(Y|X)-\lim_{x\to c^-}E(Y|X)}{\lim_{x\to c^+}E(D|X)-\lim_{x\to c^-}E(D|X)} \tag{7.223}$$

现在分子和分母都可以用精确断点回归得到，只需一步计算就可以得到 ATE。模糊断点回归实际上是一种工具变量方法，$IV=I(X_i>c)$,I 为指示函数，IV 在断点出相当于局部随机实验，满足外生性和相关性。

第三种政策评估方法是处理效应模型，这就是样本选择模型在政策评估方面的应用。

$$Y_i=X_i'\beta+\gamma D_i+\varepsilon_i \tag{7.224}$$

$$D_i=I(Z_i'\delta+u_i) \tag{7.225}$$

$$\begin{bmatrix}\varepsilon_i\\u_i\end{bmatrix}=\left[\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\rho_\varepsilon^2 & \rho\sigma_\varepsilon\\ \rho\sigma_\varepsilon & 1\end{bmatrix}\right] \tag{7.226}$$

$$and,Cov(Z_i,\varepsilon_i)=0,Z_i\to IV \tag{7.227}$$

如果 $\rho=0$，则不存在内生性可用 OLS 得到一致估计，处理效应为 $\hat{\gamma}$，如果 $\rho\neq 0$，则存

在内生性，可用 heckit 估计（罗伊模型），分别对处理组和对照组求条件均值为：

$$E(Y_i|D_i=1,X_i,Z_i)=X_i'\beta+\gamma+\rho\sigma_\varepsilon\lambda(Z_i'\delta) \tag{7.228}$$

$$E(Y_i|D_i=0,X_i,Z_i)=X_i'\beta-\rho\sigma_\varepsilon\lambda(-Z_i'\delta) \tag{7.229}$$

$$So, E(Y_i|D_i=1,X_i,Z_i)-E(Y_i|D_i=0,X_i,Z_i)=\gamma+\rho\sigma_\varepsilon[\lambda(Z_i'\delta)+\lambda(-Z_i'\delta)] \tag{7.230}$$

从而处理效应 $ATE=\gamma+\rho\sigma_\varepsilon[\lambda(Z_i'\delta)+\lambda(-Z_i'\delta)]$。Heckit 方法的两步估计过程如下：

- Probit 模型估计 $P(D_i=1|Z_i)=\Phi(Z_i'\delta)$, 得到 $\hat{\delta}$，计算 $\hat{\lambda}$
- OLS 回归 $Y_i\to X_i,D_i,\hat{\lambda}$，得到 $\hat{\beta},\hat{\gamma},\hat{\rho\sigma_\varepsilon}$

Heckit 方法虽然计算方便，但是有一定的效率损失。我们设置调整项：

$$\lambda=\begin{cases}\lambda(Z_i'\delta),D_i=1\\-\lambda(-Z_i'\delta),D_i=0\end{cases} \tag{7.231}$$

得到统一方程为：

$$E(Y_i|X_i,Z_i)=X_i'\beta+\gamma D_i+\rho\sigma_\varepsilon\lambda_i \tag{7.232}$$

现在可以采用 MLE 直接估计上述模型得到处理效应 $\hat{\gamma}$。

最后一种估计方法为双重差分倾向分匹配 $DID-PSM$，即在倾向分匹配的基础上进一步消除不可观测变量的影响。具体操作是先计算倾向分，再根据匹配方法进行匹配，然后使用 DID 估计方法。

政策评估的总体思路可以分两步，第一步将处理效应化为处理组的处理效应 ATT，依赖可忽略假定，第二步是建立反事实框架求出 ATT，即得到 ATE。

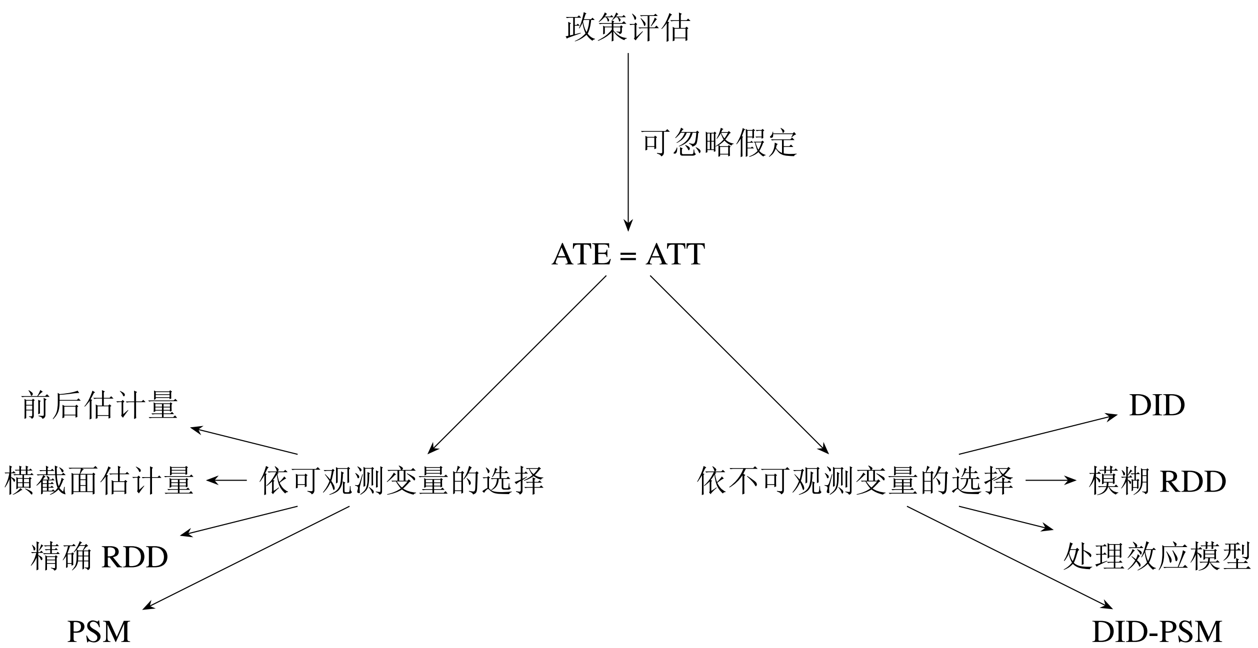

图 7.5: 政策评估方法

7.4 时间序列分析

时间序列数据的生成过程与截面数据、面板数据有很大的差别，在时间序列中，观测有前后顺序，一般具有自相关性。$ARMA$ 模型是最基本的时间系列模型，平稳性和遍历性是时间序列的基本特征，掌握这两点才能初步理解时间序列。

7.4.1 差分方程

设变量在 t 期的值为 y_t，满足如下动态方程

$$y_t = \phi y_{t-1} + \omega_t \tag{7.233}$$

该方程称为线性一阶差分方程。我们关心的是给定初始值和各期误差 ω，y_t 的值为多少，以及各期误差（冲击）对 y_t 的边际效应有多大。可以用迭代法表示出 y_t，

$$\begin{cases} y_0 = \phi y_{-1} + \omega_0, t = 0 \\ y_1 = \phi y_0 + \omega_1, t = 1 \\ \cdots \\ y_t = \phi y_{t-1} + \omega_t, t = t \end{cases} \Rightarrow y_t = \phi^{t+1} y_{-1} + \phi^t \omega_0 + \phi^{t-1}\omega_1 + \cdots + \phi\omega_{t-1} + \omega_0 \tag{7.234}$$

如果 y_{t-1} 已知，那么 y_{t+j} 可以表示为

$$y_{t+j} = \phi^{j+1} y_{t-1} + \phi^j \omega_t + \phi^{j-1}\omega_{t+1} + \cdots + \phi\omega_{t+j-1} + \omega_{t+j} \tag{7.235}$$

此时 ω_t 对 y_{t+j} 的边际效应为

$$\frac{\partial y_{t+j}}{\partial \omega_t} = \phi^j \tag{7.236}$$

ϕ^j 称为动态乘子，它只取决于 y_{t+j} 与 ω_t 之间的时间间隔，不取决于冲击 ω_t 所在的时间。如果在 t 期及以后，冲击 ω 永久提高一个单位，则这一永久改变对 y_{t+j} 的影响为

$$\frac{\partial y_{t+j}}{\partial \omega_t} + \frac{\partial y_{t+j}}{\partial \omega_{t+1}} + \cdots + \frac{\partial y_{t+j}}{\partial \omega_{t+j}} = \phi^j + \phi^{j-1} + \cdots + \phi + 1 = \frac{1-\phi^{j+1}}{1-\phi} \tag{7.237}$$

当 $|\phi| < 1$ 且 $j \to \infty$，上式的极限称为 ω 对 y 的长期效应，即

$$\lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial \omega_t} + \frac{\partial y_{t+j}}{\partial \omega_{t+1}} + \cdots + \frac{\partial y_{t+j}}{\partial \omega_{t+j}} = \frac{1}{1-\phi} \tag{7.238}$$

现在将一阶差分方程推广到 p 阶，即 y_t 取决于其 p 阶滞后变量和当期冲击，

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \omega_t \tag{7.239}$$

上式称为 p 阶差分方程。根据迭代法无法直接求出 y_{t+j} 的表达式，但是可以运用向量表

示将 p 阶差分方程转化为一阶差分方程，即

$$\varepsilon_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \cdots \\ y_{t-p+1} \end{bmatrix}, F = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, v_t = \begin{bmatrix} \omega_t \\ 0 \\ \cdots \\ 0 \end{bmatrix} \Rightarrow \varepsilon_t = F\varepsilon_{t-1} + v_t \tag{7.240}$$

根据迭代法可以求出 ε_t 的表达式，

$$\varepsilon_t = F^{t+1}\varepsilon_{-1} + F^t v_0 + \cdots + F v_{t-1} + v_t \tag{7.241}$$

带入向量定义得 ε_t 的第一项表达式为

$$y_t = f_{11}^{(t+1)} y_{-1} + f_{12}^{(t+1)} y_{-2} + \cdots + f_{1p}^{(t+1)} y_{-p} + f_{11}^{(t)}\omega_0 + f_{11}^{(t-1)}\omega_1 + \cdots + f_{11}^{(1)}\omega_{t-1} + \omega_t \tag{7.242}$$

这里 $f_{ij}^{(k)}$ 是 F^k 矩阵的 (i,j) 元素。同样，当初始值为 ε_{t-1} 时，ε_{t+j} 的表达式为

$$\varepsilon_{t+j} = F^{j+1}\varepsilon_{t-1} + F^j v_t + \cdots + F v_{t+j-1} + v_{t+j} \tag{7.243}$$

则

$$y_{t+j} = f_{11}^{(j+1)} y_{t-1} + f_{12}^{(j+1)} y_{t-2} + \cdots + f_{1p}^{(j+1)} y_{t-p} + f_{11}^{(j)}\omega_t + f_{11}^{(j-1)}\omega_{t+1} + \cdots + f_{11}^{(1)}\omega_{t+j-1} + \omega_{t+j} \tag{7.244}$$

p 阶差分方程的动态乘子为

$$\frac{\partial y_{t+j}}{\partial \omega_t} = f_{11}^{(j)} \tag{7.245}$$

特别的，低阶的 $f_{11}^{(j)}$ 容易求，如 $f_{11}^{(1)} = \phi_1, f_{11}^{(2)} = \phi_1^2 + \phi_2$。高阶的 $f_{11}^{(j)}$ 可以采用特征值来求。

定理 7.1

矩阵 F 的特征值是满足如下特征方程的 λ 值，

$$\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \cdots - \phi_{p-1}\lambda - \phi_p = 0 \tag{7.246}$$

如果矩阵 $F_{p\times p}$ 的特征值各不相同，则存在非奇异矩阵 $T_{p\times p}$ 使得 $F = T\Lambda T^{-1}$，其中 Λ 为由特征值组成的对角矩阵，即 $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$，从而 $F^j = (T\Lambda T^{-1})^j = T\lambda^j T^{-1}$，设 $T = t_{ij}, T^{\{} - 1\} = t^{\{} ij\}$，则

$$f_{11}^{(j)} = [t_{11}t^{11}]\lambda_1^j + [t_{12}t^{21}]\lambda_2^j + \cdots + [t_{1p}t^{p1}]\lambda_p^j \tag{7.247}$$

可以写成

$$f_{11}^{(j)} = c_1\lambda_1^j + c_2\lambda_2^j + \cdots + c_p\lambda_p^j, c_i = t_{1i}t^{i1}, i = 1, 2, \cdots, p \tag{7.248}$$

特别的，$\sum_{i=1}^p c_i = \sum_{i=1}^p t_{1i}t^{i1}$ 为单位矩阵 TT^{-1} 的第一个元素，故 $\sum_{i=1}^p c_i = 1$。p 阶差分方程的动态乘子可以写为

$$\frac{\partial y_{t+j}}{\partial \omega_t} = c_1\lambda_1^j + c_2\lambda_2^j + \cdots + c_p\lambda_p^j \tag{7.249}$$

特征值可以通过定义求出，即解 $|F - \lambda I| = 0$，I 为单位矩阵。c_i 的值由以下定理给出

定理 7.2

如果矩阵 $F_{p\times p}$ 具有 p 个不同的特征值，则 c_i 的值为

$$c_i = \frac{\lambda_i^{p-1}}{\prod_{k=1,k\neq i}^{p}(\lambda_i - \lambda_k)}, i = 1, 2, \cdots, p \tag{7.250}$$

一般的，记 p 阶差分方程的动态乘子为 ψ_j，则

$$\frac{\partial y_{t+j}}{\partial \omega_t} = \psi_j = f_{11}^{(j)} \tag{7.251}$$

如果矩阵 F 的特征值各不相同，分别设为 $\lambda_1, \lambda_2, \cdots, \lambda_p$ 则

$$\psi_j = c_1\lambda_1^j + c_2\lambda_2^j + \cdots + c_p\lambda_p^j, c_i = \frac{\lambda_i^{p-1}}{\prod_{k=1,k\neq i}^{p}(\lambda_i - \lambda_k)}, i = 1, 2, \cdots, p \tag{7.252}$$

如果 $\lambda_i, i = 1, 2, \cdots, p$ 均为实数，则设 λ_{absmax} 是绝对值最大的特征值，动态乘子由 λ_{absmax} 主导，

$$\lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial \omega_t} \frac{1}{\lambda_{absmax}^j} = c_{absmax} \tag{7.253}$$

如果 $|\lambda_{absmax}|$ 大于 1，则系统是爆炸性的，即冲击 ω_t 对 y 的影响随着时间增加而增加。如果存在共轭的特征值，如 λ_1, λ_2 是共轭复数，设 $\lambda_1 = a + b_i, \lambda_2 = a - bi$，则根据欧拉定理，可以写为

$$\lambda_{1,2} = R[cos(\theta) \pm isin(\theta)] = R[e^{\pm i\theta}], R = \sqrt{a^2 + b^2}, tan(\theta) = b/a \tag{7.254}$$

则 $\lambda_{1,2}^j = R^j[e^{\pm i\theta j}] = R^j[cos(\theta j) \pm isin(\theta j)]$。这里 c_1, c_2 为共轭复数，可以设为 $\alpha \pm i\beta$，则共轭复数对对动态乘子的贡献为

$$\begin{aligned}
c_1\lambda_1^j + c_2\lambda_2^j &= c_1R^j[cos(\theta j) + isin(\theta j)] + c_2R^j[cos(\theta j) - isin(\theta j)] \\
&= R^j[(c_1 + c_2)cos(\theta j) + i(c_1 - c_2)sin(\theta j)] \\
&= 2R^j[\alpha cos(\theta j) - \beta sin(\theta j)] \\
&= 2R^j\sqrt{\alpha^2 + \beta^2}sin(\theta j + \psi), tan(\psi) = \alpha/\beta
\end{aligned} \tag{7.255}$$

从而，共轭复数根对动态乘子的贡献是幂函数和正弦函数的积，如果 $R > 1$，这一贡献是爆炸性的。因而，在存在复数根时，动态乘子是由 $Max\{$实数根的绝对值, 复数根的模$\}$ 项主导的，如果该值大于 1，则 p 阶差分方程是爆炸性的。这一结果也可以表达为如果特征方程的根均在单位圆内，线性差分方程系统是稳定的，如果某个根在单位圆外，则线性差分方程是爆炸性的。

特别的，考虑二阶线性差分方程的参数值如何决定系统的爆炸性。二阶线性差分方程的特征方程为 $\lambda^2 - \phi_1\lambda - \phi_2 = 0$，解及分类如下：

当特征值有重复根时，需采用约当分解，暂时不考虑这一内容。

$\Delta=\phi_1^2+4\phi_2$	解	分类 1	分类 2	爆炸性
$\Delta>0$	$\lambda=\frac{\phi_1\pm\sqrt{\phi_1^2+4\phi_2}}{2}$	$\lambda_{max}>1$	$\phi_1\geq 2$	爆炸
			$\phi_1<2,\phi_1+\phi_2>1$	爆炸
		$\lambda_{min}<-1$	$\phi_1\leq -2$	爆炸
			$\phi_1>-2,\phi_2-\phi_1>1$	爆炸
		$\lvert\lambda\rvert\leq 1$	$\phi_1+\phi_2\leq 1,\phi_2-\phi_1\leq 1,-2<\phi_1<2$	稳定
$\Delta<0$	$\lambda=\frac{\phi_1\pm\sqrt{-(\phi_1^2+4\phi_2)i}}{2}$	$R=\sqrt{-\phi_2}$	$\phi_2<-1$	爆炸
			$-1\leq\phi_2<0$	稳定

表 7.4: 二阶差分方程的动态性

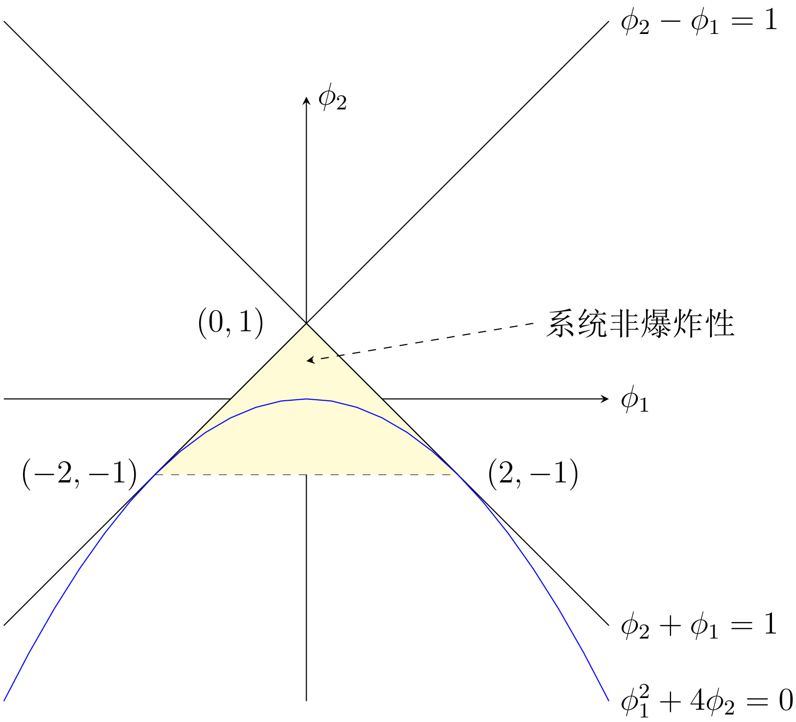

图 7.6: 二阶差分方程的动态性

7.4.2 滞后算子

定义 7.4

时间序列算子是把一个或一组时间序列转换为一个新的时间序列的操作。

加法算子和乘法算子遵循所有代数运算法则。滞后算子是取滞后一致的值生成新序列，记为 L，则滞后算子作用于 $\{x_t\}_{-\infty}^{+\infty}$ 得

$$Lx_t \equiv x_{t-1} \tag{7.256}$$

滞后算子可以叠加，$\forall k \in Q, L^k x_t = x_{t-k}$。滞后算子满足以下性质，

- 交换律：$L(\beta x_t) = \beta L x_t, \beta$ 为常数
- 分配率：$L(x_t + \omega_t) = x_{t-1} + \omega_{t-1}$

故滞后算子满足乘法的代数运算法则，因此可以将对 $\{x_t\}_{-\infty}^{+\infty}$ 作滞后运算表达为 L乘x_t。形如 $(aL + bL^2)$ 的式子称为滞后算子多项式，可以用代数运算法则做变换。特别的，如果对常数序列做滞后运算，结果仍为原序列。可以用滞后算子来对线性差分方程做运算。

采用滞后算子表达一阶差分方程，

$$y_t = \phi L y_t + \omega_t \tag{7.257}$$

可以做代数运算表达出 y_t，

$$(1 - \phi L) y_t = \omega_t \tag{7.258}$$

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L) y_t = (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)\omega_t \tag{7.259}$$

$$(1 - \phi^{t+1} L^{t+1}) y_t = (1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)\omega_t \tag{7.260}$$

$$y_t - \phi^{t+1} y_{-1} = \omega_t + \phi\omega_{t-1} + \cdots + \phi^t \omega_0 \tag{7.261}$$

$$y_t = \phi^{t+1} y_{-1} + \omega_t + \phi\omega_{t-1} + \cdots + \phi^t \omega_0 \tag{7.262}$$

这一结果与迭代法结果完全一致。如果 y_{-1} 有限，且 $|\phi| < 1$，随着 t 增大，有

$$(1 + \phi L + \phi^2 L^2 + \cdots + \phi^t L^t)(1 - \phi L) y_t = y_t - \phi^{t-1} y_{-1} \cong y_t, t\text{足够大} \tag{7.263}$$

从而，如果时间序列 $\{y_t\}_{-\infty}^{+\infty}$ 有界，即 $\forall t, |y_t| < M, M$ 为有限的数。则当 $|\phi| < 1$ 时，可以定义 $(1 + \phi L + \phi^2 L^2 + \cdots + \phi^j L^j)$ 作为滞后算子 $1 - \phi L$ 的估计值，它满足

$$(1 - \phi L)^{-1} = \lim_{j \to \infty} (1 + \phi L + \phi^2 L^2 + \cdots + \phi^j L^j) \tag{7.264}$$

算子 $(1 - \phi L)^{-1}$ 满足 $(1 - \phi L)^{-1}(1 - \phi L) = 1, 1$ 是恒等算子。从而，满足 $|\phi| < 1$ 的一阶差分方程（需要有界或平稳）的解可以表达为

$$y_t = (1 - \phi L)^{-1} \omega_t \tag{7.265}$$

如果将二阶差分方程改写为滞后方程，得

$$(1 - \phi_1 L - \phi_2 L^2) y_t = \omega_t \tag{7.266}$$

要找出 $(1 - \phi_1 L - \phi_2 L^2)$ 的“逆算子”，需要解方程 $1 - \phi_1 z - \phi_2 z^2 = 0$，设解为 z_1, z_2，即

$(1-\phi_1 z_1)(1-\phi_2 z_2)=0$，则可以将滞后算子多项式分解为

$$(1-\phi_1 L-\phi_2 L^2)=(1-\lambda_1 L)(1-\lambda_2 L), \lambda_1=z^{-1}, \lambda_2=z^{-1} \tag{7.267}$$

这里的 λ_1, λ_2 恰好是前一节提到的特征方程的解，$\lambda^2-\phi_1\lambda-\phi_2=0$。

定理 7.3

滞后算子多项式 $(1-\phi_1 L-\phi_2 L^2)$ 的因式分解 $(1-\lambda_1 L)(1-\lambda_2 L)$ 和求解矩阵 F 的特征值是同一个运算过程，并且二者的 λ 完全相同。

现在，对同一个线性差分方程，有两种根，需要注意区分。一般用特征值指代由矩阵 F 的求特征值过程导出的特征方程的根，记为 λ。

假设二阶差分方程的特征根均在单位圆内，则动态系统是稳定的，如果 $\lambda_1\neq\lambda_2$，则

$$(1-\lambda_1 L)(1-\lambda_2 L)y_t=\omega_t \tag{7.268}$$

$$\begin{aligned} y_t &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\omega_t \\ &= (\lambda_1-\lambda_2)^{-1}[\lambda_1(1-\lambda_1 L)^{-1}-\lambda_2(1-\lambda_2 L)^{-1}]\omega_t \\ &= (\lambda_1-\lambda_2)^{-1}(\lambda_1+\lambda_1^2 L+\cdots-\lambda_2-\lambda_2^2 L-\cdots)\omega_t \\ &= [c_1+c_2]\omega_t+[c_1\lambda_1+c_2\lambda_2]\omega_{t-1}+\cdots, c_1=\frac{\lambda_1}{\lambda_1-\lambda_2}, c_2=\frac{\lambda_2}{\lambda_2-\lambda_1} \end{aligned} \tag{7.269}$$

$$\Rightarrow \frac{\partial y_{y+j}}{\partial \omega_t}=c_1\lambda_1^j+c_2\lambda_2^j \tag{7.270}$$

这与前一节的结果完全一致。下面把这一结果推广到 p 阶差分方程，注意理解 $c_i, i=1,2,\cdots,p$ 的构造过程。

p 阶差分方程的滞后形式为

$$(1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p)y_t=\omega_t \tag{7.271}$$

假设其特征根 $\lambda_i, i=1,2,\cdots,p$ 均在单位圆内（动态系统稳定），且各不相同，则 $(1-\lambda_i L)^{-1}, i=1,2,\cdots,p$ 存在，即

$$(1-\lambda_1 L)(1-\lambda_2 L)(1-\lambda_p L)y_t=\omega_t \tag{7.272}$$

$$\begin{aligned} y_t &= (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\cdots(1-\lambda_p)^{-1}\omega_t \\ &= \sum_{i=1}^{p} c_i(1-\lambda_i L)^{-1}\omega_t \\ &= \sum_{i=1}^{p} c_i(1+\lambda_1 L+\lambda_1^2 L^2+\cdots)\omega_t \\ &= \sum_{i=1}^{p} c_i(\omega_t+\lambda_1\omega_{t-1}+\lambda_1^2\omega_{t-2}+\cdots) \end{aligned} \tag{7.273}$$

$$\Rightarrow \frac{\partial y_{t+j}}{\partial \omega_t}=\sum_{i=1}^{p} c_i\lambda_i^j \tag{7.274}$$

这里的 $c_i, i = 1, 2, \cdots, p$ 满足

$$\frac{1}{(1-\lambda_1 z)(1-\lambda_2 z)\cdots(1-\lambda_p z)} = \frac{c_1}{1-\lambda_1 z} + \frac{c_2}{1-\lambda_2 z} + \cdots + \frac{c_p}{1-\lambda_p z} \tag{7.275}$$

这一等式成立的依据是有理函数可以化简为最简分数（唯一），可以使用的是待定系数法分解。通过给 z 赋值可以解出待定系数 $c_i, i = 1, 2, \cdots, p$。

上式可化为，

$$c_1 = \frac{1}{(1-\lambda_2 z)(1-\lambda_3 z)\cdots(1-\lambda_p z)} - \left(\frac{c_2}{1-\lambda_2 z} + \cdots + \frac{c_p}{1-\lambda_p z}\right)(1-\lambda_1 z) \tag{7.276}$$

对该式可以设 $1-\lambda_1 z = 0$，即 $z = 1/\lambda_1$，得到

$$c_1 = \frac{\lambda_1^{p-1}}{\prod_{i=1, i\neq 1}^{p}(\lambda_1 - \lambda_i)} \tag{7.277}$$

虽然对于式(7.275)，$z \neq 1/\lambda_1$，否则该式没有定义，但是式(7.276)表明，当 $\lim\limits_{z\to 1/\lambda_1} c_1 = \frac{\lambda_1^{p-1}}{\prod_{i=1,i\neq 1}^{p}(\lambda_1-\lambda_i)}$，并且有理函数的分解是唯一的，故式(7.277)的结果是唯一成立的。同理得出所有的 $c_i, i = 1, 2, \cdots, p$，

$$c_i = \frac{\lambda_i^{p-1}}{\prod_{k=1, k\neq i}^{p}(\lambda_i - \lambda_k)}, i = 1, 2, \cdots, p \tag{7.278}$$

现在，可以将 p 阶差分方程记为

$$y_t = \psi(L)\omega_t, \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \tag{7.279}$$

这里 $\psi_i, i = 0, 1, \cdots$ 为动态乘子，$\psi(L)$ 为滞后算子多项式，相应的多项式为 $\psi(z)$，值为

$$\psi(z) = \frac{1}{1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p} \tag{7.280}$$

因而 p 阶差分方程的长期乘子为

$$\lim_{j\to\infty} \frac{\partial y_{t+j}}{\partial \omega_t} + \frac{\partial y_{t+j}}{\partial \omega_{t+1}} + \cdots + \frac{\partial y_{t+j}}{\partial \omega_{t+j}} = \lim_{j\to\infty} \psi_0 + \psi_1 + \cdots + \psi_j = \psi(1) = \frac{1}{1 - \phi_1 - \phi_2 - \cdots - \phi_p} \tag{7.281}$$

以上的分析是在给定 p 阶差分方程、p 个初始值和各期冲击 ω 以后，确定变量 y 的取值和性质。现在，分析的问题变为，给定 p 阶差分方程的性质，初始值应该满足什么条件？给定未来的值，如何求出初始值？这里采用股利折现估值模型来分析这一问题。

设股票价格为 P_t，股利支付为 D_t，投资者在 t 期购入股票，在 $t+1$ 期卖出股票，总回报率为

$$r_{t+1} = \frac{P_{t+1} - P_t}{P_t} + \frac{D_t}{P_t} \tag{7.282}$$

假设总回报率恒为 r，则股票价格预测方程为

$$P_{t+1} = (1+r)P_t - D_t \tag{7.283}$$

这是一个一阶的线性差分方程系统，可以根据初始值 P_0 和各期股利预测 $t+1$ 期的股票价格为（时间从 0 开始）

$$P_{t+1} = (1+r)^{t+1}P_0 - (1+r)^t D_0 - (1+r)^{t-1}D_1 - \cdots - D_t \tag{7.284}$$

进一步假设各期股利不变，均为 D，则

$$\begin{aligned}P_{t+1} &= (1+r)^{t+1}P_0 - (1+r)^t D - (1+r)^{t-1}D - \cdots - D \\ &= (1+r)^{t+1}P_0 - \frac{1-(1+r)^{t+1}}{1-(1+r)}D \\ &= (1+r)^{t+1}[P_0 - D/r] + D/r \end{aligned} \tag{7.285}$$

如果 $P_0 = D/r$，则 $P_t = D/r, t = 1, 2, \cdots$，此时投资者的收入只有股利支付。$P_0 = D/r$ 被称为市场基本面。如果 $P_0 > D/r$，则股票价格 P_t 将呈指数上涨。投资者相信股票会一直上涨，这是不可持续的，此时出现投机泡沫。

如果没有出现投机泡沫，股票价格有上限，即 $\forall t |P_t| < M, M$ 为常数。假设股利 $\{D_t\}_{-\infty}^{+\infty}$ 为有限序列，初始的股票价格决定于未来的股票价格和股利，递推方程为

$$P_t = \frac{1}{1+r}[P_{t+1} + D_t] \tag{7.286}$$

采用迭代法得到

$$P_t = [\frac{1}{1+r}]^T P_{t+T} + [\frac{1}{1+r}]^T D_{t+T-1} + [\frac{1}{1+r}]^{T-1} D_{t+T-2} + \cdots + [\frac{1}{1+1}]D_t \tag{7.287}$$

由于 $\{P_t\}_{-\infty}^{+\infty}$ 有界,则 $\lim\limits_{T\to\infty}[\frac{1}{1+r}]^T P_{t+T} = 0$,由于 $\{D_t\}_{-\infty}^{+\infty}$ 也有界,则极限 $\lim\limits_{T\to\infty}\sum_{j=0}^{T}[\frac{1}{1+r}]^{j+1}D_{t+j}$ 存在，即

$$P_t = \sum_{j=0}^{\infty}[\frac{1}{1+r}]^{j+1}D_{t+j} \tag{7.288}$$

这被称为市场基本面解式。设 $t = 0$，将上式带入式(7.284)，得到的结果满足上式。而将初始值设为其他值都将使得股票违背有限条件。

下面用滞后算子来表示股价估值模型。定义滞后算子的逆为 L^{-1}，则 $L^{-1}x_t = x_{t+1}$，且 $L^{-k}L^j = L^{j-k}$，L^0 为恒等算子。则

$$P_{t+1} = (1+r)P_t - D_t \tag{7.289}$$

$$(1 - \frac{1}{1+r}L^{-1})P_t = \frac{1}{1+r}D_t \tag{7.290}$$

$$(1 - \frac{1}{1+r}L^{-1})(\sum_{i=0}^{k}(\frac{1}{1+r})^k L^{-k})P_t = (\sum_{i=0}^{k}(\frac{1}{1+r})^i L^{-i})\frac{1}{1+r}D_t \tag{7.291}$$

$$(1 - (\frac{1}{1+r})^{k+1}L^{-(k+1)})P_t = (\sum_{i=0}^{k}(\frac{1}{1+r})^i L^{-i})\frac{1}{1+r}D_t \tag{7.292}$$

$$P_t - (\frac{1}{1+r})^{k+1}P_{t+k+1} = \sum_{i=0}^{k}(\frac{1}{1+r})^{i+1}D_{t+i} \tag{7.293}$$

假设股价有上限，则 $\lim\limits_{k\to\infty}(\frac{1}{1+r})^{k+1}P_{t+k+1} = 0$，令 $k \to +\infty$，

$$P_t = \sum_{i=0}^{\infty}(\frac{1}{1+r})^{i+1}D_{t+i} \tag{7.294}$$

这一结果与直接采用迭代法算出的结果完全一致。值得注意的是，这里定义了滞后算子的逆的多项式的逆，如 $(1-\gamma L^{-1}) = \sum_{i=0}^{\infty}\gamma^i L^{-i}, |\gamma| < 1$，这表明 $(1-\phi L)^{-1}, |\phi| > 1$ 的逆也

存在，因为 $(1-\phi L)^{-1} = -(\phi^{-1}L^{-1})(1-\phi^{-1}L^{-1})$（滞后算子满足代数运算定律，看做实数变量来推导这个等式），从而滞后算子多项式的逆为

$$(1-\phi L)^{-1} = \begin{cases} \sum_{i=0}^{\infty} \phi^i L^i, |\phi| < 1 \\ -\phi^{-1}L^{-1}\sum_{i=0}^{\infty} \phi^{-i}L^{-i}, |\phi| > 1 \end{cases} \tag{7.295}$$

一般的，在上式时需要假设时间序列有限，以避免系统爆炸性的情况。

7.4.3 ARMA 过程

定义 7.5

时间序列在 t 期的期望是其概率分布的均值，假设该期望存在，则

$$E(Y_t) \equiv \int_{-\infty}^{+\infty} y_t f_{Y_t}(y_t)dy_t = \underset{I\to\infty}{plim}(1/I)\sum_{i=1}^{I} Y_t^{(i)} \tag{7.296}$$

这里的 Y_t 是时间序列在 t 期的总体，y_t 是 Y_t 的所有可能取值，$Y_t^{(i)}$ 是 Y_t 的第 i 个观测。♣

定义 7.6

随机变量 Y_t 的方差 γ_{0t} 定义为

$$\gamma_{0t} = E(Y_t-\mu_t)^2 = \int_{-\infty}^{+\infty} (y_t-\mu_t)^2 f_{Y_t}(y_t)dy_t \tag{7.297}$$

随机变量 Y_t 的第 j 个自协方差 γ_{jt} 为

$$\begin{aligned} \gamma_{jt} &= E(Y_t-\mu_t)(Y_{t-j}-\mu_{t-j}) \\ &= \underset{R^j}{\iint\cdots\int} (y_t-\mu_t)(y_{t-j}-\mu_{t-j}) f_{Y_t,Y_{t-1},\cdots,Y_{t-j}}(y_t,y_{t-1},\cdots,y_{t-j})dy_t dy_{t-1}\cdots dy_{t-j} \\ &= \underset{I\to\infty}{plim}(1/I)\sum_{i=1}^{I}[Y_t^i-\mu_t][Y_{t-j}^i-\mu_{t-j}] \end{aligned} \tag{7.298}$$

♣

定义 7.7

对于过程 Y_t，若无论均值 μ_t 还是自协方差 γ_{jt} 都不依赖时间 t，则称此过程是协方差平稳的或弱平稳的，即

$$\begin{cases} \mu_t = E(Y_t) = \mu, t \in Z \\ \gamma_{jt} = E(Y_t-\mu)(Y_{t-j}-\mu) = \gamma_j, t, j \in Z \end{cases} \tag{7.299}$$

对于协方差平稳过程，$\gamma_j = \gamma_{-j}$。♣

如果 $\forall j_k, k=1,2,\cdots,n, (Y_t, Y_{t+j_1},\cdots,Y_{t+j_n})$ 的联合分布只取决于时间间隔 $j_k, k=1,2,\cdots,n$ 而不取决于时间 t，则该过程是严平稳的。如果一个过程是严平稳的，则一定是弱平稳的。

定义 7.8

对于协方差平稳过程，如果当 $T \to \infty$ 时，样本均值依概率收敛于总体均值，则该过程具有均值遍历性，记为

$$\lim_{T\to\infty}(1/T)\sum_{t=1}^{T} y_t^{(1)} \xrightarrow{P} E(Y_t) \tag{7.300}$$

如果协方差平稳过程满足 $\sum_{j=0}^{\infty}|\gamma_j| < \infty$，则 $\{Y_t\}$ 关于均值是遍历的。♣

定义 7.9

零均值、同方差且跨期不相关的序列为白噪声，即白噪声序列 $\{\varepsilon_t\}_{-\infty}^{+\infty}$ 满足

$$\begin{cases} E(\varepsilon_t) = 0 \\ E(\varepsilon_t^2) = \sigma^2 \\ E(\varepsilon_t\varepsilon_\tau) = 0, t \neq \tau \end{cases} \tag{7.301}$$

如果白噪声序列同时各期独立，则称为独立白噪声序列，如果独立白噪声序列同时服从正态分布，则称为高斯白噪声。♣

ARMA 过程是移动平均过程和自回归过程的联合过程，下面分别分析这两类过程。一阶移动平均过程 $MA(1)$ 的动态方程如下

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1} \tag{7.302}$$

这里 $\{\varepsilon_t\}$ 是白噪声序列（下同），方差为 σ^2。$MA(1)$ 的性质如下

- 期望：$\mu_t = E(Y_t) = \mu$
- 方差：$\gamma_{0t} = E(Y_t - \mu)^2 = (1+\theta^2)\sigma^2$
- 自协方差：$\gamma_{jt} = \begin{cases} \theta\sigma^2, j = 1 \\ 0, j > 1 \end{cases}$

从而 $MA(1)$ 过程是协方差平稳的，且 $\sum_{j=0}^{\infty}|\gamma_j| = (1+\theta^2)\sigma^2 + |\theta\sigma^2|$，则 $MA(1)$ 过程也是均值遍历的。

协方差平稳过程的第 j 个自相关系数定义为 $\rho_j \equiv \gamma_j/\gamma_0, |\rho_j| \leq 1$。则 $MA(1)$ 过程的一阶自相关系数为 $\rho_1 = \frac{\theta}{1+\theta^2}$，高阶自相关系数为 0。当 $|\theta| = 1$ 时，ρ_1 取最值 ± 0.5，并且 $\theta, 1/\theta$ 对应的 ρ_1 相同。

$MA(q)$ 的形式如下

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \tag{7.303}$$

则

- 均值：$E(Y_t) = \mu$
- 方差：$\gamma_0 = E(Y_t - \mu)^2 = (1+\theta_1^2+\theta_2^2+\cdots+\theta_q^2)\sigma^2$
- 自协方差：$\gamma_j = \begin{cases} [\theta_j + \theta_{j+1}\theta_1 + \cdots + \theta_q\theta_{q-j}]\sigma^2, j = 1, 2, \cdots, q \\ 0, j > q \end{cases}$

$MA(q)$ 过程具有弱平稳性和遍历性。无限阶移动平均过程 $MA(\infty)$ 的形式为

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi \varepsilon_{t-j} = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \cdots \tag{7.304}$$

如果序列 $\{\psi_j\}$ 平方可加，则 $MA(\infty)$ 是弱平稳的，一般使用更强的条件，$\{\psi_j\}$ 绝对可加。$MA(\infty)$ 过程的均值为 μ,方差为 $\lim_{T\to\infty} \sum_{i=0}^{T} \psi_i^2 \sigma^2$,自协方差系数为 $\gamma_j = (\psi_j\psi_0 + \psi_{j+1}\psi_1 + \cdots)\sigma^2$。系数绝对可加的 $MA(\infty)$ 过程，自协方差也绝对可加，即 $\sum_{j=0}^{\infty} |\gamma_j| < \infty$。

一阶自回归过程 $AR(1)$ 的形式为

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \tag{7.305}$$

这里 $\{\varepsilon_t\}$ 是白噪声序列（下同），自回归过程具有线性差分方程的形式。可以递推出

$$Y_t = c/(1-\phi) + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots \tag{7.306}$$

这等价于 $MA(\infty)$ 过程，且 $\psi_j = \phi^j$，则该过程协方差平稳条件为 ψ_j 平方可加。如果 $|\phi| < 1$，则 $\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi|^j = 1/(1-|\psi|) < \infty$，即 ψ_j 绝对可加，协方差平稳条件也成立。如果 $|\phi| \geq 1$，$\sum_{j=0}^{\infty} \psi_j^2 = \sum_{j=0}^{\infty} \phi^{2j}$ 无限，则 Y_t 序列不具有弱平稳性。因而，在 $|\psi| < 1$ 的条件下，Y_t 序列的数字特征如下

- 期望：$E(Y_t) = \mu = c/(1-\phi)$
- 方差：$\gamma_0 = \sigma^2/(1-\phi^2)$
- 自协方差：$\gamma_j = \frac{\phi^j}{1-\phi^2}\sigma^2$
- 自相关系数：$\rho_j = \gamma_j/\gamma_0 = \phi^j$

也可以假设该过程是协方差平稳的，然后由 $AR(1)$ 的线性差分方程形式直接求出上述数字特征，此时需要构造出对应的项，如两边取期望、平方再取期望、同乘 $Y_{t-j} - \mu$ 项等。

二阶自回归过程 $AR(2)$ 的形式为

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{7.307}$$

写成滞后算子的形式为 $(1 - \phi_1 L - \phi_2 L^2)Y_t = c + \varepsilon_t$，这是二阶线性差分方程的形式，差分系统平稳的条件是特征方程的特征根均在单位圆内，这也是 $AR(2)$ 过程弱平稳的条件。如果弱平稳条件成立，即 $\psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1}$，则

$$Y_t = \psi(L)c + \psi(L)\varepsilon_t = \frac{c}{1-\phi_1-\phi_2} + \psi(L)\varepsilon_t \tag{7.308}$$

此时 $\sum_{j=0}^{\infty} |\psi_j| < \infty$，即 $AR(2)$ 过程是均值遍历的。假设 $AR(2)$ 过程是协方差平稳的，则 $E(Y_t) = c + \phi_1 E(Y_{t-1}) + \phi_2 E(Y_{t-2}) + E(\varepsilon_t) \Rightarrow \mu = c/(1-\phi_1-\phi_2)$，这就是期望，下面求二阶

矩

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{7.309}$$

$$Y_t = \mu(1-\phi_1-\phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t \tag{7.310}$$

$$Y_t - \mu = \phi_1(Y_{t-1}-\mu) + \phi_2(Y_{t-2}-\mu) + \varepsilon_t \tag{7.311}$$

$$E(Y_t-\mu)(Y_{t-j}-\mu) = \phi_1 E(Y_{t-1}-\mu)(Y_{t-j}-\mu) + \phi_2 E(Y_{t-2}-\mu)(Y_{t-j}-\mu) + E\varepsilon_t(Y_{t-j}-\mu) \tag{7.312}$$

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2}, j = 1, 2, \cdots \tag{7.313}$$

$$/\gamma_0 \Rightarrow \rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} \tag{7.314}$$

$$\tag{7.315}$$

当 $j=1$，$\rho_1 = \phi_1 + \phi_2\rho_1, \rho_1 = \phi_1/(1-\phi_2)$，当 $j=2$，$\rho_2 = \phi_1\rho_1 + \phi_2 = \phi_1^2/(1-\phi_2) + \phi_2$。最后求方差，

$$E(Y_t-\mu)^2 = \phi_1 E(Y_{t-1}-\mu)(Y_t-\mu) + \phi_2 E(Y_{t-2}-\mu)(Y_t-\mu) + E(\varepsilon_t)(Y_t-\mu) \tag{7.316}$$

$$\begin{aligned} E(\varepsilon_t)(Y_t-\mu) &= \phi_1 E(\varepsilon_t)(Y_{t-1}-\mu) + \phi_2 E(\varepsilon_t)(Y_{t-2}-\mu) + E(\varepsilon_t^2) \\ &= \sigma^2 \end{aligned} \tag{7.317}$$

$$\Rightarrow \gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2 \tag{7.318}$$

$$\rho_1, \rho_2, \gamma_0 \Rightarrow \gamma_0 = \frac{(1-\phi_2)\sigma^2}{(1+\phi_2)[(1-\phi_2)^2-\phi_1^2]}\sigma^2 \tag{7.319}$$

类似的，$AR(p)$ 过程的协方差平稳条件为方程 $\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2}\cdots - \phi^p = 0$ 的根都在单位圆内，等价于 $1 - \phi_1 z - \phi_2 z^2 - cdots - \phi_p z^p = 0$ 的根都在单位圆外。假设该过程弱平稳，则

- 期望：$\mu = c/(1-\phi_1-\phi_2-\cdots-\phi_p)$
- 协方差迭代式：$\gamma_j = \begin{cases} \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p}, j = 1, 2, \cdots \\ \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2, j = 0 \end{cases}$

当 $j>0$，协方差迭代式是一个 p 阶线性差分方程，一般解可以写作 $\gamma_j = \sum_{i=1}^p g_i\lambda_i^j, \lambda$ 是 $\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \cdots - \phi_p = 0$ 的特征值。

$ARMA(p,q)$ 的形式如下

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \tag{7.320}$$

滞后形式如下

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t \tag{7.321}$$

当 $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ 的根都在单位圆外，$ARMA(p,q)$ 过程是弱平稳的，此时可对左侧的滞后算子多项式求逆，则

$$Y_t = \mu + \psi(L)\varepsilon_t \tag{7.322}$$

这里

$$\psi(L)=\frac{1+\theta_1L+\theta_2L^2+\cdots+\theta_qL^q}{1-\phi_1L-\phi_2L^2-\cdots-\phi_pL^p},\sum_{j=0}^{\infty}|\psi_j|<\infty \tag{7.323}$$

$$\mu=\frac{c}{1-\phi_1-\phi_2-\cdots-\phi_p} \tag{7.324}$$

协方差平稳的 $ARMA(p,q)$ 过程的离差形式为

$$Y_t-\mu=\phi_1(Y_{t-1}-\mu)+\phi_2(Y_{t-2}\mu)+\cdots+\phi_p(Y_{t-p}-\mu)+\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\cdots+\theta_q\varepsilon_{t-q} \tag{7.325}$$

上式两边同乘 $Y_{t-j}-\mu$ 并取期望可得到自协方差，当 $j>q$

$$\gamma_j=\phi_1\gamma_{j-1}+\phi_2\gamma_{j-2}+\cdots+\phi_p \tag{7.326}$$

这也是一个 p 阶线性差分方程。当 $j\le q$ 时，$E(Y_{t-j}-\mu)(\theta_j\varepsilon_{t-j})\ne 0$，自协方差的形式比较复杂。

$ARMA(p,q)$ 过程存在过度参数化的可能，如果 AR 过程和 MA 过程的滞后算子多项式有相同的特征根，如 λ_0，则两边可以同时乘以 $(1-\phi_0L)^{-1}$ 消去这一冗余。

如果需进一步把握 $ARMA(p,q)$ 的协方差，则需引入自协方差生成函数。

第 8 章 运筹学

8.1 线性规划与单纯形法

经典线性规划模型（LP）的一般形式可以概况如下：

$$目标函数 Max/Min : Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \tag{8.1}$$

$$约束条件 s.t. \begin{cases} a_{11}x_1 + \ldots + a_{1n}x_n \le (=, \ge)b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n \le (=, \ge)b_2 \\ \ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n \le (=, \ge)b_m \\ x_1, x_2, \ldots, x_n \ge 0 \end{cases} \tag{8.2}$$

$$决策变量 x_1, x_2, \ldots, x_n \tag{8.3}$$

其 Σ 形式为：

$$Max/Min : Z = \sum_{j=1}^{n} c_jx_j \tag{8.4}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{ij}x_j \le (=, \ge)b_i, i = 1, \ldots, m \\ x_1, x_2, \ldots, x_n \ge 0 \end{cases} \tag{8.5}$$

其向量形式为：

$$Max/Min : Z = CX \tag{8.6}$$

$$s.t. \begin{cases} AX \le (=, \ge)b \text{ OR } P_1x_1 + P_2x_2 + \ldots + P_nx_n \le (=, \ge)b \\ X \ge 0 \end{cases} \tag{8.7}$$

where, $C = [c_1, \ldots, c_n], X = [x_1, \ldots, x_n]', A = [P_1, \ldots, P_n], P_i = [a_{1i}, \ldots, ami], b = [b_1, b_2, \ldots, b_m]$。

提炼出标准形式：

$$MAX : Z = CX \tag{8.8}$$

$$s.t. \begin{cases} AX = b \\ X, b \ge 0 \end{cases} \tag{8.9}$$

其中，b 可以看做市场资源，C 是价值系数，A 是技术系数/消耗系数/工艺系数。由一般形式转化为标准形式的方法如下：

- $MinZ = CX$ TO $MAXZ' = -CX = C'X$
- $s.t. :\le$ TO $+x'$(松弛变量) and $\ge$ TO $-x'$(剩余变量)
- $b_i \le 0$ TO 方程两边同乘以-1

- $x_j < 0$ TO $-x'_j$
- x_j无约束 TO $x_j = x'_j - x''_j, x'_j$ and $x''_j \geq 0$

对于线性规划问题（LP），首先定义术语：

- 可行域：满足约束 $AX = b, X, b \geq 0$ 的 X 的集合
- 可行解：$X \in s.t$
- 最优解：使得 CX 最大的 X^*
- 最优值：$Z^* = CX^*$
- 基：A 中的可逆 m 阶子矩阵 B,$A = [B_{m\times m}, N_{m\times(n-m)}]$, N 为非基。

对应的 $AX = [B, N][X_B, X_N]'$, X_B 为基变量，X_N 为非基变量。令 $X_N = 0$，则 $BX_B = b, X_B = B^{-1}b$，得到基解 $X = [B^{-1}b, 0]'$。在可行域中的基解是基可行解，它对应的基为可行基。这些概念的图形化表示如下：

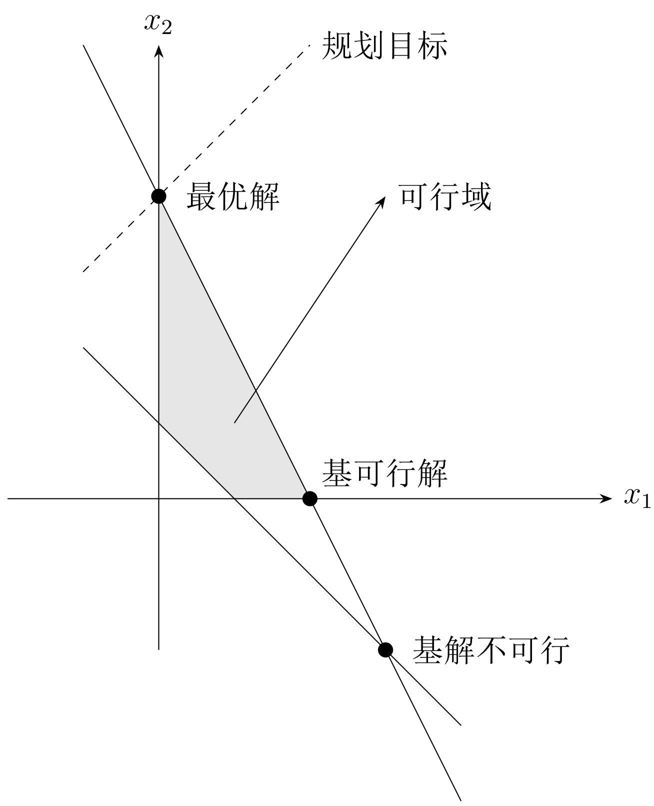

图 8.1: 可行域、最优解和基

可以证明，线性规划问题具有如下特征：

定理 8.1

LP 问题的可行域是 n 维凸集 (单纯形，有界，有限顶点)。

定理 8.2

LP 问题的基可行等价于可行域的顶点。

定理 8.3

LP 问题一定可以在某顶点处取得最优解（若有）。

定理8.1，8.1，8.1表明线性规划问题是凸规划的问题，可能的最优解是有限的，因此可以通过测试和换点的方式“枚举”出最优解，这就是单纯形法的基本思想。

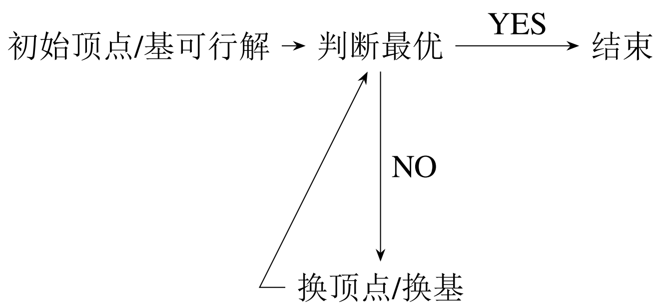

图 8.2: 单纯形法的基本思想

现在，解线性规划问题的核心在于：(1) 找出初始可行解；(2) 判断最优和换基。首先解决第二个问题。假设可行解 $X=[X_B,X_N]'$，则用 X_N 来决定 X_B 有 $AX=BX_B+NX_N=b$，即 $X_B=B^{-1}b-B^{-1}NX_N$，目标值为 $Z=CX=C_BX_B+C_NX_N=C_BB^{-1}b+(C_N-C_BB^{-1}N)X_N$，设 $\sigma=C_N-C_BB^{-1}N$，则:

$$Z=C_BB^{-1}b+\sigma X_N \tag{8.10}$$

这里，令 $X_N=0$，得到基 B 对应的基可行解 $B^{-1}b$，因此该顶点的目标值为 $Z_0=C_BB^{-1}b$，从而 $Z-Z_0=\sigma X_N$ 是判断是否最优的关键，这里遍历 X_N 即代表了遍历可行解。此时，引入单纯形表将计算过程可视化：

表 8.1: 单纯形表

	C_B	C_N	
C_BX_B	B	N	b
C_BX_B	E	N'	b'
		C_N-C_BN'	

这里，

$$\begin{aligned}\sigma =& C_N-C_BB^{-1}N && (8.11)\\ =& C_N-C_B(P'_{m+1},\ldots,P'_n) && (8.12)\\ =& (C_N-C_BP'_{m+1},\ldots,C_N-C_BP'_n) && (8.13)\\ =& (\sigma_{m+1},\ldots,\sigma_n) && (8.14)\end{aligned}$$

根据 σ 的值可以判断是否最优以及如何调整，σ 被称为检验数。

1. $\sigma_{m+1},\ldots,\sigma_n<0$，由 $X_N\geq 0$，得到 $\sigma<0$，此时的解为最优解。这里，保持目标值换基得到的解不会改变。
2. 存在 $\sigma_j=0$，则存在无数个可行解的目标值与最优值相同，即存在无穷个最优解。

3. 存在 $\sigma_j > 0$，则可以通过调整基增加目标值，基的调整服从如下的法则：（1）x_j 进基以后必须选择 x_l 出基；（2）新的基解必须是可行解。

假设 $\sigma_j > 0$，则假设前 m 个变量为可行基，单纯形表如下：

表 8.2: 单纯形表换基

		$c_1 \dots c_m$	$c_{m+1} \dots c_j \dots c_n$	
c_1	x_1	$1 \dots 0$	$a'_{1,m+1} \dots a'_{1,j} \dots a'_{1,n}$	b'_1
$\dots$		$\dots$	$\dots$	$\dots$
c_l	x_l	$\dots$	$a'_{l,m+1} \dots a'_{lj} \dots a'_{l,n}$	b'_l
$\dots$		$\dots$	$\dots$	$\dots$
c_m	x_m	$0 \dots 1$	$a'_{m,m+1} \dots a'_{m,j} \dots a'_{m,n}$	b'_m
			$\sigma_{m+1} \dots \sigma_j \dots \sigma_n$	

取 $X_N = [0, \dots, \theta_j, \dots, 0]'$，则 $Z = C_B B^{-1} b + \sigma_j \theta_j$，如果 θ_j 可以取正值，则目前的解不是最优。考虑 θ_j 需要满足的条件：

$$X_B = B^{-1} b - B^{-1} N X_N \tag{8.15}$$

$$= b' - (P'_{m+1}, \dots, P'_n) X_N \tag{8.16}$$

$$= [b'_1 - \theta_j a'_{1,j}, \dots, b'_m - \theta_j a'_{m,j}] \geq 0 \tag{8.17}$$

$$\tag{8.18}$$

从而：

1. 所有的 $a_{i,j} \geq 0, i = 1\ m$，则 θ_j 无限制，目标值可以无限增加或减少，无界解
2. 存在 $a_{i,j} > 0$，则 $0 \leq \theta_j \leq min\{\frac{b'_i}{a'_{i,j}}, b'_i > 0, a'_{i,j} > 0\} = \frac{b'_l}{a'_{l,j}}$

即需要选择 x_l 出基，x_j 进基，此时 $x_l = 0$，相应的目标值上浮 $\theta_j \sigma_j$。如果 $b_i = 0$，此时换基可以进行，但是得到的解是一致的。这里取到等号的原因是基中增加新变量 j，必须要从原基中消除一个变量，即使其值为 0，取等号是换基必要条件。相应的，如果 θ_j 取到小于 $\frac{b'_l}{a'_{l,j}}$，目标值虽然增加，但并不是基解，用到的变量为 $m+1$ 个。如果 θ_j 取到大于 $\frac{b'_l}{a'_{l,j}}$，则某个基变量必然小于 0，因此不是可行解。

在换基的过程中实际上只涉及一次进基变量的选择，一旦选定了进基变量，出基变量也就由上述的规则确定。进基变量的选择方法：

1. 如果存在 $P'_j \leq 0$，选 x_j 进基，**无界解**
2. 不存在 $P'_j \leq 0$，$\sigma_j > 0$，选择 $max|\sigma_j|$ 进基，**可调整**
3. Blank 选择：$\sigma_j > 0$ 中选择最大的 j 进基，相同的 θ_j 选择最小的 j 出基，避免退化
4. 随意选择一个进基

现在可以统一在有初始基可行解的情况下完成单纯形法：

这里特别要注意按照判断流程进行，但是理解单纯形法的判断过程要从后向前看，可以取得无界解就不再换基，可以换基增加目标值就不取无穷解，直到得到三种情况中的一种。

现在剩下的问题是：初始基可行解如何取得？当然可以随意指定初始基，但是可能出现不可行的情况，也就无法保证一定取得基可行解。利用换基的思想——从一个基可行解转换到

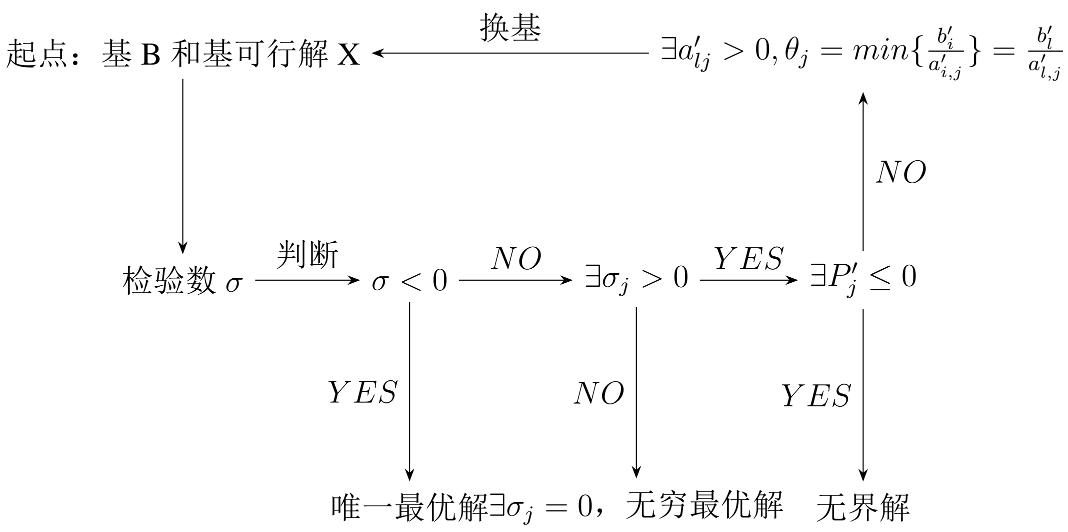

图 8.3: 单纯形法步骤演示

另一个基可行解，可以采用两阶段法和大 M 法构造基可行解，然后将人工变量逐个出基，得到原问题的基可行解。出基的办法就是使人工变量在最优解中取值为 0，这是容易做到的。

- 大 M 法：

$$Max : Z = \sum_{j=1}^{n} c_j x_j - \sum_{j=1}^{m} M y_j \tag{8.19}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{ij} x_j + y_i \leq (=, \geq) b_i, i = 1, \ldots, m([A, E][X, Y]' = b) \\ x_1, x_2, \ldots, x_n, y_1, \ldots, y_m \geq 0 \end{cases} \tag{8.20}$$

逐个将 y_j 出基，如果取得最优解时仍然存在 y_j 未出基，则原问题无解。

- 两阶段法：

1. 由初始基逐步构造可行基，

$$Max : Z = -\sum_{j=1}^{m} M y_j \tag{8.21}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{i,j} x_j + y_i = b_i, i = 1 - m([A, E][X, Y]' = b) \\ x_1, x_2, \ldots, x_n, y_1, \ldots, y_m \geq 0 \end{cases} \tag{8.22}$$

2. 求解原问题，

$$Max : Z = -\sum_{j=1}^{n} c_j x_j \tag{8.23}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{i,j} x_j = b_i, i = 1, \ldots, m \\ x_1, x_2, \ldots, x_n \geq 0 \end{cases} \tag{8.24}$$

8.2 对偶问题

8.2.1 建立对偶问题

对偶问题换个角度分析资源的配置问题。LP 模型是在给定资源和组合方式和产品的价值系数以后求解组合价值最大的方案，对偶问题则尝试给这些资源定价，即每种资源应该以多少的价格在市场上出售，这里需要考虑两点：（1）出售资源的收益要高于自己生产产品。（2）优化目标是最小化资源的总价值，这是由于市场是完全竞争的，产品的价格应该等于其边际价值。这样，对称形式的 LP 问题和对偶问题为：

$$L.P.MAX : Z = CX \tag{8.25}$$

$$s.t. \begin{cases} AX \leq b \\ X, b \geq 0 \end{cases} \tag{8.26}$$

$$L.P.'MIN : W = Yb \tag{8.27}$$

$$s.t. \begin{cases} YA \geq C \\ Y \geq 0 \end{cases} \tag{8.28}$$

$$OR \quad L.P.'Min : W = b_1y_1 + b_2y_2 + \ldots + b_ny_n \tag{8.29}$$

$$s.t. \begin{cases} a_{11}y_1 + \ldots + a_{m1}y_n \geq c_1 \\ a_{12}y_1 + \ldots + a_{m2}y_n \geq c_2 \\ \ldots \\ a_{1n}y_1 + \ldots + a_{mn}x_n \geq c_n \\ y_1, y_2, \ldots, y_m \geq 0 \end{cases} \tag{8.30}$$

在对偶单纯形表中表示如下：

表 8.3: 对偶单纯形表

<table>
<tr><td></td><td>C_B</td><td>C_N</td><td>C_r</td><td></td><td></td></tr>
<tr><td>C_BX_B</td><td>B</td><td>N</td><td>b</td><td>b</td><td>Y</td></tr>
<tr><td>C_BX_B</td><td>E</td><td>N'</td><td>b'</td><td>$B^{-1}b$</td><td>Y'</td></tr>
<tr><td></td><td></td><td>$C_N - C_BB^{-1}N'$</td><td>$-C_BB^{-1}$</td><td></td><td></td></tr>
</table>

从对称形式的原问题转换到对偶问题需要注意（L.P. 一般指最大值问题，L.P.’ 一般指最小值问题）：（1）优化目标从最大值转换为最小值，原问题的价值系数转换为约束系数，原问题的约束系数转换为新问题的价值系数；（2）原问题的约束方程符号变号为对偶问题的变量符号，原问题的变量符号和对偶问题的方程符号一致。而将一个一般形式的线性规划问题转换为对称形式的线性规划问题，需要处理：

- 约束方程：（1）$\geq$，方程同乘以-1 变号（2）$=$，方程分解为两个方程 $\geq, \leq$
- 变量：（1）$x \leq 0$，令 $x = -x', x' \geq 0$（2）x 无约束，令 $x = x' - x'', x', x'' \geq 0$

$$\text{L.P.:MAX} \xleftrightarrow{\quad\text{对偶}\quad} \text{L.P.':MIN}$$

$$\text{价值系数} \xleftrightarrow{\quad\text{互换}\quad} \text{约束常量}$$

$$\text{L.P 约束条件} \xleftrightarrow{\quad\text{异号}\quad} \text{L.P' 变量}$$

$$\text{L.P 变量} \xleftrightarrow{\quad\text{同号}\quad} \text{L.P' 约束条件}$$

$$\text{等号} \xleftrightarrow{\text{方程、变量的符号转换}} \text{无约束}$$

图 8.4: 原问题和对偶问题

将一般形式的问题转换为它的对偶问题，可以先写成对称问题，再转换为对偶问题。也可以一步到位，特别需要注意的是对符号的处理，也就是：

8.2.2 对偶理论

已经了解了对偶问题的经济含义和基本形式，现在进一步从问题层次、可行域、目标值、最优解、可行解约束条件、迭代过程等角度分析。

定理 8.4

对偶问题的对偶问题是原问题，即 (L.P.')' = L.P.。

证明定理8.2.2可以将对偶问题转化为最大值问题，再写出对偶问题，得到的最小值问题和原问题一致。这一定理说明适用于原问题的结论也适用于对偶问题。

定理 8.5

若 L.P. 问题可行域有界，则一定有最优解。

证明 可行域可能有三种情况：不存在、有界、无界。可行域有界则是闭合的凸集，那么就只有有限的顶点，因此在某个顶点上可以取得最优解。

定理 8.6

若 L.P. 目标值有界，则存在最优解。

定理8.2.2意味着最大化问题目标值有上界则存在最优解，最小化问题目标值有下界则存在最优解。

定理 8.7

若 L.P. 有最优解 X^*，则 L.P' 有可行解 $Y = C_B B^{-1}$，且 $Z^* = C_B B^{-1} b = Yb = W \geq W^*$。

证明 这里将证明原问题标准型的非基变量和松弛变量的检验数的相反数是对偶问题标准型的基解，当原问题取最优解时，对偶问题取到可行解。原问题标准型的单纯形表和运算过程如下：L.P. 问题的检验数 $\sigma_N = C_N - C_B N', \sigma_r = -C_B B^{-1}$。对于对偶问题的标准型，有

表 8.4: 原问题对偶单纯形表

	C_B	C_N	C_r	
C_BX_B	B	N	b	b
C_BX_B	E	N'	b'	$B^{-1}b$
		$C_N - C_BB^{-1}N'$	$-C_BB^{-1}$	-z = -w

$[Y, Y^r][A, E]' = C, YA + Y^r = C, Y[B, N] + [Y^r_B, Y^r_N] = [C_B, C_N]$，即:

$$YB = Y^r_B + C_B, YN = Y^r_N + C_N \tag{8.31}$$

$$Set, Y^r_B = 0, \tag{8.32}$$

$$So, Y = C_BB^{-1}, Y^r_N = C_N - C_BB^{-1}N \tag{8.33}$$

这里 Y^r 是剩余变量,Y, Y^r_N 是基变量,这里 B 对应的变量是非基变量,定义 $B^{-1}b$ 为单纯形算子。可以发现,这里对偶问题的解和原问题的检验数是相反数:$\sigma = [0, C_N - C_BB^{-1}N, -C_Bb^{-1}], Y = [0, C_Bb^{-1}, C_BB^{-1}N - C_N]$。而且，依照检验数的计算方法在最后一列增加检验数得到目标值的相反数 $-z = -C_BB^{-1}b$，特别需要说明的是：当检验数小于等于 **0** 时，对偶问题取可行解，可以取 **w**；当 b 大于等于 **0** 时原问题取可行解，可以取 **z**；当以上两个条件只满足一个时，原问题和对偶问题一个对应可行解一个对应非可行解；当以上两个条件同时满足时，原问题和对偶问题取相同的目标值。这里的解都是基解。因此可以通过 b, σ 同时判断原问题和对偶问题的解的情况。

定理 8.8

弱对偶定理：对于 L.P 和 L.P' 的任意可行解 X, Y，有 $CX \leq Yb$。

弱对偶定理意味着如果 L.P 和 L.P' 均存在可行解，则（1）L.P 问题的任意可行解的目标值是 L.P.' 问题的目标值的下限，L.P' 问题的任意可行解的目标值是 L.P. 问题的目标值的上限。（2）L.P 和 L.P' 都有最优解（3）由定理8.2.2和8.2.2，$Z^* \geq W^*$ and $Z^* \geq W^*$，L.P 和 L.P' 的最优值一致。（4）如果 $CX = Yb$，则由 $CX \leq CX^* \leq Y^*b \leq Yb$ 知 X, Y 为最优解且最优值一致。如果考虑所有可能的解的情况，可以将 L.P. 和 L.P' 的解的关系概括如下：

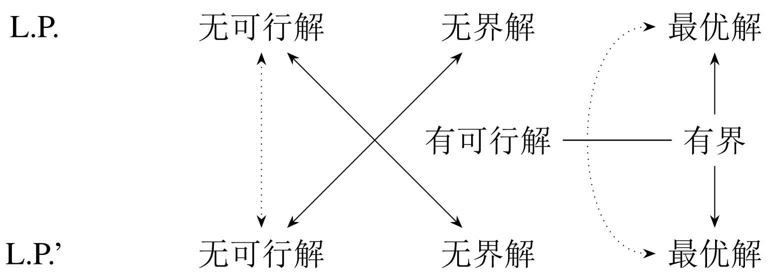

图 8.5: L.P. 和 L.P.' 解的关系

定理 8.9

互补松弛定理（松紧定理）：对于最优解 X^*, Y^*，有对应任意的 i，$y^*_i(\sum_{j=1}^n a_{i,j}x^*_j - b_i) = 0, i = 1, \ldots, m$ 成立；对应任意的 j，$x^*_j(\sum_{i=1}^m a_{i,j}y^*_i - c_j) = 0, j = 1, \ldots, n$ 成立。

松紧定理意味着如果 $y_i^* > 0$，则 $\sum_{j=1}^{n} a_{i,j}x_j^* - b_i = x_i^{r*} = 0$，即对偶问题中资源的价格为正，说明资源应当在原问题中完全分配，资源是稀缺的。反过来如果 $\sum_{j=1}^{n} a_{i,j}x_j^* - b_i > 0$，则 $y_i^* = 0$，即如果资源没有完全分配，资源是充足的，市场定价为 0。同样，如果 $x_j^* > 0$，则 $\sum_{i=1}^{m} a_{i,j}y_i^* - c_j = y_i^{r*} = 0$，即如果生产产品 j，则说明产品的价值和生产成本一致。反过来，如果 $\sum_{i=1}^{m} a_{i,j}y_i^* - c_j > 0$，则 $x_j = 0$，即生产成本大于产品价值，不生产而出售资源是合适的。松紧定理可以记为：前非零，后取等。一非零，二取等。进一步的，在对称形式的 L.P. 中加入松弛变量作为人工变量，可以得到松弛变量的检验数是对偶问题相同基的解的相反数。

最后需要了解的是换基过程会如何影响原问题和对偶问题的解，注意这里非基变量和松弛变量的检验数的相反数就是对偶问题的同基解，并且换基时总是在原问题或者对偶问题的可行域中移动顶点，需要保证原问题可行（单纯形法）或者对偶问题可行（对偶单纯形法），这里我们同时关注 σ, b。

表 8.5: 对偶单纯形表换基

		c_l	c_k	c_j	
c_1	x_1	0	$a_{1,k}$	$a_{1,j}$	b_1
$\dots$		$\dots$	$\dots$	$\dots$	$\dots$
c_l	x_l	1	$a_{l,k}$	$a_{l,j}$	b_l
$\dots$		$\dots$	$\dots$	$\dots$	$\dots$
c_m	x_m	0	$a_{m,k}$	$a_{m,j}$	b_m
		0	σ_k	σ_j	$-z = -w$
c_1	x_1	$-\frac{a_{1,j}}{a_{l,j}}$	$a_{1,k} - \frac{a_{l,k}}{a_{l,j}}a_{1,j}$	0	$b_1 - \frac{b_l}{a_{l,j}}a_{1,j}$
$\dots$		$\dots$	$\dots$	$\dots$	$\dots$
c_j	x_j	$-\frac{1}{a_{l,j}}$	$a_{l,k} - \frac{a_{l,k}}{a_{l,j}}$	1	$\frac{b_j}{a_{l,j}}$
$\dots$		$\dots$	$\dots$	$\dots$	$\dots$
c_m	x_m	0	$a_{m,k} - \frac{a_{l,k}}{a_{l,j}}a_{m,j}$	0	$b_m - \frac{b_l}{a_{l,j}}a_{m,j}$
		$-\frac{\sigma_j}{a_{l,j}}$	$\sigma_k - \frac{a_{l,k}}{a_{l,j}}\sigma_j$	0	$-z + \frac{b_l}{a_{l,j}}\sigma_j$

在表8.5中，只列出了进基 j，出基 l 和其他变量 k。换基会同事影响 σ, b，并且在单纯形法中刚出基的变量 l，其检验数为会小于 0，不可能马上进基。在单纯形法中换基需要保证 b 为正，原问题的解可行；相应的，在对偶单纯形法中需要保证 $\sigma \leq 0$，即对偶问题的解可行。此时依然是先选出基变量，后选进基变量。即：

$$[Y, Y^r][A, -I^r] = C, YA - Y^r = C, Y[B, N] - [Y_B^r, Y_N^r] = [C_B, C_N] \tag{8.34}$$

$$Y = C_B B^{-1} + Y_B^r B^{-1} \tag{8.35}$$

$$W = Yb = C_B B^{-1} b + Y_B^r B^{-1} b = C_B B^{-1} b + Y_B^r b' \tag{8.36}$$

$$W = C_B B^{-1} b + (y_1^r b_1' + \ldots + y_m^r b_m') \tag{8.37}$$

这里，$y_1^r, \ldots, y_m$ 是对偶问题非基变量，需要从中选择一个进基，这里只需要要求 $b_l < 0$ 即可，但是通常都会选择最小的 b_l 以有可能减少迭代次数。即选择 $b_l = min b_i, i = 1, \ldots, m$ 对应的变量 x_l 出基。并且，上面的运算意味着换基可以降低目标值（$b_l < 0$，否则已经是最优

解)。选择进基变量的条件即保证对偶问题在可行域即可，即：

$$\sigma_k - \frac{a_{l,k}}{a_{l,j}}\sigma_j \leq 0, k = 1, \ldots, m+n \tag{8.38}$$

$$Set, \phi_k = \frac{\sigma_j}{a_{l,j}} \tag{8.39}$$

$$So, \phi_j = min\{\frac{\sigma_j}{a_{l,j}}, a_{l,j} < 0\} > 0 \tag{8.40}$$

即选择最小的 $\frac{\sigma_j}{a_{l,j}}, a_{l,j} < 0$ 进基即可。

到现在，可以阐述解线性规划问题的对偶单纯形法的整体框架。

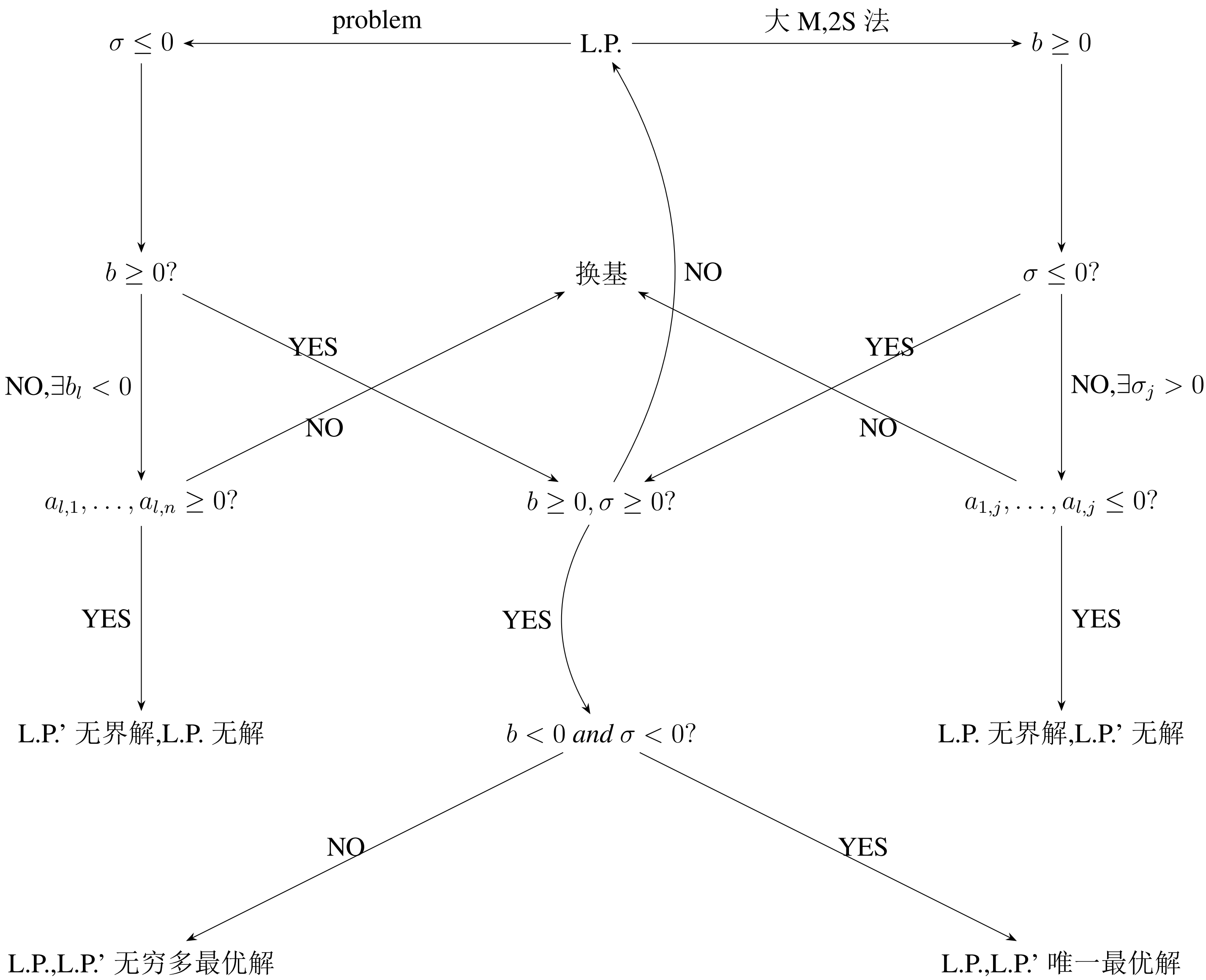

图 8.6: 对偶单纯性法运算

需要说明的是，在对偶单纯形法的运算演示中，σ 均是非基变量的检验数，并且假设线性规划问题是有可行解的，判断是否存在可行解可以用大 M 方法。

8.2.3 对偶问题经济意义

$$L.P, MAX : Z = CX, s.t. : AX \leq b, X, b \geq 0 \tag{8.41}$$

$$L.P', MIN : W = Yb, s.t. : YA \geq C, Y \geq 0 \tag{8.42}$$

原问题是资源利用问题，对偶问题是资源定价问题，二者同时取得最优解:$Z^* = CX^* = Y^*b = W^*$，Y^* 可以理解为资源的影子价格，即边际资源的价格。

- $CX^* = Y^*b$，Y^* 是决定于拥有量和资源利用能力的价格
- $Y^*A \geq C$，资源是稀缺的，取到等号的资源可能是充足的
- $MIN : W = Yb$，资源是非垄断自由竞价的
- $CX \leq Yb$ 是一种合作价格，购买和出售双方均有剩余（生产成本小于买价）
- $y_i^* x_i^{r*} = 0$，Y^* 是资源充分利用的价格
- $Z^* = W^*, \frac{\partial Z^*}{\partial b_i} = y_i^*$，按边际成本定价

8.2.4 灵敏度分析

灵敏度分析是在其他参数不变的情况下，分析某一参数的变化如何影响最终的结果。线性规划的参数包括 C, b, A，分析的目标是最优解和最优值是否变化以及变化的范围，本质上是找出最优解的基，核心是迭代。

8.2.4.1 C 的变化

设 $c_j' = c_j + \Delta c_j$，在原有最优解上 c_j 变化不影响 B，因此解不变，但可能不再是最优。c_j 变化的唯一影响就是变量 x_j 的检验数，即:

$$\sigma' = C' - C_B' B^{-1} A = C + \Delta C - (C_B + \Delta C_B) B^{-1} A = C + \Delta C - \Delta C_B B^{-1} A \tag{8.43}$$

需要判断两个问题：（1）c_j 变化以后最优解是否改变？（2）c_j 在什么范围内变化最优解不会改变？

8.2.4.2 b 的变化

设 $b' = b + \delta b$，如果 $b' \geq 0$，则就是最优解，基不变。否则需要对偶单纯形法换基，即基变化，最优解会改变。

8.2.4.3 A 的变化

1. 增加一个变量：增加的变量在现有解中是非基变量，只要其检验数非正则基不变，最优解不变，即需要 $\sigma_k = c_k - C_B B^{-1} P_k$ 为非正，否则需要换基，最优解变化

2. 改变 $a_{i,j}$ 的值：如果 x_j 是非基变量，只需要判断检验数是否非正，非正则最优解不变，否则最优解变化；如果 x_j 是基变量，最优解改变
3. 增加一个约束条件：可行域将变小，将原最优解带入新的约束方程中检验，成立则最优解不变，否则最优解改变

总结：灵敏度分析的思考路径（1）B 是否改变（若改变，重算）（2）可行解是否改变（若改变，重算）（3）最优解是否改变，判断检验数是否非正。

8.3 运输问题

8.3.1 建立运输问题模型

运输问题：将产品从产地运往销地，求最低成本的运输方案。从最简单的运输问题开始分析，即将同一种产品从 m 个产地运往 n 个销地，运费为 $c_{i,j}$，途中无转运、无限制及其他约束并且产销平衡。运输问题的线性规划形式如下：

$$MIN : z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} x_{i,j} \tag{8.44}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} x_{i,j} = a_i, i = 1, \ldots, m \\ \sum_{i=1}^{m} x_{i,j} = b_j, j = 1, \ldots, n \\ x_{i,j} \geq 0 \end{cases} \tag{8.45}$$

对应的运输作业表如下：

表 8.6: 运输作业表

	B_1	$\ldots$	B_n	
A_1	$x_{1,1}^{c_{1,1}}$	$\ldots$	$x_{1,n}^{c_{1,n}}$	a_1
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
A_m	$x_{m,1}^{c_{m,1}}$	$\ldots$	$x_{m,n}^{c_{m,n}}$	a_m
	b_1	$\ldots$	b_n	$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$

运输问题的单纯形表如下：

表 8.7: 运输问题单纯形表

	$c_{1,1} \ldots c_{1,n}$	$\ldots$	$c_{m,1} \ldots c_{m,n}$	
	$1, \ldots, 1$	$\ldots$	$0, \ldots, 0$	a_1
	$\ldots$	$\ldots$	$\ldots$	$\ldots$
	$0, \ldots, 0$	$\ldots$	$1, \ldots, 1$	a_m
	$1, \ldots, 0$	$1, \ldots, 0$	$1, \ldots, 0$	b_1
	$\ldots$	$\ldots$	$\ldots$	$\ldots$
	$0, \ldots, 1$	$0, \ldots, 1$	$0, \ldots, 1$	b_n

8.3.2 运输作业表解法

求解运输问题当然可以采用单纯形法迭代出最优解，但是针对运输问题有更为精炼的方法。我们依然沿着线性规划的思路，考虑解的存在性、确定基、检验是否最优，接下来逐步分析这些问题。

定理 8.10 解的存在性：运输问题一定存在可行解，并且存在最优解，可能有多个最优解。

证明 设 $x_{i,j} = \frac{a_i b_j}{Q}, Q = \sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$，则 $x_{i,j}$ 是可行解，并且运输问题有下界 0（$c_{i,j} > 0$，否则没有经济含义），则运输问题存在最优解。根据单纯形法，只要某个检验数为 0，就可能存在多个最优解。

定义 8.1 闭回路：由横线、竖线组成的回路。数学表达为由 $x_{i1,j1}, x_{i1,j2}, x_{j2,i2}, \ldots, x_{in,jn}, x_{in,j1}$ 构成的回路，这里所有的 x 均不同。

由闭回路的定义可以发现：（1）闭回路上的点一定是偶数个（2）闭回路上任一点都有对应的横点和竖点。并定义在链上没有对应横点或者竖点的点是**孤立点**。

定理 8.11 闭回路上没有孤立点；链上没有闭回路则一定有孤立点。

证明 反证法：如果闭回路上没有孤立点，则从任何一点出发，一次经过链上的其他点，一定可以回到曾经过的点，即出现闭回路。

定理 8.12 链上存在闭回路等价于对应向量组线性相关；链上不存在闭回路等价于对应向量组线性无关；

证明 (1) 存在闭回路，线性相关。设一条链上的闭回路为 $x_{i1,j1}, x_{i1,j2}, x_{j2,i2}, \ldots, x_{in,jn}, x_{in,j1}$，则这条闭回路对应的向量组为 $P_{i1,j1}, P_{i1,j2}, P_{i2,j2}, \ldots, P_{in,jn}, P_{in,j1}$，并且根据单纯形法，$P_{i,j} = e_i + e_j$。设 $K = [1, -1, 1, -1, \ldots, 1, -1]$，则 $KP' = e_{i1} + e_{j1} - e_{i1} - e_{j2} + e_{i2} + e_{j2} + \ldots + e_{in} + e_{jn} - e_{in} - e_{j1} = 0$ (2) 线性相关，存在闭回路。设 $\exists K \neq 0, KP' = 0$，如果不存在闭回路，则存在孤立点，它没有对应的行或列匹配点，不妨设为 P_1，则孤立点对应的向量系数为 0，基 $k_1 = 0$。去掉向量 P_1，剩下的向量仍是线性相关，此时不存在闭回路则可以去掉向量 P_2，不断进行判断，直到 $K = 0$，与线性相关的假设违背。因此与原定理成立。

定理 8.13 运输问题中，$r(A) = m + n - 1$。

运输矩阵的秩可以通过初等行变化计算，这里不再列出。直观地看，对 mn 个变量有 $m+n$ 个运输方程和一个平衡方程，平衡方程决定了行变换可以消掉系数矩阵 A 的一行，因此秩就

是 $m+n-1$。同时这决定了基变量只有 $m+n-1$ 个，而不是 $m+n$ 个。下面分析换基，换基依然遵循两个要求：（1）减少目标值（2）新的基解可行。在运输表中，首先需要考虑的是换基如何发生，即选定一个入基以后如何寻找合适的出基。

在运输作业表中，根据定理8.3.2，基是满秩矩阵，即向量组线性不相关，即表中的点无法构成闭回路。而从非基中选择一个变量进基，加上原来的变量，向量组中一共有 $m+n$ 个向量，而秩不会增加，因此 $[P_{new}, B]$ 是线性相关向量组，因此在运输作业表中一定可以构成闭回路，并且这个闭回路是唯一的。在闭回路中考虑换基和求目标值的变化就相对容易很多，这在单纯形表中去计算则显得相对困难。即考虑如下的闭回路：

$$x_{i1,j1}, x_{i1,j2}, x_{j2,i2}, \ldots, x_{in,jn}, x_{in,j1} \tag{8.46}$$

其中设进基变量 $x_{i1,j1}$ 为 0，其他变量大于等于 0。现在要增加进基变量的值，同时要保持作业表平衡，其他变量需要依次变动，变化如下：

$$OLD: 0, x_{i1,j2}, x_{j2,i2}, \ldots, x_{in,jn}, x_{in,j1} \tag{8.47}$$

$$NEW: \Delta, x_{i1,j2}-\Delta, x_{j2,i2}+\Delta, \ldots, x_{i,j}+(-1)^{k+1}\Delta, \ldots, x_{in,j1}-\Delta \tag{8.48}$$

这里的 k 是从进基点开始的计数，闭环的计数方向无关紧要。要保持新的解仍然是基解，那么这个闭环中必须仅其仅仅去掉一个点，即需要满足：

$$x_{i,j}+(-1)^{k+1}\Delta \geq 0 \tag{8.49}$$

$$\Delta = min\{x_{i,j}, k=2,4,...\} \tag{8.50}$$

新的进基变量必须取到最小的偶数点的值，这里不考虑奇数点是因为其一定大于等于 0。相应的，目标值的变化量为：

$$\Delta z = \Delta(x_{i1,j1} - x_{i1,j2} + x_{j2,i2} - \ldots + x_{in,jn}, -x_{in,j1}) \tag{8.51}$$

检验数可以设定为 $\sigma = \frac{z}{\Delta} = \sum(-1)^{k+1}c_{i,j}, x_{i,j} \in$ 闭合路。当 σ 大于 0，对应变量进基将降低目标值，可以调整；当 σ 小于 0，对应变量进基将增加目标值，应该维持现状；当 σ 等于 0，对应变量进基将不影响目标值，即存在无穷多解。还可能发生一种情况，就是 $\Delta=0$，进基的取值只能取 0，这意味着换基不换解，可以不再考虑包含值为 0 的基变量的闭回路，如果无法再换解，说明当前的解是唯一最优解。

现在了解了如何在现有的基解下检验最优和换基的方法，但是初始基解如何取得呢？俗话说好的开始是成功的一半，选择一个良好的基解可以降低检验和换基的工作量。初始基解的选择可以采用：西北角法、最小费用法和 Vogel 法。这里不再详细说明。现在可以将运输问题标准形式的解的步骤总结如下：

值得注意的是：（1）这里的初始点常常选择最小运价点（2）划线法先求出正值的变量数量小于 $m+n-1$ 并且线性无关，这里选出的即为基变量（3）划线法得到的正值变量数量不足，选 0 补足但不要形成闭环（4）从空格点出发和部分基变量组成的闭环是唯一存在的（5）调整需要保证目标有优化并且仍然是可行解。

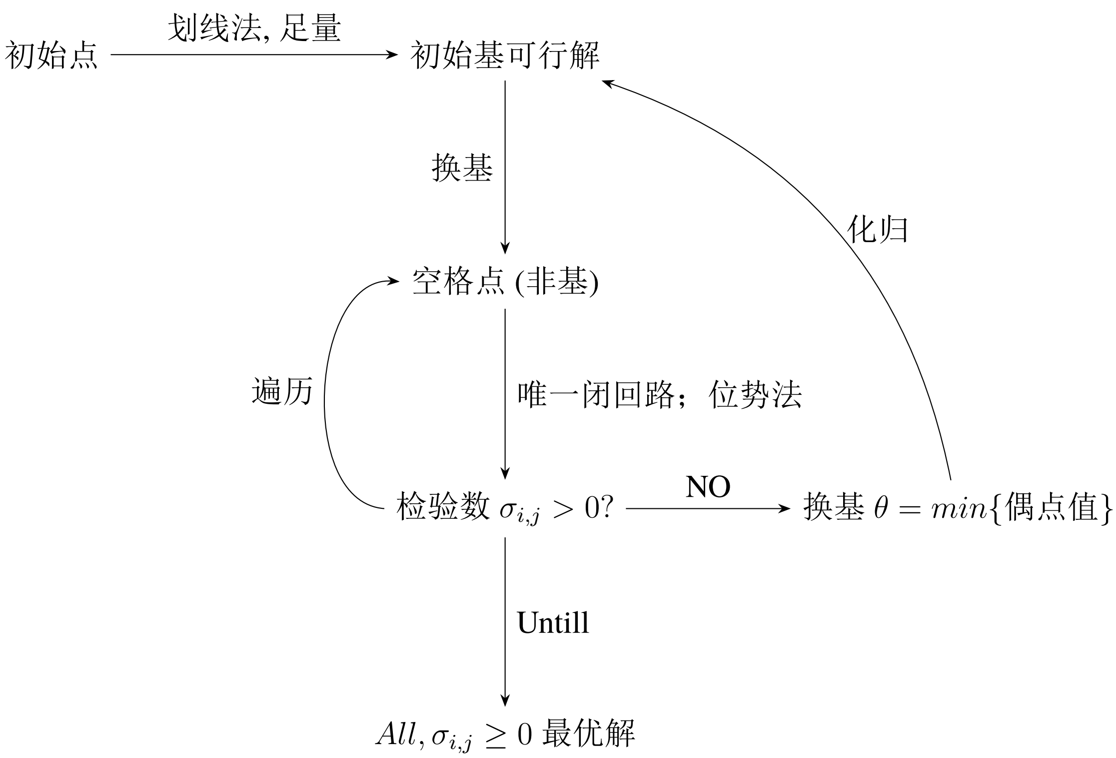

图 8.7: 运输问题解法

8.3.3 运输对偶解法

运输问题的对偶问题：

$$MAX\ :w = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \tag{8.52}$$

$$s.t. \begin{cases} u_i + v_j \leq c_{i,j}, i = 1, \dots, m, j = 1, \dots, n \\ u_i, v_j \text{无约束} \end{cases} \tag{8.53}$$

这里给出原问题的检验数为：

$$\sigma_{i,j} = c_{i,j} - C_B B^{-1} P_{i,j} \tag{8.54}$$

$$= c_{i,j} - (u_i + v_j) \tag{8.55}$$

回想一下对偶问题的解 $y = C_B B^{-1}$,这里的 $C_B B^{-1}$ 就是对应的可行解 $[u_1, \dots, u_m, v_1, \dots, v_n]$，而前面已经说明了 $P_{i,j}$ 就是第 $i, i+j$ 位取 1，其他元素为 0 的向量，可以看一下运输问题的单纯形表。当运输问题取最优解时，$\sigma_{i,j} \geq 0$，即 $c_{i,j} \geq (u_i + v_j)$，这对应于对偶问题的约束方程。根据松紧定理，如果原问题的解是非零，则对偶问题约束方程（实际 $m+n-1$ 个，平衡方程）取等，因此只需要随意确定一个 u_i 或 v_j 就可以取到所有的 u_i, v_j（$m+n$ 个）。因此，非基变量的检验数可以采用位势计算：

$$\sigma_{i,j} = c_{i,j} - (u_i + v_j) \tag{8.56}$$

位势表如下：

位势法实际上是在利用对偶问题求原问题的检验数，它在运输问题解法中的地位和闭回

表 8.8: 运输问题位势表

	$B_1\dots$	B_j	$\dots B_n$		
A_1	$c_{1,1}\dots$	$c_{1,j}$	$\dots c_{1,n}$	a_1	u_1
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
A_i	$c_{i,1}\dots$	$c_{i,j}$	$\dots c_{i,n}$	a_i	u_i
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	
A_m	$c_{m,1}\dots$	$c_{m,j}$	$\dots c_{m,n}$	a_m	u_m
	$b_1\dots$	b_j	$\dots b_n$		
	$v_1\dots$	v_j	v_n		

路一样。

8.3.4 运输问题的拓展

这里分析，在运输问题上增加约束条件将如何改变运输作业表和结果，这里考虑的如何将问题的化归为标准问题。

8.3.4.1 多个产品的规划

分别为每一种产品规划运输方案。

8.3.4.2 产销非平衡

增加虚拟的产地或者销地，即：

- $Q_a > Q_b$，产大于销，增加虚拟的销地，销量为 $Q_a - Q_b$
- $Q_a < Q_b$，产小于销，增加虚拟的产地，产量为 $Q_b - Q_a$

表 8.9: 虚拟产地

	$B_1\dots B_n$	
A_1	$c_{1,1}\dots c_{1,n}$	a_1
$\dots$	$\dots$	$\dots$
A_m	$c_{m,1}\dots c_{m,n}$	a_m
A_{m+1}	$0\dots 0$	$Q_b - Q_a$
	$b_1\dots b_n$	

8.3.4.3 可以转运

不可以转运、没有专门转运地和有专门转运地的示意图如下，其中 A_1, A_2 是产地，B_1, B_2, B_3 是销地，C_1, C_2 是转运地，虚线代表自己转给自己，这一想象在解决运输问题时非常有用。由于中途有转运（可以在产地和产地、销地和销地之间转），因此无法确定中途转运的运量，作业表无法取得，但是有了自己转给自己的虚拟过程，可以将中途转运的量设置为绝对大 (如

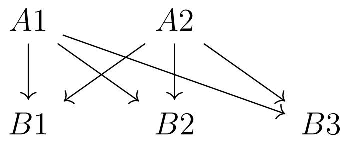

图 8.8: 不可转运

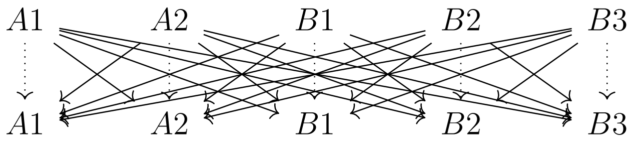

图 8.9: 无专门转运地

Q_a, Q_b)，而多出的量由自己转给自己来消耗，因此运输作业表如下，其中运输作业表对角线元素取 0，有专门转运地可以在产地和销地中转，这里的示意图简化了这些中转。

8.3.4.4 运输工具限制

限制运输工具相当于只能在转运地中转，运输作业表如下，其中的不能转运的部分用足够大的罚数表现。

8.3.4.5 路线限制

A_i 无法向 B_j 运输，只需要设置 $c_{i,j} = M, M$ 足够大即可。

8.3.4.6 运量限制

如果 $x_{i,j} \geq k$，只需先做计划即可，即 $a_i' = a_i - k, b_j' = b_j - k$；如果 $x_{i,j} \leq k$，将产地 A_i 分为两部分，一部分 A_i' 的产量为 k，一部分 A_i 的产量为 $a_i - k$，只需要设置 A_i 到 B_j 的运输成本 $c_{i,j}$ 为足够大的 M 即可。

8.3.4.7 配比、规格限制

先完成配比工作。

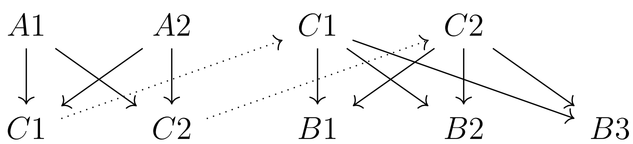

图 8.10: 有专门转运地

表 8.10: 无专门转运地

	$A_1 \dots A_m$	$B_1 \dots B_n$	
A_1	$0 \dots c_{1,m}$	$c_{1,m+1} \dots c_{1,m+n}$	$a_1 + Q$
$\dots$	$\dots$	$\dots$	$\dots$
A_m	$c_{m,m} \dots 0$	$c_{m,m+1} \dots c_{m,m+n}$	$a_m + Q$
B_1	$c_{m+1,1} \dots c_{m+1,m}$	$0 \dots c_{m+1,m+n}$	Q
$\dots$	$\dots$	$\dots$	$\dots$
B_n	$c_{m+n,1} \dots c_{m+n,m}$	$c_{m+n,m+1} \dots 0$	Q
	$Q \dots Q$	$b_1 + Q \dots b_n + Q$	

表 8.11: 有专门转运地

	$A_1 \dots A_m$	$T_1 \dots T_k$	$B_1 \dots B_n$	
A_1	$0 \dots c_{1,m}$	$c_{1,m+1} \dots c_{1,m+k}$	$c_{1,m+k+1} \dots c_{1,m+k+n}$	$a_1 + Q$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
A_m	$c_{m,m} \dots 0$	$c_{m,m+1} \dots c_{m,m+k}$	$c_{m,m+k+1} \dots c_{m,m+k+n}$	$a_m + Q$
T_1	$c_{m+1,1} \dots c_{m+1,m}$	$0 \dots c_{m+1,m+k}$	$c_{m+1,m+k+1} \dots c_{m+1,m+k+n}$	Q
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
T_m	$c_{m+k,m} \dots c_{m+k,m}$	$c_{m+k,m+1} \dots 0$	$c_{m+k,m+k+1} \dots c_{m+k,m+k+n}$	Q
B_1	$c_{m+l+1,1} \dots c_{m+k+1,m}$	$c_{m+k+1,m+1} \dots c_{m+k1,m+k}$	$0 \dots c_{m+k+1,m+k+n}$	Q
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
B_n	$c_{m+k+n,1} \dots c_{m+k+n,m}$	$c_{m+k+n,m+1} \dots c_{m+k+n,m+k}$	$c_{m+k+n,m+k+1} \dots 0$	Q
	$Q \dots Q$	$Q \dots Q$	$b_1 + Q \dots b_n + Q$	

表 8.12: 运输工具限制

	$T_1 \dots T_k$	$B_1 \dots B_n$	
A_1	$0 \dots c_{1,k}$	$M \dots M$	a_1
$\dots$	$\dots$	$\dots$	$\dots$
A_m	$c_{m,1} \dots 0$	$M \dots M$	a_m
T_1	$0 \dots M$	$c_{m+1,k+1} \dots c_{m+1,m+k}$	Q
$\dots$	$\dots$	$\dots$	$\dots$
T_k	$M \dots 0$	$c_{m+k,k+1} \dots c_{m+k,m+k}$	Q
	$Q \dots Q$	$b_1 \dots b_n$	

8.3.4.8 产量限制

产量限制问题可以从在原有的基础上某个产地的产量发生变化来理解。第一种限制是 $a_{1,new} >\geq a_1$，这又分为两种情况，即原来的产销力量对比。如果 $Q_{a_1} \geq Q_b$，则 $Q_{a_1,new} \geq Q_b$，但是这里未知 $a_{1,new}$，因此需要采用虚拟销地来消除这个未知量的影响，运输作业表如下：

表 8.13: 产量限制:$a_{1,new} > a_1, Q_{a_1} \geq Q_b$

	$B_1 \dots B_n$	B_{n+1}	
A_1	$c_{1,1} \dots c_{1,n}$	0	$a_1 + Q_{a_1}$
A_2	$c_{2,1} \dots c_{2,n}$	M	a_2
$\dots$	$\dots$	$\dots$	$\dots$
A_m	$c_{m,1} \dots c_{m,n}$	M	a_m
	$b_1 \dots b_n$	Q_{a_1}	

如果 $Q_{a_1} < Q_b$，此时的产销关系并不清楚，因此需要增加虚拟产地和销地达到平衡，运输作业表如下：

表 8.14: 产量限制:$a_{1,new} > a_1, Q_{a_1} < Q_b$

	$B_1 \dots B_n$	B_{n+1}	
A_1	$c_{1,1} \dots c_{1,n}$	0	$a_1 + Q_b$
A_2	$c_{2,1} \dots c_{2,n}$	M	a_2
$\dots$	$\dots$	$\dots$	$\dots$
A_m	$c_{m,1} \dots c_{m,n}$	M	a_m
A_{m+1}	$0 \dots 0$	M	$Q_b - Q_{a_1}$
	$b_1 \dots b_n$	Q_b	

第二种限制是 $a_{1,new} < a_1$，这又分为两种情况，即原来的产销力量对比。如果 $Q_{a_1} \geq Q_b$，则运输作业表如下：

表 8.15: 产量限制:$a_{1,new} < a_1, Q_{a_1} \geq Q_b$

	$B_1 \dots B_n$	B_{n+1}	
A_1	$c_{1,1} \dots c_{1,n}$	0	a_1
A_2	$c_{2,1} \dots c_{2,n}$	M	a_2
$\dots$	$\dots$	$\dots$	$\dots$
A_m	$c_{m,1} \dots c_{m,n}$	M	a_m
	$b_1 \dots b_n$	$Q_{a_1} - Q_b$	

如果 $Q_{a_1} < Q_b$，则产一定小于销，认为无解。

8.4 整数规划

8.4.1 割平面法

整数规划问题是在线性规划问题中加入整数条件，原来的连续可行解退化为离散的可行解。表示如下 (IP):

$$MAX : Z = CX \tag{8.57}$$

$$s.t. \begin{cases} AX = b \\ X \in N^+ \end{cases} \tag{8.58}$$

这里主要分析松弛问题的整数规划，即单纯性表为：

表 8.16: 松弛问题单纯形表

	$x_1 \dots x_m$	$x_{m+1} \dots x_n$	
x_1	$1 \dots 0$	$a_{1,m+1} \dots a_{1,n}$	b_1
$\dots$	$\dots$	$\dots$	$\dots$
x_m	$0 \dots 1$	$a_{m,m+1} \dots a_{,,n}$	b_n

对于第 i 个方程有:

$$x_i + \sum_{j=m+1}^{n} a_{i,j} x_j = b_i \tag{8.59}$$

$$Apart, a_{i,j} = [a_{i,j}] + f_{i,j}, b_i = [b_i] + f_i \tag{8.60}$$

$$So, x_i + \sum_{j=m+1}^{n} [a_{i,j}] x_j - [b_i] = f_i - \sum_{j=m+1}^{n} f_{i,j} x_j \tag{8.61}$$

这个方程的左边是整数部分，右边是小数部分，并且 $f_i, f_{i,j}$ 非负。从左边来看，左边部分的值是整数，从右边的小数部分来看，

$$f_i - \sum_{j=m+1}^{n} f_{i,j} x_j \leq f_i < 1 \tag{8.62}$$

即小数部分小于 1 并且结果是整数，因此小数部分是小于等于 0，即：

$$-\sum_{j=m+1}^{n} f_{i,j} x_j \leq -f_i \tag{8.63}$$

这个方程就**割平面方程**。在松弛问题的单纯形表中加入这个方程即可以割掉非整数的最优解。具体操作是取整、写余、加负。同样的，在单纯形表中带入整数解条件等价。

8.4.2 分枝定界法

分枝定界法的思路是如果当前解不是整数解，则将当前的可行域从该非整变量的上下整数界两端分开，分别计算两个小的可行域的最优解，逐步得到最优解。如图，分枝定界法：

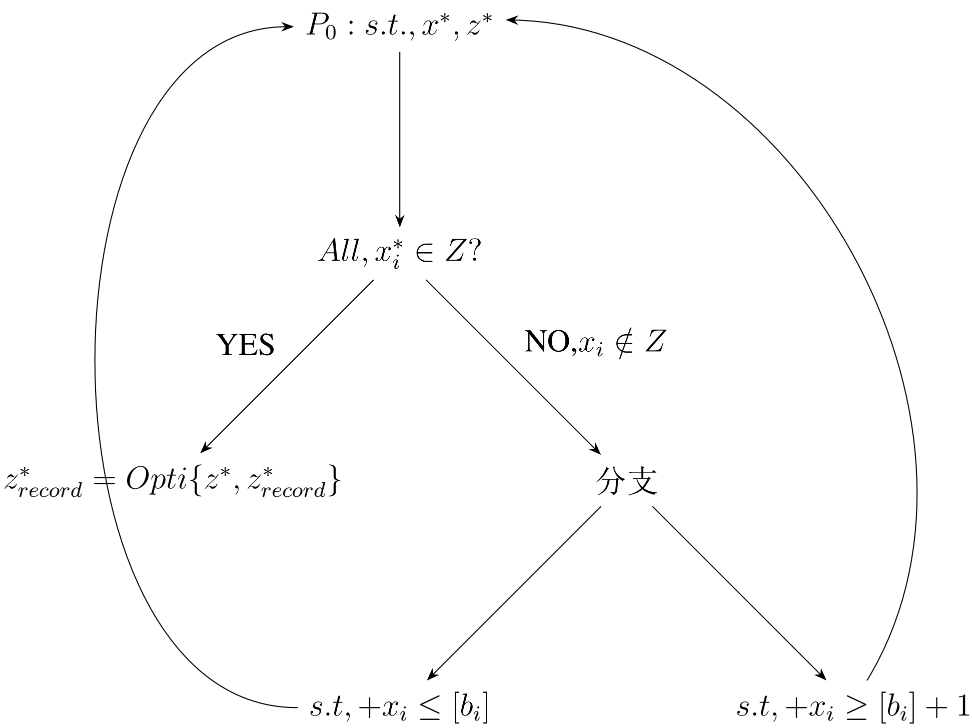

图 8.11: 分枝定界法

8.5 0-1 规划

0-1 规划即决策变量只能是 0 和 1，其经典形式为：

$$MAX : Z = CX \tag{8.64}$$

$$s.t. \begin{cases} AX \leq b \\ x_i = 0, 1, x_i \in X \end{cases} \tag{8.65}$$

对于 0-1 规划问题，其解的个数是有限的 2^n，n 为变量个数。可能的解数量很少的时候可以采用**枚举**的办法逐个求出解的目标值并取出最优的目标值。但是，枚举法在变量很多的时候工作量很大，下面介绍采用图论的方法解 0-1 规划的经典问题——指派问题。在介绍解法之前，首先了解图论的基础知识。

8.5.1 图论基础

定义 8.2

顶点：$V = \{v_1, \ldots, v_n\}$ 表示顶点的集合。

定义 8.3

边：$E = \{e_1, \ldots, e_m\}$ 表示边的集合，$e_i = v_{j_1}v_{j_2}$。

定义 8.4
图：$G=(V,E)$，图是顶点和边的集合。♣

定义 8.5
（两点）连接：$\exists e=v_iv_j$。♣

定义 8.6
（点和边）关联：For e,v_i,$\exists v_j$ set $e=v_iv_j$。♣

定义 8.7
（顶点的）次/度：和顶点关联的边数。♣

定义 8.8
环：$\{e,v_i|e=v_iv_i\}$。♣

定义 8.9
多重连接/弧：$\exists e_1,e_2,\ldots$ *and* $e_i=v_{j_1}v_{j_2},i=1,2,\ldots$。♣

定义 8.10
简单图：无环和多重连接的图。♣

定义 8.11
无向图：连接没有方向的图。♣

定义 8.12
有向图：连接有方向的图。♣

定义 8.13
完全图：所有结点之间都有连接的图。♣

定义 8.14
二分图：$G=G_1+G_2$，G_1 和 G_2 之间有连接，各自内部没有连接。♣

定义 8.15
链/路/道路/路径/链路：由点依次连接形成的路径上的点和连接的集合。♣

定义 8.16
圈/回路：闭合链。♣

定义 8.17

连通图：任意两点间都有道路连接的图。♣

定义 8.18

树：无圈的连通图。♣

定义 8.19

生成子图：在图 G 中取一部分得到的新图 G'，$G' = \{V, E'\}, E' \subset E$。♣

定义 8.20

生成树：如果一个图的子图是一个树，此树为生成树。♣

定义 8.21

最小生成树：具有最小权重和的生成树。♣

树的性质：

- $\Leftrightarrow$ 无圈，$m = n - 1$
- $\Leftrightarrow$ 连通，$m = n - 1$
- $\Leftrightarrow$ 无圈，加一条边就有圈
- $\Leftrightarrow$ 连通，减去一条边就不连通

求最小生成树的方法：

- 破圈法：找圈，去掉最大权重线，直到无圈 $(m = n - 1)$
- 避圈法：找最小权重边，逐步加边增点，避圈舍线直到遍历所有节点

图的矩阵表示：

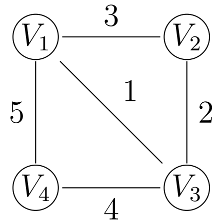

图 8.12: 无向图示例

图的邻接阵和权重阵：

$$A_{connect} = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \tag{8.66}$$

$$A_{weight} = \begin{bmatrix} 0 & 3 & 1 & 5 \\ 3 & 0 & 2 & 0 \\ 1 & 2 & 0 & 4 \\ 5 & \infty & 4 & 0 \end{bmatrix} \tag{8.67}$$

$A^2_{connect}, A^3_{connect}$ 分别代表矩阵的 $x_{i,j}$ 的度和所在三角形数量的二倍。

8.5.2 指派问题

指派问题是给 n 个人分配 n 个任务，每个人一个任务，达到最优化。指派问题的标准形式是：

$$MIN : z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{i,j}x_{i,j} \tag{8.68}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} x_{i,j} = 1 \\ \sum_{i=1}^{n} x_{i,j} = 1 \\ x_{i,j} = 0, 1; i, j = 1, \ldots, n \end{cases} \tag{8.69}$$

其中，$x_{i,j} = 1$ 代表将任务 j 指派个人 i，并且这里考虑的 $c_{i,j} >\geq 0$。将一个指派问题转换成标准问题可能需要考虑：

- MAX 问题 $\rightarrow c'_{i,j} = -c_{i,j}$
- 人数不等于任务数 $\rightarrow$ 虚拟人或者任务
- 每个人可以做多个任务或者每个任务可以多个人同时做 $\rightarrow$ 虚拟分身
- 某人不能做某个任务 $\rightarrow c_{i,j} = M, M = \infty$

对于指派问题的标准形式，它的可行解是由不同行不同列的元素构成，一共有 n 个独立零元素，总共有 $n!$ 个可行解。设 $j_1 j_2 \ldots j_n$ 是 $1, 2, \ldots, n$ 的全排列，则指派问题的目标值 $z = MIN\{c_{1,j_1} + \ldots + c_{n,j_n}\}$。解决指派问题可以采用单纯形法、运输作业表、整数规划、枚举法等方法解决，但是指派问题可以转化为一个统一的形式，这个形式的指派问题的最优解一目了然。这个统一问题就是：

$$MIN : z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{i,j}x_{i,j} \tag{8.70}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} x_{i,j} = 1 \\ \sum_{i=1}^{n} x_{i,j} = 1 \\ x_{i,j} = 0, 1; i, j = 1, \ldots, n \\ c_{1,j_1} = \ldots = c_{n,jn} = 0 \end{cases} \tag{8.71}$$

这里的 $j_1 j_2 \ldots j_n$ 是全排列，上述问题的下界是 0，而解 $x_{i,j_i} = 1, i = 1, \ldots, n$ 的目标值恰好是 0，该解就是最优解，但不一定是唯一解。将标准问题转化为统一问题依赖指派问题的特性，接下来说明如何调整指派问题以及其理论证明。

定理 8.14

系数矩阵一行（列）加上同一个数 k，最优解不变。

证明 设原指派问题的可行解 $X_1 = [x_{1,j_1}, \ldots, c_{x,j_n}]'$（只列出独立元素），目标值为 $Z_1 = C_1X_1, C_1 = [c_{1,j_1}, \ldots, c_{n,j_n}]$，行 i 加上一个数 k 以后对于同样的解 X_1，目标值变为

$$Z_2 = [c_{1,j_1}, \ldots, c_{i,j_i} + k, \ldots, c_{n,j_n}]X_1 = Z_1 + k \tag{8.72}$$

因此原问题和调整后的问题的解和目标值一一对应，原问题取到最优解等价于新问题取到最优解，需要注意的是这里一般要保持系数矩阵非负。

定理 8.15

如果 n 个独立元素全为 0，则一定为最优解。

证明 独立元素全为 0 时该问题对应统一问题，这些独立零元素对应的解就是最优解，也是原问题的最优解。

有了定理8.5.2,8.5.2，现在只需要通过行调整和列调整找出 n 个独立零元素即可找到最优解，问题在于如何判断选择的独立零元素集是最大的集合并且如何调整，下面这个定理将解决这个问题并且引出匈牙利法。

定理 8.16

独立零元素的最大数等于覆盖所有 0 的直线的最小数。

证明 对于系数矩阵 A，每行、列的均逐步减去其最小值，生成每行每列均有 0 元素的系数矩阵，并从中选择 m 个独立零元素，记独立零元素为 $\odot$, 非独立零元素为 $\oslash$，下面是找覆盖直线的流程：上述划线流程考虑了所有可能发生的情况，这里特别要注意的是调整的过程，在现有的划线下仍有 $\oslash$ 未覆盖，则选择该元素为独立零元素而将竖线上的独立零元素调整为非独立零元素，如果竖线上无独立零元素，则 $\oslash$ 直接加入集合，独立零元素的数量加一。并且这里换元素总是加竖去横而没有加横去竖，这是因为初始的划线是横线，在调整中加横去竖是逆运算，会重复之前的情形，导致退化无法求解。

现在导出匈牙利算法：

其中，系数矩阵调整方法为：

$$\theta = MIN\{Uncoveredfators\} \tag{8.73}$$

$$Row'_{uncovered} = Row_{uncovered} + \theta \tag{8.74}$$

$$Col'_{covered} = Col_{covered} - \theta \tag{8.75}$$

系数调整其实是在考虑如何生成 0 元素，但是为什么要这样调整暂时不清楚。下面用图论的方法补充找最小覆盖直线的方法。

8.5.3 匈牙利法原理

采用二分图描述指派问题：

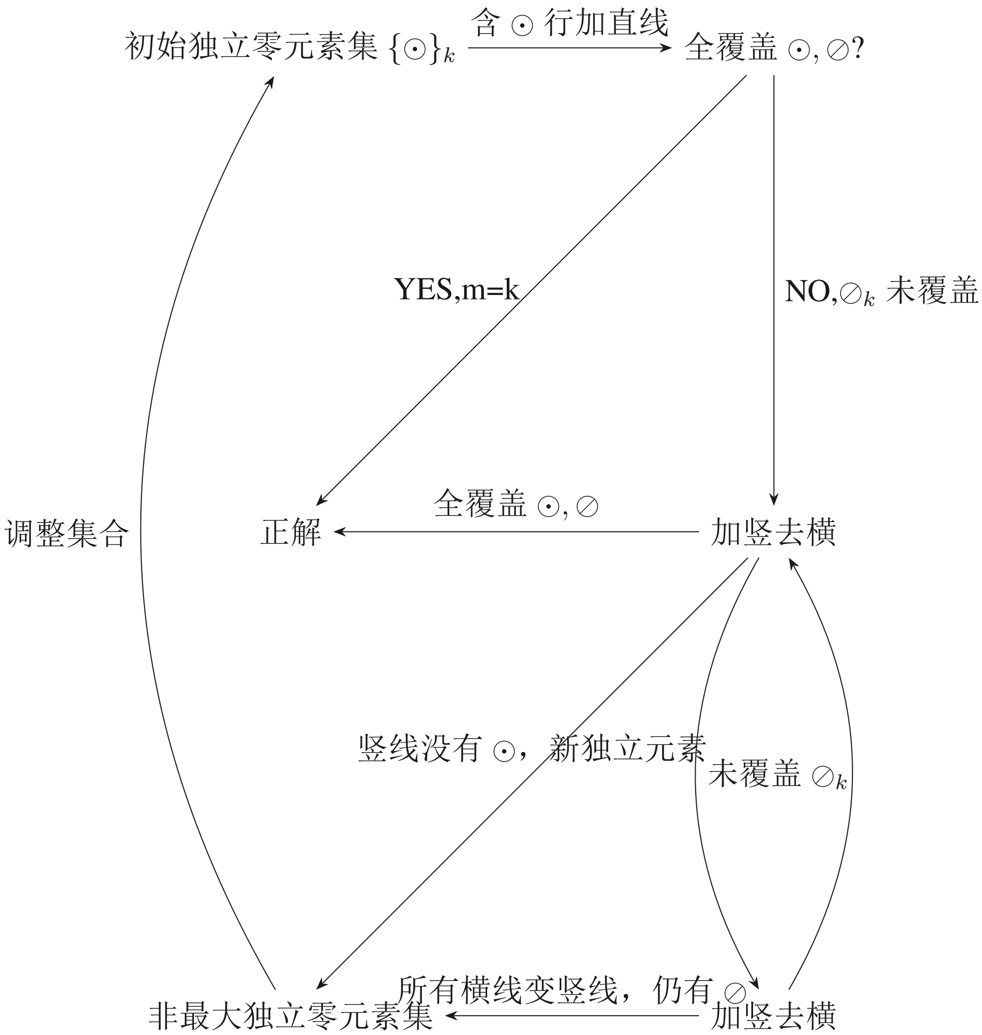

图 8.13: 划线流程

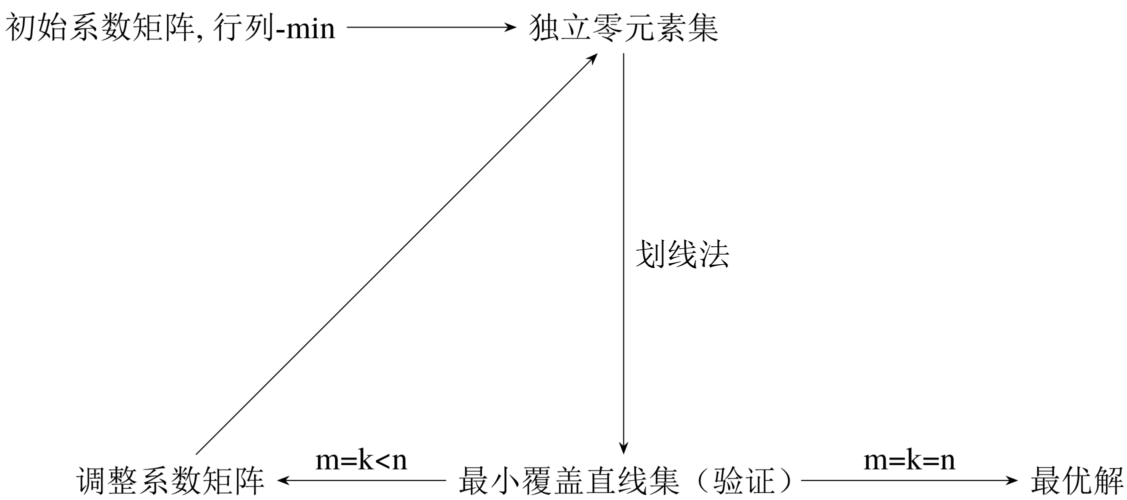

图 8.14: 匈牙利法

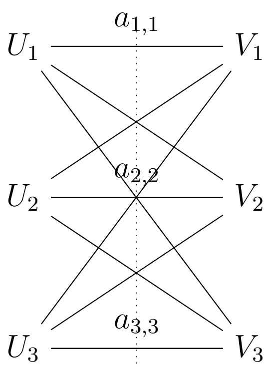

图 8.15: 指派问题二分图表示

$$U_1 \text{ —— } V_1$$

$$U_2 \text{ —— } V_2$$

$$U_3 \text{ —— } V_3$$

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8.76}$$

图 8.16: 指派问题可行解表示

用二分图表示指派问题，连接 $a_{i,j}$ 是第 i 个人完成任务 j 的成本（系数矩阵元素）。图的作用主要是判断调整后的系数矩阵的独立零元素集是否最大，这里直接基于调整后系数矩阵画出二分图，如下：

在指派问题分析示例中，连接 $a_{i,j}$ 表示元素 $c_{i,j}$ 为零元素，一个顶点可以有多条连接，但是只能有一个独立零元素，而且在可行解中只能出现一个关联的独立零元素，也就是说需要在所有连接中选择 k 条连接使得每条连接的关联点都不同。在二分图中如何表示以及如何调整独立零元素集呢？

定义 8.22

匹配：独立零元素在二分图中对应的连接。 ♣

定义 8.23

匹配点：匹配对应连接的顶点。 ♣

定义 8.24

交错链：依次经过 $N, M, N, \ldots$ 成的链，N 是未匹配点，M 是匹配点。 ♣

定义 8.25

增广链：由未匹配点出发到达匹配点的交错链。 ♣

$$C = \begin{bmatrix} 0 & 3 & 0 & 4 & 8 \\ 0 & 1 & 7 & 7 & 3 \\ 0 & 2 & 3 & 2 & 1 \\ 0 & 0 & 5 & 0 & 4 \\ 0 & 7 & 3 & 9 & 0 \end{bmatrix} \tag{8.77}$$

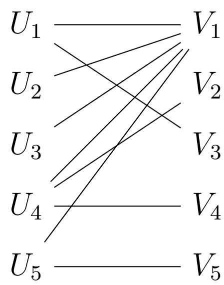

图 8.17: 指派问题分析示例

直观来看，增广链在系数矩阵中对应于可调整的情形，如：

$$C = \begin{bmatrix} \oslash & \odot \\ 1 & \oslash \end{bmatrix} \Rightarrow \begin{bmatrix} \odot & \oslash \\ 1 & \odot \end{bmatrix} \tag{8.78}$$

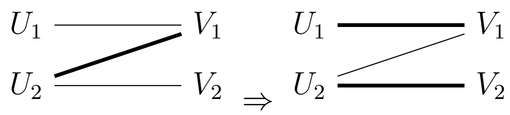

图 8.18: 增广链示例

下面这个基于交错链的定理是运用二分图寻找最大零元素集的关键。

定理 8.17

二分图中存在增广链等价于匹配可改进（独立零元素集可调整）。

证明 （1）先证必要性，匹配可改进则一定存在增广链。采用反证法，即证明不存在增广链匹配也可以改进不成立。设原匹配为 UV，新匹配为 $U'V'$，如图。

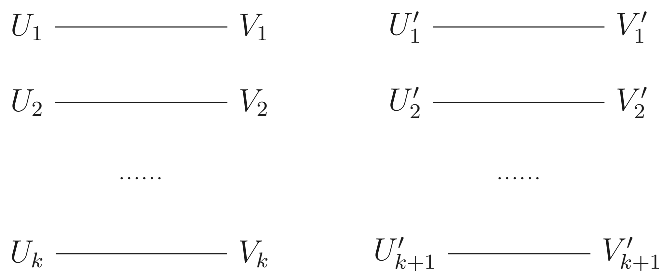

图 8.19: 无增广链 $\Rightarrow$ 匹配不可改进

原匹配为 $U_1 - V_1, \ldots, U_k - V_k$，匹配改进以后为 $U'_1 - V'_1, \ldots, U'_{k+1} - V'_{k+1}$。首先对于任意的匹配都可以重排命名以符合上述的形式，因此现在关心的是两个匹配之间的关联。假设新

匹配中多出的点是 U'_M 和 V'_N，如果 U'_M 和 V'_N 匹配，则原匹配中没有穷尽所有匹配，矛盾。故 U'_M 和 V'_N 不匹配，那就可以假设 U'_M 和 V'_{P_0} 匹配，而 V'_{P_0} 是原匹配中的匹配点，如 V_{Q_0}，V_{Q_0} 在原匹配中匹配点 U_{Q_0}，显然 U_{Q_0} 也对应新匹配中的 V_{P_1}，这样循环下去最终只能出现原匹配中的增广链 $U'_M - V_{Q_0} - V_{P_1} - \ldots - V'_N$。这一逻辑链条的核心在于原匹配和新匹配对应相同的顶点但是却有不同的匹配方式，一旦匹配不一致就会完全互换，由于这里加入了不能直接匹配的顶点，所以这里的互换是中间换两端接，也就出现增广链。

（2）再证充分性，即存在增广链则匹配可以改进。对于匹配 $U_1 - V_1, \ldots, U_k - V_k$，存在增广链 $U_M - V_{P_0} - \ldots - V_N$，这里我们只关心根据增广链改进匹配而不调整无关匹配。增广链中包含的匹配为 $V_{P_0} - U_{P_0}, \ldots, V_{P_k} - U_{P_k}$，共有 k 个匹配。将增广链从始点开始逐个拆分为匹配 $U_M - V_{P_0}, U_{P_0} - V_{P_1}, \ldots, U_k - V_N$，则出现了 $k+1$ 个匹配，这样一条增广链就可以且仅能增加一个匹配，也就是匹配改进了。

基于定理8.5.3，可以得出根据二分图找最小覆盖直线的方法：在二分图中，从左侧未匹配点出发走过所有的交错链，并标记节点，所有左侧节点未标记而右侧节点标记的点构成最小覆盖直线（左侧横线右侧竖线）。这一方法的思想是左侧未匹配点出发，经过交错链就是对定理8.5.2的证明中划线流程重现。如果匹配边的左侧和右侧均标记，说明需要加竖去横，以覆盖非独立零元素，如果匹配边的左侧和右侧均未标记，说明该匹配边无需调整，依然是划横线。匹配边的左侧是匹配点，不可能只标记左边不标记右边，并且是交错链不可能只标记右边不标记左边。对于未匹配点，从左侧标记是为了在加竖去横中将其划去。

总的来说，解决指派问题的匈牙利法如图8.14，只需强调的是运用二分图找最小覆盖直线是划线法的替代方法，两种方法的内核是统一的，都是在确认和改进独立零元素集，而且调整的方法也一致。

8.6 网络优化

前面所研究的是在凸集上的优化问题，借助上一节提供的图论知识，这里分析图的优化问题，最基本的图优化是有向图最短路、无向图最短距离和有向图最大流量问题。

8.6.1 有向图最短路

有向图最短路是在有向图 (G) 中两个顶点的最短通路 $(V_1 \ldots V_n)$。如：

定理 8.18

最短路上的子路径仍然是最短路。

证明 假设当前的最短路的子路是 $V_i, \ldots, V_j$ 并且存在更短的 V_i, V_j 的通路 $V_i, \ldots', V_j$，那么将 $V_i, \ldots', V_j$ 替代最短路的子路 $V_i, \ldots, V_j$ 将得到更短的最短路，这与最短路的前提矛盾。

从定理8.6.1可以推出求最短路的方法：从起点开始，逐个记录到达图上各点的最短路，直到达到终点。具体的算法如下：

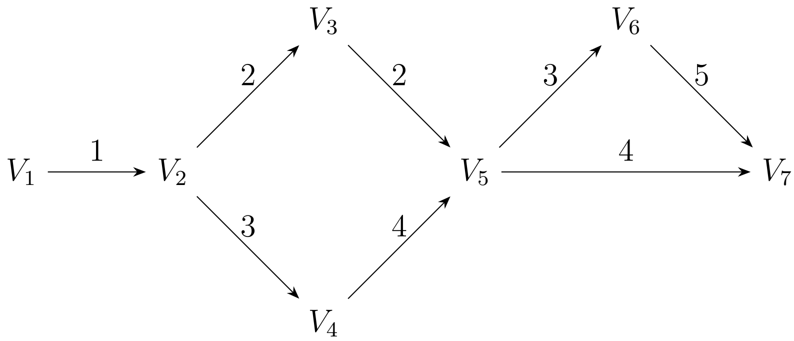

图 8.20: 有向图最短路示例

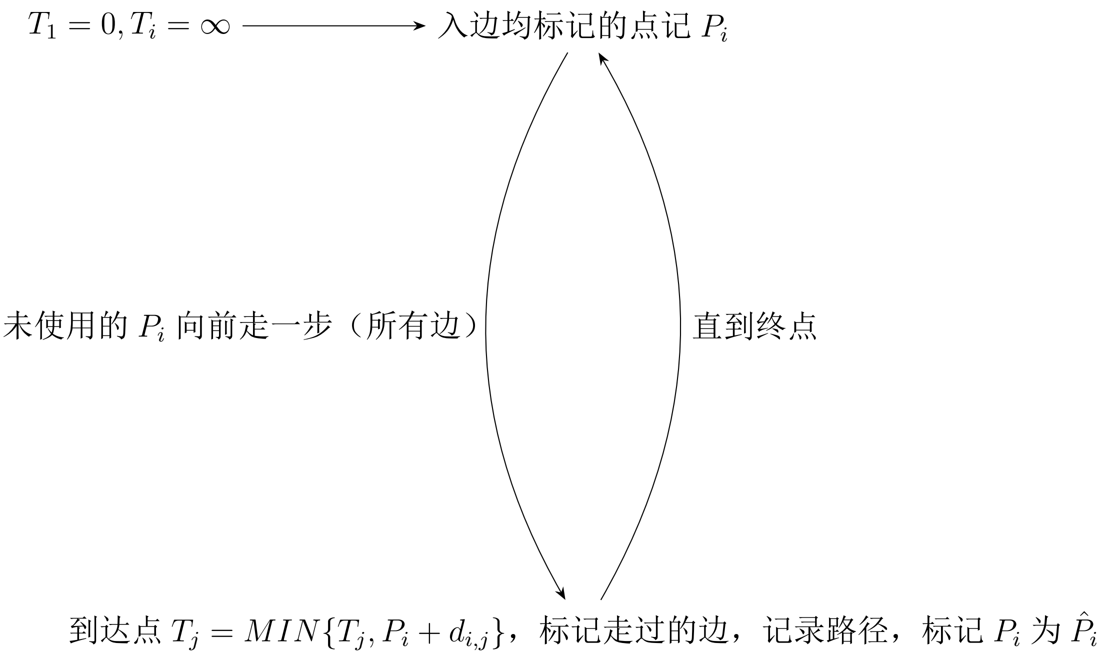

图 8.21: Djikstra 算法

	T_1 V_1	T_2 V_2	T_3 V_3	T_4 V_4	T_5 V_5	T_6 V_6	T_7 V_7
1	$\hat{0}$	∞	∞	∞	∞	∞	∞
2		$\hat{1}_{1,2}$	∞	∞	∞	∞	∞
3			$\hat{3}_{1,2,3}$	$\hat{4}_{1,2,4}$	∞	∞	∞
4					$\hat{5}_{1,2,3,5}$	∞	∞
5						$\hat{8}_{1,2,3,5,6}$	$\hat{9}_{1,2,3,5,7}$

表 8.17: Djikstra 作业表

其中 T_i 是点 i 的暂时值，用于比较和记录路径长度，P_i 是点 i 的永久值，记录从始点到达点 i 的最短距离，这里入边均标记表示可能的路径都试过了，可以确定下来最短路。其作业表如表8.17。

8.6.2 无向图最小距离

无向图最小距离即在无向图中任意两点的最小距离。

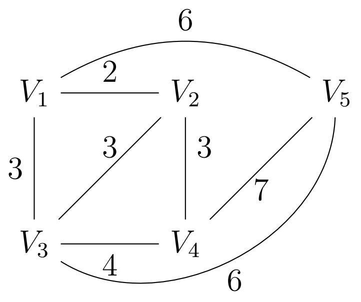

图 8.22: 无向图最短距离示例

设无向图的系数矩阵 D 为:

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,n} \\ d_{2,1} & d_{2,2} & \dots & d_{2,n} \\ \dots & \dots & \dots & \dots \\ d_{1,1} & d_{1,2} & \dots & d_{1,n} \end{bmatrix} \tag{8.79}$$

D 是一个对称阵，$d_{i,j}$ 代表的是从 i 一步到达 j 的距离。如果要考虑两步到达，应该如何计算呢？

$$d_{i,j}^{(2)} = MIN\{d_{i,1} + d_{1,j}, \dots, d_{i,n} + d_{n,j}\} = MIN_k\{d_{i,k} + d_{k,j}\}; \tag{8.80}$$

现在可以定义运算 $\odot$，

$$D^{(2)} = D \odot D, d_{i,j}^{(2)} = MIN_k\{d_{i,k} + d_{k,j}\} \tag{8.81}$$

在实际运算中 $d_{i,j} = d_{j,i}$，进行 $\odot$ 运算可以变为对应的行或者列相加的最小值，并且 $D^{(k)}$ 也是对称阵，对角线上为 0，每次只需进行 $\frac{n(n-1)}{2}$ 次计算。这一求无向图最短距离的方法是 Floyd 算法，总结如下：对最长链长为 n 的无向图，最多需要进行 k 次 $\odot$ 运算，$k = [log_2 n] + 1$，当 $\odot$ 运算后系数矩阵不变，则已经达到了最长链 (增加连接的是 $v_i - v_i$)。

8.6.3 有向图最大流量

有向图最大流量问题是在流量限制的情况下在各条通路上分配流量以获得从始点到终点的最大流量的问题。即对于有向图 $G = (V, \vec{E}, C)$，通路 $V_i\vec{V_j}$ 的限制流量为 $c_{i,j}$，记 $f_{i,j}$ 为 $V_i\vec{V_j}$ 上的流，$w_{i,j}$ 为 $f_{i,j}$ 的流量，则对于可行流的限制包括：

- $0 \leq w_{i,j} \leq c_{i,j}$

- 点 j 的进出流平衡：$\sum_i w_{i,j} = \sum_k w_{j,k}$
- 流平衡：$\sum_j w_{1,j} = \sum_i w_{i,n}$

解决有向图最大流量问题的关键在于找出割集，这是流通的瓶颈。

定义 8.26

割集：如果 E' 满足以下条件：

1. $G = (V, E - E', C)$ 不连通
2. $G = (V, E - E', C)$ 可分为两个连通图 $G_1 = (S, E_1, C_1), G_2 = (\bar{S}, E_2, C_2)$
3. $S \cup \bar{S} = V, S \cap \bar{S} = \oslash, E_1 \cup E_2 = E, E_1 \cap E_2 = \oslash$
4. $V_1 \in S, V_n \in \bar{S}$
5. $\forall V_i\vec{V_j} \in E', V_i \in S$ *and* $V_j \in \bar{S}$

则称 $(S, \bar{S})$ 为割集。♣

定义 8.27

最大流量：可行流中的流量最大的流量，记为 $w^*(f^*)$。♣

定义 8.28

最大流：最大流量对应的流，记为 $f^*(w^*)$。♣

定义 8.29

割集容量：割集所有起点终点的荣量之和。♣

定义 8.30

最小割集：具有最小容量的割集，记为 $(S^*, \bar{S}^*)$，容量为 $C^*(S^*, \bar{S}^*)$。♣

定理 8.19

$w(f) \leq c(S, \bar{S})$ ♡

证明 该定理非常符合直觉：可行流量小于限制流量。对于任何流 f 和任意的割集 $(S, \bar{S})$，该流总是要通过割集的通路 E'，因此流量要小于割集的容量。

基于定理8.6.3，可以推出：

$$w(f) \leq C^*(S^*, \bar{S}^*) \tag{8.82}$$

$$w^*(f^*) \leq C(S, \bar{S}) \tag{8.83}$$

$$w^*(f^*) \leq C(S^*, \bar{S}^*) \tag{8.84}$$

定理 8.20

$w^*(f^*) = c^*(S^*, \bar{S}^*)$ ♡

证明 首先区别正向边和反向边，相对于一条选定的链来说，连接的方向和链的方向一致的是正向边，相反的是反向边。如果正向边上的 $w_{i,j} = c_{i,j}$，称该流是饱和的，否则是不饱和的，余量

为 $c_{i,j}-w_{i,j}$。如果反向边上的 $w_{i,j}=0$，则称该流是饱和的，否则是不饱和的，余量为 $w_{i,j}$。饱和的含义是无法增加流量。只需要找出一个可行的 $w(f)=c(S,\bar{S})$ 即可证明 $w^*(f^*)\le c^*(S^*,\bar{S}^*)$，因为二者相等时都同时取到极值。

如果 $w^*(f^*)$ 是最大流，从起点开始标记，对于已标记的点 V_i，找出不饱和的关联边 V_iV_j（正向边 $w<c$，反向边 $w>0$），对 V_j 进行标记使尽可能其达到饱和，直到到达 V_n。现在令所有已标记点构成的集合为 S，一定有 $V_n\notin S$，否则在 S 中存在一条可增广链 u，可以构造

$$f^{*'}=\begin{cases}w^*_{i,j}+\sigma,\text{正向边}\\w^*_{i,j}-\sigma,\text{反向边}\\w^*_{i,j},\textit{not in u}\end{cases}\tag{8.85}$$

这里，$\sigma=min\{\sigma_{i,j}\}=min\{c_{i,j}-w_{i,j}\}$。此时 $f^{*'}$ 也是可行流，但是流量比 f^* 增加 σ，因此 f^* 不是最大流，矛盾。

现在，令 $\bar{S}=V-S$，则得到割集 $(S,\bar{S})$，现在割集上的流量都是饱和的，并且有 $c(S,\bar{S})=w^*(f^*)$，因此 $w^*(f^*)=c^*(S^*,\bar{S}^*)$ 。

现在可以总结有向图最大流问题的解法。（1）增广链法：从 0 流开始，找可增广链进行调整，直到没有可增广链为止境（存在不饱和边）。（2）最大流标号法，如图。

初始可行流 (0 流)

$\downarrow$

标记 V_1 为 $\oplus$

$\downarrow$

已标记点 V_i，找未标记的有余量的 V_j

$\downarrow$

正向边标记 $\oplus V_i = min\{\sigma_i, c_{i,j} - w_{i,j}\}$, 反向边标记 $\ominus V_i = min\{\sigma_i, c_{i,j}\}$

$\downarrow$

若 V_n 未标记，得到增广链 u 可调整

$\downarrow$

正向边 $+\sigma$, 反向边 $-\sigma$

$\downarrow$

得到新可行流，去掉标记

$\downarrow$

无增广链，S^* 为标记点集合，$\bar{S}^*$ 为未标记点集合

$\downarrow$

$w^*(f^*) = c^*(S^*, \bar{S}^*)$

图 8.23: 最大流标号法

第 9 章 政治经济学

政治经济学是以历史的生产关系或一定的社会生产关系为研究对象的经济学。本文主要参考逄锦聚编写的《政治经济学》一书，梳理了政治经济学中的一般理论，意在对政治经济学有一个较为全面且基础的把握。政治经济学的一般理论主要有生产力、生产关系和生产方式之间的关系；商品、劳动和价值之间的联系；货币、资本、信用等概念的联系；社会再生产的问题；垄断与竞争，本章依托于前述分类对政治经济学的一般理论进行介绍。在学习的过程中，也要注重与现实经济发展的联系，学以致用。

9.1 生产力、生产关系和生产方式

9.1.1 社会再生产过程中的生产关系

定义 9.1

生产关系：一个由多重关系组成的系统，既包括生产、交换、分配和消费的相互关系，又包括人们在生产、交换、分配和消费各个环节中的关系。♣

生产与消费的关系：

1. 生产决定消费。生产决定着消费的对象，如吃、穿、住、用等消费品，同时，生产也决定着消费的方式；
2. 消费也决定着生产。一方面，只有通过消费，才能使产品成为现实的产品，另一方面，消费是生产的目的和动力，消费又创造出现实的生产。

生产与交换的关系：

1. 生产决定交换。生产的性质决定交换的性质，生产发展的程度决定交换的发展程度；
2. 交换对生产也具有反作用。随着交换的发展，市场的扩大，从而对用来交换的产品需求的增长，生产也会随之发展。

生产与分配的关系：

1. 分配处于生产与消费之间，但分配也并不独立于生产之外。在产品分配上，生产决定分配；
2. 分配对生产也有反作用，分配不仅仅是被动的生产成果的分配。

人们在生产、交换、分配和消费过程中的关系：

1. 人们在直接生产过程中的相互关系，一方面是由生产资料的所有制形式所决定的，另一方面也与生产的基本方式有关；
2. 人们在交换过程中的相互关系包括三个方面，直接生产过程内部活动的交换、再生产过程中国民经济各个部门、地区、企业、个体劳动者之间的产品交换、在不同的利益主体之间的交换；

3. 人们在分配过程中的相互关系包括生产资料分配中人们之间的关系、产品分配过程中人们之间的关系。

9.1.2 所有制和产权

定义 9.2

所有制：所有是所有制关系的基础，所有制是一个经济制度范畴。♣

定义 9.3

生产资料所有制：作为经济范畴，其内部结构由人们对生产资料的所有、占有、支配、使用等经济关系所组成。♣

定义 9.4

所有权：第一，作为法律用语是指对象的排他性，一个人或某个集团垄断地占有某物，从而可以按照自己的意志对该物处理；第二，作为经济用语是指生产资料的归属、拥有关系，也即生产资料归谁所有的问题。依据所有权，人们实现其经济利益关系。♣

定义 9.5

经济学分析中的所有权：包括资本、劳动力、土地、技术和企业家才能等在内的生产要素的所有权。♣

所有制与所有权的区别与联系：

1. 所有权是所有制的法律表现形式；
2. 所有制的性质和内容决定着所有权的性质和内容。

定义 9.6

产权：在社会经济运行过程中，人们在生产资料上形成的所有、占有、支配、使用、处分、转让、收益等关系，受到成文或非成文的法律的承认和保护，在法权方面就分别表现为所有权、占有权、支配权、使用权、处分权、转让权和收益权。以上权利都同一定的财产关系相联系，也被称为财产权利、财产权或产权。♣

马克思主义产权理论（特别关注产权的公有和私有的属性）：

1. 所有权表现为在一定经济关系中的个体或团体对生产条件的排他的占有或归属关系；
2. 财产权利关系的实质是人与人之间的经济关系；
3. 产权对财产主体有实现利益的要求

现代西方产权理论（出发点是私人产权）：

1. 产权是人们在交易过程中获取的一定的收益权利，具有可转让的特征；
2. 产权是由一组或一束权利组成的，它包括所有权、使用权、收益权、转让权等，形成产权结构。我们将这种由一束产权组成的结构成为权能结构。

9.1.3 生产力

定义 9.7

生产力：人们改造自然和控制自然界的能力，它反映人和自然界之间的关系。包含三要素：人的劳动、劳动资料和劳动对象。劳动是生产力的决定性要素，但划分经济时代的标志是生产工具。♣

定义 9.8

生产方式：生产力和生产关系的对立统一，构成社会的生产方式，生产方式包括生产力和生产关系两个方面：生产力是生产的物质内容，生产关系则是生产的社会形式。♣

生产力与生产关系：

1. 在生产力与生产关系的对立统一中，生产力起决定性作用，一方面，生产力在社会生产中是最活跃、最革命的因素，处在不断发展变化过程之中，它的这种发展变化会引起生产关系的变化和发展；另一方面，生产力的状况如何，决定或要求有什么样的生产关系；
2. 生产关系对生产力具有反作用：当生产关系与生产力相适应时，它就是生产力的发展形式，就能促进生产力的发展；当生产关系与生产力状况不相适应时，它是生产力的桎梏，就会阻碍生产力的发展。先进的社会生产方式应该是先进的社会生产力与先进的社会制度相适应。

9.2 劳动、商品和价值

9.2.1 商品的使用价值和价值

定义 9.9

商品：用来交换的劳动产品，包含使用价值和价值两个要素，是使用价值和价值的统一体。♣

定义 9.10

使用价值：指物品和服务能够满足人们某种需要的属性，即物品和服务的有用性。使用价值构成社会财富的物质内容。♣

商品的使用价值与一般物品的使用价值的差别：

1. 它必须是劳动产品；
2. 商品的使用价值是为他人、为社会有用的；
3. 必须通过交换让渡给他人才能进入消费。

定义 9.11

交换价值：表现为一种使用价值同另一种使用价值相交换的量的比例关系。♣

定义 9.12

价值：凝结在商品中的无差别的一般人类劳动，价值是商品的本质的、社会属性。♣

价值与交换价值的关系：价值是交换价值的基础；交换价值是价值的表现形式。

使用价值和价值的关系：使用价值是价值的物质承担者。一方面，商品作为使用价值，在质上各不相同，在量上难以比较；另一方面，商品作为价值，在质上是相同的，在量上是可以比较的，因此，各种不同的商品能按一定的比例进行交换。

9.2.2 抽象劳动和具体劳动

商品的两要素与劳动二重性：商品的两要素是由生产商品的劳动二重性决定的，具体劳动创造商品的使用价值，抽象劳动创造价值。

定义 9.13

具体劳动：即从劳动的具体形态考察的劳动。具体劳动反映了人和自然的关系，是人类社会生存和发展的条件，是劳动的自然属性，是一个永恒性范畴。♣

定义 9.14

抽象劳动：无差别的一般人类劳动。抽象劳动是同质的、无差别的形成商品价值的劳动，抽象劳动形成商品的价值实体，是形成商品价值的惟一源泉。抽象劳动是劳动的社会属性，是商品生产所特有的范畴，商品生产的历史性，决定了形成价值的抽象劳动的历史性。♣

具体劳动和抽象劳动的关系：具体劳动和抽象劳动是生产商品的同一劳动的两个方面，二者在时间上、空间上都是不可分割的。只有在商品生产和商品交换的经济关系中，具体劳动才需要还原为抽象劳动，人类脑力和体力的耗费才形成价值。在不存在商品经济的条件下，人们的劳动产品不用来交换，他们的劳动就只表现为具体劳动，而没有必要将其还原为同一的抽象劳动来计算其劳动量。

9.2.3 个别劳动和社会劳动

定义 9.15

个别劳动：指属于不同所有制的劳动，在私有制条件下，个别劳动即私人劳动；同一个所有制内部，个别劳动指不同的有独立利益的企业的劳动。♣

定义 9.16

社会劳动：劳动的社会性是人类劳动的本质特征。在商品经济中，由于社会分工的存在，商品生产者之间是相互联系、相互依存的，每个商品生产者的劳动都是社会总劳动的一部分。♣

个别劳动与社会劳动

1. 个别劳动和社会劳动的矛盾是商品经济内在矛盾的根源。个别劳动和社会劳动的对立在商品生产中表现为具体劳动和抽象劳动的对立，进而产生使用价值和价值的对立；
2. 个别劳动要转化为社会劳动必须经过流通过程。

定义 9.17

个别劳动时间：各个商品生产者实际耗费的劳动时间。♣

定义 9.18

社会必要劳动时间：是指在现有的正常的生产条件下，在社会平均的劳动熟练程度和劳动强度下生产某种使用价值所需要的劳动时间（这种含义的社会必要劳动时间，是在生产同种商品的不同生产者之间形成的，它涉及的是同种商品生产上的劳动耗费）。第二种含义是指社会总劳动中按一定比例用来生产社会需要的某种商品所耗费的劳动时间。涉及的是社会总劳动时间在各种商品上的分配。♣

两种社会必要劳动时间含义的联系：社会必要劳动时间的两种含义具有相关性，共同决定商品的价值。第一种含义是价值的决定，第二种含义则是价值的实现。

两种社会必要劳动时间含义的区别：

1. 第一种含义是从社会生产条件的角度来说明社会必要劳动时间的，第二种含义则是从社会需要的角度来说明社会必要劳动时间的；
2. 第一种含义所决定的是单位商品的价值，第二种含义所决定的则是部门总商品的价值；
3. 第一种含义涉及劳动消耗，第二种含义则涉及社会规模的使用价值。

9.2.4 价值量和价值规律

定义 9.19

劳动生产率：指劳动者在一定时间内生产某种使用价值的效率。♣

决定劳动生产率高低的因素：

1. 劳动者的平均熟练程度；
2. 科学技术的发展水平及其在生产中的应用程度；
3. 生产过程的社会结合（分工协作、劳动组织、生产管理）形式；
4. 劳动对象的状况以及自然条件等。

劳动生产率和商品的价值量的关系：

1. 不管生产力发生了什么变化，同一劳动在同样的时间内提供的价值量总是相同的；
2. 劳动生产率同商品的使用价值成正比，同商品的价值量成反比。

定义 9.20

价值规律：是商品经济的基本规律。既是价值量如何决定的规律，也是价值量如何实现的规律。♣

价值规律的基本内容：商品的价值量是由生产商品的社会必要劳动时间决定的，商品的交换要依据商品的价值量来进行，实行等价交换。

价值规律作用的形式（商品经济条件下）：价格受供求关系的影响，围绕价值上下波动。

9.3 货币和货币流通量

9.3.1 货币的本质

货币的起源和发展：货币是随着商品的交换过程、价值形式的发展而发展的，它的产生经历了以下的历程： 简单的、个别的或偶然的价值形式； 总和的或扩大的价值形式； 一般价值形式； 货币形式。

货币的本质：货币是固定地充当一般等价物的特殊商品。

9.3.2 货币的职能

货币的职能：价值尺度、流通手段、贮藏手段、支付手段、世界货币。价值尺度和流通手段是货币的基本职能。

- 价值尺度：货币表现、衡量、计算商品价值的尺度。
- 流通手段：指货币充当商品交换媒介的职能。
- 贮藏手段：指货币退出流通领域当作独立的价值形式和社会财富贮藏起来的职能。
- 支付手段：指货币用来清偿债务或支付赋税、租金、利息、工资等的职能。
- 世界货币：指货币在世界市场充当一般等价物的职能。

9.3.3 货币的形式

货币的形式：

- 金属货币：金属货币之所以能够被纸币代替，是因为商品价值观念地表现在一个金量上，这个金量则由纸象征地可感觉地体现出来。
- 纸币：纸币本身没有价值，仅仅是价值的符号；纸币只有代表金量（金量同其他一切商品量一样，也是价值量），才成为价值符号；纸币是由国家通过法定的途径发行的，纸币一旦进入流通，就要受到流通规律的支配。
- 信用货币：指代替金属货币充当支付手段和流通手段的信用证券。银行券是信用货币的主要形式，它是由发行银行发行的用以代替商业票据的银行票据；存款货币是信用货币的另一种重要形式，是指能够发挥货币作用的银行存款。

- 外汇：是国际汇兑的简称。具有两个本质特征：是以外币表示的资产；所表示的资产必须具有可兑换性。

9.3.4 货币流通量

货币层次：

- $M0 =$ 现金（纸币和硬币）；
- $M1 = M0+$ 所有金融机构的活期存款；
- $M2 = M1+$ 商业银行的定期存款和储蓄存款；
- $M3 = M2+$ 其他金融机构的定期存款和储蓄存款；
- $M4 = M3+$ 其他短期流动资产（如国库券、商业票据、短期公司债券、人寿保单等）。

定义 9.21

货币乘数：指在货币供给的过程中，中央银行的初始货币供给与最终形成的社会货币流通量之间存在着倍数扩张或收缩关系，它反映了商业银行通过贷款（投资）创造派生存款的扩张（或收缩）倍数。下式中，K 代表货币乘数，D 为货币存量，R 为基础货币存量：

$$K = \frac{1}{r} = \frac{D}{R} \tag{9.1}$$

定义 9.22

货币流通规律：货币流通同商品流通相适应的规律，即流通中的货币量必须满足商品流通的需要。

$$\text{一定时间内流通中需要的货币量} = \frac{\text{流通中商品价格总额}}{\text{同一货币单位的平均流动速度（次数）}} \tag{9.2}$$

$$\text{一定时期内流通中需要的货币量} = \tag{9.3}$$

$$\frac{\text{流通中商品价格总额} - \text{赊销商品价格总额} + \text{到期支付总额} - \text{互相抵销的支付总额}}{\text{同一货币单位的平均流动速度（次数）}} \tag{9.4}$$

9.4 资本

9.4.1 资本的形态和构成

定义 9.23

资本：是价值的一种特殊形式，是不断地在运动中谋求自身增殖的价值。

资本的特点：

1. 增殖性；

2. 资本的运动性；
3. 资本的返还性；
4. 资本的风险性。

资本形态：

1. 货币资本即以货币形态表现的资本，它是资本最一般的和初始的形态；
2. 实物资本是以物质形态表现的资本，包括投入商品生产过程和流通过程的一切物的要素和待售的产出品，亦称物质资本；
3. 无形资本是指市场主体所占有和使用的以知识形态存在的特有经济资源，它可以较长期地为该主体提供一定的特殊权利或有助于该主体取得相应的收益，通常包括专利权、商标权、版权、著作权、特许经营权、商誉、技术秘密（亦称专有技术或非专利技术）等；
4. 虚拟资本形态。

9.4.2 资本价值增殖

定义 9.24

必要劳动：指劳动者用以实现劳动力再生产而付出的劳动。♣

定义 9.25

剩余劳动：指一定时期内劳动者的劳动中超出必要劳动的部分，是劳动生产率提高的产物。♣

定义 9.26

剩余价值：剩余劳动创造的价值即剩余价值。♣

价值增殖是资本运动的原动力和目标。

定义 9.27

价值增殖额：

- 从它活的源泉考察，其为活劳动在剩余劳动时间里创造的价值，是剩余价值；
- 把它与全部预付资本比较即看作大于预付资本的增殖额，其是利润。

$$\text{利润率} = \frac{\text{利润额}}{\text{预付资本}} \times 100\% \tag{9.5}$$

$$\text{剩余价值率} = \frac{\text{剩余价值}}{\text{必要劳动价值}} \times 100\% \tag{9.6}$$

♣

9.4.3 资本积聚和资本集中

定义 9.28

资本积累：把剩余价值即利润转化为资本，使资本规模得以扩大的过程，资本积累是资本扩大再生产的重要源泉，而剩余价值或利润是资本积累的唯一源泉。♣

定义 9.29

资本积聚：指单个资本依靠自身的积累来使实际资本在价值形态和生产要素形态上实现量的扩大。♣

资本积累与资本积聚的关系：资本积累是资本积聚的前提和基础，资本积聚则能通过资本规模的扩大来增强资本积累的能力。

定义 9.30

资本集中：指把若干个已有的规模相对较小的资本合并重组为规模较大的资本。♣

资本积聚和资本集中的关系：

1. 都能使单个资本的规模增大，并且二者有相互促进的作用；
2. 资本积聚以剩余价值的积累为前提，而资本集中不以积累为必要前提；
3. 资本积聚的实现会受到社会所能提供的实际生产要素增长的制约，资本集中若采取原有资本之间的兼并和联合的途径，则较少受社会实际生产要素增长的限制；
4. 资本积聚在增大单个资本的同时，也增大了社会总资本；资本集中若采取原有资本之间的兼并或联合的途径，则一般不能直接增大社会总资本，但可以改变资本的结构和质量；
5. 通过资本积聚方式扩大单个资本规模，一般速度较慢，资本集中则可以用很快的速度实现资本规模的扩大。

9.4.4 资本循环和周转

定义 9.31

产业资本循环：产业资本从其最初的形态即货币资本的职能形态出发，依次经过购买、生产、售卖等阶段，并相应变换资本的职能形态，然后又回到原来的出发点的过程。♣

产业资本循环的三个阶段以及三种职能形式：

- 三个阶段：购买阶段、生产阶段、售卖阶段；
- 三种职能形式：货币资本、生产资本、商品资本；
- 产业资本的连续进行的现实循环，不仅是流通过程和生产过程的统一，而且是它的所有三个循环的统一。

资本循环正常进行必须具备的基本条件：

1. 产业资本的三种职能形式在空间上是的同时并存的；
2. 产业资本的三种职能形式的循环在时间上是相互继起的。

定义 9.32

资本周转：资本连续不断的、周而复始的循环运动。♣

9.5 社会总资本再生产和市场实现

9.5.1 社会总资本

定义 9.33

社会资本：以社会分工和市场交换为条件，相互联系、相互依存、相互制约的全社会各单个资本的总和。♣

定义 9.34

社会资本运动：互为条件、相互交替的单个资本运动的总和。社会资本运动的过程是生产和流通过程的统一，并且要在运动中变换资本的职能形式。♣

定义 9.35

社会总产品：社会在一定时期内（通常为一年）所生产的全部物质资料的总和。社会总产品在物质形态上，根据其最终用途，区分为用于生产性消费的生产资料和用于生活消费的消费资料。♣

定义 9.36

社会总产值：社会总产品的价值形态称为社会总产值，包括在产品中的生产资料的转移价值（c），凝结在产品中的由雇佣工人必要劳动创造的价值（v），以及凝结在产品中的由雇佣工人在剩余劳动时间里创造的价值（m）。♣

社会生产划的两大部类：第一部类（I）是由生产生产资料的部门所构成，其产品进入生产消费领域；第二部类 II）是由生产消费资料的部门所构成，其产品进入生活消费领域。

9.5.2 社会再生产

社会再生产的核心问题是社会总产出的实现：

1. 社会资本再生产反映了资本的现实运动，而资本从商品形态向货币形态的转化，是资本运动之“惊险的跳跃”。如果社会总产出不能或部分不能顺利销售，所耗费的资本就不能得到或不能充分得到价值补偿，社会再生产就不能正常进行；
2. 社会资本再生产的顺利进行，不仅要求生产中所耗费的资本在价值上得到补偿，而且要求资本价值所反映的实际生产过程所耗费的生产资料和消费资料得到实物替换。如果社会生产的物质产品和服务在实物构成上与所耗费和要补偿的不一致，则社会再生产的进行就会遇到困难。

定义 9.37

社会简单再生产：指社会剩余产品是用于消费而不是用于积累的，生产在维持原来的规模上重复进行。♣

社会简单再生产两大部类内部的实现条件：

$$I(c+v+m) = Ic + IIc, II(c+v+m) = I(v+m) + II(v+m) \tag{9.7}$$

社会简单再生产的基本实现条件：第一部类的消费需求得到补偿，第二部类的生产耗费得到替换，即：$I(v+m) = IIc$。

定义 9.38

社会扩大再生产：指社会生产在社会总资本循环运动中不断扩大规模，它可以通过生产要素投入的增加或生产要素产出效率的提高或二者同时发生而得到实现。♣

社会扩大再生产的前提条件：

1. 第一部类的产出在补偿了当年两大部类的生产资料和服务耗费以后，必须有一个余额，用来满足两大部类扩大再生产对追加生产资料和生产服务的需要，即：$I(v+m) > IIc$；
2. 第二部类的产出在补偿了当年两大部类的劳动者和资本所有者已消费的生活资料和生活服务后，也要有一个余额，用以满足两大部类扩大再生产对追加劳动力所引起的追加生活资料和生活服务的需要，即：$II[c+(m-m/x)] > I(v+m/x)$。

社会扩大再生产基本的实现条件：

$$I(v+\Delta v+m/x) = II(c+\Delta c) \tag{9.8}$$

9.6 信用制度与虚拟资本

9.6.1 信用

定义 9.39

信用：经济学中信用是指一种借贷行为，是以偿还付息为条件的价值运动的特殊形式。商品经济的产生和发展是信用产生的基础。♣

信用分类（以信用主体的不同）：商业信用、银行信用、企业信用、国家信用、消费信用等。

信用促进市场经济发展的积极功能：

1. 信用可以增加投资机会、促进资本的自由转移，推动社会资源的优化配置；
2. 信用可加速资本的积聚和集中；
3. 信用可以加快商品流通速度；
4. 信用可以给居民提供新的投资渠道和金融资产的持有方式；
5. 信用有效地调节着国民经济运行。

信用的消极功能：

1. 信用可能造成虚假繁荣，加深生产与消费的矛盾，触发生产过剩的危机；
2. 信用还可能引发货币信用危机；
3. 信用刺激投机，可能引起流通秩序混乱，使市场信号失真，生产盲目发展，加剧经济危机。

9.6.2 信用制度

定义 9.40

银行资本：银行资本所有者为经营银行获取利润所投入的自有资本和通过各种途径集中到银行的货币资本。包括：银行自有资本和银行借入的资本（银行吸收的存款）。♣

定义 9.41

银行业务：银行的负债业务主要是以吸收存款的方式借入资金；银行的资产业务就是银行通过发放贷款贷出资金。♣

银行存款的主要来源是：

1. 职能资本暂时闲置的货币资本；
2. 借贷资本或食利者阶层为获取利息而存入银行的货币资本；
3. 其他不同阶层居民的储蓄存款等等。

银行的发放贷款主要方式：

1. 票据贴现，即票据持有者为了融通资金，将未到期的票据交给银行，兑取现金；
2. 抵押贷款和信用贷款；
3. 投资业务，即从事购买有价证券的经营活动，如通过购买股票，成为其他企业的股东；或通过购买政府公债、公司债券等，形成一种长期贷款形式。

定义 9.42

股份公司：以入股方式筹集社会资本而形成的一种企业的资本或财产组织形式。股份公司是资本集中的途径和形式，按股东所负责任的不同，可分为：无限公司、有限公司、两合公司和股份有限公司。♣

定义 9.43

股票：股份公司发给股东的借以证明其股份数额并用以取得股息的凭证。具有三个特点：不返还性；风险性；流通性。♣

定义 9.44

股票价格：股票价格形成的基本因素有预期股息和银行利息率。股票价格与预期股息的大小成正比，而与利息率的高低成反比。股票价格 = 预期股息/利息率 = 股票面额 × 预期股息率/利息率。♣

9.6.3 虚拟资本和虚拟经济

定义 9.45

虚拟资本：指能定期带来收入的，以有价证券形式表现的资本。主要有两种形式：一种是信用形式上的虚拟资本（主要有期票、汇票、银行券、纸币、国家债券、各种证券抵押贷款等）；另一种是收入资本化形式上产生的虚拟资本（主要由股票、债券构成）。♣

虚拟资本是信用制度发展的结果：

1. 货币的虚拟化是虚拟资本产生的前提；
2. 借贷资本信用关系是虚拟资本产生的根据；
3. 社会信用制度的逐步完善是虚拟资本产生的基础。

实体资本与虚拟资本之间存在着对立统一的关系：

1. 统一性：虚拟资本的存在和运动必然要以它所表现的实体资本为基础；
2. 矛盾性：虚拟资本虽然是实体资本的价值表现和纸制副本，但其价格却不是实体资本价值决定的，而是由预期收入和平均利息率决定的。虚拟资本的价格变动会与实体资本价值变动相背离，其价格不随实体资本价值变动而变动，呈现出相对独立的运动。

定义 9.46

虚拟经济：是市场经济中信用制度与资本证券化的产物，是源于实体经济又相对独立于实体经济的虚拟资本的活动。♣

虚拟资本与虚拟经济：虚拟资本是信用制度和资本证券化的产物，但各种信用工具和证券化的资本形式并不必然是虚拟经济。

实体经济与虚拟经济的关系：实体经济与虚拟经济是现代市场经济运行同一过程的两个方面，二者是对立统一的关系。

1. 实体经济是虚拟经济的基础。实体经济的发展阶段、水平、结构和规模，决定着虚拟经济发展的状况。虚拟经济应该适应而不能脱离实体经济发展的需要而盲目发展，过度膨胀。
2. 虚拟经济也不是被动地决定于实体经济。在现代市场经济条件下，虚拟经济对实体经济具有越来越明显的反作用，并表现为虚拟经济对实体经济的正、负两个方面的效应。

9.7 竞争和垄断

9.7.1 竞争和垄断关系

定义 9.47

竞争：指商品生产者或其它经济利益主体，为了争取有利的生产、销售等条件，从而获取更多的经济利益而进行的角逐。♣

竞争的类型：

1. 根据市场结构特征：完全竞争与不完全竞争（寡头垄断与垄断竞争）；
2. 根据竞争主体所在的领域：同一产业内部企业之间的竞争及不同产业企业之间的竞争；
3. 根据竞争主体关系：卖方之间的竞争、买方之间的竞争、买者双方之间的竞争；
4. 根据竞争的手段：价格竞争与非价格竞争。

定义 9.48

垄断：指少数企业凭借其控制的巨额资本、足够的生产经营规模和市场份额，通过协定、同盟、联合、参股等方法，操纵与控制一个或几个部门的商品生产和流通，以获取高额利润。♣

垄断的类型：

1. 按照垄断主体存在的不同领域，可分为经济垄断（自然垄断、技术垄断、市场垄断及由产品差别而形成的垄断）与行政垄断；
2. 按照垄断行为，包括价格垄断行为、非价格垄断行为及组织垄断行为等等。

竞争引起垄断：竞争引起的生产和资本集中为垄断的产生提供了可能性与必要性：

1. 生产集中，使企业规模扩大，大企业的生产能力迅速膨胀；
2. 生产集中使大企业规模巨大，这对于中小企业进入大企业的生产经营领域就构成了较高的进入壁垒；
3. 少数大企业之间为了避免过度竞争而造成两败俱伤的灾难性后果，必然会寻求某种妥协，最终达成垄断协定。

9.7.2 不完全竞争市场

不完全竞争市场的类型：完全垄断市场、寡头垄断、垄断竞争。

定义 9.49

完全垄断市场：指某个产业（市场）上，卖方只有一家企业，其产品没有替代品，新企业进入为不可能的市场结构。♣

定义 9.50

寡头垄断：指由提供相似或相同产品的少数几家企业所组成的市场结构。它主要包括两种情况：一是少数几家企业生产完全相同的产品；二是少数几家企业生产有差别但可以替代的产品。♣

定义 9.51

垄断竞争：是一种介于垄断与竞争之间，既具有某些竞争的特点，又具有某些垄断特征的市场结构。♣

定义 9.52

垄断利润：指那些垄断企业凭借其对生产要素、技术专利、品牌等排他性的独占权和市场势力而所获得的高额利润。♣

9.7.3 有效竞争与反垄断

有效竞争是既有利于维护竞争又有利于发挥规模经济作用的竞争格局。具体含义包括：

1. 从短期来看，如果现实的市场背离了完全竞争的某个条件，那么最好在其它条件下也有所背离，这样的竞争才是有效的；
2. 从长期来看，存在着潜在竞争和替代品竞争的产业或市场，就是有效竞争的产业或市场；
3. 工业生产经营的多样化也具有竞争的性质。

反垄断政策：

1. 预防形成垄断性市场结构的政策；
2. 对垄断性市场结构的事后调节政策；
3. 禁止或限制企业间的共谋、卡特尔和不正当的价格歧视，对欺骗、行贿和压制竞争者的行为进行裁定政策。

第三部分

阅读书单

第 10 章 笔者已读书单

在大学入学的第一堂课，陈卓老师要求我们大学四年读够 40 本书，但是并没有给出明确的书单。回首大学四年，我应该读了不少书，达到了这个标准，但是仍然觉得书读得不够多，不够广。本章汇总了本书三位作者在本科期间读过的有价值的书籍，按照内容分类为经济理论与实践、研究与写作、哲学与文化、程序设计和其他，其中大部分书籍的阅读难度并不高，但是要掌握好阅读的方法，才能享受书籍带来的精神享受和知识收获。特别推荐读者优先阅读《如何阅读一本书》，该书是关于阅读的方法论的经典之作。读者不必苛求完成这个书单，我们提供书单的目的是方便读者在不知道读什么书和想了解某个问题但缺少资源的时候可以在本书单中找到参考。

10.1 经济理论与实践

1. 《微观经济学十八讲》平新乔
2. 《微观经济学：现代观点》范里安
3. 《金融学》黄达、张杰
4. 《新卖桔者言》张五常
5. 《理解现代经济学》钱颖一
6. 《中国是部金融史》陈雨露、杨忠恕
7. 《数理经济学的基本方法》蒋中一
8. 《德鲁克全书》彼得・德鲁克
9. 《量化投资：策略与技术》丁鹏
10. 《量化投资分析与策略——基于中国股票市场》肖刚
11. 《金融经济学二十五讲》徐高
12. 《宏观经济学二十五讲》徐高
13. 《宏观大势与市场逻辑》海闻
14. 《解读中国经济》林毅夫
15. 《置身事内》兰小欢
16. 《结构性改革》黄奇帆
17. 《魔鬼经济学》斯蒂芬・列维特
18. 《一本书读懂财报》肖星
19. 《金字塔原理大全集》芭芭拉・明托
20. 《估值建模》诚迅金融培训公司
21. 《投行职业进阶指南：从新手到合伙人》王大力

10.2 研究与写作

1.《经济学研究入门指南》史蒂文 • A. • 格林劳
2.《实证论文写作八讲》刘西川
3.《学术论文写作手册》Anthony，C.，Winkler，McCuen-Metherell
4.《芝加哥大学论文写作指南》杜拉宾
5.《如何成为学术论文写作高手》史帝夫 • 华乐丝
6.《原因与结果的经济学》中室牧子、津川友介
7.《发论文拿项目其实很简单》老踏
8.《博雅光华》张志学、徐淑英
9. *On Writing Well* William Zinsser
10.《如何阅读一本书》莫提默 • J. 德勒、查尔斯 • 范多伦
11. *The Elements of Style* William Strunk Jr.
12.《写作是门手艺》刘军强
13.《研究是一门艺术》Wayne C.Booth

10.3 哲学与文化

1.《乡土中国》费孝通
2.《文明的冲突》亨廷顿
3.《问学 • 与北大学生谈中国文化》余秋雨
4.《中国文化课》余秋雨
5.《有教养的中国人》孙正聿
6.《哲学通论》孙正聿
7.《菊与刀》鲁思 • 本尼迪克特
8.《大败局》吴晓波
9.《雪国》川端康成
10.《不能承受的生命之轻》米兰 • 昆德拉
11.《世说新语》
12.《莫泊桑短篇小说集》居伊 • 德 • 莫泊桑
13.《契诃夫短篇小说集》安东 • 巴甫洛维奇 • 契诃夫
14.《欧 • 亨利短篇小说集》欧 • 亨利
15.《苏东坡传》林语堂
16.《马背上的男孩》路伯特 • 伊萨克森
17.《红星照耀中国》埃德加 • 斯诺
18.《乞力马扎罗的雪》海明威
19.《再见，美丽堡》罗恩 • 莱瑟姆

20.《悟空传》今何在
21.《莎士比亚喜剧悲剧集》威廉・莎士比亚
22.《环球人物》杂志
23.《看天下》杂志
24.《中国景色》单之蔷
25. *Tuesdays with Morrie* Mitch Albom
26. *Harry Potter* 全集 J.K. Rowling
27. *The Chronicles of Narnia* 全集 C.S. Lewis
28.《月亮与六便士》毛姆
29.《刀锋》毛姆
30.《巴黎圣母院》维克多・雨果
31.《高老头》巴尔扎克
32.《欧也妮葛朗台》巴尔扎克
33.《百年孤独》加西亚・马尔克斯
34.《傲慢与偏见》简・奥斯汀
35.《简爱》夏洛蒂・勃朗特
36.《飘》玛格丽特・米切尔
37.《活着》余华
38.《杀死一只知更鸟》哈珀・李
39.《你当像鸟飞往你的山》塔拉・韦斯特弗
40.《撒哈拉的故事》三毛
41.《万水千山走遍》三毛
42.《我的灵魂骑在纸背上》三毛
43.《荆棘鸟》考琳・麦卡洛
44.《追风筝的人》卡勒德・胡赛尼
45.《当呼吸华为空气》保罗・卡拉尼什
46.《我与地坛》史铁生
47.《成为》米歇尔・奥巴马
48.《你想活出怎样的人生》吉野源三郎

10.4 程序设计

1.《计量经济分析方法与建模》高铁梅、王金明、陈飞、刘玉红
2.《金融风险建模及投资组合优化——使用 R 语言》伯恩哈德・拜福
3.《计量经济学及 Stata 应用》陈强
4.《高级计量经济学及 Stata 应用》陈强
5.《Python 金融大数据分析》YvesHilpisch

6.《R 数据科学》哈德利·威克姆、加勒特·格罗勒芒德
7.《MATLAB 编程》Stehen J. CHapman
8.《一份（不太）简短的 LaTeX 介绍》Tobias Oetiker, Hubert Partl, Irene Hyna and Elisabeth Schlegl
9. *STATA: A Really Short Introduction* Felix Bittman

10.5 其他

1.《掌控谈话》Chris Voss、Tahl Raz
2.《演讲的力量》克里斯·安德森等
3.《李银河说爱情》李银河
4.《刻意练习》安德斯·艾利克森、罗伯特·普尔
5.《沟通圣经：听说读写全方位沟通技巧》尼基·斯坦顿
6.《心理学导引》田媛
7.《传染：为什么疾病、金融危机和社会行为会流行》亚当·库哈尔斯基
8.《人类简史》尤瓦尔·赫拉利
9.《当下的力量》埃克哈特·托利

参考文献

[1] N. 格里高利 • 曼昆. 宏观经济学 [M]. 第 10 版. 卢远瞩译. 北京: 中国人民大学出版社, 2020.

[2] 陈强. 计量经济学及 Stata 应用 [M]. 北京: 高等教育出版社, 2015.

[3] 戴维 • 罗默. 高级宏观经济学 [M]. 第 5 版. 吴化斌, 龚关译. 上海: 上海财经大学出版社, 2021.

[4] 哈尔 R. 范里安. 微观经济学：现代观点 [M]. 第 9 版. 费方域, 朱保华译. 上海: 格致出版社, 2015.

[5] 李子奈, 潘文卿. 计量经济学 [M]. 第 5 版. 北京: 高等教育出版社, 2020.

[6] 逄锦聚. 政治经济学 [M]. 第 6 版. 北京: 中国人民大学出版社, 2018.

[7] 钱颖一. 理解现代经济学 [M]. 上海: 东方出版中心, 2021.

[8] 斯蒂芬 • P • 罗宾斯, 玛丽 • 库尔特. 管理学 [M]. 第 13 版. 刘刚, 程熙鎔, 梁晗译. 北京: 中国人民大学出版社, 2017.

[9] 伍德里奇. 计量经济学导论 [M]. 第 4 版. 北京: 中国人民大学出版社, 2010.

[10] 余章宝. 西方经济学理论的经验论哲学基础 [J]. 哲学研究, 2007: 121-126.

[11] 詹姆斯 • D • 汉密尔顿. 时间序列分析 [M]. 北京: 中国人民大学出版社, 2015.