
High School Mathematics Done Right

中学数学这样学

吴金闪



为了理解数学而写：知识简单，思考深刻，体现数学
从现实到结构，从结构到现实
思维的语言，严密的论证，纯粹理性的愉悦
有系统，有联系，知识网络
教的更少，学得更多

List of changes

批注清单

1	除法竖式需要改成中国模式，目前是法国模式。	39
3	考虑是否印单独一页	51
4	编制一些集合交和并运算，补集运算的习题。	125
5	编制一些三段论的应用题	149
6	编制一些反证法的应用题	151
7	添加参考文献	152
8	添加参考文献	152
10	这张图折页印刷，一定要够大。	194
11	平面几何概念地图，本书自己的，其他体系的用其他颜色显示	314
13	这张图采用折页打印，要足够大。	315
15	这张图折页印刷，一定要够大。	349
16	增加选择最优检测顺序的文献。	403

目录

第零章 对小学数学的要求和复习	29
0.1 对小学数学的要求	29
0.2 小学数学复习	33
0.2.1 从数数到数和运算律	34
0.2.2 现实和数学, 从关系到数学结构	43
0.2.3 数学知识的系统性	44
0.2.4 从小学数学学习和复习到理解型学习	50
0.3 小学数学知识总结以及本章小结	50
 第一部分 初中篇	 53
第一章 数轴上的点和数	55
1.1 集合的复习	55
1.2 数的复习——运算封闭性和数的出现	58
1.2.1 从 $\sqrt{2}$ 不是分数到 $\sqrt{\frac{q}{p}}$	61
1.2.2 无理数以及开方数之外的无理数	62
1.2.3 等式的性质	62
1.2.4 绝对值、相反数、倒数和运算律	64
1.3 数轴和用数轴表示数	65
1.3.1 非负整数的数轴表示	65
1.3.2 分数的数轴表示	65
1.3.3 无理数的数轴表示?	67
1.3.4 实数, 以及数轴上还有其它数吗?	67
1.3.5 数轴上的区间	69

1.4	运算在数轴上的表示 *	69
1.5	把数和运算合起来表示操作 *	71
1.6	平面直角坐标系上的点	73
1.7	代数的思想——用符号表示对象	76
1.8	推荐阅读材料	79
1.9	作业	79
1.10	本章小结	79
第二章	幂运算	81
2.1	数的有理数次幂运算	81
2.2	数的无理数次幂	83
2.3	关于平方和立方的几个公式	84
2.4	推荐阅读材料	86
2.5	作业	86
2.6	本章小结	87
第三章	变量和常数	89
3.1	完全已知的数到给定的任意的数	89
3.2	未知的或者重点关注的数	90
3.3	单项式和多项式的元数和阶数	92
3.4	幂次方多项式这么特殊吗?	93
3.5	代数式的分类	95
3.6	推荐阅读材料	96
3.7	作业	96
3.8	本章小结	96
第四章	因式分解	97
4.1	为什么需要因式分解	97
4.2	简单因式分解	98
4.3	几种类型混合的因式分解 *	99
4.4	因式分解的应用	101
4.5	推荐阅读材料	104

4.6	作业	104
4.7	本章小结	105
第五章	一元一次方程和一元一次不等式	107
5.1	为什么要学习一元一次方程和一元一次不等式?	107
5.2	一元一次方程的求解	108
5.3	一元一次方程的应用	109
5.4	一元一次不等式的求解	109
5.5	一元一次不等式组的求解	113
5.6	一元一次不等式的应用	114
5.7	推荐阅读材料	114
5.8	作业	114
5.9	本章小结	114
第六章	集合初步	117
6.1	集合的朴素概念	117
6.2	集合的交和并	120
6.3	子集、全集和补集	122
6.4	不等式的解当作集合	125
6.5	推荐阅读材料	125
6.6	作业	126
6.7	本章小结	126
第七章	映射初步	129
7.1	从命名、分类、操作到映射	129
7.2	映射	131
7.3	映射的复合, 最简单的群和表示理论	133
7.4	三个元素的集合到自身的映射 *	137
7.5	推荐阅读材料	141
7.6	作业	141
7.7	本章小结	141
第八章	逻辑学初步	143

8.1	从集合到命题	143
8.2	什么是和为什么需要逻辑?	144
8.3	逻辑三律和演绎逻辑	147
8.3.1	三段论	148
8.3.2	逆否命题的等价关系和反证法	149
8.3.3	逻辑变量的运算	151
8.4	逻辑是共通的普适的吗?	152
8.5	逻辑和批判性思维	152
8.6	逻辑和数学、逻辑和科学	155
8.7	推荐阅读材料	157
8.8	作业	157
8.9	本章小结	157
第九章	一元函数	159
9.1	一元函数当作一个吞吐机	159
9.2	一元函数当作一个关系	160
9.3	平面直角坐标系	161
9.4	一元函数当作一个图形	162
9.5	一元一次函数	163
9.6	一元一次方程的几何含义	164
9.7	一元二次函数和一元二次方程	166
9.8	负一次方函数	170
9.9	为什么需要函数?	171
9.10	推荐阅读材料	173
9.11	作业	173
9.12	本章小结	174
第十章	二元一次方程和二元一次不等式	175
10.1	为什么要学习二元一次方程和二元一次不等式	175
10.2	换元法: 化二元一次方程为一元一次方程	177
10.3	消元法: 等式的计算	178
10.4	二元一次方程的几何含义	179

10.5 二元一次方程的应用	180
10.6 二元一次不等式组的求解和几何含义	180
10.7 二元一次不等式组的应用	181
10.8 推荐阅读材料	181
10.9 作业	181
10.10 本章小结	181
第十一章 一元二次方程和一元二次不等式	183
11.1 为什么要学习一元二次方程和一元二次不等式	183
11.2 一元二次方程的几何含义	185
11.3 配方法求解一元二次方程	186
11.4 韦达定理和因式分解法求解一元二次方程	187
11.5 关于三阶和更高阶的多项式方程	188
11.6 一元二次方程的应用	189
11.7 一元二次不等式的几何含义和求解	189
11.8 一元二次不等式组的应用	191
11.9 推荐阅读材料	191
11.10 作业	191
11.11 本章和初中代数部分小结	191
第十二章 锐角三角函数	195
12.1 复习: 勾股定理	195
12.2 从直角三角形的等比例扩大到相似、全等和三角函数的定义	199
12.3 锐角的正弦余弦函数 *	202
12.4 一般三角形的相似以及三角形中的正弦定理和余弦定理	204
12.5 三角函数的应用	210
12.6 推荐阅读材料	210
12.7 作业	210
12.8 本章小结	210
第十三章 平面几何: 基本定义	211
13.1 点和距离的定义	211

13.2 圆和角的定义	218
13.2.1 半圆、直径、平角和直角	221
13.2.2 对顶角、同位角、内错角和同旁内角以及它们之间的 关系、关系的关系	226
13.2.3 直线和坐标轴相交的角度	231
13.3 三角形和四边形的定义	232
13.4 推荐阅读材料	234
13.5 作业	234
13.6 本章小结	234
第十四章 平面几何：平行初步	235
14.1 平行的定义、平行公理	235
14.2 平行基本定理	237
14.3 推荐阅读材料	240
14.4 作业	240
14.5 本章小结	240
第十五章 平面几何：三角形	241
15.1 三角形定义和为什么研究三角形?	241
15.2 三角形的内角和	242
15.3 三角形的全等	245
15.4 两条直线相交的夹角	249
15.5 三角形的相似	251
15.6 相似三角形的应用——证明勾股定理	256
15.7 全等三角形的应用——尺规作图	257
15.8 全等三角形的应用——等腰三角形	264
15.9 推荐阅读材料	266
15.10 作业	266
15.11 本章小结	266
第十六章 平面几何：平行	269
16.1 平行线的判定和性质	270

16.2 从平行线和全等三角形看相似三角形	275
16.3 本书平面几何的知识体系	278
16.4 平行和全等三角形的应用——四边形	279
16.5 推荐阅读材料	285
16.6 作业	285
16.7 本章小结	285
第十七章 平面几何：圆	287
17.1 圆的定义和平移	287
17.2 圆的旋转 *	291
17.3 圆中三角形的相关定理	293
17.4 共圆定理：把不是圆的问题变成圆的问题	296
17.5 推荐阅读材料	299
17.6 作业	299
17.7 本章小结	300
第十八章 平面几何：解析方式	301
18.1 代数表达式和几何图形	301
18.2 通过把几何问题转变成代数表达式来解决几何学问题	306
18.3 通过把代数表达式转变成几何问题来解决代数问题	308
18.4 推荐阅读材料	310
18.5 作业	310
18.6 本章小结	310
第十九章 平面几何：综合	313
19.1 数学解题 WHWM 四问和概念地图用于几何证明	313
19.2 几个平面几何知识体系赏析	313
19.2.1 Euclid 体系	313
19.2.2 Hilbert 体系	313
19.2.3 MSG 体系	313
19.2.4 本书的体系	314
19.3 推荐阅读材料	314

19.4 作业	314
19.5 本章小结	314
第二十章 概率论初步	317
20.1 古典概型的例子	317
20.2 事件和事件集合的运算	318
20.3 概率：从事件集合到 $[0, 1]$ 的映射	321
20.4 概率分布	323
20.5 独立事件和互斥事件	324
20.6 初识条件概率	325
20.7 概率论的应用	328
20.8 推荐阅读材料	330
20.9 作业	330
20.10 本章小结	331
第二十一章 统计学初步	333
21.1 经验分布图的绘制	333
21.2 描述性统计	338
21.3 统计推断和人类知识	343
21.4 条件概率用于统计推断	344
21.5 统计学的应用	347
21.6 推荐阅读材料	347
21.7 作业	347
21.8 本章小结	348
第二十二章 排列组合	351
22.1 组合和排列	351
22.2 排列组合的应用	352
22.3 推荐阅读材料	355
22.4 作业	355
22.5 本章小结	355
第二十三章 逻辑学和数学论证	357

23.1 命题	357
23.2 定义	362
23.3 论证	363
23.3.1 逻辑三律的复习	364
23.3.2 命题关系当作命题以及反证法	365
23.3.3 三段论	367
23.3.4 数学归纳法	370
23.3.5 不完全归纳法	373
23.4 推荐阅读材料	375
23.5 作业	375
23.6 本章小结	375
第二十四章 数学建模	377
24.1 什么是数学建模和数学模型	378
24.2 从数学建模的角度看数学知识的发展	379
24.3 从把数学知识用于描述现实看真正的数学应用题——WHWM 求解数学应用题	382
24.4 数学建模五步和 WHWM 四问及其应用	396
24.5 推荐阅读材料	404
24.6 作业	404
24.7 本章和本卷小结	405
第二部分 高中篇	409
第二十五章 函数之幂函数、指数函数、对数函数	411
25.1 幂函数	411
25.2 指数函数	411
25.3 对数函数	411
25.4 函数和反函数的图形和表达式	411
25.5 推荐阅读材料	411
25.6 作业	411
25.7 本章小结	411

第二十六章 函数之半圆、半椭圆函数	413
26.1 半圆函数	413
26.2 半椭圆函数	413
26.3 圆	413
26.4 椭圆	413
26.5 双曲线	413
26.6 推荐阅读材料	413
26.7 作业	413
26.8 本章小结	413
 第二十七章 函数之三角函数	 415
27.1 余弦函数	415
27.2 余弦定理	415
27.3 正弦余弦函数关系	415
27.4 正切反切函数	415
27.5 周期函数	415
27.6 推荐阅读材料	415
27.7 作业	415
27.8 本章小结	415
 第二十八章 平面和空间矢量	 417
28.1 直角坐标系和矢量的分量形式	417
28.2 矢量加法和数乘	417
28.3 矢量内积	417
28.4 坐标变换和描述坐标变换的矩阵	417
28.5 矢量的抽象记号语言	417
28.6 矢量的应用	417
28.7 推荐阅读材料	417
28.8 作业	417
28.9 本章小结	417
 第二十九章 矩阵和量子力学——数学建模	 419

29.1 概率论的矩阵描述	419
29.2 量子现象	419
29.3 量子现象的经典波解释及其问题	419
29.4 量子现象的矩阵和矢量描述	419
29.5 解释更多的实验现象?	419
29.6 推荐阅读材料	419
29.7 作业	419
29.8 本章小结	419

第三十章 复数 421

30.1 负数的开平方	421
30.2 负数的开三次方	421
30.3 复数当作矢量使用	421
30.4 Euler 公式	421
30.5 复数的加法和乘法	422
30.6 复数的应用	422
30.7 推荐阅读材料	422
30.8 作业	422
30.9 本章小结	422

第三十一章 不等式初步 423

31.1 几个恒成立不等式	423
31.2 从恒成立不等式到函数极值	423
31.3 不等式和极值的应用	423
31.4 推荐阅读材料	423
31.5 作业	423
31.6 本章小结	423

第三十二章 导数和极值初步 425

32.1 导数的定义初步	425
32.2 简单函数的导数	425
32.3 函数的导数用于求极值	425

32.4 导数和极值的应用	425
32.5 推荐阅读材料	425
32.6 作业	425
32.7 本章小结	425
第三十三章 离散变量的概率论和统计学	427
33.1 从观测数据得到频率再到概率	427
33.2 从已知离散概率到观测数据的频率	427
33.3 用 Monte Carlo 和 Bootstrap 配合数值计算来沟通频率和概率	428
33.4 概率论和统计学的应用	428
33.5 推荐阅读材料	428
33.6 作业	428
33.7 本章小结	428
第三十四章 立体几何	429
34.1 空间直线的平行和垂直	429
34.2 空间平面的平行和垂直	429
34.3 推荐阅读材料	429
34.4 作业	429
34.5 本章小结	429
第三十五章 表示论：数学是对操作的研究	431
35.1 当作对象属性或者过程的数	431
35.2 数当作加法的表示	431
35.3 数当作乘法的表示	431
35.4 数和矩阵当作旋转的表示	431
35.5 数学是对操作的研究对关系的研究	431
35.6 把数学看作是操作的研究的意义	432
35.7 推荐阅读材料	432
35.8 作业	432
35.9 本章小结	432
第三十六章 数学知识网络和学科大图景	433

目录	15

36.1 数学一二层知识的系统性	433
36.2 数学学科大图景, 及其和一二层知识的联系	433
36.3 从数学知识体系到科学知识体系	433
36.4 高等数学学什么	433
36.5 推荐阅读材料	434
36.6 作业	434
36.7 本章小结	434

第三十七章 再论数学建模	435

37.1 从数学建模的角度看数学知识的发展	435
37.2 从把数学知识用于描述现实看真正的数学应用题	435
37.3 数学建模五步及其应用	435
37.4 推荐阅读材料	435
37.5 作业	435
37.6 本章和本卷小结	435

参考文献	437

名词索引	439

人名索引	441

插图目录	446

举例目录	448

作业目录	452

献给

吴立心、吴逸兮，以及所有的中学生们。愿你们永远可以体验到思考的快乐，讲道理的快乐，把知识搞通透的快乐，运用知识解决问题的快乐，而不是被教“就是这样，你记住”，“你只需要会做题”。

致谢

序

前言

本书是《小学数学这样学》[1] 的后续篇，也是“人类知识高速公路上以培养掌握高层知识生成器的创造者为目标的理解型学习”（简称“理解型学习”[2]）应用于具体学科的另一例子。在这里我们把中学，包含初级中学和高级中学，的数学知识做了梳理，用新的方式做了反映数学学科大图景的呈现。一个学科的学科大图景指的这个学科的典型研究对象、典型研究问题、典型思维方式、典型分析方法。以及这个学科和世界以及其他学科的关系。在这个新的呈现里面，我们强调了学习是为了创造知识、创造性地使用知识，以及欣赏知识的创造和创造性使用，并且企图用理解型学习的方式来通过具体知识的学科之中体会好数学学科大图景，从而达成上面的学习的目的。

我们来补充一点点关于理解型学习的知识。在这个学习方法之中，最重要概念是知识的 $1+5$ 个层次：第零层前知识，经验和体验；第一层，事实性程序性知识；第二层，学科概念知识；第三层，学科大图景知识；第四层，一般性人类思维；第五层，教和习的方法。在学习三角形之前就看过摸过三角形，在学习加法之前就在生活中把两堆东西加起来数过，这成为经验和体验。乘法口诀、三角函数公式、题型和解题步骤口诀属于第一层知识。乘法是重复加法的简便记号、三角形是三条线段首尾相连构成的封闭图形、三角函数的定义以及各个三角函数之间的关系属于第二层知识。数学是思维的语言，加法来自于对“合起来数一数”的抽象，数来自于对“数 + 量 + 物”的模式抽象，每个题型及其解题步骤的梳理总结的过程（提出或者被提出问题；找到这个问题中的对象，识别对象之间的关系，把关系转变成数学表达式，把所要解决的问题变成数学问题；求解问题；检验答案；进一步推广到一类问题。有的时候也被称为数学建模五步，或者简称数学

五步 [1]) 属于第三层知识。例如从加法 (和起来数一数, 从经验和体验到学科概念) 和乘法 (重复的加法, 为了更加简便的语言和记号) 的概念的定义方式体会到抽象这个思维方式在数学中的重要性, 体会到概念之间的联系, 体会到不同层次知识之间的联系, 进一步看到系联性思考 (看到联系, 通过联系从个体看到整体, 从整体的角度来看个体), 从证明中体会到批判性思维 (在数学上, 任何公理以外的命题必须经过我自己的严格证明, 才能被我接受), 进而把抽象、系联性思考、批判性思维迁移到其他学科的知识的理解和创造之中, 属于第四层知识。通过知识之间的联系从而实现上下左右贯通来学习, 通过具体知识的学习体会到学科大图景, 以及从是否能够把学科大图景体现好的角度来看每一个具体知识的学习, 是有关“理解型学习”这个教和习的方法的知识 (甚至, 从这样通过具体数学知识的学习体会到理解型学习, 从理解型学习的角度来设计具体数学知识的学习, 也是理解型学习的例子), 属于第五层知识。

更进一步, 我们把这样的通过不同层知识之间的关系来学习和思考, 也就是从对相对低层知识的总结提炼中学习、体会甚至创造相对高层知识, 从相对高层知识来指导相对低层知识的学习、理解和创造, 称为上下贯通; 我们把通过同一层知识之间的联系来理解好这一层知识, 以及把来自于一个领域的高层知识用于学习着解决或者创造着解决另一个领域的问题称为左右贯通。我们可以借助概念地图 (概念和概念通过连词相连, 例如“加法—是—合起来数一数”、“合起来数一数—基于—数”、“合起来数一数—基于—数数”、“数—来自于—对‘数 + 量 + 物’这个模式的抽象—体现了—抽象”) 来呈现和运用上下左右贯通。各个学科的概念地图的总和被称为人类知识高速公路。正是由于第三层和第四层知识具有帮助解决本领域或者其他领域问题, 创造知识的能力, 我们把它们称为高层知识生成器。

因此, 知识的层次中的第三、四、五层知识, 也就是高层知识生成器; 人类知识高速公路, 也就是各个学科的概念地图的总和; 上下左右贯通, 也就是从低层知识中总结体会到高层知识, 从高层知识来看低层知识, 运用高层知识来学习着或者创造着解决本领域或者其他领域的问题从而学习或者创造知识, 是理解型学习最重要的核心概念和理念。更多信息可参见吴金闪的《教的更少, 学得更多》[2]。

因此, 本书的第二个目的就是用理解型学习来帮助学习者学科中学数

学，从而也学会理解型学习，进而用理解型学习来学习其他学科，以及给创造和创造性使用各个学科的知识做好准备。

在这样的一个定位和写作意图的过程中，我们设想的直接的读者是教师、家长和教育管理者。希望这些读者能够从本书所选择的教什么、怎么教之中体会到理解型学习，进而来开展理解型学习指导下的课程教学、课程设计和考察方式。但是，为了达到这个目的，我们希望能够通过“做中学、创造中学”来帮助教师、家长和教育管理者，也就是从看如何把理解型学习的理念和工具用于实际课程的设计来领会理解型学习，而不是理解型学习的理念和概念的直接传达。因此，我们打算通过在理解型学习的指导下来具体编撰这样一本中学数学教材，当例子，来帮助这些读者得到更大的收获，而不是被说教。

另一方面，我们也希望写出来一本中学数学教材，能够帮助中学生对数学具体知识的理解更透彻，甚至能够进一步理解到数学的学科大图景，学会像一个数学家一样来思考，或者至少能够欣赏数学家的思考，而不是仅仅学会了一堆算来算去的知识。这就要求我们第一、把数学材料梳理好，突出联系，可以帮助学习者更容易更透彻地实现上下左右贯通，看到数学学科大图景；第二、帮助学习者学会理解型学习，甚至在其中一部分知识的学习上，设计一些问题来做“创造体验式学习”——通过“假装着”把所要学习的知识创造出来而学会这个知识，从而更高地为将来创造知识做准备。

因此，本书的前言分成了“写给中学生们的”和“写给教师、家长和教育管理者的”两个部分。请读者们各取所需。

写给中学生们的

上面是写给中学生们的前言。下面是写给教师、家长和教育管理者的前言。

写给教师、家长和教育管理者的

本书所体现的数学学科大图景

前面我们提到了本书的核心思想通过“人类知识高速公路上以高层知识生成器为目标的理解型学习”来实现“教和学的更少，学得更多”。对于数学学习，这里的高层知识生成器指的是第三层知识数学学科大图景和第四层超越学科的一般性人类思维，以及教和学的方法。为了实现这个“人类知识高速公路上以高层知识生成器为目标的理解型学习”，我们大概需要完成以下步骤 [2]：

1. 梳理这个学科的学科大图景；
2. 选择体现这个学科大图景的核心概念、概念间关系来构建学科概念网络；
3. 选择合适的案例来体现每个概念的提出和运用，促进对概念的理解；
4. 标注书籍、论文、项目、案例、习题等到这个学科概念网络上；
5. 结合概念网络上的分析算法（例如学习顺序算法、自适应诊断性检测算法）来设计和开展教学实践。

这其中最关键的就是对学科大图景的梳理。那么，数学学科大图景，也就是数学的典型研究对象、典型研究问题、典型思维方式、典型分析方法、和世界还有其他学科的关系，到底是什么呢？我们在《小学数学这样学》[1]中有专门的一章来展开讨论。在这里，我们把其中的主要意思复述一下。更完整的版本请参考吴金闪的《小学数学这样学》[1]。

- 数学观
 - 数学是思维的语言
 - 数学是描述世界的语言，“双向”（从现实提炼数学结构，把数学结构用于描述现实）数学建模
 - 数学是研究关系和模式的学科

- 数学思维
 - 抽象是数学的核心思维
 - 批判性思维在数学中的地位：证明和计算是数学最重要的论证方式
 - 表示数学结构和研究表示间的转换是数学研究的目标和研究方式
 - 分解和综合是数学结构研究的典型方式
 - 数学知识有系统：公理化体系，层次性和跨越层次的联系，不同数学分支的融合
 - 化归的思想
- 分析方法
 - 数学建模五步
 - 数学解题 WHWM 四问，及其嵌套合并
 - 试错和分析错误是数学的重要分析方法之一
 - 近似是数学重要分析方法之一
 - 微积分——累积量和变化率的关系
 - 无穷小和极限的概念
 - 连续性及其分析方法
 - 变换和不变量的思想和相应的分析方法
 - 约束下求极值
- 最基础的概念（其他高级结构的基础）
 - 集合和映射
 - 范畴论
 - 数理逻辑：什么是演绎证明，为什么能用

本书会围绕以上内容来决定教什么怎么教，在可能的情况下把数学学科大图景体现在具体知识的学习之中，把理解型学习体现在数学的学习之中。

其实，这个具体知识和学科大图景以及超越学科的一般性人类思维还有教和学的方法的关系，就是系统思维 [3]：通过直接和间接联系，从个体看到整体，从整体的角度来看个体。我们把它叫做“系联性思考”。不仅仅是学科学习，在面对任何系统的任何问题的时候，系联性思考——也就是通过把这个系统在各个层次做分解和综合，看到在各个层次上各个部分之间的联系，进而从整个系统的整体功能的角度来看其每一个部分的作用和地位，从这个系统的每一个部分来看到这个系统的整体的功能——往往可以给我们提出和解决问题的启发。而且，更进一步，系统科学还在此思维方式的基础上，发展了一些分析方法，解决了一些典型问题 [3]。因此，我们也希望通过这本书，你不仅仅学会数学，体会到数学学科大图景，学会理解型学习，还能体会到系统思维的核心“系联性思考”，而且都是通过做中学、创造中学，而不是通过这些高层知识的被说教，或者被强迫记住，来学习。

第零章 对小学数学的要求和复习

在我们开始中学数学的学习之前，我们希望你最好有下面的小学数学的知识和思维的基础。如果没有这个基础，你可以先试着学下去，遇到问题了再想办法补上，或者你也可以先去学习《小学数学这样学》[1]，然后再来看本书。

0.1 对小学数学的要求

首先，我们希望你知道的第一、二层知识有：

1. 认识非负整数（正整数和零），会数数；
2. 会整数的加法和乘法的计算，理解整数的加法和乘法的含义；
3. 会整数的减法和除法的计算，理解整数的减法和除法的含义；
4. 整除、因数、倍数、最大公约数、最小公倍数的概念和简单计算；
5. 从非负整数通过不断的减法拓展到包含负的整数的所有整数；
6. 通过除法从整数拓展到分数，以及分数的四则运算；
7. 从分数到等价的循环小数（包含有限小数——循环位是 0，和无限循环小数），以及小数的四则运算；
8. 平方、立方，开平方，开立方的含义和简单计算；
9. 从 $\sqrt{2}$ 不是分数，初步了解无理数（无限不循环小数）的存在性；

10. 可以用十进制来表达数，通过数十进制表达来理解乘法竖式计算；
11. 整数、分数、循环小数的数轴表示；
12. 负数的数轴表示，负数和正数在数轴上的关系；
13. 针对具体数的加法和乘法（减法和除法分别通过负数和分数成了加法和乘法）的运算律，以及通过运算律来给算式做变形；
14. 用字母表示数，从关注算出来的得数，到关注代表事物的数量的字母之间的关系，以及关系对应的数学操作；
15. 看得懂用字母表示的加法和乘法的运算律；
16. 等式的性质；
17. 不等式的性质；
18. 方程的思想，以及通过等式的性质来求解一元一次方程；
19. 三角形、四边形（包含其特殊情况梯形、平行四边形、长方形、正方形）的内角、周长、面积的定义和计算，圆的周长和面积的定义和计算；
20. 长方体、正方体的表面积和体积的定义和计算；
21. 具有等几率基本事件（古典概型）的事件的概率的定义和简单计算；
22. 最基本的描述统计（均值、直方图）的定义和计算；
23. 把以上知识用于求解应用题，或者自己提出和解决日常生活中的问题。

以上知识中，如果有还没有学到的部分，可以尝试从已经学过的知识来自己想出来，或者去从《小学数学这样学》[1]中学习。针对其中的下列可能一部分学生没有理解到位的部分，我们也会做一个复习：负数、分数和循环小数的关系以及无限不循环小数、字母表示的运算律、方程的思想以及通过等式的性质来求解一元一次方程（这部分我们会在本书的正文来重复学习），等式的性质和不等式的性质。这是我们学习中学数学的基础。

但是,更加重要的是,获得这些第一二层知识的过程,或者说这些一二层知识之间的联系及其背后的三四层知识。具体来说,

1. 从“数 + 量 + 物”经过“抽象”,也就是忽略“量”和“物”(从而适用于任何“量”和“物”),而得到非负整数;
2. 更进一步从上面的非负整数的抽象过程,领会到数学概念、数学思维和现实的关系——受现实启发;
3. 从“合起来数一数”到加法,从重复多次的加法到乘法,体现了数学对重复模式的抽象:抓住重复模式的核心并给个名字;
4. 建立加法和减法,乘法和除法之间的互为逆运算的关系,进一步对比从加法到乘法、从减法到除法的这两个逻辑关系的关系;
5. 从加减乘除的定义、相互关系,体会到逻辑关系在数学中的核心地位,体会到数学知识的系统性;
6. 从整个非负整数到包含负整数的所有整数、分数、小数的过程,体会到运算的封闭性的重要性,结合出现这些运算和数的场景,体会到数学和现实的关系——现实使得这些运算会出现,于是要求数学给这些运算的结果一个表示;
7. 从整数和小数的加减乘除四则运算总结出来有且只有加法和乘法的四则运算规律的过程,体会到数学知识的系统性,体会到抽象、模式发现和总结这样的思维在数学中的重要性;
8. 从分数和循环小数的等价性的证明中体会到逻辑思维在数学中的重要地位;
9. 从把运算封闭性运用于平方和平方根从而得到 $\sqrt{2}$ 的含义和存在性的过程,体会到上下贯通,也就是从之前的学习中得到了运算的封闭性,现在把运算的封闭性用于新的运算;
10. 从把运算封闭性运用于平方和平方根从而得到 $\sqrt{2}$ 的含义和存在性的过程,结合实际场景(面积等于 2 的正方形的边长),体会到数学和现实的关系——数学需要给现实提供描述它的数学结构;

11. 从 $\sqrt{2}$ 不是分数的证明中, 以及从 $\sqrt{2}$ 不是分数这个事实到提出无理数的过程中, 中体会到逻辑思维在数学中的重要地位;
12. 通过可以用数的十进制来理解乘法竖式计算步骤进一步领会到数学知识的系统性;
13. 通过数的数轴表示初步体会到数和形这些看起来不同的数学对象之间的关系, 从而更进一步体会到左右贯通——不同领域之间的知识之间的关系;
14. 依托于从四则运算中抽象总结出来四则运算律及其字母表示的过程, 更进一步地理解抽象在数学中的重要地位;
15. 从用字母表示任意的数代入算式反而集中体现了算式中的运算, 也就是关系, 这一点, 体会到数学开始从关注数, 到开始关注关系;
16. 从运用等式的性质来解一元一次方程, 进一步领会到数学知识的系统性;
17. 从几何初步进一步领会到数学知识的系统性, 而且是从数的知识到形的知识的迁移, 左右贯通这个系统性;
18. 在求解数学问题过程中, 不依靠题型和口诀, 而是每次去做数学解题 WHWM 四问: 这个问题里面有哪些东西 (What, 有什么), 这些东西之间具有什么关系 (是如何相互联系的, How), 这个联系如何用数学来表达 (How), 为什么它们之间有这个关系 (Why), 为什么这个关系对于解决这个问题是重要的 (Why), 通过解决了这个问题我学习到了什么体会到了什么, 这个问题对我来说意味着什么 (Meaningful);
19. 通过把数学用于提出和解决实际生活中的问题, 以及前面提到的这些数学概念和数学思维的提出和发展的过程, 领会到数学建模五步: 提出有关实际对象的问题, 把问题转化为数学问题, 求解数学问题, 通过实验或者实践来检验得到的答案, 对问题、答案、解答、这个过程的分析方法思维方式做总结提炼然后尝试系统化和推广;

20. 从以上的证明、抽象、解决实际问题等过程中体会到数学是思维的语言，数学是描述世界的语言。

如果更进一步，你已经掌握了用包含字母的算式来表示运算律，并且还能够从以下几条基本的运算律“加法结合律、加法交换律、乘法结合律、乘法交换律、乘法对加法的左侧分配律（乘法在左，加法在右）”推导出来（也知道推导是什么意思，也就是依靠定义和更加基本的已经知道是正确的命题，加上逻辑，来得到新的命题）更多的运算律，例如乘法对加法的右侧分配律，见例 0.1，那就更好了。如果你还进一步掌握了集合的概念，明白了实际上证明的过程依靠的是逻辑学当中的叫做演绎逻辑的东西，而这个东西的背后是集合那就更更好了。不过，这些知识——包括前面的列表中的一二层和三四层知识，以及这里的字母表示数和算式、带字母的算式表示运算律、从定义到基本运算律、从基本运算律到导出运算律（以及背后的知识的系统性）、演绎逻辑用于证明数学命题、演绎逻辑和集合的关系、集合我们都会在本章以及本书后续章节中再一次学习。

例 0.1 (乘法对加法的右侧分配律). 已知加法结合律、加法交换律、乘法结合律、乘法交换律、乘法对加法的左侧分配律，证明乘法对加法的右侧分配律 $(a + b) \times c = a \times c + b \times c$ 。

证明.

$$(a + b) \times c = c \times (a + b) \text{ (乘法交换律)} \quad (1)$$

$$= c \times a + c \times b \text{ (乘法对加法的左侧分配律)} \quad (2)$$

$$= a \times c + b \times c \text{ (乘法交换律)} \quad (3)$$

□

0.2 小学数学复习

这一章，我们从重新构建整个小学数学学科概念知识以及学科思维等学科大图景的角度来复习一下小学数学。如果通过这个复习，你发现仍然有不太能理解的地方，请参考《小学数学这样学》[1]。

0.2.1 从数数到数和运算律

从数数的经验体验，到自然数，非负整数，所有整数，分数，小数，无理数的存在，实数初步；从概念到基本运算律，到导出运算律；用字母表示数，用字母表示算式（算式是数）。

我们假设咱们的读者已经具有数 1, 2, 3 三个数的经验和体验，也就是看到篮子里有一个苹果，知道对应的符号是 1，单位是个；看到盘子里面有两块饼干，知道对应的符号是 2，单位是块；看到桌子周围坐了三个，知道对应的符号是 3，单位是个；看到餐桌上放了三双筷子，知道对应的符号是 3，单位是双，等等等等。从这个朴素的数的概念开始，我们每次给前面的数增加 1 来获得一个新的数——指的是含义上。知道含义之后，我们给每一个这些新增加的数一个名称，例如 $3 + 1$ 记作 4，读作四； $4 + 1$ 记作 5，读作五； $5 + 1$ 记作 6，读作六； $6 + 1$ 记作 7，读作七； $7 + 1$ 记作 8，读作八； $8 + 1$ 记作 9，读作九。原则上，我们还可以有 $9 + 1, 9 + 1 + 1, \dots$ 。但是，我们稍后再给它们一套记号和读法。也就是说，所有的自然数的含义都可以基于 1, 2, 3（其实，2, 3 都不用，只需要知道朴素的 1 的含义和加 1 的朴素含义——一个数后面加上 1 所代表的对应到实际对象的含义，不需要知道严格的 1 的定义和加法的定义），但是 9 以后的自然数的符号和名称我们会稍后再介绍。

有了 1, 2, 3, $\dots$, 9 这九个符号以后，我们来看这些符号在记号、比大小和加法上如何发挥作用。首先，我们说，有了每一个数所代表的在特定单位下的某种东西的数量的含义，我们就可以在同样的单位下把同质的东西和起来看数量。例如，一篮子里面装了 2 个苹果，另一个篮子里面装了 3 个苹果，把两篮子苹果装在同一个篮子里面，就得到这个篮子里面有 5 个苹果。在这里，实际上，我们都不用数数：我们把 5 的含义——在“个”当作单位来看“苹果”时候的数量——用于这个合起来的篮子里面的苹果，则自然就是 5 个，然后，我们发现，实际上这个两个篮子合到同一个篮子里面的操作，就是把两堆苹果合并。当以后，这样的合并的操作我们见多了，就会发现，背后的数之间存在着脱离开了对象的关系—— $2 + 3 = 5$ 。这个表达式，无论是苹果，还是饼干，还是筷子，无论是相应的个、块还是双，都是成立的。于是，我们把这个对应着合起来以后再看数量的操作定义为加法。

有了加法，也就是合起来数一数，自然我们就可以考虑反过来的操作，

整体中去掉一部分，被称为减法。减法也是可以超越对象和单位的，只要坚持同一个单位同质的对象。例如，从一个装了 3 个苹果的篮子里面拿出 2 个苹果，则剩下的是一个苹果，算式记作 $3 - 2 = 1$ 。同时，我们也能够理解，减法是加法的逆运算，反向操作——因为把剩下的和拿走的合起来的数量自然就是一开始的整体的数量。

一旦有了加法我们就可以不断地加下去，所以，我们可以在含义上得到从 1 开始的所有的自然数。类似地，一旦有了减法，我们就可以不断地减下去，从而得到从例如 3 开始的所有的数。我们试试： $3 - 1 = 2$ ， $2 - 1 = 1$ ， $1 - 1 = ?$ 。我们发现，我们还没有这个 $1 - 1$ 的含义的记号和读音。我们给这个含义发明一个记号和读音，记作 $1 - 1 = 0$ ，读作零。我们其实可以继续减下去，例如 $0 - 1$ ， $0 - 1 - 1$ ， $\dots$ 。于是我们就有了负的整数。对应到现实当中，负数当作具有参考点和方向的数的表示是最好理解的。例如我们规定某个温度叫做零度，则比零度大的用正数表示，比零小的用负数表示。例如，我们规定某个高度叫做零点，更高的用正数，更低的用负数。例如，我们如果规定往东为正方向用正数表示，那么往西就用负数表示。一旦有了负数，则减法就也成了加法，相当于加上一个负数。例如， $4 - 2 = 4 + (-2)$ 。如果我们把所有的正整数排列在一条直线（先不要管直线的精确定义，以后会知道）上，那么，自然地 0, 1 之间的间隔应该和 1, 2 之间，2, 3 之间等等相同。也就是说，0, 1, 2, 3, $\dots$ 排出来的是一条从零开始的有一个端点的延伸到某个方向（一般取向右为正整数增加的方向）的直线（后来我们会知道这叫做射线）。那么，这个时候，负数会出现在数轴的哪里呢？看过温度计就能够猜出来负数应该出现在相反的一侧，并且延伸出去，也就是一般来说，从 0 这个端点开始向左延伸。同时， $-1, 0$ 之前见的间隔应该和 0, 1 之间的相同， $-2, 0$ 之间的间隔应该和 0, 2 之间的相同。

现在，我们有了正负整数、零，有了加法，以及化归成了加法的减法。

有了加法，我们可以从加法的含义得到加法的结合律和交换律：对于合起来数一数，如果要把三堆东西合起来数那么先把哪两堆合起来数一数不影响得到总数；如果是把两堆东西合起来数一数，则先数清楚哪一堆也不影响得到的总数。例如， $(1 + 2) + 3 = 6 = 1 + (2 + 3)$ ， $2 + 3 = 5 = 3 + 2$ 。这些运算律，在将来我们学习了用字母表示数（任何一个数都可以代入到这个字母的位置上去）之后，就可以表示为 $(a + b) + c = a + (b + c)$ ， $a + b = b + a$ 。

有了整数和加法（注意减法已经通过负数统一为加法），我们来看乘法。当我们需要把多个同样数量的对象合起来数一数的时候，我们可以给这个“重复多次的加法”一个自己的名字——乘法。例如，一箱子羽毛球，每个箱子里面有 50 筒，每筒有 12 个羽毛球。我们可以算 49 次加法来得到羽毛球的总数，也可以用这个新的计算——乘法来表示和得到相应的规则，然后重复使用这个规则来完成羽毛球总数的计算。这个规则在个位数乘以个位数的时候被称为乘法口诀表。例如 $2 \times 3 = 6$ （二三得六）， $3 \times 7 = 21$ （三七二十一）。这些口诀本身是通过加法来得到的，例如 $3 \times 7 = 7 + 7 + 7 = 21$ 。

有了乘法，我们可以从乘法的含义得到乘法的结合律和交换律： $(a \times b) \times c = a \times (b \times c)$ ， $a \times b = b \times a$ 。在这里我们直接用字母来代笔了一个任意的整数¹。乘法的交换律可以通过把乘法算式 $a \times b$ 看作是一个平面上有排好的 a 行 b 列个同样的东西要合起来数一数来理解。这样一个平面上的所有东西，也可以通过两种方式来计算，得到相同的结果：第一种，把每一行看作一个单元——每一行里面有 b 个，总共有 a 行，于是总数为 $a \times b$ ；第二种，把每一列看作一个单元——每一列里面有 a 个，总共有 b 列，于是总数为 $b \times a$ 。乘法的结合律可以通过把乘法算式 $(a \times b) \times c$ 看作是一个三维的长方体上有排好的 a 行 b 列 c 层个同样的东西要合起来数一数来理解。这样一个长方体上的所有东西，也可以通过两种方式（其实还有第三种）来计算，得到相同的结果：第一种，先计算 a, b 所在的面，总共有 $a \times b$ 个，然后这样的面有 c 层，于是总数为 $(a \times b) \times c$ ；第二种，先计算 b, c 所在的面，总共有 $b \times c$ 个，然后这样的面有 a 层，于是总数为 $a \times (b \times c)$ 。如果你还是不习惯同字母表示数，你可以通过把字母换成具体的数的方式来理解上面的论证过程。

顺便，我们来看看乘法对加法的左侧分配律，也就是： $a \times (b + c) = a \times b + a \times c$ 。这同样可以通过乘法和加法的含义来理解。我们这样来看 $a \times (b + c)$ ——每一个单元里面有 $(b + c)$ 个同样单位下的同样的东西，这样的单元有 a 个。自然把每一个单元拆分为 a 个东西和 b 个东西这两个小单元，然后分别计算 a 份这样的两种小单元内的东西的数量，再合起来，得到 $a \times b + a \times c$ 。自然，这样的拆分不影响总量。

有了乘法，我们可以来定义乘法的逆运算——除法。也就是，如果 $3 \times 7 =$

¹未来会发现拓展到分数小数之后，这些运算律仍然成立

21, 那么, 我们就可以问 $3 \times ? = 21$ 呢? 答案就是 7, 记作 $21 \div 3 = 7$ 。类似地, $2 \times 3 = 6$ 用除法表示就是 $6 \div 3 = 2$ 。我们把 $2 \times 3 = 6$ 对应到现实场景——把两个每个篮子里面装了 3 个苹果的篮子合起来的苹果的总数, 把 $6 \div 3 = 2$ 看作是, 把总数是 6 的苹果分成每个篮子里面放 3 个苹果可以放几个篮子, 或者说每次从 6 中取出来 3 可以取多少次。从另一个角度, 我们可以把 $6 \div 2 = 3$ 看作把 6 个苹果平均放到两个篮子里面, 每个篮子里面会有几个苹果。把这个平均分的问题和前面的每次取出多少可以取多少次联系起来的办法是: 每次我们取出来足够的苹果给每个篮子里面放一个, 由于总篮子数量是 2 则, 每次需要取出来 2 个。这就回到了从整体中每次取出一部分可以取多少次的问题。于是, 就可以用除法来计算了。

当然, 反过来, 把每个篮子里面相同数量的苹果合起来数一数, 自然就是乘法。于是, 乘法和除法就互为逆运算。注意, 我们约定除法的除数不能为零——无论多少个零加起来还等于零, 不会等于一个非零的数, 因此 $\frac{q}{0}$ 当 $q \neq 0$ 的时候没有意义²。

从另一个角度, 我们来看 $2 \times 3 = 6$ 。我们把这 6 个苹果当成整体, 我们说, 每一个篮子里面的苹果, 占这个总苹果的多少呢? 我们发现, 这里有两个篮子, 其中的一个篮子自然就是占两个之中的一个。把一个整体分成 n 份, 取其中的 m 份, 就记作 $\frac{m}{n}$ 。例如这里的 $\frac{1}{2}$ 。 $\frac{m}{n}$ 也可以看作是 $m \times \frac{1}{n}$ ——从整体 n 份取出 m 份相当于是 m 次重复地取出来整体 n 份中的 1 份再合起来。这样我们就有了分数的定义, 包含真分数 $m < n$ 两者都是整数, 以及假分数 $m > n$ 。其中假分数相当于重复取 n 份中的 1 份的次数非常多, 多过了 n 次。这个时候, 我们需要想象以下拿出来另一个整体也分成 n 份, 然后来满足这个取 m 次 ($m > n$) 的要求。

下面, 我们来把分数和除法联系起来—— $\frac{q}{p} = q \div p$ ($p \neq 0$)。我们以假分数并且可以整除为例, 不能整除和真分数的的情形道理相同但是论证过程会稍微复杂一点。 $q \div p$ ($q > p$ 并且可以整除) 有一个性质 $(q \div p) \times p = q$ 。这是上面的乘法和除法就互为逆运算的直接结果。或者说, 先把一个整体 (q 个东西) 平均分成 p 份, 得到每一份的数量, 然后再取这样的数量的东西 p 份, 合起来自然还是 q 个东西。 $\frac{q}{p}$ 也具有同样的性质: $\frac{q}{p} \times p = q$ 。我们把这个整体分成 p 份取其中的 q 份, 然后把 p 个这样的份合起来, 自然

²未来会学习到无穷大, 以及 $\frac{0}{0}$ 的情形。

就是 p 个原始的整体。我们可以这样来理解：先把 p 当作 1，自然 $\frac{1}{p}$ 就是一个整体的 q 份中的 1 份，然后我们现在有 q 份合起来，自然就得到了整体；接着，如果我们一开始取走的不是一份而是 p 份，则需要把前面取走一份的操作重复 p 次合起来，自然就得到原始的整体 p 倍。现在我们得到两个对象具有同一个性质： $(q \div p) \times p = \frac{q}{p} \times p$ ，两个数乘以同一个不等于零的数得到的得数相同，则这两个数也相同。

现在，除法和分数已经联系起来了，我们说，除法和乘法就统一了， $q \div p = \frac{q}{p} = q \times \frac{1}{p}$ 。这样任何一个除法算式（只要分母不为零）就可以转化为乘法算式了。我们也可以来计算分数乘法了。例如， $\frac{q}{p} \times \frac{m}{n} = q \times \frac{1}{p} \times m \times \frac{1}{n} = m \times q \times \frac{1}{p} \times \frac{1}{n} = (m \times q) \times \left(\frac{1}{p} \times \frac{1}{n}\right) = (m \times q) \times \left(\frac{1}{p \times n}\right) = (m \times q) \times \left(\frac{1}{p \times n}\right) = \frac{m \times q}{p \times n}$ 。其中，我们用到了 $\frac{1}{p} \times \frac{1}{n} = \frac{1}{p \times n}$ 。这个可以从分数的含义上来理解：把一个整体先分成 p 份取其中的 1 份，然后，把这个取好的部分再看作整体分成 n 份取其中的 1 份，自然相当于把原始的整体分成了 $p \times n$ 份取其中的 1 份。

分数的加法在同样的分母下就很容易计算，相当于直接计算分子的加法。对于分布不相同的分数，我们只需要把分母变成相同的数——称作通分，就可以回到分母相同的情形。通分需要计算整数的最小公倍数，以及需要分数的性质——分子分母同时乘以一个数分数值不变，也就是 $\frac{q}{p} = \frac{q \times a}{p \times a}$ ($a \neq 0$)。回到分数和乘法的统一，我们有 $\frac{q \times a}{p \times a} = q \times a \times \frac{1}{p \times a} = q \times a \times \frac{1}{p} \times \frac{1}{a} = q \times \frac{1}{p} \times a \times \frac{1}{a} = q \times \frac{1}{p} \times (a \times \frac{1}{a}) = \frac{q}{p}$ 。反过来，如果我们的分数的分子和分母之间存在公因数，则需要通过计算最大公因数的方法来对分数作化简，得到分子和分母没有公因数的最简分数。其原理仍然是分子分母同时乘以一个不等于零的数，分数值不变。

这样我们就有了分数的加法和乘法的计算方法，也理解了“分数就是除法（分数线就是除号）”。于是，除法就通过分数成了乘法，也就是 $q \div p = q \times \frac{1}{p} = \frac{q}{p}$ 。

这样，我们就从数一二三的经验开始构建了整数、整数的加减乘除、负数以及通过负数把减法统一到加法、分数及其加减乘除运算以及通过分数把除法统一为乘法。我们也会了分数的加减乘除。

有了分数我们来学习小数的概念。首先，我们可以把任何一个分数通过除法竖式运算转化为无限循环小数。注意，这里，当有整数部分的除法算

完还有余数的时候，中间引入一个小数点来表示整数部分用完了，然后我们还是继续往下计算（1 相当于 $1.00000\cdots$ ，因此只要继续往下算每到下一位就跟整数的除法一样，拉下来一个下一位上的数字——这里就是 0）。例如我们来算一下 $\frac{1}{2}$ 和 $\frac{1}{3}$ 试试。除法竖式需要改成中国模式，目前是法国模式。

[illegible]

数 $\frac{q}{p}$, q, p 都是有限大小的整数。我们只要按照上面的除法不断地算下去, 每次出现的余数 r 必定是小于 p 的一个数, $r < p$ 。这样的整数只有有限多个 (p 个, $0, 1, 2, \dots, p-1$)。于是, 只要我们不断地除下去, 次数足够大, 余数肯定会重复。当然很早就出现了余数为 0 那么之后就都是 0 了, 也就是除尽了。余数一旦重复, 我们可以得到后面的商也将重复的结论 (因为后面的商的每一位仅仅依赖于前一个余数)。于是, 我们发现, 只要进入余数的计算也就是整数部分已经用完了还继续往下算, 商必然会重复。这样我们就得到了分数必然是无限循环小数。

那么是不是所有的无限循环小数都是分数呢? 整数 n 它自己就是分数 $\frac{n}{1}$ 。因此, 我们只需要来看没有整数部分的小数。一个最一般的循环小数可以记为 $0.n\dot{m}$, 其中 n 部分假设有 N 位, m 部分假设有 M 位。我们来把这个数转变为一个分数。

$$0.n\dot{m} \times 10^N = n.\dot{m}, 0.n\dot{m} \times 10^{N+M} = nm.\dot{m}, \quad (7)$$

两者相减得到

$$0.n\dot{m} \times 10^{N+M} - 0.n\dot{m} \times 10^N = nm.\dot{m} - n.\dot{m} = nm - n, \quad (8)$$

对左边做一个乘法对加法的分配律的反向应用得到,

$$0.n\dot{m} \times (10^{N+M} - 10^N) = nm - n, \quad (9)$$

于是,

$$0.n\dot{m} = \frac{nm - n}{10^{N+M} - 10^N}. \quad (10)$$

这是一个分数——分子分母都是整数。

习题 0.1 (无限循环小数化为分数). 把 $0.\dot{3}$ 和 $0.\dot{9}$ 分别转化为分数。

那么, 这世界的数 (包含人们通过思考设想出来的数和实际会遇到的数) 里面会不会出现无限不循环小数呢? 只要是有这样的数, 那这个数就不再是分数了。我们发现, 还真的。例如 $\sqrt{2}$, 定义为那个平方以后等于 2 的数, $(\sqrt{2})^2 = 2$, 或者说边长为 1 的正方向的对角线的长度。关于 $\sqrt{2}$ 不是分数的证明我们也简单重复在这里。

数学这样学》[1], 稍后我们也会复习 0.2.3 节, 和再次学习, 见第八章和第二十三章。

定理 0.1: $\sqrt{2}$ 不是分数 $\sqrt{2}$ 不是分数] $\sqrt{2}$ 不是分数。

证明 0.1. 我们先假设 $\sqrt{2}$ 还是一个分数, 也就是 $\sqrt{2} = \frac{p}{q}$ 。我们对分子分母已经做好了做约分, 保证 $\frac{p}{q}$ 是最简分数, 也就是 p, q 没有公约数。由于 $(\sqrt{2})^2 = 2$, 我们得到

$$\begin{aligned}\frac{p}{q} \times \frac{p}{q} &= 2 \\ \Rightarrow \frac{p^2}{q^2} &= 2 \\ \Rightarrow p^2 &= 2q^2.\end{aligned}\tag{11}$$

所以 p^2 是一个偶数, 于是, p 也是一个偶数。如果 p 是奇数的话, 则 p^2 肯定是奇数, 不满足要求⁴我们设偶数 $p = 2k$, 则

$$\begin{aligned}(2k)^2 &= 2q^2 \\ \Rightarrow 2k \times 2k &= 2q^2 \\ \Rightarrow 4k^2 &= 2q^2 \\ \Rightarrow 2k^2 &= q^2.\end{aligned}\tag{12}$$

所以 q^2 是一个偶数, 于是, q 也是一个偶数。所以, 我们得到 p, q 都是偶数, 至少存在公约数 2。这个和 $\frac{p}{q}$ 是最简分数矛盾。因此, 这一开始的猜测 $\sqrt{2} = \frac{p}{q}$, 即 $\sqrt{2}$ 是一个分数, 是错的。既然是一个分数是错的, 那么正确的只能是不是一个分数⁵。

于是, 我们就有了更加丰富的数的概念——是无限循环小数的分数(整数也是分数)和不是分数的小数。我们把分数称作有理数, 或者说比例数, 也就是可以通过调整测量单位来变成整数的数, 或者说可以表达成整数的比例的数; 反过来, 把不是分数的小数称为无理数。当然, 它们各自有正数和负数。

⁴实际上这里用到了反证法, 不满足要求的情况需要被剔除。我们也会在后面复习。

⁵这里我们用到了逻辑学知识, 一个命题只能正确或者错误, 两者比取其一, 只能取其一。后面我们也会学习到这个知识

有了小数的概念，则更进一步，在之前只有整数的数轴上，我们添加进去表示小数的点：例如 0.1 (0.2) 是就是把 1 分成十个小格取其中一 (两) 个小格的地方，并且 0.1, 0.2 之间的间隔和 0.2, 0.3 之间的间隔相同。0.01, 0.02 也类似可以找到对应的点。那么，我们能不能在数轴上找到一个表示某个无限循环小数的点呢？当然可以，只要我们能不断地把上面的每一位的对应关系做下去。甚至同样的方式我们可以找到一个表示无限不循环小数的点。唯一担心的问题是，当真的需要我们不断地一位位找下去的时候，是否真的我们可以实现找无限多位。讨论这个问题需要极限的概念，因此留到以后大学数学再来讨论。只要假设可以无限找下去，那么，任何有理数和无理数，我们都可以找到数轴上的点。

数轴上是不是还有除了有理数和无理数之外的其他数呢？这个问题也留到大学再来讨论。我们先拿过来结论用一下：所有数轴上的点不是有理数就是无理数。由于这个很好的性质，我们给有理数和无理数合起来一个名称——实数。合起来，我们有，数轴上的点和实数具有一一对应关系。一定要注意，现阶段这个命题我们并没有证明，而是暂时接受的，未来总是要来检验它的。

对于数，我们从自然数拓展到了正整数、分数、负数、实数，我们也有了实数的四则运算的计算方法，还有了四则运算律及其字母表示。那我们的数学阶段的复习基本就完成了。当然，考虑到大部分小学数学不包含无理数，在本书中，我们仍然在正文的第一章做了这个数的扩张。

小数的计算可以采用两种方式。一种是把小数转化为分数来作计算。不过我们发现，只有无限循环小数才能这样做。另一种是参考整数的计算，也就是基于数位对齐来作计算。例如先对齐小数点，再来算加法（包含了减法）。有限位小数的乘法（包含了除法）可以先转化为整数的乘法，然后，最后再来处理转化为乘法的过程中需要做的 $\times 10^m$ 的问题。无限位小数的计算可以考虑先在某个位置截断，当做有限位小数来计算，然后在需要的时候，可以保留更多位小数，也就是使得这个有限小数和无限小数更加接近。

习题 0.2 (无限小数的有限位计算). 算一下 $0.\dot{3} \times 3$ 取百分位、千分位、万分位的时候的结果，看看哪一个更接近真实值（这里是 1）。

对于无限不循环小数的计算，我们大概可以这样做：在任何一个有限位截断，然后，按照有限小数（也就是循环体为零的无限循环小数）来转化

为分数计算，最后，我们可以把这个截断位放到所需要的任何一个足够小的分数位上。例如保留到 $\frac{1}{1000}$ 不够的时候，就取到 $\frac{1}{10000}$ 、 $\frac{1}{100000}$ ，等等等等。第一这样的保留到越来越小的分数位的计算我们肯定可以做下去，第二只要我们保证这样的计算算出来的结果离我们想要的结果越来越接近，那么，这个计算就是合理的。不过，是不是这样的计算会离我们的结果越来越接近，目前我们只能直观感受一下（例如，你去算一下 $0.\dot{3} \times 3$ 取百分位、千分位、万分位的时候的结果，看看哪一个更接近真实值），严格的证明需要极限的概念，也留到大学再来完成。

这样我们就有了所有实数的概念和实数的计算规则，以及运算律及其字母表示。原则上，我们需要重新来证明实数上仍然有加法结合律、加法交换律、乘法结合律、乘法交换律、乘法对加法的左侧分配律。不过，这个任务也留给将来大学阶段再说。我们暂时接受这些运算律在实数范围内仍然成立。反正在分数范围内是存在的，从分数到实数就差一个从有限位到无限位，也就是说这仍然是一个关于截断和极限的问题。

有了数和运算律，我们还缺一个等式的性质，也就是：等式的两边同时作加法和乘法（包含除以同一个不为零的数）等号仍然成立。而等式的性质的理解就特别简单了：无论一个算式多么的复杂，等号成立，意味着左边的得数等于右边的得数。对于已经相等的两个数，或者说实际上同一个数，则无论对这个数作什么样的运算，只要这个运算合法而且结果唯一，得到的结果左右两边仍然相同。

最后，运算律和等式的性质是对任何数都成立的，因此有必要从以具体数为例的表示过渡到可以适用于任何数的表达式。为了这个目的，我们用字母来表示任意的一个数，就有了字母所表示的运算律和等式的性质。在中学阶段，这个用字母表示的运算律和等式的性质是我们数学学习的起点。

0.2.2 现实和数学，从关系到数学结构

数学 WHWM 四问，数学五步，数学建模

0.2.3 数学知识的系统性

如果这部分对你不是复习，没关系，我们提供以下的解决方案：第一，你可以去看一下《小学数学这样学》[1]；第二，你可以看一下下面的复习，其实很多时候可以当作第一次学习的材料看；第三，在后续章节中会有再一次专门学习集合和逻辑的部分，见第八章和第二十三章。

集合是一个满足下面的一个叫做“确定性”的要求的一堆东西：给定任何一个东西，我们必须能够明确地判断这个东西是否属于这一堆。未来在大学阶段我们会学习更加抽象的定义。在那之前，我们先采用这个描述性定义。我们还给这个描述性定义的核心“是否属于”一个数学符号：属于用 $\in$ ，不属于用 $\notin$ 。例如，1 是自然数之一，也就是属于自然数，记作“ $1 \in \mathbb{N}$ ”（其中 $\mathbb{N}$ 代表了所有自然数）；1 不是质数（也不是合数），记作“ $1 \notin \mathbb{P}$ ”（其中 $\mathbb{P}$ 代表了自然数中的所有质数）。于是，我们可以用这个符号来描述集合：一个集合 A 就是给定任何对象 x ，我们都可以得到 $x \in A$ 和 $x \notin A$ 有并且只有一个成立。

这个确定性的要求其实很高。例如，秃头、高个子、好苹果、水果等等都不是一个集合。比如，尽管我们大多数时候可以对于一个人是否属于秃头、高个子达成共识，但是也会经常出现分歧。甚至有些词语本身的含义就是不够明确的，例如，智慧、聪明、才干、气质等等（注意这里的“等等”的含义，原则上本来是应该清楚的，也就是那些含义不明确的词语，但是在生活中往往“等等”不一定能够具有这样明确的意思，甚至“含义不明确”这个短语本身的含义也不一定是明确的）。那么，什么才是真的明确呢？最严格的标准就是可操作可测量的。例如，如果我们规定只有头发少于 1000 根的叫做秃头，身高大于 180cm 的叫做高个子，那没问题，就是明确的。类似地，我们可以说，智商测量量表不一定测量的就是聪明程度，而且不同的量表可能测出来的东西和数值都不一样也不能转换，但是，如果我们忽略这些问题，仅仅用某一套量表来测量，然后规定一个阈值，我们就有了“聪明人”的明确含义，就可以构成集合了。因此，实际上，集合背后是可操作可测量。

如果我们把集合 A 包含的所有对象——称作集合 A 的元素——写下来，我们用下面的记号 $A = \{a, b, c\}$ 。一般来说，我们用大写字母表示集合，小写字母表示元素。例如，一个骰子扔下去可能出现的向上的面的集合

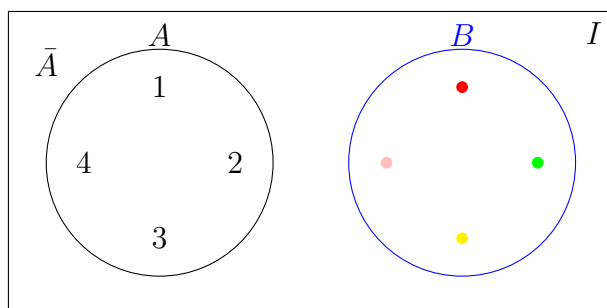


图 1: **集合和元素**。展示集合和元素的图示法。用圆形或者长方形表示集合，用放在集合里面的点或者其他符号表示元素。

记作 $D = \{1, 2, 3, 4, 5, 6\}$ 。注意这里的六个数仅仅是记号，表示哪一个面向上。如果我们约定每个出现的面可以获得和面的符号数量相同的钱（一人民币元为单位），那么，这里的六个数就有了可以比大小和作加减运算的含义。

我们还可以通过一个叫做 Venn 图的东西，如图 1，来把集合可视化。我们画一个圆圈或者长方形表示集合 A ，用点（可以带上记号）来表示集合里面的元素。为了表示属于和不属于集合 A ，我们画一个全集 I ，其中的一部分被包含在 A 之中，另一部分没有被包含在 A 之中，称作 A 的补集 $\bar{A}$ 。其含义是 $\bar{A} = \{x | x \notin A\}$ 。一般来说，当我们说补集的时候，我们会约定一个全集（这可能是因为简单地说“所有的其他对象”所指的对象不够明确，也就是不一定构成集合） I 。于是，我们通常把补集写作 $\bar{A} = \{x | x \notin A, x \in I\}$ 。其中 $\{x | p(x)\}$ 的后半部分 $p(x)$ 表示 x 所需要满足的条件。

定义 0.1 (补集). 给定一个集合 A ，以及一个当作全集的集合 I ，集合 A 的补集定义为 $\bar{A} = \{x | x \notin A, x \in I\}$ 。

到此，我们就有了集合、补集和全集的概念。下面我们来学习交集和并集——关于两个集合的关系。我们定义两个集合的交集为同时属于两个集合的元素。

定义 0.2 (交集). 给定两个集合 A, B ，其交集定义为 $A \cap B = \{x | x \in A \text{ 并且 } x \in B\}$ 。

其中“并且”可以用记号 $\wedge$ 或者 $\&$ 来表示。

定义 0.3 (并集). 给定两个集合 A, B ，其并集定义为 $A \cup B = \{x | x \in A \text{ 或者 } x \in B\}$ 。

其中“或者”可以用记号 $\vee$ 或者 $|$ 来表示。

很容易我们就可以把交集和并集通过两两集合的操作来定义多个集合的交集和并集。

如果我们发现，两个集合的交集和其中之一相等，例如 $A \cap B = A$ ，那就是 A 完全被包含于 B 之中，或者说，所有 A 中的元素肯定都在 B 中。这个可以通过画图来理解。将来我们也可以严格证明。比如说，我们假设有至少一个元素在 A 中但是没有在 B 中，我们就可以证明这个元素不会在 $A \cap B$ 中，于是 $A \cap B \neq A$ 。这里我们用到了反证法。后面我们会回到用集合的语言来理解反证法。我们把这样的关系称为包含关系。

定义 0.4 (被包含于). 给定两个集合 A, B ，我们称 A 被包含于 B ，记作 $A \subseteq B$ ，如果 $x \in B, \forall x \in A$ 。

其中“ $\forall$ ”的含义是“任意”，也就是随便取一个满足后面的条件。数学家经常把条件放后面，因此，这里的含义是“取任何一个集合 A 中的元素 x ，都满足 x 在集合 B 中”。反过来，称作 B 包含 A 。

定义 0.5 (包含). 给定两个集合 A, B ，我们称 A 包含于 B ，记作 $B \supseteq A$ ，如果 $x \in B, \forall x \in A$ 。

这里的包含实际上既可以是真包含也可以是相等，也就是 $A \subseteq B$ 可以是 $A \subset B$ 还可以是 $A = B$ 。其中前者表示 B 中有不属于 A 的元素。

顺便，我们增加一个两个集合相等的定义和一个集合相等和集合包含关系之间的对应关系的定理。

定义 0.6 (集合相等). 两个集合完全相等的含义是两个集合中的元素完全相同，也就是 $x \in B, \forall x \in A$ 并且 $x \in A, \forall x \in B$ 。

从这个定义我们就看出来，刚好这意味着 $A \subseteq B$ 和 $B \subseteq A$ 。

定理 0.2: 集合相等和双向包含

集合相等和双向包含] $A = B \iff (A \subseteq B) \wedge (B \subseteq A)$.

证明. $A = B$ 意味着 $x \in B, \forall x \in A$ ，也就是 $A \subseteq B$ 。 $A = B$ 还意味着 $x \in A, \forall x \in B$ ，也就是 $B \subseteq A$ 。

反过来， $A \subseteq B$ 意味着 $x \in B, \forall x \in A$ ； $B \subseteq A$ 意味着 $x \in A, \forall x \in B$ 。加起来，正好就是 $A = B$ 。□

注意集合相等的定义是一个额外强加的要求。不一定任何对象都满足这个要求。例如，如果我们用一个人的祖先构成的集合来代表这个人。那么，兄弟姐妹具有完全相同的祖先，也就是集合完全相同，但是兄弟姐妹并不一定只有完全相同的一个人。当然，我们仍然可以说，这些兄弟姐妹有完全相同的祖先的集合（甚至这些祖先的顺序）。因此，集合相同仅仅表示集合相同，不一定能够延伸到这个集合整体所指代的对象之间的关系，如果集合整体所指代的对象有一个定义的话。这个问题牵涉到把集合当作一个更大层次的集合的元素，然后建立映射的关系。未来我们会再来学习这个集合的集合以及映射。很有意思的是，除了未来往往在研究生阶段会学习到非常特殊的情况（叫做范畴），绝大多数数学对象的定义都可以建立在集合的基础上。也就是粗糙地说，数学基本上只需要集合（以及集合的集合，集合的集合的集合，这些仍然都是集合）就够了。集合在数学语言中具有非常强大的威力。

有了上面这一点点集合论的知识，我们来理解一下演绎逻辑。首先，我们来看三段论。三段论的原始形式是这样的：大前提，小前提，结论。其中大前提就是类似于“所有的人要是会死的”这样一个对集合里面的所有元素的断言。小前提是说某个对象属于这个集合，例如“苏格拉底是人”。结论就是把对象和大前提的断言联系起来，例如“于是，苏格拉底是会死的”。用集合论的语言，现在我们就很好理解这个三段论了：我们从一个集合中的所有的元素都具有某性质开始，自然也就得到任何一个这个集合中的元素具有这个性质。注意，这看起来就跟废话一样，但是，正是这样的看起来特别简单自然的命题和推理成了构建宏伟的数学知识大厦的基石。

理解了三段论之后，我们来用集合理解逻辑论证中的另外一个主题——命题、逆命题、否命题和逆否命题及其之间的关系。还是用“所有的人要是会死的”、“苏格拉底是会死的”当作原始命题，否命题就是“不是所有人（有的人）都不会死”，“不是苏格拉底就不会死”；逆命题就是“会死的是人”、“会死的是苏格拉底”；逆否命题是“不会死的不是人”、“不会死的不是苏格拉底”。通过对这些话的具体含义作思考，我们自然可以来判断这些命题是否成立。例如，我们发现，逆命题不对，会死的不仅仅是人，以及人之一苏格拉底，还有猫、马、兔子等等。否命题也类似，例如苏格拉底之外的人也会死。比较麻烦的是逆否命题。不过，如果我们可以使用反证法，我

们可以这样来论证：假设你否命题不成立，也就是“不会死的还是人”，根据“所有人都会死”，我们得到“不会死的是人，会死”。这里“不会死的会死”直接就矛盾了，因此，这个原始的假设，也就是你否命题不成立和原命题直接矛盾。因此，如果原命题为真，则逆否命题肯定为真。也就是说，原命题和逆否命题完全等价。但是，是不是我们可以脱离这些命题的实际内容，不需要去找出来满足或者不满足这些命题的事实，直接来作判断，或者至少找到作判断的方法呢？这就是逻辑学所要解决的问题。

我们用集合来当作逻辑学的基础来解决这个原命题、否命题、逆命题、逆否命题之间的关系的问題。我们说一个命题必然有一个前提 A 一个结论 B ，记作 $A \Rightarrow B$ 。我们把所有的满足前提 A 中的要求的对象组成的集合也记作 A^6 ，满足结论 B 中的性质的对象组成的集合也记作 B^7 。我们说，原命题 $A \Rightarrow B$ 就意味着集合包含关系 $A \subseteq B$ ，也就是任何一个 A 中的元素必然属于集合 B 。集合的包含关系就是命题中的导出关系。这一点非常重要。通过集合论的语言，我们就可以把命题中的条件到结论的导出关系变成集合的包含关系。

有了这个导出关系和包含关系的对应，我们就可以来看原命题、否命题、逆命题、逆否命题之间的关系了。我们假设原命题 $A \Rightarrow B$ 成立，也就是包含关系 $A \subseteq B$ 成立。我们说“否”就对应着补集，否命题就对应着补集之间的包含关系。也就是说，否命题就是 $\bar{A} \Rightarrow \bar{B}$ ，对应着 $\bar{A} \subseteq \bar{B}$ 。如图 2 所示，由于集合 A 被包含在 B 的内部，补集 $\bar{A}$ 实际上会把补集 $\bar{B}$ 包含在内部。因此，我们有 $\bar{A} \supseteq \bar{B}$ ，或者说 $\bar{B} \subseteq \bar{A}$ 。对应到命题，我们有 $\bar{B} \Rightarrow \bar{A}$ 。因此，我们得不到否命题，但是能够得到逆否命题。也就是说，逆否命题和原命题等价，但是，原命题和否命题之间没什么直接的关系，原命题成立的同时逆命题可以成立也可以不成立。

我们再来看原命题和逆命题之间的关系。首先，我们从逆命题和否命题之间的关系是逆否命题关系，我们就可以得到逆命题和否命题完全等价。于是，原命题和逆命题之间的关系，就是原命题和否命题之间的关系，也就是

⁶或者 $\mathcal{A} = \{x | P_A(x) = 1\}$ 表示满足 $P_A(x)$ ——也就是 x 满足条件 A ——的所有的 x 的集合，但是只要没有歧义我们就用同样的符号指代条件和集合

⁷或者 $\mathcal{B} = \{x | P_B(x) = 1\}$ 表示满足 $P_B(x)$ ——也就是 x 满足性质 B ——的所有的 x 的集合，但是只要没有歧义我们就用同样的符号指代性质和集合

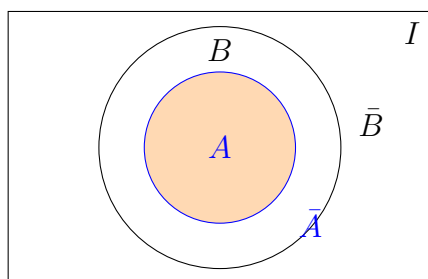


图 2: 原命题和逆否命题等价。展示原命题和逆否命题等价的集合图式法:
 $A \subseteq B \Leftrightarrow \bar{B} \subseteq \bar{A}$ 。

没什么直接关系，原命题成立的同时逆命题可以成立也可以不成立。其次，我们也可以直接从集合来看原命题和逆命题之间的关系。原命题 $A \Rightarrow B$ 成立意味着集合包含关系 $A \subseteq B$ 成立；而逆命题 $B \Rightarrow A$ 成立意味着集合包含关系 $B \subseteq A$ 成立；但是， $A \subseteq B$ 既不能保证也不能否定 $B \subseteq A$ 成立。

当集合 $A = B$ 的时候，我们既有 $A \subseteq B$ 也有 $B \subseteq A$ ，于是我们有 $A \Rightarrow B$ 和 $B \Rightarrow A$ ，也就是原命题和逆命题同时成立。反之亦然。这个时候，我们称 A 和 B 等价。

总结一下：通过集合的语言，以及把集合和命题相结合，我们得到——逆否命题和原命题等价，当且仅当对应的集合 $A = B$ 的时候，原命题和逆命题也同时成立，逆命题和否命题由于是逆否命题关系也等价。有了这个知识，我们就可以更加方便地来处理逻辑演绎的问题了。例如，我们可以不加判断就得到“苏格拉底是要死的”和“不会死的不是苏格拉底”完全等价，以及如果要判断逆命题“会死的就是苏格拉底”、否命题“不是苏格拉底就不会死”是否和原命题“苏格拉底是要死的”等价需要去检查集合“苏格拉底”、“会死的”是否完全相同。这个时候，我们自然发现，两个集合不同，于是不等价。

有了这个命题和逆否命题等价，我们就可以进一步来证明反证法是可以用的的一种逻辑演绎证明方式。反证法指的就是，为了证明 $A \Rightarrow B$ ，我们先假设 B 不成立，也就是从 $\bar{B}$ 开始，然后证明一个和 A 直接矛盾的结果，也就是证明 $\bar{A}$ ，也就是得到了 $\bar{B} \Rightarrow \bar{A}$ 。这正好就是 $A \Rightarrow B$ 的逆否命题。因此，反证法相当于通过证明逆否命题成立来证明原命题成立。回到 $\sqrt{2}$ 不是有理数的证明，我们从假设 $\sqrt{2} = \frac{q}{p}$ 是一个有理数并且整数 p, q 之间没

有公约数（有就化简掉，直到最简分数）开始，证明了 p, q 之间有公约数 2。这个结果和假设有直接矛盾。因此，假设不能成立，也就是说， $\sqrt{2}$ 不是有理数。

另外，初看起来，这里好像没有了前提。实际上，这里的前提是 $a^2 = 2$ ，也就是说这里的命题的完整表述是： $a^2 = 2 \Rightarrow a \notin \mathbb{Q}$ （ $\mathbb{Q}$ 表示有理数），或者说 $a^2 = 2 \Rightarrow a \neq \frac{q}{p}, \forall p, q \in \mathbb{Z}$ 。在我们的证明过程中，我们需要多次使用这个前提条件。

在这一小节，我们学习了一点点集合的语言，然后我们用集合的语言理解了三段论，理解了原命题、否命题、逆命题、逆否命题之间的关系，理解了反证法。三段论、反证法和命题之间的关系是我们后面要用到的演绎证明的基础。将来我们还会把更多的其他的数学概念也建立在集合之上。这体现了集合这个数学基本概念的重要性，也体现了数学知识的系统性。

0.2.4 从小学数学学习和复习到理解型学习

人类知识高速公路上以高层知识生成器为目标的理解型学习

0.3 小学数学知识总结以及本章小结

我们还可以把以上的知识之间的联系通过一张图来表达，如图 3。这张图可以用来帮助你检查一下是否具备了学习本书所要求的小学阶段的知识。如果你差不多能够走通这张图上的各条连边（概念间关系）和各个顶点（概念，数学学科概念和学科思维），则你就基本上具有了学习本书的基础。

在本章中，我们明确了为了学习本书所需要的小学阶段的数学知识，给出了要求列表，做个这些知识的复习。我们再一次强调，本书的这个前置知识的要求不算低，主要问题不在数学计算上，而是在于对数学知识以及这些知识的被创造的过程还有过程背后的思维的理解。如果从列表和复习中你发现好多知识对你新的，你可以选择把这部分复习多看几遍，也可以回去看一下《小学数学这样学》[1]，也可以顺着本书先看下去，然后时不时地回到本章再一次来阅读。

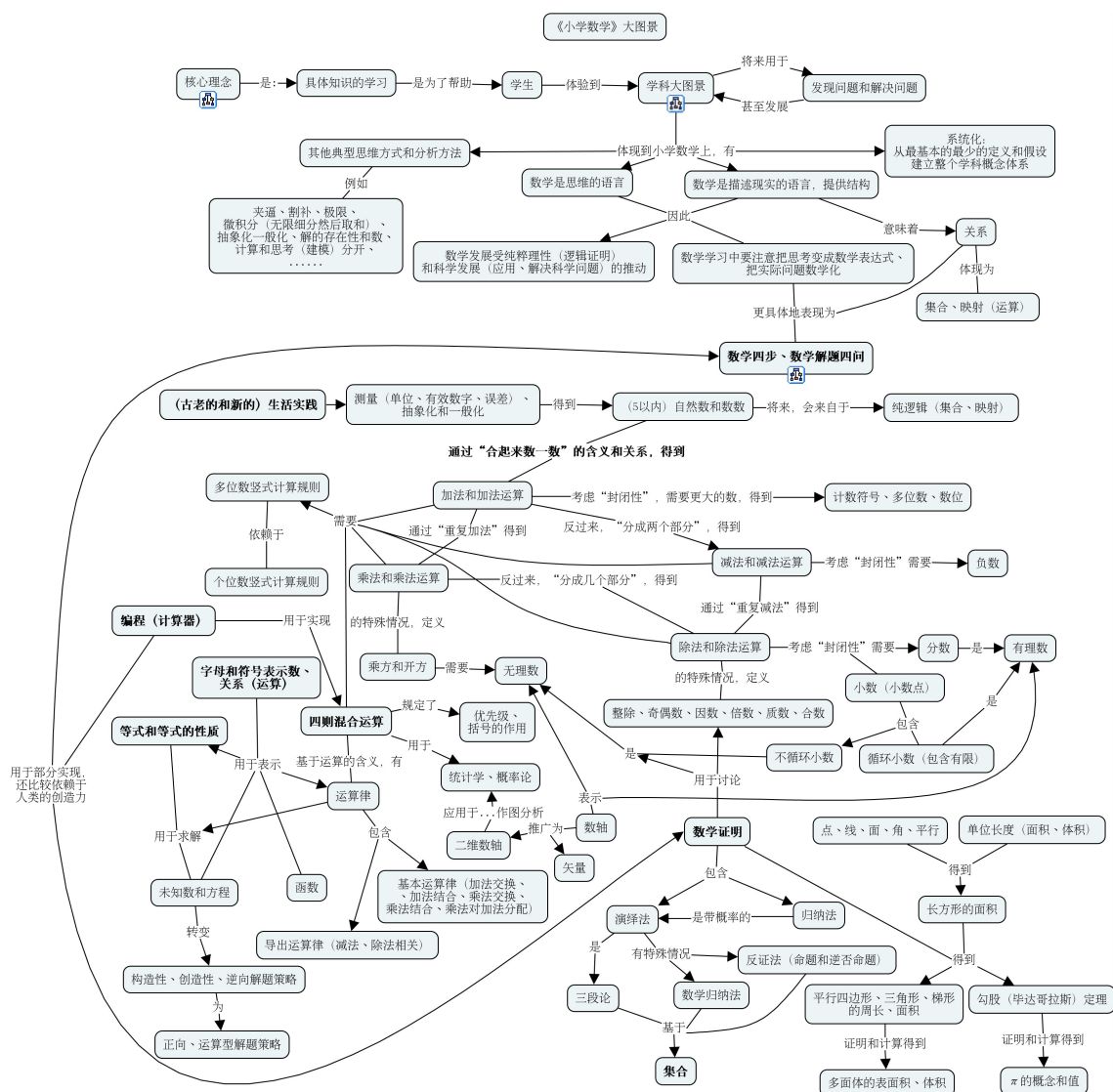


图 3: **小学数学整体概念地图**。可以用来检查一下是否具备了学习本书所要求的小学阶段的知识。[考虑是否印单独一页](#)

第一部分

初中篇

第一章 数轴上的点和数

We do not describe the world we see, we see the world we can describe. The only thing we have power over in the universe is our own thoughts. It is not enough to have a good mind. The main thing is to use it well.

并不是我们描述我们看到的世界，而是我们见到我们能够描述的世界。在这个世界中我们唯一能够自主的就是我们的思想。仅仅有一个好脑子是不够的，重要的是用好它。

– Rene Descartes

在这一章中，我们将从小学学习过的自然数或者说非负整数开始，走到所有的整数，然后走到分数，再到小数（包含无限循环和不循环小数），然后在数轴上来表示这些数。

最后我们将引入直角坐标系，也称为笛卡尔坐标系（Descartes Coordinates）——用两个相互垂直的数轴来表示平面上的所有点。

1.1 集合的复习

我们先来复习以下集合。本书后面的内容为了表达方便会经常使用“集合”这个词。未来我们会专门来学习集合。在本书的前置书《小学数学这样学》中，我们也已经初步学习过集合。未来到了大学或者研究生阶段，我们还会重新严格地来学习集合的定义。在大学或者研究生院将要学习的集合的公理化定义之前，集合的概念的学习都是粗糙的，不准确的。这里，我们也暂时按照这样的粗糙的不准确但是大概够用的方式来复习。

在这里，集合的含义就是一团对象，或者说一群对象构成的这个群体，

只需要满足这个群体做包含的对象的明确性。这些对象的每一个被称为这个集合的元素。有的时候，我们也强调无序性（集合元素之间没有顺序）和唯一性（每一个元素仅仅在集合里面出现一次）。明确性指的是我们可以判断任何一个对象是否是这个群体的元素。例如所有的自然数是一个集合¹，一个硬币的两面（可以记作 H, T 表示正面和反面），一个骰子的所有面（用 1, 2, 3, 4, 5, 6 可以来标记这些面）。反过来，对象不明确的就不能称作集合，例如所有聪明人、所有头发多的人、所有好吃的²。

在所有的集合中，有一个特别特殊的集合，它里面什么元素也没有，被称为空集，记作 ϕ 。你可能会觉得为什么会需要一个什么元素也没有的集合。那你可以想象一下一个你家冰箱里面的鸡蛋格子里的鸡蛋用完了，你把一袋子糖吃完了，里面就没有所要存储的那个你关心的东西了。当你表示“那里面没有（我所关心的那个）东西”的时候，你就需要这个空集的概念——里面没有元素的集合。

当然，你说那个时候，还是有一个鸡蛋格子和糖袋子的。是的，我们说这就是你用 $\{\phi\}$ 来指代那个装着“啥也没有”——也就是空集——的对象³。注意， $\{\phi\}$ 有元素啊，只不过其包含的元素是空集这个集合。

类似地，你说，你是不是还可以发明这样的符号，问这是什么—— $\{\phi, \{\phi\}\}$ 。我说这里面装了“ ϕ ”和“ $\{\phi\}$ ”这两个数学对象啊。一个是空集，一个是包含把空集当作元素的集合。顺着这个思路，你还可以继续构建下去：

$$\phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}, \dots \quad (1.1)$$

于是，我们就从内部元素个数是 0 的集合，走到了内部元素个数是 1 的集合，内部元素个数是 2 的集合，内部元素个数是 3 的集合，等等等等。

¹ 尽管不同的书不同的人会做不同的约定——例如有的书的自然数包含零有的书不包含零，但是对于给定的书，给定的人，或者两个人讨论之后选择了遵从某个人的约定之后，自然数包含哪些数是明确的。

² 当然，也不是没有补救措施来使得这些对象明确起来。例如，都去测量一下某种智力测试量表，规定大于多少的算从聪明人。我们还需要额外假设这个量表非常可靠重复测量的误差很小很小。例如把每个人的头发数量都数清楚，然后约定大于多少根算多。例如把还吃与否也做一个量表来检测，只要这个量表的测量结果可重复，则约定大于多少算好吃。

³ 仅仅从你关注的鸡蛋和糖的角度来说是如此啊，如果你直接就是关注糖袋子，例如你收集它们，则直接 $\{\text{糖袋子}\}$ 就是一个包含了东西的非空集合。

你看，我们好像不再需要从自然语言出发来构建自然数了，我们有了基于集合的自然数的定义——自然数好像就可以看作是前面构建的集合的大小。未来等到了大学阶段，我们会更加严格地来学习这部分内容——自然数的纯粹数学定义而不是基于生活经验和自然语言的定义，以及自然数和集合的关系。现阶段，你只要看到，其实自然数也是可以通过纯粹理性的想象力来定义的就行了。在后面使用自然数的时候，你仍然可以按照你之前从生活经验和自然语言中做抽象得来的那样去使用。

有了集合的概念，我们还学习了集合之间的运算，包含集合之间的交集、并集和补集（注意，补集实际上也是一种关系，牵涉到全集这个集合和要做补的那个集合）的概念。如果你实际上在小学没有学习过这部分，你可以去看看《小学数学这样学》，或者，我在下面也给你准备了一个非常粗糙的但是够用的关于什么是集合之间的交集、并集和补集的说明。

两个集合交集指的是把两个集合中共同具有的元素找出来，组成的那个集合。两个集合并集指的是把两个集合中所有的元素都找出来，组成的那个集合。一个集合在另一个更大的集合中的补集指的是在更大的集合中剔除掉更小的集合中的所有元素，把剩下的组成的那个集合。这里一个集合比另一个集合大的含义是，前者包含了后者的所有元素。

有了集合的含义（关键是明确性）、集合的交集、并集、补集和包含关系，我们就暂时够用了。

习题 1.1 (我家的人). 你自己来列出来你家都有谁，当作集合 F_S 。

习题 1.2 (我爸爸家的人). 让你爸爸列出来他家都有谁，当作集合 F_D 。

习题 1.3 (我妈妈家的人). 让你爸爸列出来他家都有谁，当作集合 F_M 。

习题 1.4 (我爷爷家的人). 让你爷爷列出来他家都有谁，当作集合 F_G 。

习题 1.5 (比较我、爸爸、妈妈、爷爷家的人). 把上面三个集合比较一下，有没有包含关系，各自有多少个元素，找出来它们的交集、并集。如果有包含关系，也算一下补集。看看这些交集、并集、补集在现实世界中有什么含义。例如，想一想有没有那个谁满足“我在谁的家里，但是谁不在我家里”。如果有，找出来，再想想怎么会出现的这个情况，我或者那个谁要做什么吗？

顺便，通过这几个问题，我希望你看到，数学的研究对象可以是任何东

西。顺便，千万不要害怕“集合”这个词，这个词在日常语言层面的解释非常容易懂——不过就是所包含的对象明确的一群对象，或者索性可以简称“一群”（一团、一堆）。很多时候，我们学习一个概念，先会基于日常经验来形成初步的理解，然后后来再接触到更进一步的知识的时候重新再来把这个概念学习一遍，形成更加深刻的理解。第一次没有完全把握住完全不用担心。例如，自然数咱们就是这样学习的，而你看到了一旦初步会了集合，你就可以把自然数理解的得更好。将来集合也会是如此。

1.2 数的复习——运算封闭性和数的出现

我们再来复习一下数。

自然数是在小学阶段不交代来源和定义的东西。依赖于自然语言和日常生活经验，我们就有了自然数：大概来说，就是给定一个计量单位（可能是精确的，例如“一斤糖”，也可能是大概的，例如“一个人”、“一个苹果”），一群或者一个对象里面有几个这样的计量单位，这个“几”就成了自然数。本质上自然数就是对“数 + 量 + 物”的表达方式的抽象，忽略其中的“量 + 物”而剩下的部分。

然后，通过对自然数（在这里，我们取包含 0）的加法的计算，我们可以发现，永远可以在现有的数的基础上增加 1 来获得更大的自然数；反过来，对一个给定的自然数做不断的减法的操作，总会得到 0，而继续做减法就会得到小于 0 的数。由于加法和减法可以对应着现实世界中有意义的操作——例如减法大概就相当于下山（当然，受实际地形所限，在地球上的任何地方这个下山也有一个最低点，上山也是如此。不过数学是受现实启发，但是不受现实制约的学科），如果我们要求数学尽可能地描述所有的现实对象，则我们就会要求数学中也得有这些小于 0 的数。我们称之为负数。

正整数、零和负整数合起来，我们称为整数。通过负数，我们可以把减法和加法认同 $a - b = a + (-b)$ 。我们稍后会再一次回到这个加减法的认同。

类似地，通过整数除法我们就可以得到分数（ $\frac{p}{q} = p \div q$ ），包含对应着整数的分数和不对应着整数的分数（因此，原则上，分数包含了整数）。正是不对应着整数的除法计算的结果的出现，使得在原来的整数范围内除法计算不封闭了。所以，我们需要引入新的数。

通过对分数的进一步研究就会发现，分数必然是循环小数（因为整数的余数必然小于这个整数因此只有有限个，这时候如果能够除尽则最终余数会等于零，得到有限小数相当于循环的部分是零，如果除不尽则余数必然会重复出现，只要重复出现了一次，则以后得到的结果都相同，因此就是无限循环小数）。以后我们把有限小数（循环体是 0）和无限循环小数合起来称为无限循环小数。

我们也可以反过来证明无限循环小数都是分数。在这里，我们仅仅通过一个例子来体现。例如， $0.\dot{3}4\dot{5} = a$ ，则 $1000 \times a = 345.\dot{3}4\dot{5}$ ，于是 $999a = 1000a - a = 345$ ，得到 $a = \frac{345}{999}$ 。

把分数看作是除法算式（ $\frac{p}{q} = p \div q$ ）我们就可以得到分数的性质——分子分母同时乘以一个不为零的数分数值不变，

$$\frac{b}{a} = \frac{b \times q}{a \times q}. \quad (1.2)$$

这很容易证明：

$$\frac{b \times q}{a \times q} = (b \times q) \div (a \times q) = b \div a = \frac{b}{a}. \quad (1.3)$$

这里我们运用了除法的性质：除数和被除数同时扩大相同的倍数商不变。

有了分数我们可以把四则运算推广到分数。例如 $\frac{b}{a} + \frac{p}{q}$ ，其中 a, b, p, q 是整数。我们的推广的方法是，找一个更小的单位，使得在这个单位下，所有的数可以重新成为整数，然后用这些整数来得到计算结果，最后把单位换回来。例如，因为 $\frac{1}{a}$ 的含义是 把一个单位分成新的 a 个单位来看， $\frac{1}{q}$ 的含义是 把一个单位分成新的 q 个单位来看，那么，如果我们把一个单位分成 $a \times q$ 个单位来看，则原来的 b 一个单位现在成了新的单位下的 $b \times (a \times q)$ 个单位（每个单位是 $\frac{1}{a \times q}$ ），原来的 p 一个单位现在成了新的单位下的 $p \times (a \times q)$ 个单位（每个单位是 $\frac{1}{a \times q}$ ）。于是在统一单位（ $\frac{1}{a \times q}$ ）下，上面的算式成了

$$\frac{b}{a} + \frac{p}{q} \rightarrow \frac{b \times (a \times q)}{a} + \frac{p \times (a \times q)}{q} = bq + pa, \quad (1.4)$$

注意现在的每一个单位是 $\frac{1}{a \times q}$ ，因此把单位换回来，就得到

$$\frac{b}{a} + \frac{p}{q} = \frac{bq + pa}{a \times q}. \quad (1.5)$$

我们也可以把上面的换成更小的单位先计算再换回来的过程看作是分数性质的运用,

$$\frac{b}{a} + \frac{p}{q} = \frac{b \times q}{a \times q} + \frac{p \times a}{q \times a} = \frac{b \times q}{a \times q} + \frac{p \times a}{a \times q} = \frac{b \times q + p \times a}{a \times q}. \quad (1.6)$$

类似地基于分数的含义,也可以把乘法计算推广到分数,可以先从分数的含义上证明 $\frac{1}{a} \times \frac{1}{q} = \frac{1}{a \times q}$: 把一个单位分成 a 份, 取其中的 1 份, 然后对这个对象再分成 q 份取其中的 1 份, 这相当于把原始的一个单位分成了 $a \times q$ (原来的每一份分成 a 份以后都做相同的再次分成 q 份的操作), 取其中的一份, 也就是 $\frac{1}{a \times q}$ 。然后, 运用乘法交换律和结合律——假设对于分数参与乘法它们还成立, 就有

$$\begin{aligned} \frac{b}{a} \times \frac{p}{q} &= b \times \frac{1}{a} \times p \times \frac{1}{q}, \\ &= (b \times p) \times \left(\frac{1}{a} \times \frac{1}{q} \right) = (b \times p) \times \frac{1}{a \times q}, \\ &= \frac{b \times p}{a \times q}. \end{aligned} \quad (1.7)$$

有了分数乘法的计算规则和乘法的含义, 我们很容易得到 $a \times \frac{1}{b} = \frac{a}{b}$: a 个 $\frac{1}{b}$ 合起来就相当于把原来的单位一分成了 b 份取了其中的 a 份⁴; 也可以这样看 $a \times \frac{1}{b} = \frac{a}{1} \times \frac{1}{b} = \frac{a \times 1}{1 \times b} = \frac{a}{b}$ 。这样我们就把除法和乘法认同, 通过分数。

有了分数的乘法和除法和乘法的认同 ($a \div b = a \times \frac{1}{b}$), 就可以除法计算推广到分数了,

$$\frac{b}{a} \div \frac{p}{q} = \frac{b}{a} \times \frac{1}{\frac{p}{q}} = \frac{b}{a} \times \frac{q}{\frac{p}{q} \times q} = \frac{b}{a} \times \frac{q}{p}. \quad (1.8)$$

这里我们展示了一个把数扩大以后把运算也拓展到这些数上的例子。在我们其他拓展数的时候, 例如前面拓展到负数, 后面拓展到无理数, 我们也要完成这样的运算的拓展。但是, 我们就不再展示这些运算是如何拓展的了。

如果我们有意挑选小数使得其无限但是不循环——一会儿我们还会看到其他的方式来得到无限不循环小数, 我们就有可能得到分数之外的数。

⁴如果乘法的定义 $a \times b = \underbrace{a + a + \cdots + a}_b$ 而不是 $a \times b = \underbrace{b + b + \cdots + b}_a$, 则还得先用一下包含分数的乘法交换律。

另一个得到无限不循环小数的方式还是运算封闭性——也就是通过运算来得到新的数——例如开方。也就是定义 $a^2 = m$ ，找一个数 a 使得 $a^2 = m$ ，例如 $m = 2$ 。我们把这个数记作 $a = \sqrt{m}$ 。我们来看看这个数是不是分数。

1.2.1 从 $\sqrt{2}$ 不是分数到 $\sqrt{\frac{q}{p}}$

之前我们已经用逻辑证明了 $\sqrt{2}$ 不是分数，我们来看一看更加一般的 $\sqrt{\frac{q}{p}}$ 是不是分数。

之前的证明过程如下：假设 $\sqrt{2} = \frac{q}{p}$ 是一个分数（为了唯一性和简单性，我们假设 $\frac{q}{p}$ 是最简分数，也就是 p, q 没有公约数），则得到 $q = \sqrt{2}p$ 。接着，等式两边同时做平方得到 $q^2 = 2p$ 。我们得到结论 q 必定是偶数，因为如果 $q = 2k + 1$ 是一个奇数，则 $q^2 = (2k + 1) \times (2k + 1) = 4k^2 + 4k + 1$ 必定是奇数，而 $2p$ 是一个偶数，两者不可能相等。于是，得到 $q^2 = 2k \times 2k = 2p$ ，也就是 $2k^2 = p$ 。同样的道理，我们得到 p 肯定也是一个偶数。这样 p, q 都是偶数，就有了公约数 2。这和一开始假设的有矛盾。从一个假设开始通过每一步都严密的逻辑得到和假设矛盾的结论。这只能说明一开始的假设是错的。于是， $\sqrt{2}$ 不是一个分数。顺便，我们把这个先假设再得到矛盾的证明过程称作反证法。

我们再来试试 $\sqrt{4}$ 。我们知道它是一个分数 $\sqrt{4} = 2\frac{2}{1}$ 。我们来试着重复上面的论证过程。假设 $\sqrt{4} = \frac{q}{p}$ ，平方以后得到 $q^2 = 4p^2$ ，于是 $q = 2k$ 必须是一个整数。注意，这里不再要求 $q = 4k$ ！这个结果是一个关键。于是， $4k^2 = 4p^2$ ，也就是 $k^2 = p^2$ 。我们不再能够得到类似于 $4k^2 = p^2$ 的结论。于是，没有矛盾。

我们看到其中最关键的地方在于 $2^2 = 4$ ，于是我们只能从 $q^2 = 4p^2$ 得到 $q = 2k$ 而不是 $q = 4k$ ，于是后面也只能得到 $k^2 = p^2$ 而不是 $4k^2 = p^2$ 。

因此，很简单，对于整数 m ， $\sqrt{m}$ 是一个整数的必要条件就是 m 是一个平方数，也就是存在一个整数 n 使得 $n^2 = m$ 。同时这也是充分条件，也就是只要 $n^2 = m$ ，则 $\sqrt{m} = n$ （实际上，还有另一个整数 $-n$ 也满足要求）。

对于更加一般的分数来说， $\sqrt{\frac{q}{p}}$ 是分数的必要和充分条件很可能就是分别有两个整数满足 $m^2 = q, n^2 = p$ 。首先，这是充分条件： $\sqrt{\frac{q}{p}} = \sqrt{\frac{m^2}{n^2}} = \frac{m}{n}$ ，

这个可以通过等式两边同时做平方来验证。当然，跟之前的一样，实际上 $-\frac{m}{n}$ 也满足这个要求。其次，这是必要条件：假设其中之一不是一个平方数，另一个是平方数，则我们就回到了前面已经讨论的 $\sqrt{m}$ 的情况， $\sqrt{\frac{q}{p}}$ 肯定不是分数；假设两个都不是平方数而且没有公约数，则把 $\sqrt{\frac{q}{p}} = \frac{m}{n}$ 平方一下以后得到 $\frac{q}{p} = \frac{m^2}{n^2}$ ，最后根据最简分数的唯一性，得到 $q = m^2, p = n^2$ 。

这样通过开平方得到的无限不循环小数的问题就清楚了。那么，是不是还有其他方式来得到无限不循环小数呢？有。例如，开三次方 $\sqrt[3]{m}$ ，以及更高次方。未来我们还会遇到更多其他的得到无限不循环小数的情形。

1.2.2 无理数以及开方数之外的无理数

我们之前当作事实性知识学习过，园周长和面积公式中出现的 $\pi \approx 3.1415926$ 是一个无限不循环小数。有人的还以能够记住足够多位的数为乐或者为傲。但是，我们怎么肯定 π 真的不会循环呢？证明这个结论可不是一件简单的事情。

有兴趣的同学还可以自己进一步去了解一下欧拉数 e （自然对数的底数，利滚利的故事），Omega 常数 Ω ，黄金分割数，Apéry 常数 $\zeta(3)$ 。

分数 $\frac{q}{p}$ 可以看作是整数的比例数，同时分数是无限循环小数同时无限循环小数是分数，而无限不循环小数不能看作是分数，也就是不能看作是比例数。比例数很容易通过调整计量单位来理解。但是非比例数不能这样来理解。人们就把分数称作有理数，把无限不循环小数称作无理数。

这样，我们就从自然数拓展到了负数、整数、分数和无理数。

1.2.3 等式的性质

根据等号的含义——两个数相等就是这两个数在最终得数的意义下可替换，我们可以得到等式的性质：

$$\text{两边同加等号成立：} a = b \Rightarrow a + c = b + c, \quad (1.9a)$$

$$\text{两边同乘等号成立：} a = b \Rightarrow a \times b = b \times c. \quad (1.9b)$$

实际上，两边同乘也可以看作是两边同加的导出结论，因为乘法是重复的加法。我们没有写出来两边同减和两边同除（以一个不为零的数），因为这

两个计算分别可以和加法和乘法认同。

实际上，维持等号不变的操作远远比四则运算要多：只要两边进行的是结果唯一的操作，等号都成立。例如，我们可以做平方、立方等计算， $a = b \Rightarrow a^2 = b^2$ 。还可以把这些操作合起来，例如 $a = b \Rightarrow a^2 + 3a = b^2 + 3b$ 。一般地我们把这个等式的性质记作

$$a = b \Rightarrow f_u(a) = f_u(b). \quad (1.10)$$

其中 f_u 的含义就是一个结果唯一的操作，也就是对于给定的 x ， $f_u(x)$ 唯一。用这个记号，前面算式 (1.9) 中的加法和乘法操作可以记为

$$f_u^{(+c)}(a) = a + c, f_u^{(\times c)}(a) = a \times c. \quad (1.11)$$

习题 1.6 ($f_u^{(+c)}$ 和 $f_u^{(\times c)}$ 的唯一性). 证明 $f_u^{(+c)}$ 和 $f_u^{(\times c)}$ 的唯一性。也就是去看看如果有两个不同的数 x, y 使得 $f_u(a) = x \neq y = f_u(a)$ 会怎样。

通过这道题，我还想给你传达一个信息：不要害怕任何数学符号，数学就是给你的思考提供语言符号的，你需要表达什么，就去找相应的已经提供好的表达这个意思的符号，如果没有就自己去创造这些符号。

如果计算结果不唯一，我们就要小心。根据正数和负数的平方都是正数， $x^2 = (-x)^2 = y$ ，对 y 开平方这个操作其实可能有两个结果 x 和 $-x$ 。因此，如果我们将开平方记作 $\sqrt{y}$ ，则 $a = b \Rightarrow \sqrt{a} = \sqrt{b}$ 不一定成立。为此，人们约定 $\sqrt{y}$ 仅仅表示正的那个平方根，负的那个记作 $-\sqrt{y}$ 。这样区分以后，对 y 开平方的操作 $\sqrt{y}$ 的结果仍然唯一，因此， $a = b \Rightarrow \sqrt{a} = \sqrt{b}$ 仍然成立。

习题 1.7 ($\sqrt{\cdot}$ 唯一和不唯一). 证明 $f_u^{(\sqrt{\cdot})}$ 在不区分成 $\sqrt{\cdot}$ 仅仅表示正的那个，负的那个表示为 $-\sqrt{\cdot}$ 之前，不唯一；区分之后唯一。

1.2.4 绝对值、相反数、倒数和运算律

有了数轴和原点，我们定义了一个数 x 的绝对值，记作 $|x|$ ，为

$$|x| = \begin{cases} x & x > 0, \\ 0 & x = 0, \\ -x & x < 0, \end{cases} \quad (1.12)$$

用来表示这个点到原点的距离。这样距离就是一个非负数。我们把数轴上位于原点两侧距离相同的点对应的两个数称为互为相反数，也就是 x 的相反数是 $-x$ 。把乘起来等于 1 的两个数称为互为倒数，也就是 x 的倒数是 $\frac{1}{x}$ (注意，这时候要求 $x \neq 0$)。

通过相反数我们认同了减法和加法，也就是当 $a, b \geq 0$ 的时候，我们有

$$a - b = a + (-b). \quad (1.13)$$

我们甚至进一步发现，对于一般的 a, b 这个仍然成立。

通过倒数我们认同了除法和乘法，也就是当 $b \neq 0$ 的时候，我们有

$$a \div b = a \times \frac{1}{b}. \quad (1.14)$$

这样四则运算就只剩下了两则运算——加法和乘法。我们把运算律写成包含这两则运算的一般形式，

$$\text{加法结合律: } (a + b) + c = a + (b + c), \quad (1.15a)$$

$$\text{加法交换律: } a + b = b + a, \quad (1.15b)$$

$$\text{乘法结合律: } (a \times b) \times c = a \times (b \times c), \quad (1.15c)$$

$$\text{乘法交换律: } a \times b = b \times a, \quad (1.15d)$$

$$\text{乘法对加法的左侧分配律: } a \times (b + c) = a \times b + a \times c. \quad (1.15e)$$

这些运算律都可以基于这些运算的定义证明得到。等上了大学，我们会通过这些性质来定义运算：满足形如算式 (1.15) 这样的性质（也会稍微改变一下，例如乘法可以不要求交换律）的叫做加法和乘法。如果还满足两个运算之间的分配律的，还会给另一个专门的名称——未来会学习到叫做数域。

其实，如果你足够细心就已经发现，在前面讨论把运算从整数拓展到分数的时候，我们很可能就是冲着保持运算律这个目标去的。所以，未来我们以是否满足运算律为运算的定义就可以理解了。

1.3 数轴和用数轴表示数

把数和几何的点对应起来是数学非常重要的一步。今天，我们来初步体会这一步，或者至少为这一步做准备。未来在解析几何和平面几何部分我们会充分地体会到这个对应关系的威力。

1.3.1 非负整数的数轴表示

一方面，本来数就有描述线段长度的作用。另一方面，如果我们约定一个一定长度的线段当作一个单位长度，就可以用这个单位长度的自然数倍，也就是把所要表示的自然数那个多个单位长度累计向着右侧相连（往右侧 x 轴正方向一个接一个），来代表任何一个自然数。现在，我们来把这个线段长度和数之间的关系一般化。

有了自然数，我们来看负的自然数。利用负的自然数是正的自然数的相反数，我们就可以直接从正的自然数在数轴上的表示得到负的自然数在数轴上的表示——向着正方向累计相连之后，转过来到负的一侧。当然，实际上这也和直接在负的一侧累计相连的结果相同。

这样我们就可以在数轴上表示所有的整数了。实际上，给定一个数轴，我们通常就会标出来其正方向、单位长度和若干整数点。

1.3.2 分数的数轴表示

一个分数 $\frac{p}{q}$ 可以看作是 $p \times \frac{1}{q}$ 。于是，我们只需要先表示出来 $\frac{1}{q}$ 在数轴上，然后累计相连这一段相应的次数，就能得到分数 $\frac{p}{q}$ 在数轴上的表示。那么， $\frac{1}{q}$ 如何表示呢？我们需要把数轴上一个单位的长度均匀分成 q 份，取其中一份。这个均匀划分可以大概来做。当作示意图，这就够了。数轴绝大多数都是当作一个示意图，因此，够用了。

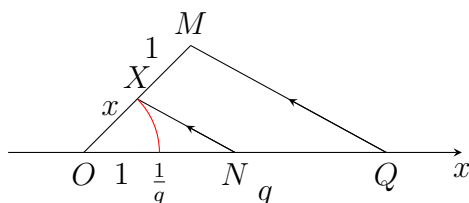


图 1.1: 分数在数轴上的表示。基于未来要学习的平行基本定理或者相似三角形的判定定理和性质，可以得到分数 $\frac{1}{q}$ 在数轴上的表示。

但是，如果我们要严格来操作呢，怎么做？下面的操作可以实现这个目标，但是，会用到后面在平面几何部分才能证明的相似三角形的判定定理和性质（见 15.5 节），或者等价地平行基本定理（见 14.2 节）。因此，在这里，你的目标是看一下，知道能做。未来再来检验是否真的可以这样做。

如图 1.1，在数轴上先找到 1 和 q 这两个整数对应的点，分别记作 N 和 Q 。过原点 O 做一条不和数轴重合的直线段，取其长度为 1，记其另一个终点为 M 。连接 QM ，过 P 做平行于 QM 的直线交 OM 于 X 。则 OX 的长度为 $x = \frac{1}{q}$ 。最后我们通过做一个以原点为圆心， OX 为半径的圆，交 x 轴于一点。那个点就是 $\frac{1}{q}$ 所对应的点。

我们运用后面才能学到的知识——尽管后面才学但是这部分知识不依赖于其所要证明的分数可以这样表示的结论，因此没有逻辑上的循环论证——来证明这个结论。

证明 1.1. 根据平行基本定理，有 $\frac{x}{1} = \frac{OX}{OM} = \frac{QN}{OQ} = \frac{1}{q}$ ，因此， $x = \frac{1}{q}$ 。

在数轴上表示分数还有一个近似的办法：先把分数转化成小数，然后，在数轴上来表示这个小数。小数的整数部分我们已经解决。现在我们只关心小数部分的表示。未来把小数部分的表示和整数部分的表示“加起来”（相当于首尾相连，见 1.4 节）就可以。

在数轴上表示纯小数的办法是，对于每一个小数位上的整数（必然是个位数，在 $0 \sim 9$ 之间，包含 $0, 9$ ），相应地把上一个小数位（对于十分位其上一个数位是个位）的一个单位均匀划分成十份，找到下一个小数位上的整数对应的位置，就行了。对于有限小数，这样做原则上是准确的。如果遇到无限小数，这样做就会没完没了。不过，随着小数位原来越高，再把那个小数位上的一个单位划分成十份就会很困难，而且如果是把数轴当作

示意图，只要小数位高到一定程度也会越来越没有必要再高下去。

这里面唯一需要的技术就是把一份分成十份。刚好，我们之前已经学会了从 q 得到 $\frac{1}{q}$ 的表示方法。这里只需要取 $q = 10$ 就行了。

1.3.3 无理数的数轴表示？

我们已经学会了严格表示一个分数，原则上严格但实际上近似地表示一个有限小数和一个无限小数。其中无限小数自然也就包含了无理数。也就是说，如果我们能够把无理数的小数位不断地写下来，到任意高位，那么，我们只需要前面表示小数的方法就可以得到无理数在数轴上的表示。也就是说，无理数的表示其实是一个没完没了的过程。在近似条件下，可以在满足表示精度要求的时候截断。但是，精确地表示一个无理数，得永远不断地一位位小数位做下去。

那问题是：既然就没有停止的时候，得不断地做下去，尽管一个无理数在任何时候截断都是数轴上的点，那么，一个无理数真的是数轴上的一个点吗？

将来在大学和研究生阶段，我们会再一次来学习数，从构建自然数开始，一直到构建实数，然后来证明一个无理数真的是一个过程，也是数轴上的一个点。这个过程称为“极限”。目前，我们暂时接受这个结论“任何一个无理数都可以表示为数轴上的一个点”。

有了所有这些数在数轴上的表示，我们就可以来研究一下运算在数轴上的表示。如果两者都可以用数轴表示，则我们实际上就有了另一套描述数和运算的语言——完全转化成用点和线以及相应的操作来表述。这套语言称为几何的语言，简称几何。

1.3.4 实数，以及数轴上还有其它数吗？

我们把所有的有理数（无限循环小数⁵）和无理数（无限不循环小数）的集合称作实数。到这里为止，实数仅仅是把有理数和无理数合在一起的那

⁵无限循环小数也包含了有限小数。有限小数的循环体是最后那一位不为零的小数位的下一位上的 0。

个集合的名称。结合前面的数的表示，我们已经可以把任意一个实数表示为数轴上的一个点了。

下面还有两个问题：这个从数到点的表示唯一吗，反过来所有点都能找到一个数来表示吗？

按照前面的找有限小数在数轴上的点的方式，肯定是唯一的。回顾一下有限小数的表示方式就能看到。如果两个有限小数对应了同一个点，则这两个有限小数在之前的小数位和整数位上对应的值完全相同，之后为零因此也相同。因此就是两个相同的有限小数。

问题在于无限小数，包含无限循环小数和无限不循环小数。不过，至少我们可以这样看：如果两个数不同，那么至少其在某一位上会不同。因此，只要我们给这两个小数找对对应点的时候不断地走，走到比这个不同的位更细致的位上，则这两个小数所对应的点至少在那个时候就不同了。因此，看起来，从那以后，这两个小数都应该不同。当然，如果后面我们要学习的“极限”的过程很神奇，可以把之前不同的变成相同的话，那就再说了。也就是，差不多，我们对于唯一性的回答是——是的，是唯一的。

我们再来回答另一个问题：任何一个点都能找到一个数，并且这个数就是实数，也就是包含有理数和无理数了吗？填满了吗？不可能再有其它数了吗？

从小数的角度来讲，有理数和无理数是一个二分法——无限循环小数（有限小数当作了它的特殊情况）和无限不循环小数。二分法是完整的，包含一切上一级的集合的元素的。也就是说，如果数轴上每个点都可以表示为一个小数，则数轴上每一个点就都可以表示为有理数和无理数。那，数轴上的每一个点都可以表示为一个小数吗？

按照小数在数轴上的表示方法——每一个当前位都把上一位的一个单位均匀分成十份，按照相应数位上的值（在 $0 \sim 9$ 之间的整数）找到相应的位置，然后继续看下一位——来说，好像也没有别的了。你看，每次你划分一个单位为十份，找到一个位置点，然后继续这个划分和找位置。看起来，数轴上的任何一个点，你都可以通过这样的方式来“偶遇”这个划分过程。但是，由于我们没有学习过极限，我们还是说不清楚这个问题。不过，看起来，好像还能想象——每个点都会对应到上面的一个划分过程。因此，我们也暂时接受，数轴上所有点都有一个对应的实数。

合起来，我们暂时接受实数和数轴上的点的一一对应关系。未来会有机会进一步来学习、质疑和思考它，尤其是思考那个永远不停止的过程和一个特定点的关系——也就是那个称为“极限”的我们当前还不清楚的概念。

1.3.5 数轴上的区间

有了数轴和实数的一一对应关系，我们引入数轴上的开区间和闭区间来表示两个实数之间的所有的实数。

定义 1.1 (开区间). 给定任意两个实数 $a \neq b$ ，不失一般性假设 $a < b$ ， a, b 之间的所有实数的集合，不包含 a, b 自身，被称为 a, b 之间的开区间，记作 (a, b) 。

定义 1.2 (闭区间). 给定任意两个实数 $a \neq b$ ，不失一般性假设 $a < b$ ， a, b 之间的所有实数的集合，包含 a, b 自身，被称为 a, b 之间的闭区间，记作 $[a, b]$ 。

类似地可以拓展到半开半闭区间，例如 $(a, b]$ 和 $[a, b)$ 。

利用数轴上的点和数的对应关系，以及集合的交集、并集、补集和包含关系的定义，还可以证明以下命题：

习题 1.8 (区间关系). 当 $a < b < c < d$ 的时候， $(b, c) \subset (a, d)$ ， $(a, c) \cap (b, d) = (b, c)$ ， $(a, c) \cup (b, d) = (a, d)$ 。

1.4 运算在数轴上的表示 *

我们已经把数表示为数轴上的点，或者说，从原点开始到数轴上的点的线段。实际上，这个线段还带方向，要么是从原点到 x 轴正方向的，要么是从原点到 x 轴负方向的。现在我们来把运算也表示为数轴上的操作。这样我们就可以把数和运算都表示为基于数轴的操作了。

实数的加法 $(a + b)$ 表示为从原点到 a 的线段和从原点到 b 的线段首尾相连——把后者起点移动一下连到前者的末尾，合起来得到的线段就是 $(a + b)$ 对应的线段。如图 1.2 所示， $2 + 3$ 表示为从原点 (O 对应着实数 0)

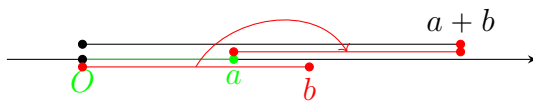


图 1.2: 加法在数轴上的表示。加法表示为首尾相连。为了避免重叠在一起, 我们把 x 轴上的线段分开呈现。

到 2 的线段, 加上——也就是首尾相连——从原点 (O 对应着实数 0) 到 3 的线段。首尾相连就是把第二条线段的起点连到第一条线段的尾部。

乘法是加法, 原则上有了加法就可以表示乘法了。整数的乘法就是重复相加。有理数的乘法可以最后表示为整数的乘法和除法 (整数的除法之前我们已经在图 1.1 展示过)。因此, 有理数乘法在数轴上的表示不是问题。那无理数的表示呢? 要么考虑用有理数去无限逼近, 要么考虑运用之前得到的无理数在数轴上的表示⁶。

如图 1.3, 在数轴上先找到 a 对应的点 A 使得 $OA = a$, 再找到 b 对应的点, 使得 $AB = b$ 。过 A 做 $AN \perp AB$, 并且 $AN = 1$ 。连接 ON 。在 B 点做 $\angle ABM = \angle ANO$, 使得这个角的一边交 NA 的延长线于 M 。 AM 的长度就是 $a \times b$ 。最后我们通过做一个以原点为圆心, AM 为半径的圆, 交 x 轴于一点。那个点就是 $a \times b$ 所对应的点。

证明 1.2. 在三角形 $\triangle ONA$ 和 $\triangle BMA$ 中, $\angle MBA = \angle ONA, \angle BAM = \angle NAO$, 则 $\triangle ONA \sim \triangle BMA$ (三角形相似 AA 判定), 因此 $\frac{NA}{AO} = \frac{BA}{AM}$ (三角形相似性质), 也就是 $NA \times AM = BA \times AO$, 也就是 $MA = \frac{AO \times AB}{AN} = \frac{a \times b}{1} = a \times b$ 。

这里我们用来还没有学习过的相似三角形的判定定理和性质 (见 15.5 节)。也可以等价用四点共圆的判定定理和圆内的相交弦定理 (见第十七章) 来证明这个结论。同样, 由于这几个用到的命题不依赖于乘法在数轴长的表示, 这里没有循环论证。

这里的证明部分你暂时不需要完全学会, 只需要留下一个映像: 加减乘除运算是可以在数轴上表示的, 就行了。等学习了平面几何你再来完成这几个证明, 见习题 15.5 和习题 17.7。

⁶其实, 我们也只是假装得到。但是, 假装一次总比假装两次好。因此, 我们把这一次运算的数轴表示建立在上一次关于无理数的数轴表示的基础上。

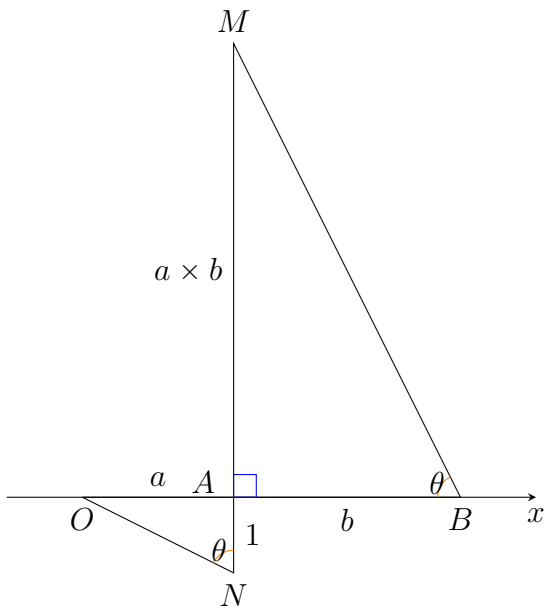


图 1.3: 乘法在数轴上的表示。基于未来要学习的相似三角形的判定定理和性质，或者圆的判定和性质，可以得到乘法 $a \times b$ 在数轴上的表示。

注意，很多其他计算，纯粹靠把数表示为点以及点对应的有向线段，然后对线段做几何操作，还是很难完成的。例如求一个数的平方根的操作。未来我们定义了函数，有了函数的几何表示——函数图像，我们会有其他办法来完成这些更加复杂的操作。

1.5 把数和运算合起来表示操作 *

我们到目前为止是这样来看数和运算的：数是自己已经存在的，运算是两个数之间，或者对于求一个数的相反数和倒数、绝对值这些运算来说，是对于一个数的运算。现在，我们来把数和运算直接联系起来——我们说数和运算合起来就是操作的表示。

例如，我们这样来看 $2 \rightarrow +2$ ：其含义是对于任何一个给定的数 a ，对这个数做 $+2$ 的操作，得到 $a + 2$ 。回到现实世界的例子，以位置为例， $+2$ 的操作意味着

- 1. 对于原来在 a 位置的某个事物，把这个事物往 x 轴正方向——通常

是右侧——移动两个单位；

2. 对于原来在 a 位置的某个事物不动，但是坐标原点做了移动，使得 a 的位置在新的坐标下的坐标是 $a + 2$ 。

目前，我们采用第一种含义。未来我们会发现第二种也是合理的。于是， -3 的含义是：对于任何一个数 a ，这个操作使得其成为 $a + (-3) = a - 3$ 。我们也顺便发现了，在数和运算的这个含义下，加法和减法的认同完全是自然的。

这样我们就把任何一个数 x 变成了一个加法操作： $a \rightarrow a + x$ ，其含义是给定任何一个数 a ，这个操作的结果是，得到 $a + x$ 。

我们来看看在这个视角下 $a + x + y$ 是什么含义：“ $a + x + y$ ”把 a 操作成了 $a + x + y$ ，这其实就是“ $x + y$ ”的作用，也就是 $+(x + y)$ 的作用。 $a + x + y$ 还可以这样来看先把 a 操作成 $a + x$ ，再把 $a + x$ 操作成 $a + x + y$ ，也就是说，相当于 $+x + y$ 。于是，我们得到 $+(x + y) = +x + y$ 。也就是两个数加起来在作用到这个数 a 上，相当于先把第一个数代表的加法操作作用到 a 上，再把第二个数代表的加法操作作用到第一个作用完的结果上。这非常自然，正好就是 $(x + y)$ 这个数的含义。

因此，当我们把 x 看作是操作 $+x$ 的表示之后， $x + y$ 的含义就是先作用 $+x$ 再作用 $+y$ 这两个操作的结合——数的加法自然成了操作的结合。

有了这个理解，我们再来看数轴上的数的加法： $+2$ （也就是 2）表示一个操作，对于任何数往右边移动 2 个单位长度， $+3$ （也就是 3）表示另一个操作，对于任何数往右边移动 3 个单位长度，于是， $+2 + 3$ （也就是 $2 + 3$ ）也就自然地表示先把任何一个其所作用的数先右边移动 2 个单位长度再往右边移动 3 个单位长度，合起来往右边移动 $2 + 3$ 个单位长度。因此，两个数的加法 $(2 + 3)$ 操作成了两个数各自的操作的合并：数的加法成了数轴上的首尾相接（ $+3$ 的操作接着 $+2$ 的操作的结果继续操作）。

注意到负数是从原点指向 x 轴负方向的数，因此， $+(-3)$ （也就是 -3 ）代表了把一个数往左侧移动 3 个单位长度的操作。上面的首尾相接代表操作的合并也可以用于有负数的加法。例如， $+2 + (-3)$ （也就是 $2 + (-3) = 2 - 3$ ）也就自然地表示先把任何一个其所作用的数先右边移动 2 个单位长度再往左边移动 3 个单位长度，合起来往右边移动 $2 - 3$ 个单位长度。因此，两个

数的加法 ($2 + (-3)$) 操作成了两个数各自的操作的合并：数的加法成了数轴上的首尾相接 ($+(-3)$ 的操作接着 $+2$ 的操作的结果继续操作)。

有两个负数的时候，两个数的加法成了这两个数各自做代表的操作的首尾相接，仍然成立。

未来这个数和运算合起来表示操作，数的加法成了首尾相接，会成为矢量和矢量加法的基础。

数学对象可以表示操作——操作是一种关系，将来也会成为更加抽象的数学的基础。

1.6 平面直角坐标系上的点

既然一条直线上的点和实数是一一对应的，那么，是不是可以给整个平面铺满直线来代表平面上的所有点呢？

比如说，如图 1.4 所示，我们取每一条直线都跟 x 轴平行，其中每一条这样的直线按照某种方式来编号，例如记作 y （以后会看到为什么会用 y ），把这样的直线铺满整个平面。然后我们说这条标记为 y 的直线上的那个叫做 x 的点，这样就得到了标记了这个点的一对数 (y, x) （以后我们会交换一下顺序）。

这样的平行于 x 轴的直线用 y 的值来标记，而这些可以看作是这些直线在和 x 轴垂直的直线的交点，如图 1.4 所示。于是，我们就把其中经过原点的垂直于 x 轴的，在图 1.4 中用来显示标记的，称为 y 轴。

习惯上，我们把 x_P 写前面， y_P 写后面。这相当于如图 1.5 所示，我们先用平行于 y 轴的用 x 来标记直线铺满平面，然后在某一条记号为 $x = x_P$ 的这样的直线上选一个由 y_P 所标记的点，合起来记作 $P = (x_P, y_P)$ 。

这样我们就得到了平面上的点 P 和一对数 (x_P, y_P) 的一一对应关系。从平面上的给定点 P 找一对数 (x_P, y_P) 的方法是：过 P 点分别做垂直于 x 轴和 y 轴的直线，其分别会和 x 和 y 轴相交于 x_P 和 y_P 处。从给定的一对数 (x_P, y_P) 找到平面上对应点 P 的方法是：在 x 轴上的 x_P 处做垂直于 x 轴的直线，在 y 轴上的 y_P 处做垂直于 y 轴的直线，两条垂线的交点就是 P 点。未来我们学习几何学之后就可以明白垂线如何做，并且是唯一的，因此，两个过程所确定的数值和交点也唯一。

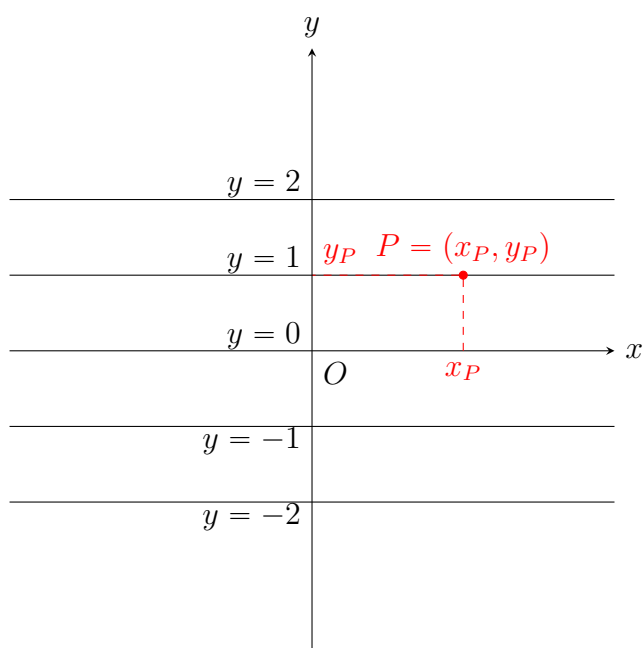


图 1.4: 平面上的点用一对数来表示。让平行于 x 轴的直线铺满平面，用 y 值来标记这些直线，然后在这个特定的 y 值为 y_P 的直线上取一个对应着 x_P 的点，就可以确定一个平面上的点。

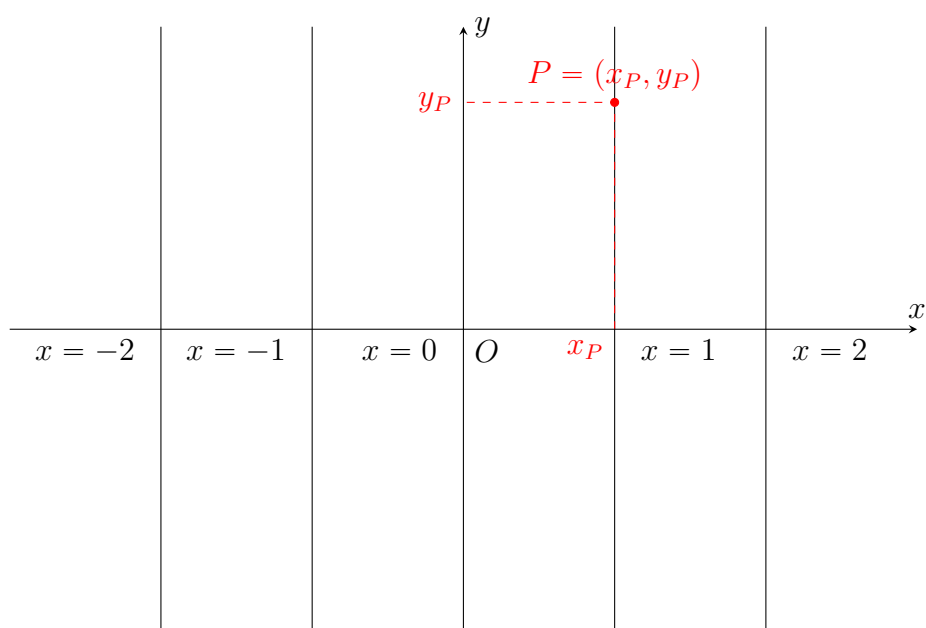


图 1.5: 平面上的点用一对数来表示。让平行于 y 轴的直线铺满平面，用 x 值来标记这些直线，然后在这个特定的 x 值为 x_P 的直线上取一个对应着 y_P 的点，就可以确定一个平面上的点。

1.7 代数的思想——用符号表示对象

对于如下的自然语言，我们可以分别用下面的数学符号来表示，

- “篮子里面有两个苹果”： $Q_a^{in} = 2$ ，或者 量_{苹果}^{篮内} = 2，或者 $L_P^{ln} = 2$ ，或者 $A^i = 2$
- “篮子外面有三个苹果”： $Q_a^{out} = 3$ ，或者 量_{苹果}^{篮外} = 3，或者 $L_P^{lw} = 3$ ，或者 $A^o = 2$
- “我家有四个人——爸爸、妈妈、我和妹妹”： $F_I = \{D, M, S, I\}$ ， $|F_I| = 4$ ，或者 $P^{F_I} = 4$
- “农场里面有猪、牛、羊、萝卜和西瓜”： $F = \{P, C, L, R, W\}$ ，或者 $F = \{, , , , \}$ ，或者 $F = \{P, C, L, R, W\}$
- “农场里面有两只猪、一头牛、四只羊、一百斤萝卜和三个西瓜”， $F = \{P, C, L, R, W\}$ ， $P^F = 2$ (只)， $C^F = 1$ (头)， $L^F = 4$ (只)， $R^F = 100$ (斤)， $W^F = 3$ (个)

我们一个个来分析，这些符号都表达了什么，怎么表达的。 $Q_a^{in} = 2$ 的 Q 是英文单次数量的意思，上标 in 表示“在篮子里面的意思”，其中“篮子”可以被放进去 ($in \rightarrow InB$ ，这里“B”是篮子的英文单次的首字母)也可以省略，下标 a 是英文单词 *apple* 的首字母。因此，左边的含义就是“篮子内苹果的数量”。剩下的右边是这个数量的具体数，中间是等号。

我们再来看 量_{苹果}^{篮内} = 2。“量”是“数量”的中文词，上标“篮内”就表示“篮子内”，下标“苹果”就表示“苹果”，合起来，左边就是“篮子内苹果的数量”。剩下的右边是这个数量的具体数，中间是等号。

$L_P^{ln} = 2$ 和 量_{苹果}^{篮内} = 2，就是把每一个汉字变成了这个汉字的拼音的首字母。

$A^i = 2$ 值得单独说明一下。当我们本来就是为了表示数量关系的时候，每次都出现 Q 、“量”这些符号有点多余，而且无论是什么东西的数量都要出现这些符号。因此，不如直接就抛弃这些冗余信息。这样就成了“ A ”代表“苹果”，上标“ i ”代表“内”（篮子也省略了，因为下面还有一个类似

的信息，主要强调篮子内外)，下标就用不着了，留下来以后用来表示其他信息。这样 $A^i = 2$ 仍然表示了“篮子内的苹果的数量是 2”。

第二个例子完全是对第一个例子的重复，就不再解释了。

我们来看第三个例子。第一式 $F_I = \{D, M, S, I\}$ 的左边“ F_I ”表示“我(I)的家(Family)”，右边是一个集合的表达式，列出来所有的元素：爸爸(Dad)、妈妈(Mom)、我(I)和妹妹(Sister)。第二式 $|F_I| = 4$ 表示“我家里的人的这个集合的大小是 4”。合起来，两句话表达了“我家里有四个人——爸爸、妈妈、我和妹妹”。其中的第二式，按照前两个例子可以表达为 $P^{F_I} = 4$ ，其中 P 是人的英文单次(Person)的首字母，上标 F_I 表示“我(I)的家(Family)”。左边合起来就是“我家里的人的数量”。右边是数 4，中间是等号，表示“我家里的人的数量是 4”。

第四个例子是前面三个例子的组合。也不再详细展开解释。

现在我们来对这些例子中的数学表达式做一个总结。首先，数学可以描述“我的爸爸、我的妈妈、我的妹妹、我、农场里的猪”等等具体的事物。其次，数学还可以描述这些事物在某个单位下的量，例如“胡萝卜的斤数”、“西瓜的个数”、“苹果的个数”。再次，数学可以描述前者中那些量的具体数值。最后，数学还可以描述一个关系表达式，或者说用关系表达式来表达一个意思，例如表达式

$$A^i = 2 \text{ (个)} \quad (1.16)$$

表示“篮子里苹果的数量”“是”“2”。这是四个不同层次的数学对象：事物(通常用名称来代表)、事物在某个单位下的量这个名称(也就是算式(1.16)的左半部分)、事物在某个单位下的量的具体数值(通常表示为数，或者“数 + 单位”)以及关系表达式(这里我们看到左边是符号，右边是数，中间是等号)。

把具体数和这个数的含义分开表达，都明确出来，是一个很好的数学中描述对象的习惯。这样分开具体数和这个数的含义之后，我们只需要关心左边的表示含义的符号，未来我们可以用这个具体数的任何值代入到这个符号里面去。例如，如果我们知道苹果每一个五元钱——表示为

$$P_A = 5 \text{ (元/个)}, \quad (1.17)$$

则无论篮子里面有多少苹果，

$$\text{¥} = A^i \times P_A \quad (1.18)$$

都能给你正确的购买总的那一篮子苹果所需要的钱的数量。

有的时候，甚至我们会给某些表达式一个专门的符号或者名字，例如我们给了算式 (1.16)、算式 (1.17)、算式 (1.18) 它们自己的方程的编号，甚至我们可以叫它们分别为“苹果的数量表达式”、“苹果的价格表达式”、“买苹果的钱的计算公式”。这样我们就把关系表达式都有了自己的名称。

有的时候，我们还会对不同的关系表达式作计算。例如，算式 (1.18) 实际上就是把算式 (1.16) 和算式 (1.17) 做了相乘。这就是把表达式看作对象以后的好处，我们可以操作这些表达式。

总结一下，第一、“用符号表示对象”中的“对象”在数学有多个层次，目前看来至少有有以下四个层次：“数、量、物、算式”四个层次的对象。未来还会有什么未来遇到再加入进来。第二、在构建数学表达式的时候，我们往往把对象所对应的数和这个对象的含义写成等号的两侧。未来遇到不等号等其他关系运算符的时候也是如此。

把对象的含义（以及代表这个含义的符号）和对象对应的数分开是非常有深意的。这被称为代数的思想——关注数对应的对象的含义，而不是关注数。我们很快将开始我们的代数之旅。其中一个目的就是深刻体会代数的思想，学会运用代数的思想。

还有一个要注意的小地方：在本书中，我们尽量把单位和数写在一起代表量，也就是说，我们倾向于用 $A^i = 2$ （个）或者 $A^i = 2\text{个}$ ，而不是 $A^i = 2$ 。这不仅仅是一个习惯问题。养成带着单位做计算和思考的习惯对于将来学习科学，构建现实世界的数学模型有好处。例如，如果你带着单位作计算就会发现，其实算式 (1.18) 给出的最后的量的单位正好是“元”。你可能需要注意到（元/个）就是分数 $\frac{\text{元}}{\text{个}}$ 。

习题 1.9 (表达式的运算). 按照前面给出的 A^i 和 A^0 的含义，猜一猜 $A^T = A^i + A^0$ 的含义是什么？

习题 1.10 (练习“数、量、物、算式”四个层次的数学对象). 参考书里面的例子，你自己想一些其他的例子，表示成“数、量、物、算式”这四个层次的对象。

1.8 推荐阅读材料

1.9 作业

习题 1.11 (家里的人). 把你之前得到的你的家人、你爸爸的家人、你妈妈的家人、你爷爷的家人的集合表示出来，算一算这些集合以及这些集合的交集、并集、补集（注意算补集的时候一定要先确定全集）它们的大小。用符号表示出来。

习题 1.12 (对硬币的操作). 你有一个理想的硬币——只有两面忽略其立起来的情况，你对这个硬币的状态可以进行哪些操作，列出来，写成一个集合。

习题 1.13 (对骰子的操作). 你有一个理想的骰子——只有六个面忽略其它状态，你对这个骰子的状态可以进行哪些操作，列出来，写成一个集合。

习题 1.14 (硬币状态的概率). 你有一个理想的硬币——只有两面忽略其立起来的情况，并且这个硬币在扔出去以后每一面出现的概率相同。请你发明一套符号来表示这件事情。注意，你不需要之前就学习过概率论。你就按照你自己想要的符号的样子来发明你的符号。

习题 1.15 (硬币状态概率的检验). 你有一个真实的硬币。请你去扔上若干次，最好每次的次数不同，例如先扔两次，再扔十次、一百次、一千次、一万次，记录下来正面出现的次数，算一下正面的比例是多少。基于这个观察和计算，你来看看下一次扔出去这个硬币，正面的可能性高一点还是反面的可能性高？注意，你不需要之前就学习过概率论。你就按照你自己想要的符号的样子来发明你的符号，用你自己的到的数据来计算。

1.10 本章小结

在这一章中，我们复习了集合、数和运算，然后把数和运算都表示在了数轴上。然后，我们把数轴拓展到平面直角坐标系，以及用平面直角坐标系来表示所有平面上的点。这一部分我们花了大量的时间和精力挑战来说明，利用数轴——也就是带方向的线段和对这些线段的操作，也就是几何——

可以表示所有的数和相应的这些数的加法乘法运算。

当然，我们也留下了几个问题：实数真的和数轴上的点一一对应吗，实数是一个数轴上不断地找对应点的过程还是一个数轴上的点；我们在构建分数和乘法在数轴上的表示的过程中，用到了一些相似三角形等几何知识，这部分几何知识真的不依赖于今天所提到的分数和乘法可以表示在数轴上吗？这些问题留给你后面学习，或者自己探索。注意，在数学学习中，你没有必要接受任何不能证明的结论。

最后，我们学习了用符号表示对象，而数学中的对象包含了数、量、物、表达式这四个层次。这体现了代数的思想，也是在为后续学习多项式做准备。

用符号表示对象这部分的更深层的目的是帮助你看到数学的研究对象可以是任何可以表示为含义明确的符号的东西，并且这个符号你也可以自己来发明。你可以看到，在本章中我就发明了不少自己的符号。当然，往往发明完之后，你会发现，你发明的都已经有了。这个时候，你就回到前人已经常用的符号。

我甚至怀疑用几何方式表示分数和乘法也是我发明的。我不确定，大概率应该不是的。这个发明当然仅仅是为了数学教学，而不是数学研究，算不得什么。不过，我也不打算去考证是不是我自己发明的。如果有前人已经发明过，那就我恰好也再一次发现了它而已。

数学就是促进你思考，增加长度，做到含义明确的语言，往往因为我们要用思维描述世界也就成了描述世界的语言。“数学就是那个你需要它——这种语言——的时候就去创造出来的东西”——据彭桓武转述于吴金闪的来自于Paul Dirac（保罗·狄拉克）的一句话。

第二章 幂运算

有了实数以及实数之间的加法和乘法运算，我们来把另一个重要计算——一个数的正整数次方——推广到一个实数的实数次方，也称为一个实数的实数次幂。从中我们也可以再一次体会到数学知识的系统性。

2.1 数的有理数次幂运算

之前我们已经学习了实数的乘法。从实数的乘法开始，我们可以定义一个实数的平方、三次方，以及任意正整数次方，也就是对于任意正整数 n ,

$$a^n = \underbrace{a \times a \times \cdots \times a}_n. \quad (2.1)$$

正整数范围内的幂次方有如下的运算规律，

$$a^m \times a^n = a^{m+n}. \quad (2.2)$$

$$(a^m)^n = a^{m \times n}. \quad (2.3)$$

我们只要用上面的定义和乘法的结合律就能证明这两个等式。

$$a^m \times a^n = \underbrace{a \times a \times \cdots \times a}_m \times \underbrace{a \times a \times \cdots \times a}_n = \underbrace{a \times a \times \cdots \times a}_{m+n} = a^{m+n}. \quad (2.4)$$

算式 (2.3) 的证明留作练习。

习题 2.1 ($(a^m)^n = a^{m \times n}$ 的证明). 证明 $(a^m)^n = a^{m \times n}$ 。

注意, 这两个性质实际上不独立: 第二个是第一个的多次重复。

当 $a \neq 0$ 的时候 (当 $a = 0$ 的时候, 我们暂时只能定义其正整数次方, 因为负整数次方会牵涉到倒数, 而我们约定零不能做分母), 有了实数的正整数次方, 我们可以来定义其负整数次方的含义,

$$a^{-n} = \underbrace{a^{-1} \times a^{-1} \times \cdots \times a^{-1}}_n. \quad (2.5)$$

于是, 我们只需要定义 $b = a^{-1}$ 就可以。我们把 $b = a^{-1}$ 定义为和 a^1 相乘等于 1 的数,

$$b \times a^1 = 1 \rightarrow b = \frac{1}{a}. \quad (2.6)$$

我们也可以直接定义 $c = a^{-n}$ 为 a^n 的倒数, 也就是

$$c \times a^n = 1 \rightarrow b = \frac{1}{a^n}. \quad (2.7)$$

这样定义的 a^{-n} 仍然满足原来在正整数范围内适用的幂次方的运算规律算式 (2.2), 无论对于正的还是负的 m, n 。我们来证明一下这个结论。

我们先假设整数 $n > 0$, $m < 0$ 是负的整数的情况。不失一般性, 我们假设 $|m| < n$,

$$a^m \times a^n = a^{-|m|} \times a^{|n|} = \frac{a^n}{a^{|m|}} = a^{n-|m|} = a^{m+n}, \quad (2.8)$$

倒数第二步用了分数的性质——分子分母同时除以 $a^{|m|}$ 。当 m, n 同时小于零的时候,

$$a^m \times a^n = a^{-|m|} \times a^{-|n|} = \frac{1}{a^{|m|}} \times \frac{1}{a^{|n|}} = \frac{1}{a^{|m|+|n|}} = a^{-(|m|+|n|)} = a^{m+n}. \quad (2.9)$$

我们再来把正负整数方拓展到零次方。我们还是约定算式 (2.2) 对包含零次方的算式也适用。则,

$$a^m \times a^0 = a^{m+0} = a^m. \quad (2.10)$$

因此, 看起来应该有

$$a^0 = 1. \quad (2.11)$$

有了 a 的所有整数的幂次方的定义, 我们来考虑分数次方。我们约定算式 (2.3) 对于分数次方仍然适用。我们先来试试最简单的分数次方, $a^{\frac{1}{n}}$ 。其中 n 为大于零的整数。按照算式 (2.3) 我们应该有

$$\left(a^{\frac{1}{n}}\right)^n = a^{\frac{1}{n} \times n} = a^1 = a, \quad (2.12)$$

因此,

$$a^{\frac{1}{n}} = \sqrt[n]{a}, \quad (2.13)$$

但是, 这里有个小问题: 对一个数开 n 次方, 有可能会得到多个值。例如, 一个正数的二次方根就有两个。为了解决这个问题, 人们约定对于正数 a 和偶数 n , $\sqrt[n]{a}$ 是那个大于零的那个。我们把下面的结论的证明当作练习: 对于奇数 n 在实数范围内, $\sqrt[n]{a}$ 是唯一的; 对于负数 a 和偶数 n , 则实数范围内无解。有了这个结论和前面的约定, $\sqrt[n]{a}$ 就唯一了。

习题 2.2 (多次方根的问题). 对于奇数 n 在实数范围内, $\sqrt[n]{a}$ 是唯一的; 对于负数 a 和偶数 n , 则实数范围内没有对应值。

对于一般的分数次幂, $a^{\frac{m}{n}}$, 既然 $a^{\frac{1}{n}}$ 有定义了, 那再次保证算式 (2.3) 的适用性, 就有

$$a^{\frac{m}{n}} = \left(a^{\frac{1}{n}}\right)^m = \left(\sqrt[n]{a}\right)^m. \quad (2.14)$$

同样, 对于负的分数的, 我们仍然去按照保证算式 (2.2) 适用来定义就行, 也就是仍然有

$$a^{-q} \times a^q = 1 \rightarrow a^{-q} = \frac{1}{a^q}. \quad (2.15)$$

至此, 我们定义了 a 的任意有理数次方。

2.2 数的无理数次幂

那能定义实数的无理数次幂吗? 无理数 q 在任何一位阶段都是一个有限小数 q_L ——表示小数的位数为 L , 也就是有理数。拿着这个有限小数 q_L , 我们总是可以得到在这个有限小数下的结果 a^{q_L} 。那么, 是不是随着 L 的增加, a^{q_L} 也会越来越接近一个数呢? 我们已经知道尽管是暂时接受的 q_L 会越来越接近 q (L 越大小数位越高, 能够有的差别 $q_L - q$ 越小。我们还知道

这个过程的名字叫“极限”，尽管我们还不知道这个过程在数学上如何操作如何表示)。我们来算一下在得到 q_L 之后再下一位可能造成的差别，

$$q_{L+1} - q_L < 10^{-L}. \quad (2.16)$$

这个数会随着 L 变大越来越小。因此，看起来很可能 a^{q_L} 也会越来越接近一个数。如果确是如此，则我们把这个数称为 $a^{q_L} \rightarrow a^q$ 。

这是实数的无理数次幂的定义。实际计算的时候我们可以依靠计算机来完成。下面，在表2.1中，我们列出来了几个计算机算出来的值， $a^2, a^3, a^{\frac{3}{2}}, a^{\frac{2}{3}}, a^{\sqrt{2}}$ ，其中 $a = 2, 1, 0, -2, -1$ 。注意这里我们没有包含负数次幂，因为有了算式(2.15)，负数次幂的结果可以从正数次幂的结果计算出来。

表 2.1: **实数的幂 (a^q) 的值的几个例子**。列出来几个用计算机算出来的实数的幂 (a^q) 的值。小数部分最多保留了小数点之后三位。其中 *NA* 表示在实数范围内没有对应值。

a	a^2	a^3	$a^{\frac{3}{2}}$	$a^{\frac{2}{3}}$	$a^{\sqrt{2}}$	a^0
2	4	8	2.828	1.587	2.665	1
1	1	1	1	1	1	1
0	0	0	0	0	0	NA
-1	1	-1	NA	1	NA	1
-2	4	-8	NA	1.587	NA	1

2.3 关于平方和立方的几个公式

有了实数的幂次方，我们来把幂次运算和之前的加法、乘法混起来，来研究几个后面会用到的特定的幂次方的公式。为了符号简单，我们把乘号简化，也就是

$$a \times b \rightarrow a \cdot b, \quad (2.17)$$

或者更简化的符号，

$$a \times b \rightarrow ab. \quad (2.18)$$

定理 2.1: 完全平方公式

完全平方公式] $(a + b)^2 = a^2 + 2ab + b^2$ 。

证明 2.1.

$$\begin{aligned}
 (a + b)^2 &= (a + b) \times (a + b), \\
 &= a^2 + ab + ba + b^2, \\
 &= a^2 + 2ab + b^2. \quad (2.19)
 \end{aligned}$$

如果我们把其中的 b 换成 $-b$, 可以得到

$$(a + (-b))^2 = a^2 + 2a(-b) + (-b)^2 = a^2 - 2ab + b^2. \quad (2.20)$$

这两个完全平方经常被合起来记作,

$$(a \pm b)^2 = a^2 \pm 2ab + b^2. \quad (2.21)$$

定理 2.2: 完全立方公式

完全立方公式] $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$ 。

证明 2.2.

$$\begin{aligned}
 (a + b)^3 &= (a + b) \times (a + b) \times (a + b), \\
 &= (a^2 + ab + ba + b^2) \times (a + b), \\
 &= a^3 + a^2b + aba + ab^2 + ba^2 + bab + b^2a + b^3, \\
 &= a^3 + 3a^2b + 3ab^2 + b^3. \quad (2.22)
 \end{aligned}$$

定理 2.3: 平方差公式

平方差公式] $(a + b)(a - b) = a^2 - b^2$ 。

证明 2.3.

$$\begin{aligned}(a + b)(a - b) &= a^2 - ab + ba - b^2, \\ &= a^2 - b^2.\end{aligned}\quad (2.23)$$

定理 2.4: 立方和公式

立方和公式] $(a + b)(a^2 - ab + b^2) = a^3 + b^3$ 。

证明 2.4.

$$\begin{aligned}(a + b)(a^2 - ab + b^2) &= a^3 - a^2b + ab^2 + ba^2 - bab + b^3, \\ &= a^3 + b^3.\end{aligned}\quad (2.24)$$

如果我们把其中的 b 换成 $-b$ ，可以得到

$$\begin{aligned}&(a + (-b))(a^2 - a(-b) + (-b)^2), \\ &= (a - b)(a^2 + ab + b^2) = a^3 + (-b)^3 = a^3 - b^3.\end{aligned}\quad (2.25)$$

有时候这也被称为立方差公式。实际上，这个公式和立方和公式不是独立的，只需要把其中的 b 换成 $-b$ 就可以得到。

2.4 推荐阅读材料

2.5 作业

习题 2.3 (另一个完全立方公式). 证明另一个完全立方公式: $(a - b)^3 = a^3 - 3a^2b + 3ab^2 + b^3$ 。

习题 2.4 (用计算机作幂次方的计算). 自己编程，或者用数学计算软件来把表 2.1 中的结果验算一遍。

习题 2.5 (用计算机作幂次方的画图). 自己编程, 或者用数学软件来选取一些 a 的值 (推荐多选几个), 计算 $a^3, a^{\frac{4}{3}}, a^{\frac{3}{4}}, a^{\sqrt{3}}$ 的结果, 并尝试把这些结果连成一条曲线。如果你的计算机数学软件支持直接画图, 你也可以画出来图, 跟你自己连出来的曲线比较一下。

2.6 本章小结

在本章中, 我们学习了实数的实数次幂。在这个过程中, 我们从实数的正整数次幂开始, 向着保证幂次方运算律 $a^p a^q = a^{p+q}$ 对于任何 p, q 都成立的目标, 拓展到了实数的负整数、零、正分数次幂和负分数次幂, 最后到实数次幂。最后我们证明了几个常用公式: 完全平方公式、完全立方公式、平方差公式、立方和公式。

通过这个拓展的过程, 我们希望你能感受到: 数学不是对规则的记忆和使用, 规则的定义本身是讲道理的, 规则的出现是有动机的。也就是, 数学知识是有系统性的。

在这一章中, 我们也留下了一个尾巴: 从实数的有理数次方到实数的无理数次方, 我们还是用到了用有理数逼近无理数的这个我们目前还没有真的学习的“极限”过程。这个问题和之前留下来的数轴上的无理数的表示的问题是同一个。将来大学阶段学习了这个极限过程的概念和操作, 我们才能真正回答这个问题。

第三章 变量和常数

有了代数的思想，我们就需要对常量和变量做一个区分，知道在一个代数表达式中，哪些是量——未来可以被具体化为任何一个数，哪些是数——原则上或者实际上就已经确定的具体的数。这一章，我们来完成这个任务。

3.1 完全已知的数到给定的任意的数

让我们从计算购买一定数量的苹果需要的总的钱数的算式 (1.18) 开始。我们把算式 (1.18) 重新写成，

$$V = P \times Q. \quad (3.1)$$

其中 V 表示购买苹果的总价 (Value 的首字母)， P 表示苹果的单价 (Price 的首字母)， Q 表示苹果的数量 (Quantity 的首字母)。也可以写成，

$$Z = J \times S \quad (3.2)$$

用汉语拼音的首字母 (分别代表总价，价格，数量) 来取代英文单词的首字母。

当我们已知所要购买的苹果的数量——例如三个或者两斤，也知道某一家卖苹果的商店的苹果的单价——例如五元每个或者十元每斤，则我们

就可以得到以这样的价格来购买这些苹果的总价，

$$V_1 = 3 \text{ (个)} \times 5 \left(\frac{\text{元}}{\text{个}} \right) = 15 \text{ (元)}, \quad (3.3)$$

$$V_2 = 2 \text{ (斤)} \times 10 \left(\frac{\text{元}}{\text{斤}} \right) = 20 \text{ (元)}. \quad (3.4)$$

这里，所有的数——尽管写成 P, Q, V 的数学符号——都完全已知。

如果我们要去多家不同的苹果商店，其卖苹果的单价不同，购买相同的确定数量的苹果，例如 3 个或者 2 斤，则上面的表达式成为

$$V_1 = 3 \text{ (个)} \times P \left(\frac{\text{元}}{\text{个}} \right) = 3P \text{ (元)}, \quad (3.5)$$

$$V_2 = 2 \text{ (斤)} \times P \left(\frac{\text{元}}{\text{斤}} \right) = 2P \text{ (元)}. \quad (3.6)$$

其中 P 未来可以被替换为代表苹果单价的一个具体数。

如果我们要去同一家的苹果商店，其卖苹果的单价只有一个，例如十元每斤，我们可以购买不同数量的苹果，则上面的表达式成为

$$V = 10 \left(\frac{\text{元}}{\text{斤}} \right) \times Q \text{ (斤)} = 10Q \text{ (元)}. \quad (3.7)$$

其中 Q 未来可以被替换为代表所购买的苹果的量的一个具体数。

注意，这里的所有的字母的含义都是相应的量，都可以未来用一个这个量对应的具体的数代入进去。也就是说，这些字母不过就是未来所代入的数所要替换的相应的位置的表示。这个时候，我们说，尽管写成了字母的样子，但是其含义是完全已知的数，或者任意给定的数。

3.2 未知的或者重点关注的数

但是，当我们要去给一家商店设计一个付款计算器的时候，我们其实不知道未来这家商店所卖的苹果的价格和顾客所购买的苹果的数量，甚至，这家商店也不仅仅卖苹果。这个时候，我们说，对于给定的产品和这个产品的给定的价格，例如十元每斤，则顾客购买数量是一个待定的变量。我们希望我们的付款计算器可以用于顾客购买任何数量的这个给定价格的给定产品的情况，则

$$V = 10Q \text{ (元)} \quad (3.8)$$

中的 Q 是一个变量——具体来一个决定了买多少拿了多少来称的顾客之后，这个 Q 才成为一个具体的数值，例如两斤。

更进一步，如果也允许商品变，为了适用于多种价格的多种商品，

$$V = PQ \quad (3.9)$$

中的 P, Q 都是变量。只有将来遇到一个决定了买什么买多少的顾客之后，这些 P, Q 才会成为具体的数。

我们把这样的未来可以成为但是现在还不是一个个不同的可能的具体数的，我们当前特别关注的量称为变量。

另一种情况是，如果有一个顾客总共的预算是 100 元，不管苹果价格是多少，这个顾客都希望把所有预算换成苹果。这个时候，我们还是有

$$100(\text{元}) = V = PQ, \quad (3.10)$$

但是，这个时候总结 V 是一个确定值。如果这个时候顾客进去购买苹果的商店的苹果价格是 10 元每斤，那么，其购买的数量就是，

$$100(\text{元}) = V = 10\left(\frac{\text{元}}{\text{斤}}\right) Q(\text{斤}), \quad (3.11)$$

如果我们需要知道 Q 的具体数值，则等式两边同时乘以 $\frac{1}{10}$ 得到，

$$Q = 100 \times \frac{1}{10}(\text{斤}) = 10(\text{斤}). \quad (3.12)$$

我们也可以来面对这个场景：小明花了 100 块钱，去一家商店买了 10 斤苹果，但是没注意看苹果的价格，你能帮他看看他所购买的苹果的价格吗？这个时候，我们有

$$100(\text{元}) = V = P\left(\frac{\text{元}}{\text{斤}}\right) 10(\text{斤}), \quad (3.13)$$

如果我们需要知道 P 的具体数值，则等式两边同时乘以 $\frac{1}{10}$ 得到，

$$P = 100 \times \frac{1}{10}(\text{元}) = 10\left(\frac{\text{元}}{\text{斤}}\right). \quad (3.14)$$

我们把这样的其实对应着确定的具体数的，但是目前具体数值还是未知的量，当作（这个量的）未知数。

为了对这些数和量有所区别,我们引入下面一套习惯上经常用的符号:用 x, y, z 表示未知数或者变量,用 a, b, c 表示给定的任意的数。实际上,这套符号不能区分未知数和变量。如果我们愿意,我们可以用 x, y, z 表示未知数,用希腊字母 ξ, η 表示变量,用 a, b, c 表示给定的任意的数,例如 $\xi^2 + x = a$ 。或者,我们每次在使用中明确一下,或者自然就明确,哪些是变量哪些是未知数,哪些是给定数。

对于未知数也确实可以做变量来理解。未知数一般出现在方程中。我们把方程分成出现未知数的一侧,记作 $f(x)$,和不出现未知数的一侧,记作给定数 a ,也就是

$$f(x) = a. \quad (3.15)$$

这个时候,对于任何一个 x ——这个时候的 x 被看作了一个变量, $f(x)$ 都会有一个值,所谓方程的解,就是这个值刚好和 a 相等。也就是我们把找到未知数的值的过程看作是对一个变量不断地取值代进去作计算,一直到左侧算出来的结果刚好就是右侧的数。在这个意义下,未知数和变量就不需要去区分了。因此,上面的符号习惯也够用。

顺便,出现在包含未知数和变量的表达式里面的给定数被称为常数。这样我们就有了常数、变量和未知数。当我们要做一个一般性的商品结算计算器的时候,表达式 $V = PQ$ 里面的 P, Q 都是变量。未来我们还会知道这个时候 V 也叫做变量——因变量,按照别的变量的变化而变化的量。那个时候,为了区分因变量和自己可以变的变量,我们把自己可以变的变量称作自变量。

3.3 单项式和多项式的元数和阶数

按照上一节约定的符号习惯, $ax^3 + bx^2 + cx + d$ 中 x 是变量或者未知数, a, b, c, d 是常数,是给定的任意数。 $x^3 + 2x^2 + 4x + 3$ 也是如此, x 是变量或者未知数, $a = 1, b = 2, c = 4, d = 3$ 是常数。显然,我们主要关注的是变量或者未知数,而不是常数。因此,我们把 $ax^3 + bx^2 + cx + d$ 这样的,由一个或者多个 x^n 次方—— n 约定为非负整数——组成的数学表达式,称作关于 x 的多项式。

如果一个关于 x 的多项式其实只出现了 x 的某一次方项 x^{n^*} ，则称为 x 的单项式。例如 $2x, 3x^2, (a+b)x^3, a^6bx^5, x^2ax^3 = ax^5$ 都是 x 的单项式。

对于单项式 ax^n —— n 约定为非负整数，其中的 n 称为这个单项式的阶，其中的 a 称为这个单项式的系数常数，简称系数。

如果一个多项式中出现了多个变量或者未知数，则把买一个变量和未知数单独处理。也就是说 ax^2y^3 看作是关于 x 的 $(ay^3)x^2$ 和关于 y 的 $(ax^2)y^3$ 。

因此， ax^2y^3 是关于 x 的单项式，是关于 y 的单项式，因此也被称为关于 x, y 的单项式。对于包含多个变量或者未知数单项式，我们把所有变量的阶数之和称为这个单项式的阶。因此 ax^2y^3 是一个 $2+3=5$ 阶单项式，而 ax^4 是一个单项式。其中的变量或者未知数的个数被称为这个单项式的元。例如 ax^2y^3 是二元五阶单项式，而 ax^4 是一元四阶单项式。

如果两个单项式出现的每一个变量和每一个变量的阶都完全相同，则称这两个单项式为同类项，也就是同一个类型的单项式。可以把常量系数相加得到一个合起来的单项式。这称为合并同类项。例如 $ax + bx = (a+b)x$ ， $2x^2y + 4yx^2 = 6x^2y$ 。但是，我们不能把 $2x^2y$ 和 $3xy^2$ 合并，尽管其总的元和阶都相同（每个元的阶不同）。合并同类项是简化代数式的一种方式。

习题 3.1 (合并同类项)，用合并同类项的方式再次推导完全平方公式和立方和公式。

对于一个多项式，我们把其中所包含的每一个单项式的阶中最高的那个称为多项式的阶，把其中包含的所有变量或者未知数的数量称为这个多项式的元。例如 $ax + b$ 是一元一次多项式， $ax^4 + bx^3 + cx^2 + dx + e$ 是一元四次多项式，而 $ax + by + c$ 是二元一次多项式， ax^2y^3 是一个二五阶多项式（也是单项式）。

有了多项式中的每一个单项式的阶的概念，我们把多项式中的每一个单项式按照其阶数从高到底的加和称作多项式的降序格式，反过来从低到高的加和称作升序格式。

3.4 幂次方多项式这么特殊吗？

为什么变量之间的关系千千万万种，我们非得先研究幂次方多项式呢？首先， $y = ax$ 是乘法，出现很正常——只要 a 和 x 之间的关系是多次重复

的加法,自然就会出现。例如,匀速行走车的速度和路程、时间的关系。接着, $y = x + b$ 的出现也很正常,只要 b 和 x 之间是合起来数一数的关系。例如,车不是从原点出发的,或者问车开了一段时间以后离想要去的终点的距离,都会用到这个表达式。于是,合起来,就会出现 $y = kx + b$ 这样的关系式。除法原则上就是乘法。其出现的场合就是乘法出现的场合。

问题是,为什么 $y = ax^2$ 这样的表达式会是在常数、乘法、加法之后就需要学习的东西呢?我们来举个例子。

例 3.1 (长方形的面积). 一段长度为 $C = 8$ 的绳子,围成一个长方形,求这个长方形面积最大的时候的面积。

假设长方形的长和宽分别为 a, b , 则 $C = 2a + 2b$, $S = ab$ 。相当于已知 $a + b = \frac{C}{2} = 4$ 的情况下,求 $S = ab$ 的最大值。

根据 $a + b = \frac{C}{2}$, 我们得到在 $a > 0, b > 0$ 的条件下, $a = \frac{C}{2} - b$ 。于是, $S = a \times (\frac{C}{2} - a) = -a^2 + \frac{C}{2}a$ 。

你看就出现了 a^2 的项。在这里其来自于乘法和加法的组合—— $a \times (\frac{C}{2} - a)$ 。

至于这个问题的答案和求解过程,等到我们学习了二次函数之后再来展示。为什么这里会有最大值呢?一开始假设 b 很小,这时候 a 确实大(最大也不过就是 $\frac{C}{2}$),但是,因为这时候 b 几乎为零,则 ab 的乘积也几乎为零;当 b 特别大接近 $\frac{C}{2}$ 的时候, a 几乎为零, ab 的乘积也几乎为零;两头都几乎为零,中间显然有大于零的点,例如 $a = b = \frac{C}{4}$ 的时候, $S = ab = \frac{C^2}{16}$ 。既然中间有大的地方,两边几乎为零,那么中间肯定有至少一个最大值所在的点。

我们再来看另一个场景,

例 3.2 (长方形的面积). 一段长度为 $L = 100$ 米的路。当匀速运动的小明的运动速度增加了 5 米每秒的时候,其所用的时间减少了 4 分钟。问其原始速度是多少?也不知道小明为什么这么愚蠢不直接测量一下其原始的速度和时间。

假设其原始速度是 v , 则原来所用时间为 $\frac{L}{v}$, 现在所用时间为 $\frac{L}{v+v_a}$ 。其中 $v_a = 5$ 米每秒。两者的差值为 $4 = \Delta t = \frac{L}{v} - \frac{L}{v+v_a}$ 。

对于这个分数表达式,方程的两边同时乘以两个分母,得到 $\Delta t v (v + v_a) = L (v + v_a - v)$, 也就是 $\Delta t v^2 + \Delta t v_a v - L v_a = 0$ 。

你看这里也出现了 v^2 的项。其来源是两个包含未知数或者是变量的分数加法算式。因此，只要把加减乘除混合起来，很容易就会出现平方项以及更高阶的项。

未来我们还会学习到，对于任何一条曲线，在曲线上的某个点附近，我们总是可以用多项式去逼近这条曲线，近似表示这条曲线。因此，有很多的理由使得我们在常数、线性表达式之后，就来研究这个幂次方的函数。

3.5 代数式的分类

用字母——当前我们把字母看作代表了数的变量或者未知数——和数以及这些对象的加法（包含了减法）、乘法（包含了除法）、任意实数次幂次方运算构成的数学表达式称为代数式。例如 $a, ax, ax^2 + bx, x^{-1}, (ax + b)^{-3}, (ax + b)^{\frac{2}{3}}, (ax + b)^{-\frac{2}{3}}$ 。

其中，只出现了变量的非负整数次幂的，并且要么没有加法符号的，要么出现在加法中的每一项中所包含的每一个变量的阶，和其他参与加法的项中的相应变量的阶，完全相同，也就是可以做合并同类项的操作的，被称为单项式。注意，系数中的出现加法项不影响，例如 $(a + b)x^3$ 仍然是单项式。

只出现了变量的非负整数次幂的，可以表达成一个单项式或者多个单项式相加的形式的，称为多项式。例如 $ax^2 + b$ 。

原则上，多项式包含了单项式。因此，传统上有一个名称叫做整式——其是单项式和多项式的合称——就不再被需要了。如果要保留这个名字，则在我们这里整式和多项式完全就是一个意思。

如果代数式之中出现了变量或者变量的多项式的负整数次幂，则被称为分式。例如 $(ax^2 + b)^{-2}$ 。整式和分式统称有理式。类比于整数和分数统称有理数。

如果代数式之中出现了变量或者变量的多项式的非整数型的分数次幂，或者无理数次幂，则被称为无理式。例如 $(ax^2 + b)^{\frac{1}{2}}, (ax^2 + b)^{-\frac{1}{2}}, (ax^2 + b)^{\sqrt{2}}$ 。

3.6 推荐阅读材料

3.7 作业

习题 3.2 (多项式的元和阶). 写下来以下多项式的元和阶, 如果是单项式注明这是一个单项式: 3 , $2x + 1$, $2x + 3y$, a , $ax + b$, $ax + by$, $2x^2 - 1$, $2(x + 3y)^2$, $ax^5 + 3x$, x^{-1} 。

习题 3.3 (多项式降序格式). 写下来以下多项式的降序格式: $3x + 2x^2 - 1$, $2(x + 3y)^2 + (x - y)^2$, $ax^5 + 3(x - 1)^3$ 。

习题 3.4 (变量和多项式的元和阶). 找出来一个类似买苹果的生活场景, 列出来其中的核心数学表达式。按照生活场景的具体情况, 确定常量、变量、未知数, 给出这个表达式的相应的元和阶。

3.8 本章小结

在本章中, 我们学习了常量、变量和未知数, 以及由它们构成的单项式和多项式。学习了按照变量和未知数来判断多项式的元和阶。

同一个数学表达式, 当我们把其中不同的量看作是变量的时候, 其含义是不同的, 我们的主要关注点也是不同的。变量和未知数也是两个不同的概念, 尽管通过求解未知数的过程把未知数看作变量。

未来, 我们还会注意到, 用字母表示的变量的系数——目前我们当作给定的常数, 在有必要的时候也可以当作变量来对待。例如, 我们可以问 $ax = b$ 的解是什么, 以及什么时候这个方程有解。前者我们把 a 当作一个常数, 得到其解是 $x = \frac{b}{a}$ 。而后者, 我们发现只有 $a \neq 0$ 的时候, 这个解 $x = \frac{b}{a}$ 才有意义。因此, 在回答后一个问题的时候, 我们其实要考虑不同的 a 带来的对解的影响。这时候, 相当于我们把 a 当作了变量——决定方程的变量。这个时候我们的眼光已经超越了具体的一个给定 a 的值的方程, 而是看到了所有的线性方程。随着进一步学习, 我们会更深入地了解关于变量和常量之间的关系。

第四章 因式分解

在本章中我们将学习什么是因式分解，因式分解可以怎么做，以及体会为什么要学习因式分解。从数学知识本身来说，因式分解也为后面解方程做了铺垫。

4.1 为什么需要因式分解

在我们之前的绝大多数计算中，都是把包含乘积项的数学表达式变成不包含乘积项，而是仅仅包含加法项的数学表达式。例如 $(x+a)(x-a) = x^2 - a^2$ 。但是，我们也看到，这个计算是可逆的，也就是我们显然也可以这样来使用上面的等式 $x^2 - a^2 = (x+a)(x-a)$ 。这样的把一个整体是多项相加关系的多项式变成若干个多项式相乘的形式的变换被称为因式分解。其中每一个参与乘法的多项式被成为因式。

对于整数，我们学习过整数的因子分解，把一个整数变换为多个整数的乘积。因子分解是对整数的拆分，因式分解则是对整式也就是多项式的拆分。

为什么我们要来学习因式分解呢？将来在求解方程的时候，我们会把方程变形为多个因子相乘的形式，例如 $f(x) = 0 \rightarrow g(x)l(x) = 0$ 。于是，根据“零乘以任何数都等于零，乘积项等于零其中一项必然为零”，求解原始的方程 $f(x) = 0$ ，就等价于求解要么 $g(x) = 0$ ，要么 $l(x) = 0$ ，也就是说把 $g(x) = 0$ 的解和 $l(x) = 0$ 的解的集合放到一起，做成并集就是原始方程 $f(x) = 0$ 的解的集合。往往 $g(x) = 0$ 和 $l(x) = 0$ 的阶要比原始的方程 $f(x) = 0$ 的阶更低。低阶方程一般更容易求解。因此，只要能够不断地这

样做下去，方程就能够更容易求解。

更重要的是，这个因式分解代表了一种思维：分而治之，简称分治。一个复杂的问题可以转化为多个更加简单的问题，从而如果解决了这些更加简单的问题，则原始的问题也就可以解决了。这个分治的思想也可以表述为，一个问题里面有多个独立的因素联合起来发挥作用，那么先搞清楚每一个因素如何发挥作用，然后再合起来，有助于解决问题。

4.2 简单因式分解

最简单的因式分解就是反用乘法计算。例如，反用乘法对加法分配律 $(a(b+c) = ab+ac)$,

$$ab+ac=a(b+c). \quad (4.1)$$

我们称之为提取公因式。例如，反用完全平方公式 $((a+b)^2 = a^2+2ab+b^2)$,

$$a^2+2ab+b^2=(a+b)^2. \quad (4.2)$$

例如，反用平方差公式 $((a+b)(a-b) = a^2-b^2)$,

$$a^2-b^2=(a+b)(a-b). \quad (4.3)$$

例如，反用立方和公式 $((a+b)(a^2-ab+b^2) = a^3+b^3)$,

$$a^3+b^3=(a+b)(a^2-ab+b^2). \quad (4.4)$$

例如，反用完全立方公式 $((a+b)^3 = a^3+3a^2b+3ab^2+b^3)$,

$$a^3+3a^2b+3ab^2+b^3=(a+b)^3. \quad (4.5)$$

例如，反用乘法公式 $((ax+b)(cx+d) = acx^2+(ad+bc)x+bd)$,

$$acx^2+(ad+bc)x+bd=(ax+b)(cx+d). \quad (4.6)$$

这一种也有一个自己的名称——十字相乘法。原因是我们往往可以如图 4.1 把这个公式展示成“十字相乘”的形式：左侧把 x^2 的系数做因子分解得到 a, c ，右侧把常数项做因子分解得到 b, d ，使得交差十字相乘以后得到的结果 $ad+bc$ 刚好等于一次项 x 的系数。

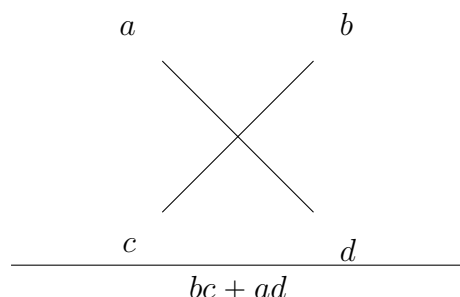


图 4.1: 十字相乘法。用十字相乘的方式来呈现因式分解公式 $acx^2 + (ad + bc)x + bd = (ax + b)(cx + d)$ 。

4.3 几种类型混合的因式分解 *

我们先来做个简单的,

例 4.1 $(ax + ay + bx + by)$. $ax + ay + bx + by$ 的因式分解。

我们发现了公因式,

$$ax + ay + bx + by = a(x + y) + b(x + y) = (a + b)(x + y), \quad (4.7)$$

$$ax + ay + bx + by = x(a + b) + y(a + b) = (x + y)(a + b). \quad (4.8)$$

我们再来一个稍微复杂一点的,

例 4.2 $(x^4 + 4)$. $x^4 + 4$ 的因式分解。

我们先用凑完全平方公式的方式,

$$x^4 + 4 = x^4 + 4x^2 + 4 - 4x^2, \quad (4.9)$$

$$= (x^2 + 2)^2 - (2x)^2, \quad (4.10)$$

$$= (x^2 + 2x + 2)(x^2 - 2x + 2). \quad (4.11)$$

后一步使用了平方差公式。

我们再来一个更加复杂的, 但是可以用统一的方法的——这里是凑完全平方方法。

例 4.3 $(x^2 + x - 6)$. $x^2 + x - 6$ 的因式分解。我们先用凑完全平方公式的方

$$\begin{array}{cc}
 1 & 3 \\
 & \diagdown \quad \diagup \\
 & \times \\
 & \diagup \quad \diagdown \\
 1 & -2 \\
 \hline
 1 \times 3 + 1 \times (-2) = 1
 \end{array}$$

图 4.2: 十字相乘法举例。用十字相乘的方式来呈现因式分解公式 $x^2 + x - 6$ 。

式,

$$x^2 + x - 6 = x^2 + 2 \times \frac{1}{2}x + \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 - 6, \quad (4.12)$$

$$= \left(x + \frac{1}{2}\right)^2 - \frac{25}{4}, \quad (4.13)$$

$$= \left(x + \frac{1}{2}\right)^2 - \left(\frac{5}{2}\right)^2, \quad (4.14)$$

$$= \left(x + \frac{1}{2} + \frac{5}{2}\right) \left(x + \frac{1}{2} - \frac{5}{2}\right), \quad (4.15)$$

$$= (x + 3)(x - 2). \quad (4.16)$$

我们再来试试十字相乘法。第一项最简单的就是拆分成 $1 = 1 \times 1$, 最后一项可以拆分为 $-6 = -2 \times 3 = -3 \times 2 = -1 \times 6 = -6 \times 1$ 。经过交叉相乘, 我们发现, 如图 4.2 所示的分解是可以的。因此得到 $x^2 + x - 6 = (x + 3)(x - 2)$ 。

我们也注意到, 凑完全平方不需要猜测和技巧, 但是十字相乘法需要一切猜测和试探。中间我们还运用了一次平方差公式。

例 4.4 $((x^2 - 6)^2 - x^2) \cdot (x^2 - 6)^2 - x^2$ 的因式分解。我们注意到整体上这是平方差公式,

$$(x^2 - 6)^2 - x^2 = (x^2 - 6 + x)(x^2 - 6 - x), \quad (4.17)$$

$$= (x + 3)(x - 2)(x - 3)(x + 2). \quad (4.18)$$

其中, 我们还用到了上一题的结果, 以及上一题的方法——要么是十字相乘法要么是凑完全平方再用平方差公式。

我们还可以把这个题的表达式直接先彻底展开,

$$(x^2 - 6)^2 - x^2 = x^4 - 12x^2 + 36 - x^2, \quad (4.19)$$

$$= x^4 - 13x^2 + 36, \quad (4.20)$$

$$= (x^2 - 4)(x^2 - 9), \quad (4.21)$$

$$= (x + 2)(x - 2)(x + 3)(x - 3). \quad (4.22)$$

这里我们先把 x^2 看作一个整体来运用上一题的方法——要么是十字相乘法要么是凑完全平方再用平方差公式, 然后在得到 $(x^2 - 4)(x^2 - 9)$ 之后, 再使用平方差公式。

因式分解的计算可以非常需要技巧。不过随着我们计算机数学软件的发展, 我们已经很大程度上可以依赖计算机数学软件来完成数学计算了。我们人类仍然需要掌握因式分解是什么——把什么东西变成什么东西, 每一步操作的原理是什么, 通常用在哪里, 还可以用在哪里, 背后的思想是什么。于是将来遇到问题的时候能够想得起来去用它。

4.4 因式分解的应用

解方程, 再找找其他例子。

解方程是因式分解的典型应用。前面已经提到, 按照“零乘以任何数都等于零, 乘积项等于零其中一项必然为零”, 只要多项式 $f(x)$ 能够被表示为 $f(x) = g(x)l(x)$ 其中 $g(x), l(x)$ 也要求是多项式, 则原始方程 $f(x) = 0$ 的解必然是方程 $g(x) = 0$ 的解集和方程 $l(x) = 0$ 的解集的并集。特别地, 如果多项式 $f(x)$ 能够被表示为 $f(x) = (x - x_1)l(x)$, 则 x_1 必然是原始方程 $f(x) = 0$ 的解。

例 4.5 (求解 $x^2 + x - 6 = 0$). $x^2 + x - 6 = 0$ 的求解。

我们之前已经得到

$$x^2 + x - 6 = (x + 3)(x - 2). \quad (4.23)$$

按照“零乘以任何数都等于零, 乘积项等于零其中一项必然为零”得到原始方程的解为以下两个方程的解的并集,

$$x + 3 = 0, \quad (4.24)$$

和

$$x - 2 = 0. \quad (4.25)$$

对于前者, 等式两边加上 -3 得到

$$x = -3. \quad (4.26)$$

对于后者, 等式两边加上 2 得到

$$x = 2. \quad (4.27)$$

合起来, 我们得到原始方程的解为 $\{-3, 2\}$ 。

未来学习到函数的时候, 因式分解还可以用来画出来函数的大概形状。未来我们会回到这个问题。目前, 我们暂时非常粗糙地来展示一下因式分解在看到函数形状上的作用。

假设我们有一个多项式分解好了是这样的

$$f(x) = (x - x_1)(x - x_2)(x - x_3), \quad (4.28)$$

并且 x_1, x_2, x_3 是从小到大排列的。我们说, $f(x)$ 在每一个边界点上为零; 对于 $x < x_1$, 三项都为负, 整体结果为负, 并且, x 越往左边 $f(x)$ 的绝对值越大; 对于 $x_1 < x < x_2$, 第一项为正, 其他两项为负, 整体结果为正, 并且, 到下一个零点 x_2 之前没有其他零点; 对于 $x_2 < x < x_3$, 第一项为正, 第二项为正, 第三项为负, 整体结果为负, 并且, 到下一个零点 x_3 之前没有其他零点; 对于 $x_3 < x$, 三项都为正, 整体结果为正, 并且, 再也没有其他零点, 而且随着 x 越往右边 $f(x)$ 的绝对值越大。合起来, 就是这样一个形状, 如图 4.3。

对于 $f(x) = (x - x_1)^2(x - x_2)(x - x_3)$ 来说, 其在每一个边界点上为零; 对于 $x < x_1$, 第一项为正, 后两项为负, 整体结果为正, 并且, x 越往左边 $f(x)$ 的绝对值越大; 对于 $x_1 < x < x_2$, 第一项为正, 其他两项为负, 整体结果为正, 并且, 到下一个零点 x_2 之前没有其他零点; 对于 $x_2 < x < x_3$, 第一项为正, 第二项为正, 第三项为负, 整体结果为负, 并且, 到下一个零点 x_3 之前没有其他零点; 对于 $x_3 < x$, 三项都为正, 整体结果为正, 并且, 再也没有其他零点, 而且随着 x 越往右边 $f(x)$ 的绝对值越大。合起来, 就是这样一个形状, 如图 4.4。

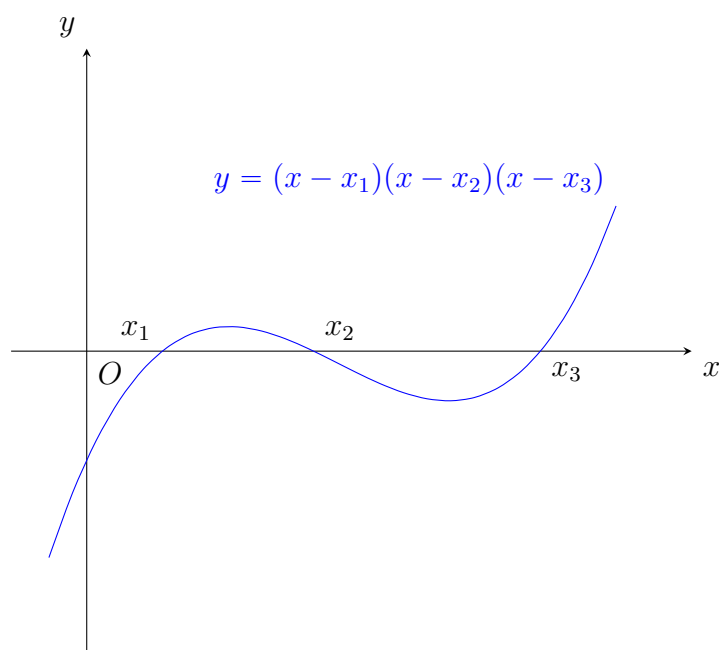


图 4.3: 通过因式分解看函数形状。把一个函数写成因式分解之后的样子就可以大概知道其形状。

在这里，我们还没有严格学习函数的概念。这里我们仅仅把“函数”当作一个多项式表达式，其作用就是给一个 x 值，就能得到一个相应的 y 值。把这样的 (x, y) 值对应到平面上的点，就会得到一条平面上的曲线——很多的这样的 (x, y) 值构成的集合。未来我们会严肃地来学习函数这个概念。

对比图 4.3 和图 4.4，还可以发现，前者是穿过了 x_1 的点，后者是拐了一个弯没有穿过。这其实就是 $(x - x_1)$ 和 $(x - x_1)^2$ 的差别：前者在 x_1 前后符号相反于是穿过这个零点，后者在 x_1 前后符号相同于是只能在这个零点处拐弯。你可以猜测 $(x - x_1)^3$ 和 $(x - x_1)^4$ 的效果。

习题 4.1 (通过因式分解看函数形状). 类比对 $f(x) = (x - x_1)(x - x_2)(x - x_3)$ 和 $f(x) = (x - x_1)^2(x - x_2)(x - x_3)$ 的分析，尝试描绘一下 $f(x) = (x - x_1)^3(x - x_2)(x - x_3)$ 和 $f(x) = (x - x_1)^4(x - x_2)(x - x_3)$ 的图象。用计算机数学软件画出来图看看。

也就是说，只要知道一个函数的每个零点的位置，零点左侧右侧的正负，我们大概就可以把这个函数凑起来。这背后是这个函数的连续性。

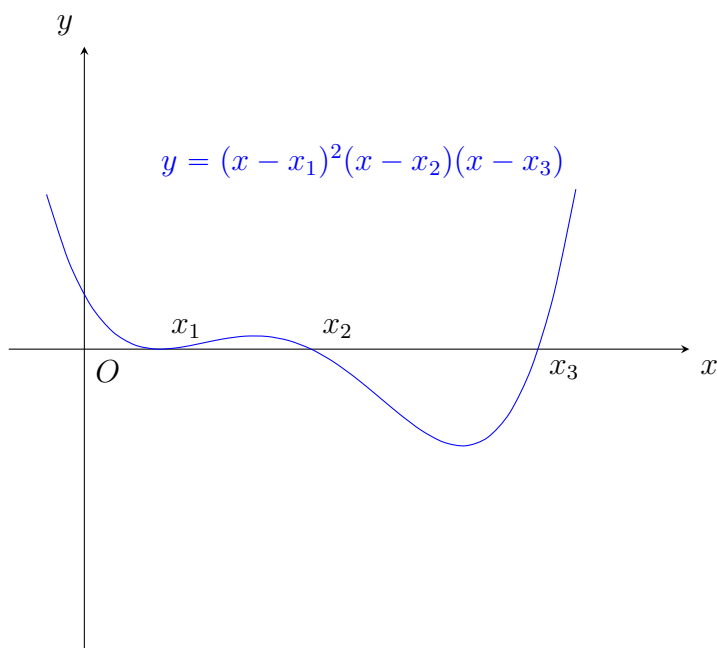


图 4.4: **通过因式分解看函数形状**。把一个函数写成因式分解之后的样子就可以大概知道其形状。

未来我们会看到更多因式分解及其背后的思维方式——分而治之——在数学和在思考问题中的作用。

4.5 推荐阅读材料

4.6 作业

习题 4.2 (因式分解). 尝试对以下代数式做因式分解: $a^3b - ab^3$, $x^2 - 5x + 6$, $x^2 - y^2 - z^2 + 2yz$, $x^4 + x^3 + x^2 + x$, $x^3 + x^2 + x + 1$, $x^3 + y^3 + x^2 + y^2 + 2xy$, $(a + b)^3 + (b + c)^3 + (c + a)^3 + a^3 + b^3 + c^3$ 。

习题 4.3 (因式分解求解方程). 尝试通过因式分解求以下方程的解: $x^4 + x^3 + x^2 + x$, $x^3 + x^2 + x + 1$, $x^2 - 5x + 6$ 。

4.7 本章小结

在本章中，我们学会了什么是因式分解，简单的因式分解大概可以有哪些方法来做。更重要的是我们体会了一下为什么要学习因式分解，从因式分解有什么用的角度——例如求解方程和大概描绘函数图象，以及从因式分解背后的思维方式——分而治之——的角度。

注意在本章范围内，实际上我们是在有理数的范围内来做因式分解。也就是，我们架设在原始的多项式和分解以后的多项式中所有出现的常数——不管是常数项还是常数系数——都是有理数。其实，因式分解其实是在实数范围内做的——也就是可以出现无理数做为常数。以后等学习了复数，展开以后的多项式里面可能还会出现复数。不过，无理数和复数常数的因式分解怎么做就是另一回事了。目前，我们只学习有理数常数的因式分解。

顺便，本书每一章的小结都会对本章的学科概念知识内容做一个简短总结，然后提一提留下来的关于学科概念知识的尾巴如果本章有这样的只有未来才能真正解决的尾巴的话。接着，小结会提炼一下本章学科概念知识的发展形成背后的学科大图景——这个学科的典型研究对象、典型研究问题、典型思维方式、典型分析方法和这个学科的典型责任。如果可以，我们会进一步提炼一般性人类思维以及教和习的方法。知道每一章小结的写作目的以后，你要学会怎么来看每一章的小结。

第五章 一元一次方程和一元一次不等式

在这一章中我们来学习一元一次方程和一元一次不等式的求解和应用。一元一次方程是最简单的方程，而一元一次不等式是最简单的不等式。它们也是求解更加复杂的方程和不等式的计算上的基础。

在学习这部分内容的过程中，要注意求解方法和已有数学知识的联系，尤其是思考求解的每一步为什么需要这样做和为什么可以这样做的问题。

5.1 为什么要学习一元一次方程和一元一次不等式？

方程是包含未知数和等号的数学表达式。一元一次方程就是未知数出现的形式是一元一次多项式， $ax + b$ 。因此，一般的一元一次方程可以表示为 $ax + b = cx + d$ ，或者通过等式的性质简化为 $ax + b = 0$ 的形式。注意后面的系数 a, b 的取值不一定是前面的 a, b 。

例 5.1 ($ax + b = cx + d \rightarrow ax + b$). 把 $ax + b = cx + d$ ，或者通过等式的性质简化为 $ax + b = 0$ 。

对 $ax + b = cx + d$ 这个等式的两边同时加上 $-cx - d$ （等式的性质），得到

$$ax + b - cx - d = cx + d - cx - d = 0, \quad (5.1)$$

也就是 (交换律, 结合律, 分配律)

$$(a - c)x + (b - d) = 0. \quad (5.2)$$

因此, 只要重新命名变量 $(a - c) \rightarrow a, (b - d) \rightarrow b$, 就能把原始的方程变成 $ax + b = 0$ 的形式。

最一般的一元一次不等式可以表示为 $ax + b \neq cx + d, ax + b > cx + d, ax + b \geq cx + d$ 。注意, 出现小于号和小于等于号的只需要交换变量 (a, b) 和 (c, d) 。因此, 对于一元一次不等式我们只研究 “不等于号、大于号、大于等于号” 三种情况。同样地, 通过重新命名变量, 可以使得最一般的形式变为相应的 $ax + b \neq 0, ax + b > 0$ 和 $ax + b \geq 0$ 。“ $\neq$ ” 其实就是等号的相反, 因此, 只要求解出来等号对应的方程的解, 让 x 不等于所有的这些解就可以。因此, 对于一元一次不等式, 我们以后只处理 $ax + b > 0$ 和 $ax + b \geq 0$ 的情况。

那我们为什么要学习一元一次方程和一元一次不等式呢? 第一, 为了解决表示为一元一次方程和一元一次不等式的问题。第二, 为了给将来求解更加复杂的方程和不等式做准备。第三, 为了体会到知识的系统性, 在同一个知识体系内任何新的数学知识不过就是已有数学知识的发展。在这里, 求解方程的基础不过就是运算律和等式的性质, 以及不等式的性质。

第一点, 我们在稍后的应用中会体会到。第二点, 我们在稍后更加复杂的方程和不等式的求解中会体会到。第三点, 我们即将在下一节体会到。

5.2 一元一次方程的求解

求解方程的基础是运算律和等式的性质, 也就是算式 (1.15) 和 算式 (1.9)。

对于一元一次方程 $ax + b = 0$, 我们只需要完成下面的步骤, 就能得到

满足这个方程的 x ，也就是这个方程的解：

$$ax + b = 0, \quad (5.3)$$

$$\Rightarrow ax + b + (-b) = 0 + (-b) \text{ (等式的性质) }, \quad (5.4)$$

$$\Rightarrow ax + (b + (-b)) = 0 + (-b) \text{ (加法结合律) }, \quad (5.5)$$

$$\Rightarrow ax = -b \text{ (加法计算) }, \quad (5.6)$$

$$\Rightarrow \frac{1}{a} \times ax = \frac{1}{a} \times (-b) \text{ (等式的性质) }, \quad (5.7)$$

$$\Rightarrow \left(\frac{1}{a} \times a \right) x = \frac{1}{a} \times (-b) \text{ (乘法结合律) }, \quad (5.8)$$

$$\Rightarrow x = -\frac{b}{a} \text{ (乘法计算) }. \quad (5.9)$$

其中，每一步我们都给出来了为什么可以这样算的理由——要么来自于运算律，要么来自于等式的性质。这样把每一步为什么可以这样做的理由放到每一步后面的括号的的方式，未来我们会经常使用。

我们也很容易看出来为什么需要这样算：我们希望左边将来最终只出现未知数 x ，而没有其他任何对象，也就是形如 $x = x^*$ 。于是，我们把常数项先从左边拿出去，到右边；再把系数 a 变成 1。从结果我们也看到，一元一次方程必然要求 $a \neq 0$ 。

注意，这个一元一次方程的解的一般表达式是第二位的，我们是怎么来得到这个解的，中间要回答为什么可以这样做，为什么需要这样做才是第一位的。

5.3 一元一次方程的应用

5.4 一元一次不等式的求解

求解不等式的基础是运算律和不等式的性质，前者就是算式 (1.15) 我们来复习一下后者。

按照前面的分析，我们只需要处理 $ax + b > 0$ 和 $ax + b \geq 0$ 的情况。因此，下面的不等式的性质，我们也只讨论这两种情况。不等式 $a > b$ （以及 $a \geq b$ ）具有如下含义：在数轴上， a 出现在 b 的右侧。因此，如果我们

加上同一个数, 则 a, b 同时右移 (或者同时左移) 相同的距离, 其相对关系不变。也就是说

$$a > b \Rightarrow a + c > b + c. \quad (5.10)$$

其次, 如果 a 在 b 右侧, b 在 c 的右侧, 则 a 在 c 的右侧,

$$a > b, b > c \Rightarrow a > c. \quad (5.11)$$

这被称为不等号的传递性。运用传递性, 我们还可以证明,

$$a > b, c > d \Rightarrow a + c > b + d. \quad (5.12)$$

其证明过程如下

$$a > b \Rightarrow a + c > b + c, \quad (5.13)$$

$$c > d \Rightarrow c + b > d + b. \quad (5.14)$$

因此, $a > b, c > d \Rightarrow a + c > b + d$ 。最后一步我们用到不等号传递性。不等式还有一条性质: 不等式的左右两侧同时乘以一个大于零的数, 不等号不变, 同时乘以一个小于零的数不等号相反, 也就是

$$a > b \Rightarrow a \times c > b \times c, (c > 0) \quad (5.15a)$$

$$a > b \Rightarrow a \times c < b \times c. (c < 0) \quad (5.15b)$$

其成立的原因可以联系到乘法和加法的关系, 以及相反数的数轴上的表示。

假设这里的 c 就是一个整数。如果是正整数, 我们可以把乘法变成加法, 如果是负整数则是在乘以正整数的基础上, 再做一个乘以 -1 , 也就是做相反数。我们先来处理正整数 n ,

$$a \times n = \underbrace{a + a + \cdots + a}_n, b \times n = \underbrace{b + b + \cdots + b}_n. \quad (5.16)$$

于是, 把 $a > b, c > d \Rightarrow a + c > b + d$ (这里的 $c = a, b = d$) 重复 $(n - 1)$ 次就得到

$$a \times n > b \times n. \quad (5.17)$$

对于负整数 $-n$ ，相当于两边还需要乘以 -1 ，也就是求一次相反数。求相反数的操作，会颠倒左右相对顺序。于是符号相反，就得到

$$a \times (-n) < b \times (-n). \quad (5.18)$$

从整数的情况到分数也简单，我们先来证明下面的命题，

定理 5.1: 不等式除以整数

不等式除以整数] 当 n 为正整数的时候 $a > b \Rightarrow a \times \frac{1}{n} > b \times \frac{1}{n}$,
 $a > b \Rightarrow a \times \frac{1}{-n} > b \times \frac{1}{-n}$ 。

证明 5.1. 我们用反证法来证明这个命题。取 $c = a \frac{1}{n}, d = b \frac{1}{n}$ ，我们假设不等式不成立，也就是 $c \leq d$ ，我们来证明必然有 $a \leq b$ 。因此，得到和已知信息 $a > b$ 矛盾的结果，于是原假设错误，必须有 $c > d$ ，也就是 $a > b \Rightarrow a \times \frac{1}{n} > b \times \frac{1}{n}$ 。

我们来证明 $c \leq d \Rightarrow a \leq b$ 。根据之前已经证明的，对于正整数 n ，有 $c \leq d \Rightarrow c \times n \leq d \times n$ 。注意尽管之前是对小于号来证明的，只需要交换符号，自然就得到大于号也成立。等号本来就成立，基于等式的性质。而 $c \times n = a, d \times n = b$ ，因此， $c \leq d \Rightarrow a \leq b$ 。

对于 $(-n)$ 的部分，我们不再证明。

于是，和之前的乘上一个整数联合就得到了算式 (5.15) 对于分数成立。当 c 是无理数的时候，我们就需要再次借助“无理数可以用有理数来不断逼近”的事实，来“论证”既然算式 (5.15) 对于任何一个有理数都成立，则对于无理数仍然成立。严格的论证过程只有等学习过“极限”的概念才能真的来实现这个逼近从而完成了。

有了算式 (5.10)，算式 (5.11) 和算式 (5.15) 这三条，我们就可以来求解不等式了。注意，实际上这三条并不独立，第三条是前两条的导出结论。当然，这三条都是数在数轴上的表示、运算和运算律的导出结果。为了以后查阅方便，我们把这三条不等式的性质写在一起，

- 不等式加上实数不等号不变： $a > b \Rightarrow a + c > b + c$,
- 不等号的传递： $a > b, b > c \Rightarrow a > c$,

- 不等式乘上大于零的实数不等号不变, 乘上小于零的实数不等号相反:

$$a > b, c > 0 \Rightarrow ac > bc, \quad a > b, c < 0 \Rightarrow ac < bc.$$

对于一元一次不等式 $ax + b > 0$ (我们假设 $a > 0$, 否则, 可以两边乘以 -1 之后得到 $(-a)x + (-b) < 0$, 然后求解这个新的不等式), 我们只需要完成下面的步骤, 就能得到满足这个不等式的 x , 也就是这个不等式的解:

$$ax + b > 0, \quad (5.19)$$

$$\Rightarrow ax + b + (-b) > 0 + (-b) \text{ (不等式的性质)}, \quad (5.20)$$

$$\Rightarrow ax + (b + (-b)) > 0 + (-b) \text{ (加法结合律)}, \quad (5.21)$$

$$\Rightarrow ax > -b \text{ (加法计算)}, \quad (5.22)$$

$$\Rightarrow \frac{1}{a} \times ax = \frac{1}{a} \times (-b) \text{ (不等式的性质)}, \quad (5.23)$$

$$\Rightarrow \left(\frac{1}{a} \times a\right)x > \frac{1}{a} \times (-b) \text{ (乘法结合律)}, \quad (5.24)$$

$$\Rightarrow x > -\frac{b}{a} \text{ (乘法计算)}. \quad (5.25)$$

其中, 每一步我们都给出来了为什么可以这样算的理由——要么来自于运算律, 要么来自于不等式的性质。

我们也很容易看出来为什么需要这样算: 我们希望左边将来最终只出现变量 x , 而没有其他任何对象, 也就是形如 $x > x^*$ 。于是, 我们把常数项先从左边拿出去, 到右边; 再把系数 a 变成 1。

注意, 这个一元一次不等式的解的一般表达式是第二位的, 我们是怎么来得到这个解的, 中间要回答为什么可以这样做, 为什么需要这样做才是第一位的。

我们看到满足 $x > -\frac{b}{a}$ 的 x 有很多。所有的这样的 x 构成一个集合。为了表示这个集合我们用之前已经学过的数轴上的区间 (见 1.3.5 节): (a, b) 、 $(a, b]$ 、 $[a, b)$ 和 $[a, b]$ 。对于 $x > -\frac{b}{a}$, 我们可以看到左侧的边界, 但是右侧的边界怎么表示呢? x 可以任意大。我们把数轴的最右侧, 那个可以任意大的东西, 叫做“无穷大”, 记作 ∞ 。类似地, 我们把数轴的最左侧, 那个在 x 轴负方向可以任意大的东西, 叫做“负无穷大”, 记作 $-\infty$ 。因此, $x > -\frac{b}{a}$ 就可以被记作 $(-\frac{b}{a}, \infty)$ 。注意, 对于无穷大这个边界点, 永远不可能直接取到, 因此其相应的区间边界符号永远是圆括号。

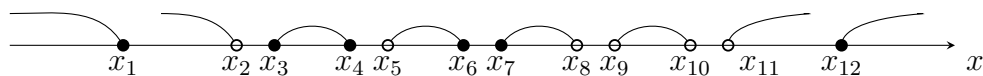


图 5.1: 区间示意图。这里我们列出来了各种区间在坐标轴上的示意图。

我们也可以把这个区间在数轴上表示出来。找到一个区间的边界上的两个点。如果遇到无穷大，就只需要找那个不是无穷大的点。然后，对于给定的边界点 x^* ，把 $x > x^*$ 画成在数轴上方从 x^* （用在这个点的空心原点表示）出发指向右侧的箭头，把 $x < x^*$ （用在这个点的空心原点表示）画成在数轴上方从 x^* 出发指向左侧的箭头。如果遇到大于等于和小于等于号，把空心原点改成实心原点。我们在图 5.1 中提供了这样的例子。

5.5 一元一次不等式组的求解

对于一元一次方程，我们求解完了直接得到一个解。如果有两个一元一次方程构成方程组，要求找到一个 $x = x^*$ 都满足，则这两个方程必须是等价的，也就是所有系数对应成比例， $\frac{a'}{a} = \frac{b'}{b}$ ，否则无解。

但是，对于一元一次不等式，由于求解出来以后是一个解的集合，我们就可以把不同的一元一次不等式联合起来求解。我们来完成一个这样的例子，再来总结一般性的方法。

例 5.2 (一元一次不等式组). 求同时满足下面的要求的 x ,

$$\begin{cases} 4x + 9 \geq 0, \\ 3x + 2 < 0. \end{cases} \quad (5.26)$$

我们把两者分别先求解出来。从第一式得到 $x \geq -\frac{9}{4}$ ，从第二得到 $x < -\frac{2}{3}$ 。把这两个解的集合都表示成区间， $[-\frac{9}{4}, \infty)$ ， $(-\infty, -\frac{2}{3})$ 。

这两个区间可以通过数轴上的表示来建立交集，如图 5.2 所示。最后，两个集合的交集就是区间交差重叠的区域，于是，我们得到这一组不等式的解为 $[-\frac{9}{4}, -\frac{2}{3})$ 。

也就是说，对于一般的不等式组，我们只需要先求出来每一个不等式的解的集合，表示为在数轴上的区间，然后，找出来这些区间的重叠区域就可以得到整体不等式组的解。

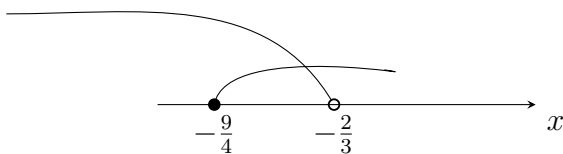


图 5.2: 不等式组的区间示意图。不等式组的区间是各自不等式的区间的交集。

5.6 一元一次不等式的应用

5.7 推荐阅读材料

5.8 作业

5.9 本章小结

在本章中，我们定义了一元一次方程和一元一次不等式的标准形式，学习了如何求解它们，学习了它们的应用。我们看到，它们求解步骤完全就是等式的性质和运算律，以及相应的不等式的性质和运算律的导出结果，而这些性质和运算律都是数和运算的含义的导出结果。导出的方式就是逻辑演绎证明。因此，从这一章的学习中看到数学知识的系统性是第一位的。

其次我们还看到对于一元一次方程和一元一次不等式我们既有来自于解决实际问题的需求，也有数学本身发展的需求——为求解未来更加复杂的方程做准备。

当我们定义一元一次方程以及它的标准形式的时候，我们没有直接从 $ax + b = 0$ 出发，而是从“等号的两边都只能出现一元一次多项式”出发。这个定义未来可以迁移到更加复杂的方程的定义和标准形式上面去。定义一元一次不等式及其标准形式的时候也相同。这是本章中体现的数学的另外一个要注意的点：任何的定义和操作，表述和定义的时候，尽量考虑到未来可以迁移到更加一般的情况。未来我们会看到，会体会到，数学是为了看到相似性，相似性的相似性，数学就是给不同的对象相同的名字和符号。

此外，在本章中，你还可以体会到“对任何一个操作，你要问为什么可

以这样做，为什么需要这样做”，而不是仅仅停留在“怎么做”。甚至，你还可以进一步追问：我为什么要做这些。例如这里就体现为，求解一元一次方程和一元一次不等式的时候，每一步怎么做，为什么可以这样做，为什么需要这样做，为什么我要学习它们以及它们的求解过程。从做什么 (What)，怎么做 (How)，为什么可以和需要这样做 (Why)，完成这个做什么或者知道这个怎么做或者知道这个为什么对我有什么意义 (Meaningful)，这四个问题 (WHWM) 的角度来思考问题，是一个对于学习和解决问题来说非常有用的思维方式。未来我们还会多次看到。

更进一步，对于本书的每一章，你需要先从数学知识本身的体系来看一遍，再进一步对这个体系的建立的每一个步骤做思考体会到背后的数学思维，然后再从一般性的对于思考和学习有意义的思维方式的角度来看一遍。

这样你就可以从启发这个数学学科概念知识的来自于现实世界的经验和体验，或者来自于前面已经学会的数学学科概念知识，得到当前所学习的数学学科概念知识，然后进一步看到数学的学科大图景，尤其是数学的学科典型责任、典型思维方式和典型分析方法，最后看到一般性人类思维，以及最后的最后看到教和习的方法 [2]。

本书之所以这样写，于是也要求你这样读，是因为本书在写作的时候只有一个目的：**一切为了你的迁移、创造和欣赏，也就是为了未来你能够创造性使用知识和创造知识，或者至少可以欣赏的知识的被创造和被创造性使用。**如果你从具体数学学科概念知识的学习中体会到了数学学科大图景，一般性人类思维以及教和习的方法，然后你选择忘掉其中的一大部分数学学科知识尤其是关于具体哪个计算怎么算，例如本章的一元一次方程和一元一次不等式求解，都不是问题。

第六章 集合初步

这一章我们来学习集合这个概念。将来继续学习数学之后，我们会发现，集合是绝大多数数学概念的基础，以前学过的数学概念都可以用集合以及下一章要学习的定义在集合之上的映射的语言来重新表述和理解。而提出集合这个概念不过就是为了一个特别简单的目标，为了把意思说清楚。这也体现了数学就是思维的语言，就是描述世界的语言。

跟小学阶段和本书前面的部分对集合的基于自然语言的介绍相比，这一章我们会引入更多的数学符号。

6.1 集合的朴素概念

在我们用语言来表达我们的意思的过程中，我们往往先要确定一个对象，然后再来说这个对象怎样怎样。例如，“这个苹果很新鲜”这里的对象就是“这个苹果”。这个对象有一个“很新鲜”的属性。在这里，我们其实还藏起来了一个新鲜程度的含义甚至确定这个新鲜程度的操作性测量——例如拿眼睛看如果发现外皮蔫了等特征就说不新鲜，或者用仪器测量这个苹果的水分含量和这个苹果或者这一类苹果刚采摘下来的时候的水分含量做一个比较从而得到一个比例值，然后规定这个比例值小于多少就是不新鲜——的方法。这里的一个测量的方法以及一个判断的标准，就给出了明确的是否新鲜的含义。更进一步，我们还可以说“这一箱子苹果、这一堆苹果很新鲜”，“这种苹果很甜（同样地通过尝一下，或者测量糖分含量，建立标准）”来把我们所指的对象从一个苹果变成一批苹果，一种苹果。在对象上，“一箱子苹果”就从特指的单个的苹果变成了特指的若干个苹果；“一种

苹果”就从单个的具体的苹果变成了抽象的一类苹果——这一类可能指的是和苹果的 DNA（脱氧核糖核酸¹）以及 DNA 所决定的外形、颜色等属性，以及和这一种苹果的生长环境例如产地等因素有关的品种。也就是说，“一种苹果”其实是一个抽象概念，它不是由具体的每一个苹果来承担的，而是由这一类苹果的共性特征来承担的。

顺便，沿着这个一个和一类的思路，当我们写作的时候说“这样的人很可爱”的时候，其实，我们是做了抽象的，原则上，这个时候我们必须至少在自己心中是明确地把这个人的某个特征提炼为这样的人的特征的。否则，我们最好还是就这个具体对象来讨论，而不是抽象为对其所代表的一类对象的讨论。因此，当你说“你这人怎么这样”的时候，你对这个人的批评远远没有“你这种~~种~~人，就会这样这样”来得狠。好了，说了这么一大段语言和表达经常遇到的事情之后，让我们回到数学。

为了更明确地表达出来所指的对象，以及用于指代一群对象或者一类对象，人们提出了“集合”的概念。一个集合指的是这样一个群体的概念：我们完全可以判断任何一个对象是否属于这个群体。这个性质被称为集合元素的确定性。群体本身的含义是一个或者多个对象合起来，往往人们还会给这个群体一个合适的名称。例如， $\{1, 3, 5, 7, 9\}$ 是一个集合，一个群体，我们可以称之为“十以内的所有奇数”。例如，所有的自然数也是一个集合²。例如，一个硬币（通常抛出去以后落下来的硬币只会显示为两面之一，立着的情况不算）的两面——正面或反面，也是集合。

那么，是不是人们日常生活中所用的所有的类别名词都满足确定性，都可以算是集合呢？不是的。例如，胖子、秃头、美景、好人都不算集合。我们很难说清楚体重大于多少，或者体重身高比例大于多少算胖子。类似地，秃头、美景、好人也一样。当然，实在有必要的时候，也可以定义一个测量

¹一个生物的脱氧核糖核酸决定了这个生物的大部分性状，是这个生物的可以遗传给后代的一定程度上决定了后代的性状的遗传物质。通常同类的生物具有相似的脱氧核糖核酸。这个类的概念在生物学中具有层次性，例如界、门、纲、目、科、属、种。更细的种类上相同的物种一般具有更大的遗传物质上的相似性。

²这是一个有趣的故事，有些数学书约定零属于自然数，有些约定 $1, 2, 3, \dots$ 才是自然数而 0 不是。因此看起来好像一个数是否属于自然数，当这个数是零的时候，是不明确的。因此，自然数不满足前面提到的确定性要求。但是，实际上，任何一本书，这个约定是唯一的。因此，只要在那本书本身的范围内，自然数仍然是满足确定性要求的。

方式和判断的阈值。例如，身体质量指数 (Body Mass Index, BMI) 就是这样的判断肥胖的方式，

$$BMI = \frac{M}{H^2}, \quad (6.1)$$

其中 M 是以公斤为单位的体重， H 是以米为单位的身高。当 $BMI \geq 28$ 的时候被称为肥胖。这个时候，忽略测量误差带来的不确定性，我们也就有了满足集合元素确定性的“胖子”集合的定义。

习题 6.1 (BMI 的定义). 请思考为什么 BMI 这样来定义，而不是体重除以身高或者身高的三次方，以及这个定义出来的 BMI 通常被用于哪些地方。

也就是说，由于我们在表达中对明确性有要求，我们尽量从集合开始来描述我们的对象。注意，这不是集合定义，而是给出来集合必须满足的一个要求——元素的确定性。其含义是：给定任意一个对象，我们可以判断这个对象是否属于这个集合。用数学符号这样来表达，就是如果 X 是一个集合，则 $\forall x$ (“对于任意一个 x ”的意思，读作“任意 x ”)， $x \in X$ (读作“ x 属于 X ”) 或者 $x \notin X$ (读作“ x 不属于 X ”) 两者有且只有一个成立。也就是，要么 x 属于 X ，要么 x 不属于 X ，必须有而且只能有其中一个成立，不能两个都不成立，也不能两个都成立。

将来我们会进一步学习到不满足这个确定性要求的数学对象，以及如何来定义集合才能使得它满足这个确定性的要求。

习题 6.2 (理发师悖论). 有兴趣的同学可以查一查“理发师悖论”。思考一下这个理发师悖论和集合的定义以及集合的元素的确定性的关系。

除了确定性，有的时候我们也说集合元素具有互异性和无序性。互异性的意思是，集合的任意两个元素是不同的。无序性的意思是，集合元素的出现顺序无关紧要。于是， $\{1, 1, 2, 3\}$ 和 $\{1, 2, 3\}$ 、 $\{1, 3, 2\}$ 是同一个集合。

在这里，我们用到了集合的一种描述方式：列出来这个集合的所有元素，习惯上用一对大括号 $\{\cdot\}$ 来把这些元素包含在里面。对于有限个元素的集合，这可以做到。对于无限个元素的集合，这需要一些定义这个集合的人和阅读这个定义的人之间的一些共同的知识基础。例如，我们说这个符号 $\{0, 1, 2, 3, 4, \dots\}$ 代表了所有的非负整数，这个符号 $\{\dots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}$ 代表了所有的整数，以及 $\{1, 3, 5, 7, \dots\}$ 代表了所有的奇数。由于我们经常用到所有整数和所有非负整数这两个集合，我们还给了它们自己的名字和

符号,

$$\mathbb{Z} = \{\cdots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \cdots\}, \quad (6.2)$$

$$N = \mathbb{Z}^* = \{0, 1, 2, 3, 4, \cdots\}. \quad (6.3)$$

集合还可以用另一种方法来描述, 给出来所有的元素满足的一个性质, 只要这个性质是否满足的条件对于任何对象都是具有明确的结论的, 就行。例如, 所有的奇数我们可以写成 $\{2n + 1, n \in \mathbb{Z}^*\}$ 。

现在我们有了对集合的几种描述方式: 给出来所有的元素, 写出来一个所有元素需要满足的性质, 约定一个符号或者名称。

有了集合, 也就是指代明确的对象(们), 仅仅是我们准确表达思想, 表达思考过程, 表达我们的意思的第一步。未来我们还会学习到如何表达这个对象具有什么性质, 这样的判断句——将来会被成为命题, 以及对象之间的关系——将来会被称为映射。为了给命题和映射的学习作准备, 我们来学习一点点集合本身的运算。

顺便, 在这里学习集合这个概念的过程中, 你已经看到了我们——其实是人类历史上的数学家们——是如何尽力把意思表达明确的。数学就是一个帮助人们把意思表达明确的语言。

6.2 集合的交和并

其实, 我们应该早就发现了, 集合 $\mathbb{Z}, \mathbb{Z}^*$ 之间, 以及它们和奇数集合之间是有一定的关系的。例如, 非负整数集合 $\mathbb{Z}^*$ 其实是整数集合 $\mathbb{Z}$ 的一部分, 奇数集合也是整数集合的一部分, 并且奇数集合可以通过非负整数的集合来定义。这一节, 我们来定义集合之间的关系, 以及基于这些关系的运算。

给定任意两个集合 X, Y , 可能有元素同时属于这两个集合, 也就是 $\exists x \in X, x \in Y$ 。我们把满足这个条件的元素放到一起, 构成一个新的集合, 也就是 $\Omega = \{x | x \in X, x \in Y\}$ 。首先, 我们来看看是否这可以看作一个集合, 也就是是否满足元素的确定性。对于任意的一个元素 y , 其是否属于集合 X, Y 只有四种情况: 第一种, 不属于两者之中的任何一个; 第二种, 属于 X , 而不属于 Y ; 第三种, 不属于 X , 而属于 Y ; 第四种, 属于 X 也属于 Y 。我们发现, 以上四种情况的任何一种都可以判断这个元素 y 是否属于

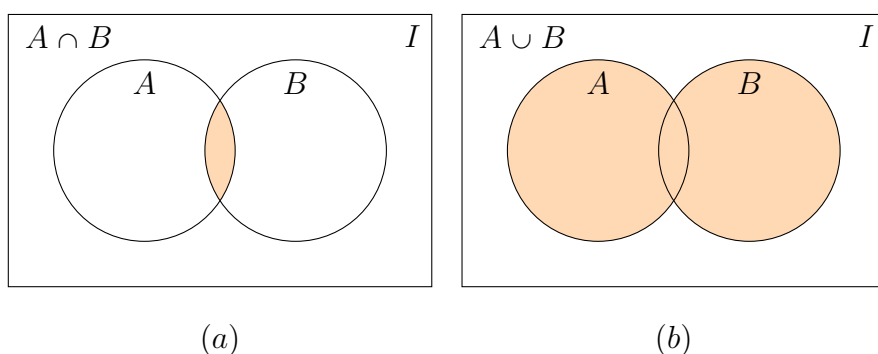


图 6.1: 两个集合的交集和并集。用 Venn 图展示了两个集合的 (a) 交集 ($A \cap B$) 和 (b) 并集 ($A \cup B$)。

Ω 。因此，满足元素的确定性。例如，只有第四种情况下， $y \in X$ 并且 $y \in Y$ 则 $y \in \Omega$ ，其他情况下都有 $y \notin \Omega$ 。直观上，这个集合 Ω 相当于把两个集合共同的元素都找了出来。

为了把这个直观上的关系清楚地表示出来，我们引入一个集合的图形表示，Venn 图：我们用一个圆形、长方形、椭圆形之类的闭合图形来表示一个集合，把集合内的元素表示成这个闭合图形内的点，集合之外的元素表示为这个闭合图形之外的点。

在这个 Venn 图的表示下，我们发现，前面的集合 Ω 就相当于两个集合 X, Y 的 Venn 图表示的交叉的部分，共同的部分。因此，我们把 Ω 称作集合 X 和 Y 的交集，记作 $X \cap Y$ ，也就是，

定义 6.1 (交集). $X \cap Y = \{x | x \in X \text{ 并且 } x \in Y\}$.

有了交集的概念，我们就很容易定义并集，

定义 6.2 (并集). $X \cup Y = \{x | x \in X \text{ 或者 } x \in Y\}$.

可以发现，前面四种情况的后三种都满足并集所要求的性质，只有第一种不满足。因此，其也满足元素的确定性要求。在直观上，并集就相当于把 X, Y 两个集合覆盖的元素合并起来，因此被称作并集。我们可以画图来表示交集和并集，如图 6.1。这样的图称为集合的 Venn 图。

6.3 子集、全集和补集

有了交集和并集的概念，我们就可以来判断两个集合的包含关系了。如果两个集合 X, Y 的交集就是其中之一，例如 X ，我们说， X 被包含在 Y 之中， X 是 Y 的子集。其中， X 是 Y 的子集的含义是 X 被包含在 Y 之中，也就是

定义 6.3 (子集). $X \subset Y$ 指的是 $\forall x \in X, x \in Y$.

我们还给出如下的一个判断句，或者说命题，如果 X 是 Y 的子集，则 X, Y 的并集是 Y 。我们甚至说，反过来也如此，也就是如果 X, Y 的并集是 Y ，则 X 是 Y 的子集。

我们来看看这些判断句是否正确：第一， $X \cap Y = X \Rightarrow X \subset Y$ ；第二， $X \subset Y \Rightarrow X \cap Y = X$ ；第三， $X \cup Y = Y \Rightarrow X \subset Y$ ；第四， $X \subset Y \Rightarrow X \cup Y = Y$ 。这里符号“ $\Rightarrow$ ”的意思就是左边的条件成立则右边的结论也成立。我们一个个来检查这些判断句是否成立。

通过这个检查，你也会发现集合的语言的另一个作用：集合和逻辑论证是紧密相连的。实际上，把含义表示明确，把对象描述明确，再加上把逻辑论证表述明确，正好就是集合的语言的作用和魅力。

定理 6.1: 集合包含关系和交集一

集合包含关系和交集一] $X \cap Y = X \Rightarrow X \subset Y$ 。

证明 6.1. 根据交集的定义 $X \cap Y = \{x | x \in X \text{ 并且 } x \in Y\}$ 。现在已知 $X \cap Y = X$ ，那么就可以推断出来，没有任何一个 Y 之外的属于 X 的元素。因为，如果有 $z \in X, z \notin Y$ ，那么，这个 $z \notin X \cap Y$ ，于是 $X \neq X \cap Y$ 。因此， $\forall z \in X$ ，必然有 $z \in Y$ 。根据子集的定义， $X \subset Y$ 。

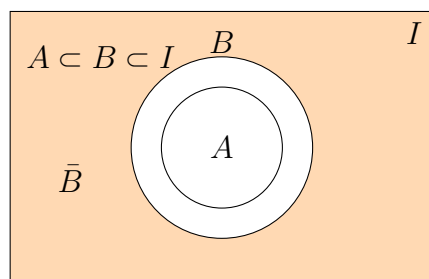


图 6.2: 子集、全集、补集。用 Venn 图展示了子集 ($A \subset B \subset I$)、全集 (I)、补集 ($\bar{B}$)。

定理 6.2: 集合包含关系和交集二

集合包含关系和交集二] $X \subset Y \Rightarrow X \cap Y = X$ 。

证明 6.2. 一方面, 根据交集的定义 $X \cap Y = \{x | x \in X \text{ 并且 } x \in Y\}$ 。另一方面, 根据子集的定义, $X \subset Y \Rightarrow \forall x \in X, x \in Y$ 。也就是说, 在 $X \subset Y$ 的条件下, 交集的条件中的第二条 $x \in Y$ 已经自然被满足。因此, 这个时候交集的条件只剩下 $X \cap Y = \{x | x \in X\} = X$ 。

第三个和第四个判断句的证明交给读者。有了这四个命题, 我们就可以非常好地建立起来 Venn 图的直觉图像和集合的交集、并集、子集之间的对应关系了——如图 6.2 所示, 你直接上看起来图像上的包含关系、合并图形、交叠图形, 正好就是集合之间的相应的子集、并集、交集的关系。

习题 6.3 (集合包含关系和并集一). 证明 $X \cup Y = Y \Rightarrow X \subset Y$ 。

习题 6.4 (集合包含关系和并集二). 证明 $X \subset Y \Rightarrow X \cup Y = Y$ 。

对于两个具有包含和被包含关系的集合, $X \subset Y$, 或者说 $Y \supset X$ (开口方向的含义类似于大于小于号), 如果我们把更大的那个集合 Y 称作全集, 则那个更小的集合 X 就称作子集。而那部分在 Y 中但是不在 X 中的元素构成的集合则称作子集 X 在全集 Y 中的补集, 记作

定义 6.4 (补集). 当 $X \subset Y$ 的时候, 子集 X 在全集 Y 中的补集 $\bar{X} \triangleq \{z | z \notin X \text{ 并且 } z \in Y\}$ 。

有了子集的概念, 我们可以来讨论两个集合的相等和子集的关系, 也就是

定理 6.3: 子集和集合相等

子集和集合相等] $X = Y \iff X \subset Y$ 并且 $Y \subset X$ 。

证明 6.3. 按照集合的定义，两个集合相等 $X = Y$ 的含义是两个集合中包含的所有元素相等，也就是 $\forall x \in X, x \in Y, \forall y \in Y, y \in X$ 。其中，前半句正好就是 $X \subset Y$ 的定义，后半句正好就是 $Y \subset X$ 的定义。因此， $X = Y \implies X \subset Y$ 并且 $Y \subset X$ 。

反过来，按照包含关系的定义， $X \subset Y$ 就是 $\forall x \in X, x \in Y$ ， $Y \subset X$ 就是 $\forall y \in Y, y \in X$ ，两者和起来正好就是 $X = Y$ 的含义。因此， $X \subset Y$ 并且 $Y \subset X \implies X = Y$ 。

其中双向箭头符号 $\iff$ 的含义是，既有 $\implies$ 还有 $\impliedby$ 。也就是，命题的两侧互为结论和条件，也就是两者相互等价。因此， $\iff$ 也被读作“等价于”。在证明等价关系的过程中，我们要作双向的命题的证明。注意，命题并不都是双向的。

有了这个集合相等和子集之间的关系，我们在以后证明两个集合相等的过程中，经常会通过证明这两个存在相互包含关系来完成。

更进一步，为了区别两个集合之间的包含关系和相等关系，我们把前面所说的“包含”关系区分为“包含” ($\subseteq$) 和“真包含” ($\subset$) 关系。两者之间的区别是“包含”关系允许两个集合之间是相等的关系或者真包含关系，而“真包含”关系不允许两个集合之间是相等的关系。也就是说，以后 $X \subset Y$ ，意味着 $\forall x \in X, x \in Y$ ，而且， $\exists y \in Y, y \notin X$ 。 $\subseteq$ 和 $\subset$ 就类似于小于等于号 ($\leq$) 和小于号 ($<$) 的关系。

到此为止，我们就学完了基本的集合知识了，包含集合要满足的元素明确性（以及不太重要的互异性、无序性），集合的交和并运算，子集（包含和被包含关系）、全集和补集，以及集合相等和相互包含的关系。为了帮助大家熟悉这些概念和符号，我们编制了需要大家完成的一些练习题，还把在一个全集中的两个集合的一般关系的每一个区域画成了一张 Venn 图，如图 6.3。

习题 6.5 (三集合相交的每一个区域). 用集合运算表达式写出来图 6.3 中的各个区域。

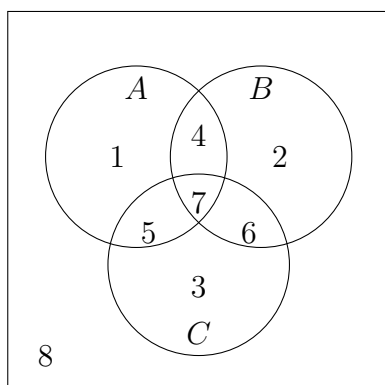


图 6.3: 三集合相交的每一个区域。用 Venn 图展示了三集合相交的每一个区域：例如 1 对应着 $(A \cap \bar{B} \cap \bar{C})$ ，4 对应着 $(A \cap B \cap \bar{C})$ 。

再次回到为什么我们要有集合的语言：为了我们在思考和表达中所指代的对象更加明确，以及形成可以直接指代一堆或者一类对象的语言和符号，也就是用来描述对象之间的层次关系、交叉关系的合适的语言和符号。我们也初步看到了，集合这个语言在表达和证明命题中的作用。

6.4 不等式的解当作集合

回到图 5.1 中用端点和连线表示的不等式的解，以及相应的区间表示： (a, b) ，其中“(,)”可以被替换为“[,]”按照端点的被包含和不被包含，“ a, b ”可以是一般的数，也可以是 $\pm\infty$ 。

这个时候，如果我们由两个不等式组成的不等式组，需要同时被满足，则我们实际上就是在计算两个不等式的解的集合的交集。这也是为什么我们在图 5.2 中得到的是两个区域的重叠的区域。因此，之前，我们实际上就朴素地运用了集合的交集运算。

6.5 推荐阅读材料

编制一些集合交和并运算，补集运算的习题。

习题 6.6 (集合运算)，自然数集合和整数之间具有包含关系吗？这两个集合的交集、并集是什么？如果有包含关系，被包含的集合在包含它的集合中

的子集是什么？

习题 6.7 (集合运算). 奇数和偶数之间具有包含关系吗？这两个集合的交集、并集是什么？如果有包含关系，被包含的集合在包含它的集合中的子集是什么？

习题 6.8 (集合运算). 质数和合数之间具有包含关系吗？这两个集合的交集、并集是什么？如果有包含关系，被包含的集合在包含它的集合中的子集是什么？

习题 6.9 (集合运算). 能被 3 整除的所有非负整数和能被 5 整除的所有非负整数之间具有包含关系吗？这两个集合的交集、并集是什么？如果有包含关系，被包含的集合在包含它的集合中的子集是什么？

习题 6.10 (集合运算). 能被 3 整除的所有非负整数和能被 9 整除的所有非负整数之间具有包含关系吗？这两个集合的交集、并集是什么？如果有包含关系，被包含的集合在包含它的集合中的子集是什么？

习题 6.11 (集合运算). 之前你在习题 1.1~ 习题 1.4 得到的各个集合，它们相互之间具有包含关系吗？画图来表示这几个集合之间的关系。

6.6 作业

6.7 本章小结

在这一章中，我们以日程生活经验和日常语言为基础，结合之前已经接触过的集合的具体例子，例如自然数、整数等等，学习了集合的概念。强调了集合最重要的特征是其包含的内部元素是明确的，也就是任何一个对象拿过来，我们都可以判断这个对象是不是属于这个集合的，是不是这个集合的元素。然后，我们学习了集合的交、并、补运算，还有集合之间的包含关系和相等关系，并且学会了画 Venn 图来表示集合之间的关系。画 Venn 图有的时候能够帮助我们思考的更深入更准确。

此外，本章留下了一个尾巴：我们的集合是用自然语言描述性地来定义的“一堆东西”，只要元素具有明确性。我们没有回答具有这性质的东西

存在吗？我们也没有回答所有的看起来具有这个性质的东西真的具有这个性质吗，可以成为数学的研究对象吗？这个问题只能留到以后再来讨论了。不过，至少你可以先去查查资料，看看包含“所有集合的集合”、“所有不包含自身的集合”之类的看起来元素明确的集合有什么问题。

第七章 映射初步

这一章我们来学习数学的另一个抽象概念——映射。我们将会发现，映射是描述对对象的操作——例如给对象命名、分类，以及把一个对象和另一个对象联系起来——的最基本的方式。而对象和操作是我们人类思考和面对世界的典型方式。因此，作为思维的语言和描述世界的语言的数学就有这个责任来给对象和操作准备好语言。上一章，我们学会了用集合来明确指代对象，这一章，我们来学习用映射指代操作。

7.1 从命名、分类、操作到映射

有三个世界时时刻刻伴随着我们：现实世界、概念世界（即存在于人类的认知结构中，也存在于各个学科的知识构成的空间中）、语言世界。我们表面上用语言世界（可以包含人类的自然语言以及数学语言、编程语言）去描述现实世界，实际上，我们是用概念世界去描述现实世界，但是需要通过语言世界这个媒介。因此，对于现实世界的对象，我们往往就得给这个对象一个语言世界的名称以及一个概念世界的含义。这个把现实世界的对象和其在概念世界和语言世界的名称和含义联系起来的过程，就是映射。

例如，如果我们有两支铅笔，而且我们经常需要按照某种方式——例如使用目的或者个人偏好——来选择用其中的哪一支。现在假设每次我们有另一个人替你去拿笔，你必须想办法明确告诉那个人每次你需要拿哪一支。实际上，你也可以把“你”看作你的大脑，把“另一个人”看作你的手。我们来设想一个如何完成这个明确告诉另一个人的过程。

首先，你可能由于某种原因选择了其中那支黑色外壳蓝色笔芯的铅笔，

而不选用另一支是白色外壳黑色笔芯的圆珠笔。接着，你怎么告诉另一个人我们的这个选择呢？按照所选择的这支笔的某种特征，而且是可以和另一支相区别的特征。例如，请你帮我取一下那支黑色外壳的笔，那支铅笔，那支蓝色笔芯的笔，以及它们这些特征的组合。不过在这里，我们不能说“黑色的笔”，这可能意味着黑色外壳，也可能意味着黑色笔芯。我们还可以先跟另一个人做一个约定，把“黑色外壳蓝色笔芯的铅笔”称作“一号笔”记作①，把“白色外壳黑色笔芯的圆珠笔”称作“二号笔”记作②。只要在这个约定下，我们就可以很方便地达到告知另一个人我们的选择的目的。

我们说，在这个过程中，你建立了映射，也就是从这两支笔到这两支笔的被指代方式之间的一个映射。例如，从这两支笔到这两支笔的两种外壳颜色的映射，从这两支笔到这两支笔的笔芯颜色的映射，从这两支笔到这两支笔的笔头材料（铅笔还是圆珠笔）的映射，从这两支笔到这两支笔的记号之间的映射。我们在下一节再来明确地学习什么是映射。但是，通过这个例子，我们大概知道了映射就是给一群东西里面的每一个东西一个名字，或者说一个记号。在下一节中，我们会发现，这个指代明确的记号名字的映射通常要求是一对一的。

其实在上面的为了指代对象明确而作的命名中，我们实际上用到了分类。例如，所有的黑色外壳的笔，所有的铅笔，所有的蓝色笔芯的笔等等。因此，当我们给一个东西一个名字，其实往往这个名字和这个对象的某种属性有关，并且满足这个属性特征的对象可能还不止这一个。这时候，我们说，其实我们对对象做了分类。从映射的角度，分类就是建立了一群对象中的每一个对象和类别名称以及这个名称代表的类别属性之间的映射。具体来说，在这里，就是建立了第一支笔和铅笔之间的映射，第二支笔到圆珠笔之间的映射。就像我们已经知道的，有很多笔都是铅笔，有很多笔都是圆珠笔，因此，这个映射实际上是给了很多东西一个共同的名字，然后把这个名字叫做这些很多个东西的类别。在下一节中，我们会学习到，这就是多对一的映射。

下面我们来假设一个更加复杂一点场景：我们有一把削铅笔的刀，它可以用于削所有的铅笔但是不能用于任何一支削圆珠笔。那么，这个削笔刀就是一个映射，把一支通常笔头磨损了不方便书写的铅笔变成一支具有再次方便书写的笔头的铅笔。任何一支钝笔头铅笔和这个削笔刀结合（其

实还要加上削的过程，但是我们先忽略这个过程)，就可以得到一支好笔头铅笔（假设这支铅笔的长度使得它还能和值得被削）。为了说起来简单，我们允许削笔刀去削本来就不钝的铅笔，也就是只要来一支铅笔，就削一下得到一支好笔头铅笔。我们还可以让这个削笔刀更加智能一点，给它增加一个判断器——如果来的是铅笔，则削，如果是圆珠笔，则不削。实际上这个判断器也是一个映射，遇到“铅笔”对应着“使用削笔刀”，遇到“圆珠笔”对应着“不使用削笔刀”。原来的削笔刀和这个判断器的组合成了一个有点智能的削笔刀。实际上，你还可以增加一个映射——一个钝笔头判断器。但是，为了说起来简单，前面我们已经允许削笔刀削本来就笔头不钝的笔。然后，我们来把这个智能削笔刀映射和前面的取用笔的映射结合起来：先取出来笔，然后，把这个取出来的笔投入到智能削笔刀。这时候，我们就得到了一个更加智能的取用削笔刀，我们只要告诉它“帮我取黑色外壳的笔”，它就会取出来，判断一下是否是铅笔，如果是，削好比头，递过来给我们用。稍后我们会发现，用映射的语言，这叫做映射的复合。

通过这一节，我希望你已经领会到，明确对象以后，对对象的命名、分类和操作这些典型的认识和改变世界的行为，实际上都是映射。现在，我们来明确定义映射。数学就是给你认识和改变世界准备的语言。

7.2 映射

定义 7.1 (映射). 两个集合 X, Y 之间的映射 f ，记作 $f: A \mapsto B$ ，满足如下条件：第一，对于任意一个 X 中的元素 x ($\forall x \in X$)，存在着唯一的一个 Y 中的 y ($\exists x \in X$)，使得 $f(x) = y$ 。

这个定义就是把前面已经形成的关于映射的直觉，从一个集合 X 里面取出来一个元素 x ，把这个元素联系到另一个集合 Y 的一个元素 $y = f(x)$ 。我们把取出来元素 x ——称作原像——的集合称作定义域 (domain)，联系到的元素 y ——称作像——的集合称作陪域 (codomain)。这个联系要满足一个条件：任何一个定义域中的元素都可以在陪域中找到一个而且只有一个像。

映射的定义对于是否会有两个不同的 $x_1, x_2 \in X$ 都被映射为同一个 $y \in Y$ 没有约束。如果这个条件满足，则映射 f 称作单射，

定义 7.2 (单射). 单射是满足如下条件的映射: $f: A \mapsto B$, 对于 X 中的任何两个元素 x_1, x_2 ($\forall x_1 \neq x_2 \in X$), $f(x_1) \neq f(x_2)$ 。

映射的定义对于是否每一个 $y \in Y$ 都被映射 f 用到没有约束。如果这个条件满足, 则映射 f 称作满射,

定义 7.3 (满射). 满射是满足如下条件的映射: $f: A \mapsto B$, 对于 Y 中的任何一个元素 y ($\forall y \in Y$), 存在者一个 X 中的 x ($\exists x \in X$), 使得 $f(x) = y$ 。

如果一个映射既是单射又是满射, 则被称为可逆映射。可逆映射的含义是, 可以构造出来一个从 Y 到 X 的映射 $g: Y \mapsto X$, 使得任何一个 $f(x) = y$ 都有 $g(y) = x$ 。以后我们会看到为什么我们特别关注映射的可逆性。现在, 我们来看一下, 为什么既是单射又是满射的映射是可逆的。映射的核心就是原像必然找得到唯一的像。要是把 $f: X \mapsto Y$ 反过来成为 $g: Y \mapsto X$, 那就是要求, 任何一个 y 也有且只有一个 x 满足 $g(y) = x$, 而且这里的 $g(y) = x$ 正好就是 $f(x) = y$ 。因为映射是满射, 所以, 任何一个像 y 都找得到至少一个原像 x 使得 $f(x) = y$ 。我们就差一个唯一性了, 也就是要求不存在这样的 $z \neq x \in X$ 使得 $f(z) = y$ 。这正好就是单射的要求: 两个不同的原像必然导致不同的像。这个时候, 我们把逆映射 g 也记作 f^{-1} , 也就是 $g = f^{-1}: Y \mapsto X$, 而 $f: X \mapsto Y$ 。

一旦两个集合之间可以建立起来可逆映射, 则这两个集合的元素之间就有了一一对应关系。这就好像是我们之前讲到的取名字的场景, 任何一个被取名字的对象都有了唯一的名字, 任何一个名字都有了唯一的对象。

那既然可逆映射这么好, 就好像完全就是给对象取个名字一样, 为什么我们不直接要求所有的映射都是可逆映射呢? 可逆映射没法满足分类的要求。比如说, 我们通过身份证给每一个人一个的号码, 通过这个号码也找得到对应的人, 但是, 我们就不能通过身份证号来把人分成男人、女人了¹。那实际上, 身份证号确实就区分标记了男人和女人。怎么做的呢? 第一, 我们建立一个集合 S 其中有两个元素 F, M 分别代表女性和男性。第二, 我们把所有的人的集合 P (或者你们班里的所有同学们) 通过映射 $\mathcal{S}$ 映射到这个性别集合 S , 使得对于任何一个 $p \in P$, $\mathcal{S}(p) = F$ 或者 $\mathcal{S}(p) = M$ (映射的定义已经要求两者不能同时成立)。当然, 我们知道这个映射本身是通

¹假设我们在这里暂时仅仅考虑男人和女人两种性别。

过每一个人的 DNA 来判断的，或者大多数时候可以通过性别特征来作粗糙的判断。因此，这个分类的背后其实有更加深刻的概念。这个分类也会带来一些额外的信息，例如可能男性普遍来说身高更高一点，力气更大一点，因此有些任务可能更加适合男性。也就是说，把人分成男人和女人，是有意义的。好了，我们已经看到了不可逆的映射，那些把一类的 x 映射为某个 y ，把另一类的 x' 映射为另一个 y' 是有意义的。实际上，将来我们会发现，可逆映射反而太平庸，这样的相当于抽象和分类的映射在更深刻的问题上才更有用。

因此，在这里，我们先定义了一般的映射，然后在定义单射和满射，再合起来定义了可逆映射。实际上，你可以追问，那是不是可以允许有些 $x \in X$ 找不到像，以及有些 $x \in X$ 找得到多个像。对于前者，我们重新定义一下集合 X 就可以。对于后者，可能我们就需要想办法把这样的多个像用某种标准分开，然后分开以后各自成为一个映射。例如，4 的平方根有两个 ± 2 ，但是，我们可以区分出来其中哪些大于零，哪些小于零，然后，分别构建映射。这就是正的平方根和负的平方根的映射，也就是 $\sqrt{x}$ 和 $-\sqrt{x}$ ，对于任何 $x > 0$ 。也就是说，很有可能，基于这里定义的映射，我们可以很好地处理各种类型的关系。这也有待于我们的后续学习和运用数学中来进一步检验这一条。

7.3 映射的复合，最简单的群和表示理论

有了集合和映射，我们可以重新定义大部分的数学概念。例如，在小学阶段，我们学习数学是从自然数开始的。我们把自然数和用自然数来数数当作“自然的天生就有的”东西，然后，从数数开始我们可以建立加法的概念，以及加法之上的减法、乘法，还有乘法和减法之上的除法的概念。接着，我们从具体运算中总结出来了运算律。最后，到了用字母表示数。这基本上就是我们目前阶段已经学习到的算术和代数。今天，我们非常粗糙和初步地来体会一下，数其实不是数学里面最重要的最基础的概念，也不是天生的自然就有的，反而是运算和关系决定了数。在高等数学里面，这叫做表示理论。我们不会去学习这个名词，但是，我们会通过例子来帮你体会一下什么是表示理论，为什么世界往往可以由数（或者说代表数的字母，

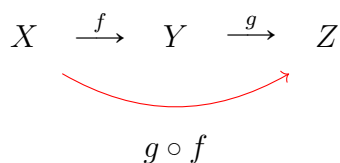


图 7.1: **复合函数**。把两个函数前后作用结合起来可以得到复合函数。

字母参与的算式) 来描述。

为了这个目的, 我们先要来学习一下映射的复合。如图 7.1 所示, 给定两个映射 $f: X \mapsto Y, g: Y \mapsto Z$, 我们可以定义一个新的映射 $g \circ f: X \mapsto Z$: 对于任意 $x \in X$, $g \cdot f(x) = g(f(x))$ 。其中我们用到了 $f(x) \in Y$, 因此, $g(f(x))$ 存在且唯一。这满足了 $g \circ f: X \mapsto Z$ 是一个映射的要求。复合函数中先作用的函数一般写在右侧。这样它会先作用于写在右侧括号中的原像上。

前面我们提到了给笔编号的映射 (记为 $L: P \mapsto N$, 从笔的集合到编号的映射), 判断是否铅笔并且如果是则削尖铅笔的映射 (记为 $S: P \mapsto P$, 从笔的集合到笔的集合的映射), 我们还提到了我们可以组合成一个决定使用哪一支笔, 告诉另一个人需要的笔的编号, 然后判断一下是否铅笔, 如果是就削尖的智能映射 (记为 $IS: N \mapsto P$, 从编号的集合到笔的集合的映射)。这其实就是映射的复合。

具体这是如何复合起来的呢? 我们看到, 复合映射 $IS: N \mapsto P$, 把一个数字编号变成一支笔而且是削尖了的笔。因此, 它肯定先要根据数字编号来找到相应的笔, 也就是利用 $L^{-1}: N \mapsto P$, 它是映射 $L: P \mapsto N$ 的逆映射。这也是为什么在之前的场景中, 我们要求编号和笔之间具有一一对应关系的原因。一一对应关系是可逆的。找到笔之后, 就可以接着用上削笔映射 $S: P \mapsto P$ 了。因此, $IS = S \circ L^{-1}$ 。

现在, 我们可以来描述一个原始场景下完全和数没有关系的问题了——我们来考虑对一个硬币的状态的可能的操作。忽略抛出以后立起来的硬币, 硬币的状态只有两个, 我们记作 $\mathbb{S} = \{H, T\}$ 。注意, 这里的 H, T 不是数, 就是代表硬币两个状态的一套符号。你可以理解为“头、尾”状态或者“正、反”状态。给定一个抛出以后静止在桌子上的硬币, 其状态肯定是 H, T 两者之一, 也就是 $S = H$ 或者 $S = T$ 。

从日常生活中，我们知道，对于一个硬币，我们只有两种操作，不翻转（保持硬币当前状态不动）和翻转。实际上，这两个操作是映射。我们分别记作 $\hat{I}$ 和 $\hat{X}$ 。上面加一个箭头符号表示这是映射不是集合。把这两个映射明确写出来，我们有， $\hat{I}(H) = H, \hat{X}(T) = T, \hat{I}(H) = T, \hat{X}(T) = H$ 。

我们来考虑一下所有的可能的从 $\mathbb{S}$ 到 $\mathbb{S}$ 的映射。由于 $\mathbb{S}$ 只有两个元素，因此，从 $\mathbb{S}$ 到 $\mathbb{S}$ 的映射只能有四个：除了上面的 $\hat{I}$ 和 $\hat{X}$ ，还有 $\hat{H}$ 和 $\hat{T}$ ，前者把所有的状态都映射为 H ，后者把所有的状态都映射为 T 。后两者不是满射，我们暂时仅仅考虑前面两个满射， $\hat{I}$ 和 $\hat{X}$ 。

我们来作一下 $\hat{I}$ 和 $\hat{X}$ 的复合映射。通过画下面的图，我们很容易就能发现， $\hat{I} \circ \hat{I} = I, \hat{I} \circ \hat{X} = X, \hat{X} \circ \hat{I} = X, \hat{X} \circ \hat{X} = I$ 。从直觉上这也很好理解，两次不翻转合起来是不翻转，先翻转后不翻转合起来是翻转，先不翻转后翻转合起来是翻转，两次翻转合起来是不翻转。到现在为止，我们的硬币问题仍然没有出现任何数。但是，你有没有发现，这里的 $\hat{I}$ 和 $\hat{X}$ 的复合映射的关系，使得 $\hat{I}$ 和 $\hat{X}$ 看起来很象很象 1 和 -1 ： $1 \times 1 = 1, 1 \times (-1) = -1, (-1) \times 1 = -1, (-1) \times (-1) = 1$ 。在这个意义上，我们就用 1 和 -1 分别来表示不翻转和翻转操作 $\hat{I}$ 和 $\hat{X}$ 。到这里，忽然我们的硬币翻转问题出现了数和乘法，而且这个数和乘法正好描述了对硬币的操作以及这些操作之间的复合映射关系。我们已经提过，映射代表了对对象的操作，远远要比数的运算适用范围大得多。但是，你看，从映射出发，考虑复合映射，自然地我们回到了数以及数的运算。这是为什么世界往往由数和运算来描述：对这个世界中的对象的操作，以及这些操作之间的复合关系，导致了数和运算的出现，或者说往往被数和运算来表示。

为什么我们优先讨论这两个可逆映射，而不是把这四个都放在一起讨论呢？将来我们会学习到一个叫做群的东西，就是对一个对象的所有的可逆操作的集合，操作之间具有复合关系。在数学和物理学的研究中，我们往往需要用到合适的数学结构来表示群（以及建立在群的基础之上的更加复杂的数学结构）。

当然，实际上，我们可以尝试给更加一般的映射找到合适的数学结构。例如，

习题 7.1 (全部四个映射之间的复合关系). 请你来写出来全部四个映射之间的复合关系，然后，试一试是不是能够找到合适的数学结构来表示这四个

映射, 从而这些数学结构之间的某种乘法运算正好就是这四个映射之间的复合关系。这个问题有点难, 你可能得学一点点矩阵运算之后再来找这样的数学表示。你暂时做不出来没关系。

小知识 7.1 (Dirac 符号). 在学习矩阵之前, 我们还可以用基于前面提到的 H, T 的符号来给这四个映射找到一个数学表示。将来, 这一套表示会在认识世界中, 尤其是量子的世界中, 发挥巨大的作用。我们考虑一个从 $\mathbb{S} = \{H, T\}$ 到 $\{0, 1\}$ 的映射。稍后我们会发现为什么需要这个映射。同样, 我们只考虑到满射。定义域和陪域都只有两个元素, 因此满射只有两种: $H \rightarrow 1, T \rightarrow 0$, 以及 $H \rightarrow 0, T \rightarrow 1$ 。

我们把前者记作 $\langle H|$, 后者记作 $\langle T|$ 。稍后我们也会看到为什么我们用这个记号。然后, 我们引入两个新的记号, $|H\rangle$ 和 $|T\rangle$ 。前者代表硬币状态 H , 也就是从定义域中选取了 H ; 后者代表硬币状态 T , 也就是从定义域中选取了 T 。然后, 我们约定下面的计算规则,

$$\langle H|H\rangle = 1, \langle H|T\rangle = 0, \quad (7.1)$$

$$\langle T|H\rangle = 0, \langle T|T\rangle = 1. \quad (7.2)$$

在这个运算规则下, 我们发现, $\langle H|$ 就起到了第一个映射的作用, $\langle T|$ 就起到了第二个映射的作用。也就是说, 运算规则算式 (7.1) 和算式 (7.2) 下, $\langle H|$ 和 $\langle T|$ 就是前面提到的映射的一种表示。

注意, 我们最终的目标, 不翻转和翻转操作, 是从 $\mathbb{S} = \{H, T\}$ 到 $\mathbb{S} = \{H, T\}$ 的映射。现在我们有从 $\mathbb{S} = \{H, T\}$ 到 $\{0, 1\}$ 的映射的表示 $\langle H|$ 和 $\langle T|$, 我们能不能在此基础上, 再把上面的不翻转和翻转映射也构建起来呢?

我们这样来想, $\langle H|$ 把 H 映射为 1, 因此, 如果再配上一个 $|H\rangle$ 在后面, 那就得到了把 H 映射为 H , $|H\rangle\langle H|$ 。

类似地, $|H\rangle\langle T|$ 把 T 映射为 H ; $|T\rangle\langle H|$ 把 H 映射为 T ; $|T\rangle\langle T|$ 把 T 映射为 T 。

有了这四个操作, 我们终于有了翻转和不翻转的表示,

$$\hat{I} = |H\rangle\langle H| + |T\rangle\langle T|, \quad (7.3)$$

$$\hat{X} = |H\rangle\langle T| + |T\rangle\langle H|. \quad (7.4)$$

用这一套记号, 算符 $\hat{T}$ 和 $\hat{H}$ 可以表示为

$$\hat{T} = |T\rangle\langle H| + |T\rangle\langle T|, \quad (7.5)$$

$$\hat{H} = |H\rangle\langle T| + |H\rangle\langle H|. \quad (7.6)$$

很容易检验算式 (7.5) 中的 $\hat{T}$ 把任意状态变成了 $|T\rangle$ 状态, 和算式 (7.6) 中的 $\hat{H}$ 把任意状态变成了 $|H\rangle$ 状态。

习题 7.2 (Dirac 符号下算赋的验证 1). 请验证算式 (7.3) 和算式 (7.4) 的算赋, 结合运算规则算式 (7.1) 和算式 (7.2), 正好就是映射不翻转 $\hat{I}$ 和翻转 $\hat{X}$ 作用到硬币状态 H, S 上的效果。

习题 7.3 (Dirac 符号下算赋的验证 2). 请验证算式 (7.3) 和算式 (7.4) 的算赋, 结合运算规则算式 (7.1) 和算式 (7.2), 正好就是映射不翻转 $\hat{I}$ 和翻转 $\hat{X}$ 。这次你来通过计算这些映射之间的复合关系完全相同的方式来验证。也就是, 验证 $\hat{I}\hat{I} = \hat{I}, \hat{I}\hat{X} = \hat{X}, \hat{X}\hat{I} = \hat{X}, \hat{X}\hat{X} = \hat{I}$ 。

一旦我们作了这个验证, 那我们就可以再一次得到 $\hat{I}, \hat{X}$ 可以分别用数 $1, -1$ 以及它们之间的乘法来表示。你看, 我们再一次从完全没有数的一堆 H, T 之类的符号, 得到了数和运算, 并且用这些数和运算来表示对硬币的翻转和不翻转操作。

这就是数学, 给现实世界准备好描述它们的数学结构, 以及受现实世界启发依靠人类想象力和逻辑来发展未来可以用于描述世界的语言。

7.4 三个元素的集合到自身的映射 *

上一节我们已经学习了两个元素的集合上的映射。这一节, 我们来试试三个元素的集合上的映射。记这个三元素集合为 $\mathbb{T} = \{1, 2, 3\}$ 。其中 $1, 2, 3$ 仅仅是代表这个集合中的三个不同元素的记号。它们之间没有任何关系。

跟之前一样, 我们主要考虑满射。由于需要是满射, 从三个元素映射到三个元素, 必然也是单射。否则, 当有两个或者以上的元素映射为同一个元素的时候, 则肯定有的元素没有原像, 也就不可能是满射。于是, 这样的映射既是满射又是单射, 合起来必须是可逆映射。那么, $\mathbb{T}$ 到其自身的映射可能有哪些呢? 我们一个个来列举一下。

例如我们从元素 1 开始, 其像可能为 1, 2, 3 三者之一。确定了 1 的像之后, 我们来确定 2 的像。这时候有两个不同选择——1 选中的像就不能再选为 2 的像了。最后, 3 的像只能是除了前两个像之外的第三个元素。也就是只有一种选择。合起来, 我们有 $3 \times 2 \times 1 = 6$ 种映射。我们把它们都写出来, 并且给一个记号。

映射 $e: \mathbb{T} \rightarrow \mathbb{T}$ 的原像和像分别为: $e(1) = 1, e(2) = 2, e(3) = 3$ 。映射 $g_1: \mathbb{T} \rightarrow \mathbb{T}$ 的原像和像分别为: $g_1(1) = 2, g_1(2) = 3, g_1(3) = 1$ 。映射 $g_2: \mathbb{T} \rightarrow \mathbb{T}$ 的原像和像分别为: $g_2(1) = 3, g_2(2) = 1, g_2(3) = 2$ 。映射 $f_1: \mathbb{T} \rightarrow \mathbb{T}$ 的原像和像分别为: $f_1(1) = 1, f_1(2) = 3, f_1(3) = 2$ 。映射 $f_2: \mathbb{T} \rightarrow \mathbb{T}$ 的原像和像分别为: $f_2(1) = 3, f_2(2) = 2, f_2(3) = 1$ 。映射 $f_3: \mathbb{T} \rightarrow \mathbb{T}$ 的原像和像分别为: $f_3(1) = 2, f_3(2) = 1, f_3(3) = 3$ 。

我们直觉上可以看到, e 相当于不变映射; g_1 大概相当于把这三个数做了一个往前推一步, 当得到的结果大于 3 的时候, 把得到的结果减去 3, 重新找到一个原始集合 $\mathbb{T}$ 的元素; g_2 大概相当于把这三个数做了一个往前推两步, 当得到的结果大于 3 的时候, 把得到的结果减去 3, 重新找到一个原始集合 $\mathbb{T}$ 的元素; f_1 相当于把第一号元素保持不变, 其他两个元素互换; f_2 相当于把第二号元素保持不变, 其他两个元素互换; f_3 相当于把第三号元素保持不变, 其他两个元素互换。

现在我们来考虑这些映射之间的复合。首先, 由于 e 是不变映射, 因此任何其他映射 g_i 或者 f_i 和 e 的复合, 都得到那个映射它自己, 也就是相应地 g_i 或者 f_i 。也就是 $e \circ g_i = g_i \circ e = g_i$, $e \circ f_i = f_i \circ e = f_i$ 。

其次, $g_1 \circ g_1 = g_2$, 连着两次往前推一步, 合起来是推两步。类似地, $g_1 \circ g_2 = e = g_2 \circ g_1$, 推三步相当于不变; $g_2 \circ g_2 = g_1$, 推四步相当于推一步。于是, 我们发现, e, g_1, g_2 之间的关系很象 $1, q, q^2$, 只要满足 $q^3 = 1$ 就行。当然, 这里我们还不知道什么样的 q , 除了 $q = 1^2$ 之外还满足 $q^3 = 1$ 。未来等我们学习到复数, 就可以发现这样的 q 确实存在。假设我们找到了这样的 q , 我们就再一次从映射、映射的复合走到了数和数之间的乘法运算。

其实, 直觉上, 我们还可以构建另一个表示: 在圆环上, 选择三个距离相等的点, 顺时针地记作点 1, 2, 3。然后, e 就是不转动圆环, g_1 就是逆时

²当 $q = 1$ 的时候, e, g_1, g_2 都被表示为 1。这个显然不是理想的情况, 因为我们知道 e, g_1, g_2 各不相同。

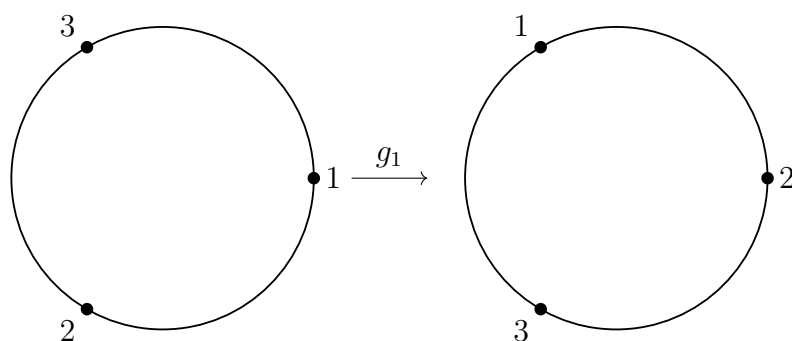


图 7.2: 在圆上的有顺序的三点被转动。 g_1 可以看作是 $(1, 2, 3) \rightarrow (2, 3, 1)$, 也就是把这三个点整体逆时针转动 120° 。

针转动一步 (如图 7.2), g_2 就是逆时针转动两步。转动三步的时候, 自然就回到原始的样子, 也就是相当于没有转动。

加入 f_i 之后, 复合映射的关系会复杂一些, 我们把它编制成下面这张表。其中行对应的映射放在左边 (后作用), 列对应的映射放在右侧 (先作用)。我们发现, 这个表的每一行和每一列, 每一个映射必须而且只能出现一次。我们来尝试理解一下这是为什么。

我们给这个 6 个元素的集合一个记号, $G_6 = \{e, g_1, g_2, f_1, f_2, f_3\}$ 。首先, 我们发现, 在复合映射下, 每一个映射都可以看作是 G_6 到 G_6 的映射。例如, $e: G_6 \mapsto G_6$ 的规则是 $e(g) = e \circ g = g, \forall g \in G_6$ 。换一个试试, $g_1: G_6 \mapsto G_6$ 的规则是 $g_1(g) = g_1 \circ g, \forall g \in G_6$ 。我们说, $g_1(g)$ 不会重复出现, 也就是不存在 $g \neq g' \in G_6$ 满足 $g_1(g) = g_1(g')$ 。

我们来用反证法, 假设存在这样的 $g \neq g'$, 使得 $g_1 \circ g = g_1 \circ g'$ 。由于 g_1 是可逆映射, 也就是在像和原像之间存在着一对一关系, 我们发现, g 必须等于 g' , 否则, 只要某一个 $j \in \mathbb{T}$ 有 $g(j) \neq g'(j)$, 则必然有 $g_1(g(j)) \neq g_1(g'(j))$ 。

不会重复出现, 又要和每一个映射去复合, 因此, 只能每一个都出现一次。也就是说, 在复合映射下, 每一个映射也都 G_6 到 G_6 的可逆映射。

列的情况也一样。因此, 我们得到了每一行每一列每一个 G_6 的元素必须而且仅仅出现一次的结果。你有没有发现, 这个和“数独” (Sudoku, 每行每列每个宫格每个数只出现一次的含义) 游戏很像很像。你可能觉得数独就是一个编出来的游戏而已。现在, 我们似乎看到了数独后面的数学结构。

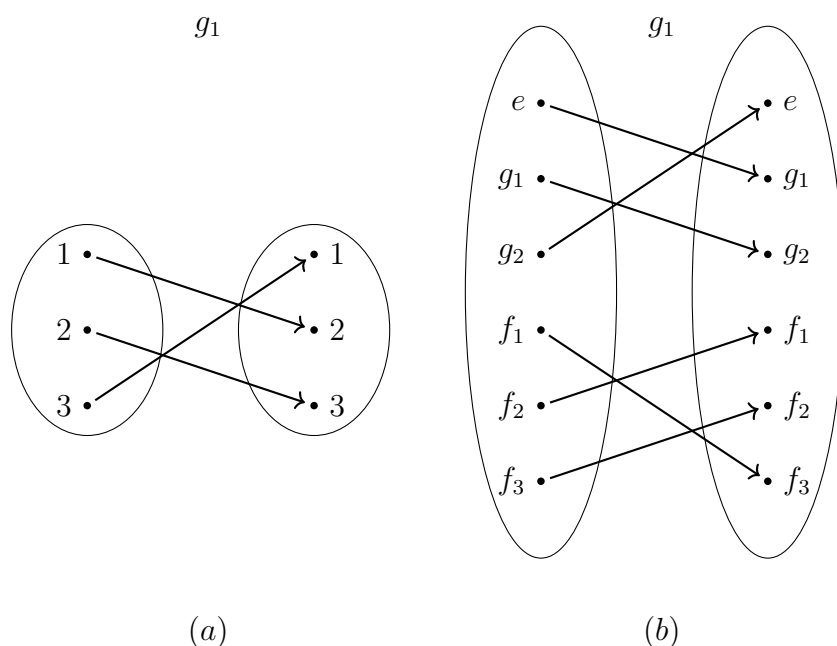


图 7.3: g_1 看作两种映射。 g_1 可以看作是 $\mathbb{T}$ 到 $\mathbb{T}$ 的映射，也可以看作是 G_6 到 G_6 的映射。

习题 7.4 (数独游戏). 请你查一查，数独游戏是什么，其背后的数学都有什么。提示：例如，你可以查“数独”（或者“*Sudoku*”）、“数独和群”（或者“*Sudoku and group*”）。

注意，这里的每一个映射有了两种含义：第一，在原始的语言下，它是 $\mathbb{T}$ 到 $\mathbb{T}$ 的映射；第二，在复合映射的语言下，它是 G_6 到 G_6 的映射。并且，看起来，我们完全可以脱离 $\mathbb{T}$ 而直接在 G_6 讨论这些映射以及这些映射的表示。

将来我们会学习到， G_6 这样的集合——这个集合上有复合映射关系，并且在每行每列中每个元素在这个复合映射关系下出现而且仅仅出现一次——叫做群。群通常建立在对一个更加基本的对象——例如这里的集合 $\mathbb{T}$ ——之上，但是，也可以忘掉这个更加基本的对象，仅仅从复合映射关系出发来研究群。

仔细观察 G_6 的复合映射结果列表，我们发现，一定程度上，这个列表是分块的： e, g_1, g_2 基本上构成了一个内部封闭的区域，在这个小区域内不牵涉到 f_1, f_2, f_3 ；另一方面，一旦加入了 f_1, f_2, f_3 ，则这两片又混合

起来了，联系起来了。

在本小节中，我们用集合 $\mathbb{T}$ 来定义其上的映射，然后把这些映射再看做一个集合 G_6 ，接着再用复合映射来定义 G_6 上的映射，而且这些映射正好是 G_6 的元素自己。最后，通过这个复合映射我们初步体验了群这个数学结构，尤其是其结构上的两面性：第一面， G_6 建立在对更加基本的对象 $\mathbb{T}$ 的操作之上，也就是 G_6 中的任何一个映射把一个 $\mathbb{T}$ 元素变成 $\mathbb{T}$ 的元素，维持 $\mathbb{T}$ 整体不变；第二面， G_6 可以完全脱离其基本作用对象 $\mathbb{T}$ ，单纯从复合映射的角度来研究。

我们已经看到群可以从依赖对象的操作和操作之间的复合关系来看，还可以直接从操作之间的复合关系来看，仅仅名义上需要一个对象实际上不再需要对象（也就是不再需要明确地看到集合 $\mathbb{T}$ 内部的元素在映射 g 之下导致的原像-像关系）；而且，这两个视角之间还可以转换。这一点非常重要，未来会启发一个非常非常的数学结构“范畴”的提出。

7.5 推荐阅读材料

7.6 作业

7.7 本章小结

在这一章中，我们学习了映射的基本概念还有单射、满射、可逆映射。我们把映射先看作是对集合的操作，或者说对集合中每一个元素的操作。进一步，我们也可以把映射看作是作用在映射组成的集合上的操作，也就是操作的复合，或者说复合映射。有意思的事情是，一旦我们把映射看作是通过复合映射对映射集合的作用，我们就可以完全脱离建立这个映射的时候的被操作的对象的集合。这样我们就有了脱离被操作对象的集合定义的抽象的映射了。

未来，我们会从中发展出来群和范畴的数学概念。这一章，我们仅仅通过这样的一个例子来展示了一下。

集合和映射是我们后续定义一切数学概念的基础。这一章非常抽象，可能需要是不是回来多学习几遍。

第八章 逻辑学初步

Pure mathematics is, in its way, the poetry of logical ideas.

纯粹数学，就其本质而言，是逻辑思想的诗篇。

– Albert Einstein

Logic will get you from A to B. Imagination will take you everywhere.

逻辑会带你从 A 点到达 B 点，想象力将把你带到任何地方。

– Albert Einstein

8.1 从集合到命题

一个命题是关于对象之间关系的一个判断句。我们已经知道集合可以用来清楚地指代对象。那么，集合可以用来指代命题吗？可以，通过集合之间的关系。既然命题是对象之间的关系，对象对应着集合，则对象之间的关系就成了集合之间的关系。

让我们从日常语言的例子开始。注意，日常语言的模糊性比较大。例如，我们说，“人总是要吃饭的”¹。这是一个判断句，是“人”和“需要吃饭的”之间的关系的一个命题。在这里，假设“人”这个集合是明确的，并且进一步明确为“活着的和还希望继续活着的人”。也就是说，我们假设这个集合是可以定义的，具有明确的元素的²。我们再来看看“需要吃饭的”这个集合是不是可以定义。首先，我们需要把“饭”拓展和明确一下。原则上，

¹这就是“人是铁，饭是钢，一顿不吃饿得慌”这句话的朴素说法。

²实际上，还真的不一定满足明确性的要求。

如果指的是大米饭,或者中餐、西餐等任何一个具体的餐饮的种类,则人其实可以不吃这样的具体的饭的种类的。因此,我们把“饭”看作是“能量输入”³。于是,“人总是要吃饭的”就成了“活着和还希望继续活着的人总是需要摄入能量的”。这里还有一个程度副词“总”需要进一步明确一下。“总”在这里有两个意思:“都”和“总有一天,也就是很长时间以后,最终”。因此合起来就是,“每一个活着和还希望继续活着的人过了足够长时间以后是需要摄入能量的”。我们把“每一个活着和还希望继续活着的人”看作集合 A ,把“过了足够长时间以后需要摄入能量的”事物看作集合 B 。然后,我们说,这个命题被表示为“ $A \Rightarrow B$ ”,其含义是任何一个来自于 A 中的元素,必然属于 B 。也就是集合具有包含关系 $A \subset B$ 。

那么,这个命题是否成立呢?也就是集合 A, B 是否真的具有包含关系 $A \subset B$ 呢?这是另外一个问题,称作命题的检验或者证明。这可以通过逻辑检验和事实性检验来完成。我们稍后正式学习逻辑的时候,我们再来讨论如何判断命题是否成立的问题。

让我们再举一个数学语言表达的例子——“所有的偶数都是整数”。其含义,用集合的语言就是,偶数的集合(记作 $A_{2\mathbb{Z}}$)中的任何一个元素,必然是整数的集合(记作 $\mathbb{Z}$)中的元素,也就是 $A_{2\mathbb{Z}} \subset \mathbb{Z}$ 。

通过这两个例子,我们发现,命题就是集合之间的包含关系。

8.2 什么是和为什么需要逻辑?

让我们从一个日常生活的例子开始。我们假设“天下雨,地就湿”成立,来看看“天不下雨,地就不湿”、“地湿了,天下雨了”、“地没有湿,天没有下雨”这三句话是否成立。

注意,我们没有一开始就去检验“天下雨,地就湿”是否成立哦,而是假设“天下雨,地就湿”成立。当然,如果你愿意,我们也可以先来做一个“天下雨,地就湿”是否成立的事实性检验。既然“天下雨”了,肯定有降水。这个水肯定都有地方去。当然,我们可以把降水区域全都保护起来,把雨水收集起来,一点都不让这些个雨水接触地面。但是,一般情况下,这样

³或者也可以更具体一点“生物形式的能量输入”,从而来和“电能”之类的能量输入区别开来。不过,为了讨论的方便,放大一点到“能量输入”也可以。

做的成本也太高了，也不太可能做到对降水区域的完全覆盖。因此，在日常语言的范围内，我们认为“如果天下雨了，那么（总有那么一小块）地（地球的表面）就湿（沾了雨水）”成立。你看，一旦进入到事实性检验，我们需要关注很多细节。不过，让我们暂时接受“天下雨，地就湿”成立。

我们关注的问题是，假设“天下雨，地就湿”成立，来看看“天不下雨，地就不湿”、“地湿了，天下雨了”、“地没有湿，天没有下雨”这三句话是否成立。

“天不下雨，地就不湿”是否成立，我们从逻辑上不能做判断。但是，从事实上，洒的水浇的水，以及更多其他方式，都可以导致地湿。因此，“天不下雨，地就不湿”是不成立的。但是，我们还是没有回答，是不是可以从假设“天下雨，地就湿”成立来得到“天不下雨，地就不湿”。我们稍后回到这个问题。

“地湿了，天下雨了”是否成立，我们从逻辑上不能做判断。但是，从事实上，地湿了可以来自于洒的水浇的水，以及更多其他方式，而不一定是天下雨。因此，“地湿了，天下雨了”是不成立的。但是，我们还是没有回答，是不是可以从假设“天下雨，地就湿”成立来得到“地湿了，天下雨了”。我们稍后回到这个问题。

“地没有湿，天没有下雨”是否成立，我们来试试从逻辑上做一个判断。这个命题在同样的前提（“地没有湿”）下，如果不成立，则肯定结论是“天下雨了”。但是，根据“天下雨，地就湿”，我们发现，这个前提“地没有湿”就不成立。任何一个命题，前提的成立是一个假设，也就是“地没有湿，天没有下雨”的含义是“如果地没有湿，那么天没有下雨”，因此前提不能不成立的⁴。于是，我们得到了“地没有湿，天没有下雨”不成立的结论是不成立的。也就是，“地没有湿，天没有下雨”只能是成立的⁵。

那么，是不是可以把前面的四个命题之间的关系，尤其是，其中第一个可以用来证明第四个这样的关系，推广到一般的命题上去，从而更多地时候我们可以不依赖于事实性检验来判断命题的真假呢？也就是说，如果我们从事实性判断已经肯定了某个命题的成立⁶，则我们就可以从这个命题出

⁴这里，我们用到了反证法，我们会稍后用集合的语言来展示反证法确实是成立的。

⁵这里我们用到了逻辑的一个性质：一个命题只能要么成立要么不成立，两者必须而且只能取其一。后面，我们也会用集合的语言来说明这个性质。

⁶事实性判断是否可以推断出概念性命题成立，这是一个问题。一般来说，事实性判断

发来论证其他命题是否成立。

这样大大增加了判断命题是否成立的效率。这也使得研究者可以实现分工，一部分人从事实性检验的工作，另一部分人从事按照已知甚至已经假设成立的命题出发来论证其他命题的工作。其实，后一部分就是数学家和各个学科的理论研究者，前一部分是各个学科的实验研究者。当然，除了进行对后一部分的研究者假设的和得到的命题做检验，前一部分研究者还给后一部分研究者提供启发，构建新的命题体系的启发。

逻辑就是由于这个更加高效率地判断命题是否成立，从而尽可能地依靠非事实性检验，而建立起来的知识体系。回到前面提到的四个命题，用数学符号来表示就是，原命题 $R = 1 \Rightarrow W = 1$ ，否命题 $R = 0 \Rightarrow W = 0$ ，逆命题 $W = 1 \Rightarrow R = 1$ ，逆否命题 $W = 0 \Rightarrow R = 0$ 。其中 $R = 1$ 表示下雨， $R = 0$ 表示没有下雨， $W = 1$ 表示地湿， $W = 0$ 表示地不湿。更一般地，从一个原命题 $A \Rightarrow B$ 出发，我们把否定了前提否定了结论的命题 $\bar{A} \Rightarrow \bar{B}$ 称作否命题，把颠倒了前提和结论的命题 $B \Rightarrow A$ 称作逆命题，把否定了前提否定了结论的又颠倒了前提和结论的命题 $\bar{B} \Rightarrow \bar{A}$ 称作逆否命题。我们可以看到，逆否命题的逆否命题就是原命题。也就是原命题和逆否命题互为逆否命题，或者称作具有逆否命题关系。可以检验，否命题和逆命题之间也是逆否命题的关系。

习题 8.1 (否命题和逆命题). 证明否命题和逆命题之间互为逆否命题。

稍后我们就会看到，逻辑告诉我们说，原命题和逆否命题完全等价。那么，逻辑是如何来展示这样的等价性的呢？或者说，更一般地，是如何从一个命题来证明或者导出另一个命题的呢？提供方法来帮助研究命题之间的关系，就是逻辑。

从上面这些例子中，我们粗糙地看到了什么是逻辑以及为什么我们需要逻辑。人们研究科学根本上，是为了得到一些和现实相符的命题，甚至在这个现实真的出现之前。我们固然可以通过把这个现实实现，然后再来和这个命题相比较的方式来检验每一个这样的命题。但是，我们还可以从基于每一个命题涉及到的具体事实的判断走到基于超越事实的一般性的判断方法——这就需要逻辑。效率更高，或者说，把逻辑判断和事实判断分开进行。如果逻辑判断都通不过，则不再需要进行事实判断。或者说，只要从事

只能推断一个概念性命题不成立。后面我们会回到这个问题。

实开始构建出来了一个和当前所有已知的事实不矛盾的命题的体系——所谓体系就是有一些基本的命题当作基础其他的命题建立在这些基本的命题之上，则先运用这个命题的体系来判断命题的真假。最后，我们只需要把这些基本的命题做一个事实性检验，而不需要每一个命题都做事实性检验。

当然，在关于现实世界对象的科学研究中，我们强调每一个依赖逻辑体系得到的命题，仍然得做事实性检验；同时，我们还强调，基于已有事实构建的逻辑体系本身一旦发现了新的和这个逻辑体系冲突的新的事实，则得重新构建包含了新的事实的逻辑体系。

8.3 逻辑三律和演绎逻辑

前面，我们已经看到，命题就是集合包含关系，命题可以成立——称做这个命题的真值为真，也可以不成立——称做这个命题的真值为假。下面，我们来介绍逻辑的基本概念。

让我们首先从逻辑变量开始。我们把一个命题 $A \Rightarrow B$ 的真值当作一个逻辑变量 l ，其可能为真记作 $l = T$ 或者 $l = 1$ ，可能为假记作 $l = F$ 或者 $l = 0$ 。用集合的语言， $l = 1$ 意味着 $A \subset B$ ， $l = 0$ 意味着 $A \not\subset B$ ，也就是存在着 $\alpha \in A, \alpha \notin B$ 。

我们首先说，任何一个逻辑变量 l 其取值必须是 $l = 1$ 或者 $l = 0$ ，并且不能两者同时成立。前者称为排中律，后者称为无矛盾律。注意 $l = 1$ 或者 $l = 0$ 互为否定 ($l = 1 \Leftrightarrow \bar{l} = 0$)，如果两者可以同时成立，就产生了矛盾。排中律的意思是，原命题 ($l = 1$) 和否命题 ($\bar{l} = 1$ ，也就是 $l = 0$) 必须有一个成立。我们可以直接从逻辑的基本性质出发来构建逻辑的知识。不过，既然我们学习了集合，我们还可以从集合来理解或者说建构逻辑的知识。

为什么 $l = 1$ 和 $l = 0$ 必须而且只能取其一？让我们回到命题 $A \Rightarrow B$ 的集合表示 $A \subset B$ ，也就是判断是否 $\forall \alpha \in A, \alpha \in B$ 的问题。首先，根据集合元素的确定性， $\forall \alpha \in A, \alpha \in B$ 的问题是一个可以明确回答的问题。如果“是”，则 $l = 1$ ，如果“否”则 $l = 0$ 。而且，这个答案不能是“是、否”两者都成立，因为集合的元素确定性说了“一个元素 α 是否属于一个集合 B 有且只有一个答案，要么是要么否”。因此，逻辑中的排中律和无矛盾律都来自于集合元素的确定性。

我们还注意到，集合元素还有一个互异性，也就是一个东西只在一个集合中出现一次，只要是这个东西（对应着的记号、名称）那么就是这个东西。这就是逻辑的同一律。

好了，我们有了逻辑三律——同一律、排中律和无矛盾律，并且它们都来自于集合的性质，也就是集合元素的确定性和互异性。

8.3.1 三段论

有了作为逻辑知识的基础的逻辑三律实际上来自于集合知识这个认知，我们就可以来讨论三段论这个演绎证明的形式了。在后面的讨论中，绝大多数情况下，我们都从集合的知识开始。

让我们从三段论的一个著名例子开始：

大前提：所有人都是要死的，

小前提：苏格拉底 (Socrate) 是人，

结论：所以苏格拉底 (Socrate) 是要死的。

所有的三段论都可以被标准化为这样的“大前提、小前提和结论”的三段的格式，所以被称为三段论。用集合的语言，我们重新表述一下，就是

大前提： $H \subset M$ ，

小前提： $s \in H$ ，

结论： $s \in M$ 。

其中 H 表示所有人的集合， M 表示所有要死的（具有有限的寿命的）事物的集合， s 代表苏格拉底 (Socrate)。也可以把这个例子画成集合表示的图，见图 8.1。

我们来看看为什么这个成立，从集合的角度。首先，已知 $H \subset M$ ，也就是 $\forall h \in H, h \in M$ 。其次，我们还已知 $s \in H$ 。于是，自然地， $s \in M$ ，也就得到了最后的结论。也就是说，所谓逻辑中的三段论演绎法不过就是集合的包含关系的基本定义。只要假设集合元素的确定性（以及平庸的互异性和无序性）是正确的，那么，自然三段论演绎法就是正确的。

注意，三段论演绎法的每一段话都是一个命题，并且从前两个命题，得到了后一个命题。也就是说，三段论演绎法可以用来建立命题之间的导出关系，做命题的论证。

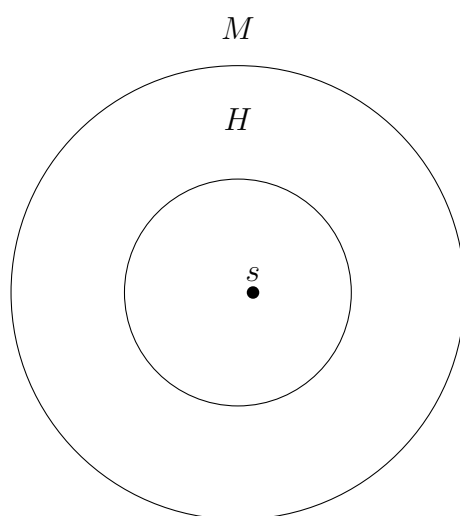


图 8.1: 苏格拉底三段论的集合表示。用 Venn 图展示了以苏格拉底为例的三段论。其中 M 表示所有要死的事物的集合, H 表示所有人的集合, s 代表苏格拉底。

编制一些三段论的应用题

8.3.2 逆否命题的等价关系和反证法

反证法是逻辑证明中另一个常用的方法。我们来看看它的基础是什么, 为什么成立。为此, 我们需要先讨论逆否命题的等价关系。

我们已经知道, 原命题 $A \Rightarrow B$ 的集合表示是 $A \subset B$ 。同样地, 命题 $\bar{B} \Rightarrow \bar{A}$ 的集合表示是 $\bar{B} \subset \bar{A}$ 。其中, $\bar{A}, \bar{B}$ 分别是集合 A, B 的补集。注意, 既然可以讨论补集, 则肯定有一个天然定义好的全集 I 。例如, 在“天下雨”的例子中, “天下雨”的反面就是“天不下雨”, 合起来的全集就是“天下雨或者不下雨”。下面, 我们来证明 $A \subset B$ 就是 $\bar{B} \subset \bar{A}$ 。

我们先通过画图——见图 8.2——来建立一个直觉, 然后再用集合的语言来证明。首先, 从 $A \subset B$ 看到, 集合 A 被包含在 B 里面, 两者都在全集 I 里面。于是, 我们自然可以看到 A 的补集更大, 包含了 B 的补集。一个集合的补集是那些不在这个集合但是在全集里面的元素的整体。

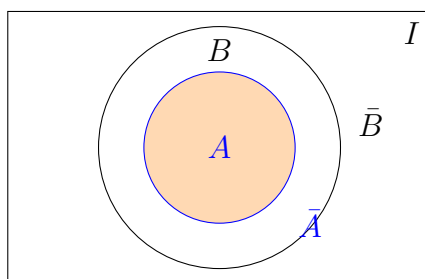


图 8.2: 原命题和逆否命题等价。用集合的 Venn 图展示了原命题和逆否命题等价。

定理 8.1: 逆否命题集合关系

逆否命题集合关系 $A \subset B \Rightarrow \bar{B} \subset \bar{A}$ 。

证明 8.1. 选择任意一个 $\bar{B}$ 中的元素 x , $x \in I, x \notin B$ 。我们来证明 $x \in \bar{A}$ 。这就要求我们证明 $x \notin A, x \in I$ 。第二条自然成立。我们来看第一条。由于任意的 x , 两种关系中 $x \in A$ 和 $x \notin A$ 有且只有一种成立。我们只要证明了 $x \in A$ 不成立, 则肯定 $x \notin A$ 成立。

根据 $A \subset B$, 任意的 $x \in A$ 有 $x \in B$ 。我们这里已经有了 $x \notin B$, 所以, $x \in B$ 不成立。因此, $x \in A$ 也不成立。

在证明过程中, 我们多次用到了集合元素的确定性, 或者说等价地, 逻辑三律。

定理 8.2: 逆否命题等价性

逆否命题等价性 $A \Rightarrow B \iff \bar{B} \Rightarrow \bar{A}$ 。

证明 8.2. 根据命题和集合的关系, $A \Rightarrow B$ 等价于 $A \subset B$, $A \subset B \Rightarrow \bar{B} \subset \bar{A}$, 而 $\bar{B} \subset \bar{A}$ 和 $\bar{B} \Rightarrow \bar{A}$ 等价, 因此, $A \Rightarrow B \iff \bar{B} \Rightarrow \bar{A}$ 。

或者, 连起来写成

$$A \Rightarrow B \iff A \subset B \iff \bar{B} \subset \bar{A} \iff \bar{B} \Rightarrow \bar{A}. \quad (8.1)$$

或者，也可以写成更方便阅读的形式，

$$(A \Rightarrow B) \Longleftrightarrow (A \subset B) \Longleftrightarrow (\overline{B} \subset \overline{A}) \Longleftrightarrow (\overline{B} \Rightarrow \overline{A}). \quad (8.2)$$

既然原命题和逆否命题是完全等价的，那么，当我们证明原命题 $A \Rightarrow B$ 有困难的时候，我们就可以尝试来证明逆否命题，也就是 $\overline{B} \Rightarrow \overline{A}$ 。也就是先试着假设结论不成立，看看能否推导出来前提也不成立。如果可以，则仍然相当于证明了原命题。

编制一些反证法的应用题

8.3.3 逻辑变量的运算

有的时候，我们需要把若干个命题合起来组成一个整体的命题，例如“天下雨，地就湿，而且地就滑”⁷，例如“偶数是整数，并且不是奇数”，“合数是整数，并且不是质数”，“偶数是整数，并且是合数”，“偶数是质数，或者合数”等等。这时候，我们就需要来做若干个逻辑变量的运算。基本的逻辑变量的计算包含三种，非 ($\neg l$)，与 ($l_1 \wedge l_2$)，或 ($l_1 \vee l_2$)。它们的计算规则分别是

$$\neg l = \begin{cases} 0 & l = 1 \\ 1 & l = 0 \end{cases} \quad (8.3)$$

就是把逻辑变量的值反过来。

$$l_1 \wedge l_2 = \begin{cases} 1 & l_1 = 1, l_2 = 1 \\ 0 & \text{其他} \end{cases} \quad (8.4)$$

就是只有两个逻辑变量的取值都为真的时候，才为真，否则为假。

$$l_1 \vee l_2 = \begin{cases} 0 & l_1 = 0, l_2 = 0 \\ 1 & \text{其他} \end{cases} \quad (8.5)$$

就是只需要两个逻辑变量的任意一个取值都为真的时候，都为真，否则为假。

⁷同样地，如果要成为命题，首先得是集合，含义要明确，因此，如果我们真的来讨论这个命题我们还得定义“滑”。

8.4 逻辑是共通的普适的吗？

人类的论证方式为什么会有一个共通的基础，也就是逻辑？人类的还是世界的什么共性决定了人类在论证方式上的共性？还有其他的逻辑形式吗？我们发现，在接受集合知识的前提下，只有这一种，是普适的。但是，问题是为什么又要接受集合论呢？集合论又为什么是共通的普适的呢？如果单纯从人类自身的角度来说，出现一个这样共同的集合或者逻辑，是一件非常神奇的事情。

但是，很有可能逻辑是人类的大脑受到现实的刺激而演化从而自洽形成的。也就是说，我们自然地要去给事物按照其属性，并且是通过人类的思维认识到的那些属性来分类的。有了分类之后，自然要考虑这些类之间的关系，不属于这些类又怎样。因此，集合的知识看起来是对人类为了认识自然而采取的最朴素的认识方式的概念化。

当然，到底是事物本身有规律可循，可以被分类，可以被认知，还是人类通过人类思维认识了事物给了事物这些属性这些组织方式，才使得事物有规律可循，可以被认知，那就是另一个问题了。

不过，还是有人在追求可能的另外一套逻辑的[添加参考文献](#)。同样，只要是自洽的，也就是自身内部不包含矛盾的命题，都可以被提出来。但是，迄今为止，人们并没找到这样一套不同于我们前面学习的逻辑体系的逻辑[添加参考文献](#)。

8.5 逻辑和批判性思维

学会了逻辑，首先就是用来给命题做效率更高更准确的判断的。也就是前面提到的通过命题来证明命题，而不是每个命题都依赖于事实性检验。毕竟，事实性检验的成本高很多，但是通过命题来证明命题是可以在我们的脑子里面来完成的，往往成本低很多。

但是，学会了逻辑，还有可以支撑一个更加重要的思维方式或者说思维习惯，那就是批判性思维，也就是“我从来不把没有经过我自己的理性检验的命题放入到我的认知结构之中”。这里的理性检验包含逻辑论证和事实性检验。事实性检验一般通过做实验来进行。

例如，我可以告诉你“地球是圆的”、“地球围着太阳转”、“三角形的内

角和是 180° ”、“两直线平行则同位角相等”等等一系列的命题，但是，只有经过你自己的逻辑论证可以被证明的，或者是通过了事实性检验的命题⁸，你才能接受它们，把它们放到你的脑子里面，成为你进一步提出问题和解决问题的基础。

那么，如何来做这个逻辑检验呢？这就要靠知识的公理化体系。在数学和科学知识中，人们往往从定义和公理来构建整个知识体系。也就是说，无论受到现实的启发，还是受到现有知识体系中的问题（例如发现，应该由某个定理，但是目前这个定理还没有得到证明，逻辑链条还不通）的挑战，还是受到来自于自身想象力的启发，人们会提出一些定义，提出一些公理也就是假设成立的命题，然后，人们企图从这些定义和公理出发来得到定理。我们学习的时候也一样，首先，至少要能够自己来从定义和公理来推导出来定理。其次，更高的要求是，我们也能够受到相应的启发自己去尝试“独立地再一次”提出定义或者公理，猜测出来定理，然后再来证明这些定理。只有这样自己来发展和构建起来这个公理化的知识体系，我们将来在遇到问题的时候，就会有方向和有意识地运用合适的定义、公理和定理来提出和解决问题，甚至运用那些启发来提出和解决问题，而不是去凑去试，看哪一个定义、公理和定理正好可以解决某个问题。更进一步，如果我们有受到启发的经验，我们将来遇到新场景可能也会再次受到启发，从而提出真正的新概念、新公理，猜测和证明新定理，甚至发现新的启发。自己构建起来和体会到知识体系，自己受到了和体会到了促使发展新概念、新公理和新定理的启发，从而具有了自己来面对新场景、再次受启发从而提出和解决问题的可能性，甚至形成这样的习惯和随时准备好这样做的意愿，就是理解型学习的目标 [2]，就掌握了高层知识生成器。

下面我们来看几个把演绎逻辑用于批判性思维的例子。

实际上，基于集合和命题的整个知识体系，有可能会有问题——并不是所有的命题都可以判断其是否成立。我们在下面这个小知识中来给大家做一个小小的展示。

小知识 8.1 (说谎者悖论和理发师悖论). 有一个人经常说谎，然后，他宣称“我在说谎”。我们来看看这句话是否成立。首先，如果这句话成立，那么

⁸再次强调，事实性检验只能否定命题，不能真的证实命题。见后续逻辑和科学的关系的讨论。

“我在谎言”本身是谎话，谎言的意思是这句话不成立。产生了矛盾。如果这句话不成立，那么，说谎的反面就是说真话，也就是“我在说真话”。只要这句话是真话，我们又回到了第一种情况。在此产生了矛盾。

在这里，我们用到了一个逻辑的基本性质，任何一个命题其要么成立要么不成立，两者必须而且只能居其一。按照集合的基本要求，给定任何一个集合，任何一个元素要么属于这个集合要么不属于这个集合。我们如果把所有的成立的命题构成一个集合 $\mathbb{T}_p$ 。那么任何一个命题 l ，根据集合元素的确定性，要么属于这个集合要么不属于这个集合。

于是，上面这个说谎者悖论表明了：“任何一个命题其要么成立要么不成立，两者必须而且只能居其一”必然导致“有的命题是否成立不能被判断”，或者“有的命题的真值不是成立和不成立两者必须而且只能取其一的”。

我们再来看另一个例子：小镇上有一个理发师，他宣称他不给自己理发的人理发，但是同时他给所有不是自己理发的人理发。这个规则在不牵涉到他自己的时候是逻辑上完全没有问题的。但是，我们来看看，他怎么把这个规则用于他自己。如果他决定给他自己理发，则他就是那个给自己理发的人，按照其规则，他不能给自己理发的人——在这里就是他自己——理发，因此，产生了矛盾。如果他决定不给自己理发，则按照其规则，他必须给不给自己理发的人——在这里就是他自己——理发，因此，有产生了矛盾。因此，这条规则是自我矛盾的。

Russell 给这个理发师悖论构建了一个集合的语言来描述。首先，定义一个集合 $\mathbb{S}$ 为“所有的不包含自身的集合”，也就是 $\mathbb{S} = \{x | x \notin x\}$ 。这个定义中用到了对于一个集合自身是否包含在这个集合之中的判断。按照集合的元素的确定性——给定一个集合，对于任何一个元素，这个元素在属于和不属于这个集合之间必须而且只能有一项成立，这个判断确实可以被执行。

当然，我们之间遇到的集合，都不会去判断这个集合本身是不是这个集合的元素。例如，自然数集合的元素是一个个的自然数，从来不会说，自然数集合本身是否属于自然数集合。但是，这个问题仍然允许被问。

接着 *Russell* 问，那么， $\mathbb{S}$ 是否属于 $\mathbb{S}$ 呢？假设 $\mathbb{S} \in \mathbb{S}$ ，那么按照 $\mathbb{S}$ 的定义 $\mathbb{S} = \{x | x \notin x\}$ ，刚好违反了这条定义，因此， $\mathbb{S} \notin \mathbb{S}$ 。但是，只要 $\mathbb{S} \notin \mathbb{S}$ ，

则按照 S 的定义，这个 S 就刚好符合这个定义，因此， $S \in S$ 。矛盾。

看起来，所有的矛盾的根源都在集合元素的确定性——也就是给定一个集合，对于任何一个元素，这个元素在属于和不属于这个集合之间必须而且只能有一项成立——这一条上。

那这个问题到底怎么解决呢？留给你去思考，或者进一步学习。

顺便，从批判性思维的角度，这其实是运用逻辑进行了对集合论和逻辑的本身的批判性思维。

8.6 逻辑和数学、逻辑和科学

我们已经看到，如果已经有了概念和命题，则数学主要靠逻辑演绎证明或者表现为计算的逻辑演绎证明来否定或者确认命题。同时我们发现，用集合的语言来表述逻辑之后，逻辑学本身也成了数学的研究对象。那问题是，一开始的数学的概念和命题从哪里来呢？

来自于受现实启发的人的思想，来自于人的想象力和直觉（其背后仍然可能是现实的启发、对某些概念和命题的长期的理解和运用导致的自动化，以及对某些待解决问题的长期的思考）。这也是Albert Einstein（爱因斯坦）说的“逻辑会带你从 A 点到达 B 点，想象力将把你带到任何地方”的意思。

因此，我们需要掌握逻辑，用好逻辑，但是不能停止于或者说仅仅依靠逻辑。知识的创造的起点是现实世界，以及你的大脑对现实世界的加工。因此，受现实或者直觉、想象力的启发，做概念的抽象提炼，做命题的猜测，在创造数学知识上是非常重要的。

我们强调逻辑，不是说受现实或者直觉、想象力的启发就不重要，而是这些才是珍贵的重要的：你要把珍贵的重要的没有太多章法可循的用在最需要它们的场合，然后其他一切地方能运用不珍贵的有章法可循的逻辑的就用逻辑，并且还需要靠逻辑去发现珍贵的东西的产物里面可能包含的错误或者悖论⁹。

当然，我们也看到了逻辑对于数学知识的创造也有另一个作用：当我

⁹悖论一方面意味着现有知识体系有错，另一方面往往意味着一个可能的知识的大的突破。

们用知识的系统性来要求数学知识的时候，我们也会创造出来一些新的概念和命题。例如，我们依靠运算的封闭性拓展了数。未来我们学习到微积分的时候，也会为了使得导数的概念更加严格而创造出来极限这个概念，以及实数的构造性定义。

对于数学知识来说，只要有了一开始的公理和定义，通过逻辑演绎证明形成了严密的知识体系，就不需要进一步检验。它不受现实世界的约束，不屑于接受现实世界的检验。

但是，科学不一样。科学同时需要来自现实、想象力和直觉的启发，严格的逻辑演绎证明和表现为计算的逻辑演绎证明，还需要实验来检验通过前面的过程建立起来的命题。

那么，是不是科学家做了实验，发现某个命题和实验相符就够了呢？比如说“重物下落更快”显然可以通过大量的实验检验：我们今天知道这个结论是错的，也知道只要让空气阻力发挥作用，则确实很多时候可以观察到和这个结论相符的实验现象。当然，你可以说，那是做的还不够。继续做更多次，只要一个命题真的是错的，总是会遇到和这个命题不相符的实验现象的。

但是，做多少次才够呢？原则上，从纯逻辑来说，做多少次都是不够的。我们不能因为观察到一千只乌鸦都是黑的，就说天下乌鸦都是黑的，一万只，一亿只，也同样如此。

Karl Popper（卡尔·波普尔）注意到了这个问题，提出来，通常科学家坚称的“可检验、可验证”其实是科学知识所做不到的，应该改为“可证伪性”：一个命题只要有被证明是错的可能性，也就是存在着一个实验其中的一个实验结果会证明这个命题不对，并且迄今为止还没有每证明是错的，那这个命题就可以成为科学知识的一部分。当然，实际上，这个区分对科学家意义不大。完全可以明白之后，继续用那些词。因此，大部分科学家仍然使用“科学的可验证性”，尽管从纯逻辑的角度确实改成“科学的可证伪性”更合适。

在科学知识的创造和学习中，我们仍然非常强调逻辑，强调知识的系统性。同样，这也不是受现实或者直觉、想象力的启发就不重要，而是这些才是珍贵的重要的，甚至相比于数学，它们在科学中更加的重要。背后的逻辑还是一样的：你要把珍贵的重要的没有太多章法可循的用在最需要它们

的场合，然后其他一切地方能运用不珍贵的有章法可循的逻辑的就用逻辑，并且还需要靠逻辑去发现珍贵的东西的产物里面可能包含的错误或者悖论。

数学和科学都是逻辑、现实、想象力、直觉一起编织的诗歌，只不过数学不需要接受现实的检验，仅提供思维和描述世界的语言，而科学需要接受现实的检验。

8.7 推荐阅读材料

8.8 作业

8.9 本章小结

第九章 一元函数

当一个映射是建立在两个关于数的集合之间的时候，我们就称这个映射为函数。在这一章中，我们来学习什么是一元函数以及几个一元函数的例子。

9.1 一元函数当作一个吞吐机

一元函数 $y = f(x)$ 是把一个数 x 变成另一个数 $f(x)$ 的一个映射。前一个数的集合称作定义域记作 X ，后一个数的集合称作陪域记作 Y 。例如， $f(x) = x + 1$ ，是一个从实数集到实数集的映射。我们可以验证其满足映射的要求：对于任意一个实数 $x \in \mathbb{R}$ ，存在着唯一的 $f(x) = x + 1 \in \mathbb{R}$ 。顺便，为了熟悉一下满射和单射的概念，我们还可以验证这是映射是满射和单射。

习题 9.1 ($f(x) = x + 1$ 是满射). 证明 $f(x) = x + 1$ 是满射，也就是对于任何一个 $y \in \mathbb{R}$ ，可以找到原像 $x \in \mathbb{R}$ 使得 $f(x) = y$ 。

习题 9.2 ($f(x) = x + 1$ 是单射). 证明 $f(x) = x + 1$ 是单射，也就是对于任何一个 $x_1 \neq x_2 \in \mathbb{R}$ ， $f(x_1) \neq f(x_2) \in \mathbb{R}$ 。

注意，当我们把 $f(x) = x + 1$ 看作是从非负整数集合 $\mathbb{Z}^*$ 到 $\mathbb{Z}^*$ 的映射的时候，就不再是满射了，因为找不到一个 $\mathbb{Z}^*$ 的元素使得 $x + 1 = 0$ 。因此，一个映射是不是满射和单射是和这个映射的定义域和陪域相关的。有的时候，人们也定义一个叫做值域的集合，也就是所有的像的集合 $f_X = \{y | y = f(x), \forall x \in X\}$ 。

习题 9.3 ($f(x) = x + 1$ 是满射). 证明 $f(x) = x + 1: \mathbb{Z}^* \mapsto \mathbb{Z}^*$ 不是满射, 并且找出来这个函数的值域。

这个一元函数收到或者说吞进去一个数, 吐出来一个数, 特别像一个吞吐机, 特别像一个盒子。很可能正是由于这一点, 函数被成为函数: “函”(古代字形为) 的意思是一个盒子¹, “数”的意思是这个盒子吞进去和吐出来的都是数。顺便, 大多数数学和科学概念的名称和其含义之间是具有相当程度的对应关系的。因此, 学习的时候, 概念的名称也不仅仅是一个任意选的记号, 而往往是可以理解的。

9.2 一元函数当作一个关系

函数, 或者说映射, 还有另一种理解方式, 那就是关系。朴素地说, 两个东西之间的关系就是一个东西总是伴随着另一个东西出现。例如, 在一元函数这里, 就是 $(x, f(x))$ 总是一起出现, 给一个 x 后面就有一个 $f(x)$ 跟着。更一般地, 关系可以定义为集合的直积集合的子集。

为此, 我们先来定义两个集合的直积集合。

定义 9.1 (集合的直积). 给定一个集合 X 和一个集合 Y (可以是同一个集合), 我们把集合 $X \times Y \triangleq \{(x, y) | \forall x \in X, \forall y \in Y\}$ 称作集合 X 和 Y 直积集合, 也称为 *Cartesian 积* (卡氏积)²。

关系指的是 $X \times Y$ 的一个子集。回到朴素的语言, 来自于 X 集合的 x 和集合 Y 的 y 的关系指的就是一个有序对 (x, y) 。于是, 一元函数就是一个特定的关系, 通过函数 $y = f(x)$ 构建的特定的关系有序对 $(x, f(x))$ 。意思就是说, 在一个 (x, y) 平面上, 或者说 $X \times Y$ 这个卡氏积里面, 对于一个给定 x 往往具有非常多个的 y 可以选择, 但是, 映射 $f(x)$ 使得 x 仅仅选择了唯一的 $f(x)$ 来组合成 $X \times Y$ 这个卡氏积里面的元素 $(x, f(x))$ 。因

¹ “函” 甲金文由倒“矢” (后讹变为“水”)、一个圈 (现写为“口”) 和上面可以手提的小环 (现写为“了”) 组成, 像收藏箭矢的袋子或盒子, 本义是箭袋、箭匣。见“汉字理解型学习”网站, <https://www.learnm.org/>。转引自见“汉语多功能字库”网站, <https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/>。

²这个名字来自于发明了用两条数轴的直积来得到平面直角坐标系的René Descartes (勒内·笛卡尔)。其拉丁语名为“Renatus Cartesius”

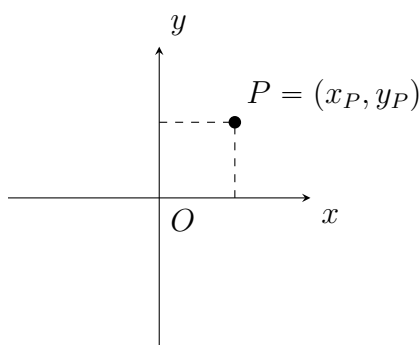


图 9.1: 平面上坐标和点之间的对应关系。给定一个平面上的点 P 可以找到一对坐标 (x_P, y_P) ，给定一对坐标 (x_P, y_P) 可以找到平面上的一个点 P 。

此， $(x, f(x))$ 或者说 $f(x)$ 可以看作一种关系。

9.3 平面直角坐标系

前面我们已经学习过数轴——实数和数轴上的点一一对应³，现在，我们用集合的直积来构建一个平面直角坐标系。平面直角坐标系的作用是让数对 (x, y) 和平面上的点一一对应。

它是这样来建立这个一一对应关系的。首先，我们已经有了实数集合 $\mathbb{R}$ 和 $\mathbb{R}$ 的卡氏积，也就是有了数对 (x, y) 的集合。接着，我们把 x 轴画成对平方向的一条直线，并且在 $x = 0$ 处画一条垂直于 x 轴的直线，称为 y 轴。同时，这个交叉点也当作 y 轴的零点。这样我们就有了坐标轴——每一个坐标轴都是一条在自己的方向上的数轴。由于这两条坐标轴成直角关系，构成了一个平面，所以称为平面直角坐标系。

现在，我们来看看，为什么叫做“坐标系”——也就是给定任何一个数对 (x, y) ，如何找到一个在上面的平面上的对应的点。我们这样来找，见图 9.1：首先，在 x 轴上，找到数对 (x, y) 中的相应的 x 值，过这个点做一条平行于 y 轴的直线；接着，在 y 轴上，找到数对 (x, y) 中的相应的 y 值，过这个点做一条平行于 x 轴的直线；两条直线的交点就是数对 (x, y) 在平面中所对应的点。根据几何知识，这个交点必然存在，并且唯一。

³尽管这个命题我们还没有证明。这个证明留给大学阶段的数学来完成。

两条坐标轴把整个平面分成了四个区域，我们也给四个区域各自一个名字： $x > 0, y > 0$ 的区域被称为第一象限， $x < 0, y > 0$ 的区域被称为第二象限， $x < 0, y < 0$ 的区域被称为第三象限， $x > 0, y < 0$ 的区域被称为第四象限。由于是基于卡氏积来定义的，平面直角坐标系也经常被称为笛卡尔坐标系。

这里我们需要用到一些几何知识，例如如何做垂线、平行线，两条不相同的直线相交要么只有一个交点，要么没有交点。我们假设这部分知识都已经在小学阶段的几何知识里面学习过。我们也会将来在中学阶段的几何知识部分中心来学习，更加严格地有系统性地来学习。

9.4 一元函数当作一个图形

我们还可以把一元函数看作一个图形，只需要借用上面提到的两个实数集合通过卡氏积构成的实数对平面，然后把每一个关系对 $(x, f(x))$ 都画到实数对平面上去。画的方法就是在 x 轴（通常选水平方向的那个实数轴）上找到 $(x, f(x))$ 中的 x 对应的点，在 y 轴（通常选垂直于水平方向的那个实数轴）上找到 $(x, f(x))$ 中的 $f(x)$ 对应的点，然后做过 x 点的平行于 y 轴的直线同时做过 $f(x)$ 点的平行于 x 轴的直线，两直线相交的点就对应着 $(x, f(x))$ 。把所有的这样的 $(x, f(x))$ 都画上去就得到了对应着函数 $y = f(x)$ 的图形。

这样，我们就从函数得到了函数的图形。现在，我们反过来，从函数图形试试来得到函数。

所谓得到一个函数，就是，对于任意的 $x \in X$ ，我们都得找到一个而且是唯一的一个 $y \in Y$ 。这里的 X, Y 是 $\mathbb{R}$ 。我们这样来完成这个任务：对于任何一个 x ，我们先找到 x 轴上对应的点，然后，过这个点做平行于 y 轴的直线。只要函数在这个点有定义，则这个直线必定和图像交于一点⁴。由于函数是映射，映射要求给定一个 x 有且只有一个 $y = f(x)$ ，这个函数的图形，在任何一个 x 点，做平行于 y 轴的直线，这条直线和函数图形只能

⁴如果函数在这个点有定义而且平行于 y 轴的直线和函数图像没有交点，则函数图像必定也是平行于 y 轴的直线。然而，平行于 y 轴的直线不构成映射——同一个 x 对应着无穷多个 y 的值。

有一个交点。找到这个交点，过这个交点做平行于 x 的直线。这个直线和 y 轴的交点就得到了函数值，也就是 x 的像， $y = y = f(x)$ 。

那为什么我们要把一元函数看作图形，而且是一个“在任何一个 x 点，做平行于 y 轴的直线，这条直线和图形只能有一个交点”的图形呢？因为从函数的表达式来学习函数，往往给我们造成一个错觉，好像只有能够写出来函数表达式的才叫做函数，例如 $f(x) = x + 1$ ， $f(x) = x^2$ 等等等等。但是，通过把函数和“在任何一个 x 点，做平行于 y 轴的直线，这条直线和图形只能有一个交点”的图形联系起来，我们发现，实际上，只要有一个满足这个要求的图形，我们就定义了函数。更进一步，为什么要求这个性质的图形呢？不过就是为了通过 x 找到 $y = f(x)$ 的上面那个作直线找交点的过程能够找到一个并且只有一个 $y = f(x)$ 而已。从现在起，只要图象满足这个性质，能不能写出来函数表达式，都是函数。

9.5 一元一次函数

前面我们已经提到了一元函数看起来很象一个吞吐机，一元函数实际上是一个从数的集合到数的集合的映射，有的时候一元函数可以表达为一个表达式 $y = f(x)$ ，一元函数实际上就是在平面直角坐标系上的一条满足特定要求的曲线图形。

在实际使用中，表达式 $y = f(x)$ 形式的定义往往更加方便。因此，下面我们开始从表达式 $y = f(x)$ 形式来讨论一元函数。这个时候，我们把 x 称作自变量， y 称作因变量。意思是 x 可以任意取，自己决定取什么，但是 y 呢随着 x 的确定就确定了。

其中最简单的是常数函数，也就是 $y = f(x) = c$ 。 $c \in R$ 是一个常数，例如 $c = 0$ ， $c = 1$ 。这个函数的因变量的取值 y 不依赖于自变量 x 。这个函数画成图形就是平行于 x 轴的直线，和 y 轴相交于 y 轴上的 $y = c$ 这个点。

不过常数函数实在太简单了，实在没什么可研究的。我们来看一个稍微复杂一点的函数——一元一次函数， $y = kx + b$ ($k \neq 0$ ，否则就回到了前面的常数函数)。首先，这个函数过点 $(0, b)$ ，和 y 轴相交。这样的交点只有一个，并且我们称 b 为截距——和 y 相交和 $y = 0$ 点距离为 b 的点。接着，我们发现，如果忽略这个截距为零，也就是暂时让 $b = 0$ ，我们得到

的是正比例函数 $y = kx$ 。我们很清楚这是一条直线，并且过原点。反过来，如果我们已经得到了正比例函数 $y = kx$ 的图象，那么我们只需要把这个图像整体加上一个 b ——这体现为把整个函数图像上下平移距离 b ，就能够得到 $y = kx + b$ 的图象。

因此，现在我们来关注 $y = kx$ 的图象。我们从一个特例 $y = x$ 开始。通过描点画图法，我们可以猜测， $y = x$ 就是平面直角坐标系中 45° 方向的一条直线：直线经过一三两个象限，而且从正中间经过。描点法的具体过程如下：任意选择一些个 x_i ，算出来每一个 x_i 对应的 $y_i = f(x_i)$ ，组成若干点对 $(x_i, f(x_i))$ ，把这些点对一一画到图上，尽量用光滑的曲线或者折线把这些点按照 x 的大小顺序相连，看看构成了什么图形。

类似地， $y = -x$ 就是平面直角坐标系中 135° 方向的一条直线：直线经过二四两个象限，而且从正中间经过。

对于更一般的 $y = kx$ ，如果 $k > 1$ ，则还是一三象限中的一条直线，但是 y 增加的更快，因此，比 $y = x$ 更加靠近 y 轴；如果 $0 < k < 1$ ，则还是一三象限中的一条直线，但是 y 增加的更慢，因此，比 $y = x$ 更加远离 y 轴而靠近 x 轴；如果 $-1 < k < 0$ ，则还是在二四象限中的一条直线，但是 y 的绝对值增加的更慢，因此，比 $y = -x$ 更加靠近 x 轴；如果 $k < -1$ ，则还是在二四象限中的一条直线，但是 y 的绝对值增加的更快，因此，比 $y = -x$ 更加靠近 y 轴。

这样，我们就得到了 $y = kx + b$ 的图象。经常我们需要在函数图像和函数表达式之间做转换。

9.6 一元一次方程的几何含义

一元一次方程是一元一次函数反过来，也就是，在函数中我们从给定的 x 来计算 $y = f(x)$ ，在方程中，我们从给定的 y_* 来找到合适的 x_* 使得 $y_* = f(x_*)$ 。我们发现，这正好是逆映射。因此，方程的求解的前提就是映射是可逆的。

我们来看看，一元一次函数对应的映射是不是可逆的。我们先从映射是否可逆开始。首先，实数集到实数集上的一元一次函数是一个满射（留作作业），也是一个单射（留作作业），因此，肯定是一个可逆映射。

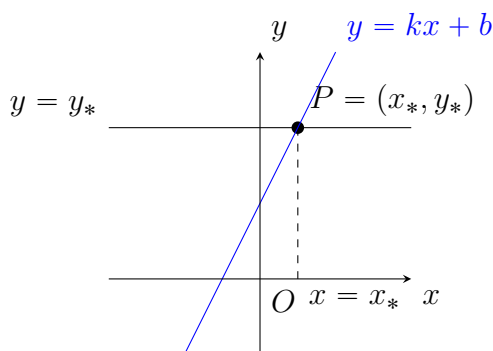


图 9.2: 一元一次方程的解的几何含义。一元一次方程 $y_* = kx_* + b$ 的解的几何含义就是其做代表的直线和另一条直线 $y = y_*$ 的交点对应的 x_* 。

习题 9.4 ($f(x) = kx + b$ 是满射). 证明 $f(x) = kx + b$ ($k \neq 0$) 是满射, 也就是对于任何一个 $y \in \mathbb{R}$, 可以找到原像 $x \in \mathbb{R}$ 使得 $f(x) = y$ 。

习题 9.5 ($f(x) = kx + b$ 是单射). 证明 $f(x) = kx + b$ ($k \neq 0$) 是单射, 也就是对于任何一个 $x_1 \neq x_2 \in \mathbb{R}$, $f(x_1) \neq f(x_2) \in \mathbb{R}$ 。

我们再从函数图象开始。也就是, 从给定的 y_* 能不能找到一个函数图象上的唯一的 x 。我们这样来找, 如图 9.2: 选 y 轴上的点 y_* , 过 y_* 点做 x 轴的平行线, 因为一元一次函数的图象也是一条直线, 因此这个过 y_* 点做 x 轴的平行线必然和函数图象有且只有一个交点 P 。除非 $k = 0$, 则 $y = b$ 是常数函数, 其和 y_* 点做 x 轴的平行线可能没有交点或者完全重合。不过, 在讨论一元一次函数的时候, 我们总是先排除 $k = 0$ 的情况。过交点 P , 做平行于 y 轴的直线, 和 x 轴交于点 Q 。交点的 Q 的 x 值就是所要找的满足 $y_* = f(x_*)$ 的 x_* 。这个交点唯一。因此, 给定 y_* 所对应的 x_* 也唯一。因此, 函数可逆。

我们再来从代数式看一元一次函数是否可逆。已知 y , 要找到 x 使得 $y = f(x) = kx + b$ 成立。可以检验, 我们只需要选 $x = \frac{y-b}{k}$, 肯定满足上面的要求。当 $k \neq 0$ 时, $x = \frac{y-b}{k}$ 存在且唯一。因此, 一元一次函数是否可逆。

我们还可以通过把每一步操作反过来来看看一元一次函数是否可逆。我们先看, 给定 y 是否可以找到 z 使得 $y = z + b$ 成立。这个时候已知的 y 就是给未知的 z 增加了 b , 因此, 反过来就是 $z = y - b$ 。这里我们用到了加法的逆运算是减法。我们再来看, 给定 z , 是否可以找到 x 使得 $z = kx$

成立。这个时候，已知的 z 不过就是未知的 z 称以 k ，只要 $k \neq 0$ 这个计算就可以反过来，得到 $x = \frac{z}{k}$ 。这里我们用到了乘法的逆运算是除法。合起来就是 $x = \frac{y-b}{k}$ 。

将来判断更一般的方程是否有解和求解的时候，也类似，可以从映射是否可逆来考虑，从函数曲线的角度来考虑，从把每一个运算逆过来考虑，或者直接检验不知道从哪里获得的解。

9.7 一元二次函数和一元二次方程

一元二次函数的表达式是 $y = ax^2 + bx + c$ ($a \neq 0$ ，否则就回到了一元一次函数)。我们也从最简单的一元二次函数 $y = x^2$ 开始。

首先，从实数集到实数集上一元二次函数 $y = x^2$ 不是一个满射，因为对于任意的 $x \in \mathbb{R}$ ， $y = x^2 \geq 0$ 。也就是说，对于 $y < 0$ 的像，我们找不到原像。

其次， $y = x^2$ 也不是一个单射，因为存在着 $x_1 = x$ 和 $x_2 = -x$ 不同 (只要 $x \neq 0$)，但是 $y_1 = x_1^2 = x^2$ ， $y_2 = x_2^2 = (-x)^2 = x^2$ ， $y_1 = y_2$ 。

因此， $y = x^2$ 不是一个可逆映射，原则上不能求解一元二次方程。一方面，我们可以把方程的含义稍微扩大一下，从找到逆映射扩大为给定 y 找到所有的 x 使得 $f(x) = y$ 就可以，不需要唯一，不需要用到逆函数。但是，另一方面，如果我们希望能够写下一个方程的来表达式形式的解，注意表达式一般来说是个函数，那么，我们就希望能够用某种方式把一个方程的多个可能的解从形式上分开，给每一个形式一个表达式。

例如， $y = x^2$ 其实可以做到，只需要分开正负两个分支，也就是 $x_1 = \sqrt{y}$ 和 $x_2 = -\sqrt{y}$ ($y \geq 0$)，或者合起来记作 $x_{1,2} = \pm \sqrt{y}$ ($y \geq 0$)。这里我们就有了两个分支，每个分支有一个函数形式，每一个分支上，函数可逆。

接着，我们再来看 $y = x^2$ 的图象。首先，我们说它的函数图象关于 y 轴对称，也就是把 x 替换为 $-x$ 函数值不变，也就是 $f(-x) = f(x)$ 。因此，我们只需要关注 $x \geq 0$ 这一侧的函数图象就行，另外一侧只需要通过轴对称图形来转化就可以得到。

为了得到 $y = x^2$ 在第一象限内的图象，我们先来试试描点法。我们选择 $x = -4, -3, -2, 0, 1, 2, 3, 4$ 这几个点，先计算出来相应的 y 值，然后把

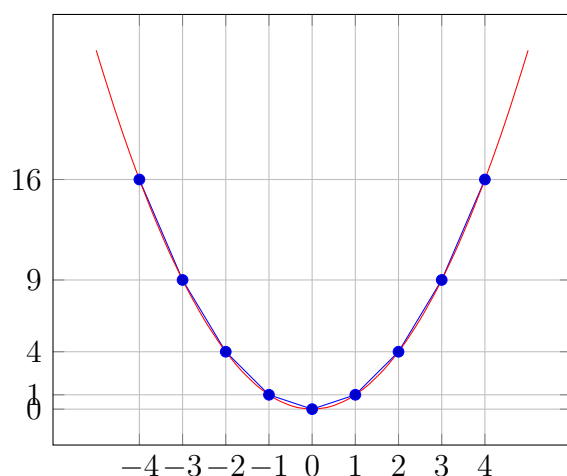


图 9.3: 一元二次函数 $y = x^2$ 的图像举例。我们选择一群 x 值，算出来相应的 y 值，画出来每个点 (x, y) ，然后连起来。为了画得更准确，可以选择更多的 x 值。

这些点画到 (x, y) -平面直角坐标系上，然后把这些点用折线（如果你对这个函数的样子已经有认知，也可以用尽可能光滑的曲线）把相邻的点连接起来连接起来，就能得到相应的函数的图象。在图 9.3 中我们看到连接起来以后的图和计算机做出来的更加准确的图——实际是上就是通过描更多的点来做出来的，这里暂时当作这个函数的准确的图象——还是有一定区别。这个时候，只要我们增加更多的点，更密集，我们总是可以更接近这个函数的准确的图象。

表 9.1: 一元二次函数值几个例子。为了画出来一元二次函数 $y = x^2$ 的图像，我们列出来几个这个函数的值。

x	-4	-3	-2	-1	0	1	2	3	4
x^2	16	9	4	1	0	1	4	9	16

通过描点法，我们发现，这是一个增函数，函数值随着 x 的增加而增加，并且增加的速度也在变大。我们也可以来证明一下这个结论：我们让 x

增加一个固定的量, 例如 1, 看看 y 的增加量⁵, 也就是

$$f(x+1) - f(x) = (x+1)^2 - x^2 = x^2 + 2x + 1 - x^2 = 2x + 1. \quad (9.1)$$

也就是说, 给 x 增加 1 个单位在 y 上的效果, 在 x 大的时候更大。注意, 在这里, 我们只关注第一象限, 也就是 $x > 0$ 的情况。

对于一元一次函数, 例如 $y = g(x) = x$ (因为在这里一次函数和二次函数需要放在一起比较, 为了区别, 把一次函数记作 g), 这个增加的速度不依赖于 x , 也就是

$$g(x+1) - g(x) = (x+1) - x = 1. \quad (9.2)$$

也就是说, 给 x 增加 1 个单位在 y 上的效果, 不依赖 x 的取值。

于是, 我们发现, 这个增加的速度要超过 $y = x$ 。我们已经知道, $y = x$ 是 45° 的直线, 直线的斜率也就是增加的速率是一个常数。那么 $y = x^2$ 的每一个点的斜率, 如果可以定义一个斜率的话, 要不断地变大, 而且要大于 $y = x$ 的斜率。也就是说, $y = x^2$ 的增加的速度, 也就是曲线往 y 轴接近的速度, 要远远大于 $y = x$ 曲线, 而且 x 越大这个差异越大。

有了这个发现, 再加上描点法, 我们就可以大概把这个函数图画出来了。

搞清楚了特殊的 $y = x^2$ 的一切, 我们再来看更一般的一元二次函数 $y = ax^2 + bx + c$ 。我们说更一般的一元二次函数 $y = ax^2 + bx + c$ 可以被转化为特殊的 $y = x^2$ 的情形, 也就是

例 9.1 (从 $y = ax^2 + bx + c$ 到 $y = x^2$). 把 $y = ax^2 + bx + c$ 被转化为接近 (这里的接近指的是, 将来可以通过平移和缩放等操作, 来从后者的函数图象得到前者) $y = x^2$ 的形式。

$$ax^2 + bx + c = a \left(x^2 + \frac{b}{a}x + \frac{c}{a} \right), \quad (9.3)$$

$$= a \left[\left(x + \frac{b}{2a} \right)^2 - \frac{b^2}{4a^2} + \frac{c}{a} \right], \quad (9.4)$$

$$= a \left(x + \frac{b}{2a} \right)^2 + \frac{4ac - b^2}{4a}. \quad (9.5)$$

⁵将来我们学习了微积分会发现, 这就是在计算导数, 或者说微分, 或者说曲线斜率。

于是,我们发现,从 $y = x^2$ 的图象开始,我们:第一、先移动一下 x 轴,使得 $x \rightarrow x + \frac{b}{2a}$,也就是把 x 轴平移 $-\frac{b}{2a}$,就得到了函数 $y = (x + \frac{b}{2a})^2$ 的图象;第二、再把得到的函数整体乘上一个 a ,就得到了 $y = a(x + \frac{b}{2a})^2$;第三,把前一步得到的图象整体上下移动也就是加上 $\frac{4ac-b^2}{4a}$,就得到了 $y = ax^2 + bx + c$ 的图象。

注意,由于 $y = x^2$ 有一个最小值,因此 $y = a(x + \frac{b}{2a})^2 + \frac{4ac-b^2}{4a}$ 也有一个最大值 ($a < 0$) 或者最小值 ($a > 0$)。这个最值点的坐标是 $(-\frac{b}{2a}, \frac{4ac-b^2}{4a})$ 。同时, $x = -\frac{b}{2a}$ 也是 $y = a(x + \frac{b}{2a})^2 + \frac{4ac-b^2}{4a}$ 的函数图象的对称轴。对于一个一元二次函数,对称轴 $x = -\frac{b}{2a}$, 开口方向和大小 (a), 最值点 $(-\frac{b}{2a}, \frac{4ac-b^2}{4a})$ 是几个最重要的信息。

类似地,对于求解方程 $y = ax^2 + bx + c$, 或者说 $ax^2 + bx + (c - y) = 0$ (这里我们可以通过重新定义 c 使得方程的右侧等于零), 由于 $y = ax^2 + bx + c$ 和 $y = x^2$ 可以通过平移和缩放联系起来而平移和缩放是可逆操作, 那么一旦我们会求解 $y = x^2$, 我们就可以把 $y = x^2$ 的解通过平移和缩放变成 $y = ax^2 + bx + c$ 的解。这里我们通过平移和缩放来求解, 未来, 见第十一章, 我们会专门来学习一元二次方程的求解。

具体来说, 第一、我们先求出来 $y = z_1 + \frac{4ac-b^2}{4a}$, 也就是 $z_1 = y - \frac{4ac-b^2}{4a}$; 第二, 我们再求出来 $z_1 = az_2$, 也就是 $z_2 = \frac{z_1}{a}$ ($a \neq 0$); 第三, 我们然后求出来 $z_2 = z_3^2$, 也就是 $z_3 = \pm \sqrt{z_2}$; 我们接着求出来 $z_3 = x + \frac{b}{2a}$, 也就是 $x = z_3 - \frac{b}{2a}$; 最后, 整合起来, 就是一元二次方程 $y = ax^2 + bx + c$ 的解,

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4a(c - y)}}{2a}. \quad (9.6)$$

通过重新定义 c ($(c - y) \rightarrow c$) 可以变成方程右侧 $y = 0$ 的标准形式, 其解为

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (9.7)$$

在实数范围内, 这个表达式有意义的条件是 $\Delta = b^2 - 4ac \geq 0$ 。如果这个条件不满足, 那么, 在实数范围内方程无解。

从函数图象的角度来说, 在实数范围内方程 $y_* = ax^2 + bx + c$ 无解意味着函数的最值点和给定的 y_* 值不匹配, 也就是, 如图 9.4 所示, 当 $a < 0$ 时, 所给定的 y_* 比函数 $y = ax^2 + bx + c$ 的最大值点的 y 坐标 $\frac{4ac-b^2}{4a}$ 还要大; 当 $a > 0$ 时, 所给定的 y_* 值比函数 $y = ax^2 + bx + c$ 的最小值点的 y

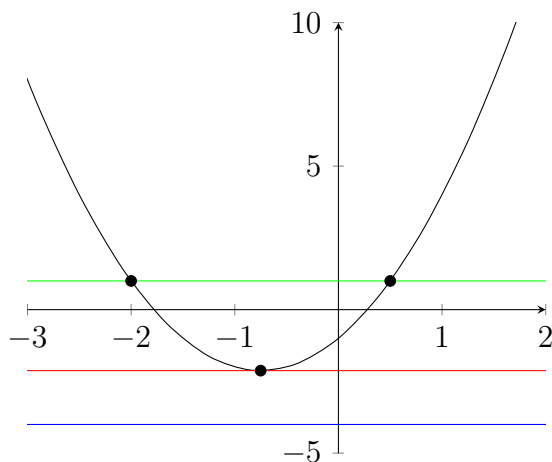


图 9.4: 一元二次函数和 $y = y_*$ 直线的交点。对于给定的一元二次函数，直线 $y = y_*$ 和这个函数的图象可能有两个、一个、或者没有交点。

坐标 $\frac{4ac-b^2}{4a}$ 还要小。因此，函数 $y = ax^2 + bx + c$ 和直线 $y = y_*$ 没有交点。回到一元二次方程的标准形式，相当于让这里的 y_* 等于零，或者说重新定义了 c 。

9.8 负一次方函数

$y = \frac{a}{x} + b = ax^{-1} + b$ 称为 x 的负一次方函数。由于通过重新命名 y 轴 ($y - b \rightarrow y$)，我们总是可以使得这个函数变形为 $y = ax^{-1}$ ，下面我们只研究 $y = ax^{-1}$ 。首先，我们注意到 a 的符号重要，大小也不重要，因为我们总是可以缩放 y ($y \rightarrow \frac{1}{|a|}y$) 来把方程变形为 $y = \frac{a}{|a|}x^{-1}$ 。所以，下面我们只研究 $y = \pm x^{-1}$ 。我们先来搞清楚 $y = x^{-1}$ ，然后 $y = -x^{-1}$ 相当于把 $y = x^{-1}$ 的 y 值做一个相反数。体现在函数图像上，相当于把函数图象做一个 y 方向上的翻转。

对于 $y = x^{-1}$ ，我们还是通过描点法来先画个图看看。当然，如果你记得小学的时候学习过反比例函数，就知道，这个函数有一个特点：在 $x > 0$ 的范围内， x 越大 y 越小，两者成反比。

在表9.2中，我们列出了一些负一次方函数值。在图 9.5 中，我们按照描点法做了负一次方函数的图象。注意，在接近 $x = 0$ 的地方， x 的取值要

表 9.2: 负一次方函数值几个例子。为了画出来负一次方函数 $y = x^{-1}$ 的图像，我们列出来几个这个函数的值。“NA”表示函数在这个点没有取值 ($\frac{1}{0}$)。

x	-2	-1.5	-1	-0.5	-0.4	-0.2	-0.1	0
x^{-1}	-0.5	$-\frac{2}{3}$	-1	-2	-2.5	-5	-10	NA
x	0.4	0.2	0.1	0.5	1	1.5	2	3
x^{-1}	10	5	2.5	2	1	$\frac{2}{3}$	0.5	$\frac{1}{3}$

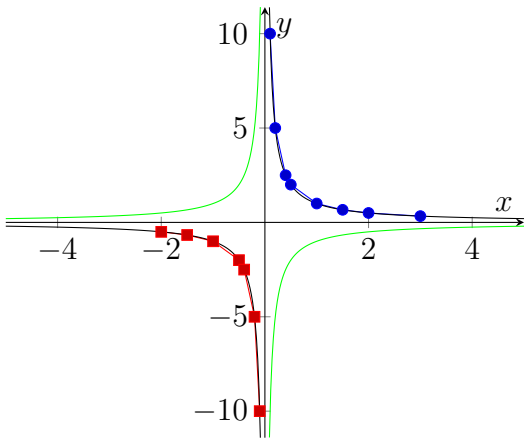


图 9.5: 负一次方函数 $y = x^{-1}$ 的图像。我们选择一群 x 值，算出来相应的 y 值，画出来每个点 (x, y) ，然后连起来。为了画得更准确，可以选择更多的 x 值。

更紧密，因为那个时候 y 变化的特别大。如果取值少一点就会不容易看到这个变化很大的趋势。此外，由于 $x = 0$ 这个点上，函数没有取值 ($\frac{1}{0}$ 没有定义)，我们只能避开这个点。于是，这个函数的 y 轴左右两侧不能相连。

有了函数 $y = x^{-1}$ 的图象，函数 $y = -x^{-1}$ 的图象就特别简单了，相应地沿着 x 轴上下翻转一下，也就是在图 9.5 的处于二四象限的绿色曲线。

9.9 为什么需要函数？

为什么我们会给某些表达式，或者说这些表达式相应的图象，一个自己的名字，也就是函数名称呢？甚至为什么会需要函数这个概念呢？

我们先来回答后者——为什么我们需要函数这个概念。其实，我们提

到过，数学是研究关系的，研究操作的。所谓操作就是一个东西进来以后经过这个操作变成另外一个东西。对于操作来说，其往往不仅作用于一个对象，而是一类对象。因此，操作具有一定程度上比其所操作的对象更加具有一般性。例如，你学会了转动，则线、三角形、四边形等都可以是被转动的对象。例如，你学会了炒、或者油炸、或者清蒸这个做菜的方式，则各种鱼、肉、菜等食材都可以是这些烹饪方式的作用对象。以学习做菜为例，当然，具体每一道菜都是要在具体材料上做的，往往食材确实是一个菜的核心，但是，当你学习做菜的时候你学习烹饪方式（以及烹饪方式和食材之间的关系、烹饪方式和出品的菜之间的关系），才是关键。学习烹饪方式的任务量也比学习每一样食材要小很多。当你来创作和制作菜品的时候，你要做的是从想要的菜品来反向思考，确定所用的食材和所用的烹饪方式。同样地，在数学中，我们要研究操作，研究关系，研究变换。因为数学的可以操作、变换的对象实在是太多了。甚至，我们会发现，我们还需要研究对操作的操作。映射就是对操作，以及对操作的操作，乃至对操作的操作的操作的抽象，因为映射的定义域可以是任何一个集合。具体到把数当作这样的定义域的时候，映射就成了函数。

有了一般的函数的概念以及这个概念的必要性，自然我们也就有了定义一些具体的函数的必要性。由于我们在解决现实问题或者在数学学科知识本身的发展中，经常会遇到某些表达式以及相应的图象，或者某些表达式以及相应的图象特别简单未来会成为更加复杂的对象的基础，我们就给了这些表达式和相应的图象自己的名字。这样当我们提起来这个名字的时候，所有人的理解都是一致的。

因此，为什么会有函数的概念，以及具体的函数的定义？其根本原因仍然是数学是思维的语言，是描述世界的语言。我们要给重复出现的模式、操作、关系、结构一个名字，这样含义明确，这样将来也可以用来构成更加复杂的数学结构。其次，数学研究操作、对操作的操作等等，而函数或者说其背后的映射是对它们的一般性的抽象。

9.10 推荐阅读材料

9.11 作业

习题 9.6 (画函数图). 用描点法画出来 $y = 4(x - 1) + 2 = 4x - 2$ 的函数图, 再用缩放和平移 $y = x$ 的方式来画图, 最后用计算机数学软件来和你画的图对比一下。

习题 9.7 (画函数图). 用描点法画出来 $y = 4(x - 1)^2 + 2 = 4x^2 - 8x + 6$ 的函数图, 再用缩放和平移 $y = x^2$ 的方式来画图, 最后用计算机数学软件来和你画的图对比一下。

习题 9.8 (画函数图求解方程). 画出来 $y = 4x - 2, y = 3$ 的函数图, 找到交点。为了更准确, 你可能得用计算机数学软件来画图和找点。或者, 如果你坚持手工操作, 你可以考虑用画图纸来帮助你话得更准确。

习题 9.9 (画函数图求解方程). 画出来 $y = 4x - 2, y = 3x + 1$ 的函数图, 找到交点。为了更准确, 你可能得用计算机数学软件来画图和找点。或者, 如果你坚持手工操作, 你可以考虑用画图纸来帮助你话得更准确。

习题 9.10 (画函数图求解方程). 画出来 $y = 4x^2 - 8x + 6, y = 3$ 的函数图, 找到交点。为了更准确, 你可能得用计算机数学软件来画图和找点。或者, 如果你坚持手工操作, 你可以考虑用画图纸来帮助你话得更准确。

习题 9.11 (画函数图求解方程). 画出来 $y = 4x^2 - 8x + 6, y = 3x^2 - 4x + 2$ 的函数图, 找到交点。为了更准确, 你可能得用计算机数学软件来画图和找点。或者, 如果你坚持手工操作, 你可以考虑用画图纸来帮助你话得更准确。

习题 9.12 (画函数图求解方程). 画出来 $y = 4x^2 - 8x + 6, y = 3x^2 - 4x + 3$ 的函数图, 找到交点。为了更准确, 你可能得用计算机数学软件来画图和找点。或者, 如果你坚持手工操作, 你可以考虑用画图纸来帮助你话得更准确。和上一题对比一下, 有什么不一样。

9.12 本章小结

在本章中，我们以映射的概念为基础，学习了函数的概念和对函数的不同角度的理解。为了研究函数，我们建立了平面直角坐标系。然后，我们研究了几个具体的函数——包括一元一次函数、一元二次函数、负一次方函数。我们从函数的表达式、图象这两个方面来研究一个函数。我们还学习了画函数图的方法——描点法。基于函数的图象，我们还研究了几个方程的几何意义，包含一元一次方程和一元二次方程。

在这个过程中，我们特别强调对函数的几个角度的理解，包含映射的角度、函数表达式的角度、函数图象的角度和关系的角度。在研究具体函数的时候，我们还多次使用了对坐标轴的平移和缩放来更好地理解函数之间的关系，从而可以实现用函数的一个简单形式来得到关于函数的一般形式的一些信息——例如从函数的简单形式的图象通过缩放平移可以得到函数的一般形式的图象。

在这个过程中，我们还特别注重代数式和函数图象——也就是这个代数式的几何含义——之间的联系。这个未来称为数形结合的东西，在未来会发挥更大的作用。更进一步，这是更具有一般性的“把看起来不同的数学分支学科融合”从而来促进理解和研究数学的一个例子。

第十章 二元一次方程和二元一次不等式

Mathematics is the art of reducing any problem to linear algebra.

数学就是一门把任何问题转化为线性代数的艺术。

– William Stein

在这一章中我们来学习二元一次方程和二元一次不等式分别是什么，以及如何求解，为什么可以这样求解，学会了它们是什么和怎么求解以后有什么用。

10.1 为什么要学习二元一次方程和二元一次不等式

二元一次方程就是等号的两边都只能出现二元一次多项式，也就是 $ax + by + c = dx + ey + f$ ，通过等式的性质可以变成其标准形式

$$ax + by + c = 0. \tag{10.1}$$

一个二元一次方程中有两个未知数， x, y 。当固定 y 的时候，它成了一个二元一次方程 $ax + (by + c) = 0$ ，而这个方程有确定的解 $x = -\frac{by+c}{a}$ 。因此，任何一个 $(-\frac{by+c}{a}, y)$ 都是原始的二元一次方程 $ax + by + c = 0$ 的解。这样的解有无穷多个。如果我们想要这样的方程最后得到一个只有唯一解或者

有限解, 则我们可以考虑多个这样的方程的组合。例如,

$$\begin{cases} ax + by + c = 0, & (10.2a) \\ dx + ey + f = 0. & (10.2b) \end{cases}$$

这个时候除了从第一式得到 $x = -\frac{by+c}{a}$, 我们还会从第二式得到 $x = -\frac{ey+f}{d}$ 。如果一个解要同时满足这两个方程, 则表示 (x, y) 的值都得相同, 也就是

$$-\frac{by+c}{a} = x = -\frac{ey+f}{d}. \quad (10.3)$$

这是一个关于未知数 y 的一元一次方程。我们之前学习过, 一般情况下, 这个方程只有一个解。也就是说, 我们需要两个二元一次方程组成的方程组来得到一个确定解。

更简单的理解为什么我们需要一个两个二元一次方程构成的方程组的方式是利用二元一次方程的几何含义。我们在小学阶段就学习过 $ax + by + c = 0$ 表示一条在 (x, y) 二维平面上的直线。后面在平面几何部分我们也会再一次学习到这一点 (见 13.1 节)。两条直线相交可以确定一个点。因此, 我们需要同时满足两个二元一次方程来得到一个解。

也就是说, 未来我们可能会遇到两条直线相交的问题, 或者可以转化成两条直线相交的问题的问题, 于是我们就要学习二元一次方程和二元一次方程的求解。

至于二元一次不等式的概念和求解, 也更容易从几何来解释。 $ax + by + c > 0$ (为简单计, 假设 $b > 0$), 或者说 $y > -\frac{a}{b}x - \frac{c}{b}$ 指的是直线 $y = -\frac{a}{b}x - \frac{c}{b}$ 的上方的在 (x, y) 平面上的区域。因此, 无论多少个二元一次不等式, 其通过这些对应的直线的相交的关系, 都会确定出来一个满足所有不等式的一个区域。当然, 这个区域也可能是空集, 也就是没有任何公共区域同时满足所有这些不等式。因此, 求解多个二元一次不等式构成的不等式组就相当于在确定满足所有要求的区域。

那什么时候会出现这样的问题场景呢? 未来在大学阶段我们会学习到一个叫做线性规划的问题, 就是在某些约束下求某一个目标函数的最大值。如果所有的约束都体现为二元一次不等式那就对应了求解多个二元一次不等式的问题。而在大量的问题中, 二元一次不等式经常会出现。例如, 我们用固定的预算总额 (m) 去购买 X 和 Y 两种物品, 其价格确定为 P_A, P_B ,

则购买的数量 (x 和 y) 必然满足以下二元一次不等式

$$P_A x + P_B y \leq m. \quad (10.4)$$

下面，我们来学习如何求解二元一次方程和二元一次不等式。

10.2 换元法：化二元一次方程为一元一次方程

其实上一节中已经包含了一种求解二元一次方程的方法，先假设其中一个变量——例如 y ——的值已知，用这个变量 y 来表示另一个变量 x 。然后，分别从两个二元一次方程得到两个用 y 的表达的等于 x 的表达式。接着，让这两个表达式相等（因为所求出来的解 (x, y) 必须同时满足这两个二元一次方程），就得到了一个一元一次方程。最后求解这个一元一次方程，就能得到同时满足两个方程的 y^* 。再代入到用 y 表达的 x 里面，就得到了同时满足两个方程的 x^* 。合起来，相当于我们用了三次的一元一次方程求解方法，用了一次等式的性质（等号的传递性，或者叫做等量替换），我们就能求解出来一个由两个二元一次方程构成的方程组。于是，因为一元一次方程的求解方法是完全由运算律和等式的性质导出的，那么，这里的二元一次方程组的解法也是完全由运算律和等式的性质导出的。

我们来求解一般的二元一次方程组算式 (10.2)。假设 y 已知，则把算式 (10.2a) 看作一元一次方程，得到

$$-\frac{by + c}{a} = x, \quad (10.5)$$

我们要找的是同时满足算式 (10.2a) 和算式 (10.2b) 的这样的 (x, y) ，因此算式 (10.5) 中用 y 表示的 x 必然也满足算式 (10.2b)。于是，把算式 (10.2b) x 换成来自算式 (10.2a) 的用 y 表示的 x ——因此这个方法被称为换元法，得到

$$d \left(-\frac{by + c}{a} \right) + ey + f = 0. \quad (10.6)$$

求解这个一元一次方程，得到

$$y = \frac{af - cd}{bd - ae} \quad (10.7)$$

代入到算式 (10.5) 得到,

$$x = \frac{ce - bf}{bd - ae}. \quad (10.8)$$

其中为了简化结果我们再计算中还用到了因式分解。合起来, 我们把二元一次方程组算式 (10.2) 记为

$$(x, y) = \left(\frac{ce - bf}{bd - ae}, \frac{af - cd}{bd - ae} \right). \quad (10.9)$$

如果多于两个方程, 则把求出来的解, 代入到其他方程中, 如果这个解也满足其他方程, 则这就是所要找的方程组的解。否则, 这个方程无解。

10.3 消元法: 等式的计算

二元一次方程组的另一种求解方法是消元法——通过对等式的计算来去掉其中一个变量, 使得方程称为一元一次方程。

我们把算式 (10.2a) 左右两边乘以 d , 把算式 (10.2b) 左右两边乘以 a , 得到

$$\begin{cases} adx + bdy + cd = 0, & (10.10a) \\ adx + aey + af = 0. & (10.10b) \end{cases}$$

因为算式 (10.10a) 的左侧等于零, 算式 (10.10b) 的左侧也等于零, 显然根据 $0 - 0 = 0$ (恒等式), 其左侧相减也等于零 (等量替换), 也就是

$$\begin{aligned} (adx + bdy + cd) - (adx + aey + af) &= 0, \\ \Rightarrow (bd - ae)y + (cd - af) &= 0, \\ \Rightarrow y &= \frac{af - cd}{bd - ae}. \end{aligned} \quad (10.11)$$

当然, 我们发现这里得到的 y 和上面的换元法得到结果相同。得到 y 之后代回到算式 (10.2a) 或者算式 (10.2b), 就能得到另一个一元一次方程, 从而求出来 x 。

或者我们重新用一遍上面的计算方法, 这次让变量 y 前面的系数为零, 也就是消掉变量 y ——因此称为消元法: 我们把算式 (10.2a) 左右两边乘以

e , 把算式 (10.2b) 左右两边乘以 b , 得到

$$\begin{cases} aex + bey + ce = 0, & (10.12a) \\ bdx + bay + bf = 0. & (10.12b) \end{cases}$$

因为算式 (10.12a) 的左侧等于零, 算式 (10.12b) 的左侧也等于零, 显然根据 $0 - 0 = 0$ (恒等式), 其左侧相减也等于零 (等量替换), 也就是

$$\begin{aligned} (aex + bey + ce) - (bdx + aby + bf) &= 0, \\ \Rightarrow (ae - bd)x + (ce - bf) &= 0, \\ \Rightarrow x &= \frac{ce - bf}{bd - ae}. \end{aligned} \quad (10.13)$$

我们再次发现这里得到的 x 和上面的换元法得到结果相同。

10.4 二元一次方程的几何含义

我们前面已经提到, 一个二元一次方程的几何含义是一条直线, 因此两个二元一次方程都要满足的几何含义是这两个方程所对应的两条的交点。未来我们会学习到, 当两条直线平行的时候没有交点, 当两条直线重叠的时候, 有无穷多个交点, 其他情况下两条直线只有一个交点。

从我们得到的解来看

$$\begin{cases} x = \frac{ce - bf}{bd - ae}, & (10.14a) \\ y = \frac{af - cd}{bd - ae}, & (10.14b) \end{cases}$$

当 $bd - ae \neq 0$ (也就是 $\frac{a}{b} \neq \frac{d}{e}$) 的时候, (x, y) 唯一; 当 $bd - ae = 0$ (也就是 $\frac{a}{b} = \frac{d}{e}$) 但是 $ce \neq bf$ (也就是 $\frac{c}{b} \neq \frac{f}{e}$) 的时候, (x, y) 中会出现 $\frac{1}{0}$, 没有意义; 当 $bd - ae = 0$ (也就是 $\frac{a}{b} = \frac{d}{e}$) 同时 $ce = bf$ (也就是 $\frac{c}{b} = \frac{f}{e}$) 的时候, (x, y) 中会出现 $\frac{0}{0}$, 现在我们还不知道这意味着什么¹。联系到二元一次方程组的几何含义, 第一种情况代表两直线相交——交点唯一, 第二种情况代表两直线平行——没有交点, 而剩下的只有第三种情况了, 因此第三种情况代表两直线重合——两直线有无穷多个共同点。

¹未来我们会知道这代表可以出现任何一个数, 方程有无穷多个解。

10.5 二元一次方程的应用

10.6 二元一次不等式组的求解和几何含义

基于二元一次不等式的几何含义来求解二元一次不等式组是非常方便的。就像再求解一元一次不等式组的时候一样，不等式组所包含的不等式的数量可以是任意多个的。其实，求解一元一次不等式组的时候我们就用几何方法——在数轴上画区间图的方式——来求解了。现在，我们来在平面上画二元一次不等式组的图来求解二元一次不等式组。

我们举例来说明这个求解过程。

例 10.1 (求解二元一次不等式组). 求解下面的二元一次不等式组，

$$\begin{cases} 2x + 3y < 0, & (10.15a) \\ 4x + 9y + 2 > 0, & (10.15b) \\ 5x - 6y + 7 \geq 0. & (10.15c) \end{cases}$$

这个不等式组中的每一个不等式的等式部分都是一条直线，不等号代表了以这条直线为边界的这条直线在 y 方向往上或者往下的区域。如图 10.1 所示， $2x + 3y < 0$ 等价于 $y < -\frac{2}{3}x$ ，因此是直线 $2x + 3y = 0$ 的 y 方向向下的部分，用一种颜色的阴影表示； $4x + 9y + 2 > 0$ 等价于 $y > -\frac{4}{9}x - \frac{2}{9}$ ，因此是相应的直线的 y 方向向上的部分，用一种颜色的阴影表示； $5x - 6y + 7 \geq 0$ 等价于 $y \leq \frac{5}{6}x - \frac{7}{6}$ ，因此是相应的直线的 y 方向向下的部分，用一种颜色的阴影表示。注意，最后那条曲线，由于是 $\leq$ 号，包含了边界上的直线自己。最后三种阴影部分重叠的就是满足不等式组要求的解的集合。

作为一般的方法，我们把不等式组中所有的直线都画到 (x, y) 平面上，然后标记出来满足每一个不等式的区域，找出来重叠的区域，就是二元一次不等式组的解的集合。

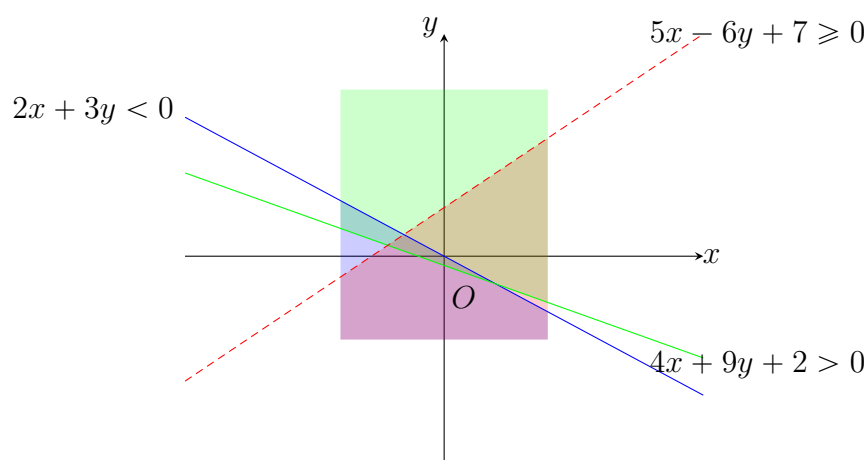


图 10.1: 二元一次不等式组的画图求解。不等式组中所有的直线都画到 (x, y) 平面上, 然后标记出来满足每一个不等式的区域, 找出来重叠的区域。

10.7 二元一次不等式组的应用

10.8 推荐阅读材料

10.9 作业

10.10 本章小结

在本章中, 我们学习了什么是二元一次方程组和二元一次不等式组, 以及它们的求解。我们看到, 我们的求解方法不过就是之前的一元一次方程和一元一次不等式组的拓展, 而它们基于的是运算律和等式的性质、不等式的性质, 而它们进一步基于的是数的含义、数在数轴上的表示以及运算(加、乘、不等号运算)在数轴上的表示。因此, 数学学科概念知识是具有系统性的: 从极少数基本定义(未来会加上一些基本的假设——在数学里面称为公理)开始可以得到所有其他的数学学科概念知识。

为了展示二元一次方程组和二元一次不等式组的含义, 也为了求解二元一次不等式组, 在本章中我们还把方程、不等式和它们的几何意义相结合。这个数形结合未来我们会再次遇到。把不同的数学分支领域结合起来

往往可以促进我们提出、理解和解决问题。

第十一章 一元二次方程和一元二次不等式

在这一章中，我们将学习一元二次方程和一元二次不等式的求解，以及思考其中的求解的方式的数学基础是什么，以及思考能不能把求解的方式或者获得求解方式背后的思路推广到更高阶的方程和不等式。

11.1 为什么要学习一元二次方程和一元二次不等式

一元二次方程指的是等号的左右两次都只能出现一元二次多项式的方程，也就是 $ax^2 + bx + c = a'x^2 + b'x + c'$ 。通过运用等式的性质和运算律，我们可以得到一元二次方程的标准形式，也就是

$$ax^2 + bx + c = 0. \quad (11.1)$$

一元二次不等式也类似，就是不等号的两次都出现一元二次多项式的不等式。经过运用不等式的性质和运算律，我们也总是可以把一元二次不等式写成标准形式，

$$ax^2 + bx + c > 0. \quad (11.2)$$

其中的不等号 $>$ 可以替换成其他不等号，包含 $\neq, <, \geq, \leq$ 。

同时，由于等式具有“等式两边同时乘以 -1 等号仍然成立”的性质，不等式具有“不等式的两边同时乘以 -1 改变不等号相反”的性质，我们总

是可以保证一元二次方程和一元二次不等式的标准形式中的 $a > 0$ 。注意，当 $a = 0$ 的时候，这个不等式退化为我们已经学习过的一元一次不等式。因此，下面我们只处理 $a > 0$ 的情况。

那人们为什么需要专门研究，我们现在为什么要专门来学习，一元二次方程和一元二次不等式呢？首先，有些实际问题场景自然地会变成一元二次方程和一元二次不等式。例如，假设我们要建造一个面积为 c 的正方形屋子，我们自然要问边长为多少合适？你看，自然地就出现了 $x^2 = c$ 。例如，如果每一期的存款利率是 r ，则数量为 S 的存款在第一年结束以后的总额为 $(1+r)S$ ，而第二年为 $(1+r)(1+r)S = (1+r)^2 S$ 。如果我们的目标是两年以后总额达到 T ，自然地出现了如下的一元二次方程 $(1+r)^2 S = T$ 。

例如，我们有一匹特别能跑的马，每天能跑 L 米的距离。有人说，你只要跑一圈，也就是跑一条首尾衔接的封闭曲线，圈起来的地都是你的。你怎么跑？¹我们可以尝试先来跑一个长方形——好看，边界清晰好管理。设这个长方形的长为 a ，则宽为 $(\frac{L}{2} - a)$ ，于是其面积为 $a(\frac{L}{2} - a) = -a^2 + \frac{L}{2}a$ 。我们希望这个面积最大。按照之前一元二次函数的性质，这个函数由最大值，在其对称轴上，也就是 $a = \frac{L}{4}$ 的时候。这个时候， $a = \frac{L}{2} = b$ 正好是一个正方形。因此，我们得到了如果想让这个长方形的面积最大，则最好圈成正方形。反过来，如果我们要圈出来的面积也是确定的，则我们有方程，

$$a\left(\frac{L}{2} - a\right) = S, \quad (11.3)$$

也就是

$$a^2 - \frac{L}{2}a + S = 0. \quad (11.4)$$

这是一个一元二次方程。通过前面的最大值的分析，我们还知道，如果这个 S 比 $a(\frac{L}{2} - a)$ 的最大值还要大，则无解。这就出现了一元二次不等式 $S > a(\frac{L}{2} - a)$ 。

习题 11.1 (正多边形圈地法). 验证一下：跑马距离一定的时候，正六边形圈地法得到的土地面积比正四边形要大。你猜是不是正多边形的边数越多，面积越大？尝试几个具体的多边形。然后试试能不能一般地证明这个结论。如果这个结论正确，那最好的圈地方式是什么样的，如果仅仅考虑面积最大，不考虑管理方便等因素的话。

¹据说这个问题实际上在人类历史上就出现过，叫做“跑马圈地”。

其次,是不是从之前的求解一元一次方程和一元一次不等式,以及现在的求解一元二次方程和一元二次不等式,可以提炼出来一般性的方法,未来可以迁移到求解多元多次方程和多元多次不等式中去?或者反过来,看到这些提炼出来的方法将来都不可能迁移过去,于是多元多次方程和多元多次不等式没有一般求解方式?比如说,对于一元一次方程我们是运用运算律和等式的性质来求解的,得到解可以表达为方程的系数的多项式。那么,未来对于多元多次方程是不是也可以找到这样的用系数的多项式所表达的解呢,其背后也许需要运用运算律和等式的性质呢?如果能,这个表达式是什么,怎么得到的;如果不能,为什么?

再次,有可能是将来更加复杂的方程或者其他数学问题要基于一元二次方程和一元二次不等式来构建和解决。这个等遇到的时候再说。

11.2 一元二次方程的几何含义

根据之前对一元二次函数的图象的研究(见 9.7 节),我们说一元二次方程算式 (11.1) 的解的含义是曲线 $y = ax^2 + bx + c$ 和 x 轴——也就是 $y = 0$ 的直线——的交点。

我们约定了 $a > 0$, 则 $y = ax^2 + bx + c$ 的图象开口像上, 当 $y = ax^2 + bx + c$ 最小值小于零的时候, $y = ax^2 + bx + c$ 的图象和 x 轴有两个交点; 当 $y = ax^2 + bx + c$ 最小值等于零的时候, $y = ax^2 + bx + c$ 的图象和 x 轴有唯一的交点; 当 $y = ax^2 + bx + c$ 最小值大于零的时候, $y = ax^2 + bx + c$ 的图象和 x 轴没有交点。于是, 我们只需求出来 $y = ax^2 + bx + c$ 最小值 ($y_{\min} = \frac{4ac-b^2}{4a}$, 见 9.7 节), 就可以判断方程是否有解, 有一个还是两个解。也就是当 $\Delta = b^2 - 4ac > 0$ 的时候, 有两个解; 当 $\Delta = 0$ 的时候, 有一个解; 当 $\Delta < 0$ 的时候无解。

在 9.7 节中, 通过 $y = x^2$ 和 $y = ax^2 + bx + c$ 之间的对坐标轴平移和缩放的关系, 我们还从 $x^2 + c = 0$ 的解为 $x = \pm\sqrt{-c}$ 得到了 $y = ax^2 + bx + c$ 的解为,

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (11.5)$$

现在, 我们来展示另一种不从函数出发的求解方式——配方法求解一元二次方程。

11.3 配方法求解一元二次方程

我们明确能够求解的一元二次方程是 $x^2 + c = 0$, 当 $c \leq 0$ 的时候, $x_{1,2} = \pm \sqrt{-c}$. 我们能不能把一般性的一元二次方程变形为这个简单形式来求解呢? 我们试试。

$$ax^2 + bx + c = a \left(x^2 + \frac{b}{a}x + \frac{c}{a} \right), \quad (11.6)$$

$$= a \left(x^2 + 2\frac{b}{2a}x + \left(\frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 + \frac{c}{a} \right), \quad (11.7)$$

$$= a \left(x^2 + \frac{b}{2a} \right)^2 - a \left(\frac{b}{2a} \right)^2 + c, \quad (11.8)$$

$$= a \left(x^2 + \frac{b}{2a} \right)^2 - \frac{b^2 - 4ac}{4a}. \quad (11.9)$$

也就是说,

$$ax^2 + bx + c = 0 \Rightarrow a \left(x^2 + \frac{b}{2a} \right)^2 = \frac{b^2 - 4ac}{4a}. \quad (11.10)$$

而 $a \left(x^2 + \frac{b}{2a} \right)^2 = \frac{b^2 - 4ac}{4a}$ 已经接近 $x^2 + c = 0$ 的形式, 因此其解为

$$x_{1,2}^2 + \frac{b}{2a} = \sqrt{\frac{b^2 - 4ac}{4a^2}}, \quad (11.11)$$

也就是

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (11.12)$$

这被称为一元二次方程求根公式。其中的 $\Delta = b^2 - 4ac$ 称作根的判别式。当 $\Delta > 0$ 的时候, 方程有两个实数解; 当 $\Delta = 0$ 的时候, 方程有一个实数解; 当 $\Delta < 0$ 的时候, 在实数范围内 $\sqrt{\Delta}$ 不能算, 方程没有实数解。

顺便, 尽管我们之前为了简单假设了 $a > 0$, 这个公式也适用于 $a < 0$ 的情况。

习题 11.2 ($a < 0$ 的时候的一元二次方程的解的公式). 证明算式 (11.12) 适用于也适用于 $a < 0$ 的情况。你可以完全从头推导一遍, 也可以注意算式 (11.12) 中的 $\pm$ 符号和对 $\sqrt{\frac{1}{4a^2}}$ 的计算结果的讨论。

例 11.1 (求解一元二次方程). 求解 $x^2 - 2x - 3 = 0$ 。

第一步, 把二次项系数整成大于零的标准形式: 已经完成。实际上, 这一步也可以不做, 因为我们也证明了 $a < 0$ 的时候, 求根公式也能用。第二步, 检查 $x^2 - 2x - 3 = 0$ 有几个解: $\Delta = b^2 - 4ac = 4 + 12 = 16 > 0$ 。因此, 有两个解。分别用求根公式求出, 得到 $x_1 = -1, x_2 = 3$ 。

例 11.2 (求解一元二次方程). 求解 $-x^2 - 3x - 2 = 0$ 。

第一步, 把二次项系数整成大于零的标准形式: $x^2 + 3x + 2 = 0$ 。第二步, 检查 $x^2 + 3x + 2 = 0$ 有几个解: $\Delta = b^2 - 4ac = 9 - 8 = 1 > 0$ 。因此, 有两个解。分别用求根公式求出, 得到 $x_1 = -2, x_2 = -1$ 。实际上, 我们也可以跳过第一步, 甚至跳过检查判别式, 直接用求根公式来求解。未来一旦出现根号下是负数, 自然得到没有实数解。

这个求解过程所运用的是一元二次方程的最简单的形式 $y^2 + c = 0$ 的解已知, 然后通过加减乘除乘方等运算——实际上是平移和缩放坐标轴——把一般的一元二次方程转化为最简单的形式。这个方法具有一般性吗? 比如说, 将来我们可以把一元三次方程转化为其最简单的形式 $y^3 + d = 0$ 来求解吗? 这个转化的过程还是只需要平移和缩放坐标轴吗? 更高阶的方程呢。可以类似求解吗? 非常遗憾, 后来人们发现这条路走不通。

11.4 韦达定理和因式分解法求解一元二次方程

根据因式分解和解方程的关系的前一半——任何一个 $f(x) = 0$ 的解 $x = x^*$ 一定会使得多项式 $f(x)$ 可以被表示为 $f(x) = (x - x_*)g(x)$ 形式, 我们得到

$$ax^2 + bx + c = a(x - x_1)(x - x_2), \quad (11.13)$$

$$= ax^2 - a(x_1 + x_2) + ax_1x_2. \quad (11.14)$$

因此

$$x_1 + x_2 = -\frac{b}{a}, x_1x_2 = \frac{c}{a}. \quad (11.15)$$

这个一元二次方程的两根之和和两根之积和方程系数的关系被称为韦达定理。

同时, 因式分解和解方程的关系的后一半——任何一个能够写成 $0 = f(x) = (x - x_*)g(x)$ 形式的多项式 (要求 $f(x), g(x)$ 都是多项式), 则 $x = x_*$ 一定是 $f(x) = 0$ 的解, 我们得到只要把一元二次方程做因式分解, 得到

$$ax^2 + bx + c = a(x - x_1)(x - x_2), \quad (11.16)$$

则这个一元二次方程的解必为上面的 $x = x_1$ 和 $x = x_2$ 。

例 11.3 (求解一元二次方程). 求解 $x^2 - 2x - 3 = 0$ 。

对 $x^2 - 2x - 3$ 运用十字相乘法可以将其因式分解为 $(x - 3)(x + 1)$ 。于是, 原始方程变形为 $(x - 3)(x + 1) = 0$ 。其解为 $x_1 = -1, x_2 = -3$ 。

例 11.4 (求解一元二次方程). 求解 $-x^2 - 3x - 2 = 0$ 。

对 $-x^2 - 3x - 2$ 运用十字相乘法可以将其因式分解为 $-(x + 1)(x + 2)$ 。于是, 原始方程变形为 $-(x + 1)(x + 2) = 0$ 。其解为 $x_1 = -2, x_2 = -1$ 。

这个求解过程所运用的是一元二次多项式的因式分解。如果一元二次多项式总是可以被因式分解, 因式分解也存在着一般性的通用的方法, 则因式分解就可以把一个二阶方程变成两个一阶方程来求解。这样的求解方程的方法可以推广到更高阶的方程吗?

一个高阶多项式总是可以被因式分解为低级多项式吗, 因式分解具有一般性的通用的方法吗? 如果是, 则这个方法可以推广到更高阶方程的求解上, 否则不能。非常遗憾, 后来人们发现这条路也走不通。

11.5 关于三阶和更高阶的多项式方程

既然因式分解没有通用方法, 每一阶的多项式也不一定能够通过坐标平移和缩放变成其最一般的形式 $y = x^n$, 那以上两条路就行不通了。

那么, 难道三次方程以及更高阶的多项式方程就没有一般性的求解方法和求出来的用方程系数表达的解了吗? 不是的。三次方程其实有这样的解的表达式。这又是怎么做的呢? 大概来说, 数学家们, 尤其是Girolamo Cardano (吉罗拉莫·卡尔达诺) 和Niccolò Tartaglia (尼科洛·塔尔塔利亚), 发现, 通过引入一个神奇的变换可以把三次方程变成一个新的变量表示的二次方程——其中用到了韦达定理, 然后先求解这个新变量的二次方

程，在把变量变回去就得到了三次方程的解的形式。那这个方法可以推广到更高阶的方程吗？关键是那个“神奇变换”不再是线性的！有可能需要对每一阶的方程来发明一个这样的变换。如果是这样，能找出来也行。关键是肯定有这样的变换可以把高阶方程变成更低阶的方程吗？

非常神奇的是Lodovico Ferrari（洛多维科·费拉里）也给四阶方程找到了这样的一个变换，可以使得其成为一个三阶方程，从而利用之前的三阶方程的一般解法来求解，得到用系数表达的解的一般公式。

那五阶方程呢？能找到这样的变换吗？如果每一阶方程的求解都依赖于这样的巧妙的神奇的变换，那么，数学研究就成了非常具有偶然性的技术性的劳动了。非常幸运地，也非常不幸地，数学家Évariste Galois（埃瓦里斯特·伽罗瓦）彻底改变了这个多项式方程求解的领域。他证明了五阶及以上的方程不存在着用系数的加减乘除加上平方开方等运算表述的通用的解的形式。那他到底是怎么来证明这个的？这个结果对数学有什么意义？

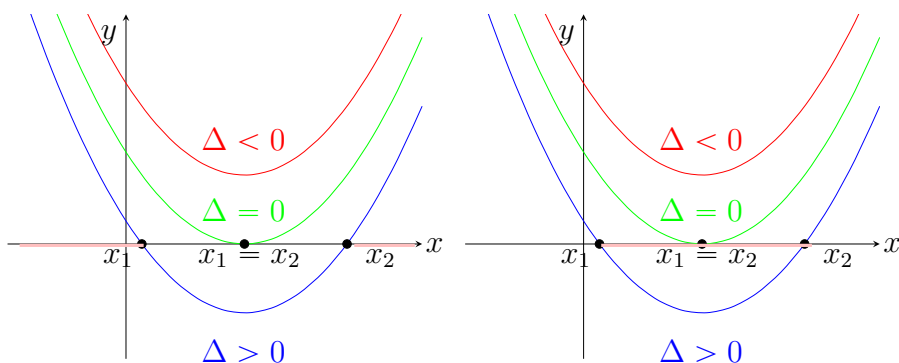
第一，未来——大学或者研究生阶段——我们会学习到他用群这个新的数学工具也是为了所有代数的基础，通过证明对方程做加减乘除乘方开方形式的运算的求解必然满足某一个性质，而五阶及以上的方程不可能都满足这个性质来得到了这个结论。因此，从数学学科概念知识上是一个开创。第二，更加重要的是，这个工作使得代数这个学科从尝试找巧妙的变换来求解方程中解脱出来了，有机会真正发展为代数了，而不是算术了。

这是值得每个人去体会一下的对数学发展的巨大的创新，巨大的突破。

11.6 一元二次方程的应用

11.7 一元二次不等式的几何含义和求解

有了一元二次函数 $y = ax^2 + bx + c$ 的图象，如图 11.1 所示，一元二次不等式 $ax^2 + bx + c > 0$ （约定 $a > 0$ ，其中的大于号可以替换为其他不等号）的解的几何意义就很简单了。 $y = ax^2 + bx + c$ 开后向上，如果 $ax^2 + bx + c = 0$ 有两个解，则其最小值小于零，在这两个解的左右两侧（ $x < x_1$ 和 $x > x_2$ ）， $y = ax^2 + bx + c > 0$ ，其他地方 $y = ax^2 + bx + c \leq 0$ ；如果 $ax^2 + bx + c = 0$ 有一个解，则其最小值等于零，在这个解的左右



$$(a) \quad ax^2 + bx + c > 0$$

$$(b) \quad ax^2 + bx + c < 0$$

图 11.1: 一元二次不等式的解。 x 轴上的粉色粗线显示一元二次不等式 $ax^2 + bx + c > 0$ (a) 和 $ax^2 + bx + c < 0$ (b) 的解的集合。对于不等号里面包含等于零的情况, 相应地再把 x_1, x_2 加上就行。

两侧 (不包含这个解 $x \neq (x_1 = x_2)$), $y = ax^2 + bx + c > 0$, 其他地方 $y = ax^2 + bx + c \leq 0$; 如果 $ax^2 + bx + c = 0$ 没有解, 则其最小值大于零, x 取任何值 ($x \in (-\infty, \infty)$), $y = ax^2 + bx + c > 0$ 。

对于一元二次不等式 $ax^2 + bx + c < 0$, 如果 $ax^2 + bx + c = 0$ 有两个解, 则其最小值小于零, 在这两个解的内侧 ($x_1 < x < x_2$), $y = ax^2 + bx + c < 0$, 其他地方 $y = ax^2 + bx + c \geq 0$; 如果 $ax^2 + bx + c = 0$ 有一个解, 则其最小值等于零, 在整个空间 $y = ax^2 + bx + c \geq 0$, 也就是没有满足不等式的解; 如果 $ax^2 + bx + c = 0$ 没有解, 则其最小值大于零, x 取任何值都有 $y = ax^2 + bx + c > 0$, 因此没有满足不等式的解。

对于不等号里面包含等于零的情况, 相应地再把一元二次方程 $ax^2 + bx + c = 0$ 的解 x_1, x_2 加上就行。如果出现多个一元二次不等式构成的一元二次不等式组, 只需要把每一个一元二次不等式的解集求出来, 然后, 做这些集合的交集, 就行了。我们来展示一个例子。

例 11.5 (求解一元二次不等式). 求 $x^2 - 2x - 3 \geq 0$ 的解。

第一步, 把二次项系数整成大于零的标准形式: 已经完成。第二步, 检查 $x^2 - 2x - 3 = 0$ 有几个解: $\Delta = b^2 - 4ac = 4 + 12 = 16 > 0$ 。因此, 有两个解。分别求出, 得到 $x_1 = -1, x_2 = 3$ 。注意, 按照约定 $x_1 \leq x_2$ 。如

图 11.1 所示, 不等号为大于等于号, 则不等式的解集为 $x \leq x_1 = -1$ 和 $x \geq x_2 = 3$, 也就是 $x \in (-\infty, -1] \cup [3, \infty)$ 。

例 11.6 (求解一元二次不等式). 求 $-x^2 - 3x - 2 \geq 0$ 的解。

第一步, 把二次项系数整成大于零的标准形式: $x^2 + 3x + 2 \leq 0$ 。第二步, 检查 $x^2 + 3x + 2 \leq 0$ 有几个解: $\Delta = b^2 - 4ac = 9 - 8 = 1 > 0$ 。因此, 有两个解。分别求出, 得到 $x_1 = -2, x_2 = -1$ 。注意, 按照约定 $x_1 \leq x_2$ 。如图 11.1 所示, 不等号为小于等于号, 则不等式的解集为 $x_1 = -2 \leq x \leq x_2 = -1$, 也就是 $x \in [-2, -1]$ 。

11.8 一元二次不等式组的应用

11.9 推荐阅读材料

11.10 作业

习题 11.3 (求解一元二次方程和一元二次不等式). 用求根公式、配方法、因式分解、计算机数学软件来求解 $x^2 - 5x - 6 = 0, x^2 - 5x + 6 = 0, -x^2 + 5x + 6 \geq 0, -x^2 + 5x - 6 \leq 0, -x^2 + 5x - 6 > 0$ 的解。用计算机数学软件的时候, 要结合相应的一元二次函数的图象。

11.11 本章和初中代数部分小结

在本章中, 我们学习了一元二次方程和一元二次不等式的求解。我们发现, 后者的求解依赖于前者的解。除了一元二次方程求根公式, 我们还得到了韦达定理。在这个过程中, 我们多次运用了一元二次函数的图象。而对于一个意愿二次函数画出来其图象, 强烈依赖于完全平方公式和配方法, 以及对坐标轴的移动和缩放。这些都是之前已经学习过的数学学科概念知识。因此, 我们再一次看到数学知识的系统性——这一章新学习的知识本质上不过仍然就是数和运算的含义, 以及基于这个含义的运算律, 等式的性质和不等式的性质。

更进一步，我们还思考了是否可以将这样的求解方法，或者获得求解方法背后的方法推广到更高阶的方程。我们发现，如果没有Galois的工作对五阶以及以上方程的代数解法的否定，数学将成为一个高度技巧性和高度偶然性的研究领域，但是，Galois的工作以及其所发明和使用的群，使得代数完全从求各阶方程的解这个问题中脱离出来，去关注更大的问题，以及为未来所有的代数结构的发展提供了学科概念知识上的基础。为了解决某个问题而提出新的数学结构，这是一个巨大的创造。同时，竟然这个新的数学结构对于整个数学的发展还是如此的重要。这是一个非凡的创造。作为一个数学创造的欣赏者，甚至未来的数学创造者，你不能错过对Galois的工作复现和赏析。

非常遗憾，我这个作者还不能用你们现阶段的语言能懂的方式来帮助你们完成这个复现和赏析。当你们学了群论，希望咱们还有机会再来试试。

一元二次方程和一元二次不等式的求解本身其实一点也不重要。你只需要知道什么是一元二次方程和一元二次不等式，通常在什么场合出现，求出来的解的含义是什么，求解的原理大概是什么，就可以调用计算机数学软件来具体求解。同理，之前的一元一次方程和一元一次不等式、二元一次方程和二元一次不等式的求解也一点都不重要。你也只需要它们是什么，通常在什么场合出现，求出来的解的含义是什么，求解的原理大概是什么，就可以调用计算机数学软件来具体求解。当然，对于这几个方程求解方法之间的联系，你还是要知道的。也就是说，概念上、逻辑上的联系，你有必要搞清楚。具体求解步骤能用计算机数学软件来完成就行。

在本书即将进入到几何部分之前，我们把代数部分的知识梳理一下，如图 11.2。我们会发现，首先我们推广了数，从分数走到了无理数，合起来得到了实数。在这里我们仍然把自然数的概念建立在经验和日常语言之上。但是，我们也稍微体会了一下，用集合的语言，我们可以重新建立起来自然数。我们还在没有完全讲清楚的“极限”的概念的帮助之下，建立起来了“数轴上的点和实数一一对应”的认知。然后，把运算也拓展到了无理数上，例如把加法、乘法和幂运算都拓展到无理数上。

接着，我们学习了代数的思想——用符号表示对象。这里的对象包含物、量、数、关系算式，以及操作（从集合和映射来定义）。结合代数的思想和幂运算可以定义多项式。多项式是所有代数式里面最简单的情形，而

一元一次多项式、一元二次多项式是最简单的多项式。因此我们学习了这些简单的多项式。为了给求解方程做准备，我们也学习了一下因式分解。

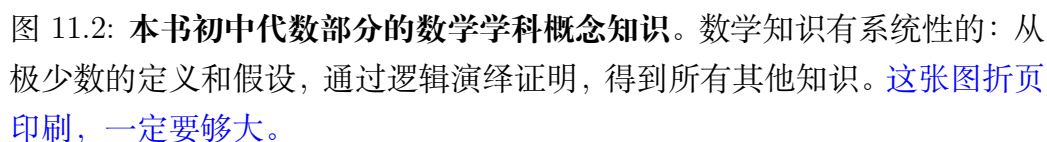
同时，基于集合和映射还可以定义函数（数集合上的映射）。有了函数的概念，结合代数的思想，我们就可以研究一元一次多项式和一元二次多项式对应的函数，也就是一元一次函数和一元二次函数。在函数的帮助下，我们更清楚地看到了方程——函数和 x 轴的交点和不等式——根据函数找出来其函数值符号不等号的区域的含义。

在具体求解方程和不等式的时候，我们运用了等式的性质、运算律和不等式的性质，通过对方程的做边做变换使得其未知数或者说变量单独出现，系数为 1。另一个求解方程的办法，或者说可以同时使用相互参考的办法是把相应的函数变成这个函数的最简单的形式，例如把 $y = ax^2 + bx + c$ 通过移动和缩放、翻转坐标轴变成 $y = x^2$ 的图象。同时，我们还发现，求解一元二次方程的时候，我们可以运用一元一次方程——通过因式分解，求解二元一次方程的时候，我们可以运用一元一次方程——通过变量替换法或者消元法。这就是知识内部的依赖关系。

通过这部分知识的学习，我们希望读者能够充分地体会到：

1. 数学知识有系统性的：从极少数的定义和假设，通过逻辑演绎证明，得到所有其他知识。例如，所有的代数部分的知识，都基于数和运算的定义（目前仍然是基于经验的，未来可以基于集合和映射），以及由此得到的运算律、等式的性质、不等式的性质；
2. 从小学到初中最重要的转变就是代数的思想——用符号表示对象，以及后来专门研究符号表示的对象，而不再关注数；
3. 逻辑言以证明和归纳总结启发直觉对于数学的不同意义：定义和假设部分可以受现实启发或者受数学家的直觉和想象力的启发，受对知识系统性的追求的启发。当然逻辑演绎证明的部分也可能得依靠数学家的直觉和想象力，但是总是更可以靠努力靠逻辑的。

跟平面直角坐标系一样，下一章将是代数和几何之间的联系——这将是非常重要的章节。



第十二章 锐角三角函数

Mathematics compares the most diverse phenomena and discovers the secret analogies that unite them.

数学把最相异的现象做比较，并且发现把它们统一起来的秘密（例如，三角函数¹）。

– Joseph Fourier

12.1 复习：勾股定理

勾股定理指的是直角三角形 $\triangle ABC$ 的斜边 c 和直角边 a, b 的长度满足如下关系

$$a^2 + b^2 = c^2. \quad (12.1)$$

这个勾股定理，之前我们可能已经从面积关系来证明过。未来在本书的几何学知识体系之中，我们会反过来把勾股定理当作最基础的知识，从这里来得到面积关系等其他知识。但是，在这里，我们先来做一个复习，把勾股定理当作面积关系等其他知识的导出知识。

假设我们已知长方形面积公式，也就是

$$S = ab. \quad (12.2)$$

而这个公式本身是通过数数和乘法得到的。也就是，先规定一个面积的标准单位，例如一个周长为一个单位长度（例如 1 厘米，或者 1 米）的正方

¹括号内为吴金闪补充。傅里叶把三角函数用于求解一般微分方程和表示一般函数，从而三角函数在数学中成了一般的工具而不仅仅用于处理三角形。

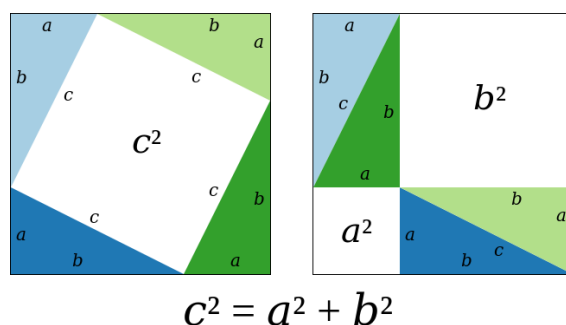


图 12.1: **Pythagoras 割补图形的证明**。把一个边长为 $a + b$ 的正方形做两种不同的分割，可以得到中间区域不同的图形，但是总面积不变，染色部分的面积也不变。图片来自于Wikipedia 的“Pythagorean Theorem”词条，感谢 William B. Faulk。2025 年 10 月 17 日访问。

形的面积为一个标准面积单位，然后，我们就看看一个长为 a 个单位宽为 b 个单位的长方形的面积是多少。只需要把这个长方形做一下分割，自然就发现，我们有 $a \times b$ 个标准面积单位。这样就得到了长方形面积公式。

其次，我们还可以朴素地得到面积可以累加，也就是如果我们把一个长方形分割成多个长方形，则整体长方形的面积等于分割出来的所有长方形的面积之和。更进一步，把任意一个图形分割成若干个任意的图形，其面积仍然满足可加的关系，也就是整体图形的面积等于分割开的面积之和。实际上，上面通过乘法和数数来得到长方形面积公式的时候，我们已经用了这一条面积可加性。

在这一章中，暂时我们接受这两条性质。有了这两条性质，我们就可以来“证明”勾股定理了。

证明 12.1. 基于割补图形证明勾股定理 这个来自于 *Pythagoras* (毕达哥拉斯) 的证明是这样的，首先，把一个边长为 $a + b$ 的正方形做两种不同的分割。第一种分割是四个相同的直角三角形把边长为 c 的正方形围在中间。第二种分割是把四个三角形两两组合起来，剩下的是一个边长为 a 的正方形和一个边长为 b 的正方形。按照总面积不变，四个三角形的面积之和也不变，我们得到

$$c^2 = S_{Total} - 4 \times S_{\triangle} = a^2 + b^2. \quad (12.3)$$

在这个图形切分的过程中，我们实际上用到了直角三角形的两个锐角之和等于一个直角的度数的性质，也就是三角形内角和等于平角的度数的性质。

我们还可以用其他方法来证明勾股定理，例如运用三角形的相似，运用量纲分析——一种基于单位的分析方法。由于量纲分析的证明和我们这一章需要学习的三角函数关系密切，我们把量纲分析的证明也在这里呈现给大家。首先，我们判断，对于直角三角形来说，其形状由两个因素决定，要么就是两条直角边的长度，要么就是斜边的长度和一条直角边的长度，要么就是斜边的长度和斜边与一条直角边的夹角。每一个上面的条件决定了一个直角三角形这一点，我们可以通过作图来验证。但是，严格来说，我们是通过几何知识来证明这一点的。为了下面的证明方便，我们取斜边的长度 c 和斜边与一条直角边的夹角 θ 当作直角三角形的形状的决定因素。

我们知道，直角三角形的面积，尽管我们还不会计算，肯定是 $S = s(c, \theta)$ 。其中，我们知道， c 的单位是长度的单位，例如米； S 的单位是面积的单位，例如平方米。按照单位可计算并且要匹配的原则，我们可以知道， S 和 c 之间的关系肯定是

$$S = \lambda(\theta) c^2. \quad (12.4)$$

注意，这里角度 θ 就是一个数，没有单位。

证明 12.2. 基于单位来证明勾股定理 我们在直角三角形 $\triangle ABC$ 里面，作一条斜边上的高 CD ，可以把这个三角形分成两个小的直角三角形 $\triangle ACD$ 和 $\triangle BCD$ ，如图 12.2 所示。它们也是一个角是 θ 的直角三角形，斜边长度分别是 b 和 a 。因此，其面积分别为

$$S_{ACD} = \lambda(\theta) b^2, S_{BCD} = \lambda(\theta) a^2. \quad (12.5)$$

这两个直角三角形合起来正好就是原来的三角形，于是面积也相同，所以

$$\lambda(\theta) c^2 = \lambda(\theta) (a^2 + b^2). \quad (12.6)$$

只要 $\lambda(\theta) \neq 0$ （对于需要计算面积的三角形，这个条件自然满足），等式两边同时除以 $\lambda(\theta)$ ，就得到

$$c^2 = a^2 + b^2. \quad (12.7)$$

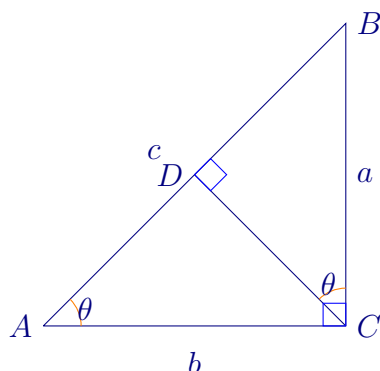


图 12.2: **勾股定理的量纲证明**。基于单位可计算而且算出来的单位要和所要求出来的量的单位一样这个理念（称为量纲定理），可以证明勾股定理。其中，直角三角形决定因素的确定和选择是关键。

这个证明，按照 Wikipedia，是 Einstein 提出来的²，也可以在赵凯华的《定性和半定量物理学》[4] 上找到。

注意，我们这个证明用到了面积关系——分割开的图形的面积之和等于合起来的图形的面积，我们还用到了三角形的内角和等于 180° ，以及直角三角形只有斜边长度 c 和一个锐角大小 θ 两个自由度，也就是说，只要两个直角三角形的斜边长度 c 和一个锐角大小 θ 相同，则这两个直角三角形肯定是完全相同的。未来这些性质，我们都是要去严格地证明的。

或者，反过来，一旦这些性质之中有一条不成立，实际上，勾股定理就有可能是不成立。在非欧几何中，勾股定理是不成立的，那么，那个时候到底上面的哪一条被破坏了呢？等你学习到非欧几何的时候，希望你还能记得去思考这个问题。

²见 Wikipedia 的 Pythagorean Theorem 词条，2025 年 10 月 17 日访问。

12.2 从直角三角形的等比例扩大到相似、全等和三角函数的定义

注意勾股定理有一个很好的性质，如果我们等比例扩大三条边，则勾股定理仍然成立，也就是当 $a^2 + b^2 = c^2$ 成立的时候，只要 $k > 0$ 就有

$$(ak)^2 + (bk)^2 = (ck)^2. \quad (12.8)$$

这意味着直角三角形可以被等比例放大，有可能这样的等比例放大保持了直角三角形的某种东西没变。我们来试试寻找这个不变的东西。

注意我们要找一个边的等比例变化不改变的东西。例如，我们可以试试边的比例 $\frac{a}{c}, \frac{b}{c}, \frac{a}{b}, \frac{b}{a}$ 。如果我们考虑到自由变量的数量，结合勾股定理，其实所有这些都是一个比例，知道其中就可以得到所有的其他的比例。例如当我们选 a, b 当作自变量，则 $\frac{a}{c} = \frac{a}{\sqrt{a^2+b^2}}, \frac{b}{c} = \frac{b}{\sqrt{a^2+b^2}}$ ，进而有

$$\left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = 1, \quad (12.9)$$

$$\frac{a}{b} = \frac{\frac{a}{c}}{\sqrt{1 - \left(\frac{a}{c}\right)^2}}. \quad (12.10)$$

例如，这里我们选 $\frac{a}{c}$ 。注意一开始我们有两个自变量 (a, b) ，定义成比例以后，我们只有一个独立的比例了。注意，这个独立的比例不随着直角三角形的放大缩小而变化，

$$\frac{a}{\sqrt{a^2 + b^2}} = \frac{ka}{\sqrt{(ka)^2 + (kb)^2}}. \quad (12.11)$$

我们进一步来看看这个不随着直角三角形的缩放而改变的东西和三角形的什么量有关。首先，在直角三角形里面，除了三条边（由于有勾股定理，只有两个独立变量），就剩下三个角。其中一个角是直角，被固定了。那只剩下两个变量。同时由于三角形内角和等于 180° ，就导致另外两个角加起来必须等于 90° 。也就是说，角度这里只有一个独立变量。比如说，在图 12.3 中的直角三角形 $Rt\triangle ABC$ （逆时针方向编号， C 为直角）中，我们选 $\angle CAB$ 当作独立变量。

这样一个三角形本来有三个角的大小和三条边的长度这 6 个变量, 经过我们分析之后, 发现对于直角三角形最多有 $a, b, \angle CAB$ 三个独立变量。我们来进一步考察这三个变量是否独立。

如果我们固定 b (也就是固定 C, A 两点, 固定 $\angle BCA = 90^\circ$), 让 a 变大 (也就是移动 B 使得 CB 长度增加), 我们自然发现, $\angle CAB$ 必须相应地变大才能连接上 AB 。同理, 固定 a , 让 b 变大, 则 $\angle ABC = 90^\circ - \angle CAB$ 必须变大, 也就是说 $\angle CAB$ 必须相应地变小。这就说明, $a, b, \angle CAB$ 不独立, $\angle CAB$ 是 (a, b) 的函数, 并且是 a 的增函数, b 的减函数,

$$\angle CAB = g(a, b). \quad (12.12)$$

前面我们已经得到一个直角三角形中不随着边长的缩放而变化的量 $\frac{a}{\sqrt{a^2+b^2}}$, 刚好也具有这个量是 a 的增函数, b 的减函数的特征。也就是说, 改变 (a, b) 会改变 $\angle CAB$, 也会改变 $\frac{a}{\sqrt{a^2+b^2}}$ 。连函数增减都有对应。那么, 是不是 $\angle CAB$ 和 $\frac{a}{\sqrt{a^2+b^2}}$ 就是一一对应的呢? 注意, 在直角三角形里面确实也已经没有其他独立变量。因此, 我们猜测

$$f(\angle CAB) = \frac{a}{\sqrt{a^2+b^2}}. \quad (12.13)$$

未来我们会发现这个函数对数学知识结构本身非常重要, 也非常有用。因此, 它有了自己的名称和符号, 称作正弦函数, 记作,

$$\sin(\angle CAB) = \frac{a}{\sqrt{a^2+b^2}}. \quad (12.14)$$

至于这个函数图象长什么样, 我们稍后再来回答。我们至少知道, 在直角三角形内, 也就是当 $0^\circ < \angle CAB < 90^\circ$ 的范围内, $\sin(\angle CAB)$ 是一个增函数 (因为当 a 增加的时候, 右侧函数值增加, 而通过前面的三角形构建过程我们也知道 $\angle CAB$ 的角度也再增加)。

我们把不独立的那几个函数也给一个专门的名称和符号,

$$\cos(\angle CAB) = \frac{b}{\sqrt{a^2+b^2}} = \sqrt{1 - \sin^2(\angle CAB)}, \quad (12.15)$$

$$\tan(\angle CAB) = \frac{a}{b} = \frac{\sin(\angle CAB)}{\cos(\angle CAB)}. \quad (12.16)$$

一旦这几个函数真的定义了, 这几个函数的图象我们也清楚了 (稍后我们就来完成这个任务), 我们就可以得到如下的关系,

$$a = \sin(\angle CAB)c, b = \cos(\angle CAB)c. \quad (12.17)$$

这告诉我们, 对于直角三角形, 只要确定斜边的长度, 确定其中一个锐角的大小 (另外两个角自然也就确定了), 则其两条直角边的长度也就确定了。也就是说, 斜边和锐角完全确定了一个直角三角形。这就补上了之前量纲分析的时候的假设。

更进一步, 这还告诉我们, 对于直角三角形, 只要其中一个锐角的大小确定, 则 $\frac{a}{c}, \frac{b}{c}, \frac{a}{b}$ 完全确定, 也就是说每条边同时缩放相同的比例, 这个比值不变

$$\sin(\angle CAB) = \frac{a'}{c'} = \frac{ak}{ck} = \frac{a}{c}, \quad (12.18a)$$

$$\cos(\angle CAB) = \frac{b'}{c'} = \frac{bk}{ck} = \frac{b}{c}, \quad (12.18b)$$

$$\tan(\angle CAB) = \frac{a'}{b'} = \frac{ak}{bk} = \frac{a}{b}. \quad (12.18c)$$

我们终于发现, 同时缩放直角三角形所有的边不变的是什么, 不变的是锐角的大小。再加上, 只要一个锐角相同, 则缩放前后的两个直角三角形的所有角都相同, 我们就得到两个三角形长度成比例, 角度相同。我们把这样的两个三角形称作相似三角形。其朴素含义就是通过缩放可以把两个三角形完全重合。

算式 (12.18) 还可以变形为

$$\frac{a'}{a} = \frac{b'}{b} = \frac{c'}{c} = k, \quad (12.19)$$

这表示两个直角三角形之间的对应边成比例。我们把满足三边对应成比例并且每一对角都对应相等的两个三角形称为相似三角形。特别地, 当这个比例系数是 1 的时候, 我们把这样的两个相似三角形称为全等三角形。于是, 我们得到有一个锐角大小相等的直角三角形相似。如果这个基础上, 再增加一条, 对应 (对应的意思是, 这条相等的边要么都是这个相等的锐角的对边, 要么都是邻边, 要么都是斜边) 边相等, 则两个直角三角形全等。为什么只需要增加一条边相等的条件? 因为根据算式 (12.19), 只要有一个锐

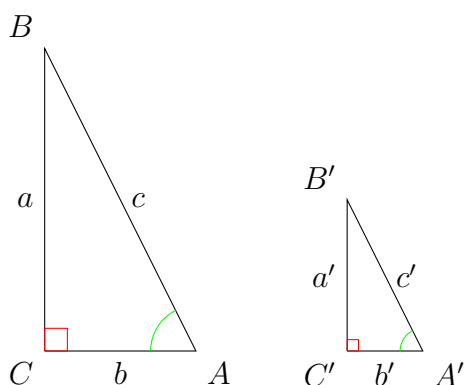


图 12.3: 直角三角形的边角关系和相似。通过勾股定理可以看到直角三角形的边角关系, 以及同比例缩放以后的两个直角三角形的关系。

角相等则这两个直角三角形对应边成比例, 只要一对对应边相等则这个比例就是 1, 因此其他边都对应相等。

习题 12.1 (相似直角三角形的边). 从算式 (12.17) 证明 算式 (12.19), 从而得到两个相似的直角三角形——这里的含义是有一个锐角大小相同的直角三角形——的三边对应成比例。

在这里, 我们已经学习一下直角三角形的相似性和全等。未来我们会学习一般的三角形的相似性和全等。再进入那部分之前, 我们来了解一下正弦函数和余弦函数所建立的从角的大小 (在锐角的范围内) 和角的正弦值和余弦值之间的一一对应关系到底是什么样的。

12.3 锐角的正弦余弦函数 *

根据正弦和余弦三角函数和对边以及邻边的关系, 我们来通过围绕单位圆 ($r = 1$ 相当于直角三角形的斜边 $c = 1$) 的画图了解更多的关于正弦和余弦三角函数的性质。

单位圆指的是在二维平面直角坐标系中以原点为中心, 半径长度为 1 的圆。圆的明确的定义见平面几何部分。在这里, 我们先粗略地运用在小学阶段已经初步具有的圆的概念。

我们把单位圆中 x 轴和弦的夹角当作 θ , 选择圆上一点 P , 连起来这

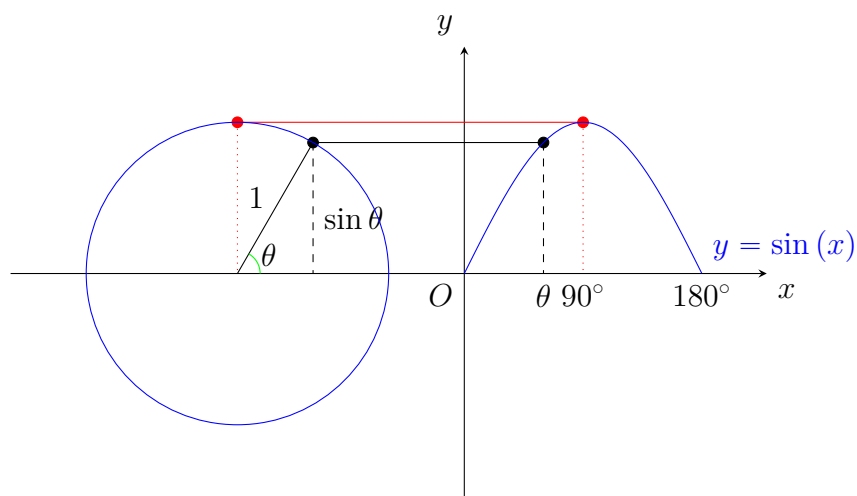


图 12.4: 通过单位圆得到正弦函数值。通过单位圆可以得到正弦函数值, 并画出来正弦函数图象。

一点和圆心, 就得到一条弦。过 P 做 PH 垂直于 x 轴, 得到 P 点的 x 坐标和 y 坐标。这样 $\sin \theta$ 就是 y 坐标, $\cos \theta$ 就是 x 坐标。

在图 12.4 中, 我们画出了 $0^\circ \sim 180^\circ$ 之间的角度对应的正弦函数值。其中大于 90° 的部分的正弦值可以看作是其补角的正弦值。于是余弦值, 从单位圆上的点的 x 坐标值也可以得到。其中大于 90° 的部分的余弦值可以看作是其补角的余弦值的相反数。

$$\sin(180^\circ - \theta) = \sin \theta \quad (0^\circ \leq \theta \leq 90^\circ), \quad (12.20a)$$

$$\cos(180^\circ - \theta) = -\cos \theta \quad (0^\circ \leq \theta \leq 90^\circ). \quad (12.20b)$$

我们之所以能够把角度推广到 $0^\circ \sim 180^\circ$ 之间, 是因为我们把 “ $\sin \theta$ 就是 y 坐标, $\cos \theta$ 就是 x 坐标” 从 $0^\circ \sim 90^\circ$ 拓展到了 $90^\circ \sim 180^\circ$ 。未来我们还会进一步把角度推广到 $0^\circ \sim 360^\circ$ 。目前, 我们只需要关注其中的锐角部分。

我们得到在 $0^\circ \sim 180^\circ$ 的范围内, 正弦函数值和角度具有一一对应关系, 余弦函数值和角度也有一一对应关系。

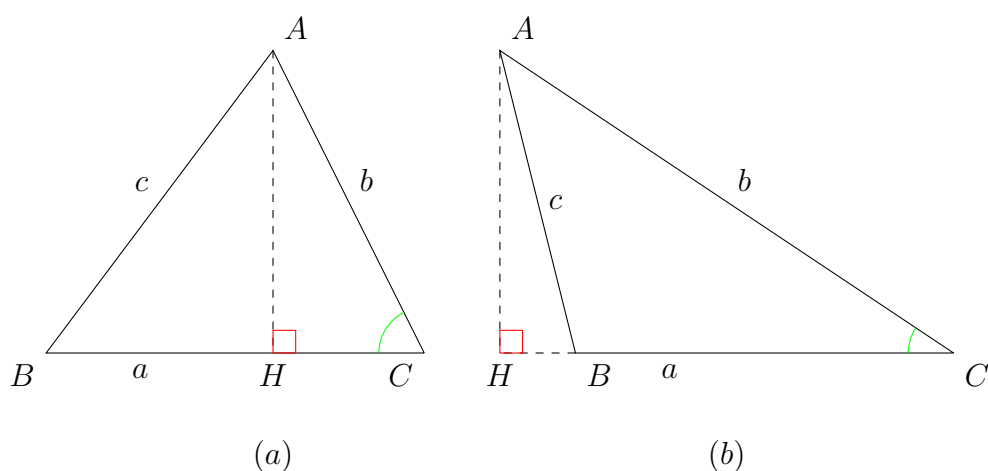


图 12.5: 余弦定理的证明。通过勾股定理可以证明余弦定理。

12.4 一般三角形的相似以及三角形中的正弦定理和余弦定理

对于一般的锐角三角形，我们总是可以通过做底边上的高的方式来转化为两个直角三角形。然后，我们就可以运用直角三角形的三角函数以及边的关系来研究一下一般的锐角三角形了。我们试试，是否可以通过这样的方式来得到一般锐角三角形的相似和全等，乃至更加一般的三角形的相似和全等，以及一般的三角形的三角函数。

定理 12.1: 余弦定理

余弦定理] 在三角形 $\triangle ABC$ 中, $a^2 + b^2 - 2ab \cos C = c^2$ 。

证明 12.3. 我们先考虑锐角三角形 $\triangle ABC$ 。由于三个角都是锐角, 每一条高都在三角形的内部。如图 12.5(a), 我们做 BC 上的高 AH , 也就是 $AH \perp BC$ 于 H 。于是三角形 $\triangle AHC, \triangle AHB$ 都是直角三角形。按照直角三角形的三角函数和边的关系, 有 $AH = b \sin C, HC = b \cos C$, 因此 $BH = a - b \cos C$ 。对直角三角形 $\triangle AHB$ 运用勾股定理得到 $c^2 = (a - b \cos C)^2 + (b \sin C)^2 = a^2 + b^2 - 2ab \cos C$ 。以上证明就算对于钝角三角形里面的锐角仍然成立。下面我们来补充专门针对钝角的证明

如图 12.5(b), 我们再考虑钝角三角形 $\triangle ABC$, 假设 C 是钝角。我们做 BC 上的高 AH , 也就是 $AH \perp BC$ 于 H 。但是, 这个时候 H 在 BC 的延长线上, 而不是 BC 之间。 $(180^\circ - C)$ 是锐角, 会出现在直角三角形里面。三角形 $\triangle AHC, \triangle AHB$ 都是直角三角形。按照直角三角形的三角函数和边的关系, 有 $AH = b \sin(180^\circ - C), HC = b \cos(180^\circ - C)$, 因此 $BH = a + b \cos(180^\circ - C)$ 。对直角三角形 $\triangle AHB$ 运用勾股定理得到 $c^2 = (a + b \cos(180^\circ - C))^2 + (b \sin(180^\circ - C))^2 = a^2 + b^2 + 2ab \cos(180^\circ - C)$ 。

按照算式 (12.20), 我们得到 $c^2 = a^2 + b^2 - 2ab \cos C$ 。于是, 不管是锐角还是钝角三角形, 我们都有 $c^2 = a^2 + b^2 - 2ab \cos C$ 。

余弦定理的表达式

$$a^2 + b^2 - 2ab \cos C = c^2 \quad (12.21)$$

可以看作是从边来计算任意一个三角形的内角角度的方式 $\cos C = \frac{c^2 - a^2 - b^2}{2ab}$ 。这时候, 我们发现, 如果三角形的三边同比例缩放, 这个角的大小不变, $\cos C = \frac{(kc)^2 - (ka)^2 - (kb)^2}{2k a k b} = \frac{c^2 - a^2 - b^2}{2ab}$ 。也就是说, 我们可以得到三边成比例, 三个角完全相等的两个三角形。或者说, 两个三角形之间可以通过缩放使得其相互重合。这样的三角形正好符合我们定义直角三角形相似的时候的相似的含义。

余弦定理的表达式也可以看作是给定两条边 (a, b) 和两条边的夹角 C

以后计算另一条边 c 的方式, 也就是 $c^2 = a^2 + b^2 - 2ab \cos C$ 。这时候我们发现, 如果 (a, b) 被缩放 k 倍但是两条边的夹角 C 不变, 则 $c' = \sqrt{(ka)^2 + (kb)^2 - 2kaka \cos C} = kc$ 。也就是说, c 也被缩放 k 倍了。这正好也是相似的含义。

甚至我们可以把两者结合起来, 把上面的情况的另外连个角也算出来, 跟缩放之前的三角形的相应角比较一下, 例如

$$\cos A = \frac{(ka)^2 + (kb)^2 - (kc)^2}{2kbkc} = \frac{a^2 + b^2 - c^2}{2bc}. \quad (12.22)$$

完全没有变化。因此, 我们得到如果 $(a', b') = (ka, kb)$ 被缩放 k 倍但是两条边的夹角 C 不变, 则 $c' = kc$, 同时角 A, B 也不变。因此, 这两个三角形也是相似三角形——三个角对应相等, 三条对应边成比例。

定理 12.2: 三角形相似的 SSS 判定定理

三角形相似的 SSS 判定定理] 给定两个三角形, 如果其对应边成比例, 则这两个三角形是相似三角形, 也就是三个角对应相等, 三条对应边成比例。

定理 12.3: 三角形相似的 SAS 判定定理

三角形相似的 SAS 判定定理] 给定两个三角形, 如果其两组对应边成比例, 其夹角相等, 则这两个三角形是相似三角形, 也就是三个角对应相等, 三条对应边成比例。

我们再来证明正弦定理。

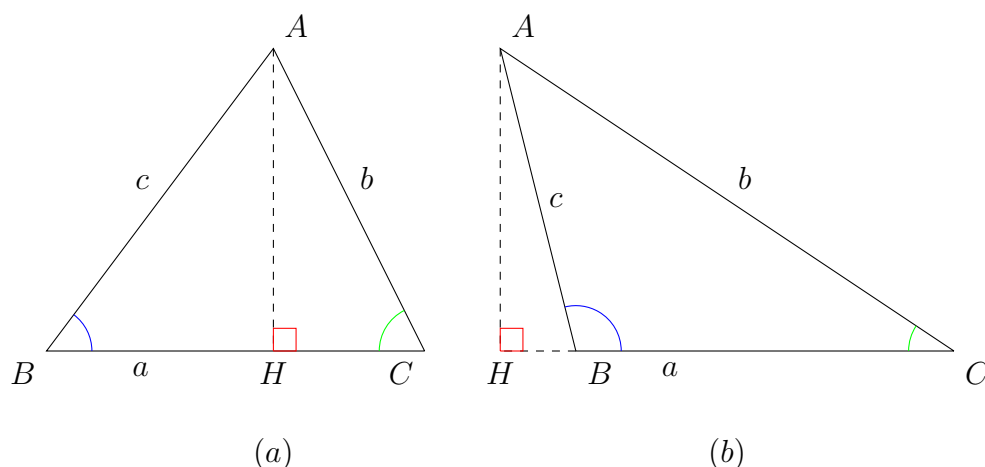


图 12.6: 正弦定理的证明。通过直角三角形内的锐角三角函数可以证明正弦定理。

定理 12.4: 正弦定理

正弦定理] 在三角形 $\triangle ABC$ 中, $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$ 。

证明 12.4. 如图 12.6(a), 我们先考虑锐角三角形 $\triangle ABC$ 。由于三个角都是锐角, 每一条高都在三角形的内部。我们做 BC 上的高 AH , 也就是 $AH \perp BC$ 于 H 。于是三角形 $\triangle AHC$, $\triangle AHB$ 都是直角三角形。按照直角三角形的三角函数和边的关系, 有 $AH = b \sin C$, $AH = c \cos B$, 因此 $b \sin C = AH = c \cos B$, 也就是 $\frac{b}{\sin B} = \frac{c}{\sin C}$ 。同理可得 $\frac{a}{\sin A} = \frac{b}{\sin B}$, 合起来就得到了 $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$ 。

以上证明就算对于钝角三角形里面的锐角仍然成立。下面我们来补充专门针对钝角的证明。

如图 12.6(b), 我们再考虑钝角三角形 $\triangle ABC$, 假设 C 是钝角。我们做 BC 上的高 AH , 也就是 $AH \perp BC$ 于 H 。但是, 这个时候 H 在 BC 的延长线上, 而不是 BC 之间。 $(180^\circ - C)$ 是锐角, 会出现在直角三角形里面。三角形 $\triangle AHC$, $\triangle AHB$ 都是直角三角形。按照直角三角形的三角函数和边的关系, 有 $AH = b \sin(180^\circ - C)$, $AH = c \sin B$, 因此 $b \sin(180^\circ - C) = AH = c \sin B$ 。

按照算式 (12.20), 我们得到 $b \sin C = c \sin B$, 也就是 $\frac{b}{\sin B} = \frac{c}{\sin C}$ 。结合另一组对角和对边的同样的表达式 $\frac{a}{\sin A} = \frac{b}{\sin B}$, 合起来就得到了 $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$ 。于是, 不管是锐角还是钝角三角形, 我们都有 $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$ 。

根据正弦定理，如果我们保持一个三角形的两个角不变（自然也就保持了三个角都不变），把其中的一条边（例如 $a' = ka$ ）缩放 k 倍，则另外两条边也会被同样缩放 k 倍，因为

$$b = \frac{\sin B}{\sin A}ka, c = \frac{\sin C}{\sin A}ka. \quad (12.23)$$

因此，有两组角对应相等的三角形是相似三角形。这样我们就证明了相似三角形的 AA 判定定理。

定理 12.5: 三角形相似的 AA 判定定理

三角形相似的 AA 判定定理] 给定两个三角形，如果其两组对应角相等，则这两个三角形是相似三角形，也就是三个角对应相等，对应边成比例。

当对应边相等，也就是比例系数 $k = 1$ 的时候，我们得到的两个三角形三个角对应相等，三条边也相等。我们把这样的三角形称为全等三角形。于是，我们有了下列全等三角形判定的定理。

定理 12.6: 三角形全等的 SSS 判定定理

三角形全等的 SSS 判定定理] 给定两个三角形，如果其对应边相等，则这两个三角形是全等三角形，也就是三个角对应相等，三条对应相等。

定理 12.7: 三角形全等的 SAS 判定定理

三角形全等的 SAS 判定定理] 给定两个三角形，如果其两组对应边相等，其夹角相等，则这两个三角形是全等三角形，也就是三个角对应相等，三条对应相等。

定理 12.8: 三角形全等的 ASA 判定定理

三角形全等的 ASA 判定定理] 给定两个三角形, 如果其两组对应角相等 (例如 $\angle A' = \angle A, \angle B' = \angle B$), 一组对边相等 (例如 $c' = c$), 则这两个三角形是全等三角形, 也就是三个角对应相等, 三条对应相等。

证明 12.5. 根据相似三角形, 我们已经得到了三个角对应相等了。我们来证明三条边也对应相等。根据 $c' = c$ 和正弦定理, 有 $a' = \frac{\sin A}{\sin 180^\circ - A - B} c' = \frac{\sin A}{\sin 180^\circ - A - B} c = a, b' = \frac{\sin B}{\sin 180^\circ - A - B} c' = \frac{\sin B}{\sin 180^\circ - A - B} c = b$ 。因此, 另外两条边也对应相等。

在这里, 我们证明过程其实没有区分这条额外给出的相等的边是和给定对应相等的角的相对关系: 第一种关系是给定的对应相等的两组角夹住了这条边, 也就是这条边和这两组对应相等的角中的两个都相邻, 第二种关系是给定的对应相等的两组角中的一组和这条边相邻。我们能够这样做的原因是在三角形中, 知道任何两个内角, 都相当于知道了三角形的所有内角, 因此, 总是可以得到三组角都对应相等。也就是说, 上面的两种情况可以相互转换, 完全等价。也就是说, 以下“三角形全等的 ASA 判定定理”也成立。我们也可以给出单独的证明。

定理 12.9: 三角形全等的 AAS 判定定理

三角形全等的 AAS 判定定理] 给定两个三角形, 如果其两组对应角相等 (例如 $\angle A' = \angle A, \angle B' = \angle B$), 一组对边相等 (例如 $a' = a$), 则这两个三角形是全等三角形, 也就是三个角对应相等, 三条对应相等。

证明 12.6. 根据相似三角形, 我们已经得到了三个角对应相等了。我们来证明三条边也对应相等。假设 $a' = a$ 是那条对应相等的边, 则 $b' = \frac{\sin B}{\sin A} a' = \frac{\sin B}{\sin A} a = b, c' = \frac{\sin 180^\circ - A - B}{\sin A} a' = \frac{\sin 180^\circ - A - B}{\sin A} a = c$ 。因此, 另外两条边也对应相等。

12.5 三角函数的应用

例 12.1 (三角形面积公式). 证明三角形面积公式 $S = \frac{1}{2}ab \sin C$ 。其中 C 为 a, b 边的夹角。

在三角形 $\triangle ABC$ 中, 过 A 做 BC 的高, 记交点也就是垂足为 H , 则 $AH = b \sin C$ 。根据三角形面积公式 $S = \frac{1}{2}BC \times AH = \frac{1}{2}a \times b \sin C$, 因此, $S = \frac{1}{2}ab \sin C$ 。

12.6 推荐阅读材料

12.7 作业

12.8 本章小结

在本章中, 我们从面积关系和量纲分析证明了勾股定理。但是, 未来我们会看到, 面积关系其实往往是依赖于勾股定理而成立的。因此, 严格来说, 这不算勾股定理的真正的证明。

但是, 我们从勾股定理成立引入了直角三角形的相似和直角三角形的三角函数。接着, 推广到了一般三角形的相似的概念以及判定定理。最后, 得到了一般三角形的全等的概念和判定定理。

中间为了实现这个从直角三角函数到一般三角函数的相似和全等的推广, 我们证明了余弦定理和正弦定理。在三角函数的应用这一节中我们看到, 三角函数以及这两个定理自身也是很重要的。

注意, 这是一套和通常的平面几何学差异很大的知识体系。一般的平面几何学从平行公理出发来构建整个知识体系, 而我们这个从勾股定理出发来构建整个平面几何的知识体系。未来我们也会建立这两套知识体系之间的等价关系。但是, 在那之前, 我们先走以勾股定理为基础的路线。

在下面几章中, 我们将严格地来建立以勾股定理为基础的平面几何知识。

第十三章 平面几何：基本定义

13.1 点和距离的定义

之前我们已经学习过“实数和数轴上的点完全一一对应”。注意这一点不是显然的，例如整数显然不能覆盖数轴上的所有点，有理数也是如此，那为什么有理数和无理数合——合起来称为实数——起来就包含了所有的点呢？未来在大学阶段我们会更加严格地来学习实数，从而真正能够明白“实数和数轴上的点完全一一对应”这个结论。目前，我们只需要知道这个结论以及明白这里其实还有个问题就可以。

有了数轴，有了数轴上的点和实数相对应，我们就可以建立平面直角坐标系。这样平面上的任意一点 P 就可以用一对数 $P = (x_P, y_P)$ 来表示了。根据勾股定理，通过构建一个直角三角形，如图 13.1，任意两点之间的距离可以表示为

$$PQ = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2}. \quad (13.1)$$

有了距离，我们来定义直线。先给定两点 $P = (x_P, y_P)$ 和 $Q = (x_Q, y_Q)$ 。注意，这里 $x_P = x_Q$ 和 $y_P = y_Q$ 不能同时满足，否则， P, Q 就是同一个点了。我们假设 $x_P \neq x_Q$ 成立。否则，我们只需要交换 x, y 轴的名称（因为当 x 坐标相等的时候， y 坐标肯定不等）。不失一般性，我们进一步假设 $x_P < x_Q$ 。如果不满足这个条件，我们就把 P, Q 的符号交换一下。

我们定义过 P, Q 两点的直线——也称为由 P, Q 决定的直线——为任意一点 $M = (x_M, y_M)$ 使得 $PM + MQ = PQ$ ($x_P \leq x_M \leq x_Q$), $MP + MQ =$

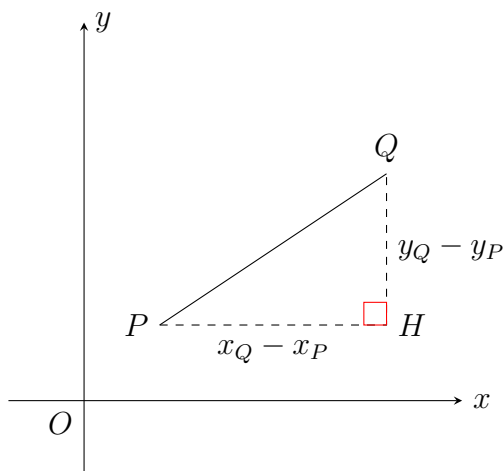


图 13.1: **距离公式 (勾股定理)**。基于坐标, 两点之间的距离可以表示为 $PQ = \sqrt{(x_Q - x_P)^2 + (y_Q - y_P)^2}$ 。

MQ ($x_M \leq x_P \leq x_Q$), $PM = PQ + QM$ ($x_P \leq x_Q \leq x_M$) 三者之一成立的点 M 的集合。这三种情况刚好对应着 M 在 PQ 之间, M 在 QP 延长线上和 M 在 PQ 延长线上三种情况。未来, 当我们学习了矢量之后, 我们可以用矢量符号把这三个表达式统一为一个。

按照距离的公式, 也就是勾股定理, 这相当于是要求在相应的情况下有

$$\begin{aligned} \sqrt{(x_P - x_M)^2 + (y_P - y_M)^2} + \sqrt{(x_M - x_Q)^2 + (y_M - y_Q)^2} &= \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2}, \\ \sqrt{(x_M - x_P)^2 + (y_M - y_P)^2} + \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2} &= \sqrt{(x_M - x_Q)^2 + (y_M - y_Q)^2}, \\ \sqrt{(x_P - x_Q)^2 + (y_P - y_Q)^2} + \sqrt{(x_Q - x_M)^2 + (y_Q - y_M)^2} &= \sqrt{(x_P - x_M)^2 + (y_P - y_M)^2} \end{aligned}$$

(13.2a)

(13.2b)

(13.2c)

对于每一种情况, 我们可以通过求解方程求出来 (x_M, y_M) 之间的关系。这

里我们仅仅验证下面的表达式满足上面要求的方程,

$$\frac{(y_M - y_P)}{(x_M - x_P)} = \frac{(y_Q - y_P)}{(x_Q - x_P)}, \quad (13.3a)$$

$$\frac{(y_Q - y_M)}{(x_Q - x_M)} = \frac{(y_Q - y_P)}{(x_Q - x_P)}, \quad (13.3b)$$

并且这两个表达式等价。我们先来验证这两个表达式等价——把算式 (13.3a) 代入算式 (13.3b)。先从算式 (13.3a) 得到

$$y_M = (x_M - x_P) \frac{(y_Q - y_P)}{(x_Q - x_P)} + y_P, \quad (13.4)$$

再代入算式 (13.3b),

$$\frac{(y_Q - y_M)}{(x_Q - x_M)} = \frac{1}{(x_Q - x_M)} \left[y_Q - (x_M - x_P) \frac{(y_Q - y_P)}{(x_Q - x_P)} - y_P \right] \quad (13.5a)$$

$$= \frac{(y_Q - y_P)}{(x_Q - x_M)} \left[1 - (x_M - x_P) \frac{1}{(x_Q - x_P)} \right] \quad (13.5b)$$

$$= \frac{(y_Q - y_P)}{(x_Q - x_M)} \frac{(x_Q - x_M)}{(x_Q - x_P)} \quad (13.5c)$$

$$= \frac{(y_Q - y_P)}{(x_Q - x_P)}. \quad (13.5d)$$

刚好验证了算式 (13.3b)。

下面, 我们来验证算式 (13.3) 刚好满足算式 (13.2), 并且反过来算式 (13.2) 的所有解都满足算式 (13.3)。也就是两者完全等价。为了简化符号, 我们记常数——称为直线的斜率 (以后我们会明白为什么叫这个名字)

$$\frac{(y_Q - y_P)}{(x_Q - x_P)} = k. \quad (13.6)$$

以算式 (13.2a) 为例,

$$\sqrt{(x_P - x_M)^2} \sqrt{1 + k^2} + \sqrt{(x_M - x_Q)^2} \sqrt{1 + k^2} \stackrel{?}{=} \sqrt{(x_P - x_Q)^2} \sqrt{1 + k^2}. \quad (13.7)$$

如果 $x_P < x_M < x_Q$, 则上式变成

$$(x_M - x_P) + (x_Q - x_M) \stackrel{?}{=} (x_Q - x_P), \quad (13.8)$$

自动满足。同理可以验证当 $x_M < x_P < x_Q$ 的时候，算式 (13.3) 自动满足算式 (13.2b)；当 $x_P < x_Q < x_M$ 的时候，算式 (13.3) 自动满足算式 (13.2c)。

反过来，算式 (13.2) 的解肯定就是算式 (13.3) 怎么证明呢？在相应的条件下，例如当 $x_P < x_M < x_Q$ 的时候，算式 (13.2a) 只有唯一的解。因此，既然算式 (13.3a) 是算式 (13.2a) 的一个解，那么算式 (13.2a) 的解就是算式 (13.3a)。

那怎么证明当 $x_P < x_M < x_Q$ 的时候，算式 (13.2a) 只有唯一的解呢？

利用记号算式 (13.6) 定义的 k （由 x_P, y_P, x_Q, y_Q 共同决定），对算式 (13.3a) 做 x_M, y_M 变量的分离，得到

$$y_M = kx_M + (y_P - kx_P), \quad (13.9)$$

记常数——称为直线的截距（以后我们会明白为什么叫这个名字）

$$b = y_P - kx_P, \quad (13.10)$$

则

$$y_M = kx_M + b, \quad (13.11)$$

我们发现，满足算式 (13.11) 这个条件的 $M = (x_M, y_M)$ 是经过 P, Q 的一条直线。由于这个时候的点 M 不是特定点，而是一群满足算式 (13.11) 的点的集合，因此，我们也往往省去角标，直接写作

$$y = kx + b. \quad (13.12)$$

这就是我们最终得到的经过 P, Q 两点的方程。注意， P, Q 的信息被包含在了常数 k, b 之中。

我们进一步来证明，如果两条直线有完全相同的 (k, b) 则这两条直线完全重合，如果有不同的 (k, b) 则这两条直线要么相交要么平行，不可能完全重合。这样我们就得到了直线和 (k, b) 之间的一一对应关系，

定理 13.1: 直线和 (k, b) 一一对应

直线和 (k, b) 一一对应] 直线和描述这条直线的方程 $y = kx + b$ 的两个参数 (k, b) 之间的一一对应关系。也就是给定一组 (k, b) 存在唯一的一条直线, 给定一条直线存在着唯一的一组 (k, b) 。

直线 l_1 的方程为 $y = k_1x + b_1$, 直线 l_2 的方程为 $y = k_2x + b_2$ 。我们来证明, 如果 $k_1 = k_2, b_1 = b_2$ 则 l_1 和 l_2 重合, 也就是任何一个 l_1 上的点必然在 l_2 上, 反之亦然。

当 $k_1 = k_2, b_1 = b_2$ 的时候, 直线 l_1 的任意点 M 必然满足 $y_M = k_1x_M + b_1$, 因此, 也必然满足 $y_M = k_2x_M + b_2$ 。根据 l_2 是所有满足 $y = k_2x + b_2$ 的点的集合, 则 M 必然在 l_2 上。因此, 任意 l_1 上的点必然在 l_2 上。同理, 任意 l_2 上的点必然在 l_1 上。因此, l_1 和 l_2 完全重合。

我们再来看看两条直线是否会有有限个交点, 有的话多少个, 在什么情况下有。直线 l_1 的任意点必然满足 $y = k_1x + b_1$; 直线 l_2 的任意点必然满足 $y = k_2x + b_2$ 。我们来看一看两者是不是会在同一个点上。如果有相同的点 R , 则必然有

$$y_R = k_1x_R + b_1, y_R = k_2x_R + b_2. \quad (13.13)$$

求解出来这组二元一次方程组, 得到

$$x_R = \frac{b_2 - b_1}{k_1 - k_2}, y_R = \frac{k_1b_2 - k_2b_1}{k_1 - k_2}. \quad (13.14)$$

我们发现, 当 $k_1 = k_2$ 但是 $b_1 \neq b_2$ 的时候, 方程无解。也就是 l_1 和 l_2 没有共同的点, 也就是两条直线不相交。我们定义不相交的直线为平行线。因此, 我们得到平行线之间具有相同的斜率。

当 $k_1 \neq k_2$ 的时候, l_1 和 l_2 之间永远有且只有一个共同的点——称为两条直线的交点。

这样我们就得到了直线和 (k, b) 之间的一一对应关系: 给一组 (k, b) 的值, 就有一条独一无二的直线; 给一条直线就有一组独一无二的 (k, b) 的值。顺便, 我们还定义了两条直线的相交和平行。

根据前面的计算, 我们发现, 只要给定两个点, 就给定了一组 (k, b) 的值。结合现在得到的直线和 (k, b) 的值的一一对应关系, 我们就得到了

定理 13.2: 两点确定一条直线

两点确定一条直线] 给定任意不重合的两点，有且只有一条直线经过这两点。

有了直线，我们来定义两点之间的直线段。给定两点，过这两点的直线中处于这两点之间的部分称为这两点之间的直线段。根据勾股定理，我们有两点之间的距离就是两点之间的直线段的长度。

有了两点间的距离和直线概念，我们来讨论一下三个点之间的距离的关系。记给定的两点为 P, Q ，我们已经知道这两点之间的距离为 PQ 。如果第三个给定点在直线段 PQ 上，那么根据 $PM + MQ = PQ$ ，有 P, M 之间的距离，加上 M, Q 之间的距离，等于 PQ 之间的距离。先从 P 到 M 再从 M 到 Q 的路径的长度为 $PM + MQ = PQ$ 。如果第三点在 P, Q 决定的直线上，但是不处于 PQ 之间，例如在 PQ 的延长线上，根据 $PQ + QM = PM$ ，先从 P 到 M 再从 M 到 Q 的路径的长度为 $PM + MQ = PQ + 2QM \geq PQ$ 。如果 M 不在 PQ 所确定的直线上，则先从 P 到 M 再从 M 到 Q 的路径的长度为 $PM + MQ$ ，

$$PM + MQ = \sqrt{(x_P - x_M)^2 + (y_P - y_M)^2} + \sqrt{(x_M - x_Q)^2 + (y_M - y_Q)^2}. \quad (13.15)$$

上面这个表达式在下面的计算中会相当复杂。为了简化和不失去一般性，我们取 $P = (0, 0)$ ， $Q = (x_Q > 0, 0)$ 。坐标原点和坐标轴我们总是可以移动和旋转的。之所以不把 PQ 放在 y 轴上，也就是让 $x_P = x_Q$ 是之前我们假设了 $x_Q > x_P$ 。可以验算这个时候 QP 确定的直线的 $k = 0, b = 0$ 。这样， PQ 之间的直线段上的点 M 满足 $M = (0 \leq x_M \leq x_Q, 0)$ 。有了这个简化计算的设置之后，我们来重新计算先从 P 到 PQ 外一点 M 再从 M 到 Q 的路径

的长度为 $PM + MQ$,

$$\begin{aligned}
 PM + MQ &= \sqrt{(x_P - x_M)^2 + (y_P - y_M)^2} + \sqrt{(x_M - x_Q)^2 + (y_M - y_Q)^2}, \\
 &= \sqrt{(0 - x_M)^2 + (0 - y_M)^2} + \sqrt{(x_M - x_Q)^2 + (y_M - 0)^2}, \\
 &= \sqrt{x_M^2 + y_M^2} + \sqrt{(x_M - x_Q)^2 + y_M^2}, \\
 &\geq \sqrt{x_M^2} + \sqrt{(x_M - x_Q)^2} = x_M + x_Q - x_M = x_Q = PQ.
 \end{aligned}$$

(13.16)

顺便, 在解决问题的时候, 我们经常会为了简化计算做一些不失一般性 (这个要证明) 的设置。

习题 13.1 ($PM + MQ \geq PQ$ 的一般性证明). 对于一般的 $P = (x_P, y_P)$, $Q = (x_Q, y_Q)$, 证明先从 P 到 PQ 见直线段外一点 M 再回到 Q 的路径长度大于 PQ 。

这样我们就证明了 “在包含了两点之间的包含一个中间点的所有折线 (中间点在端点的直线段上的情况也可以看作是折线的特殊情况) 之中, 两点之间直线段长度最短”, $PM + MQ \geq PQ$ 。

同样的道理, 我们可以证明, 有两个中间点的折线的长度大于等于有一个中间点的折线的长度。记另一个中间点为 N , 连接 MQ 我们可以直接看出来, $MN + NQ \geq MQ$, 因此 $PM + MN + NQ \geq PQ$ 。或者, 我们也可以把 NP 直接连起来, 就能得到同样的结果。

类似地, 我们可以证明, 有三个中间点的折线的长度大于等于有两个中间点的折线的长度。也就是中间点的增加只能增加路径的长度。于是, 我们得到 “两点之间直线段长度最短”¹。

¹将来, 我们会更加严格地来证明这个结论, 用 “微分、积分” 和 “极限” 的概念

定理 13.3: 两点之间直线段长度最短

两点之间直线段长度最短] 在给定的两点之间的任意连线，这两点之间的直线段的长度最短。

未来我们会学习到勾股定理的一种叫做微分的表达形式 $ds = \sqrt{d^2x + d^2y}$ ，然后任何一条从 P 到 Q 的曲线的长度可以定义为积分的表达形式 $S = \int_{(x_P, y_P)}^{(x_Q, y_Q)} \sqrt{d^2x + d^2y}$ 。通过对这个积分形式求极值，就可以得到 PQ 之间长度最短的曲线是 P, Q 之间的直线段的长度，也得到直线段所应该满足的方程——称为测地线方程。在我们这里，相当于我们只能从有限长度形式的勾股定理 $a^2 + b^2 = c^2$ 出发，来得到过给定两点的直线的方程，来证明两点之间的长度最短的曲线是这两点之间的直线段。这里的论证过程比较复杂。未来基于微分和积分来得到相同的结论会简洁很多。顺便，这样的从距离表达式——未来称为度规——来得到几何知识的数学分支学科叫做微分几何。将来，微分几何不仅仅可以研究平直空间的几何，还可以研究弯曲空间的几何。

13.2 圆和角的定义

有了距离我们来定义圆。圆就是到固定点 O 距离相等的点的集合。其中的距离称为这个圆的半径 r ，固定点称为这个圆的圆心。由于圆心的位置移动以后圆的大小形状（形状在这里就是圆形，或者说各向同性，或者说旋转对称——把一个圆无论平面内转动多少度仍然和原来的圆重合）不变，因此，忽略圆的位置则圆的大小形状完全由半径决定。

有了圆的定义，结合日常生活中看到的时钟，我们就可以来定义角度。我们约定从某一个半径开始——通常画图的时候习惯上选为从圆指向右侧的水平方向（也就是通常平面直角坐标系的 x 轴方向），沿着逆时针方向到达的另一个半径，由这两个半径在圆心处相交构成的几何图形称为角，其大小称为这个角的度数，简称角度。两条半径分别称为这个角的两边。圆心也就是这两条边的夹角称为这个角的顶点。

历史上把整个一周的度数——就是初始半径沿着圆周绕一圈回到原处——称为 360° 。因此，沿着圆绕半圈——稍后我们会发现这个时候两条

半径在同一条直线上——就是 180° 。也就是说，一个角的两边在同一条直线的两侧，相当于是从中间点分别往直线两侧的方向上——其中任意一侧的延伸都称作射线，构成的角的大小为 180° 。这样的角的两边分别处于同一直线的两个相反的射线方向上的角也被称为平角：平表示在同一条直线上，是平的，没有拐弯。平角的一半被称为直角，大小为 90° 。比如说，时钟整点十二点的时候短针的方向——简称十二点方向，和时钟整点三点的时候短针的方向之间的夹角就是直角。

我们可以猜测，把三点方向（也就是通常平面直角坐标系的 x 轴方向）当作角度起点之后，随着半径的转动，角度可以取 $[0^\circ, 360^\circ]^2$ 之间的任何一个值。我们来不太严格地证明这一点。

以圆上三点方向（也就是通常平面直角坐标系的 x 轴方向）的点 P 为基础，从圆上任意选另一点 Q ，连接 OP ，连接 OQ ，连接 PQ 。我们得到一个角 $\angle POQ$ （角的标记我们一般采用从一条边到顶点到另一条边的方式，这里就是从 PO 到 O 再到 OQ ）——由两条半径交于圆心构成的角称作圆心角。同时，还有一条边 PQ ，以及圆上的一段弧 $\widehat{PQ}$ 。由于圆的周长是 $C = 2\pi r$ ³——其对应的圆心角是整个圆也就是 360° 和，我们也很容易看到一段圆弧的长度正比于其对应的圆心角——圆心角越大则圆弧越长。如果我们承认这样的一一对应关系，那么，圆心角的大小就和圆弧的长度建立了一一对应关系。同时，圆弧的长度和实数可以建立一一对应关系，只需要把圆弧拉直就成了数轴，而数轴上的点和实数存在一一对应关系。

我们甚至可以考虑用这个一一对应关系来重新标记角度的大小。例如，在固定 $r = 1$ 的圆——称为单位圆——上，长度为 2π 的弧对应着角度为 360° 的角，长度为 π 的弧——也就是圆的一半——称为半圆——对应着角度为 180° 的角，长度为 0 的弧对应着角度为 0° 的角。这样，我们似乎可以建立起来角度的新的度量方式。这个新的度量方式和用 $[0^\circ, 360^\circ]$ 来度量

²严格来说， 0° 和 360° 这两个值对应的角完全相同。为了保持一一对应，这里只能选了其中之一。但是，只要我们知道 0° 和 360° 对应着同一个角，我们还是同时保留这两个角度值。其含义上的区别是尽管角度是同一个，但是 360° 的角表示经历了一整圈的旋转回到起点半径，而 0° 表示没有经过任何旋转，完全停留在起点半径。

³这个常数 2π 可以通过圆的内接和外切正多边形的周长来不断近似得到，随着正 n 边形的 n 的增加。

角度的关系如下，

$$\theta = \theta^\circ \frac{2\pi}{360^\circ}. \quad (13.17)$$

我们把这个新的度量方式称为角的弧度，简称弧度。

未来我们还会用更严格的方式来证明角度可以取 $[0^\circ, 360^\circ]$ 之间的任何值，或者说 $[0, 2\pi]$ 之间的任何值。

为了未来使用方便，我们还想证明角的大小和角的度数之间具有一一对应关系。也就是相同度数的角都能相互重叠起来，任何一个角其大小对应着在 $[0^\circ, 360^\circ)$ 之间的唯一的数值⁴。后者的证明好办一些，对于任何一个以 x 轴正方向为角的起始边的圆心角，其角度对应着圆弧上的一个唯一的点，圆弧上的点和弧度一一对应，弧度和角度一一对应，因此任何角度都对应着一个唯一的角度值。对于前者，我们可以先证明给一个度数都能至少找到一个角：一个以 x 轴正方向为角的起始边，从圆心到以给定的度数对应的圆弧上的点的半径为终点边形成的圆心角。至于是否存在另一个不重合的这样的角，那就是另一个问题了。如果有另一个在单位圆上的角也具有相同的度数，我们总是可以通过旋转单位圆来使得我们前面构建的角和这个角重合——毕竟角度数值相同意味着圆弧长度相同。只要圆弧长度相同，通过旋转单位圆就可以使得两个圆心角重合。

其实，这一点也很容易从逆时针方程旋转圆心角的终点边会不断放大圆心角，直到到达 360° 之前，不会回到同一个数值，而得到角度的数值和角度大小是一一对应的直接感受。不过，数学知识不能仅仅依靠直觉来证明其成立。

原则上，论证角度重合，还需要角度的另一个性质：在固定了这个角的两边之后，两边的向外延长（其实向内延长，或者说反向延长也不改变角度，不过那个时候就会多出来几个角，我们暂时避开讨论其他角），不改变角度。这其实是我们前面能够仅仅用单位圆来讨论角度的原因——改变半径长度不影响角的大小。

⁴注意，这个时候为了唯一性，我们去掉了区间边界。

13.2.1 半圆、直径、平角和直角

现在我们来讨论圆的直径和半圆。按照 180° 是 360° 的一半，我们对应的圆弧也应该是整个圆的一半的长度。我们画一条通过圆心的直线，交圆 O 于 A, B 两个点。我们发现，沿着 AB 把圆折叠起来，圆的两侧完全重合。因此，这表示相应的圆弧的长度刚好就是整圆的一半，对应的圆心角刚好是 180° 。我们发现，这个时候的圆心角是 $\angle AOB$ ，其两边刚好在同一条直线上的不同射线方向上。因此，我们得到在同一直线的不同射线的方向上构成的角——也就是平角——的大小是 180° 。如果我们继续把直线沿着角的顶点折叠起来，会发现平角再一次被分成两半，也就是说平角的一半是 90° ——称为直角。我们把角的两边构成直角，也就是 90° 角的情况称为这两条边相互垂直。进而称这两条边所在的直线相互垂直。

我们甚至可以严格地来用坐标和计算的方式证明沿着直径折叠得到的圆的两侧完全重合。不失一般性，我们把圆心放到直角坐标系的原点（因为移动圆心不改变圆的形状，不影响折叠以后是否会重合的讨论），把直径直接就设成 x 轴方向的直径。于是圆和直径表示为，

$$\sqrt{x^2 + y^2} = r, \quad (13.18)$$

$$A = (-r, 0), B = (r, 0). \quad (13.19)$$

所谓折叠就是对于圆上的一个点 $P = (x_P, y_P \leq 0)$ ，使得其变成点 $P' = (x_{P'}, y_{P'})$ ， P' 是 P 关于直径 AB 的折叠点——未来会被称为轴对称点。严格来说，我们还得先定义轴对称才能完成这个计算。不过，如果我们稍微偷一下懒，我们说： P' 和 P 的关系是—— $x_{P'} = x_P, y_{P'} = -y_P$ 。几何上很容易看出来——这就是折叠上去的几何含义。

现在我们来证明，根据 $x_{P'} = x_P, y_{P'} = -y_P$ ，圆 O 上必然存在一个点 Q 使得 Q 和 P' 重合，也就是说， P' 在圆上，

$$\sqrt{x_{P'}^2 + y_{P'}^2} = \sqrt{x_P^2 + y_P^2} = r, \quad (13.20)$$

并且在其上半部分，

$$y_{P'} = -y_P \geq 0. \quad (13.21)$$

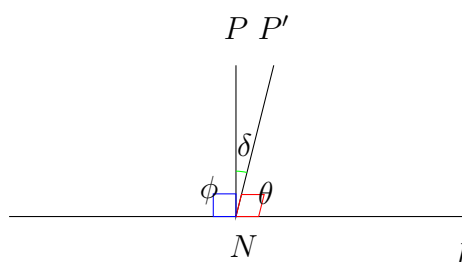


图 13.2: **线上多条垂线的假设**。假设过直线上一点有多条和这个直线垂直的直线。

于是，对于任何一个处于圆下半部分的点，折叠以后必然能够找到另一个圆上的上半部分的点，完全重合。反之亦然。因此，直径 AB 把这个圆分成完全相同的两个部分。

构成直角的两条直线被称为相互垂直。有了直角的概念我们来证明几个未来要用到的关于垂直的定理。

定理 13.4: 线上垂线唯一

线上垂线唯一] 过直线上一点有且只有一条这条直线的垂线。

证明 13.1. 我们用数学语言把待证明命题表述为：已知直线 l 和直线上一点 N ，要证明有且只有唯一一条直线 $PN \perp l$ 。

根据直角等于平角 (180°) 的一半，角度可以连续取值（之前已经通过圆弧长度证明描述角度大小的实数和角之间的一一对应关系），我们在 N 点旋转 NP 射线，总是可以找到等于其一半的角。因此，必然存在至少一条过线上一点的垂线。

现在我们来假设有两条（两条以上下面的证明也成立）不重合的过 N 的 l 的垂线，如图 13.2 所示。首先 $PN \cap P'N = N$ ，因此两者相交，不可能平行。其次，我们假设了两者不重合。因此， PN 和 $P'N$ 之间必然有一个不等于零的夹角（否则，夹角为零则重合）。于是 $\delta > 0$ 。因此，我们有 $\theta = 90^\circ, \phi = 90^\circ$ 。合起来，我们有 $\theta + \delta + \phi = 180^\circ + \delta > 0$ 。矛盾。因此原假设错误。过线上一点，没有两条以及两条以上的垂线。

定理 13.5: 线外垂线唯一

线外垂线唯一] 过直线外一点有且只有一条这条直线的垂线。

证明 13.2. 我们用数学语言把待证明命题表述为：已知直线 l 和直线外一点 P ，要证明有且只有一条直线 $PN \perp l$ 于 N 。

我们连接 P 和直线上任一点 M 。然后根据直线上一点有且只有一条垂线，就能得到 MP' 。如果 MP' 刚好过 P ，那就是要找的直线。因此，必然存在至少一条过线外一点的垂线。

现在我们来假设有两条（两条以上下面的证明也成立）不重合的过 P 的 l 的垂线， PM, PN ，如图 13.2 所示。

未来我们会直接从勾股定理证明三角形的内角和必定等于 180° 。在这里我们先用一下这个结论。注意，由于证明三角形内角和的时候没有用到过直线外一点有且只有一条垂线的结论。这里没有逻辑上的循环论证。

在三角形 $\triangle PMN$ 中， $\angle PMN = 90^\circ, \angle PNM = 90^\circ$ ，其内角和为 $\delta + \angle PMN + \angle PNM = \delta + 180^\circ > 180^\circ$ 。矛盾。因此原假设错误。过线外一点，没有两条以及两条以上的垂线。

这里我们用到了后面在 15.2 节中才会证明的一条定理——三角形的内角和是 180° 。不过那条定理的证明只需要勾股定理当作基础，不需要我们这里的“过直线外一点有且只有一条垂线”的结论，也不依赖于任何需要用这个结论来证明的其他命题。因此，这里没有逻辑上的循环论证。我们把这个证明放到后面完全就是因为这里我们还没有正式定义三角形，以及把内角和定理放到用正弦定理和余弦定理来证明三角形的性质的那部分（15.2 节）更加合适。

习题 13.2 (垂线平行). 证明垂直于同一直线的两条直线平行。

习题 13.3 (平行线之一垂直). 证明平行线之一垂直于第三条直线，则平行线中的另一条也垂直于第三条直线。（提示：反证法，平行传递性）

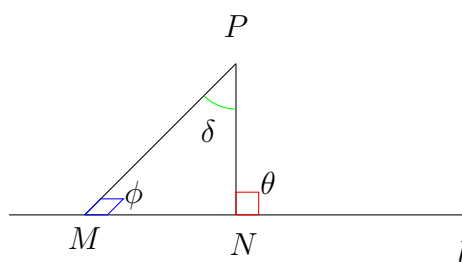


图 13.3: 线外多条垂线的假设。假设过直线外一点有多条和这个直线垂直的直线。

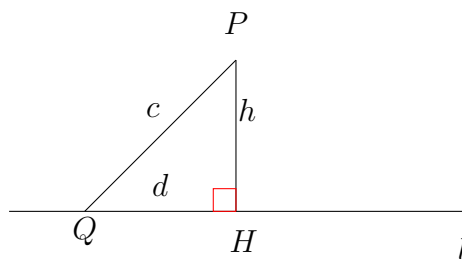


图 13.4: 垂线段距离最短。过直线外一点到直线的距离中垂线段最短。

定理 13.6: 线外垂线段最短

线外垂线段最短] 过直线外一点到这条直线的连线中，垂线段最短。

证明 13.3. 根据两点之间线段最短，我们只需要比较过线外一点到这条给定直线的所有直线段的长度。如图 13.4，根据勾股定理，对于不垂直的连线，其长度为 $c = \sqrt{h^2 + d^2}$ 。只要不垂直， $d \geq 0$ ，因此 $c \geq h$ 。

定理 13.7: 勾股定理的逆定理

勾股定理的逆定理] 如果三角形的三边长 a 、 b 、 c 有关系 $a^2 + b^2 = c^2$, 那么这个三角形是直角三角形。

证明 13.4. 我们先用数学的符号表述待证明的命题: 三角形 $\triangle ABC$ 满足 $AC^2 + BC^2 = AB^2$, 证明 $\angle ACB = 90^\circ$ 。

假设 $\angle ACB \neq 90^\circ$ 。过 A 做 $AH \perp BC$ 于 H , 则 H 和 C 不重合。否则, $\angle ACB = 90^\circ$ 。有两种可能 H 在 BC 之间, H 在 BC 之外。对于第一种情况, $BH < BC$, 三角形 $\triangle AHC$ 是直角三角形, 则 $AH^2 + HC^2 = AC^2$, 因此 $AH < AC$; $\triangle AHB$ 是直角三角形 $AB^2 = BH^2 + AH^2 < BC^2 + AC^2$, 勾股定理不成立。

针对第二种情况 $\angle ACB > 90^\circ$, 则 H 在 AC 之外也就是在射线 AC 上。根据勾股定理有: $AC^2 = AH^2 + CH^2$, $AB^2 = BH^2 + AH^2$, 我们来证明这两个等式和 $AC^2 + BC^2 = AB^2$ 冲突。把三式合并, 得到 $BC^2 + CH^2 + 2BC \cdot CH + AH^2 = (BC + CH)^2 + AH^2 = AB^2 = BC^2 + AC^2 = BC^2 + CH^2 + AH^2$, 也就是 $BC \cdot CH = 0$, 从而 $CH = 0$ 。 C 和 H 点重合, 矛盾。

也就是当 $\angle ACB \neq 90^\circ$ 的时候, 无论如何 $AB^2 = AC^2 + BC^2$ 不可能成立, 矛盾。因此, 原假设错误, $\angle ACB = 90^\circ$ 。

我们注意到在前面我们已经通过直角三角形定义了角度的正弦值和余弦值。那个时候, 我们是把角度本身的价值当作天上掉下来的知识的。今天, 我们利用单位圆来定义了角度的值。我们得保证这两种角度的概念和大小是相容的。回到角的定义和角度的定义, 这里基于圆的定义其实和之前的角和角度的定义是完全一样的。或者说, 这里是之前的模糊知识的精确化。至于通过直角三角形定义的角度的正弦值和余弦值, 只要我们在单位圆内构造一个相应的直角三角形——由 x 轴正方向也就是这个角的一条边一条半径, 这个角的另一条边也就是另一条半径, 以及这条半径到 x 轴正方向的高构成的直角三角形就是我们要找的直角三角形: 如图 13.5 所示, 在直角三角形 $Rt\triangle PHO$ 中, $PH = OP \sin \theta = \sin \theta$, $OH = OP \cos \theta = \cos \theta$ 。

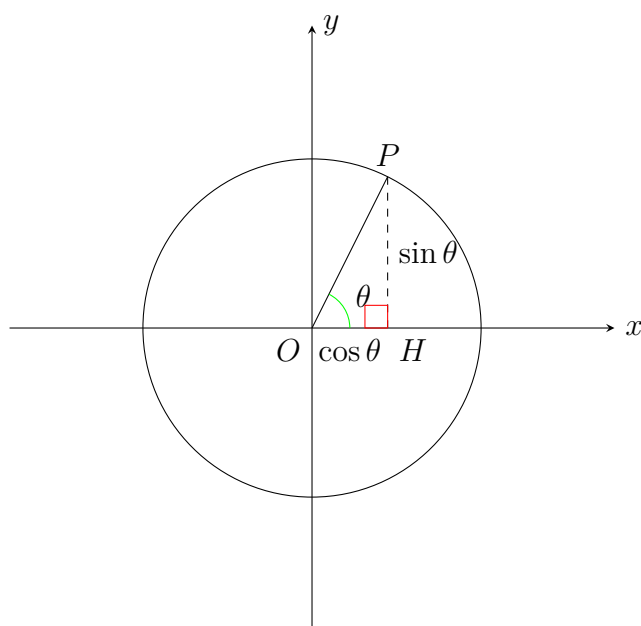


图 13.5: 角度的两种定义。基于直角三角形的和基于圆的角度定义可以统一起来。

13.2.2 对顶角、同位角、内错角和同旁内角以及它们之间的关系、关系的关系

有了角度和角的大小的一一对应关系，我们来证明几个关于角的重要定理。我们先给出一些关于角度关系的定义。首先，我们定义对顶角：两直线相交，一个角的两边（射线）正好是另一个角的两边（射线）的反向延长线，则称这两个角为对顶角。也就是说，两个角的两边刚好方向相反的角是对顶角。为什么称为对顶角？因为两个角共享一个顶点，同时方向正好相反，好像对起来的一对角。两直线相交于一点会产生四个角。其中会有两对对顶角。

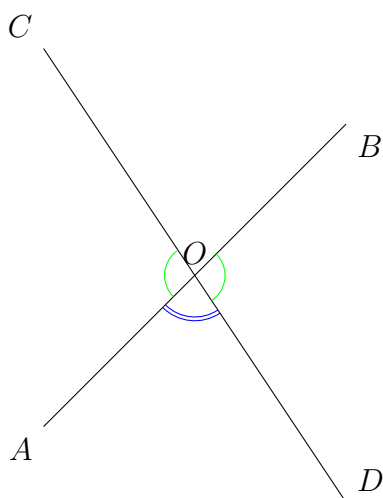


图 13.6: 对顶角相等。基于平角的定义可以证明对顶角相等。

定理 13.8: 对顶角相等

两条直线相交，对顶角大小相等。

证明 13.5. 我们用数学语言重新表述待证明命题：直线 AB 和直线 CD 相交于点 O ，则 $\angle AOC = \angle BOD$ 。

如图 13.6，因为 AB 是直线，因此 $\angle AOD + \angle DOB = 180^\circ$ 是平角；因为 CD 是直线，因此 $\angle AOD + \angle AOC = 180^\circ$ 是平角。因此， $\angle DOB = 180^\circ - \angle AOD = \angle AOC$ 。

其中两个角加起来等于平角的时候，我们称这两个角互补。

有了两条直线的关系——平行和相交，现在我们来查看一个更加复杂的情况——三条直线的关系：第一、其中两条平行，另一条也和其中一条平行（这个已经在平行公理中讨论过，三条相互平行）；第二、其中两条平行，另一条和其中一条相交（这个已经在平行公理中讨论过，另外两条也必须相交）；第三、其中两条相交，另一条和其中一条相交，和另一条也相交（如果和其中一条平行，一条相交，则就回到了前一种情况），也就是三条直线都相交。最后一种情况，我们区分一下相交于同一个点还是不同的点。只要相交于不同的点，则必然是三个交点，不能是两个交点。

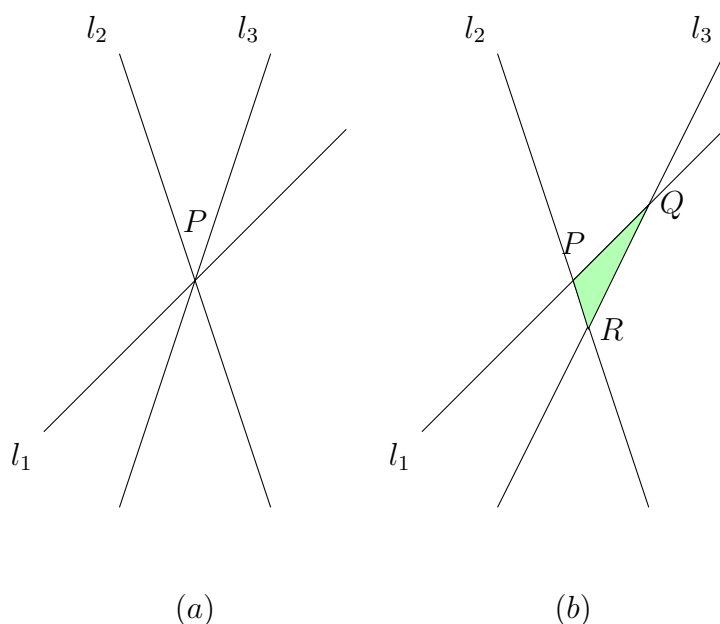


图 13.7: 三条线相交。不彼此重合的三条线只有一个或者三个交点。后者围成的封闭区域（图中阴影部分）是个三角形。

定理 13.9: 三条直线相交

三条直线相交（排除重合的情况），交点要么是一个要么是三个。

证明 13.6. 我们用数学语言重新表述待证明命题：如图 13.7 所示，三条两两不重合的直线 l_1, l_2, l_3 相交，交点要么只有一个 $P = l_1 \cap l_2 \cap l_3$ ，要么有三个不两两重合的 $P = l_1 \cap l_2$, $Q = l_2 \cap l_3$ 和 $R = l_3 \cap l_1$ 。不可能有两个，不可能有四个以及以上的交点。

我们假设只有两个交点 $P = l_1 \cap l_2$, $Q = l_2 \cap l_3$, $R = l_3 \cap l_1$ 。则， l_2 过 P, Q 两点， l_3 过 P, Q 两点。根据两点确定一条直线， l_2 和 l_3 完全重合。矛盾，因此，原假设错误。

我们假设有四个交点（多余四个的可以重复使用这个证明）。其中三个分别是两两直线的交点 $P = l_1 \cap l_2$, $Q = l_2 \cap l_3$, $R = l_3 \cap l_1$ 。我们还需要第四个交点，那这个交点是哪两条直线的交点呢？由于我们总共现在只有三条直线，这第四个交点只能是上面的三个已有组合之一的重复，例如 $S = l_1 \cap l_2$ 。则我们发现， $S = P$ 必然重合（两条不重合的直线只有一个或者零个交点，不能有两个，其根源仍然是两点确定一条直线）。矛盾，因此，原假设错误。

合起来，我们得到三条直线相交，交点要么是一个要么是三个。

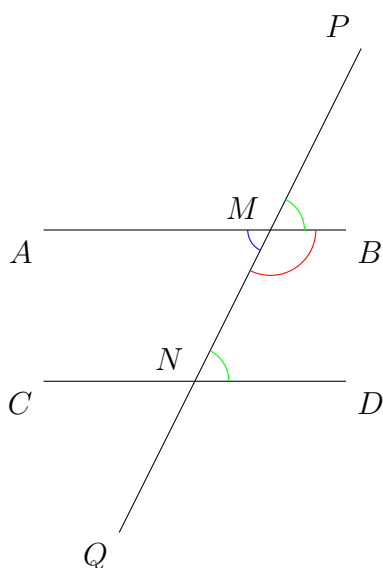


图 13.8: 平行线和第三条直线相交。通过和平行线相交的直线可以定义同位角、内错角和同旁内角。这个定义也可以推广到一般的三条线相交的情形。

其中三条直线相交于三个点的情况，我们给它一个专门的名字——三角形：三条直线相交于三个点，取其中的封闭图形加上其边界上的顶点和边，我们称为三角形。因此，三角形也可以看成是三条线段首尾相连构成的图形。如图 13.7 中的 PQR ，记作 $\triangle PQR$ ，读作三角形 PQR 。它有三条边 PQ, QR, RP 和三个角 $\angle PQR, \angle QRP, \angle RPQ$ 。

注意，在封闭区域的内部确实构成了三个角，因此才被命名为三角形。实际上，我们也可以命名其为它三边形。这三个角被称为这个三角形的内角。

第三种情况我们讨论完了，现在来回到第二种情况：三条直线中的两条相互平行 $AB \parallel CD$ ，另一条和这两条平行线相交 $PQ \cap AB = M, PQ \cap CD = N$ ，如图 13.8。我们把 $\angle PMB$ 和 $\angle MND$ 称为同位角（其字面含义是这两个角从构成其的两条边的角度来说，相对位置关系完全相同，都是把 PQ 的 QP 射线方向当成一条边，把 AB 的 AB 射线方向或者 CD 的 CD 射线方向当成另一条边）， $\angle AMN$ 和 $\angle MND$ 称为内错角（其字面含义是两条平行线内部的沿着相交线左右错开的两个角）， $\angle BMN$ 和 $\angle MND$ 称为同旁内角（其字面含义是两条平行线内部的沿着相交线左右同侧的两个

角)。

习题 13.4 (找同位角、内错角和同旁内角). 把一条直线和两条平行线相交的情况下的所有同位角、内错角和同旁内角列出来。

定理 13.10: 同位角、内错角和同旁内角的关系之间的关系

同位角、内错角和同旁内角的关系之间的关系] 如果同位角相等则内错角相等, 如果内错角相等则同旁内角互补, 如果同旁内角互补则同位角相等。

证明 13.7. 我们来用数学的符号表述要证明的命题: 如图 13.8, 三条直线中的两条相互平行 $AB \parallel CD$, 另一条和这两条平行线相交 $PQ \cap AB = M, PQ \cap CD = N$ 。如果 $\angle PMB = \angle MND$, 则 $\angle AMN = \angle MND$; 如果 $\angle AMN = \angle MND$, 则 $\angle BMN + \angle MND = 180^\circ$; 如果 $\angle BMN + \angle MND = 180^\circ$, 则 $\angle PMB = \angle MND$ 。

已知 $\angle PMB = \angle MND$, 又有 $\angle PMB = \angle AMN$ (对顶角相等), 因此 $\angle AMN = \angle MND$ (等式的性质——等号关系的传递性)。已知 $\angle AMN = \angle MND$, 又有 $\angle AMN + \angle BMN = 180^\circ$ (平角的定义), 因此 $\angle BMN + \angle MND = 180^\circ$ (等式的性质——等量替换)。已知 $\angle BMN + \angle MND = 180^\circ$, 又有 $\angle PMB + \angle MND = 180^\circ$ (平角的定义), 则 $\angle PMB = \angle MND$ (等式的性质)。

实际上, 我们发现上面的证明过程中, 完全没有用到“三条直线中的两条相互平行”——这一条在这里仅仅提供了同位角、内错角和同旁内角定义的场景。因此, 我们发现上面证明的定理, 以及同位角、内错角和同旁内角的定义可以推广到一般性的三条直线的关系上, 也就是三条直线相交于三点的情况。

因此, 当两条直线不平行, 但是第三条直线和这两条直线都相交的时候, 我们仍然可以定义同位角、内错角和同旁内角。

定义 13.1. 同位角 直线 AB 和直线 CD 分别和直线 PQ 相交于 M, N 。以交点 M 为顶点, MB 和 MP 为射线的角 $\angle PMB$ 和以交点 N 为顶点, ND (和 MB 方向协调, 所谓方向协调指的是尽管在不同的直线上但是都在直线 PQ 的同一侧) 和 NP (和 MP 方向相同) 为射线的角 $\angle DMP$,

称为同位角。

定义 13.2. 内错角直线 AB 和直线 CD 分别和直线 PQ 相交于 M, N 。以交点 M 为顶点, MQ 和 MA 为射线的角 $\angle AMQ$ 和以交点 N 为顶点, ND 和 NP 为射线的角 $\angle DNP$ ——两者被包含在直线 AB 和 CD 的内部, 分居于直线 PQ 的错开的两侧——称为内错角。

定义 13.3. 同旁内角直线 AB 和直线 CD 分别和直线 PQ 相交于 M, N 。以交点 M 为顶点, MB 和 MQ 为射线的角 $\angle BMQ$ 和以交点 N 为顶点, ND 和 NP 为射线的角 $\angle DNP$ ——两者被包含在直线 AB 和 CD 的内部, 居于直线 PQ 的同一侧——称为同旁内角。

13.2.3 直线和坐标轴相交的角度

有了直线和直线方程 $y = kx + b$ 的一一对应关系, 有了角和角度的大小的定义, 我们来看看两条直线相交的角的大小是否可以从直线方程的参数算出来。

为此, 我们先来研究直线 $y = kx + b$ 和坐标轴的夹角, 以后在 15.4 节中我们再来研究一般性的两条直线的夹角。首先, 当 $k = 0$ 的时候, 直线方程成为 $y = b$ 。这是平行于 x 轴的直线。稍后我们来证明这条直线和 y 轴垂直。当直线方程成为 $x = \frac{-b}{k} = x_0$ 为常数的时候, 直线是平行于 y 轴的直线。稍后我们来证明这条直线和 x 轴平行。其次, 对于一般的情况, 我们来计算这条直线和 x 轴、 y 轴的夹角。

我们把这条直线和 x, y 轴的交点 P, Q 先求出来,

$$0 = kx_P + b, y_Q = k \cdot 0 + b, \quad (13.22)$$

于是

$$x_P = \frac{-b}{k}, y_Q = b. \quad (13.23)$$

在直角三角形 $Rt\triangle QPO$ 中, 直线 l 和 x 轴的夹角 $\theta = \angle QPO$ 的三角函数值满足

$$\tan \theta = \frac{QO}{PO} = \frac{b}{\frac{-b}{k}} = k. \quad (13.24)$$

在 $[0^\circ, 90^\circ]$ 的范围内, $\tan \theta$ 的值和角度大小存在一一对应关系, 因此, 从这里就可以求出来直线 l 和 x 轴的夹角。 l 和 y 轴的夹角可以类似求出来。

习题 13.5 (交换直线表达式的 x, y). 替换直线表达式为 $x = k'y + b'$, 重复上面的分析。注意, 如果换回到原来的表达式则: $k' = \frac{1}{k}, b' = -\frac{b}{k}$ 。

特殊地, 当 $k = 0$ 的时候, 直线和 x 轴平行, $\theta = 0^\circ$; 当 $\frac{1}{k} = 0$ 的时候, 直线和 y 轴平行, $\theta = 90^\circ$, 和 x 轴垂直。

顺便, 从这个计算中, 我们发现, 直线 l 在 y 轴上的交点 (当 $x = 0$ 的时候) 是 $(0, b)$ 。这是为什么 b 被称为截距。我们还发现, 直线 l 和 x 轴的夹角的大小完全由 k 决定, k 越大这个夹角越大, 直线倾斜程度越高。因此, k 被称为 l 的斜率。

13.3 三角形和四边形的定义

有了点、线、角的定义, 我们就可以来三角形、四边形和多边形了。

前面我们已经在讨论三条直线的关系的时候定义了三角形。这里我们整理一下。

定义 13.4. 三角形三条直线相交于三个点, 取其中的封闭图形加上其边界上的顶点和边, 称为三角形。或者如果只看其边界, 三角形也可以定义为三条线段首尾相连构成的图形。

至于我们在使用中到底指的是其内部和边界, 其内部, 还是其边界, 一般来说在使用场景中是清楚的。定义上, 我们不再区分这三种情况。一般来说, 我们更多的时候指的是其边界。

类似地, 我们定义四边形

定义 13.5. 四边形四边形是四条线段首尾相连构成的凸图形。

这里的“凸”的字面含义是没有任何一块是“凹”进去的, 也就是其不相邻顶点之间的连线必须包含在其内部区域, 不会出现把其中的两个点连起来, 连线的一部分在这个四边形所封闭的图形的外部。未来, 我们还可以更准确地用代数式来表达这个含义。

为什么三角形不需要强调“凸”呢? 因为三角形本来就“凸”, 其根本就没有不相邻的顶点, 其内部连线必然处于其封闭区域内部。一般来说, 我

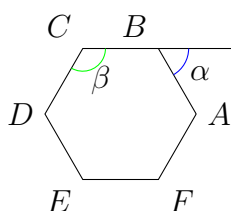


图 13.9: 多边形的内角和外角。 α 和 β 分别为多边形的内角和外角。

们研究几何图形都只研究凸图形⁵。

在四边形中，非相邻的顶点之间的连线被称为对角线，因为它们把两个对着的角——而不是本来就相邻于是相互连接——连接了起来。

类似地，我们定义一般的多边形—— n 边形对于任何一个给定的 n 。

定义 13.6. n 边形 n 边形是 n 条线段首尾相连构成的凸图形。

有的时候，我们也把 n 边形中连接不相邻的角，也就是不相邻的顶点的线，叫做对角线。

如图 13.9 所示，多边形中，以任何一个顶点为顶点，以这个顶点的两边为边，包含在这个多边形的内部的角，称为这个多边形的内角。每一个内角都有一个对顶角——反向延长构成这个内角的两条射线。我们知道其大小等于相应的内角。两条相交直线在同一个顶点上形成四个角。除了内角和其对顶角，我们把另两个角称为外角——由内角的一条边的射线方向和另外一条边的反向延长线的射线方向构成的角。这两个外角也是一对对顶角，大小相同。

有了内角我们可以定义内角和为每一个内角的度数相加。有了外角，我们可以定义多边形的外角和。注意，由于每一个外角都有一个相同大小的对顶角也称作外角，我们约定外角和只计算一次这样的每一对外角。通过在同一个顶点连接其他所有的不相邻顶点，可以把一个多边形切分为 $(n - 2)$ 个三角形。未来我们会证明，现在你也已经知道，平面几何中三角形的内角和是 180° 。因此，我们得到多边形的内角和为 $(n - 2) \times 180^\circ$ 。

习题 13.6 (多边形外角和). 基于多边形内角和算一算多边形的外角和。

⁵为什么这样呢？我也不知道。是不是非凸的图形更复杂，而且可以转化为凸的图形来研究？

13.4 推荐阅读材料

13.5 作业

13.6 本章小结

在这一章中，我们从数对——看作是平面直角坐标系上的点——定义了平面中的点，基于距离——也就是勾股定理——定义了直线和圆，基于圆定义了角，定义了平行和垂直。我们还证明了“两点确定一条直线”、“两点之间直线段最短”。

我们提到了，等未来有了微分和积分的概念，有了微分几何，我们就可以计算曲线长度，于是也可以用更加简单的方式来证明以上的命题。未来有了微积分我们还可以定义面积和计算面积大小。很多命题我们还可以从面积的角度来做更简单精巧的证明。

但是，今天，我们将以这样的比较复杂的方式，基于微积分之前的数学知识，从勾股定理来构建完整的平面几何知识体系。当前这个以勾股定理为基础的平面几何知识体系在一般的教材中并不常见。等我们走通了这个体系，我们也会回顾从其他的公理出发构建的平面几何知识体系，并且对这些知识体系做比较。那个时候，你会对整个平面几何学的知识体系有更清晰的认知。

这一章看起来只交代了几个定义，证明了一两个命题，还比较繁琐和抽象，但是，对于我们要构建的平面几何知识体系，这一章的内容是非常关键的。请你非常仔细地阅读本章，做到理解透彻。做不到，就多做几遍。或者，先明白逻辑结构，然后往下看，看清楚更大的逻辑结构，然后回来进一步理解本章的细节。

第十四章 平面几何：平行初步

在前一章，我们已经定义了线、线的相交和平行。在这一章我们来讨论线的相交和平行的判定和性质。任何一个几何对象，我们都要从定义、判定和性质三个角度来讨论。

一个对象的定义——通常表述为“满足条件 A 的被称为 B ”（记作 $A \Leftrightarrow B$ ）——既可以用于判定也是这个对象的性质。一个对象的判定——通常表述为“满足条件 A 的是（可以被判断为） B ”（记作 $A \Rightarrow B$ ）用于判断一个对象是 B 。一个对象的性质——通常表述为“ A 必定具有 B ”（记作 $A \Rightarrow B$ ）用于指定对象 A 的性质。后者往往更多地用于对象 A 的应用场合。

14.1 平行的定义、平行公理

两条直线的关系有平行、相交和重合。忽略重合，则就剩下平行和相交。相交有垂直相交和不垂直相交。在这一节，我们来研究平行的判定和性质。

平行的定义是两条直线不相交。这个定义如果不是写出来直线方程非常难以操作。有了直线方程，我们有“两条直线的方程的斜率 k 相同，截距 b 不同，则两直线平行”。当斜率和截距都相同的时候两直线重合。因此，按照前面忽略重合的说明，我们可以把两直线平行的判定简化为“斜率相同的直线相互平行”。不过，这个时候，我们为了判断两直线是否平行就必须线把直线方程写出来。

我们试着来找另一个不需要写出来直线方程的判断平行线的方法。我们先来证明

定理 14.1: 线外平行线唯一

线外平行线唯一] 过直线外一点有且只有一条直线和这条直线平行。

证明 14.1. 我们先把要证明的命题用数学语言描述一遍：已知直线 l (其方程为 $y = kx + b$) 和 l 外一点 $P = (x_P, y_P)$, 证明过 P 点有且只有一条直线和 l 平行。

我们先假设这样的直线 l' 存在, 来看看是不是唯一。根据 $l' \parallel l$, l' 的方程 $y = k'x + b'$ 必然满足 $k' = k$ (平行线的斜率相等)。根据 l' 过 P 点必然有 $y_P = k'x_P + b'$, 则 $b' = y_P - kx_P$ 。也就是说, 方程为 $y = kx + (y_P - kx_P)$ 的直线满足要求。

由于 k' 和 b' 的值唯一, 并且之前我们证明了直线和 (k', b') 的值完全一一对应, 我们就得到了直线存在, 且唯一。

我们再来证明平行的传递性。

定理 14.2: 平行传递

平行传递] 如果 $l_1 \parallel l_2$, $l_2 \parallel l_3$, 则 $l_1 \parallel l_3$ 。

证明 14.2. 同样, 我们忽略这三条直线之间的任意两条之间的重合关系。

根据直线和 (k, b) 的值完全一一对应, 由 $l_1 \parallel l_2$ 我们得到 $k_1 = k_2$, 由 $l_2 \parallel l_3$ 我们得到 $k_2 = k_3$, 因此根据等式的性质——相等关系可传递, 就得到 $k_1 = k_3$, 因此 $l_1 \parallel l_3$ 。

证明 14.3. 我们还可以通过前面已经证明的线外平行线唯一来证明平行的传递性。

我们假设 l_1 不平行于 l_3 , 则 l_1 和 l_3 必然存在一个交点 P , 也就是 $l_1 \cap l_3 = P$, 也就是 l_1 和 l_3 都经过 P 。 P 必然不在 l_2 上, 否则, l_1 不可能平行 l_2 。于是, 我们发现, 过 l_2 外一点 P 存在两条不重合的直线平行于 l_2 。矛盾。因此, 原假设错误, 也就是 $l_1 \parallel l_3$ 。

反过来成立吗？也就是我们能够从平行的传递性来证明线外平行的唯一性吗？我们来试试。

定理 14.3: 线外平行线唯一

线外平行线唯一] 过直线外一点有且只有一条直线和这条直线平行。

证明 14.4. 我们先把要证明的命题用数学语言描述一遍：已知直线 l (其方程为 $y = kx + b$) 和 l 外一点 $P = (x_P, y_P)$ ，证明过 P 点有且只有一条直线和 l 平行。

首先，过任意直线外一点肯定有至少一条平行线是已知的。否则，如果对某些直线就没有过线外某一点平行线，则也就不需要讨论平行的传递性的问题。也就是平行的传递性就假设了平行线的存在。因此，这里关键是证明只有一条。

我们假设过直线 l 外一点 P 点有至少两条直线 m, n 平行于 l ，也就是 $m \parallel l, n \parallel l$ 。则，根据平行的传递性 $m \parallel n$ ，也就是 m, n 不相交。这和已知的 $m \cap n = P$ (m, n 都经过 P 点) 矛盾。因此，原假设不成立，过直线 l 外一点 P 点最多有一条直线平行于 l 。

以上内容在一般的平面几何的书中被定位为平行公理。

14.2 平行基本定理

下面，我们来证明一个平行基本定理 (PBT) 及其逆定理。在未来讨论相似三角形的时候，这一组定理非常的重要。

定理 14.4: 平行基本定理

平行基本定理] 三角形内平行线切割线段成比例。

证明 14.5. 我们把待证明的命题用数学的符号表述为：在三角形 $\triangle ABC$ 内，过两边 AB, AC 各取一点 E, F ，若 $EF \parallel BC$ ，则所切割的线段成比例，即 $\frac{AE}{EB} = \frac{AF}{FC}$ 。

取平面直角坐标系，不失一般性，令 $A = (0, 0)$ ， $B = (x_B, 0)$ ， $C = (x_C, y_C)$ ($y_C \neq 0$ ，保证 AC 不在 AB 上)，选取 $E = (x_E, 0)$ 在 AB 上，选取 $F = (x_F, y_F = \frac{y_C}{x_C} x_F)$ 在 AC 上。

现在，我们来通过平行确定 x_F 。 $\frac{y_F - y_E}{x_F - x_E} = \frac{y_C - y_B}{x_C - x_B}$ ，得到 $x_F = \frac{x_C}{x_B} x_E$ 。有了这些坐标，我们来计算线段长度和计算比例。

$$AE = \sqrt{x_E^2} \quad (14.1)$$

$$EB = \sqrt{(x_B - x_E)^2} \quad (14.2)$$

$$\begin{aligned} AF &= \sqrt{x_F^2 + y_F^2}, \\ &= \sqrt{x_F^2 + \left(\frac{y_C}{x_C} x_F\right)^2}, \\ &= \sqrt{x_F^2 \left(1 + \frac{y_C^2}{x_C^2}\right)}, \\ &= \sqrt{\frac{x_E^2}{x_B^2}} \sqrt{x_C^2 + y_C^2} \end{aligned} \quad (14.3)$$

$$\begin{aligned} FC &= \sqrt{(x_C - x_F)^2 + (y_C - y_F)^2}, \\ &= \sqrt{\left(x_C - \frac{x_C}{x_B} x_E\right)^2 + \left(y_C - \frac{y_C}{x_C} \frac{x_C}{x_B} x_E\right)^2}, \\ &= \sqrt{\frac{(x_B - x_E)^2}{x_B^2}} \sqrt{x_C^2 + y_C^2}. \end{aligned} \quad (14.4)$$

于是，

$$\frac{AE}{EB} = \frac{\sqrt{x_E^2}}{\sqrt{(x_B - x_E)^2}} \quad (14.5)$$

$$\frac{AF}{FC} = \frac{\sqrt{\frac{x_E^2}{x_B^2}} \sqrt{x_C^2 + y_C^2}}{\sqrt{\frac{(x_B - x_E)^2}{x_B^2}} \sqrt{x_C^2 + y_C^2}} = \frac{\sqrt{x_E^2}}{\sqrt{(x_B - x_E)^2}}, \quad (14.6)$$

因此 $\frac{AE}{EB} = \frac{AF}{FC}$ 。

如果允许使用面积关系和三角形面积公式可以做如下的证明。

证明 14.6. 在三角形 $\triangle ABC$ 内, 已知两边 AB, AC 各取一点 E, F , $EF \parallel BC$, 我们有 $S_{\triangle EFB} = S_{\triangle FEC}$ (平行线间距离相等, 三角形面积公式)。因此, $\frac{S_{\triangle AEF}}{S_{\triangle EFB}} = \frac{S_{\triangle AEF}}{S_{\triangle FEC}}$ 。再根据三角形面积公式 $\frac{S_{\triangle AEF}}{S_{\triangle EFB}} = \frac{AE}{EB}$ (不同底等高的三角形), $\frac{S_{\triangle AEF}}{S_{\triangle FEC}} = \frac{AF}{FC}$ (不同底等高的三角形)。合起来, 得到 $\frac{AE}{EB} = \frac{AF}{FC}$ 。

我们之所以不优先列出来基于面积关系的证明是因为逻辑上, 三角形的面积公式需要先给出证明, 而这个证明严格来说依赖于距离公式 (勾股定理) 和微积分。当然, 这里我们先假设大家已经知道从正方形的面积公式——这个可以认为是定义——到长方形的面积公式再到三角形的面积公式这个逻辑链路。但是, 就算如此, 我们也没有从距离公式 (勾股定理) 先证明正方形面积公式。因此, 这里本质上不存在逻辑循环, 但是需要用到未来才能学习的微积分的知识。我们尽量避免以这样的论证为基础, 尽管在本书中我们也会提到或者列出来这样的论证。

此外, 这里还需要用到平行线间距离相等。这个我们会在后面讨论四边形的时候补充。逻辑上这个不是问题。在那里我们会看到, 不需要用到这些依赖于平行线间距离相等才能证明的命题。也就是不存在逻辑循环。

对于每一个定理, 我们都可以尝试去证明其逆定理。

定理 14.5: 基本平行定理的逆定理

基本平行定理的逆定理] 三角形内切割线段成比例的两条线是平行线。

证明 14.7. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 内, 过两边 AB, AC 各取一点 E, F , 如果所切割的线段成比例, 即 $\frac{AE}{EB} = \frac{AF}{FC}$, 则 $EF \parallel BC$ 。

我们假设 $EF \parallel BC$ 不成立, 我们在 AC 上找一点 F' 使得 $EF' \parallel BC$, 则根据平行基本定理, 有 $\frac{AE}{EB} = \frac{AF'}{F'C}$ 。 F' 不重合于 F , 必然在 F 更加靠近 A 或者 C 的一侧。不失一般性, 我们假设其在靠近 A 的一侧。则 $\frac{AF}{FB} = \frac{AF' + F'F}{F'C - F'F}$ 。根据已知条件, $\frac{AF' + F'F}{F'C - F'F} = \frac{AF}{FC} \frac{AE}{EB} = \frac{AF}{F'C}$ 。这个等式不可能成立, 除非 $FF' = 0$ 。这和 F 和 F' 不重合的假设矛盾。因此, 原假设错误, F' 和 F 重合, 而 $EF' \parallel BC$ 。因此, $EF \parallel BC$ 。

14.3 推荐阅读材料

14.4 作业

14.5 本章小结

在这一章中，我们定义了平行，我们从勾股定理开始证明了平行公理的两条表述及其等价性，证明了平行基本定理，以及平行基本定理的逆定理。这些定理在整个知识体系中非常重要。同时，平行基本定理的逆定理可以看作是，可以在不写下来明确的直线方程的情况下用于判断两直线平行的一个定理。

未来我们会进一步学习判断两直线平行的方式以及平行线的性质。那时候，平行的含义会更加清楚，平行这个概念也会更容易使用。未来你会看到，本书把平行部分分成两张，中间插入有关三角形的知识，是因为相似三角形的知识需要平行基本定理，而平行线的判定又需要有关三角形（尤其是三角形的内角和等于 180° ）的知识。

第十五章 平面几何：三角形

15.1 三角形定义和为什么研究三角形？

前面我们已经定义了三角形。我们的思路是从点、线、角、两条直线的关系——相交和平行的定义开始，来讨论三条直线的关系——三条都不相交、其中两条平行另一条和它们相交、以及三条都相交（其中包含相交于三点和相交于一点），而最后这个三条线相交于三点定义了三角形。

这是一条从简单对象、简单关系到复杂对象、复杂关系的道路。但是，在数学发展历史上，在现实世界中，人们并不是这样来提出和面对问题的。实际上，我们可能有一块多边形，甚至其中一部分出现曲线甚至非凸的不规则图形的土地和我们邻居的土地连在一起。我们需要大概估算一下这个土地的面积，这样好估算其产出，或者好用来和领域的土地做交换，或者某一年发了大水冲坏了边界需要重新划分而需要保证面积基本不变才能比较公平。无论如何，假设我们需要知道这个区域的面积，或者其他的属性，例如周长（比如用于估计建篱笆的投入）、内角和（这个我也暂时想不出来要干什么）。

于是，人们就发明各种方式来解决上面的问题。随着这些问题的解决，其中有一些解决方案就会成为能够重复使用的结论被人们记下来，沉淀下来。比如说，对由折线边界构成的不规则图形做切分，连起来所有的折线转折点也就是我们今天说的多边形的顶点，可以把图形转变为三角形。甚至，对于曲线构成的图形，只要我们把曲线切分的足够小，每一段都可以看作是折线。这样，我们仍然可以把曲线构成的不规则图形变成三角形。假设所

有的三角形的属性问题，例如三角形的面积、周长、内角和，对比两个三角形是否相同，等等问题，我们都可以解决，那我们就可以进一步用三角形的这些解决方案来解决更复杂的图形的问题。

于是，我们再来分解三角形，这里面又包含什么更简单的对象以及这些对象之间的关系。我们发现这里面由线段（直线）、角、顶点（直线交点）。于是，我们在构建知识体系的时候反过来走，先提出和演技简单的对象、简单的关系，然后再来研究更加复杂的对象和更加复杂的关系。

如果我们保证这个从简单到复杂的过程的每一步都是逻辑严密的——稍后我们会解释什么是逻辑严密，那么我们就说这样构建出来的知识体系具有系统性。在平面几何学的知识体系中，三角形具有非常重要的地位——稍微比点线角复杂一点，但是比其他图形都简单，并且是研究所有更复杂的图形的基础。

通过提出和面对问题，找到解决方案，梳理概念和命题，分解现象概念和命题来找到最基本的知识，再从最基本的知识构建知识体系，来得到那些应该有的概念和命题，顺便发现新的概念和命题，甚至提出新的假设公理，是一个非常普遍的知识运用和创造的过程。

15.2 三角形的内角和

在第十二章中，我们从勾股定理为基础定义了直角三角形的相似和直角三角形中的三角函数，然后进一步证明了一般三角形的余弦定理和正弦定理。现在，以此为基础，我们来讨论三角形的内角和。下一节我们同样用这部分知识来讨论三角形的全等。

那个时候，我们是把勾股定理当作是可以从面积关系、量纲分析、相似三角形来证明的定理。在几何学这部分，到目前为止，我们都是把勾股定理当作距离公式，也就是最基础的知识，也就是公理来使用的。尽管第十二章中的知识的逻辑结构不同，但是，其中所有的以勾股定理为基础所构建的知识，在这里仍然是适用的。因此，在这里，我们会按照几何学这部分的逻辑结构重新回顾一遍，但是不再重复那些基本内容相同的证明过程。

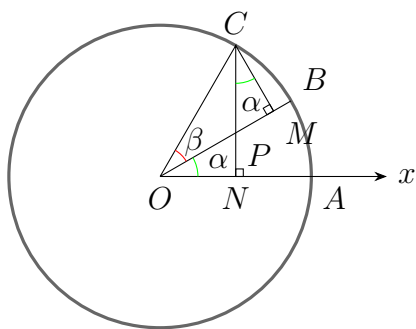


图 15.1: **和角三角函数**。图中的 ON 就是 $\cos(\alpha + \beta)$ 。通过计算出来 ON 的长度可以得到和角三角函数公式。

定理 15.1: 和角三角函数

和角三角函数] $\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$, $\sin(\alpha + \beta) = \sin \alpha \sin \beta + \cos \alpha \cos \beta$ 。

证明 15.1. 如图 15.1 所示, 在单位圆 O 中从 x 轴正方向开始 (记作 OA), 逆时针方向画一个 α 角, 终点边为 OB , 继续沿着逆时针方向画一个 β 角, 终点边为 OC 。过 C 作 $CM \perp OB$ 于 M , 作 $CN \perp OA$ 于 N 。 $CN = \cos(\alpha + \beta)$, $CM = \cos \beta$ 。于是, CN 交 OB 于 P 。则, $OP = \cos \beta - CM \tan \alpha$, $ON = OP \cos \alpha = (\cos \beta - \cos \beta \tan \alpha) \cos \alpha = \cos \alpha \cos \beta - \sin \alpha \sin \beta$ 。

其中我们用到了 $\angle PCM = \alpha$ 。这可以通过 $\angle PCN = 90^\circ - \angle CPM = 90^\circ - \angle OPN = \alpha$ 来得到。

另一半的证明留作练习。

我们下面来证明三角形的内角和是 180° 。我们已经可以从余弦定理得到每个角的大小和边长的关系, 我们来把这几个角加起来, 看看是否会等于 180° 。首先我们注意到三角形的每一个内角都小于 180° 。这只需要看到每一个内角都有一个外角, 而这个外角加上内角才是 180° , 同时外角肯定大于零 (等于零的时候三角形的两条边在同一直线上, 不能构成三角形)。有了这个观察, 我们只需证明 $\cos C = -\cos A + B$, 就可以得到 $A + B + C = 180^\circ$ 。因为在 $(0^\circ, 180^\circ)$ 的范围内, 余弦值和角度一一对应, 并且互补角的余弦值为相反数, 因此, 余弦值为相反数的两个角也必定互补。

定理 15.2: 三角形内角和

三角形内角和] 三角形内角和为 180° 。

证明 15.2. 我们用数学语言把待证明的命题表述为：已知 $\cos C = \frac{a^2+b^2-c^2}{2ab}$, $\cos B = \frac{a^2+c^2-b^2}{2ac}$, $\cos A = \frac{c^2+b^2-a^2}{2bc}$, $\sin A = \sqrt{1 - \cos^2 A}$, $\sin B = \sqrt{1 - \cos^2 B}$, 需要证明 $\cos C = -\cos(A+B)$ 。其中，我们把正弦函数也列出来是因为我们知道计算 $\cos(A+B)$ 的时候需要正弦值。

下面我们来通过计算证明上式成立。

$$\begin{aligned}
 \cos(A+B) &= \cos A \cos B - \sin A \sin B, \\
 &= \frac{c^2+b^2-a^2}{2bc} \frac{c^2+a^2-b^2}{2ac} \\
 &\quad - \sqrt{1 - \left(\frac{c^2+b^2-a^2}{2bc}\right)^2} \sqrt{1 - \left(\frac{c^2+a^2-b^2}{2ac}\right)^2}, \\
 &= \frac{c^2+b^2-a^2}{2bc} \frac{c^2+a^2-b^2}{2ac} \\
 &\quad - \frac{(b+c-a)(a+b-c)(a+c-b)(a+b+c)}{2abc}, \\
 &= -\frac{a^2+b^2-c^2}{2ab}, \\
 &= -\cos C.
 \end{aligned}$$

(15.1)

基于三角形内角和定理，很容易就能证明以下几条定理。证明部分当作练习。

定理 15.3: 三角形的内外角关系一

三角形的内外角关系一] 三角形的外角等于不相邻内角之和。

定理 15.4: 三角形的外角关系二

三角形的内外角关系二] 三角形的外角大于不相邻内角。

15.3 三角形的全等

我们把两个三角形认为全等，如果这两个三角形可以通过移动（平移、旋转、翻转）完全重合。用三角形中的对象来说，就是两个三角形的对应边的长度相等对应角的大小相同，则称这两个三角形全等。我们来看一看，实际判断两个三角形是否全等是不是真的需要同时判断这六组等式。三角形全等的判定在未来三角形的相似、四边形、圆的性质的学习中都是非常重要的。

定理 15.5: 三角形全等 SSS 判定

三角形全等 SSS 判定] 两个三角形的三条边对应相等则两个三角形全等。

证明 15.3. 我们把待证明的命题用数学的符号表述为：在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中， $AB = A'B'$, $BC = B'C'$, $CA = C'A'$ ，则这两个三角形全等，记作 $\triangle ABC \cong \triangle A'B'C'$ 。

按照余弦定理，我们可以计算出来三角形的任何一个角，例如，

$$\cos C = \frac{a^2 + b^2 - c^2}{2ab} = \frac{(a')^2 + (b')^2 - (c')^2}{2a'b'} = \cos C'. \quad (15.2)$$

然后，再根据在 $(0, 180^\circ)$ 范围内，余弦函数的值和角度一一对应，就得到 $C = C'$ 。同理可得每一个角都对应相等。

因此，这两个三角形满足所有对应边相等对应角相等的条件，属于两个全等三角形。

定理 15.6: 三角形全等 AAS 判定

三角形全等 AAS 判定] 两个三角形的两个角对应相等, 这两个角的一组对边也对应相等则两个三角形全等。

证明 15.4. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $\angle CAB = \angle C'A'B'$, $\angle BCA = \angle B'C'A'$, $BC = B'C'$, 则这两个三角形全等, 记作 $\triangle ABC \cong \triangle A'B'C'$ 。

根据三角形内角和定理我们可以通过两个对应角相等得到第三个角也对应相等, $\angle BCA = \angle B'C'A'$ 。按照正弦定理, 我们可以通过两个角和一条边算出来另一条边, 也就是

$$b = \frac{a}{\sin A} \sin B = \frac{a'}{\sin A'} \sin B' = b', \quad (15.3)$$

$$c = \frac{a}{\sin A} \sin C = \frac{a'}{\sin A'} \sin C' = c'. \quad (15.4)$$

因此, 这两个三角形满足所有对应边相等对应角相等的条件, 属于两个全等三角形。

同时我们也就证明了全等三角形 ASA 判定定理。

定理 15.7: 三角形全等 ASA 判定

三角形全等 ASA 判定] 两个三角形的两个角对应相等, 这两个角的公共边也相等则两个三角形全等。

定理 15.8: 三角形全等 SAS 判定

三角形全等 SAS 判定] 两个三角形的两条边对应相等, 这两条边的夹角也相等则两个三角形全等。

证明 15.5. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $AB = A'B'$, $BC = B'C'$, $\angle ABC = \angle A'B'C'$, 则这两个三角形全等, 记作 $\triangle ABC \cong \triangle A'B'C'$ 。

根据余弦定理, 我们可以计算出来 $\angle ABC$ 的对边 $AC = b$,

$$b = \sqrt{a^2 + c^2 - 2ac \cos B} = \sqrt{(a')^2 + (c')^2 - 2(a')(c') \cos B'} = b'. \quad (15.5)$$

因此, 这两个三角形的三条边都对应相等, 这两个三角形全等 (三角形判定的 SSS 定理)。

但是, 我们没有“三角形全等 ASS 判定——两个三角形的两条边对应相等, 这两条边的其中一组对角也相等则两个三角形全等”。为什么呢?

例 15.1 (有三角形全等 ASS 判定定理吗?). 两个三角形的两条边对应相等, 这两条边的一个对角也相等, 这两个三角形全等吗?

证明 15.6. 我们把待检验的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $AB = A'B'$, $BC = B'C'$, $\angle CAB = \angle C'A'B'$, 问这两个三角形全等吗?

这时候基于余弦定理我们有

$$b^2 + c^2 - a^2 = 2bc \cos A. \quad (15.6)$$

这里 $a, c, \angle A$ 已知。这是一个一元二次方程。这个方程可以是无解 (当 $\Delta = (2c \cos A)^2 - 4(c^2 - a^2) = 4(a^2 - c^2 \sin^2 A) < 0$), 两个解 (当 $\Delta > 0$) 和一个解 (等 $\Delta = 0$)。当方程有两个解的时候, 我们不能保证 $b = b'$ 。

根据正弦定理, 我们可以计算出来 $\angle BCA$,

$$\sin C = \frac{\sin A}{a} c = \frac{\sin A'}{a'} c' = \sin C'. \quad (15.7)$$

但是, $(0^\circ, 180^\circ)$ 之间的正弦值和角的大小之间的对应关系不是一一的, 同

一个正弦值会有两个可能的角度—— $C = C'$ ，或者 $C = 180^\circ - C'$ ，除非那个角 C 正好是 90° 才会唯一。

我们可以发现，当 $\Delta = 0$ 的时候 $a = c \sin A$ ，正好要求 $\angle C = 90^\circ$ 。因此，从正弦定理和余弦定理得到的结论相同：这两个三角形有可能全等，也有可能不全等，除非刚好其中一个角是直角。

我们也可以从构成三角形的条件来看这个结果。我们在已经给定大小的 $\angle A$ 的两条边的其中一条边的射线上确定点 B 使得 $AB = c$ 长度等于已知值。三角形 $\triangle ABC$ 中只有 C 点没有确定了。我们尝试来确定它。我们发现，使得射线 BC 能够到达射线 AC 上，从而确定 C 点，至少 $a \geq c \sin A$ ，因为那个时候 B 到 AC 正好是直线外一点到直线的垂直距离，而垂直距离最短。当 $a > c \sin A$ 的时候有两个交点，因此会出现两个可能的三角形。于是，我们就不能保证 $\triangle ABC$ 和 $\triangle A'B'C'$ 全等。只有当 $a = c \sin A$ ， $\angle C$ 刚好成为直角，只有一个这样的三角形，因此 $\triangle ABC$ 和 $\triangle A'B'C'$ 全等。

于是，我们也顺便证明了直角三角形全等判定的特殊定理——HL 定理。

定理 15.9: 直角三角形全等 HL 判定

直角三角形全等 HL 判定] 两个直角三角形的一条直角边和一条斜边对应相等，则两个三角形全等。

其实对于直角三角形，我们知道其任意两边对应相等，这两个三角形都全等。因为我们根据其中的两边都可以求出来第三边的长度，因此，可以证明第三边也对应相等。因此，根据一般三角形的全等判定 SSS 定理，得到这两个三角形全等。

我们也可以对其中一部分定理的证明采用纯几何的形式，而不是都从勾股定理和正弦定理、余弦定理出发。

15.4 两条直线相交的夹角

现在，我们终于可以来研究一般性的两条直线的夹角了。

如图 15.2 所示，假设我们有两条直线 $l_1: y = k_1x + b_1$ 和 $l_2: y = k_2x + b_2$

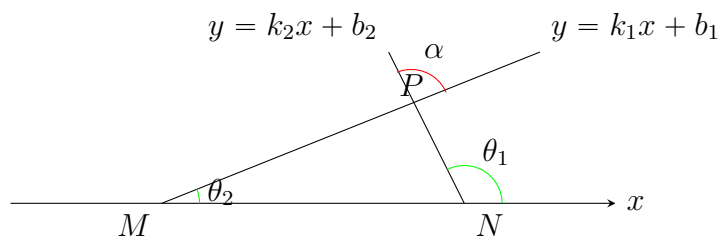


图 15.2: 两直线的夹角。通过“外角等于不相邻内角之和”可以得到两直线夹角为 $\alpha = \theta_1 - \theta_2$ 。

相交于点 P ($k_1 \neq k_2$, 否则两直线不相交), 构成了四个夹角, 其中两对对顶角, 并且从这两对中分别取一个角, 则这两个角加起来等于平角。因此, 对于两直线相交, 我们只需要算出来四个角其中的那一个就可以。习惯上, 我们选择那个小于等于 90° 的角来计算, 当作两条直线的夹角。

P 点的坐标前面已经通过求解二元一次方程算过,

$$x_P = \frac{b_2 - b_1}{k_1 - k_2}, y_P = \frac{k_1 b_2 - k_2 b_1}{k_1 - k_2}. \quad (15.8)$$

l_1 和 x 轴的交点 M 的坐标为

$$x_M = \frac{-b_1}{k_1}, y_M = 0. \quad (15.9)$$

l_2 和 x 轴的交点 N 的坐标为

$$x_N = \frac{-b_2}{k_2}, y_N = 0. \quad (15.10)$$

我们暂时忽略 x_M 或者 x_N 不存在的特殊情况。实际上, 通过把直线方程表示为 $x = k'y + b'$ 我们就可以讨论这样的情况 (相应地, 在那里, 直线和 y 轴交点不存在的属于特殊情况)。因此, 我们的证明完全可以形式上交换 x, y 之后重新做一次, 就能覆盖所有的情况。

在三角形 $\triangle PMN$ 中, l_1 和 l_2 的夹角 α 满足

$$\alpha = \theta_1 - \theta_2, \quad (15.11)$$

其中 θ_1, θ_2 分别是 l_1 和 l_2 和 x 轴的夹角。因此,

$$\tan \alpha = \tan (\theta_1 - \theta_2) = \frac{\sin (\theta_1 - \theta_2)}{\cos (\theta_1 - \theta_2)}, \quad (15.12)$$

而

$$\sin(\theta_1 - \theta_2) = \sin(\theta_1)\cos(\theta_2) - \cos(\theta_1)\sin(\theta_2), \quad (15.13)$$

$$\cos(\theta_1 - \theta_2) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2). \quad (15.14)$$

因此,

$$\tan \alpha = \frac{\tan \theta_1 - \tan \theta_2}{1 + \tan \theta_1 \tan \theta_2} = \frac{k_1 - k_2}{1 + k_1 k_2}. \quad (15.15)$$

考虑几个特殊情况: 当 $k_1 = k_2$ 的时候, $\tan \alpha = 0$ 表示两条直线平行或者重合; 当 $k_1 k_2 = -1$ 的时候, 分布为零, 分子有限, 结合 $\tan$ 函数的性质, 我们发现, 因此 $\alpha = 90^\circ$, 也就是两条直线相互垂直。

我们发现, 两条直线之间的夹角仅仅和两条直线的斜率有关系和截距没有关系。同时, 这个 $k_1 k_2 = -1$ 的时候两条直线相互垂直的结论也非常有用。例如从给定的一条直线可以直接求出来和这条直线垂直的直线。

例 15.2 (求已知直线的垂线). 已知直线 l 的方程式 $y = kx + b$ 和直线上一点 $P = (1, k + b)$ 和直线外一点 $Q = (-1, b)$ ($k \neq 0$ 保证了这个点在 l 之外), 求过 P, Q 垂直于 l 的直线。

假设过 P 垂直于 l 的直线的方程为 $y = k'x + b'$, 则 $k' = -\frac{1}{k}$, 因此 $k + b = -\frac{1}{k} + b'$, 得到 $b' = k + b + \frac{1}{k}$ 。

假设过 Q 垂直于 l 的直线的方程为 $y = k'x + b'$, 则 $k' = -\frac{1}{k}$, 因此 $b = -\frac{1}{k}(-1) + b'$, 得到 $b' = k + b - \frac{1}{k}$ 。

两条直线都唯一。这正好体现了我们前面得到的一般结论: 过直线上和直线外一点有且只有一条直线垂直于原直线。其背后是一元一次方程的解的存在性和唯一性。

有了这个从斜率到角度大小的一一对应关系 (在 $[0^\circ, 90^\circ]$ 范围内), 我们就可以用直线方程讨论任何直线相关的几何问题了: 从直线方程我们可以得到距离、直线平行、直线相交以及相交的角度。

15.5 三角形的相似

三角形相似定义为三组对应角分别相等, 每一组对应边成比例的三角形, 也就是,

定义 15.1. 三角形相似 [三角形相似] 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $\frac{AB}{A'B'} = \frac{BC}{B'C'} = \frac{CA}{C'A'}$, $\angle ABC = \angle A'B'C'$, $\angle BCA = \angle B'C'A'$, $\angle CAB = \angle C'A'B'$, 则称这两个三角形相似。

在未来几何学知识中也有重要地位。甚至, 我们的三角函数的定义等几何学知识就是从观察到直角三角形的相似性开始的: 当把其中两条边缩放 k 倍的时候, 另外一条边必须同时被缩放 k 倍, 因为 $a^2 + b^2 = c^2 \Rightarrow (ka)^2 + (kb)^2 = (kc)^2$ 。

我们先尝试从勾股定理来证明三角形相似的判定定理, 之后从前面已经证明的尤其是关于三角形全等和后面将要证明的平行线有关的定理来得到相似三角形的判定定理。这样从不同的路径来建立同一个几何学知识有助于我们未来看到几何知识的不同的系统。

定理 15.10: 三角形相似 AA 判定

三角形相似 AA 判定] 两个三角形的两个角对应相等则两个三角形相似。

证明 15.7. 我们把待证明的命题用数学的符号表述为：在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中， $\angle ABC = \angle A'B'C'$, $\angle BCA = \angle B'C'A'$ ，则这两个三角形相似，记作 $\triangle ABC \sim \triangle A'B'C'$ 。

根据三角形内角和定理，我们可以得到另一个组角也相等，也就是 $\angle CAB = \angle C'A'B'$ 。根据正弦定理，我们有，

$$\frac{b}{\sin \angle ABC} = \frac{c}{\sin \angle BCA} = \frac{a}{\sin \angle CAB} \quad (15.16)$$

则

$$\frac{a}{b} = \frac{\sin \angle CAB}{\sin \angle ABC}, \frac{b}{c} = \frac{\sin \angle ABC}{\sin \angle BCA}. \quad (15.17)$$

同理

$$\frac{a'}{b'} = \frac{\sin \angle C'A'B'}{\sin \angle A'B'C'}, \frac{b'}{c'} = \frac{\sin \angle A'B'C'}{\sin \angle B'C'A'}. \quad (15.18)$$

根据 $\angle ABC = \angle A'B'C'$, $\angle BCA = \angle B'C'A'$, $\angle CAB = \angle C'A'B'$ ，有

$$\frac{a}{b} = \frac{a'}{b'}, \frac{b}{c} = \frac{b'}{c'}. \quad (15.19)$$

于是

$$\frac{a}{a'} = \frac{b}{b'} = \frac{c}{c'}. \quad (15.20)$$

因此，这两个三角形的三条边都对应成比例三个角对应相等，这两个三角形相似（三角形相似的定义）。

定理 15.11: 三角形相似 SSS 判定

三角形相似 SSS 判定] 两个三角形的三条边对应成比例, 则这两个三角形相似。

证明 15.8. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $\frac{a}{a'} = \frac{b}{b'} = \frac{c}{c'}$, 则这两个三角形相似, 记作 $\triangle ABC \sim \triangle A'B'C'$ 。

根据余弦定理, 我们有,

$$\cos C = \frac{a^2 + b^2 - c^2}{2ab}. \quad (15.21)$$

根据 $\frac{a}{a'} = \frac{b}{b'} = \frac{c}{c'} = k$, $a = a'k, b = b'k, c = c'k$, 有

$$\begin{aligned} \cos C &= \frac{a^2 + b^2 - c^2}{2ab} = \frac{(ka')^2 + (kb')^2 - (kc)^2}{2ka'kb'}, \\ &= \frac{(a')^2 + (b')^2 - (c')^2}{2a'b'} = \cos C'. \end{aligned} \quad (15.22)$$

根据 $[0^\circ, 180^\circ)$ 范围内, 余弦值和角度一一对应, 得到 $C = C'$ 。同理 $B = B', A = A'$ 。

因此, 这两个三角形的三条边都对应成比例三个角对应相等, 这两个三角形相似 (三角形相似的定义)。

定理 15.12: 三角形相似 SAS 判定

三角形相似 SAS 判定] 两个三角形的两条边对应成比例, 这两条边的夹角对应相等, 则这两个三角形相似。

证明 15.9. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $\frac{AB}{A'B'} = \frac{AC}{A'C'}$, $\angle CAB = \angle C'A'B'$, 则这两个三角形相似, 记作 $\triangle ABC \sim \triangle A'B'C'$ 。

根据 $\triangle ABC$ 中的余弦定理, 我们有,

$$\begin{aligned} BC &= a = \sqrt{b^2 + c^2 - 2bc \cos A}, \\ &= \sqrt{(BC)^2 + (AB)^2 - 2(AB)(BC) \cos \angle CAB}, \\ &= \sqrt{k^2(B'C')^2 + k^2(A'B')^2 - 2k^2(A'B')(B'C') \cos \angle C'A'B'}, \end{aligned} \quad (15.23)$$

最后一步用了 $\frac{AB}{A'B'} = \frac{AC}{A'C'} = k$, $\angle CAB = \angle C'A'B'$ 。根据 $\triangle A'B'C'$ 中的余弦定理, 我们有

$$\begin{aligned} B'C' &= a' = \sqrt{(b')^2 + (c')^2 - 2b'c' \cos \angle C'A'B'}, \\ &= \sqrt{(B'C')^2 + (A'B')^2 - 2(A'B')(B'C') \cos \angle C'A'B'}, \end{aligned} \quad (15.24)$$

因此, $\frac{BC}{B'C'} = k = \frac{AB}{A'B'} = \frac{CA}{C'A'}$ 。因此, 这两个三角形的三条边都对应成比例, 于是这两个三角形相似 (三角形相似的 SSS 判定)。

定理 15.13: 相似三角形 HL 判定

相似三角形 HL 判定定理] 两个直角三角形的一条直角边和一条斜边对应成比例，则这两个三角形相似。

证明 15.10. 我们用数学语言先表述一下待证明命题：已知直角三角形 $Rt\triangle ABC$ 和 $Rt\triangle A'B'C'$ ，角 $\angle BCA = 90^\circ = \angle B'C'A'$ ，已知 $c' = \lambda c, a' = \lambda a$ ，求证 $\triangle ABC \sim \triangle A'B'C'$ 。

根据勾股定理，两个三角形的另一边的长度可以计算出来，分别记作 $b' = \sqrt{(c')^2 - (a')^2}, b = \sqrt{c^2 - a^2}$ ，已知 $c' = \lambda c, a' = \lambda a$ ，有 $b' = \sqrt{(c')^2 - (a')^2} = \lambda \sqrt{c^2 - a^2} = \lambda b$ ，因此，三边对应成比例。根据相似三角形判定定理 SSS 得到两三角形相似。

15.6 相似三角形的应用——证明勾股定理

在本书的知识体系中，相似三角形的判定定理是基于距离公式也就是勾股定理的。因此，我们不需要在从相似三角形来证明勾股定理。不过，第一这个证明很漂亮，第二，未来在其他平面几何学的知识体系中，有可能勾股定理或者说距离公式并不是最基础的公理，最基础的知识。因此，我们这里还是列出来这个证明。

定理 15.14: 从相似三角形证明勾股定理

从相似三角形证明勾股定理] 从相似三角形的判定定理和性质证明直角三角形的三边满足 $a^2 + b^2 = c^2$ (两直角边的平方和等于斜边的平方)。

证明 15.11. 我们先用数学的符号表述待证明的命题: 在直角三角形 $Rt\triangle ABC$ 中, 角 C 为直角, 证明 $a^2 + b^2 = c^2$, 要求证明过程只能运用相似三角形的判定定理和性质。

过 C 做 $CD \perp AB$ 于 D 。在三角形 $\triangle ACD$ 和 $\triangle ABC$ 中, $\angle ADC = 90^\circ = \angle ACB$, $\angle DAC = \angle CAB$ (三角形内角和定理), 于是 $\triangle ACD \sim \triangle ABC$ (相似三角形 AA 判定定理), 于是 $\frac{AC}{AD} = \frac{AB}{AC}$, 也就是 $AC^2 = AB \times AD$ 。同理, 三角形 $\triangle BCD \sim \triangle BAC$, 于是 $\frac{BC}{BD} = \frac{BA}{BC}$, 也就是 $BC^2 = AB \times BD$ 。合起来, 有 $AC^2 + BC^2 = AB \times BD + AB \times AD = AB \times (AD + BD) = AB^2$, 也就是 $a^2 + b^2 = c^2$ 。

将来在讨论圆的时候, 我们会在很多地方用到相似三角形。实际上, 相似的概念是平面几何学的核心, 否则, 我们就不能在草稿纸上研究几何图形, 然后把研究出来的结果用于现实了。能够这样去用就是因为相似——我们考虑的是相似变换下保持不变的那些性质。这些性质可以将来从纸面上迁移到现实中, 或者其他地方。

15.7 全等三角形的应用——尺规作图

尺规作图, 就是用如图 15.3 所示的圆规 (只能用于画圆, 也就是找等距离的点或者说辅助等长的线段) 和不带刻度的直尺 (只能用于画直线) 来画几何图形, 在平面几何中有独特的地位。很多时候, 我们证明几何题的时候, 需要添加辅助线, 例如给定直线外过一点这条直线的平行线或者垂线, 做一个给定角的角平分线, 找给定线段的中点。之前, 我们都是假设这样的几何对象存在, 就继续去做证明了。实际上, 更加严格的要求是其存在性要被证明出来, 最好这个几个对象也通过画图展示出来。

为什么要去掉刻度呢? 比如说用量角器上的角的刻度, 用直尺上带的线的刻度? 原因是, 只要用上了刻度信息, 实际上就带来了不准确——因为

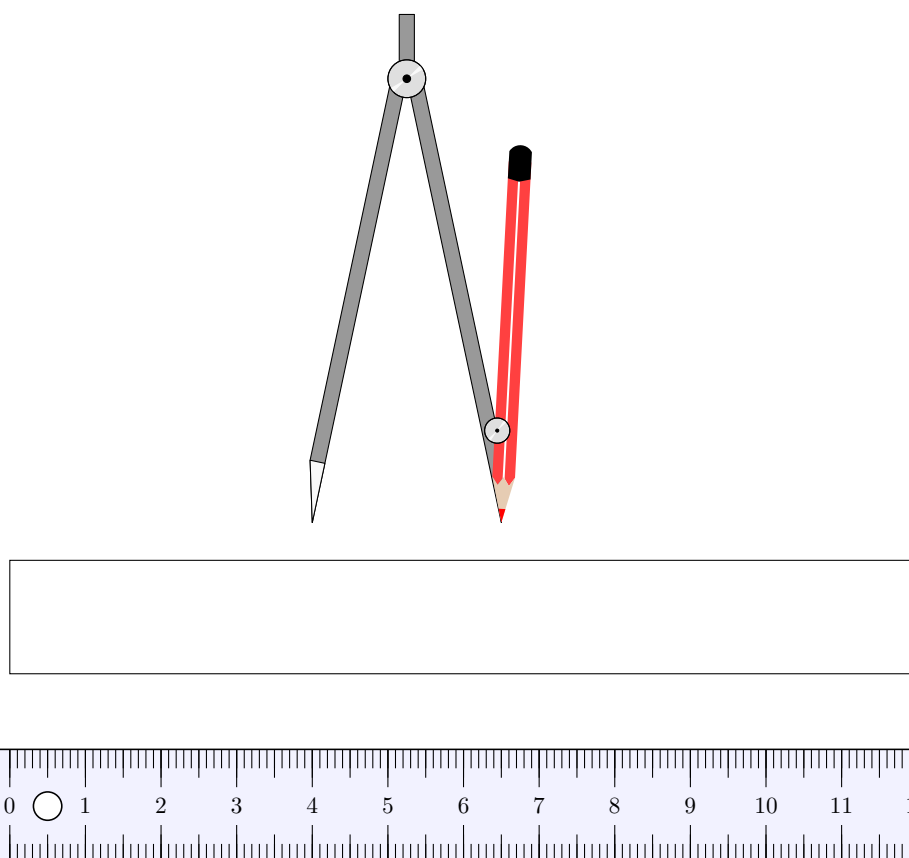


图 15.3: 直尺和圆规。圆规和不带刻度的直尺。如果你的直尺带了刻度你不去使用这个刻度就好。

你总是要对着有限精度的刻度来确定边的长度和角的大小，而几何从原则上就要避免这个有限精度带来的不准确。当然，一会儿你会看到在真实场景中，用圆规也会遇到“有限精度”，但是，在那里原则上是可以无限精确的。如果你的直尺带了刻度你不去使用这个刻度就好。

当然，回到证明的场景，只要我们能够先确认所有的辅助线等这样的我们证明过程需要的几何对象存在就够了。因此，我们主要还是关注这些对象的存在性的证明。从这个角度来说，尺规作图正好就是一种存在性证明——只要按照给定的步骤去做就能得到想要的几何对象，而且这个给定的步骤是经过严格的几何证明的。

下面，我们就来看几个几何对象的尺规作图的过程及其证明——这样做确实能够得到想要的对象。

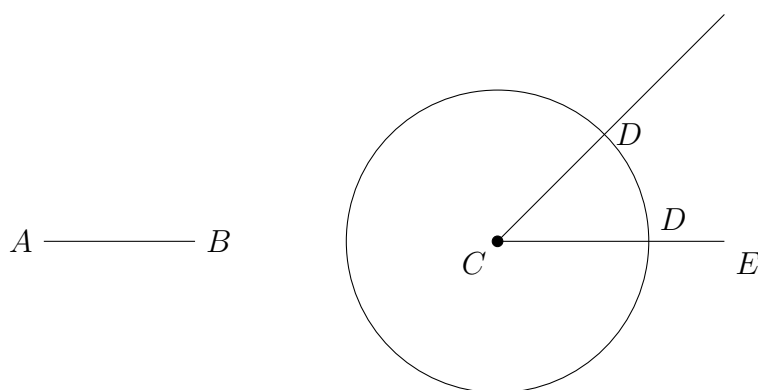


图 15.4: 尺规作图复制线段。用圆规和不带刻度的直尺可以复制线段。

首先, 我们来尝试复制一段给定的线段。已知线段 AB 和另一个点 C , 我们想绘制一段线段 CD 使得 $CD = AB$ 。我们选择圆规的一个脚的端点放在 A 另一个脚的端点放在 B , 然后移动用于固定的脚 (金属质地的坚硬的) 放置于 C , 让另外那只带有笔的脚尖画一个圆, 如图 15.4 所示。根据圆规的性质, 我们直到在这个圆上任何一点到 C 的距离都相等, 并且等于 AB 。最后选择圆上任何一个点 D , 把 C 和 D 点用直尺相连, 就得到满足要求的线段 CD 。如果要求 CD 在一个已经给定的直线 CE 上, 则只需要选择 D 为直线 CE 和所做出来的圆的交点。我们在解释这个做图过程的时候已经证明确实 $CD = AB$ 。

接着, 我们来复制一个给定角。已知角 $\angle ABC$ 和另一个点 B' , 我们希望绘制另一个角 $\angle A'B'C'$, 使得 $\angle A'B'C' = \angle ABC$ 。如果我们已知的不仅仅是另一个点 B' , 而是另一条直线 $B'C'$ (也就是角的一条边), 也可以通过下面的方法得到一个相等的角。我们用圆规在角的顶点 B 处画一个圆 (半径任意), 和角的两边 BA 和 BC 分别交于 D 、 E 两点。用同样的半径, 在 B' 处也用圆规画一个圆。在圆上随便选一点 C' 。用圆规把半径调成 ED 的长度, 用这个半径做另一个以 C' 为圆心的圆, 交圆 B' 于 A' 。连接 B' 和 A' , 连接 B' 和 C' , 则角 $\angle A'B'C'$ 就是要找的角。在三角形 $\triangle DBE$ 和 $\triangle A'B'C'$ 中, $BD = B'A'$, $BE = B'C'$, $DE = A'C'$ (圆规的作用, 圆的性质), 则 $\triangle DBE \cong \triangle A'B'C'$ (三角形全等 SSS 判定), 因此, $\angle A'B'C' = \angle ABC$ (全等三角形性质)。我们就证明了这个尺规作图的过程完成了复制一个角的任务。

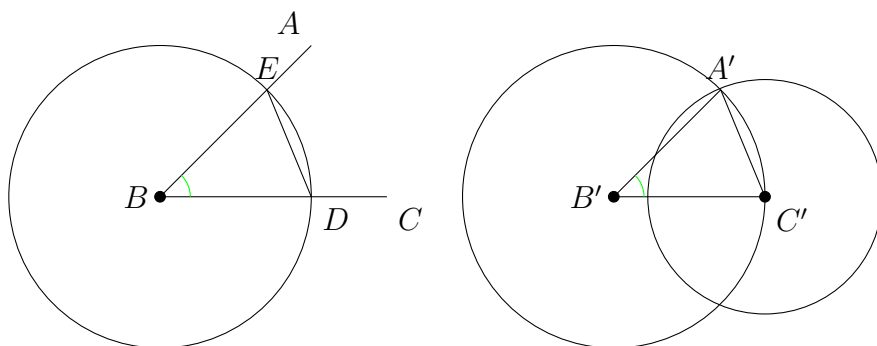


图 15.5: 尺规作图复制角。用圆规和不带刻度的直尺可以复制角。

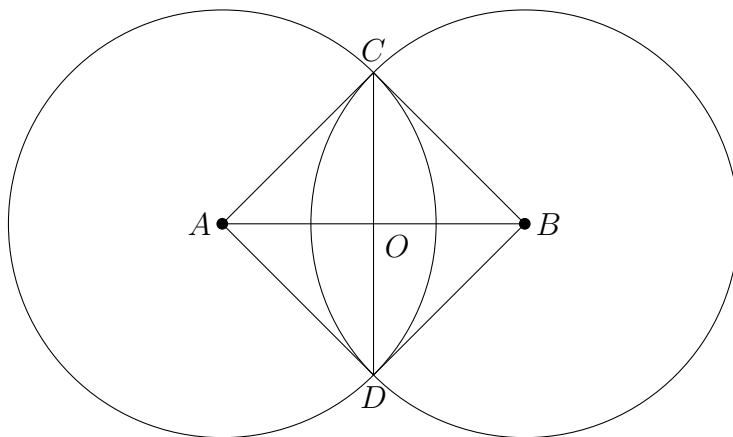


图 15.6: 尺规作图找线段中点。用圆规和不带刻度的直尺可以找到给定线段的中点。

然后，我们来找一个线段的中点。给定线段 AB ，我们要找到一个线段 AB 上的点 O 使得 $AO = OB$ 。图 15.6 所示，我们过 A 做一个半径为比较大的圆（比如就选半径为 AB 也行，一会儿就会知道，至少要选半径大于 $\frac{1}{2}AB$ ），用同样的半径过 B 做一个圆。这两个圆肯定有两个交点，记作 C 和 D 。如果没有交点就是半径还不够大。用直尺连接 C 和 D ，这条直线必然和 AB 有交点，交点 O 就是要找的 AB 的中点。同样我们通过考虑三角形 $\triangle ABC$ 和 $\triangle ABD$ 的全等，三角形 $\triangle CDA$ 和 $\triangle CDB$ 的全等，还有三角形 $\triangle ACO$ 和 $\triangle BCO$ 的全等来证明 $AO = OB$ 。证明留作练习。顺便，思考为什么半径至少要 AB 的一半。其中 CD 也是 AB 的垂直平分线。因此，这也是给定线段的垂直平分线的尺规画图方法。

习题 15.1 (线段中点的尺规作图). 给定线段 AB , 我们过 A 做一个半径为比较大的圆, 用同样的半径过 B 做一个圆, 使得这两个圆有两个交点, 记作 C 和 D . 用直尺连接 C 和 D , 这条直线必然和 AB 有交点, 记作 O . 证明交点 O 就是要找的 AB 的中点。

接着, 我们来平分一个角——做一个给定角的平分线。已知角 $\angle ABC$, 我们希望绘制一条射线 BD 使得 $\angle ABD = \angle DBC$ 。我们用圆规以 B 点为圆心任意长度 (不等于零) 为半径做一个圆交 BA 和 BC 分别与 E 和 F 两点。找到 EF 的中点 D , 连接 BD , 就是要绘制的角平分线。或者, 也可以稍微简化一下: 分别以 E, F 两点为圆心做半径大一点 (或者选择半径和 BE 一样大) 的一个圆在 $\angle ABC$ 的内侧交于一点 D' 。连接 BD' , 就是绘制的角平分线。以第二种简化做图方法为例, 我们来证明这个结论。在三角形 $\triangle BD'E$ 和 $\triangle BD'F$ 中, $BE = BF, ED = FD$ (圆的性质), 则 $\triangle BD'E \cong \triangle BD'F$ (三角形全等 SSS 判定), 因此 $\angle EBD' = \angle FBD'$ (全等三角形性质)。

下面, 我们来做在直线外一点的给定直线的平行线。如图 15.8 所示, 已知直线 AB 和直线外一点 P , 我们要画出来另一条直线 $A'B'$ 过点 P , 使得 $A'B' \parallel AB$ 。在 AB 上随便选一点 C , 连接 PC 并延长 CP 到一个点 Q 。我们得到角 $\angle PCB$ 。以 PQ 为给定边, 用尺规作图方法在 PC 的同侧画一个角 $\angle QPB'$, 使得 $\angle QPB' = \angle PCB$ 。于是, $\angle QPB' = \angle PCB$ 是一对同位角。根据平行线的判定定理, 同位角相等则两直线平行。因此, PB' 或者延长 $B'A$ 到一个点 A' , 则 $A'B'$ 就是要绘制的过 P 平行于 AB 的直线。

最后, 我们来做过直线外一点和过直线上一点做给定直线的垂线。如图 15.9 所示, 已知直线 AB 和直线外一点 P , 以及直线上一点 Q , 我们想用尺规作图找到 $PH \perp AB$ 于 H , $QM \perp AB$ 于 Q 。以 P 为圆心, 取比较大的半径画一个圆, 使得圆 P 和 AB 交于两点 C 和 D 。当半径比较小的时候可能没有交点, 可能有一个交点。分别以 C 和 D 为圆心, 同样大小的半径或者更大一点的半径画一个圆, 这两个圆有两个交点 (如果半径相同则其中一个交点就是 P), 选择和 P 点在直线 AB 另一侧的 N , 连接 PN 。 PN 和 AB 必然有交点, 记作 H 。 $PH \perp AB$ 。证明留作练习。

习题 15.2 (线外垂线的尺规作图). 已知直线 AB 和直线外一点 P 。以 P 为圆心, 取比较大的半径画一个圆, 使得圆 P 和 AB 交于两点 C 和 D 。分

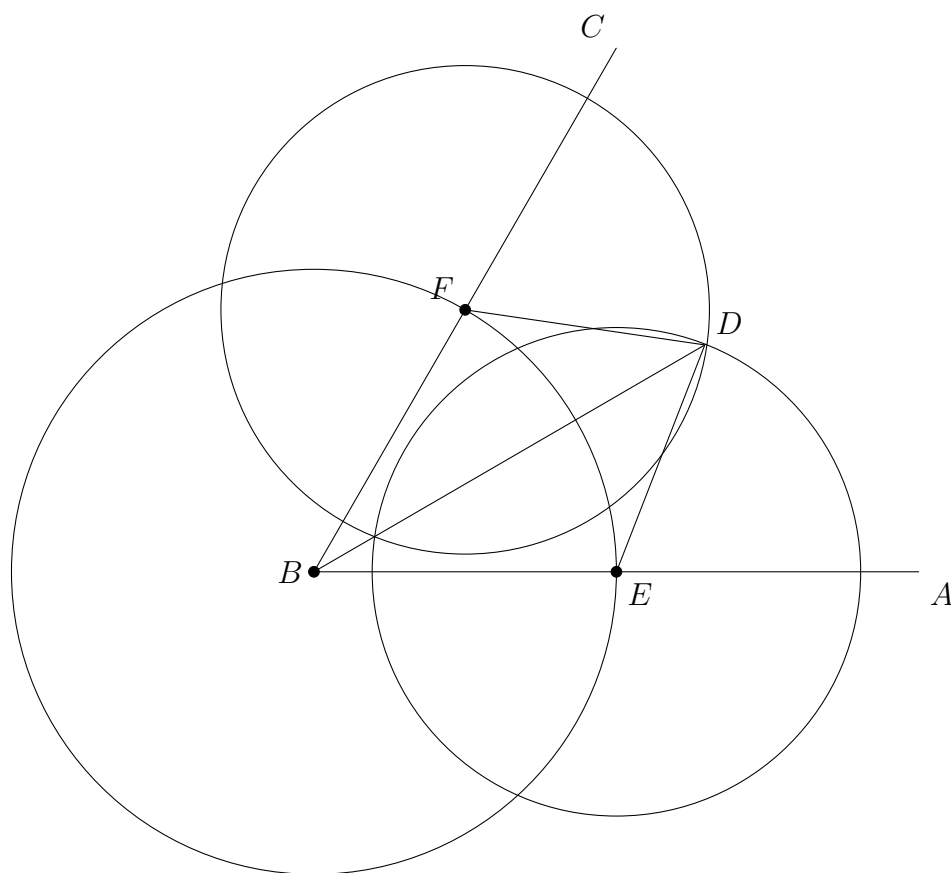


图 15.7: 尺规作图做角平分线。用圆规和不带刻度的直尺可以做一个角的角平分线。

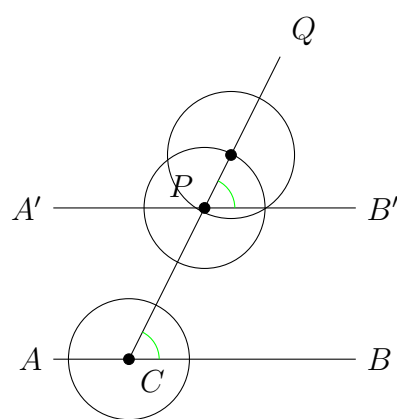


图 15.8: 尺规作图做平行线。基于尺规作图复制一个角，用圆规和不带刻度的直尺可以做过直线外一点的平行线。

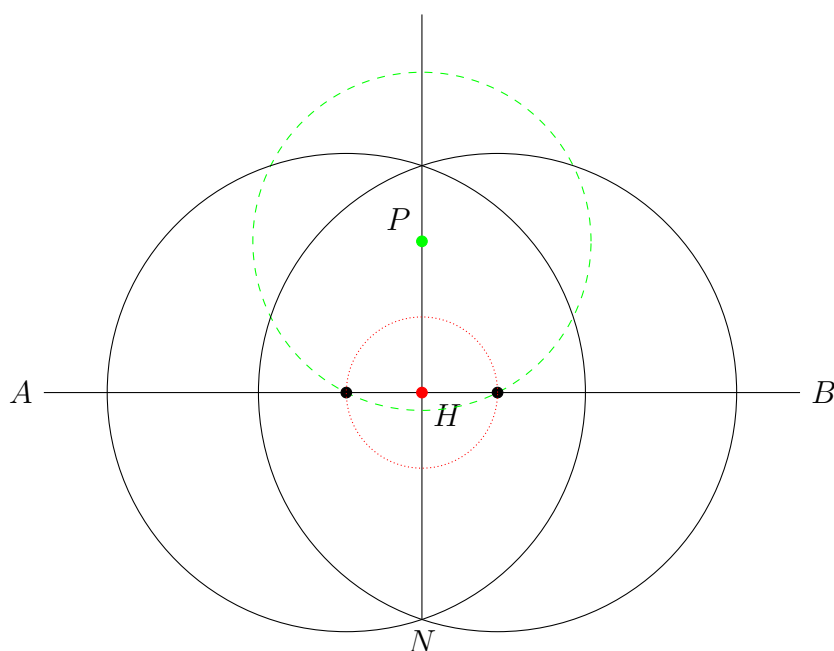


图 15.9: **尺规作图做垂线**。基于线段垂直平分线的尺规作图，可以用圆规和不带刻度的直尺做对于给定直线的过线外或者线上一点的垂线。其中绿色短横线用于过直线外一点做垂线的第一步，而红色虚线部分用于过直线上一点做垂线的第一步。

别以 C 和 D 为圆心，以更大的半径画一个圆，这两个圆有两个交点，选择和 P 点在直线 AB 另一侧的 N ，连接 PN 。 PN 和 AB 必然有交点，记作 H 。证明 $PH \perp AB$ 于 H 。

如果是直线上一点 H ，我们以 H 为圆心，任意长度（不为零）为半径，画一个圆，交 AB 于 C 和 D 两点。分别以 C 和 D 为圆心，更大一点的半径为半径画一个圆，这两个圆有两个交点，记为 P, N ，连接 PN ，则 $PN \perp AB$ 于 H 。证明留作练习。

习题 15.3 (线上垂线的尺规作图)。已知直线 AB 和直线上一点 H 。任意长度（不为零）为半径，画一个圆，交 AB 于 C 和 D 两点。分别以 C 和 D 为圆心，更大一点的半径为半径画一个圆，这两个圆有两个交点，记为 P, N ，连接 PN 。证明 $PN \perp AB$ 于 H 。

尺规作图最基本的操作就是复制一条线段。因此，我们看到实际上后

面这些更复杂的尺规作图任务本质上靠的就是复制一条线段。于是，证明方式也就自然依靠三角形全等的 SSS 判定定理了。

15.8 全等三角形的应用——等腰三角形

有了全等三角形我们就可以来研究等腰三角形了。等腰三角形的判定和性质未来在研究平行四边形、矩形、菱形和正方形的时候会经常用到。

等腰三角形指的是有至少两条边相等的三角形。其中相等的那两条称为等腰三角形的腰，另一条称为等腰三角形的底边。我们先来证明一条等腰三角形的性质定理。

定理 15.15: 等腰三角形底角相等

等腰三角形底角相等] 等腰三角形两个底角相等。

证明 15.12. 我们用数学语言把待证明命题表述为：如图 15.10 所示，在三角形 $\triangle ABC$ 中，已知 $AB = AC$ ，证明 $\angle B = \angle C$ 。

作 $AH \perp BC$ 于 H 。在直角三角形 $Rt\triangle ABH$ 和 $Rt\triangle ACH$ 中， $AB = AC$ ， $AH = AH$ ，因此 $\triangle ABH \cong Rt\triangle ACH$ （直角三角形全等 HL 判定）。因此， $\angle B = \angle C$ （全等三角形性质）。

我们来看看这个定理的逆定理是否存在，也就是证明一条等腰三角形的判定定理。

定理 15.16: 底角相等的是等腰三角形

底角相等的是等腰三角形] 两个底角相等的三角形是等腰三角形。

证明 15.13. 我们用数学语言把待证明命题表述为：如图 15.10 所示，在三角形 $\triangle ABC$ 中，已知 $\angle B = \angle C$ ，证明 $AB = AC$ 。

作 $AH \perp BC$ 于 H 。在三角形 $\triangle ABH$ 和 $\triangle ACH$ 中， $\angle B = \angle C$ ， $\angle AHB = \angle AHC = 90^\circ$ ， $AH = AH$ ，因此 $\triangle ABH \cong Rt\triangle ACH$ （三角形全等 AAS 判定）。因此， $AB = AC$ （全等三角形性质）。

我们再来证明一条以后会用到的等腰三角形的更加复杂的性质定理。

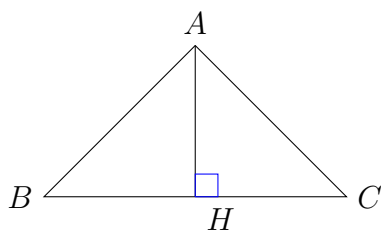


图 15.10: 等腰三角形的底角和腰。等腰三角形底角相等，底角相等的三角形是等腰三角形。

三角形的中线是一个顶点到其对边的中点的直线段。三角形的高是一个顶点到其对边的垂线段。三角形的角平分线是三角形的一个角的平分线和对边的交点形成的从这个角的顶点到交点的线段。

定理 15.17: 等腰三角形三线合一

等腰三角形三线合一] 等腰三角形底边上高、中线和顶角的角平分线三线合一。

证明 15.14. 我们用数学语言把待证明命题表述为：如图 15.10 所示，在三角形 $\triangle ABC$ 中， $AB = AC$ ，证明如果 $AH \perp BC$ 于 H 则 $BH = CH$ ，如果 $BH = CH$ 则 $\angle BAH = \angle CAH$ ，如果 $\angle BAH = \angle CAH$ 则 $AH \perp BC$ 于 H 。

根据 $AH \perp BC$ 于 H ，在直角三角形 $Rt\triangle ABH$ 和 $Rt\triangle ACH$ 中， $AB = AC$ ， $AH = AH$ ，因此 $\triangle ABH \cong \triangle ACH$ （直角三角形全等 HL 判定）。因此， $BH = CH$ ， $\angle BAH = \angle CAH$ （全等三角形性质）。根据 $BH = CH$ ，在三角形 $\triangle ABH$ 和 $\triangle ACH$ 中， $AB = AC$ ， $AH = AH$ ， $BH = CH$ ，因此 $\triangle ABH \cong \triangle ACH$ （三角形全等 SSS 判定）。因此， $AH \perp BC$ 于 H ， $\angle BAH = \angle CAH$ （全等三角形性质）。

根据 $\angle BAH = \angle CAH$ ，在三角形 $\triangle ABH$ 和 $\triangle ACH$ 中， $AB = AC$ ， $\angle BAH = \angle CAH$ ， $AH = AH$ ，因此 $\triangle ABH \cong \triangle ACH$ （三角形全等 SAS 判定）。因此， $AH \perp BC$ 于 H ， $BH = CH$ （全等三角形性质）。

我们实际上证明了其中的任意一个条件成立都可以得到另外两个条件

成立，尽管我们只要证明三个条件之间的循环关系就可以。两者是等价的。

这个定理的逆定理存在吗？更弱的形式的逆定理呢？

习题 15.4 (三 (二) 线合一则等腰). 一个三角形一条边上的高就是这条边上的中线，这个三角形是等腰三角形；一个三角形一条边上的高就是这条边的对角的角平分线，这个三角形是等腰三角形；一个三角形一条边上的中线就是这条边的对角的角平分线，这个三角形是等腰三角形。

这几个命题都成立，证明留作练习。这里我们本来可以把逆定理写成更强的形式：一个三角形一条边上的高就是这条边上的中线，并且是这条边的对角的角平分线，则这个三角形是等腰三角形。不过，我们发现，实际上，我们只需要两个条件就可以保证得到的是等腰三角形。在数学推理中，我们往往找使得命题成立的最低的条件要求。我们也往往尽可能地建立起来等价命题，也就是原名题和逆命题都成立的命题。

15.9 推荐阅读材料

15.10 作业

习题 15.5 (分数在数轴上的表示). 证明图 1.1 中所示的方法可以在数轴上表示任意一个分数。

15.11 本章小结

在这一章中，我们先定义了三角形，再从勾股定理证明了三角函数的和角公式，然后基于之前已经证明的三角形内部的余弦定理证明了三角形内角和为 180° 。接着我们用三角形内部的余弦定理和正弦定理证明了全等三角形和相似三角形的判定定理。最后，我们用全等三角形的性质和判定定理研究了尺规作图以及等腰三角形。尺规作图部分是为了补充我们证明几何定理的时候所做的几何对象的存在性，以及避开通过有限精度的长度和角度来研究几何。等腰三角形是为了给后面四边形的研究做准备。当然，从中体会到全等三角形怎么找，在证明几何定理的时候是多么有用，也是有意义的。

中间我们还学习了如何从两条直线的方程判断这两条直线是否重合、平行还是相交于一点，以及如果相交于一点，形成的角的大小又如何得到。这样我们就有了线段的长度的计算公式和角度的大小的计算公式。原则上，有了平面几何所有基础知识（点、线、角）的数学表示。

在这一章中，我们还反过来用相似三角形的判定和性质来证明了勾股定理。这个是不在本书的平面几何部分的知识结构之内的。在本书中，勾股定理是最基本的公理。其他平面几何知识都建立在这个公理之上。这是为在本书几何部分的最后，来比较不同的平面几何的知识体系来做准备的。

三角形是比点、线和角更加复杂的，比其他的几何对象更加简单的，又是可以构成其他更加复杂的几何对象的基础的对象。对三角形的研究是为后面所有更复杂的对象做的铺垫。

第十六章 平面几何：平行

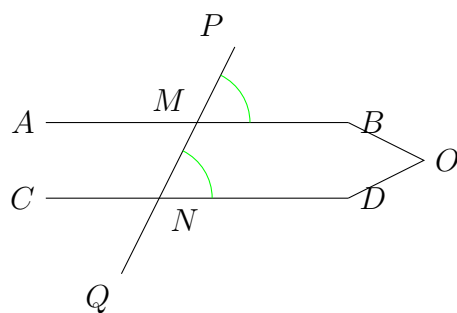


图 16.1: **同位角相等则两直线平行**。从三角形外角大于不相邻内角可以证明同位角相等则两直线平行。

16.1 平行线的判定和性质

定理 16.1: 同位角相等则两直线平行

同位角相等则两直线平行] 两条不重合的直线和第三条直线相交于两个不同点。当两条直线和第三条直线构成的一对同位角相等的时候，这两条直线平行。

证明 16.1. 我们用数学的符号语言来表述待证明的命题：如图 16.1 所示，直线 AB 和 CD 和直线 PQ 相交于 M, N ，已知 $\angle PMB = \angle PND$ ，求证 $AB \parallel CD$ 。

假设 AB 不平行于 CD ，也就是两者相交于一点 O 。则在三角形 $\triangle MNO$ 中，外角 $\angle PMO$ 等于内角 $\angle PNO$ 。这个结果和“三角形外角大于不相邻的内角”相矛盾。因此，原假设错误。 $AB \parallel CD$ 。

根据之前已经证明的同位角、内错角和同旁内角的关系的关系，我们得到

定理 16.2: 内错角相等则两直线平行

内错角相等则两直线平行] 两条不重合的直线和第三条直线相交于两个不同点。当两条直线和第三条直线构成的一对内错角相等的时候，这两条直线平行。

定理 16.3: 同旁内角互补相等则两直线平行

同旁内角互补相等则两直线平行] 两条不重合的直线和第三条直线相交于两个不同点。当两条直线和第三条直线构成的一对同旁内角互补的时候，这两条直线平行。

我们来看看这些定理的逆定理是否成立。

定理 16.4: 两直线平行则同位角相等

两直线平行则同位角相等] 两条平行线和第三条直线相交于两个不同点, 则这两条直线和第三条直线构成的一对同位角相等。

证明 16.2. 我们用数学的符号语言来表述待证明的命题: 如图 16.2 所示, 直线 $AB \parallel CD$, 分别和直线 PQ 相交于 M, N , 求证 $\angle PMB = \angle PND$ 。

不失一般性, 假设 $\angle PMB > \angle PND$, 在 $\angle PMB$ 内部作一个角 $\angle PMF$ 使得 $\angle PMF = \angle PND$ 。则 $MF \parallel CD$ (同位角相等则两直线平行)。这就出现了过同一点 M 有两条直线 AB 和 MF 都和 CD 平行, 这个结果和“线外平行线的唯一性”相矛盾。因此, 原假设错误。 $\angle PMB = \angle PND$ 。

根据之前已经证明的同位角、内错角和同旁内角的关系的关系, 我们得到

定理 16.5: 两直线平行则内错角相等

两直线平行则内错角相等] 两条平行线和第三条直线相交于两个不同点, 这两条直线和第三条直线构成的一对内错角相等。

定理 16.6: 两直线平行则同旁内角互补相等

两直线平行则同旁内角互补相等] 两条平行线和第三条直线相交于两个不同点, 这两条直线和第三条直线构成的一对同旁内角互补。

实际上, 我们也可以从平行线的性质定理来证明平行线的判定定理。

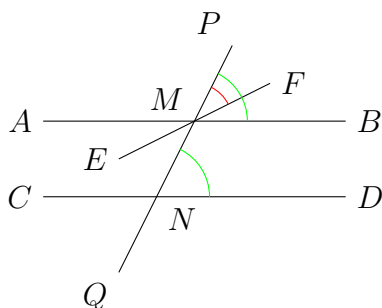


图 16.2: 用于平行性质和平行判定定理证明。同一张图可以用于平行线性质定理 (16.1) 和平行线判定定理 (16.1) 的证明。

定理 16.7: 从两直线平行则同位角相等导出同位角相等则两直线平行

从两直线平行则同位角相等导出同位角相等则两直线平行] 已知：两条平行线和第三条直线相交于两个不同点，则这两条直线和第三条直线构成的一对同位角相等，求证：两条不重合的直线和第三条直线相交于两个不同点，当两条直线和第三条直线构成的一对同位角相等的时候，这两条直线平行。

证明 16.3. 我们用数学的符号语言来表述待证明的命题：如图 16.2 所示，直线 AB 和 CD ，分别和直线 PQ 相交于 M, N ，已知 $\angle PMB = \angle PND$ ，求证 $AB \parallel CD$ （中间得运用平行线的性质定理，也就是两直线平行则同位角相等）。

假设 AB 不平行于 CD ，过 M 作 CD 的平行线 MF 。于是，根据平行线的性质， $\angle PMF = \angle PND$ 。结合 $\angle PMB = \angle PND$ ，有 $\angle PMF = \angle PNB$ 。注意，这是两个锐角（直线间的夹角习惯上选为锐角）。在锐角范围内，同一个边当作角的起点边（这里是 PM ），我们出现了两个相等的角，但是角的终点边不同，这和“同一个起点边，角的大小和角一一对应”相矛盾。因此，原假设错误。因此， $AB \parallel CD$ 。

有了平行线的性质，我们可以来做几个练习，

习题 16.1 (垂直于同一条直线). 垂直于同一条直线的两条直线平行。

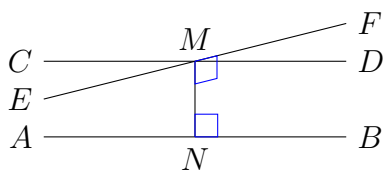


图 16.3: 线外平行线的唯一性。从平行线的性质定理可以证明线外平行线的唯一性。

习题 16.2 (分别垂直于一组平行线). 分别垂直于一组平行线中的一条的两条直线平行。

之前我们不用平行线的性质和判定就证明过前者。你可以对比一下两种证明。

定理 16.8: 从平行线性质证明三角形内角和

从平行线性质证明三角形内角和] 求证三角形内角和为 180° ，证明过程只能用平行线性质。

证明 16.4. 我们把待证明的命题用数学的符号表述为：在三角形 $\triangle ABC$ 中，基于平行线的性质，证明 $\angle A + \angle B + \angle C = 180^\circ$ 。
 延长 BC 射线到 D ，过 C 点做 AB 的平行线 CE 。则 $\angle ECA = \angle CAB$ (两直线平行内错角相等)， $\angle ECD = \angle ABC$ (两直线平行同位角相等)。 $180^\circ = \angle BCD = \angle BCA + \angle ACE + \angle ECD = \angle BCA + \angle BAC + \angle ABC$ 。

到此，我们从“三角形内角和为 180° ”证明了“平行线的判定定理”，从平行线的判定定理结合“线外平行线的唯一性”证明了平行线的性质定理，从平行线的性质定理证明了“三角形内角和为 180° ”。

我们还可以从平行线的性质定理来证明线外平行线的唯一性。

定理 16.9: 从平行线的性质定理来证明线外平行线的唯一性

从平行线的性质定理来证明线外平行线的唯一性] 已知：两条平行线和第三条直线相交于两个不同点，则这两条直线和第三条直线构成的一对同位角相等，求证：过直线外一点有且只有一条平行线。

证明 16.5. 我们用数学的符号语言来表述待证明的命题：直线 AB 和 AB 外一点 M ，求证过 M 有且只有一条直线平行于 AB （中间得运用平行线的性质定理，也就是两直线平行则同位角相等）。

根据线外和线上垂线的存在性和唯一性，如图 16.3，我们过 M 做 $MN \perp AB$ 于 N ，过 M 做 $CD \perp MN$ 于 M 。我们发现 CD 是存在并且唯一的。现在我们来证明 $CD \parallel AB$ 。

假设 $CD \cap AB = R$ 。首先，我们来证明 R 不在直线 MN 上。否则， MN 和 CD 具有两个交点 M, R 。两条不重合的直线只有一个交点或者没有交点。矛盾。于是，过 MN 外一点 R 就存在着两条垂直于 MN 的直线 CD 和 AB 。矛盾。因此， $CD \parallel AB$ 。因此，过直线外一点存在至少一条平行线。

我们再来证明这样的平行线唯一。我们假设过 M 还有另外至少一条不和 CD 重合平行线（下面的证明也适用于更多条） $EF \parallel AB$ 。根据“两直线平行则同旁内角互补”，有 $\angle FMN = \angle BNM = 90^\circ$ 。因此， $EF \perp MN$ ，于是过直线 MN 上的一点 M 有两条不重合的直线 CD 和 EF 垂直于直线 MN 。这个和“过直线上一点有且只有一条垂线”矛盾。因此，原假设错误。过 M 没有其他平行线。

合起来，过 M 有且只有一条直线平行于 AB 。

我们补充这几条定理的证明是为了说明，并不是所有的定理的证明都需要直接和间接地依赖于距离公式（勾股定理）。实际上，我们也可以选择从“三角形的内角和是 180° ”开始，或者选择从“线外平行线唯一”开始。甚至，从相似三角形的判定定理我们也证明了“勾股定理”。也就是说，如果从“三角形的内角和是 180° ”或者“线外平行线唯一”能够证明相似三角形的判定定理，那么，勾股定理在这个知识结构中就是一个定理，而不是像本书一样是所有的平面几何学知识的逻辑基础。具体的从“三角形的内角和是 180° ”或者“线外平行线唯一”能够证明相似三角形的判定定

理证明过程，我们就不在本书中展示了。但是，这是做得到的¹。

这也是为什么人们经常说“勾股定理”、“三角形的内角和是 180° ”和“平行公理”等价的原因。

16.2 从平行线和全等三角形看相似三角形

有了平行线的判定和性质定理，有了平行线基本定理，有了全等三角形的判定定理和性质，我们来重新证明相似三角形的判定定理。

¹通常的这样的平面几何学的教材需要通过三角形面积公式来证明“平行基本定理”，而三角形面积公式的地位比较特殊，看起来是不一定能够从“三角形的内角和是 180° ”或者“线外平行线唯一”得到的。凡是用面积关系来证明命题的，在我看来，都相当于用了距离公式和基于距离公式的微积分。因此，这也是本书没有选择另一套平面几何知识体系的原因之一——另一个原因是通过坐标来定义点和线本身也更合理更自然。如果有一天，我自己建立了或者看到了一个更干净的基于“三角形的内角和是 180° ”或者“线外平行线唯一”平面几何知识体系的构建方式，我会更新这个部分。

定理 16.10: 三角形相似 AA 判定

三角形相似 AA 判定] 两个三角形的两个角对应相等则两个三角形相似。

证明 16.6. 我们把待证明的命题用数学的符号表述为：在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中， $\angle ABC = \angle A'B'C'$, $\angle BAC = \angle B'A'C'$ ，则这两个三角形相似，记作 $\triangle ABC \sim \triangle A'B'C'$ 。

首先，通过两个角对应相等可以得到第三个角也对应相等（三角形内角和定理）。

假设两个三角形没有一组边对应相等，如果有，按照全等三角形的 AAS 或者 ASA 判定定理，这两个三角形全等，而全等三角形必然相似（全等三角形的性质和相似三角形的定义）。不失一般性，假设 $AB > A'B'$ 。

我们在 AB 上截取 $AD = A'B'$ ，在 AC 上截取 $AE = A'C'$ ，已知 $\angle DAE = \angle BAC = \angle B'A'C'$ 则 $\triangle ADE \cong \triangle A'B'C'$ （三角形全等 SAS 判定）。于是， $\angle ADE = \angle A'B'C'$ （三角形全等的性质）。于是， $DE \parallel BC$ （平行线的判定）。因此， $\frac{AD}{AB} = \frac{AE}{AC}$ ，也就是 $\frac{A'B'}{AB} = \frac{A'C'}{AC}$ （平行基本定理）。

同理可得 $\frac{B'A'}{BA} = \frac{B'C'}{BC}$ 。合起来得到 $\frac{A'B'}{AB} = \frac{A'C'}{AC} = \frac{B'C'}{BC}$ 。三边对应成比例。

因此，这两个三角形的三条边都对应成比例三个角对应相等，这两个三角形相似（三角形相似的定义）。

定理 16.11: 三角形相似 SSS 判定

三角形相似 SSS 判定] 两个三角形的三条边对应成比例, 则这两个三角形相似。

证明 16.7. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $\frac{a}{a'} = \frac{b}{b'} = \frac{c}{c'}$, 则这两个三角形相似, 记作 $\triangle ABC \sim \triangle A'B'C'$ 。

假设两个三角形没有一组边对应相等, 如果有, 按照全等三角形的 SSS 判定定理, 这两个三角形全等, 而全等三角形必然相似 (全等三角形的性质和相似三角形的定义)。不失一般性, 假设 $AB > A'B'$ 。

我们在 AB 上截取 $AD = A'B' = c'$, 在 AC 上截取 $AE = A'C' = b'$, 已知 $\frac{c'}{c} = \frac{b'}{b}$ 则 $DE \parallel BC$ (平行基本定理的逆定理)。因此 $\frac{DE}{a} = \frac{DE}{BC} = \frac{c'}{c} = \frac{b'}{b}$, 因此 $DE = B'C'$ 。于是 $\triangle ADE \cong \triangle A'B'C'$ (三角形的 SSS 判定定理)。因此, $\angle BAC = \angle DAE = \angle B'A'C'$, $\angle ADE = \angle A'B'C'$, $\angle AED = \angle C'B'A'$ 。

因为 $DE \parallel BC$, 所以 $\angle ADE = \angle ABC$, $\angle AED = \angle ACB$ (平行线的性质)。因此 $\angle BAC = \angle B'A'C'$, $\angle ABC = \angle A'B'C'$, $\angle ACB = \angle C'B'A'$ 。三个角对应相等。

因此, 这两个三角形的三条边都对应成比例三个角对应相等, 这两个三角形相似 (三角形相似的定义)。

定理 16.12: 三角形相似 SAS 判定

三角形相似 SAS 判定] 两个三角形的两条边对应成比例, 这两条边的夹角对应相等, 则这两个三角形相似。

证明 16.8. 我们把待证明的命题用数学的符号表述为: 在三角形 $\triangle ABC$ 和 $\triangle A'B'C'$ 中, $\frac{AB}{A'B'} = \frac{AC}{A'C'}$, $\angle CAB = \angle C'A'B'$, 则这两个三角形相似, 记作 $\triangle ABC \sim \triangle A'B'C'$ 。

假设两个三角形的对应比例值 $\frac{AB}{A'B'} = \frac{AC}{A'C'} = k \neq 1$ 。如果 $k = 1$, 按照全等三角形的 SAS 判定定理, 这两个三角形全等, 而全等三角形必然相似 (全等三角形的性质和相似三角形的定义)。不失一般性, 假设 $k > 1$, 也就是 $AB > A'B'$ 。

我们在 AB 上截取 $AD = A'B' = c'$, 在 AC 上截取 $AE = A'C' = b'$ 。加上 $\angle CAB = \angle C'A'B'$, 有 $\triangle ADE \cong \triangle A'B'C'$ (三角形的 SAS 判定定理)。因此, $\angle ADE = \angle A'B'C'$, $\angle AED = \angle A'C'B'$, $DE = B'C'$ 。又已知 $\frac{c'}{c} = \frac{b'}{b}$ 也就是 $\frac{AD}{AB} = \frac{AE}{AC}$ 则 $DE \parallel BC$ (平行基本定理的逆定理)。因此, $\angle ADE = \angle ABC$ (平行线的性质)。

于是, $\angle A'B'C' = \angle ADE = \angle ABC$, 加上已知的 $\angle CAB = \angle C'A'B'$, 这两个三角形的两个角对应相等, 于是这两个三角形相似 (三角形相似的 AA 判定)。

16.3 本书平面几何的知识体系

我们从距离公式证明了线外平行线的唯一性 (等价地平行的传递性, 或者说传统教材中的平行公理); 从距离公式定义了直角三角形中的三角函数, 推广到了一般三角形中的三角函数, 进而证明了一般三角形的余弦定理和正弦定理, 以及证明了三角函数的和差公式; 从距离公式 (间接通过余弦定理) 证明了三角形内角和等于 180° 这个公式 (传统教材的三角形内角和公理); 从距离公式证明了两点确定一条直线 (传统教材的直线公理) 以及直线段最短 (传统教材的直线最短公理); 从距离公式证明了线上以及线外垂线的唯一性 (传统教材的垂线唯一性公理); 从距离公式证明了平行基本定理——平行线切割线段成比例。

从三角形内角和公式证明了平行线的判定定理，以平行公理为基础证明了判定定理和性质定理的等价性。从正弦定理和余弦定理证明了全等三角形的判定定理。从平行基本定理证明了相似三角形的判定定理。从相似三角形的判定定理进而可以得到勾股定理。

后面将会看到，我们用平行线的性质和判定定理，相似三角形的判定和性质（全等三角形的判定和性质当作相似的特殊情况），平行基本公理，来得到其他的几何知识，例如四边形和圆中的几何知识。

我们发现，在本书的这个体系中，所有的其他命题都是距离公式的导出结论。也就是说，我们把距离公式当作平面几何知识体系的基本公理。但是，实际上，在很多其他书中，公理的选择是不同的。

当然，像绝大部分中国初中的几何教材那样，把“三角形内角和公式、平行公理、平行线的性质、平行线的判定、三角形全等的判定”都当作公理，那是完全胡闹。这样的几何教材，只能学习到一题条条的命题，而不能学习到命题通过命题之间的证明和被证明关系所构成几何知识的体系，也不能体会到逻辑演绎证明在其中的关键作用，更加不能体会到数学家对尽可能少地使用公理把更多的知识建立在更少的公理的基础上的艰苦卓绝的努力，以及这个努力过程中的痛苦和快乐。

放开这样的低层次的教材不谈，在国外数学家和数学教育家的努力下，其实是存在着几套不同的几何知识体系的。不同的体系选择不同的命题当作公理，通过受现实启发对现实做抽象或者纯粹人类想象力和逻辑来提出（猜测出）概念和命题，通过严格的逻辑演绎证明来建立命题之间的关系。在本书几何部分的最后一节（19.2 节）中，我们会来欣赏一下除了本书的以距离公式为公理的平面几何知识体系之外，还有哪一些。

16.4 平行和全等三角形的应用——四边形

四边形是由四条边首尾相连构成的凸图形。有一组对边平行的四边形被称为梯形。有的时候，我们也要求梯形有且只有一组对边平行。注意这两个梯形的定义并不相同。在日常语言的习惯上，我们往往意味着后者。但是，我们这里采用前者，也就是允许梯形的另一组对边也平行。

两组对边对应相互平行的四边形称为平行四边形。因此，按照我们选

择的梯形的定义，平行四边形是梯形，属于梯形的一种特殊情况。

四个角是直角的平行四边形称为矩形。四条边相等的矩形被称为正方形。四条边相等的四边形被称为菱形。注意菱形的定义本身对边是否平行没有要求。

下面我们来证明几条关于四边形的判定和性质定理。这样未来更容易使用各种四边形。

定理 16.13: 平行四边形对边相等

平行四边形对边相等] 平行四边形对边相等。

证明 16.9. 我们把待证明的命题用数学的符号表述为：在四边形 $ABCD$ 中， $AB \parallel DC$ ， $AD \parallel BC$ ，求证 $AB = DC$ ， $AD = BC$ 。
连接 AC ，因为 $AB \parallel DC$ ，所以 $\angle BAC = \angle DCA$ (平行线性质)，因为 $AD \parallel BC$ ，所以 $\angle DAC = \angle BCA$ (平行线性质)。在三角形 $\triangle ABC$ 和 $\triangle ADC$ 中， $\angle BAC = \angle DCA$ ， $AC = CA$ ， $\angle DAC = \angle BCA$ ，因此， $\triangle ABC \cong \triangle ADC$ (全等三角形 ASA 判定)。因此， $AB = DC$ ， $AD = BC$ (全等三角形性质)。

我们来看看这条定理的逆定理是否成立。

定理 16.14: 对边相等的四边形

对边相等的四边形] 两组对边对应相等的四边形是平行四边形。

证明 16.10. 我们把待证明的命题用数学的符号表述为：在四边形 $ABCD$ 中， $AB = DC$ ， $AD = BC$ ，求证 $AB \parallel DC$ ， $AD \parallel BC$ 。
连接 AC ，在三角形 $\triangle ABC$ 和 $\triangle ADC$ 中， $AB = CD$ ， $BC = DA$ ， $AC = CA$ ，因此， $\triangle ABC \cong \triangle ADC$ (全等三角形 SSS 判定)。因此， $\angle BAC = \angle DCA$ ， $\angle DAC = \angle BCA$ (全等三角形性质)。因此， $AB \parallel DC$ ， $AD \parallel BC$ (平行线判定)。

有了这个逆定理，我们就有了定义之外的方法来判断一个四边形是不是平行四边形了。接着，我们来证明其他判定定理。

定理 16.15: 一组对边平行且相等的四边形

一组对边平行且相等的四边形] 一组对边平行且相等的四边形是平行四边形。

证明 16.11. 我们把待证明的命题用数学的符号表述为：在四边形 $ABCD$ 中， $AB = DC, AB \parallel DC$ ，求证 $AD \parallel BC$ 。

连接 AC ，因为 $AB \parallel CD$ ， $\angle BAC = \angle DCA$ （平行线的性质）。在三角形 $\triangle ABC$ 和 $\triangle ADC$ 中， $AB = CD, \angle BAC = \angle DCA, AC = CA$ ，因此， $\triangle ABC \cong \triangle ADC$ （全等三角形 SAS 判定）。因此， $\angle DAC = \angle BCA$ （全等三角形性质）。因此， $AD \parallel BC$ （平行线判定）。

注意，联系到三角形全等判定没有 SSA 定理，我们可能也没有下面的这条平行四边形判定定理。检验这一条是否成立留作作业。

习题 16.3 (一组对边相等另一组对边平行的四边形). 一组对边相等一组对边平行的四边形是平行四边形吗？

平行四边形还有更复杂的性质，例如

定理 16.16: 平行四边形角平分线

平行四边形角平分线] 平行四边形角平分线相互平分。

同时其逆定理也成立，

定理 16.17: 角平分线相互平分的四边形

角平分线相互平分的四边形] 角平分线相互平分的四边形是平行四边形。

这些都可以用全等三角形的判定定理和性质来证明。留作练习。

有了四边形，我们还可以把平行基本定理（定理 (14.2)）做一个从三角形到四边形的推广。

定理 16.18: 平行线切割线段成比例

平行线切割线段成比例] 三条平行线切割两条直线，形成的两组线段对应成比例。

证明 16.12. 我们把待证明的命题用数学的符号表述为：直线 $l_1 \parallel l_2 \parallel l_3$ ，交直线 l_4, l_5 分别于 A, B （在 l_1 和 l_4, l_5 上）， C, D （在 l_2 和 l_4, l_5 上）， E, F （在 l_3 和 l_4, l_5 上）。证明 $AC : CE = ED : DF$ 。

假设 l_4, l_5 平行，则四边形 $ABDC$ 和 $CDFE$ 都是平行四边形，则对边相等，也就是 $AC = BD, CE = DF$ ，于是 $\frac{AC}{CE} = \frac{BD}{DF}$ 。如果 l_4, l_5 相交，延长两直线到其交点 G ，则三角形 $\triangle GCD$ 和 $\triangle GEF$ 刚好满足基本平行定理的要求，也就是有 $\frac{GA}{AC} = \frac{GB}{BD}, \frac{GA}{AE} = \frac{GB}{BF}$ 。合起来，得到 $\frac{GA}{GB} = \frac{AC}{BD} = \frac{AE}{BF}$ ，于是 $\frac{AC}{AE} = \frac{BD}{BF}$ ，也就是 $\frac{AC}{CE} = \frac{BD}{DF}$ 。

我们这里就用到了平行四边形的性质。

有了平行四边形的判定和性质定理，我们来研究一下矩形。同样地，我们来从矩形的定义出发研究矩形的判定和性质。

定理 16.19: 一个角是直角的平行四边形

一个角是直角的平行四边形] 有一个角是直角的平行四边形是矩形。

证明 16.13. 我们把待证明的命题用数学的符号表述为：在平行四边形 $ABCD$ 中， $AB \parallel DC, AD \parallel BC, \angle BAD = 90^\circ$ 。求证 $\angle ABC = \angle BCD = \angle CDA = 90^\circ$ 。

根据 $AB \parallel DC$ ，有 $\angle BAD + \angle CDA = 180^\circ$ （平行线的性质），又已知 $\angle BAD = 90^\circ$ ，得到 $\angle CDA = 90^\circ$ 。同理， $\angle ABC = 90^\circ$ 。根据四边形的内角和等于 360° ，或者我们也可以再一次使用平行线的性质，得到另一个角 $\angle BCD = 90^\circ$ 。

这样我们在判断矩形的时候，只需要看平行四边形是否有一个内角是直角就可以。

再次运用相邻的角是直角以及平行四边形对边相等、全等三角形的判定和性质，可以证明矩形对角线相等。

定理 16.20: 矩形对角线相等

矩形对角线相等] 矩形对角线相等。

我们还可以考虑其逆命题。注意对角线相互平分是平行四边形的性质和判定条件。

定理 16.21: 矩形对角线

矩形对角线] 对角线相等且相互平分的四边形是矩形。

证明过程中，需要先利用对角线相互平分证明这个四边形是平行四边形，得到对边相等且平行，再运用全等三角形的判定来得到临角相等。再结合平行线同旁内角互补，就得到其中有两个角是直角。于是，证明了这个四边形是矩形。具体证明留作练习。

有了平行四边形的判定和性质定理，我们来研究一下菱形。同样地，我们来从菱形的定义出发研究菱形的判定和性质。首先，菱形四条边都相等，则对边相等，因此肯定是平行四边形。那菱形比平行四边形多什么呢？

定理 16.22: 菱形对角线

菱形对角线] 菱形对角线相互垂直平分。

证明 16.14. 我们把待证明的命题用数学的符号表述为：在四边形 $ABCD$ 中， $AB = BC = CD = DA$ ， $AC \cap BD$ 于点 O 。求证 $AC \perp BD$ ，并且 $AO = OC, BO = OD$ 。

根据 $AB = CD, BC = AD$ 得到四边形 $ABCD$ 是平行四边形（平行四边形判定），于是对角线相互平分 $AO = OC, BO = OD$ 。在三角形 $\triangle ABD$ 中， $AB = AD$ 是等腰三角形（等腰三角形判定），因此 BD 上的高和中线合一，也就是从 $BO = OD$ 得到 $BD \perp AO$ 。

我们也可以直接从菱形的定义开始，运用两次全等三角形来得到菱形的对角线相互垂直平分。我们也看看这个定理的逆定理是否成立。通过运用等腰三角形底边三（二）线合一的逆定理，可以证明这个逆定理也成立。也可以运用三次全等三角形的判定定理和性质来得到这个四边形的四边都

相等，于是是一个菱形。具体证明留作练习。

定理 16.23: 对角线相互垂直平分

对角线相互垂直平分] 对角线相互垂直平分的四边形是菱形。

正方形相当于是矩形和菱形的结合，因此很容易我们猜测

习题 16.4 (正方形的对角线). 正方形的对角线相等而且相互垂直平分；

以及

习题 16.5 (对角线相互垂直平分且相等). 对角线相等而且相互垂直平分的四边形是正方形。

这两条是否成立，如何证明，留作练习。

再一节所证明的命题本身是关于四边形的，但是，在这里我们更多地把这些命题看作是展示三角形全等判定定理和性质、平行的判定和性质的案例。未来我们会看到，在平面几何的知识体系内，最基本的也就这几样东西——点的定义、线的定义、圆和角的定义、平行的定义、垂直的定义、平行的判定定理和性质、平行基本定理（推广为平行线切割线段成比例）、相似三角形（把全等三角形看做了特殊情况）的判定和性质，其它东西全都可以从这几样东西得到。

学习知识不是为了积累足够数量的结论，而是为了看到这些结论背后的知识体系，进而体会到这个知识体系是怎么建构起来的，从而实现：第一、有必要的时候自己来按照先有的结构重建建构这个知识体系，或者按照自己的新的方式来建构这个知识体系；第二、发现现有知识体系里面的问题，例如逻辑上的概念上的漏洞，或者新的值得提出的概念和命题，从而发展这个知识体系；第三、把这个知识体系中的概念和命题——由于我们是自己构建起来的理解得更深刻——创造性地用到看起来不像这些概念和命题但是实际上可以转化为这些概念和命题的问题中去。

16.5 推荐阅读材料

16.6 作业

16.7 本章小结

在本章中，我们终于把平行线的判定定理和性质定理的证明完成了。加上前面的三角形全等的判定定理的证明、三角形相似判定定理的证明、平行基本定理的证明，我们从勾股定理当作逻辑起点，已经建构起来了平面几何知识的核心定理。基于勾股定理和点、线、角、平行、垂直的定义，本书中所有命题都是被证明出来的。

从决定走以勾股定理（距离公式）当作平面几何的知识基础²，到真的把这个知识体系构建起来，并不是一件简单的工作。但是，我决定这么去走，并且尽量严格，尽管这条路和绝大多数教材都不同。从中，你可以体会到我对常见的中国几何教材是多么地痛恨：我宁愿付出额外的大量的努力，也不愿意我的学生在学习几何的时候大量的命题是受现实启发，做作总结归纳，就当作真的数学知识的！数学确实受现实的启发，但是现实的启发和检验，不是数学知识确立的方式！数学也不受现实的制约——数学命题在现实世界中检验不通过的不能当作否定数学命题的理由！检验通过也不能当作数学命题成立的理由！

几何知识的系统性，也就是从少数的定义和公理出发，通过逻辑演绎证明来得到所有的其他知识，构建起来一个知识系统，是所有的后来的各个自然科学和社会科学的学科都尽可能地构建知识系统的典范。

实际上，你可以把具体本书是如何构建平面几何的知识系统忘了，你可以把本书中的平面几何知识体系中的任何一个概念、命题和证明命题的过程忘了，只要你仍然对构建知识系统这个目标以及构建知识系统的方式主要是逻辑演绎证明，以及所构建的系统里面的概念要么来自于对现实世界的抽象要么来自于纯粹理性的想象力，这几条印象深刻，那也完全达到了学习平面几何学的目的。

²那是在十来年前我写《小学数学这样学》的时候。那时候我梳理了平面几何学的大多数教材的知识结构和平面几何学知识本身的结构，决定要写一本《中学数学这样学》。

第十七章 平面几何：圆

17.1 圆的定义和平移

之前我们就定义了圆，为平面上到固定点 O （称为这个圆的圆心）距离（称为这个圆的半径）相等的点的集合。按照距离公式，这就是下面的方程的解，

$$\sqrt{(x - x_O)^2 + (y - y_O)^2} = r. \quad (17.1)$$

或者等价地

$$(x - x_O)^2 + (y - y_O)^2 = r^2. \quad (17.2)$$

根据这个定义，我们得到“不在同一直线上的三个点确定一个圆”，

定理 17.1: 三点定圆

过不在同一直线上的三点有且只有一个圆。

证明 17.1. 首先，在同一直线上的三个点不可能确定一个圆，因为没有这样一个圆同时过同一直线上的三个点。假设 A, B, C 在同一直线上，则到 A, B 距离相等的点在 AB 的垂直平分线 l_1 上，到 B, C 距离相等的点在 BC 的垂直平分线 l_2 上。 l_1, l_2 都垂直于直线 ABC ，因此 $l_1 \parallel l_2$ （垂直于同一条直线的两条直线平行），也就是两者没有交点。因此，没有一个点能够做到到同一直线上的三个点 A, B, C 距离都相同。

其次，我们来证明经过不在同一直线上的三个点有且只有一个圆。我们来找到和证明圆心和半径存在并且唯一。在三角形 ABC 中，我们做 AB 上的垂直平分线 l_{AB} ， BC 上的垂直平分线 l_{BC} ，我们来证明两者必然有交点 O ，并且交点 O 就是圆心。我们先来证明 l_{AB} 不可能平行于 l_{BC} 。如果两者平行，则 $AB \parallel BC$ （分别垂直于一组平行线的两条直线必然平行）。但是， AB, BC 显然相交于 B 矛盾。我们记这个交点为 O 。根据垂直平分线的性质，我们有 $AO = BO$ ， $BO = CO$ ，因此 $AO = CO$ ，于是 O 也在 AC 的垂直平分线上。 O 唯一确定（两条垂直平分线各自唯一确定，两条不重合的直线的交点确定一个唯一的交点），半径长度为 $OA = OB = OC$ 也唯一确定。

这里我们用到了垂直于同一条直线的两条直线平行，以及分别垂直于一组平行线中的一条的两条直线平行。这些定理已经之前留作作业。

根据这个定义，我们还可以得到，通过移动圆心，可以使得两个半径相同的圆重合。为了严格地表达清楚什么是移动什么是重合，为什么会重合，我们先来学习一下移动。

在 x 轴上，我们说移动可以表示为

$$x \rightarrow x + a, \quad (17.3)$$

其含义上把原来坐标轴上 x 对应的点，移动到 $x + a$ 的地方。当 $a > 0$ 的时候，就是往右移动了 $|a|$ ；当 $a < 0$ 的时候，就是往左移动了 $|a|$ 。也可以把这个移动看成是坐标原点的移动：当 $a > 0$ 的时候，坐标原点往左移动



图 17.1: 平移的两种视角。平移可以看作是坐标原点的移动，几何点不变，也可以看作是坐标原点没动，几何点的移动。

了 $|a|$ ；当 $a < 0$ 的时候，就是往右移动了 $|a|$ ，如图 17.1 所示。

后者，因为原点移动了，相当于给每个点（这些点几何上没有动）引入了新的坐标，可以表达为一套新的坐标，

$$x' = x - a. \quad (17.4)$$

当然，我们再次强调这两个视角——固定原点几何上移动点和固定几何上的点看作这些点的坐标在变动——是完全等价的。由于坐标变换的语言更适合写成代数表达式，我们往往更倾向于用后一种视角。

假设另一个圆的圆心在不和 O 重合的 $P = (x_P, y_P)$ ，其可以表述为如下方程的解，

$$(x - x_P)^2 + (y - y_P)^2 = r^2. \quad (17.5)$$

我们来这样移动坐标原点——保持每个点几何上不变但是改变其坐标的值，

$$x' = x - x_P + x_O, y' = y - y_P + y_O. \quad (17.6)$$

则算式 (17.5) 可以被表示为新的方程，

$$(x' - x_O)^2 + (y' - y_O)^2 = r^2. \quad (17.7)$$

这正好形式上和算式 (17.2) 完全相同。

我们回到移动几何点的视角。按照两种视角的关系，这相当于

$$x \rightarrow x + x_P - x_O, y \rightarrow y + y_P - y_O. \quad (17.8)$$

例如，一个原来在 $O = (x_O, y_O)$ 的点（圆 O 的圆心），就被移动到了 $O = (x_O + x_P - x_O = x_P, y_O + y_P - y_O = y_P)$ 的点（圆 P 的圆心）。同理，圆 O 圆周上的所有点也就被移动成了圆 P 圆周上的所有点。因此，两个圆完全重合。

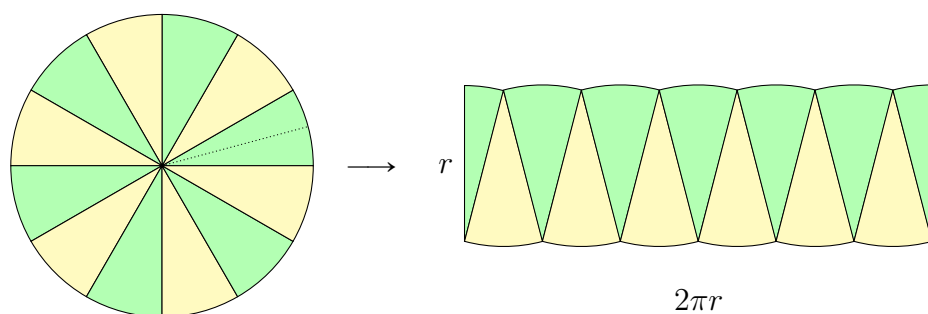


图 17.2: 圆面积的切割计算。通过把圆切割成多个扇形, 然后分成两片, 上下合起来, 可以得到近似的长方形, 从而得到圆的面积。

于是, 我们发现, 只要两个圆半径相等, 总是可以通过移动使得两个圆重合。因此, 半径相同的圆全等。所以, 圆的周长、面积等属性, 也就只依赖于圆的半径, 因为圆的周长和面积属于不随着移动圆而改变的性质, 而在决定圆的变量——只有圆心和半径——里面, 不随着移动而改变的性质只有圆的半径。再根据单位的一致性, 我们得到周长和面积分别为 $C \sim r, S \sim r^2$ 。

如果我们记其中一个系数为 $C = \Pi r$, 则可以得到面积为 $S = \frac{\Pi}{2} r^2$ 。具体计算将来我们仍然可以用微积分基于距离公式来完成。但是, 我们也可以引入下面的包含了极限的思想的割补来完成 $C = \Pi r \Rightarrow S = \frac{\Pi}{2} r^2$ 。

如图 17.2 所示, 我们把整个圆的圆心角均匀分成 N 个小圆心角, 每一份都对应着一个扇形——两条半径用一条圆弧连起来的形状。我们知道分的越细 N 越大, 则每一份越接近一个三角形。其中三角形的总的底边长度是圆弧的总长度 C 。我们把切好的这样的周长上长了的 N 个三角形拉开, 可以发现每一片都是一个三角形。我们把这样的周长对折, 用一侧的三角形插入另一侧的三角形, 就能得到长方形。长方形的面积为 $S = \frac{C}{2} \times r$ 。其中 $\frac{C}{2}$ 是因为我们要对折, r 是因为当 N 足够大的时候小三角形的边长和半径接近。因此, 我们得到 $S = \frac{\Pi}{2} r^2$ 。

习惯上, 我们不想写着个分数在公式里面, 因此, 把系数改为 $C = 2\pi r, S = \pi r^2$ 。至于 π 的值, 我们可以通过从圆的内接正多边形和外切正多边形的周长来逼近圆的周长的方式来无限逼近。

看到一个对象一个现象和哪些变量有关, 一个系统里面有哪些自由变量, 哪些变量是由这些自由变量决定的, 是一个非常有用的思维习惯, 甚至比知道具体系统里面具体变量之间的关系还重要。

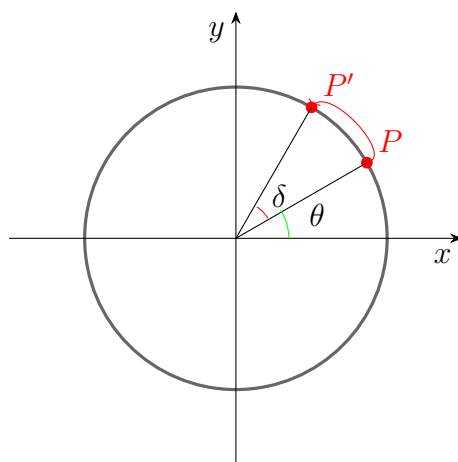


图 17.3: **旋转圆上的点**。我们把旋转看作是坐标轴和坐标原点没动, 几何点的变动。

如果我们定义旋转对应的坐标变换操作, 我们还可以证明, 一个圆无论旋转多少度, 其和自身重合。这是下面拓展知识里面的内容。目前, 这部分知识不做要求。等到我们学习了矢量会重新来正式学习旋转。

17.2 圆的旋转 *

现在我们来考虑圆的旋转的时候, 由于我们不会移动圆心, 我们同样在单位圆——把圆心放到坐标原点——里面来讨论。这里, 我们采用变换几何点不改变坐标轴和原点的视角。我们要找到围着原点来转动一个几何点 $P = (x_P, y_P)$, 成为新的几何点 $P' = (x_{P'}, y_{P'})$ 。我们要发现它们之间的关系 $x_{P'} = f(x_P, y_P)$, $y_{P'} = g(x_P, y_P)$ 。然后我们要来证明, 当 (x_P, y_P) 满足以原点为圆心的方程的时候, $(x_{P'}, y_{P'})$ 也满足同一个方程。这时候, 我们就证明了旋转之后的圆和旋转之前的圆重合——任何一个旋转之前在圆上的点旋转之后仍然在同一个圆上, 以及反过来, 任何一个旋转之后在圆上的点, 旋转之前肯定在同一个圆上。

如图 17.3 所示, 我们把原来在圆上的点 $P = (x_P, y_P)$ 转动到 $P' = (x_{P'}, y_{P'})$, 中间转动的角度为 δ 。如果原来 P 和 x 轴正方向的夹角为 θ , 则

P' 和 x 轴正方向的夹角为 $(\theta + \delta)$ 。这个时候， P' 点的坐标可以表示为，

$$x_{P'} = r \cos(\theta + \delta), y_{P'} = r \sin(\theta + \delta). \quad (17.9)$$

利用和角三角函数，我们得到

$$x_{P'} = r \cos \theta \cos \delta - r \sin \theta \sin \delta = x_P \cos \delta - y_P \sin \delta, \quad (17.10)$$

$$y_{P'} = r \sin \theta \cos \delta + r \cos \theta \sin \delta = y_P \cos \delta + x_P \sin \delta. \quad (17.11)$$

经过计算，我们可以验证这个新点 P' 仍然还在原来的圆上，

$$x_{P'}^2 + y_{P'}^2 = r^2. \quad (17.12)$$

其实以上的计算有一个特别直观的解释：围绕原点的旋转不改变任何一个点到圆心的距离，而圆的定义就是到圆心（这里就是原点）的距离为一个固定长度的点的集合，因此，旋转以后的点到圆心的距离仍然和旋转之前其到圆心的距离相等，因此，仍然在圆上。不过，在这里，我们通过计算来证明这一点，还是比观察结论要可靠的。

圆围绕其圆心旋转以后重合这个结论其实很重要。这样我们就可以把圆上的两个点一起旋转，可以发现，这两个点的相对关系——例如用这两个点之间的角度来度量——是不变的。这其实意味着同样的夹角对应的圆弧的长度不变，因为你总是可以把那两段圆弧重合起来。我们把这个结果写成一个定理。

定理 17.2: 等圆心角对等弧

在圆上同样大小的圆心角对应的弧长相等。

同样的道理，等弧对应的圆心角相等。

定理 17.3: 等弧对等圆心角

在圆上相等的弧长对应圆心角相等。

除了上面的依赖于旋转的证明方式，未来我们也可以基于距离公式来用微积分计算弧长，从而证明这个等圆心角对等弧。

扇形的面积和弧长的公式，按照整个圆的周长和面积的公式来等比例划分，可以得到

$$l_{\theta^\circ} = \frac{2\pi r}{360^\circ} \theta^\circ, \quad (17.13)$$

$$S_{\theta^\circ} = \frac{\pi r^2}{360^\circ} \theta^\circ. \quad (17.14)$$

之前我们引入了弧度的定义——见算式 (13.17)。利用弧度来度量角度，有

$$l_\theta = r\theta, \quad (17.15)$$

$$S_\theta = \frac{1}{2}r^2\theta. \quad (17.16)$$

17.3 圆中三角形的相关定理

前面我们已经定义了圆心角为圆心到圆上两点到的半径构成的角，圆周角为圆上两点到圆周上任意一点构成的角。

定理 17.4: 圆周角圆心角

一条弧所对的圆周角等于它所对的圆心角的一半。

证明 17.2. 我们用数学语言把待证明命题表述为：在圆 O 中， AB 是一条弦，证明其对应的圆心角 $\angle AOB$ 和圆周角 $\angle AMB$ 满足 $\angle AOB = 2\angle AMB$ 。

连接 OM 反向延长交圆 O 于 P ，根据外角等外不相邻的内角之后以及等腰三角形底角相等，可以得到 $\angle AOP = 2\angle AMP$ ， $\angle BOP = 2\angle BMP$ 。圆心角 $\angle AOB$ 要么等于 $\angle AOP$ 和 $\angle BOP$ 之和，要么等于两者之差。不失一般性，我们选两者之和的情况， $\angle AOB = \angle AOP + \angle BOP = 2\angle AMP + 2\angle BMP = 2\angle AMB$ 。

基于这个圆周角和圆心角的关系，可以得到以下两个命题。

习题 17.1 (直径对直角). 直径对应的圆周角是 90° 。

和

习题 17.2 (对角互补). 圆的内接四边形的对角互补。

证明留作练习。这两个命题的逆命题是什么，成立吗？也留作练习。

定理 17.5: 等弦对等圆心角

在同圆中，两段相同长度的弦对应的两个圆心角相等。

证明 17.3. 我们用数学语言把待证明命题表述为：在圆 O 中，弦长 $AB = CD$ ，证明其对应的圆心角 $\angle AOB = \angle COD$ 。

在三角形 $\triangle AOB$ 和 $\triangle COD$ 中 $OA = OB = OC = OD, AB = CD$ ，则 $\triangle AOB \cong \triangle COD$ (全等三角形 SSS 判定)，因此， $\angle AOB = \angle COD$ (全等三角形性质)。

再根据圆周角等于圆心角的一半，就可以得到等弦对等圆周角。类似地可以证明“等圆心角对等弦”、“等圆周角对等弦”、“等弧对等弦”、“等弦对等弧”。这些都留作练习。

定理 17.6: 相交弦定理

圆内的两条相交弦，被交点分成的两条线段长的积相等。

证明 17.4. 我们用数学语言把待证明命题表述为： AB 和 CD 是圆 O 的两条弦，已知其相交于一点 P 。证明 $AP \times PB = CP \times PD$ 。
根据同弧对应于等圆周角得到 $\angle CAP = \angle BDP, \angle ACP = \angle DBP$ ，因此 $\triangle ACP \sim \triangle DBP$ (相似三角形 AA 判定定理)，于是 $\frac{AP}{PD} = \frac{CP}{PB}$ ，也就是 $AP \times PB = CP \times PD$ 。

切线是一个非常重要的概念。未来我们会有专门研究切线的数学的子学科。圆上的切线是我们第一次正式学习切线。

定义 17.1. 切线 [切线] 圆上的切线指的是和圆有且只有一个交点的直线。

未来我们会把切线的概念拓展到和某个曲线在局部有且只有一个交点。我们来看看切线的判定和性质。

定理 17.7: 切线判定

经过半径的外端并且垂直于这条半径的直线是圆的切线。

证明 17.5. 我们用符号语言来表述待证明的命题： OH 是圆 O 的一条半径， $OH \perp AB$ 于 H ， $OH = r$ 为圆 O 的半径大小。证明 AB 是圆 O 的切线。

切线的定义是，一条和圆有且只有一个交点的直线。首先， H 在圆上 ($OH = r$)，也在 AB 上，因此 AB 和圆 O 有交点。我们接着来证明 AB 和圆 O 不可能有另一个交点。假设， AB 和圆 O 有另一个交点 P ，则 OHP 构成一个等腰三角形 ($OH = r = OP$)。等腰三角形底角相等，因此两个底角都是 90° 。合起来，加上顶角肯定大于 180° 。矛盾。因此，原假设错误， AB 和圆 O 不可能有另一个交点。因此， AB 和圆 O 有且只有一个交点。因此， AB 是圆 O 的切线。

切线的第一条性质是上面判定定理的逆定理。我们来看看其是否成立。

定理 17.8: 切线性质

圆的切线垂直于经过切点的半径。

证明 17.6. 我们用数学的符号语言来表述待证明的命题： AB 是圆 O 的切线， H 为其和圆 O 的唯一交点，我们来证明 $OH \perp AB$ 。

我们假设 OH 不垂直于 AB 。我们来做 $OP \perp AB$ 于 P ，然后延长射线 HP 到 AB 上的另一点 H' 使得 $H'P = HP$ 。在三角形 $\triangle HOP$ 和 $\triangle H'OP$ 中， $H'P = HP$ ， $\angle OPH = 90^\circ = \angle OP'H$ ， $OP = OP$ ，因此 $\triangle HOP \cong \triangle H'OP$ (全等三角形 SAS 判定定理)，于是 $OH = OH'$ 。因此， H' 同时在圆 O 上和在直线 AB 上。因此，直线 AB 和圆 O 有两个交点。这个和已知条件直线 AB 是圆 O 的切线矛盾。因此，原假设错误， $OH \perp AB$ 。

定理 17.9: 弦切角定理

弦切角等于它所夹的弧对的圆周角。

证明 17.7. 我们用数学的符号语言来表述待证明的命题： P 是圆 O 的外的一点， PA 是圆 O 的切线 A 是切点， AB 是圆 O 的一条弦。我们来证明 $\angle PAB = \frac{1}{2}\angle AOB$ 。

连接 OA ，延长 AO 交圆 O 于 M 。则 OM 是直径，并且 $OM \perp AP$ 于 A 。连接 BM 。根据 OM 是直径，得到 $\angle ABM = 90^\circ$ ，因此 $\angle AMB + \angle OAB = 90^\circ$ 。根据 $OM \perp AP$ 于 A 得到 $\angle PAB + \angle OAB = 90^\circ$ 。因此， $\angle AMB = \angle PAB$ 。再根据 $\angle AMB = \frac{1}{2}\angle AOB$ ，得到 $\angle PAB = \frac{1}{2}\angle AOB$ 。于是，弦切角等于它所夹的弧对的任意圆周角。

我们还可以把切线和割线——也就是弦的延长——联系起来。

定理 17.10: 切割线定理

从圆外一点引圆的切线和割线，切线长是这点到割线与圆交点的两条线段长的比例中项。

证明 17.8. 我们用数学的符号语言来表述待证明的命题： P 是圆 O 的外的一点， PN 是圆 O 的切线 N 是切点，连接 PO 交圆 O 于 B ，延长 PO 交圆 O 于 A 。我们来证明 $PN^2 = PA \times PB$ 。

连接 ON ，连接 AN ，连接 BN 。则根据弦切角定理有 $\angle PNB = \angle NAP$ ，同时 $\angle NPB = \angle APN$ ，因此 $\triangle PNB \sim \triangle NAP$ （相似三角形 AA 判定定理），因此 $\frac{PB}{PN} = \frac{PN}{PA}$ ，也就是 $PN^2 = PA \times PB$ 。

17.4 共圆定理：把不是圆的问题变成圆的问题

把其他几何图形中的问题看作圆的问题，有的时候能够简化问题。在这一节中，我们给出几个这样例子。

定理 17.11: 直角三角形斜边中线

直角三角形斜边上中线的长度等于斜边的一半。

证明 17.9. 我们用数学语言把待证明命题表述为：在直角三角形 $Rt\triangle ABC$ 中， $\angle C = 90^\circ$ 。D 是 AB 中点。证明： $CD = AD = DB$ 。过 $DH \perp AC$ 于 H。于是 $DH \parallel BC$ (平行线判定)。因此 $1 = \frac{AD}{DB} = \frac{AH}{HC}$ (平行基本定理)。因此在三角形 $\triangle ADC$ 中，H 既是 AC 上的高也是 AC 上的中线，因此 $\triangle ADC$ 是等腰三角形 $CD = DA$ (等腰三角形三线合一逆定理)。

最后一步，如果不用等腰三角形三线合一逆定理，也可以用全等三角形来完成。如果我们把这个直角三角形看作圆内的直角三角形，这个证明会简单很多。

证明 17.10. 我们用数学语言把待证明命题表述为：在直角三角形 $Rt\triangle ABC$ 中， $\angle C = 90^\circ$ 。D 是 AB 中点。证明： $CD = AD = DB$ 。

不在同一直线上的 A, B, C 肯定存在一个通过这三点的圆。根据圆内“直角对直径”，AB 必然是直径，则 AB 的中点 D 是这个圆的圆心。因此，A, B, C 到 D 的距离相等，也就是 $AD = BD = CD$ 。

对比两个证明，可以发现，利用圆的证明简单很多。

类似地我们可以证明其逆定理，

定理 17.12: 直角三角形斜边中线逆定理

如果一个三角形斜边上中线的长度等于斜边的一半则这个三角形是直角三角形。

证明 17.11. 我们用数学语言把待证明命题表述为：在三角形 $Rt\triangle ABC$ 中， $CD = AD = DB$ ，求证 $\angle C = 90^\circ$ 。

根据 $AD = DB = CD$ ，ABC 在以 D 为圆心的圆上，因此 AB 是直径， $\angle ACB$ 是直径对应的圆周角，根据圆周角是圆心角（这里是 180° ）的一半，得到 $\angle ACB = 90^\circ$ 。

为了更好地把其他几何图形问题转化成圆的问题，我们可以来证明几

个转化用到的定理。

定理 17.13: 四点共圆一

点 C, D 在线段 AB 的同侧, 如果 $\angle ACB = \angle ADB$, 则 A, B, C, D 四点共圆。

证明 17.12. 按照三点确定一个圆, 我们过 ABC 做一个圆。连接 AD 和这个圆相交于一点 E 。如果 D 就在这个圆上, 则 $E = D$ 。我们假设 E, D 不重合, 也就是 D 不在这个圆上。这有两种情况: D 在这个圆外面, D 在这个圆内部。

如果 D 在这个圆外面, 则在三角形 $\triangle BDE$ 中, $\angle AEB$ 是外角, $\angle ADB$ 是内角, 则 $\angle AEB > \angle ADB$ 。但是, 根据同弧所对圆周角相等得到 $\angle AEB = \angle ACB$, 又已知 $\angle ACB = \angle ADB$, 则 $\angle AEB = \angle ADB$, 矛盾。

如果 D 在这个圆内部, 则在三角形 $\triangle BDE$ 中, $\angle ADB$ 是外角, $\angle AEB$ 是内角, 则 $\angle ADB > \angle AEB$ 。但是, 根据同弧所对圆周角相等得到 $\angle AEB = \angle ACB$, 又已知 $\angle ACB = \angle ADB$, 则 $\angle AEB = \angle ADB$, 矛盾。

因此, 原假设错误, A, B, C, D 四点共圆。

定理 17.14: 四点共圆二

线段 AD 和线段 BC 相交于点 P , 如果 $AP \cdot DP = BP \cdot CP$, 则 A, B, C, D 四点共圆。

证明 17.13. 在三角形 $\triangle CPA$ 和 $\triangle DPB$ 中, $\frac{AP}{CP} = \frac{BP}{DP}$ (已知), $\angle CPA = \angle DPB$ (对顶角), 则 $\triangle CPA \sim \triangle DPB$ (相似三角形 SAS 判定), 因此, $\angle ACP = \angle BDP$ (相似三角形性质)。满足四点共圆判定定理一, 因此四点共圆。

例 17.1 (四点共圆应用). 在直角三角形 $Rt\triangle ABC$ 中, $\angle ACB = 90^\circ$, $AC = 4$, $BC = 3$. O 是 AB 边的中点。在平面上选一点 P 使得 $OP = OC$ 。连接 CP , 作 $CQ \perp CP$ 于 C , Q 是 CQ 和 PB 的交点。求 CQ 的最大值。

$\triangle ABC$ 是直角三角形, 则如果我们把 ABC 看作一个圆, AB 是其直

径,其中点是 O 圆心。因此 $OA = OB = OC$ 。已知 $OC = OP$, P 也在这个圆 O 上。因此, $ACBP$ 四点共圆。于是 $\angle CAB = \angle BPC$ (同弦对等圆周角)。已知 $CQ \perp CP$, $90^\circ = \angle PCQ$ 。于是三角形 $\triangle ABC \sim PQC$ (三角形相似 AA 判定)。因此, $\frac{QC}{CP} = \frac{BC}{AC}$, 得到 $QC = \frac{4}{3}PC$ 。

要求 QC 的最大值,我们只需要找出来 PC 的最大值。而 P 必须保证在圆 O 上, PC 必须是一条弦,其最大的长度为直径,那时候 $PC = AB = \sqrt{3^2 + 4^2} = 5$ 。因此, QC 的最大值为 $\frac{20}{3}$ 。

在这个问题的字面表述中,我们看不到圆。但是,无论相似三角形的证明过程,还是后来 PC 最大值的确定,我们都用了圆的信息。把一个问题变成另一个问题,或者说看到两个初步看起来不同的问题其实结构上是同一个问题,是数学的巨大作用之一。

17.5 推荐阅读材料

17.6 作业

习题 17.3 (直角对直径). 在圆中直角的圆周角对应的弦是直径。

习题 17.4 (对角互补则共圆). 一个四边形的对角互补则这个四边形的四个顶点共圆。

习题 17.5 (垂直于弦的直径). 垂直于弦的直径平分这条弦;弦的垂直平分线必定是直径。

习题 17.6 (同点切线等长). 从圆外一点引圆的两条切线,它们的长度相等。

习题 17.7 (乘法在数轴上的表示). 证明图 1.3 中所示的方法可以在数轴上表示任意一对实数的乘法,在假设实数都可以表示为数轴上的数的前提下。

17.7 本章小结

在本章中，我们以三角形全等的判定和性质、三角形相似的判定和性质、垂直的定义，圆的定义和这个定义的方程表示，有的时候还用上了反证法，得到了圆里面的一些几何对象的判定和性质，例如圆周角、圆心角、圆弧、弦、切线、弦切角。我们还看到了可以把一个看起来跟圆无关的问题转化为圆上的问题来简化这个问题的求解。

但是，在这里，我们要再一次强调：所有的这些圆上的具体对象的知识，其实都是第二位的，看到之前学习过的概念和命题可以用来研究圆上的几何对象的性质和命题，构建几何学的知识体系才是重要的，第一位的。

当然，切线的概念未来很重要。这可能是你们第一次接触这个概念。圆心角在我们的知识体系中是用来定义角的。因此，如果说还能学会点具体知识，那就除了圆的定义，再留下一点关于切线和圆心角的印象就可以了。

第十八章 平面几何：解析方式

The power of mathematics is often to change one thing into another, to change geometry into language.

数学的力量往往在于把一个东西变成另一个东西，例如把几何变成语言。

– Marcus du Sautoy

我们已经把实数和数轴上的点一一对应起来，把平面上的点和二维笛卡尔直角坐标系下的一对对的数联系起来，我们也已经学习过直线可以在二维笛卡尔直角坐标系下用 $y = kx + b$ 来表示。在这一章中，我们来学习更多的这样的代数表达式和几何图形的对应关系，以及用这个对应关系来解决几何问题，或者偶尔也用这个关系来解决代数问题。

18.1 代数表达式和几何图形

我们把之前已经学会过的代数表达式和几何图形之间的对应关系都复习一边，然后，我们来学习几个新的这样的对应关系。

首先，我们来把直线方程拓展为更一般地的形式： $lx + my = c$ 。

前面（13.1 节）我们已经学习过一次函数 $y = kx + b$ 和直线一一对应。首先，当 $k = 0$ 的时候，一次函数成了 $y = b$ 。把二维平面上这样的点 $(x, y = b)$ 都取出来，描成黑体，我们发现，就是平行于 x 轴的直线——把 x 轴向上平移了 b （如果实际上， $b = -|b|$ ，则向上移动负的，自然就成了向下移动）。

接着，按道理，我们应该去关心一下 y 轴的平移。我们发现， y 轴本身是 $(x = 0, y)$ 的所有的点的集合。写成代数表达式就是 $x = 0$ 。那么，更一般

的平行于 y 轴的直线，很自然就应该是向右平移 y 轴来得到，也就是 $x = c$ 。我们也可以通过描点法来把满足 $(x = c, y)$ 这个条件的所有点都找出来看一看。不过，这里有个问题，如何把一般的直线的代数形式 $y = kx + b$ 变成 $x = c$ 的形式呢？如果我们一定要尝试，可以这样，

$$x = \frac{1}{k}y - \frac{b}{k} \stackrel{?}{\Rightarrow} x = c \quad (18.1)$$

看起来必须 k 是一个比所有的 y 都要大的数才行，但是 y 是一个任意的数啊，而且 b 呢还跟这个比所有的数都要大的数成比例。将来我们会学习到，这个比任意给定的数都大的“数”是“无穷大”。但是，在你们现在的知识范围内这个“无穷大”和“跟无穷大成正比”的概念还解释不清楚。因此，我们只能说，所谓的直线的一般代数表达式 $y = kx + b$ 其实是不能用来表示 $x = c$ 的。这个最好用 $x = ky + b$ 来表示。同样，我们看到 $x = ky + b$ 不能用来表示 $y = c$ 的直线。

为了把 $y = kx + b$ 和 $x = ky + b$ 两种情况结合起来，我们可以把直线的代数表达式拓展为 $lx + my = c$ 。除了 $l = 0$ 或者 $m = 0$ 的特殊情况， $y = kx + b$ 和 $lx + my = c$ 等价，相当于重新定义了一下这些系数。

例 18.1 (直线的不同代数表达式). 除了 $l = 0$ 或者 $m = 0$ 的特殊情况，把代数表达式 $lx + my = c$ 变形为 $y = kx + b$ ，找到这样的变换的条件，以及相应的变换，也就是后面代数式的 k, b 和前面的代数式中的 l, m, c 的关系。

当 $m \neq 0$ 的时候，把 $lx + my = c$ 的两边同时乘以 $\frac{1}{m}$ ，可以得到

$$y = -\frac{l}{m}x + \frac{c}{m}. \quad (18.2)$$

因此，令 $k = -\frac{l}{m}, b = \frac{c}{m}$ 就可以得到等价的表达式 $y = kx + b$ 。

反过来，对于 $y = kx + b$ ，我们可以等价变形为 $kx + (-1)y = b$ ，或者更一般地

$$\lambda kx + (-\lambda)y = \lambda b. \quad (18.3)$$

也就是，只要 $\lambda \neq 0$ ，令 $l = \lambda k, m = (-\lambda), c = \lambda b$ 可以使得 $lx + my = c$ 和 $y = kx + b$ 等价。

注意这里的不同的 λ 值，只要 $\lambda \neq 0$ ，不会影响任何 (x, y) 的值。也就是说，任何两个不为零的不同的 λ 对应了同一个组点因此也就对应着同一

条直线，同一条曲线对应着所有的不为零的 λ 所确定的上面的方程。也就是说，一条直线其实对应了一类 l, m, c 。任何两组系数如果可以通过同时乘以一个不为零的常数变成相同——也就是相应成比例——都算同一类。在这个类的意义下，也就是把同一类看作同一个方程，则直线和方程仍然一一对应。

不过这个说法还是稍微有点不方便。并且只要排除 $k = 0$ 或者 $l = 0$ 的特殊情况， $y = kx + b$ 确实够用并且满足了直线和方程的一一对应关系。因此，在通常情况下，我们也接受“直线的一般代数表达式是 $y = kx + b$ ”这个不太准确的说法。实在有必要，我们就把另一种互补的情况 $x = ky + b$ 也用一下，就完全够了。现在我们来做一个需要用到这三个表达式的例题。

之前 (13.1 节) 我们已经用 $y = kx + b$ 在“不失一般性的条件下”证明了“两点确定一条直线”。现在有了直线的更一般的数学表达式，我们再来针对最一般的情况证明一次“两点确定一条直线”：平面上给定任何两个不同的点，确定了一条并且只有一条直线。有了“直线和直线方程一一对应”，我们只需要进一步证明两点确定了唯一的直线方程。

例 18.2 (两点确定一条直线). 证明过两点 (x_P, y_P) 和 (x_Q, y_Q) 有且只有一条直线。

我们先尝试用 $y = kx + b$ 来试试通过这两个点，也就是

$$\begin{cases} y_P = kx_P + b, \\ y_Q = kx_Q + b. \end{cases} \quad (18.4)$$

这是二元一次方程组，求解出来 k, b 得到

$$k = \frac{y_Q - y_P}{x_Q - x_P}, b = \frac{y_P x_Q - y_Q x_P}{x_Q - x_P}. \quad (18.5)$$

当 $x_Q - x_P = 0$ 的时候，这个表达式会出问题，分子是有限大小，分母是零。这个时候，实际上，我们就要用 $x = ky + b$ 来表示这个直线了，于是，

$$\begin{cases} x_P = ky_P + b, \\ x_Q = ky_Q + b. \end{cases} \quad (18.6)$$

求解出来 k, b 得到

$$k = \frac{x_Q - x_P}{y_Q - y_P}, b = \frac{x_P y_Q - x_Q y_P}{y_Q - y_P}. \quad (18.7)$$

这个表达式只有当 $y_Q - y_P = 0$ 的时候出问题。但是, $x_Q - x_P = 0$ 和 $y_Q - y_P = 0$ 不可能同时发生, 如果同时发生了就意味着 P, Q 实际上是同一个点, 而这个和我们要求的两个不同点矛盾。因此, 上面求出来的两个直线的表达式, 至少总有一个是成立, 不可能同时不成立。

下面, 我们来证明, 如果两者都成立, 也就是 $x_Q - x_P \neq 0$ 和 $y_Q - y_P \neq 0$, 其实它们是同一个表达式。

$$y = \frac{y_Q - y_P}{x_Q - x_P}x + \frac{y_P x_Q - y_Q x_P}{x_Q - x_P} \quad (18.8)$$

$$\implies \frac{x_Q - x_P}{y_Q - y_P}y = x + \frac{x_Q - x_P}{y_Q - y_P} \frac{y_P x_Q - y_Q x_P}{x_Q - x_P} \quad (18.9)$$

$$\implies x = \frac{x_Q - x_P}{y_Q - y_P}y + \frac{y_Q x_P - y_P x_Q}{y_Q - y_P}. \quad (18.10)$$

因此, 合起来我们证明了, 两点确定一条直线。

这个时候, 这条直线的一般表达式为

$$\lambda(y_Q - y_P)x + \lambda(x_P - x_Q)y = \lambda(y_Q x_P - y_P x_Q). \quad (18.11)$$

注意到 λ 只要不为零, 不影响直线上的点的 (x, y) 值。因此, 这个方程(类)对应着一条唯一的直线, 这个直线由 $P = (x_P, y_P)$ 和 $Q = (x_Q, y_Q)$ 唯一确定。

上面的证明方法是纯代数计算形式的, 也就是说, 我们把直线表达式完全地计算了出来, 从给定的两点开始。这个计算过程, 未来对于从给定的两点得到直线表达式是非常有用的。此外, 通过计算来证明, 尤其是通过解方程来证明, 往往比通过定理的逻辑演绎操作起来更简单。第一, 这说明了计算就是演绎推理; 第二, 实际上, 这也启发了定理的机器证明的研究[5, 6]。

这样的把几何问题转化为代数问题, 采用代数计算的方法来求解问题的方式就成为解析几何的方式。在本书中, 我们其实不区分纯几何和解析几何, 因为我们的基本几何对象和基本几何关系, 包含点、线、圆、角、平行、垂直都是通过代数表达式来定义的。之前的平面几何部分的很多命题的论证过程也是通过代数计算来进行的。

其他的几何对象的代数表示, 也就是解析几何表示, 不需要做拓展, 我们只需要列在这里。角由两条相交直线构成, 也就是交点 (x, y) 同时满足以

下两个直线方程¹,

$$y = k_1x + b_1, \quad (18.12)$$

$$y = k_2x + b_2. \quad (18.13)$$

当两条直线斜率相同 $k_1 = k_2$ ($b_1 \neq b_2$, 否则两条直线是同一条), 没有交点, 两条直线平行。相交角的大小由这两条直线的斜率通过下式决定 (算式 (15.15)),

$$\tan \alpha = \frac{\tan \theta_1 - \tan \theta_2}{1 + \tan \theta_1 \tan \theta_2} = \frac{k_1 - k_2}{1 + k_1 k_2}. \quad (18.14)$$

已知 k_1, k_2 可以得到两条直线夹角 α 的正切值 $\tan \alpha$ 。然后, 我们通过正切值对应就能找到角度的大小, 因为正切函数在 $[0, 90^\circ]$ 范围内函数值和角度之间存在着——对应关系。

尤其是, 当两直线垂直的时候, $\alpha = 90^\circ$, $\tan 90^\circ = \infty$, 刚好对应着 $k_1 k_2 + 1 = 0$ 的时候——那时候分母为零, 分子有限, $\frac{k_1 - k_2}{1 + k_1 k_2} = \frac{k_1 - k_2}{0}$ 。一个有限数除以 0 等于无穷大²。

当 $k_1 = k_2$ 的时候, $\tan \alpha = 0$, 也就对应着 $\alpha = 0^\circ$, 正好就对应着两条直线平行——夹角为 0° 。这样我们就完全定义了角度和角度的大小。

原则上, 我们也可以把三角形写成代数表达式——三条相互不平行而且不交于一点的直线所确定的三个交点和三角线段围出来的封闭区域。有了直线的方程, 我们就不明确列出来三角形的方程了。

圆是我们关心的另一个几何对象, 由圆心 $O = (x_O, y_O)$ 和半径 r 确定, 其代数方程为算式 (17.2), 我们也重新写在这里:

$$(x - x_O)^2 + (y - y_O)^2 = r^2. \quad (18.15)$$

至此, 我们就有了点、直线、角、圆、平行、垂直这些对象和关系的解析表达式了。

¹一旦问题中出现 $x = c$ 平行于 y 轴的直线, 我们用 $x = ky + b$ 重新讨论一遍, 或者直接用 $lx + my = c$ 重新讨论一遍。

²未来会更加严格地学习到无穷大。目前, 可以这样来看, 假设 $k_1 - k_2$ 不变, 随着 $k_1 k_2$ 越来越接近 -1 , $\frac{k_1 - k_2}{1 + k_1 k_2}$ 肯定会越来越大, 也就越来越接近 $\tan 90^\circ$ 。

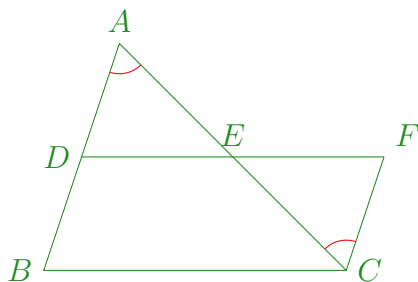


图 18.1: 三角形中位线定理。几何方法证明三角形中位线定理。其中需要添加的辅助线 EF 有一定难度。

18.2 通过把几何问题转变成代数表达式来解决几何学问题

我们来用上面的解析几何知识，也就是把几何图形写成代数表达式的知识，来解决一个完全是几何图形的问题。

例 18.3 (三角形中位线定理的解析几何). 三角形中位线平行于底边，并且其长度是底边的一半。

我们用数学的符号语言把待证明命题表述为：在任意三角形 $\triangle ABC$ 中， D, E 分别是 AB 和 AC 的中点，证明 $DE \parallel BC$ ，并且长度 $DE = \frac{1}{2}BC$ 。

我们先用纯几何方法来证明这个命题。延长 DE 到 F 使得 $EF = DE$ 。在三角形 $\triangle CEF$ 和 $\triangle AED$ 中， $AE = EC, EF = ED, \angle CEF = \angle AED$ ，因此， $\triangle CEF \cong \triangle AED$ (三角形全等 SAS 判定)。于是， $\angle EAD = \angle ECF$ ， $FC = AD$ (全等三角形性质)。因此， $AB \parallel FC$ (平行线的判定)。又有 $FC = AD = DB$ ，得到四边形 $BCFD$ 中， $DB = FC, DB \parallel FC$ 。因此， $BCFD$ 是平行四边形 (平行四边形判定)。于是， $DF = BC, FD \parallel BC$ (平行四边形性质)。再根据 $DE = EF = \frac{1}{2}DF$ (已知)，得到 $DE = \frac{1}{2}BC, DE \parallel BC$ 。

在这个证明中，我们先做了辅助线 EF, FC (其中 EF 的添加有一定难度想到)，然后利用了全等三角形的判定和性质，再用了平行四边形的判定和性质。注意，如果还原到等基础的知识，实际上平行四边形的判定和性质也需要用到多次全等三角形。因此，这里实际上相当于多次使用全等三

角形的判定和性质。而全等三角形的判定来自于余弦定理和正弦定理，而余弦定理和正弦定理来自于直线和角的代数表达式。

有了解析几何，我们试试直接用直线和角的代数表达式来证明这个命题。我们先求出来 D, E 的坐标，再计算出来过 D, E 的直线的代数表达式，然后和直线 BC 的代数表达式去相比较，看是否有相同的斜率。然后，再计算出来 DE 的长度，和 BC 的长度做比较。下面我们来完成这个计算。

我们先来建立一个坐标系：不失一般性，以 B 为坐标原点， BC 方向为 x 轴正方向。于是，我们把 A, B, C 分别表示为，

$$(x_A, y_A), (0, 0), (x_C, 0). \quad (18.16)$$

用这些信息，我们来计算 D, E 的坐标。我们先来计算直线 AB 的代数表达式，然后在这条直线上，选择一个点 D 使得 $BD = \frac{1}{2}BA$ 。用前面例 18.2 中的方法，我们得到直线 AB 的代数表达式为

$$y = \frac{y_A}{x_A}x. \quad (18.17)$$

于是 D 的坐标为

$$\left(x_D, \frac{y_A}{x_A}x_D\right). \quad (18.18)$$

下面我们来确定 x_D ，其条件是

$$\begin{aligned} BD = \frac{1}{2}BA &\Rightarrow \sqrt{x_D^2 + \left(\frac{y_A}{x_A}x_D\right)^2} = \frac{1}{2}\sqrt{x_A^2 + y_A^2}, \\ &\Rightarrow \frac{x_D}{x_A}\sqrt{x_A^2 + y_A^2} = \frac{1}{2}\sqrt{x_A^2 + y_A^2}, \\ &\Rightarrow x_D = \frac{1}{2}x_A. \end{aligned} \quad (18.19)$$

于是 D 的坐标为

$$\left(\frac{1}{2}x_A, \frac{1}{2}y_A\right). \quad (18.20)$$

同理可得 E 的坐标为，

$$\left(\frac{1}{2}(x_A + x_C), \frac{1}{2}y_A\right). \quad (18.21)$$

我们再来计算 DE 的代数表达式，会发现， D, E 的 x 坐标不同，但是 y 坐标相同，于是直线为 $y = y_A$ ，其斜率为零，也就是平行于 x 轴 $y = 0$ 。我们再来计算 DE 的距离，

$$DE = \sqrt{\left(\frac{1}{2}x_A - \frac{1}{2}(x_A + x_C)\right)^2 + \left(\frac{1}{2}y_A - \frac{1}{2}y_A\right)^2} = \frac{1}{2}x_C = \frac{1}{2}BC. \quad (18.22)$$

完毕。

前面我们已经看到，不使用解析几何，三角形的中位线定理是一个比较难证明的定理。我们发现，一旦用了解析几何，这个就很简单。

你还可以尝试一下下面的几个问题。你会发现，解析几何的证明在这几个问题上都比平面几何简单很多。当然，也有一些代数问题，通过几何的方法来求解会更简单。

习题 18.1 (三角形三条中线交于一点). 用平面几何和解析几何的方法证明三角形三条中线交于一点。

习题 18.2 (三角形三条高交于一点). 用平面几何和解析几何的方法证明三角形三条高交于一点。

18.3 通过把代数表达式转变成几何问题来解决代数问题

我们来用上面的解析几何知识，也就是把代数表达式看作几何图形的知识，来解决一个完全是代数的问题。

例 18.4 (方程的解的几何). 已知如下两个方程无解 $x^2 + y^2 = 4, y = x + b$ ，求其中的系数 b 的取值。

传统的代数解法如下：

$$\begin{aligned} x^2 + (x + b)^2 = 4 &\Rightarrow 2x^2 + 2bx + b^2 - 4 = 0, \\ &\Rightarrow \Delta = (2b)^2 - 4 \times 2(b^2 - 4) < 0, \\ &\Rightarrow b \in (-\infty, -2\sqrt{2}) \cup (2\sqrt{2}, \infty). \end{aligned} \quad (18.23)$$

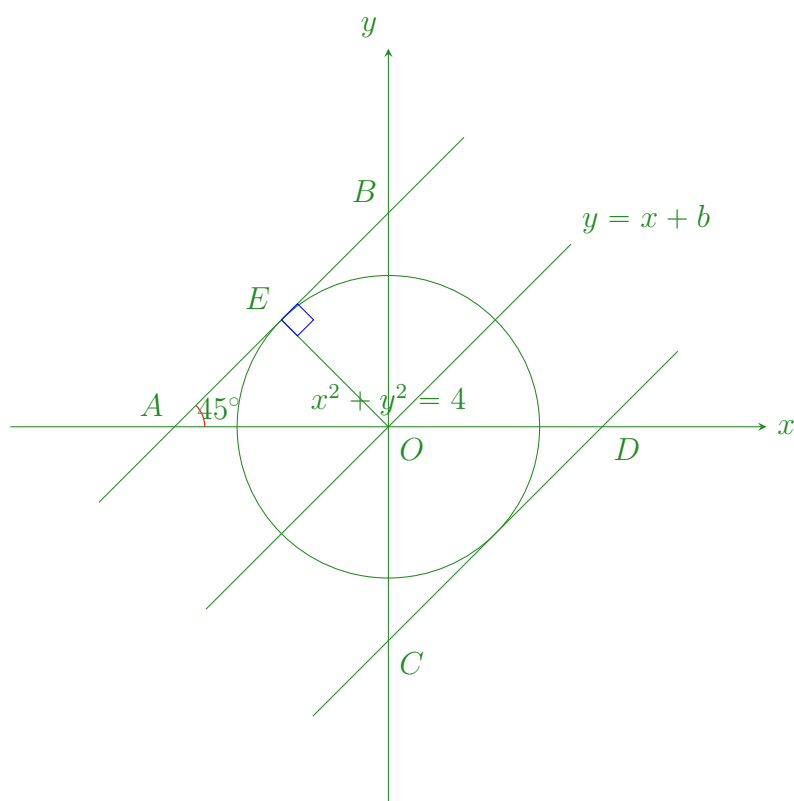


图 18.2: 圆和直线的交点。随着截距的不同, 直线 $y = x + b$ 和圆 $x^2 + y^2 = 4$ 的交点可以有两个、一个和零个。

现在我们来用几何解法。我们注意到 $x^2 + y^2 = 4$ 是一个以原点为圆心半径为 2 的圆, $y = x + b$ 是 45° 的直线。我们画成图。很自然就会发现, 随着直线截距的变化, 两者的交点会发生变化。两点就是同时满足这两个方程的解。因此, 我们只需要找出来直线和圆只有一个交点, 也就是直线和圆相切的情形, 就能够得到合适的参数 b 的边界值。

于是问题就成了完全一个几何题, 给定一个以原点为圆心半径为 2 的圆, 求其 45° 方向的切线。如图 18.2 所示, 我们发现, 在直角三角形 $\triangle AOB$ 中, OE 是垂足, 并且 OE 的长度是 2。按照 45° 的信息, 我们发现, $\angle ABO = 45^\circ$ 。于是, $\triangle AOB$ 中是等腰直角三角形, 直角边长为 $|b|$ 。于是, 三角形 $\triangle OEB$ 也是一个直角等腰三角形。其直角边为 2。因此, 其斜边长度为 $|b| = 2\sqrt{2}$ 。进一步, 我们发现, 如果我们让 $b < -2\sqrt{2}$, 则这里的直线和圆就没有交点了。

同理我们对直角三角形 $\triangle COD$ 做计算, 就发现, 当 $b > 2\sqrt{2}$ 时这里的直线和圆就没有交点了。合起来, 满足要求的 b 是, $(-\infty, -2\sqrt{2}) \cup (2\sqrt{2}, \infty)$ 。

我们发现两种解法得到的答案一致。你可能会觉得, 是不是代数解法更简单呢? 可能是的。套用一元二次方程根的判别式就可以。但是, 你也会感受到, 从几何图形来理解这个问题更清晰: 就是通过改变截距来移动一条直线, 在这个过程中看直线和圆的交点。未来遇到更加复杂的计算问题, 甚至可能到底如何计算不是特别明显的时候, 用几何图形来辅助思考, 来形成更加深刻的认知, 甚至一定程度上的图形认知、直觉认知, 对于解决问题也是有意义的。

此外, 在之前的一元二次方程、一元二次不等式(组)、二元一次方程方程组、二元一次不等式(组)、一元一次不等式(组)的求解中, 我们也已经体会到了几何方法对代数解法的指导意义。因此, 在解决问题的时候, 代数解法、解析几何解法、纯几何解法, 本来就是需要相互转化相互配合的。这也被称为“数形结合”。其实任何一个数学分支学科, 都是应该尽可能地相互结合起来来解决问题的, 只不过有的时候这个结合能看到能想到, 有的时候不容易看到。

18.4 推荐阅读材料

18.5 作业

18.6 本章小结

在本章中, 我们复习和建立了几种几何对象和几种关系——包含点、线、角、圆、平行、垂直、相交的夹角——的代数表达式, 然后用这些联系来求解了几个例题, 包含用代数方法求解几何问题和用几何方法求解代数问题。我们可以把这样的解题方法粗略地称作数形结合。未来, 我们会学习到更多几何对象和代数表达式之间的对应关系, 例如椭圆、双曲线、抛物线、其他函数的曲线和表达式、球、椭球等等, 于是可以更多地用数形结合来求解更多的更复杂的问题。

但是, 其实, 这不是本章真正的目的, 那仅仅是过程。本章的目的是帮

助你进一步理解到：代数式和几何图形本来就是一个东西。在咱们的平面几何体系中，点本来就是通过坐标来定义的，更加复杂的几何对象，也不过就是点的集合，用坐标表达的点的集合。在咱们的平面几何体系中，直线是通过距离来定义，而距离公式被当作了整个几何学知识的基础。你看距离公式，自然就是一个代数表达式。因此，代数表达式和几何图形之间本来就没有区隔。

在这个角度，未来如果距离公式改了，那自然整个几何学的知识也就相当于被改了。如果你未来会学习到非欧几何，其根源就在改变了距离公式。咱们这样的学习，用简单的知识，来体会到高层知识，高视角的对这个学科是什么的理解，可以给未来的学习和创造做更好的铺垫。

第十九章 平面几何：综合

In mathematics the art of proposing a question must be held of higher value than solving it.

– Georg Cantor

19.1 数学解题 WHWM 四问和概念地图用于几何证明

19.2 几个平面几何知识体系赏析

知识梳理的目标：用最少数量的公理，结合定义，用严格的逻辑演绎证明，给出所有的平面几何的定理。顺便，看看去掉或者题换掉其中的一部分公理，整个知识结构会怎样。

19.2.1 Euclid 体系

[7]

19.2.2 Hilbert 体系

[8]

19.2.3 SMSG 体系

[9]

George Birkhoff (乔治·伯克霍夫) 的平面几何公理体系, 以及基于 Birkhoff 公理体系发展而来的 School Mathematics Study Group (学校教学用数学研究小组, SMSG) 公理¹

19.2.4 本书的体系

平面几何概念地图, 本书自己的, 其他体系的用其他颜色显示

我们来梳理一下本书的平面几何的知识体系, 然后来跟其他的平面几何的知识体系做一个对比。从距离公式 (勾股定理) 开始我们建立了本书的所有定义和定理。任何定理、定义和公理都可以受到现实世界的启发, 但是定理不能停留在通过了有限次数的和有限精度的现实世界的检验, 必须是建立在一个公理化体系的内部, 被严格证明的。

19.3 推荐阅读材料

19.4 作业

19.5 本章小结

在使用平面几何知识解决问题的时候, 我们仍然使用 WHWM 思维方式。也就是问: 这个问题中有哪些对象——只不过这里是几何对象或者几何对象的代数表达式, 这些对象之间有什么关系, 这些关系用什么数学表达式来表达, 为什么要关心这些对象, 为什么这些对象之间有这样的关系, 为什么这些关系用这样的表达式来表达, 以及完成这个问题之后的总结反思——从解决这个问题中我学会了什么。WHWM 思维方式的背后是分解和综合, 是看到对象, 看到对象之间的关系, 看到概念和场景的联系, 看到做题和思维提升之间的联系。这个思维方式对于提出和解决问题是非常有帮助的。

¹SMSG 公理的全文可以从 <http://www.math.uaa.alaska.edu/~afmaf/classes/math305/SMSG.html> 看到。关于平面几何公理以及 SMSG 公理的更多信息可以参考 Wikipedia “Foundations of geometry” 词条。

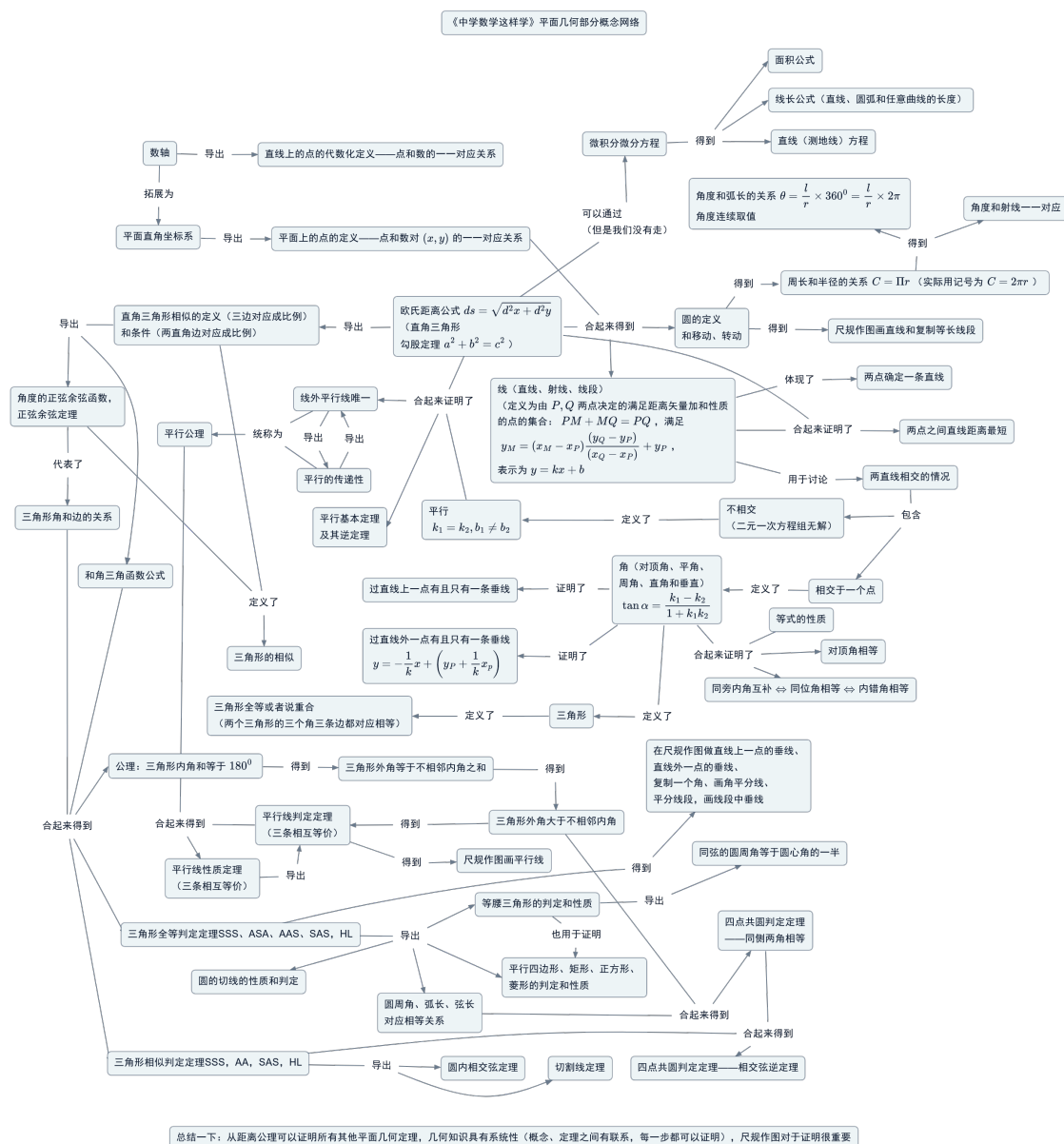


图 19.1: 本书平面几何的知识体系。从距离公式 (勾股定理) 开始我们建立了本书的所有定义和定理。这张图采用折页打印, 要足够大。

本书的本章之前的几何部分主要呈现了我们构建的几何学的知识体系。在本章中，我们还希望帮助学习者跳出来这个我们选择的平面几何的知识体系，看到其他的平面几何的知识体系，进行对比、欣赏、批判。

第二十章 概率论初步

在这一章中，我们来给一次操作或者一个事件可能出现多个可能结果的这一类现象提供数学语言——概率。我们还会把其中的事件之间的两种简单的典型关系（独立性和互斥性）也数学化——也就是用含义明确符号来表示。

20.1 古典概型的例子

一个硬币扔出去，忽略侧面立着的可能性，其最终落地以后的状态只能是正面（记作 H ）或者反面（记作 T ）向上。我们甚至往往会基于过去的观察，进一步提炼出来：下一次扔出去正面出现的可能性为 $P_H = \frac{1}{2}$ ，等于反面出现的可能性。其含义是，如果这样的“下一次扔出去”被重复了很多很多（记为 N ）次，则其中正面出现的次数（记为 M_H ）会正好或者至少接近总数的一半（ $M_H \approx \frac{1}{2}N$ ）。

一个骰子也类似：扔出去以后，一般情况下最终落地以后的状态只能是正面显示为 $1, 2, 3, 4, 5, 6$ 之一。我们甚至往往会基于过去的观察，进一步提炼出来：下一次扔出去某一记号为 n 的面出现的可能性为 $P_n = \frac{1}{6}$ ，每一面出现的可能性相同。其含义是，如果这样的“下一次扔出去”被重复了很多很多（记为 N ）次，则其中记号为 n 出现的次数（记为 M_n ）会正好或者至少接近总数的六分之一（ $M_n \approx \frac{1}{6}N$ ）。

注意最后这个命题是可检验的：只要我们重复这个“下一次扔出去”的操作或者说事件足够多次，我们可以算出来 $\frac{M}{N}$ ，看看这个数值和我们的猜测是不是相符，多大程度上相符。关于从实际数据中得到我们的猜测以及

把猜测和实际数据对比来判断多大程度上相符的这部分的数学知识，被称为统计学。我们也会稍后学习到。

或者，如果这个硬币或者骰子有偏被专门设计过，则我们可以修改相应的 P_H 和 P_n 的值，但是整体用这样一个“可能性”的符号和值来描述这个事情还是可以继续使用的。

我们来把这个受硬币和骰子的启发的 $P_n = \frac{1}{N}$ 的描述称为古典概型。我们来给这个古典概型一个定义。

定义 20.1 (古典概型). 对于一个可能出现 N 个结果的事件，如果我们假设每个结果的出现的可能性相同，并且任何一次事件的结果都只能是这 N 个结果之一，那么，我们称这个事件的数学模型为一个古典概型。具体地说，每一个事件 $n \in \{1, 2, \dots, N\}$ 出现的可能性为 $P_n = \frac{1}{N}$ 。

这个 $P_n = \frac{1}{N}$ 是怎么得到的呢？我们的第一个假设意味着

$$p_n = c, \forall n \in \{1, 2, \dots, N\}. \quad (20.1)$$

第二个假设意味着

$$\sum_{n=1}^N p_n = 1. \quad (20.2)$$

其中 1 约定为所有事件出现的概率，也可以表示为 100%。于是，根据算式 (20.1) 和算式 (20.2) 可以得到

$$p_n = \frac{1}{N}. \quad (20.3)$$

我们也可以拓展一下这个古典概型，仍然维持事件有有限个（记为 N ）可能的结果，只不过每一个结果的概率可能不同，也就是 p_n 不一定就是常数，但是仍然是给定的 $[0, 1]$ 之间的数，仍然满足算式 (20.2)。中学所学的概率都满足这个条件。

20.2 事件和事件集合的运算

对于一个硬币问题，除了问正面出现的概率、反面出现的概率，我们还可以问正面或者反面出现的概率。对于一个骰子，除了问编号为 n 的

面出现的概率，我们还可以问编号为 1 或者 2 出现的概率，编号为偶数（也就是 2, 4, 6）出现的概率，编号大于等于 3（也就是 3, 4, 5, 6）出现的概率，等等等等。因此，不是说，我们只关注单次事件的单个结果，也就是 $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$ 这些，我们还关注这样的结果出现的可能性： $\{1, 2\}, \{2, 4, 6\}, \{3, 4, 5, 6\}$ 等等等等。

这些更加复杂的结果显然是前面那个更加基本的单次事件的单个结果的某种组合，某种运算。我们能不能直接就把复杂的结果定义为简单的结果的数学计算呢？为此，我们把简单结果对应着的单次事件称为基元事件，把复杂结果对应着的事件称为复合事件，把简单事件和复合事件之间的关系定义为复合运算——这个复合运算表现为集合的子集。

定义 20.2 (基元事件). 记我们所关注的事件在单次发生的条件下的可能出现的各种结果——在中学只关注有限个结果的情况——的集合为 $\Omega = \{\omega\}$ ，称为样本空间。每次出现的结果对应着样本空间 Ω 中的任意一个元素构成的集合——也就是 $\{\{\omega\}, \omega \in \Omega\}$ 的元素——的事件称为基元事件。

例如，在硬币的例子中，样本空间 $\Omega = \{H, T\}$ ，基元事件是 $\{H\}, \{T\}$ 对应的事件（翻译成自然语言就是：出现的结果是 H 的事件，出现的结果是 T 的事件），基元事件的集合是 $\{\{H\}, \{T\}\}$ 。

习题 20.1 (扔骰子的基元事件). 参照硬币的例子，按照基元事件的定义，写出来扔骰子的样本空间、基元事件以及所有基元事件的集合。

有了基元事件我们来定义复合事件。注意，前面在概率论这部分的语言上我们已经把事件和集合统一。

定义 20.3 (复合事件). 在古典概型中，任意有限个基元事件（相应的集合）的交并补运算得到的集合被称作复合事件。

也可以把复合事件定义为

定义 20.4 (复合事件). 在古典概型中，复合事件定义为样本空间的任意一个子集。

我们来看看复合事件都是什么，然后来证明两者等价。还是以扔硬币为例。我们先从第一个定义出发。基元事件是 $\{H\}, \{T\}$ ，其交集为空集 $\{H\} \cap \{T\} = \phi$ ，继续做交集运算都只能得到空集；其并集为样本空间也就是全集 $\Omega = \{H\} \cup \{T\}$ ，继续做并集运算都只能得到全集 Ω ； $\{H\}$ 和 $\{T\}$

互为补集，新得到的集合 ϕ 的补集为全集 Ω ，全集 Ω 的补集为 ϕ 。至此不能再产生新的集合。因此，扔硬币的复合事件有： $\phi, \{H\}, \{T\}, \{H, T\}$ ，其集合为 $\mathcal{F}_\Omega = \{\phi, \{H\}, \{T\}, \{H, T\}\}$ 。

我们再从第二个定义出发。样本空间为 $\Omega = \{H, T\}$ ，其所有子集为 $\phi, \{H\}, \{T\}, \{H, T\}$ ，这些子集构成的集合为 $\{\phi, \{H\}, \{T\}, \{H, T\}\} = \mathcal{F}_\Omega$ 。

我们发现两者完全相同。更一般地，对于一个有限集合 Ω ，其所有单元元素集合通过任意有限次交并补运算得到的集合构成的集合正好就是 Ω 的所有子集构成的集合 $\mathcal{F}_\Omega$ 。我们只需要先证明 Ω 的任意一个子集都可以通过单元元素集合的有限次交并补运算得到，然后再证明单元元素集合的有限次交并补运算都只能得到 Ω 的一个子集，就得到了上面的命题。具体证明过程留作练习。

习题 20.2 (扔骰子的复合事件). 给出来扔骰子的复合事件的集合 $\mathcal{F}_\Omega$ 。可以考虑用两个定义分别给出，然后做一个对比。

习题 20.3 (子集来自于交并补). 对于有限集合 Ω ，其任意一个子集都可以通过其单元元素集合 $\{\omega\}$ 的有限次交并补运算得到。

习题 20.4 (交并补必然得到子集). 对于有限集合 Ω ，其单元元素集合 $\{\omega\}$ 的有限次交并补运算都只能得到 Ω 的一个子集。

有了复合事件的集合 $\mathcal{F}_\Omega$ ，我们就可以说事件了：事件就是 $\mathcal{F}_\Omega$ 的一个元素，也就是说不加修饰词的“事件”指的就是复合事件。注意，复合事件（的集合）实际上包含了基元事件（的集合），也就是 $\Omega \in \mathcal{F}_\Omega$ 。按照 $\mathcal{F}_\Omega$ 对其内部元素的交并补运算封闭， $\phi \in \mathcal{F}_\Omega$ 。

为什么我们要求事件的集合对交并补运算封闭呢？因为我们经常会这样说：假设已知事件 A 和事件 B 发生的概率，我们会问“事件 A 和事件 B 都发生的概率”，“事件 A 和事件 B 都不发生的概率”，“事件 A 或者事件 B 发生的概率”，分别相当于问：事件 A 和事件 B 的交集，事件 A 和事件 B 的交集的补集，事件 A 和事件 B 的并集。

例如，我们如果知道了骰子问题中偶数面出现的概率，我们可以问其补集——非偶数也就是奇数面出现的概率；如果还知道了大于等于 3 的面出现的概率，还可以问这两个事件同时出现的概率——比如说我的两位好朋友参与游戏，一位赌会出现偶数面，一位赌会出现大于等于 3 的面，我

想看看最后他们都赢的概率（于是看到两位开心的朋友）——那就去看这两个集合的交集，或者我关心至少其中一位赢了的概率（于是至少看到一位朋友开心）——那就去看这两个集合的并集。

顺便，设想一下，如果其中有一位朋友惹我生气了，我想看到他不开心另一位开心的概率，这时候我应该算的是哪个集合？留作作业。

习题 20.5 (“开心不开心”的集合运算). 我有两位朋友 A, B 参与了扔骰子游戏。记每一位朋友压注的集合分别为 S_A 和 S_B 。如果其中 A 惹我生气了，我想看到他不开心另一位开心的概率，这时候我应该关注的是哪个集合。得到一般公式以后，按照等概率古典概型，假设 A 压注了偶数面出现， B 压注了大于等于 3 的面出现，算出来你看到 A 不开心 B 开心的概率。

你看，从已知的概率和对应的事件开始，问这些事件的交并补得到的集合的概率，于是这些集合也就成了已知概率的集合，可以继续问这些事件以及这些事件和之前的事件合起来的事件的交并补的事件的概率，是一件特别自然的事情，在概率这件事情上。

20.3 概率：从事件集到 $[0, 1]$ 的映射

我们最后关注的概率的数值肯定是一个 $[0, 1]$ 的数，这个数是关于某事件，也就是 $\mathcal{F}_\Omega$ 中的一个元素 S 的。因此，概率其实是一个从 $\mathcal{F}_\Omega$ 到 $[0, 1]$ 映射 P ，满足

$$0 \leq P(A) \leq 1, \forall A \in \mathcal{F}_\Omega, \quad (20.4)$$

对于代表了所有可能出现的结果的集合 Ω 有

$$P(\Omega) = 1, \quad (20.5)$$

对于两个不相交的集合 $A \cap B = \emptyset, A, B \in \mathcal{F}_\Omega$ ，有

$$P(A \cup B) = P(A) + P(B). \quad (20.6)$$

这样我们就可以得到

$$P(\bar{A}) = 1 - P(A), \quad (20.7)$$

尤其是

$$P(\phi) = 0. \quad (20.8)$$

例 20.1 (计算空集的概率). 根据概率映射要满足的性质计算空集的概率 $P(\phi)$ 。

根据算式 (20.6), 对于一个集合 A 和它的补集 $\bar{A}$ 有 $A \cap \bar{A} = \phi$, 因此 $P(\bar{A}) = 1 - P(A)$ 。当 $A = \Omega$ 的时候, $\bar{A} = \phi$, 因此 $P(\phi) = 1 - P(\Omega) = 1 - 1 = 0$ 。

更一般地我们还可以通过基元事件的概率 P_ω 算出来任何一个复合事件的概率,

$$P(A) = \sum_{\omega \in A} P_\omega. \quad (20.9)$$

注意, 严格来说这里的 P_ω 其实是 $P_{\{\omega\}}$ 。

例 20.2 (从基元事件的概率算一般事件的概率). 证明算式 (20.9)。

取 A 中的每一个元素 $\omega \in A$, 构成一个单元素集合 $\{\omega\}$ 。首先, 任意取其中的两个不同的这样的集合, 如果有的话, 根据算式 (20.6), 有 $P(\{\omega_1\} \cup \{\omega_2\}) = P(\{\omega_1\}) + P(\{\omega_2\})$, 因为 $\{\omega_1\} \cap \{\omega_2\} = \phi$ 。如果这样的集合就不到两个, 则 S 本身已经满足算式 (20.9)。

同理, 取第三个这样的集合 $\{\omega_3\}$, 结合已经得到的关于 $\{\omega_1\} \cup \{\omega_2\}$ 的概率, 可以得到 $P(\{\omega_1\} \cup \{\omega_2\} \cup \{\omega_3\}) = P((\{\omega_1\} \cup \{\omega_2\}) \cup \{\omega_3\}) = P(\{\omega_1\}) + P(\{\omega_2\}) + P(\{\omega_3\})$ 。

不断重复这个过程, 直到穷尽 A 中的元素, 就得到了算式 (20.9)。

你可能认为, 这本来就是一个非常简单事实: 一个多个可能性的事件的概率是其中的各个可能性对应的的概率相加。我们把这个关系表示为集合以及定义了集合的分解式算式 (20.6) 之后, 未来对于更加复杂的, 没有单个可能性这个事件的时候, 例如从离散状态的概率过渡到连续取值的概率 (比如说, 两个人约会的时间点和由于堵车等不确定性导致的真正相见的时间点的问题) 的时候, 我们仍然可以使用集合的语言。

20.4 概率分布

对于基元事件出现的结果是有限个的情况,如果每一个结果可以用数来表示¹,那这个数就被称为随机数。因为有限个,每一个数和另一个数不连着,被称为离散随机数。这个时候概率就成了一个函数。例如,我们用 $1, 2, \dots, 6$ 来表示骰子的正面状态,则相应地我们就可以来计算 $P_1, P_2, \dots, P_6$; 我们用 $0, 1$ 而不是 H, T 来表示硬币的正面状态,则就有了 P_0, P_1 。

未来我们会遇到这些用来表示的数连续取值的情况,例如身高原则上可以取数轴上大于零的任何数。我们也可以把这样的连续取值的离散化:把取值范围找出来,均匀划分成若干个(记作 N)小区间,算出来一个代表了基元事件的数落在第 n 个区间里面的概率 P_n 。这样我们就又得到了一个离散取值的函数。

数学上严格的概率分布函数的定义为 $P(\xi \leq x)$ ——代表了基元事件的数落在区间 $(-\infty, x]$ 的概率。如果这个 x 刚好是第 n 个区域的边界,则前面的离散概率和这里的概率分布之间有如下的联系

$$P(\xi \leq x) = \sum_{l=1}^n P_l. \quad (20.10)$$

以后我们会看到,之所以定义为 $(-\infty, x]$ 是因为这样的定义可以同时适用于离散和连续变量。

在概率论中一旦有了随机数,还有几个重要的量值得一提:均值和方差。它们分别定义为

$$\mu_x = \langle x \rangle = \sum_{n=1}^N x_n P_n, \quad (20.11)$$

和

$$\sigma_x^2 = \langle (x - \langle x \rangle)^2 \rangle = \sum_{n=1}^N \left(x_n - \sum_{m=1}^N x_m P_m \right)^2 P_n, \quad (20.12)$$

¹最好还要求这个数的表示一定程度上保距离——非常接近的基元事件其数的表示也接近。不过在这里,我们暂时不关心这个细节。

20.5 独立事件和互斥事件

我们把两个不相交的集合 $A \cap B = \phi, A, B \in \mathcal{F}_\Omega$ 称作互斥事件。其字面含义是这两个事件不可能同时出现。对于互斥事件我们有算式 (20.6)。我们已经知道了互斥事件的例子，硬币的正反面，骰子的两面之间，骰子的奇数面和偶数面之间。

对于 $A, B \in \mathcal{F}_\Omega$ ，当

$$P(A \cap B) = P(A)P(B) \quad (20.13)$$

成立的时候，我们把 A, B 称为独立事件。

下面我们来列出几个独立事件的例子。例如，同时扔两个完全无关的两面出现的概率相同的硬币的时候， $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$ 。其中 “ (C_1, C_2) ” 的第一个字符表示第一个硬币的状态，第二个字符表示第二个硬币的状态。我们来计算下面的两个状态的概率， $A = \{(H, H), (H, T)\}$ ， $B = \{(H, T), (T, T)\}$ 。根据等概率假设以及出现的基元事件的数量占样本空间中总的基元事件的数量比例可以得到，

$$P(A) = \frac{2}{4} = \frac{1}{2}, \quad (20.14)$$

$$P(B) = \frac{2}{4} = \frac{1}{2}, \quad (20.15)$$

$$P(A \cap B) = \frac{1}{4}. \quad (20.16)$$

于是，我们确实有

$$P(A \cap B) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A)P(B). \quad (20.17)$$

也就是说，这里的 A, B 是独立事件。这里的 A 其实只关注了第一个硬币的状态，对第二个硬币处于什么状态其实是不关心的。这里的 B 其实只关注了第二个硬币的状态，对第一个硬币处于什么状态其实是不关心的。这样的事件之间的关系，再加上两个硬币之间确实无关，导致我们找到了满足事件独立性要求的两个事件 A, B 。

如果我们要求这两个硬币其实是状态完全相同的，则这时候样本空间是 $\Omega = \{(H, H), (T, T)\}$ 。我们来计算下面的两个状态的概率， $A = \{(H, H), (H, T)\}$ ，

$B = \{(H, T), (T, T)\}$ 。注意这里那些两个硬币不同的状态——例如 (H, T) ——其实不会出现。根据等概率假设以及出现的基元事件的数量占样本空间中总的基元事件的数量比例可以得到,

$$P(A) = \frac{1}{2}, \quad (20.18)$$

$$P(B) = \frac{1}{2}, \quad (20.19)$$

$$P(A \cap B) = P(\{(H, T)\}) = 0. \quad (20.20)$$

于是, 我们显然有

$$P(A \cap B) = 0 \neq \frac{1}{4} = P(A)P(B). \quad (20.21)$$

也就是说, 这里的 A, B 不是独立事件。

20.6 初识条件概率

对于一般的不一定独立的两个事件, 我们定义条件概率

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (20.22)$$

其含义为 B 已经发生的条件下 A 发生的概率。对于独立事件, 则

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A). \quad (20.23)$$

其含义为 B 发生的条件下 A 发生的概率和 B 是否已经发生没有关系。这正好就是独立的字面含义。

有了算式 (20.22), 我们可以得到一个看起来非常平庸但是实际上非常有用的关系式,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}. \quad (20.24)$$

其中除了条件概率的定义, 我们还用到了把事件拆分为两个互斥事件的和

$$B = (B \cap A) \cup (B \cap \bar{A}). \quad (20.25)$$

这个公式的具体证明过程留作练习。这个公式被称为贝叶斯公式。稍后我们会通过一个例子看到它是如此地应用广泛和含义深刻。其中的

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) \quad (20.26)$$

被称为全概率公式。

习题 20.6 (贝叶斯公式的证明). 证明贝叶斯公式算式 (20.24)。

例 20.3 (从穿裙子推断性别). 我在大街上看到一个穿裙子的人。这个人的性别是女性的概率是多少？需要哪些信息才能作出这个判断。

在这个问题中，我们想知道的是已经直到一个人穿裙子这个人是女性的概率，也就是条件概率 $P(S=f|C=s)$ 。这个时候，我们去采访一下大量的穿裙子的人，确认一下性别（假设人家会告诉咱们真实情况），做一个统计就好了。

另一个方法是，去看看男性和女性中有多少人穿裙子，也就是先得到一些反过来的条件概率的信息 $P(C=s|S=f)$ 和 $P(C=s|S=m)$ ，然后根据算式 (20.24)，有

$$P(S=f|C=s) = \frac{P(C=s|S=f)P(S=f)}{P(C=s|S=f)P(S=f) + P(C=s|S=m)P(S=m)}. \quad (20.27)$$

这时候我们发现，如果男性从来不穿裙子，则 $P(C=s|S=m)=0$ ，于是不管女性穿裙子的概率是多少，只要大于零，就得到 $P(S=f|C=s)=1$ ，也就是凡是看到的穿裙子的人都是女性。当然，实际上可能有一部分男性也是会传裙子的。比如说，概率为 $P(C=s|S=m)=0.01$ ，而女性在你观察的地方和气候时间点等情况下穿裙子的概率为 0.5，则我们就需要素出现在左右表达式的三个量（其中 $P(S=m)=1-P(S=f)$ ）。假设在你观察的地方男性女性出现的概率相同 $P(S=f)=0.5=P(S=m)$ ，则表达式可以进一步简化为

$$\begin{aligned} P(S=f|C=s) &= \frac{P(C=s|S=f)}{P(C=s|S=f) + P(C=s|S=m)}, \\ &= \frac{0.5}{0.5 + 0.01} = 0.98. \end{aligned} \quad (20.28)$$

注意，数值上 0.5 和 0.98 是很大的差别。因此，条件概率和反向的条件概率在数值上可不相等。

这个把条件概率的条件和结果转换了一下的公式为什么这么有用呢？因为它可以把某些便宜的信息变废为宝，来代替贵的难获得的信息。例如，我们把这里的女性改称得某一种疾病 D ，把穿裙子改成这个疾病的患者比较明显的一种症状 S 。通常的情况下，我们去医院检查其实是确认是不是有这个症状 S ，然后想通过这个症状来获得是否得了相应的疾病的信息，也就是我们想知道条件概率 $P(D|S)$ 。为此，我们可以提前做好一个这样的统计分析：找出来一大群具有症状 S 的人，看看其是否得了疾病 D 。同样地，往往我们也关心漏症的可能性，也就是 $P(D|\bar{S})$ 没有观测到症状 S 但是有 D 病的概率。这个时候，我们也需要做一个相关的统计分析来得到这个条件概率。

另外一个办法是通过贝叶斯公式算式 (20.24) 来从另外的那个概率值 $(P(D), P(\bar{D}) = 1 - P(D))$ 和两个条件概率 $(P(S|D) = 1 - P(\bar{S}|D), P(S|\bar{D}) = 1 - P(\bar{S}|\bar{D}))$ 得到这两个条件概率。

纯粹从数据处理的角度来说，两者没有区别。但是，从机理的角度来说，我们认为 $P(S|D)$ 一般来说是背后有机理的：一个得了疾病 D 的人，某些器官地方发生了异常，这些异常造成了症状 S 。如果我们可以把这些机理模型化，我们甚至就可以得到理论上的那堆需要的条件概率。同时，由于有机理，我们也可以认为从数值上这些条件概率更稳定，而 $P(D)$ 却可能随着我们关注的对象的基因和生活环境的改变而改变。因此，只要用上适合这个对象的 $P(D)$ ，就算从更普适的 $P(S|D)$ 开始，我们也能得到个性化的 $P(D|S)$ 。更进一步，有的时候 D 是不可直接观测的，但是从 D 到 S 的机理是已知的，那我们就更加依赖 $P(S|D)$ 这个条件概率的模型了。

习题 20.7 (从信号推断真值). 有一个信号发生器其自身的信号具有随机性：以 q 的概率为 $w = 1$ ，否则为零。它发出的信号也具有随机性，以 p 的概率发出等同于自身的信号 $s = w$ ，以 $1 - q$ 的概率发出跟自身相反状态的信号这里是 $s = 1 - w$ 。请你估计一下，当收到一个信号 s 的时候，这个信号发生器的真实信号是 w 的概率 $P(w|s)$ 。如果你收到了 n 个这样的信号 $s_1 s_2 \cdots s_n$ 呢？

20.7 概率论的应用

我们先来解决一个概率计算的问题和一个条件概率推断的问题。这两个问题将来会发展成为一个专门的工程学科——可靠性工程，一个专门分析和提升由多个元器件构成的系统的失效概率，并且找出来最需要提升的元器件的工程学科。

例 20.4 (串联元器件状态推断). 如图 20.1(a) 所示，假设有一个有两个元器件 A, B 通过串联构成的整体器件 S 。假设每一个元器件的失效概率为 $P(A = 0) = P(B = 0) = q$ 。请你计算整体器件失效的概率。如果我们发现整体器件失效了，请你推断其中每一个元器件失效的概率。

我们先来计算整体器件失效的概率 $P(S)$ 。 S 的状态由 A, B 决定： $S = AB$ 。也就是说，只要 A, B 有一个等于零——失效， S 就会等于零——失效。因此， S 的状态有两个，对应的 A, B 的状态分别为： $\{(A = 1, B = 1)\}$ (简化记号为 $\{(1, 1)\}$) 和 $\{(1, 0), (0, 1), (0, 0)\}$ 。因此，相应的概率为 $P(S = 1) = q^2$, $P(S = 0) = q(1 - q) + (1 - q)q + (1 - q)^2 = 1 - q^2$ 。

我们再来计算 $P(A = 0 | S = 0)$ 。按照贝叶斯公式算式 (20.24)，有

$$P(A = 0 | S = 0) = \frac{P(S = 0 | A = 0) P(A = 0)}{P(S = 0 | A = 0) P(A = 0) + P(S = 0 | A = 1) P(A = 1)}, \quad (20.29)$$

$$= \frac{1 \times (1 - q)}{1 \times (1 - q) + P(B = 0) P(A = 1)}, \quad (20.30)$$

$$= \frac{(1 - q)}{(1 - q) + (1 - q)q} = q. \quad (20.31)$$

其中我们用了 $P(S = 0 | A = 1) = P(B = 0)$ 。这是串联器件的性质——一个元器件没有失效的时候，整体器件的状态由另一个决定。

在调查敏感问题的问卷中，如果直接让被查询的对象回答敏感问题，例如问“你是否坐公共汽车逃过票”、“你是否讨厌你的数学老师”，有些受调查者可能会担心被知道自己的答案而不真心回答这些问题。为了让被查询对象放心提供其真实回答，我们可以这样做：让被查询对象自己抛一次硬币²，正面向上则回答问题“你是否坐公共汽车逃过票”，反面向上则再抛出

²或者也可以从一大堆所有的问卷中随机抽一张问卷，让被查询对象看到这些问卷问的

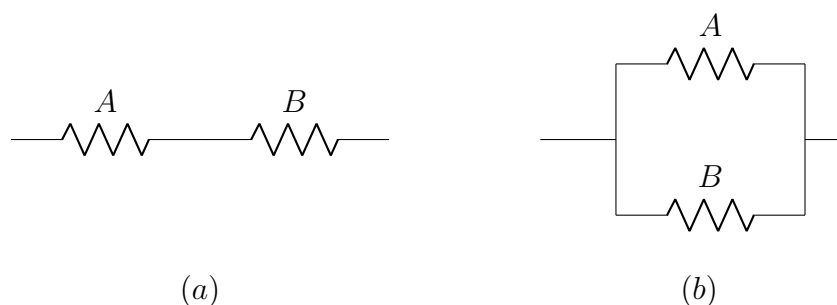


图 20.1: 串联和并联元器件。(a) 整体器件有两个元器件串联构成；(b) 整体器件有两个元器件并联构成。

一个硬币回答”这次你扔出来的硬币是正面吗”。我们来看看从这样的问卷设置的结果，能否通过单个被查询对象的答案推断出来，通过一大群被查询对象的答案能否知道我们想知道的信息。

如果被查询对象扔出第一个硬币以后，得到正面，则其回答的问题是我们真正关注的问题，其答案可能是“是”或者“否”，记作 $1, 0$ ，相应的概率为 $p_1, p_0 = 1 - p_1$ ；得到反面，则其回答的问题是下一次硬币是不是正面，其答案可能是“是”或者“否”，记作 $1, 0$ ，相应的概率为 $q_1 = 0.5, q_0 = 0.5$ 。这时候，我们不能从单个被查询对象的答案——我们只看到其答案为 1 或者 0 不知道其回答了哪一个问题——得到其对于敏感问题的答案。

但是，如果有一个很大的群体，都回答了这个问题，我们来看看从这个群体的答案中 1 的比例，记作 $\hat{P}_1$ ，是否能够得到我们关注的 P_1 的值。这里有一个概率论的问题和一个统计学的问题。我们先来解决其中的概率论的问题。

例 20.5 (敏感问题调查的概率论问题). 已知人们关于两个“是、否”两答案——记作 $(1, 0)$ ——问题 A 和 B 回答是 1 的概率分别为 P_1^A 和 P_1^B 。还知道随机选一个人之后，其选择回答 A 的概率是 q ，选择回答 B 的概率是 $(1 - q)$ 。请问，合起来，算一算其答案是 1 的概率 $P(s = 1)$ 。

按照全概率公式算式 (20.26)：

$$P(s = 1) = P(s = 1 | A) P(A) + P(s = 1 | \bar{A}) P(\bar{A}), \quad (20.32)$$

问题是不同的。

得到

$$P(s=1) = P_1^A \times q + P_1^B \times (1-q). \quad (20.33)$$

其中我们用到了 $P(\bar{A}) = P(B) = 1 - P(A) = 1 - q$ ，其含义是人们回答的问题不是 A 就是 B ，而且其概率分别为 q 和 $1 - q$ 。

换回到前面的扔硬币的记号，我们有

$$P_1 = \frac{1}{2}p_1 + \frac{1}{2} \times \frac{1}{2}. \quad (20.34)$$

于是，如果 P_1 已知，则

$$p_1 = 2P_1 - 0.5. \quad (20.35)$$

至于如果从对一大群人的调查中估计出来 P_1 ，也就是得到 $\hat{P}_1$ ，以及需要多大的人群才足够可靠，那就是统计学要解决的问题了。

概率论还有非常神奇的应用：它可以把非概率的问题转化为概率问题来研究。例如，通过一个叫做蒲丰投针的实验可以估算出来 π 的值——原本这是一个关于圆的性质的常数和随机事件没关系。遗憾的是讲清楚这个原理需要关于微积分的知识。例如，算一个曲线下的面积——原本这是微积分中的积分的任务，也可以通过扔随机数数一数有多少个随机数被包含曲线面积覆盖的区域——称作 Monte Carlo 方法——来完成。但是，遗憾的这还是需要关于微积分的知识。

20.8 推荐阅读材料

20.9 作业

习题 20.8 (并联元器件状态推断)，如图 20.1(b) 所示，假设有一个有两个元器件 A, B 通过并联构成的整体器件 S 。假设每一个元器件的失效概率为 $P(A=0) = P(B=0) = q$ 。请你计算整体器件失效的概率。如果我们发现整体器件失效了，请你推断其中每一个元器件失效的概率。

20.10 本章小结

在本章中，我们在古典概型和拓展了的古典概型的范围内，学习了什么是样本空间、事件、独立事件、互斥事件和概率映射，学会了怎么计算概率，学习了什么是条件概率以及贝叶斯公式，还通过例子体会了初步如何使用这些概念和公式。

在这个过程中，我们尤其体会了一下为什么复合事件的集合得满足对交并补运算封闭，得包含样本空间和空集，以及概率映射为什么会有 $P(\Omega) = 1$ 。复合事件的定义为什么是这个样子的，对于体会到什么概率，未来拓展到超越古典概型的概率，体会到什么是数学建模，都是非常重要的。其背后的思想仍然是——数学是思维的语言是描述世界的语言。只要描述世界有这些要求，我们就要在定义或者性质里面达成这些要求。

这里其实有一个藏起来的尾巴：未来我们会发现基于基元事件的等概率或者特定概率值来定义任意事件的概率是有问题的。有的场景中，等概率的基元事件的选择非唯一，甚至可能都不对。这个时候，怎么办？这是未来更为一般的概率论的定义所要解决的问题。不过这个问题由于需要微积分也只能等到以后再来学习了。

第二十一章 统计学初步

All knowledge is, in final analysis, history. All sciences are, in the abstract, mathematics. All judgements are, in their rationale, statistics.

在终极的分析中，一切知识都是历史；在抽象的意义下，一切科学都是数学；在理性的基础上，所有的判断都是统计学。

– C. R. Rao

统计学所解决的问题是如何从给定的关于随机事件数据来得到关于这个随机事件的性质的认知，来估计某个关于这个随机事件的性质的结论正确的可能性。在这一章中，我们来学习如何解决几个最简单的这样的问题。

21.1 经验分布图的绘制

为了解决这些问题，我们先来学会绘制经验分布图。经验分布指的就是从一堆给定数据画出来一个每一个数据区间上有多少个点或者说有占有点的百分之多少的点的图。其中的数据区间一般来说是对数据中的取值范围——也就是其中的最大的数减去最小的数——的等间隔划分。

因为这个离散化的概率分布来自于实际数据，所以这样得到的分布函数被称为经验分布。这里的经验 (Empirical) 是“来自于数据”的意思，更好的翻译是“实证”。

我们来做一个从数据绘制经验分布图的例子。下面是我们产生的一个取值在 $[0, 1]$ 之间的均匀分布随机数¹。我们来把这个区间划分成 10 个小格

¹你暂时不需要管这里的“均匀”是什么意思。未来会学习到。如果你特别想提前知道，大概就是一会儿得到的 P_n 是一个不依赖于 n 的常数，到处相等。所以叫“均匀”。

子，数一数每个小格子里面的数据点的数量（遇到刚好在区间边界上的数据点可以随机处理或者约定边界属于左侧的格子），计算百分比，然后画图。

$$U = \{0.631, 0.655, 0.114, 0.706, 0.361, 0.00437, 0.649, 0.470, 0.601, 0.996, \\ 0.903, 0.474, 0.0905, 0.848, 0.724, 0.0455, 0.384, 0.203, 0.657, 0.0153, \\ 0.509, 0.671, 0.278, 0.589, 0.792, 0.188, 0.893, 0.342, 0.407, 0.885, \\ 0.356, 0.802, 0.583, 0.964, 0.276, 0.122, 0.604, 0.667, 0.319, 0.0468, \\ 0.381, 0.577, 0.780, 0.532, 0.413, 0.909, 0.0459, 0.103, 0.505, 0.380, \\ 0.641, 0.841, 0.786, 0.811, 0.106, 0.0288, 0.634, 0.988, 0.0682, 0.171, \\ 0.884, 0.603, 0.566, 0.685, 0.109, 0.781, 0.478, 0.0453, 0.269, 0.213, \\ 0.950, 0.571, 0.780, 0.764, 0.817, 0.730, 0.756, 0.853, 0.138, 0.617, \\ 0.823, 0.721, 0.514, 0.370, 0.962, 0.958, 0.969, 0.511, 0.672, 0.206, \\ 0.617, 0.402, 0.402, 0.0400, 0.0981, 0.462, 0.0690, 0.521, 0.495, 0.610\}$$

(21.1)

对数据 U 做画格子和统计以后得到表21.1。图 21.1 把表21.1中的数据呈现为直方图——由一个个长方形构成的图。其中每一个长方形在 x 轴上的边界是所划分的格子的边界，长方形的高对应着出现次数或者百分比。这里我们呈现了出现的次数。

我们再来完成一个例子。

例 21.1 (正态分布的数据和经验分布). 我们生成了一个未来要学习的称为

表 21.1: **从数据到经验分布**。按照划分小格子，统计出现次数和计算出现比例的方式可以从一堆数据——这里是数据 U ——得到经验分布 P_n 。 x 对应着每个格子的中心。

n	0	1	2	3	4
x	0.1	0.3	0.5	0.7	0.9
$M(N = 100)$	18	19	22	18	23
P_n	0.18	0.19	0.22	0.18	0.23

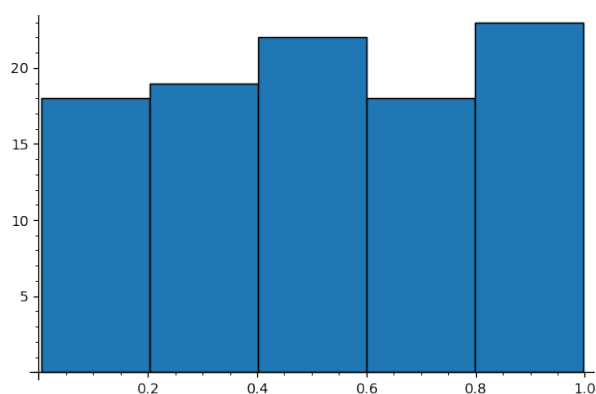


图 21.1: **数据 U 的经验分布图**。把表21.1中的次数或者百分比画成直方图。

“正态分布”的随机数的一堆数据 $\mathcal{N}$ 。做这个数据的经验分布图。

$$\begin{aligned} \mathcal{N} = \{ & -0.194, -0.449, -1.498, 0.15, 1.01, 0.834, -0.326, 0.703, -1.235, 0.298, \\ & 0.439, 0.084, -0.255, 0.701, 0.895, -0.945, 0.862, -0.721, 0.622, 0.983, \\ & 1.209, 0.952, 0.199, 0.397, -1.094, 1.233, 0.282, 1.279, 0.171, 0.835, \\ & 0.193, -0.57, -0.786, -1.791, -0.719, 1.428, 0.563, -0.249, -0.004, 0.29, \\ & -0.806, -0.448, -1.302, 1.432, -0.968, -0.025, -1.216, -0.628, -0.554, 0.468, \\ & -0.871, -0.501, 0.913, -1.79, -0.634, -1.061, 1.139, 0.021, 1.203, 0.148, \\ & 1.147, 1.76, -0.138, 0.049, -0.311, -0.659, 2.685, -0.471, -0.004, 1.55, \\ & 1.547, -0.619, -0.882, 0.276, 1.11, -0.937, 0.432, -0.383, -0.168, 1.062, \\ & -1.057, -0.419, -0.251, 0.509, -0.359, 0.333, 0.162, 1.098, 1.079, 0.794, \\ & 0.701, -0.231, 0.293, 0.23, -0.231, 0.268, 0.455, -0.765, 1.577, -0.203 \} \end{aligned}$$

(21.2)

对数据 $\mathcal{N}$ 做画格子 and 统计以后得到 21.2。图 21.3 把表 21.2 中的数据呈现为直方图——由一个个长方形构成的图。其中每一个长方形在 x 轴上的边界是所划分的格子的边界，长方形的高对应着出现次数或者百分比。这里我们呈现了出现的次数。

为了看看数据点更多格子更多以后的效果会怎样，我们从正态分布中生成 10000 个数据点，分成 30 个格子、数数、画直方图。我们发现图 21.3 和上一个图图 21.2 相比，离红色的正态分布函数的样子就接近了很多。

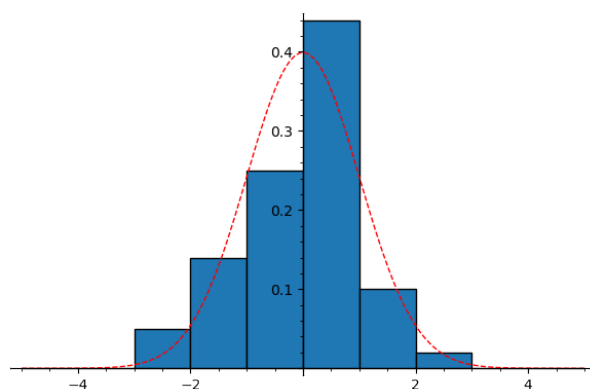


图 21.2: 数据 $\mathcal{N}$ 的经验分布图。把表21.2中的次数或者百分比画成直方图。

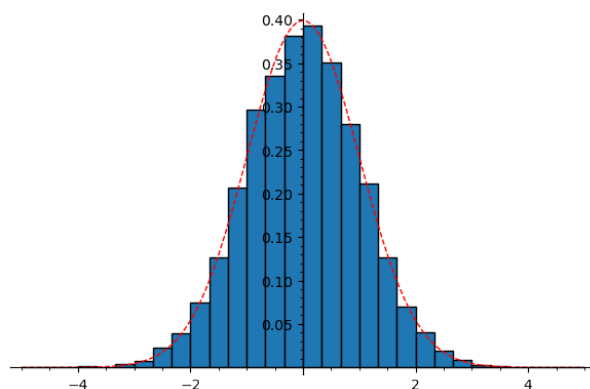


图 21.3: 一万个数据点的 $\mathcal{N}$ 的经验分布图。从正态分布中生成一万个数据点，分格子、数数、画直方图。

表 21.2: **从数据到经验分布**。按照划分小格子, 统计出现次数和计算出现比例的方式可以从一堆数据——这里是数据 U ——得到经验分布 P_n 。 x 对应着每个格子的中心。

n	0	1	2	3	4	5	6	7	8	9
x	-4.5	-3.5	-2.5	-1.5	-0.5	0.5	1.5	2.5	3.5	4.5
$M(N = 100)$	0	0	0	9	37	36	17	1	0	0
P_n	0	0	0	0.09	0.37	0.36	0.17	0.01	0	0

习题 21.1 (自己完成绘制经验分布图). 你自己来运用这一节的所有数据绘制相应的经验分布图。

实际上, 这个分格子、数落在格子里面的次数和计算百分比、画图的工作, 可以由计算机数学软件来完成。甚至, 有些计算机数学软件直接可以从数据得到经验分布图。请你自己学习这样的软件, 用计算机数学软件来绘制这一节的所有数据的经验分布图。

习题 21.2 (用计算机数学软件绘制经验分布图). 用计算机数学软件来绘制这一节的所有数据的经验分布图。用数学软件的时候, 你不一定要用书里面的数据。你可以自己来生成这些数据: 一个是 $[0, 1]$ 之间的均匀随机数, 一个是均值为 0 方差为 1 的正态分布随机数。

21.2 描述性统计

有了经验分布图, 我们就可以从经验分布图中直观地获取这个数据背后的随机事件的很多信息了。不过为了获得这些信息的数字特征, 以及判断这些的数字特征的可靠性, 我们还需要描述性统计和统计推断。以拓展古典概型的离散分布为例, 第一个要解决的问题是, 我们如何由一堆关于这个随机现象的数据来得到 P_n 的估计值, 以及所得到的估计值有多么可靠。

比如, 我们下面是一个硬币扔出来以后的实际数据, $c = 1$ 表示正面, $c = 0$ 表示反面。我们想看看 $P_1 = 0.5$ 是一个好的估计值吗, 也就是这个硬币和无偏硬币有很大的差别吗?

$$\begin{aligned}
C = \{ & 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, \\
& 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, \\
& 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, \\
& 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, \\
& 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0 \} \quad (21.3)
\end{aligned}$$

数一数以后我们发现总共有 $N = 100$ 个数，其中 $c = 1$ 出现了 $M = 50$ 次。我们直接算一个比例，得到在这个实际数据 D 中， $c = 1$ 出现的频率是 $P_1^* = \frac{M}{N} = 0.5$ 。这刚好等于无偏硬币的 P_1 。但是这是偶然的。

我们还尝试了扔 $N = 1000$ 次和 $N = 10000$ 次，从数据记录中分别得到 $P_1^* = 0.494$ 和 $P_1^* = 0.503$ 的结果。那么，这意味着什么呢？有的时候等于 0.5 有的时候不等于，但是整体看起来比较接近 0.5。我们还需要一个衡量接近程度的量。我们来试试计算这几组数据的均值和方差。

均值为

$$\mu_C = \langle c \rangle_C = \frac{1}{N} \sum_{n=1}^N c_n, \quad (21.4)$$

我们发现 $\mu_C = \frac{M}{N} = P_1^*$ 正好成立，而方差为

$$\sigma_C^2 = \langle (c - \langle c \rangle_C)^2 \rangle = \frac{1}{N} \left(c_n - \frac{1}{N} \sum_{m=1}^N c_m \right)^2. \quad (21.5)$$

由于均值非常接近 0.5，于是上式中的每一项都非常接近 0.5^2 ，平均以后仍然非常接近 0.5^2 。于是，我们得到近似值

$$\sigma_C^2 \approx 0.25. \quad (21.6)$$

或者说等价地记为

$$\sigma_C \approx 0.5. \quad (21.7)$$

稍后我们会看到，为什么对于均值我们区分 $N = 100, 1000, 10000$ 的结果，对于方差和标准差，我们不区分。

未来有一个叫做中心极限定理的数学定理，告诉我们，我们需要把标准差除以样本数量的平方根 $\frac{\sigma}{\sqrt{N}}$ 来看一组样本中的表现为某个平均值的统

计量的真实值到其平均值的偏离,

$$\delta = \frac{\sigma_C}{\sqrt{N}}, \quad (21.8)$$

对于上面三个不同的 N , 分别大约等于 0.05, 0.016, 0.005。那是不是说, 大概可以认为当 $N = 100$ 的时候, 大约 $P_1 \in [0.5 - 0.05 = 0.45, 0.5 + 0.05 = 0.55]$, 当 $N = 1000$ 的时候, 大约 $P_1 \in [0.494 - 0.016 = 0.478, 0.494 + 0.016 = 0.510]$, 当 $N = 10000$ 的时候, 大约 $P_1 \in [0.503 - 0.005 = 0.498, 0.503 + 0.005 = 0.508]$ 呢? 当 $N = 10000$ 的时候, 我们发现, $P_1 \in [0.503 - 0.005 = 0.498, 0.503 + 0.005 = 0.508]$ 已经和参考值非常接近了。一方面, 我们可以继续增加 N 。看看是不是会更接近。另一方面, 如果我们能够用某种概率大小, 或者其他的可靠性的度量, 来给出来, 真实 P_1 的取值落在相应的范围内的概率, 那我们就基本上实现了给一个 P_1 的估计值和参考值的差距的愿望了。

这其实是因为我们往往对一组数据就计算其均值和标准差: 均值往往对应着我们关注的某个量的估计值, 标准差给出了这个估计值和其真实值之间的差 (以某个概率或者可靠性的形式)。

将来的中心极限定理的数学定理大概说的是, 一个来自于均值存在方差有限的概率的 N 个样本 x_n , 其均值 $\mu_N = \frac{1}{N} \sum_{n=1}^N x_n$, 注意由于 x_n 是随机数 μ 也是随机数, 每次从原始概率中抽取 N 次随机数才能得到一个 μ 的值——符合一个叫做“正态分布”的特定函数, 其均值就是原始随机数 x 的均值, 其方差是原始随机数的方差的 $\frac{1}{N}$, 也就是

$$\langle \mu_N \rangle = \mu, \sigma_{\mu_N} = \frac{\sigma}{\sqrt{N}}. \quad (21.9)$$

至于中心极限定理的具体内容怎么表达, 其为什么成立, 这些问题我们留待以后再说。这是一个概率论的问题——从一个变量 x 的已知的概率开始, 要求出来关于 $\mu_N = \frac{1}{N} \sum_{n=1}^N x_n$ 的概率。这个问题只有等到大学阶段学习了微积分以后才能解决了。

不过, 尽管我们现在还不能证明, 我们可以暂时用一个数值结果来做中心极限定理的演示。我们取前面的 $N = 32, 64, 128$ 个取值为 0, 1 的离散分布随机数加在一起算平均, 看看算出来的随机数 μ_N 的经验分布。从图

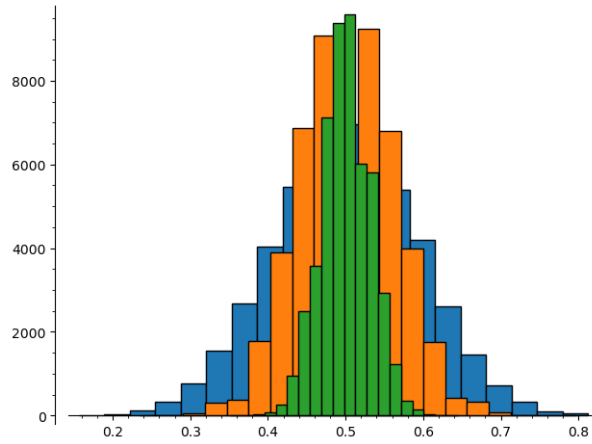


图 21.4: **多个随机数平均的经验分布**。取 $N = 32, 64, 128$ 个取值为 0, 1 的离散分布随机数加在一起算平均。随着 N 增加, 得到的均值越来越接近正态分布, 分布的中心不变, 分布的范围越来越窄。

21.4 中看出来, 随着 N 增加, 得到的均值越来越接近之前我们在图 21.3 中见过的正态分布的样子, 分布的中心不变, 分布的范围越来越窄。

不过, 其实, 我们存在着另一条道路来解决这个问题——直接从样本数据开始的叫做“非参数估计”的一条道路——Bootstrap 方法。其具体过程如下:

1. 从样本数据 C 中抽取 M 个样本点记作 $C^{(M)}$, 采用有放回的抽象, M 不一定等于 N ;
2. 计算均值 $\mu^{(M)} = \frac{1}{M} \sum_{m=1}^M c_m^{(M)}$;
3. 重复这样的过程 L 次, 得到每一次的 $\mu_l^{(M)}$;
4. 计算 $\{\mu_l^{(M)}, l = 1, 2, \dots, L\}$ 的均值和标准差。

这里的均值和标准差可以代替前面的均值的估计值和这个估计值的标准差。这个背后仍然可以看作是中心极限定理。但是, 这个过程完全可以得到和展示数值结果。图 21.5 就是这样得到的 $\mu^{(M)}$ 的经验分布。

我们先生成大小为 $N = 10000$ 的数据 C (没有展示在这里)。再从 C 中取 $M = 32, 64, 128$ 样本数据加在一起算平均得到 $\mu^{(M)}$, 然后绘制 $\mu^{(M)}$

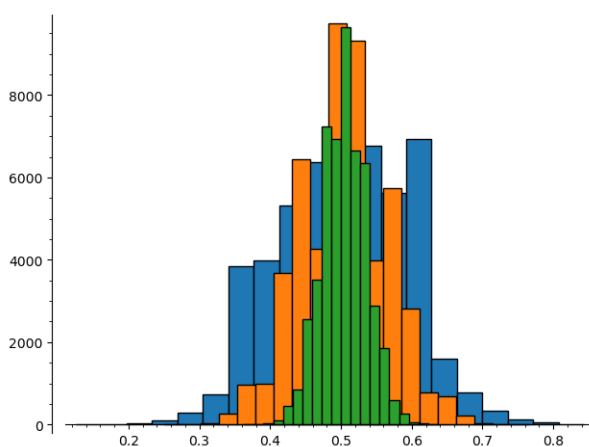


图 21.5: **多个随机数平均的经验分布**。从已经生成的取值为 0,1 的离散分布样本集合 C 中随机取 $M = 32, 64, 128$ 个加在一起算平均。随着 M 增加, 得到的均值越来越接近正态分布, 分布的中心不变, 分布的范围越来越窄。

的经验分布图。从图 21.5 中看出来, 随着 M 增加, 得到的均值越来越接近之前我们在图 21.3 中见过的正态分布的样子, 分布的中心不变, 分布的范围越来越窄。

如果你看仔细一点, 你还是可以发现图 21.4 和图 21.5 的不同的: 前者看起来更接近正态分布一些, 尤其是当样本数比较少的时候。不过这个差别不影响我们的结论: 无论是从分布函数直接抽样, 还是从给定的样本数据中 Bootstrap 抽样, 随着取出来的样本数量的增加, 我们得到的均值都会接近离一个共同中心越来越近²。

这就是我们为什么可以用样本均值 μ_C 来当作 P_n 的估计值以及用 $\frac{\sigma_C}{\sqrt{N}}$ 来当作这个估计值和真实值之间的偏离的度量的原因。

正是有了这个性质——有限大小的样本的均值可以用来估计分布函数的均值, 并且两者的误差随着样本大小而越来越小——才使得我们可以从有限大小的抽样中来获取一个现象的整体性质。例如, 假设我们想知道中国人的身高的分布, 尤其是平均身高和标准差, 我们通常通过选择一定数量的人——称作抽样——来测量其身高, 获得一个有限大小的身高数据。然后, 我们希望通过这个数据——比如说计算出来这个抽样样本的平均值和

²前提是原始的分布函数的均值和方差有限。未来在大学阶段真的来证明这个结果的时候才能明白为什么有这个要求。

标准差——来一定程度上反应所有中国人的身高——尤其是总样本的平均值和标准差。

除了平均值，我们有的时候还会从数据中找出来最大值、最小值、众数、中位数。前两者简单。众数是数据中出现次数最多的数，体现的也是一种典型性——下一个数据以最大的相对可能性表现为众数。中位数是居中的那个数，也就是比它大的数的数量和比它小的数的数量一样多。如果总数是奇数，这样的数只有一个。如果总数是偶数，则把中间的两个都取出来，做一个平均。如果整体数据中有一些极端值，平均值会很大程度上受这些极端值影响，但是中位数不太受它们的影响，只要数的相对位置不变，中位数就不变。也就是说，本质上，提出和使用中位数仍然是在考虑对整个分布的某种代表性典型性，是为了帮助做“统计推断”——知道中位数这个值以后可以知道关于这个数据的一些什么结论。

习题 21.3 (独立画出来这一节的图). 用计算机数学软件重复这一节的两张图的生成过程，画出来图，做一个比较。

习题 21.4 (降低极端值的影响增加典型性). 在体育运动的评分中，由另一种降低极端值的影响增加典型性的方式：去掉最高分和最低分，再来算平均。你思考一下，为什么需要和可以这么做呢？你能在设计出来一种方法吗？如果能，可以检验其效果吗？怎么检验？

21.3 统计推断和人类知识

在上一节中，在我们对描述性统计的目的的介绍中，我们已经采用了统计推断的角度——我们想从数据中获得分布函数的信息，无论是从经验分布图来看真实的分布函数的图，还是从数据中通过计算均值估算出来 P_n 然后用 $\frac{\sigma_n}{\sqrt{N}}$ 来当作这个估计值和真实值之间的偏离的度量。

实际上，一旦我们有了经验分布或者有了所有的 P_n ，我们就可以来判断对于这个随机数某些结论对不对。例如，在 $0,1$ 分布中我们可以判断 $P_0 > 0.5$ 对不对。例如，拿到一个数据集，我们想判断一下这个数据是否来自于某个已知的分布函数。例如，如果有两个分布函数，我们可以取比较这两个分布函数的差异。比如说，药物效果问题就会用到这样得比较。我们把所有的参与实验的被试分成使用这个药物的实验组和不使用这个药物

(例如使用安慰剂, 不能完全不吃“药”, 因为如果一组吃一组不吃不吃的人自然就会知道没有得到治疗, 而这个主观认知可能影响实验效果), 去测量这两组被试在一段时间之内和之后的各种指标。取其中的一个指标, 画出来这个指标的经验分布, 我们就有可能可以看出来其差异的大小。如果进一步还有一个量能够用来度量这两个经验分布之间的差异, 那就更好了。那统计学就是为了提出和解决这些问题而发展起来的。

实际上, 人们确实找到了这样的量——成为 Kolmogorov-Smirnov 统计量——来度量两个经验分布之间的区别。这个检验被称为 Kolmogorov-Smirnov 检验。详细讲述这个检验的原理和使用就不是本书的目的了。

实际上, 我们可以认为绝大部分的人类知识——尤其是自然科学和社会科学的知识, 数学知识也很大程度上是如此——都是通过这个方式获得的: 先有数据, 或者观察 (可以看作非常粗糙的数据), 然后对这个数据做分析以后提炼出来一些概念从而形成一些猜想, 最后通过统计推断来看这些命题是否正确。

有了概念和命题之后的部分, 我们做得是统计推断, 或者说带着概率的逻辑推理, 或者对于完全确定的问题做得是不带概率的逻辑推理。有概念和命题之前, 我们做的是数据分析、经验分析、总结归纳、猜测和抽象提炼。你看无论前后, 统计学都是非常重要的。甚至, 在我们能够充分肯定这个数据背后的现象不牵涉到随机性之前, 我们也应该那这个现象先靠做是随机现象做分析。因此, 待概率的逻辑推理, 也就是统计学, 应该被放到完全确定的逻辑推理之前。

21.4 条件概率用于统计推断

我们来用条件概率尝试解决一些稍微复杂一点统计推断的问题。例如, 我们希望从学习已经掌握某些知识来得到其掌握另一个知识的概率, 以及反过来从没有掌握某些知识来得到没有掌握另一个知识的概率。这些信息有助于设计考试题, 也有助于帮助学生的学习, 帮助到老师的教。

为了简化这些问题, 我们仅仅考虑图 21.6 所示的几个知识点以及它们之间的关系: 设未知数、等式的性质、变量替换、一元一次方程的求解、等式计算、二元一次方程的计算。当然, 这只是展示一下条件概率可以如何用

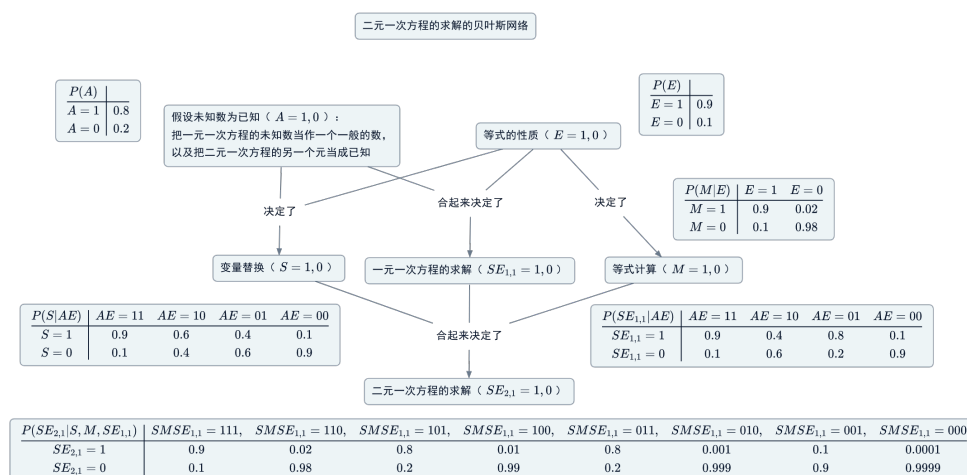


图 21.6: 贝叶斯网络用于检测学习效果。贝叶斯网络的顶点在这里是知识点的掌握与否的状态，连边是条件概率上的依赖关系——在这里是要求更高的因果关系。每一个连边上放了条件概率。

于统计推断，而不是真正的知识结构和诊断模型。把条件概率用网络图的方式展示出来被称为贝叶斯网络 [10]。

例 21.2 (贝叶斯网络计算任意顶点组合的任意状态的概率). 根据图 21.6 中给出的贝叶斯网络, 随便选一个被试计算其掌握一元一次方程的求解的概率 ($P(SE_{1,1} = 1)$), 以及已知一个被试其掌握了一元一次方程的求解 ($SE_{1,1} = 1$), 推断其掌握二元一次方程的求解的概率 ($P(SE_{2,1} = 1|SE_{1,1} = 1)$), 还有已知一个被试没有掌握一元一次方程的求解 ($SE_{1,1} = 0$), 推断其没有掌握等式的性质的概率 ($P(E = 0|SE_{1,1} = 0)$)。

从贝叶斯网络图 21.6, 我们直接读出来, $P(E = 1) = 0.9$, 也就是如果我们在这些数据获取的样本所代表的世界中随便选取一个人, 其掌握等式的性质的概率是 0.9。我们再来计算

$$P(SE_{2,1}|SE_{1,1} = 1) = \sum_{S, M} P(SE_{2,1}|S, M, SE_{1,1} = 1)P(S)P(M), \quad (21.10)$$

其中条件概率表 $P(SE_{2,1} = 1|S, SE_{1,1}, M)$ 已知, 而

$$P(S) = \sum_{A,E} P(S|A, E)P(A)P(E) = \begin{cases} 0.77 & S = 1 \\ 0.23 & S = 0 \end{cases}, \quad (21.11)$$

$$P(M) = \sum_E P(M|E)P(E) = \begin{cases} 0.826 & SE_{1,1} = 1 \\ 0.174 & SE_{1,1} = 0 \end{cases}. \quad (21.12)$$

因此, 我们得到

$$P(SE_{2,1}|SE_{1,1} = 1) = \begin{cases} 0.836 & SE_{2,1} = 1 \\ 0.164 & SE_{2,1} = 0 \end{cases}. \quad (21.13)$$

$$P(E = 0|SE_{1,1} = 0) = \frac{P(SE_{1,1} = 1|E = 0)P(E = 0)}{P(SE_{1,1} = 0|E = 0)P(E = 0) + P(SE_{1,1} = 1|E = 1)P(E = 1)} \quad (21.14)$$

而

$$P(SE_{1,1} = 1|E)P(E) = \sum_A P(SE_{1,1} = 1|AE)P(E)P(A). \quad (21.15)$$

合起来得到

$$P(E = 0|SE_{1,1} = 0) = 0.379. \quad (21.16)$$

实际上所有的这些计算在已知贝叶斯网络图 21.6 之后都可以通过软件直接计算出来。这称为贝叶斯网络上的概率推断问题。那贝叶斯网络和每一条边上的条件概率又是怎么来的呢? 这被称为贝叶斯网络的构建的问题: 从数据获得网络里面有哪些连边以及每条连边对应的条件概率。这相当于已知每个变量的分布、各个变量合起来的分布, 需要凑出来合适的条件概率, 使得每个变量的分布以及各个变量合起来的分布刚好和从这些条件概率计算出来的相同。当然, 实际上不可能这样来求所有的条件概率——需要尝试的可能性太多了。将来会有更高效的方法。将来在大学或者研究生阶段, 你会学习到贝叶斯网络的构建以及贝叶斯网络上的概率推断问题都怎么做。

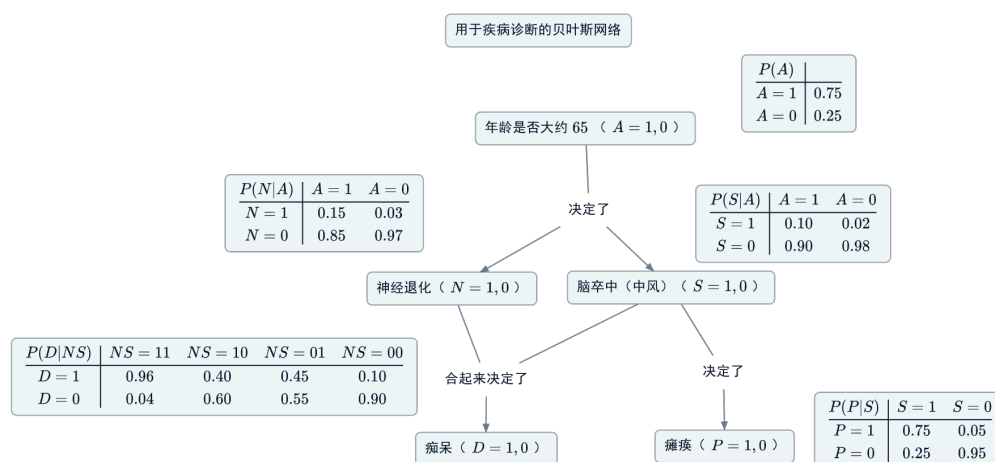


图 21.7: 贝叶斯网络用于疾病诊断。贝叶斯网络的顶点在这里是与否得了某种疾病或者年龄是否大于六十五的状态，连边是条件概率上的依赖关系每一个连边上放了条件概率。这个案例来自于 [11]。

习题 21.5 (贝叶斯公式用于疾病诊断). 假设“年龄、神经退化疾病、脑卒中、痴呆、瘫痪”之间的关系如图 21.7 所示——注意是对现实的一个极大的简化现实远远更为复杂，如果一位医生遇到了一位病人，其年龄大于六十五岁，已经诊断得了神经退化，其发展为痴呆的可能性式多大？反过来，如果一个医生遇到一个痴呆的患者，其大于六十五岁的概率是多大？

21.5 统计学的应用

21.6 推荐阅读材料

21.7 作业

习题 21.6 (男性女性寿命问题). 据说平均来说女性比男性长寿。你看看能不能找到一组数据来检验一下这个结论。

习题 21.7 (抽烟和健康问题). 据说抽烟的人比不抽烟的人更健康，尤其在肺部疾病上。你看看能不能找到一组数据来检验一下这个结论。

习题 21.8 (抽烟和健康问题). 据说有运动习惯的人比没有运动习惯的人更健康。你看看能不能找到一组数据来检验一下这个结论。

21.8 本章小结

在本章中，我们特别关注统计学的学科责任——这个数学的分支学科（未来会发现其基本上成了一个独立学科）是干什么的：是从数据来得到数据背后的现象的性质的，例如其分布函数、分布函数的参数、某些基于这些分布函数的结论。当然，其在数学中的这部分是给这样的学科责任提供分析方法的。其在数学之外的部分，则是通过面对现实世界的问题，运用已有的分析方法或者反过来推动其在数学之内的部分来发展相应的分析方法，从而解决问题的。

当然，其数学之内的部分的知识基础是概率论。例如，我们发现对分布函数参数 P_n 的估计的背后是中心极限定理，而中心极限定理是概率论的内容。这就是统计学的理论基础、核心内容和应用之间的关系。我们还特别强调了统计学做为认识世界的带概率的逻辑推理在获取人类知识上的作用。这也是数学是描述世界的语言，是思维的语言，以及双向数学建模的体现。

在这一章，我们留下了好多尾巴：正态分布是什么，为什么多个随机变量加起来做平均以后得到的随机变量会符合正态分布，比较两个经验分布的具体方法都有什么怎么做，非参数估计还有哪些方法，参数估计除了均值如果我们想估算方差以及比较估算出来的方差和真实方差之间的关系又怎么做。由于这些知识都依赖于一门叫做微积分的数学知识，我们都只能留到以后来学习。

通过能够搞懂的例子，来搞明白一个学科的典型责任、核心思维方式、核心分析方法是最重要的。当然，能大致搞明白为什么需要和可以这样分析这样思考就更好了。具体的更多的分析方法，以及更透彻地搞明白为什么可以这样做，可以随着进一步学习和解决问题的需要来补充。

我们把这部分关于概率论和统计学的知识也做了梳理，见图 21.8。概率论所解决的是为随机事件的描述提供数学结构——概率，顺便也解决从随机变量的概率分布函数得到随机变量的函数的概率分布函数的问题。统计学所解决的是从数据得到相应的分布函数或者判断某个命题是否成立（成

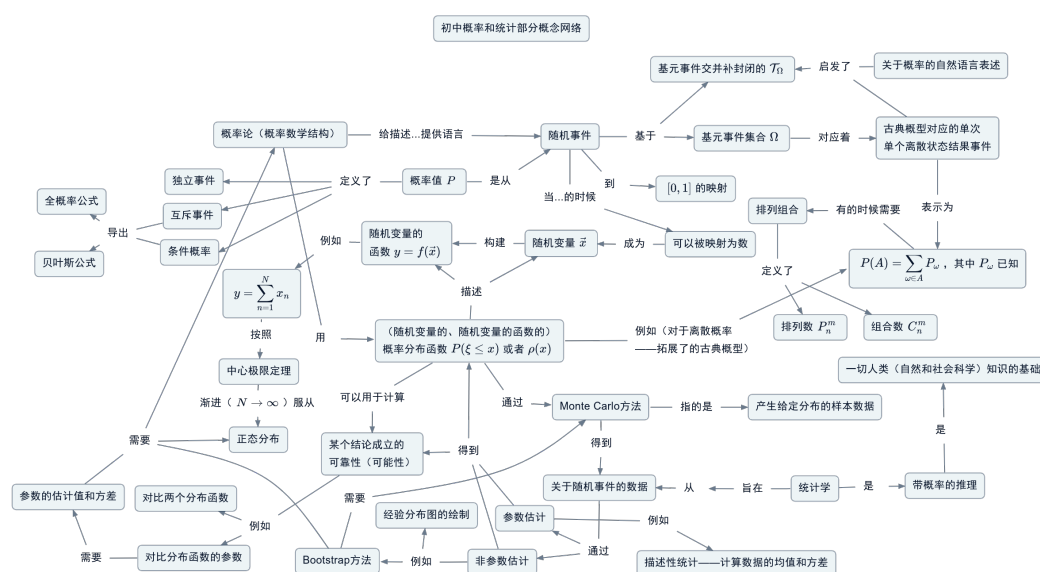


图 21.8: **本书初中概率论和统计学知识的梳理**。概率论为统计学提供概念和定理的基础，统计学解决从实际数据中获得结论，以及如何来获得结论的问题，顺便推动概率论的发展。[这张图折页印刷，一定要够大。](#)

立的可能性可靠性)的问题,以及提供接着这些问题的方法。这些方法的理论基础是概率论。因此概率论为统计学提供概念和定理的基础,统计学当需要发展方法的时候也推动统计学的理论部分或者概率论往前发展。

此外，Monte Carlo 方法也使得概率论和统计学两边沟通起来——从给定分布函数得到对应这个分布的数据，从实际样本数据得到分布函数的参数或者所关注的量的分布。

第二十二章 排列组合

Mathematics is a language.

– Josiah Willard Gibbs

排列组合未来在群论的学习中具有非常重要的地位：置换群可以用来表示任何一个有限群。我们之前提到过的人类知识的非凡创造之一——关于高阶方程求根的伽罗瓦理论——也是基于置换群来研究的。也有一些实际问题，尤其是牵涉到概率的问题，需要用排列组合来思考和计算。此外，未来组合学也是专门的数学分支学科之一。因此，这一章我们来学习排列组合。

22.1 组合和排列

对于给定的元素个数为 n 的集合 Ω ，如果我们从中取出来 m 个不同的元素构成新的集合 $\{\omega_1, \omega_2, \dots, \omega_m\}$ ，则这些新集合的数量就称为组合数 C_n^m ，这个相应的过程称为组合。如果我们关注这些取出来的元素的顺序，也就是构成的是序列，记为 $(\omega_1, \omega_2, \dots, \omega_m)$ ，而不是集合（因为集合不在乎其内部元素的顺序），则这些新集合的数量就称为排列数 P_n^m ，这个相应的过程称为排列。

例如，我们从单元素的集合中只能取出来一个元素 $\Omega = \{\omega\}$ ，得到一个集合 Ω ，得到一个排列 (ω) ，因此 $C_1^1 = 1$ ， $P_1^1 = 1$ 。我们从包含两个元素的集合中能取出来一个元素或者两个元素。它们分别是集合： $\{\omega_1\}$ ， $\{\omega_2\}$ ， $\{\omega_1, \omega_2\}$ ，序列 (ω_1) ， (ω_2) ， (ω_1, ω_2) ， (ω_2, ω_1) 。因此 $C_2^1 = 2$ ， $P_2^1 = 2$ ， $C_2^2 = 1$ ， $P_2^2 = 2$ 。

更一般地，我们说对于排列，第一个元素有 n 个选择，第二个元素有 $(n-1)$ 个选择，直到第 m 个元素有 $(n-m+1)$ 个选择，因此总共的排列

数为 $P_n^m = n(n-1)\cdots(n-m+1)$ 。我们引入一个记号 $n! = n(n-1)\cdots 1$, 称为 n 的阶乘, 也就是把 1 到 n 都连乘起来, 则我们可以把排列数记作

$$P_n^m = n(n-1)\cdots(n-m+1) = \frac{n!}{(n-m)!}. \quad (22.1)$$

现在让我们建立排列数和组合数之间的联系。如果我们已经得到一个 m 个元素的组合, 我们说这些元素总共可以有 $m!$ 个顺序。理由仍然是: 第一个元素有 m 个选择, 第二个元素有 $(m-1)$ 个选择, 直到第 1 个元素有 1 个选择。因此,

$$P_n^m = C_n^m m!, \quad (22.2)$$

也就是

$$C_n^m = \frac{P_n^m}{m!} = \frac{n!}{(n-m)!m!}. \quad (22.3)$$

考虑到记号的通用性, 我们约定 $0! = 1$, 这样就保持了下式正确,

$$1 = C_n^n = \frac{n!}{0!n!}. \quad (22.4)$$

22.2 排列组合的应用

我们来用排列数公式证明二项式定理, 对于非负整数 n :

$$(a+b)^n = \sum_{m=0}^n C_n^m a^m b^{n-m}. \quad (22.5)$$

在这里我们需要把乘法算式转化为加法算式。其中的每一项都是 a 和 b 的整数次幂, 而且其幂次相加必然等于 n , 因为乘法计算的每一项相当于从每一项因式 $(a+b)$ 中必须选出来一个 a 或者一个 b 跟来自于其他项的 a 或者 b 乘在一起。也就是说, 我们只需要确定从 n 个 $(a+b)$ 中选择 m 个 a , 无论用什么样的顺序选择出来的, 最后得到的乘法之后的表达式都是 $a^m b^{n-m}$ 。那么, 这样的项有多少呢? 每一个组合都是同一项, 因此组合数 C_n^m 就是项的数目, 也就是 $a^m b^{n-m}$ 前面的系数。例如, 第一项相当于一直不选 a 都选 b , 有 $C_n^0 = 1$ 项。类似地, 最后一项相当于都选 a 一直不选 b , 有 $C_n^n = 1$ 项。你算几个具体的数的例子就会发现 m 越接近 n 的中间 C_n^m 越大。

结合等概率假设, 排列组合还经常用来计算概率。例如,

例 22.1 (随机选袜子凑一对). 我有两种袜子, 每种四双。我随便从这些袜子中取出来两只刚好构成同种的一对的可能性是多少。这里我们假设袜子部分左右脚。如果分那就要更复杂了。

首先, 总共 $2 \times 4 \times 2 = 16$ 只袜子, 因此随机从中选两只的方式有 C_{16}^2 。能够选出来凑成一对, 表示要么从第一种袜子里面选出了两只, 要么从第二种袜子里面选出了两只, 各自有 C_8^2 种方式。因此, 能凑成一对的概率为 $P = \frac{C_8^2 + C_8^2}{C_{16}^2} = \frac{7}{15}$ 。

习题 22.1 (随机选袜子凑一对). 我有两种袜子, 每种四双。我随便从这些袜子中取出来两只刚好构成同种的一对的可能性是多少。这里我们假设袜子部分左右脚。

我们来把排列组合和置换群联系起来。首先, 我们来看把一串字符 ABC 任意交换顺序可以得到哪些字符串。它们分别是: $ABC, BCA, CAB, ACB, CBA, BAC$ 。合起来它们是 A, B, C 这三个字母的全排列, 因此有 $3! = 6$ 个。没错。

这样的改变字符串中的字符的顺序但是维持所有字符都出现的变换被称作是置换, 也就是 ABC 的意思是把 ABC 变成 ABC ——不变; BCA 的意思是把 ABC 变成 BCA ——也就是 $A \rightarrow B, B \rightarrow C, C \rightarrow A$ 。在这个理解下, 我们来把它们分成两组: ABC, BCA, CAB 和 ACB, CBA, BAC 。其中前者相当于保持 ABC 的相对顺序不变, 但是把其中的第二个元素变成第一个、第三个元素变成第二个, 第一个元素变成第三个——可以看作是第 0 个同时当第零个和第三个等同起来。也就是说, 如果我们将 ABC 看作首尾相连的——也就是三角形的三个角的记号, 则这几个排列之间的关系就像转动。其中后者是对自己做两次之后会回到 ABC 的变换: 例如, $ABC \rightarrow ACB \rightarrow ABC$, 其中的箭头的含义是“交换第二第三个字符”。回到三角形的粒子, 这正好就代表了三角形沿着中线翻转。

因此, 我们看到, $ABC, BCA, CAB, ACB, CBA, BAC$ 相当于是对正三角形 ABC 假设逆时针编号的保持正三角形形状完全相同的六个操作: 不动, 顺时针转动 60° , 顺时针转动 120° , 沿着 BC 边上的中线 (或者说高、角平分线¹) 翻转, 沿着 AC 边上的高翻转, 沿着 AB 边上的高翻转。

¹未来我们会证明对于等腰三角形, 底边上的中线、底边上的高、顶角的角平分线, 三线合一。

在这个例子中，我们看到了置换和对三角形的操作之间的联系。顺便，我们也再一次看到了数学是研究对象、对对象的操作、对操作的操作的，等等等等。

物理学中也经常使用排列组合。我们再来看几个简单的例子，然后来尝试不那么简单的例子。

例 22.2 (经典粒子放置问题——可区分). 我们将 n 个可区分的编号为 $1, 2, \dots, n$ 的粒子放到 g 个可区分的编号为 $1, 2, \dots, g$ 的盒子里面。如果每个盒子里面放的粒子没有限制，有多少种方式。

第一个粒子有 g 个选择，第二个粒子仍然有 g 个选择，因此放置方式的总数为 g^n 。

例 22.3 (经典粒子放置问题——最多放一个). 我们将 n 个可区分的编号为 $1, 2, \dots, n$ 的粒子放到 g 个可区分的编号为 $1, 2, \dots, g$ 的盒子里面。如果每个盒子里面最多放一个粒子，有多少种方式？

由问题所知， $n \leq g$ ，则否必然有至少一个盒子里面会出现多于一个粒子。这问题正好就是排列数问题：从 g 个盒子里面选择 n 个，并且交换顺序以后是另一种放置的方式。因此，有 P_g^n 种方式。

例 22.4 (费米子放置问题). 我们将 n 个不可区分的粒子放到 g 个可区分的编号为 $1, 2, \dots, g$ 的盒子里面。如果每个盒子里面最多放一个粒子，有多少种方式？

由问题所知， $n \leq g$ ，则否必然有至少一个盒子里面会出现多于一个粒子。这问题正好就是组合数问题：从 g 个盒子里面选择 n 个，但是交换顺序以后是同一种放置的方式。因此，有 C_g^n 种方式。

有的时候，你可能还需要转化一下排列组合问题。例如，

例 22.5 (玻色子放置问题). 我们将 n 个不可区分的粒子放到 g 个可区分的编号为 $1, 2, \dots, g$ 的盒子里面。如果每个盒子里面放的粒子没有限制，有多少种方式？

这相当于看下面的算式 $\sum_{l=1}^g n_l = n$ 的非负整数解的个数。但是，这个问题仍然比较难以直接运用现有对象的排列组合来解决。历史上的科学家发明了一个非常巧妙的求解方法。 n 个不可区分粒子可以这样看：把所有粒子排成一根线，给所有粒子的首尾先各自放一块隔板，然后每一个粒子

之间都可以插入或者不插入一块隔板，任意两块隔板之间的粒子表示它们在同一个盒子里面。这样，我们只需要决定在所有的可能的隔板和粒子的位置—— $n + g - 1$ 个位置——中，选择 $n - 1$ 块隔板的方式有多少种就行，于是得到有 C_{n+g-1}^{g-1} 种方式。

22.3 推荐阅读材料

22.4 作业

习题 22.2 (汉诺塔问题). 汉诺塔 (*Tower of Hanoi*) 是一种玩具。你去查查那是什么，自己动手做一套或者画一套。看看要完成汉诺塔的任务——把这个塔的所有层都从第一个柱子移动到最后的第三个柱子上需要多少步，在遵循每一个柱子上的层只能下面大上面小的顺序的前提下。

习题 22.3 (玻色子放置问题——盒子不可区分). 我们将 n 个不可区分的粒子放到 g 个不可区分的盒子里面。如果每个盒子里面放的粒子没有限制，有多少种方式？

例 22.6 (费米子放置问题——都不可区分). 我们将 n 个不可区分的粒子放到 g 个不可区分的盒子里面。如果每个盒子里面最多放一个粒子，有多少种方式？

22.5 本章小结

在本章中，我们学习了排列和组合的概念，排列数和组合数的概念。我们还把它们联系到了古典概型中概率的计算，还有二项式定理、对对象的操作，以及未来会学习到的统计物理学中的费米子和玻色子的问题。关于排列组合我们要学习的暂时就这么多。等到未来学习置换群的时候我们还会再次回到排列组合这个主题。

顺便，我们已经在本书中多个场合呼唤群论和微积分的出现了。我们还是得等等。等你们更加具备学习它们的基础了，再来学习。

第二十三章 逻辑学和数学论证

Mathematics is the most beautiful and most powerful creation of the human spirit.

– Stefan Banach

In formal logic, a contradiction is the signal of defeat, but in the evolution of real knowledge it marks the first step in progress toward a victory.

在形式逻辑中，一个悖论表示哪里出问题了失败了，但是在真正的知识的演化发展中，悖论往往是通向突破和胜利的第一步。

– Alfred North Whitehead

23.1 命题

在经过几何部分的学习之后，我们终于可以更深入地来认识逻辑，以及逻辑和数学的关系了。在几何的学习中，我们遇到了“三角形的内角和是 180° ”、“过直线外一点有且只有一条直线和这条直线平行（翻译成：过直线外一点和这条直线平行的直线的数量等于 1）”、“三角形内平行线切割线段成比例”、“三角形内切割线段成比例的两条线是平行线”这样的命题，我们还遇到了从“三角形内平行线切割线段成比例”可以通过“反证法”来证明“三角形内切割线段成比例的两条线是平行线”的证明过程。

我们把一个可以判断是否成立的陈述句叫做命题。但是这是一个依赖于自然语言的不可操作的定义，尽管实际上我们很多时候也能用这个定义来判断几个陈述句是不是命题。我们想把这个定义变得可操作，或者说可

以基于已有的数学概念来定义。那我们学习过的最基本的数学概念是什么？是集合和映射，尽管其实集合我们也没有可操作的仅仅有基于自然语言的描述性定义¹。那为什么不允许命题的定义是另一个不可操作的自然语言的定义呢？我们还是希望这样的定义越少越好。因此，如果用其中一个定义了另一个，至少，将来解决了那个更基本的也就解决了那个用更基本的来定义的。

那我们试试用集合和映射来定义命题。其实，在第八章中我们已经把命题表示为集合之间的包含关系。这次，我们更加严格地来建立起来这个关系。

首先，我们来引入一个集合上的示性函数 $I_A(\omega)$ 。假设我们关心的“世界”已经被一个集合 $\mathcal{W}$ 描述。例如，一个班级里面的所有人，一棵树上的所有苹果，所有的自然数，所有的实数，等等等等。然后，在这个世界全集里面有一个集合 A ，则集合 A 上的示性函数被定义为一个从 $\mathcal{W}$ 到 $\{0, 1\}$ 的映射，

$$I_A(\mu) = \begin{cases} 1 & \mu \in A, \\ 0 & \mu \notin A. \end{cases} \quad (23.1)$$

可以见其含义就是如果一个元素属于 A 则得到 1，否则得到 0。

反过来，如果有一个命题 T ，例如“三角形的内角和是 180° ”，就相当于其是一个示性函数——对于那些确实满足“内角和是 180° 的三角形”给出答案 1——表示成立，对于那些不满足“内角和是 180° 的三角形”给出答案 0——表示不成立。也就是说，这个命题 T 看作一个示性函数的话，其对应的集合是 A_T ，定义为“所有内角和是 180° 的三角形”，用集合表示为

$$A_T = \left\{ \triangle \left| \sum_{\alpha \in \triangle} \alpha = 180^\circ \right. \right\}. \quad (23.2)$$

其中 $\sum_{\alpha \in \triangle} \alpha$ 表示三角形的内角和。我们可以拓展这个记号和这个集合，例如把任意的满足内角和为 180° 多边形都当作一个集合，记作

$$P_{180^\circ} = \left\{ P \left| \sum_{\alpha \in P} \alpha = 180^\circ \right. \right\}. \quad (23.3)$$

¹就像我们之前提到的，现代数学通过范畴论、集合的公理化定义等是解决了集合的定义的问题的。

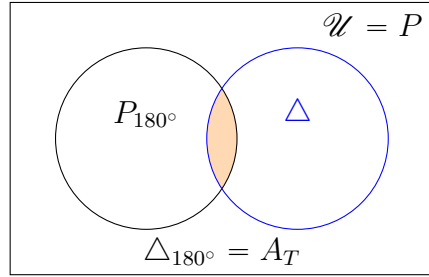


图 23.1: P, P_{180°, Δ 和 Δ_{180° 之间定义上的关系。用 Venn 图展示了 P, P_{180°, Δ 和 Δ_{180° 之间定义上的关系。

这时候全集，也就是世界，就是所有多边形，其中的一个元素用 P 来表示。 P_{180° 是这个多边形全集的满足内角和为 180° 的一个子集。我们把所有的三角形的集合记为 Δ 。顺便，在这个记号下

$$A_T = \Delta_{180^\circ}. \quad (23.4)$$

我们先来看看这几个集合从定义来得到的相互关系，再来看实际上，按照我们已经学到平面几何知识，这些集合之间是什么关系，从而明白命题和集合之间的关系。如图 23.1 所示， $P_{180^\circ} \subset P$ ， $\Delta \subset P$ ， $\Delta_{180^\circ} = \Delta \cap P_{180^\circ}$ 。

这个时候， T 这个映射的作用就很明确了，除了中间的交集部分 $T(\mu) = 1, \forall \mu \in \Delta_{180^\circ}$ 因为其既满足内角和是 180° 又满足是三角形，其他的任何元素 $T(\mu) = 0, \forall \mu \notin \Delta_{180^\circ}$ 。因此，这个映射的原像 $A_T = T^{-1}(\{1\}) = \Delta_{180^\circ}$ 。

于是，我们这样来定义一个命题：

定义 23.1 (命题). 对于给定的世界——也就是全集，如果存在这一个从这个全集到从 $\mathcal{U}$ 到 $\{0, 1\}$ 的映射 T ，对任何一个 $\mathcal{U}$ 中的元素 μ 都有 $T(\mu) = 1$ 或者 $T(\mu) = 0$ ，则我们称这个 T 为一个命题。

相应地，把原像集合 $A_T = T^{-1}(\{1\})$ 称为命题 T 对应的集合。

由于我们总是可以建立起来 T 和 $T^{-1}(\{1\})$ 的一一对应关系，以后我们不再区分命题和命题对应的集合。对于给定命题 T ，当我们说“集合 T ”的时候我们指的就是 $T^{-1}(\{1\})$ ，对于给定集合 A 当我们说命题 A 的时候，我们指的就是把其示性函数 I_A 当作命题。

我们再来看实际上在平面几何中这几个集合之间的关系。我们事实上知道在平面几何的范围内，所有三角形的内角和都是 180° ，所有内角和为

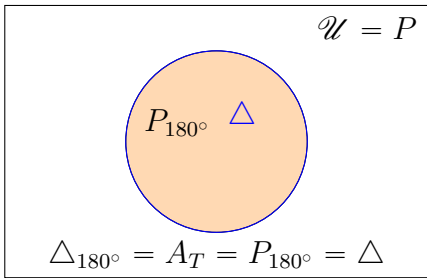


图 23.2: P, P_{180°, Δ 和 Δ_{180° 之间实际上的关系。用 Venn 图展示了 P, P_{180°, Δ 和 Δ_{180° 之间实际上的关系: P_{180° 和 Δ 这两个集合完全重合。

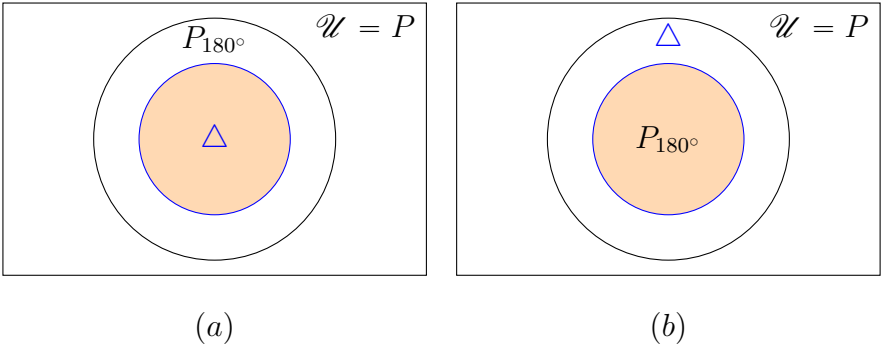


图 23.3: $\Delta \rightarrow P_{180^\circ}$ 和 $P_{180^\circ} \rightarrow \Delta$ 的集合关系。用 Venn 图展示了命题 $\Delta \rightarrow P_{180^\circ}$ 和 $P_{180^\circ} \rightarrow \Delta$ 之间的关系。

180° 的多边形也都是三角形,也就是如图 23.2 所示的关系:由于 $P_{180^\circ} = \Delta$, 两者的交集就是这个集合本身,也就是 $\Delta_{180^\circ} = A_T = P_{180^\circ} = \Delta$ 。这正好就是命题“三角形内角和是 180° , 记作 $\Delta \rightarrow P_{180^\circ}$ ”和“内角和为 180° 的多边形是三角形, 记作 $P_{180^\circ} \rightarrow \Delta$ ”合起来的意思。

从这里我们看到,一个定理如果成立,则相当于从图 23.1 中的集合之间关系的一般情况到了图 23.2 中的关注这些集合之间的关系的真实情况——一般情况的一种具体实现。这其实也正好就是定理的作用,通过集合的方式给出来背后的真实世界中所包含的元素之间的关系。

如果我们把上面那两个命题分开,各自用集合来表示,则其关系可以分别表示为图 23.3 中的 (a) 和 (b)。 $\Delta \rightarrow P_{180^\circ}$ 表示 $\Delta \subset P_{180^\circ}$, 而 $P_{180^\circ} \rightarrow \Delta$ 表示 $P_{180^\circ} \subset \Delta$ 。命题的语言和集合的语言完全等价。

这样我们得到了命题的另一个定义,

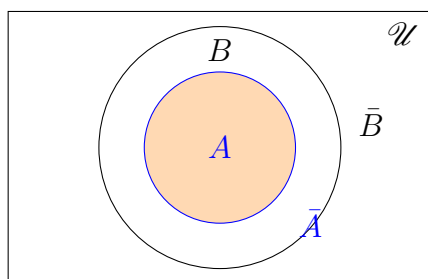


图 23.4: 原命题和逆否命题等价。用 Venn 图展示了原命题和逆否命题等价。

定义 23.2 (命题). 对于给定的世界——也就是全集，一个命题往往可以表示为两个集合之间的包含关系 $A \subset B$ ，或者等价地 $I_A \rightarrow I_B$ 。

其中 I_A, I_B 表示集合 A, B 上的示性函数。 $A \subset B$ 的含义是，对于世界中任何一个属于集合 A 的元素， $\mu \in A$ 则有 $\mu \in B$ ，也就是必然属于集合 B 。因此满足条件 I_A 的就会有结论 I_B 。按照我们之前的不再区分集合和相应的命题的符号的约定，我们把 $I_A \rightarrow I_B$ 也记作 $A \rightarrow B$ 。

有了命题的集合包含关系表示，则 $B \rightarrow A$ 就是 $B \subset A$ ，就是把两个集合之间的包含关系反过来逆过来，因此 $B \rightarrow A$ 称作是 $A \rightarrow B$ 的逆命题。如果原命题和逆命题同时成立，则意味着集合 $A = B$ ，也就是两个集合完全相同。这个时候称这个命题的两个条件为等价条件。

利用集合运算，我们还可以引入否命题。给定了集合 A, B 和全集 $\mathcal{U}$ ，也就给定了 A, B 的补集 $\bar{A}, \bar{B}$ 。因此，既然我们可以判断 $A \subset B$ 是否成立，就自然也可以判断 $\bar{A} \subset \bar{B}$ 是否成立。因此， $\bar{A} \rightarrow \bar{B}$ 就被成为 $A \subset B$ 的否命题。“否”就来自于补集，或者说对示性函数 I_A 做逻辑“否”运算，就得到示性函数 $I_{\bar{A}}$ 。

有了逆命题和否命题，我们还可以得到逆否命题： $\bar{B} \rightarrow \bar{A}$ 就被成为 $A \subset B$ 的逆否命题，因为可以通过先做逆命题（ $A \rightarrow B$ 变成 $B \rightarrow A$ ），再做否命题（ $B \rightarrow A$ 变成 $\bar{B} \rightarrow \bar{A}$ ）的操作而得到逆否命题，或者先做否命题，再做逆命题的操作。

通过集合运算，见图 23.4，我们还可以得到原命题和逆否命题等价，也就是，如果 $A \rightarrow B$ 成立，则 $\bar{B} \rightarrow \bar{A}$ 也成立，反之亦然。我们还可以发现，逆命题和否命题之间也是逆否命题的关系。因此，逆命题和否命题也等价。

习题 23.1 (逆命题和否命题). 用集合的语言和先做逆再做否 (或者先做否再做逆) 两种方式证明逆命题和否命题之间是逆否命题的关系。

有了逆否命题和原命题等价, 我们来定义几个名称: 必要条件、充分条件和充分必要条件。如果 $A \rightarrow B$ 成立, 我们就说 A 是 B 的充分条件。其含义是, 只要满足了条件 A 也就有了结论 B 。如果 $B \rightarrow A$ 成立, 我们就说 A 是 B 的必要条件。其含义是, $B \rightarrow A$ 意味着 $\bar{A} \rightarrow \bar{B}$, 也就是说如果 A 不成立, 则 B 肯定不成立。至于 A 成立的时候, B 是否成立, 那 $B \rightarrow A$ 不能给出。

当 $A \rightarrow B$ 和 $B \rightarrow A$ 两者都成立的时候, 我们就说 A 是 B 的充分必要条件, 简称充要条件。

如果原命题和逆命题同时成立就代表了这个命题的条件和结论对应着相同的集合, 因此, 逆否命题就相当于说 A 就是 B , B 就是 A 。这样的条件特别好用。比如说“直角三角形满足 $a^2 + b^2 = c^2$ ”、“满足 $a^2 + b^2 = c^2$ 的三角形是直角三角形”就是同时成立的一对原命题和逆命题。有了这个, 我们就可以把 $a^2 + b^2 = c^2$ 看作是直角三角形的另一个名称, 只不过前者需要检验三条边来看是否成立, 后者就看一个角是否是直角。具体用那一条来判断一个三角形是不是直角三角形, 就看当时的具体情况。我们特别喜欢这样把一个对象从另一个角度完全对应到另一个对象的定理。因此, 对于任何定理, 我们都可以去问一问其逆定理成立吗。

不过, 显然, 不是所有的成立的命题, 其逆命题都成立。比如说, “对顶角相等”成立, 但是我们没有“相等的角是对顶角”。也就是说, 如果两个角相等, 这两个角不一定就是对顶角。

习题 23.2 (相等的角和对顶角). 用集合的语言表示出来相等的角和对顶角的关系。注意, 你需要选用合适的全集。

在数学的知识体系中, 我们把假设成立的命题叫做公理, 把通过公理可以证明其成立的命题叫做定理。

23.2 定义

数学中的概念的定义, 通常有这样的阐述“一个满足条件 q 的对象 A 被称为 C ”, 往往被定义在之前的更加简单的概念 (这里的对象 A 和条件 q

都是之前就定义好的)的基础上,也是一个命题,并且是一个可逆命题。

例如,“有一个角是直角的三角形被称为直角三角形”有:第一、条件 q ——“至少有一个角是直角”;第二,对象 A ——三角形,第三,概念 C ——直角三角形。有的时候,条件 q 是一个有多个子条件的复合条件。例如,三条边和三个角分别对应相等的两个三角形被称为全等三角形。有的时候,也可以没有条件 q 。例如,等边三角形也被称为正三角形。

我们把“一个满足条件 q 的对象 A ”记作 A ,则定义就可以表达为 $A \leftrightarrow B$,其含义是 $A \rightarrow B$ 和 $B \rightarrow A$ 这一对可逆命题同时成立。因此,从“两个三角形有三条边和三个角分别对应相等”可以得到“这两个三角形全等”,“两个三角形全等”可以得到“这两个三角形的三条边和三个角分别对应相等”。

也就是说,出现在定义中的“是、被称为”是集合符号中的相等的意思,相互包含的意思,是逻辑符号中的双向箭头的意思,是充分必要条件的意思。出现在一般定理中的“是”只具有“被包含于”的意思。例如“直角三角形是三角形”,表示直角三角形这个集合被包含于三角形这个集合之中。

有了定义、公理、定理的概念,那我们现在就差论证方法,就可以来构建数学知识的体系了。

23.3 论证

现在,我们来讨论论证。一个命题 $\mathcal{A}$,如果可以通过逻辑演绎证明,得到一个命题 $\mathcal{B}$ (通常 $\mathcal{A} \neq \mathcal{B}$),则我们称这个 $\mathcal{A} \Rightarrow \mathcal{B}$ 的过程为论证。

这里就产生了一个问题:我们能使用的逻辑演绎证明的方式有哪些,为什么它们能用,如果它们的数量有限为什么只有那些?稍后我们来回答这些个论证部分要回答的主要问题。我们先在例举这些证明方式之前来做一些概念上的准备。

当我们来论证 $\mathcal{A} \Rightarrow \mathcal{B}$ 的时候,我们从来不管 $\mathcal{A}$ 本身是否成立。我们关注的是:如果 $\mathcal{A}$,是否可以证明 $\mathcal{B}$ 成立。这样我们就可以把 $\mathcal{A}$ 的检验的问题和 $\mathcal{A} \Rightarrow \mathcal{B}$ 的论证的问题完全分开。前者可以通过事实来检验,也可以反向找到证明 $\mathcal{A}$ 成立的那些直接的和间接的 $\mathcal{C}$ (例如,直接地 $\mathcal{C} \Rightarrow \mathcal{A}$),然后去对它们做事实性的检验,或者看看它们是不是就是假设成立的公理

或者定义。

逻辑学的核心任务就是把对前提和结论的检验和对论证过程的检验分开。正是这个分离，使得逻辑学可以关注做正确的论证过程的方法，而不去关注前提或者结论是否成立——那是一个个具体学科在具体领域里面的任务。反过来，只要人们检验通过了前提，论证方法也正确，则结论必然成立。

23.3.1 逻辑三律的复习

我们还是从集合开始。对于一个集合 A ，我们有元素明确性的要求，也就是对于给定的世界 $\mathcal{U}$ 中的任何一个元素 $\mu \in \mathcal{U}$ ，我们必须给出来 $\mu \in A$ 和 $\mu \notin A$ 两者必须和只能取其一。前面我们已经把命题和集合对应起来：对于命题 T 定义了原像 $A_T = T^{-1}(\{1\})$ ，对于集合 A 定义了其上的示性函数 $I_A(\mu)$ 来当作命题。

一个逻辑变量就是取值等于 0 或者 1 的变量，并且一个逻辑变量往往是对世界中的某些对象的判断，也就是说是一个从世界 $\mathcal{U}$ 到 $\{0, 1\}$ 的映射。于是，一个逻辑变量自然就和一个集合相对应。

既然集合要求元素的明确性，就相当于要求集合上定义的示性函数的相应的性质，也就是逻辑变量的相应的性质。我们之前在 8.3 节中已经学习过逻辑变量要满足逻辑三律，也就是

同一律 一个变量的符号和名称 a 在任何场合中都表示 a 。不能出现同一个变量指代不同对象。

排中律 $a = 1$ 和 $a = 0$ 必须有一个成立。

无矛盾律 $a = 1$ 和 $a = 0$ 不能同时成立。

通过示性函数我们可以把这些要求和集合的性质联系起来。排中律相当于要求 $I_A(\mu) = 1$ 和 $I_A(\mu) = 0$ 必须有一个成立，也就是 $\mu \in A$ 和 $\mu \notin A$ 必须有一个成立。当然，这样没有防止 $\mu \in A$ 和 $\mu \notin A$ 都成立。于是，无矛盾律保证了 $I_A(\mu) = 1$ 和 $I_A(\mu) = 0$ 不能同时成立，也就是 $\mu \in A$ 和 $\mu \notin A$ 不能同时成立。这正好就是“ $\mu \in A$ 和 $\mu \notin A$ 两者必须和只能取其一”。因此，排中律和无矛盾律完全就对应着集合元素的明确性。

至于同一律，这是用符号来指代对象的最起码的要求：同一个名称或者符号的变量，其含义必须是同一个。否则，就没法讨论任何问题。

顺便，正是集合和逻辑变量之间的这个关系，才使得可以用数学来研究逻辑，也才使得我们可以用集合来表示逻辑变量、命题和证明。

23.3.2 命题关系当作命题以及反证法

在那之前，我们来讨论一个更加抽象的概念性问题： $\mathcal{A} \Rightarrow \mathcal{B}$ 这个证明过程是否可以看作是一个命题，记作 $\mathcal{A} \rightarrow \mathcal{B}$ 。例如，“从‘三角形内角和是 180° ’可以证明‘同位角相等则两直线平行’”、“从‘同位角相等则两直线平行’可以证明‘内错角相等则两直线平行’”。显然，这也是一个可以判断是否成立的陈述句，那么，按照命题的自然语言的定义，这是一个命题。

但是， $\mathcal{A}$ （记为 $a_1 \rightarrow a_2$ ）对应着一个集合 a_1 和 a_2 之间的关系， $\mathcal{B}$ （记为 $b_1 \rightarrow b_2$ ）对应着一个集合 b_1 和 b_2 之间的关系。这样来看， $\mathcal{A} \Rightarrow \mathcal{B}$ 就相当于是集合关系之间的关系。这个和集合之间的关系，也就是命题 $a \rightarrow b$ 这种，还是不一样的。

回顾“从‘三角形内角和是 180° ’可以证明‘同位角相等则两直线平行’”的证明过程，我们采用了反证法：先假设“有同位角相等的两条直线不平行”，然后，得到这个假设和“三角形内角和是 180° ”矛盾。因此，如果我们接受“三角形内角和是 180° ”则就不能接受“有同位角相等的两条直线不平行”，也就是只能接受“所有同位角相等的两条直线都平行”。

用集合的语言表示，如图 23.5 所示，我们假设同位角相等（记作集合 $a_{\angle_1=\angle_2}$ ）则两直线平行（记作集合 $a_{l_1 \parallel l_2}$ ），也就是 $a_{\angle_1=\angle_2} \subset a_{l_1 \parallel l_2}$ 不成立，于是必然存在一个元素 Q ， $Q \in a_{\angle_1=\angle_2}$ 但是 $Q \notin a_{l_1 \parallel l_2}$ 。接着我们发现，这相当于构成了一个内角和不等于 180° 的三角形——交出来同位角的那条直线和两条相交（我们假设其不平行，则肯定有交点）的直线，构成的三角形中，外角等于其中一个内角，于是这个三角形的内角和肯定大于 180° 。这样，我们就发现，其实 $Q \in \Delta$ 是一个三角形，但是 $Q \notin P_{180^\circ}$ ，也就是其内角和不等于 180° 。这个和已知的结论“ $\Delta \subset P_{180^\circ}$ ”矛盾。

也就是说，我们证明了 $\bar{\mathcal{B}} \Rightarrow \bar{\mathcal{A}}$ ，而在我们的知识体系中 $\mathcal{A}$ 不可能是成立的，因此 $\bar{\mathcal{B}}$ 不可能是成立的，因此 $\mathcal{B}$ 是成立的。在这里，我们用到了逻辑学三律中的排中律和矛盾律。

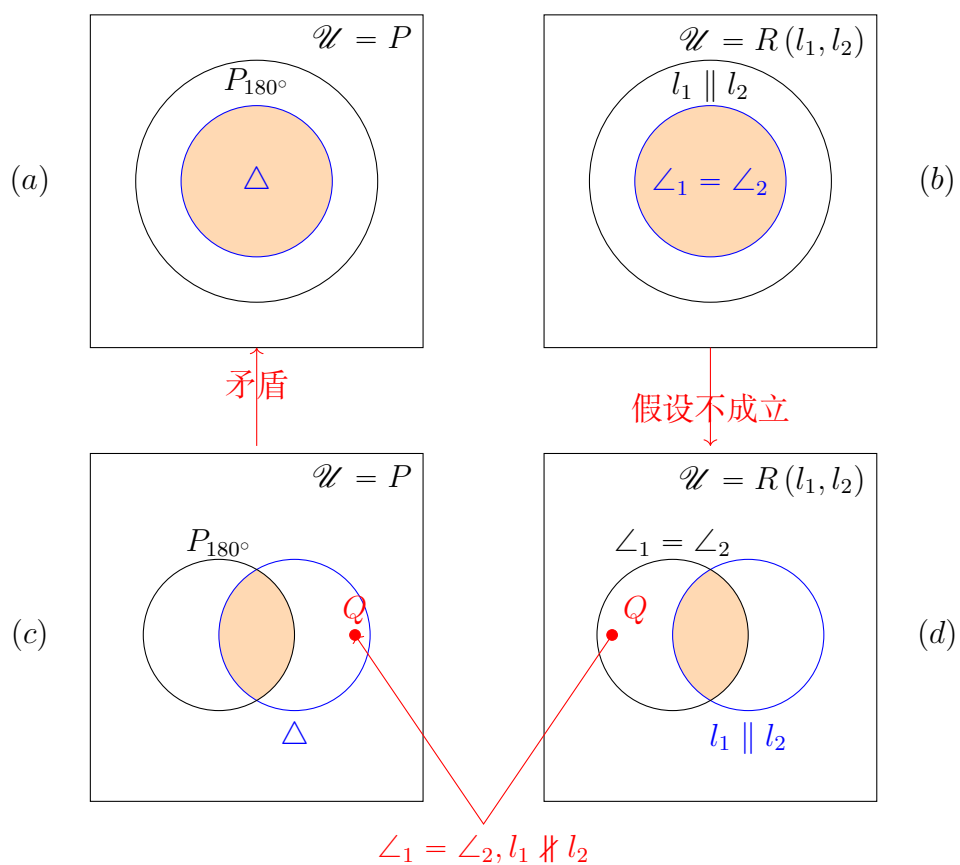


图 23.5: 集合关系的关系和反证法。论证讨论的是集合关系的关系, 其远远比集合关系也就命题要复杂。但是, 看起来反证法仍然能够使用。

通过这个例子，我们看到论证过程是从一个集合关系——也就是一个命题，到另一个集合关系——也就是另一个命题的关系的讨论。这个讨论远远要比判断命题是否成立——那时候就是讨论集合之间的关系。因此，简单地把论证过程 $\mathcal{A} \Rightarrow \mathcal{B}$ 看作一个命题 $\mathcal{A} \rightarrow \mathcal{B}$ 看起来似乎是不合适的。

但是，按照命题的自然语言定义， $\mathcal{A} \Rightarrow \mathcal{B}$ 显然是可以判断是否成立的，甚至把 $\mathcal{A}$ 当作条件，把 $\mathcal{B}$ 当作结论也很说得过去。甚至，我们看到，如果把 $\mathcal{A} \Rightarrow \mathcal{B}$ 看作一个命题 $\mathcal{A} \rightarrow \mathcal{B}$ ，则我们可以用逆否命题和原命题的等价性来： $\mathcal{A} \rightarrow \mathcal{B}$ 是否成立和 $\bar{\mathcal{B}} \rightarrow \bar{\mathcal{A}}$ 是否成立完全相同。

于是，我们发现，把 $\mathcal{A} \Rightarrow \mathcal{B}$ 看作 $\mathcal{A} \rightarrow \mathcal{B}$ ，除了用集合表示的时候要注意，在完全命题、条件的层面，完全是可以的。因此，从今往后，在不回到命题和证明过程的集合表示的场合，我们不再区分 $\Rightarrow$ 和 $\rightarrow$ 这两个符号。

按照原命题 $A \rightarrow B$ 和逆否命题 $\bar{B} \rightarrow \bar{A}$ 的之间的等价关系，当我们证明原命题遇到困难的时候，我们可以选择去证明其逆否命题。现在，我们不再区分命题和证明过程了，意味着，我们可以把证明过程 $\mathcal{A} \Rightarrow \mathcal{B}$ 用其等价形式 $\bar{\mathcal{B}} \Rightarrow \bar{\mathcal{A}}$ 来替换。从今往后，我们把这两种都称为反证法。

23.3.3 三段论

我们要学习的证明方法中最重要的是三段论。其实，我们已经用过很多次三段论。例如，从“三角形内角和是 180° ”，我们可以得到“直角三角形两个锐角互余”，那是因为“直角三角形是三角形”于是其内角和也等于 180° ，再根据定义知道其有一个角是直角，也就是 90° ，于是，另外两个角之和等于 $180^\circ - 90^\circ = 90^\circ$ 。更一般地，三段论表现为：

大前提 A 具有某性质 P_A ；

小前提 b 是一个 A ，或者 B 是一类 A ；

结论 b, B 都具有性质 P_A 。

我们来翻译成集合的语言，

大前提 $A \subset P_A$ ；

小前提 $b \in A$ ，或者 $B \subset A$ ；

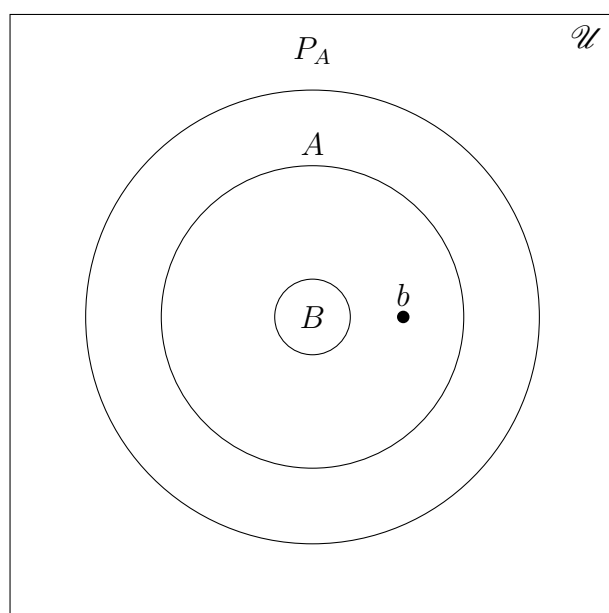


图 23.6: 三段论的集合表示。用 Venn 图展示了三段论。

结论 $b \in P_A$, $B \subset P_A$ 。

例如,

大前提 所有人都会死。 $P \subset D$, P 代表人的集合, D 代表会死的东西的集合;

小前提 苏格拉底是人 $s \in P$, 或者男人是人 $M \subset P$, 其中 s 代表苏格拉底, M 代表所有的男人的集合;

结论 $s \in D$, $M \subset D$ 。

我们来做个练习。

习题 23.3 (直角三角形锐角互余的三段论). 把上面的“直角三角形两个锐角互余”的证明过程呈现为三段论的形式。

我们来从集合的角度看三段论。我们把一般三段论中的大前提和小前提, 画成集合图, 就得到图 23.6。于是, 很自然, 就有其中列出来的结论。例如, 我们有 $b \in A \subset P_A$, 自然 $b \in P_A$ 。因为 $A \subset P_A$ 的含义就是对于任意的 $b \in A$, $b \in P_A$ 。也就是说, 从集合来看, 三段论是特别简单的和必然成立的。

在证明过程中，我们还把三段论用这样的格式来表示，

$$\begin{aligned} & \therefore \text{小前提,} \\ & \therefore \text{结论 (大前提).} \end{aligned} \quad (23.5)$$

如果最后命题证明完毕，我们还用专门的证明结束符号 $\square$ 来表示。例如，记直角三角形的三个角分别为 $\angle A, \angle B, \angle C$ 其中 $\angle C = 90^\circ$ ，我们来表示之前的证明：

$$\begin{aligned} & \therefore \text{直角三角形是三角形,} \\ & \therefore \angle A + \angle B + \angle C = 180^\circ (\text{三角形内角和是 } 180^\circ). \end{aligned} \quad (23.6)$$

然后，

$$\begin{aligned} & \therefore \angle C = 90^\circ, \\ & \therefore \angle A + \angle B + 90^\circ = 180^\circ (\angle A + \angle B + \angle C = 180^\circ, \text{等量替换}). \end{aligned} \quad (23.7)$$

最后，

$$\begin{aligned} & \therefore \angle A + \angle B + 90^\circ = 180^\circ, \\ & \therefore \angle A + \angle B = 90^\circ (\text{等式的性质}) \quad \square \end{aligned} \quad (23.8)$$

我们也可以把上面的步骤整合起来，成为

$$\begin{aligned} & \therefore \text{直角三角形是三角形,} \\ & \therefore \angle A + \angle B + \angle C = 180^\circ (\text{三角形内角和是 } 180^\circ), \\ & \quad \text{又 } \therefore \angle C = 90^\circ, \\ & \therefore \angle A + \angle B + 90^\circ = 180^\circ (\text{等量替换}), \\ & \therefore \angle A + \angle B = 90^\circ (\text{等式的性质}) \quad \square \end{aligned} \quad (23.9)$$

这样做的好处是我们可以跳过一些小前提的重复书写，逻辑上连起来，自然表示依赖了前面已经得到的结论来当作后面一步的小前提。

你需要学会通过 $\therefore, \because, (\cdot)$, $\square$ 这几个关键词来表达证明过程。建议你重新去看一下本书的很多命题的证明过程。

那问题来了：为什么基于这么简单的一个集合包含关系的三段论，会是证明命题的主要方式呢？注意，甚至我们用了反证法，从 $\mathcal{A} \Rightarrow \mathcal{B}$ 替换成 $\bar{\mathcal{B}} \Rightarrow \bar{\mathcal{A}}$ 的时候，后者的证明也还是需要通过三段论来完成的。

那是因为我们给事物下定义——概念化——也是这个模式：现有一个基本概念 A ，再加上额外条件 q ，定义为导出概念 B ，而下定义是我们描述事物描述世界的最主要的方式。当然，概念有了之后要研究命题，有概念之前我们需要接受现实世界、想象力和直觉的启发，有了命题之后，在科学中还要接受现实世界的检验。但是，无论如何我们都是从对事物概念化开始的对事物的研究。

例如，我们先有了集合，再得到两个集合的卡氏积，进而把卡氏积满足一定条件的子集定义为映射。例如，我们先有了平面直角坐标系（那之前得有实数的定义）、点和距离，然后就可以定义线、圆这些几何对象——它们都是满足某些要求的点的集合。例如，有了线，加上是否有交点可以定义相交和平行，再加上夹角可以定义垂直，加上三条线首尾相连的条件，可以定义三角形。有了三角形加上有一个角是直角就可以定义直角三角形，加上一段线段长度相等这个条件就可以定义等腰三角形。等等等等。

于是，我们就发现，当我们需要来研究直角三角形的（判定和性质）时候，肯定得依靠三角形（的判定和性质）来当大前提；当我们来研究等腰三角形的时候，肯定得依靠三角形肯定得依靠三角形（的判定和性质）来当大前提；当我们来研究三角形的时候（例如证明三角形内角和是 180° ），肯定得依靠线和距离公式来当大前提²。等等等等。

也就是说，概念化，这个人类认识世界描述世界的最重要的方式之一，和三段论的结构是相同的。这是三段论如此威力巨大的原因。

23.3.4 数学归纳法

我们要学习的下一个论证方法是数学归纳法。其表述为：

第一步 某个命题 A 对于 $N = N_0$ 成立；

第二步 如果这个命题对于 $N = k$ 成立，则对于 $N = k + 1$ 也成立；

²在实际证明这个定理的时候，我们还需要引入中间概念——角的三角函数和三角函数的和角公式、三角形内的余弦定理，才能建立起来从大前提到要证明的定理之间的联系。

结论 则这个命题对于所有的 $N \geq N_0$ 都成立。

我们先通过一个例子来学习一下这个论证方法的使用，然后来思考为什么这个论证方法是有效的。

例 23.1 (数学归纳法用于证明). 用数学归纳法证明 $1+2+\cdots+N = \frac{N(N+1)}{2}$ 。

我们来证明第一步: $N = 1$ 的时候, 等式成立: $1 = \frac{1 \times 2}{2} = 1$ 。

我们再来证明第二步: 假设 $N = k$ 是等式成立, 也就是 $1+2+\cdots+k = \frac{k(k+1)}{2}$, 则对于 $N = k+1$, 有

$$\begin{aligned} 1+2+\cdots+k+(k+1) &= \frac{k(k+1)}{2} + (k+1), \\ &= (k+1) \left(\frac{k}{2} + 1 \right), \\ &= \frac{(k+1)(k+2)}{2}. \end{aligned} \quad (23.10)$$

因此, 待证明等式也成立。

第三步, 我们得到结论, $1+2+\cdots+N = \frac{N(N+1)}{2}$ 对于所有 $N \geq 1$ 都成立。

当然, 我们可以通过首尾组合, 把 $1+2+\cdots+N = (1+N) + (2+N-1) + \cdots + \left(\left[\frac{N}{2}\right], \left[\frac{N}{2}\right] + 1\right)$, 来得到上面的待证明等式。其中 $[q]$ 表示取 q 的整数部分。但是, 采用数学归纳法的证明往往就不需要“首尾组合”这一类如此巧妙的构造, 只需要照着这个三个步骤来做。

例 23.2 (数学归纳法用于证明). 用数学归纳法证明 $1+3+5+\cdots+(2N+1) = (N+1)^2$ 。

我们来证明第一步: $N = 0$ 的时候, 等式成立: $1 = (0+1)^2 = 1$ 。

我们再来证明第二步: 假设 $N = k$ 是等式成立, 也就是 $1+3+5+\cdots+(2k+1) = (k+1)^2$, 则对于 $N = k+1$, 有

$$\begin{aligned} 1+3+5+\cdots+(2k+1)+[2(k+1)+1] &= k^2+2k+1+(2k+3), \\ &= k^2+4k+4, \\ &= (k+2)^2. \end{aligned}$$

(23.11)

因此,待证明等式也成立。

第三步,我们得到结论, $1 + 3 + 5 + \cdots + (2N + 1) = (N + 1)^2$ 对于所有 $N \geq 0$ 都成立。

当然,我们还是可以通过首尾组合来得到待证明等式。注意,这里我们从 $N = 0$ 开始。数学归纳法往往把所有证明的难点都集中到了第二步,并且数学归纳法一般用于有了对带证明表达式的猜想之后。

例 23.3 (数学归纳法用于求数列通项). 已知一系列数 $a_1, a_2, a_3, \cdots, a_N$ 满足 $a_{n+1} = a_n + (2n + 1)$, $a_1 = 1$ 。求 a_n 的表达式。

我们来求出来几项,用来提出猜想: $a_1 = 1$, $a_2 = a_1 + (2 \times 1 + 1) = 1 + 3 = 4$, $a_3 = a_2 + (2 \times 2 + 1) = 4 + 5 = 9$, $a_4 = a_3 + (2 \times 3 + 1) = 9 + 7 = 16$ 。我们提出猜想 $a_n = n^2$ 。

我们来证明第一步: $n = 1$ 的时候,等式成立: $a_1 = 1 = 1^2$ 。

我们再来证明第二步: 假设 $n = k$ 是等式成立,也就是 $a_k = k^2$, 则对于 $n = k + 1$, 有

$$\begin{aligned} a_{k+1} &= a_k + (2k + 1), \\ &= k^2 + 2k + 1, \\ &= (k + 1)^2. \end{aligned} \quad (23.12)$$

因此,待证明等式也成立。

第三步,我们得到结论, $a_n = n^2$ 对于所有 $n \geq 1$ 都成立。

习题 23.4 (数学归纳法用于证明). 用数学归纳法证明 $1^2 + 2^2 + 3^2 + \cdots + N^2 = \frac{N(N+1)(2N+1)}{6}$ 。

习题 23.5 (数学归纳法用于求数列通项). 已知一系列数 $a_1, a_2, a_3, \cdots, a_N$ 满足 $a_{n+1} = a_n + \frac{8(n+1)}{(2n+1)^2(2n+3)^2}$, $a_1 = \frac{8}{9}$ 。求 a_n 的表达式。

那数学归纳法为什么成立呢? 因为对于任何 $N \geq N_0$, 我们总是可以运用第二步的结论发现, 只要 $N - 1$ 成立, 待证明结论就成立; 于是, 继续运用第二步的结论发现, 只要 $N - 2$ 成立, 待证明结论就成立; 不断重复, 直到遇到只要 N_0 成立, 待证明结论就成立。而自然数天生满足这个不断能够减去 1 来得到更小的数直到某个起点的要求。因此, 数学归纳法能用完全是因为自然数的性质, 以及三段论。只不过, 第二步是大前提, 第一步是

小前提，然后反过来每次从 N 到 $(N + 1)$ 待证明结论成立是结论，也是下一步的小前提。因此，数学归纳法完全是逻辑演绎推理，而不是归纳，尽管出现了“归纳”这个词在其名称里面。

下面，我们来讨论什么是归纳。

23.3.5 不完全归纳法

通过对某一个对象的有限个样本的观察（或者测量，或者其他研究）而得到关于这个对象的一般性结论的过程，叫做归纳。这样来作结论的方式被称为不完全归纳。不完全在这里的意思就是“有限个这个对象的样本”，而不是“所有的这个对象的样本”。如果对所有的这个对象的样本都做了这个结论是否成立的检验，则称为列举法，穷举法，就是把每一个样本都穷尽都列出来的意思。

比如说，当我们看到很多只乌鸦都是黑色的，就说“天下乌鸦都是黑色的”。例如，当我们看到的每一个人都最后会死亡，就是“所有人都会死亡”。例如，我们前面求解 a_n 的表达式的时候，通过看前面的四项而总结出来其一般形式。

不完全归纳法是不能用来证明一个结论成立的。因为从任何有限个样本总结出来的这个对象的特性都有可能不被后面的其他样本遵从。除非，在总结归纳出来可能的结论之后，再通过数学归纳法或者三段论、反证法等方法来证明这个结论。

不过，既然如此，为什么会有不完全归纳法呢？既然有这个所谓的方法，肯定还是有用的。比如说，在总结猜测出来结论这一点上，不完全归纳法还是很有用的。我们已经在求解 a_n 的表达式的时候见识过。其实，还有一点，把不完全归纳法看作是带概率的逻辑演绎推理，则不完全归纳法完全就有了它在整个论证方法、逻辑学以及自然科学中的适当的位置。

用条件概率，我们可以这样来表示不完全归纳法：已知从有限个样本 $\{a_1, a_2, \dots, a_N\}$ 中提炼出来结论 S 成立的可能性 $P(S(a_N), S(a_{N-1}), \dots, S(a_1))$ ，我们想推断出来对于下一个样本结论 S 成立的可能性 $P(S(a_{N+1}))$ 。为了

记号简单, 我们约定 $S(a_n) = S_n$ 。按照全概率公式算式 (20.26), 我们有

$$\begin{aligned} & P(S_{N+1}) \\ &= P(S_{N+1}|S_N, S_{N-1}, \dots, S_1) P(S_N, S_{N-1}, \dots, S_1) \\ &+ P(S_{N+1}|\overline{S_N}, S_{N-1}, \dots, S_1) P(\overline{S_N}, S_{N-1}, \dots, S_1). \end{aligned} \quad (23.13)$$

考虑之前的样本结论 S 都成立的情况, 则第二项为零, 得到

$$P(S_{N+1}) = P(S_{N+1}|S_N, S_{N-1}, \dots, S_1). \quad (23.14)$$

运用贝叶斯公式算式 (20.24) 可以得到

$$P(S_{N+1}|S_N, S_{N-1}, \dots, S_1) = \frac{P(S_N, S_{N-1}, \dots, S_1|S_{N+1}) P(S_{N+1})}{P(S_N, S_{N-1}, \dots, S_1|S_{N+1}) P(S_{N+1}) + P(S_N, S_{N-1}, \dots, S_1|\overline{S_{N+1}}) P(\overline{S_{N+1}})} \quad (23.15)$$

如果我们进一步假设这些观察到的样本都是独立的, 那么

$$P(S_{N+1}|S_N, S_{N-1}, \dots, S_1) = \frac{P^N(S_n|S_{N+1}) P(S_{N+1})}{P^N(S_n|S_{N+1}) P(S_{N+1}) + P^N(S_n|\overline{S_{N+1}}) P(\overline{S_{N+1}})}. \quad (23.16)$$

其中既然假设了独立性则 $P(S_n|S_{N+1})$ 和具体的那两个样本就没关系, 也就是说, 就是一次看到结论 S 成立的样本的条件下, 下一次看到的样本结论 S 还成立的可能性。我们把这个概率记作 q 。把 $P(S_n|\overline{S_{N+1}})$ 记作 p , 其含义是一次看到结论 S 不成立的样本的条件下, 下一次看到的样本结论 S 成立的可能性。于是,

$$\begin{aligned} P(S_{N+1}|S_N, S_{N-1}, \dots, S_1) &= \frac{q^N P(S_{N+1})}{q^N P(S_{N+1}) + p^N P(\overline{S_{N+1}})} \\ &= \frac{1}{1 + \left(\frac{p}{q}\right)^N \frac{P(\overline{S_{N+1}})}{P(S_{N+1})}}. \end{aligned} \quad (23.17)$$

因此, 只要 $p < q$, 则必然有随着 N 增加, 结论成立的可能性不断地增加的结论。在 $N \rightarrow \infty$ 的“极限”下, 结论成立。那么, $p < q$ 会成立吗? $P(S_n|S_{N+1}) > P(S_n|\overline{S_{N+1}})$, 看到一次结论成立的样本之后, 看到下一个样本结论还是成立的概率, 比看到一个结论不成立的样本之后看到下一个样

本成立的概率，要大。当我们不知道任何关于这个系统的其他信息的时候，这个结论很有道理。比如说，一个袋子里面装了不知道比例的很多很多的红球和黑球。当我们随机摸出来一个球发现是红球，然后我们推测下一个球还是红球的概率，显然比随机摸出来一个球发现是黑球，然后我们推测下一个球还是红球的概率要高。

未来，我们可以通过对于具体例子作计算来真的算出来这些概率（见例如 [12]），看到其随着 N 变大而增加。今天我们只需要有一个大概的概念性的认知就可以。

23.4 推荐阅读材料

吴金闪《系统科学导引》第二卷。

23.5 作业

23.6 本章小结

在本章中，我们从集合和映射开始，学习了什么是命题，什么是定义，什么是论证，都有哪些论证方法——这里具体来说我们学习了反证法、三段论（三段论的 $\therefore, \because, (\cdot)$ 符号表示）、数学归纳法。它们都是逻辑演绎推理。我们还看了看什么是不完全归纳法——带概率的逻辑演绎推理。

我们说，数学知识的确立一定是要通过逻辑演绎证明的。但是，并不是说其他的思维方式，例如不完全归纳法，就不重要。逻辑仅仅帮助你判断论证的对错，做更严格的论证，而证明什么命题，提出什么概念和哪些公理³，并不是逻辑要完成的事情。这些靠的是对知识结构的深刻的认知，尤其是长程联系的建立，靠直觉、想象力和现实世界的启发。

在这一章中，我们再一次体会到，一切都是从集合开始的。

³这个提出哪些公理的问题更复杂一些：一方面公理来自于现实的启发和想象力，另一方面也来自于尝试从不同公理出发来建立整个知识体系，然后选择那些体系更加自然更加严格的来当作公理。

实际上，如果你数学知识（和学科知识）理解的透彻，其实逻辑演绎证明是不需要专门学习的，你完全可以通过学习那些数学证明题的证明（那些科学知识怎么获得）来学会逻辑演绎证明。

第二十四章 数学建模

What distinguishes a mathematical model from, say, a poem, a song, a portrait or any other kind of "model," is that the mathematical model is an image or picture of reality painted with logical symbols instead of with words, sounds or watercolors.

数学模型和其他的模型，例如诗、歌、绘画，不同的地方在于，数学模型是用逻辑符号而不是词、声音、水彩画出来的现实世界的像。

– John L. Casti

This model will be a simplification and an idealization, and consequently a falsification. It is to be hoped that the features retained for discussion are those of greatest importance in the present state of knowledge.

模型将会是一个简化，一个理想化，因此也就是一个错误化。只不过，人们希望，这些被模型保留下来的特征对于当前的（关于这个模型所描述的对象）知识来说仍然是重要的。

– Alan Turing

All models are wrong, but some are useful.

所有模型都是错的，但是其中有些它管用。

– George Box

If you want to learn about nature, to appreciate nature, it is necessary to understand the language that she speaks in. The language is mathematics.

如果你想了解自然，欣赏自然，你就有必要理解自然所说的语言——
这个语言是数学。

– Richard Feynman

在本书的每一个知识的学习的过程中，我们都尽可能地渗透了数学建模的思想。在这一章中，我们专门来学习数学建模。

24.1 什么是数学建模和数学模型

数学模型指的是用数学符号以及这些符号所代表的数学结构，例如集合、三角形、实数和实数上的运算、概率，来得到的对现实世界中的相应的被建模对象的描述。所谓对被建模对象的描述指的是在这个建好的模型上，能够对对象做的操作，也能够用这个模型的数学操作来描述，而且数学操作得到的结果，得在无差范围内以及人们关注的这个对象的属性范围内，和对这个对象的操作的结果一致。一般来说，一个数学模型仅仅抓住相应的被建模对象的一部分属性和一部分操作。只要在这一部分数学操作的结果和实际操作的结果能够在无差范围内对上，我们就说这是一个合理的数学模型。

数学建模指的是给现实世界中的被建模的对象构建出来这样的数学模型的过程。对于我们学习数学模型和数学建模的任务来说，数学建模才是重点。数学模型不过就是学会数学建模的素材。一会儿我们会看到，在数学建模的过程中，我们肯定是要扔掉被建模对象的一些属性的。这是一个简化、理想化的过程，也使得我们解决问题更专注，解决这个问题可能性更高。这个过程也被称为模型化。

数学建模的主要任务，就是搞清楚是什么（What）层面的问题：我们要解决关注这个现实世界的被建模对象的什么问题，这些问题和这个对象的什么状态和什么操作有关系，这些状态和操作可以用什么样的数学结构来描述；以及怎么操作怎么计算（How）层面的问题：这些数学结构现有数学知识中已经有了吗，如果有了的话我们相应的状态和操作的数学表达式和数学计算是什么样的，有可以实现这些操作的数学计算的方式吗？

当然，背后我们也得搞清楚为什么（Why）层面的问题：为什么要我们要解决这个对象的这个问题，为什么我们要解决这个问题关注这些对象和操

作，为什么这些对象和操作需要用这些的数学结构来描述。一旦解决了这个问题，我们还得反思总结梳理，问对数学和你自己这个问题解决者的意义（Meaningful）的问题：解决这个问题的过程中有创造新的数学对象吗，有已有数学对象的创造性使用（第一次把这个数学结构用来描述这种现实世界的对象）吗，模型管用吗，模型可以推广吗，建模的方法和背后的思想可以推广吗，我自己从中收获了什么？

从人类文明发展的角度来看，或者说假设我们每一个学习者都有足够的时间和创造力，则我们不需要为数学建模提前准备好关于数学结构的知识。如果我们已经掌握的数学知识足够我们来描述世界解决问题，就用；否则，如果出现我们掌握的数学知识还不能描述我们关注的对象解决相应的问题，就创造出来新的数学结构来描述它从而解决问题。因此，数学建模是双向的：一方面是运用现有的数学结构来解决现实世界的问题，一方面是通过解决现实世界的问题来推动创造新的数学结构。

我们先从后者的角度来看数学知识的发展。

24.2 从数学建模的角度看数学知识的发展

我们知道数来自于对身边的事物的计数的实际生活需求。比如说，我们排除了几个人去打仗或者狩猎，回来了多少人，损失了多少人，带回来了多少战利品多少猎物，怎么根据派出去的人和回来的人以及部落里面所有的人的数量来分配。比如说，过去了几天（可能是按照每一次太阳出来或者月亮出来来算一天），去了几个不同的地方，其中有几个地方可以有猎物或者可以采集到食物，再过去几天有可以看到月亮了于是可以开个聚会、休息休息或者某种食物刚好隔几个月亮的时候就可以再次去获取了。所有的这些都会启发自然数和加减法的提出。

如果我们要对同一个单元来重复做加减法，比如说我们可以把若干个天称作月，表示月亮型态的一个循环，那么当我们说若干个月的时候，我们就需要来转化成相应的天。当我们直到某种食物需要过若干个月就会再次可采集的时候，我们就可以不用天需要出现很大的数的这个单位，而用月这个用很小的数就可以来描述这个食物再次被采集的间隔的单位。这就是后来的乘法和除法。

因此，自然数和自然数上的加减乘除就是从实际生活的需要中被创造出来的，是数学建模的结果，是做抽象的结果。

至于做完这个抽象以后，什么时候人们意识到这些抽象出来的结构其实不依赖于所关注的对象是什么，其单位是什么，那就是另一个问题了。这个从具体事物中提炼出来数学结构，然后看到这个数学结构数学结构具有一般性可以被用到其他对象上，应该不是一件很容易的事情。当然，这里我们假设我们的原始人足够聪明和有闲，会做抽象思维，会形成概念给相应的名称和记号，也会把抽象以后的概念用于描述具体事物。

自然数一般来只论个，并且往往伴随着一个自然的粗糙的单位，比如说，“一个人”其实个子和强壮程度不同，“一个苹果”其实大小不同。当我们需要关注的更细致的时候，这个粗糙的单位和只论个就不够用了。例如，这个人的力气可以扛起来一只羊，另一个人可以扛起来两只羊。如果我们要决定派多少人去把另一群人已经猎好的羊带回来，我们就得考虑人之间的这个差异。如果我们猎到的是相当于五只羊的一头牛，则我们还需要按照这个新的比例再来安排人手。因此，比例数——尤其是处理由整数分不完的比例数也就是我们今天说的分数——的出现就是很自然的事情了。

顺便，非比例数，也就是无理数的出现确实更加抽象。很大程度上，这是需要数学概念——例如乘方、勾股定理——形成以后才能出现的数学结构。但是，你看，勾股定理本身可能也是受现实世界的启发得来的。因此，无理数至少间接上也是数学建模的结果。

随着我们的数学语言的丰富，随着科学技术的发展，我们能够接触到的现实世界也在拓展，我们需要解决的关于现实世界的问题也在拓展。我们一旦有了交易和用于交易的某种货币单位，例如贝壳、羊、黄金、银子、铜币、纸币，我们就有了去表达交易的金额和交易的量和交易的单位价格之间的关系 $V = P \times Q$ 的需要。我们希望在给定物品和物品的价格的时候，对于任何一个交易量都能够得到交易金额，这就是 $V = P_* \times Q$ ，其中价格 P_* 是一个常数；在给定物品和物品数量的时候，对于任何一个价格（例如讨价还价的时候，换一个卖家的时候，考虑我是不是买得起的时候）都能够得到交易金额，这就是 $V = P \times Q_*$ ，其中交易量 Q_* 是一个常数；在给定总的预算和特定的购买物品的时候，针对任何价格我们都能得到能够购买的数量，或者反过来，这就是 $P = \frac{V_*}{Q}$ 或者 $Q = \frac{V_*}{P}$ ，其中交易量 V_* 是一个

常数。这个时候，我们就需要变量、表达式和函数。我们甚至还会希望对于任何物品，只要得到其价格，得到人们要购买的数量，都可以用 $V = P \times Q$ 得到总的交易金额。这个时候， P, Q 都是自变量。你看很自然，我们需要有多个变量的代数式和函数、还有方程。方程指的是一旦关系给定，例如 $V = P \times Q$ ，这个关系不仅仅能够用于正面计算——已知等式的右边来求出来左边的结果，还可以用于反面计算——等式任何地方的已知和未知都可以算（例如 $V_* = P_* \times Q$ ，其中 Q 未知其他已知）。函数的背后，经过数学家的梳理，发现是集合和映射的概念。

在考虑对一个硬币的所有的操作的时候，我们就会发现：硬币有两个完全不会同时出现的状态，正面和反面；相应的操作有翻转——使得硬币的正面状态变成反面状态以及使得反面状态成为正面状态，以及不翻转——保持硬币的状态不变。如果我们用这样的硬币来做两个人之间的一个决定输赢的游戏，则我们就要提供一套可以记录和预期我的对家和我的行为的符号语言。这就是集合和群。

如果我们所描述的现实世界的对象经过一个给定的操作以后每次都出现一个结果，但是每次出现的结果不一定一样，则我们还需要提炼出来新的数学结构——概率。这个概念包含基元事件、基元事件的集合、集合上对交并补运算封闭的集合、集合到 $[0, 1]$ 之间的映射。当然，这些更基础和更抽象的概念都是后来的数学家在已经体会到需要概率这个数学结构来描述现实之后经过反向梳理才得到的。

注意，当这里我们把硬币用一个只有两个取值 $C \in \{0, 1\}$ 的变量来描述而且假设两面出现的概率值相等的时候，实际上，这是一个很大的近似。实际的硬币里面偶然还有立着的状态，有硬币的厚度、材质、重量两面不同的图案等等各种因素。但是对于我们前面的那个游戏输赢的场景来说，我们这个近似够用了。实际上，所有的模型都是为了解决问题而通过对所描述的对象简化得到的。

通过这些例子的复习，我们希望你看到：第一、本书的所有数学学科概念知识的引入基本上都是通过数学建模的方式——企图去描写某一类现实——而完成的；第二、整个数学也是，其发展过程中双向数学建模是发挥了巨大的作用的。

当然，数学里面也有一些是依靠对知识的系统性的追求，依靠人类的

想象力而建立的。例如集合、映射、从定义和公理到定理的知识体系、论证方法（逻辑学）都只能看作是间接受双向数学建模的推动，直接主要受到对知识系统性的追求。

今天的数学研究也是如此。有一部分数学家的研究工作是在解决数学知识内部的问题、提出按照数学知识的内部结构很可能有的命题、证明那些命题，另一部分数学家的研究工作是和物理学等其他自然科学和工程类的学科相结合，从中获得新的命题的灵感，以及提炼出来新概念，或者研究物理学里面的方程和问题的求解方法。

24.3 从把数学知识用于描述现实看真正的数学应用题——WHWM 求解数学应用题

按道理，数学应用题的目的就是培养数学建模的能力和养成数学建模的习惯。当然，跟真正的数学建模相比，数学应用题更偏向把已有的数学结构用于解决现实世界的问题，也就是双向数学建模的应用向。在这一小节，我们从数学建模的角度来看几道常见的数学应用题以及几道本书中的数学场景题。我们在学习完这一小节之后，你都把数学应用题来当作应用向数学建模，而不是通过“分题型、背口诀、套公式”来求解数学应用题。那要怎么来求解数学应用题呢？稍后我们会介绍应用题求解的 WHWM 方法。

例 24.1 (存款方式问题——应用题版本). 小明过年的时候收到了 3000 元的压岁钱。他想去银行存起来。银行有一个三年期的存款方案，每年的利率是 $r = 0.02$ (或者说 2%)。算一下三年以后小明可以获得多少钱。

最开始的钱是 $M_0 = 3000$ 元，存了一年以后的钱是 $M_1 = M_0 + M_0 \times r = M_0(1 + r) = 3060$ 元，再存一年到了第二年结束的时候的钱是 $M_2 = M_1(1 + r) = 3121.2$ 元，再存一年到了第三年结束的时候的钱 $M_3 = M_2(1 + r) = 3183.6$ 元。

也可以最后再来算具体数值，中间通过符号来运算，得到 $M_3 = M_0(1 + r)^2 = 3183.6$ 元。

注意，这里我们得到了一个一般的复利计算的表达式 $M_n = M_0(1 + r)^n$,

我们在下面将直接使用它。之后，我们回来关注为什么这个公式成立，因此可以直接使用。

例 24.2 (存款方式问题——数学建模版本). 小明过年的时候收到了 3000 元的压岁钱。他想去银行存起来。到了银行，他发现银行有提供不同的存款选项，有活期的（为简单计，假设这个活期没利息。实际上还有三个月的、半年的，咱们在此假装没有这些低于一年的存款选项¹⁾）、一年的、两年的、三年的、五年的。银行对每一个不同周期都给了一个年化收益率（每存一年的利息）分别为 $r = 0.02, 0.03, 0.04, 0.06$ 。按照那一种方案存？我要怎么来做这个决定？

当我们面对这样一个决策问题的时候，一开始，我们需要考虑哪些因素，抽象成哪些变量是不清楚的。我们的第一步就是搞清楚目标和所需要考虑的因素，第二步就是用数学结构（可能很多时候实数就够用了，甚至整数就够用了）来描述它们，第三步思考这些变量之间的关系以及这些关系对于你的目标来说的意义，看是否通过这个关系就解决了你的问题实现了你的目标，第四步去实践中检验一下或者换个角度考虑一下，或者反思一下决定主要因素的时候忽略的东西真的不重要吗，最后第五步如果检验通过对这个建模的过程所建立的模型作总结梳理。

我们来逐个完成以上的步骤。小明去存款的目的是什么：第一、保存在银行随时还能够用，安全和方便；第二、可以得到一些利息，相当于增加了钱。对于小明决策到底按照几年来存这笔钱的问题，第一个目的都能满足，因此不用考虑。也就是说，主要需要考虑的就是每个方案最后增加的钱的多少。

如果这样的话，我们把一开始的钱记作 M_0 ，存了 n 年结束以后的钱记作 M_n ，采用哪一种方案记作 $r^{(l)}$ ($l = 1, 2, 3, 5$)。我们简单做一个下面的计算，

$$M_1^{(l)} = M_0 (1 + r^{(l)}). \quad (24.1)$$

于是，肯定选那个 $r^{(l)}$ 最大的。在这里就是五年期的。

但是，问题是，五年期是什么意思呢？它并不是单纯一个名称。其含义是：如果你存够了五年，就按照这个 $r^{(l)}$ 来算每年的利息，但是，如果中间

¹⁾由于银行有一个所谓的“报价利率”的问题，导致低于一年的计算起来很别扭。

需要取出来, 则取出来的部分按照活期²计算。

这个时候, 我们发现, 有一个很重要的因素我们没有考虑: 我们预期多长时间内会用到这笔钱呢? 这个量用什么数学结构来描述呢? 这个问题又是由哪些因素决定的呢? 这些因素可以数学化吗, 这些因素如何决定这个量的关系可以数学化吗?

小明觉得: 用 $P(l)$, 其中 l 近似以年为单位, 来表示从今天开始过了 l 年之后我需要用到这笔钱的概率, 可能可以来描述我前面的“预期多长时间内会用到这笔钱”的问题。对于每一年来说, 我要么需要用到, 要么不需要用到。那一年之后到下一年的几天之前都算到前面的那一年。也就是到下一年的今天之前相当于 $l = 0$, 第一年结束之后到第二年的今天之前相当于 $l = 1$ 。

那么, 什么因素决定了我会某一年用到这笔钱呢? 例如, 大概四年之后我是要用到这笔钱的, 因为爸爸允许我小学毕业了购买一位我自己的机器人助手。那大概会用到我每一年的压岁钱的总和。那也就是说 $P_n = 1$ 对于 $n \geq 4$ 。剩下的, 我可能会考虑自己购买一些爸爸妈妈不一定同意给我买的玩具, 给姐姐、妈妈和爸爸买点节日和生日的小礼物。不过, 这些大概每年的压岁钱和妈妈给的零花钱就够用。所以, 我打算留下来 1000 元钱来当作这部分钱。但是, 估计这些因素不影响 P_n 。不对, 我想起姐姐三年以后估计要上大学了, 我肯定想要给她送个她自己的笔记本 (或者帮她出一部分钱)。按照每年差不多 3000 的压岁钱, 我估计这第一年的钱不会被用到或者说被用到的概率很小, 但是明年我存钱的时候得考虑这个可能性。

没准还有别的因素, 不过我现在能够想到的大概就这么多, 那大概来说: 我把 $M_0 = 3000$ 修改成 $M_0 = 2000$, 或者说另外的 1000 存活期; 然后, $P_1 = 0$, $P_2 = 0$, $P_3 = 0.1$, $P_4 = 1$ 。

有了这些量, 结合 $r^{(l)}$ 我们来计算每种方案的最后钱的数量。如果我

²具体银行规定不同, 有得可以按照已经存满的周期算。为了简单, 我们这里按照活期算。

每年都存一年期，有

$$\begin{aligned}M_1^{(1)} &= M_0 (1 + r^{(1)}) (1 - P_1) + M_0 P_1, \\M_2^{(1)} &= M_1^{(1)} (1 + r^{(1)}) (1 - P_2) + M_1^{(1)} P_2, \\M_3^{(1)} &= M_2^{(1)} (1 + r^{(1)}) (1 - P_3) + M_2^{(1)} P_3.\end{aligned}\quad (24.2)$$

注意，我只需要算到第四年也就是存满三年，因为第四年我估计之后都得拿出来用。如果我先都存一个两年期，再存一个一年期，

$$\begin{aligned}M_2^{(2)} &= M_0 (1 + r^{(2)})^2 (1 - P_1) (1 - P_2) + M_0 [1 - (1 - P_1) (1 - P_2)], \\M_3^{(2)} &= M_2^{(2)} (1 + r^{(1)}) (1 - P_3) + M_2^{(2)} P_3.\end{aligned}$$

(24.3)

如果我直接存一个三年期，则

$$M_3^{(3)} = M_0 (1 + r^{(3)})^3 (1 - P_1) (1 - P_2) (1 - P_3) + M_0 [1 - (1 - P_1) (1 - P_2) (1 - P_3)].$$

(24.4)

算出来之后，得到 $M_3^{(1)} = 2118.25$ ， $M_3^{(2)} = 2159.99$ ， $M_3^{(3)} = 2224.76$ 。因此，我要选择存三年的方案。

对比“存款方式问题——应用题版本”和“存款方式问题——数学建模版本”我们发现，最大的区别是：前者是一个已知和求明确的数学问题，而后者是一个连目标、主要因素、目标和主要因素的数学描述都不清楚的粗糙的问题。我们必须对因素做分析和简化以后才能抓住主要因素，进而考虑加上目标和主要因素之间的关系，才能解决问题。当然，目标、主要因素、目标和主要因素的数学描述都确定之后，基本上就成了一个数学应用题。也就是说，数学模型是把问题数学化的过程，而应用题是把数学问题求解出来的过程。

原则上，数学应用题的本来的目标就应该是包含把问题数学化、求解数学问题，甚至包含检验答案，做求解方法和过程的总结反思梳理。但是，实际上，现在的所谓的数学应用题已经就成了求解已经数学化之后的问题。这是完全不对的。通过面对和解决粗糙的问题你能够收获的东西，远远比面对已经数学化之后的问题要多很多很多。

有了这个理解,我希望读者们,你们可以回去把本书的例题和作业题重新看一遍,其中哪些其实是需要你们自己来做数学化这一步的。例如,家庭成员的问题其实是需要定义家庭成员这个集合的标准的,不同的家庭成员的比较其实是要联系到集合的交并补运算的,从穿裙子推断性别其实是要自己来决定需要哪些数据哪些变量的,敏感问题调查其实是一个相当具有挑战性和粗糙的问题尽管在本书中已经被数学化了,随机袜子凑一对的问题其实是要你自己想到是否袜子区分左右脚然后做出相应的假设的,练习“数、量、物、算式”四个层次的数学对象的问题其实是要你自己来走同从现实世界的对象到数学对象的,“开心不开心”的集合运算更强调从问题到数学语言,从信号推断真值本来是一个非常粗糙的问题但是在本书中也被初步数学化了。

当然,在本书的例题和习题中,还有一种类型的问题——概念理解型测试题:如果你理解这个题的场景相应的概念就能做出来,否则,必须先去学会相应的概念才能做出来。这样的问题是为了考察和促进你学习概念。这也是本书中另一种需要你体会的例题和作业题。有的时候,有的问题刚好同时扮演了数学建模和概念理解的角色。

为了给下一节数学建模做准备,也为了帮大家找到求解数学题的一般方法,我们在这里跟大家分享一个叫做 WHWM 的思维方式。对于求解数学题,WHWM 指的是问以下四个问题:

第零, Trans: 尽量把题中的每一句话都用数学表达式来表达。原题中是否有某个词,你需要先明确一下其对应的现象是什么,明确表述一下关于这个现象的事实性知识?

第一, What: 这个问题里面有哪几个对象。这些对象之中哪些是已知的,哪些是要求出来的。在应用题里面一个对象可以是一个物理对象,可以具体到一个物理对象的一个物理量,甚至进一步具体到一个对象的一个物理量的一个特定的取值。一般来说,在数学应用题中,我们选取中间层次——一个物理对象的一个物理量——来当作对象

第二, How: 这些对象之间有什么关系,这些关系用什么样的数学表达式来表达,运用这个关系如何从已知的对象得到要求出来的对象。

第三, Why: 为什么我们关注这些对象,为什么这些对象之间有这些关

24.3 从把数学知识用于描述现实看真正的数学应用题——WHWM求解数学应用题387

系(这时候,你往往可以追问:这些关系是基于什么样的物理规律,为什么这些对象服从这样的物理规律),为什么这些关系用这样的数学表达式来表达(这时候你往往需要走到数学概念)。

第四, Meaningful: 从这个问题的求解里面我收获了什么? 分成四个子问题子步骤。

- 第一步 Validation 验算, 验算一下问题是否已经解决。
- 第二步 Logic Summary 逻辑梳理, 上面的求解方式的逻辑总结。
- 第三步 Reflection 反思, 上面这个解决问题的逻辑正确吗, 还是有漏洞? 还有别的求解这个问题的方法吗?
- 第四步 Lesson Learned 总结提炼, 从这个问题的求解里面我学会了什么, 增加了对什么概念的理解, 增加了对什么分析方法和思维方式的理解?

在这里, 第零步是为了把 What 和 How 提炼的更准确做的准备。第一第二步就解决了问题。第三步是对为什么这样来解决问题做的追问, 走到学科概念。第四步是对整个解决问题的过程做的反思, 走到学科思维和一般性人类思维。只要这样来问和回答问题, 解题思路自动就能得到, 还能促进数学概念的学习, 促进数学思维的体会。因此, 这不仅仅是获得数学题答案的方法(可以直接给答案, 也可以给解题思路), 还是获得解题思路的方法。

我们来展示几个这样的例子。

例 24.3 (存款方式问题——WHWM 版本). 小明过年的时候收到了 3000 元的压岁钱。他想去银行存起来。银行有一个三年期的存款方案, 每年的利率是 $r = 0.02$ (或者说 2%)。算一下三年以后小明可以获得多少钱。

第零, Trans: 尽量把题中的每一句话都用数学表达式来表达。原题中是否有某个词, 你需要先明确一下其对应的现象是什么, 明确表述一下关于这个现象的事实性知识?

答: “小明过年的时候收到了 3000 元的压岁钱” 用数学语言表述为 $M_0 = 3000$ (元)。“算一下三年以后小明可以获得多少钱” 用数学语言表述为 $M_3 = ?$ 。“银行有一个三年期的存款方案, 每年的利率是 $r = 0.02$

(或者说 2%)”这里面有一个专用知识和专用词“每年的利率”，这表示 $M_{t+1} = M_t \times (1 + r)$ 。

第一, What: 这个问题里面有哪几个对象。这些对象之中哪些是已知的, 哪些是要求出来的。对象可以是事物、事物的量、事物的量对应的数、关系表达式、方程。

答: 这里有对象: 现在的钱 $M_0 = 3000$ (元) (已知), 三年以后的钱 M_3 (未知), 每年的利率 $r = 0.02$ (已知), 以及“每年的利率”的含义——下一年的钱和今年的钱的关系 $M_{t+1} = M_t \times (1 + r)$ (已知)。

第二, How: 这些对象之间有什么关系, 这些关系用什么样的数学表达式来表达, 运用这个关系如何从已知的对象得到要求出来的对象。

答: 运用 $M_{t+1} = M_t \times (1 + r)$ 可以得到 $M_1 = M_0 \times (1 + r)$, $M_2 = M_1 \times (1 + r)$, $M_3 = M_2 \times (1 + r)$, 合起来得到 $M_3 = M_0 \times (1 + r)^3 = 3183.6$ 元。

第三, Why: 为什么我们关注这些对象, 为什么这些对象之间有这些关系, 为什么这些关系用这样的数学表达式来表达。

答: 需要从现在的钱算出来存了三年以后的钱, 因此就需要每隔一年的钱之间的关系, 才能直到从第零年到第三年之间的关系, 而每一年的关系已经由年利率的值和含义给出——每一年的钱会增加一定的比例。按照比例增加, 表示用乘法来计算, 多次重复, 因此需要连续多次乘以同一个比例数, 变成乘方。

第四, Meaningful: 从这个问题的求解里面我收获了什么? 分成四个子问题子步骤。

- 第一步 *Validation* 验算, 验算一下问题是否已经解决。
- 第二步 *Logic Summary* 逻辑梳理, 上面的求解方式的逻辑总结。
- 第三步 *Reflection* 反思, 上面这个解决问题的逻辑正确吗, 还是有漏洞? 还有别的求解这个问题的方法吗?

24.3 从把数学知识用于描述现实看真正的数学应用题——WHWM求解数学应用题389

- 第四步 *Lesson Learned* 总结提炼：从这个问题的求解里面我学会了什么，增加了对什么概念的理解，增加了对什么分析方法和思维方式的理解？

答： 以下是我的反思总结提炼收获，

- 第一步 *Validation* 验算：这里的验算没什么好做的，可以每一步算一个结果而不是整体合起来算一次，做个比较。不过这个验算仅仅是对计算的验算，不是对解题思路和算式的验算，不值得做。
- 第二步 *Logic Summary* 逻辑梳理：这道题从年利率的含义出发得到每隔一年的钱之间的关系，然后重复运用这个关系得到相隔三年的钱之间的关系，最后代入具体数作计算。
- 第三步 *Reflection* 反思：这个问题的逻辑正确，完全基于年利率的含义，没有漏洞，也没有别的求解方法。
- 第四步 *Lesson Learned* 总结提炼：从这个问题中，我体会到了 WHWM 解题方法，以后可以尝试用这个方法来解决别数学题；尤其是在其中的第二步，我体会到了从关系的角度去思考和求解数学题；而在第三步，我还建立起来了问题场景和数学概念（比例、乘法、乘方）的联系从而明白了为什么有这个关系，而不仅仅是关注关系是什么和怎么算。

为了能够更清楚地体会到 WHWM 求解数学题的方法，我们用最简单的数学题再来展示一遍。我们来展示几个这样的例子。

例 24.4 (两次的钱合起来——WHWM 版本). 我已经有了两块钱，爸爸再给了我三块钱，请问我有多少钱？

第零, Trans: 尽量把题中的每一句话都用数学表达式来表达。原题中是否有某个词，你需要先明确一下其对应的现象是什么，明确表述一下关于这个现象的事实性知识？

答：“我已经有了两块钱”表示已有的钱的数量是两块，记作 $O = 2(\text{元})$ ；“爸爸再给了我三块钱”表示我新增加的钱的数量是三块，记作 $N = 3(\text{元})$ ；“我有多少钱”表示我总的钱的数量要求出来，记作 $T = ?(\text{元})$ 。这里没有我需要专门去明确表达一下的词和现象。

第一, What: 这个问题里面有哪几个对象。这些对象之中哪些是已知的, 哪些是要求出来的。对象可以是事物、事物的量、事物的量对应的数、关系表达式、方程。

答: 这里有我之前已经有的钱, 具体到数量就是已有钱的数量 (2 元, 已知, 可以记作 $O = 2(\text{元})$); 这里有爸爸新给我的钱, 具体到数量就是新给的钱的数量 (3 元, 已知, 可以记作 $N = 3(\text{元})$); 这里有我总共有的钱, 具体到数量就是总共的钱的数量 (未知, 要求出来, 可以记作 $T = ?(\text{元})$)。

第二, How: 这些对象之间有什么关系, 这些关系用什么样的数学表达式来表达, 运用这个关系如何从已知的对象得到要求出来的对象。

答: 之前已有的钱的数量 (O) 加上新给的钱的数量 (N) 是总的钱的数量 (T), 数学上, 用加法来表示, 也就是: 已有的钱的数量 + 新给的钱的数量 = 总的钱的数量。结合这题目的信息也就是: $2(\text{元}) + 3(\text{元}) = 5(\text{元})$ 。

第三, Why: 为什么我们关注这些对象, 为什么这些对象之间有这些关系, 为什么这些关系用这样的数学表达式来表达。

答: 因为要求出来的是总的钱的数量, 这个量需要把已有的钱的数量和新给的钱的数量合起来, 所以需要这些对象。之所以用加法来表示是因为关系是“合起来数一数”。

第四, Meaningful: 从这个问题的求解里面我收获了什么? 分成四个子问题子步骤。

- 第一步 *Validation* 验算, 验算一下问题是否已经解决。
- 第二步 *Logic Summary* 逻辑梳理, 上面的求解方式的逻辑总结。
- 第三步 *Reflection* 反思, 上面这个解决问题的逻辑正确吗, 还是有漏洞? 还有别的求解这个问题的方法吗?
- 第四步 *Lesson Learned* 总结提炼, 从这个问题的求解里面我学会了什么, 增加了对什么概念的理解, 增加了对什么分析方法和思维方式的理解?

24.3 从把数学知识用于描述现实看真正的数学应用题——WHWM求解数学应用题391

答： 以下是我的反思总结提炼收获，

- 第一步 *Validation* 验算：得到了我总共有 5 元，减去一开始有的 2 元，得到 $5 - 2 = 3$ 元，也就是爸爸给了我 3 元。正确。问题解决了。
- 第二步 *Logic Summary* 逻辑梳理：要求出来总的钱的数量，根据总的钱的数量是已有的钱和新增加的钱合起来数一数，得到用加法。
- 第三步 *Reflection* 反思：这个逻辑完全正确，没有其他关系和解法。这个问题的 *Trans* 翻译这一步比较简单，和 *What* 基本上重复。不过，为了以后养成先 *Trans* 翻译的习惯，我们还是来完成这一步。
- 第四步 *Lesson Learned* 总结提炼：我更好地理解了加法的概念，也知道在解决实际问题中要把场景和概念结合，我还学会了用 *WHWM* 来求解数学问题。

例 24.5 (路程问题——WHWM 版本). 有两个人从总长度为 100 米的跑道上的两个端点相向而行。已知一个人的行走速度为 4 米每秒，另一个人的速度是 6 米每秒。请问前一个人需要多长时间走完全程从一个端点到达另一个端点？

第零，Trans： 尽量把题中的每一句话都用数学表达式来表达。原题中是否有某个词，你需要先明确一下其对应的现象是什么，明确表述一下关于这个现象的事实性知识？

答：“有两个人从总长度为 100 米的跑道上的两个端点相向而行”、“走完全程从一个端点到达另一个端点”表示这个人要走的总路程是 $L = 100$ (米)， “一个人的行走速度为 4 米每秒，另一个人的速度是 6 米每秒”表示这个人的行走速度是 $v_1 = 4$ (米/秒)， $v_2 = 6$ (米/秒)， “前一个人需要多长时间”说明我们需要求出来前一个人走完总长度所需要的时间 $t_1 = ?$ (秒)。这里的现象是“相向而行的两个人”，其含义就是两个人向着彼此越来越接近的方向前行。

第一, What: 这个问题里面有哪几个对象。这些对象之中哪些是已知的, 哪些是要求出来的。对象可以是事物、事物的量、事物的量对应的数、关系表达式、方程。

答: 这里有跑道的长度 (已知, $L = 100$ (米)), 两个人行走的速度 (已知, $v_1 = 4$ (米/秒), $v_2 = 6$ (米/秒)), 前一个人走完全程需要的时间 (未知, 要求出来, 记作 $t_1 = ?$ (秒))。

第二, How: 这些对象之间有什么关系, 这些关系用什么样的数学表达式来表达, 运用这个关系如何从已知的对象得到要求出来的对象。

答: 时间 = 距离 $\div$ 速度, 用算式表示就是 $t_1 = L \div v_1 = 100(\text{米}) \div 4(\text{米/秒}) = 25(\text{秒})$ 。

第三, Why: 为什么我们关注这些对象, 为什么这些对象之间有这些关系, 为什么这些关系用这样的数学表达式来表达。

答: 要算出来走完全程的时间, 我们得去看这个时间和什么有关系。首先, 因为要走完全程, 因此, 总的路程就是跑道的长度。其次, 这个人每过一个单位时间就会走其速度这么大的距离, 总的路程就会减少这么大的距离, 重复多次的减法用除法表示, 重复的次数就是总的路程被减完的时间。这关系就是数学基本关系“除法”所代表的关系。同时我们看到, 我们不需要关注第二个人的信息。因为这个问题问的就是第一个人走完全程的时间。

第四, Meaningful: 从这个问题的求解里面我收获了什么? 分成四个子问题子步骤。

- 第一步 *Validation* 验算, 验算一下问题是否已经解决。
- 第二步 *Logic Summary* 逻辑梳理, 上面的求解方式的逻辑总结。
- 第三步 *Reflection* 反思, 上面这个解决问题的逻辑正确吗, 还是有漏洞? 还有别的求解这个问题的方法吗?
- 第四步 *Lesson Learned* 总结提炼, 从这个问题的求解里面我学会了什么, 增加了对什么概念的理解, 增加了对什么分析方法和思维方式的理解?

24.3 从把数学知识用于描述现实看真正的数学应用题——WHWM求解数学应用题393

答：以下是我的反思总结提炼收获，

- 第一步 *Validation* 验算：得到了需要 $t_1 = 25$ (秒) 走完全程，其速度是 $v_1 = 4$ (米/秒)，这段时间内其行走的路程为 $L = v_1 \times t_1 = 100$ (米)，正好和题目信息相符。没错。问题解决了。
- 第二步 *Logic Summary* 逻辑梳理：我们识别出来第一个人的路程和速度，然后运用路程问题的基本公式 $t = L \div v$ 就得到了其走完全程所需要的时间。
- 第三步 *Reflection* 反思：我们已经注意到了因为路程问题的背后是重复的加法（每个单位时间行走一样的路程）或者说重复的减法（每个单位的时间所需要完成的剩余路程减少相同的量），因此，在这里，我们直接就用了除法。这个逻辑完全正确，没有其他关系和解法。注意，我们已经把重复的减法和除法认同。但是，这里其实隐藏了一个容易出错的地方：我们给了两个相向而行的人的速度和总路程，可能有的人就会直接套用相遇问题的公式 $t = L \div (v_1 + v_2)$ ，得到 10(秒)。这说明，在解决问题的时候一定要建立场景、对象、关系、数学表达式之间的内在关联，不能直接套用公式。
- 第四步 *Lesson Learned* 总结提炼：我更好地理解了除法的概念（重复的减法），也知道在解决实际问题中要把场景和概念结合，我还学会了用 WHWM 来求解数学问题。此外，通过求解这道题我还发现，求解数学题的时候不能直接套用公式，要搞清楚场景、对象、关系、数学表达式之间的内在关联。

例 24.6 (平行四边形判定——WHWM 版本). 证明：一组对边平行一组对角（非相邻）相等的四边形是平行四边形。

第零，Trans： 尽量把题中的每一句话都用数学表达式来表达。原题中是否有某个几何对象，你需要先明确写出来这个对象相应的公理、定义和定理，为后续的解题做准备。

答： 四边形 $ABCD$ （顶点依次逆时针编号，见图 24.1），“一组对边平行”翻译为 $AB \parallel CD$ ，“一组对角（非相邻）相等”翻译为

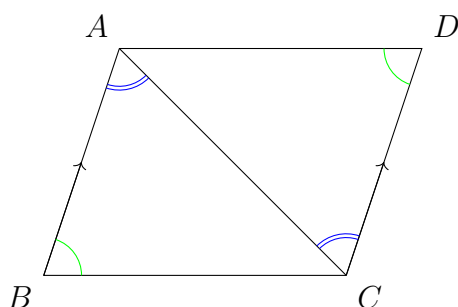


图 24.1: 平行四边形判定——WHWM。辅助例 24.6 中证明的图: 四边形 $ABCD$ 中有 $AB \parallel CD$, $\angle ABC = \angle CDA$, 需要证明 $ABCD$ 是平行四边形。

$\angle ABC = \angle CDA$, 待证明 $ABCD$ 是平行四边形, 这时候只需要补充 $BC \parallel AD$, 即另一组对边也平行。能需用到的初中几何概念和定理有: 平行线的判定和性质、平行四边形的定义和判定。

第一, What: 这个问题里面有哪几个对象。这些对象之中哪些是已知的, 哪些是要求出来的。对象可以是事物、事物的量、事物的量对应的数、关系表达式、方程。

答: 这里有以下对象: 四边形 $ABCD$ (已知), 也就有四边形的每条边 AB, BC, CD, DA 和每个角 $\angle DAB, \angle ABC, \angle BCD, \angle CDA$, 一组平行线 $AB \parallel CD$ (已知), 一组对角相等 $\angle ABC = \angle CDA$ (已知), $ABCD$ 是平行四边形或者说另一组对边平行 $BC \parallel AD$ (待证明)。

第二, How: 这些对象之间有什么关系, 这些关系用什么样的数学表达式来表达, 运用这个关系如何从已知的对象得到要求出来的对象。

答: 连接 AC , 从 $AB \parallel CD$ 得到内错角相等, 然后考虑三角形 $\triangle ABC$ 和 $\triangle ADC$, 有两组角对应相等, 一条公共边, 因此全等, 就能得到 $AB = CD$, 满足平行四边形定义。写成三段论的形式为: 连接 AC , $\because AB \parallel CD, \therefore \angle BAC = \angle DCA$ 。在三角形 $\triangle ABC$ 和 $\triangle ADC$ 中, 有 $\angle BAC = \angle DCA, \angle ABC = \angle CDA, AC = CA, \therefore \triangle ABC \cong \triangle ADC$ (全等三角形判定 AAS), $\therefore AB = CD$ (全等三角形性质), 因此, 四边形 $ABCD$ 是平行四边形 (平行四边形判定)。

第三, Why: 为什么我们关注这些对象, 为什么这些对象之间有这些关系, 为什么这些关系用这样的数学表达式来表达。

答: 我们要证明平行四边形, 发现已有一组对边平行, 则我们只需要证明还有另一组对边平行, 或者之前那组对边相等。我们尝试选择后者。从一组对边平行我们联系到一组内错角相等, 接着联系到另一个已知条件——另一组对角相等, 以及一条公共边, 就得到启发去证明 $\triangle ABC \cong \triangle ADC$, 然后得到 $AB = CD$, 也就是平行的那组对边相等。于是, 已知和待证明就完全联系起来了。这里的平行、相等、全等的数学表达式都是其标准表达。

第四, Meaningful: 从这个问题的求解里面我收获了什么? 分成四个子问题子步骤。

- 第一步 *Validation* 验算, 验算一下问题是否已经解决。
- 第二步 *Logic Summary* 逻辑梳理, 上面的求解方式的逻辑总结。
- 第三步 *Reflection* 反思, 上面这个解决问题的逻辑正确吗, 还是有漏洞? 还有别的求解这个问题的方法吗?
- 第四步 *Lesson Learned* 总结提炼, 从这个问题的求解里面我学会了什么, 增加了对什么概念的理解, 增加了对什么分析方法和思维方式的理解?

答: 以下是我的反思总结提炼收获,

- 第一步 *Validation* 验算: 对于证明题部分, 我们发现这个论证的起点成立和论证过程的每一步是否都严格, 基于已有的定义和命题, 采用逻辑演绎证明。验算通过。
- 第二步 *Logic Summary* 逻辑梳理: 在本题中, 我们先通过平行线得到内错角相等, 然后通过另一组对角相等和公共边, 得到两个三角形全等, 接着得到平行的对边相等, 得到待证明命题。
- 第三步 *Reflection* 反思: 上面的思路清晰, 逻辑严密。可能还有其他的证明和求解方法。

- 第四步 *Lesson Learned* 总结提炼：通过求解这道题，更好地理解了解了平行线的性质、平行四边形的判定、全等三角形的判定和性质；更好地体会到了从已知条件和待证明结论、待求出结果这样两端往中间凑，直到联系上的思考方式，并且注意到把这些信息标注在几何图上有助于启发思考；深刻体会了 WHWM 法有助于剖析（看到对象、对象之间的关系）、校验、提炼和反思整个问题解决链条，以后我自己也可以用 WHWM 来求解数学题，做完题之后也要来这样做反思总结。

WHWM 数学解题方法不仅仅是为数学建模准备的尽管它特别适合用于数学建模，还是一个求解普通数学题的方法。应用题和几何题都可以用这个方法来获得解题思路，进而解题。未来你甚至会发现，物理题、化学题都可以使用。

24.4 数学建模五步和 WHWM 四问及其应用

我们把下面这个

1. 从现象提出问题，
2. 把问题转化为数学问题：搞清楚目标和主要变量以及它们的数学描述，
3. 求解数学问题，
4. 通过实验或者实证来检验答案，
5. 对求解方法、问题、概念、模型做总结形成系统化的理论

的过程称为数学建模五步，或者简称数学五步。其中求解数学问题的那一步，可以使用 WHWM 方法来完成。

上两节中，我们已经从数学发展和本书中对数学知识的呈现方式、应用题以及粗糙版的应用题感受到了数学建模。尤其是其中的例 24.2 中的从粗糙的数学题或者说场景问题到数学应用题的转化的过程，可以帮助我们感受到我们是如何把一个问题数学化的，也就是搞清楚目标和主要变量，以

及它们的数学描述。至于求解数学题的过程，倒是相对简单，找出来目标和主要变量之间的关系用数学表达式来表达，也就完成了。

下面，我们再来通过一个例子来感受一下这个数学五步。我们选择的例子是概念诊断，或者说高效率的考试方式。通常的考试可以看作是对概念的随机抽样：从一个待考查的概念集合中随机或者有意识地抽取一个概念，来让被试尝试完成关于这个概念的一个问题，如果给出正确回答则认为被试掌握了这个概念，否则则认为没掌握，然后就再一次抽取下一个概念。这样的话，如果我们想给被试做一个比较全面的诊断，例如想知道这个被试在所有的概念上是否掌握，我们只能把所有的概念都测试一遍了。我们想在此基础上有一个改进：第一，我们不想所有的概念都测试一遍就想知道一个被试在所有或者绝大多数概念上是否掌握；第二，我们希望下一个要考察的概念是根据已经考察过的概念和被试的回答情况做的某种最优选择，而不是随机抽样，或者某种主观性的衡量。顺便，一旦我们实现了第一点，评价性测量的问题就解决了，因为我们解决的是覆盖绝大多数概念的诊断性测量。因此，如果可以达到只用测一个概念集合里面的 10% 就可以知道这个集合里面的绝大多数概念的掌握情况，那我们就可以大大提升测试的效率了。从评价性测量到诊断性测量，测试的效果，尤其对于被试的意义，就会得到很大的提升。

注意，关于这个问题我们是有一定的经验和体验的：我们知道如果检测汉字“木”（或者“直角”这个数学概念）发现这个被试不认识，则很可能其也不认识“林、森”（相应地“直角三角形”）；反过来如果检测汉字“森”（或者“直角三角形”这个数学概念）发现这个被试认识，则可能其也认识“林、木”（相应地“直角，三角形”）。这个经验体验启发了我们这个目标是有可能可以达成的，也启发了我们大概可以如何达成——通过检测少数和基于这些少数的统计推断。

好了，问题场景和模糊的目标我们都有了，经验体验和有了。我们完成了数学五步的第一步。注意，对于完成第一步来说，对问题场景的经验体验是非常重要的。以后我们会看到对于第二步也是如此。下面，怎么来解决这个问题？我们先来数学化，搞清楚目标和主要变量以及它们的数学描述。

前面用语言交代了目标，我们来看看这个目标能不能变成数学表达式。我们的目标是：对于给定的 N 个概念，尽可能用最少的 M 个概念就能

够得到一个被试在所有或者绝大多数概念上的掌握状态。被试对于一个概念的掌握状态用什么数学结构描述呢？可能是掌握，可能是不掌握，也可能是不完全掌握。我们用随机变量，对于一个概念 C_n ，用户的掌握状态 $L_n = 1, 0$ 有两个状态值，分别表示完全掌握和完全没掌握。然后用一个概率值 $P_n \in [0, 1]$ 来表示其掌握这个概念的概率。只要我们设计测试题的时候把一个概念的更基础的概念明确分解出来，然后做到对每一个更基础的概念以及这个概念本身的检测，则做到就用 $L_n = 1, 0$ 两个状态来表示是否掌握是没问题的。当信息不足的时候，那就用一个随机数来衡量掌握程度。

这样，我们就可以把目标写成，

$$\begin{aligned} & \min M, \\ & \text{s.t. } K = \sum_n |2P_n - 1| \geq N - \epsilon. \end{aligned} \quad (24.5)$$

其中 ϵ 是一个给定的比较小的数。当 $\epsilon = 0$ 的时候， $K = N$ 表示每一个概念，对于给定的被试必须做到完全已知其掌握状态，也就是 $P_n = 1, 0$ 。因此，在这个极限下很可能我们需要把每一个概念都测一遍。也就是 M 很可能得等于 N （除非一会儿遇到我们将要看到的条件概率的特殊情况。不过，注意，目前我们还不知道这个目标如何实现，因此，我们假装还不知道即将要用到条件概率）。这里的“s.t.”符号表示“以至于”、“达到”。这是我们整个检测方案的约束，也就是必须满足的一个条件。

那我们如何来实现这个目标呢？第一，我们来选择合适的检测序列 $\vec{C}_M = (C_1, C_2, \dots, C_M)$ 来得到这些概念上的检测结果 $L_{\vec{C}_M} = \{L_{C_1}, L_{C_2}, \dots, L_{C_M}\}$ ；第二，我们可能得想办法从 $\vec{C}_M$ 来得到其他没有检测过的概念的掌握状态， $\bar{L}_{\vec{C}_M}$ ，因为我们的目标里面包含了一个必须对所有的概念的掌握程度都有足够的信息的约束。

这样我们的问题就数学化了：通过选择 M 尽可能小的 $\vec{C}_M$ （也就是我们要找到确定这样的检测序列的算法），从这个检测直接的结果 $L_{\vec{C}_M}$ 得到未检测概念的掌握情况 $\bar{L}_{\vec{C}_M}$ ，从而来使得 K 满足要求 $K \geq N - \epsilon$ 。

你看，完成这个第二步的过程中，你必须对检测、概念掌握状态等等具有非常好的经验体验。

这里有几个数学问题：如何从检测直接的结果 $L_{\vec{C}_M}$ 得到未检测概念的掌握情况 $\bar{L}_{\vec{C}_M}$ 保证把检测过的信息尽可能地用好发挥其最大作用（这样 M

才会尽可能地小), 如何确定最高效率的 $\vec{C}_M$ (这样才能使得 M 尽可能小)。这两个问题都是明确的数学问题。

我们先来解决第一个试试。这是数学建模的第三步。这一步可以使用求解数学题的 WHWM 方法。这个问题是要从一部分对象的已知信息来得到另一部分对象的位置信息。按照我们之前学习过的概率论和统计学, 这显然是一个统计推断问题。和“从看到穿裙子来推断性别”的问题很象。于是, 我们发现, 如果这些概念之间的掌握状态之间有依赖关系, 例如之前提到过的“木、林、森”和“直角、三角形、直角三角形”这样的, 那确实可以做推断。为了做这个推断, 我们需要什么呢? 需要各个条件概率, 例如 $P(L_{\text{木}} = 1 | L_{\text{森}} = 1)$, $P(L_{\text{森}} = 0 | L_{\text{木}} = 0)$, $P(L_{\text{直角}} = 1 | L_{\text{直角三角形}} = 1)$, $P(L_{\text{直角三角形}} = 0 | L_{\text{直角}} = 0)$ 。有了这些条件概率, 已知一个被试真实数据, 例如 $L_{\text{直角三角形}} = 1$, 就可以运用贝叶斯公式和全概率公式从这个已知信息来计算各个概率了。这其实就是例 21.2 中展示的计算。我们把上面的思考过程按照 WHWM 方法做了一个整理。

例 24.7 (掌握状态的推断——WHWM 版本). 对于给定的一个被试, 在一群概念里面, 某些概念已经被直接检测过其掌握状态已知, 能否得到其他未检测概念的掌握情况。

第零, Trans: 尽量把题中的每一句话都用数学表达式来表达。原题中是否有某个词, 你需要先明确一下其对应的现象是什么, 明确表述一下关于这个现象的事实性知识?

答: “在一群概念里面”被翻译成总的包含 N 个元素的概念集合 C 已知。“某些概念已经被直接检测过其掌握状态已知”被翻译成数学语言对于 $\vec{C}_M = (C_1, C_2, \dots, C_M)$, $L_{\vec{C}_M} = \{L_{C_1}, L_{C_2}, \dots, L_{C_M}\}$ 已知。“能否得到其他未检测概念的掌握情况”表示对于补集 $\bar{\vec{C}}_M$ 中的元素的掌握状态 $\bar{L}_{\bar{\vec{C}}_M}$ 需要计算出来。

第一, What: 这个问题里面有哪几个对象。这些对象之中哪些是已知的, 哪些是要求出来的。对象可以是事物、事物的量、事物的量对应的数、关系表达式、方程。

答: 已经被检测的概念集合 $\vec{C}_M = (C_1, C_2, \dots, C_M)$ 及其是否被被试掌握

的状态 $L_{\vec{C}_M} = \{L_{C_1}, L_{C_2}, \dots, L_{C_M}\}$ 已知, 其补集 $\bar{\vec{C}}_M$ 中的元素的掌握状态 $\bar{L}_{\vec{C}_M}$ 未知, 需要计算出来。

第二, How: 这些对象之间有什么关系, 这些关系用什么样的数学表达式来表达, 运用这个关系如何从已知的对象得到要求出来的对象。

答: 对于有联系的概念, 如果已知条件概率, 运用贝叶斯公式和全概率公式, 可以从以检测概念的掌握情况的概率分布得到未检测概念的掌握情况的概率分布。

第三, Why: 为什么我们关注这些对象, 为什么这些对象之间有这些关系, 为什么这些关系用这样的数学表达式来表达。

答: 前面的数学建模问题需要求解数学题这一步, 而这一步需要从已检测概念的掌握情况推断出来未检测概念的掌握情况, 因此我们关注这些对象。这些对象之间的关系刚好符合条件概率关系, 所以我们用贝叶斯公式和全概率公式来做这个推断。

第四, Meaningful: 从这个问题的求解里面我收获了什么? 分成四个子问题子步骤。

- 第一步 *Validation* 验算, 验算一下问题是否已经解决。
- 第二步 *Logic Summary* 逻辑梳理, 上面的求解方式的逻辑总结。
- 第三步 *Reflection* 反思, 上面这个解决问题的逻辑正确吗, 还是有漏洞? 还有别的求解这个问题的方法吗?
- 第四步 *Lesson Learned* 总结提炼, 从这个问题的求解里面我学会了什么, 增加了对什么概念的理解, 增加了对什么分析方法和思维方式的理解?

答: 以下是我的反思总结提炼收获,

- 第一步 *Validation* 验算: 真正的验算部分留到数学建模第四步——实验检验再来完成。不过, 从逻辑上, 这个问题确实就是基于条件概率的推断问题, 数学结构的选择合理。

- 第二步 *Logic Summary* 逻辑梳理：我们发现已知的信息可以看作概率分布，未知的要算出来的信息也可以看作概率分布，一旦有了条件概率，我们就可以从已知的概率分布算出来未知的概率分布。逻辑合理。
- 第三步 *Reflection* 反思：这里我们从已知和未知开始，把两者联系起来，发现需要的数学学科知识是条件概率、贝叶斯公式和全概率公式。但是，实际上，这里有一个还不明确的问题——原题中没有给出是否可以获得所有必要的条件概率。这是和通常的数学解题不同的地方。在数学建模中，需要什么信息也是需要我们来思考的问题，而通常的数学解题中，给的信息一般是充分的够用的。
- 第四步 *Lesson Learned* 总结提炼：我们通过在实际问题中要把场景和概念结合进一步理解了贝叶斯公式、全概率公式和条件概率，我还学会了用 WHWM 来求解数学问题。

在例 21.2 中，我们把这个问题当作一个关于条件概率的数学应用题来感受了一下：已知描述这些对象的数学结构是什么——随机变量，解决这个问题的数学结构是什么——一群相互联系在一起的条件概率或者说贝叶斯网络，怎么计算——运用贝叶斯公式和全概率公式。只不过我们今天，其实是一开始不知道这个数学问题，通过经验体验来把问题数学化——目标和主要变量数学化，然后才发现，原来求解模型这部分可以依靠例 21.2 中这个样子的贝叶斯网络。

有了这个模型，至少等我们给定一个检测顺序 $\vec{C}_M$ 和得到相应的测量结果之后我们就可以得到所有的 P_n 的值，也就能得到 K 的值。这样就可以检验是否满足要求的约束 $K \geq N - \epsilon$ 了。这里其实还有一个数学计算问题：对于只有少数概念需要检测的贝叶斯网络，我们用贝叶斯公式和全概率公式手算就可以，对于非常复杂的网络，我们还得确定一个最优的计算顺序来把未知的概率从已知的已经检测过的概念的信息得到。这个问题也不是很简单，甚至是不是所有的贝叶斯网络针对所有的已检测概念的选择这个计算都可以完成也是一个问题。这个问题后来发展成了一门专门的数学或者说统计学分支学科——贝叶斯网络分析或者随机图理论 [10]。现在，假设我们意识到我们需要这个学科的专门知识以后，我们也就学会了它。或

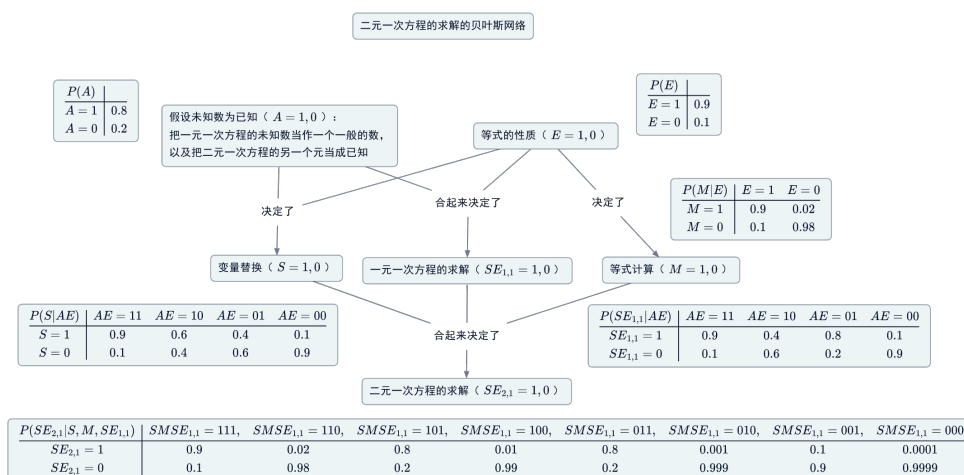


图 24.2: 贝叶斯网络用于诊断性检测。为了方便阅读, 我们在这里复制了图 21.6。已知条件概率和被检测知识点的掌握状态, 我们可以计算出来所有知识点掌握的概率。

者, 其实, 如果还没有这样的学科的知识, 你就自己通过受解决这个问题的启发来自己创造和发展它。

于是, 就剩下一个问题, 能不能设计一个方式来找到足够好的 $\vec{C}_M$? 当然, 实在没有这样的办法的时候, 就把所有可能的 $\vec{C}_M$ 对于给定的 M 都做一遍, 然后增加 M 直到找到那个满足约束的最小的 M 。也就是说, 穷举法可以完成我们希望完成的任务。

不过, 如果真的使用穷举法就麻烦了。当你有 N 个概念的时候, 对于固定的 M , 你得尝试 C_N^M 个可能的检测顺序, 考虑到你还得变化很多次 M , 你的尝试总次数会接近 2^N 。这个数会随着 N 急剧增加。那么, 能不能找到效率更高的来找到合适的 $\vec{C}_M$ 的办法呢? 这又是另一个数学问题。

比如说, 你可以尝试“贪心算法”——一个通常不算差但是不一定能保证最优的方法: 每一次尝试, 你随便从剩下的每检测的点当中选一个点, 计算出来如果检测这个点加上基于这个点的推断以后可以增加的 K 值, 然后选择那个增加 K 值最多的——因此也是最贪心的——那个点来当作实际检测点。这样最多需要尝试 $\frac{N(N+1)}{2}$ 次。这比 2^N 小多了。当然, 如果能找到其他方法来确定这个最优顺序就更好了。毕竟每次只往前探测一步不一定能找到最优解。以找到山谷的最低点为例, 有的山谷从某个地方开始找的

话，需要先翻过一个山峰才能到。贪心算法会使得这样的山谷不能被找到。目前这个问题也有相关的研究，但是还没有形成成熟的知识体系。[增加选择最优检测顺序的文献。](#)

其实，还有一个问题要解决：我们是如何得到这些条件概率和这些概念之间的联系的呢？我们目前的概念之间的依赖关系是通过专家——这里就是我这个本书作者——来手动基于专家对学科的认知绘制的。这样的概念之间的依赖关系肯定正确吗？还真不一定。比如说，假设出现了教材经常把“金、木、水、火、土”这几个字放在一起学习，那么就算“木、金”之间没有任何直接字形字义和字音上的联系，但是因为教材上经常放在一起学，也会出现关联性。我们的专家概念网络就忽略了这样的外在联系。因此，我们这里做了简化假设，概念之间的联系必然是概念之间的内在联系。在数学概念上，这个假设会更加合理一些。

不过，我们总得想个办法来从数据获得这个网络的结构和所有的条件概率的值。这个问题被成为贝叶斯网络的构建。这也是已经得到贝叶斯网络分析或者随机图理论 [10] 专门研究的问题。

假设这里的贝叶斯网络上的计算问题、贝叶斯网络的构建问题、检测顺序的优化问题都解决了，那我们就能得到一个满足约束的 M 值最小的检测顺序了。下面，就剩下实验检验的问题了。这是数学建模的第四步。

如果实验检验通过，则我们还需要从这里面提炼出来学科概念知识、解决问题的知识、思维层面的知识。这就是数学建模的第五步。比如说，假设之前这些具体数学问题的知识都是通过我们解决这个问题创造出来的，则我们就得到了后来被叫做贝叶斯网络分析的学科的很多概念：贝叶斯网络、贝叶斯网络的构建算法、贝叶斯网络上的推断算法。同时，我们所得到的检测顺序的优化问题的解决方式可以直接推广到其他的贝叶斯网络的状态检测上面去。我们得到这个检测顺序的优化算法背后没准也有一些可以推广到其他问题里面的东西，比如说贪心算法，尽管其还不是最优的。将来找到更好的，没准背后也会有一个类似的值得推广的思想在。

好了，通过这个优化检测顺序的问题，我们基本上把数学建模五步都体验了一下，也体验了一下求解数学问题的 WHWM 方法。

在这里我们把 WHWM 方法当作数学建模中的求解数学题的那一步来完成。实际上，我们还可以在更一般的意义上使用 WHWM 方法，问

What: 我要解决什么问题;

How: 我要怎么解决这个问题;

Why: 我为什么要解决什么问题,为什么要这样解决;

Meaningful: 通过这样来解决这个问题,对我有什么意义。

这样,WHWM 其实是可以放到数学建模的具体步骤之前来思考的:我遇到的现象和问题是什么,为什么我决定用数学建模的方式来解决它,具体的数学建模又怎么做,建完模型之后管用吗,从这个建模过程中“我——这个建模的人,这个解决这个问题的人”收获了什么。

24.5 推荐阅读材料

24.6 作业

习题 24.1 (存款问题数学建模版本的 WHWM). 按照从应用题的普通解答到应用题的 WHWM 解答,做一个上述存款问题数学建模版本的 WHWM 版本。也就是每一个子步骤,提炼出来:有哪些对象(其中哪些已知,哪些未知),对象之间的关系如果描述,为什么关注这些对象,为什么它们之间的关系用这个数学表达式来描述,这样来解决问题我收获了什么。

习题 24.2 (银行所谓的“报价利率”). 去差查资料,或者去银行问问,搞清楚报价年化利率、有效年化利率和计息期利率之间的差别。想想为什么会有“报价年化利率”这个东西。

习题 24.3 (路程问题 WHWM 版). 用 WHWM 方法回答以下问题:有两个人从总长度为 100 米的跑道上的两个端点相向而行。已知一个人的行走速度为 4 米每秒,另一个人的速度是 6 米每秒。请问两人在什么地方相遇?

习题 24.4 (重复相加问题 WHWM 版). 用 WHWM 方法回答以下问题:我要买三瓶矿泉水,每一瓶矿泉水的价格是两块,请问我需要花多少钱?

习题 24.5 (年龄问题 WHWM 版). 用 WHWM 方法回答以下问题:我今年 3 岁,我姐姐的年龄是我的年龄的两倍。请问我 5 岁的时候我的姐姐几岁?

习题 24.6 (抽烟和健康问题). 据说抽烟的人比不抽烟的人更健康, 尤其在肺部疾病上。你看看能不能找到一组数据来检验一下这个结论 (你可以查一查相关的研究论文里面的数据)。在这里, 你需要考虑用哪些指标来描述健康。你可能还得考虑如何描述抽烟的多少。

习题 24.7 (抽烟和健康问题). 据说有运动习惯的人比没有运动习惯的人更健康。你看看能不能找到一组数据来检验一下这个结论 (你可以查一查相关的研究论文里面的数据)。在这里, 你需要考虑用哪些指标来描述健康。你可能还得考虑如何描述运动的多少。

24.7 本章和本卷小结

在本章中, 我们学习了什么是数学建模, 以及能够帮助数学建模的方法——数学建模 WHWM 五步、WHWM 数学解题方法以及一般性的 WHWM 思考方式。我们希望你看到数学建模在数学知识的发展中的重要性之后, 还能初步学会运用数学建模 WHWM 五步、WHWM 方法来构建数学模型。当然, 也顺便初步学会用 WHWM 方法来求解数学题。

数学的发展来源于现实世界的启发、数学家的想象力以及数学研究者对数学知识结构本身的逻辑严密性的追求。其中, 在现实启发方面, 主要通过双向的数学建模来完成, 通过把数学结构用于描述现实来解决现实问题, 以及通过把数学结构用来描述现实启发提出新的数学结构。因此, 数学建模绝对不是应用数学结构来描述世界而已。甚至可以反过来, 认为数学学科概念知识的发展就是为了给数学建模也就是描述世界提供语言的, 以及这个过程中也顺便给人类思考提供一下语言因为描述世界不得不用到人类思考。当然, 有的人会把这两条的关系倒过来, 认为数学是思维的语言, 顺便因为思维可以用于描述世界所以才是描述世界的语言。

数学家的想象力有可能也是受到现实世界的启发的, 例如为了描述物理现象而提出的 δ 函数和微积分。当然也可能是受现有数学知识体系的启发, 例如对平面几何平行公理的替换而提出的其他形式的平行公理。当然, 也可能就是偶然得之。

对数学知识结构本身的逻辑严密性的追求也能促进数学知识的创造。一方面, 证明命题运用的是逻辑。另一方面, 也可以从考察现有数学知识

结构本身的逻辑严密性来提出新的概念和命题。例如，极限的概念的提出、实数的概念的完善、自然数的公理化定义都属于这样的例子。

本书特别强调数学知识的系统性，也就是从某一个逻辑起点开始——这个起点可能是受现实启发从现实中抽象出来的或者靠数学家的想象力提出来的——通过严密的逻辑来构建数学的知识体系。这样只要人们掌握了这个起点和从这个起点发展到其他知识的方法，注意这里的方法不仅仅有命题已经有了之后的证明命题所用的逻辑演绎证明还有其他的思维方式和分析方法（见本书前言中关于“数学学科大图景的”部分），那么人们就可以掌握这些知识，甚至进一步创造知识了。这是很好的事情。

但是，问题来了：能够找出来一组公理，然后运用逻辑演绎证明可以得到现有的所有的数学学科概念知识——称为完备性，而且公理之间是相容的没有矛盾（在一套体系里面成立，在另一套体系里面不成立）的吗？David Hilbert（大卫·希尔伯特）和Bertrand Russell（伯特兰·罗素）就企图来构建一个这样的数学知识的体系。尤其Russell的研究使得这个问题本身也成了—一个可以通过数学来研究的问题。直到数学发展遇到了Kurt Gödel（库尔特·哥德尔）证明的“哥德尔不完全性定理”——相容性和完备性不可能同时满足³。

既然统一的基于公理的知识体系已经不可能了，反过来，也就是说至少存在着一系列的不统一的公理体系，每个公理体系得管着自己的事情，有着自己不能证明的命题。这个时候就是现实和想象力进入一个知识体系的时候了：选择关注哪些现实，基于这些现实抽象出来哪些概念，猜出来哪些命题，进而反向提炼出来哪些公理。这些就是相应的那个现实世界的研究者和数学家可以一起努力的事情了。通过测量总结归纳来受到现实的启发，在这个启发下通过对现实的抽象来提炼概念和命题——这称为经验主义，通过逻辑演绎证明来反向猜测公理和在已有公理的情况下得到更多的命题——这称为逻辑主义或者理性主义，这些合起来才能推动一个学科的发展，才能解决描述相应的现实的需求。科学从来都是在经验主义和逻辑主义的结合下发展的。数学也是。

顺便，这时候，你就再一次看到了双向数学建模的重要性。数学建模不

³这个命题还存在着一个前提，这个公理体系能够定义自然数。不过，这个细节等到你以后学习到哥德尔不完全性定理的时候再说。

仅仅是运用数学知识解决现实世界中的问题，还是推动数学结构的发现的力量。数学建模五步和其中用到的求解数学题 WHWM 四问是帮助数学建模和数学问题求解的重要工具。

数学为思维和描述世界提供语言，数学依靠对现实世界做抽象看到相似性以及以及已有的数学结构做进一步的抽象看到相似性的相似性得到概念和命题，也依靠现实的启发和数学家的想象力来作出概念和命题的猜测，依靠逻辑演绎证明来确认数学命题和构建数学的知识体系。

学习数学的目标是帮助你体会到这些数学学科的大图景：数学学科的典型责任、典型研究对象、典型研究问题、典型思维方式和典型分析方法。请你从这个角度，在掌握了前面学习到的数学学科知识之后，再一次去阅读本书，但是这一次主要关注数学的学科大图景。

第二部分

高中篇

第二十五章 函数之幂函数、指数函数、对数函数

25.1 幂函数

25.2 指数函数

25.3 对数函数

25.4 函数和反函数的图形和表达式

25.5 推荐阅读材料

25.6 作业

25.7 本章小结

第二十六章 函数之半圆、半椭圆 函数

26.1 半圆函数

26.2 半椭圆函数

26.3 圆

26.4 椭圆

26.5 双曲线

26.6 推荐阅读材料

26.7 作业

26.8 本章小结

第二十七章 函数之三角函数

27.1 余弦函数

27.2 余弦定理

27.3 正弦余弦函数关系

27.4 正切反切函数

27.5 周期函数

27.6 推荐阅读材料

27.7 作业

27.8 本章小结

第二十八章 平面和空间矢量

28.1 直角坐标系和矢量的分量形式

正交归一基矢量

28.2 矢量加法和数乘

28.3 矢量内积

28.4 坐标变换和描述坐标变换的矩阵

28.5 矢量的抽象记号语言

28.6 矢量的应用

28.7 推荐阅读材料

28.8 作业

28.9 本章小结

第二十九章 矩阵和量子力学—— 数学建模

29.1 概率论的矩阵描述

29.2 量子现象

29.3 量子现象的经典波解释及其问题

29.4 量子现象的矩阵和矢量描述

29.5 解释更多的实验现象？

29.6 推荐阅读材料

29.7 作业

29.8 本章小结

第三十章 复数

30.1 负数的开平方

30.2 负数的开三次方

代数基本定理（多项式方程的解的个数和方程的阶数相同）

30.3 复数当作矢量使用

30.4 Euler 公式

复数的模和复角表示，Euler 公式。

30.5 复数的加法和乘法

30.6 复数的应用

30.7 推荐阅读材料

30.8 作业

30.9 本章小结

第三十一章 不等式初步

31.1 几个恒成立不等式

31.2 从恒成立不等式到函数极值

31.3 不等式和极值的应用

31.4 推荐阅读材料

31.5 作业

31.6 本章小结

第三十二章 导数和极值初步

32.1 导数的定义初步

极限和导数，曲线斜率

32.2 简单函数的导数

32.3 函数的导数用于求极值

32.4 导数和极值的应用

32.5 推荐阅读材料

32.6 作业

32.7 本章小结

第三十三章 离散变量的概率论和 统计学

33.1 从观测数据得到频率再到概率

独立样本，概率极值条件

33.2 从已知离散概率到观测数据的频率

大数定律

33.3 用 Monte Carlo 和 Bootstrap 配合数值计算来沟通频率和概率

33.4 概率论和统计学的应用

33.5 推荐阅读材料

33.6 作业

33.7 本章小结

第三十四章 立体几何

34.1 空间直线的平行和垂直

从三位直角坐标系到平行和垂直

34.2 空间平面的平行和垂直

34.3 推荐阅读材料

34.4 作业

34.5 本章小结

第三十五章 表示论：数学是对操作的研究

35.1 当作对象属性或者过程的数

35.2 数当作加法的表示

35.3 数当作乘法的表示

35.4 数和矩阵当作旋转的表示

35.5 数学是对操作的研究对关系的研究

任何一个操作的集合，再保持某个研究对象的某个特征不变，例如一个皮球还是皮球没有被剪开，的前提下所有的可能的操作的集合。研究这个操作集合的元素之间的关系，从而从已知的具体状态经过操作得到新的状态，或者得到所有新状态的可能，或者研究选择什么样的操作来调控这个对象会满足对这个对象的状态的某个特定要求。这样的操作构成群。

因此，从这个角度可以认为，所有的数学，建立在群论的基础上。不过就是某些群表现出来的结构适合用整数、分数、实数、矩阵等来表示。

35.6 把数学看作是操作的研究的意义

35.7 推荐阅读材料

35.8 作业

35.9 本章小结

第三十六章 数学知识网络和学科 大图景

Mathematics is the tool specially suited for dealing with abstract concepts of any kind and there is no limit to its power in this field.

数学是一个尤其适合描述各种抽象概念的工具。在这方面它的能力无限。

– Paul Dirac

36.1 数学一二层知识的系统性

36.2 数学学科大图景，及其和一二层知识的联系

36.3 从数学知识体系到科学知识体系

公理化方案

36.4 高等数学学什么

为什么叫做高等数学。数学知识上的高等，数学思维上的高等，数学知识体系性上的高等。从数学知识的学习，到数学知识的创造性运用和创造。

主要数学分支子学科。

（基本）不再出现数而是出现关系的数学。

36.5 推荐阅读材料

36.6 作业

36.7 本章小结

第三十七章 再论数学建模

37.1 从数学建模的角度看数学知识的发展

从实数到复数，逻辑道路和现实道路

37.2 从把数学知识用于描述现实看真正的数学应用题

37.3 数学建模五步及其应用

37.4 推荐阅读材料

37.5 作业

37.6 本章和本卷小结

数学的发展来源于现实世界的启发和数学研究者对数学知识结构本身的逻辑严密性的追求。其中，在现实启发方面，主要通过双向的数学建模来完成，通过把数学结构用于描述现实来解决现实问题，以及通过把数学结构用来描述现实启发提出新的数学结构。

参考文献

- [1] 吴金闪. 小学数学这样学 [M]. 浙江人民出版社, 2022.
- [2] 吴金闪. 教的更少, 学得更多 [M]. 科学出版社, 2021.
- [3] 吴金闪. 《系统科学导引》第一卷 [M]. 科学出版社, 2018.
- [4] 赵凯华. 定性与半定量物理学 [M]. 高等教育出版社, 2008.
- [5] 吴文俊. 几何定理机器证明的基本原理 [M]. 科学出版社, 1984.
- [6] 吴文俊. 数学机械化 [M]. 科学出版社, 2003.
- [7] Euclid, edited by Heath T L. The Thirteen Books of Euclid's Elements, Books 1 and 2 (《几何原本》) [M]. New York, NY, USA: Dover Publications, Inc., 1956.
- [8] Hilbert D. The Foundations of Geometry (《希尔伯特几何基础》) [M]. CreateSpace Independent Publishing Platform, 2013.
- [9] Moise E E. Elementary Geometry from an Advanced Standpoint [M]. ADDISON-WESLEY PUBLISHING COMPANY, 1990.
- [10] Koller D, Friedman N. Probabilistic Graphical Models: Principles and Techniques [M]. MIT Press, 2009.
- [11] Bielza C, Larrañaga P. Data-Driven Computational Neuroscience: Machine Learning and Statistical Models [M]. Cambridge University Press, 2020.
- [12] 吴金闪. 《系统科学导引》第二卷 [M]. 科学出版社, 2019.

名词索引

B | M

B

贝叶斯公式 Bayesian Formula. 325, 327, 328

M

理解型学习 Meaningful Learning, 人类知识高速公路上以
高层知识生成器为目标的理解型学习. 23

人名索引

B | C | D | E | F | G | H | P | R | S | T | W | Z

B

Birkhoff George Birkhoff (乔治·伯克霍夫) . 313

C

Cardano Girolamo Cardano (吉罗拉莫·卡尔达诺) .
188

D

Descartes René Descartes (勒内·笛卡尔) . 160

Dirac Paul Dirac (保罗·狄拉克) . 80

E

Einstein Albert Einstein (阿尔伯特·爱因斯坦) . 155,
197

F

Ferrari Lodovico Ferrari (洛多维科·费拉里) . 189

G

Galois Évariste Galois (埃瓦里斯特·伽罗瓦) . 189,
191

Gödel	Kurt Gödel (库尔特·哥德尔) . 406
H	
Hilbert	David Hilbert (大卫·希尔伯特) . 406
P	
彭桓武	中国当代物理学家. 80
Popper	Karl Popper (卡尔·波普尔) . 156
Pythagoras	Pythagoras of Samos (毕达哥拉斯) . 196
R	
Russell	Bertrand Russell (伯特兰·罗素) . 406
S	
数学五步	数学提出和解决问题的四个典型步骤: 提出问题、把问题数学化、求解数学化以后的问题、解的检验、问题和解法的系统化和推广. 396
SMSG	School Mathematics Study Group (学校教学用数学研究小组, SMSG). 313
T	
Tartaglia	Niccolò Tartaglia (尼科洛·塔尔塔利亚) . 188
W	
吴金闪	北京师范大学系统科学学院教师. 24, 26, 80
Z	
赵凯华	北京大学物理系教师. 197

插图

1	集合和元素	45
2	原命题和逆否命题等价	49
3	小学数学整体概念地图	51
1.1	分数在数轴上的表示	66
1.2	加法在数轴上的表示	70
1.3	乘法在数轴上的表示	71
1.4	平面上的点用一对数来表示	74
1.5	平面上的点用一对数来表示	75
4.1	十字相乘法	99
4.2	十字相乘法举例	100
4.3	通过因式分解看函数形状	103
4.4	通过因式分解看函数形状	104
5.1	区间示意图	113
5.2	不等式组的区间示意图	114
6.1	两个集合的交集和并集	121
6.2	子集、全集、补集	123
6.3	三集合相交的每一个区域	125
7.1	复合函数	134
7.2	在圆上的有顺序的三点被转动	139
7.3	g_1 看作两种映射	140
8.1	苏格拉底三段论的集合表示	149

8.2	原命题和逆否命题等价	150
9.1	平面上坐标和点之间的对应关系	161
9.2	一元一次方程的解的几何含义	165
9.3	一元二次函数 $y = x^2$ 的图像	167
9.4	一元二次函数和 $y = y_*$ 直线的交点	170
9.5	负一次方函数 $y = x^{-1}$ 的图像	171
10.1	二元一次不等式组的画图求解	181
11.1	一元二次不等式的解	190
11.2	本书初中代数部分的数学学科概念知识	194
12.1	Pythagoras 割补图形的证明	196
12.2	勾股定理的量纲证明	198
12.3	直角三角形的边角关系和相似	202
12.4	通过单位圆得到正弦函数值	203
12.5	余弦定理的证明	204
12.6	正弦定理的证明	207
13.1	距离公式 (勾股定理)	212
13.2	线上多条垂线的假设	222
13.3	线外多条垂线的假设	224
13.4	垂线段距离最短	224
13.5	角度的两种定义	226
13.6	对顶角相等	227
13.7	三条线相交	228
13.8	平行线和第三条直线相交	229
13.9	多边形的内角和外角	233
15.1	和角三角函数	243
15.2	两直线的夹角	250
15.3	直尺和圆规	258
15.4	尺规作图复制线段	259

插图	445
15.5 尺规作图复制角	260
15.6 尺规作图找线段中点	260
15.7 尺规作图做角平分线	262
15.8 尺规作图做平行线	262
15.9 尺规作图做垂线	263
15.10 等腰三角形的底角和腰	265
16.1 同位角相等则两直线平行	269
16.2 用于平行性质和平行判定定理证明	272
16.3 线外平行线的唯一性	273
17.1 平移的两种视角	289
17.2 圆面积的切割计算	290
17.3 旋转圆上的点	291
18.1 三角形中位线定理	306
18.2 圆和直线的交点	309
19.1 本书平面几何的知识体系	315
20.1 串联和并联元器件	329
21.1 数据 U 的经验分布图	335
21.2 数据 $\mathcal{N}$ 的经验分布图	337
21.3 一万个数据点的 $\mathcal{N}$ 的经验分布图	337
21.4 多个随机数平均的经验分布	341
21.5 多个随机数平均的经验分布	342
21.6 贝叶斯网络用于检测学习效果	345
21.7 贝叶斯网络用于疾病诊断	347
21.8 本书初中概率论和统计学知识的梳理	349
23.1 $P, P_{180^\circ}, \triangle$ 和 $\triangle_{180^\circ}$ 之间定义上的关系	359
23.2 $P, P_{180^\circ}, \triangle$ 和 $\triangle_{180^\circ}$ 之间实际上的关系	360
23.3 $\triangle \rightarrow P_{180^\circ}$ 和 $P_{180^\circ} \rightarrow \triangle$ 的集合关系	360
23.4 原命题和逆否命题等价	361

23.5 集合关系的关系和反证法	366
23.6 三段论的集合表示	368
24.1 平行四边形判定——WHWM	394
24.2 贝叶斯网络用于诊断性检测	402

举例目录

0.1 例 (乘法对加法的右侧分配律)	33
3.1 例 (长方形的面积)	94
3.2 例 (长方形的面积)	94
4.1 例 ($ax + ay + bx + by$)	99
4.2 例 ($x^4 + 4$)	99
4.3 例 ($x^2 + x - 6$)	99
4.4 例 ($(x^2 - 6)^2 - x^2$)	100
4.5 例 (求解 $x^2 + x - 6 = 0$)	101
5.1 例 ($ax + b = cx + d \rightarrow ax + b$)	107
5.2 例 (一元一次不等式组)	113
9.1 例 (从 $y = ax^2 + bx + c$ 到 $y = x^2$)	168
10.1 例 (求解二元一次不等式组)	180
11.1 例 (求解一元二次方程)	187
11.2 例 (求解一元二次方程)	187
11.3 例 (求解一元二次方程)	188
11.4 例 (求解一元二次方程)	188
11.5 例 (求解一元二次不等式)	190

11.6 例 (求解一元二次不等式)	191
12.1 例 (三角形面积公式)	210
15.1 例 (有三角形全等 ASS 判定定理吗?)	248
15.2 例 (求已知直线的垂线)	251
17.1 例 (四点共圆应用)	298
18.1 例 (直线的不同代数表达式)	302
18.2 例 (两点确定一条直线)	303
18.3 例 (三角形中位线定理的解析几何)	306
18.4 例 (方程的解的几何)	308
20.1 例 (计算空集的概率)	322
20.2 例 (从基元事件的概率算一般事件的概率)	322
20.3 例 (从穿裙子推断性别)	326
20.4 例 (串联元器件状态推断)	328
20.5 例 (敏感问题调查的概率论问题)	329
21.1 例 (正态分布的数据和经验分布)	334
21.2 例 (贝叶斯网络计算任意顶点组合的任意状态的概率)	345
22.1 例 (随机选袜子凑一对)	353
22.2 例 (经典粒子放置问题——可区分)	354
22.3 例 (经典粒子放置问题——最多放一个)	354
22.4 例 (费米子放置问题)	354
22.5 例 (玻色子放置问题)	354
22.6 例 (费米子放置问题——都不可区分)	355
23.1 例 (数学归纳法用于证明)	371
23.2 例 (数学归纳法用于证明)	371
23.3 例 (数学归纳法用于求数列通项)	372
24.1 例 (存款方式问题——应用题版本)	382

24.2 例 (存款方式问题——数学建模版本)	383
24.3 例 (存款方式问题——WHWM 版本)	387
24.4 例 (两次的钱合起来——WHWM 版本)	389
24.5 例 (路程问题——WHWM 版本)	391
24.6 例 (平行四边形判定——WHWM 版本)	393
24.7 例 (掌握状态的推断——WHWM 版本)	399

作业目录

0.1 习题 (无限循环小数化为分数)	40
0.2 习题 (无限小数的有限位计算)	42
1.1 习题 (我家的人)	57
1.2 习题 (我爸爸家的人)	57
1.3 习题 (我妈妈家的人)	57
1.4 习题 (我爷爷家的人)	57
1.5 习题 (比较我、爸爸、妈妈、爷爷家的人)	57
1.6 习题 ($f_u^{(+c)}$ 和 $f_u^{(\times c)}$ 的唯一性)	63
1.7 习题 ($\sqrt{\cdot}$ 唯一和不唯一)	63
1.8 习题 (区间关系)	69
1.9 习题 (表达式的运算)	78
1.10 习题 (练习“数、量、物、算式”四个层次的数学对象)	78
1.11 习题 (家里的人)	79
1.12 习题 (对硬币的操作)	79
1.13 习题 (对骰子的操作)	79
1.14 习题 (硬币状态的概率)	79
1.15 习题 (硬币状态概率的检验)	79
2.1 习题 ($(a^m)^n = a^{m \times n}$ 的证明)	82

2.2	习题 (多次方根的问题)	83
2.3	习题 (另一个完全立方公式)	86
2.4	习题 (用计算机作幂次方的计算)	86
2.5	习题 (用计算机作幂次方的画图)	87
3.1	习题 (合并同类项)	93
3.2	习题 (多项式的元和阶)	96
3.3	习题 (多项式降序格式)	96
3.4	习题 (变量和多项式的元和阶)	96
4.1	习题 (通过因式分解看函数形状)	103
4.2	习题 (因式分解)	104
4.3	习题 (因式分解求解方程)	104
6.1	习题 (BMI 的定义)	119
6.2	习题 (理发师悖论)	119
6.3	习题 (集合包含关系和并集一)	123
6.4	习题 (集合包含关系和并集二)	123
6.5	习题 (三集合相交的每一个区域)	124
6.6	习题 (集合运算)	125
6.7	习题 (集合运算)	126
6.8	习题 (集合运算)	126
6.9	习题 (集合运算)	126
6.10	习题 (集合运算)	126
6.11	习题 (集合运算)	126
7.1	习题 (全部四个映射之间的复合关系)	135
7.2	习题 (Dirac 符号下算赋的验证 1)	137
7.3	习题 (Dirac 符号下算赋的验证 2)	137
7.4	习题 (数独游戏)	139
8.1	习题 (否命题和逆命题)	146
9.1	习题 ($f(x) = x + 1$ 是满射)	159

9.2	习题 ($f(x) = x + 1$ 是单射)	159
9.3	习题 ($f(x) = x + 1$ 是满射)	160
9.4	习题 ($f(x) = kx + b$ 是满射)	164
9.5	习题 ($f(x) = kx + b$ 是单射)	165
9.6	习题 (画函数图)	173
9.7	习题 (画函数图)	173
9.8	习题 (画函数图求解方程)	173
9.9	习题 (画函数图求解方程)	173
9.10	习题 (画函数图求解方程)	173
9.11	习题 (画函数图求解方程)	173
9.12	习题 (画函数图求解方程)	173
11.1	习题 (正多边形圈地法)	184
11.2	习题 ($a < 0$ 的时候的一元二次方程的解的公式)	186
11.3	习题 (求解一元二次方程和一元二次不等式)	191
12.1	习题 (相似直角三角形的边)	202
13.1	习题 ($PM + MQ \geq PQ$ 的一般性证明)	217
13.2	习题 (垂线平行)	223
13.3	习题 (平行线之一垂直)	223
13.4	习题 (找同位角、内错角和同旁内角)	230
13.5	习题 (交换直线表达式的 x, y)	232
13.6	习题 (多边形外角和)	233
15.1	习题 (线段中点的尺规作图)	260
15.2	习题 (线外垂线的尺规作图)	261
15.3	习题 (线上垂线的尺规作图)	263
15.4	习题 (三 (二) 线合一则等腰)	266
15.5	习题 (分数在数轴上的表示)	266
16.1	习题 (垂直于同一条直线)	272
16.2	习题 (分别垂直于一组平行线)	272
16.3	习题 (一组对边相等另一组对边平行的四边形)	281

16.4 习题 (正方形的对角线)	284
16.5 习题 (对角线相互垂直平分且相等)	284
17.1 习题 (直径对直角)	293
17.2 习题 (对角互补)	293
17.3 习题 (直角对直径)	299
17.4 习题 (对角互补则共圆)	299
17.5 习题 (垂直于弦的直径)	299
17.6 习题 (同点切线等长)	299
17.7 习题 (乘法在数轴上的表示)	299
18.1 习题 (三角形三条中线交于一点)	308
18.2 习题 (三角形三条高交于一点)	308
20.1 习题 (扔骰子的基元事件)	319
20.2 习题 (扔骰子的复合事件)	320
20.3 习题 (子集来自于交并补)	320
20.4 习题 (交并补必然得到子集)	320
20.5 习题 (“开心不开心”的集合运算)	321
20.6 习题 (贝叶斯公式的证明)	326
20.7 习题 (从信号推断真值)	327
20.8 习题 (并联元器件状态推断)	330
21.1 习题 (自己完成绘制经验分布图)	336
21.2 习题 (用计算机数学软件绘制经验分布图)	338
21.3 习题 (独立画出来这一节的图)	343
21.4 习题 (降低极端值的影响增加典型性)	343
21.5 习题 (贝叶斯公式用于疾病诊断)	347
21.6 习题 (男性女性寿命问题)	347
21.7 习题 (抽烟和健康问题)	347
21.8 习题 (抽烟和健康问题)	348
22.1 习题 (随机选袜子凑一对)	353
22.2 习题 (汉诺塔问题)	355

22.3 习题 (玻色子放置问题——盒子不可区分)	355
23.1 习题 (逆命题和否命题)	361
23.2 习题 (相等的角和对顶角)	362
23.3 习题 (直角三角形锐角互余的三段论)	368
23.4 习题 (数学归纳法用于证明)	372
23.5 习题 (数学归纳法用于求数列通项)	372
24.1 习题 (存款问题数学建模版本的 WHWM)	404
24.2 习题 (银行所谓的“报价利率”)	404
24.3 习题 (路程问题 WHWM 版)	404
24.4 习题 (重复相加问题 WHWM 版)	404
24.5 习题 (年龄问题 WHWM 版)	404
24.6 习题 (抽烟和健康问题)	405
24.7 习题 (抽烟和健康问题)	405