

高中数学（选修）

——学习笔记

程琰

2025 年 7 月 1 日

目录

第零章 前言	1
第一章 空间向量与立体几何	3
1.1 空间向量及其运算	4
1.1.1 空间向量及其线性运算	4
1.1.2 空间向量的数量积运算	4
1.1.3 拓展讨论：向量外积	6
1.1.4 习题	6
1.2 空间向量基本定理	10
1.2.1 习题	10
1.3 空间向量及其运算的坐标表示	13
1.3.1 空间直角坐标系	13
1.3.2 空间向量运算的坐标系表示	13
1.3.3 习题	13
1.4 空间向量的应用	15
1.4.1 用空间向量研究直线、平面的位置关系	15
1.4.2 用空间向量研究距离、夹角问题	16
1.4.3 习题	17
1.5 本章小结	20
1.5.1 习题	20
第二章 直线和圆的方程	23
2.1 直线的倾斜角与斜率	24
2.1.1 倾斜角和斜率	24
2.1.2 两条直线平行和垂直的判定	24
2.2 直线的方程	25
2.2.1 拓展讨论：直线的定义	25
2.2.2 习题	26

2.3	直线的交点坐标与距离公式	27
2.3.1	习题	27
2.4	圆的方程	29
2.4.1	拓展讨论: 圆的定义	29
2.4.2	习题	30
2.5	直线与圆、圆与圆的位置关系	31
2.5.1	直线与圆的位置关系	31
2.5.2	圆与圆的位置关系	31
2.5.3	拓展讨论: 两个圆方程相减的图形	31
2.6	本章小结	33
2.6.1	习题	33
第三章	圆锥曲线的方程	35
3.1	椭圆	36
3.1.1	椭圆及其标准方程	36
3.1.2	椭圆的简单几何性质	36
3.1.3	拓展讨论: 坐标系变换	37
3.1.4	拓展讨论: 斜率之积	37
3.1.5	拓展讨论: 两个焦点	38
3.1.6	拓展讨论: 动点距离直线	39
3.1.7	拓展讨论: 椭圆的一般方程	39
3.1.8	习题	41
3.2	双曲线	43
3.2.1	双曲线及其标准方程	43
3.2.2	双曲线的简单几何性质	43
3.2.3	拓展讨论: 坐标系变换	44
3.2.4	拓展讨论: 斜率之积	44
3.2.5	拓展讨论: 两个焦点	44
3.2.6	拓展讨论: 动点距离直线	45
3.2.7	拓展讨论: 双曲线的一般方程	46
3.2.8	拓展讨论: $y = \frac{1}{x}$ 和 $y = x + \frac{1}{x}$	46
3.2.9	习题	47
3.3	抛物线	50
3.3.1	抛物线及其标准方程	50
3.3.2	抛物线的简单几何性质	50
3.3.3	拓展讨论: 焦点	50
3.3.4	拓展讨论: 抛物线的一般方程	51

目 录	III
3.4 本章小结	52
第四章 数列	53
4.1 数列的概念	54
4.1.1 习题	54
4.2 等差数列	56
4.2.1 等差数列的概念	56
4.2.2 等差数列的前 n 项和公式	56
4.2.3 习题	57
4.3 等比数列	58
4.3.1 等比数列的概念	58
4.3.2 等比数列的前 n 项和公式	58
4.3.3 习题	59
4.4 数学归纳法	63
4.5 本章小结	64
4.5.1 习题	64
第五章 一元函数的导数及其应用	67
5.1 导数的概念及其意义	68
5.1.1 变化率的问题	68
5.1.2 导数的概念及其几何意义	68
5.2 导数的运算	70
5.2.1 基本初等函数的导数	70
5.2.2 导数的四则运算法则	71
5.2.3 简单的复合函数的导数	71
5.2.4 拓展讨论：反函数的求导	71
5.2.5 拓展讨论：隐函数的求导	72
5.2.6 拓展讨论：参数式函数的求导	72
5.2.7 拓展讨论：高阶导数	73
5.2.8 习题	73
5.3 导数在研究函数中的应用	74
5.3.1 函数的单调性	74
5.3.2 函数的极值与最大（小）值	74
5.3.3 拓展讨论：目标函数优化	75
5.3.4 拓展讨论：凸函数和拐点	77
5.3.5 拓展讨论：琴生（Jensen）不等式	77
5.3.6 拓展讨论：渐近线	78

5.4	本章小结	80
5.4.1	习题	80
第六章	计数原理	81
6.1	分类加法计数原理与分步乘法计数原理	82
6.2	排列与组合	83
6.3	二项式定理	84
6.4	本章小结	85
第七章	随机变量及其分布	87
7.1	条件概率与全概率公式	88
7.1.1	条件概率	88
7.1.2	全概率公式	88
7.1.3	习题	90
7.1.4	拓展习题	92
7.2	离散型随机变量及其分布列	95
7.2.1	拓展讨论: 密度函数的物理意义	97
7.3	离散型随机变量的数字特征	99
7.3.1	离散型随机变量的均值	99
7.3.2	离散型随机变量的方差	100
7.3.3	拓展讨论: 数学期望的数学意义	100
7.3.4	拓展讨论: 数学期望的物理意义	101
7.3.5	拓展讨论: 数学期望的性质	101
7.3.6	拓展讨论: 数学期望的范围	103
7.3.7	拓展讨论: 中位数的概念	103
7.3.8	拓展讨论: 方差的数学由来和统计意义	103
7.3.9	拓展讨论: 方差的物理意义	103
7.3.10	拓展讨论: 方差的性质	104
7.3.11	拓展讨论: 方差的取值范围	105
7.3.12	拓展讨论: 随机变量的标准化	105
7.3.13	习题	106
7.4	二项分布与超几何分布	109
7.4.1	二项分布	109
7.4.2	超几何分布	110
7.4.3	习题	111
7.5	正态分布	113
7.5.1	习题	114

7.6 本章小结	115
7.6.1 习题	115
第八章 成对数据的统计分析	119
8.1 成对数据的统计相关性	120
8.2 一元线性回归模型及其应用	121
8.3 列联表与独立性检验	122
8.4 扩展知识：概率论和数理统计的知识结构	123
8.5 本章小结	124

第零章 前言

数学选修，上中下三册，讲了 3 大部分内容：

- 几何：用代数的方法讨论立体几何和平面几何；
- 代数：介绍导数的概念；
- 概率论：简单介绍概率论的基础知识。

章节编排：

- **第一章 空间向量与立体几何**，49 页：用代数的方法讨论几何，立体几何到此介绍完毕；
- **第二章 直线和圆的方程**，54 页：用代数的方法讨论平面的直线和圆；
- **第三章 圆锥曲线的方程**，43 页：用代数的方法讨论平面的椭圆、双曲线、抛物线，平面几何到此介绍完毕；
- **第四章 数列**，57 页：介绍等差数列和等比数列；
- **第五章 一元函数的导数及其应用**，47 页：简单介绍微积分的基础——导数，至此代数介绍完毕；
- **第六章 计数原理**，38 页：主要讲排列组合，为下一章铺垫；
- **第七章 随机变量及其分布**，49 页：介绍概率论的基础知识；
- **第八章 成对数据的统计分析**，50 页：继续介绍概率论的基础知识和数理统计中的假设检验，至此概率论介绍完毕。

本笔记选取有价值的课后习题，给出求解和难度评价。

难度评价标准：

- ★：简单，考察基本定义等概念，解题方法有章可循；
- ★★：中等，考察“定理—定义—性质”的数学范式，解题方法依然有章可循，可能多绕几个弯，类似“定理—定义 1—性质—定义 2—性质……”；
- ★★★：较难，需融汇数个知识点，要求对这几个知识点在方法论层面融会贯通。

- ***: 最难, 考察实际问题, 并没有直接给出明确的数学问题, 需自行建模, 所以我们需要在方法论和哲学层面融会贯通所有知识点。

第一章 空间向量与立体几何

本章介绍空间上的向量，对必修第六章（平面向量）的拓展，大致路数、重点和难点和第六章平面向量类似。

本章要点：

- 空间向量及其几何意义。

自行对比第六章的目录，看看结构是不是高度相似。

大学数学中关于三维空间的向量问题在微积分和线性代数中讲述，特别是线性代数，一整本书都在讲述向量。对比之下，高中课本的 49 页篇幅实在过于粗浅，加之向量接触时间短，所以需要反复仔细阅读课本，深刻领会。

1.1 空间向量及其运算

本节要点:

- 掌握向量的概念;
- 掌握向量运算的概念及其法则;
- 深刻理解向量运算的几何意义。

1.1.1 空间向量及其线性运算

定义 1.1.1. 我们称空间内既有大小又有方向的量为**向量**, 常用粗体小写记作 $\mathbf{a}, \mathbf{b}, \mathbf{c}$, 同时称向量的大小为**长度**或**模**。若 $|\mathbf{a}| = 0$, 称为**零向量**, 记作 $\mathbf{0}$; 若 $|\mathbf{a}| = 1$, 称为**单位向量**。若数个向量的方向一致或反向, 则称它们为**共线向量**或**平行向量**; 若数个向量平行于同一个平面, 则称它们为**共面向量**, 显然任意两个向量必共面。

注意, 向量的定义中并没有规定起点。

几何上, 向量的加法满足平行六面体法则, 向量的数乘表示伸缩, 且有如下运算法则:

$$\begin{aligned}\mathbf{a} + \mathbf{b} &= \mathbf{b} + \mathbf{a} \\ (\mathbf{a} + \mathbf{b}) + \mathbf{c} &= \mathbf{a} + (\mathbf{b} + \mathbf{c}) \\ \lambda(\mu\mathbf{a}) &= (\lambda\mu)\mathbf{a} \\ (\lambda + \mu)\mathbf{a} &= \lambda\mathbf{a} + \mu\mathbf{a} \\ \lambda(\mathbf{a} + \mathbf{b}) &= \lambda\mathbf{a} + \lambda\mathbf{b}\end{aligned}$$

定理 1.1.1. 不难得到以下两个定理:

- 平行定理: $\mathbf{a} \parallel \mathbf{b} \Leftrightarrow \mathbf{a} = \lambda\mathbf{b}$
- 共面定理: \mathbf{p} 与 \mathbf{a}, \mathbf{b} 共面 $\Leftrightarrow \mathbf{p} = x\mathbf{a} + y\mathbf{b}$

1.1.2 空间向量的数量积运算

定义 1.1.2 (数量积). 设 \mathbf{a}, \mathbf{b} 为两个空间向量, 我们称它们各自的模的乘积和它们的夹角的余弦的乘积为它们的**数量积**, 也称**内积**, 记作 $\mathbf{a} \cdot \mathbf{b}$, 即:

$$\mathbf{a} \cdot \mathbf{b} := |\mathbf{a}| |\mathbf{b}| \cos \theta$$

几何上，内积表示一向量在另一向量上的投影的乘积，有交换律和分配律，如下：

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= \mathbf{b} \cdot \mathbf{a} \\ \mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) &= \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c} \end{aligned}$$

注意，内积没有结合律。

定理 1.1.2. 不难得到以下三个定理：

- 垂直定理： $\mathbf{a} \perp \mathbf{b} \Leftrightarrow \mathbf{a} \cdot \mathbf{b} = 0$
- 模长定理： $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$
- 平行定理： $\mathbf{a} = \lambda_1 \mathbf{b}_1 + \lambda_2 \mathbf{b}_2 \Leftrightarrow \mathbf{a}$ 平行于 $\mathbf{b}_1, \mathbf{b}_2$ 构成的平面

定理 1.1.3. 使用内积不难证明以下两个重要的不等式：

- 三角不等式： $|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$ ，当且仅当 \mathbf{a}, \mathbf{b} 同向时等号成立；
- 柯西—施瓦茨不等式： $|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}| \cdot |\mathbf{b}|$ ，当且仅当 \mathbf{a}, \mathbf{b} 平行时等号成立。

证明. 简单证明第一个不等式：

$$\begin{aligned} \therefore |\mathbf{a} + \mathbf{b}|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) = \mathbf{a} \cdot \mathbf{a} + 2(\mathbf{a} \cdot \mathbf{b}) + \mathbf{b} \cdot \mathbf{b} \\ \therefore (|\mathbf{a}| + |\mathbf{b}|)^2 &= |\mathbf{a}|^2 + 2|\mathbf{a}||\mathbf{b}| + |\mathbf{b}|^2 \\ \therefore \mathbf{a} \cdot \mathbf{b} &= |\mathbf{a}||\mathbf{b}|\cos\theta \\ \therefore |\mathbf{a} + \mathbf{b}|^2 &\leq (|\mathbf{a}| + |\mathbf{b}|)^2 \end{aligned}$$

□

三角不等式表示的是加法，柯西—施瓦茨不等式表达的是乘法，两者反映了线性赋范空间上的守恒，XML。

1.1.2 中讨论空间向量内积的几何意义，我们将两个空间向量挪到一个平面内，用平面向量的内积解释空间向量，这反映了人类拓展知识的方法，即用已有的知识拓展研究未知并获得新的知识。

例 3 很有意思，仔细体会，例题如何从定义出发用代数的方法证明，例题并没有用纯几何的方法。

两个向量还有外积，定义略，我看外积的模：

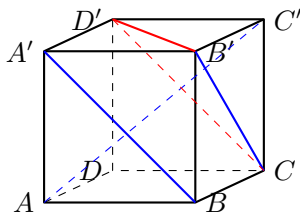
$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| \sin \theta$$

表示 a, b 构成的平行四边形的面积, 和内积有关系:

$$|\mathbf{a} \times \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = |\mathbf{a}| |\mathbf{b}|$$

例 1.1.1 (综合运用 7, 难度 ★). 如图, 正方体 $ABCD - A'B'C'D'$ 的棱长为 a 。

1. 求 $A'B$ 和 $B'C'$ 的夹角;
2. 求证: $A'B \perp AC'$ 。



(1)

$$\cos \theta = \frac{\overrightarrow{A'B} \cdot \overrightarrow{B'C}}{|\overrightarrow{A'B}| \cdot |\overrightarrow{B'C}|} = \frac{(\overrightarrow{AB} - \overrightarrow{AA'}) \cdot (\overrightarrow{BC} - \overrightarrow{BB'})}{|\overrightarrow{A'B}| \cdot |\overrightarrow{B'C}|} = \frac{\overrightarrow{AA'} \cdot \overrightarrow{BB'}}{\sqrt{2}a \cdot \sqrt{2}a} = \frac{1}{2}$$

(2)

$$\overrightarrow{A'B} \cdot \overrightarrow{AC'} = (\overrightarrow{AB} - \overrightarrow{AA'}) \cdot (\overrightarrow{AB} + \overrightarrow{AD} + \overrightarrow{CC'}) = a^2 - a^2 = 0$$

(1) 可以用纯几何的方法, $A'B$ 和 $B'C$ 的夹角也即 $D'C$ 和 $B'C$ 的夹角, 添加如图红色辅助线, 不难发现 $\triangle D'B'C$ 为等边三角形, 后略。

我可以试着分析若要 $A'B \perp AC'$ ，则该几何体需要满足的条件。

$$\begin{aligned}\overrightarrow{A'B} \cdot \overrightarrow{AC'} &= (\overrightarrow{AB} - \overrightarrow{AA'}) \cdot (\overrightarrow{AB} + \overrightarrow{AD} + \overrightarrow{CC'}) \\ &= \overrightarrow{AB} \cdot \overrightarrow{AB} - \overrightarrow{AA'} \cdot \overrightarrow{AB} + \overrightarrow{AB} \cdot \overrightarrow{AD} - \overrightarrow{AA'} \cdot \overrightarrow{AD} + \overrightarrow{AB} \cdot \overrightarrow{CC'} - \overrightarrow{AA'} \cdot \overrightarrow{CC'}\end{aligned}$$

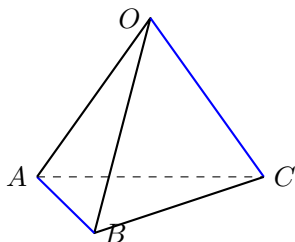
当对棱平行且等长时：

$$\begin{aligned}
 &= -\overrightarrow{AA'} \cdot \overrightarrow{AB} + \overrightarrow{AB} \cdot \overrightarrow{AD} - \overrightarrow{AA'} \cdot \overrightarrow{AD} + \overrightarrow{AB} \cdot \overrightarrow{CC'} \\
 &= \overrightarrow{AB} \cdot (\overrightarrow{AD} - \overrightarrow{AA'}) + \overrightarrow{AA'} \cdot (\overrightarrow{AB} - \overrightarrow{AD}) \\
 &= \overrightarrow{AB} \cdot \overrightarrow{A'D} + \overrightarrow{AA'} \cdot \overrightarrow{DB}
 \end{aligned}$$

侧面均为菱形。

本题考察向量内积的几何意义。

例 1.1.2 (拓广探索 9, 难度: ★). 如图, 在四面体 $OABC$ 中, $OA \perp BC$, $OB \perp AC$. 求证: $OC \perp AB$.



解：

$$\begin{aligned}
 \overrightarrow{OC} \cdot \overrightarrow{AB} &= (\overrightarrow{AC} - \overrightarrow{AO}) \cdot (\overrightarrow{CB} - \overrightarrow{CA}) \\
 &= \overrightarrow{AC} \cdot \overrightarrow{CB} - \overrightarrow{AO} \cdot \overrightarrow{CB} - \overrightarrow{AC} \cdot \overrightarrow{CA} + \overrightarrow{AO} \cdot \overrightarrow{CA} \\
 &= \overrightarrow{AC} \cdot \overrightarrow{CB} + \overrightarrow{AC} \cdot \overrightarrow{AC} + \overrightarrow{AO} \cdot \overrightarrow{CA} \\
 &= \overrightarrow{AC} \cdot \overrightarrow{AB} - \overrightarrow{AO} \cdot \overrightarrow{AC} \\
 &= \overrightarrow{AC} \cdot \overrightarrow{OB} = 0
 \end{aligned}$$

深入分析：

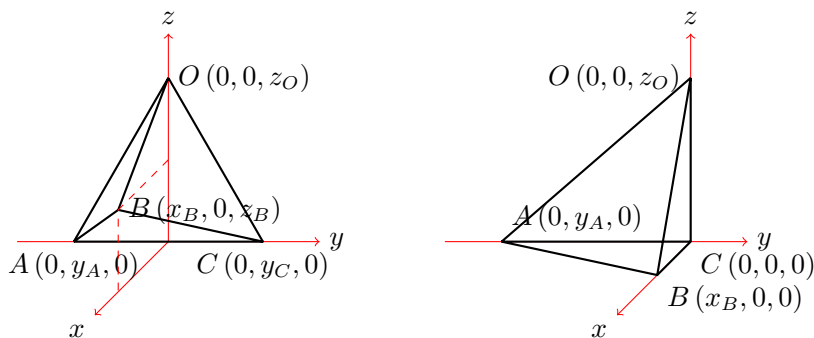
我们证明了在任意的四面体中, 只要两组对棱垂直, 则第三组对棱一定垂直。我们可以进一步分析, 这样三组对棱垂直的情况下, 四面体会有什么性质。

以 $OB \perp AC$ 为例, 意思就是 B, O 两点在 AC 上的投影重合, 也即任意两个顶点在对面棱上的投影重合。根据这个结论建立如下左图的坐标系,

可得:

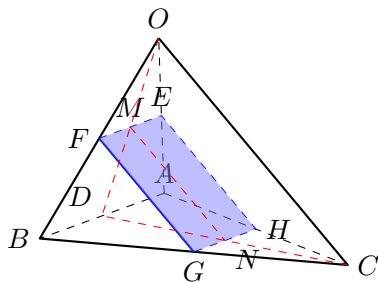
$$\begin{aligned} \therefore \begin{cases} OA \perp BC \Rightarrow (0, y_A, -z_O) \cdot (-x_B, y_C, -z_B) = 0 \\ OC \perp BA \Rightarrow (0, y_C, -z_O) \cdot (-x_B, y_A, -z_B) = 0 \end{cases} \\ \therefore y_A y_C + z_O z_B = 0 \end{aligned}$$

四面体不一定非要正四面体, 特别地, 当 $z_B = 0$ 时, 要么 $y_A = 0$ 要么 $y_C = 0$, 四面体为长方体的一个角, 如下右图。



本题依然考察向量内积的几何意义。

例 1.1.3 (拓广探索 10, 难度: $\star\star$). 如图, 在四面体 $OABC$ 中, $OA = OB, CA = CB$, E, F, G, H 分别是 OA, OB, BC, CA 的中点。求证: 四边形 $EFGH$ 是矩形。



解:

首先, 不难通过 $\overrightarrow{FG} = \overrightarrow{EH}$ 证明四边形 $EFGH$ 是平行四边形。

再证明 $\angle EFG$ 为直角。直接用向量的方法证明 $\overrightarrow{FG} \cdot \overrightarrow{FE} = 0$ 较复杂。找到 AB 中点 D , 连接 OD, CD , 不难证明 OD, CD 是三角形的中垂线,

而且 FG, MN 平行且相等, 所以也即证明 $\overrightarrow{MN} \cdot \overrightarrow{FE} = 0$ 。

$$\overrightarrow{MN} \cdot \overrightarrow{FE} = (\overrightarrow{DN} - \overrightarrow{DM}) \cdot \overrightarrow{FE} = \overrightarrow{DN} \cdot \overrightarrow{GH} - \overrightarrow{DM} \cdot \overrightarrow{FE}$$

深入分析:

不难发现, 当 AB 长度一定时, OD, CD 决定了矩形面积, 我们可以考察一下当矩形面积 S 和 DO, DC 长度的关系。

$$\begin{aligned} \because S &= \frac{AB}{2} \cdot \frac{OC}{2} \\ \because \cos \angle ODC &= \frac{DO^2 + DC^2 - OC^2}{2 \cdot DO \cdot DC} \\ \therefore S &= \frac{AB}{4} \cdot (DO^2 + DC^2 - 2 \cos \angle ODC \cdot DO \cdot DC) \end{aligned}$$

易得:

- 当 S 恒定时, OD, CD 是一个椭圆关系, 可设 $\angle ODC \in (0, \frac{\pi}{2}]$, 则 $2 \cos \angle ODC \in [0, 2)$, 椭圆转了 45° , 特别地, 当 $DO \perp DC$ 时, 为圆;
- 当 OD, CD 相互制约时, S 有最值。

本题需要添加辅助线, 用纯几何 + 向量的方法, 事半功倍。

1.2 空间向量基本定理

本节要点：

- 掌握空间向量基本定理。

定理 1.2.1 (空间向量基本定理). 设 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 为三个不共面的空间向量, 则对于任意向量 \mathbf{p} , 存在唯一的有序实数组 (x, y, z) , 使得

$$\mathbf{p} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

我们把 $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ 称作空间的一个基, 把 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 称作基向量, 对应的 (x, y, z) 称为 \mathbf{p} 在该组基下的坐标。

该定理和平面向量基本定理一样, 指的是向量可以由基组合得到。也要注意, 定理表达的意思是给定一个基就有一个坐标与之对应, 所以不同的基有不同的坐标, 但 \mathbf{p} 还是那个 \mathbf{p} , 也即:

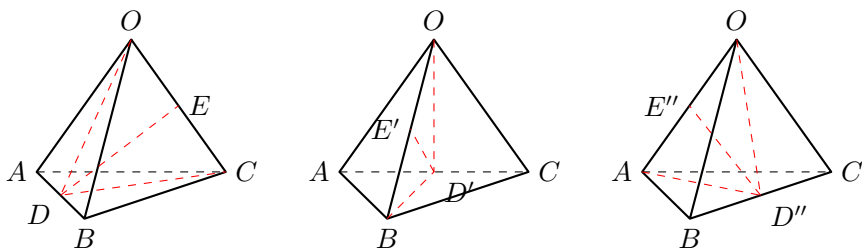
$$\mathbf{p} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c} = l\mathbf{d} + m\mathbf{e} + n\mathbf{f}$$

1.2.1 习题

例 1.2.1 (拓广探索 8, 难度: $\star\star$). 已知四面体中三组相对棱的中点间的距离都相等, 求证: 这个四面体相对的棱两两垂直。

解:

如下图, 不难得到三组相对棱的中点连接的向量:



$$\begin{aligned}\overrightarrow{DE} &= -\frac{1}{4}(\overrightarrow{OB} + \overrightarrow{OA} + \overrightarrow{CB} + \overrightarrow{CA}) \\ \overrightarrow{D'E'} &= -\frac{1}{4}(\overrightarrow{OA} + \overrightarrow{OC} + \overrightarrow{BA} + \overrightarrow{BC}) \\ \overrightarrow{D''E''} &= -\frac{1}{4}(\overrightarrow{AB} + \overrightarrow{AC} + \overrightarrow{OB} + \overrightarrow{OC})\end{aligned}$$

若要平方进行长度相等，非常复杂，我们简化，使之都用 OX 表示：

$$\begin{aligned}\overrightarrow{DE} &= -\frac{1}{2}(\overrightarrow{OB} + \overrightarrow{OA} - \overrightarrow{OC}) \\ \overrightarrow{D'E'} &= -\frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OC} - \overrightarrow{OB}) \\ \overrightarrow{D''E''} &= -\frac{1}{2}(\overrightarrow{OB} + \overrightarrow{OC} - \overrightarrow{OA})\end{aligned}$$

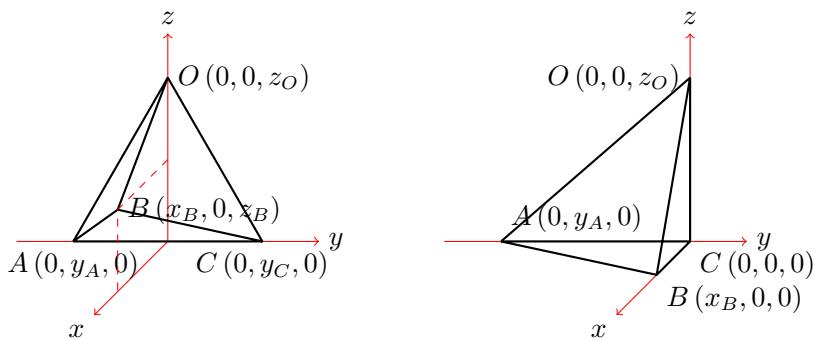
取 $|\overrightarrow{DE}| = |\overrightarrow{D'E'}|$ 分析：

$$\begin{aligned}\because |\overrightarrow{DE}| &= |\overrightarrow{D'E'}| \\ \therefore (\overrightarrow{OB} + \overrightarrow{OA} - \overrightarrow{OC})^2 &= (\overrightarrow{OA} + \overrightarrow{OC} - \overrightarrow{OB})^2 \\ \therefore \overrightarrow{OB} \cdot \overrightarrow{OA} - \overrightarrow{OA} \cdot \overrightarrow{OC} &= \overrightarrow{OA} \cdot \overrightarrow{OC} - \overrightarrow{OA} \cdot \overrightarrow{OB} \\ \therefore \overrightarrow{OB} \cdot \overrightarrow{OA} - \overrightarrow{OA} \cdot \overrightarrow{OC} &= 0 \\ \therefore \overrightarrow{OA} \cdot (\overrightarrow{OB} - \overrightarrow{OC}) &= 0 \\ \therefore \overrightarrow{OA} \cdot \overrightarrow{CB} &= 0\end{aligned}$$

余下略。

深入分析：

结合上一节的拓广探索 9，若三对棱垂直了，计算一下对棱的中点距。



以 OA, BC 为例，它们的中点距为 $(\frac{x_B}{2}, \frac{y_A}{2}, \frac{z_B}{2}), (0, \frac{y_C}{2}, \frac{z_O}{2})$ 的距离，令 l ，则有：

$$\begin{aligned}\because OA \perp BC &\Rightarrow (0, y_A, -z_O) \cdot (-x_B, y_C, -z_B) = 0 \\ \therefore y_A y_C + z_O z_B &= 0 \\ \therefore l^2 &= \left(\frac{x_B}{2}\right)^2 + \left(\frac{y_A - y_C}{2}\right)^2 + \left(\frac{z_B - z_O}{2}\right)^2 \\ \therefore 4l^2 &= x_B^2 + y_A^2 + y_C^2 + z_B^2 + z_O^2 = OA^2 + BC^2\end{aligned}$$

而且有：

$$OA^2 + BC^2 = OC^2 + AB^2 = OB^2 + AC^2$$

这也是对棱垂直的四面体的性质，不难发现，正四面体和长方体一角都是该式的特殊形式。

本题思路简单，从定义出发，但是计算量大，如果不将一开始的表达式化简，几乎无解。

1.3 空间向量及其运算的坐标表示

本节要点：

- 掌握向量运算的坐标表示。

1.3.1 空间直角坐标系

和平面向量一样，当我们定义了空间直角坐标系后，可以得到空间向量的 3 个等价表示方法：

- \overrightarrow{AB} ：表示三维空间的有向线段；
- $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ ：表示 xyz 坐标系中的有向线段；
- (x, y, z) ：表示 xyz 坐标系中的点。

这里要注意一点，直角坐标系的右手性是人为规定的。

1.3.2 空间向量运算的坐标系表示

加法、数乘、内积的坐标表示如下：

$$\mathbf{a} + \mathbf{b} = (x_a + x_b)\mathbf{i} + (y_a + y_b)\mathbf{j} + (z_a + z_b)\mathbf{k}$$

$$\lambda\mathbf{a} = \lambda x_a\mathbf{i} + \lambda y_a\mathbf{j} + \lambda z_a\mathbf{k}$$

$$\mathbf{a} \cdot \mathbf{b} = x_a \cdot x_b + y_a \cdot y_b + z_a \cdot z_b = |\mathbf{a}| |\mathbf{b}| \cos \alpha$$

1.3.3 习题

例 1.3.1 (拓广探索 9, 难度: **). $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ 是空间的一个单位正交基底, 向量 $\mathbf{p} = \mathbf{a} + 2\mathbf{b} + 3\mathbf{c}$, $\{\mathbf{a} + \mathbf{b}, \mathbf{a} - \mathbf{b}, \mathbf{c}\}$ 是空间的另一个基底, 用 $\{\mathbf{a} + \mathbf{b}, \mathbf{a} - \mathbf{b}, \mathbf{c}\}$ 表示向量 \mathbf{p} 。

解：

\mathbf{p} 在新坐标系下可表示为：

$$\mathbf{p} = x(\mathbf{a} + \mathbf{b}) + y(\mathbf{a} - \mathbf{b}) + z\mathbf{c} = (x + y)\mathbf{a} + (x - y)\mathbf{b} + z\mathbf{c}$$

于是：

$$\begin{cases} x + y = 1 \\ x - y = 2 \\ z = 3 \end{cases} \Rightarrow \begin{cases} x = \frac{3}{2} \\ y = -\frac{1}{2} \\ z = 3 \end{cases}$$
$$\boldsymbol{p} = \frac{3}{2}(\boldsymbol{a} + \boldsymbol{b}) - \frac{1}{2}(\boldsymbol{a} - \boldsymbol{b}) + 3\boldsymbol{c}$$

本题其实是对空间向量基本定理的拓展，任意向量在不同基下有不同的表示，但向量还是那个向量。

1.4 空间向量的应用

本节要点：

- 掌握空间点线面的方程；
- 熟练掌握使用向量判断空间点线面的关系。

1.4.1 用空间向量研究直线、平面的位置关系

首先定义点、线、面的向量表达式。

空间中的点

直接用向量或其坐标表示：

$$\overrightarrow{OP} = \mathbf{p} = (x_p, y_p, z_p)$$

空间中的直线

直线的定义是和给定点 P_0 构成的向量与给定向量 \mathbf{n} 平行的所有点的集合，即：

$$\overrightarrow{P_0P} = \lambda \mathbf{n} \quad \text{或} \quad (x - x_0, y - y_0, z - z_0) = \lambda (A, B, C)$$

展开后得空间直线方程：

$$\frac{x - x_0}{A} = \frac{y - y_0}{B} = \frac{z - z_0}{C}$$

其中：

- $\mathbf{n} = (A, B, C)$ ：直线的方向；
- (x_0, y_0, z_0) ：直线上一点 P_0 的坐标。

空间中的平面

平面的定义是和给定点 P_0 构成的向量与给定向量 \mathbf{n} 垂直得所有点得集合，即：

$$\overrightarrow{P_0P} \cdot \mathbf{n} = 0 \quad \text{或} \quad A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

展开后得到空间平面的一般方程：

$$Ax + By + Cz + D = 0$$

其中：

- $\mathbf{n} = (A, B, C)$ ：平面的法线方向；
- (x_0, y_0, z_0) ：平面上一点 P_0 的坐标。

特别地，当平面和 xyz 轴分别交于 x_0, y_0, z_0 时，则可以写成截距式：

$$\frac{x}{x_0} + \frac{y}{y_0} + \frac{z}{z_0} = 0$$

以上定义中，关键在于点和方向。点好理解，直线或平面上的点，直线的方向也好理解，面的方向是第一次碰到，我们取平面的法线的方向作为平面的方向。还要注意，高中阶段的方向，不分正反！

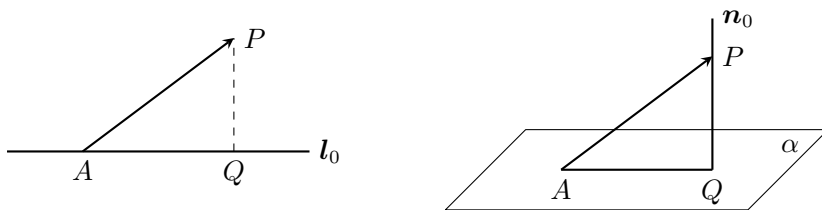
定理 1.4.1. 根据以上定义，我们不难得到直线和平面关系的判定定理：

- 若 $l_1 = \lambda l_2$ ，则两直线 l_1, l_2 平行；
- 若 $l \cdot n = 0$ ，则直线 l 平面 α 平行；
- 若 $n_\alpha = \lambda n_\beta$ ，则两平面 α, β 平行；
- 若 $l_1 \cdot l_2 = 0$ ，则两直线 l_1, l_2 垂直；
- 若 $l = \lambda n$ ，则直线 l 平面 α 垂直；
- 若 $n_\alpha \cdot n_\beta = 0$ ，则两平面 α, β 垂直。

	平行	垂直
线线 l_1, l_2	$l_1 = \lambda l_2$	$l_1 \cdot l_2 = 0$
线面 l, α	$l \cdot n = 0$	$l = \lambda n$
面面 α, β	$n_\alpha = \lambda n_\beta$	$n_\alpha \cdot n_\beta = 0$

1.4.2 用空间向量研究距离、夹角问题

如下图，点 P 到直线和平面的距离表达式：



$$PQ = \sqrt{AP^2 - (\overrightarrow{AP} \cdot \vec{l}_0)^2}$$

$$PQ = |\overrightarrow{AP} \cdot \vec{n}_0|$$

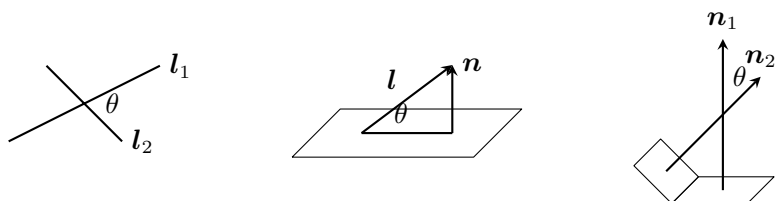
其中：

- \vec{l}_0 : 直线的单位方向向量；

- \mathbf{n}_0 : 平面的单位法向方向向量。

如下图, 线线、线面、面面夹角的余弦表达式:

$$\cos \theta = \frac{|\mathbf{l}_1 \cdot \mathbf{l}_2|}{|\mathbf{l}_1| |\mathbf{l}_2|} \quad \sin \theta = \frac{|\mathbf{l} \cdot \mathbf{n}|}{|\mathbf{l}| |\mathbf{n}|} \quad \cos \theta = \frac{|\mathbf{n}_1 \cdot \mathbf{n}_2|}{|\mathbf{n}_1| |\mathbf{n}_2|}$$

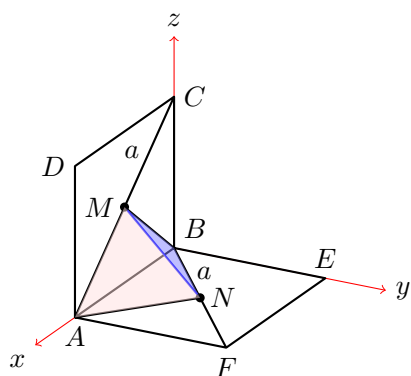


本节的例 1、例 2、例 3、例 9、例 10, 都值得反复阅读。

1.4.3 习题

例 1.4.1 (拓广探索 18, 难度: ***). 在如图所示的实验装置中, 两个正方形框架 $ABCD$, $ABEF$ 的边长都是 1, 且它们所在的平面互相垂直。活动弹子 M, N 分别在正方形对角线 AC 和 BF 上移动, 且 CM 和 BN 的长度保持相等, 记 $CM = BN = a, 0 < a < \sqrt{2}$ 。

1. 求 MN 的长;
2. a 为何值时, MN 的长最小?
3. 当 MN 的长最小时, 求平面 MNA 和 MNB 夹角的余弦值。



解:

(1) 以 B 为原点建立直角坐标系。根据关系可得 M, N 坐标 $M = (x, 0, 1-x), N = (x, x, 0)$, 其中 $\sqrt{2}x = a$, 于是:

$$MN = \sqrt{x^2 + (1-x)^2} = \sqrt{2x^2 - 2x + 1} = \sqrt{\left(\sqrt{2}x - \frac{1}{\sqrt{2}}\right)^2 + \frac{1}{2}}$$

(2) 可见, 当 $x = \frac{1}{2}$ 时, MN 最小, 为 $\sqrt{\frac{1}{2}}$, 此时 $a = \frac{\sqrt{2}}{2}$ 。

(3) 不难发现 $\triangle AMN, \triangle BMN$ 均为等边三角形, 后略。

深入分析:

探讨一下 MN 最值。首先有一个直觉, 高中阶段的不等值最终都可以归结为基本不等式, MN 两头都是 $CB = AF = 1$, 一定是对称点取到最值, 事实也是如此。另一个直觉, 异面直线距离最短, MN 应该是距离, 但显然结果的 MN 并不是距离, 这点从 MN 和 AC, BF 都不垂直可以得出。两个直觉发生矛盾了。

假设 M, N 各自运动, 我们看看什么时候 MN 最短。另 M, N 坐标 $M = (x_m, 0, 1-x_m), N = (x_n, x_n, 0)$, 于是:

$$\begin{aligned} MN &= \sqrt{(x_m - x_n)^2 + x_n^2 + (1-x_m)^2} \\ &= \sqrt{2x_m^2 - 2x_mx_n + 2x_n^2 + 1 - 2x_m} \end{aligned}$$

另 $y = 2x_m^2 - 2x_mx_n + 2x_n^2 + 1 - 2x_m$, 计算一阶和二阶偏导:

$$\begin{aligned} \therefore \begin{cases} \frac{\partial y}{\partial x_m} = 4x_m - 2x_n - 2 = 0 \\ \frac{\partial y}{\partial x_n} = -2x_m + 4x_n = 0 \end{cases} & \begin{cases} A = \frac{\partial^2 y}{\partial x_m^2} = 4 \\ C = \frac{\partial^2 y}{\partial x_n^2} = 4 \\ B = \frac{\partial^2 y}{\partial x_m \partial x_n} = -2 \end{cases} \\ \therefore x_m = \frac{2}{3}, x_n = \frac{1}{3} \end{aligned}$$

由于 $AC > B, A > 0$, 所以该值为最小值, 此时:

$$\begin{aligned} M &= \left(\frac{2}{3}, 0, \frac{1}{3}\right), N = \left(\frac{1}{3}, \frac{1}{3}, 0\right) \\ MN &= \sqrt{(1/3)^2 + (1/3)^2 + (1/3)^2} = 1/\sqrt{3} \end{aligned}$$

再考察 $\triangle AMN$:

$$\begin{aligned} A &= (1, 0, 0) \\ AM &= \sqrt{(1/3)^2 + (1/3)^2} = \sqrt{2}/\sqrt{3} \\ AN &= \sqrt{(2/3)^2 + (1/3)^2} = \sqrt{5}/\sqrt{3} \end{aligned}$$

可见 $\triangle AMN$ 为直角三角形, 同理 $\triangle BMN$ 也为直角三角形, 此刻 MN 垂直于 AC, BF , MN 确实是距离。

分析可得, 这两个直觉的矛盾在于 M, N 两点的运动方式, 若无约束确实可以取到距离, 但题目对它们的运动方式有了约束, 所以取不到距离。

本题采用向量的方法将几何问题转化为代数问题, 需要将几何和代数融会贯通, 但总体上思路还是明确的。

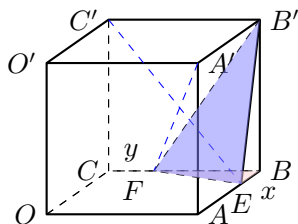
1.5 本章小结

引入向量后,所有的几何问题都有了明显的思路,只是计算量大小的问题。任何几何问题采用暴力计算的方法一定能求解,但这显然不是高效的方法。要学好几何,则不可偏用向量,适当结合一定的纯几何可大大降低计算量。

1.5.1 习题

例 1.5.1 (拓广探索 16, 难度: **). 如图, 在棱长为 a 的正方体 $OABC - O'A'B'C'$ 中, E, F 分别是棱 AB, BC 上的动点, 且 $AE = BF$ 。

- 求证: $A'F \perp C'E$;
- 当三棱锥 $B' - BEF$ 的体积取得最大值时, 求平面 $B'EF$ 与平面 BEF 夹角的正切值。



解:

(1) 在 E, F 均是动点的前提下, 显然不能用纯几何的方法求证。考虑采用向量的方法, 以 O 为原点建立直角坐标系, 于是:

$$\begin{aligned}
 &\therefore \begin{cases} F = (0, y, 0), E = (x, a, 0) \\ A' = (a, a, a), C' = (0, 0, a) \end{cases} \\
 &\therefore \begin{cases} \overrightarrow{A'F} = (-a, y-a, -a) \\ \overrightarrow{C'E} = (x, a, -a) \end{cases} \\
 &\therefore \overrightarrow{A'F} \cdot \overrightarrow{C'E} = -ax + a(y-a) + a^2 = -ax + ay \\
 &\therefore AE = BF \\
 &\therefore EB = CF \\
 &\therefore x = y \\
 &\therefore \overrightarrow{A'F} \cdot \overrightarrow{C'E} = 0
 \end{aligned}$$

(2) 三棱锥 $B' - BEF$ 的体积

$$V_{B' - BEF} = \frac{1}{3} \cdot S_{\triangle BEF} \cdot BB'$$

显然当 $S_{\triangle BEF}$ 最大时, $V_{B' - BEF}$ 最大, 于是:

$$\therefore S_{\triangle BEF} = \frac{1}{2} \cdot BF \cdot BE = \frac{1}{2} \cdot (a - y) \cdot x$$

$$\therefore x = y$$

$$\therefore S_{\triangle BEF} = \frac{1}{2} (-x^2 + ax) \quad x \in (0, a)$$

显然当 $x = y = 1/2$ 时, $V_{B' - BEF}$ 最大, 此时, $B'EF$ 和 BEF 均为等腰三角形, 后略。

深入分析:

有兴趣可以计算一下 $V_{E - FA'B'}$, 提示 $V_{E - FA'B'} = V_{B - FA'B'}$ 。再看一下 $V_{E - FA'B'}$ 什么时候取得最大值。

该题是典型的通过向量联系几何和代数的题目, 整体思路还是明显的, 通过向量将几何问题转化为代数问题, 剩下的就是函数了。

第二章 直线和圆的方程

本章介绍直线和圆的方程，并讨论其位置关系。

本章要点：

- 直线和圆。
- 位置关系。

2.1 直线的倾斜角与斜率

本节要点：

- 掌握倾斜角的概念；
- 掌握斜率的概念。

2.1.1 倾斜角和斜率

定义 2.1.1 (倾斜角和斜率). 我们将直线 l 的向上的方向和 x 轴正向所成的角称为倾斜角, 通常用希腊字母表示, 不难得到 $\alpha \in [0, \pi)$ 。我们称倾斜角的正切值为该斜线的斜率, 通常记作 k , 即:

$$k := \tan \alpha$$

注意, 垂直于 x 轴的直线是没有斜率的。

想两个问题。首先, 既然有了倾斜角, 为什么还要定义斜率? 其次, 既然要定义斜率, 为什么要用正切值? 用正弦或余弦不行吗? 何况正弦和余弦在 $\alpha \in [0, \pi)$ 上都有定义, 不会像正切一样存在不定义点。XML。

2.1.2 两条直线平行和垂直的判定

定理 2.1.1. 设直线 l_1, l_2 的斜率为 k_1, k_2 , 则有:

$$l_1 \parallel l_2 \Leftrightarrow k_1 = k_2$$

$$l_1 \perp l_2 \Leftrightarrow k_1 k_2 = -1$$

注意, 各自平行于两个轴的直线是没有办法用上述定理判断垂直的。

2.2 直线的方程

本节要点:

- 掌握 5 种直线的代数形式;
- 深刻理解它们代表的几何意义。

点斜式, k 必须有定义, 即不能垂直于 x 轴:

$$y - y_0 = k(x - x_0)$$

斜截式, k 必须有定义, 即不能垂直于 x 轴:

$$y = kx + b$$

两点式, 对两点有要求 $x_1 \neq x_2$ 且 $y_1 \neq y_2$:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

截距式, 任意一个截距都不能为 0:

$$\frac{x}{a} + \frac{y}{b} = 1$$

一般式, 任何情况都能用:

$$Ax + By + C = 0$$

2.2.1 拓展讨论: 直线的定义

其实我们可以参考上一章中对空间直线的描述, 定义平面上的直线。

定义 2.2.1 (直线). 若二维平面中有一个给定点 P_0 和一个给定的向量 \mathbf{n} , 若平面上的点和 P_0 构成的向量与 \mathbf{n} 平行, 则称这些点组成的集合为**直线**, 常用 l 表示, 即:

$$l := \left\{ P \mid \overrightarrow{P_0 P} = \lambda \mathbf{n} \right\}$$

若令 $P = (x, y), P_0 = (x_0, y_0), \mathbf{n} = A\mathbf{x} + B\mathbf{y}$, 则直线可以表示为下列的向量方程:

$$(x - x_0, y - y_0) = \lambda(A, B)$$

若保留 λ , 可得到参数方程:

$$\begin{cases} x = \lambda A + x_0 \\ y = \lambda B + y_0 \end{cases} \quad \lambda \in \mathbb{R}$$

若约去 λ , 则可得到一般式:

$$\frac{x-x_0}{A} = \frac{y-y_0}{B}$$

$$(-Bx) + (Ay) + (Bx_0 - Ay_0) = 0$$

2.2.2 习题

例 2.2.1 (综合运用 12, 难度: ★). 若直线 l 沿 x 轴向左平移 3 个单位长度, 再沿 y 轴向上平移 1 个单位长度后, 回到原来的位置, 试求直线 l 的斜率.

解一:

从定义出发, 设直线 $y - y_0 = k(x - x_0)$, 有:

$$\begin{aligned}\because y - y_0 - 1 &= k(x - x_0 + 3) \\ \therefore y - y_0 - 1 &= k(x - x_0) + 3k \\ \therefore k &= -\frac{1}{3}\end{aligned}$$

解二:

从向量的角度, 设点直线上的点 (x_0, y_0) , 平移后到达 $(x_0 - 3, y_0 + 1)$, 又这两点都在直线上, 于是:

$$\frac{(y_0 + 1) - y_0}{(x_0 - 3) - x_0} = k$$

本题考察斜率的定义。

2.3 直线的交点坐标与距离公式

本节要点:

- 从二元一次方程组的解出发理解直线的几何关系;
- 熟练掌握点到直线的距离公式;
- 掌握平行线间的距离公式。

点线距离公式:

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

线线距离公式:

$$d = \frac{|C_1 - C_2|}{\sqrt{A^2 + B^2}}$$

仔细阅读距离公式的推导过程。

2.3.1 习题

例 2.3.1 (综合运用 14, 难度: ★). 已知 $A(-3, -4), B(6, 3)$ 两点到直线 $l: ax + y + 1 = 0$ 的距离相等, 求 a 的值。

解一:

根据点线距离公式:

$$\begin{aligned} \because d_A &= \frac{|-3a - 4 + 1|}{\sqrt{a^2 + 1}} \\ \because d_B &= \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}} = \frac{|6a + 3 + 1|}{\sqrt{a^2 + 1}} \\ \therefore (-3a - 3)^2 &= (6a + 4)^2 \\ \therefore 27a^2 + 30a + 7 &= 0 \\ \therefore a &= -\frac{1}{3} \text{ or } -\frac{7}{9} \end{aligned}$$

解二:

根据几何关系, 若两点不在同一侧, 则线段 AB 的中点 $(\frac{3}{2}, -\frac{1}{2})$ 在直线上, 带入直线方程易得:

$$a = -\frac{1}{3}$$

或者两点在同侧, 易得向量 $\overrightarrow{AB} = (9, 7)$ 和直线平行, 于是:

$$k = -a = \frac{7}{9}$$

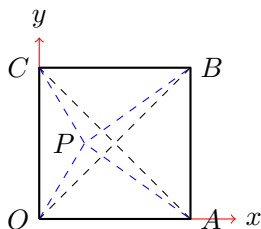
观察几何关系可以降低计算量。

例 2.3.2 (拓广探索 11, 难度: **). 已知 $0 < x < 1, 0 < y < 1$ 。

1. 求证: $\sqrt{x^2 + y^2} + \sqrt{x^2 + (1-y)^2} + \sqrt{(1-x)^2 + y^2} + \sqrt{(1-x)^2 + (1-y)^2} \geq 2\sqrt{2}$, 并求使等式成立的条件。
2. 说明上述不等式的几何意义。

解:

构建如下单位正方形 $OABC$, P 在正方形内部。



根据三角形三边长度关系易得:

- $\sqrt{x^2 + y^2} + \sqrt{(1-x)^2 + (1-y)^2} \geq \sqrt{2}$, 当且仅当 P 在直线 OB 上时等号成立;
- $\sqrt{x^2 + (1-y)^2} + \sqrt{(1-x)^2 + y^2} \geq \sqrt{2}$, 当且仅当 P 在直线 AC 上时等号成立;

所以, 当且仅当 P 在 OB, AC 的交点时求证的等式的等号成立。

使用几何可以非常容易证明此题。

2.4 圆的方程

本节要点:

- 掌握 2 种圆的代数形式;
- 深刻理解它们代表的几何意义。

标准方程:

$$(x-a)^2 + (y-b)^2 = r^2$$

一般方程:

$$x^2 + y^2 + Dx + Ey + F = 0$$

详细阅读所有例题。例 2 和例 4 都是已知求三角形的外接圆, 分别用标准方程和一般方程, 体会两者的计算量。例 5 中 M 是中点, 若 M 不是中点, 而是某个等比例点, 则 M 的运动轨迹是什么样子? 再如果 A 是任意轨迹, 则 M 的运动轨迹和 A 的轨迹有什么关系? 早期钟表内部的细小零件的制作就是依据这个原理, XML。

2.4.1 拓展讨论: 圆的定义

我们给出圆的定义, 然后推导圆的公式。

定义 2.4.1 (圆). 设二维平面中有一定点 P_0 , 若平面上的点和 P_0 的距离都是 $r > 0$, 则称这些点组成的集合为圆, 常用 C 表示, 即:

$$C := \{P \mid |\overrightarrow{P_0P}| = r\}$$

若令 $P = (x, y), P_0 = (x_0, y_0)$, 则圆可以表示为下列的向量方程:

$$(x - x_0)^2 + (y - y_0)^2 = r^2$$

使用复数的概念, 可得到参数方程:

$$\begin{cases} x - x_0 = r \cos \theta \\ y - y_0 = r \sin \theta \end{cases} \quad \theta \in [0, 2\pi)$$

再考虑一个问题, 考察课本给出的圆的一般方程

$$x^2 + y^2 + Dx + Ey + F = 0$$

A, B, C 去哪里了?

事实上, 关于 x, y 的二次曲线, 更一般的应该是:

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0$$

是一个倾斜的椭圆, 当 $C = 0$ 时, 椭圆归正, 即长短轴平行于 xy 轴, 进一步地, 当 $A = B$ 时, 为圆。

2.4.2 习题

例 2.4.1 (综合运用 5, 难度: $\star\star$). 已知圆的一条直径的端点分别是 $A(x_1, y_1)$, $B(x_2, y_2)$, 求证此圆的方程是

$$(x - x_1)(x - x_2) + (y - y_1)(y - y_2) = 0$$

解:

用向量的方法, 向量表示的点在圆上, 则构成直角三角形, 两个直角边的内积为 0。

本题联合向量内积的几何意义, 一句话就能证明, 非常容易。

例 2.4.2 (综合运用 7, 难度: \star). 已知等腰三角形 ABC 的一个顶点为 $A(4, 2)$, 底边的一个端点为 $B(3, 5)$, 求底边的另一个端点 C 的轨迹方程, 并说明它是什么图形。

解:

显然 $AB = AC$, 但 A, B, C 不能构成直线, 于是

$$(x - 4)^2 + (y - 2)^2 = 10 \quad x \neq 3, 5$$

此题不难, 但需注意 x 的取值。

2.5 直线与圆、圆与圆的位置关系

本节要点：

- 从代数角度深刻理解线圆位置关系；
- 从代数角度深刻理解圆圆位置关系。

2.5.1 直线与圆的位置关系

我们可以根据线圆位置关系的定义总结出判定定理。

定理 2.5.1. 设二维平面有直线 l 和圆 C ，将直线方程带入圆方程得到一个关于 x 或 y 的一元二次方程，则有：

- $\Delta > 0 \Leftrightarrow$ 直线和圆相交；
- $\Delta = 0 \Leftrightarrow$ 直线和圆相切；
- $\Delta < 0 \Leftrightarrow$ 直线和圆相离。

2.5.2 圆与圆的位置关系

同上，我们可以总结出判定定理。

定理 2.5.2. 设二维平面有两个圆 C_1, C_2 ，将两个圆方程合并后得到一个关于 x 或 y 的一元二次方程，则有：

- $\Delta > 0 \Leftrightarrow$ 圆和圆相交；
- $\Delta = 0 \Leftrightarrow$ 圆和圆相切；
- $\Delta < 0 \Leftrightarrow$ 圆和圆相离。

这里需要注意：

- 两个圆的方程可能是一模一样，说明两个圆重合，该判定定理自然无法使用；
- 相切可能是外切，也可能是内切，需要结合圆心距离进一步判定；
- 相离也是有外部和内部两个可能，同样需要结合圆心距判定。

2.5.3 拓展讨论：两个圆方程相减的图形

设两圆 C_1, C_2 相交，将它们的一般方程相减必然得到一直线方程，也即两个交点确定一条直线，这点毋庸置疑，而且几何意义也十分清楚。

那么将两圆方程相加呢？由于二次项不能抵消，必然也是一个圆，用标

准方程推导可以直观地看清圆心和半径。设两圆 C_1, C_2 如下：

$$\begin{cases} (x - x_1)^2 + (y - y_1)^2 = r_1^2 \\ (x - x_2)^2 + (y - y_2)^2 = r_2^2 \end{cases}$$

相加，化简后：

$$\left(x - \frac{x_1 + x_2}{2}\right)^2 + \left(y - \frac{y_1 + y_2}{2}\right)^2 = \frac{r_1^2 + r_2^2}{2} - \frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{4}$$

其中 $(x_1 - x_2)^2 + (y_1 - y_2)^2$ 为圆心距，所以也可以表示为：

$$\left(x - \frac{x_1 + x_2}{2}\right)^2 + \left(y - \frac{y_1 + y_2}{2}\right)^2 = \frac{r_1^2 + r_2^2}{2} - \frac{C_1 C_2^2}{4}$$

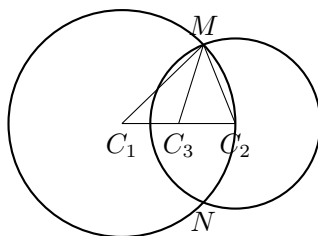
所以，两个圆方程相加的方程表示一个圆 C_3 ，圆心和半径如下：

$$(x_3, y_3) = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right)$$

$$r_3^2 = \frac{r_1^2 + r_2^2}{2} - \frac{C_1 C_2^2}{4} \quad C_1 C_2 \in [|r_1 - r_2|, r_1 + r_2]$$

$r_3 \in \left[\frac{|r_1 - r_2|}{2}, \frac{r_1 + r_2}{2}\right]$ ，分别当两圆内切和外切时，取到区间两侧。圆心非常好理解，半径较难理解，我们从向量的角度出发研究。

如下图，圆 C_1, C_2 交于 M, N ，新圆的圆心为 C_3 ：



$$\begin{aligned} \therefore |\overrightarrow{MC_3}| &= |\overrightarrow{MC_1} + \overrightarrow{MC_2}| / 2 \\ \therefore |\overrightarrow{MC_3}|^2 &= \frac{r_1^2 + r_2^2 + \overrightarrow{MC_1} \cdot \overrightarrow{MC_2}}{4} \\ \therefore |\overrightarrow{C_1 C_2}| &= |\overrightarrow{MC_2} - \overrightarrow{MC_1}| \\ \therefore |\overrightarrow{C_1 C_2}|^2 &= r_1^2 + r_2^2 - \overrightarrow{MC_1} \cdot \overrightarrow{MC_2} \\ \therefore |\overrightarrow{MC_3}|^2 &= \frac{r_1^2 + r_2^2 + \left(r_1^2 + r_2^2 - |\overrightarrow{C_1 C_2}|^2\right)}{4} \end{aligned}$$

2.6 本章小结

本章用代数的方法讨论了直线和圆。注意，我们有了第一章的知识，学习时更要注重数形结合。学习和做题时，必须从几何出发，在理解几何关系的基础上运用代数的方法。

2.6.1 习题

例 2.6.1 (拓广探索 20, 难度: ★). 已知圆 $C: (x-1)^2 + (y-2)^2 = 25$, 直线 $l: (2m+1)x + (m+1)y - 7m - 4 = 0$.

1. 求证: 直线 l 恒过定点。
2. 直线 l 被圆 C 截得的弦何时最长、何时最短? 并求截得的弦长最小时 m 的值以及最短弦长。

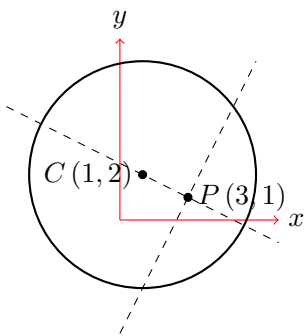
解一:

(1) 恒过定点, 表示取 $m_1 \neq m_2$ 生成两条直线, 交点必须和 m_1, m_2 无关, 联立直线方程:

$$\begin{cases} (2m_1+1)x + (m_1+1)y - 7m_1 - 4 = 0 \\ (2m_2+1)x + (m_2+1)y - 7m_2 - 4 = 0 \end{cases}$$

消元得 $(m_2 - m_1)y - (m_2 - m_1) = 0$, 可见只有当 $y = 1$ 时与 m_1, m_2 无关, 带入直线方程得 $(2m+1)x - 6m - 3 = 0$, 易得恒过 $(3, 1)$ 。

(2) 易得恒定点在园内, 于是弦有最短且垂直于直径方向, 否则园外点可以相切, 最短为 0, 最长弦自然是直径。略。



解二:

(1) 过恒定点的意思是有一点 (x_0, y_0) 必然在直线上, 而且不依赖于 m 的值, 也就是说该点坐标的表达式中不能出现 m , 将直线写成:

$$(2x_0 + y_0 - 7)m + x_0 + y_0 - 4 = 0$$

要使之不出现 m ，则有：

$$\begin{cases} 2x_0 + y_0 - 7 = 0 \\ x_0 + y_0 - 4 = 0 \end{cases}$$

可得 $(x_0, y_0) = (3, 1)$ 。

理解“恒过一点”的几何含义，本题没有难度。

第三章 圆锥曲线的方程

本章介绍 3 类二次曲线——椭圆、双曲线和抛物线。它们都可以通过圆锥获得，所以通称圆锥曲线，事实上，圆也是一类圆锥曲线。

本章要点：

- 椭圆。
- 双曲线。
- 抛物线。

3.1 椭圆

本节要点：

- 掌握椭圆的概念；
- 从代数角度掌握椭圆的几何性质。

3.1.1 椭圆及其标准方程

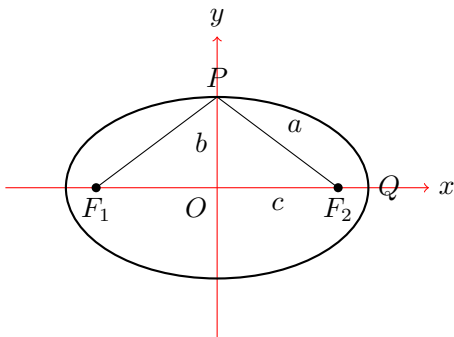
定义 3.1.1 (椭圆). 设二维平面中有两个定点 F_1, F_2 , 若平面上的点和 F_1, F_2 的距离之和都是 $r > 0$, 则称这些点组成的集合为椭圆, 常用 E 表示, 即:

$$E := \left\{ P \mid \left| \overrightarrow{F_1 P} \right| + \left| \overrightarrow{F_2 P} \right| = r \right\}$$

其中, F_1, F_2 称为**焦点**, $\left| \overrightarrow{F_1 F_2} \right|$ 称为**焦距**。若令 $F_1 = (-c, 0), F_2 = (c, 0)$, 则椭圆可以表示为下列的**标准方程**:

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

其中 a, b, c 满足 $a^2 - b^2 = c^2$ 。



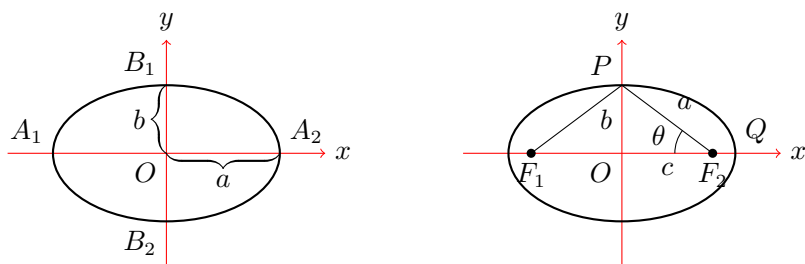
详细阅读课本关于椭圆标准方程的推导过程, 体会两次平方的意义。
详细阅读例 2、例 3、例 6, 可以认为是椭圆的另外三种定义。

3.1.2 椭圆的简单几何性质

设有如下椭圆, 我们称:

- O, A_1, A_2, B_1, B_2 : 一个中心和四个顶点;
- $A_1 A_2 = 2a$: 长轴;

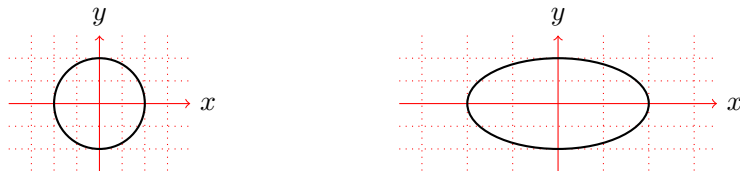
- $B_1B_2 = 2b$: 短轴;
- $F_1F_2 = 2c$: 焦距;
- $e = \frac{c}{a} = \cos \theta \in [0, 1)$: 离心率, e 越大椭圆越扁平, $e = 0$ 时退化为圆, $e \rightarrow 1$ 时变成两条平行线 $y = \pm b$ 。



3.1.3 拓展讨论：坐标系变换

对课本例 2 深入讨论。

椭圆可以通过圆的拉伸获得，如下左图，圆方程为 $x^2 + y^2 = 2^2$ ：



我们进行坐标变换，将 x 轴拉长一倍，如上右图：

$$\begin{cases} u = 2x \\ v = y \end{cases} \Rightarrow \begin{cases} x = 0.5u \\ y = v \end{cases}$$

则有：

$$\frac{u^2}{4^2} + \frac{v^2}{2^2} = 1$$

可见，将坐标轴拉伸就可将圆变成椭圆，反之，将坐标轴压缩就可将椭圆变成圆。

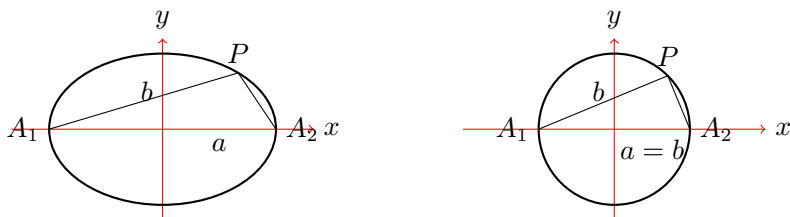
3.1.4 拓展讨论：斜率之积

对课本例 3 深入讨论。

设椭圆 E 及椭圆上一点 P , 如下左图, 则 PA_1, PA_2 的斜率之积有:

$$\begin{aligned} \because \frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1 \\ \therefore k_{PA_1} k_{PA_2} &= \frac{y_P}{x_P + a} \cdot \frac{y_P}{x_P - a} = \frac{b^2 \left(1 - \frac{x_P^2}{a^2}\right)}{x_P^2 - a^2} = -\frac{b^2}{a^2} = -\tan^2 \theta \end{aligned}$$

特别地, 当 $a = b$ 时, $k_{PA_1} k_{PA_2} = -1$ 即 $PA_1 \perp PA_2$, 此时椭圆退化为圆, 如下右图。



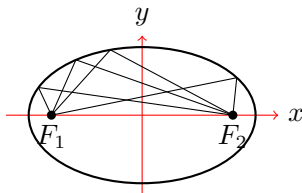
可见:

- $k_1 k_2 \in (-\infty, -1)$: 长条形椭圆;
- $k_1 k_2 = -1$: 圆;
- $k_1 k_2 \in (-1, 0)$: 扁宽形椭圆。

3.1.5 拓展讨论: 两个焦点

对课本例 5 深入讨论。

例 5 说明了从任意一个焦点出发的光线都能聚焦到另一个焦点, 证明这点需要微积分的知识, 略。但我们可以思考, 当两个焦点重合的时候, 从焦点出发的任意光线都能返回到焦点, 这是一个圆, 而且圆的半径总是垂直于切线。

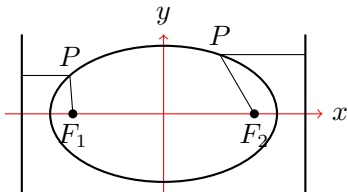


有些音乐厅就设计成椭圆, 乐队位于一个焦点附近, VIP 位于另一个焦点附近, 使得 VIP 在听觉上能位于乐队中。

3.1.6 拓展讨论：动点距离直线

对课本例 6 深入讨论。

我们考察，对于任意一个椭圆 E ，是否必有垂直坐标轴的直线 $x = x_0$ ，且椭圆上的点 P 距离焦点和直线的比是定值 k 。



距离焦点和直线的比：

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

$$k = \frac{\sqrt{(x \pm c)^2 + y^2}}{|x - x_0|} \Rightarrow k^2(x - x_0)^2 = (x \pm c)^2 + y^2$$

化简后得：

$$\left(1 - k^2 - \frac{b^2}{a^2}\right)x^2 + 2(k^2x_0 \pm c)x + (a^2 - k^2x_0^2) = 0$$

若要对所有 x 都成立，必须：

$$\begin{cases} 1 - k^2 - \frac{b^2}{a^2} = 0 \\ k^2x_0 \pm c = 0 \\ a^2 - k^2x_0^2 = 0 \end{cases} \Rightarrow \begin{cases} k = \frac{c}{a} \in (0, 1) \\ x_0 = \pm \frac{a^2}{c} \end{cases}$$

不难发现特点：

- 对于左焦点有左边的直线 $x = -\frac{a^2}{c}$ ，对于右焦点有右边的直线 $x = \frac{a^2}{c}$ ，且两条直线一定在椭圆外侧；
- 动点距离焦点更近；
- 当 $c = 0$ 时椭圆退化为圆，且 $x_0 = \pm\infty, k = 0$ ，表示直线为无穷远处，自行脑补。

3.1.7 拓展讨论：椭圆的一般方程

教材中对于直线和圆都有一般形式的方程，为什么椭圆就没有？我们来推导一下椭圆方程的一般形式，从最一般的二次曲线开始，如下：

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0$$

A, D 和 B, E 可以凑出 x, y 的平方项, 一方面只影响 F , 另一方面 x, y 的平方项可以通过坐标平移解决, 所以我们可以之分析如下形式:

$$Ax^2 + By^2 + Cxy + F = 0$$

易得该方程对应的曲线关于原点对称。

我们先考察另一个问题, 将椭圆绕中心点沿顺时针转 α 后方程是什么样子的。设椭圆如下:

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

转换为极坐标形式:

$$\frac{(r \cos \theta)^2}{a^2} + \frac{(r \sin \theta)^2}{b^2} = 1$$

将椭圆绕中心点沿顺时针转 α :

$$\frac{[r \cos(\theta - \alpha)]^2}{a^2} + \frac{[r \sin(\theta - \alpha)]^2}{b^2} = 1 \quad \alpha \in [0, \pi)$$

注意我们这里是顺时针转动图形, 相当于逆时针转动坐标系, 所以是 $-\alpha$ 而不是 $+\alpha$, 展开后:

$$(b^2 \cos^2 \alpha + a^2 \sin^2 \alpha) x^2 + (b^2 \sin^2 \alpha + a^2 \cos^2 \alpha) y^2 - c^2 \sin 2\alpha xy = a^2 b^2$$

$$Ax^2 + By^2 + Cxy + F = 0$$

$$\begin{cases} A = b^2 \cos^2 \alpha + a^2 \sin^2 \alpha = b^2 + \frac{c^2}{2} (1 - \cos 2\alpha) \\ B = b^2 \sin^2 \alpha + a^2 \cos^2 \alpha = b^2 + \frac{c^2}{2} (1 + \cos 2\alpha) \\ C = -c^2 \sin 2\alpha \\ F = -a^2 b^2 \end{cases}$$

不难发现:

- $F < 0$, 若 $F > 0$ 则曲线不存在, 若 $F = 0$ 则有可能是一点有可能是直线;
- $A, B > 0$, 否则变成双曲线, 且有 $A + B = a^2 + b^2$, 当 $A = B$ 时 $\alpha = \pi/4$ 或 $3\pi/4$;
- $A, F > 0$ 共同决定了椭圆和 x 轴的交点, $B, F > 0$ 共同决定了椭圆和 y 轴的交点。

特别地, 当 $\alpha = \pi/4$ 时, 方程为在圆的基础上加一个 xy 项, 如下:

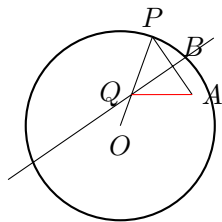
$$(a^2 + b^2) x^2 + (a^2 + b^2) y^2 - 2c^2 xy = 2a^2 b^2$$

当 $\alpha = \pi/2$ 时, 表现为长短轴互换, 如下:

$$a^2 x^2 + b^2 y^2 = a^2 b^2$$

3.1.8 习题

例 3.1.1 (复习巩固 6, 难度: ★). 如图, 圆 O 的半径为定长 r , A 是圆 O 内一个定点, P 是圆 O 上任意一点. 线段 AP 的垂直平分线 l 和半径 OP 相交于点 Q , 当点 P 在圆上运动时, 点 Q 的轨迹是什么? 为什么?



解:

连接 QA , 有:

$$QO + QA = QO + QP = r$$

仔细观察图形, 本题没有难度。不要盲目建立坐标系, 椭圆的中心未知。

例 3.1.2 (拓广探索 14, 难度: ★). 已知椭圆 $\frac{x^2}{4} + \frac{y^2}{9} = 1$, 一组平行直线的斜率是 $\frac{3}{2}$ 。

1. 这组直线何时与椭圆有两个公共点?
2. 当它们与椭圆有两个公共点时, 证明这些直线被椭圆截得的线段的中点在同一条直线上。

解:

(1) 略。

(2) 一般的做法是假设直线 $y = \frac{3}{2}x + c$, 带入椭圆求交点:

$$18x^2 + 12cx + 4c^2 - 36 = 0$$

$$x_{1,2} = \frac{-12c \pm \sqrt{144c^2 - 4 \cdot 18 \cdot (4c^2 - 36)}}{36}$$

得到线段中点:

$$\begin{cases} x_0 = -\frac{1}{3}c \\ y_0 = \frac{1}{2} \end{cases}$$

是一个参数方程, 就是一条过原点, 斜率是 $-3/2$ 的直线。

但是，理解了坐标系变换，可以先压缩 xy 两个轴，使得椭圆变成一个标注圆，此时，这一组平行线将变成：

$$y = x + c$$

不难发现割线的中点是直线 $y = -x$ ，最后通过拉伸 xy 两个轴变回来，略。

本题没有难度，但从坐标系变换的角度分析，还是有点意思的。

3.2 双曲线

本节要点：

- 掌握双曲线的概念；
- 从代数角度掌握双曲线的几何性质。

3.2.1 双曲线及其标准方程

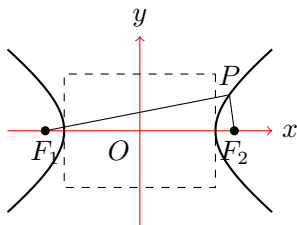
定义 3.2.1 (双曲线). 设二维平面中有两个定点 F_1, F_2 , 若平面上的点和 F_1, F_2 的距离之差的绝对值都是 $r > 0$, 则称这些点组成的集合为双曲线, 常用 H 表示, 即:

$$H := \left\{ P \mid \left| \left| \overrightarrow{F_1 P} \right| - \left| \overrightarrow{F_2 P} \right| \right| = r \right\}$$

其中, F_1, F_2 称为焦点, $\left| \overrightarrow{F_1 F_2} \right|$ 称为焦距。若令 $F_1 = (-c, 0), F_2 = (c, 0)$, 则双曲线可以表示为下列的标准方程:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

其中 a, b, c 满足 $a^2 + b^2 = c^2$ 。



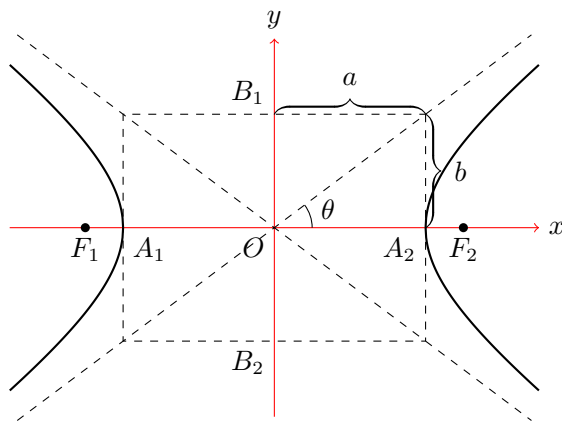
详细阅读课本关于双曲线标准方程的推导过程, 体会两次平方的意义。详细阅读例 6, 可以认为是双曲线的另外的定义。

3.2.2 双曲线的简单几何性质

双曲线图形如下, 我们称:

- O, A_1, A_2, B_1, B_2 : 一个中心和四个顶点;
- $A_1 A_2 = 2a$: 实轴;
- $B_1 B_2 = 2b$: 虚轴;
- $F_1 F_2 = 2c$: 焦距;

- $e = \frac{c}{a} = \frac{1}{\cos \theta} \in (1, +\infty)$: 离心率, e 越大双曲线开口越大;
- $y = \pm \frac{b}{a}x$: 渐近线。



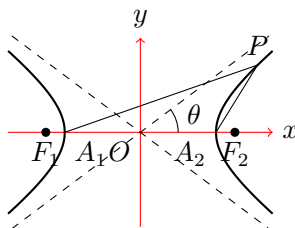
3.2.3 拓展讨论：坐标系变换

同椭圆，略。

3.2.4 拓展讨论：斜率之积

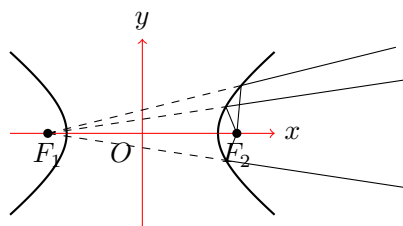
设双曲线 H 及双曲线上一点 P ，则 PA_1, PA_2 的斜率之积有：

$$\begin{aligned} \because \frac{x^2}{a^2} - \frac{y^2}{b^2} &= 1 \\ \therefore k_{PA_1} k_{PA_2} &= \frac{y_P}{x_P + a} \cdot \frac{y_P}{x_P - a} = \frac{b^2 \left(\frac{x_P^2}{a^2} - 1 \right)}{x_P^2 - a^2} = \frac{b^2}{a^2} = \tan^2 \theta \end{aligned}$$



3.2.5 拓展讨论：两个焦点

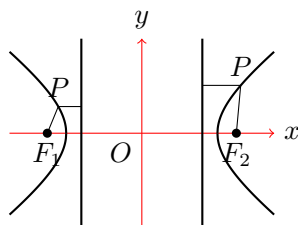
从任意一个焦点发出的光线都可以视作从另一个焦点发出，证明这点需要微积分的知识，略。



3.2.6 拓展讨论：动点距离直线

对课本例 5 深入讨论。

我们考察，对于任意一个椭圆 E ，是否必有垂直坐标轴的直线 $x = x_0$ ，且椭圆上的点 P 距离焦点和直线的比是定值 k 。



距离焦点和直线的比：

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

$$k = \frac{\sqrt{(x \pm c)^2 + y^2}}{|x - x_0|} \Rightarrow k^2(x - x_0)^2 = (x \pm c)^2 + y^2$$

化简后得：

$$\left(1 - k^2 + \frac{b^2}{a^2}\right)x^2 + 2(k^2x_0 \pm c)x + (a^2 - k^2x_0^2) = 0$$

若要对所有 x 都成立，必须：

$$\begin{cases} 1 - k^2 + \frac{b^2}{a^2} = 0 \\ k^2x_0 \pm c = 0 \\ a^2 - k^2x_0^2 = 0 \end{cases} \Rightarrow \begin{cases} k = \frac{c}{a} \in (1, +\infty) \\ x_0 = \mp \frac{a^2}{c} \end{cases}$$

不难发现特点：

- 对于左焦点有左边的直线 $x = -\frac{a^2}{c}$ ，对于右焦点有右边的直线 $x = \frac{a^2}{c}$ ，且两条直线一定在双曲线内侧；
- 动点距离直线更近。

3.2.7 拓展讨论：双曲线的一般方程

我们来推导一下双曲线方程的一般形式。中心点平移的讨论见椭圆，我们直接从旋转开始。双曲线的采用极坐标的形式，双曲线绕中心点沿顺时针转 α 后的方程如下：

$$\frac{[r \cos(\theta - \alpha)]^2}{a^2} - \frac{[r \sin(\theta - \alpha)]^2}{b^2} = 1 \quad \alpha \in [0, \pi)$$

注意我们这里是顺时针转动图形，相当于逆时针转动坐标系，所以是 $-\alpha$ 而不是 $+\alpha$ ，展开后：

$$(b^2 \cos^2 \alpha - a^2 \sin^2 \alpha) x^2 + (b^2 \sin^2 \alpha - a^2 \cos^2 \alpha) y^2 + c^2 \sin 2\alpha xy = a^2 b^2$$

$$Ax^2 + By^2 + Cxy + F = 0$$

$$\begin{cases} A = b^2 \cos^2 \alpha - a^2 \sin^2 \alpha = \frac{c^2}{2} (1 + \cos 2\alpha) - a^2 \\ B = b^2 \sin^2 \alpha - a^2 \cos^2 \alpha = \frac{c^2}{2} (1 - \cos 2\alpha) - a^2 \\ C = c^2 \sin 2\alpha \\ F = -a^2 b^2 \end{cases}$$

不难发现：

- $F < 0$ ，若 $F > 0$ 则曲线不存在，若 $F = 0$ 则有可能是一点有可能是直线，这点同椭圆；
- $A + B = b^2 - a^2, A - B = c^2 \cos 2\alpha$ ，当 $A = B$ 时 $\alpha = \pi/4$ 或 $3\pi/4$ 。

特别地，当 $\alpha = \pi/4$ 时，方程为在圆的基础上加一个 xy 项，如下：

$$\frac{b^2 - a^2}{2} x^2 + \frac{b^2 - a^2}{2} y^2 + c^2 xy = a^2 b^2$$

当 $\alpha = \pi/2$ 时，表现为实轴虚轴互换，如下：

$$-a^2 x^2 + b^2 y^2 = a^2 b^2$$

3.2.8 拓展讨论： $y = \frac{1}{x}$ 和 $y = x + \frac{1}{x}$

继续上面的讨论，若有 $a = b = \sqrt{2}$ ，则双曲线为：

$$\frac{x^2}{2} - \frac{y^2}{2} = 1$$

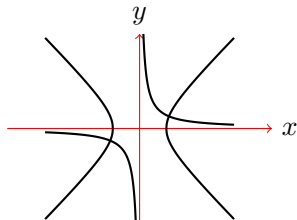
显然 $y = \pm x$ 是它的两条渐近线。当我们沿顺时针旋转 $\pi/4$ 后：

$$\frac{b^2 - a^2}{2} x^2 + \frac{b^2 - a^2}{2} y^2 + c^2 xy = a^2 b^2$$

$$\because a = b = \sqrt{2}, c = 2$$

$$\therefore xy = 1$$

这也是为什么之前碰到的 $y = \frac{1}{x}$ 也称为双曲线的原因。



对于 $y = x + \frac{1}{x}$, 两边同乘以 x 可得:

$$x^2 - xy + 1 = 0$$

其实就是压缩了开口, 再沿顺时针转 $3\pi/8$, 具体分析略, 可借助 GeoGebra 研究其变化。

3.2.9 习题

例 3.2.1 (综合运用 11, 难度: ★). M 是一个动点, MA 与直线 $y = x$ 垂直, 垂足 A 位于第一象限, MB 与直线 $y = -x$ 垂直, 垂足 B 位于第四象限。若四边形 $OAMB$ (O 为原点) 的面积为 3, 求动点 M 的轨迹方程。

解:

本题一步步解不难, 略。实际为图象 $xy = 3$ 绕原点按顺时针转 $\pi/4$, 自然是双曲线。

我们要考察的是若这两条线为任意的 $y_1 = k_1x, y_2 = k_2x$, 以动点和原点加上这两条直线组成的平行四边形 $OAMB$ 的面积为定值 S , 动点 M 的轨迹方程是否为双曲线。根据上面的讨论, 其实是双曲线, $y_1 = k_1x, y_2 = k_2x$ 是两条渐近线。

本题本身没有难度, 但之后我们的思考很有意思。

例 3.2.2 (拓广探索 13, 难度: ★). 已知双曲线 $x^2 - \frac{y^2}{2} = 1$, 过点 $P(1, 1)$ 的直线 l 与双曲线相交于 A, B 两点, P 能否是线段 AB 的中点? 为什么?

解:

易得 P 在双曲线中间, 于是可设直线:

$$y - 1 = k(x - 1)$$

带入双曲线方程,求得交点坐标:

$$x = \frac{k(1-k) \pm \sqrt{k^2(1-k)^2 + (2-k^2) \cdot [(1-k)^2 + 2]}}{(2-k^2)}$$

若有焦点,则通过中心点求得 k :

$$x = \frac{k(1-k)}{2-k^2} = 1 \Rightarrow k = 2$$

验证此时:

$$\Delta = 4 + (-2) \cdot 3 < 0$$

实际无交点,所以没有这样的直线。

本题没有难度。

例 3.2.3 (拓广探索 14, 难度: ★★). 一直双曲线 $\frac{x^2}{4} - \frac{y^2}{16} = 1$ 与直线 $l: y = kx + m, k \neq \pm 2$ 有唯一的公共点 M , 过点 M 且与 l 垂直的直线分别交 x 轴、 y 轴于 $A(x, 0), B(0, y)$ 两点。当点 M 运动时, 求点 $P(x, y)$ 的轨迹方程, 并说明轨迹是什么曲线。如果推广到一般双曲线, 能得到什么相应的结论?

解:

将直线带入双曲线得到:

$$(4 - k^2)x^2 - 2kmx - (m^2 + 16) = 0$$

根据题意要求:

$$\Delta = 16(m^2 - 4k^2 + 16) = 0 \Rightarrow m^2 = 4(k^2 - 4)$$

$$x_{1,2} = \frac{km \pm 2\sqrt{m^2 - 4k^2 + 16}}{4 - k^2}$$

得到 M 坐标:

$$M = (x_M, y_M) = \left(-\frac{4k}{m}, -\frac{16}{m}\right)$$

于是过 M 垂直于 l 的直线:

$$y + \frac{16}{m} = -\frac{1}{k} \left(x + \frac{4k}{m}\right)$$

P 坐标:

$$P = (x_{y=0}, y_{x=0}) = \left(-\frac{20k}{m}, -\frac{20}{m}\right)$$

是一个参数方程，结合 $\Delta = 0$ 可化简为：

$$\frac{x^2}{100} - \frac{y^2}{25} = 1$$

题中的一个交点显然 l 是双曲线的切线，又由于 l 是通过斜距式给出，所以 $k \in (-\infty, -2) \cup (2, +\infty)$ ，所以 P 的轨迹是双曲线但是不包括顶点。

对于一般的双曲线方程，有：

$$H: \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

$$H \cap l: (b^2 - a^2 k^2) x^2 - 2a^2 k m x - (a^2 m^2 + a^2 b^2) = 0$$

$$x_{1,2} = \frac{a^2 k m \pm ab \sqrt{b^2 + m^2 - a^2 k^2}}{b^2 - a^2 k^2} \text{ 且 } -m^2 = b^2 - a^2 k^2$$

$$M = (x_M, y_M) = \left(-\frac{a^2 k}{m}, -\frac{b^2}{m} \right)$$

$$l_{\perp}: y + \frac{b^2}{m} = -\frac{1}{k} \left(x + \frac{a^2 k}{m} \right)$$

$$P = \left(-\frac{(a^2 + b^2)k}{m}, -\frac{a^2 + b^2}{m} \right)$$

本题思路明显，但是计算量大。

3.3 抛物线

本节要点：

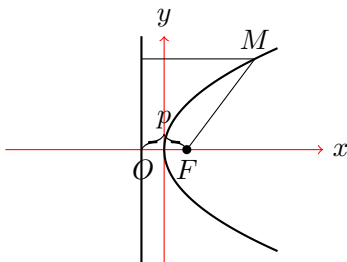
- 掌握抛物线的概念；
- 从代数角度掌握抛物线的几何性质。

3.3.1 抛物线及其标准方程

定义 3.3.1 (抛物线). 设二维平面中有两个定点 F 和直线 l , 设它们距离为 $p > 0$, 若平面上的点和 F, l 的距离相等, 则称这些点组成的集合为**抛物线**, 常用 P 表示, 若令 $F = (\frac{p}{2}, 0), l: x = -\frac{p}{2}$, 则抛物线可以表示为下列的标准方程:

$$y = 2px \quad p > 0$$

其中, F 称为**焦点**, l 称为**准线**。



详细阅读课本关于双曲线标准方程的推导过程, 体会两次平方的意义。详细阅读例 2、例 4、例 5。

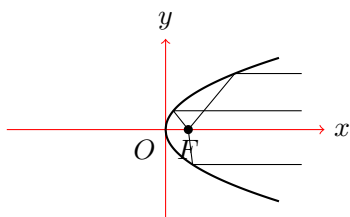
3.3.2 抛物线的简单几何性质

我们称：

- $y = 0$: 轴;
- $(0, 0)$: 顶点;
- $e = 1$: 离心率, 恒为 1。

3.3.3 拓展讨论：焦点

从焦点出发的光线经抛物线反射成平行光, 沿抛物线的轴入射的平行光汇聚于焦点, 证明这点需要微积分的知识, 略。



反射式望远镜就是由抛物面和双曲面构成，抛物面的焦点和双曲面的一个焦点重合。遥远的星体发出的光线可以认为是平行光，到达抛物面反射后汇聚于焦点，再由双曲面反射，形成的虚像位于双曲面的另一个焦点，这样，星体就像位于双曲面的另一个焦点一样，使得星体看起来更近了，详见教材本章小结的拓广探索 15。

对于焦点，我们还可以这么思考。设有 F_1, F_2 两个焦点，若从 F_1 射出的光线经曲线反射汇聚于 F_2 ，则曲线构成椭圆。当 F_2 逐渐远离 F_1 ，椭圆变得越来越扁平，当 F_2 置于无穷远处时，光线经曲线反射称为平行线，此时椭圆变成了抛物线。对于离心率：

$$\lim_{c \rightarrow a} e = \lim_{c \rightarrow a} \frac{c}{a} = 1$$

3.3.4 拓展讨论：抛物线的一般方程

采用极坐标的形式，抛物线绕顶点沿顺时针转 α 后的方程如下：

$$y^2 = 2px$$

$$[r \sin(\theta - \alpha)]^2 = 2pr \cos(\theta - \alpha)$$

$$\cos^2 \alpha \cdot y^2 - \sin 2\alpha \cdot xy + \sin^2 \alpha \cdot x^2 = 2p \cos \alpha \cdot x + 2p \sin \alpha \cdot y$$

注意我们这里是顺时针转动图形，相当于逆时针转动坐标系，所以是 $-\alpha$ 而不是 $+\alpha$ 。对比椭圆和双曲线，不难发现，抛物线的特征是有关于 x, y 的一次项。

3.4 本章小结

本章讨论了三类圆锥曲线，椭圆、双曲线和抛物线，它们都是由一个平面切圆锥构成，所以统称为圆锥曲线。学习时要多画图，多从几何理解曲线的方程。

第四章 数列

本章介绍数列。

本章要点：

- 等差数列。
- 等比数列。

数列可以看出是函数的离散化，所以学习过程中特别要注意和函数的联系。

4.1 数列的概念

本节要点:

- 掌握数列的概念;
- 掌握数列前 n 项和的概念。

数列及其前 n 项的和没有难度。反复体会本节第 2 个思考题, 如何从 S_n 推导 a_n , 具有典型意义。

注意两点:

- 数列可以看作是函数的离散化, 所以所有对函数的考察和工具, 都应用于数列;
- 在微积分中, 我们一般不关心 a_n , 而是关心当 $n \rightarrow \infty$ 时 S_n 的敛散性, 称为级数, 因为级数可以描述一个数。

4.1.1 习题

例 4.1.1 (综合运用 5, 难度: \star). 传说古希腊毕达哥拉斯学派的数学家用沙砾和小石子来研究数。他们根据沙砾或小石子所排列的形状把数分成许多类, 如图中第一行的 1, 3, 6, 10 称为三角形数, 第二行的 1, 4, 9, 16 称为正方形数, 第三行的 1, 5, 12, 22 称为五边形数。请你分别写出三角形数、正方形数、五边形数所构成的数列的第 5 项和第 6 项。

解:

三角形数。观察图形发现规律:

$$a_n - a_{n-1} = n$$

也就是说它们的差构成公差为 1 的等差数列。于是, 可另 $b_n = n$, 得:

$$a_n = S_n = \sum_{i=1}^n b_i = \frac{n(1+n)}{2}$$

正方形数。不难发现是正方形的面积, 于是:

$$a_n = n^2$$

五边形数。令 $\Delta a_n = a_n - a_{n-1}$, 不难发现

$$\Delta a_n = 3(n-1) + 1$$

也是一个等比数列, 于是

$$a_n = S_n = \sum_{i=1}^n \Delta a_i = \frac{n[1 + 3(n-1) + 1]}{2} = \frac{n(3n-1)}{2}$$

本题只是求解具体的项，没有难度。若是求通项公式，需要仔细观察。

例 4.1.2 (拓广探索 7, 难度: ★). 已知函数 $f(x) = \frac{2^x-1}{2^x}, x \in \mathbb{R}$, 设数列 $\{a_n\}$ 的通项公式为 $a_n = f(n), n \in \mathbb{N}^+$.

1. 求证 $a_n \geq \frac{1}{2}$ 。
2. $\{a_n\}$ 是递增数列还是递减数列? 为什么?

解:

(1、2) 易得 $a_n = \frac{2^n-1}{2^n} = 1 - \left(\frac{1}{2}\right)^n$, 可令 $n_1 < n_2$

$$a_{n_1} - a_{n_2} = \left[1 - \left(\frac{1}{2}\right)^{n_1}\right] - \left[1 - \left(\frac{1}{2}\right)^{n_2}\right] = \left(\frac{1}{2}\right)^{n_2} - \left(\frac{1}{2}\right)^{n_1} < 0$$

可见 $a_n \geq \frac{1}{2}$ 递增, 于是

$$a_n \geq a_1 = \frac{1}{2}$$

本题没有难度, 但本题是一个很好的数列和函数关系的例子。

4.2 等差数列

本节要点：

- 掌握等差数列的概念；
- 掌握等差数列的通项公式和求和公式。

4.2.1 等差数列的概念

$$a_n = a_1 + (n-1)d \quad n = 1, 2, 3, \dots$$

需要注意：

- 确定一个等差数列需要解决两个未知量，所以有两个条件就能求解一个等差数列；
- 没有说 $d > 0$ ，事实上， d 可以是任意实数；
- 等差数列有 $a_n = (a_{n-1} + a_{n+1})/2$ ，对应了算术平均；
- 等差数列可以视作直线 $f(x) = kx + c$ 的离散化， d 可视为“斜率”。

定理 4.2.1. 设两个等差数列 $\{a_n\}, \{b_n\}$ ，则 $\{pa_n + qb_n\}$ 也是等差数列。

4.2.2 等差数列的前 n 项和公式

$$S_n = \frac{n(a_1 + a_n)}{2} = na_1 + \frac{n(n-1)}{2}d \quad n = 1, 2, 3, \dots$$

反复体会求和公式的推导过程。

或许写成如下方式更能发现问题：

$$S_n = \frac{d}{2}n^2 + \left(a_1 - \frac{d}{2}\right)n$$

$$\frac{S_n}{n} = \frac{d}{2}n + \left(a_1 - \frac{d}{2}\right)$$

可见 S_n 是一条二次曲线，其次 $\frac{S_n}{n}$ 是一个等差数列。如果写成如下形式：

$$a_n = \frac{S_n - S_{n-1}}{n - (n-1)}$$

体会“ a_n 是 S_n 在 n 点的斜率”这个说法，这就是差分方程的雏形。

4.2.3 习题

例 4.2.1 (综合运用 8, 难度: ★). 已知两个等差数列 $2, 6, 10, \dots, 190$ 及 $2, 8, 14, \dots, 200$, 将这两个等差数列的公共项按从小到大的顺序组成一个新数列. 求这个新数列的各项之和.

解:

易得:

$$\begin{aligned}a_n &= 2 + 4(n-1) & n &= 1, 2, \dots, 48 \\b_m &= 2 + 6(m-1) & m &= 1, 2, \dots, 34\end{aligned}$$

求解公共项:

$$\begin{aligned}\because 2 + 4(n-1) &= 2 + 6(m-1) \\ \therefore 2n + 1 &= 3m\end{aligned}$$

可见 $3m$ 是奇数, 考虑到 $2n+1$ 最大取到 97, 于是 m 取值:

$$m = 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31$$

新数列为首项 2 公差 12 的等差数列, 共 16 项, 于是:

$$16 \cdot 2 + \frac{16 \cdot (16-1)}{2} \cdot 12 = 1472$$

本题考察项的表达式。

4.3 等比数列

本节要点：

- 掌握等比数列的概念；
- 掌握等比数列的通项公式和求和公式。

4.3.1 等比数列的概念

$$a_n = a_1 q^{n-1} \quad n = 1, 2, 3, \dots$$

需要注意：

- 确定一个等比数列需要解决两个未知量，所以有两个条件就能求解一个等比数列；
- 没有说 $q > 0$ ，事实上， q 可以是任意实数；
- 等比数列有 $a_n = \sqrt{a_{n-1}a_{n+1}}$ ，对应了几何平均；
- 等比数列可以视作指数函数 $f(x) = a^x$ 的离散化， q 可视为“底数”。

定理 4.3.1. 设两个等比数列 $\{a_n\}, \{b_n\}$ ，则 $\{a_n b_n\}, \left\{\frac{a_n}{b_n}\right\}$ 也是等比数列。

4.3.2 等比数列的前 n 项和公式

$$S_n = \frac{a_1(1-q^n)}{1-q} = \frac{a_1 - a_n q}{1-q} \quad q \neq 1, n = 1, 2, 3, \dots$$

反复体会求和公式的推导过程。

或许写成如下方式更能发现问题：

$$\begin{aligned} S_n &= \frac{a_1 - a_1 q^n}{1-q} = \frac{a_1}{1-q} + \frac{-a_1}{1-q} q^n \\ S_n - \frac{a_1}{1-q} &= \frac{-a_1}{1-q} q^n \end{aligned}$$

S_n 减去一个偏移量后依然是等比数列，事实上，指数函数的积分依然是指数函数：

$$\int a^x dx = \frac{a^x}{\ln a} + C \quad a > 0, a \neq -1$$

4.3.3 习题

例 4.3.1 (复习巩固 3, 难度: ★). 求和:

$$1. (2 - 3 \times 5^{-1}) + (4 - 3 \times 5^{-2}) + \cdots + (2n - 3 \times 5^{-n})$$

$$2. 1 + 2x + 3x^2 + \cdots + nx^{n-1}$$

解一:

(1) 将式子化为:

$$[2 + 4 + \cdots + 2n] - \left[3 \times \left(\frac{1}{5}\right)^1 + 3 \times \left(\frac{1}{5}\right)^2 + \cdots + 3 \times \left(\frac{1}{5}\right)^n \right]$$

即一个等差数列和与一个等比数列和的差, 略。

(2)

$$T_1 = 1 + x + x^2 + \cdots + x^{n-1} = \frac{1 - x^n}{1 - x}$$

$$T_2 = x + x^2 + \cdots + x^{n-1} = \frac{x(1 - x^{n-1})}{1 - x} = \frac{x - x^n}{1 - x}$$

$$T_3 = x^2 + \cdots + x^{n-1} = \frac{x^2(1 - x^{n-2})}{1 - x} = \frac{x^2 - x^n}{1 - x}$$

⋮

$$T_n = x^{n-1} = \frac{x^{n-1} - x^n}{1 - x}$$

$$S_n = T_1 + T_2 + \cdots + T_n = \frac{\frac{1-x^n}{1-x} - nx^n}{1-x} = \frac{nx^{n+1} - (n+1)x^n + 1}{(1-x)^2}$$

解二:

(2) 仔细观察, 不难发现:

$$S_n = 1 + 2x + 3x^2 + 4x^3 + \cdots + nx^{n-1}$$

$$xS_n = x + 2x^2 + 3x^3 + \cdots + (n-1)x^{n-1} + nx^n$$

$$S_n - xS_n = 1 + x + x^2 + \cdots + x^{n-1} - nx^n = \frac{1 - x^n}{1 - x} - nx^n$$

后略。

本题仔细观察表达式的项即可。

例 4.3.2 (综合运用 7, 难度: ★★). 已知数列 $\{a_n\}$ 的首项 $a_1 = 1$, 且满足 $a_{n+1} + a_n = 3 \times 2^n$ 。

1. 求证: $\{a_n - 2^n\}$ 是等比数列。

2. 求数列 $\{a_n\}$ 的前 n 项和 S_n 。

解:

(1)

$$\frac{a_{n+1} - 2^{n+1}}{a_n - 2^n} = \frac{3 \times 2^n - a_n - 2 \times 2^n}{a_n - 2^n} = \frac{2^n - a_n}{a_n - 2^n} = -1$$

证毕。

(2) 令 $b_n = a_n - 2^n$, 则 $b_1 = a_1 - 2 = -1$, 于是:

$$S_{n,a} = S_{n,b} + S_{n,2^n} = \frac{(-1) \cdot [1 - (-1)^n]}{1 - (-1)} + \frac{2 \cdot [1 - 2^n]}{1 - 2}$$

虽然 a_n 不构成等差或等比, 但是其变形构成等比。

例 4.3.3 (综合运用 8, 难度: ★). 若数列 $\{a_n\}$ 的首项 $a_1 = 1$, 且满足 $a_{n+1} = 2a_n + 1$, 求数列 $\{a_n\}$ 的通项公式及前 10 项的和。

解:

$$\because a_{n+1} = 2a_n + 1$$

$$\therefore a_{n+1} + 1 = 2a_n + 2 = 2(a_n + 1)$$

可见 $\{a_n + 1\}$ 构成了等比数列, 略。

仔细观察条件, 不难得到 a_n 之间的关系。

例 4.3.4 (拓广探索 10, 难度: ★). 已知数列 $\{a_n\}$ 为等比数列, $a_1 = 1024$, 公比 $q = \frac{1}{2}$ 。若 T_n 是数列 $\{a_n\}$ 的前 n 项的积, 求 T_n 的最大值。

解:

$$\because a_n = 1024 \cdot \left(\frac{1}{2}\right)^{n-1}$$

$$\therefore T_n = 1024^n \cdot \left(\frac{1}{2}\right)^{0+1+\cdots+(n-1)} = 1024^n \cdot \left(\frac{1}{2}\right)^{\frac{n(n-1)}{2}} = 2^{\frac{-n^2+n}{2}+10n}$$

指数将乘除化成加减, 所以本题实则考察等差数列。

例 4.3.5 (拓广探索 11, 难度: ★). 已知数列 $\{a_n\}$ 的首项 $a_1 = \frac{3}{5}$, 且满足

$$a_{n+1} = \frac{3a_n}{2a_n+1}.$$

1. 求证: 数列 $\left\{\frac{1}{a_n} - 1\right\}$ 为等比数列。
2. 若 $\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \cdots + \frac{1}{a_n} < 100$, 求满足条件的最大整数 n 。

解:

(1)

$$\begin{aligned}\because a_{n+1} &= \frac{3a_n}{2a_n+1} \\ \therefore \frac{1}{a_{n+1}} &= \frac{2a_n+1}{3a_n} = \frac{2}{3} + \frac{1}{3a_n} \\ \therefore \frac{1}{a_{n+1}} - 1 &= \frac{1}{3a_n} - \frac{1}{3}\end{aligned}$$

(2) 注意到:

$$\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} + \cdots + \frac{1}{a_n} = \left(\frac{1}{a_1} - 1\right) + \cdots + \left(\frac{1}{a_n} - 1\right) + n$$

本题分析 a_n 需要仔细。

例 4.3.6 (拓广探索 12, 难度: ★). 已知数列 $\{a_n\}$ 为等差数列, $a_1 = 1, a_3 = 2\sqrt{2} + 1$, 前 n 项的和为 S_n , 数列 $\{b_n\}$ 满足 $b_n = \frac{S_n}{n}$, 求证:

1. 数列 $\{b_n\}$ 为等差数列;
2. 数列 $\{a_n\}$ 中点任意三项均不能构成等比数列。

解:

(1)

$$\begin{aligned}\because a_n &= 1 + \sqrt{2} \cdot (n-1) \\ \therefore S_n &= \frac{n[1 + 1 + \sqrt{2} \cdot (n-1)]}{2} = \frac{n[\sqrt{2}n + 2 - \sqrt{2}]}{2} \\ \therefore \frac{S_n}{n} &= \frac{\sqrt{2}n + 2 - \sqrt{2}}{2}\end{aligned}$$

(2) 任取三项 a_l, a_m, a_n , 且 $l < m < n$, 假设能构成等比数列, 则:

$$\begin{aligned}\because a_l \cdot a_n &= a_m^2 \\ \therefore [1 + \sqrt{2} \cdot (l-1)] \cdot [1 + \sqrt{2} \cdot (n-1)] &= [1 + \sqrt{2} \cdot (m-1)]^2 \\ \therefore 2 \cdot (l-1) \cdot (n-1) + \sqrt{2} \cdot (l+n-2) &= 2(m-1)^2 + 2\sqrt{2} \cdot (m-1)\end{aligned}$$

考察该等式，由于 l, m, n 为整数，且 $\sqrt{2}$ 是无理数，所以要使等式成立，必有：

$$\begin{cases} 2 \cdot (l-1) \cdot (n-1) = 2(m-1)^2 \\ \sqrt{2} \cdot (l+n-2) = 2\sqrt{2} \cdot (m-1) \end{cases}$$

稍加分析就可以知道不可能，略。

本题考察等比数列的性质。

4.4 数学归纳法

略。

归纳法是一种由个别到一般的推理方法，是人类从经验中获得知识的方法。由于对个别的考察缺乏普遍性，所以来讲，真实的个别并不能一定获得真实的结论。而数学归纳法因为推理过程严谨，所以必然能获得真实的结论。

4.5 本章小结

本章介绍数列，特别介绍了两个数列：等差数列和等比数列。本章有两个难点， a_n 一般不会直接是等差或等比，而是加上一个偏移量或倒数后构成等差或等比，其次要注意等差和等比的平均数。

4.5.1 习题

例 4.5.1 (综合运用 12, 难度: $\star\star$). 已知数列 $\{a_n\}$ 的前 n 项的和为 S_n , 且 $a_{n+1} = 2S_n + 2, n \in \mathbb{N}^+$ 。

1. 求数列 $\{a_n\}$ 的通项公式。
2. 在 a_n 与 a_{n+1} 之间插入 n 个数, 使这 $n+2$ 个数组成一个公差为 d_n 的等差数列, 在数列 $\{d_n\}$ 中是否存在 3 项 d_m, d_k, d_p (其中 m, k, p 成等差数列) 成等比数列? 若存在, 求出这样的 3 项; 若不存在, 请说明理由。

解:

(1) 先分析表达式:

$$\begin{aligned}\because a_{n+1} &= 2S_n + 2 & a_n &= 2S_{n-1} + 2 \\ \therefore a_{n+1} - a_n &= 2a_n \\ \therefore a_n &= a_1 \cdot 3^{n-1}\end{aligned}$$

可见是等比数列, 再求首项:

$$\begin{aligned}\because a_{n+1} &= 2S_n + 2 \\ \therefore a_2 &= 2a_1 + 2 \\ \therefore 3a_1 &= 2a_1 + 2 \\ \therefore a_1 &= 2\end{aligned}$$

最终得到 $a_n = 2 \cdot 3^{n-1}$ 。

(2)

$$\begin{aligned}
&\because a_n = 2 \cdot 3^{n-1}, a_{n+1} = 2 \cdot 3^n \\
&\therefore d_n = \frac{2 \cdot 3^n - 2 \cdot 3^{n-1}}{n+1} = \frac{4 \cdot 3^{n-1}}{n+1} \\
&\because d_m \cdot d_p = d_k^2 \\
&\therefore \frac{4 \cdot 3^{m-1}}{m+1} \cdot \frac{4 \cdot 3^{p-1}}{p+1} = \frac{4 \cdot 3^{k-1}}{k+1} \frac{4 \cdot 3^{k-1}}{k+1} \\
&\because m+p=2k \\
&\therefore \frac{3^{2k-2}}{mn+2k+1} = \frac{3^{2k-2}}{k^2+2k+1} \\
&\therefore mn = k^2
\end{aligned}$$

得到 $m+p=2k$ 且 $mp=k^2$, 联立两式得:

$$(m-p)^2 = 0$$

不存在。

本题难度不大, 但较为复杂。

第五章 一元函数的导数及其应用

本章介绍导数。导数是分析函数的一个系统化的工具，但不可偏重导数。

本章要点：

- 导数。

导数是最后的工具，别啥都一上来求导！

5.1 导数的概念及其意义

本节要点:

- 掌握导数的概念;
- 理解导数的几何意义。

5.1.1 变化率的问题

定义 5.1.1 (增量和变化率). 我们定义, 自变量的增量是两个自变量的差, 因变量的增量是两个因变量的差, 即:

$$\Delta x := x_1 - x_0$$

$$\Delta y := y_1 - y_0 = f(x_1) - f(x_0)$$

我们继续定义两个增量的比值为函数的变化率, 即:

$$\frac{\Delta y}{\Delta x} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

显然, 函数的变化率是一个和自变量及其增量区间有关的新函数。当取相同的 Δx 时, 变化率描述了函数的变化快慢。变化率越大, 说明函数在同等 Δx 下的变化越大。

5.1.2 导数的概念及其几何意义

下面我们考察函数在一个点上的变化率, 即当 $x_1 \rightarrow x_0$ (或 $\Delta x \rightarrow 0$), 函数的变化率的存在性和取值。

定义 5.1.2 (导数). 设函数 $f(x)$ 在点 x_0 的某邻域 $N(x_0)$ 内有定义, 若当 $\Delta x \rightarrow 0$ 时 $\Delta y/\Delta x$ 的极限存在, 则称函数 $f(x)$ 在 x_0 处可导, 并称此极限值为函数 $f(x)$ 在点 x_0 的导数 (derivative), 记为 $f'(x_0)$, 即:

$$f'(x_0) := \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} \quad \text{或} \quad f'(x_0) := \lim_{x_1 \rightarrow x_0} \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

也可用莱布尼兹 (Leibniz) 记号记为:

$$y'|_{x=x_0} \quad \text{或} \quad \left. \frac{dy}{dx} \right|_{x=x_0}$$

从定义上来讲, 函数 $f(x)$ 在 x_0 处的导数是一个极限, 是一个可计算的确定的数。具体还有左导数和右导数, 略。

定义 5.1.3 (导函数). 如果函数 $f(x)$ 在区间 D 内每一点都可导, 则称每个导数构成的新函数为 $f(x)$ 的**导函数**, 记为 $f'(x)$ (或 y')。显然, 导函数 $f'(x)$ 在 x_0 的值就是导数 $f'(x_0)$ 。

除非特别指明, 一般我们将导函数简称为导数。区间 (a, b) (或 $[a, b]$) 内所有可导函数的集合通常记作 $D(a, b)$ (或 $D[a, b]$)。所以 $f(x)$ 在区间 D 内可导也可记作 $f(x) \in D(a, b)$ (或 $f(x) \in D[a, b]$)。

求导是对函数的一种运算。运算对象是函数, 得到的结果是另一个函数。从集合角度, 求导是一个函数集合到另一个函数集合的映射。

在讨论导数的时候, 紧紧抓住导数的定义, 从定义出发。后续关于导数的定理和公式都是用定义证明和推导。

从几何上来讲, 导数的意义是切线的斜率。物理上, 导数描述了一个物理量的瞬时变化速度, 比方说路程的导数是速度, 速度的导数是加速度。导数为求解物理量的变化速度提供了强大的数学工具。

5.2 导数的运算

本节要点:

- 掌握基本初等函数的求导公式;
- 掌握导数的运算法则;
- 掌握复合函数的求导方法。

5.2.1 基本初等函数的导数

$$\begin{array}{ll}
 (C)' = 0 & (x^a)' = a \cdot x^{a-1} \\
 (a^x)' = a^x \ln a & (e^x)' = e^x \\
 (\log_a x)' = \frac{1}{x \ln a} & (\ln x)' = \frac{1}{x} \\
 (\sin x)' = \cos x & (\cos x)' = -\sin x \\
 (\tan x)' = \sec^2 x & (\cot x)' = -\csc^2 x \\
 (\sec x)' = \sec x \cdot \tan x & (\csc x)' = -\csc x \cdot \cot x \\
 (\arcsin x)' = \frac{1}{\sqrt{1-x^2}} & (\arccos x)' = -\frac{1}{\sqrt{1-x^2}} \\
 (\arctan x)' = \frac{1}{1+x^2} & (\operatorname{arccot} x)' = -\frac{1}{1+x^2} \\
 (\operatorname{sh} x)' = \operatorname{ch} x & (\operatorname{ch} x)' = \operatorname{sh} x
 \end{array}$$

以上基本公式中, 除了 $(C)'$, $(a^x)'$, $(\sin x)'$, 其余都可以从这三个公式导出。

计算 $(C)'$

$$(C)' = \lim_{\Delta x \rightarrow 0} \frac{C - C}{\Delta x} = 0$$

计算 $(a^x)'$

$$(a^x)' = \lim_{\Delta x \rightarrow 0} \frac{a^{x+\Delta x} - a^x}{\Delta x} = a^x \lim_{\Delta x \rightarrow 0} \frac{a^{\Delta x} - 1}{\Delta x} = a^x \ln a$$

计算 $(\sin x)'$

$$\begin{aligned}
 (\sin x)' &= \lim_{\Delta x \rightarrow 0} \frac{\sin(x + \Delta x) - \sin x}{\Delta x} \\
 &= \lim_{\Delta x \rightarrow 0} \frac{2 \sin \frac{\Delta x}{2} \cos \frac{2x + \Delta x}{2}}{\Delta x} = \lim_{\Delta x \rightarrow 0} \cos \frac{2x + \Delta x}{2} = \cos x
 \end{aligned}$$

5.2.2 导数的四则运算法则

$$\begin{aligned}(C \cdot f)' &= C \cdot f' \\(f \pm g)' &= f' \pm g' \\(f \cdot g)' &= f'g + fg' \\ \left(\frac{f}{g}\right)' &= \frac{f'g - fg'}{g^2} \\ \left(\frac{1}{g}\right)' &= -\frac{g'}{g^2}\end{aligned}$$

前两个公式表明导数运算是线性的。

5.2.3 简单的复合函数的导数

定理 5.2.1 (复合函数求导定理). 若函数 $y = y(u)$, $u = u(x)$ 在对应区间可导, 则复合函数 $y = y[u(x)]$ 在对应区间也可导, 且有:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

复合求导法则是接下去三种求导法则(反函数求导、隐函数求导、参数式函数求导)的基础。

5.2.4 拓展讨论: 反函数的求导

定理 5.2.2 (反函数求导定理). 若函数 $y = f(x)$ 在某区间内单调可导, 且 $y' \neq 0$, 则它的反函数 $x = f^{-1}(y)$ 在对应区间内也可导, 且有:

$$f'(x) = \frac{1}{(f^{-1})'(y)} \quad \text{或写成} \quad \frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}$$

证明.

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{1}{\frac{\Delta x}{\Delta y}} = \frac{1}{\lim_{\Delta y \rightarrow 0} \frac{\Delta x}{\Delta y}} = \frac{1}{(f^{-1})'(y)}$$

□

例 5.2.1. 设 $y = \arcsin x$, 求 y' 。

解:

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{(\sin y)'} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}} = \frac{1}{\sqrt{1 - x^2}}$$

5.2.5 拓展讨论：隐函数的求导

定理 5.2.3 (一元隐函数求导定理). 若方程 $F(x, y) = 0$ 确定唯一的单值连续可导函数 $y = f(x)$, 则有:

$$\frac{dy}{dx} = -\frac{F_x}{F_y}$$

其中:

- F_x : 表示 $F(x, y) = 0$ 对 x 求导, x 是自变量, y 是常量;
- F_y : 表示 $F(x, y) = 0$ 对 y 求导, y 是自变量, x 是常量。

该定理在多元函数微积分中证明, 这里只需知道如何计算一元隐函数的导数。

例 5.2.2. 假设方程 $e^y = e - xy$ 确定了函数 $y = f(x)$, 求 y' 。

解:

令 $F(x, y) := e^y - e + xy = 0$, 得:

$$\begin{aligned} \therefore \begin{cases} F_x = y \\ F_y = e^y + x \end{cases} \\ \therefore y' = -\frac{F_x}{F_y} = -\frac{y}{e^y + x} \end{aligned}$$

5.2.6 拓展讨论：参数式函数的求导

定理 5.2.4 (参数函数求导定理). 若参数方程

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

可导, $x = x(t)$ 存在可导的反函数, 且 $x'(t) \neq 0$, 则由该参数方程确定的函数 $y = f(x)$ 可导, 且有:

$$\frac{dy}{dx} = \frac{y'(t)}{x'(t)}$$

例 5.2.3. 若有摆线

$$\begin{cases} x = t - \sin t \\ y = 1 - \cos t \end{cases}$$

求 $t = \pi/2$ 处的切线方程。

解:

切线方程 $y = f'(x_0)x + [f(x_0) - f'(x_0)x_0]$, 可得:

$$\begin{aligned} \left. \frac{dy}{dx} \right|_{t=\pi/2} &= \left. \frac{dy}{dt} / \frac{dx}{dt} \right|_{t=\pi/2} = \left. \frac{\sin t}{1 - \cos t} \right|_{t=\pi/2} = 1 \\ y &= 1 \cdot x + \left[\left(1 - \cos \frac{\pi}{2} \right) - 1 \cdot \left(\frac{\pi}{2} - 1 \right) \right] = x + 2 - \frac{\pi}{2} \end{aligned}$$

5.2.7 拓展讨论: 高阶导数

定义 5.2.1 (二阶导数). 设函数 $f(x)$ 在点 x_0 的某邻域 $N(x_0)$ 内可导, 若极限

$$\lim_{\Delta x \rightarrow 0} \frac{f'(x_0 + \Delta x) - f'(x_0)}{\Delta x}$$

存在, 则称该极限值为 $f(x)$ 在点 x_0 处的二阶导数, 记作:

$$f''(x_0) \quad \text{或} \quad y''|_{x=x_0} \quad \text{或} \quad \left. \frac{d^2y}{dx^2} \right|_{x=x_0}$$

同样, 我们可以定义三阶导数、四阶导数、直至 n 阶导数。

5.2.8 习题

例 5.2.4 (综合运用 6, 难度: ★). 已知函数 $f(x)$ 满足 $f(x) = f'(\frac{\pi}{4}) \sin x - \cos x$, 求 $f(x)$ 在 $x = \frac{\pi}{4}$ 处的导数。

解:

对函数求导:

$$\begin{aligned} f'(x) &= f'(\frac{\pi}{4}) \cos x + \sin x \\ f'(\frac{\pi}{4}) &= f'(\frac{\pi}{4}) \cdot \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} \end{aligned}$$

本题没有难度。

5.3 导数在研究函数中的应用

本节要点:

- 掌握函数单调性的判断方法;
- 掌握函数极值的求解方法。

5.3.1 函数的单调性

定理 5.3.1. 若 $f(x) \in D(a, b)$, 则:

- $f'(x) \geq 0 (\leq 0) \Leftrightarrow f(x)$ 在 (a, b) 上单调增 (减);
- $f'(x) > 0 (< 0) \Leftrightarrow f(x)$ 在 (a, b) 上严格单调增 (减)。

推论 5.3.2. 若 $f(x) \in C[x_0, x] \cap D(x_0, x)$, 且 $f(x_0) = 0$, 则:

- 当 $x > x_0$ 时 $f'(x) > 0$, 则 $f(x) > 0$;
- 当 $x > x_0$ 时 $f'(x) < 0$, 则 $f(x) < 0$ 。

推论要表达的是当我们知道函数的某个起点和趋势, 我们就可以判断函数的在该点后的值域。

5.3.2 函数的极值与最大(小)值

定义 5.3.1 (极值和极值点). 设函数 $f(x)$ 在区间 D 上有定义, 若 $x_0 \in D$ 存在邻域 $N(x_0, \delta) \subseteq D$, 使得 $\forall x \in N$ 有:

$$f(x) < f(x_0) \quad \text{或} \quad f(x) > f(x_0)$$

则称 $f(x_0)$ 为 $f(x)$ 在邻域 N 上的一个极大值(或极小值, 统称为极值), 而点 x_0 称为 $f(x)$ 在邻域 N 上的极大值点(或极小值点, 统称为极值点)。如果 $f(x)$ 在区间 D 上可导, 则极值点 x_0 处必有 $f'(x_0) = 0$, 但 $f'(x_0) = 0$ 并不代表 x_0 为极值点, 我们称 $f'(x_0) = 0$ 时的 x_0 为 $f(x_0) = 0$ 的驻点。如果 $f(x_0) = 0$ 在 x_0 处不可导, 则 x_0 也有可能是极值点。

定理 5.3.3. 设 $f(x) = 0$ 在点 x_0 处连续, 在 $N(\hat{x}_0, \delta)$ 内可导, 则:

- 若 $\forall x \in (x_0 - \delta, x_0)$ 有 $f'(x) < 0$, $\forall x \in (x_0, x_0 + \delta)$ 有 $f'(x) > 0$, 则 $f(x_0)$ 为极小值;
- 若 $\forall x \in (x_0 - \delta, x_0)$ 有 $f'(x) > 0$, $\forall x \in (x_0, x_0 + \delta)$ 有 $f'(x) < 0$, 则 $f(x_0)$ 为极大值;
- 若 $\forall x \in N(\hat{x}_0, \delta)$ 有 $f'(x)$ 恒为正或恒为负, $f(x_0)$ 为非极值。

定理 5.3.4. 设 $f(x) = 0$ 在点 x_0 处存在二阶导数, 且 $f'(x) = 0$, 则:

- 若 $f''(x) > 0$, 则 $f(x_0)$ 为极小值;
- 若 $f''(x) < 0$, 则 $f(x_0)$ 为极大值;
- 若 $f''(x) = 0$, 不能判定。

定理 5.3.5. 设 $f(x) = 0$ 在点 x_0 处存在二阶及以上导数, 且

$$\begin{aligned} f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0 \\ f^{(n)}(x_0) \neq 0 \end{aligned}$$

则:

- n 为偶数时, 若 $f^{(n)}(x_0) > 0$ 则 $f(x_0)$ 为极小值, 若 $f^{(n)}(x_0) < 0$ 则 $f(x_0)$ 为极大值;
- n 为奇数时, $f(x_0)$ 非极值。

上述 3 个定理是对微分中值定理中的费马定理的推广, 费马定理给出的是极值点的必要条件, 这里给出了极值的充要条件。但必须注意, 这些定理的前提是可导, 不可导点也是潜在的极值点, 需要判断。

第 1 个定理可以从几何上理解, 通过描述一阶导数的走势判断极值。第 2 个定理通过二阶导数量化地描述了函数走势。第 3 个定理是第 2 个定理的推广。

5.3.3 拓展讨论: 目标函数优化

极值是一个小范围内的最值, 如果将考察区域扩大, 则对极值的考察就扩展为对最值的考察。最值点将会出现在极值点、不可导点和边界点上。工程上有很多类似考察最值的问题, 数学上归结为求解目标函数的最值问题, 有时又称为目标函数的最优化。

若 $f(x)$ 在 $[a, b]$ 内连续, (a, b) 内存在有限个不可导点, 则首先根据连续函数的性质, 必然存在最值, 且两个端点 a, b 、驻点和不可导点 x_i 都是最值可能的点, 将这些点的函数值计算就可以判断最值, 或者根据一阶导数判断最值。

例 5.3.1. 若 $f(x) = \sqrt[3]{(x^2 - 2x)^2}$, 求 $[0, 3]$ 上的最值。

解:

易得 $f(x)$ 连续, 考察一阶导数:

$$f'(x) = \frac{4}{3} \frac{x-1}{\sqrt[3]{x^2-2x}}$$

有 $x = 1$ 为驻点, $x = 2$ 为不可导点, 于是端点、驻点、不可导点集合:

$$\{0, 3, 1, 2\}$$

分别求解:

$$f(0) = 0 \quad f(3) = \sqrt[3]{9} \quad f(1) = 1 \quad f(2) = 0$$

于是可得最值:

$$f_{\min} = 0 \quad f_{\max} = \sqrt[3]{9}$$

例 5.3.2. 若要做一个容积为 V_0 的圆柱形储罐, 怎样设计用料最省。

解:

圆柱形储罐容积 $V = \pi r^2 h$, 用料为表面积:

$$S(r) = 2\pi r^2 + 2\pi r h = 2\pi r^2 + 2\pi r \frac{V}{\pi r^2} = 2\pi r^2 + 2\frac{V_0}{r}, r \in (0, +\infty)$$

即求 $S(r)$ 的最值, 考察一阶导数:

$$S'(r) = 4\pi r - 2V_0 \frac{1}{r^2} = \frac{4\pi r^3 - 2V_0}{r^2}$$

在 $(0, +\infty)$ 上可导且只有一个驻点

$$\begin{aligned} \because \frac{4\pi r^3 - 2V_0}{r^2} &= 0 \\ \therefore r_0 &= \sqrt[3]{\frac{V_0}{2\pi}} \end{aligned}$$

考察二阶导数 $S''(r) = \frac{4\pi r^3 + 4V_0}{r^3} > 0$, 所以该驻点为极小值点。由于 $S(r)$ 在 $(0, +\infty)$ 上只有一个极小值点, 且无不可导点, 无端点, 所以 r_0 为 $S(r)$ 的最小值点, 此时:

$$h_0 = \frac{V_0}{\pi r_0^2} = 2\sqrt[3]{\frac{V_0}{2\pi}} = 2r_0$$

即储罐高和底面直径相等时, 用料最少。

例 5.3.3. 假设一个物理试验中, 一共进行了 n 次测量, 得到 x_1, x_2, \dots, x_n , 若用 \bar{x} 表示测量结果, 问 \bar{x} 为多少使得测量的总平方误差

$$TSE(\bar{x}) = \sum_{i=1}^n (\bar{x} - x_i)^2$$

最小。

解:

首先易得 $TSE(x)$ 在 $(-\infty, +\infty)$ 连续, 考察一阶和二阶导数:

$$TSE'(x) = 2 \sum_{i=1}^n (x - x_i) = 2nx - 2 \sum_{i=1}^n x_i$$

$$TSE''(x) = 2n > 0$$

可得, 当

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

时, $TSE(x)$ 有极小值, 且是最小值。

这里可以看出, 当取算术平均值时, 总平方误差最小。所以用算术平均值代替测量值, 在总平方误差的角度是最可靠的。

5.3.4 拓展讨论: 凸函数和拐点

定义 5.3.2 (凸函数). 设 $f(x) \in C[a, b]$, 若 $\forall x_1, x_2 \in (a, b), x_1 \neq x_2$ 和 $\forall t_1, t_2 > 0, t_1 + t_2 = 1$:

- 若 $f(t_1x_1 + t_2x_2) < t_1f(x_1) + t_2f(x_2)$, 则称 $f(x)$ 在 (a, b) 上下凸;
- 若 $f(t_1x_1 + t_2x_2) > t_1f(x_1) + t_2f(x_2)$, 则称 $f(x)$ 在 (a, b) 上上凸;

通常我们将下凸函数称为凸函数。

定理 5.3.6. 若 $f(x)$ 在 $[a, b]$ 内连续, (a, b) 内二阶可导且恒有 $f''(x) > 0$ (或 $f''(x) < 0$), 则 $f(x)$ 在 (a, b) 内下凸 (或上凸)。

几何上, 凸函数表示 $[a, b]$ 上任取两点作连线, $f(x)$ 都在连线之下。注意这里取点的任意性。

定义 5.3.3 (拐点). 设 $f(x)$ 在 $N(x_0)$ 内连续, 若在 x_0 的左右两侧凸性相反, 则称 x_0 为 $f(x)$ 的拐点。

定理 5.3.7. 若 $f(x)$ 在 (a, b) 内二阶可导, $x_0 \in (a, b)$:

- 若 x_0 是 $f(x)$ 的一个拐点, 则有 $f''(x_0) = 0$;
- 若 $f''(x_0) = 0$ 且 $f''(x_0^+) \cdot f''(x_0^-) < 0$ (即 x_0 两侧 $f''(x)$ 异号), 则 x_0 是 $f(x)$ 的一个拐点。

5.3.5 拓展讨论: 琴生 (Jensen) 不等式

反过来我们用凸函数的定义可以获得一个比较重要的不等式。

定义 5.3.4 (琴生 (Jensen) 不等式). 若 $f(x)$ 在 (a, b) 内下凸, 则对于

$$\forall x_1, x_2, \dots, x_n \in (a, b)$$

$$\forall t_1, t_2, \dots, t_n > 0$$

$$\sum_{i=1}^n t_i = 1$$

必有:

$$f\left(\sum_{i=1}^n t_i x_i\right) \leq \sum_{i=1}^n [t_i f(x_i)]$$

其中 x_1, x_2, \dots, x_n 不全相等. 特别的, 当 $t_1 = t_2 = \dots = t_n = 1/n$ 时, 有:

$$f\left(\frac{x_1 + x_2 + \dots + x_n}{n}\right) \leq \frac{f(x_1) + f(x_2) + \dots + f(x_n)}{n}$$

5.3.6 拓展讨论: 渐近线

定义 5.3.5 (函数的渐近线). 对于曲线 $f(x)$, 若存在直线 $y = kx + b$ 使得:

$$\lim_{x \rightarrow +\infty} [f(x) - y(x)] = 0 \quad \text{或} \quad \lim_{x \rightarrow -\infty} [f(x) - y(x)] = 0$$

则称 $y = kx + b$ 为 $f(x)$ 的渐近线, 特别的当 $k = 0$ 时, 称 $y = b$ 为 $f(x)$ 的水平渐近线. 若曲线 $f(x)$ 有:

$$\lim_{x \rightarrow x_0^-} f(x) = \infty \quad \text{或} \quad \lim_{x \rightarrow x_0^+} f(x) = \infty$$

则称 $x = x_0$ 为 $f(x)$ 的垂直渐近线。

渐近线判断方法:

1. 考察 $(-\infty, +\infty)$ 上未定义点 x_0 处的极限 $\lim_{x \rightarrow x_0} f(x)$, 若为 ∞ , 说明有垂直渐近线 $x = x_0$ 。
2. 计算 $k_1 = \lim_{x \rightarrow +\infty} \frac{f(x)}{x}$ 、 $k_2 = \lim_{x \rightarrow -\infty} \frac{f(x)}{x}$, 和 $b_1 = \lim_{x \rightarrow +\infty} [f(x) - kx]$ 、 $b_2 = \lim_{x \rightarrow -\infty} [f(x) - kx]$, 获得斜渐近线或水平渐近线。

例 5.3.4. 判断逻辑斯蒂 (Logistic) 函数 $f(x) = \frac{c}{1+be^{-ax}}$ 的渐近线。

解:

$f(x)$ 在 $(-\infty, +\infty)$ 内均有定义, 所以没有垂直渐近线, 考察斜渐近线

和水平渐近线:

$$\begin{aligned} k_{1,2} &= \lim_{x \rightarrow \pm\infty} \frac{f(x)}{x} = \lim_{x \rightarrow \pm\infty} \frac{c}{(1 + be^{-ax})x} = 0 \\ b_1 &= \lim_{x \rightarrow +\infty} [f(x) - kx] = \lim_{x \rightarrow +\infty} \frac{c}{1 + be^{-ax}} = c \\ b_2 &= \lim_{x \rightarrow -\infty} [f(x) - kx] = \lim_{x \rightarrow -\infty} \frac{c}{1 + be^{-ax}} = 0 \end{aligned}$$

可见 $f(x) = \frac{c}{1+be^{-ax}}$ 有两条渐近线:

$$y = c$$

$$y = 0$$

例 5.3.5. 考察 $f(x) = \frac{(x-3)^2}{4(x-1)}$ 的渐近线。

解:

显然, $x = 1$ 处未定义, 考察:

$$\lim_{x \rightarrow 1} \frac{(x-3)^2}{4(x-1)} = \infty$$

可得 $x = 1$ 为 $f(x)$ 的一条垂直渐近线。再考察斜渐近线和水平渐近线:

$$\begin{aligned} k_{1,2} &= \lim_{x \rightarrow \pm\infty} \frac{f(x)}{x} = \lim_{x \rightarrow \pm\infty} \frac{(x-3)^2}{4(x-1)x} = \frac{1}{4} \\ b_{1,2} &= \lim_{x \rightarrow \pm\infty} [f(x) - kx] = \lim_{x \rightarrow \pm\infty} \left[\frac{(x-3)^2}{4(x-1)} - \frac{x}{4} \right] = -\frac{5}{4} \end{aligned}$$

可得只有一条斜渐近线:

$$y = \frac{1}{4}x - \frac{5}{4}$$

5.4 本章小结

本章介绍了导数。导数的基础是极限与连续，高中阶段不详细介绍极限与连续，所以对导数的推导和应用都比较简单，我只需要掌握求导公式和了解导数能用于求解极值即可。

5.4.1 习题

例 5.4.1 (拓广探索 19, 难度: ★★). 已知函数 $f(x) = ae^{2x} + (a-2)e^x - x$.

1. 讨论 $f(x)$ 的单调性;
2. 若 $f(x)$ 有两个零点, 求 a 的取值范围.

解:

(1) 对函数求导:

$$f'(x) = 2ae^{2x} + (a-2)e^x - 1$$

$$f''(x) = 4ae^{2x} + (a-2)e^x$$

令 $f'(x) = 0$ 求得 $e^x = 1/a$, 讨论:

- 当 $a = 0$ 时, $f(x) = -2e^x - x$, 显然单调递减;
- 当 $a < 0$ 时, $f'(x) < 0$, 显然单调递减;
- 当 $a > 0$ 时, 求得极值点 $x = \ln \frac{1}{a}$, $f''(\ln \frac{1}{a}) = \frac{2}{a} + 1 > 0$, 显然为极小值, 函数先减后增。

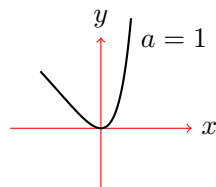
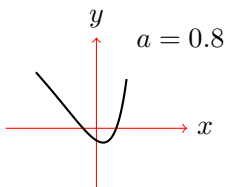
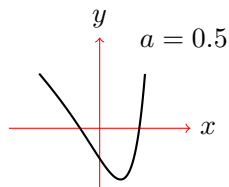
(2) 显然至少需要 $a > 0$, 而且最小值必须小于 0:

$$\because f\left(\ln \frac{1}{a}\right) = a \frac{1}{a^2} + (a-2) \frac{1}{a} - \ln \frac{1}{a} = 1 - \frac{1}{a} - \ln \frac{1}{a} < 0$$

$$\therefore 1 - \frac{1}{a} < \ln \frac{1}{a}$$

$$\therefore \frac{1}{a} > 1$$

得到取值范围 $0 < a < 1$, a 分别取 0.5、0.8、1 时 $f(x)$ 的图形如下:



本题没有难度。函数形状的讨论都是针对其一阶和二阶导数的分段讨论, 讨论的是另两个函数。

第六章 计数原理

本章主要介绍排列组合。

本章要点：

- 计数原理。
- 排列组合。
- 二项式定理。

6.1 分类加法计数原理与分步乘法计数原理

本节要点：

- 深刻理解和熟练掌握分类加法计数原理；
- 深刻理解和熟练掌握分步乘法计数原理。

本节是本章的基础，要熟练掌握。

分类加法计数原理体现的是空间上的并列性，步骤间相互独立；分步乘法计数原理体现了时间上的先后性，步骤间可能有相关性，后面几步的取值范围可能受上几步的影响。

6.2 排列与组合

本节要点：

- 掌握排列的概念及其计算公式；
- 掌握组合的概念及其计算公式。

定义 6.2.1 (排列和组合). 从 n 个元素中任取 m 个元素 ($m \leq n$) 排成一有序列, 称从 n 个元素中取出 m 个元素的一个排列, 所有排列的个数称作排列数, 记作 P_n^m , 有:

$$P_n^m := \frac{n!}{(n-m)!} = n(n-1)(n-2)\cdots[n-(m-1)]$$

从 n 个元素中任取 m 个元素 ($m \leq n$) 组成一无序组, 称从 n 个元素中取出 m 个元素的一个组合, 所有组合的个数称作组合数, 记作 C_n^m , 有:

$$C_n^m := \frac{n!}{(n-m)! \cdot m!} = \frac{P_n^m}{m!}$$

特别地, $P_n^1 = C_n^1 = n, P_n^n = n!, C_n^n = 1$ 。

排列组合均是对“从 n 个元素中抽取 m 个元素”的计算, 排列突出有序, 组合突出无序。排列组合的关系还可以这么理解, 当 m 个元素排列完成后, 若不考虑其顺序, 就是组合, 由于每个排列有 P_m^m 种不同的顺序, 所以:

$$P_n^m = C_n^m \cdot P_m^m = C_n^m \cdot m!$$

定理 6.2.1. 组合公式有以下性质:

$$\begin{aligned} C_n^m &= C_n^{n-m} \\ C_{n+1}^m &= C_n^m + C_n^{m-1} \end{aligned}$$

6.3 二项式定理

本节要点：

- 熟练掌握二项式定理。

$$\begin{aligned}(a+b)^n = & C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + \cdots + \\ & C_n^k a^{n-k} b^k + \cdots + \\ & C_n^{n-1} a^1 b^{n-1} + C_n^n a^0 b^n\end{aligned}$$

6.4 本章小结

本章介绍了计数原理和排列组合，是后一章的基础，需要认真掌握。

第七章 随机变量及其分布

本章介绍随机变量。

本章要点：

- 条件概率。
- 全概率公式。
- 离散型随机变量。
- 3 类分布。

7.1 条件概率与全概率公式

本节要点:

- 掌握并理解条件概率的概念;
- 掌握并理解全概率公式;
- 深刻理解贝叶斯公式的哲学意义。

7.1.1 条件概率

定义 7.1.1 (条件概率). 在事件 A 发生的情况下, 事件 B 发生的概率称为条件概率, 记作 $P(B|A)$, 当 $P(A) > 0$ 时有:

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

$P(B|A)$ 和 $P(B)$ 的大小关系是不定的。也就是说, 事件 A 的发生并不一定有利于 B , 也可能没有影响。切不能想当然地认为 $P(B|A) > P(B)$!

定理 7.1.1 (乘法公式). 若 $P(A) > 0$, 则

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

推论 7.1.2 (乘法公式). 若 $P(A_1 A_2 \cdots A_n) > 0, n \geq 2$, 则

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

7.1.2 全概率公式

定义 7.1.2 (完备事件组). 若事件组 $A_1, A_2, \cdots, A_i, \cdots$ 两两互斥, 且 $A_1 + A_2 + \cdots + A_i + \cdots = \Omega$, 则 $A_1, A_2, \cdots, A_i, \cdots$ 构成样本空间 Ω 的一个完备事件组, 简称完备组。

虽然教材上没提, 但是我们还是需要理解完备组的概念, 运用全概率的前提就是找到一个完备组。

定义 7.1.3 (全概率公式). 若事件组 $A_1, A_2, \cdots, A_i, \cdots$ 构成样本空间 Ω 的一个完备组, 则对于任何事件 B 有:

$$P(B) = \sum_{i=1}^{\infty} P(A_i B) = \sum_{i=1}^{\infty} [P(A_i) \cdot P(B|A_i)]$$

完备组“瓜分”了整个样本空间，是事件对立关系的扩充，其意义在于将一个复杂的大问题分解为若干简单的小问题。

全概率可以认为是条件概率的延申，如果事件组满足完备条件，则事件 B 的发生总是伴随某个组员的发生，或者说 B 是通过某个路径发生的。当 $P(B)$ 的计算较为复杂时，我们可以构造一个合适的完备组，简化 $P(B)$ 的计算。其次，该公式从数学上还可以理解为 $P(B)$ 是 $P(B|A_i)$ 的加权和，权重为 $P(A_i)$ 。最后还需注意， $P(B) \leq 1$ ，但 $\sum_{i=1}^{\infty} P(B|A_i)$ 几乎可能大于 1。

全概率公式 $P(B) = \sum_{i=1}^n [P(A_i) \cdot P(B|A_i)]$ 中， $P(B|A_i)$ 可以认为是事件 B 在基 A_i 下的坐标分量，该分量越大，表示越可能发生。

定义 7.1.4 (贝叶斯公式). 若事件 $A_1, A_2, \dots, A_i, \dots$ 构成样本空间 Ω 的一个完备事件组，且 $P(A_i) > 0$ ，若 B 为一随机事件，且 $P(B) > 0$ ，则：

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty} [P(A_i) \cdot P(B|A_i)]} = \frac{P(A_i)P(B|A_i)}{P(B)}$$

又称**逆概率公式**。通常，我们将 $P(A_i)$ 称为**先验概率**，将条件概率 $P(A_i|B)$ 称为**后验概率**。也可以结合向量内积理解。

证明. 由条件概率的概念可得：

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

结合全概率公式，且 $P(A_i B) = P(BA_i) = P(A_i)P(B|A_i)$ 。 □

$P(A_i)$ 称为先验概率的原因是 A_i 作为一种假定而不依赖于 B 。 B 称为**证据**，一种出现的客观事物。在证据出现前， A_i 有自己的概率。当证据发现时， A_i 的概率变成了 $P(A_i|B)$ ，所以我们称为后验概率。“先”和“后”指的是证据出现前和出现后。其中 $P(B|A_i)$ 描述了 A_i 和 B 的关系，机器学习中称为**似然 (likelihood)**。

全概率公式体现了由因求果，而贝叶斯公式反映了执果寻因的一种辩证思想，我们知道完备组员的概率 $P(A_i)$ 和 B 在各个完备组员的条件概率 $P(B|A_i)$ ，就可以计算 B 大概会发生在哪个组员。很多情况下，我们知道各个原因发生的可能性，但是一旦结果出现，这些原因的可能性又会出现变化，贝叶斯公式揭示了两者之间的关系。

7.1.3 习题

例 7.1.1 (复习巩固 3, 难度: ★). 甲、乙两人向同一个目标各射击 1 次, 已知甲命中目标的概率为 0.6, 乙命中目标的概率为 0.5。已知目标至少被命中 1 次, 求甲命中目标的概率。

解:

设甲命中目标为 A , 乙命中目标为 B , 目标至少被命中 1 次为 C , 由于甲乙相互独立, 有:

$$P(C) = P(A) + P(\bar{A}B) = P(A) + P(\bar{A}) \cdot P(B)$$

甲命中目标的概率可表示为条件概率:

$$\begin{aligned} P(A|C) &= \frac{P(AC)}{P(C)} = \frac{P(A) \cdot P(\bar{B}) + P(A) \cdot P(B)}{P(A) + P(\bar{A}) \cdot P(B)} \\ &= \frac{0.6 \cdot 0.5 + 0.6 \cdot 0.5}{0.6 + 0.4 \cdot 0.5} = 0.75 \end{aligned}$$

拓展讨论:

C 的概率可以有如下三种算法:

- 甲命中乙不命中 + 甲不命中乙命中 + 甲乙都命中;
- 甲命中 + 甲不命中乙命中;
- 乙命中 + 乙不命中甲命中。

以上三种算法是三种完备组, 易得这三种方法的结果是一样的。

$$\begin{aligned} P(C) &= P(A) + P(\bar{A}B) = P(A) + P(\bar{A}) \cdot P(B) \\ &= P(A) \cdot [P(B) + P(\bar{B})] + P(\bar{A}) \cdot P(B) \\ P(C) &= P(B) + P(\bar{B}A) = P(B) + P(\bar{B}) \cdot P(A) \\ &= P(B) [P(A) + P(\bar{A})] + P(\bar{B}) \cdot P(A) \\ P(C) &= P(A\bar{B}) + P(\bar{A}B) + P(AB) \\ &= P(A) \cdot P(\bar{B}) + P(\bar{A}) \cdot P(B) + P(A) \cdot P(B) \end{aligned}$$

本题关键在于理解 $P(C), P(AC)$ 。

例 7.1.2 (复习巩固 4, 难度: ★). 甲乙两个箱子中各装有 10 个球, 其中甲箱中有 5 个红球、5 个白球, 乙箱中有 8 个红球、2 个白球。掷一枚质地均匀的骰子, 如果点数为 1 或 2, 从甲箱子中随机摸出 1 个球; 如果点数为 3, 4, 5, 6, 从乙箱子中随机摸出 1 个球。求摸到红球的概率。

解:

设骰子点数为 1 或 2 为 A , 从甲箱子中摸出红球为 B , 从乙箱子中摸出红球为 C , 则:

$$\begin{aligned} P &= P(A) \cdot P(\bar{B}) + P(A) \cdot P(C) \\ &= \frac{2}{6} \cdot \frac{5}{10} + \frac{4}{6} \cdot \frac{8}{10} = \frac{7}{10} \end{aligned}$$

拓展讨论:

红球要么从甲箱要么从乙箱中获得, 构成一个完备组, 而骰子的出现可以认为是给甲箱和乙箱加了一个加权系数, 而且加权系数和为 1。

本题运用全概率公式即可。

例 7.1.3 (复习巩固 5, 难度: ★). 在 A, B, C 三个地区爆发了流感, 这三个地区分别有 6%, 5%, 4% 的人患了流感。假设这三个地区的人口数的比为 5:7:8, 现从这三个地区中任意选取一个人。

1. 求这个人患流感的概率;
2. 如果此人患流感, 求此人选自 A 地区的概率。

解:

显然地区 A, B, C 构成完备组。

(1)

$$P(A) = \frac{5}{5+7+8} \cdot 6\% + \frac{7}{5+7+8} \cdot 5\% + \frac{8}{5+7+8} \cdot 4\% = 4.85\%$$

(2)

$$P = \frac{\frac{5}{5+7+8} \cdot 6\%}{P(A)} = \frac{0.015}{0.0485} = 0.309278$$

本题考查条件概率, 不难。

例 7.1.4 (综合运用 7, 难度: ★). 一批产品共有 100 件, 其中 5 件为不合格品。收货方从中不放回地随机抽取产品进行检验, 并按以下规则判断是否接受这批产品: 如果抽检的第 1 件产品不合格, 则拒绝整批产品; 如果抽检的第 1 件产品合格, 则再抽 1 件, 如果抽检的第 2 件产品合格, 则接受整批产品, 否则拒绝整批产品。求这批产品被拒绝的概率。

解:

易得, 这批产品要么整批接收, 要么整批拒绝, 显然整批接收的条件为两次抽检均合格, 比较好计算, 有:

$$1 - P(A) = 1 - \frac{95}{100} \cdot \frac{94}{99} = 0.0979798$$

本题用补集清晰简单。

7.1.4 拓展习题

例 7.1.5. 设 10 件产品中存在 2 件次品, 现不放回地抽取 2 次, 每次 1 件, 求第 2 件是次品的概率。

解:

显然第 2 件是否为次品和第 1 件是否为次品有关, 是典型的条件概率。更进一步, 第 1 件产品的合格和次品两个事件可以构成完备组, 所以可以用全概率公式求解。

设第 1 次抽取为次品为事件 A , 于是 A, \bar{A} 可视为完备组, 且易得 $P(A) = \frac{2}{10}, P(\bar{A}) = \frac{8}{10}$ 。再设第 2 次抽到次品为事件 B , 于是 $P(B)$ 可以用这个完备组分解:

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

其中:

- $P(B|A)$: 表示第 1 次抽到次品时第 2 次也抽到次品的概率, 易得 $\frac{1}{9}$;
- $P(B|\bar{A})$: 表示第 1 次抽到合格时第 2 次抽到次品的概率, 易得 $\frac{2}{9}$ 。

最终得到:

$$P(B) = \frac{1}{5} \cdot \frac{1}{9} + \frac{4}{5} \cdot \frac{2}{9} = \frac{1}{5}$$

解二, 不用全概率的概念, 直接使用古典概型的方法, 抽取 2 件排序, 共 P_{10}^2 种, 第 2 件是次品有两种互不相同可能:

- 第 1 次抽到次品时第 2 次也抽到次品, $C_2^1 \cdot C_1^1$ 种;
- 第 1 次抽到合格时第 2 次抽到次品, $C_8^1 \cdot C_2^1$ 种。

于是:

$$P = \frac{C_2^1 \cdot C_1^1 + C_8^1 \cdot C_2^1}{P_{10}^2} = \frac{18}{90} = \frac{1}{5}$$

例 7.1.6. 假设血清甲蛋白法测定肝癌阳性准确率为 95%、阴性准确率为 90%, 若已知社会人群的患癌率为 0.04%, 若某人检测阳性, 分析其实际患癌的概率。

解:

设事件 A 表示被检者实际患癌, 事件 B 表示用该测定法判定阳性, 由题干可知:

$$P(A) = 0.0004$$

$$P(B|A) = 0.95$$

$$P(\bar{B}|\bar{A}) = 0.90$$

由于 A, \bar{A} 构成完备组, 则有:

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})} \\ &= \frac{0.0004 \cdot 0.95}{0.0004 \cdot 0.95 + 0.9996 \cdot (1 - 0.9)} = 0.00378712 \end{aligned}$$

一方面可见即便测定阳性, 实际患癌也不到 1%, 但比一般人群已经高了 $0.3787/0.04 = 9.4675$, 将近 10 倍。另一方面, 通过贝叶斯公式我们了解到, 血清甲蛋白测定法不是判断患癌的概率, 而是判断比正常人患癌概率高了多少。

该例子很具典型性。首先 0.04% 的患癌概率是作为一种先天的经验存在。测试结果则是一种证据。证据的出现可以让我们对这个先天经验进行修正。由于这个证据本身带有概率, 而非 100%, 所以它实际上是提高了患癌概率。

例 7.1.7. 某水文站测量水流量, 令 B 为测量不准的事件, 并总结了几个测量不准的原因 A_i , 原因的发生概率 $P(A_i)$, 以及某个原因发生时测量不准的概率 $P(B|A_i)$, 如下表, 分析哪几个原因主要导致了测量不准。

测不准的原因 A_i	$P(A_i)$	$P(B A_i)$	$P(A_i B)$
流速计故障	0.01	0.90	0.05
绕道设备误差	0.05	0.80	0.24
水中有杂质	0.04	0.10	0.02
气象条件不佳	0.10	0.30	0.18
人员配合问题	0.10	0.80	0.47
其他	0.70	0.01	0.04

解:

“那几个原因主要导致了测量不准”也即测量不准的情况下,各个原因的发生概率 $P(A_i|B)$, 可以用贝叶斯公式解决。

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty} P(A_i)P(B|A_i)}$$

计算结果如上表最右列。可见,单看每个已知的明确的原因,它们的发生概率都不大,但是一旦问题出现,最有可能是人员配合导致。所以若要提高测量准确度,首先得从提高人员配合度着手。

例 7.1.8. 甲箱中有 2 个白球 4 个黑球,乙箱中有 6 个白球 2 个黑球,现从这两箱中各任取一球,再从所取出的两球中任取一球,求最终该球是白球的概率。

解:

令从甲箱中取得白球为事件 A , 从乙箱中取得白球为事件 B , 易得甲乙两个箱子中取球为独立事件, 于是:

$$\begin{aligned} P(A) &= \frac{1}{3}, P(B) = \frac{3}{4} \\ P(AB) &= \frac{1}{4}, P(A\bar{B}) = \frac{1}{12}, P(\bar{A}B) = \frac{1}{2}, P(\bar{A}\bar{B}) = \frac{1}{6} \end{aligned}$$

且 $AB, \bar{A}B, A\bar{B}, \bar{A}\bar{B}$ 构成完备组, 于是可以用全概率公式得到:

$$\begin{aligned} P(C) &= P(AB)P(C|AB) + P(A\bar{B})P(C|A\bar{B}) + \\ &\quad P(\bar{A}B)P(C|\bar{A}B) + P(\bar{A}\bar{B})P(C|\bar{A}\bar{B}) \end{aligned}$$

而易得:

$$\begin{aligned} P(C|AB) &= 1 \\ P(C|A\bar{B}) &= P(C|\bar{A}B) = \frac{1}{2} \\ P(C|\bar{A}\bar{B}) &= 0 \end{aligned}$$

于是:

$$P(C) = \frac{1}{4} + \frac{1}{12} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{13}{24}$$

7.2 离散型随机变量及其分布列

本节要点：

- 掌握一维离散型随机变量及其分布律的概念；
- 掌握 0-1 分布的概念。

定义 7.2.1 (一维随机变量). 设随机试验 E 的样本空间为 Ω , 若每个样本点 ω 均有唯一确定的实数 X 与之对应, 则称 X 为定义在 Ω 上的一维随机变量, 也称随机变量或简称变量。

随机变量将对样本点的文字描述投射到实数, 注意, 这里投射的是样本点, 或者说是基本事件, 而非任意事件。 X 的值在有些情况下可直观地赋予, 如次品数量、人流数量、等待时间, 但在有些情况下是人为赋予的, 并没有特别的物理意义, 如抛硬币的正反面, 此时只要保证“唯一”、“确定”两个要求即可。

于是, 我们将事件 A 及其概率 $P(A)$ 和变量 X 的取值范围 D 的概率 $P\{X \in D\}$ 等价起来:

$$\begin{aligned} \omega & \quad X \\ A & \quad \Leftrightarrow X \in D \\ P(A) & \quad P\{X \in D\} \end{aligned}$$

一维随机变量可视作一种特殊的变量, 虽然其取值范围可以为 \mathbb{R} , 但与普通变量相比有如下特殊性:

- 根据具体试验, 随机变量有意义的取值是人为设定且事先已知的;
- 但是在试验前, 具体的取值无法预知;
- 我们能知道的是每个取值的概率。

定义 7.2.2 (分布函数). 设有随机变量 X , 对于 $\forall x \in \mathbb{R}$, 若 $P\{X \leq x\}$ 均有唯一的值与之对应, 则称该对应关系为 X 的分布函数, 记作 $F(x)$, 即:

$$F(x) := P\{X \leq x\}, \quad x \in \mathbb{R}$$

分布函数描述了 X 每个取值之前 (包括该值) 的所有概率之和, 所以分布函数的差可以描述变量在一个范围内的概率, 即 X 在 $(a, b]$ 的概率可以表示为:

$$P\{a < X \leq b\} = F(b) - F(a)$$

分布函数的定义还隐含了判断一个函数能否被视为分布函数的充要条件:

$$F(X) \text{ 是分布函数} \Leftrightarrow \begin{cases} 0 \leq F(x) \leq 1 \\ \lim_{x \rightarrow -\infty} F(x) = P(\emptyset) = 0 \\ \lim_{x \rightarrow +\infty} F(x) = P(\Omega) = 1 \end{cases}$$

分布函数还有以下性质:

- $F(x)$ 必单调增, 即 $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$;
- $F(x)$ 每个点 x_0 均右连续, 即 $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$ 。

这里需要理解 X, x 两者的区别。 x 是普通变量, 取值范围为 \mathbb{R} , 本身不承担物理意义, 而 X 是随机变量, 有意义的取值范围通过事件规定。如投硬币中, 我们规定 $X = 1$ 表示正面向上, $X = 0$ 表示反面向上, X 取 0、1 之外的值是没有意义的, 或者说 $P\{X \neq 0 \text{ or } 1\} = 0$ 。

至此, 我们将之前的概念和函数的概念统一了起来, 我们用变量代替样本, 用函数描述概率。

事件	变量
样本点 ω	随机变量 X
样本空间 Ω	\mathbb{R}
随机事件 A	取值范围, 如 $\{X = 1\}, \{X \geq b\}, \{a < X \leq b\}$
概率 $P(A)$	分布函数的差, 如 $P\{a < X \leq b\} = F(b) - F(a)$

定义 7.2.3 (一维离散型随机变量). 若一维随机变量 X 的取值为有限个或可列无穷多个, 则称 X 为一维离散型随机变量, 简称离散变量。设可能的取值为 $x_1, x_2, \dots, x_i, \dots$, 且对应的概率记作:

$$P\{X = x_i\} = p_i$$

称序列 p_i 为 X 的分布律或概率分布, 可用表格展现, 或记作:

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_i & \cdots \\ p_1 & p_2 & \cdots & p_i & \cdots \end{pmatrix}$$

其中:

- x_i : 纯数学记号, 对应基本事件 A_i , 通常为简化讨论, x_i 取值为非负整数, 分布律的计算通常基于上一章的讨论;
- $p_i = P\{X = x_i\} = P(A_i)$: 表示基本事件 A_i 的发生概率。

若离散变量 X 有分布律 p_i , 则其分布函数为:

$$F(x) = P\{X \leq x\} = \sum_{x_i \leq x} p_i$$

由分布函数的充要条件可得 p_i 成为分布律的充要条件:

$$p_i \text{ 为分布律} \Leftrightarrow \begin{cases} p_i \geq 0 \\ \sum_i p_i = 1 \end{cases}$$

0-1 分布 又称**两点分布**, 记作 $B(1, p)$, 描述最简单的分布:

$$\begin{aligned} P\{X = 1\} &= p \\ P\{X = 0\} &= 1 - p \end{aligned}$$

更普遍地记作:

$$P\{X = x\} = p^x (1 - p)^{1-x}, \quad \begin{matrix} x = 0, 1 \\ p \in (0, 1) \end{matrix}$$

其中:

- x : 一般取 0 表示事件不发生, 1 表示事件发生;
- p : 表示事件发生的概率。

特别地, 当 $p = 0$ 或 1 时称为**常数分布**。常数分布表示了事件 100% 发生, 所以无实际用途, 只是为了完善概念而提出。

一般如果试验中只有两个样本点, 非黑即白、非正即负, 我们就可以用两点分布描述。两点分布的使用场景非常少。

7.2.1 拓展讨论: 密度函数的物理意义

离散变量的分布律的意义是清晰的, 每个点代表基本事件的发生概率。但连续变量的密度函数的意义并不十分清晰, 本节讨论密度函数的物理意义, 以求能直观地把握这个概念。

若有一维无限长细线, 总质量为 M 但分布不均匀, 已知线密度 $\rho(l)$, l 为距原点的距离, 任取一段 $[l_1, l_2]$, 该段质量占比为:

$$\frac{m}{M} = \int_{l_1}^{l_2} \frac{\rho(l)}{M} \cdot dl$$

若考虑一维无限长带电细线, 总电荷量为 Q 但分布不均匀, 已知电密度 $\rho(l)$, 任取一段 $[l_1, l_2]$, 该段电荷占比为:

$$\frac{q}{Q} = \int_{l_1}^{l_2} \frac{\rho(l)}{Q} \cdot dl$$

再考虑变速运动, 若速度随时间变化且有规律 $v = v(t)$, 则在某时段走过的路程占总路程 S :

$$\frac{s}{S} = \int_{t_1}^{t_2} \frac{v(t)}{S} \cdot dt$$

对比概率计算公式:

$$P\{a < X \leq b\} = \int_a^b f(x) \cdot dx$$

容易发现, 概率 $P\{a < X \leq b\}$ 表示了一段范围内某个物理量的占比, 密度函数 $f(x)$ 刻画了该物理量的分布情况。根据 $\rho(l)$ 我们可以知道哪里重, 根据 $v(t)$ 我们可以知道哪个时段走得更多。同样道理, 根据密度函数, 我们可以知道 X 在哪里出现的概率大。关联到具体事件, 我们就可以知道哪些事件容易发生。所以, 我们通常用“占比大”、“出现多”来指代概率大。

密度函数的物理意义是反映了一个物理量在某个范围内的占比。这个“占比”就是概率, 而“范围”可以是空间上的一个段, 也可以是时间上的一个段。以质量为例, 可以认为密度函数具有量纲 $[\text{m}^{-1}]$ 。

7.3 离散型随机变量的数字特征

本节要点：

- 掌握数学期望的概念；
- 理解数学期望的数学意义和物理意义；
- 掌握方差的概念；
- 理解方差的统计意义和物理意义。

7.3.1 离散型随机变量的均值

定义 7.3.1 (数学期望). 设离散变量 X 有分布律 p_i , 若无穷级数 $\sum_{i=1}^{+\infty} x_i \cdot p_i$ 绝对收敛 (即 $\sum_{i=1}^{+\infty} |x_i \cdot p_i|$ 收敛), 则称该无穷级数为 X 的数学期望或均值, 记作 $E(X)$ 或 EX :

$$EX := \sum_{i=1}^{+\infty} x_i \cdot p_i$$

若离散变量取值为有限个, 则必有数学期望。

一般, 我们用 $E(X)$ 表示谓词, 即计算 X 的数学期望这个过程, 用 EX 表示名词, 即数学期望这个值。 E 是 Expectation 的首字母。

级数有个现象, 如果调换各项的顺序, 收敛性可能发生改变, 即便收敛, 也可能收敛到不同的值, 而在概率论中, 随机变量的取值和顺序无关, 所以这里要求绝对收敛。

定义 7.3.2 (随机变量函数的数学期望). 设离散变量 X 有分布律 p_i , 若有另一随机变量 $Y = g(X)$, 且无穷级数 $\sum_{i=1}^{+\infty} g(x_i) p_i$ 绝对收敛, 则称 $E[g(X)]$ 为函数 $Y = g(X)$ 的数学期望:

$$EY := \sum_{i=1}^{+\infty} g(x_i) \cdot p_i$$

该公式的意义在于, 有时候我们很难获得 Y 的具体分布规律, 但可以通过已知的 X 和 $Y = g(X)$ 获得 Y 的数学期望。

7.3.2 离散型随机变量的方差

定义 7.3.3 (方差). 设离散型随机变量 X , 若函数 $(X - EX)^2$ 的数学期望存在, 则称该数学期望为 X 的方差, 记作 $D(X)$ 或 DX , 即:

$$DX := E[(X - EX)^2] = \sum_{i=1}^{+\infty} (x_i - EX)^2 \cdot p_i$$

同时称 \sqrt{DX} 为 X 的平方差。

方差的英文为 Variance, 有些教材使用 $\text{Var}(X)$ 作为记号。

方差的求解较为繁琐, 由于 $(X - EX)^2 = X^2 - 2 \cdot X \cdot EX + (EX)^2$, 且 EX 是常数, 利用数学期望的性质可得便捷公式:

$$\begin{aligned} DX &= E[X^2 - 2 \cdot X \cdot EX + (EX)^2] \\ &= E(X^2) - 2 \cdot EX \cdot EX + (EX)^2 \\ &= E(X^2) - (EX)^2 \end{aligned}$$

从便捷公式可以得到 $[E(X)]^2$ 和 $E(X^2)$ 的关系:

$$E(X^2) = DX + (EX)^2$$

由于 $DX \geq 0$, 所以 $E(X^2) \geq (EX)^2$ 。

由于数学期望的线性性, 便捷公式的推导没有任何前置条件, 所以被广泛使用, 定义式则更多用来理解概念。

7.3.3 拓展讨论: 数学期望的数学意义

数学期望是随机变量的加权和, 权重是概率, 由于总概率为 1, 于是 $EX < x_{+\infty}$, 所以期望也称加权平均。任何一个给定的随机变量, 若有数学期望, 则都是一个固定的、可计算的常量。

从线性代数看, 数学期望是随机变量构成的向量及其分布规律构成的向量的标量积。以离散型随机变量为例:

$$\begin{aligned} EX &:= \sum_{i=1}^{+\infty} x_i p_i = \mathbf{x}^T \mathbf{p} \\ \begin{cases} \mathbf{x} = (x_1, x_2, \dots, x_n, \dots) \\ \mathbf{p} = (p_1, p_2, \dots, p_n, \dots) \end{cases} \end{aligned}$$

EX 越大表示 p 和 x 的契合度越高, 或者说越靠后的事件越容易发生。

这里的数学意义更多是为意义而意义, 是一个“形象的比喻”, 不用太过深究。

7.3.4 拓展讨论: 数学期望的物理意义

考虑一维物体的转动。长度为 L 的细长杆质量 M 且分布不均匀, 有线密度 $\rho(x), x \in [d, d+L]$, 一端距原点 d , 绕原点匀速转动且有角速度 ω , 分析所需的向心力。可以认为是同质量的质点在 x_0 处以通常角速度转动所需的向心力。

可以假设细长杆的一个微元可视为质点, 则向心力微元和整体向心力:

$$dF = \rho(x) dx \cdot \omega^2 \cdot x \quad [\text{N}]$$

$$F = \int_d^{d+L} dF = \int_d^{d+L} \rho(x) dx \cdot \omega^2 \cdot x = \omega^2 \int_d^{d+L} x \rho(x) \cdot dx \quad [\text{N}]$$

若将其视为质点, 则向心力等价于质量 M 的质点小球在 x_0 处以角速度 ω 匀速转动, 有:

$$\omega^2 \int_d^{d+L} x \rho(x) \cdot dx = M \cdot \omega^2 \cdot x_0 \quad [\text{N}]$$

于是可得:

$$x_0 = \frac{\int_d^{d+L} x \rho(x) \cdot dx}{M} = \int_d^{d+L} x \frac{\rho(x)}{M} \cdot dx \quad [\text{m}]$$

对比一维变量的数学期望定义式:

$$EX = \int_{-\infty}^{+\infty} x f(x) \cdot dx$$

数学期望的物理意义为绕原点匀速转动时的质心位置, 有量纲 $[\text{m}]$ 。

7.3.5 拓展讨论: 数学期望的性质

性质 1: 常数的数学期望为其本身, 即 $E(c) = c$ 。

性质 2: $X \geq a \Rightarrow EX \geq a, X \leq a \Rightarrow EX \leq a$

性质 3: $E(kX) = k \cdot E(X)$

性质 4: $E(kX + c) = k \cdot E(X) + c$, 由性质 1+3 可证。

性质 5: 设 (X, Y) 为二维随机变量, 则 $E(X \pm Y) = E(X) \pm E(Y)$

性质 6: 设 (X, Y) 为二维随机变量, 若 X, Y 相互独立 $\Rightarrow E(XY) = E(X) \cdot E(Y)$

性质 7: 设 (X, Y) 为二维随机变量, 则 $[E(XY)]^2 \leq E(X^2) \cdot E(Y^2)$

性质 1 表示了常数 c 的分布, 其实就是质点小球, 自然质心位置为 c 。

性质 2 表示了质心的范围, 物理意义也很明显, 质心位置肯定不会落到物体之外。

性质 3 表达当 x 轴拉伸时, 质心也有同样的拉伸, 压缩亦然。

这三个性质可结合物理意义理解。

性质 5 用定义即可证明:

$$\begin{aligned} E(X \pm Y) &= \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} (x_i \pm y_j) \cdot p_{ij} = \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} (x_i p_{ij} \pm y_j p_{ij}) \\ &= \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} x_i p_{ij} \pm \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} y_j p_{ij} \\ &= E(X) \pm E(Y) \end{aligned}$$

性质 6, 独立则有 $p_{ij} = p_{i \cdot} \cdot p_{\cdot j}$, 于是:

$$\begin{aligned} E(XY) &= \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} (x_i y_j) \cdot p_{ij} = \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} (x_i y_j) \cdot p_{i \cdot} \cdot p_{\cdot j} \\ &= \sum_{i=1}^{+\infty} \left(x_i p_{i \cdot} \cdot \sum_{j=1}^{+\infty} y_j p_{\cdot j} \right) = \sum_{i=1}^{+\infty} [x_i p_{i \cdot} \cdot E(Y)] \\ &= E(X) \cdot E(Y) \end{aligned}$$

数学期望的性质常用于随机变量组合成新变量的数学期望的计算。由于数学期望是线性运算, 所以有

$$E(aX + bY + c) = aEX + bEY + c$$

但对于非线性组合, 会有一些前置条件或者为不等式。

$(EX)^2$ 不大于 $E(X^2)$, 证明关系留到介绍方差后再给出。

7.3.6 拓展讨论：数学期望的范围

数学期望的范围是变量的范围，这点从数学期望的物理意义也很好理解，特别是当变量的分布为常数分布时，取到最大值。此时，随机事件变成确定性事件。对应地，函数的数学期望的范围是函数的值域。

7.3.7 拓展讨论：中位数的概念

定义 7.3.4 (中位数). 若随机变量 X 有分布函数 $F(x)$, 若 m 满足

$$F(m) = \frac{1}{2}$$

则称 m 为随机变量 X 的中位数。

数学期望和中位数都反映了随机变量的一个均值情况。中位数表示了该点前后的概率是一样的，优点是受特大值、特小值影响小，而且必然存在，缺点是处理不方便（如两个随机变量加减乘除后的中位数），而且不唯一，特别在离散型随机变量，由于取值的离散型，中位数大多不会精确的存在。

从物理意义上理解，数学期望反映的是质心，考虑了质点的力矩，而中位数不考虑力矩的影响。

与中位数相关的还有一个概念，上侧 α 分位点。

7.3.8 拓展讨论：方差的数学由来和统计意义

方差定义中， $X - EX$ 是根本，表示了随机变量 X 和它的均值 EX 的偏离程度。由于 $E(X - EX) = E(X) - EX = 0$ ，总是为 0，所以我们取绝对值 $|X - EX|$ ，又由于绝对值函数在 0 点不可导，所以我们替换成平方 $(X - EX)^2$ 定义方差。

方差描述了随机变量 X 偏离其数学期望的程度， DX 越小，说明离散程度越小。

7.3.9 拓展讨论：方差的物理意义

考虑一维物体的转动。长度为 L 的细长杆质量 M 且分布不均匀，有线密度 $\rho(x), x \in [d, d + L]$ ，一端距原点 d ，绕原点转动。

其转动惯量 I 有：

$$I = \int x^2 \cdot dM = \int_d^{d+L} x^2 \cdot \rho(x) \cdot dx \quad [\text{kg} \cdot \text{m}^2]$$

若将转轴移到质心 x_0 ，根据平行轴定理，新的转动惯量 I_{x_0} 有：

$$\begin{aligned}\therefore I &= I_{x_0} + Mx_0^2 \\ \therefore \int_d^{d+L} x^2 \cdot \rho(x) \cdot dx &= I_{x_0} + Mx_0^2 \\ \therefore \int_d^{d+L} x^2 \cdot \frac{\rho(x)}{M} \cdot dx &= \frac{I_{x_0}}{M} + x_0^2\end{aligned}$$

对比方差的便捷计算公式：

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) \cdot dx = DX + (EX)^2$$

显然，方差的物理意义可以理解为物体以质心为原点时的单位质量的转动惯量，有量纲 $[m^2]$ 。跟物理的转动惯量不一样，没有质量这个因子，所以描述的是一种和具体质量无关的惰性，这种惰性和分布有关， DX 越大，说明分布越靠外，转动惯量也越大。

7.3.10 拓展讨论：方差的性质

性质 1：常数的方差为 0，即 $D(c) = 0$ 。

推论： $D(X) \geq 0$ ，当且仅当 $P\{X = EX\} = 1$ 时（即常数分布），等号成立。

性质 2： $D(kX) = k^2 \cdot DX$

性质 3： $D(kX + c) = k^2 \cdot DX$ ，由性质 1, 2 可得。

性质 4：设 (X, Y) 为二维随机变量，

若 X, Y 相互独立 $\Rightarrow D(X \pm Y) = DX + DY$

进一步地，若 X, Y 相互独立 $\Rightarrow D(aX \pm bY) = a^2 \cdot DX + b^2 \cdot DY$

性质 1，对于常数分布，质心就是 c ，显然绕质心的转动惯量为 0。

性质 2，若 $X \rightarrow kX$ ，相当于直线被拉长 k 倍，自然转动惯量扩大了 k^2 倍。

上述性质可结合物理意义理解。

性质 4, 证明如下:

$$\begin{aligned}
 \therefore DX &= E(X^2) - (EX)^2 \\
 \therefore D(X \pm Y) &= E[(X \pm Y)^2] - E[(X \pm Y)] \cdot E[(X \pm Y)] \\
 &= E[X^2 \pm 2XY + Y^2] - [EX \pm EY]^2 \\
 &= [E(X^2) \pm 2E(XY) + E(Y^2)] - \\
 &\quad [(EX)^2 \pm 2(EX)(EY) + (EY)^2] \\
 \therefore E(XY) &= (EX)(EY) \\
 \therefore D(X \pm Y) &= [E(X^2) - (EX)^2] + [E(Y^2) - (EY)^2] \\
 &= DX + DY
 \end{aligned}$$

7.3.11 拓展讨论: 方差的取值范围

变量的方差的取值范围根据不同的分布有不同的范围。如常数分布, 中心点转动惯量必然为 0。又如 $P\{X = \pm\infty\} = 0.5$, 显然转动惯量 $+\infty$ 。

7.3.12 拓展讨论: 随机变量的标准化

定义 7.3.5 (随机变量的标准化). 设随机变量 X 有数学期望 EX 和方差 $DX > 0$, 若将 X 进行运算 $\frac{X-EX}{\sqrt{DX}}$, 获得的随机变量称为 X 的标准化随机变量, 记作 X^* , 即:

$$X^* = \frac{X - EX}{\sqrt{DX}}$$

且有:

$$\begin{aligned}
 E(X^*) &= E\left(\frac{X - EX}{\sqrt{DX}}\right) = \frac{1}{\sqrt{DX}} \cdot E(X - EX) = 0 \\
 D(X^*) &= D\left(\frac{X - EX}{\sqrt{DX}}\right) = \frac{1}{DX} \cdot D(X - EX) = 1
 \end{aligned}$$

考察标准化过程:

- $X - EX$: 表示将坐标原点拉到期望 EX 的位置, 即原点对称;
- $\frac{1}{\sqrt{DX}}$: 拉伸或收缩 x 轴, 使得分布不过于稀疏或稠密。

排除这两个“干扰”后, 很多初看起来不同的分布, 实际是一类分布, 有利于后续的进一步分析。如不同的正态分布, 标准化之后就是标准正态分布。

7.3.13 习题

例 7.3.1 (复习巩固 3, 难度: ★). 随机变量 X 的分布列为 $P(X=0)=0.2, P(X=1)=a, P(X=2)=b$. 若 $E(X)=1$, 求 a 和 b .

解:

由均值的定义可得:

$$E(X) = 0.2 \cdot 0 + a \cdot 1 + b \cdot 2 = 1$$

$$a + 2b = 1$$

作为概率还必须满足:

$$0.2 + a + b = 1$$

解得 $a = 0.6, b = 0.2$.

本题考察均值的概念, 没有难度。

例 7.3.2 (综合运用 6, 难度: ★★). 有 A 和 B 两道谜语, 张某猜对 A 谜语的的概率为 0.8 , 猜对得奖金 10 元; 猜对 B 谜语的的概率为 0.5 , 猜对得奖金 20 元。规则规定: 只有在猜对第一道谜语的情况下, 才有资格猜第二道。如果猜谜顺序由张某选择, 他应该选择先猜哪一道谜语?

解:

若从 A 谜语开始, 平均能获得:

$$P\{X=0\} = (1-0.8) = 0.2$$

$$P\{X=10\} = 0.8 \cdot (1-0.5) = 0.4$$

$$P\{X=30\} = 0.8 \cdot 0.5 = 0.4$$

$$E(X) = 0 \cdot 0.2 + 10 \cdot 0.4 + 30 \cdot 0.4 = 16$$

若从 B 谜语开始, 平均能获得:

$$P\{X=0\} = (1-0.5) = 0.5$$

$$P\{X=20\} = 0.5 \cdot (1-0.8) = 0.1$$

$$P\{X=30\} = 0.5 \cdot 0.8 = 0.4$$

$$E(X) = 0 \cdot 0.5 + 20 \cdot 0.1 + 30 \cdot 0.4 = 14$$

可见从 A 谜语开始获得总金额可能高一些。

本题考察均值的实际应用。

例 7.3.3 (综合运用 7, 难度: ★). 甲、乙两种品牌的手表, 它们的日走时误差分别为 X 和 Y (单位: s), 其分布列为

X	-1	0	1
P	0.1	0.8	0.1

X	-2	-1	0	1	2
P	0.1	0.2	0.4	0.2	0.1

试比较甲、乙两种品牌手表的性能。

解:

$$E(X) = E(Y) = 0$$

$$\begin{aligned} D(X) &= E(X^2) - (EX)^2 \\ &= (-1)^2 \cdot 0.1 + 0^2 \cdot 0.8 + 1^2 \cdot 0.1 - 0 \\ &= 0.2 \end{aligned}$$

$$\begin{aligned} D(Y) &= E(Y^2) - (EY)^2 \\ &= (-2)^2 \cdot 0.1 + (-1)^2 \cdot 0.2 + 0^2 \cdot 0.4 + 1^2 \cdot 0.2 + 2^2 \cdot 0.1 - 0 \\ &= 1.2 \end{aligned}$$

甲乙平均误差一样, 但甲的稳定性高。

本题考察方差的意义。

例 7.3.4 (拓广探索 8, 难度: ★★). 设 $E(X) = \mu$, a 是不等于 μ 的常数, 探究 X 相对于 μ 的偏离程度与 X 相对于 a 的偏离程度的大小关系, 并说明结论的意义。

解:

X 相对于 μ 的偏离程度即为方差 $D(X) = E(X^2) - (EX)^2 = E(X^2) - \mu^2$, X 相对于 a 的偏离程度如下:

$$\begin{aligned} E[(X-a)^2] &= E[(X-\mu+(\mu+a))^2] \\ &= E[X^2 - 2 \cdot X \cdot a + a^2] \\ &= E(X^2) - 2a \cdot EX + E(a^2) \\ &= E(X^2) - 2a \cdot \mu + a^2 \end{aligned}$$

易得:

$$\begin{aligned} D(X) - E[(X-a)^2] &= [E(X^2) - \mu^2] - [E(X^2) - 2a \cdot \mu + a^2] \\ &= -\mu^2 + 2a \cdot \mu - a^2 \\ &= -(\mu - a)^2 < 0 \end{aligned}$$

意义可以根据转动惯量说明, 略。

本题考查方差的意义——偏离程度, 抓住这一点, 就可以列出“相对于 a 的偏离程度”的表达式。

7.4 二项分布与超几何分布

本节要点：

- 掌握二项分布及其意义；
- 掌握超几何分布及其意义。

7.4.1 二项分布

二项分布又称伯努利分布，记作 $B(n, p)$ ，因其分布律表达式中包含二项式系数而得名，描述 n 重伯努利试验中具有发生概率 p 的事件发生 x 次的概率的分布规律：

$$P\{X = x\} = C_n^x p^x (1-p)^{n-x}, \quad \begin{array}{l} x = 0, 1, \dots, n \\ p \in (0, 1) \end{array}$$

其中：

- x ：表示事件发生的次数，包括一次也不发生（ $x = 0$ ）和每次都发生（ $x = n$ ）；
- p ：表示事件发生的概率；
- C_n^x ： n 次子试验中，事件发生了 x 次的次数，注意，这里用组合，表示我们只关心 x 次，而不关心具体是哪 x 次；
- p^x ：表示事件发生 x 次的概率；
- $(1-p)^{n-x}$ ：表示其余 $n-x$ 次子试验中事件不发生的概率。

伯努利试验中，小概率事件倾向于少发生，大概率事件倾向于多次发生，且发生概率最大的在 np 次。

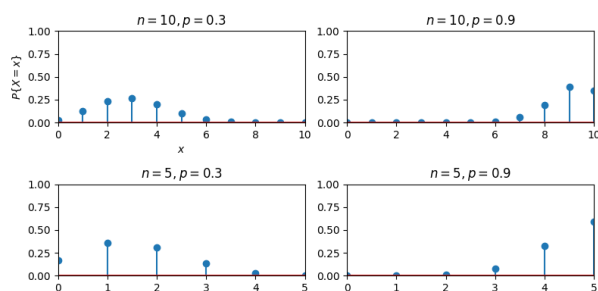


图 7.1: 二项分布

二项分布描述伯努利试验，我们可以知道在固定 n 次子试验中，事件 A 发生几次的概率是最大的。最典型的应用就是 n 次放回抽样，计算抽到 x 个次品的概率。

7.4.2 超几何分布

超几何分布，记作 $H(M, N, n)$ ，因其分布律与超几何函数的级数展开式的系数类似而得名，描述不放回抽取试验，从 N 个样品中抽取 n 个，出现 x 个次品的概率分布：

$$P\{X=x\} = \frac{C_{N-M}^{n-x} C_M^x}{C_N^n}, \quad N > 1, M \leq N, n \leq N, x \in [\max\{0, M+n-N\}, \min\{M, n\}]$$

其中：

- N, M ：共 N 个样品，其中有 M 个次品；
- n, x ：抽取 n 个，其中 x 个为次品。

超几何分布中，抽到 $\frac{M}{N} \cdot n$ 的概率最大。

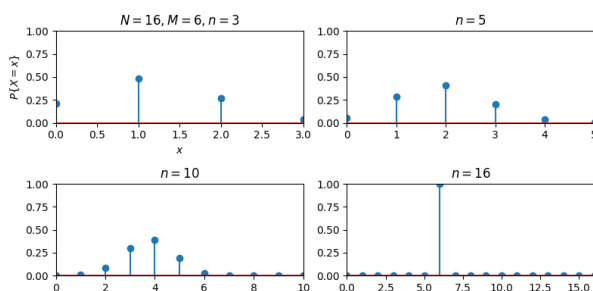


图 7.2: 超几何分布，16 个样品，6 个次品

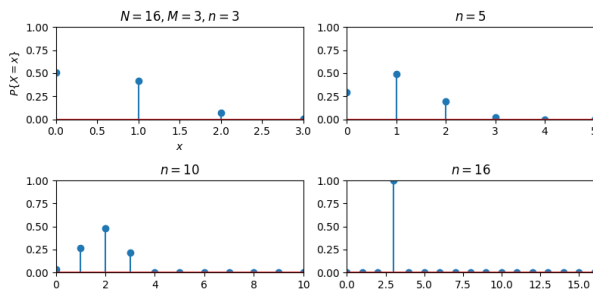


图 7.3: 超几何分布，16 个样品，3 个次品

超几何分布描述“不放回”抽取试验，不是基于伯努利试验，描述抽到 x 个次品的概率分布。

7.4.3 习题

例 7.4.1 (综合运用 7, 难度: ★). 一个车间有 3 台车床, 它们各自独立工作。设在一段时间内发生故障的车床数为 X , 在下列两种情形下分别求 X 的分布列。

1. 假设这 3 台车床型号相同, 它们发生故障的概率都是 20%;
2. 这 3 台车床中有 A 型号 2 台, B 型号 1 台, A 型车床发生故障的概率为 10%, B 型车床发生故障的概率为 20%。

解:

(1) 典型的二项分布:

$$P\{X=0\} = C_3^0 \cdot 0.2^0 \cdot (1-0.2)^{3-0} = 0.512$$

$$P\{X=1\} = C_3^1 \cdot 0.2^1 \cdot (1-0.2)^{3-1} = 0.384$$

$$P\{X=2\} = C_3^2 \cdot 0.2^2 \cdot (1-0.2)^{3-2} = 0.096$$

$$P\{X=3\} = C_3^3 \cdot 0.2^3 \cdot (1-0.2)^{3-3} = 0.008$$

(2) 不是二项分布, 需要分别计算:

$$P\{X=0\} = (1-0.1) \cdot (1-0.1) \cdot (1-0.2) = 0.648$$

$$P\{X=1\} = 0.1 \cdot (1-0.1) \cdot (1-0.2)$$

$$+ (1-0.1) \cdot 0.1 \cdot (1-0.2) + (1-0.1) \cdot (1-0.1) \cdot 0.2 = 0.306$$

$$P\{X=2\} = 0.1 \cdot 0.1 \cdot (1-0.2) + 0.1 \cdot (1-0.1) \cdot 0.2 + (1-0.1) \cdot 0.1 \cdot 0.2 \\ = 0.044$$

$$P\{X=3\} = 0.1 \cdot 0.1 \cdot 0.2 = 0.002$$

本题考察二项分布的概念, 没有难度。

例 7.4.2 (拓广探索 8, 难度: ★★★). 某药厂研制一种新药, 宣称对治疗某种疾病的有效率为 90%, 随机选择了 10 名患者, 经过使用该药治疗后, 治愈的人数不超过 6 人, 你是否怀疑药厂的宣传?

解:

假设新药的有效率 90% 为真, 则治愈人数不超过 6 人的概率:

$$P\{X=0\} = C_{10}^0 \cdot 0.9^0 \cdot (1-0.9)^{10-0} = 1 \times 10^{-10}$$

$$P\{X=1\} = C_{10}^1 \cdot 0.9^1 \cdot (1-0.9)^{10-1} = 9 \times 10^{-9}$$

$$P\{X=2\} = C_{10}^2 \cdot 0.9^2 \cdot (1-0.9)^{10-2} = 3.645 \times 10^{-7}$$

$$P\{X=3\} = C_{10}^3 \cdot 0.9^3 \cdot (1-0.9)^{10-3} = 8.748 \times 10^{-6}$$

$$P\{X=4\} = C_{10}^4 \cdot 0.9^4 \cdot (1-0.9)^{10-4} = 0.000137781$$

$$P\{X=5\} = C_{10}^5 \cdot 0.9^5 \cdot (1-0.9)^{10-5} = 0.00148803$$

$$P\{X=6\} = C_{10}^6 \cdot 0.9^6 \cdot (1-0.9)^{10-6} = 0.0111603$$

$$\sum_{i=0}^6 P\{X=i\} = 0.0127952$$

也即在有效率 90% 的情况下治愈人数不超过 6 人的概率仅为 1.28%, 显然这个有效率 90% 是小概率事件, 有理由怀疑宣传的真实性。

本题是数理统计的检验假设问题, 稍微有点超纲。

7.5 正态分布

正态分布, 也称高斯分布, 记作 $N(\mu, \sigma^2)$, 类似泊松分布, 描述一个范围内的分布情况:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \mu \in \mathbb{R}, \sigma > 0$$

其中:

- μ : 数学期望, 规定了中心点的横坐标, $f(x)$ 以 $x = \mu$ 为对称, 并在 $x = \mu$ 取得最大值 $\frac{1}{\sqrt{2\pi}\sigma}$;
- σ : 标准差, 一方面规定了 $f(x)$ 的两个拐点 $\mu \pm \sigma$, 另一方面也规定了平坦度, σ 越大, 曲线越平坦。

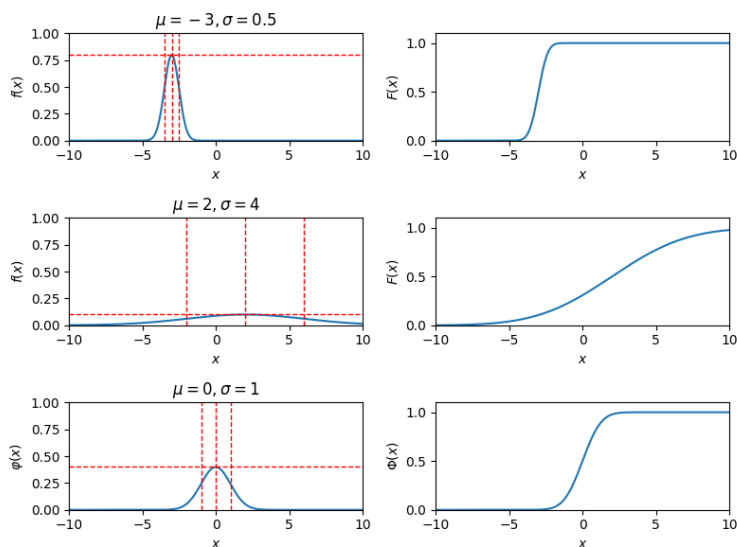
特别地, 当 $\mu = 0, \sigma = 1$ 时, 称为标准正态分布, 记作 $N(0, 1)$, 此时密度函数和分布函数通常记为 $\varphi(x), \Phi(x)$, 有:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) \cdot dt$$

根据密度函数的偶函数定理, 有性质:

$$\Phi(x) + \Phi(-x) = 1$$

$$\Phi(0) = 0.5$$



正态分布常用于描述身高、体重、成绩、元器件指标等的离散程度。

定理 7.5.1. 若 $X \sim N(\mu, \sigma^2)$, 则有:

$$P\{a < X \leq b\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

可用积分变量替换进行证明, 略。该定理使得我们可以用标准分布计算任何正态分布。

定理 7.5.2. 设 $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, 则 $Y = X_1 + X_2$ 也服从正态分布, 且有:

$$Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

$$f(y) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{[y - (\mu_1 + \mu_2)]^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

反之, 若正态分布 Y 能表示成两个变量的和 $Y = X_1 + X_2$, 则 X_1, X_2 也必然是正态分布, 这点被称为正态分布的再生性。

正态分布有很多特别的表现。一方面, 若足够多的独立分布叠加, 则总体可看成正态分布; 另一方面, 在相同方差的情况下, 正态分布具有最大的不确定性。所以正态分布不但非常常见, 而且也是优先考虑使用的模型。

7.5.1 习题

例 7.5.1 (综合运用 4, 难度: ★). 袋装食盐标准质量为 400g, 规定误差的绝对值不超过 4g 就认为合格。假设误差服从正态分布, 随机抽取 100 袋食盐, 误差的样本均值为 0, 样本方差为 4。请你估计这批袋装食盐的合格率。

解:

设每袋食盐的质量为随机变量 X , 则 $X \sim N(0, 4^2)$, 误差的绝对值不超过 4g 就认为合格, 于是食盐质量在 ± 4 内的概率为:

$$P\{X \in [-4, 4]\} = P\{X \in [-\sigma, \sigma]\} = 0.6827$$

后略。

本题考察正态分布的概念, 没有难度。

7.6 本章小结

本章介绍了随机变量和分布。我们将事件和概率进行数学上的抽象，上升为函数，用函数刻画概率，描绘生活中常见的概率的分布规律。

常见分布如下。

分布	含义、用途
二项分布 $B(n, p)$	n 重伯努利试验中发生 x 次的概率分布。
超几何分布 $H(M, N, n)$	不放回抽取中，抽到含 x 个次品的概率分布。
正态分布 $N(\mu, \sigma^2)$	描述某指标的离散度。

此外还有几何分布、均匀分布、指数分布等，略。

7.6.1 习题

例 7.6.1 (复习巩固 3, 难度: ★). 假设有两箱零件，第一箱内装有 10 件，其中有 2 件次品；第二箱内装有 20 件，其中有 3 件次品。现从两箱中随意挑选一箱，然后从该箱中随机取 1 个零件。

- 1. 求取出的零件是次品的概率；
- 2. 已知取出的是次品，求它是从第一箱取出的概率。

解：

(1) 全概率问题，易得：

$$P(A) = 0.5 \cdot \frac{2}{10} + 0.5 \cdot \frac{3}{20} = 0.175$$

(2)

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{0.5 \cdot \frac{2}{10}}{0.175} = 0.571429$$

本题考察全概率公式和条件概率，没有难度。

例 7.6.2 (复习巩固 5, 难度: ★★). 已知随机变量 X 取所有的值 $1, 2, \dots, n$ 是等可能的，且 $E(X) = 10$ ，求 n 的值。

解:

$$\begin{aligned} E(X) &= 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \cdots + n \cdot \frac{1}{n} \\ &= \frac{1}{n} \cdot \sum_{i=1}^n i = \frac{1}{n} \cdot \frac{n(1+n)}{2} = \frac{1+n}{2} = 10 \end{aligned}$$

得 $n = 19$ 。

本题结合了数列的知识。

例 7.6.3 (复习巩固 6, 难度: ★). 已知每门大炮击中目标的概率都是 0.3, 现在 n 门大炮同时对某一目标各射击一次。

1. 当 $n = 10$ 时, 求恰好击中目标 3 次的概率 (精确到 0.001);
2. 如果使目标至少被击中一次的概率超过 95%, 至少需要多少门大炮?

解:

(1) 是典型的二项分布, 易得:

$$P\{X = 3\} = C_{10}^3 \cdot 0.3^3 \cdot (1 - 0.3)^{10-3} = 0.266828$$

(2) 题目也即没有击中的概率小于 5%, 有:

$$\therefore P\{X = 0\} = C_n^0 \cdot 0.3^0 \cdot (1 - 0.3)^{n-0} < 0.05$$

$$\therefore 1 \cdot 1 \cdot 0.7^n < 0.05$$

$$\therefore n > \log_{0.7} 0.05 = \frac{\ln 0.05}{\ln 0.7} = 8.39905$$

至少需要 9 门。

本题 (2) 转换成不击中能极大减小计算量。

例 7.6.4 (综合运用 9, 难度: ★). 假设一份某种意外伤害保险费为 20 元, 每次赔付金额为 50 万元。一家保险公司一年能销售 10 万份保单, 而每一份保单需要赔付的概率为 10^{-5} 。利用计算工具求 (精确到 0.0001):

1. 这家保险公司在这个险种上亏本的概率;
2. 这家保险公司在这个险种上一年内获利不少于 100 万元的概率?

解:

本题为典型的二项分布问题。

(1) 易得收益 $20 \cdot 100000 = 2000000$, 也即出现 4 次以上 (不包含 4 次) 赔付就亏本, 于是:

$$P\{X=0\} = C_{100000}^0 \cdot (10^{-5})^0 \cdot (1-10^{-5})^{100000-0} = 0.367878$$

$$P\{X=1\} = C_{100000}^1 \cdot (10^{-5})^1 \cdot (1-10^{-5})^{100000-1} = 0.367881$$

$$P\{X=2\} = C_{100000}^2 \cdot (10^{-5})^2 \cdot (1-10^{-5})^{100000-2} = 0.183941$$

$$P\{X=3\} = C_{100000}^3 \cdot (10^{-5})^3 \cdot (1-10^{-5})^{100000-3} = 0.0613129$$

$$P\{X=4\} = C_{100000}^4 \cdot (10^{-5})^4 \cdot (1-10^{-5})^{100000-4} = 0.0153279$$

$$1 - \sum_{i=0}^4 P\{X=i\} = 1 - 0.99634 = 0.00365962$$

亏本的概率为 0.366%。

(2) 即赔付 0、1 次的概率:

$$\sum_{i=0}^1 P\{X=i\} = 0.367878 + 0.367881 = 0.735759$$

本题考察二项分布的概念。

例 7.6.5 (拓广探索 10, 难度: ★★★). 甲、乙、丙三人相互做传球训练, 第 1 次由甲将球传出, 每次传球时, 传球者都等可能地将球传给另外两个人中的任何一人。求 n 次传球后球在甲手中的概率。

解:

即求 $P\{X=n\}$ 的表达式, 假设我们知道这个表达式了, 则 $n+1$ 次后球在甲手中的概率:

$$P\{X=n+1\} = \frac{1-P\{X=n\}}{2} \cdot \frac{1}{2} \cdot 2 = \frac{1-P\{X=n\}}{2}$$

其中, $1-P\{X=n\}/2$ 表示乙丙两人每人得球概率, 除以 2 表示他们各自都有一半的可能传给甲, 再乘以 2 表示两人一共传给甲的概率。稍作化简:

$$P\{X=n+1\} + C = -\frac{1}{2} \cdot (P\{X=n\} - 2C - 1)$$

$$C = -2C - 1$$

$$C = -\frac{1}{3}$$

$$P\{X=n+1\} - \frac{1}{3} = -\frac{1}{2} \cdot \left(P\{X=n\} - \frac{1}{3}\right)$$

可见 $P\{X=n\} - 1/3$ 是一个等比数列, 于是:

$$\begin{aligned}\because P\{X=1\} - \frac{1}{3} &= 0 - \frac{1}{3} = -\frac{1}{3} \\ \therefore P\{X=n\} - \frac{1}{3} &= \left(-\frac{1}{3}\right) \cdot \left(-\frac{1}{2}\right)^{n-1} \\ \therefore P\{X=n\} &= \left(-\frac{1}{3}\right) \cdot \left(-\frac{1}{2}\right)^{n-1} + \frac{1}{3}\end{aligned}$$

本题是马尔科夫链, 略有超纲, 但结合数列的知识还是能解。

例 7.6.6 (拓广探索 11, 难度: ★★). 某单位有 10000 名职工, 想通过验血的方法筛查乙肝病毒携带者。假设每人携带乙肝病毒的概率为 5%, 如果对每人的血样逐一化验, 就需要化验 10000 次。统计专家提出了一种化验方法: 随机地按 5 人一组分组, 然后将各组 5 人的血样混合再化验。如果混合血样呈阴性, 说明这 5 人全部阴性; 如果混合血样呈阳性, 说明其中至少有一人的血样呈阳性, 就需要对每人再分别化验一次。

1. 按照这种化验方法能减少化验次数吗?
2. 如果每人携带乙肝病毒的概率为 2%, 按照 k 人一组, k 取多大时化验次数最少?

解:

(1) 可以通过考察数学期望判断化验次数。易得单位有 500 名携带者, 于是 5 人一组存在阳性的概率:

$$1 - P\{X=0\} = 1 - 0.95^5 = 0.226219$$

化验次数:

$$\frac{10000}{5} + 5 \cdot 0.226219 = 2001.13$$

可见确实能减少次数。

(2) 概率为 2% 时, 有 200 人携带, 于是 k 人一组的化验次数:

$$\begin{aligned}P(k) &= 1 - P\{X=0\} = 1 - 0.98^k \\ f(k) &= \frac{10000}{k} + P(k) \cdot k = \frac{10000}{k} + k - 0.98^k \cdot k\end{aligned}$$

即求函数 $f(k)$ 的最值, 求导即可, 有点复杂, 略。

$$f'(x) = \frac{-10000}{k^2} + 1 - 0.98^k (\ln 0.98 \cdot k + 1)$$

本题关键捋清分组的过程。

第八章 成对数据的统计分析

本章粗略介绍数理统计的相关知识。

本章要点：

- 相关性。
- 线性回归。
- 独立性。

8.1 成对数据的统计相关性

详细了解相关性需要微积分和线性代数的知识，高中课本只是粗略介绍，XML。可详见我的《概率论与数理统计——学习笔记》。

值得一提的是 P98 将相关和向量类比，反复阅读体会。

8.2 一元线性回归模型及其应用

详细了解相关性需要微积分和线性代数的知识，高中课本只是粗略介绍，XML。可详见我的《概率论与数理统计——学习笔记》。

8.3 列联表与独立性检验

详细了解相关性需要微积分和线性代数的知识，高中课本只是粗略介绍，XML。可详见我的《概率论与数理统计——学习笔记》。

8.4 扩展知识：概率论和数理统计的知识结构

概率论和数理统计这门学科，国外分为两本教材，分别是“概率论”和“数理统计”，国内作为理工科入门教材一般都合在一本。学科分为两部分，概率论和数理统计，前者是后者的理论基础，后者是前者的工程应用。

概率论一般分 4 大部分：

- 随机事件：主要介绍概率的相关知识，如条件概率、全概率、贝叶斯公式、古典概型、几何概型等；
- 随机变量：将随机事件抽象，用函数这一工具讨论概率，主要讲随机变量的分布，如二项分布、泊松分布、正态分布等，若干个随机变量的关系，如边缘分布、条件分布、独立性等；
- 数字特征：讲述数学期望、方差、协方差、相关系数、矩等；
- 大数定理和中心极限定理：作为概率论的补完。

数理统计一般分 3 大部分：

- 基础：主要介绍基本概念，如总体、样本、抽样；
- 参数估计：讲述如何用样本估计总体的参数，如估计总体的期望、方差等，以及对估计方法的评价；
- 假设检验：对总体的一些假设，我们如何用已知的样本进行对这个假设的判断，或者说我们如何用数学的方法定量地判断一个假设的可信度。

可见，高中选修教材第三册中，除了本章第三节是属于数理统计的假设检验部分，其他都是概率论部分的内容。概率论和数理统计虽然视作数学，但其实是建立在微积分和线性代数的基础上。若缺乏这两个基础，所有知识点都不可能进行有效深入地讨论。

8.5 本章小结

本章介绍了两个随机变量的相关概念——相关性，线性回归拟合，独立性检验。