国家教材委员会专家委员会审核通过
2019

普通高中教科书

数学

选择性必修

第二册

人民教育出版社

B版

普 通 高 中 教 科 书

数学

选择性必修

第二册

人民教育出版社　课程教材研究所
中 学 数 学 教 材 实 验 研 究 组　编著

人民教育出版社
·北京·

B版

主　　编：高存明
副 主 编：王殿军　龙正武　王旭刚
本册主编：赵　亮　罗声雄
其他编者：王雅琪　李现勇　毕政之　程凤霞　谭国文　周　琳

普通高中教科书　数学（B 版）　选择性必修　第二册
人民教育出版社　课程教材研究所
中 学 数 学 教 材 实 验 研 究 组　编著

出　　版　人民教育出版社
（北京市海淀区中关村南大街 17 号院 1 号楼　邮编：100081）
网　　址　http://www.pep.com.cn

前　言

人们喜欢音乐，是因为它拥有优美和谐的旋律；人们喜欢美术，是因为它描绘了人和自然的美；人们喜欢数学，是因为它用空间形式和数量关系刻画了自然界和人类社会的内在规律，用简洁、优美的公式与定理揭示了世界的本质，用严谨的语言和逻辑梳理了人们的思维……

我国著名数学家华罗庚先生曾经指出：数学是一切科学的得力助手和工具；任何一门科学缺少了数学这一工具便不能确切地刻画出客观事物变化的状态，更不能从已知数据推出未知的数据来，因而就减少了科学预见的可能性，或者减弱了科学预见的准确度.

事实上，任何一项现代科学技术的出现与发展，背后都一定有数学知识的支撑. 互联网的普及、共享经济的繁荣、网络支付的便利、物联网的兴起、人工智能的发展、大数据的应用，离开了数学知识都是不可能的！并且，现代生活中，类似“逻辑”“函数”“命题”“线性增长”“指数增长”“概率”“相关性”等数学术语，在政府文件、新闻报道中比比皆是.

正如《普通高中数学课程标准（2017 年版）》（以下简称“课程标准”）所指出的：数学在形成人的理性思维、科学精神和促进个人智力发展的过程中发挥着不可替代的作用. 数学素养是现代社会每一个人应该具备的基本素养. 高中生学习必需的数学知识，能为自身的可持续发展和终身学习奠定基础.

为了帮助广大高中生更好地学习相关数学知识，我们按照课程标准的要求编写了这套高中数学教材. 在编写过程中，我们着重做了以下几项工作.

1. 关注学生成长，体现时代特征

教材在选取内容的背景素材时，力图从学生熟悉的情境出发，着力体现时代特征，并为学生的成长提供支撑. 例如，以下内容在本套教材中都有所体现：利用数学知识破解魔术的“秘密”，用生活中的例子说明学习逻辑知识以及理性思考的重要性，从数学角度理解报刊上有关人工智能、新兴媒体等报道中出现的“线性增长”“爆炸式增长”等名词.

教材中还提到了“网络搜索”“人工智能”“自主招生”“环境保护”“大数据”“按揭贷款”“电子商务”“创业创新”等. 我们相信，这些能引起大家的共鸣.

此外，教材中多处出现了借助现代信息技术学习数学知识的内容，包括怎样借助数学软件解方程、不等式，怎样借助信息技术呈现统计结果、展示模拟过程，等等.

在体现时代特征的同时，我们也特别注重对中华优秀传统文化的展示. 例如，教材中精选了多道我国古代数学名题，启发大家从数学角度去理解“失败乃成功之母”“三个臭皮匠，顶个诸葛亮”等语句的含义，呈现了与二十四节气、古典诗词等有关的调查数据，介绍了《九章算术》在代数上的成就以及我国古代的统计工作，等等.

2. 吸收先进理念，改变呈现方式

在教材编写过程中，编者认真学习和讨论了当前教育学、心理学等学科的先进理念，并通过改变教材呈现方式来加以体现，力图真正做到“以学习者为中心”.

例如，教材每一章都引用了一段名人名言，旨在为大家的数学学习提供参考和指引；通过“情境与问题”栏目，展示相关数学知识在现实生活等情境中的应用；利用“尝试与发现”栏目，鼓励大家大胆尝试，并在此基础上进行猜想、归纳与总结；通过填空的方式，培养大家学习数学的信心；选择与内容紧密联系的专题，设置拓展阅读，以拓宽大家的知识面，了解数学应用的广泛性；等等.

3. 遵循认知规律，力求温故知新

数学学习必须循序渐进是一种共识. 基础不扎实是很多人学不好数学的重要原因. 本套教材在编写时特别考虑了这一点.

事实上，教材一方面按照课程标准的要求，讲解和复习了高中数学必备的集合、等式、不等式等内容；另一方面，在呈现新知识时，教材注重从已有知识出发，在回顾的基础上通过实际例子逐步引入，尽力展现新旧知识的联系，以达到温故知新的效果.

例如，教材在复习了变量以及初中函数概念的基础上介绍了函数中的对应关系，在回顾了整数指数幂、二次根式等后引入了分数指数幂，等等.

正因为如此，即使是初中数学基础比较薄弱的同学，使用本套教材也能顺利地进行学习，并最终达到理想的效果. 这在本套教材试教过程中已得到印证.

4. 揭示内容本质，重视直观理解

数学知识具有客观性，但数学知识的理解有多种方式与途径. 揭示内容本质，培养大家对数学内容的直观理解，是我们编写本套教材时特别注意的方面之一.

首先，教材内容的安排突出主线，强调“通性通法”. 例如，多次强调了配方法的使用，自始至终贯彻函数的研究应从特殊到一般、从性质到图象，等等.

其次，尽量自然地引入新内容或新方法. 例如，通过实例说明学习中位数、百分位数的必要性，通过对比说明用样本估计总体的合理性，等等.

最后，注重培养大家的数学学科核心素养. 课程标准提出的数学抽象、逻辑推理、数学建模、直观想象、数学运算和数据分析，在教材中都得到了落实. 仅以数学抽象为例，教材处处强调了自然语言与符号语言之间的相互转化等.

总的来说，“引导学生会用数学眼光观察世界，会用数学思维思考世界，会用数学语言表达世界”并不容易. 为此，我们在编写教材时做了很多新的尝试，力图给大家提供一套有时代特色、易教易学的数学教材，以帮助大家学习.

本书是这套教材选择性必修部分的第二册，呈现了排列、组合与二项式定理，概率与统计的内容. 通过本书的目录与每章的“本章导语”，可以大致了解本书的全貌，这里不再重复.

由于编写时间有限等原因，书中难免会有疏漏之处，敬请大家多提宝贵意见，以使教材日臻完善.

编者

2019 年 4 月

目录

本书拓展阅读目录

正规的数学训练很有必要. 好的基础训练对人一生的影响是很大的，因为它很大程度上决定了一个人将来思维的习惯，可以帮助克服轻率，至少会很严谨.

——张益唐

第三章

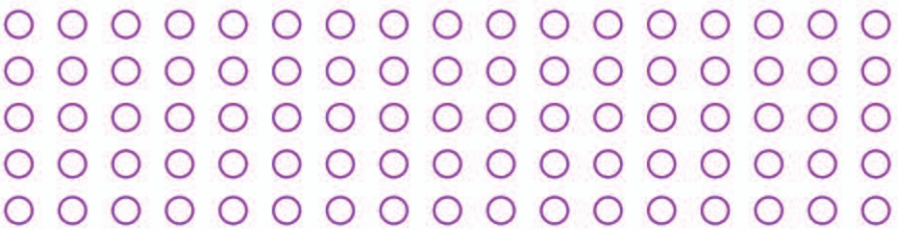
排列、组合与二项式定理

本章导语

“排列”“组合”这两个词语或许大家并不陌生，比如，大家可能看到过“按字母顺序排列”“士兵们排列在两旁”“这本书由诗、散文、短篇小说三部分组合而成”等. 首先要说明的是，我们这一章里所要学习的“排列”“组合”与“计数”有关，同上述句子中对应词语的意义有差别.

“计数”就是数事物的个数，这是数学学科发展的起点，也是我们从小学开始就在学习的. 可以说，随着大家掌握的内容越来越多，我们的计数能力也变得越来越强大.

小学一年级，我们是通过“1，2，3，4，5，…”这种不断加 1 的方法来计数的. 学习了乘法之后，我们就可以借助乘法来计数了. 比如，要数出下图中圆的个数，你会怎么数呢?



相信大家肯定不会一个一个地数，而是会先数出每一行有 17 个，每一列有 5 个，最后得出共有 $17\times5=85$ 个.

我们这里要学习的“排列”“组合”是更强大的计数方法，利用它们可以快速地解决一些看起来很难的计数问题. 比如，高考不分文理科后，思想政治、历史、地理、物理、化学、生物这 6 大科目是选考的，如果考生可以从中任选 3 科作为自己的高考科目，那么选考的组合方式一共有多少种可能的情况呢?

排列、组合的知识还与我们后续要学习的概率知识密切相关，它们能帮助我们解决一些复杂的概率问题.

二项式定理讨论的是 $(a+b)^n$ 在 n 为任意正整数时的展开式 ($n=2$ 的情形大家已经很熟悉了吧?)，利用二项式定理不仅可以证明“$99^{98}-1$ 能被 100 整除”，而且还能帮助我们研究一类重要的概率问题. 你知道吗? 我国古代数学家发现二项式定理的有关结论比西方早好几百年呢!

3.1 排列与组合

3.1.1 基本计数原理

情境与问题

在数学学习和日常生活中，我们经常会遇到类似“共有多少种情况”的计数问题. 例如：

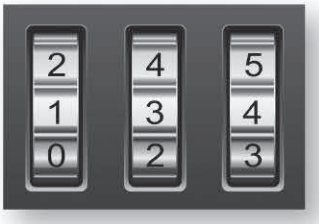


图3-1-1

(1) 一个由 3 个元素组成的集合，共有多少个不同的子集？

(2) 由 3 个数字组成的密码锁，如图 3-1-1 所示，如果忘记了密码，最多要试多少次才能打开密码锁？

(3) 有 4 位同学和 1 位老师站成一排照相，如果老师要站在正中间，如图 3-1-2 所示，则有多少种不同的站法？

图3-1-2

你能解答上述问题吗？对于比较简单的计数问题，我们可以通过列举法求得结果，例如上述的问题 (1)；但是，如果问题比较复杂，那么只借助列举法可能就难以求得问题的答案了，例如上述的问题 (2) 和 (3). 有没有

其他方法可以帮助我们计数呢？答案是肯定的.

1. 分类加法计数原理

尝试与发现

你能解答下述两个问题吗？试着由此归纳出一般的规律.

(1) 已知某天从北京到上海的G字头列车有43班，D字头列车有2班，其他列车有3班，小张想在这一天坐火车从北京到上海旅游，不考虑其他因素，小张有多少种不同的选择？

(2) 从甲地到乙地，可以乘火车，也可以乘汽车，还可以乘轮船. 假定火车每日有1班，汽车每日有3班，轮船每日有2班，那么一天中从甲地到乙地有多少种不同的走法呢？

尝试与发现中的问题(1)，小张乘坐的列车可以分为3类，即G字头列车、D字头列车或其他列车，其中任何一类的任何一班车都可以让小张从北京到达上海，因此不同的选择有

$$43+2+3=48$$

种.

类似地，问题(2)中，从甲地到乙地，可乘坐三类交通工具：火车、汽车或轮船，每类交通工具又各有若干个班次，选择其中任何一类的任何一个班次都可以从甲地到达乙地，因此一天中不同的走法有

1

种.

把上述解法推广到一般情况，就可以得出：

分类加法计数原理 完成一件事，如果有n类办法，且：第一类办法中有m_1种不同的方法，第二类办法中有m_2种不同的方法……第n类办法中有m_n种不同的方法，那么完成这件事共有

$$N=m_1+m_2+\cdots+m_n$$

种不同的方法.

例1 在某设计活动中，李明要用红色和蓝色填涂四个格子（如图3-1-3所示），要求每种颜色都用两次，李明共有多少种不同的填涂方法？

图3-1-3

尝试与发现

试给出一种满足条件的涂法，在明确要完成的事情是什么的前提下思考：

（1）怎样用符号表示填涂结果？

（2）可以将填涂结果分类吗？

解 用 R 表示红色，用 B 表示蓝色，RBRB 表示第一个和第三个格子涂红色，第二个和第四个格子涂蓝色.

因为红色和蓝色都要用两次，为了简化问题，考虑涂红色的格子是否相邻，则填涂结果可以分为两类：涂红色的格子相邻，涂红色的格子不相邻.

涂红色的格子相邻的方法有：RRBB，BRRB，BBRR，共 3 种；

涂红色的格子不相邻的方法有：RBRB，BRBR，RBBR，共 3 种.

依据分类加法计数原理，李明共有

2

种不同的涂法.

2. 分步乘法计数原理

尝试与发现

已知某公园的示意图如图 3-1-4 所示，其中从西门到景点 A 共有 3 条不同的路，从景点 A 到东门共有两条不同的路. 王瑞从公园的西门进入公园后，想去 A 景点游玩，然后从东门出公园. 只考虑路的选择，王瑞共有多少种不同的走法？你能用适当的符号表示出所有的情况吗？

图 3-1-4

如果把从西门到景点 A 的三条路分别记为 a_1，a_2，a_3，把从景点 A 到东门的路记为 b_1，b_2，用 a_1b_1 表示王瑞经 a_1 到景点 A，然后经 b_1 到东门. 注意到不管王瑞选择哪条路到景点 A，其去东门都有两种不同的选择方法，因此不同的走法为

a_1b_1，a_1b_2，3 ，

共有 6 种. 可以看出，这里的 6 能看成 3 和 2 的乘积，即

$$3\times 2=6,$$

这样，不同的走法可以用图 3-1-5 直观地表示出来.

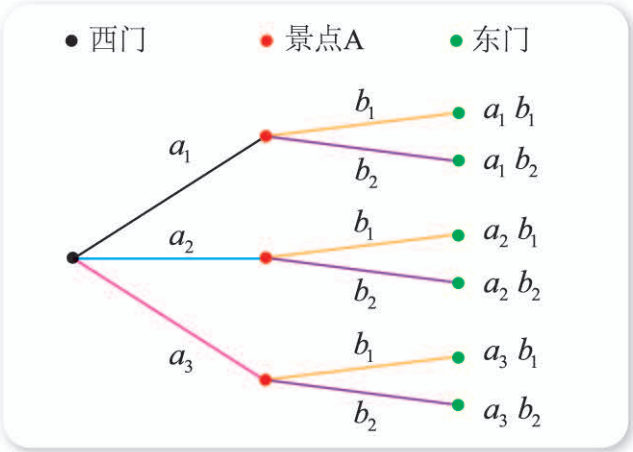


图 3-1-5

把上述解法推广到一般情况，就可以得出：

分步乘法计数原理 完成一件事，如果需要分成 n 个步骤，且：做第一步有 m_1 种不同的方法，做第二步有 m_2 种不同的方法……做第 n 步有 m_n 种不同的方法，那么完成这件事共有

$$N=m_1\times m_2\times\cdots\times m_n$$

种不同的方法.

例 2 用 1，2，3，4，5 可以排成多少个数字不重复的三位数?

分析 要排成一个三位数，只需分别指定这个三位数的百位、十位、个位上的数字即可，因此可以分为三步完成.

解 排成一个三位数，可以分为三步：

第一步，确定百位上的数字，共有 5 种方法；

第二步，确定十位上的数字，因为数字不能重复，所以不能是百位上已有的数字，共有 4 种方法；

第三步，确定个位上的数字，共有 3 种方法.

依据分步乘法计数原理，可以排成数字不重复的三位数的个数为

$$5\times 4\times 3=60.$$

想一想

例 2 的解答有其他的分步方法吗? 如果有，得到的结果一样吗?

本节一开始的情境与问题中的问题 (2)，可借助分步乘法计数原理来解答，因为共有 3 位数字，每一位数字都有 10 种可能，所以密码的设定方法共有

$$10\times 10\times 10=1\ 000$$

种. 这就意味着，遗忘密码时，最多要试 1 000 次才能打开密码锁.

情境与问题中的问题（3），可以转化为将 4 位同学安排在从左到右的 4 个位置上，也可以借助分步乘法计数原理解答，请读者自行完成.

分类加法计数原理和分步乘法计数原理合称为**基本计数原理**.

3. 基本计数原理的应用

在有关问题的解决中，我们往往需要综合使用分类加法计数原理和分步乘法计数原理.

例 3 某班班委由 2 位女同学、3 位男同学组成，现要从该班班委里选出 2 人去参加学校组织的培训活动，要求至少要有 1 位女同学参加，则不同的选法共有多少种？

解 按照选择的女同学人数分为两种情况，即 2 位都是女同学和只有 1 位女同学.

2 位都是女同学的选法显然只有 1 种.

只有 1 位女同学的选法，可以分为两步完成：先从 2 位女同学中选出 1 人，有 2 种选法；再从 3 位男同学中选出 1 人，有 3 种选法. 依据分步乘法计数原理，共有不同的选法 $2\times3=6$ 种.

依据分类加法计数原理，不同的选法共有

4

种.

值得注意的是，例 3 中的选择，不能分为如下两步来完成：首先选择 1 位女同学，然后在剩下的 4 人中选择 1 人. 事实上，如果用 a_1，a_2 代表 2 位女同学，b_1，b_2，b_3 代表 3 位男同学，则这种分步的结果可用图 3-1-6 表示，你能看出其中的问题吗？

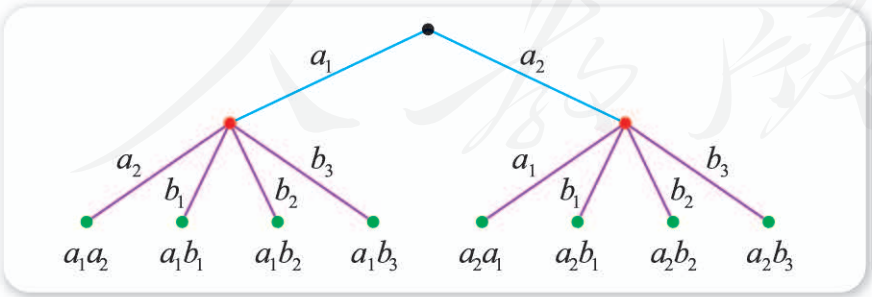


图 3-1-6

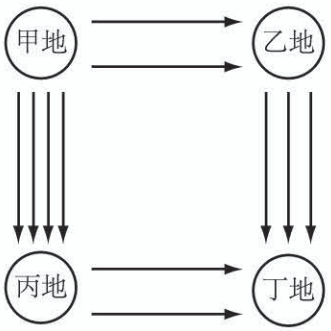
练习A

1. 张丽的书桌上有 3 本不同的语文课外读物和 2 本不同的数学课外读物. 现在她想从中取出一本随身携带，以便外出时阅读，有多少种不同的取法？如果她想从语文课外读物和数学课外读物中各取一本随身携带，有多少种不同的取法？
2. 用 0，1，2，…，9 这十个数字，可以组成多少种不同的银行卡密码？（每个银行卡密码均由六位数字组成，数字可以重复，不考虑其他因素.）
3. (1) 用 1，2，3，4，5，6 可以排成多少个数字不重复的两位数？

   (2) 用 1，2，3，4，5，6 可以排成多少个数字可以重复的两位数？
4. 将代数式$(x+y+z)(a+b+c+d+e)$展开后，共有多少项？
5. 某城市电话局管辖范围内的电话号码由八位数字组成，其中前面四位数字是固定的，后面四位的每一个数字都是 0 到 9 这十个数字中的任意一个. 该电话局管辖范围内的不同的电话号码最多能有多少个？

练习B

1. 如图所示，从甲地到乙地有 2 条路，从乙地到丁地有 3 条路；从甲地到丙地有 4 条路，从丙地到丁地有 2 条路. 要从甲地去丁地，共有多少种不同的走法？



（第 1 题）

2. 已知 A，B，C，D，E 这五位司机中，A，B 既能开大客车，也能开小客车，但 C，D，E 这三位司机都只能开小客车. 现要从这五位司机中选用两人，分别去开一辆大客车和一辆小客车，共有多少种不同的选用方法？
3. 已知 n 是一个小于 10 的正整数，且由集合$A=\{x \mid x \in \mathbf{N}_+, x \leqslant n\}$中的元素可以排成数字不重复的两位数共 20 个，求 n 的值.
4. 如图所示，把硬币有币值的一面称为正面，有花的一面称为反面. 抛一次硬币，得到正面记为 1，得到反面记为 0. 现抛一枚硬币 5 次，按照每次的结果，可得到由 5 个数组成的数组（例如，若第一、二、四次得到的是正面，第三、五次得到的是反面，则结果可记为(1，1，0，1，0)），则可得不同的数组共有多少个？


正面　　反面
（第 4 题）

5. 已知 A 是一个有限集，且 A 中的元素个数为 n，求 A 的子集的个数.

1 $1+3+2=6$　2 $3+3=6$　3 a_2b_1，a_2b_2，a_3b_1，a_3b_2　4 $1+6=7$

3.1.2 排列与排列数

1. 排列与排列数

尝试与发现

试解答下列三个计数问题：

(1) 小张要在 3 所大学中选择 2 所，分别作为自己的第一志愿和第二志愿，小张共有多少种不同的选择方式？

(2) 在 3 名学生中选出 2 名，分别在某话剧表演中扮演 A 和 B 两个角色，共有多少种不同的选择方式？

(3) 学校要在 3 名教师中指派 2 人，分别去上海和浙江交流教学经验，共有多少种不同的指派方案？

它们的答案是否一致？

如果用 A，B，C 分别表示上述问题 (1) 中的三所大学，用(A，B)表示第一志愿是 A，第二志愿是 B，你能列出小张所有的选择方式吗？上述问题 (2) (3) 的结果是否也能用类似的方法表示？

不难看出，以上三个问题虽然实际背景不同，但所求的本质上都是“从 3 个不同对象中选出 2 个并排成先后顺序，有多少种不同的排法”，因此它们的答案肯定是一致的. 事实上，根据分步计数原理可知，方法种数都是 1 .

一般地，从 n 个不同对象中，任取 m ($m \leqslant n$) 个对象，按照一定的顺序排成一列，称为从 n 个不同对象中取出 m 个对象的一个**排列**. 特别地，$m=n$ 时的排列 (即取出所有对象的排列) 称为**全排列**.

上述尝试与发现的问题 (1) 中，用(A，B)表示第一志愿是 A，第二志愿是 B，则(A，B)就是一个排列. 两个排列，如果组成排列的对象是相同的，并且对象的排列顺序也相同，那么就称这两个排列是相同的；否则，就称为是不同的. 因此，(A，B)与(A，C)是不同的排列，(A，B)与(B，A)也是不同的排列.

从 n 个不同对象中取出 m 个对象的所有排列的个数，称为从 n 个不同对象中取出 m 个对象的**排列数**，用符号 A_n^m ①表示.

① A 是英语单词 arrangement (排列) 的第一个字母的大写.

注意：(1) 所谓排成一列，是指与顺序有关，例如，排列 AB 与排列 BA 是不同的排列，可以把一个排列看成一个类似点坐标的有序数对.

(2) 符号 A_n^m 中，总是要求 n 和 m 都是正整数，且 $m \leqslant n$，以后不再声明.

前面三个计数问题实际上就是求 A_3^2，我们已经知道，$A_3^2=6$.

一般情况下，A_n^m 等于多少呢?

A_n^1 等于从 n 个不同对象中取出 1 个的方法种数，显然 $A_n^1=n$.

A_n^2 等于从 n 个不同对象中取出 2 个并排成先后顺序的方法种数. 分两步完成：第一步，选一个排在第一个位置，有 n 种选法；第二步，在剩下的对象中选一个排在第二个位置，有 $n-1$ 种选法. 因此共有 $n(n-1)$种选法，即 $A_n^2=n(n-1)$.

用类似的方法可知

$$A_n^3=n(n-1)(n-2),$$

$$A_n^4=n(n-1)(n-2)(n-3),$$

……

一般地，我们有

$$A_n^m=\underbrace{n(n-1)\cdots[n-(m-1)]}_{m\text{个数}}=n(n-1)\cdots(n-m+1),$$

这个公式称为**排列数公式**.

例如，$A_5^3=5\times4\times3=60$，$A_4^4=$ 2 .

例 1 求从 A，B，C 这 3 个对象中取出 3 个对象的所有排列的个数，并写出所有的排列.

解 所求排列数为 $A_3^3=3\times2\times1=6$.

所有的排列可用图 3-1-7 表示.

图 3-1-7

由图 3-1-7 可知，所有排列为 ABC，ACB，3 .

例 1 计算的其实是 3 个对象的全排列数.

一般地，在 A_n^m 中，当 $m=n$ 时，排列数公式为

$$A_n^n=n\times(n-1)\times\cdots\times2\times1,$$

通常将上式的右边简写为 $n!$（读作“n 的阶乘”），从而上式可以简写为

$$A_n^n=n!.$$

例如，如果 5 名同学要排成一排照相，那么不同的排法种数为

$$A_5^5=5!=\boxed{4}\ \_\_\_\_\_.$$

当 $0<m<n$ 时，注意到

$$n!=\underbrace{n\times(n-1)\times\cdots\times[n-(m-1)]}_{A_n^m}\times\underbrace{(n-m)\times[n-(m+1)]\times\cdots\times2\times1}_{(n-m)!},$$

所以此时排列数公式可以改写为

$$A_n^m=\frac{n!}{(n-m)!}.$$

为了使得上式对 $m=n$ 时也成立，我们规定 $0!=1$. 另外，为了方便起见，也规定 $A_n^0=1$.

例 2 求证：$A_n^m+mA_n^{m-1}=A_{n+1}^m$.

证明 由排列数公式可知

$$\begin{aligned}A_n^m+mA_n^{m-1}&=\frac{n!}{(n-m)!}+m\,\frac{n!}{[n-(m-1)]!}\\&=\frac{n!}{(n-m)!}\times\left[1+\frac{m}{n-(m-1)}\right]\\&=\frac{n!}{(n-m)!}\times\frac{n+1}{n-(m-1)}\\&=\frac{(n+1)!}{[(n+1)-m]!}=A_{n+1}^m.\end{aligned}$$

探索与研究

假设有 $n+1$ 个不同的对象，甲是其中一个，从这 $n+1$ 个对象中取出 m 个做成的排列，可以分成两类：

(1) 不包括对象甲的；

(2) 包括对象甲的.

分别计算每一类的排列个数.

注意到第 (2) 类排列中，甲可以占据 m 个位置中的任何一个，也就是说，甲的位置有 m 种可能. 你能由此给出例 2 的结果的一个直观解释吗？

2. 排列数的应用

例 3 某地区足球比赛共有 12 个队参加，每队都要与其他各队在主客场分别比赛一次，则共要进行多少场比赛？

解 如果把每一场比赛都看成主场队在前、客场队在后的一个排列，则不难看出，所求比赛数等于从 12 个对象中取出 2 个的排列数，即

$$A_{12}^2=12\times11=132.$$

例 3 的关键是，把所给问题转化为等价的排列问题.

例 4 某信号兵用红、黄、蓝三面旗从上到下挂在竖直的旗杆上表示信号，每次可以只挂 1 面旗，也可以挂 2 面旗或 3 面旗，旗数或顺序不同时，表示信号不同，则一共可表示多少种不同的信号？

解 按照所挂旗数，可以分为三类：

第一类是只挂 1 面旗，此时可表示 A_3^1 种不同的信号；

第二类是挂 2 面旗，此时可表示 A_3^2 种不同的信号；

第三类是挂 3 面旗，此时可表示 A_3^3 种不同的信号.

按照分类加法计数原理，一共可表示不同的信号

$$A_3^1+A_3^2+A_3^3=3+3\times2+3\times2\times1=15$$

种.

例 4 说明，解题过程中，可以将基本计数原理与排列知识有机结合.

例 5 用 0，1，2，…，9 这 10 个数字，可以排成多少个没有重复数字的三位数？

解 （方法一）要组成一个没有重复数字的三位数，可以分为两步：

第一步，确定百位上的数字，因为只能是 1，2，…，9 这 9 个数字中的某一个，所以有 A_9^1 种方法；

第二步，确定十位和个位上的数字，因为数字不能重复，所以只能从百位以外的数字来选取，因此共有 A_9^2 种方法.

由分步乘法计数原理可知，满足条件的三位数个数为

$$A_9^1A_9^2=9\times9\times8=648.$$

（方法二）从 0，1，2，…，9 这 10 个数字中，取出 3 个做排列的排列数为 A_{10}^3. 所有的这些排列中，0 排在首位的都不能对应一个三位数，而其他的都对应一个三位数. 又因为 0 排在首位的排列共有 A_9^2 个，所以可知所求三位数的个数为

$$A_{10}^3-A_9^2=10\times9\times8-9\times8=648.$$

例 5 的方法二，通常称为“排除法”，也就是先算出无限制条件的所有排法种数，然后再减去不符合条件的排法种数.

例 6 用 0，1，2，…，9 这 10 个数字，可以排成多少个没有重复数字的四位偶数？

尝试与发现

给出几个满足条件的四位数，并对所有满足条件的四位数进行分类.

解 满足条件的四位数可以分为两类：

第一类的末位数字是 0，有 A_9^3 个.

第二类的末位数字不是 0. 要排成这样的四位数，可以分成三个步骤来完成：第一步，确定末位数字，因为只能是 2，4，6 或 8，所以有 A_4^1 种方法；第二步，确定首位数字，因为数字不能重复，所以有 A_8^1 种方法；第三步，确定中间两位数字，有 A_8^2 种方法. 由分步乘法计数原理可知，这样的数字有 $A_4^1A_8^1A_8^2$ 个.

由分类加法计数原理可知，满足条件的四位数个数为

$$A_9^3+A_4^1A_8^1A_8^2=9\times8\times7+4\times8\times8\times7=41\times56=2\ 296.$$

从例 6 可以看出，利用排列数公式，可以简化思维过程.

例 7 有 3 位男生和 2 位女生，在某风景点前站成一排拍合照，要求 2 位女生要相邻，有多少种不同的站法？

尝试与发现

用适当的符号表示男生和女生，给出几种满足条件的排法，由此尝试发现解题思路.

解 分成两步来完成：第一步，先让两位女生站好，有 A_2^2 种方法；第二步，把两位女生当成一个整体，与 3 位男生去站成一排，有 A_4^4 种方法. 根据分步乘法计数原理可知，共有 $A_2^2A_4^4=48$ 种不同的站法.

例 7 的解法，相当于把两位女生捆绑在了一起，因此也常被称为“捆绑法”.

例 8 某晚会要安排 3 个歌唱节目（记为 A，B，C）和 2 个舞蹈节目（记为甲、乙），要求舞蹈节目不能相邻，共有多少种不同的安排方法？

尝试与发现

用题中的符号给出几种满足条件的排法，由此尝试发现解题思路.

解 分成两步来完成：第一步，先确定 3 个歌唱节目的先后顺序（不考虑舞蹈节目），总共有 A_3^3 种排法；第二步，歌唱节目的先后顺序确定之后，舞蹈节目共有 A_4^2 种排法（例如，如果第一步确定的歌唱节目先后顺序为 ABC，则舞蹈节目只能安排在如图 3-1-8 所示的 4 个空格中）. 由

分步乘法计数原理可知，共有

5

种不同的安排方法.

图 3-1-8

值得注意的是，例 8 中所有符合条件的安排方法都可用解法中的方式得到，如“AB 甲 C 乙”，只要在图 3-1-8 中的第三个、第四个空格分别填上甲、乙即可. 这种解题方法通常称为“插空法”. 在解决类似的要求不相邻的问题中，用插空法往往简单、有效.

3. 用信息技术计算排列数

利用 Excel 软件中的 PERMUT(permutation，排列)命令可以计算排列数. 例如，要计算 A_6^4，只要在任意一个单元格输入“=PERMUT(6，4)”，如图 3-1-9 所示，然后按回车键，就能显示出想要的结果，如图 3-1-10 所示.

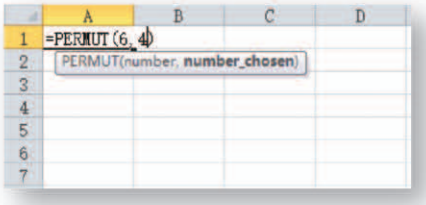


图 3-1-9

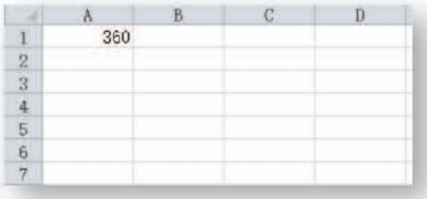


图 3-1-10

在 GeoGebra 中，输入“nPr[6，4]”或“排列数[6，4]”也可得到 A_6^4 的值，请感兴趣的读者自行尝试.

练习A

1. 写出所有由 1，2，3，4 这四个数字排成的没有重复数字的四位数.
2. 计算：

   (1) A_4^3；　(2) A_6^3；　(3) A_{15}^1；　(4) A_{20}^2；　(5) A_{100}^2.
3. 计算 1～8 的阶乘，并填入下表中：

n	1	2	3	4	5	6	7	8
$n!$								

4 从 5 种不同的蔬菜品种中选出 2 种分别种植在不同土质的土地上进行试验，共有多少种不同的种植方法？

5 从 5 名乒乓球运动员中，选出 3 名并确定出场顺序，以参加某场团体比赛，共有多少种不同的方法？

6 有 6 个人想在某风景区门口站成前后两排（各 3 人）照相，共有多少种不同的排法？

练习B

1 计算：

(1) $A_8^5-2A_8^2$；　(2) $A_4^1+A_4^2+A_4^3+A_4^4$；　(3) $\dfrac{A_{10}^5}{A_5^5}$.

2 (1) 将 2 封不同的信投入 4 个邮箱，每个邮箱最多投 1 封，共有多少种不同的投法？

(2) 将 2 封不同的信随意投入 4 个邮箱，共有多少种不同的投法？

3 用 0，1，2，3，4，5 可组成多少个：

(1) 没有重复数字的四位数？

(2) 没有重复数字且被 5 整除的四位数？

(3) 比 2 000 大且没有重复数字的自然数？

4 四对夫妇坐成一排照相：

(1) 每对夫妇都不能被隔开的排法有多少种？

(2) 每对夫妇都不能被隔开，且同性别的人不能相邻的排法有多少种？

5 将 2 个男生和 4 个女生排成一排：

(1) 男生排在中间的排法有多少种？

(2) 男生不在头尾的排法有多少种？

(3) 男生不相邻的排法有多少种？

(4) 男生不相邻且不在头尾的排法有多少种？

(5) 2 个男生都不与女生甲相邻的排法有多少种？

1 $3\times2=6$　2 $4\times3\times2\times1=24$　3 BAC，BCA，CAB，CBA

4 $5\times4\times3\times2\times1=120$　5 $A_3^3A_4^2=6\times12=72$

3.1.3 组合与组合数

情境与问题

高考不分文理科后，思想政治、历史、地理、物理、化学、生物这 6 大科目是选考的，如果考生可以从中任选 3 科作为自己的高考科目，那么选考的组合方式一共有多少种可能的情况呢？

如果用｛思想政治，历史，地理｝表示其中一种选考的组合，你能用类似的方法表示出所有的组合方式吗？你有更简单的表示方法吗？

上述问题可以用本小节我们要学习的组合知识来解.

1. 组合与组合数

尝试与发现

下面这两个计数问题的答案一样吗？

(1) 小张要在 3 所大学中选择 2 所，分别作为自己的第一志愿和第二志愿，小张共有多少种不同的选择方式？

(2) 小张要在 3 所大学中选择 2 所，作为自己努力的目标，小张共有多少种不同的选择方式？

选择合适的符号，分别表示出上述两题中所有的选择方式，并总结两者之间的关系.

不难看出，尝试与发现的两个问题中：前者选出两所学校后，还要指定一所作为第一志愿，另一所作为第二志愿；而后者只需要选出两所学校即可. 换句话说，前者选出的学校是要排列顺序的，而后者选出的学校不需要排列顺序.

这也就是说，尝试与发现 (1) 中的事情，可以分成两步来完成：第一步，完成 (2) 中的事情，即选择两所学校；第二步，将选出的两所学校做全排列（共有 A_2^2 种方法）. 因为问题 (1) 的答案是 A_3^2，所以如果设问题 (2) 的答案是 x，那么根据上述分析和分步乘法计数原理可知 $A_3^2 = xA_2^2$，从而

$$x = \frac{A_3^2}{A_2^2} = \underline{\quad 1 \quad}.$$

一般地，从 n 个不同对象中取出 m（$m\leqslant n$）个对象并成一组，称为从 n 个不同对象中取出 m 个对象的一个**组合**.

尝试与发现的问题（2）中，如果用 A，B，C 表示 3 所学校，{A，B} 表示选择学校 A 和 B 作为目标，则 {A，B} 就是一个组合，且（2）中的 3 种选择方式也就是 3 种组合分别为：{A，B}，{A，C}，{B，C}.

从 n 个不同对象中取出 m 个对象的所有组合的个数，称为从 n 个不同对象中取出 m 个对象的**组合数**，用符号 C_n^m[①]表示.

注意：（1）所谓并成一组是指与顺序无关，例如，组合 a，b 与组合 b，a 是同一组合，可以把一个组合看成一个集合.

（2）同符号 A_n^m 一样，在符号 C_n^m 中，总是要求 n 和 m 都是正整数，且 $m\leqslant n$，以后也不再声明.

尝试与发现的问题（2）实际上就是求组合数 C_3^2，我们已经知道，$C_3^2=\dfrac{A_3^2}{A_2^2}=3$.

实际生活中，有大量涉及组合数的情况. 例如，从 15 名学生中选择 3 人去参加学生代表大会，共有 C_{15}^3 种不同的选择方法；从 10 名教师中指派 7 人去参加教学交流，有 C_{10}^7 种不同的指派方法等.

尝试与发现

仿照求出 C_3^2 的过程，探讨一般情况下，组合数 C_n^m 该怎样计算.

考虑到从 n 个不同对象中取出 m 个做排列，可以分成两个步骤来完成：第一步，从 n 个不同对象中取出 m 个，有 C_n^m 种选法；第二步，将选出的 m 个对象做全排列，有 A_m^m 种排法. 由分步乘法计数原理有 $A_n^m=C_n^mA_m^m$，所以

$$C_n^m=\frac{A_n^m}{A_m^m}=\frac{n(n-1)\cdots[n-(m-1)]}{m\times(m-1)\times\cdots\times2\times1}=\frac{n!}{(n-m)!m!}.$$

上述公式称为**组合数公式**.

尝试与发现

由组合数公式，分别取 $m=0$，$m=1$，$m=n$，可得

$C_n^0=$ 2 ，$C_n^1=$ 3 ，$C_n^n=$ 4 .

试利用组合的概念直观地理解上述特殊组合数.

① C 是英语单词 combination（组合）的第一个字母的大写.

例 1 已知一个平面内有 10 个点，其中任意 3 点都不共线，且任意两点所连成的线段中，任意两条线段的长度都不相等：

（1）这些点共可以连成多少条不同的线段？

（2）以这些点为端点共可以作出多少个不同的非零向量？

解 （1）因为已知的点中，任意 3 点都不共线，而任意两点都能连成一条线段，所以共可以连成的不同线段条数为

$$C_{10}^{2}=\frac{10\times 9}{2\times 1}=45.$$

（2）因为以任意一点为始点、另一点为终点，均可作出一个非零向量，而且连成的所有线段中，任意两条线段的长度都不相等，因此共可以作出不同的非零向量个数为

$$A_{10}^{2}=10\times 9=90.$$

例 2 计算：

（1）$C_{7}^{3}+C_{7}^{4}$；　　（2）$C_{10}^{5}C_{10}^{0}-C_{10}^{10}$.

解 （1）$C_{7}^{3}+C_{7}^{4}=\frac{7\times 6\times 5}{3\times 2\times 1}+\frac{7\times 6\times 5\times 4}{4\times 3\times 2\times 1}=35+35=70.$

（2）$C_{10}^{5}C_{10}^{0}-C_{10}^{10}=\frac{10\times 9\times 8\times 7\times 6}{5\times 4\times 3\times 2\times 1}\times 1-1=252-1=251.$

本节一开始的情境与问题中，不难看出，选考组合方式共有

$$C_{6}^{3}=\frac{6\times 5\times 4}{3\times 2\times 1}=20$$

种可能的情况. 若用字母或数字表示科目的名称（如下表所示），并用 {T，H，G} 表示选考思想政治、历史、地理，则可以方便地表示出所有的选考组合.

科　目	思想政治	历史	地理	物理	化学	生物
字　母	T	H	G	P	C	B

2. 组合数的性质

尝试与发现

在了解敬老院可以进行哪些爱心活动的走访中，老师要将 5 位同学分成两组，一组 2 人，另一组 3 人. 老师完成分组，有两种不同的做法：

（1）选出 2 人作为一组，另外 3 人是另一组；

（2）选出 3 人作为一组，另外 2 人是另一组.

用组合数符号分别表示（1）和（2）所得的分法种数，说明所得结果之间的关系，并将结果推广到一般情况.

根据组合和组合数公式可知，尝试与发现中（1）和（2）所得的分法种数分别为 C_5^2 和 C_5^3，而且

$$C_5^2=\boxed{5}\quad，\quad C_5^3=\boxed{6}\quad.$$

因此 $C_5^2=C_5^3$.

一般地，我们有

$$C_n^m=\frac{n!}{(n-m)!m!},$$

$$C_n^{n-m}=\frac{n!}{[n-(n-m)]!(n-m)!}=\frac{n!}{m!(n-m)!},$$

因此

$$C_n^m=C_n^{n-m}.$$

想一想

不难知道，从 n 个对象中，取出 m 个对象后，将剩下 $n-m$ 个对象. 你能用这一事实来直观理解有关结论吗？

从这一性质与前面计算 C_5^2 和 C_5^3 的过程可知，当 $m>\frac{n}{2}$ 时，将计算 C_n^m 转化为计算 C_n^{n-m} 会更简便，例如

$$C_{10}^7=C_{10}^{10-7}=C_{10}^3=\frac{10\times9\times8}{3\times2\times1}=120.$$

例 3 一个口袋里有 7 个不同的白球和 1 个红球，从中取 5 个球：

（1）共有多少种不同的取法？

（2）如果不取红球，共有多少种不同的取法？

（3）如果必须取红球，共有多少种不同的取法？

解 （1）因为共有 8 个球，所以共有不同的取法种数为 C_8^5，且

$$C_8^5=C_8^{8-5}=C_8^3=\frac{8\times7\times6}{3\times2\times1}=56.$$

（2）因为不取红球，所以只要在 7 个白球中取 5 个球即可，所以共有不同的取法种数为

$$C_7^5=C_7^{7-5}=C_7^2=\frac{7\times6}{2\times1}=21.$$

（3）因为必须取红球，所以只需在 7 个白球中再取 4 个球即可，所以共有不同的取法种数为

$$C_7^4=C_7^{7-4}=C_7^3=\frac{7\times6\times5}{3\times2\times1}=35.$$

观察例 3 的解答可知 $C_7^5+C_7^4=C_8^5$，这一结论是否具有普遍性呢？答案是肯定的，事实上，我们有（证明留作练习）

$$C_n^{m+1}+C_n^m=C_{n+1}^{m+1}.$$

探索与研究

假设有 $n+1$ 个不同的对象，甲是其中一个，从这 $n+1$ 个对象中选出 $m+1$ 个的组合，可以分成两类：

(1) 不包括对象甲的；

(2) 包括对象甲的.

你能用这一事实直观地理解上述组合数的性质吗？

3. 组合数的应用

在从事产品检验时，经常要从产品中抽取一部分进行检查，这其中就牵涉很多计数问题.

例 4 现有 30 件分别标有不同编号的产品，且除了 2 件次品外，其余都是合格品，从中取出 3 件：

(1) 一共有多少种不同的取法？

(2) 若取出的 3 件产品中恰有 1 件次品，则不同的取法共有多少种？

(3) 若取出的 3 件产品中至少要有 1 件次品，则不同的取法共有多少种？

解 (1) 所求的取法总数，就是从 30 件产品中取出 3 件的组合数

$$C_{30}^{3}=\frac{30\times29\times28}{3\times2\times1}=4\ 060.$$

(2) 抽取可以分成两步完成：第一步，在 2 件次品中取出 1 件，有 C_2^1 种方法；第二步，在 28 件合格品中取出 2 件，有 C_{28}^2 种方法. 因此取法种数为

$$C_2^1C_{28}^2=2\times\frac{28\times27}{2\times1}=756.$$

(3) 满足条件的取法可以分成两类：恰有 1 件次品的取法和恰有 2 件次品的取法.

恰有 1 件次品的取法有 $C_2^1C_{28}^2$ 种，恰有 2 件次品的取法有 $C_2^2C_{28}^1$ 种.

因此取法种数为

$$C_2^1C_{28}^2+C_2^2C_{28}^1=2\times\frac{28\times27}{2\times1}+1\times28=784.$$

例 4 说明，解题过程中，可以将基本计数原理与组合知识有机结合.

例 5 要把 9 本不同的课外书分给甲、乙、丙 3 名同学：

(1) 如果每个人都得 3 本，则共有不同的分法多少种？

(2) 如果要求一人得 4 本，一人得 3 本，一人得 2 本，则共有不同的

分法多少种？

解 （1）要完成分配任务，可以分为三步：第一步，分给甲 3 本书，有 C_9^3 种方法；第二步，分给乙 3 本书，因为只能在剩下的 6 本书里选，所以有 C_6^3 种方法；第三步，分给丙 3 本书，因为只能在剩下的 3 本书里选，所以有 C_3^3 种方法．因此共有不同的分法数为

$$C_9^3C_6^3C_3^3=\frac{9\times8\times7}{3\times2\times1}\times\frac{6\times5\times4}{3\times2\times1}\times1=1\ 680.$$

（2）要完成分配任务，可以分为两步：第一步，将 9 本书按照 4 本、3 本、2 本分为三组，有 $C_9^4C_5^3C_2^2$ 种方法；第二步，将分好的 3 组书分别分给 3 个人，有 A_3^3 种方法．因此共有不同的分法数为

$$C_9^4C_5^3C_2^2A_3^3=\frac{9\times8\times7\times6}{4\times3\times2\times1}\times\frac{5\times4}{2\times1}\times1\times3\times2\times1=7\ 560.$$

例 5 的（2）中，因为没有指定谁得 4 本书，谁得 3 本书，谁得 2 本书，所以第二步需要做一个排列．

例 6 现要从 A，B，C，D，E，F 这 6 人中选出 4 人安排在甲、乙、丙、丁 4 个岗位上，如果 A 不能安排在甲岗位上，那么一共有多少种不同的安排方法？

解 安排方法可以分成两类：选出的 4 人中有 A 和没有 A．

有 A 的安排方法可以分成两步完成：第一步，在乙、丙、丁 3 个岗位中选择一个给 A，共 C_3^1 种方法；第二步，在 B，C，D，E，F 这 5 人中选出 3 人安排在其他 3 个岗位上，共 A_5^3 种方法．所以此类安排方法共有 $C_3^1A_5^3$ 种．

没有 A 的安排方法共有 A_5^4 种．

因此安排方法种数为

$$C_3^1A_5^3+A_5^4=300.$$

4.用信息技术计算组合数

利用 Excel 软件中的 COMBIN(combination，组合)命令可以计算排列数．例如要计算 C_{100}^3，只要在任意一个单元格输入“=COMBIN(100，3)”，如图 3-1-11 所示，然后按回车键，就能显示出想要的结果，如图 3-1-12 所示．

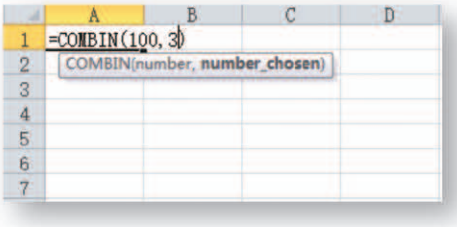


图 3-1-11

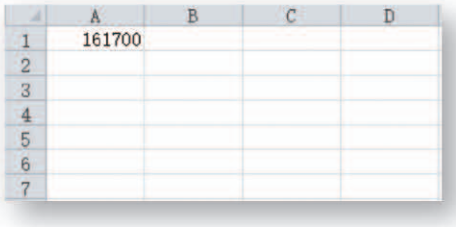


图 3-1-12

在 GeoGebra 中，输入“Binomial[100，3]”或“二项式系数[100，3]”也可得到 C_{100}^3 的值，请感兴趣的读者自行尝试.

拓展阅读

把相同的物品分给不同对象的分法种数

把 8 个相同的篮球分发给甲、乙、丙、丁 4 人，共有多少种不同的分法？

由于每个篮球都相同，因此只要指出每人所得篮球的个数即可，比如，甲得 2 个、乙得 3 个、丙得 3 个、丁得 0 个，就是一种满足条件的分法. 可能有人会想到通过列举来求解上述问题，但是，经过简单的尝试之后，你就会发现，这个问题可能比想象中的难.

注意到每一种满足条件的分法本质上就是把 8 个球分为了 4 堆，为此可借助 3 块隔板来实现. 例如，前述满足条件的分法可以用图 1 表示，其中第一块隔板前的篮球是分给甲的，第一块和第二块隔板之间的篮球是分给乙的，第二块和第三块隔板之间的篮球是分给丙的，第三块隔板后的篮球是分给丁的.

图 1

容易知道，任何一种类似图 1 的排列都对应一种分法，例如，图 2 对应的分法为：甲得 1 个，乙得 0 个，丙得 0 个，丁得 7 个.

图 2

这样一来，问题就转化为 8 个相同的篮球和 3 块相同的隔板，可以有多少种不同的排列方法.

因为总共有 8＋3＝11 个位置，而且我们只需要从这 11 个位置中选出 3 个放置隔板（其余放置篮球）即可，因此不同的排列方法种数为

$$C_{11}^3=\frac{11\times10\times9}{3\times2\times1}=165.$$

也就是说，我们有 165 种不同的分法.

有意思的是，如果设甲、乙、丙、丁 4 人所得篮球个数分别为 x_1，x_2，x_3，x_4，则不难看出，我们得到了方程

$$x_1+x_2+x_3+x_4=8$$

的非负整数解$(x_1，x_2，x_3，x_4)$个数为 165.

类似地，可以得到把 n 个相同的物品分给 r 个不同对象的方法数（其中 r 和 n 均为正整数），也就是方程 $x_1+x_2+\cdots+x_r=n$ 的非负整数解$(x_1，x_2，\cdots，x_r)$的个数，请自己尝试一下吧！

练习A

1. 北京队、上海队、天津队、广东队四个足球队举行友谊比赛，每两个队要比赛一场：

   (1) 列出所有各场比赛的双方；

   (2) 最终产生冠、亚军各一个队，列出所有可能的冠、亚军情况.

2. 写出：

   (1) 从 a，b，c，d，e 五个元素中取两个元素的所有组合；

   (2) 从 a，b，c，d，e 五个元素中取三个元素的所有组合.

3. 计算：

   (1) C_{17}^{1}；　(2) C_{6}^{3}；　(3) C_{23}^{0}；　(4) C_{100}^{98}.

4. 某校举行排球赛，每两个队赛一场，有 8 个队参加，共需比赛多少场？

5. 现有 10 件产品（除了 2 件一等品外，其余都是二等品），从中抽取 3 件：

   (1) 一共有多少种不同的抽法？

   (2) 抽出的 3 件中恰有 1 件一等品的抽法共有多少种？

   (3) 抽出的 3 件中至少有 1 件一等品的抽法共有多少种？

练习B

1. 计算：

   (1) $C_{7}^{3}+C_{7}^{4}+C_{8}^{5}+C_{9}^{6}+C_{10}^{7}$；　(2) $C_{5}^{0}+C_{5}^{1}+C_{5}^{2}+C_{5}^{3}+C_{5}^{4}+C_{5}^{5}$.

2. 解方程：$C_{18}^{x}=C_{18}^{3x-6}$.

3. 利用组合数公式证明 $C_{n}^{m+1}+C_{n}^{m}=C_{n+1}^{m+1}$.

4. 甲、乙、丙、丁、戊五名同学参加某项竞赛，决出了第一名到第五名的 5 个名次. 甲、乙两人去询问成绩，组织者对甲说："很遗憾，你和乙都未拿到冠军."对乙说："你当然不会是最差的."从组织者的回答分析，这五名同学的名次排列共有多少种不同的情况.

5. 将 6 名中学生分到甲、乙、丙 3 个不同的公益小组：

   (1) 要求有 3 人分到甲组，2 人分到乙组，1 个人分到丙组，共有多少种不同的分法？

   (2) 要求三个组的人数分别为 3，2，1，共有多少种不同的分法？

1 $\dfrac{3\times2}{2\times1}=3$　2 $\dfrac{n!}{(n-0)!0!}=1$　3 $\dfrac{n!}{(n-1)!1!}=n$　4 $\dfrac{n!}{(n-n)!n!}=1$

5 $\dfrac{5\times4}{2\times1}=10$　6 $\dfrac{5\times4\times3}{3\times2\times1}=10$

习题3-1A

1. 有不同的红球 8 个，不同的白球 7 个：
   (1) 从中取出 1 个球，共有多少种不同的取法？
   (2) 从中取出 2 个颜色不同的球，共有多少种不同的取法？
2. 求下列各式中的正整数 n：
   (1) $A_{2n}^3=10A_n^3$；　　(2) $A_{10}^n=10\times9\times8\times7\times6\times5$.
3. 已知从 n 个不同对象中取出 2 个对象的排列数等于从 $n-4$ 个不同对象中取出 2 个对象的排列数的 7 倍，求正整数 n 的值.
4. 一部影片在 4 个单位轮流放映，每一个单位放映一场，共有多少种不同的放映次序？
5. (1) 已知圆上有 10 个点，过任意 3 个点都可画一个圆内接三角形，一共可画多少个圆内接三角形？
   (2) 已知空间中有 10 个点，且任意 4 个点都不共面，即以任意 4 个点为顶点都可构造一个四面体，则一共可以构造多少个四面体？
6. (1) 平面内有两组平行线，一组有 m 条，另一组有 n 条，不同组的平行线都相交，其中 m，n 都是大于 1 的正整数，这些平行线一共构成了多少个平行四边形？
   (2) 空间中有三组平行平面，第一组有 m 个，第二组有 n 个，第三组有 l 个，不同组的平面都互相垂直，其中 m，n，l 都是大于 1 的正整数，这些平行平面一共构成了多少个长方体？
7. 将 4 封不同的信全部投入 3 个邮筒：
   (1) 不加任何限制，有多少种不同的投法？
   (2) 每个邮筒至少投 1 封信，有多少种不同的投法？
8. 某乒乓球邀请赛，参加的有三个组，第一、第二组各有 7 个队，第三组有 6 个队，首先各组进行单循环赛，然后各小组的第一名共 3 个队分主客场进行决赛，最终决出冠、亚军，该乒乓球邀请赛一共需要比赛多少场？

习题3-1B

1. 某教师上午要给 3 个班上课，每班 1 节. 如果每个班上午只能排 4 节课，并且该教师不能连上 3 节课，那么该教师上午的课表有多少种不同的排法？
2. 在不小于 3 000 且不大于 7 000 的正整数中，有多少个没有重复数字的 5 的倍数？
3. 有 6 个人分成两排就座，每排 3 人：
   (1) 共有多少种不同的坐法？
   (2) 如果甲不能坐在第一排，乙不能坐在第二排，共有多少种不同的坐法？

(3) 如果甲和乙必须在同一排且相邻，共有多少种不同的坐法？

(4) 如果甲和乙必须在同一排且不相邻，共有多少种不同的坐法？

4 某班有 35 名学生，其中正、副班长各 1 名，现要从该班选派 5 名学生参加某种活动：

(1) 如果正、副班长必须在内，共有多少种不同的选派方法？

(2) 如果正、副班长必须有一人在内，且只能有一人在内，共有多少种不同的选派方法？

(3) 如果正、副班长都不在内，共有多少种不同的选派方法？

(4) 如果正、副班长至少有一人在内，共有多少种不同的选派方法？

5 有 6 个座位连成一排，安排 3 个人就座，恰有两个空位相邻的不同坐法共有多少种？

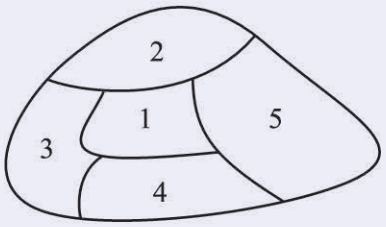
6 有 10 个人围着一张圆桌坐成一圈，共有多少种不同的坐法？

习题3-1C

1 求 $C_2^2+C_3^2+C_4^2+\cdots+C_{100}^2$ 的值.

2 求证：$A_m^m+A_{m+1}^m+A_{m+2}^m+\cdots+A_{2m}^m=A_{2m+1}^m$. (提示：考察排列数与组合数的关系.)

3 如图所示，一个地区分为 5 个行政区域，现给地图着色，要求相邻区域不得使用同一颜色，有 4 种颜色可供选择，则不同的着色方法共有多少种？



(第 3 题)

4 要把 9 本不同的课外书分别装到三个相同的手提袋里，每个袋中至少一本，一共有多少种不同的装法？

5 把分别标有 1 号、2 号、3 号、4 号的 4 个不同的小球放入分别标有 1 号、2 号、3 号的 3 个盒子中，不许有空盒子且任意一个小球都不能放入标有相同标号的盒子中，则不同的放法共有多少种？

3.2 数学探究活动：生日悖论的解释与模拟

1. 活动背景介绍与要求

假设你所在的班级共有 30 人，那么你们班至少有两个人生日相同的概率是多少？因为每个人的生日有可能是 365 天[①]中的任意一天，这样一来，只有人数超过 365 时，我们才能百分之百地肯定至少有两个人的生日相同，因此感觉上前述问题中的概率应该不会太大. 不过，令人惊讶的是，利用排列组合的知识可以算出，30 个人中，至少有两个人生日相同的概率约为 71%!

事实上，当人群的人数达到 23 时，至少有两个人生日相同的概率就超过 50%了！而当人数达到 41 时，概率就超过 90%了！这一结论与人们的直觉相差比较远，因此常被称为“生日悖论”.

生日悖论可以在日常生活中找到很多实例. 例如，2014 年世界杯中，有 32 支球队，每支球队恰好就有 23 名球员. 如果生日悖论是真的，可能会有半数球队拥有同生日球员. 从国际足联 2014 年 6 月 10 日给出的官方数据中可以看到，瑞士、伊朗、法国、阿根廷和韩国的代表队各有两对生日相同的球员；西班牙、哥伦比亚、美国、喀麦隆、澳大利亚、波黑、俄罗斯、荷兰、巴西、洪都拉斯和尼日利亚的代表队各有两名球员生日相同. 也就是说，32 支球队中，正好有 16 支球队至少有两人生日相同，所占比例正好为 50%!

也许大家还是会对生日悖论心存疑惑，因为在日常生活中，我们每个人很难遇到一个与自己生日相同的人. 再看以下事实：指定一年中的一天，253 个人中，才有 50%的概率能找到一个人的生日在指定的那天；要想使概率提高到 80%，需要 587 个人才行. 因此，如果你真遇到了一个跟你生日相同的人，那你们确实是“有缘”的. 需要注意的是，这里涉及的问题与生日悖论涉及的问题并不相同.

请与其他同学一起分工合作，完成下列任务，并填写活动记录表：

（1）通过世界杯球员的有关数据或其他数据，验证生日悖论是否属实.

（2）得出由 n 个人组成的人群中至少有两个人生日相同的概率计算公式.

（3）利用计算机软件或计算器，分别给出 $n=15, 16, \cdots, 60$ 时，（2）

① 为了简单起见，假设一年只有 365 天，下同.

中的概率值，并用适当的图象表示结果.

（4）选定一个特殊的 n 值，利用计算机软件模拟验证生日悖论中的概率.

（5）得出由 m 个人组成的人群中至少有一个人生日是指定日期的概率计算公式.

（6）利用计算机软件或计算器，分别给出 $m=200$，201，…，2 200 时，（5）中的概率值，并用适当的图象表示结果.

（7）选定一个日期和一个特殊的 m 值，利用计算机模拟验证（6）中的概率.

生日悖论的解释与模拟活动记录表

活动开始时间：_____

（1）成员与分工

姓　名	分　工

（2）验证生日悖论的实际数据

（3）n 个人组成的人群中至少有两个人生日相同的概率计算公式

（4）$n=15$，16，…，60 时，（3）中的概率值以及图象表示

（5）生日悖论模拟的方法与结果

续表

（6）m 个人组成的人群中至少有一个人生日是指定日期的概率计算公式
（7）$m=200, 201, \cdots, 2\ 200$ 时，（6）中的概率值以及图象表示
（8）模拟（7）中概率的方法及结果
（9）活动总结（可包括活动感受等）

活动结束时间：_____

2. 活动提示

（1）除了利用世界杯球员的数据验证生日悖论之外，也可利用学校中各班级的人员信息等.

（2）所要计算的概率都可借助古典概型来完成，其中需要借助排列组合的有关知识. 例如，n 个人组成的人群，生日的所有可能情形有 365^n 种，而这 n 个人生日各不相同的情形共有 A_{365}^n（$n\leqslant 365$）种，生日都不在某个指定日期的情形共有

$$(365-1)^n=364^n$$

种.

（3）计算机模拟可以借助随机函数来完成.

例如，在验证生日悖论时，可以用 Excel 中的随机函数随机产生多组数据，然后统计其中有哪些组出现了重复数据，最后计算比例.

在图 3-2-1 中，每一个有数据的单元格，输入的都是

“=RANDBETWEEN(1，365)”，

共产生了 20 组随机数，每一组都由 23 个数组成，每一组数中重复的数都标成了红色.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	243	364	121	120	87	100	273	189	211	15	321	239	365	342	185	136	51	263	9	257
2	44	290	268	100	103	50	305	170	148	112	348	49	305	156	138	6	250	343	105	204
3	288	364	226	337	14	211	150	319	105	108	45	266	242	266	17	135	117	295	147	160
4	227	336	127	80	236	13	341	299	48	198	142	128	130	250	280	234	186	345	269	136
5	79	26	304	148	311	128	115	114	192	263	225	229	288	48	157	42	48	50	6	216
6	96	213	83	136	78	266	180	329	75	256	340	136	65	13	200	271	210	285	133	106
7	60	122	363	44	50	64	48	108	167	216	265	38	349	42	139	173	58	141	180	192
8	99	21	72	9	38	240	206	270	232	112	279	205	131	202	234	228	347	19	294	208
9	37	297	40	292	68	75	23	283	217	105	268	301	244	39	106	76	105	84	138	260
10	15	300	345	51	16	245	100	197	326	28	317	337	254	254	323	115	38	148	293	31
11	107	35	335	190	162	194	248	155	311	259	150	48	314	26	140	201	87	356	46	79
12	228	341	168	249	108	181	241	137	356	352	261	217	360	45	340	194	130	92	106	53
13	148	320	130	50	305	280	78	332	338	288	218	194	287	322	316	57	77	116	335	144
14	277	309	12	1	49	167	191	83	212	11	304	124	18	331	198	161	142	304	225	80
15	303	63	229	102	39	297	25	341	12	106	103	188	180	350	147	157	293	118	207	262
16	28	320	343	290	29	247	357	330	122	140	74	231	173	267	92	87	305	81	296	310
17	277	232	273	42	324	332	351	234	12	148	74	145	118	74	207	81	356	198	305	284
18	101	28	8	56	184	60	4	135	354	257	273	2	162	360	68	89	64	279	291	30
19	175	123	324	119	40	177	128	84	205	148	361	319	123	136	290	60	49	3	128	171
20	34	1	168	117	120	299	54	309	254	63	274	94	339	118	102	53	166	99	285	58
21	358	104	121	70	180	227	297	250	288	181	122	46	41	246	31	170	359	53	196	8
22	151	363	143	242	80	317	329	8	271	237	186	167	170	271	3	209	241	154	220	315
23	92	215	191	223	306	120	109	118	96	338	25	365	148	246	136	349	266	41	201	30

图 3-2-1

在验证 m 个人组成的人群中至少有一个人生日是指定日期的概率时，可先随机指定一个不大于 365 的正整数，然后用类似的方法产生多个随机数，并查找指定的数是否在产生的随机数中，最后计算比例.

模拟可下载课件“生日悖论的模拟. xlsm”作为参考.

3.3 二项式定理与杨辉三角

情境与问题

小张在进行投篮练习，共投了 10 次，只考虑是否投中，那么不难知道，投篮结果可以分成 11 类：投中 0 次，投中 1 次，投中 2 次……投中 10 次. 而投中 0 次只有 1（即 C_{10}^0）种情况，投中 1 次有 C_{10}^1 种情况，投中 2 次有 C_{10}^2 种情况……投中 10 次有 C_{10}^{10} 种情况. 因此，小张投篮 10 次，结果共有

$$C_{10}^0+C_{10}^1+C_{10}^2+\cdots+C_{10}^{10}$$

种情况. 那么上式的结果是多少呢?

利用本节我们要学习的二项式定理，可以快速地解答这个问题.

1. 二项式定理

我们知道

$$(a+b)^1=a+b,$$
$$(a+b)^2=a^2+2ab+b^2,$$

而且

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

容易看到，上述得到$(a+b)^3$ 的展开式的过程是烦琐的，如果要用这样的方法去得到$(a+b)^{10}$，$(a+b)^{20}$ 等的展开式是很麻烦的. 那么我们有没有其他办法来得出$(a+b)^3=a^3+3a^2b+3ab^2+b^3$呢?

尝试与发现

从

$$(a+b)^3=(a+b)(a+b)(a+b)$$

出发，观察$(a+b)^3=a^3+3a^2b+3ab^2+b^3$中右边各项是如何形成的，由此总结出一般规律.

注意到

$$(a+b)^3=(a+b)(a+b)(a+b), \qquad ①$$

而展开式中的任何一项都是在右边 3 个括号中各取一个字母相乘得到的（例如，第一个括号取 a，第二个取 b，第三个取 a，则得到 a^2b），因此展开式中每一项都一定是 3 次项，即展开式中只能含有

$$a^3,\ a^2b,\ \boxed{1}\_\_\_\_\_.$$

①式右边展开后有多少个 a^2b 呢？要得到 a^2b，①式右边的 3 个括号中，要有 1 个取 b（剩下的 2 个均取 a），因此共有 C_3^1 种取法，所以有 C_3^1 个 a^2b.

同理可知，①式右边展开后有 $\boxed{2}$_____ 个 ab^2.

类似地，a^3 可以看成①式右边的 3 个括号中取 0 个 b 得到的结果，而 b^3 可以看成①式右边的 3 个括号中取 3 个 b 得到的结果，因此

$$(a+b)^3=C_3^0a^3+C_3^1a^2b+C_3^2ab^2+C_3^3b^3.$$

用同样的方法可知

$$(a+b)^4=C_4^0a^4+C_4^1a^3b+C_4^2a^2b^2+C_4^3ab^3+C_4^4b^4.$$

一般地，当 n 是正整数时，有

$$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+\cdots+C_n^ka^{n-k}b^k+\cdots+C_n^nb^n.$$

上述公式称为**二项式定理**，等式右边的式子称为$(a+b)^n$ 的**展开式**，它共有 $n+1$ 项，其中 $C_n^ka^{n-k}b^k$ 是展开式中的第 $k+1$ 项（通常用 T_{k+1} 表示），C_n^k 称为第 $k+1$ 项的**二项式系数**，我们将 $T_{k+1}=C_n^ka^{n-k}b^k$ 称为二项展开式的**通项公式**.

注意：通项公式 $T_{k+1}=C_n^ka^{n-k}b^k$ 中，要求 n 是正整数，k 是满足 $0\leqslant k\leqslant n$ 的自然数，以后不再声明.

例 1 写出$(2-x)^5$的展开式.

解 在二项式定理中令 $a=2$，$b=-x$，$n=5$，可得

$$\begin{aligned}(2-x)^5&=C_5^02^5+C_5^12^4(-x)+C_5^22^3(-x)^2+C_5^32^2(-x)^3+C_5^42(-x)^4+C_5^5(-x)^5\\&=32-80x+80x^2-40x^3+10x^4-x^5.\end{aligned}$$

例 1 的展开式中，可以看出常数项是 32，x 的系数是 -80，注意到展开式中第 1 项的二项式系数是 $C_5^0=1$，第 2 项的二项式系数为 $C_5^1=5$，由此

可知展开式中某一项的系数与二项式系数，一般情况下并不相等.

例 2 求$\left(x-\frac{1}{x}\right)^9$的展开式中含$x^3$的项.

解 因为$\left(x-\frac{1}{x}\right)^9=[x+(-x^{-1})]^9$，所以展开式中的第$k+1$项为

$$T_{k+1}=C_9^k x^{9-k}(-x^{-1})^k=(-1)^k C_9^k x^{9-k-k}=(-1)^k C_9^k x^{9-2k}.$$

要使此项含x^3，必须有$9-2k=3$，从而有$k=3$，因此含x^3的项为

$$T_4=(-1)^3 C_9^3 x^3=\boxed{3}.$$

例 3 求$\left(2\sqrt{x}+\frac{1}{\sqrt{x}}\right)^6$的展开式中常数项的值和对应的二项式系数.

解 因为$\left(2\sqrt{x}+\frac{1}{\sqrt{x}}\right)^6=(2x^{\frac{1}{2}}+x^{-\frac{1}{2}})^6$，所以展开式中的第$k+1$项为

$$T_{k+1}=C_6^k(2x^{\frac{1}{2}})^{6-k}(x^{-\frac{1}{2}})^k=C_6^k 2^{6-k}x^{\frac{6-k}{2}-\frac{k}{2}}=C_6^k 2^{6-k}x^{3-k}.$$

要得到常数项，必须有$3-k=0$，从而有$k=3$，因此常数项是第 4 项，且

$$T_4=C_6^3 2^{6-3}x^{3-3}=160.$$

从而可知常数项的值为 160，其对应的二项式系数为$C_6^3=20$.

2. 二项式系数的性质

尝试与发现

在二项式定理中，分别令a，b为以下特殊值，写出所得到的等式：

(1) $a=b=1$；

(2) $a=1$，$b=-1$.

在二项式定理中，如果令$a=b=1$，则有

$$2^n=C_n^0+C_n^1+\cdots+C_n^k+\cdots+C_n^{n-1}+C_n^n;$$

如果令$a=1$，$b=-1$，则有

$$0=C_n^0-C_n^1+C_n^2-C_n^3+C_n^4-C_n^5+\cdots,$$

也就是说

$$C_n^0+C_n^2+C_n^4+\cdots=C_n^1+C_n^3+C_n^5+\cdots.$$

由此可知，本节一开始的情境与问题中，

$$C_{10}^0+C_{10}^1+C_{10}^2+\cdots+C_{10}^{10}=\boxed{4}.$$

例 4 已知$(x^2-1)^n$的展开式中，所有的二项式系数之和为 1 024，求展开式中含x^6的项.

解 依题意可知$2^n=1\,024$，因此$n=10$.

从而可知展开式的通项为

$$T_{k+1}=C_{10}^{k}(x^{2})^{10-k}(-1)^{k}=(-1)^{k}C_{10}^{k}x^{20-2k},$$

要使此项含 x^{6}，必须有 $20-2k=6$，从而有 $k=7$，因此含 x^{6} 的项为

$$T_{8}=(-1)^{7}C_{10}^{7}x^{6}=-C_{10}^{3}x^{6}=-120x^{6}.$$

3. 杨辉三角

因为$(a+b)^{0}=1$，所以可以把 $n=0$ 对应的二项式系数看成是 1. 把 $n=0$，1，2，3，4，5，6 对应的二项式系数逐个写出，并排成数表的形式.

$(a+b)^{0}$ …… 1

$(a+b)^{1}$ …… 1 1

$(a+b)^{2}$ …… 1 2 1

$(a+b)^{3}$ …… 1 3 3 1

$(a+b)^{4}$ …… 1 4 6 4 1

$(a+b)^{5}$ …… 1 5 10 10 5 1

$(a+b)^{6}$ …… 1 6 15 20 15 6 1

图 3-3-1

我国古代数学家贾宪（北宋人）在 1050 年前后就给出了类似的数表，并利用数表进行高次开方运算，如图 3-3-1 所示，这一成果在南宋数学家杨辉著的《详解九章算术》中得到摘录. 因此，这一数表在我国称为“贾宪三角”或“杨辉三角”. 西方文献中，一般称其为“帕斯卡三角”，这些文献认为类似的数表是数学家帕斯卡于 1654 年发现的.

尝试与发现

观察杨辉三角中的数，尽可能多地总结其中的规律，并用二项式系数的性质加以说明.

杨辉三角至少具有以下性质：

（1）每一行都是对称的，且两端的数都是 1；

（2）从第三行起，不在两端的任意一个数，都等于上一行中与这个数相邻的两数之和.

另外，观察杨辉三角，还可以发现对于给定的 n 来说，其二项式系数满足中间大、两边小的特点. 这一结论是否具有普遍性呢？

假设 $C_n^{k+1}>C_n^k$，则

$$\frac{n!}{(n-k-1)!(k+1)!}>\frac{n!}{(n-k)!k!},$$

化简可得 $\frac{1}{k+1}>\frac{1}{n-k}$，从而有 $k<$ 5 _____.

利用二项式系数的对称性可知，二项式系数

$$C_n^0,\ C_n^1,\ C_n^2,\ \cdots,\ C_n^{n-2},\ C_n^{n-1},\ C_n^n,$$

是先逐渐变大，再逐渐变小的，当 n 是偶数时，中间一项的二项式系数最大，当 n 是奇数时，中间两项的二项式系数相等且最大.

4. 二项式定理的应用

例 5 求证：$99^{98}-1$ 能被 100 整除.

证明 因为 $99^{98}-1=(100-1)^{98}-1$，由二项式定理可知

$$(100-1)^{98}=C_{98}^0 100^{98}+C_{98}^1 100^{97}(-1)+C_{98}^2 100^{96}(-1)^2+\cdots+$$
$$C_{98}^{96}100^2(-1)^{96}+C_{98}^{97}100(-1)^{97}+C_{98}^{98}(-1)^{98},$$

注意到上述右边的展开式中，前面 98 项都是 100 的倍数，最后一项为 1，由此可知 $99^{98}-1$ 能被 100 整除.

例 6 当 n 是正整数且 $x>0$ 时，求证：$(1+x)^n\geqslant 1+nx$.

证明 由二项式定理可知

$$(1+x)^n=C_n^0+C_n^1x+C_n^2x^2+\cdots+C_n^{n-1}x^{n-1}+C_n^nx^n$$
$$=1+nx+C_n^2x^2+\cdots+C_n^{n-1}x^{n-1}+C_n^nx^n,$$

因为 $x>0$，所以上式右边的项都是正数，从而可知 $(1+x)^n\geqslant 1+nx$.

例 6 的结论可以用在近似计算中. 例如，假设某地区现有人口 100 万，且人口的年平均增长率为 1.2%，那么 6 年后该地区的人口应为 $100(1+1.2\%)^6$，直接计算这个数并不容易，但是利用例 6 的结果可知

$$100(1+1.2\%)^6\geqslant 100(1+6\times 1.2\%)=107.2,$$

注意到 $(1.2\%)^n$ 在 $n\geqslant 2$ 时都是很小的数，因此，如果我们认为 $100(1+1.2\%)^6\approx 107.2$ 的话，近似程度应该是比较好的. 实际上，$100(1+1.2\%)^6$ 保留 6 位有效数字的近似值是 107.419.

想一想

不借助计算器等工具，你能给出 $100(1+1.2\%)^6$ 的比 107.2 更精确的近似值吗？

习题3-3A

1 已知小张练习了 3 次投篮，如果用 1 代表投中，0 代表未投中，001 代表前两次未投中，第三次投中. 试写出小张所有可能的投篮结果，并说出共有多少种

可能.

2. 求$\left(x+\frac{1}{x}\right)^9$的展开式中$x^3$的系数.

3. 指出下列各二项式的展开式中，二项式系数最大的分别是哪一项：

(1) $(a+b)^7$；　　(2) $(2+x)^{16}$.

4. 化简$(1+\sqrt{x})^5+(1-\sqrt{x})^5$.

5. 用二项式定理证明$101^{10}-1$能被10整除.

6. 求下列各式的值：

(1) $C_6^1+C_6^2+C_6^3+C_6^4+C_6^5$；　　(2) $C_{11}^1+C_{11}^3+C_{11}^5+C_{11}^7+C_{11}^9+C_{11}^{11}$.

习题3-3B

1. 求$(1-x)^{13}$的展开式中含x的奇数次项的系数之和.

2. 已知$(1+x)^n$的展开式中，第4项和第6项的系数相等，求这个展开式所有二项式系数之和.

3. 写出$\left(x-\frac{2}{\sqrt{x}}\right)^6$的展开式.

4. 已知$\left(\frac{1}{\sqrt[3]{x}}-\frac{1}{\sqrt[5]{x^2}}\right)^n$的展开式中，所有奇数项的系数和等于1 024，求展开式中二项式系数最大的项.

5. 求$(a^{\frac{1}{3}}b^{-\frac{1}{6}}+a^{-\frac{1}{6}}b^{\frac{1}{4}})^{11}$的展开式中$a$和$b$的指数相等的项.

6. 求$(1+a)(1+b)^2(1+c)^3$的展开式中各项系数的和.

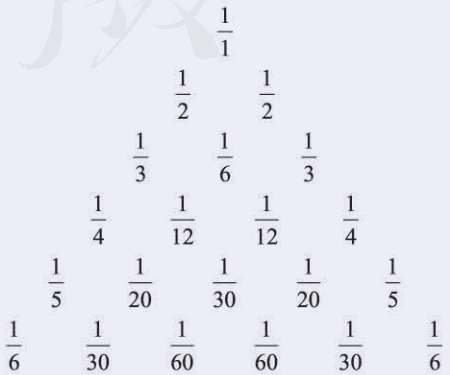
7. 求$(1+x)(2-x)^6$的展开式中的常数项和含x的项.

习题3-3C

1. 将杨辉三角中的每一个数C_n^r都换成分数$\frac{1}{(n+1)C_n^r}$，可得到一个如图所示的分数三角形，称为“莱布尼茨三角形”，从莱布尼茨三角形可看出，存在x使得

$$\frac{1}{(n+1)C_n^r}+\frac{1}{(n+1)C_n^x}=\frac{1}{nC_{n-1}^r},$$

求x的值.



（第1题）

2. 设$(3x-1)^8=a_8x^8+a_7x^7+\cdots+a_1x+$

a_0，求：

(1) $a_8+a_7+\cdots+a_1$；

(2) $a_8-a_7+a_6-a_5+a_4-a_3+a_2-a_1+a_0$；

(3) $a_8+a_6+a_4+a_2+a_0$.

3 求$(2+x-x^2)^6$的展开式中含x的项和含x^3的项.

4 当n是大于1的正整数且$x>0$时，求证：$(1+x)^n\geqslant 1+nx+\dfrac{n(n-1)}{2}x^2$.

1 ab^2，b^3　2 C_3^2　3 $-84x^3$　4 $2^{10}=1\,024$　5 $\dfrac{n-1}{2}$

本章小结

01 知识结构图设计与交流

本章我们学习了排列组合的有关知识，如果以本章所涉及的数学式子为基础，可以作出如下的知识结构图.

$$C_n^0+C_n^2+C_n^4+\cdots=C_n^1+C_n^3+C_n^5+\cdots$$

$$2^n=C_n^0+C_n^1+\cdots+C_n^k+\cdots+C_n^{n-1}+C_n^n$$

$$(a+b)^n=C_n^0a^n+C_n^1a^{n-1}b+\cdots+C_n^ka^{n-k}b^k+\cdots+C_n^nb^n$$

$$C_n^m=C_n^{n-m}$$

$$C_n^m=\frac{A_n^m}{A_m^m}=\frac{n(n-1)\cdots[n-(m-1)]}{m\times(m-1)\times\cdots\times2\times1}=\frac{n!}{(n-m)!m!}$$

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

$$N=m_1\times m_2\times\cdots\times m_n$$

$$N=m_1+m_2+\cdots+m_n$$

请充分运用自己的想象力和创造力，为本章知识设计一份独特的、专属于自己的知识结构图吧！设计好之后与同学分享，并交流学习完本章后的所得和所思.

02 课题作业

通过书籍或者网络查找有关数学史材料，了解贾宪用“杨辉三角”进行高次开方的方法，并给出实例进行说明. 将有关材料整理成小论文，然后与其他同学进行交流.

03 复习题

A 组

1. 某学生要从 5 门选修课中选择 3 门，从 4 个课外活动小组中选择 1 个，则其有多少种不同的选择方法？

2. 要安排 6 位同学表演文艺节目的顺序，要求甲既不能第一个出场，也不能最后一个出场，则共有多少种不同的安排方法？

3. “回文数”是指从左到右读与从右到左读都一样的正整数. 如 22，121，3 443，94 249 等. 显然，2 位数的回文数有 9 个，即 11，22，33，…，99；3 位数的回文数有 90 个：101，111，121，…，191，202，…，999. 求：

（1）4 位数的回文数个数；

（2）$2n+1$ 位数的回文数个数（其中 n 为正整数）.

4. 将 3 名医生和 6 名护士分配到 3 所学校为学生体检，每校分配 1 名医生和 2 名护士，共有多少种不同的分配方案？

5. 某小组有 3 名女生、4 名男生，从中选出 3 名代表，要求女生与男生都至少要有一名，共有多少种不同的选法？

6. 已知 30 件产品中有 27 件合格品，3 件次品，从中抽取 5 件进行检查：

（1）都是合格品的抽法有多少种？

（2）恰好有 2 件次品的抽法有多少种？

（3）至少有 2 件次品的抽法有多少种？

7. 从 10 名学生中选出 3 人担任课代表，则甲、乙两人中至少有 1 人入选，而丙没有入选的不同选法共有多少种？

8. 在一个小组中有 8 名女同学和 4 名男同学，从中挑选两名同学担任交通安全宣传志愿者，那么：

（1）选到的两名同学都是女同学的选法有多少种？

（2）选到的两名同学至少有一名女同学的选法有多少种？

9. 求 $\left(9x-\dfrac{1}{3\sqrt{x}}\right)^{18}$ 的展开式中的常数项，并说明它是展开式中的第几项.

10. 已知 a 是实常数，且二项式 $\left(2x+\dfrac{a}{x}\right)^7$ 的展开式中 $\dfrac{1}{x^3}$ 的系数是 84，求 a 的值.

11. 设 i 为虚数单位，求 $(\sqrt{2}-\mathrm{i})^7$ 的实部.

12. 已知 $\left(\sqrt{x}+\dfrac{3}{\sqrt[3]{x}}\right)^n$ 的展开式中，各项系数的和与其各项二项式系数的和之比为 64，求正整数 n 的值.

B 组

1. 某 4 位同学排成一排准备照相时，又来了 2 位同学要加入，如果保持原来 4 位同学的相对顺序不变，则不同的加入方法有多少种？

2. 有 5 个身高均不相等的学生要排成一排合影留念，最高的人站在中间，从中间到左边和从中间到右边身高都递减，则不同的排法共有多少种？

3. 某赛季足球比赛的计分规则是：胜一场，得 3 分；平一场，得 1 分；负一场，得 0 分. 一球队打完 15 场比赛后，积 33 分，若不考虑顺序，该队胜、负、平的情况共有多少种？

4. 书架上有 4 本不同的数学书，5 本不同的物理书，3 本不同的化学书，将这些书全部竖起排成一排：

（1）如果同类书不能分开，一共有多少种不同的排法？

（2）如果要使任意两本物理书都不相邻，一共有多少种不同的排法？

5. （1）已知 $\dfrac{1}{C_5^n}-\dfrac{1}{C_6^n}=\dfrac{7}{10C_7^n}$，求 C_8^n；

（2）已知 $\dfrac{C_n^{m-1}}{2}=\dfrac{C_n^m}{3}=\dfrac{C_n^{m+1}}{4}$，求 n，m.

6. 设

$$(1-2x)^9=a_0+a_1x+a_2x^2+\cdots+a_9x^9,$$

求 $|a_0|+|a_1|+|a_2|+\cdots+|a_9|$.

7. 已知

$$(x^2+1)(2x+1)^9=a_0+a_1(x+2)+a_2(x+2)^2+\cdots+a_{11}(x+2)^{11},$$

求 $a_0+a_1+a_2+\cdots+a_{11}$ 的值.

8. 已知

$$(1-x)^5=a_0+a_1x+a_2x^2+a_3x^3+a_4x^4+a_5x^5,$$

求 $(a_0+a_2+a_4)(a_1+a_3+a_5)$ 的值.

9. 在 $(\sqrt{2}+\sqrt[4]{3})^{50}$ 的展开式中，有多少个有理项？

10. 求 $\left(|x|+\dfrac{1}{|x|}-2\right)^3$ 的展开式中的常数项.

11. 求 $(1+x+x^2)(1-x)^{10}$ 的展开式中 x^4 的系数.

12. 求 $(1-2x)^5(1+3x)^4$ 的展开式中，按 x 的升幂排列的前三项.

C 组

1. 如图所示，有些共享单车的密码锁是由 4 个数字组成的，你认为共享单车的

密码锁能设置成由 3 个数字组成吗？5 个数字呢？为什么？

（第 1 题）

2. 把 6 张座位编号为 1，2，3，4，5，6 的电影票全部分给 4 个人，每人至少分 1 张，至多分 2 张，且这两张票具有连续的编号，那么不同的分法共有多少种？

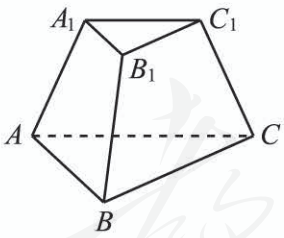
3. 设 n 是正整数，化简 $\mathrm{C}_n^1+\mathrm{C}_n^2 6+\mathrm{C}_n^3 6^2+\cdots+\mathrm{C}_n^n 6^{n-1}$.

4. 求 $(1+x)^3+(1+x)^4+(1+x)^5+\cdots+(1+x)^{19}+(1+x)^{20}$ 的展开式中含 x^3 的项.

5. 过三棱柱任意两个顶点的直线共 15 条，其中异面直线有多少对？

6. 将 a，b，c 填入 3×3 的方格中，要求每行、每列都没有重复的字母，则不同的填写方法共有多少种？

7. 某人有 3 种颜色的灯泡（每种颜色的灯泡足够多），要在如图所示的 6 个点 A，B，C，A_1，B_1，C_1 上各安装一个灯泡，要求同一条线段两端的灯泡颜色不同，则不同的安装方法共有多少种？



（第 7 题）

随机与概率

熙熙人群朋友不期而遇，茫茫宇宙陨星意外撞击.
随机事件发生并非随意，概率破解其中奥秘玄机.
情境重复催生稀有事件，历史长河沉淀自然奇迹.
同班同学常有生日相同，彩民两次中奖并不神奇.
抵押贷款房产汽车按揭，精巧设计需要借助概率.
保费计算基于概率模型，期权定价有赖随机分析.
概率技巧有助破解密码，人工智能需用概率逻辑.
日常生活常遇概率问题，学点概率知识终身受益.

——严加安

第四章

概率与统计

本章导语

我们前面已经学过一些概率与统计的知识，对现实世界中随机现象的不确定性有了初步的了解. 例如，我们已经知道，事件的概率是一个数，它描述了事件发生的可能性大小；通过古典概型与用频率估计概率，可以得到一些事件发生的概率；等等. 但是，如果要解决一些更复杂的问题，我们还需要进一步学习概率统计的知识.

例如，按照先后顺序抽奖（或抽签）在我们日常生活中很常见，但经常有人会提出疑问：如果只有一个特等奖，而恰巧特等奖被第一个人抽到了的话，那么后面抽奖的人就不可能抽到特等奖了，这是不是意味着，先抽奖的人抽到特等奖的概率更大一些？

再例如，通过必修课程中的知识我们已经能够算出，如果做某件事，每次成功的概率只有 0.1，那么只要尝试 22 次，就能保证至少成功一次的概率不小于 90%. 在同样的情境以及同样是尝试 22 次的前提下，恰好成功 1 次的概率与恰好成功 2 次的概率应该怎样计算呢？恰好成功多少次的概率最大呢？

又例如，在新闻报道中，现在时常都可以见到“相关系数”这个词，如图所示是《人民日报》2016 年 1 月 27 日一篇报道的截图，你知道其中的“相关系数”的准确含义吗？

另外，我们还经常会听到“英语学习，女生比男生更擅长”之类的说法. 怎样通过统计数据来判断这些说法是否有道理？

学完本章之后，类似的问题都能解答.

“扶贫先扶智”决定了教育扶贫的基础性地位，“治贫先治愚”决定了教育扶贫的先导性功能，“脱贫防返贫”决定了教育扶贫的根本性作用。联合国教科文组织研究表明，不同层次受教育者提高劳动生产率的水平不同：本科 300%、初高中 108%、小学 43%，人均受教育年限与人均 GDP 的相关系数为 0.562。“积财千万，不如薄技在身”“一技在手，终身受益”，教育在促进扶贫、防止返贫方面的作用，可说是根本性的、可持续的。

4.1 条件概率与事件的独立性

4.1.1 条件概率

情境与问题

金融界的人经常需要计算不同投资环境下获利的概率，因此金融投资公司在招聘新员工时，通常会考查应聘人员计算概率的能力．以下是某金融投资公司的一道笔试题，你会做吗？

从生物学中我们知道，生男、生女的概率基本是相等的，都可以近似地认为是$\frac{1}{2}$．如果某个家庭中先后生了两个小孩：

（1）当已知较大的小孩是女孩的条件下，较小的小孩是男孩的概率为多少？

（2）当已知两个小孩中有女孩的条件下，两个小孩中有男孩的概率为多少？

情境与问题中的两个概率，直觉上大家可能会觉得答案都是$\frac{1}{2}$．但是，第（2）个问题的答案并不是$\frac{1}{2}$，这可能会出乎某些同学的意料！学完本小节的条件概率之后，对此我们就可以有一个比较透彻地理解了．

尝试与发现

已知某班级中，有女生 16 人，男生 14 人，而且女生中喜欢长跑的有 10 人，男生中喜欢长跑的有 8 人．现从这个班级中随机抽出一名学生：

（1）求所抽到的学生喜欢长跑的概率；

（2）若已知抽到的是男生，求所抽到的学生喜欢长跑的概率．

可以看出，尝试与发现中的这两个问题可以借助古典概型来处理，其中样本空间 Ω 是由班级中所有学生组成的集合，共包含 $14+16=30$ 个样本点．设事件“所抽到的学生喜欢长跑”对应的集合是 A，则 A 是由所有喜欢长跑的人组成的，其中包含 $10+8=18$ 个样本点；设事件“抽到的是男生”

对应的集合是 B，则 B 包含 14 个样本点.

问题（1）中，可以看出所求概率为

$$P(A)=\boxed{1}\quad\_\_\_\_\_.$$

问题（2）中，因为已知事件 B 发生了，也就是相当于是从男生中任意抽取了一人. 此时要使得事件 A 发生，必须抽取 AB（即 $A\cap B$）中的样本点，因此所求概率应该是 $\frac{8}{14}=\frac{4}{7}$. 这里的 $\frac{4}{7}$ 称为已知事件 B 发生的条件下事件 A 发生的概率，记作 $P(A\mid B)$，即

$$P(A\mid B)=\frac{4}{7}.$$

这样的概率称为条件概率.

尝试与发现

观察上述 A 与 B 之间的关系，试探讨怎样才能求出 $P(A\mid B)$.

可以看出，上述 $P(A\mid B)$ 可以用 $P(A\cap B)$ 与 $P(B)$ 表示出来，即

$$P(A\mid B)=\boxed{2}\quad\_\_\_\_\_.$$

一般地，当事件 B 发生的概率大于 0 时（即 $P(B)>0$），已知事件 B 发生的条件下事件 A 发生的概率，称为**条件概率**，记作 $P(A\mid B)$，而且

$$P(A\mid B)=\frac{P(A\cap B)}{P(B)}①.$$

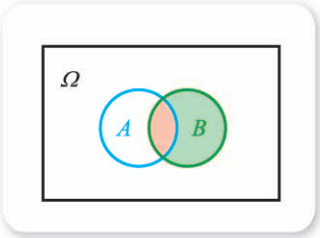


图 4-1-1

条件概率可以借助图 4-1-1 来理解. 需要注意的是，$P(A\mid B)$ 与 $P(B\mid A)$ 的意义不一样，一般情况下，它们也不相等.

例 1 掷红、蓝两个均匀的骰子，设

A：蓝色骰子的点数为 5 或 6；

B：两骰子的点数之和大于 7.

求已知事件 A 发生的条件下事件 B 发生的概率 $P(B\mid A)$.

解 用数对 (x, y) 来表示抛掷结果，其中 x 表示红色骰子的点数，y 表示蓝色骰子的点数，则样本空间可记为

$$\Omega=\{(x, y)\mid x, y=1, 2, 3, 4, 5, 6\},$$

而且样本空间可用图 4-1-2 直观表示，图中每一个点代表一个样本点. 样本空间中，共包含 36 个样本点.

不难看出，A 包含的样本点即图 4-1-2 中绿色矩形框中的点，共 12

① 如不特别声明，以后谈到类似 $P(A\mid B)$ 等条件概率时，总是默认 $P(B)>0$.

个，因此

$$P(A)=\frac{12}{36}=\frac{1}{3};$$

B 包含的样本点即图 4-1-2 中紫色三角框中的点，$B\cap A$ 共包含 9 个样本点，从而

$$P(B\cap A)=\frac{9}{36}=\frac{1}{4}.$$

因此

$$P(B\mid A)=\frac{P(B\cap A)}{P(A)}=\boxed{3}\quad.$$

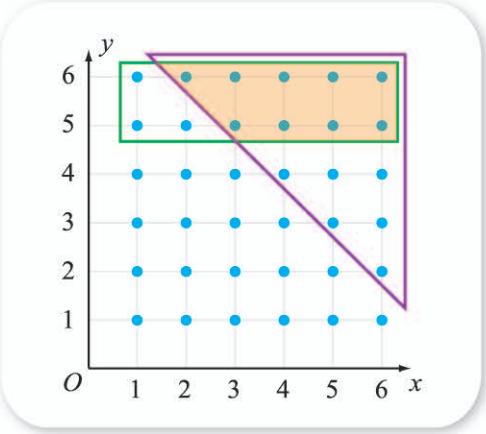


图 4-1-2

例 1 中的 $P(B\mid A)$，也可以通过 A 中的样本点个数 12 与 $B\cap A$ 中的样本点个数 9 直接得到，即

$$P(B\mid A)=\frac{9}{12}=\frac{3}{4}.$$

注意到 $P(B)=\frac{5}{12}\neq\frac{3}{4}$，这说明事件 A 的发生影响了事件 B 发生的概率.

另外，从必修内容中我们已经知道，两个事件 A 与 B 独立的充要条件是 $P(A\cap B)=P(A)P(B)$，其直观理解是，事件 A 是否发生不会影响事件 B 发生的概率，事件 B 是否发生也不会影响事件 A 发生的概率. 从例 1 可以看出，事件的独立性与条件概率有着紧密的联系，我们将在 4.1.3 中详细讨论这两者之间的关系.

例 2 已知春季里，每天甲、乙两地下雨的概率分别为 20%与 18%，且两地同时下雨的概率为 12%. 求春季的一天里：

(1) 已知甲地下雨的条件下，乙地也下雨的概率；

(2) 已知乙地下雨的条件下，甲地也下雨的概率.

解 记 A：甲地下雨，B：乙地下雨，则由已知可得

$$P(A)=20\%,\ P(B)=18\%,\ P(A\cap B)=12\%.$$

(1) 需要求的是 $P(B\mid A)$，因此

$$P(B\mid A)=\frac{P(A\cap B)}{P(A)}=\frac{12\%}{20\%}=\frac{3}{5}.$$

(2) 需要求的是 $P(A\mid B)$，因此

$$P(A\mid B)=\frac{P(A\cap B)}{P(B)}=\boxed{4}\quad.$$

例 3 已知某地区内狗的寿命超过 15 岁的概率为 0.8，超过 20 岁的概率为 0.2. 那么该地区内，一只寿命超过 15 岁的狗，寿命能超过 20 岁的

概率为多少?

解 设 A：狗的寿命超过 15 岁，B：狗的寿命超过 20 岁，则所要求的就是 $P(B \mid A)$.

依题意有 $P(A)=0.8$，$P(B)=0.2$. 又因为 $B \subseteq A$，所以 $B \cap A=B$，从而

$$P(B \cap A)=P(B)=0.2,$$

因此

$$P(B \mid A)=\frac{P(B \cap A)}{P(A)}=\boxed{5}\quad\_\_\_\_\_.$$

前面的情境与问题中，如果用(F，M)表示较大的小孩是女孩，较小的小孩是男孩，则样本空间可以表示为

$$\Omega=\{(\mathrm{F},\ \mathrm{M}),\ (\mathrm{F},\ \mathrm{F}),\ (\mathrm{M},\ \mathrm{F}),\ (\mathrm{M},\ \mathrm{M})\}.$$

则“较大的小孩是女孩”对应的是 $A=\{(\mathrm{F},\ \mathrm{M}),\ (\mathrm{F},\ \mathrm{F})\}$，“较小的小孩是男孩”对应的是 $B=\{(\mathrm{F},\ \mathrm{M}),\ (\mathrm{M},\ \mathrm{M})\}$，从而“已知较大的小孩是女孩的条件下，较小的小孩是男孩”的概率为

$$P(B \mid A)=\frac{P(B \cap A)}{P(A)}=\frac{1}{2}.$$

而“两个小孩中有女孩”对应的是 $C=\{(\mathrm{F},\ \mathrm{M}),\ (\mathrm{F},\ \mathrm{F}),\ (\mathrm{M},\ \mathrm{F})\}$，“两个小孩中有男孩”对应的是 $D=\{(\mathrm{F},\ \mathrm{M}),\ (\mathrm{M},\ \mathrm{F}),\ (\mathrm{M},\ \mathrm{M})\}$，从而“已知两个小孩中有女孩的条件下，两个小孩中有男孩”的概率为

$$P(D \mid C)=\frac{P(D \cap C)}{P(C)}=\frac{2}{3}.$$

探索与研究

假设 A，B，C 都是事件，且 $P(A)>0$. 根据条件概率的定义，探索条件概率是否满足下列性质：

(1) $0 \leqslant P(B \mid A) \leqslant 1$；

(2) $P(A \mid A)=1$；

(3) 如果 B 与 C 互斥，则 $P((B \cup C) \mid A)=P(B \mid A)+P(C \mid A)$.

练习A

1. 已知 $P(A)=0.5$，$P(B)=0.3$，$P(B \cap A)=0.1$，求：

   (1) $P(B \mid A)$；　　(2) $P(A \mid B)$.

2. 某同学算出条件概率 $P(B \mid A)=1.6$，这可能吗?

3 盒子里有 25 个形状、大小、质地相同的球，其中有 10 个白色的，5 个黄色的，10 个黑色的. 从盒子中任意取出一个球，已知这个球不是黑球，求取出的球是黄球的概率.

4 已知一种节能灯使用寿命超过 10 000 h 的概率为 0.95，而使用寿命超过 12 000 h 的概率为 0.9. 则已经使用了 10 000 h 的这种节能灯，使用寿命能超过 12 000 h 的概率为多少?

5 举出 $P(A)>0$ 而且 $P(B \mid A)=0$ 的实例.

练习B

1 利用 $P(B \mid A)=\dfrac{P(B \cap A)}{P(A)}$ 证明 $0 \leqslant P(B \mid A) \leqslant 1$.

2 已知 $P(B \mid A)=0.6$，求 $P(\overline{B} \mid A)$ 的值.

3 已知某班级中，有女生 15 人，男生 17 人，而且女生中不戴眼镜的有 6 人，男生中戴眼镜的有 5 人. 现从这个班级中随机抽出一名学生：

(1) 求所抽到的学生戴眼镜的概率；

(2) 若已知抽到的是女生，求所抽到的学生戴眼镜的概率.

4 抛一枚均匀的硬币两次，记 A：第一次出现正面，B：第二次出现正面，求 $P(B \mid A)$.

5 掷红、蓝两个均匀的骰子，设 A：蓝色骰子的点数为 1 或 2，B：两骰子的点数之和小于 5，求 $P(B \mid A)$ 与 $P(A \mid B)$.

1 $\dfrac{18}{30}=\dfrac{3}{5}$　2 $\dfrac{P(A \cap B)}{P(B)}$　3 $\dfrac{\frac{1}{4}}{\frac{1}{3}}=\dfrac{3}{4}$　4 $\dfrac{12\%}{18\%}=\dfrac{2}{3}$　5 $\dfrac{0.2}{0.8}=0.25$

4.1.2 乘法公式与全概率公式

情境与问题

学校的“我为祖国献计献策”演讲比赛共有 20 名同学参加，学校决定让参赛选手通过抽签决定出场顺序. 不过，张明对抽签的公平性提出了质疑，他的理由是，如果第一个人抽的出场顺序是 1 号，那么其他人就抽不到 1 号了，所以每个人抽到 1 号的概率不一样. 张明的想法正确吗？特别地，第一个抽签的人抽到 1 号的概率与第二个抽签的人抽到 1 号的概率是否相等？为什么？

抽签的公平性如果仅仅只是从直观上来理解的话，可能并不容易说清楚，但这可利用本节我们要学习的全概率公式来解释.

1. 乘法公式

尝试与发现

(1) 在 $P(B \mid A)$，$P(BA)$（即 $P(B \cap A)$，下同），$P(A)$这三者中，如果已知 $P(A)$与 $P(B \mid A)$，能不能求出 $P(BA)$？

(2) 某人翻开电话本给自己的一位朋友打电话时，发现电话号码的最后一位数字变得模糊不清了，因此决定随机拨号进行尝试. 你能用 (1) 中所得的结论，得出此人尝试两次但都拨不对电话号码的概率吗？

由条件概率的计算公式 $P(B \mid A)=\dfrac{P(BA)}{P(A)}$可知，

$$P(BA)=P(A)P(B \mid A),$$

这就是说，根据事件 A 发生的概率，以及已知事件 A 发生的条件下事件 B 发生的概率，可以求出 A 与 B 同时发生的概率. 一般地，这个结论称为**乘法公式**.

例如，对于尝试与发现中的 (2) 来说，如果设 A 表示第一次没有拨对，B 表示第二次没有拨对. 则 $P(A)$是容易求出的：总共有 10 种可能，拨不对电话号码的情况有 9 种，因此 $P(A)=\dfrac{9}{10}$. $P(B \mid A)$也是容易算出

来的：如果第一次拨不对，那么第二次会从第一次尝试的数以外的数中随机选取一个进行尝试，总共有 9 种可能，拨不对电话号码的情况有 8 种，因此 $P(B \mid A)=\frac{8}{9}$. 从而根据乘法公式可知，两次都拨不对电话号码的概率为

$$P(BA)=P(A)P(B \mid A)=\frac{9}{10}\times\frac{8}{9}=\frac{4}{5}.$$

想一想

比较尝试与发现中问题（2）的两种解法，思考：什么情况下利用乘法公式会有优势？

值得注意的是，尝试与发现中的（2）也可以借助排列组合来解：问题可转化为“用 10 个数字排成数字不重复的 2 位数，求某个特定数字不出现的概率”，因为总共有 A_{10}^2 种排法，特定数字不出现的排法有 A_9^2 种，因此所求概率是

$$\frac{A_9^2}{A_{10}^2}=\frac{9\times8}{10\times9}=\frac{4}{5}.$$

例 1 已知某品牌的手机从 1 m 高的地方掉落时，屏幕第一次未碎掉的概率为 0.5，当第一次未碎掉时第二次也未碎掉的概率为 0.3. 试求这样的手机从 1 m 高的地方掉落两次后屏幕仍未碎掉的概率.

解 设 A_i 表示第 i 次掉落手机屏幕没有碎掉，$i=1$，2，则由已知可得 $P(A_1)=0.5$，$P(A_2 \mid A_1)=0.3$，因此由乘法公式可得

$$P(A_2A_1)=P(A_1)P(A_2 \mid A_1)=0.5\times0.3=0.15.$$

即这样的手机从 1 m 高的地方掉落两次后屏幕仍未碎掉的概率为 0.15.

例 2 在某次抽奖活动中，在甲、乙两人先后进行抽奖前，还有 50 张奖券，其中共有 5 张写有“中奖”字样. 假设抽完的奖券不放回，甲抽完之后乙再抽，求：

（1）甲中奖而且乙也中奖的概率；

（2）甲没中奖而且乙中奖的概率.

解 设 A 表示甲中奖，B 表示乙中奖，则

$$P(A)=\frac{5}{50}=\frac{1}{10}.$$

（1）因为抽完的奖券不放回，所以甲中奖后乙抽奖时，有 49 张奖券且其中只有 4 张写有“中奖”字样，此时乙中奖的概率为 $P(B \mid A)=\frac{4}{49}$.

根据乘法公式可知，甲中奖且乙也中奖的概率为

$$P(BA)=P(A)P(B \mid A)$$

$$=\frac{1}{10}\times\frac{4}{49}=\frac{2}{245}.$$

(2) 因为 $P(A)+P(\overline{A})=1$，所以

$$P(\overline{A})=\boxed{1}\_\_\_\_\_.$$

因为抽完的奖券不放回，所以甲没中奖后乙抽奖时，还有 49 张奖券且其中还有 5 张写有“中奖”字样，此时乙中奖的概率为 $P(B\mid\overline{A})=$ $\boxed{2}\_\_\_\_\_$.

根据乘法公式可知，甲没中奖而且乙中奖的概率为

$$P(B\overline{A})=P(\overline{A})P(B\mid\overline{A})$$

$$=\boxed{3}\_\_\_\_\_.$$

例 2 也可用排列组合的知识求解，请读者自行尝试.

探索与研究

假设 A_i 表示事件，$i=1,2,3$，且 $P(A_1)>0$，$P(A_1A_2)>0$. 证明

$$P(A_1A_2A_3)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2)$$

一定成立，其中 $P(A_3\mid A_1A_2)$ 表示已知 A_1 与 A_2 都发生时 A_3 发生的概率，而 $P(A_1A_2A_3)$ 表示 A_1，A_2，A_3 同时发生的概率. 并通过具体实例来理解上式.

2. 全概率公式

尝试与发现

(1) 在例 2 中，如果想求乙中奖的概率 $P(B)$，该怎样计算？

(2) 一般地，如果已知 $P(BA)$ 与 $P(B\overline{A})$，能否求出 $P(B)$？如果已知 $P(B\mid A)$，$P(A)$，$P(B\mid\overline{A})$，$P(\overline{A})$，能否求出 $P(B)$？

在例 2 中，乙中奖可以分为两种情况：甲中奖且乙中奖，甲没中奖且乙中奖，即 $B=BA+B\overline{A}$. 这两种情况是不能同时发生的（即是互斥的），因此由互斥事件概率的加法公式可得

$$P(B)=P(BA+B\overline{A})=P(BA)+P(B\overline{A})=\frac{2}{245}+\frac{9}{98}=\frac{1}{10}.$$

一般地，如果样本空间为 Ω，而 A，B 为事件，则 BA 与 $B\overline{A}$ 是互斥的，且

$$B=B\Omega=B(A+\overline{A})=BA+B\overline{A},$$

如图 4-1-3 所示，从而

$$P(B)=P(BA+B\overline{A})=P(BA)+P(B\overline{A}).$$

更进一步，当 $P(A)>0$ 且 $P(\overline{A})>0$ 时，因为

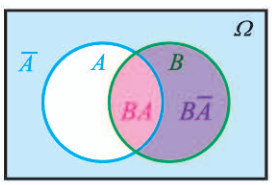


图 4-1-3

由乘法公式有

$$P(BA)=P(A)P(B\mid A),\ P(B\overline{A})=P(\overline{A})P(B\mid \overline{A}),$$

所以

$$P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid \overline{A}).$$

这称为**全概率公式**.

例 3 某次社会实践活动中，甲、乙两个班的同学共同在一个社区进行民意调查. 参加活动的甲、乙两班的人数之比为 5∶3，其中甲班中女生占 $\frac{3}{5}$，乙班中女生占 $\frac{1}{3}$. 求该社区居民遇到一位进行民意调查的同学恰好是女生的概率.

尝试与发现

用适当的符号表示例 3 中的已知条件，并思考解题的方法.

解 如果用 A 与 $\overline{A}$ 分别表示居民所遇到的一位同学是甲班的与乙班的，B 表示是女生. 则根据已知，有

$$P(A)=\frac{5}{5+3}=\frac{5}{8},\ P(\overline{A})=\boxed{4}\ ,$$

而且

$$P(B\mid A)=\frac{3}{5},\ P(B\mid \overline{A})=\frac{1}{3}.$$

题目所要求的是 $P(B)$.

由全概率公式可知

$$P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid \overline{A})=\frac{5}{8}\times\frac{3}{5}+\frac{3}{8}\times\frac{1}{3}=\frac{1}{2}.$$

例 3 也可以这样来理解：假设参加活动的甲班人数为 $5n$，则乙班人数为 $3n$，而且甲班中有女生 $3n$ 人，乙班中有女生 n 人. 从而可知参加活动的总共有 $5n+3n=8n$ 人，而女生有 $3n+n=4n$ 人，因此所求概率为 $\frac{4n}{8n}=\frac{1}{2}$.

在上述记号下，我们还可以利用

$$P(\overline{B})=P(A)P(\overline{B}\mid A)+P(\overline{A})P(\overline{B}\mid \overline{A})$$

来得到该社区居民遇到一位进行民意调查的同学恰好是男生的概率. 例 3 中的信息可借助如图 4-1-4 所示的树形图来理解.

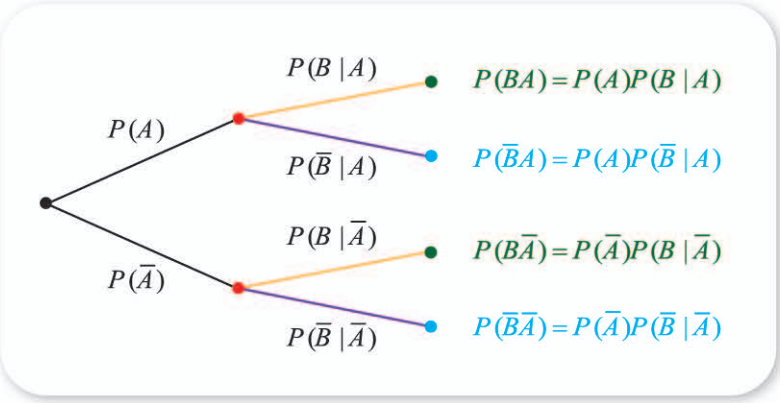


图 4-1-4

利用全概率公式，可以解决情境与问题中的抽签问题. 如果设 A_i 表示第 i 个抽签人抽到 1 号，$i=1$，2. 则可以看出

$$P(A_1)=\frac{1}{20},\ P(\overline{A}_1)=\frac{19}{20}.$$

如果第 1 个抽签人抽到 1 号，那么第 2 个人抽到 1 号的概率为 0，即 $P(A_2 \mid A_1)=0$；如果第 1 个抽签人抽到的不是 1 号，那么第 2 个人抽到 1 号的概率为 $\frac{1}{19}$，即 $P(A_2 \mid \overline{A}_1)=\frac{1}{19}$. 因此

$$P(A_2)=P(A_1)P(A_2 \mid A_1)+P(\overline{A}_1)P(A_2 \mid \overline{A}_1)=\frac{1}{20}\times 0+\frac{19}{20}\times\frac{1}{19}=\frac{1}{20}.$$

这就是说 $P(A_1)=P(A_2)$. 因此抽签是公平的.

前面提到的全概率公式，本质上是将样本空间分成互斥的两部分（即 A 与 $\overline{A}$）后得到的. 不难想到，可以将样本空间分成更多互斥的部分，从而得到如下更一般的结论.

定理 1　若样本空间 Ω 中的事件 A_1，A_2，…，A_n 满足：

(1) 任意两个事件均互斥，即 $A_iA_j=\varnothing$，i，$j=1$，2，…，n，$i\neq j$；

(2) $A_1+A_2+\cdots+A_n=\Omega$；

(3) $P(A_i)>0$，$i=1$，2，…，n.

则对 Ω 中的任意事件 B，都有 $B=BA_1+BA_2+\cdots+BA_n$，且

$$P(B)=\sum_{i=1}^{n}P(BA_i)=\sum_{i=1}^{n}P(A_i)P(B \mid A_i).$$

上述公式也称为**全概率公式**. $n=3$ 时的情形可借助图 4-1-5 来理解.

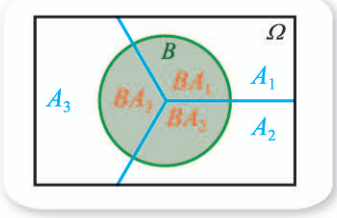


图 4-1-5

例 4 假设某市场供应的智能手机中，市场占有率和优质率的信息如下表所示.

品牌	甲	乙	其他
市场占有率	50%	30%	20%
优质率	95%	90%	70%

在该市场中任意买一部智能手机，求买到的是优质品的概率.

解 用 A_1，A_2，A_3 分别表示买到的智能手机为甲品牌、乙品牌、其他品牌，B 表示买到的是优质品，则依据已知可得

$$P(A_1)=50\%,\ P(A_2)=30\%,\ P(A_3)=20\%,$$

且

$$P(B\mid A_1)=95\%,\ P(B\mid A_2)=90\%,\ P(B\mid A_3)=70\%.$$

因此，由全概率公式有

$$\begin{aligned}P(B)&=P(A_1)P(B\mid A_1)+P(A_2)P(B\mid A_2)+P(A_3)P(B\mid A_3)\\&=50\%\times 95\%+30\%\times 90\%+20\%\times 70\%=88.5\%.\end{aligned}$$

由例 4 还可以得出全概率公式的一个直观解释：已知事件 B 的发生有各种可能的情形 $A_i(i=1,2,\cdots,n$，且任意两种情形均互斥$)$，事件 B 发生的可能性，就是各种可能情形 A_i 发生的可能性与已知在 A_i 发生的条件下事件 B 发生的可能性的乘积之和. 在实际问题中，由于随机事件的复杂性，有时很难直接求得事件 B 发生的概率，因此我们可以分析事件 B 发生的各种可能情形，化整为零地去分解事件 B，然后借助于全概率公式间接求出事件 B 发生的概率.

*3. 贝叶斯公式

尝试与发现

用适当的符号表示出下列描述中的已知与未知，并探索问题的解法：已知某厂生产的食盐，优质品率为 90%. 优质品中，包装达标的占 95%；非优质品中，包装达标的占 80%. 如果从该厂生产的食盐中，随机取一袋，发现包装是达标的，那么这袋食盐是优质品的概率为多少（精确到 0.1%）?

尝试与发现的描述中，可以用 A 表示优质品，B 表示包装达标，则 $\overline{A}$ 表示不是优质品，而且有

$$P(A)=90\%,\ P(B\mid A)=95\%,\ P(B\mid \overline{A})=80\%;$$

问题中所要求的是 $P(A\mid B)$.

由条件概率可知

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

不过，已知条件中并没有直接给出 $P(AB)$ 与 $P(B)$ 的值. 但由乘法公式和全概率公式可得

$$\begin{aligned}P(AB)&=P(BA)=P(A)P(B \mid A)=90\%\times 95\%=85.5\%,\\ P(B)&=P(A)P(B \mid A)+P(\overline{A})P(B \mid \overline{A})\\ &=90\%\times 95\%+(1-90\%)\times 80\%\\ &=93.5\%.\end{aligned}$$

因此一袋包装达标的食盐是优质品的概率为

$$P(A \mid B)=\frac{85.5\%}{93.5\%}\approx 91.4\%.$$

一般地，当 $1>P(A)>0$ 且 $P(B)>0$ 时，有

$$P(A \mid B)=\frac{P(A)P(B \mid A)}{P(B)}=\frac{P(A)P(B \mid A)}{P(A)P(B \mid A)+P(\overline{A})P(B \mid \overline{A})}.$$

这称为**贝叶斯公式**.

例 5 某生产线的管理人员通过对以往数据的分析发现，每天生产线启动时，初始状态良好的概率为 80%. 当生产线初始状态良好时，第一件产品合格的概率为 95%；否则，第一件产品合格的概率为 60%. 某天生产线启动时，生产出的第一件产品是合格品，求当天生产线初始状态良好的概率（精确到 0.1%）.

解 用 A 表示生产线初始状态良好，B 表示产品为合格品. 则由已知有

$$P(A)=80\%,\ P(B \mid A)=95\%,\ P(B \mid \overline{A})=60\%.$$

从而 $P(\overline{A})=$ 5 _____，因此由贝叶斯公式可知

$$P(A \mid B)=\text{6}\ \underline{\qquad\qquad}.$$

例 5 中的概率 80%（即 $P(A)$）是根据历史数据发现的，通常称为先验概率；获取了新信息（即第一件产品是合格品）后算出的概率 $P(A \mid B)$，通常称为后验概率. 贝叶斯公式指出的是，通过先验概率以及其他信息，可以算出后验概率. 实际上，贝叶斯公式可以看成要根据事件发生的结果找原因，看看这一结果由各种可能原因导致的概率是多少.

贝叶斯公式在日常生活中有着广泛的应用，以下是一个实际的例子.

情境与问题

已知某地居民肝癌的发病率为 0.000 4. 通过对血清甲胎蛋白进行检验可以检测一个人是否患有肝癌，但这种检测方法可能出错，具体是：患有肝癌但检测显示正常的概率为 0.01，未患有肝癌但检测显示有肝癌的概率为 0.05. 目前情况下，肝癌的致死率比较高，肝癌发现得越早，治疗越有效，因此有人主张对该地区的居民进行普查，以尽早发现肝癌患者. 这个主张是否合适？

上述情境中，如果患有肝癌，那么检测出来的概率为 99%. 然而，普查的主张是否合适，主要取决于检测结果显示患有肝癌时，实际上患有肝癌的概率.

设 A 表示患有肝癌，B 表示检测结果显示患有肝癌，则

$$P(A)=0.000\ 4,\ P(\overline{B}\mid A)=0.01,\ P(B\mid \overline{A})=0.05,$$

从而有

$$P(\overline{A})=1-P(A)=1-0.000\ 4=0.999\ 6,$$

$$P(B\mid A)=1-P(\overline{B}\mid A)=1-0.01=0.99.$$

根据贝叶斯公式，则检测显示患有肝癌的居民确实患有肝癌的概率为

$$\begin{aligned}P(A\mid B)&=\frac{P(A)P(B\mid A)}{P(A)P(B\mid A)+P(\overline{A})P(B\mid \overline{A})}\\&=\frac{0.000\ 4\times 0.99}{0.000\ 4\times 0.99+0.999\ 6\times 0.05}\\&\approx 0.007\ 9.\end{aligned}$$

这就表明，检测结果显示患有肝癌但实际上患有肝癌的概率还不到 0.8%！也就是说，如果进行普查的话，在现有条件下，100 个显示患有肝癌的人中，可能只有 1 个人是真正患有肝癌的. 从这个意义上来说，进行普查并不是一个好主意.

值得注意的是，这并不能说明对应的检测方法精度不够高，更不能说明在实际诊断时不能使用对应的检测办法.

仔细观察上述计算过程，可以发现 $P(A)$ 对 $P(A\mid B)$ 的影响很大. 用 Excel 或计算机软件进行试算也可以看出这一点，如下表所示.

$P(A)$	0.000 4	0.001	0.01	0.05	0.1	0.2	0.5
$P(A\mid B)$	0.007 9	0.019 4	0.166 7	0.510 3	0.687 5	0.831 9	0.951 9

但是，需要注意的是，在实际诊断过程中，医生往往会先观察患者的症状，只有当医生通过其他症状怀疑病人患有肝癌时，才会建议进行血清甲胎

蛋白检测. 也就是说，此时疑似患病人群的 $P(A)$ 值已经远远大于 0.000 4，甚至可能已经达到了 0.5，因此检测显示患有肝癌而实际也患有肝癌的概率 $P(A \mid B)$ 也就会比 0.007 9 大很多了.

同全概率公式一样，贝叶斯公式也可以进行推广.

定理 2 若样本空间 Ω 中的事件 A_1，A_2，…，A_n 满足：

(1) 任意两个事件均互斥，即 $A_iA_j=\varnothing$，i，$j=1$，2，…，n，$i\neq j$；

(2) $A_1+A_2+\cdots+A_n=\Omega$；

(3) $1>P(A_i)>0$，$i=1$，2，…，n.

则对 Ω 中的任意概率非零的事件 B，有

$$P(A_j \mid B)=\frac{P(A_j)P(B \mid A_j)}{P(B)}=\frac{P(A_j)P(B \mid A_j)}{\sum\limits_{i=1}^{n}P(A_i)P(B \mid A_i)}.$$

上述公式也称为**贝叶斯公式**.

拓展阅读

人工智能中的贝叶斯公式

人工智能正在逐渐改变着我们日常生活的方方面面，作为交叉学科，它融合了信息技术、数学、哲学、生物学等领域的知识. 特别地，数学在其中起到了重要的作用. 不过，人工智能所用到的数学知识并非都是遥不可及的高深理论，贝叶斯公式就被广泛地用于人工智能的分类算法中. 下面我们简要地介绍拼音输入法给出候选词时是如何应用贝叶斯公式的.

当我们在电子设备上使用拼音输入法输入汉字时，输入法会根据已经输入的信息给出一定数量的候选词. 更神奇的是，即使错误地输入了某些信息，“智能化”的输入法依然会给出若干个候选词，而其中往往会有我们想要输入的那个词语.

例如，如果想输入“信息”，在键盘上输入“xinxi”，候选词中就出现了正确的选项，如图 1 所示；即使不小心错误地输入了“xinxxi”，输入法依然会将“信息”作为候选词之一，如图 2 所示. 这能极大地提升汉字的输入速度. 那么，输入法是如何实现这种“智能”提示的呢？

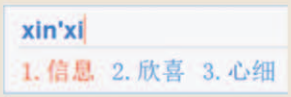


图 1

图 2

秘诀就在贝叶斯公式中！用上面的例子来说，假设有一个词库，包含了日常输入中所有可能出现的词语，输入法把词库中的第 i 个词语作为候选词记为事件 B_i. 当通过键盘输入“xinxxi”（记为事件 A）时，计算条件概率 $P(B_i \mid A)$（就是已知输入信息为 A 的条件下，输入者希望输出的词语是词库中第 i 个词的概率）. 当得到词库中所有词语对应的条件概率后，取其中概率最大的若干个词作为候选词，这样就完成了候选词自动筛

选的过程.

在这个过程中，如何计算条件概率 $P(B_i \mid A)$呢？这就要用到贝叶斯公式

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

因为只关心不同的 i 对应的条件概率的相对大小，而对同一输入信息 A，上述公式的分母不变，所以实际计算时，只要求分子 $P(B_i)P(A \mid B_i)$的值就可以达到目的，这能够减少计算的复杂程度，而且，其中的 $P(B_i)$可以通过统计第 i 个词语在大量文本中出现的频率来估计；通过衡量不同词语之间的某种特殊距离，可以得到 $P(A \mid B_i)$的值.

有意思的是，当长期使用某种“智能”输入法时，输入法给出的候选词会越来越贴近用户的使用习惯. 例如，如果用户在输入“xinxxi”后多次选择“新消息”这个词语，那么再次输入相同的字母后，输入法就会将“新消息”自动排在最前面，如图 3 所示. 实际上，这是因为输入法根据用户的选择习惯，更新了相应的词频，并重新计算了有关条件概率的缘故.

图 3

在垃圾邮件过滤、图像识别等人工智能应用场景中，贝叶斯公式都能发挥重要的作用，只不过情况会更复杂一些，感兴趣的同学可以自行查阅有关资料进一步了解.

练习A

1. 分别在下列各条件下，求 $P(BA)$：

   (1) $P(A)=0.2$，$P(B \mid A)=0.15$；　(2) $P(A)=0.6$，$P(B \mid A)=0.3$.

2. 已知 $P(BA)=0.35$，$P(B\overline{A})=0.1$，求 $P(B)$.

3. 分别在下列各条件下，求 $P(B)$，$P(A \mid B)$：

   (1) $P(A)=0.4$，$P(B \mid A)=0.25$，$P(B \mid \overline{A})=0.3$；

   (2) $P(A)=0.5$，$P(B \mid A)=0.2$，$P(B \mid \overline{A})=0.4$.

4. 某人翻开电话本给自己的一位朋友打电话时，发现电话号码的最后一位数字变得模糊不清了，因此决定随机拨号进行尝试. 求这个人正好尝试两次就拨对电话号码的概率.

5. 已知某学校中，经常参加体育锻炼的学生占 40%，而且在经常参加体育锻炼的学生中，喜欢篮球的占 25%. 从这个学校的学生中任意抽取一人，则抽到的学生经常参加体育锻炼而且喜欢篮球的概率是多少？

练习B

1. 已知 $P(A)=0.5$，$P(B \mid A)=0.2$，求 $P(BA)$ 与 $P(\overline{B}A)$.

2. 已知 $P(BA)=0.35$，$P(B)=0.72$，求 $P(B\overline{A})$.

3 已知 $P(B \mid A)=P(B)$，且 $P(A)=0.6$，$P(B)=0.3$，求 $P(AB)$.

4 在某次抽奖活动中，在甲、乙两人先后进行抽奖前，还有 20 张奖券，其中共有 3 张写有“中奖”字样. 假设抽完的奖券不放回，甲抽完之后乙再抽，求：

(1) 甲中奖而且乙也中奖的概率；

(2) 甲没中奖而且乙中奖的概率.

5 假设某市场供应的灯泡中，甲厂产品占 60%，乙厂产品占 40%，甲厂产品的合格率是 95%，乙厂产品的合格率是 80%. 在该市场中随机购买一个灯泡，已知买到的是合格品，求这个灯泡是甲厂生产的概率（精确到 0.1%）.

1 $1-\frac{1}{10}=\frac{9}{10}$　2 $\frac{5}{49}$　3 $\frac{9}{10}\times\frac{5}{49}=\frac{9}{98}$　4 $1-\frac{5}{8}=\frac{3}{8}$　5 $1-80\%=20\%$

6 $\frac{P(A)P(B \mid A)}{P(A)P(B \mid A)+P(\overline{A})P(B \mid \overline{A})}=\frac{80\%\times95\%}{80\%\times95\%+20\%\times60\%}\approx86.4\%$

4.1.3 独立性与条件概率的关系

情境与问题

从必修的内容中我们已经知道，A 与 B 相互独立（简称为独立）的充要条件是

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

而且 A 与 B 独立的直观理解是，事件 A 是否发生不会影响事件 B 发生的概率，事件 B 是否发生也不会影响事件 A 发生的概率. 那么，这个直观理解的数学含义是什么呢？

考察独立性与条件概率的关系可以得出相互独立的直观理解.

尝试与探究

假设 $P(A)>0$ 且 $P(B)>0$，在 A 与 B 独立的前提下，通过条件概率的计算公式考察 $P(A \mid B)$ 与 $P(A)$ 的关系，以及 $P(B \mid A)$ 与 $P(B)$ 的关系.

当 $P(B)>0$ 且 $P(AB)=P(A)P(B)$ 时，由条件概率的计算公式有

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

即 $P(A\mid B)=P(A)$. 这就是说，此时事件 A 发生的概率与已知事件 B 发生时事件 A 发生的概率相等. 也就是事件 B 的发生，不会影响事件 A 发生的概率.

类似地，可以看出，如果 $P(A\mid B)=P(A)$，那么一定有 $P(AB)=P(A)P(B)$.

因此，当 $P(B)>0$ 时，A 与 B 独立的充要条件是

$$P(A\mid B)=P(A).$$

这也就同时说明，当 $P(A\mid B)\neq P(A)$ 时，事件 B 的发生会影响事件 A 发生的概率，此时 A 与 B 是不独立的. 事实上，“A 与 B 独立”也经常被说成“A 与 B 互不影响”等.

例 1 已知某大学数学专业二年级的学生中，是否有自主创业打算的情况如下表所示.

	男生/人	女生/人
有自主创业打算	16	15
无自主创业打算	64	60

从这些学生中随机抽取一人：

(1) 求抽到的人有自主创业打算的概率；

(2) 求抽到的人是女生的概率；

(3) 若已知抽到的人是女生，求她有自主创业打算的概率；

(4) 判断“抽到的人是女生”与“抽到的人有自主创业打算”是否独立.

解 由题意可知，所有学生人数为

$$16+15+64+60=155.$$

记 A 为“抽到的人有自主创业打算”，B 为“抽到的人是女生”.

(1) 因为有自主创业打算的人数为 $16+15=31$，所以抽到的人有自主创业打算的概率为

$$P(A)=\boxed{1}\quad\_\_\_\_\_.$$

(2) 因为女生人数为 $15+60=75$，所以抽到的人是女生的概率为

$$P(B)=\boxed{2}\quad\_\_\_\_\_.$$

(3) 所要求的是 $P(A\mid B)$，注意到 75 名女生中有 15 人有自主创业

打算，因此

$$P(A \mid B)=\boxed{3}\quad .$$

（4）由（1）和（3）的计算结果可知 $P(A \mid B)=P(A)$，因此“抽到的人是女生”与“抽到的人有自主创业打算”独立.

多个事件之间的相互独立也可借助条件概率来理解，“A_1，A_2，…，A_n 相互独立”也可说成“A_1，A_2，…，A_n 相互不影响”. 需要强调的是，同以前一样，实际问题中，我们常常依据实际背景去判断事件之间是否存在相互影响，若可认为事件之间没有影响，则认为它们相互独立；已知事件相互独立时，根据每个事件发生的概率可以方便地求出它们同时发生的概率.

例 2 已知甲、乙、丙 3 人参加驾照考试时，通过的概率分别为 0.8，0.9，0.7，而且这 3 人之间的考试互不影响. 求：

（1）甲、乙、丙都通过的概率；

（2）甲、乙通过且丙未通过的概率.

解 用 A，B，C 分别表示甲、乙、丙驾照考试通过，则可知 A，B，C 相互独立，而且 $P(A)=0.8$，$P(B)=0.9$，$P(C)=0.7$.

（1）甲、乙、丙都通过可用 ABC 表示，因此所求概率为

$$\begin{aligned}P(ABC)&=P(A)P(B)P(C)\\&=0.8\times0.9\times0.7\\&=0.504.\end{aligned}$$

（2）甲、乙通过且丙未通过可用 $AB\overline{C}$ 表示，因此所求概率为

$$\begin{aligned}P(AB\overline{C})&=P(A)P(B)P(\overline{C})\\&=P(A)P(B)[1-P(C)]\\&=0.8\times0.9\times(1-0.7)\\&=0.216.\end{aligned}$$

例 3 在一个系统中，每一个部件能正常工作的概率称为部件的可靠度，而系统能正常工作的概率称为系统的可靠度. 现有甲、乙、丙 3 个部件组成的一个如图 4-1-6 所示的系统，已知当甲正常工作且乙、丙至少有一个能正常工作时，系统就能正常工作，各部件的可靠度均为 r（$0<r<1$），而且甲、乙、丙互不影响. 求系统的可靠度.

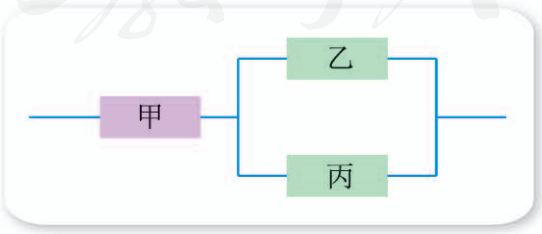


图 4-1-6

尝试与发现

例 3 中：

(1) 各个部件是否正常工作是相互独立的吗？

(2) 用合适的符号把系统能正常工作表示为互斥事件的和，并尝试给出解题思路.

解 用 A，B，C 分别表示甲、乙、丙能正常工作，D 表示系统能正常工作.

由题意知，系统能正常工作时，可分为三种互斥的情况：甲、乙、丙都正常工作，即 ABC；甲、丙正常工作，且乙不正常工作，即 $A\overline{B}C$；甲、乙正常工作，且丙不正常工作，即 $AB\overline{C}$. 因此

$$D=ABC\cup A\overline{B}C\cup AB\overline{C}.$$

因为甲、乙、丙互不影响，所以 A，B，C 相互独立，而且

$$P(A)=P(B)=P(C)=r.$$

由互斥事件概率的加法公式以及独立性可知

$$\begin{aligned}P(D)&=P(ABC\cup A\overline{B}C\cup AB\overline{C})\\&=P(ABC)+P(A\overline{B}C)+P(AB\overline{C})\\&=P(A)P(B)P(C)+P(A)P(\overline{B})P(C)+P(A)P(B)P(\overline{C})\\&=r^3+2r^2(1-r)\\&=2r^2-r^3.\end{aligned}$$

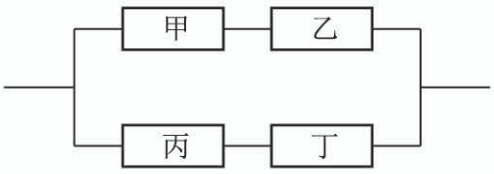
练习A

1. 已知 $P(A\mid B)=0.6$，$P(A)=0.59$，判断 A 与 B 是否独立.
2. 已知 A 与 B 独立，且 $P(A)=0.75$，求 $P(A\mid B)$.
3. 已知 $P(A\mid B)=0.6$，$P(\overline{A})=0.4$，判断 A 与 B 是否独立.
4. 已知 A 与 B 独立，且 $P(\overline{A})=0.3$，求 $P(A\mid B)$.
5. 加工某一零件需经过三道工序，设第一、二、三道工序的次品率分别为 $\frac{1}{70}$，$\frac{1}{69}$，$\frac{1}{68}$，且各道工序互不影响，求加工出来的零件的次品率.

练习B

1. 已知 A 与 B 独立，且 $P(AB)=\frac{5}{12}$，$P(B)=\frac{5}{6}$，求 $P(\overline{A}\mid B)$.

2 如图所示，已知一个系统由甲、乙、丙、丁4个部件组成. 当甲、乙都正常工作，或丙、丁都正常工作时，系统就能正常工作. 若每个部件的可靠性均为 r $(0<r<1)$，而且甲、乙、丙、丁互不影响. 求系统的可靠度.



（第 2 题）

3 针对某种突发性的流感病毒，各国的医疗科研机构都在研制疫苗. 已知甲、乙两个机构各自研制成功的概率为 $\frac{1}{5}$，$\frac{1}{3}$，而且两个机构互不影响，求：

(1) 甲、乙都研制成功的概率；

(2) 甲机构研制成功且乙机构研制不成功的概率；

(3) 甲、乙两个机构中，至少有一个研制成功的概率.

4 一批产品的次品率为 10%，进行有放回地重复抽样检查. 共抽取 3 件产品，求恰有 2 件次品的概率.

5 证明：当 $P(A)>0$，$P(B)>0$ 且 $P(B\mid A)=P(B)$ 时，有

$$P(\overline{B}\mid A)=P(\overline{B}),\ P(B\mid \overline{A})=P(B),\ P(\overline{B}\mid \overline{A})=P(\overline{B}).$$

你能给出这个结论的直观解释吗？

1 $\frac{31}{155}=\frac{1}{5}$　2 $\frac{75}{155}=\frac{15}{31}$　3 $\frac{15}{75}=\frac{1}{5}$

习题4-1A

1 已知某产品的品质是由 A，B 两项指标决定的，现有 100 件这样的产品，其中 A 指标达到优秀的有 80 件，B 指标达到优秀的有 75 件，A，B 两项指标都达到优秀的有 70 件. 从这批产品中任取一件，当已知所抽取的产品 A 指标优秀时，求 B 指标也优秀的概率.

2 已知 10 件产品中有 3 件是一等品，其余都是二等品. 从这些产品中不放回地抽取两次，若已知第一次取到的是一等品，求第二次取到的也是一等品的概率.

3 已知 $P(BA)=0.5$，$P(B\overline{A})=0.2$，求 $P(\overline{B})$.

4 某班级的学生中，是否有外地旅游经历的人数情况如下表所示.

	男生	女生
有外地旅游经历	6	9
无外地旅游经历	9	8

从这个班级中随机抽取一名学生：

(1) 求抽到的人是男生的概率；

(2) 求抽到的人是女生且无外地旅游经历的概率；

(3) 若已知抽到的人是女生，求她有外地旅游经历的概率；

(4) 若已知抽到的人有外地旅游经历，求其是男生的概率；

(5) 判断“抽到的人是女生”与“抽到的人有外地旅游经历”是否独立.

5 有 3 台机床，已知每台机床不需要照看的概率均为 0.8，且互不影响，求下列事件的概率：

(1) 3 台机床都不需要照看；

(2) 至少有 1 台机床需要照看；

(3) 3 台机床都需要照看.

6 李明早上上学的时候，可以乘坐公共汽车，也可以乘坐地铁. 已知李明乘坐公共汽车的概率为 0.3，乘坐地铁的概率为 0.7，而且乘坐公共汽车与地铁时，李明迟到的概率分别为 0.2 与 0.05.

(1) 求李明上学迟到的概率；

(2) 如果某天早上李明上学迟到了，那么他乘公共汽车的概率为多少？

习题4-1B

1 袋中有 a 个白球，b 个黑球，且 a，b 均为正整数，从中任意取一球，不放回，然后再取一球，求第二次取到白球的概率.

2 掷红、蓝两个均匀的骰子，已知两个骰子的点数不同，求其中至少有一个 6 点的概率.

3 某地区空气质量监测资料表明，一天的空气质量为优良的概率为 0.75，连续两天为优良的概率是 0.6，已知某天的空气质量为优良，则随后一天的空气质量为优良的概率是多少？

4 分别在下列各条件下，求 $P(BA)$：

(1) $P(\overline{A})=0.3$，$P(B\mid A)=0.6$；　(2) $P(\overline{A})=0.6$，$P(\overline{B}\mid A)=0.8$.

5 已知 $P(\overline{A})=0.6$，$P(B\mid A)=0.35$，$P(B\mid \overline{A})=0.2$，求 $P(\overline{B})$，$P(A\mid B)$.

习题4-1C

1 当 $0<P(A)<1$ 时，求证：$P(B\mid A)=P(B)$的充要条件是 $P(B\mid \overline{A})=P(B)$.

2 当 $P(A)>0$ 且 $P(B)>0$ 时，求证：$P(B\mid A)=P(B)$的充要条件是 $P(A\mid B)=P(A)$.

4.2 随机变量

4.2.1 随机变量及其与事件的联系

1. 随机变量的概念

我们已经知道，可以通过随机试验的样本空间来理解随机事件，在了解样本空间和随机事件所包含的样本点数目时，可以借助排列组合的知识.

情境与问题

为了督促各地做好环境保护工作，环保部门决定在 31 个省（自治区、直辖市）和新疆生产建设兵团中，随机抽取 6 个进行突击检查，抽得的结果只要有一个不同就认为是不同的试验结果，记样本空间为 Ω.

(1) Ω 中包含的样本点数目是多少？

(2) 设抽得的结果中直辖市个数为 X，那么对 Ω 中的每一个样本点，X 都有唯一确定的值吗？X 的取值是固定不变的吗？如果不是，X 可取的值有哪些？

借助组合的知识，可知情境与问题中 Ω 所包含的样本点数目为 C_{32}^{6}. 不过，因为我国只有北京市、上海市、天津市、重庆市这 4 个直辖市，而且是随机抽取，因此对样本空间 Ω 中的每一个样本点，变量 X 都有唯一的取值，但对不同的样本点，X 的取值可能不同，其值可以是 0，1，2，3，4 中的任意一个. 数学中，X 这样的变量称为随机变量.

一般地，如果随机试验的样本空间为 Ω，而且对于 Ω 中的每一个样本点，变量 X 都有唯一确定的实数值与之对应，就称 X 为一个**随机变量**. 随机变量一般用大写英文字母 X，Y，Z，…或小写希腊字母 ξ，η，ζ，…表示. 随机变量所有可能的取值组成的集合，称为这个随机变量的取值范围.

由定义可知，随机变量的取值由随机试验的结果决定.

例 1 先后抛两枚均匀的硬币，设正面朝上的硬币数为 X，样本空间为 Ω.

（1）借助合适的符号，用列举法写出样本空间 Ω；

（2）求出随机变量 X 的取值范围.

解 （1）用 FZ 表示第一枚硬币反面朝上，第二枚硬币正面朝上，则样本空间

$$\Omega=\{\mathrm{FF},\ \mathrm{FZ},\ \mathrm{ZF},\ \mathrm{ZZ}\}.$$

（2）因为有可能没有硬币正面朝上，也有可能恰有一枚硬币正面朝上，还有可能两枚硬币都正面朝上，所以 X 的取值范围是

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

尝试与发现

在例 1 中：

（1）$X=1$ 与样本空间 Ω 中的样本点之间有什么关系？

（2）记事件 A 为“恰有一枚硬币正面朝上”，写出 A 所包含的样本点，说明 $X=1$ 与事件 A 的关系；

（3）$X=1$ 与 $X=2$ 能同时成立吗？

不难看出，$X=1$ 的充要条件是试验结果为 FZ 或 ZF. 根据题意有

$$A=\{\mathrm{FZ},\ \mathrm{ZF}\}.$$

因此，$X=1$ 表示的就是“恰有一枚硬币正面朝上”，所以 $X=1$ 与事件 A 等价.

另外，因为 $X=2$ 表示的是“两枚硬币都正面朝上”，所以 $X=1$ 与 $X=2$ 是不能同时成立的，即事件 $X=1$ 与 $X=2$ 互斥.

更进一步，利用古典概型的知识可知

$$P(A)=\frac{2}{4}=\frac{1}{2},$$

由于 $X=1$ 与事件 A 等价，因此上述概率的表达式也可记作

$$P(X=1)=\frac{1}{2}.$$

由于这里的随机变量 X 只能取 0，1，2 中的某一个，所以 $0<X<2$ 与 $X=1$ 也是等价的，从而事件 A 也可用 $0<X<2$ 表示，因此

$$P(0<X<2)=P(X=1)=\frac{1}{2}.$$

这就是说，在引入了随机变量之后，可以利用随机变量来表示事件.

一般地，如果 X 是一个随机变量，a，b 都是任意实数，那么 $X=a$，$X\leqslant b$，$X>b$ 等都表示事件，而且：

（1）当 $a\neq b$ 时，事件 $X=a$ 与 $X=b$ 互斥；

（2）事件 $X\leqslant a$ 与 $X>a$ 相互对立，因此

$$P(X\leqslant a)+P(X>a)=1.$$

在用随机变量表示事件及事件的概率时，有时可不写出样本空间.

例如，抛一枚均匀硬币，如果正面朝上，取 $Z=1$；如果反面朝上，取 $Z=0$. 那么 Z 是一个随机变量，而且 Z 的取值范围是

$$\{1,\ 0\}.$$

此时，$Z=1$ 表示“正面朝上”，因此

$$P(Z=1)=\frac{1}{2};$$

$Z>-1$ 表示“正面朝上或者反面朝上”，因此

$$P(Z>-1)=1.$$

掷一个均匀的骰子，如果设朝上的点数为 Y，则 Y 是一个随机变量，且 Y 的取值范围是

1 _____.

此时，$Y=2$ 表示“朝上的点数为 2”，因此

$P(Y=2)=$ 2 _____；

$Y>3$ 表示“朝上的点数大于 3”，即“朝上的点数为 4，5，6 中的某一个”，因此

$P(Y>3)=$ 3 _____.

用 ξ 表示某网页在一天内（即 24 h 内）被浏览的次数，则 ξ 是一个随机变量，ξ 的取值范围可以认为是

$$\{0,\ 1,\ 2,\ 3,\ \cdots\}=\mathbf{N}.$$

若已知该网页在一天内被浏览的次数不超过 1 000 的概率为 0.3，则

$P(\xi\leqslant 1\ 000)=0.3$，$P(\xi>1\ 000)=$ 4 _____.

以上所介绍的随机变量，其所有可能的取值，都是可以一一列举出来的，它们都是**离散型随机变量**. 与离散型随机变量对应的是**连续型随机变量**，一般来说，连续型随机变量的取值范围包含一个区间. 例如，用 η 表示某品牌节能灯的寿命，则 η 的取值范围可以认为是$[0,\ +\infty)$，这里的 η 是一个连续型随机变量.

2. 随机变量之间的关系

尝试与发现

为了调动员工的积极性，某厂某月实行超额奖励制度，具体措施是：每超额完成 1 件产品，奖励 100 元. 假设这个月中，该厂的每名员工都完成了定额，而且超额完成的产品数都不超过 50. 从该厂员工中随机抽出一名，记抽出的员工该月超额完成的产品数为 X，获得的超额奖励为 Y 元，则 X 与 Y 均为随机变量.

(1) 当 $X=3$ 时，Y 的值是多少？总结 X 与 Y 之间的关系.

(2) 分别写出 X 与 Y 的取值范围.

因为 $X=3$ 表示超额完成了 3 件产品，所以按照奖励制度可知 $Y=100\times3=300$. 依照题意可知

$$Y=100X.$$

另外，由于 X 的取值范围是

$$\{0,1,2,3,\cdots,50\},$$

因此 Y 的取值范围是

$$\{0,100,200,300,\cdots,5\ 000\}.$$

上述 X 与 Y，虽然都是随机变量，但是它们之间的关系却是确定的：当 X 的值确定之后，Y 的值也就确定了；反之亦然.

一般地，如果 X 是一个随机变量，a，b 都是实数且 $a\neq0$，则

$$Y=aX+b$$

也是一个随机变量. 由于 $X=t$ 的充要条件是 $Y=at+b$，因此

$$P(X=t)=P(Y=at+b).$$

例 2 某快餐店的小时工是按照下述方式获取税前月工资的：底薪 1 000 元，每工作 1 h 再获取 30 元. 从该快餐店中任意抽取一名小时工，设其月工作时间为 X h，获取的税前月工资为 Y 元.

(1) 当 $X=110$ 时，求 Y 的值；

(2) 写出 X 与 Y 之间的关系式；

(3) 若 $P(X\leqslant120)=0.6$，求 $P(Y>4\ 600)$ 的值.

解 (1) 当 $X=110$ 时，表示工作了 110 个小时，所以

$$Y=110\times30+1\ 000=4\ 300.$$

(2) 根据题意有

$$Y=30X+1\ 000.$$

(3) 因为

$$X\leqslant120\Leftrightarrow30X\leqslant3\ 600\Leftrightarrow30X+1\ 000\leqslant4\ 600\Leftrightarrow Y\leqslant4\ 600,$$

所以

$$P(Y\leqslant4\ 600)=P(X\leqslant120)=0.6,$$

从而

$$P(Y>4\ 600)=\boxed{5}\quad.$$

练习A

1. 如果一批产品共有 100 件，其中恰有 3 件次品，现从这批产品中随机抽取 5 件进行检测，记样本空间为 Ω.

   (1) 用组合数表示样本空间中样本点的总数；

   (2) 假设所抽到的 5 件产品中次品数为 X，那么 X 可能的取值有多少个？

2. 写出下列随机变量的取值范围：

   (1) 某足球队在 5 次点球中，射进的球数为 X；

   (2) 从 10 张标号分别为 1，2，…，10 的卡片中随机抽取 1 张，所抽得的卡片标号为 Y；

   (3) 同时抛 5 枚硬币，正面朝上的硬币数为 Z.

3. 掷一个均匀的骰子，设朝上的点数为随机变量 Y，求 $P(Y\geqslant 5)$.

4. 已知随机变量 X 的取值范围是 $\{1, 0\}$，且 $Y=X+2$，求 Y 的取值范围.

5. 已知随机变量 X 的取值范围是 $\{1, 2, 3, 4, 5, 6\}$，且 $Y=2X$，求 Y 的取值范围.

练习B

1. 先后抛均匀硬币两次，如果两次都是正面朝上得 5 分，两次都是反面朝上得 3 分，其他结果得 0 分. 设 X 表示所得分数，求 $P(X=0)$.

2. 已知 X 是一个随机变量，a 是任意一个实数，分别说明下列各组事件之间的关系，并写出它们的概率之间的关系：

   (1) $X=a$，$X\neq a$；

   (2) $X<a$，$X\geqslant a$；

   (3) $X<a$，$X<a+1$.

3. 已知 $P(X\leqslant 0)=0.25$，求 $P(X>0)$的值.

4. 已知 $P(X=1)=0.3$，$P(X=-1)=0.1$，求 $P(|X|=1)$的值.

5. 某商场的促销员是按照下述方式获取税前月工资的：底薪 500 元，每工作 1 h 再获取 35 元. 从该商场促销员中任意抽取一名，设其月工作时间为 X h，获取的税前月工资为 Y 元.

   (1) 当 $X=80$ 时，求 Y 的值；

   (2) 写出 X 与 Y 之间的关系式；

   (3) 若 $P(Y>2\ 950)=0.27$，求 $P(X\leqslant 70)$的值.

1 $\{1, 2, 3, 4, 5, 6\}$　2 $\frac{1}{6}$　3 $\frac{3}{6}=\frac{1}{2}$　4 $1-0.3=0.7$

5 $1-P(Y\leqslant 4\ 600)=1-0.6=0.4$

4.2.2 离散型随机变量的分布列

1. 离散型随机变量的分布列

尝试与发现

已知随机变量 X 的取值范围是 $\{0, 1, 2\}$，而且

$$P(X=0)=0.2,$$

$$P(X=1)=0.4,$$

$$P(X=2)=0.4.$$

(1) 求出 $P(-1\leqslant X\leqslant 1)$ 与 $P(1\leqslant X\leqslant 2)$ 的值；

(2) 如果 a，b 是给定的实数，则 $P(a\leqslant X\leqslant b)$ 一定可以算出来吗？

(3) 探讨怎样才能对离散型随机变量有比较全面的了解.

由于 X 只能在 0，1，2 中取值，所以 $-1\leqslant X\leqslant 1$ 等价于 $X=0$ 或 $X=1$，又因为 $X=0$ 与 $X=1$ 互斥，所以

$$P(-1\leqslant X\leqslant 1)=P(X=0)+P(X=1)=0.2+0.4=0.6;$$

类似地，$1\leqslant X\leqslant 2$ 等价于 1 _____，而且

$$P(1\leqslant X\leqslant 2)=\text{2}\ \_\_\_\_\_.$$

因此，当实数 a，b 给定时，只要检查 0，1，2 是否满足 $a\leqslant X\leqslant b$ 就可以求出 $P(a\leqslant X\leqslant b)$.

由此可以看出，对于离散型随机变量来说，如果已知其每一个取值的概率，那么也就对其有了比较全面的了解.

一般地，当离散型随机变量 X 的取值范围是 $\{x_1, x_2, \cdots, x_n\}$ 时，如果对任意 $k\in\{1, 2, \cdots, n\}$，概率

$$P(X=x_k)=p_k$$

都是已知的，则称 X 的**概率分布**是已知的. 离散型随机变量 X 的概率分布可以用如下形式的表格表示，这个表格称为 X 的概率分布或**分布列**.

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

离散型随机变量 X 的概率分布还可以用图 4-2-1 或图 4-2-2 来直观表示，其中，图 4-2-1 中，x_k 上的矩形宽为 1、高为 p_k，因此每个矩形的面积也恰为 p_k；图 4-2-2 中，x_k 上的线段长为 p_k.

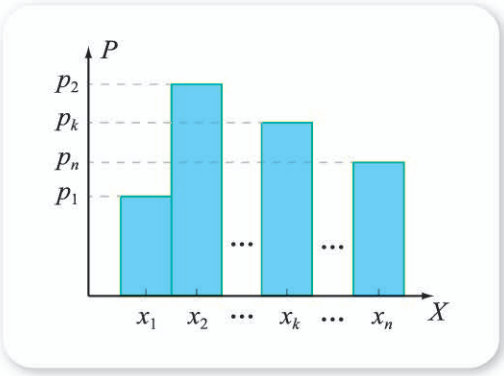


图 4-2-1

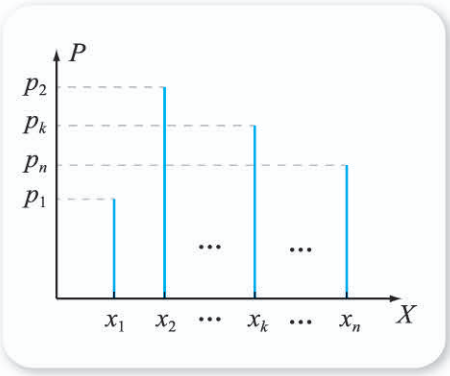


图 4-2-2

例如，对于尝试与发现中的随机变量 X 来说，其分布列如下表所示，而且 X 的概率分布可用图 4-2-3 直观表示.

X	0	1	2
P	0.2	0.4	0.4

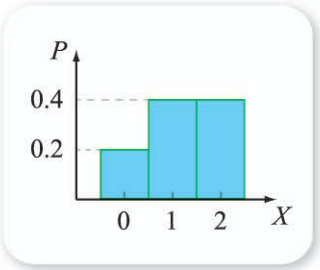


图 4-2-3

尝试与发现

观察上述分布列的实例，总结离散型随机变量 X 的分布列中 p_k 应具有的性质.

注意到 p_k 表示的是事件 $X=x_k$ 发生的概率，因此每一个 p_k 都是非负数；另外，因为分布列给出了随机变量能取的每一个值，而且随机变量取不同的值时的事件是互斥的，因此所有的 p_k 之和应该等于 1. 这就是说，离散型随机变量的分布列必须满足：

(1) $p_k \geqslant 0$，$k=1$，2，…，n；

(2) $\sum\limits_{k=1}^{n} p_k = p_1 + p_2 + \cdots + p_n =$ 3 .

例 1 掷一个均匀的骰子，记所得点数为 X.

(1) 求 X 的分布列；

(2) 求“点数大于 3”的概率.

解 (1) 因为 X 的取值范围是

$$\{1, 2, 3, 4, 5, 6\},$$

而且

$$P(X=n) = \boxed{4}\ ,\ n=1, 2, 3, 4, 5, 6,$$

因此 X 的分布列如下表所示.

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

（2）“点数大于 3”等价于 $X>3$，也就是说，X 可取 4，5，6 中的任何一个值，因此所求概率为

$$P(X>3)=P(X=4)+P(X=5)+P(X=6)=\boxed{5}.$$

例 2 抛一枚均匀的硬币 3 次，设正面朝上的次数为 X.

（1）说明 $X=2$ 表示的是什么事件，并求出 $P(X=2)$；

（2）求 X 的分布列.

解 （1）$X=2$ 表示的事件是“恰有 2 次正面朝上”.

因为抛一枚均匀的硬币 3 次，总共有 $2\times2\times2=8$ 种不同的情况，其中恰有两次正面朝上的情况共 $C_3^2=3$ 种，所以

$$P(X=2)=\frac{3}{8}.$$

（2）根据题意可知，X 的取值范围是

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

又因为用（1）中的方法可知

$$P(X=0)=\frac{1}{8},\ P(X=1)=\boxed{6},\ P(X=3)=\boxed{7}.$$

因此 X 的分布列如下表所示.

X	0	1	2	3
P	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

尝试与发现

在上一小节中我们已经看到，如果 X 是一个离散型随机变量，a，b 都是实数且 $a\neq0$，则 $Y=aX+b$ 也是一个离散型随机变量. 那么，它们的分布列之间有什么联系呢？

容易看出，当 X 与 Y 都是离散型随机变量而且 $Y=aX+b$（$a\neq0$）时，X 与 Y 的分布列分别如下表所示，它们的第二行的概率值是一样的.

X	x_1	x_2	…	x_k	…	x_n
P	p_1	p_2	…	p_k	…	p_n

$Y=aX+b$	ax_1+b	ax_2+b	…	ax_k+b	…	ax_n+b
P	p_1	p_2	…	p_k	…	p_n

2. 两点分布

尝试与发现

分别写出下列各随机变量的分布列，并分析它们的共同点.

(1) 篮球运动员在比赛中每次罚球得分的规则是：命中得 1 分，不中得 0 分. 已知某篮球运动员罚球命中的概率为 0.6，设其罚球一次的得分为 X.

(2) 假设某人寿保险的投保人年龄超过 50 岁的占 70%，从投保人中随机抽取 1 人，设 Y 表示抽到的年龄超过 50 岁的投保人人数.

(3) 从含有 3 件次品的 100 件产品中随机抽取 1 件，设抽到的次品数为 Z.

不难看出，上述尝试与发现中的 3 个随机变量，它们的取值范围均为 $\{1, 0\}$，而且分布列都能写成如下的表格形式（其中 $0<p<1$）.

W	1	0
P	p	$1-p$

但是，对 X 来说，$p=0.6$；对 Y 来说，$p=$ 8 _____；对 Z 来说，$p=$ 9 _____.

一般地，如果随机变量的分布列能写成上述表格的形式，则称这个随机变量服从参数为 p 的**两点分布**（或 **0-1 分布**）.

另外，一个所有可能结果只有两种的随机试验，通常称为**伯努利试验**. 不难看出，如果将伯努利试验的结果分别看成“成功”与“不成功”，并设“成功”出现的概率为 p，一次伯努利试验中“成功”出现的次数为 X，则 X 服从参数为 p 的两点分布，因此两点分布也常称为**伯努利分布**，两点分布中的 p 也常被称为成功概率.

上述尝试与发现中所涉及的 3 个试验均为伯努利试验. 日常生活中，还有很多随机试验都可看成伯努利试验. 例如，观察火车是否晚点、新生婴儿的性别、考试是否及格等.

练习A

1 分别判断下列表格是否可能是随机变量 X 的分布列，并说明理由：

(1)

X	−1	0	1	2	3
P	0.2	0.2	0.2	0.2	0.3

(2)

X	0	1	2	3	4	5
P	0.1	−0.2	0.3	0.4	0.2	0.2

2 已知离散型随机变量 X 的分布列如下表所示，求 a 的值.

X	1	2	3
P	0.3	a	0.5

3 抛一枚均匀的硬币，设 $X=\begin{cases}1，出现正面，\\0，出现反面，\end{cases}$ 写出 X 的分布列.

4 已知 X 服从参数为 0.3 的两点分布.

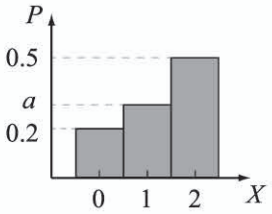
(1) 求 $P(X=0)$；

(2) 若 $Y=2X+1$，写出 Y 的分布列.

5 从装有 6 个白球和 4 个红球的口袋中任取 1 个球，用 X 表示取得的白球数，求 X 的分布列.

练习B

1 如图所示是离散型随机变量 X 的概率分布直观图，求 a 的值.



（第 1 题）

2 某商店购进一批西瓜，预计晴天西瓜畅销，可获利 1 000 元；阴天销路一般，可获利 500 元；下雨天西瓜滞销，会亏损 500 元. 根据天气预报，未来数日晴天的概率为 0.4，阴天的概率为 0.2，下雨的概率为 0.4，试写出销售这批西瓜获利的分布列.

3 某射击运动员射击一次所得环数 ξ 的分布列如下表所示.

ξ	4	5	6	7	8	9	10
P	0.03	0.05	0.07	0.08	0.26	a	0.23

(1) 求常数 a 的值；

(2) 求 $P(\xi>6)$.

4 抛一枚均匀的硬币 2 次，设正面朝上的次数为 X.

(1) 说明 $X=1$ 表示的是什么事件，并求出 $P(X=1)$；

(2) 求 X 的分布列.

5 同时掷两个均匀的骰子，设所得点数之和为 X.

(1) 写出 X 的分布列；

(2) 求 $P(X<5)$；

(3) 求“点数和大于 9”的概率.

1 $X=1$ 或 $X=2$　2 $P(X=1)+P(X=2)=0.4+0.4=0.8$　3 1　4 $\frac{1}{6}$

5 $\frac{1}{2}$　6 $\frac{3}{8}$　7 $\frac{1}{8}$　8 70%　9 $\frac{3}{100}$

4.2.3 二项分布与超几何分布

情境与问题

为了增加系统的可靠性，人们经常使用“备用冗余设备”（即正在使用的设备出故障时才启动的设备）. 已知某计算机网络的服务器采用的是“一用两备”（即一台正常设备，两台备用设备）的配置，这三台设备中，只要有一台能正常工作，计算机网络就不会断掉. 如果三台设备各自能正常工作的概率都为 0.9，它们之间相互不影响，那么这个计算机网络不会断掉的概率是多少呢？

情境中的问题，利用本节所要学习的知识，可以快速地得到解决.

1. n 次独立重复试验与二项分布

我们已经知道，一个伯努利试验是试验结果可记为“成功”与“不成功”的试验. 现实生活中，经常需要在相同的条件下将一个伯努利试验重复多次. 例如，为了观察抛硬币时出现的统计规律性，可多次重复进行抛硬币这个伯努利试验；为了了解支持改革的人的比例，可随机向多人进行访问，询问他们的态度是“支持”还是“不支持”；等等. 在相同条件下重复 n 次伯努利试验时，人们总是约定这 n 次试验是相互独立的，此时这 n 次伯努利试验也常称为 **n 次独立重复试验.**

例如，对一批产品进行抽样检查，每次取一件来判断是否合格，有放回地抽取 5 次，就是一个 5 次独立重复试验；篮球运动员练习投篮 10 次，可以认为每次投中的概率都相同，这是一个 10 次独立重复试验.

在 n 次独立重复试验中，人们经常关心的是“成功”出现的次数.

尝试与发现

已知某种药物对某种疾病的治愈率为 $\frac{3}{4}$，现有甲、乙、丙、丁 4 个患有该病的患者服用了这种药物，观察其中有多少患者会被这种药物治愈.

(1) 这能否看成独立重复试验？

(2) 求出甲、乙、丙都被治愈而丁没被治愈的概率；

(3) 求出恰有 3 个患者被治愈的概率；

(4) 设有 X 人被治愈，求 X 的分布列.

不难想到，4 个患者是否会被治愈是相互独立的，因此尝试与发现中的情形可以看成 4 次独立重复试验.

如果用 A_1，A_2，A_3，A_4 分别表示甲被治愈、乙被治愈、丙被治愈、丁被治愈，则不难看出

$$P(A_i)=\frac{3}{4},\ P(\overline{A}_i)=1-P(A_i)=\frac{1}{4},\ i=1,\ 2,\ 3,\ 4.$$

此时，甲、乙、丙都被治愈而丁没被治愈可以表示为 $A_1A_2A_3\overline{A}_4$，因此由独立性可知

$$\begin{aligned}P(A_1A_2A_3\overline{A}_4)&=P(A_1)P(A_2)P(A_3)P(\overline{A}_4)\\&=\frac{3}{4}\times\frac{3}{4}\times\frac{3}{4}\times\frac{1}{4}=\frac{27}{256}.\end{aligned}$$

注意到恰有 3 个患者被治愈的情况共有 C_4^3 种（4 个人中，选出 3 个是被治愈的，剩下的那个是没被治愈的），即

$$\overline{A}_1A_2A_3A_4,\ A_1\overline{A}_2A_3A_4,\ A_1A_2\overline{A}_3A_4,\ A_1A_2A_3\overline{A}_4,$$

这四种情况两两都是互斥的，而且每一种情况的概率均为 $\left(\frac{3}{4}\right)^3\times\frac{1}{4}=\frac{27}{256}$，因此所求概率为

$$\begin{aligned}&P(\overline{A}_1A_2A_3A_4+A_1\overline{A}_2A_3A_4+A_1A_2\overline{A}_3A_4+A_1A_2A_3\overline{A}_4)\\&=P(\overline{A}_1A_2A_3A_4)+P(A_1\overline{A}_2A_3A_4)+P(A_1A_2\overline{A}_3A_4)+P(A_1A_2A_3\overline{A}_4)\\&=C_4^3\times\left(\frac{3}{4}\right)^3\times\frac{1}{4}=\frac{27}{64}.\end{aligned}$$

因为共有 4 名患者服用了药物，所以 X 的取值范围应该是

$$\{0,\ 1,\ 2,\ 3,\ 4\},$$

而且我们已经算出

$$P(X=3)=C_4^3\times\left(\frac{3}{4}\right)^3\times\frac{1}{4}=\frac{27}{64},$$

用类似的办法可知

$$P(X=0)=C_4^0\times\left(\frac{3}{4}\right)^0\times\left(\frac{1}{4}\right)^4=\frac{1}{256},$$

$$P(X=1)=C_4^1\times\frac{3}{4}\times\left(\frac{1}{4}\right)^3=\frac{3}{64},$$

$$P(X=2)=\boxed{1}\_\_\_\_\_,$$

$$P(X=4)=\boxed{2}\_\_\_\_\_,$$

因此 X 的分布列为

X	0	1	2	3	4
P	$\frac{1}{256}$	$\frac{3}{64}$	$\frac{27}{128}$	$\frac{27}{64}$	$\frac{81}{256}$

一般地，如果一次伯努利试验中，出现“成功”的概率为 p，记 $q=1-p$，且 n 次独立重复试验中出现“成功”的次数为 X，则 X 的取值范围是

$$\{0,\ 1,\ \cdots,\ k,\ \cdots,\ n\},$$

而且

$$P(X=k)=C_n^k p^k q^{n-k},\ k=0,\ 1,\ \cdots,\ n,$$

因此 X 的分布列如下表所示.

X	0	1	$\cdots$	k	$\cdots$	n
P	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	$\cdots$	$C_n^k p^k q^{n-k}$	$\cdots$	$C_n^n p^n q^0$

注意到上述 X 的分布列第二行中的概率值都是二项展开式

$$(q+p)^n=C_n^0 p^0 q^n+C_n^1 p^1 q^{n-1}+\cdots+C_n^k p^k q^{n-k}+\cdots+C_n^n p^n q^0$$

中对应项的值，因此称 X 服从参数为 n，p 的**二项分布**，记作

$$X\sim B(n,\ p).$$

由此可以看出，上述尝试与发现中的随机变量 X 服从参数为 4，$\frac{3}{4}$ 的二项分布，即 $X\sim B(4,\ \frac{3}{4})$. 服从二项分布的随机变量，其概率分布也可用图直观地表示，如图 4-2-4 所示.

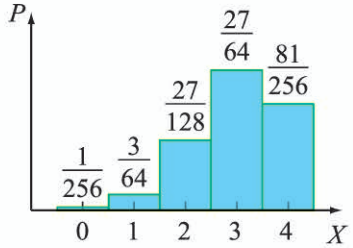


图 4-2-4

例 1 设本节一开始的情境与问题中，能正常工作的设备数为 X.

（1）写出 X 的分布列；

（2）求出计算机网络不会断掉的概率.

解 （1）可以看出，X 服从参数为 3，0.9 的二项分布，即

$$X\sim B(3,\ 0.9).$$

因此

$$P(X=0)=C_3^0\times 0.9^0\times(1-0.9)^3=0.001,$$

$$P(X=1)=C_3^1\times 0.9^1\times(1-0.9)^2=0.027,$$

$$P(X=2)=C_3^2\times 0.9^2\times(1-0.9)^1=0.243,$$

$$P(X=3)=C_3^3\times 0.9^3\times(1-0.9)^0=0.729,$$

从而 X 的分布列为

X	0	1	2	3
P	0.001	0.027	0.243	0.729

（2）要使得计算机网络不会断掉，也就是要求能正常工作的设备至少有一台，即 $X \geqslant 1$，因此所求概率为

$$P(X \geqslant 1)=1-P(X<1)=1-P(X=0)=1-0.001=0.999.$$

例 2 假设某种人寿保险规定，投保人没活过 65 岁时，保险公司要赔偿 100 万元；活过 65 岁时，保险公司不赔偿. 已知购买此种人寿保险的每个投保人能活过 65 岁的概率都为 0.8. 随机抽取 3 个投保人，设其中活过 65 岁的人数为 X，保险公司要赔偿给这三人的总金额为 Y 万元.

（1）指出 X 服从的分布；

（2）写出 Y 与 X 的关系；

（3）求 $P(Y=300)$.

解 （1）不难看出，X 服从参数为 3，0.8 的二项分布，即

$$X \sim B(3, 0.8).$$

（2）因为 3 个投保人中，活过 65 岁的人数为 X，则没活过 65 岁的人数为 $3-X$，因此

$$Y=100(3-X).$$

（3）因为

$$Y=300 \Leftrightarrow 100(3-X)=300 \Leftrightarrow X=0,$$

所以

$$\begin{aligned} P(Y=300) &= P(X=0) \\ &= C_3^0 \times 0.8^0 \times (1-0.8)^3 \\ &= 0.008. \end{aligned}$$

2. 用信息技术计算二项分布的概率值

利用二项分布的知识可以求解很多问题. 例如，将一枚均匀的硬币抛 100 次，求正好出现 50 次正面的概率时，可以设正面出现的次数为 X，则 X 服从参数为 100，0.5 的二项分布，即 $X \sim B(100, 0.5)$，因此所求概率为

$$P(X=50)=C_{100}^{50} \times 0.5^{50} \times (1-0.5)^{50}=C_{100}^{50} \times 0.5^{100}.$$

不过，要手工算出这个概率的小数形式并不容易，但我们可以借助计算机软件来完成. 例如，在 Excel 中，只要在任何一个单元格输入

=BINOM.DIST(50，100，0.5，FALSE)

即可得到上述概率的小数形式，如图 4-2-5 所示.

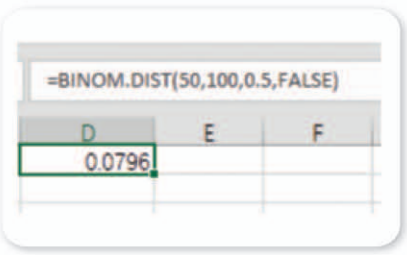


图 4-2-5

利用 GeoGebra 的概率与统计功能，选择二项分布后，一样可以得到有关的概率值，如图 4-2-6 所示.

图 4-2-6

不管用什么软件，我们都能算出，将一枚均匀的硬币抛 100 次，正好出现 50 次正面的概率大约只有 7.96%！这是不是让你觉得有些意外？需要强调的是，抛一枚均匀的硬币，出现正面的概率为 0.5，但这并不能保证，抛均匀硬币 100 次时，一定会出现 50 次正面，只能说出现的正面次数在 50 左右的概率比较大，这也可以通过下表的计算结果看出来.

范围	概率	范围	概率
$49\leqslant X\leqslant 51$	23.6%	$44\leqslant X\leqslant 56$	80.7%
$48\leqslant X\leqslant 52$	38.3%	$43\leqslant X\leqslant 57$	86.7%
$47\leqslant X\leqslant 53$	51.6%	$42\leqslant X\leqslant 58$	91.1%
$46\leqslant X\leqslant 54$	63.2%	$41\leqslant X\leqslant 59$	94.3%
$45\leqslant X\leqslant 55$	72.9%	$40\leqslant X\leqslant 60$	96.5%

3. 超几何分布

尝试与发现

某校组织一次认识大自然的夏令营活动，有 10 名同学参加，其中有 6 名男生、4 名女生，现要从这 10 名同学中随机抽取 3 名去采集自然标本.

(1) 抽取的人中恰有 1 名女生的概率是多少？

(2) 设抽取的人中女生有 X 名，写出 X 的分布列.

注意到从 10 名同学中随机抽取 3 人，共有 C_{10}^3 种不同的抽法，也就是说，样本空间中样本点的数量是 4 . 另外，抽取的人中恰有 1 名女生，等价于抽取的是 1 名女生和 2 名男生，因此包含的样本点数为 $C_4^1C_6^2$，因此所求概率为

$$\frac{C_4^1C_6^2}{C_{10}^3}=\frac{1}{2}.$$

如果抽取的人中女生数为 X，则 X 的取值范围是

$$\{0，1，2，3\}，$$

而且我们已经算出

$$P(X=1)=\frac{C_4^1C_6^2}{C_{10}^3}=\frac{1}{2}，$$

用类似的办法可知

$$P(X=0)=\frac{C_4^0C_6^3}{C_{10}^3}=\frac{1}{6}，$$

$$P(X=2)=\boxed{5}\_\_\_\_\_，$$

$$P(X=3)=\boxed{6}\_\_\_\_\_，$$

因此 X 的分布列为

X	0	1	2	3
P	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{3}{10}$	$\frac{1}{30}$

一般地，若有总数为 N 件的甲、乙两类物品，其中甲类有 M 件（$M<N$），从所有物品中随机取出 n 件（$n\leqslant N$），则这 n 件中所含甲类物品数 X 是一个离散型随机变量，X 能取不小于 t 且不大于 s 的所有自然数，其中 s 是 M 与 n 中的较小者，t 在 n 不大于乙类物品件数（即 $n\leqslant N-M$）时取 0，否则 t 取 n 减乙类物品件数之差（即 $t=n-(N-M)$），而且

$$P(X=k)=\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}，\ k=t，t+1，\cdots，s，$$

这里的 X 称为服从参数为 N，n，M 的**超几何分布**，记作

$$X\sim H(N，n，M).$$

特别地，如果 $X\sim H(N，n，M)$ 且 $n+M-N\leqslant 0$，则 X 能取所有不大于 s 的自然数，此时 X 的分布列如下表所示.

X	0	1	$\cdots$	k	$\cdots$	s
P	$\frac{C_M^0C_{N-M}^n}{C_N^n}$	$\frac{C_M^1C_{N-M}^{n-1}}{C_N^n}$	$\cdots$	$\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}$	$\cdots$	$\frac{C_M^sC_{N-M}^{n-s}}{C_N^n}$

由此可以看出，上述尝试与发现中的随机变量 X 服从参数为 10，3，4 的超几何分布，即 $X\sim H(10，3，4)$. 服从超几何分布的随机变量，其概率分布也可用图直观地表示，如图 4-2-7 所示.

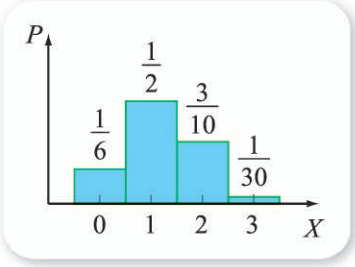


图 4-2-7

例 3 学校要从 5 名男教师和 2 名女教师中随机选出 3 人去支教，设抽取的人中女教师的人数为 X，求 $P(X\leqslant 1)$.

解 由题意知，X 服从参数为 7，3，2 的超几何分布，即 $X\sim H(7,3,2)$，因此

$$P(X\leqslant 1)=P(X=0)+P(X=1)=\frac{C_2^0C_5^3}{C_7^3}+\frac{C_2^1C_5^2}{C_7^3}=\frac{6}{7}.$$

例 4 袋中有 8 个白球、2 个黑球，从中随机地连续抽取 3 次，每次取 1 个球.

（1）若每次抽取后都放回，设取到黑球的个数为 X，求 X 的分布列；

（2）若每次抽取后都不放回，设取到黑球的个数为 Y，求 Y 的分布列.

解 （1）若每次抽取后都放回，则每次抽到黑球的概率均为 $\frac{2}{8+2}=\frac{1}{5}$.

而 3 次取球可以看成 3 次独立重复试验，因此 $X\sim B(3,\frac{1}{5})$，所以

$$P(X=0)=C_3^0\times\left(\frac{1}{5}\right)^0\times\left(\frac{4}{5}\right)^3=\frac{64}{125},$$

$$P(X=1)=C_3^1\times\left(\frac{1}{5}\right)^1\times\left(\frac{4}{5}\right)^2=\frac{48}{125},$$

$$P(X=2)=C_3^2\times\left(\frac{1}{5}\right)^2\times\left(\frac{4}{5}\right)^1=\frac{12}{125},$$

$$P(X=3)=C_3^3\times\left(\frac{1}{5}\right)^3\times\left(\frac{4}{5}\right)^0=\frac{1}{125}.$$

因此 X 的分布列为

X	0	1	2	3
P	$\frac{64}{125}$	$\frac{48}{125}$	$\frac{12}{125}$	$\frac{1}{125}$

（2）若每次抽取后都不放回，则随机抽取 3 次可看成随机抽取 1 次，但 1 次抽取了 3 个，因此黑球数 Y 服从参数为 10，3，2 的超几何分布，即

$Y\sim$ 7 ，

因此

$$P(Y=0)=\frac{C_2^0C_8^3}{C_{10}^3}=\frac{7}{15},$$

$$P(Y=1)=\frac{C_2^1C_8^2}{C_{10}^3}=\frac{7}{15},$$

$$P(Y=2)=\frac{C_2^2C_8^1}{C_{10}^3}=\frac{1}{15}.$$

因此 Y 的分布列为

Y	0	1	2
P	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$

例 4 中的（2）也可直接使用有关排列组合的知识求解，感兴趣的读者可以自行尝试. 由例 4 可知，若 N 件产品中共有 M 件次品，当我们从这些产品中每次抽取一件，共抽取 n 次进行检查时，若是有放回地抽样，则抽到的次品数 X 服从的是二项分布；若是不放回地抽样且 $n \leqslant N$，则抽到的次品数 X 服从的是超几何分布.

探索与研究

若 N 件产品中共有 M 件次品（$N>1$，$M>1$，$N>M$），则不放回地抽样中，第一次抽到次品的概率为 $\frac{M}{N}$，而第二次抽到次品的概率与第一次抽到的是否为次品有关：若第一次抽到的是次品，则第二次抽到次品的概率为 $\frac{M-1}{N-1}$；若第一次抽到的不是次品，则第二次抽到次品的概率为 $\frac{M}{N-1}$. 不过，当 N 相对 M 来说很大时，$\frac{M}{N-1}$ 与 $\frac{M-1}{N-1}$ 都可以近似为 $\frac{M}{N}$.

由以上信息出发，探索二项分布与超几何分布之间的联系.

4. 用信息技术计算超几何分布的概率值

如果 $X \sim H(N, n, M)$，则不难看出，当 N 与 n 的值都比较大时，计算有关的概率值并不容易，例如，计算 $\frac{C_{20}^{6} C_{80}^{4}}{C_{100}^{10}}$ 就会非常烦琐. 不过，同二项分布的概率值一样，超几何分布的概率值也可通过计算机软件求出来.

例如，在 Excel 中，只要在任何一个单元格输入

$$=\text{HYPGEOM.DIST}(6, 10, 20, 100, \text{FALSE})$$

即可得到上述概率的小数形式，如图 4-2-8 所示.

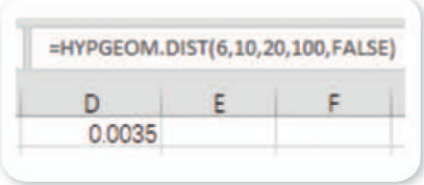


图 4-2-8

利用 GeoGebra 的概率与统计功能，选择超几何分布后，一样可以得到有关的概率值，如图 4-2-9 所示.

图 4-2-9

练习A

❶ 分别指出下列随机变量服从的分布：

(1) 即将出生的 100 个新生婴儿中，男婴的个数 X；

(2) 已知某幼儿园有 125 个孩子，其中男孩有 62 人，从这些孩子中随机抽取 10 人，设抽到男孩的人数为 X.

❷ 一个车间有 5 台同类型的且独立工作的机器，假设每天启动时，每台机器出故障的概率均为 0.1. 设某天启动时，出故障的机器数为 X.

(1) 写出 X 的分布列；

(2) 求该天机器启动时，至少有 3 台机器出故障的概率.

❸ 市教育局决定在所管辖的 20 所中学中随机抽取 3 所进行教学质量检测，已知 20 所中学中农村中学共有 8 所，设抽到的农村中学共有 X 所，指出 X 服从的分布，并求出 $P(X=3)$的值.

❹ 张明从家坐公交车到学校的途中，会通过 3 个有红绿灯的十字路口，假设在每个十字路口遇到红灯的概率均为 0.25，而且在各路口是否遇到红灯是相互独立的. 设 X 为张明在途中遇到的红灯数，求随机变量 X 的分布列.

❺ 袋中有 6 个白球、3 个黑球，从中随机地连续抽取 2 次，每次取 1 个球.

(1) 若每次抽取后都放回，设取到黑球的个数为 X，求 X 的分布列；

(2) 若每次抽取后都不放回，设取到黑球的个数为 Y，求 Y 的分布列.

练习B

❶ 已知某气象站天气预报的准确率为 80%，求 3 次预报中：

(1) 恰有 2 次预报准确的概率；

(2) 至少有 2 次预报准确的概率；

(3) 恰有 2 次预报准确且其中第 3 次预报准确的概率.

❷ 分别指出下列随机变量服从的分布，并用合适的符号表示：

(1) 某班级共有 30 名学生，其中有 10 名学生戴眼镜，随机从这个班级中抽取 5 人，设抽到的不戴眼镜的人数为 X；

(2) 已知女性患色盲的概率为 0.25%，任意抽取 300 名女性，设其中患色盲的人数为 X；

(3) 学校要从 3 名男教师和 4 名女教师中随机选出 3 人去支教，设抽取的人中男教师的人数为 X.

❸ 从 4 名男生和 2 名女生中任选 3 人参加演讲比赛，用 X 表示所选 3 人中女生的人数.

(1) 求 X 的分布列；　　(2) 求 $P(X\leqslant 1)$.

❹ 已知 $X\sim B(3,\frac{1}{4})$，且 $Y=2X+1$，求 Y 的分布列.

❺ 设某种疾病的发病率为 0.001，且每个人是否患有这种疾病是相互独立的. 已知一个单位有 1 000 名员工，求这个单位至少有 1 人患有这种疾病的概率.

❶ $C_4^2\times\left(\frac{3}{4}\right)^2\times\left(\frac{1}{4}\right)^2=\frac{27}{128}$　❷ $C_4^4\times\left(\frac{3}{4}\right)^4\times\left(\frac{1}{4}\right)^0=\frac{81}{256}$　❸ $B(3,0.8)$

❹ C_{10}^3　❺ $\frac{C_4^2C_6^1}{C_{10}^3}=\frac{3}{10}$　❻ $\frac{C_4^3C_6^0}{C_{10}^3}=\frac{1}{30}$　❼ $H(10,3,2)$

4.2.4 随机变量的数字特征

1. 离散型随机变量的均值

情境与问题

一家投资公司在决定是否对某创业项目进行资助时，经过评估后发现：如果项目成功，将获利 5 000 万元；如果项目失败，将损失 3 000 万元. 设这个项目成功的概率为 p，而你是投资公司的负责人，如果仅从平均收益方面考虑，则 p 满足什么条件时你才会对该项目进行资助？为什么？

上述情境中，平均收益显然与 p 的取值有关. 例如，当 $p=1$ 时，平均收益应为 5 000 万元；而当 $p=0$ 时，平均收益应为−3 000 万元. 一般情形下的平均收益该怎样确定呢?

注意到成功的概率为 p，指的是如果重复这个创业项目足够多次（设为 n 次），那么成功的次数可以用 np 来估计，而失败的次数可以估计为

$$n-np=n(1-p).$$

因此，在这 n 次试验中，投资方收益（单位：万元）的 n 个数据可以估计为

$$\underbrace{5\ 000,5\ 000,\cdots,5\ 000}_{np\text{个}},\ \underbrace{-3\ 000,-3\ 000,\cdots,-3\ 000}_{n(1-p)\text{个}},$$

这一组数的平均数为

$$\frac{5\ 000np+(-3\ 000)n(1-p)}{n}=5\ 000p+(-3\ 000)(1-p).$$

因为上述平均数体现的是平均收益，所以不难想到，当

$$5\ 000p+(-3\ 000)(1-p)>0,$$

即 $p>0.375$ 时，就应该对创业项目进行资助.

另一方面，如果设投资公司的收益为 X 万元，则 X 这个随机变量的分布列如下表所示.

X	5 000	−3 000
P	p	$1-p$

从上面的分析可以看出，式子

$$5\ 000p+(-3\ 000)(1-p)$$

刻画了 X 取值的平均水平.

一般地，如果离散型随机变量 X 的分布列如下表所示.

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

则称

$$E(X)=x_1p_1+x_2p_2+\cdots+x_np_n=\sum_{i=1}^{n}x_ip_i$$

为离散型随机变量 X 的**均值**或**数学期望**（简称为**期望**）.

离散型随机变量 X 的均值 $E(X)$ 也可用 EX 表示，它刻画了 X 的平均取值. 在离散型随机变量 X 的分布列的直观图中，$E(X)$ 处于平衡位置. 例如，情境与问题中收益 X 的均值为

$$E(X)=5\ 000p+(-3\ 000)(1-p)=8\ 000p-3\ 000,$$

分别取 $p=0.5$ 与 $p=0.7$，则 X 的分布列可分别用 4-2-10（1）与（2）表示. 而且，在图 4-2-10（1）中，$E(X)=1\ 000$；在图 4-2-10（2）中，$E(X)=2\ 600$.

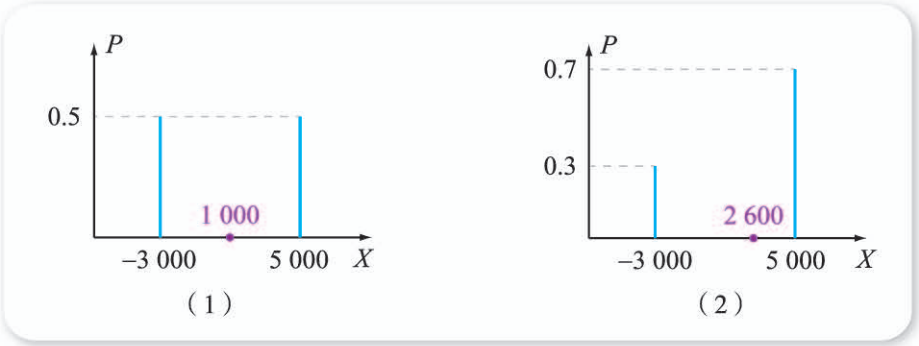


图 4-2-10

例 1 已知随机变量 X 服从参数为 p 的两点分布，求 $E(X)$.

解 因为 X 只能取 1，0 这两个值，而且 $P(X=1)=p$，所以

$$E(X)=\boxed{1}.$$

类似地，由离散型随机变量均值的定义，可以算出离散型随机变量服从二项分布、超几何分布时的均值，即：

（1）若 X 服从参数为 n，p 的二项分布，即 $X\sim B(n,p)$，则

$$E(X)=np;$$

（2）若 X 服从参数为 N，n，M 的超几何分布，即 $X\sim H(N,n,M)$，则

$$E(X)=\frac{nM}{N}.$$

尝试与发现

已知 X 是一个随机变量，且分布列如下表所示.

X	x_1	x_2	…	x_k	…	x_n
P	p_1	p_2	…	p_k	…	p_n

设 a，b 都是实数且 $a\neq 0$，则 $Y=aX+b$ 也是一个随机变量. 那么，这两个随机变量的均值之间有什么联系呢?

若 X 与 Y 都是随机变量，且 $Y=aX+b$（$a\neq 0$），则由 X 与 Y 之间分布列的关系可知

$$\begin{aligned}E(Y)&=(ax_1+b)p_1+(ax_2+b)p_2+\cdots+(ax_n+b)p_n\\&=a(x_1p_1+x_2p_2+\cdots+x_np_n)+b(p_1+p_2+\cdots+p_n)\\&=aE(X)+b.\end{aligned}$$

事实上，在前述的情境与问题中，如果项目成功，记 $W=1$；如果项目失败，记 $W=0$. 则可知 W 服从参数为 p 的两点分布，$E(W)=\boxed{2}$.

另一方面，W 与收益 X 之间的关系可以写成 $X=8\,000W-3\,000$，因此

$$E(X)=8\,000E(W)-3\,000=8\,000p-3\,000.$$

例 2 体检时，为了确定体检人是否患有某种疾病，需要对其血液进行化验，若结果呈阳性，则患有该疾病；若结果呈阴性，则未患有该疾病. 已知每位体检人患有该疾病的概率均为 0.1，化验结果不会出错，而且各体检人是否患有该疾病相互独立. 现有 5 位体检人的血液待检查，有以下两种化验方案：

方案甲：逐个检查每位体检人的血液；

方案乙：先将 5 位体检人的血液混在一起化验一次，若呈阳性，则再逐个化验；若呈阴性，则说明每位体检人均未患有该疾病，化验结束.

（1）哪种化验方案更好？

（2）如果每次化验的费用为 100 元，求方案乙的平均化验费用.

解 （1）方案甲中，化验的次数一定为 5 次.

方案乙中，若记化验次数为 X，则 X 的取值范围是 $\{1,6\}$. 因为 5 人都不患病的概率为

$$(1-0.1)^5=0.590\,49,$$

所以

$$P(X=1)=0.590\,49,$$

$$P(X=6)=\boxed{3}\quad\_\_\_\_\_.$$

从而

$$E(X)=1\times0.590\,49+6\times0.409\,51=3.047\,55.$$

这就是说，方案乙的平均检查次数不到 5 次，因此方案乙更好.

（2）若记方案乙中，检查费用为 Y 元，则 $Y=100X$，从而可知

$$E(Y)\approx\boxed{4}\quad\_\_\_\_\_,$$

即方案乙的平均化验费用为 304.76 元.

2. 离散型随机变量的方差

情境与问题

某省要从甲、乙两名射击运动员中选出一人参加全国运动会（简称“全运会”），根据以往数据，这两名运动员射击环数的分布列分别如下. 如果从平均水平和发挥稳定性角度来考虑，要你来决定谁参加全运会，你会怎样决定？说明理由.

甲的环数 X_1	8	9	10
P	0.2	0.6	0.2

乙的环数 X_2	8	9	10
P	0.4	0.2	0.4

上述情境中，不难算出 $E(X_1)=E(X_2)=9$，这就是说，如果仅从平均水平的角度考虑，是不能决定选谁参加的. 怎样来衡量他们的发挥稳定性呢?

设甲、乙两人每人都重复射击足够多次（设为 n 次），则甲所得环数可估计为

$$\underbrace{8,8,\cdots,8}_{0.2n\text{个}},\ \underbrace{9,9,\cdots,9}_{0.6n\text{个}},\ \underbrace{10,10,\cdots,10}_{0.2n\text{个}},$$

乙所得环数可估计为

$$\underbrace{8,8,\cdots,8}_{0.4n\text{个}},\ \underbrace{9,9,\cdots,9}_{0.2n\text{个}},\ \underbrace{10,10,\cdots,10}_{0.4n\text{个}}.$$

我们已经知道，这两组数的平均数是相等的，都是 9. 而甲这组数的方差为

$$\begin{aligned}&\frac{(8-9)^2\times 0.2n+(9-9)^2\times 0.6n+(10-9)^2\times 0.2n}{n}\\&=(8-9)^2\times 0.2+(9-9)^2\times 0.6+(10-9)^2\times 0.2\\&=0.4,\end{aligned}$$

类似地，乙这组数的方差为

$$\begin{aligned}&\frac{(8-9)^2\times 0.4n+(9-9)^2\times 0.2n+(10-9)^2\times 0.4n}{n}\\&=(8-9)^2\times 0.4+(9-9)^2\times 0.2+(10-9)^2\times 0.4\\&=0.8,\end{aligned}$$

由于 $0.4<0.8$，因此可以认为甲的发挥更稳定，从这一角度来说，应该派甲参加全运会.

由上可以看出，如果离散型随机变量 X 的分布列如下表所示.

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

因为 X 的均值为 $E(X)$，所以

$$\begin{aligned}D(X)&=[x_1-E(X)]^2p_1+[x_2-E(X)]^2p_2+\cdots+[x_n-E(X)]^2p_n\\&=\sum_{i=1}^{n}[x_i-E(X)]^2p_i\end{aligned}$$

能够刻画 X 相对于均值的离散程度（或波动大小），这称为离散型随机变量 X 的**方差**.

离散型随机变量 X 的方差 $D(X)$ 也可用 DX 表示. 一般地，$\sqrt{D(X)}$ 称为离散型随机变量 X 的**标准差**，它也可以刻画一个离散型随机变量的离散程度（或波动大小）.

由此可知，情境与问题中，$D(X_1)=0.4$，$D(X_2)=0.8$. 更进一步，如果将甲、乙射击环数的分布列用图直观地表示，如图 4-2-11（1）（2）所示，则从图上也可看出 $D(X_1)$ 与 $D(X_2)$ 的相对大小.

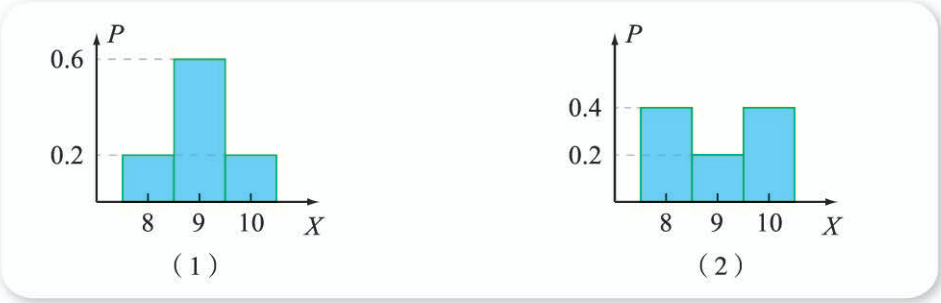


图 4-2-11

例 3 已知随机变量 X 服从参数为 p 的两点分布，求 $D(X)$.

解 因为 X 只能取 1，0 这两个值，而且 $P(X=1)=p$，$E(X)=p$，所以

$$D(X)=(1-p)^2\times p+(0-p)^2\times(1-p)=p(1-p).$$

类似地，若 X 服从参数为 n，p 的二项分布，即 $X\sim B(n,p)$，则由离散型随机变量方差的定义，可以算得

$$D(X)=np(1-p).$$

尝试与发现

已知 X 是一个随机变量，且分布列如下表所示.

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

设 a，b 都是实数且 $a\neq 0$，则 $Y=aX+b$ 也是一个随机变量，而且 $E(Y)=aE(X)+b$. 那么，这两个随机变量的方差之间有什么联系呢？

若 X 与 Y 都是离散型随机变量，且 $Y=aX+b$（$a\neq 0$），则由 X 与 Y 之间分布列和均值之间的关系可知

$$\begin{aligned}D(Y)&=\sum_{i=1}^{n}[(ax_i+b)-aE(X)-b]^2p_i\\&=a^2\sum_{i=1}^{n}[x_i-E(X)]^2p_i\\&=a^2D(X).\end{aligned}$$

例 4 已知一批产品的次品率为 0.02，从这批产品中每次随机取一件，有放回地抽取 50 次，用 X 表示抽到的次品数.

(1) 求 $D(X)$；

(2) 假设抽出的产品需要专门检测，检测费用 Y 元与次品数 X 有关，且 $Y=10X+300$，求 $D(Y)$.

解 (1) 因为 X 服从的是参数为 50，0.02 的二项分布，即 $X\sim B(50,0.02)$，所以

$$D(X)=\boxed{5}.$$

(2) 由 $Y=10X+300$ 可知

$$D(Y)=D(10X+300)=10^2D(X)=100\times 0.98=98.$$

练习A

1. 掷一个均匀的骰子，设出现的点数为 X，求 X 的数学期望与方差.
2. 一台机器生产某种产品，如果生产一件甲等品可获利 50 元，生产一件乙等品可获利 30 元，生产一件次品会亏损 20 元，已知这台机器生产甲等品、乙等品和次品的概率分别为 0.6，0.3 和 0.1，求这台机器每生产一件产品的平均预期收入.
3. 一批产品的二等品率为 0.02，从这批产品中每次随机取一件，有放回地抽取 100 次，用 X 表示抽到的二等品件数，求 $E(X)$，$D(X)$.
4. 从 8 名男生和 6 名女生中任选 5 人去阳光敬老院参加志愿服务，用 X 表示所选 5 人中女生的人数，求 $E(X)$.
5. 医学上发现，某种病毒侵入人体后，人的体温会升高. 记病毒侵入后人体的平均体温为 X ℃（摄氏度）. 医学统计发现，X 的分布列如下.

X	37	38	39	40
P	0.1	0.5	0.3	0.1

(1) 求出 $E(X)$，$D(X)$；

(2) 已知人体体温为 X ℃时，相当于 $Y=1.8X+32$ ℉（华氏度），求 $E(Y)$，$D(Y)$.

练习B

1. 已知随机变量 X 服从参数为 n，p 的二项分布，即 $X\sim B(n,p)$，且 $E(X)=7$，$D(X)=6$，求 p 的值.
2. 篮球运动员在比赛中，每次罚球命中得 1 分，不命中得 0 分. 已知某运动员罚球命中的概率为 0.7，求：

(1) 他罚球 1 次的得分 ξ 的数学期望；

(2) 他罚球 2 次的得分 η 的数学期望.

3. 若离散型随机变量 X 的概率分布是 $P(X=x_k)=p_k$，其中 $k=1,2,\cdots,n$，求证：

$$\sum_{i=1}^{n}[x_i-E(X)]^2p_i=\sum_{i=1}^{n}x_i^2p_i-[E(X)]^2.$$

4 已知 5 只动物中有 1 只患有某种疾病，需要通过血液化验来确定患病的动物，血液化验结果呈阳性的为患病动物. 下面是两种化验方案：

方案甲：将各动物的血液逐个化验，直到查出患病动物为止.

方案乙：先取 3 只动物的血液进行混合，然后检查，若呈阳性，对这 3 只动物的血液再逐个化验，直到查出患病动物；若不呈阳性，则检查剩下的 2 只动物中 1 只动物的血液.

分析哪种化验方案更好.

1 $1\times p+0\times(1-p)=p$　2 p　3 $1-0.590\ 49=0.409\ 51$

4 $100E(X)\approx 304.76$　5 $50\times 0.02\times(1-0.02)=0.98$

4.2.5 正态分布

1. 二项分布与正态曲线

尝试与发现

已知 X 服从参数为 100，0.5 的二项分布，即 $X\sim B(100,\ 0.5)$，你能手工计算出 $P(X=50)$ 的值吗？

因为

$$P(X=50)=C_{100}^{50}0.5^{50}(1-0.5)^{50}=C_{100}^{50}0.5^{100},$$

所以，如果要手工计算 $P(X=50)$，是一个“几乎不可能”完成的任务，即使是一般的计算器也难以胜任类似的计算. 事实上，利用计算机软件可知

$$C_{100}^{50}\approx 1\times 10^{29},\ 0.5^{100}\approx 7.9\times 10^{-31},$$

因此 $C_{100}^{50}0.5^{100}\approx 0.079$.

由此可以看出，如果随机变量 $X\sim B(n,\ p)$，那么 n 较大时，直接计算

$$P(X=k)=C_n^k p^k(1-p)^{n-k}$$

将是十分困难的. 有没有其他办法能得到上式的近似值呢？这正是 18 世纪 30 年代数学家棣莫弗所研究过的问题. 在讨论这个问题的过程中，棣莫弗发

现了本小节我们要学习的正态曲线.

我们已经知道，服从二项分布的随机变量，其分布列可以用图直观地表示出来. 例如，若 $X\sim B\left(6,\frac{1}{2}\right)$，则 X 的分布列如下.

X	0	1	2	3	4	5	6
P	$\frac{1}{64}$	$\frac{3}{32}$	$\frac{15}{64}$	$\frac{5}{16}$	$\frac{15}{64}$	$\frac{3}{32}$	$\frac{1}{64}$

X 的分布列可以用图 4-2-12 直观地表示出来，其中每一个矩形的宽为 1，高为对应的概率值. 此时，图 4-2-12 具有以下性质：

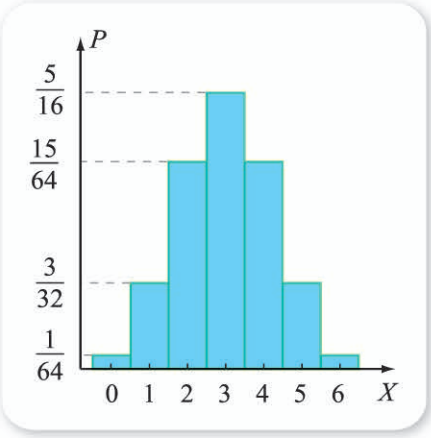


图 4-2-12

（1）中间高、两边低；

（2）图形关于直线 $X=3$ 对称，而且 $E(X)=$ 1 ；

（3）某一整数 k 上方的矩形面积正好等于 $P(X=k)$，其中，$k=0$，1，2，3，4，5，6；

（4）所有矩形的面积之和为 1.

事实上，很多服从二项分布的随机变量分布列的直观图都具有类似的特点. 例如，若 $X\sim B\left(50,\frac{1}{2}\right)$，则其分布列可用图 4-2-13（1）表示；若 $X\sim B\left(100,\frac{1}{2}\right)$，则其分布列可用图 4-2-13（2）表示.

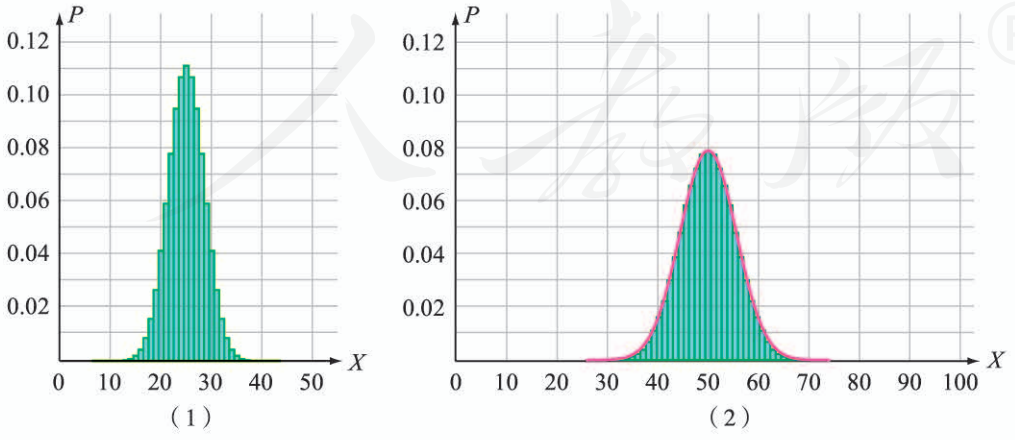


图 4-2-13

由图 4-2-12 与图 4-2-13 可以看出，当 n 充分大时，$X\sim B(n,p)$ 的直

观表示总是具有中间高、两边低的“钟形”. 而且，对不同的参数，只是钟形的宽度和高度不一样而已. 那么，是否存在一个函数 $\varphi(x)$，它对应的图象能够近似这些钟形（如图 4-2-13（2）所示）呢？如果这样的函数存在的话，要计算 X 落在某区间内的概率，只需计算对应曲线与 x 轴在适当区间所围成的面积即可.

事实上，这样的函数 $\varphi(x)$是存在的，这也正是棣莫弗的发现，具体地，

$$\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$\varphi(x)$的解析式中含有 μ 和 σ 两个参数，其中：$\mu=E(X)$，即 X 的均值；$\sigma=\sqrt{D(X)}$，即 X 的标准差. 一般地，$\varphi(x)$对应的图象称为正态曲线（也因形状而被称为“钟形曲线”，$\varphi(x)$也常常记为 $\varphi_{\mu,\sigma}(x)$）.

由此可看出正态曲线的一些性质：

（1）正态曲线关于 $x=\mu$ 对称（即 μ 决定正态曲线对称轴的位置），具有中间高、两边低的特点；

（2）正态曲线与 x 轴所围成的图形面积为 1；

（3）σ 决定正态曲线的“胖瘦”：σ 越大，说明标准差越大，数据的集中程度越弱，所以曲线越“胖”；σ 越小，说明标准差越小，数据的集中程度越强，所以曲线越“瘦”.

更进一步，利用计算机软件可计算出，正态曲线与 x 轴在区间$[\mu,\ \mu+\sigma]$内所围面积约为 0.341 3，在区间$[\mu+\sigma,\ \mu+2\sigma]$内所围面积约为 0.135 9，在区间$[\mu+2\sigma,\ \mu+3\sigma]$内所围面积约为 0.021 5，如图 4-2-14 所示.

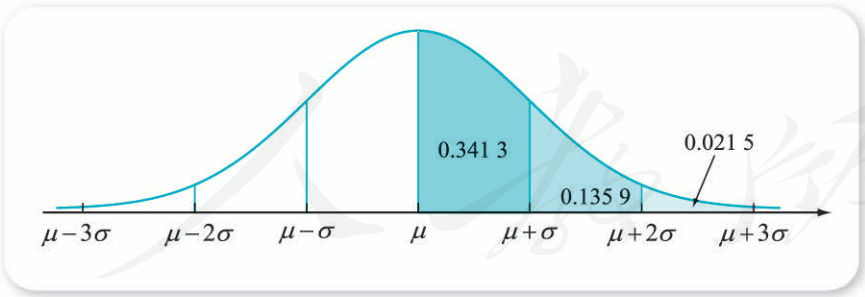


图 4-2-14

例 1 求正态曲线与 x 轴在下列区间内所围的面积（精确到 0.001）：

（1）$[\mu,\ +\infty)$；　　（2）$[\mu-\sigma,\ \mu+\sigma]$；

（3）$[\mu-2\sigma,\ \mu+2\sigma]$；　　（4）$[\mu-3\sigma,\ \mu+3\sigma]$.

解 （1）因为正态曲线是关于 $x=\mu$ 对称的，而且正态曲线与 x 轴所围成的图形面积为 1，因此所求面积为 0.5.

（2）利用对称性可知，所求面积为区间$[\mu,\ \mu+\sigma]$内面积的 2 倍，即约为

$$0.341\ 3\times2=0.682\ 6\approx0.683.$$

（3）利用对称性可知，所求面积约为

$$(0.341\ 3+0.135\ 9)\times2=0.954\ 4\approx0.954.$$

（4）利用对称性可知，所求面积约为

2 _____.

2. 正态分布

正态曲线被发现后并没有立刻得到人们的重视，这一情况直到 19 世纪初，拉普拉斯和高斯开始利用它来研究“随机误差”时才有所改善. 高斯发现，如果设 X 为测量时的误差，那么 $a\leqslant X\leqslant b$ 的概率等于 $\varphi_{\mu,\sigma}(x)$ 的图象与 x 轴在区间$[a,\ b]$内围成的面积.

一般地，如果随机变量 X 落在区间$[a,\ b]$内的概率，总是等于 $\varphi_{\mu,\sigma}(x)$ 对应的正态曲线与 x 轴在区间$[a,\ b]$内围成的面积，则称 X 服从参数为 μ 与 σ 的**正态分布**，记作

$$X\sim N(\mu,\ \sigma^2),$$

此时 $\varphi_{\mu,\sigma}(x)$称为 X 的概率密度函数. 更进一步的研究表明，此时 μ 是 X 的均值，而 σ 是 X 的标准差，σ^2 是 X 的方差.

例如，当 $X\sim N(3,\ 2)$时，X 的均值是 3，方差是 2，而标准差为$\sqrt{2}$.

由正态曲线的性质及例 1 不难得出，如果 $X\sim N(\mu,\ \sigma^2)$，那么

$$P(X\leqslant\mu)=P(X\geqslant\mu)=\text{3}\_\_\_\_\_,$$

$$P(|X-\mu|\leqslant\sigma)=P(\mu-\sigma\leqslant X\leqslant\mu+\sigma)\approx68.3\%,$$

$$P(|X-\mu|\leqslant2\sigma)=P(\mu-2\sigma\leqslant X\leqslant\mu+2\sigma)\approx95.4\%,$$

$$P(|X-\mu|\leqslant3\sigma)=P(\mu-3\sigma\leqslant X\leqslant\mu+3\sigma)\approx99.7\%.$$

最后的式子意味着，X 约有 99.7%的可能会落在距均值 3 个标准差的范围之内，也就是说只有约4 _____的可能会落入这一范围之外（这样的事件可看成小概率事件），这一结论通常称为正态分布的“3σ 原则”.

现实生活中，很多随机变量都服从或近似地服从正态分布，例如随机误差、同一地区同龄人的身高、正常条件下生产出来的产品尺寸等. 也正因为如此，正态分布在概率统计中有着广泛的应用.

例 2 假设某个地区高二学生的身高服从正态分布，且均值为 170（单位：cm，下同），标准差为 10. 在该地区任意抽取一名高二学生，求这名学生的身高：

（1）不高于 170 的概率；

（2）在区间[160，180]内的概率；

（3）不高于 180 的概率.

解 设该学生的身高为 X，由题意可知 $X\sim$ 5 .

（1）易知 $P(X\leqslant 170)=50\%$.

（2）因为均值为 170，标准差为 10，而 $160=170-10$，$180=170+10$，所以

$$P(160\leqslant X\leqslant 180)=P(|X-170|\leqslant 10)\approx 68.3\%.$$

（3）由概率的加法公式可知

$$P(X\leqslant 180)=P(X<170)+P(170\leqslant X\leqslant 180).$$

又由（2）以及正态曲线的对称性可知

$$P(170\leqslant X\leqslant 180)=\frac{1}{2}P(160\leqslant X\leqslant 180)\approx\frac{1}{2}\times 68.3\%=34.15\%,$$

因此

$$P(X\leqslant 180)=P(X<170)+P(170\leqslant X\leqslant 180)$$

$$\approx$$ 6 .

例 3 假设某厂包装食盐的生产线，正常情况下生产出来的食盐质量服从正态分布 $N(500，5^2)$（单位：g），该生产线上的检测员某天随机抽取了两包食盐，称得其质量均大于 515 g.

（1）求正常情况下，任意抽取一包食盐，质量大于 515 g 的概率为多少；

（2）检测员根据抽检结果，判断出该生产线出现异常，要求立即停产检修，检测员的判断是否合理？请说明理由.

解 设正常情况下，该生产线上包装出来的食盐质量为 X g，由题意可知 $X\sim N(500，5^2)$.

（1）由于 $515=500+3\times 5$，所以根据正态分布的对称性与“3σ 原则”可知

$$P(X>515)=\frac{1}{2}P(|X-500|>3\times 5)\approx\frac{1}{2}\times 0.3\%=0.15\%.$$

（2）检测员的判断是合理的. 因为如果生产线不出现异常，由（1）可知，随机抽取两包检查，质量都大于 515 g 的概率约为

$$0.15\%\times 0.15\%=2.25\times 10^{-6},$$

几乎为零，但这样的事件竟然发生了，所以有理由认为生产线出现了异常，检测员的判断是合理的.

实际生活中，人们常常根据类似例 3 的过程来应用正态分布的知识.

3. 标准正态分布

$\mu=0$ 且 $\sigma=1$ 的正态分布称为**标准正态分布**，其在正态分布中扮演着核心角色，这是因为如果 $Y\sim N(\mu,\sigma^2)$，那么令 $X=\dfrac{Y-\mu}{\sigma}$，则可以证明 $X\sim N(0,1)$，即任意正态分布通过变换都可化为标准正态分布.

如果 $X\sim N(0,1)$，那么对于任意 a，通常记

$$\Phi(a)=P(X<a),$$

也就是说 $\Phi(a)$表示 $N(0,1)$对应的正态曲线与 x 轴在区间$(-\infty,a)$内所围的面积，如图 4-2-15 所示.

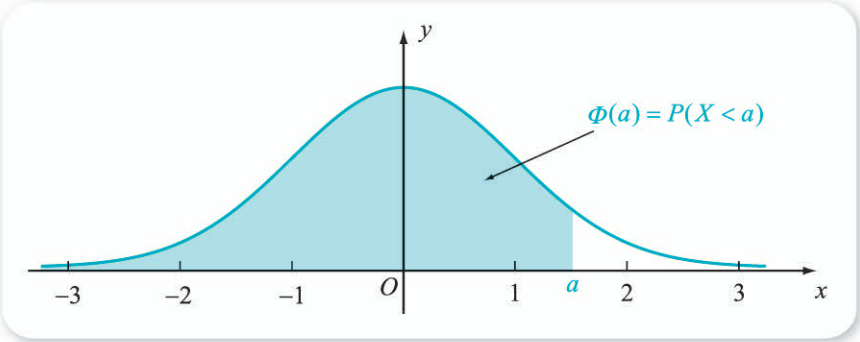


图 4-2-15

根据正态曲线的对称性，可以知道 $\Phi(a)$具有性质

$$\Phi(-a)+\Phi(a)=1,$$

由前面的知识可以得到 $\Phi(k)$ $(k=-3,-2,-1,0,1,2,3)$ 的值，例如，

$$\Phi(1)\approx 0.5+0.341\ 3=0.841\ 3.$$

为了方便起见，人们将 $a\geqslant 0$ 时部分 $\Phi(a)$的值制成了专门的表格，可供查询，下表是部分数据.

$\Phi(a)=P(X<a)$										
a	0	1	2	3	4	5	6	7	8	9
0.0	.500 0	.504 0	.508 0	.512 0	.516 0	.519 9	.523 9	.527 9	.531 9	.535 9
0.1	.539 8	.543 8	.547 8	.551 7	.555 7	.559 6	.563 6	.567 5	.571 4	.575 3
0.2	.579 3	.583 2	.587 1	.591 0	.594 8	.598 7	.602 6	.606 4	.610 3	.614 1
0.3	.617 9	.621 7	.625 5	.629 3	.633 1	.636 8	.640 6	.644 3	.648 0	.651 7
0.4	.655 4	.659 1	.662 8	.666 4	.670 0	.673 6	.677 2	.680 8	.684 4	.687 9
0.5	.691 5	.695 0	.698 5	.701 9	.705 4	.708 8	.712 3	.715 7	.719 0	.722 4

例如，从上表中可以查出，$\Phi(0.16)=0.563\ 6$，$\Phi(0.58)=0.719\ 0$.

例 4 已知 $X\sim N(0,1)$，利用上述表格求以下概率值：

(1) $P(X<0.28)$；　　(2) $P(X<-0.36)$；

(3) $P(0.18\leqslant X<0.57)$.

解 (1) $P(X<0.28)=\Phi(0.28)=0.610\ 3$.

(2) 因为 $P(X<-0.36)=\Phi(-0.36)$，而 $\Phi(-0.36)+\Phi(0.36)=1$，且由上表可知 $\Phi(0.36)=0.640\ 6$，所以

$$\Phi(-0.36)=1-\Phi(0.36)=1-0.640\ 6=0.359\ 4.$$

(3) 由概率的加法公式以及上表可知

$$\begin{aligned}P(0.18\leqslant X<0.57)&=P(X<0.57)-P(X<0.18)\\&=\Phi(0.57)-\Phi(0.18)\\&=0.715\ 7-0.571\ 4\\&=0.144\ 3.\end{aligned}$$

最后，我们来看用正态分布近似二项分布的一个实例.

假设 $X\sim B(100,0.5)$，那么 $E(X)=50$，$D(X)=$ 7 ，用正态分布近似二项分布的话有 $X\sim N(50,5^2)$，那么

$$P(X=50)\approx P(49.5\leqslant X<50.5)=P\left(-0.1\leqslant\frac{X-50}{5}<0.1\right).$$

又因为 $\dfrac{X-50}{5}\sim N(0,1)$，所以

$$\begin{aligned}P\left(-0.1\leqslant\frac{X-50}{5}<0.1\right)&=\Phi(0.1)-\Phi(-0.1)=\Phi(0.1)-[1-\Phi(0.1)]\\&=2\Phi(0.1)-1=2\times0.539\ 8-1=0.079\ 6.\end{aligned}$$

这与我们前面通过直接计算得到的 $P(X=50)\approx0.079$ 相差无几.

4. 用信息技术求正态分布的概率值

利用信息技术，可以方便地求出与正态分布有关的概率值.

例如，如果 $X\sim N(\mu,\sigma^2)$，那么 $P(X<x)$ 的值可以利用 Excel 中的函数 NORMDIST 方便地求得. 事实上，打开 Excel 表格，在任意一个单元格输入“=NORMDIST(0.28，0，1，1)”即可得到例 4 (1) 的近似值，如图 4-2-16 所示.

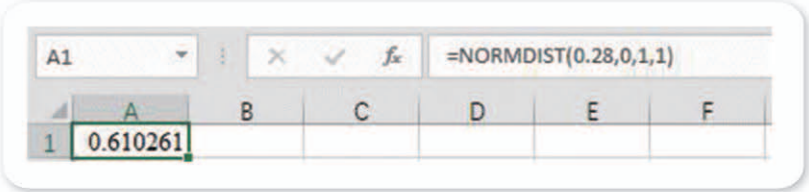


图 4-2-16

练习A

1. 已知随机变量 ξ 服从正态分布 $N(3, \sigma^2)$，求 $P(\xi<3)$.

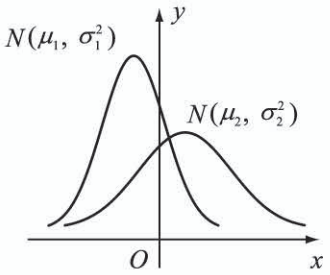
2. 设两个正态分布 $N(\mu_1, \sigma_1^2)$ $(\sigma_1>0)$ 和 $N(\mu_2, \sigma_2^2)$ $(\sigma_2>0)$ 的密度函数图象如图所示，则（　　）.

(A) $\mu_1<\mu_2$，$\sigma_1<\sigma_2$

(B) $\mu_1<\mu_2$，$\sigma_1>\sigma_2$

(C) $\mu_1>\mu_2$，$\sigma_1<\sigma_2$

(D) $\mu_1>\mu_2$，$\sigma_1>\sigma_2$



（第 2 题）

3. 若 $X\sim N(\mu, \sigma^2)$，根据

$$P(\mu\leqslant X\leqslant\mu+\sigma)=0.341\ 3,$$

$$P(\mu+\sigma\leqslant X\leqslant\mu+2\sigma)=0.135\ 9,$$

写出下列各概率值：

(1) $P(\mu-\sigma\leqslant X\leqslant\mu)$；　　(2) $P(\mu-2\sigma\leqslant X\leqslant\mu)$.

4. 设随机变量 ξ 服从标准正态分布 $N(0, 1)$，已知 $\Phi(-1.96)=0.025$，求 $P(|\xi|<1.96)$.

5. 利用 GeoGebra 分别作出 $X\sim B(50, \frac{1}{3})$，$X\sim B(100, \frac{1}{4})$，$X\sim B(500, \frac{1}{6})$ 时分布列的直观图，观察所得图象是否对称等.

练习B

1. 设随机变量 ξ 服从正态分布 $N(2, 9)$，若 $P(\xi>c+1)=P(\xi<c-1)$，求 c 的值.

2. 已知随机变量 X 服从正态分布 $N(3, 1)$，且 $P(2\leqslant X\leqslant 4)=0.682\ 6$，求 $P(X>4)$.

3. 已知随机变量 ξ 服从正态分布 $N(0, \sigma^2)$，若 $P(\xi>2)=0.023$，求 $P(-2\leqslant\xi\leqslant 2)$.

4. 已知随机变量 ξ 服从正态分布 $N(2, \sigma^2)$，且 $P(\xi<4)=0.8$，求 $P(0<\xi<2)$.

5. 一商场经营的某种包装的大米质量 X（单位：kg）服从正态分布 $N(10, \sigma^2)$，且 $P(X<10.5)=0.8$. 从该商场中任意抽取一袋该种大米，求其质量在 9.5～10.5 kg 之间的概率.

1 $6\times\frac{1}{2}=3$　　2 $(0.341\ 3+0.135\ 9+0.021\ 5)\times 2=0.997\ 4\approx 0.997$　　3 50%

4 0.3%　　5 $N(170, 10^2)$　　6 $50\%+34.15\%=84.15\%$　　7 25

习题4-2A

1. 写出下列随机变量的取值范围：

   (1) 某网页一年内的点击数；

   (2) 一个袋子里装有5个白球和5个黑球，从中任取3个，其中所含白球的个数；

   (3) 某地铁站台每间隔3 min有1辆列车通过，某人在该站台等车的时间；

   (4) 掷两个骰子，所得点数之和.

2. 已知X与Y都是随机变量，且$Y=2X-1$，$P(X=3)=0.2$，求$P(Y=5)$的值.

3. 已知$P(X=1)=0.3$，求$P(X\neq 1)$的值.

4. 假设每次测量中，出现正误差与负误差的概率都是0.5，设3次测量中，出现正误差的次数为X，写出X的分布列.

5. 甲、乙两人进行象棋比赛时，每一局甲赢的概率是0.51，乙赢的概率是0.49. 设甲、乙一共进行了10次比赛，且各次比赛的结果相互独立. 求甲赢的局数的期望.

6. 某班有35个学生，假设每个学生早上到校的时间相互没有影响，并且每个人迟到的概率均为0.05，设下周二这个班迟到的人数为X，指出X服从的分布列，并求$P(X=3)$.

7. 根据气象预报，某地区近期有小洪水的概率为0.25，有大洪水的概率为0.01. 该地区某工地上有一台大型设备，遇到大洪水时要损失60 000元，遇到小洪水时要损失10 000元. 为保护设备，有以下3种方案：

   方案1：运走设备，搬运费为3 800元.

   方案2：建保护围墙，建设费为2 000元，但围墙只能防小洪水.

   方案3：不采取措施，希望不发生洪水.

   如果你是工地的负责人，你会采用哪种方案？说明理由.

习题4-2B

1. 掷两个均匀的骰子，记两个骰子的点数差的绝对值为ξ.

   (1) 写出ξ的取值范围；　　(2) 求$P(\xi=0)$的值.

2. 某次科技知识竞赛中，需回答20个问题，计分规则是：每答对一题得5分，答错一题扣3分. 从参加这次科技知识竞赛的学生中任意抽取一名，设其答对的题数为X，最后得分为Y分.

   (1) 当$X=10$时，求Y的值；

   (2) 写出X与Y之间的关系式；

   (3) 若$P(X\leqslant 15)=0.3$，求$P(Y>60)$的值.

3. 重复掷一个均匀骰子 3 次，得到点数为 6 的次数记为 ξ，求 $P(\xi>2)$.

4. 某射击运动员在一次射击训练中，共有 5 发子弹，如果命中就停止射击，否则继续射击，直到命中或子弹用尽. 若已知每次射击命中的概率均为 0.9，求该运动员这次训练耗用的子弹数 X 的分布列.

5. 已知某类种子每粒发芽的概率都为 0.9，现播种了 1 000 粒，对于没有发芽的种子，每粒需再补种 2 粒，补种的种子数记为 X，求 $E(X)$，$D(X)$.

6. 已知随机变量 ξ 服从正态分布 $N(2,\sigma^2)$，$P(\xi\leqslant 4)=0.84$，求 $P(\xi\leqslant 0)$.

7. 在某项测量中，测量结果 ξ 服从正态分布 $N(1,\sigma^2)(\sigma>0)$. 若 ξ 在 $(0,1)$ 内取值的概率为 0.4，求 ξ 在 $(0,2)$ 内取值的概率.

4.3 统计模型

4.3.1 一元线性回归模型

情境与问题

以下是几则与“相关系数”有关的新闻报道：

(1)“1999—2008 年，俄罗斯 GDP 增长率与国际石油价格的相关系数为 0.86，2009—2014 年该系数达到 0.98.”(《人民日报》2017 年 8 月 9 日)

(2)“瑞士洛桑国际管理学院对企业国际竞争力的研究也显示，公司文化与企业管理竞争力的相关系数在几个因子中是最高的.”(《中国青年报》2009 年 9 月 11 日)

(3)“分析表明 1990 年至 2011 年我国财政收入与企业注册资本之间的关系呈高度线性相关，其相关系数高达 0.987，而斜率竟为 0.148.”(《人民日报》2014 年 5 月 21 日)

你能猜出相关系数的含义吗？(3) 中“斜率”表示的是什么？

学完本节的内容后，你就会对相关系数有一个比较完整的了解.

1. 相关关系

尝试与发现

自行选择标准，将下列变量之间的关系分为两类，并分别阐述每一类中变量关系的特点：

(1) 圆的面积 S 与半径 r 之间的关系；

(2) 16 岁学生的体重 w 与身高 h 之间的关系；

(3) 商品销售量 Q 与销售价格 P 之间的关系；

(4) 匀速运动的物体，其运动的路程 S 与时间 t 之间的关系；

(5) 平均学习时间 t 与学习成绩 f 之间的关系；

(6) 科技创新能力 y 与人才培养近亲繁殖率 x 之间的关系.

我们研究的很多问题中，两个变量之间经常存在着相互影响、相互依赖的关系. 这些关系常见的有两类，一类是变量之间的关系具有确定性，当一个变量确定后，另一个变量就确定了，尝试与发现中（1）（4）所描述的就是这一类的代表；另一类是变量之间确实有一定的关系，但没有达到可以互相决定的程度，它们之间的关系带有一定的随机性，尝试与发现中（2）（3）（5）（6）所描述的关系都是如此.

例如，一般情况下，已知一名 16 岁学生的身高 h，不能确定其体重 w，但身高越高的人体重可能越重，不过同样身高的人体重往往存在差异；商品的销售价格 P 越低，买这种商品的人可能会越多，从而会导致销售量 Q 增长，但同样的销售价格可能会有不同的销售量；平均学习时间 t 越长，学习成绩 f 可能越好，但学习时间相同不能确保学习成绩相同；人才培养近亲繁殖率 x 越大，科技创新能力 y 可能越弱，但繁殖率确定时，创新能力并不是确定的. 这些两个变量之间的关系，统计学上都称为**相关关系**.

怎样从统计数据中确定与描述两个变量之间的相关关系呢？下面我们借助一个实例来进行说明.

已知某班级学生数学成绩与物理成绩的对应表如下.

数学	88	82	79	89	92	78	79	56	83	66	61	74	68	77	78	80
物理	87	91	66	86	88	79	91	53	93	77	62	69	62	58	79	68
数学	51	58	98	73	75	84	69	76	71	43	74	82	70	75	65	63
物理	50	65	82	79	59	98	59	87	76	42	75	91	46	68	83	55

从上表中，想直接观察出数学成绩与物理成绩之间是否存在相关关系等，是有一定困难的. 为此，我们可以对数据进行整理，比如在保持成绩配对的方式不变的前提下，将数据按照数学成绩从小到大的顺序排列，如下表所示.

数学	43	51	56	58	61	63	65	66	68	69	70	71	73	74	74	75
物理	42	50	53	65	62	55	83	77	62	59	46	76	79	69	75	59
数学	75	76	77	78	78	79	79	80	82	82	83	84	88	89	92	98
物理	68	87	58	79	79	66	91	68	91	91	93	98	87	86	88	82

从表中可以看出，当数学成绩增加时，物理成绩大致也是增加的. 我们还可以根据给定数据作出图象来进行直观判断. 例如，以数学成绩为横坐标，物理成绩为纵坐标，则可以在平面直角坐标系中将每一对数据都表示出来，如图 4-3-1 所示. 这样作出来的图称为散点图，从散点图上也可得出数学成绩增加时物理成绩大致也增加的结论. 在图 4-3-1 中可以作出直线，使

得该班级学生数学成绩与物理成绩构成的点分布在直线上或直线的两侧，且都在直线附近. 这就提示我们，数学成绩与物理成绩之间是有相关关系的，而且在容许一定误差的前提下，可以用一次函数来描述它们之间的关系.

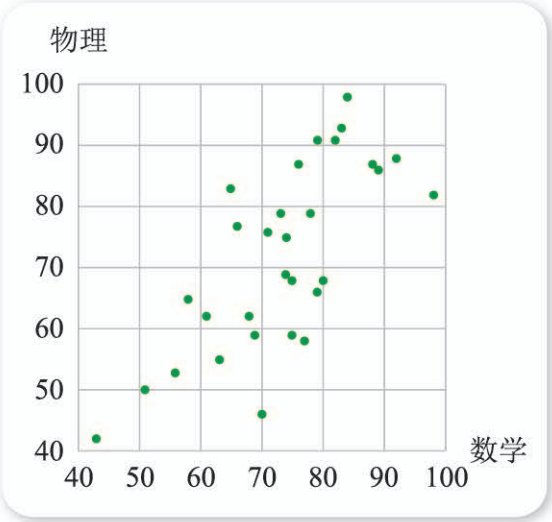


图 4-3-1

一般地，如果收集到了变量 x 和变量 y 的 n 对数据（简称为成对数据），如下表所示.

序号 i	1	2	3	…	n
变量 x	x_1	x_2	x_3	…	x_n
变量 y	y_1	y_2	y_3	…	y_n

则在平面直角坐标系 xOy 中描出点 (x_i, y_i)，$i=1, 2, 3, \cdots, n$，就可以得到这 n 对数据的**散点图**. 如果由变量的成对数据、散点图或直观经验可知，变量 x 与变量 y 之间的关系可以近似地用一次函数来刻画，则称 x 与 y **线性相关**. 此时，如果一个变量增大，另一个变量大体上也增大，则称这两个变量**正相关**；如果一个变量增大，另一个变量大体上减少，则称这两个变量**负相关**.

前面所提到的班级学生的数学成绩与物理成绩是线性相关的，而且是正相关. 另外，如果将尝试与发现的（2）（3）（5）（6）中变量之间的关系看成线性相关，则 16 岁学生的体重 w 与身高 h、平均学习时间 t 与学习成绩 f 都是1_____相关的，而商品销售量 Q 与销售价格 P、科技创新能力 y 与人才培养近亲繁殖率 x 都是2_____相关的.

正相关、负相关在日常生活中也经常使用. 例如，“反腐败、转作风，事关民心所向，事关党的生死存亡. 历史和实践表明，反腐的力度与民心的聚合度正相关”，这就说明，反腐的力度越大，民心的聚合度就越有可能更强；“网络谣言的危害性与公众判断能力负相关”已经成为一项共识，这就

意味着，公众媒介素养越高，网络谣言的危害性越有可能更小.

2. 回归直线方程

我们已经知道，线性相关的两个变量之间的关系可以用一次函数来近似刻画，那么怎样找出对应函数的表达式呢？这就是接下来要讨论的问题.

尝试与发现

某地区从某一年开始进行了环境污染整治，得到了如下数据：

第 x 年	1	2	3	4	5	6	7
污染指数 y	6.1	5.2	4.5	4.7	3.8	3.4	3.1

(1) 作出这些成对数据的散点图，直观地判断污染指数 y 与 x 是否线性相关. 如果是，进一步判断是正相关还是负相关.

(2) 在知道 y 与 x 线性相关的前提下，你能找出近似描述 y 与 x 之间关系的一次函数表达式吗？根据所得到的关系式，你能估计出该地区第 8 年的污染指数吗？

根据尝试与发现中的数据，可作出散点图如图 4-3-2 所示. 可以看出，y 与 x 之间的关系可近似地用一次函数表示，而且随着时间 x 的增加，污染指数 y 大致是减少的，因此 y 与 x 线性相关，而且是负相关的.

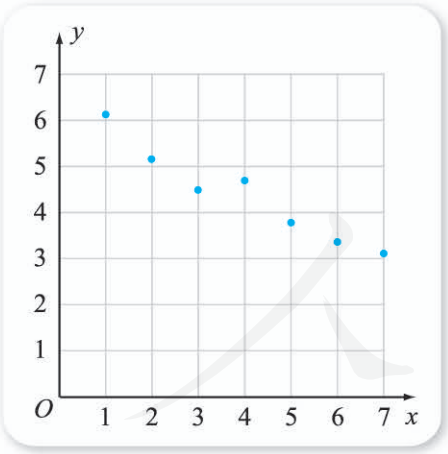


图 4-3-2

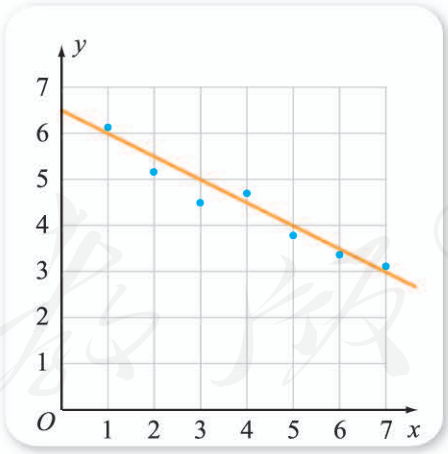


图 4-3-3

为了找出近似描述 y 与 x 之间关系的一次函数表达式，我们可先在图 4-3-2 中作出一条直线，使得成对数据构成的点分布在直线的附近. 例如，通过点(1，6)和(7，3)的直线就满足条件，如图 4-3-3 所示. 根据已知的两点就可以得出所要的函数关系式

$$y=-0.5x+6.5.$$

更进一步，代入 $x=8$，就能得到第 8 年污染指数的估计值 $y=2.5$.

当然，类似的直线我们可以找出很多条（从而也可以得到很多个一次函数关系式），现在这一条是不是“最好”的直线呢？用什么样的标准来衡量好还是不好呢？

注意到函数表达式 $y=-0.5x+6.5$ 确定之后，我们不仅可以算出 $x=8$ 的值，而且还可以算出 $x=1$，2，3，…，7 的值，也可以得到已知数据的实际值（也称为观测值）与预测值之间的误差（一般称为残差），如下表所示.

第 x 年	1	2	3	4	5	6	7
污染指数 y	6.1	5.2	4.5	4.7	3.8	3.4	3.1
预测值 $-0.5x+6.5$	6	5.5	5	4.5	4	3.5	3
误差	0.1	−0.3	−0.5	0.2	−0.2	−0.1	0.1

这也可以用图 4-3-4 来表示，图中橙色的点就是预测值对应的点，误差的绝对值就是蓝色的点与相应的橙色的点之间的距离.

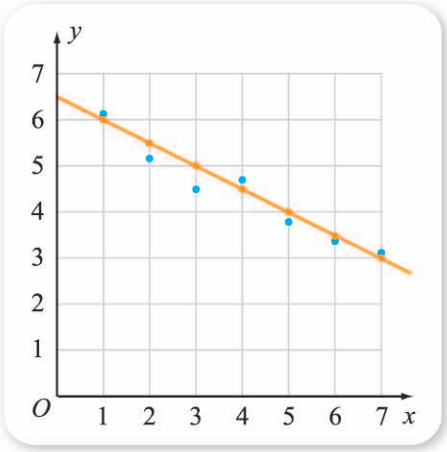


图 4-3-4

统计学意义上“最好”的直线，指的是所有误差平方和最小的直线. 可以证明，对于上述污染指数与时间的数据，误差平方和最小的直线为

$$\hat{y}=-0.475x+6.3,$$

这称为 y 关于 x 的回归直线方程，其中 $\hat{y}$ 读作“y 估”，表示 y 的估计值. 根据这个方程，可以得到第 8 年的污染指数估计值为

$$-0.475\times 8+6.3=2.5.$$

一般地，已知变量 x 与 y 的 n 对成对数据 (x_i, y_i)，$i=1$，2，3，…，n. 任意给定一个一次函数 $y=bx+a$，对每一个已知的 x_i，由直线方程可以得到一个估计值

$$\hat{y}_i=bx_i+a,$$

如果一次函数 $\hat{y}=\hat{b}x+\hat{a}$ 能使残差平方和即

$$(y_1-\hat{y}_1)^2+(y_2-\hat{y}_2)^2+\cdots+(y_n-\hat{y}_n)^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

取得最小值，则

$$\hat{y}=\hat{b}x+\hat{a}$$

称为 y 关于 x 的**回归直线方程**（对应的直线称为**回归直线**）. 因为是使得平

方和最小，所以其中涉及的方法称为**最小二乘法**.

可以证明，给定两个变量 y 与 x 的一组数据之后，回归直线方程 $\hat{y}=\hat{b}x+\hat{a}$ 总是存在的，而且

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum\limits_{i=1}^{n}(x_i-\overline{x})^2}=\frac{\sum\limits_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum\limits_{i=1}^{n}x_i^2-n\overline{x}^2},$$

$$\hat{a}=\overline{y}-\hat{b}\overline{x}.$$

其中，$\hat{b}$ 称为**回归系数**. 它实际上也就是回归直线方程的斜率. 回归直线方程确定之后，就可用于预测.

需要注意的是，上述公式中，$\overline{x}$ 指的是 x_1，x_2，x_3，…，x_n 的平均数，即

$$\overline{x}=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i;$$

类似地，$\overline{y}$ 是 y_1，y_2，y_3，…，y_n 的平均数，即 $\overline{y}=\frac{1}{n}\sum\limits_{i=1}^{n}y_i$. 另外，由计算公式可以看出，回归系数 $\hat{b}$ 的计算并不容易，实际计算过程中，我们可以通过列表的方法逐步进行计算.

例如，对于上述尝试与发现中的数据来说，可以首先算得 $\overline{x}=4$，$\overline{y}=4.4$，接着列表如下.

x	1	2	3	4	5	6	7
y	6.1	5.2	4.5	4.7	3.8	3.4	3.1
$x-\overline{x}$	−3	−2	−1	0	1	2	3
$y-\overline{y}$	1.7	0.8	0.1	0.3	−0.6	−1	−1.3
$(x-\overline{x})(y-\overline{y})$	−5.1	−1.6	−0.1	0	−0.6	−2	−3.9
$(x-\overline{x})^2$	9	4	1	0	1	4	9

从而可知

$$\sum_{i=1}^{7}(x_i-\overline{x})(y_i-\overline{y})=-5.1-1.6-0.1-0.6-2-3.9=-13.3,$$

$$\sum_{i=1}^{7}(x_i-\overline{x})^2=9+4+1+1+4+9=28.$$

因此

$$\hat{b}=\frac{-13.3}{28}=-0.475,\ \hat{a}=4.4-(-0.475)\times4=6.3.$$

所以 y 关于 x 的回归直线方程为

$$\hat{y}=-0.475x+6.3.$$

拓展阅读

“回归”一词的由来

《现代汉语词典（第 7 版）》中，“回归”的解释是“回到（原来的地方）”；地理学中，“回归线”是指地球赤道南北各 $23°26'$ 处的纬线，太阳直射点在南回归线与北回归线之间来回移动. 看了这些，你是不是感觉到回归直线方程中的“回归”与上面这些说法相差很大？

统计学中的“回归”一词，是统计学家高尔顿引入的. 早在 19 世纪 80 年代，高尔顿就开始了亲代与子代（即父母亲与子女）之间相似特征（身高、性格等）的研究. 他收集了一些亲代的身高 x 与子代的身高 y 的成对数据，并作出了散点图，发现 y 与 x 的关系可以借助一次函数来近似表示，而且总体上亲代的身高增加时，子代的身高也增加.

但是，高尔顿在研究过程中，发现了一个有趣的现象. 他收集的数据显示，总体上亲代的平均身高为 68 英寸（约为 172.72 cm），子代的平均身高为 69 英寸，子代的平均身高比亲代的平均身高高 1 英寸（约为 2.54 cm）. 于是，一个自然的推想是：平均身高为 63 英寸的亲代，其子代的平均身高应约为 64 英寸；平均身高为 72 英寸的亲代，其子代的平均身高应约为 73 英寸. 但实际数据显示：平均身高为 63 英寸的亲代，其子代的平均身高为 67 英寸，增加量为 4 英寸；平均身高为 72 英寸的亲代，其子代的平均身高为 71 英寸，增加量为 -1 英寸. 也就是说，平均身高不同的亲代，其子代的平均身高增加量并不相等，但子代的平均身高有回归于中心（即总体平均值）的趋势.

正是由于这种现象的存在，高尔顿引入了“回归”一词. 虽然不是所有相关关系中都会发生类似的现象，但从那以后，“回归”就成了相关关系讨论中一个约定俗成的词了.

3. 回归直线方程的性质

尝试与发现

假设 y 与 x 具有相关关系，而且回归直线方程为 $\hat{y}=\hat{b}x+\hat{a}$，完成下列任务：

（1）将 $\hat{a}=\overline{y}-\hat{b}\overline{x}$ 代入回归直线方程，并求出 $x=\overline{x}$ 时 $\hat{y}$ 的值；

（2）判断一次函数 $\hat{y}=\hat{b}x+\hat{a}$ 的单调性由谁决定，指出函数的单调性与正相关、负相关之间的联系；

（3）通过计算说明，当 x 每增大一个单位时，$\hat{y}$ 将如何变化，并总结出这一结论的实际意义.

将 $\hat{a}=\overline{y}-\hat{b}\overline{x}$ 代入 $\hat{y}=\hat{b}x+\hat{a}$ 后，整理可得

$$\hat{y}-\overline{y}=\hat{b}(x-\overline{x}),$$

这说明回归直线一定过点$(\overline{x},\overline{y})$.

一次函数 $\hat{y}=\hat{b}x+\hat{a}$ 的单调性当然是由 $\hat{b}$ 的符号决定的，函数递增的充要条件是 $\hat{b}>0$，这说明：y 与 x 正相关的充要条件是 $\hat{b}>0$；y 与 x 负相关的充要条件是 3 _____.

另外，如果$(x_1,\hat{y}_1)$和$(x_2,\hat{y}_2)$都是回归直线上的点，则

$$\begin{cases}\hat{y}_1=\hat{b}x_1+\hat{a}, & ①\\ \hat{y}_2=\hat{b}x_2+\hat{a}, & ②\end{cases}$$

①式减去②式可得

$$\hat{y}_2-\hat{y}_1=\hat{b}(x_2-x_1),$$

这就说明，若 $x_2-x_1=1$，则 $\hat{y}_2-\hat{y}_1=\hat{b}$. 也就是说，当 x 增大一个单位时，$\hat{y}$ 增大 $\hat{b}$ 个单位，这就是回归系数 $\hat{b}$ 的实际意义.

例 1 如果某位同学 10 次考试的物理成绩 y 与数学成绩 x 如下表所示.

数学成绩 x	76	82	72	87	93	78	89	66	81	76
物理成绩 y	80	87	75	86	100	79	93	68	85	77

已知 y 与 x 线性相关：

(1) 判断是正相关还是负相关；

(2) 求出 y 关于 x 的回归直线方程；

(3) 该同学的数学成绩每提高 3 分，物理成绩估计能提高多少分？

解 (1) 根据数据可作出散点图，如图 4-3-5 所示.

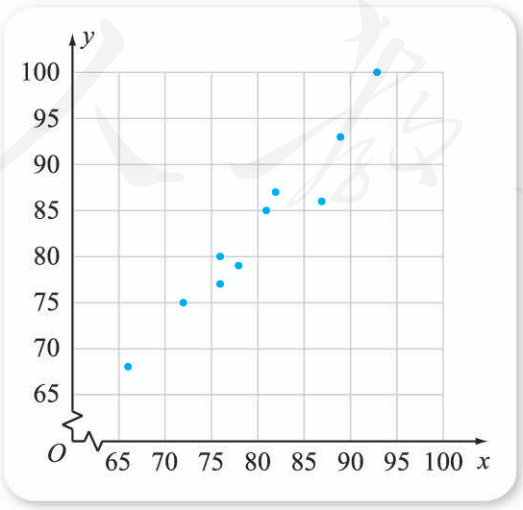


图 4-3-5

从图上可以直观地看出，y 与 x 正相关.

(2) 首先计算 $\overline{x}$ 与 $\overline{y}$.

将每个 x 的值都减去 80 可得

$$-4，2，-8，7，13，-2，9，-14，1，-4，$$

这些数的平均数为 0，因此 $\overline{x}=$ 4 .

将每个 y 的值都减去 80 可得

$$0，7，-5，6，20，-1，13，-12，5，-3，$$

这些数的平均数为 3，因此 $\overline{y}=$ 5 .

通过列表计算可得

$$\begin{aligned}&\sum_{i=1}^{10}(x_i-\overline{x})(y_i-\overline{y})\\=&12+8+64+21+221+8+90+210+2+24\\=&660，\end{aligned}$$

$$\begin{aligned}&\sum_{i=1}^{10}(x_i-\overline{x})^2\\=&16+4+64+49+169+4+81+196+1+16\\=&600.\end{aligned}$$

因此

$$\hat{b}=\frac{660}{600}=1.1，\ \hat{a}=83-1.1\times 80=-5.$$

回归直线方程为 $\hat{y}=1.1x-5$.

(3) 由回归系数 $\hat{b}=1.1$ 可知，x 每增大 1 个单位时，$\hat{y}$ 增大 1.1 个单位. 因此，数学成绩每提高 3 分，物理成绩估计能提高的分值为

$$1.1\times 3=3.3.$$

4. 相关系数

尝试与发现

如下是某班级学生数学成绩与英语成绩的对应表.

数学	43	51	56	58	61	63	65	66	68	69	70	71	73	74	74	75
英语	81	76	67	78	65	73	71	74	76	62	64	77	80	81	68	72
数学	75	76	77	78	78	79	79	80	82	82	83	84	88	89	92	98
英语	85	69	71	70	76	62	89	69	76	84	94	84	79	81	85	68

从这些数据中，你能直接看出该班级学生的数学成绩与英语成绩之间是否存在线性相关关系吗？作出这些数据的散点图，并与图 4-3-1 对比，你能得出什么结论？

根据尝试与发现中的数据，可以作出散点图，如图 4-3-6 所示. 通过与图 4-3-1 对比，直观上可以看出，相对于数学成绩与物理成绩来说，数学成绩与英语成绩之间线性相关关系要弱一些.

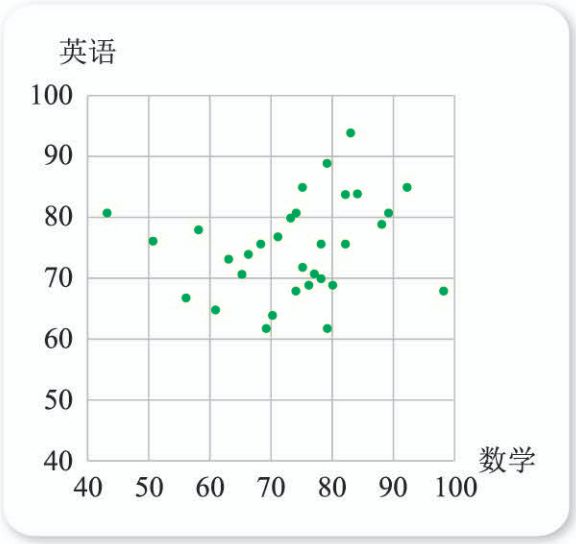


图 4-3-6

由前面可知，给定一组成对数据后，总能得到一个回归直线方程. 不难想到，如果变量之间本身的线性相关关系很弱，那么得到的回归直线方程价值是有限的，甚至是没有价值的.

怎样来刻画两个变量之间线性相关关系的强弱呢？也就是说，给定两个变量 y 与 x 的成对数据 (x_i, y_i)，$i=1, 2, 3, \cdots, n$，我们要寻找一个量来刻画 y 与 x 之间线性相关关系的强弱. 假设由成对数据得到的回归直线方程为 $\hat{y}=\hat{b}x+\hat{a}$，由前面可知，这条直线是通过点 $(\overline{x}, \overline{y})$ 的. 不难想到，我们所要寻找的量，在所有点越靠近回归直线时，特征要越明显.

如图 4-3-7 所示，在平面直角坐标系 xOy 中，作出成对数据的散点图以及回归直线，并且标出点 $(\overline{x}, \overline{y})$. 然后以 $(\overline{x}, \overline{y})$ 为原点，建立新的平面直角坐标系 $x'O'y'$，则回归直线在 $x'O'y'$ 中是过原点的，而且：如果 y 与 x 正相关，则回归直线过 $x'O'y'$ 的一、三象限；如果 y 与 x 负相关，则回归直线过 $x'O'y'$ 的二、四象限.

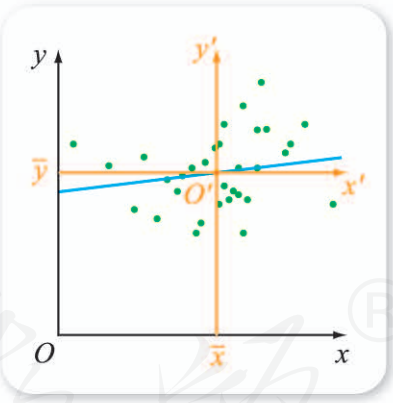


图 4-3-7

因此，从直观上可知，如果 y 与 x 正相关（或负相关），那么成对数据中，在 $x'O'y'$ 的一、三象限（或二、四象限）内的点越多，y 与 x 的线性相关关系可能会越强. 注意到 (x_i, y_i) 在 $x'O'y'$ 中的坐标为 $(x_i-\overline{x}, y_i-\overline{y})$，因此可用 $(x_i-\overline{x})(y_i-\overline{y})$ 来判定成对数据是否落在一、三象限（或二、四象限）. 这样一来，就可用含有

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$$

的量来判定 y 与 x 的线性相关性强弱.

尝试与发现

结合下列两组成对数据，判断能否直接用 $\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})$ 来衡量 y 与 x 的线性相关性强弱？并说明理由.

(1)	x	1	2	3	4	5	6	7
	y	6.1	5.2	4.5	4.7	3.8	3.4	3.1

(2)	x	10	20	30	40	50	60	70
	y	61	52	45	47	38	34	31

注意到现实生活中的数据，由于度量对象和单位的不同等，数值会有大有小，为了去除这些因素的影响，统计学里一般用

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2\sum_{i=1}^{n}(y_i-\overline{y})^2}}=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sqrt{\left(\sum_{i=1}^{n}x_i^2-n\overline{x}^2\right)\left(\sum_{i=1}^{n}y_i^2-n\overline{y}^2\right)}}$$

来衡量 y 与 x 的线性相关性强弱，这里的 r 称为**线性相关系数**（简称为**相关系数**）.

可以证明，相关系数 r 具有以下性质：

(1) $|r|\leqslant 1$，且 y 与 x 正相关的充要条件是 $r>0$，y 与 x 负相关的充要条件是 6 ；

(2) $|r|$ 越小，说明两个变量之间的线性相关性越弱，也就是得出的回归直线方程越没有价值，即方程越不能反映真实的情况；$|r|$ 越大，说明两个变量之间的线性相关性越强，也就是得出的回归直线方程越有价值；

(3) $|r|=1$ 的充要条件是成对数据构成的点都在回归直线上.

利用类似计算回归系数的方法可以算出相关系数的值. 例如，图 4-3-1 中成对数据所对应的相关系数约为 0.73，图 4-3-6 中成对数据所对应的相关系数约为 0.21. 需要注意的是，相关系数的绝对值越大，只能说明两个变量之间的关系用一次函数刻画时，效果越好，但这并不能保证两个变量之间存在因果关系. 例如，数学成绩与物理成绩的相关系数为 0.73，不能说是因为数学成绩好，所以物理成绩好，实际上，数学成绩与物理成绩之间的相关关系可能是由这两个学科的相似性造成的.

日常生活的新闻报道中，经常出现的相关系数指的就是这个意思. 例如，“分析表明 1990 年至 2011 年我国财政收入与企业注册资本之间的关系呈高度线性相关，其相关系数高达 0.987，而斜率竟为 0.148”，其中的

0.987 就是按照上面的公式计算出来的，而斜率 0.148 指的就是回归系数的大小.

例 2 某人工智能公司从某年起 7 年的利润情况如下表所示.

第 x 年	1	2	3	4	5	6	7
利润 y/亿元	2.9	3.3	3.6	4.4	4.8	5.2	5.9

(1) 计算出 y 与 x 之间的相关系数（精确到 0.01），并求出 y 关于 x 的回归直线方程；

(2) 根据回归直线方程，分别预测该人工智能公司第 8 年和第 9 年的利润.

解 (1) 可以算得 $\overline{x}=4$，$\overline{y}=4.3$.

通过列表计算可得

$$\sum_{i=1}^{7}(x_i-\overline{x})(y_i-\overline{y})=14,\quad \sum_{i=1}^{7}(x_i-\overline{x})^2=28,\quad \sum_{i=1}^{7}(y_i-\overline{y})^2=7.08,$$

因此

$$r=\frac{14}{\sqrt{28\times 7.08}}\approx 0.99,\quad \hat{b}=\frac{14}{28}=0.5,\quad \hat{a}=2.3.$$

回归直线方程为 $\hat{y}=0.5x+2.3$.

(2) 在回归直线方程中令 $x=8$，得

$$\hat{y}=0.5\times 8+2.3=6.3,$$

因此预测第 8 年的利润为 6.3 亿元.

类似地，可预测第 9 年的利润为 6.8 亿元.

拓展阅读

相关系数与向量夹角的余弦

当 $n=2$ 时，相关系数的计算公式可改写为

$$r=\frac{(x_1-\overline{x})(y_1-\overline{y})+(x_2-\overline{x})(y_2-\overline{y})}{\sqrt{(x_1-\overline{x})^2+(x_2-\overline{x})^2}\times\sqrt{(y_1-\overline{y})^2+(y_2-\overline{y})^2}}.$$

此时，如果令 $\boldsymbol{a}=(x_1-\overline{x},\ x_2-\overline{x})$，$\boldsymbol{b}=(y_1-\overline{y},\ y_2-\overline{y})$，则相关系数 r 等于向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 的夹角的余弦，即

$$r=\cos\langle\boldsymbol{a},\ \boldsymbol{b}\rangle=\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|}.$$

类似地，当 $n=3$ 时，相关系数 r 仍等于两个向量夹角的余弦，只不过此时两个向量分别为

$$\boldsymbol{a}=(x_1-\overline{x},\ x_2-\overline{x},\ x_3-\overline{x}),\quad \boldsymbol{b}=(y_1-\overline{y},\ y_2-\overline{y},\ y_3-\overline{y}).$$

一般地，

$$\boldsymbol{a}=(x_1-\overline{x},\ x_2-\overline{x},\ \cdots,\ x_n-\overline{x}),\ \boldsymbol{b}=(y_1-\overline{y},\ y_2-\overline{y},\ \cdots,\ y_n-\overline{y})$$

都称为 n 维向量，如果按照类似 2 维与 3 维的情况定义向量的内积和模，则相关系数 r 总是等于两个向量夹角的余弦.

5. 非线性回归

尝试与发现

设某幼苗从观察之日起，第 x 天的高度为 y cm，测得的一些数据如下表所示.

第 x 天	1	4	9	16	25	36	49
高度 y/cm	0	4	7	9	11	12	13

作出这组数的散点图，并通过散点图思考：近似描述 y 与 x 的关系，除使用一次函数外，还可以用其他函数吗？具体应该怎样操作？

尝试与发现中数据的散点图如图 4-3-8 所示.

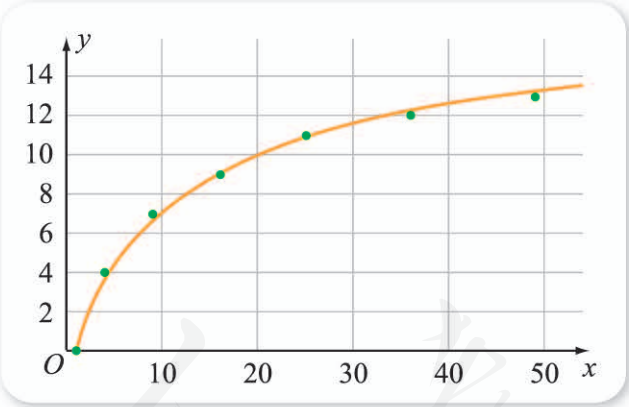


图 4-3-8

从散点图中可以看出，这些数据集中在图 4-3-8 中橙色的曲线附近，而且曲线的形状与函数 $y=\sqrt{x}$ 的图象很相似. 因此不难想到，可以用类似

$$y=b\sqrt{x}+a$$

的表达式来描述尝试与发现中 y 与 x 的关系. 那么，“最好”的曲线对应的未知系数如何求呢？如果令 $u=\sqrt{x}$，则上式可变为

$$y=bu+a,$$

这就是说，y 与 u 的关系可看成线性相关关系. 因此，我们可借助线性相关

的内容来求出“最好”的曲线. 具体过程如下.

令 $u=\sqrt{x}$，构造新的成对数据，如下表所示.

x	1	4	9	16	25	36	49
$u=\sqrt{x}$	1	2	3	4	5	6	7
y	0	4	7	9	11	12	13

容易算出，$\overline{u}=4$，$\overline{y}=8$.

通过列表计算可得

$$\sum_{i=1}^{7}(u_i-\overline{u})(y_i-\overline{y})=59,\quad \sum_{i=1}^{7}(u_i-\overline{u})^2=28,\quad \sum_{i=1}^{7}(y_i-\overline{y})^2=132.$$

因此

$$r=\frac{59}{\sqrt{28\times 132}}\approx 0.97,\ \hat{b}=\frac{59}{28},\ \hat{a}=-\frac{3}{7}.$$

故 y 关于 u 的回归直线方程为 $\hat{y}=\frac{59}{28}u-\frac{3}{7}$，代入 $u=\sqrt{x}$，则可知

$$\hat{y}=\frac{59}{28}\sqrt{x}-\frac{3}{7}.$$

这里的 y 与 x 的关系，因为不再是线性相关关系，所以称为**非线性相关关系**，所得到的方程称为**非线性回归方程**（也简称为**回归方程**）. 一般地，就像上面的实例一样，非线性回归方程的曲线类型可以通过作出散点图进行猜测，而回归方程有时可以通过变量替换后，借助求回归直线的过程确定. 当然，确定了非线性回归方程之后，也可以利用它进行预测. 例如，如果要预测尝试与发现中幼苗第 64 天的高度，可以直接将 $x=64$ 代入，从而预测高度为

$$\hat{y}=\frac{59}{28}\sqrt{64}-\frac{3}{7}=\frac{115}{7}\ \text{cm}.$$

6. 用信息技术求回归方程

回归系数和相关系数的计算步骤多，计算过程烦琐. 不过，借助计算器或者计算机软件，可以迅速地得出回归方程等.

例如，利用 GeoGebra 的“表格区”输入成对数据后，选中这些数据，然后点击“双变量回归分析”（如图 4-3-9 所示），确认数据后，就能得到数据的散点图. 在“回归模型”中选择对应的回归类型后，就能自动显示对应的方程. 输入自变量的值后，可以得到估计值. 另外，点击“显示统计”后，可以看到相关系数等值，如图 4-3-10 所示.

图 4-3-9

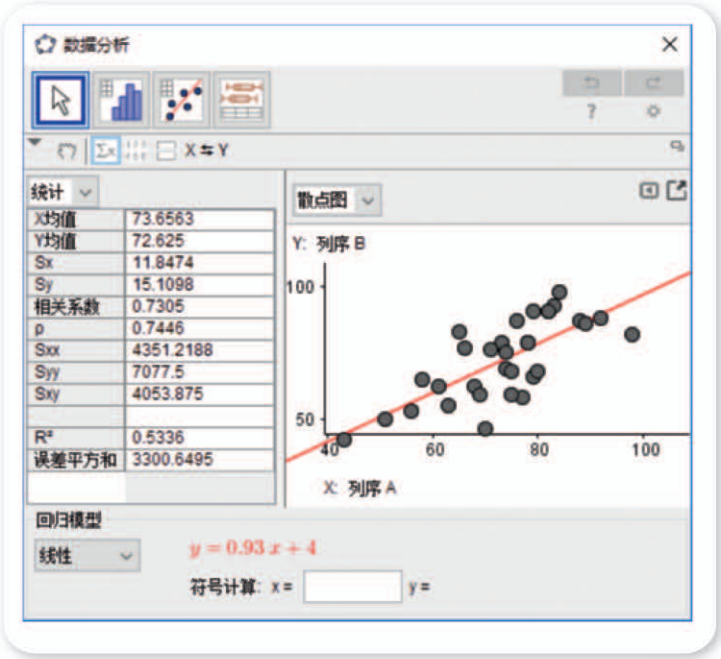


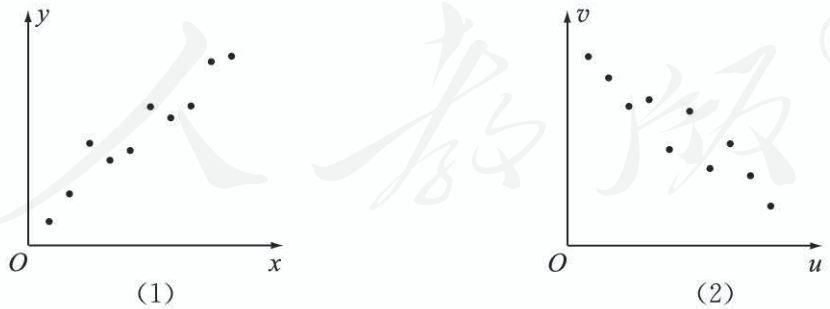
图 4-3-10

利用 Excel 也可得出回归方程等，请读者自行尝试.

练习A

1. 据报道：“一项在上海市 9 000 多名中小学生中进行的调查显示，打游戏时间和学业成绩呈明显的负相关.”依据这个结论，打游戏时间越多学习成绩就越有可能不好，对吗?

2. 根据变量 x，y 的观测数据可得散点图 (1)；根据变量 u，v 的观测数据可得散点图 (2). 由这两个散点图判断 x 与 y，u 与 v 之间的相关关系类型（即指出是正相关还是负相关).



（第 2 题）

3. 如果 x 与 y 线性相关，那么 y 与 x 也线性相关吗? 为什么?

4. 在一组样本数据 (x_1, y_1)，(x_2, y_2)，…，$(x_n, y_n)(n \geqslant 2, x_1, x_2, \cdots, x_n$ 不全相等) 的散点图中，若所有样本点 $(x_i, y_i)(i = 1, 2, \cdots, n)$ 都在直线 $y = \frac{1}{2}x + 1$ 上，求这组样本数据的相关系数.

5 已知变量 x 与 y 相对应的一组数据为 $(10,1)$，$(11.3,2)$，$(11.8,3)$，$(12.5,4)$，$(13,5)$；变量 u 与 v 相对应的一组数据为 $(10,5)$，$(11.3,4)$，$(11.8,3)$，$(12.5,2)$，$(13,1)$. 设 r_1 表示变量 y 与 x 之间的线性相关系数，r_2 表示变量 v 与 u 之间的线性相关系数，判断 r_1 与 r_2 的符号.

6 已知 y 与 x 具有相关关系，且利用 y 关于 x 的回归直线方程进行预测，$x=8$ 时 $\hat{y}=96$，且 $x=9$ 时 $\hat{y}=99$，求 y 关于 x 的回归直线方程中的回归系数.

练习B

1 已知 y 关于 x 的回归直线方程为 $\hat{y}=3x-13$，且 $u=10x$，求 y 关于 u 的回归直线方程.

2 已知 y 与 x 及 u 与 v 的成对数据如下，且 y 关于 x 的回归直线方程为 $\hat{y}=1.2x+0.6$，求 u 关于 v 的回归直线方程.

x	1	2	3	4	5
y	2	3	4	5	7

v	10	20	30	40	50
u	25	35	45	55	75

3 已知 y 与 x 具有相关关系，且 y 关于 x 的回归直线方程中，回归系数为 1.3，则当 x 每减少 3 个单位时，y 将怎样变化？

4 根据如下样本数据可得到的回归方程为 $\hat{y}=\hat{b}x+\hat{a}$，判断 $\hat{b}$ 与 $\hat{a}$ 的符号.

x	3	4	5	6	7	8
y	4.0	2.5	−0.5	0.5	−2.0	−3.0

5 为了解篮球爱好者小李的投篮命中率与打篮球时间之间的关系，下表记录了小李某月 1 号到 5 号每天打篮球时间 x（单位：h）与当天投篮命中率 y 之间的关系：

时间 x	1	2	3	4	5
命中率 y	0.4	0.5	0.6	0.6	0.4

(1) 求小李这 5 天的平均投篮命中率；

(2) 用线性回归分析的方法，预测小李该月 6 号打篮球 6 h 的投篮命中率.

6 如图所示，GeoGebra 软件中，回归类型分为“线性”“对数”等，分别选择每一种类型进行实验，总结出每一种类型的回归方程的形式.

（第 6 题）

1 正　2 负　3 $\hat{b}<0$　4 80　5 83　6 $r<0$　7 4

4.3.2 独立性检验

1. 独立性检验

我们已经知道，事件 A 与 B 独立的充要条件是

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

利用这一点，我们可以通过概率的计算来判断两个事件是否独立. 然而，如果要判断现实生活中两个随机事件是否独立，并不是一件容易的事.

情境与问题

任意抽取某市的一名学生，记 A：喜欢长跑，B：是女生.

(1) 你能得出 $P(A)$，$P(B)$，$P(AB)$这三者的准确值吗?

(2) 如果要判断 A 与 B 是否独立，该怎么办?

类似情境与问题中 $P(A)$，$P(B)$，$P(AB)$准确值的确定，都是比较难的，甚至是不可能的. 然而，因为可以利用频率估计概率，所以通过抽样调查获得样本数据后，我们就可以得到上述三个值的近似值.

例如，假设通过调查，我们获取了下述数据：抽查了 110 人，其中女生有 50 人；且这 110 人中，喜欢长跑的有 60 人，其中女生有 20 人.

首先，为了方便起见，我们可以把这些数据整理成如下的表格形式.

	喜欢长跑	不喜欢长跑	总计
女	20	30	50
男	40	20	60
总计	60	50	110

因为这个表格中，核心的数据是中间的 4 个格子，所以这样的表格通常称为 2×2 列联表.

由 2×2 列联表可知：

喜欢长跑的概率 $P(A)$可以估计为$\frac{60}{110}=\frac{6}{11}$；

是女生的概率 $P(B)$可以估计为 1 ；

喜欢长跑且是女生的概率 $P(AB)$可以估计为 2 .

想一想

利用已有的数据，你能估计出 $P(A\mid B)$的值吗?

尝试与发现

此时，可以利用 $P(AB)=P(A)P(B)$ 是否成立来判断 A 与 B 是否独立吗？为什么？

因为 $P(A)$，$P(B)$，$P(AB)$都是根据样本数据得到的估计值，而估计是有误差的，因此直接用 $P(AB)=P(A)P(B)$是否成立来判断 A 与 B 是否独立是不合理的.

但是，如果 A 与 B 独立，那么 $P(A)P(B)$应该可以作为 $P(AB)$的近似值. 因此理论上可知，喜欢长跑的女生数可以估计为 $110P(A)P(B)$，注意到实际数为 20（即 $110P(AB)$），因此

$$\frac{[110P(AB)-110P(A)P(B)]^2}{110P(A)P(B)}$$

应该不会太大.

类似地，考虑 $\overline{A}$ 与 B，A 与 $\overline{B}$，$\overline{A}$ 与 $\overline{B}$，可知

$$\frac{[110P(\overline{A}B)-110P(\overline{A})P(B)]^2}{110P(\overline{A})P(B)},$$

$$\frac{[110P(A\overline{B})-110P(A)P(\overline{B})]^2}{110P(A)P(\overline{B})},$$

$$\frac{[110P(\overline{A}\overline{B})-110P(\overline{A})P(\overline{B})]^2}{110P(\overline{A})P(\overline{B})}$$

都应该不会太大.

若记上述四项的和为 χ^2（读作“卡方”），则代入有关数据可以算得 $\chi^2\approx 7.8$.

不过，概率学上可以证明，如果 A 与 B 独立，则 $\chi^2\geqslant 6.635$ 的概率只有 1%，即 $P(\chi^2\geqslant 6.635)=1\%$. 因为算出的 χ^2 值 7.8 大于 6.635，所以若 A 与 B 独立（即“喜欢长跑”与“是女生”独立），那么我们就观察到了一件概率不超过 1%的事件. 这也可以说成，在犯错误的概率不超过 1%的前提下，可以认为“喜欢长跑”与“是女生”不独立（也称为是否喜欢长跑与性别有关）；或说有 99%的把握认为是否喜欢长跑与性别有关.

上述 1%通常称为显著性水平，而 6.635 称为显著性水平 1%所对应的分位数.

一般情况下，可以用完全类似的方法来检验两个随机事件是否独立.

如果随机事件 A 与 B 的样本数据的 2×2 列联表如下.

	A	$\overline{A}$	总计
B	a	b	$a+b$
$\overline{B}$	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

记 $n=a+b+c+d$，则由表可知：

(1) 事件 A 发生的概率可估计为 $P(A)=\dfrac{a+c}{n}$；

(2) 事件 B 发生的概率可估计为 $P(B)=$ 3 ；

(3) 事件 AB 发生的概率可估计为 $P(AB)=$ 4 .

如果 A 与 B 独立，那么上述 $P(AB)$ 与 $P(A)P(B)$ 的估计值相差不会太大，注意到总数为 n，因此利用后者可以估计出，理论上既是 A 又是 B 的数据有 $nP(A)P(B)$ 个，注意到实际的数据为 a（即 $nP(AB)$）个，因此

$$\frac{[nP(AB)-nP(A)P(B)]^2}{nP(A)P(B)}=\frac{[na-(a+c)(a+b)]^2}{n(a+c)(a+b)}$$

不会太大.

类似地，考虑 $\overline{A}$ 与 B，A 与 $\overline{B}$，$\overline{A}$ 与 $\overline{B}$，可知

$$\frac{[nb-(b+d)(a+b)]^2}{n(b+d)(a+b)},\ \frac{[nc-(a+c)(c+d)]^2}{n(a+c)(c+d)},\ \frac{[nd-(b+d)(c+d)]^2}{n(b+d)(c+d)}$$

都不会太大，因此这四个数的和

$$\chi^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

也不会太大.

另外，任意给定一个 α（称为**显著性水平**，通常取为 0.05，0.01 等），可以找到满足条件

$$P(\chi^2\geqslant k)=\alpha$$

的数 k（称为显著性水平 α 对应的**分位数**）. χ^2 是一个随机变量，其分布能够求出，上面的概率是可以计算的. 因此，如果根据样本数据算出 χ^2 的值后，发现 $\chi^2\geqslant k$ 成立，就称在犯错误的概率不超过 α 的前提下，可以认为 A 与 B 不独立（也称为 A 与 B 有关）；或说有 $1-\alpha$ 的把握认为 A 与 B 有关. 若 $\chi^2<k$ 成立，就称不能得到前述结论. 这一过程通常称为**独立性检验**.

A 与 B 独立时，也称为 A 与 B 无关. 当 $\chi^2<k$ 成立时，一般不直接说 A 与 B 无关. 也就是说，独立性检验通常得到的结果，或者是有 $1-\alpha$ 的把握认为 A 与 B 有关，或者没有 $1-\alpha$ 的把握认为 A 与 B 有关.

统计学中，常用的显著性水平 α 以及对应的分位数 k 如下页表所示.

$\alpha=P(\chi^2\geqslant k)$	0.1	0.05	0.01	0.005	0.001
k	2.706	3.841	6.635	7.879	10.828

例 1 为了了解阅读量多少与幸福感强弱之间的关系，一个调查机构得到了如下调查数据.

	幸福感强	幸福感弱	总计
阅读量多	54	18	72
阅读量少	36	42	78
总计	90	60	150

根据调查数据回答：在犯错误的概率不超过 1%的前提下，可以认为阅读量多少与幸福感强弱有关吗？

解 由题意可知

$$\chi^2=\frac{150\times(54\times42-18\times36)^2}{72\times78\times90\times60}=\frac{675}{52}\approx12.981.$$

又因为查表可得

$$P(\chi^2\geqslant6.635)=0.01,$$

由于 $12.981>6.635$，所以在犯错误的概率不超过 1%的前提下，可以认为阅读量多少与幸福感强弱有关.

例 2 某报刊对男女学生是否喜欢书法进行了一个随机调查，调查的数据如下表所示.

	喜欢书法	不喜欢书法
男学生	24	32
女学生	16	24

根据调查数据回答：有 95%的把握认为性别与是否喜欢书法有关吗？

解 由题意可知

$$\chi^2=\frac{(24+32+16+24)\times(24\times24-16\times32)^2}{(24+32)\times(16+24)\times(24+16)\times(32+24)}=\frac{96}{1\,225}\approx0.078.$$

又因为 $1-95\%=5\%$，而且查表可得

$$P(\chi^2\geqslant3.841)=0.05,$$

由于 5 _____，所以没有 95%的把握认为性别与是否喜欢书法有关.

2. 用信息技术进行独立性检验

很多软件都能进行独立性检验. 例如，运行 GeoGebra 后，打开“概率

统计”窗口，转到“统计”那一页，选定“卡方检验”后，设定“行”“列”的值都为2，在表格中输入数字后即可得到独立性检验的结果.

如图 4-3-11 所示是例 2 的检验结果，其中显示了 χ^2 的值，而 0.779 5 表示意思是，最多只有

$$1-0.779\ 5=0.220\ 5=22.05\%$$

的把握认为所检验的两件事情有关.

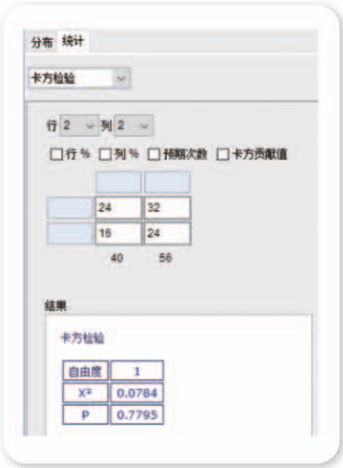


图 4-3-11

练习A

1. 某企业为了了解员工是否支持企业改革，抽查了 190 名员工进行调查，其中支持企业改革的有 80 人. 已知该企业共有员工 950 人，试估计该企业员工中支持企业改革的人数.
2. 为了探究成年人晕车与性别是否有关，调查了 320 名成年人，其中男士 152 人，而且男士与女士中，晕车的分别有 28 人与 32 人. 用 2×2 列联表表示这些数据.
3. 如果通过独立性检验发现，有 99%的把握认为 A 与 B 有关，那么是否一定有 95%的把握认为 A 与 B 有关?
4. 在检验 A 与 B 是否有关的过程中，根据所得数据算得 $\chi^2=12.5$，又

$$P(\chi^2\geqslant 6.635)=0.01,$$

那么是否有 99%的把握认为 A 与 B 有关?
5. 已知 $P(\chi^2<7.897)=0.095$，求 $P(\chi^2\geqslant 7.897)$.

练习B

1. 某学校在一次调查“体育迷”的活动中，获得了如下数据.

	男	女
体育迷	30	15
非体育迷	45	10

判断是否有 95%的把握认为是否体育迷与性别有关.
2. 某企业有甲、乙两个分厂生产同一种零件，在检查产品的优质品率时，从甲、乙两厂分别抽取了 500 件产品，其中甲厂有优质品 360 件，乙厂有优质品 320 件.
(1) 分别估计甲、乙两厂的优质品率;
(2) 是否有 99%的把握认为两个分厂生产的零件优质品率有差异?
3. 已知学生性别与考试是否及格无关，在抽样调查中，共调查了 52 人，其中女生有 32 人，且 52 人中考试及格的有 39 人. 试估计有多少女生考试是及格的.

4 为调查某地区老人是否需要志愿者提供帮助，用简单随机抽样的方法从该地区调查了 500 位老年人，结果如下：

	男	女
需要志愿者	40	30
不需要志愿者	160	270

(1) 估计该地区老年人中，需要志愿者提供帮助的老年人的比例；

(2) 能否有 99% 的把握认为该地区的老年人是否需要志愿者提供帮助与性别有关？

(3) 根据 (2) 的结论，能否提供更好的调查方法来估计该地区老年人中，需要志愿者帮助的老年人的比例？说明理由.

1 $\frac{50}{110}=\frac{5}{11}$　2 $\frac{20}{110}=\frac{2}{11}$　3 $\frac{a+b}{n}$　4 $\frac{a}{n}$　5 $0.078<3.841$

习题4-3A

1 某公司根据以往的数据发现，销售收入 y 万元与广告费 x 万元线性相关，且回归直线方程为 $\hat{y}=120x+60$，试估计广告费用每增加 2 万元时，销售收入的增加量.

2 已知根据某样本数据可得到回归方程为 $\hat{y}=4x+\hat{a}$，且 $\overline{x}=3$，$\overline{y}=6$，求 $\hat{a}$ 的值.

3 已知变量 x 和 y 满足关系 $y=-0.1x+1$，变量 y 与 z 正相关. 下列结论中正确的是 (　　).

(A) x 与 y 负相关，x 与 z 负相关　　(B) x 与 y 正相关，x 与 z 正相关

(C) x 与 y 正相关，x 与 z 负相关　　(D) x 与 y 负相关，x 与 z 正相关

4 某电视台在一次对收看文艺节目和新闻节目观众的抽样调查中，随机抽取了 100 名电视观众，相关的数据如下表所示：

	文艺节目	新闻节目
20～40 岁	40	18
大于 40 岁	15	27

(1) 由表中数据分析，是否有 95% 的把握认为收看新闻节目的观众与年龄有关？

(2) 用分层抽样的方法在收看新闻节目的观众中随机抽取 5 名，大于 40 岁的观众应该抽取几名？

(3) 在上述抽取的 5 名观众中任取 2 名，求恰有 1 名观众的年龄为 20～40 岁的概率.

习题4-3B

1 已知 x 与 y 之间的几组数据如下表所示.

x	1	2	3	4	5	6
y	0	2	1	3	3	4

假设根据上表数据所得线性回归直线方程为 $\hat{y}=\hat{b}x+\hat{a}$，若某同学根据上表中的前两组数据 (1，0) 和 (2，2)，求得一次函数表达式为 $y=b'x+a'$. 判断 $\hat{b}$ 与 b' 的相对大小，以及 $\hat{a}$ 与 a' 的相对大小.

2 某地最近十年粮食需求量逐年上升，下表是部分统计数据.

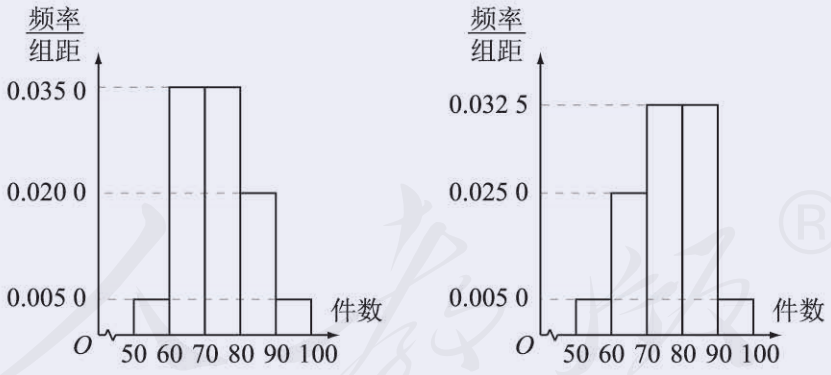
年份 x	2012	2014	2016	2018	2020
需求量 y/万吨	236	246	257	276	286

(1) 利用所给数据求年需求量与年份之间的回归方程；

(2) 利用 (1) 中所求出的线性回归方程预测该地 2022 年的粮食需求量.

3 已知 $P(\chi^2<6)=a$，$P(\chi^2<7)=b$，判断 $a\geqslant b$ 与 $a\leqslant b$ 哪个一定成立.

4 某工厂有 25 周岁及以上的工人 300 名，25 周岁以下的工人 200 名. 为研究工人的日平均生产量是否与年龄有关，现采用分层抽样的方法，从中抽取了 100 名工人，先统计了他们某月的日平均生产件数，然后按工人年龄在“25 周岁及以上”和“25 周岁以下”分为两组，再将两组工人的日平均生产件数分成 5 组：[50，60)，[60，70)，[70，80)，[80，90)，[90，100]，分别加以统计，得到如图所示的频率分布直方图.



(25 周岁及以上组)　　(25 周岁以下组)

(第 4 题)

(1) 从样本中日平均生产件数不足 60 件的工人中随机抽取 2 人，求至少抽到一名“25 周岁以下”的工人的概率；

(2) 规定日平均生产件数不少于 80 件者为“生产能手”，请你根据已知条件列出 2×2 列联表，并判断是否有 90% 的把握认为生产能手与工人所在的年龄组有关.

4.4 数学探究活动：了解高考选考科目的确定是否与性别有关

1. 活动背景介绍与要求

按照党的十八届三中全会审议通过的《中共中央关于全面深化改革若干重大问题的决定》，高考招生制度改革后，考生在报考时，可以根据报考高校提前发布的招生报考要求和自身特长，从思想政治、历史、地理、物理、化学、生物 6 科中自主选择 3 个科目的成绩，计入高考总分.

由此产生的一个问题是，高考选考科目的确定是否与性别有关呢？例如，有人认为，在选择物理的同学中，男生所占的比例更大；选择历史的人中，女生会更多. 这些看法是否有道理呢？我们能不能通过收集相关数据，并利用有关概率统计知识来说明？

请与其他同学一起分工合作，完成下列任务，并填写活动记录表：

（1）确定要研究的科目与要调查的人群范围；

（2）选定合适的收集数据的方法，并收集数据；

（3）根据有关统计知识和获得的数据，得出结论；

（4）对结论进行分析，并给出科目选择的建议.

“了解高考选考科目的确定是否与性别有关”的活动记录表

活动开始时间：_____

（1）成员与分工

姓　名	分　工

（2）待研究的科目与待调查的人群范围

续表

（3）收集数据的方法以及获得的数据
（4）所依据的统计知识以及有关结论
（5）对结论的分析以及科目选择的建议
（6）活动总结（可包括活动感受等）

活动结束时间：_____

2. 活动提示

活动过程中，要特别注意样本的代表性. 例如，若仅选择某个班级的学生进行调查，则获得的样本可能不具备代表性.

另外，科目的确定不仅可能与性别有关，也有可能与考生想要报考的专业等有关. 除了了解科目的确定与性别是否有关外，我们还可以了解科目的确定是否与其他因素有关.

本章小结

01 知识结构图设计与交流

本章我们首先学习了条件概率的有关知识，在此基础上得到了乘法公式与全概率公式，并利用条件概率重新理解了事件的独立性；随后，我们学习了有关随机变量的知识，并研究了离散型随机变量的分布列与数字特征等，还了解了二项分布、超几何分布、正态分布；最后，还学习了成对数据的统计相关性，了解了一元线性回归模型和独立性检验.

依照知识之间的联系，我们可以作出如下的知识结构图.

条件概率 $P(A\mid B)=\dfrac{P(AB)}{P(B)}$
乘法公式
全概率公式 $P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid \overline{A})$
离散型随机变量
二项分布
超几何分布
正态分布 $X\sim N(\mu,\sigma^2)$　$E(X)=\mu$　$D(X)=\sigma^2$
一元线性回归模型
独立性检验

请在上述图中补充更多的内容吧！你能作出更有特色的知识结构图吗？试着与同学们交流一下吧！

02 课题作业

（1）概率统计的有关知识已经渗透到了现代生活的方方面面. 在新闻报纸、科普著作、学术论文等中经常可以见到概率统计知识. 与其他同学合作，连续 7 天收集生活中所用到的概率统计知识，并以调研报告或小论文的形式进行整理，然后与其他同学交流.

（2）概率统计的知识目前在人工智能中有着广泛的应用. 例如，条件概率的知识在语音识别、机器学习等中是不可缺少的. 查阅网络或向有关专业人士请教，了解概率统计知识在人工智能中的应用情况，并整理成演讲材料.

03 复习题

A 组

1. 两个实习生每人加工一个零件，加工为一等品的概率分别为$\frac{2}{3}$和$\frac{3}{4}$，两个零件是否加工为一等品相互独立，求这两个零件中恰有一个一等品的概率.

2. 某班从 6 名班干部（男生 2 人，女生 4 人）中，任选 3 人参加学校的义务劳动. 设“男生甲被选中”为事件 A，“女生乙被选中”为事件 B，求 $P(B \mid A)$.

3. 某射击运动员进行射击训练时，假设每次击中目标的概率均为 0.6，且每次射击的结果互不影响，已知射手射击了 5 次，求：

（1）其中恰有 3 次击中目标的概率；

（2）其中恰有 3 次连续击中目标，而其他两次没有击中目标的概率.

4. 乒乓球单打决赛在甲、乙两名运动员间进行，比赛采用 7 局 4 胜制，假设两人在每一局比赛中获胜的可能性相等.

（1）求甲以“4∶1”获胜的概率；

（2）求乙获胜且比赛局数多于 5 局的概率.

5. 已知 X 服从两点分布，且 $P(X=0)=0.3$，求 $P(X=1)$.

6. 已知 X 与 Y 都是随机变量，X 的取值范围是$\{-1, 0, 1, 2\}$，而且 $Y=X^2$，求 Y 的取值范围.

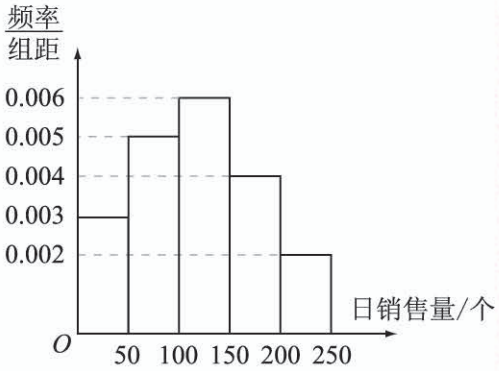
7. 已知某种疗法的治愈率为 90%，若有 10 位患者采用了这种疗法，其中被治愈的人数为 X，指出 X 满足的分布列，并求 $P(X=10)$.

8. 设某鞋店的每位顾客需要 25 号鞋的概率均为 0.2，鞋店上午开门营业后，前 5 名顾客中需要 25 号鞋的人数为 X，指出 X 满足的分布列，并求 $P(X\geqslant 1)$.

9. 假设每一年都只有 365 天，而且每人在任意一天中出生的概率都相等. 设一

个有 30 人的班级中，恰有 X 位同学在元旦出生，指出 X 满足的分布列，并求 $P(X\geqslant 2)$.

10. 一家面包房根据以往某种面包的销售记录，绘制了日销售量的频率分布直方图，如图所示. 将日销售量落入各组的频率视为概率，并假设每天的销售量相互独立.



（第 10 题）

（1）求在未来 3 天里，有连续两天的日销售量都不低于 100 个且另一天的日销售量低于 50 个的概率；

（2）用 X 表示在未来 3 天里日销售量不低于 100 个的天数，求随机变量 X 的分布列，期望 $E(X)$及方差 $D(X)$.

11. 在调查男女学生购买食品时是否阅读营养成分说明时，调查了 36 位男生、38 位女生，而且阅读营养成分说明的人有 46 位，阅读营养成分说明的人中有 28 位女生. 用 2×2 列联表表示上述数据.

B 组

1. 掷红、蓝两个均匀的骰子，设 A：红色骰子的点数为 4，B：蓝色骰子的点数是偶数，求 $P(A\mid B)$.

2. 甲、乙两队进行排球决赛，现在的情形是甲队只要再赢一局就得冠军，乙队需要再赢两局才能得冠军，若两队每局赢的概率相等，求甲队获得冠军的概率.

3. 在三次独立重复试验中，事件 A 在每次试验中发生的概率相等，若事件 A 至少发生一次的概率为$\dfrac{63}{64}$，求事件 A 恰好发生一次的概率.

4. 已知某电脑卖家只卖甲、乙两个品牌的电脑，其中甲品牌的电脑占 70%. 甲品牌的电脑中，优质率为 80%；乙品牌的电脑中，优质率为 90%. 从该电脑卖家中随机购买一台电脑：

（1）求买到优质电脑的概率；

（2）若已知买到的是优质电脑，求买到的是甲品牌电脑的概率（精确到 0.1%）.

5. 已知 $P(\overline{A}\mid B)=0.7$，$P(A)=0.3$，判断 A 与 B 是否独立.

6. 已知

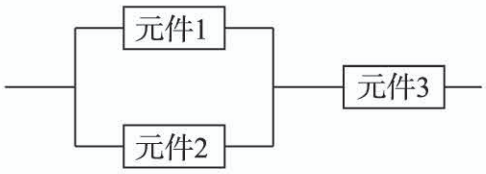
$$P(\overline{A})=\frac{1}{2},\ P(\overline{B}\mid A)=\frac{2}{3},\ P(B\mid \overline{A})=\frac{1}{4},$$

求 $P(\overline{B})$，$P(\overline{A}\mid B)$.

7. 一个布袋中共有 50 个完全相同的球，其中标记为 0 号的有 5 个，标记为 n 号的分别有 n 个（$n=1$，2，…，9），求从布袋中任取一球所得号数的分布列.

8. 甲、乙两名选手进行比赛，假设每局比赛中，甲胜的概率为 0.6，乙胜的概率为 0.4. 那么，“三局两胜制”与“五局三胜制”，哪个对甲来说更有利？由此你能得到怎样的一般结论？

9. 某一部件由三个电子元件按如图方式连接而成，元件 1 或元件 2 正常工作，且元件 3 正常工作时，部件正常工作. 设三个电子元件的使用寿命（单位：h）均服从正态分布 $N(1\,000, 50^2)$，且各个元件能否正常工作相互独立，求该部件的使用寿命超过 1 000 h 的概率.

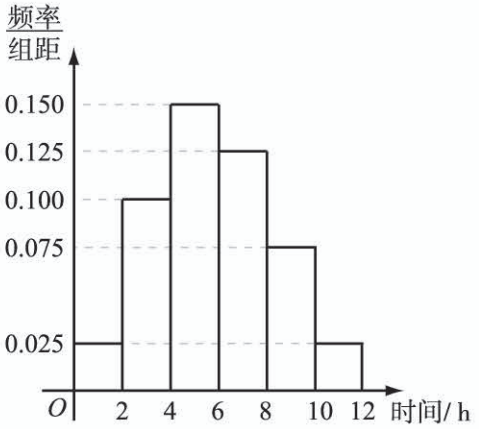


（第 9 题）

10. 某高校共有 15 000 人，其中男生 10 500 人，女生 4 500 人，为调查该校学生每周平均体育运动时间的情况，采用分层抽样的方法，收集 300 位学生每周平均体育运动时间的样本数据（单位：h）.

（1）应收集多少位女生样本数据？

（2）根据这 300 个样本数据，得到学生每周平均体育运动时间的频率分布直方图如图所示，其中样本数据分组区间为：[0，2]，(2，4]，(4，6]，(6，8]，(8，10]，(10，12]. 估计该校学生每周平均体育运动时间超过 4 h 的概率.

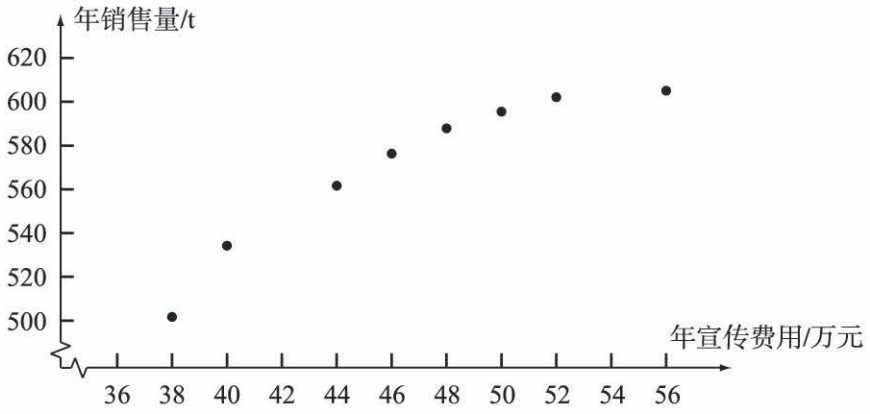


（第 10 题）

（3）在样本数据中，有 60 位女生的每周平均体育运动时间超过 4 h. 请制作每周平均体育运动时间与性别的 2×2 列联表，并判断是否有 95% 的把握认为该校学生的每周平均体育运动时间与性别有关.

C 组

1. 某公司为确定下一年度投入某种产品的宣传费，需了解年宣传费 x（单位：万元）对年销售量 y（单位：t）和年利润 z（单位：万元）的影响. 对近 8 年的年宣传费 x_i 和年销售量 y_i（$i=1$，2，…，8）数据进行了初步处理，得到下面的散点图及一些统计量的值.



（第 1 题）

$\overline{x}$	$\overline{y}$	$\overline{w}$	$\sum_{i=1}^{8}(x_i-\overline{x})^2$	$\sum_{i=1}^{8}(w_i-\overline{w})^2$	$\sum_{i=1}^{8}(x_i-\overline{x})(y_i-\overline{y})$	$\sum_{i=1}^{8}(w_i-\overline{w})(y_i-\overline{y})$
46.6	563	6.8	289.8	1.6	1 469	108.8

表中 $w_i=\sqrt{x_i}$，$\overline{w}=\frac{1}{8}\sum_{i=1}^{8}w_i$.

（1）根据散点图判断，$y=a+bx$ 与 $y=c+d\sqrt{x}$ 哪一个适宜作为年销售量 y 关于年宣传费 x 的回归方程类型？

（2）根据（1）的判断结果及表中数据，建立 y 关于 x 的回归方程；

（3）已知这种产品的年利润 z 与 x，y 的关系为 $z=0.2y-x$. 根据（2）的结果回答下列问题：年宣传费 $x=49$ 时，年销售量及年利润的预测值是多少？年宣传费 x 为何值时，年利润的预测值最大？

2. 已知 A，B 两个投资项目的利润率分别为随机变量 X_1 和 X_2. 根据市场分析，X_1 和 X_2 的分布列如下.

X_1	5%	10%
P	0.8	0.2

X_2	2%	8%	12%
P	0.2	0.5	0.3

（1）在 A，B 两个项目上各投资 100 万元，Y_1 和 Y_2 分别表示投资项目 A 和 B 所获得的利润，求 $D(Y_1)$ 和 $D(Y_2)$；

（2）将 $x(0\leqslant x\leqslant 100)$ 万元投资 A 项目，$100-x$ 万元投资 B 项目，$f(x)$ 表示投资 A 项目所得利润的方差与投资 B 项目所得利润的方差之和. 求 $f(x)$ 的最小值，并指出 x 为何值时，$f(x)$ 取到最小值.

后　记

本套教科书是人民教育出版社课程教材研究所中学数学教材实验研究组依据教育部《普通高中数学课程标准（2017年版）》编写的，经国家教材委员会专家委员会2019年审核通过.

本套教科书的编写，集中反映了我国十余年来普通高中课程改革的成果，吸取了2004年版《普通高中课程标准实验教科书 数学（B版）》的编写经验，凝聚了参与课改实验的教育专家、学科专家、教材编写专家、教研人员和一线教师，以及教材设计装帧专家的集体智慧.

我们衷心感谢2004年版《普通高中课程标准实验教科书 数学（B版）》的所有编写人员，尤其是因为种种原因未能参加此次教材修订的专家、学者：丁尔陞、江守礼、房艮孙、张润琦、高尚华、万庆炎、魏榕彬、邱万作、陈研、段发善、李沄岸、陈亦飞、刘长明、郭鸿、王池富……

本套教科书在编写过程中，得到了《普通高中数学课程标准（2017年版）》制定组、国家教材委员会专家委员会等的大力支持. 借此机会，向所有制定组成员、专家委员会成员以及其他为我们教材编写提供过帮助的专家表示衷心的感谢！

我们感谢所有对本套教科书的编写、出版、试教等提供过帮助与支持的同仁和社会各界朋友：王跃飞、胡细宝、邵丽云、王晓声、曹付生、侯立伟、王中华、王光图、王秀梅、卞文、邓艳强、田媛、史洪波、付一博、吕希、吕晶、朱明鲜、刘超、闫旭、池洪清、阮征、孙国华、牟柏林、李刚、李广勤、李洪岩、何艳国、张伟、张羽、张明、张文刚、张春青、张晶强、金永涛、郑继平、常丽艳、潘戈、薛达志、郑海军、赵争鸣、吴晖湘、戴莉、金盈、舒凤杰、李祥广、胡文亮、王玉洁、杨长智、徐会吉、尹玉柱、尹燕花……

本套教科书出版之前，我们通过多种渠道与教科书选用作品（包括照片、画作）的作者进行了联系，得到了他们的大力支持. 对此，我们表示衷心的感谢！同时也向为本书提供照片的单位表示感谢！

我们真诚地希望广大教师、学生及家长在使用本套教科书的过程中提出宝贵意见．我们将集思广益，不断修订，以使教科书日臻完善．

本书责任编辑：周琳；美术编辑：史越；插图绘制：郑海军．

联系方式

电话：010–58758527，010–58758866

电子邮箱：mathb@pep.com.cn，jcfk@pep.com.cn

人民教育出版社　课程教材研究所

中 学 数 学 教 材 实 验 研 究 组

2019年4月

PUTONG GAOZHONG JIAOKESHU

SHUXUE

绿色印刷产品

ISBN 978-7-107-34579-6

定价：00.00 元