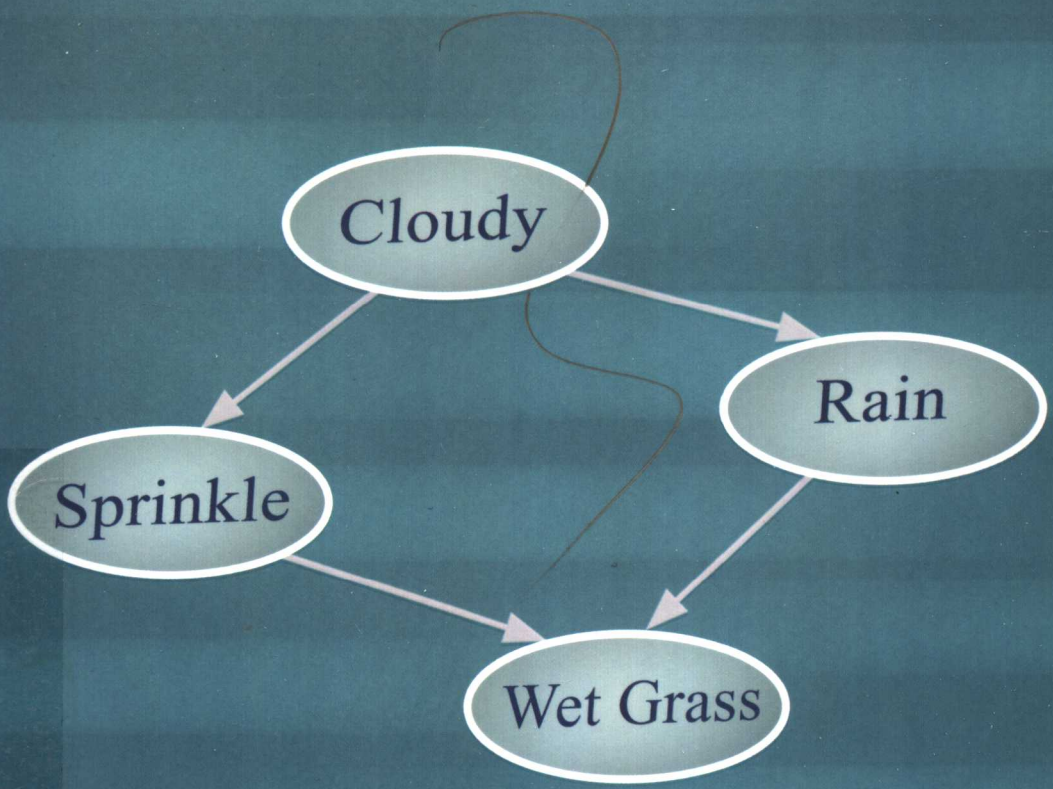
智能数据分析

Intelligent Data Analysis

刘惟一　李维华　岳　昆 著



科学出版社
www.sciencep.com

智能数据分析

刘惟一　李维华　岳　昆　著

国家自然科学基金资助项目（项目编号：60263006）

云南省自然科学基金资助项目（项目编号：2002F0011M）

科学出版社

北京

内 容 简 介

本书以数据分析为主线，旨在利用模糊集、粗糙集、遗传算法和机器学习等不确定人工智能方法分析数据间的依赖关系、概率因果关系、数据分类与聚类，并用于决策、对策及融合分析。

本书的主要特点：在阐述相关领域最基本、最重大的成果时，也介绍这些领域的最新进展，以及作者在这方面的工作；对大多数理论问题给出了证明、直观论据和实例，对问题的实现给出了具体的算法。

本书可作为知识发现、智能信息处理、决策分析等领域的研究、开发人员的参考书，也可作为计算机、信息系统等专业研究生的教材。

图书在版编目(CIP)数据

智能数据分析/刘惟一，李维华，岳昆著. —北京：科学出版社，2007
ISBN 978-7-03-019902-7

Ⅰ.智… Ⅱ.①刘… ②李… ③岳… Ⅲ.智能-数据-分析
Ⅳ.TP311.13

中国版本图书馆 CIP 数据核字（2007）第 137416 号

责任编辑：鞠丽娜/责任校对：赵 燕
责任印制：吕春珉/封面设计：三函设计

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2007 年 9 月第 一 版 开本：787×1092 1/16
2007 年 9 月第一次印刷 印张：21 1/2
印数：1—2 500 字数：510 000

定价：39.00 元

（如有印装质量问题，我社负责调换〈环伟〉）
销售部电话 010-62136131 编辑部电话 010-62138978-8002（BI08）

前　言

进入信息时代，面临知识爆炸和信息泛滥，人们力图从海量数据中挖掘出有用的信息，获取所需的知识，这就需要利用人工智能的方法对数据进行分析。本书的主题是智能化和数据分析。面对这个含义宽泛的主题，我们不可能，也不想去涉猎所有的相关领域。本书旨在利用模糊集、粗糙集、遗传算法和机器学习等不确定人工智能方法去讨论数据间的依赖关系、概率因果关系、数据分类与聚类，并用于决策、对策及融合分析。本书在试图阐述相关领域的最基本、最重大的成果时，也介绍这些领域的最新进展，以及作者在这些方面的工作。

本书以数据分析为主线，将数据间的关系与决策应用联系起来，这就涉及理论分析与实际实现两方面的内容。本书对大多数理论问题给出了证明，对部分没有证明的重要结论提供了直观的论据和实例，以及相关的引用文献。本书对问题的实现几乎都给出了具体的算法，但不陷入问题的细节，因为枝节会使解决问题的思路模糊。虽然由算法和示例过渡到实现并不总是十分容易的，但并没有实质性困难。

本书从不同角度就以下几个方面的内容展开了讨论。

第一章介绍不确定信息处理的理论与方法，主要涉及模糊集、粗糙集、概率论、信息熵及遗传算法，所介绍的理论是以后章节的基础。本章只对本书要用到的知识做扼要介绍，不进一步地展开讨论。

第二章介绍关系数据理论。数据依赖是定义在数据库上的语义约束，它反映了属性集间的客观语义联系。本章介绍了函数依赖、多值依赖、连接依赖以及非圈连接依赖，强调依赖理论的和谐性；专门讨论了数据依赖间的蕴涵关系，给出了蕴涵问题的判定方法；进一步将数据依赖扩展到模糊环境，建立了精确值、模糊值统一的数据依赖系统。

第三章讨论分类与聚类。聚类是将数据对象划分成若干个类，使同一类对象具有较高的相似度，而与其他类中的对象有较大差异。分类的目的是将指定的对象分配到合适的类别中。本章介绍了相似性度量和各种聚类方法；讨论了分类模型和不同的分类方法。

第四章讨论事件间的概率因果关系。贝叶斯网是概率因果关系的表示和推理的有效工具。本章介绍了贝叶斯网的基本概念，进而讨论了贝叶斯网的多种构造方法和推理算法。利用数据依赖构造贝叶斯网，以及网的结点聚集是作者的工作成果。

第五章讨论基于影响图的决策分析。影响图是基于贝叶斯网的一种决策模型，可以利用贝叶斯推断计算决策行为的期望效用。本章首先介绍基本统计决策原理和影响图的概念；其次讨论了影响图决策的遗传算法以及影响图结构、效用参数的机器学习算法，最后讨论了关于对手决策模型估计的修正方法。

第六章讨论博弈问题。在多决策系统影响图中，讨论各个系统相互制约的策略选择时，就要用到博弈论。博弈论是关于策略相互作用的理论。本章首先介绍了 Nash 等人

关于博弈的基本理论；进而，作者讨论了求解 Nash 均衡的遗传算法、多步博弈的增强学习算法，并利用粗糙集等方法讨论了各个对局者在博弈中的地位。

第七章讨论融合分析。数据融合是 20 世纪 70 年代提出的概念，随着它在军事及社会生活中的应用，融合概念被扩展。将本书各章节中存在的融合问题归纳起来就是融合分析一章讨论的内容。对于不同信息源提供的证据，将它们统一起来得出综合的判断，这就是身份与证据的融合。人们注意到概率推理与逻辑推理是不相容的，将二者统一起来就是推理融合问题。将多个信念模型——贝叶斯网合并起来，这是模型融合。对于多目标、多人决策问题，根据不同的决策模型相互取长补短，就是决策融合。本章根据具体的融合问题分别给出不同的融合方法与算法。

在成书过程中得到中国人民大学王珊教授的鼓励和支持。云南大学的研究生何盈捷、张忠玉、赵云、王晓峰、李劲、白磊、徐阿进等同学参加了课题的研究，并做出了很好的成果。本书所反映的研究成果得到国家自然科学基金资助项目（项目编号：60263006）和云南省自然科学基金资助项目（项目编号：2002F0011M）的资助，在此一并表示衷心的感谢。

由于作者水平有限，书中不妥之处在所难免，恳请读者批评指正。

目　录

第一章　不确定性理论与方法

1.1　概率基础

1. 随机事件及其运算

在一定条件下必然出现(或不出现)某种结果的现象称为确定性现象。例如:向上抛掷的重物必然自由下落。另一类是在相同条件下可能得到多种不同结果的现象。例如抛掷一枚硬币,落下后可能正面朝上,也可能反面朝上。这类现象虽然在个别试验中,其结果呈现出不确定性,但经多次重复试验,其结果呈现出某种客观规律性。例如多次抛掷同一硬币,正面朝上的次数大致占抛掷总次数的一半。我们把在个别试验中呈现不确定性,而在大量重复试验中又具有统计规则性的现象称为随机现象。

对随机现象进行观察叫做随机试验,随机试验具有如下特征:

1) 在同样的条件下,这种试验可以重复进行;

2) 试验的结果不止一个,每次试验只能出现其中的一个结果,并且事先不能断定必然要出现哪一个结果;

3) 能够明确指出这种试验可能出现的一切结果。

概率论研究的就是随机试验,并简称为试验。

定义 1.1.1　试验中每个可能出现的结果都叫做样本点,全体样本点构成的集合叫做样本空间。我们用 $w_1,w_2,\cdots$ 表示样本点,用 Ω 表示样本空间。

例如掷一硬币,观察哪个面朝上,试验有两个可能结果:正面和反面。用 w_1 表示正面,w_2 表示反面,则 w_1,w_2 都是样本点,而样本空间 $\Omega=\{w_1,w_2\}$。

又如对目标进行射击,直到击中目标为止。用"0"表示未击中,"1"表示击中,这个试验的可能结果有

w_1="1"(第一次射击就击中目标)

w_2="01"(第一次未击中,第二次才击中)

……

w_n="00…01"(前 $n-1$ 次皆未击中,第 n 次才击中)

……

这个试验有无穷多个可能结果,样本空间 $\Omega=\{w_1,w_2,\cdots,w_n,\cdots\}$

定义 1.1.2　样本空间 Ω 中,具有某种性质的样本点构成的集合称为随机事件,简称事件。随机事件是样本空间 Ω 的子集,常用字母 $A,B,\cdots$ 或 $A_1,A_2,\cdots$ 表示。

例如一次掷两枚硬币。设 A 表示"至少一个正面"的事件。这个试验的样本点有

$$w_1=(\text{正},\text{正}),w_2=(\text{正},\text{反}),w_3=(\text{反},\text{正}),w_4=(\text{反},\text{反})$$

事件 $A=\{w_1,w_2,w_3\}$,这里注意到 $\{w_1\}$ 与 w_1 不同,w_1 表示样本点,而单点集 $\{w_1\}$ 是事件。

事件 A 发生，当且仅当 A 中的某一个样本点发生。我们用全集 Ω 和空集 $\varnothing$ 分别表示必然事件和不可能事件。可以这样理解：每次试验必有 Ω 中某一个样本点发生，所以 Ω 是必然事件；由于空集 $\varnothing$ 不含样本点，每次试验，事件 $\varnothing$ 一定不发生，所以 $\varnothing$ 是不可能事件。

下面讨论事件间的关系及运算。

1) 包含：若事件 A 发生必然导致事件 B 发生（即 A 中每个样本点都包含在 B 中，$w \in A$ 必有 $w \in B$），称事件 B 包含事件 A，记为 $B \supseteq A$。

2) 相等：$A=B$ 指每次试验中 A 与 B 同时发生，或同时不发生，即 $B \supseteq A$ 且 $A \supseteq B$。

3) 并：$A \cup B$ 指“事件 A 与 B 至少有一个发生”构成的事件，或者记作 $A+B$。

4) 交：$A \cap B$ 指“事件 A 与 B 同时发生”构成的事件，$A \cap B$ 可写成 AB。

5) 差：$A-B$ 是“事件 A 发生而 B 不发生”构成的事件。

6) 互斥：$A \cap B=\varnothing$，即 A 与 B 设有公共的样本点，指事件 A 与 B 在任何次试验中不同时发生。

7) 逆：$\Omega-A$ 称为 A 的逆事件，即 A 不发生构成的事件，记为 $\neg A$。A 与 $\neg A$ 显然有 $A \cup \neg A=\Omega, A \cap \neg A=\varnothing$。

例 1.1.1 掷一颗骰子，观察它六个点的情况。设 A 事件为出现 1,3,5 点，记 $A=\{1,3,5\}$，同样设 $B=\{1,2,3,4\}, C=\{2,4\}$。

样本空间

$$\Omega = \{1,2,3,4,5,6\}$$
$$A \cup B = \{1,2,3,4,5\}$$
$$A \cap B = \{1,3\}$$
$$\neg(AB) = \{2,4,5,6\}$$

2. 随机事件的概率及其性质

对随机事件来说，在相同条件下大量重复进行试验，人们发现一个随机事件发生的可能性是确定的，不同事件发生的可能性有大小之分。这种可能性的大小是事件本身固有的性质，为了定量描述它，我们将给出事件概率的概念。首先介绍频率的概念。

定义 1.1.3 在相同条件下，重复 n 次试验 E，随机事件 A 在 n 次试验中出现的次数 m 称为频数，比值 m/n 称为事件 A 的频率，记作 $f_n(A)$，即 $f_n(A)=m/n$。

可以验证，当试验次数 n 固定时，$f_n(A)$ 有如下性质：

1) $0 \leqslant f_n(A) \leqslant 1$；

2) $f_n(\Omega)=1$；

3) 若事件 $A_1, A_2, \cdots, A_k$ 互不相容，即 $A_i \cap A_j=\varnothing$ $(i,j=1,\cdots,k, i \neq j)$，则

$$f_n\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} f_n(A_i)$$

定义 1.1.4 在相同条件下，重复进行 n 次试验，事件 A 发生的频率 $f_n(A)$ 在某个常数值 p 附近摆动。一般说来，n 越大，摆动幅度越小，称频率的稳定值 p 为事件 A 发生的概率，记作 $p(A)$。

例 1.1.2　某市的电话号码由七位数字组成，每位数可以是 0，1，…，9 中任一个，求由完全不同的数码组成的电话号的概率。

解　从十个数字中任取一个，允许重复的七位数码共有 10^7 种不同的排列方式，这样基本事件总数 $n=10^7$。设事件 A 为七个数码全不同。这相当于从第一次中的十个数字中任取一个，第二次在第一次取后剩下的九个数中任取一个，……如此取七次，A 包含的基本事件数是：$m=10\times9\times8\times7\times6\times5\times4=604800$，因此 $p(A)=604800/10^7=0.06048$。

概率的统计定义虽然能适合一般情况，但是它主要依据大量重复试验中频率所呈现的稳定性这一事实，然而重复试验的次数究竟应该大到怎样的程度，以及怎样理解频率在某个常数值附近摆动，这些都无法用确切的数学术语来表达。为此需要抽象地给出概率的定义，建立概率的公理化体系。

每个事件都是样本空间 Ω 的子集，这些子集构成的集合称为事件域。

定义 1.1.5　设 F 是由样本空间 Ω 的一些子集组成的集合，若它满足以下条件：

1) $\Omega\in F$；

2) 若 $A\in F$，则 $\neg A\in F$；

3) 若 $A_i\in F(i=1,2,\cdots)$，则 $\sum\limits_{i=1}^{\infty}A_i\in F$；

那么则称 F 为事件域，称 F 中的元素为事件。

把概率看作是定义在事件域 F 上的函数，且这一函数应当满足一定的要求。

定义 1.1.6　设 Ω 是一个给定的样本空间，F 是 Ω 的一事件域，$A\in F$，若实值函数 $p(A)$ 满足：

1) 对每一个 $A\in F$，有 $0\leqslant p(A)\leqslant1$；

2) 对必然事件 Ω，有 $p(\Omega)=1$；

3) 若 $A_i\in F(i=1,2,\cdots)$，$A_iA_j=\varnothing(i\neq j)$，则 $p(\sum\limits_{i=1}^{\infty}A_i)=\sum\limits_{i=1}^{\infty}p(A_i)$；

那么则称函数 $p(A)$ 为事件 A 的概率。

Ω,F,p 描述了一个随机试验的三个组成部分，我们称三元组 (Ω,F,p) 为一个概率空间。

由概率的公理化定义可以导出它的一系列基本性质。

性质 1　$p(\varnothing)=0$

性质 2　若 $A_1,A_2,\cdots,A_n$ 两两互不相容，则 $p(A_1+A_2+\cdots+A_n)=p(A_1)+p(A_2)+\cdots+p(A_n)$

性质 3　若 A 与 $\neg A$ 互为对立事件，则 $p(\neg A)=1-p(A)$

性质 4　若 $A\subseteq B$，则 $p(B-A)=p(B)-p(A)$

性质 5　设 A,B 为两个任意事件，则 $p(A+B)=p(A)+p(B)-p(AB)$

由性质 1—5，我们还可以得到如下推论。

推论 1　若 $A\subset B$，则 $p(A)\leqslant p(B)$

推论 2　设 A,B,C 为三个任意事件，则 $p(A+B+C)=p(A)+p(B)+p(C)-p(AB)-p(AC)-p(BC)+p(ABC)$

推论 3 若 A,B 是两个互斥事件，则 $p(A+B)=p(A)+p(B)$

利用这些性质和推论可以帮助我们计算事件的概率。

例 1.1.3 甲、乙两人向同一目标射击，已知甲命中目标的概率为 0.6，乙命中目标的概率为 0.5，两人都命中的概率为 0.3，求目标被命中的概率。

解 设 A,B 分别表示甲，乙命中目标的事件，目标被击中即为 $A+B$。由于 A,B 两事件是相容的，所以

$$p(A+B)=p(A)+p(B)-p(AB)=0.6+0.5-0.3=0.8$$

例 1.1.4 某个聚会上有 n 个人（$n<365$），求这 n 个人中至少有两个人的生日是同一天的概率。

解 此试验有 $(365)^n$ 个可解结果，并且所有的结果是等可能的。

设 A 表示“至少两个人的生日在同一天”。直接计算 A 包含的样本点数相当困难，我们考虑逆事件 $\neg A$。

$\neg A$ 表示“任何两个人的生日都不在同一天”，那么 $\neg A$ 包含的样本点有 $p_{365}^n=365(365-1)\cdots(365-n+1)$ 个，由性质 3 可得

$$p(A)=1-p(\neg A)=1-p_{365}^n/(365)^n$$

3. 条件概率

前面讨论事件 B 的概率 $p(B)$ 时，都是指 $p(B)$ 为无附加条件概率。在实际问题中，除了考虑 $p(B)$ 外，有时还需考虑“事件 A 已发生”这一附加条件，我们将 A 发生的条件下事件 B 的条件概率记为 $p(B/A)$。

定义 1.1.7 设 A,B 为随机试验 E 的两个事件，且 $p(A)>0$，称

$$p(B\mid A)=\frac{p(AB)}{p(A)}\qquad (p(A)>0)$$

为事件 A 发生条件下事件 B 的条件概率。

例 1.1.5 袋中有两个白球三个黑球。从袋中取两个球，问两个都是白球的概率是多少？

解 设 B 表示事件“第一个是白球”，A 表示事件“第二个是白球”。

$p(B)=2/5$。取了一个白球后，剩下四个球，其中一个为白球，这样 $p(A|B)=1/4$。利用条件概率公式 $p(AB)=p(B)\cdot p(A|B)=2/5\times 1/4=1/10$。

根据条件概率的定义，可以知道它有下面的性质：

1) $p(B|B)=1$；

2) 若 A 与 B 互斥，则 $p(A|B)=0$；

3) $p(A|\Omega)=p(A)$；

4) 若 $p(BC)>0$，则对任意 $A\in F$ 有 $p((A|C)|B)=p(A|BC)$。

性质 1),2),3) 都是显然的，下面我们证明性质 4)。

证明 因为 $B\supset BC$，所以 $p(B)\geqslant p(BC)>0$，即 $p(B)$ 和 $p(BC)$ 可以作为分母。

$$p((A\mid C)\mid B)=\frac{p(AC\mid B)}{p(C\mid B)}=\frac{p(ABC)}{p(B)}\Big/\frac{p(BC)}{p(B)}=\frac{p(ABC)}{p(BC)}=p(A\mid BC)$$

□

我们可以把条件概率的定义推广到多个事件的情况，即多个事件的乘法公式：

$$p(A_1A_2\cdots A_n) = p(A_1)\cdot p(A_2 \mid A_1)\cdot p(A_3 \mid A_1A_2)\cdots p(A_n \mid A_2\cdots A_{n-1})$$

证明　$p(A_1A_2)=p(A_1)\cdot p(A_2|A_1)$

$$p(A_1A_2A_3) = p(A_1A_2)\cdot p(A_3 \mid A_1A_2) = p(A_1)\cdot p(A_2 \mid A_1)\cdot p(A_3 \mid A_1A_2)$$

……

如此类推，立刻可得上述乘法公式。　□

例 1.1.6　袋中有三个黑球两个白球。随机取出排成一行，其位置记为 1,2,3,4,5。问头一个与最后一个为白的概率。

解　以 W 代表白球，B 代表黑球，例中要求的是事件 $W_1B_2B_3B_4W_5$ 的概率。

$$p(W_1B_2B_3B_4W_5) = p(W_1)\cdot p(B_2 \mid W_1)\cdot p(B_3 \mid W_1B_2)\cdot p(B_4 \mid W_1B_1B_2)\cdot p(W_5 \mid W_1B_1B_2B_3)$$

取第一个球为白球的概率 $p(W_1)=2/5$；取了白球之后剩下的球中取为黑球的概率 $p(B_2|W_1)=3/4$；同理可得 $p(B_3|W_1B_2)=2/3$，$p(B_4|W_1B_2B_3)=1/2$，$p(W_5|W_1B_2B_3B_4)=1$。于是 $p(W_1B_2B_3B_4W_5)=2/5\times3/4\times2/3\times1/2\times1=1/10$。

4. 全概率公式与贝叶斯(Bayesian)公式

对一个随机试验，其结果发生的可能性有多种原因，每一原因对该结果的发生有一定的“影响”。当计算这类比较复杂的事件的概率时，往往需要同时利用概率的加法公式与乘法公式，在此基础上我们给出全概率公式。

定义 1.1.8　设(Ω,F,p)为概率空间，如果 $A_i\in F(i=1,2,\cdots,n)$，$A_iA_j=\varnothing(i\neq j)$，且 $\sum_{i=1}^{n}A_i=\Omega$，称 $A_1,A_2,\cdots,A_n$ 为Ω 的一个有限划分。

定理 1.1.1　设(Ω,F,p)为一概率空间，$A_1,A_2,\cdots,A_n$ 是Ω 的一个有限划分，且 $p(A_i)>0(i=1,2,\cdots,n)$，则对任意事件 $B\in F$，有 $p(B)=\sum_{i=1}^{n}p(A_i)\cdot p(B|A_i)$，这个公式被称为全概率公式。

证明　因为 $A_i(i=1,2,\cdots,n)$相互独立，所以 $B=\bigcup_{i=1}^{n}B\cap A_i$。

$$p(B)=p(BA_1)+p(BA_2)+\cdots+p(BA_n)$$
$$=p(A_1)p(B \mid A_1)+p(A_2)\cdot p(B \mid A_2)+\cdots+p(A_n).\,p(B \mid A_n)\ \square$$

例 1.1.7　袋内装有十个乒乓球，其中七个新球，三个旧球。比赛时任意取走两球，问从剩下的球中任取一个，它是新球的概率。

解　设 A 表示从剩下的球中任取一个是新球的事件，B_1 表示取出的两球皆新球，B_2 表示取出的两球是一新一旧，B_3 表示取出两球都是旧球。我们有

$$A = AB_1 \cup AB_2 \cup AB_3$$

$$p(A) = p(B_1)\cdot p(A_1 \mid B_1)+p(B_2)\cdot p(A \mid B_2)+p(B_3)\cdot p(A \mid B_3)$$

由古典概率可知

$$p(B_1) = C_7^2/C_{10}^2, p(A \mid B_1) = 5/8$$

$$p(B_2)=C_7^1\cdot C_3^1/C_{10}^2, p(A\mid B_2)=6/8$$
$$p(B_3)=C_3^2/C_{10}^2, p(A\mid B_3)=7/8$$
$$p(A)=\frac{5}{8}\times\frac{21}{45}+\frac{6}{8}\times\frac{21}{45}+\frac{7}{8}\times\frac{3}{45}=7/10$$

在实际问题中还会碰到这样一类问题：已知随机试验的某一结果是由许多“原因”导致的，每个“原因”导致这个结果发生的可能性有多大？解决这类问题的方法就是应用贝叶斯公式。

定理 1.1.2 设 $\Omega=\bigcup_{i=1}^{n}A_i$，其中 $A_i\cap A_j=\phi(i\neq j)$，$p(A_i)>0, i=1,2,\cdots,n$，对任何事件 B(设 $p(B)>0$)，则

$$p(A_i\mid B)=\frac{p(A_i)\cdot p(B\mid A_i)}{\sum_{j=1}^{n}p(A_j)\cdot p(B\mid A_j)}$$

证明 对固定的 A_i，$p(A_i\mid B)=p(A_iB)/p(B)\cdots$

$$p(A_iB)=p(A_i)\cdot p(B\mid A_i)$$
$$p(A_i\mid B)=\frac{p(A_i)\cdot p(B\mid A_i)}{p(B)}$$

将全概率公式代入 $p(B)$得

$$p(A_i\mid B)=\frac{p(A_i)\cdot p(B\mid A_i)}{\sum_{j=1}^{n}p(A_j)\cdot p(B\mid A_j)}$$ □

例 1.1.8 有三个袋，记为 1,2,3。它们分别装有一个白球、两个黑球、三个红球，两个白球、一个黑球、一个红球和四个白球、五个黑球、三个红球。随机抽取一个袋，并从中取两个球，它们是一红一白。问它们取自第二个袋的概率是多少？

解 用 A_1,A_2,A_3 分别表示抽选了 1,2 和 3 号袋子。

$$p(A_1)=p(A_2)=p(A_3)=1/3$$

用 B 表示事件“取一个白球一个红球”。

$$p(B\mid A_1)=\frac{C_1^1\cdot C_3^1}{C_6^2}=1/5$$
$$p(B\mid A_2)=\frac{C_2^1\cdot C_1^1}{C_4^2}=1/3$$
$$p(B\mid A_3)=\frac{C_4^1\cdot C_3^1}{C_{12}^2}=2/11$$
$$p(A_2\mid B)=\frac{p(A_2)\cdot p(B\mid A_2)}{p(A_1)\cdot p(B\mid A_1)+p(A_2)\cdot p(B\mid A_2)+p(A_3)\cdot p(B\mid A_3)}$$
$$=\frac{\frac{1}{3}\times\frac{1}{3}}{\frac{1}{3}\times\frac{1}{5}+\frac{1}{3}\times\frac{1}{3}+\frac{1}{3}\times\frac{2}{11}}=55/118$$

5. 事件的独立性

条件概率反映了某一事件 A 对另一事件 B 的影响，一般说来 $p(B)$ 与 $p(B|A)$ 是不相等的，但在某些情况下，事件 A 的发生与否对某件 B 不产生影响，我们看下面的例子。

例 1.1.9　设 100 件产品中有 4 件是次品，我们抽取一件然后放回去，再抽取一件，求两件都是合格品的概率。

解　设 A 表示“第一次取到合格品”，B 表示“第二次取到合格品”，A,B 的概率可以表示为

$$p(A)=96/100,\quad p(B\mid A)=96/100,\quad p(AB)=96/100\times 96/100=0.922$$

由于是“放回抽取”，第二次抽取的条件与第一次抽取时完全相同，也就是说第一次抽取的结果完全不影响第二次抽取。

定义 1.1.9　设(Ω,F,p)为一概率空间，事件 $A\in F,B\in F$ 且 $p(A)>0$，若 $p(B|A)=p(B)$，则称事件 B 独立于事件 A。

关于事件的独立性有如下性质：

1) 若事件 B 独立于 A，且 $p(A)>0,p(B)>0$，则事件 A 也独立于事件 B；

2) 设(Ω,F,p)为一概率空间，$A\in F,B\in F$，且 $p(A)>0,p(B)>0$，则事件 A 与 B 相互独立的充要条件是 $p(AB)=p(A)\cdot p(B)$；

3)若事件 A 与 B 相互独立，则下列三对事件$(A,\neg B),(\neg A,B),(\neg A,\neg B)$分别相互独立。

我们注意到事件的独立性与事件的互斥，相容等概念是完全不同的。事件的互斥、相容仅说明样本空间中的两个子集不相交或相交，它们属于集合的概念，而独立性要涉及到概率，不是集合概念。

例 1.1.10　掷两个骰子，A 表示两个骰子点数之和为 7 的事件，B 表示第一个骰子为 4 点的事件，C 表示第二个骰子为 3 点的事件。判定 A 与 B，A 与 C 及 A 与 BC 的独立性。

解　易知 $p(B)=p(C)=1/6$。

掷两个骰子共 36 个样本点，可知 $p(A)=1/6$。

$$p(AB)=p(AC)=1/36,\quad p(AB)=p(A).p(B),\quad p(AC)=p(A)\cdot p(C)$$

所以 A 与 B，A 与 C 相互独立。

由于 $p(A|BC)=1\neq p(A)$，故 A 与 BC 不独立。

两个事件相互独立的概念可以推广到任意有限个事件的情况。

定义 1.1.10　设(Ω,F,p)是概率空间，$A,B,C,\in F$。若

$$p(AB)=p(A)\cdot p(B),\quad p(AC)=p(A)\cdot p(C)$$

$$p(BC)=p(B)\cdot p(C),\quad p(ABC)=p(A)\cdot p(B)\cdot p(C)$$

则称 A,B,C 三事件相互独立。若仅有前三个式子成立，只称 A,B,C 三事件两两独立。

定义 1.1.11　设 $A_1,A_2,\cdots,A_n$ 是同一概率空间中的 n 个事件。如果其中任意 k 个$(k=2,3,\cdots,n)$事件交的概率等于各件概率之积，则称这 n 个事件相互独立。

这个定义中要求任意 $1\leqslant i<j<k<\cdots\leqslant n$ 都有

$$p(A_iA_j) = p(A_i)p(A_j)$$
$$p(A_iA_jA_k) = p(A_i)p(A_j)p(A_k)$$
$$\cdots\cdots$$
$$p(A_1A_2\cdots A_n) = p(A_1)p(A_2)\cdots p(A_n)$$

其中第一行代表 C_n^2 个等式，第二行代表 C_n^3 个等式，……最后一行只有一个等式。

n 个事件相互独立并有如下性质：

若 $A_1, A_2, \cdots, A_n$ 相互独立，则

$$p(A_1A_2\cdots A_n) = p(A_1)p(A_2)\cdots p(A_n)$$
$$p(A_1 \cup A_2 \cdots \cup A_n) = 1 - p(\neg A_1) \cdot p(\neg A_2)\cdots p(\neg A_n)$$

例 1.1.11 一个元件能正常工作的概率称为它的可靠性。元件组成系统，系统有系统的可靠性。现有 $2n$ 个元件构成如下并串联系统（见图 1.1.1）。如果每个元件可靠性均为 $r(0<r<1)$，并且各元件能否正常工作的事件是独立的，试求整个系统的可靠性。

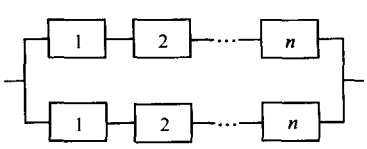


图 1.1.1 元件串并联系统

解 设 A_1, A_2 分别表示两条道路正常工作的事件，而每条通路正常工作必须 n 个元件正常工作，于是

$$p(A_1) = p(A_2) = r^n$$

整个系统正常工作的事件 $A = A_1 \cup A_2$。由于 A_1, A_2 独立，我们有

$$p(A) = 1 - p(\neg A_1) \cdot p(\neg A_2) = 1 - (1 - r^n)^2 = r^n(2 - r^n)$$

这说明比只有一条道路（其可靠性为 r^n）的可靠性提高了。

以上关于独立性的定义对于条件概率仍然适用。

定义 1.1.12 设 A_1, A_2, B 是同一概率空间中的事件，且 $p(B)>0$。如果

$$p(A_1A_2 \mid B) = p(A_1 \mid B) \cdot p(A_2 \mid B)$$

则称 A_1, A_2 关于 B 是条件独立的。

条件独立还可以有下面的表达形式：

1)
$$p(A_1A_2B) = \frac{p(A_1B)p(A_2B)}{p(B)}$$

由条件概率可知

$$p(A_1A_2B) = p(B) \cdot p(A_1A_2 \mid B)$$

由条件独立性得到

$$p(A_1A_2 \mid B) = p(A_1 \mid B) \cdot p(A_2 \mid B)$$

于是

$$p(A_1A_2B) = p(B) \cdot p(A_1 \mid B) \cdot p(A_2 \mid B)$$

因为

$$p(A_1 \mid B) = p(A_1B) \mid p(B), p(A_2 \mid B) = p(A_2B) \mid p(B)$$

所以

$$p(A_1A_2B) = \frac{p(A_1B)p(A_2B)}{p(B)}$$