Statistics and its application

统计学及其应用

赵高长 主编

西北工業大學出版社

统计学及其应用

主　编　赵高长
副主编　马继丰　杨　慧

西北工業大學出版社

【内容简介】 本书内容选材符合高等学校《工科硕士研究生统计学教学大纲》基本要求.内容包括参数估计、假设检验、相关分析、回归分析、方差分析、时间序列分析、Matlab 及 R 统计功能介绍，每章附有习题，书后有习题参考答案.

本书不仅可作为工科院校研究生统计学课程教材，也可作为经济学、管理学专业课程教材，还可供社会科学研究人员、管理及经济工作者使用.

图书在版编目（CIP）数据

统计学及其应用/赵高长主编.—西安：西北工业大学出版社，2013.2
ISBN 978-7-5612-3589-8

Ⅰ.①统…　Ⅱ.①赵…　Ⅲ.①统计学—研究生—教材　Ⅳ.①C8

中国版本图书馆 CIP 数据核字(2013)第 025829 号

出版发行：西北工业大学出版社
通信地址：西安市友谊西路 127 号　　邮编：710072
电　　话：(029)88493844　88491757
网　　址：www.nwpup.com
印 刷 者：陕西向阳印务有限公司
开　　本：727 mm×960 mm　1/16
印　　张：23.25
字　　数：433 千字
版　　次：2013 年 3 月第 1 版　　2013 年 3 月第 1 次印刷
定　　价：46.00 元

前　言

统计学是一门收集、整理、描述、显示和分析统计数据的科学，是一套关于由数据到结论的科学理论、方法和技术，也可以说统计学是关于在数据中学习的科学，目的在于探索数据内在的数量规律性或从中得到关于总体的和过程的结论. 本书强调了理论与实践的紧密结合，对重要的统计方法，都配有必要的实例，而且，无论是在理论阐述上还是在方法应用的实例上，力求贴近社会、贴近生活，尽量紧密结合我国的社会经济发展. 在编写过程中力求体系合理、内容新颖、方法简捷实用，突出统计方法的应用性，尽量避免烦琐的数学推导.

本书具备以下特点：

(1)原理阐述严谨、循序渐进，内容深入浅出，逻辑清晰，讲解到位. 对基础理论以“必需，够用”为原则，由浅入深进行阐述，以求简明易懂. 重点对应用理论及操作技能方面进行介绍，突出应用性和实践性.

(2)资料丰富、内容全面，结构完整，覆盖面广. 对统计描述、推断、回归分析、方差分析，时间序列等均有阐述，注重教材的整体性和衔接性. 相关内容已进行多轮教学讲解，并经不断修改完善，体系比较成熟.

(3)总结了以往的统计学教材特点，取长补短. 结合实际，每个知识点都带有对应的实际问题，操作性强，每个知识点后都有相关统计软件的具体操作过程及结论分析，所选例题具有代表性，能够反映相关章节的理论，达到复习及进一步理解的目的. 所有习题精挑细选，内容与结构层次丰富，能使学生得到必要的训练.

(4)案例多，问题多，引导读者思考，语言平实而又不失严谨. 通过丰富的案例，强化对知识点的牵引、辅助解说及综合考查作用，突出了教材的时代性、生活性和情趣性.

(5)关键名词都有英语解释，某些内容引进了数学背景，增强学生对概念来龙去脉的了解，从而提高学生的理解力.

(6)强化数学软件的使用. 不但将 Excel 统计学内容融入到每章的原理及实例中，在最后一章还给出了 Matlab 及 R 统计软件在统计学中的使用技巧，统计学原理及应用在教材中有机地结合在一起.

全书共分 8 章，第 1，2 章分别介绍了概率论及数理统计的基础知识；第 3，4 章介绍了参数估计与假设检验；第 5～7 章讲述了方差分析、相关分析、回归分析及时

间序列分析；第 8 章着重介绍了 Matlab 及 R 统计软件在统计学中的使用技巧.

本书主编为赵高长，副主编为马继丰、杨慧.具体编写分工如下：第 1～3 章及附录部分由赵高长执笔，第 4,5,7 章由杨慧执笔，第 6,8 章由马继丰执笔.最后由赵高长统稿定稿.在本书编写过程中，得到了西安科技大学褚维盘教授、丁正生教授、乔宝明教授的鼓励与支持，以及中国人民大学赵君等硕士的帮助，在此表示感谢！在本书编写过程中参考了大量的统计学书籍和资料，在此一并向有关作者表示感谢！

本书列为 2011 年度西安科技大学研究生教材建设项目.

因水平有限，书中一定存在缺点和疏漏之处，恳请广大读者批评、指正！

编　者

2012 年 10 月

目　录

第 1 章　概率论基础知识

统计学是阐述统计工作基本理论和基本方法的科学，是对统计工作实践的理论概括和经验总结. 它以现象总体的数量方面为研究对象，阐明统计设计、统计调查、统计整理和统计分析的理论与方法，是一门方法论科学.

《大英百科全书》对统计的定义是：统计学是收集、分析、表述和解释数据的科学.

统计工作、统计资料和统计学相互依存、相互联系，共同构成了完整的一个整体，这就是我们所说的统计.

由于统计推断是建立在概率论基础上的统计方法，因此本章对这方面的知识作基本介绍，浓缩通常概率论教材的主要内容，重点是基本概念与主要结论，部分证明从略. 本章为以后各章介绍一些近代方法提供必备的基础. 已学过一般概率论的读者可略过.

概率史界认为，帕斯卡与费马的通信标志着概率论的诞生，然而他们的通信直至 1679 年才完全公布于世，故惠更斯的《论赌博中的计算》标志着概率论的诞生. 因此，不少学者宣称惠更斯为概率论的正式创始人. 惠更斯的《论赌博中的计算》不仅是第一部概率论著作，而且是第一个把该学科建立在公理、命题和问题上而构成一个较完整的理论体系，第一次对以前概率论知识系统化、公式化和一般化的著作.

从 17 世纪中叶概率沦诞生至 1812 年，概率计算主要以代数组合方法为主，这一时期称为古典概率论——组合概率论——时期. 从 1812 年至 20 世纪初，概率论的研究主要采用分析方法，如特征函数、微分方程、差分方程等，这一时期称为分析概率论时期，开创了近代概率论时代，拉普拉斯总结了古典概率论，并使它发展到新的历史阶段，他的著作《分析概率论》就是承上启下的代表作. 1933 年以后，概率论的研究主要采用测度论方法，这一时期称为测度概率论时期，这时概率论公理化，开创了现代概率论的新时期.

概率论的发展史说明了理论与实际之间的密切关系. 许多研究方向的提出，归根到底是有其实际背景的. 反过来，在这些方向被深入研究后，又可指导实践，进一步扩大和深化应用范围. 正如拉普拉斯所说："生活中最重要的问题，其中绝大多数

在实质上只是概率的问题".

1.1 随机事件

1.1.1 基本概念

在自然界和现实生活中，根据一些事物是否有必然的因果联系，可以分成截然不同的两大类：一类是确定性现象，这类现象是在一定条件下，必定会导致某种确定的结果；另一类是不确定性现象，此类现象是在一定条件下，它的结果是不确定的.正因为这样，我们在不确定现象中，就无法用必然性的因果关系，对个别现象的结果事先做出确定的答案.事物间的这种关系是属于偶然性的，这种现象叫做**偶然现象**(accidental phenomenon)，或者叫做**随机现象**(random phenomenon).

随机现象从表面上看，似乎是杂乱无章的，没有什么规律的现象.但实践证明，如果同类的随机现象大量重复出现，它的总体就呈现出一定的规律性.大量同类随机现象所呈现的这种规律性随着观察的次数的增多而更加明显.对随机现象，在基本相同的条件下，重复进行试验或观察可能出现各种不同的结果，试验共有哪些结果事前是知道的，但每次试验出现哪一种结果却是无法预见的，这种试验称为**随机试验**(random experiment).随机试验的所有可能结果组成的集合称为**样本空间**(sample space)，随机试验的每一种可能结果称为**样本点**(sample point)，样本空间的子集称为**随机事件**(random event)，简称**事件**.样本空间Ω是**必然事件**(certain event)，空集$\varnothing$是**不可能事件**(impossible event).一次试验中，某事件A可能发生，也可能不发生，发生的可能性有大有小，这一可能性大小的数量指标就是所要研究事件的**概率**(probability).

1.1.2 随机事件的关系和运算

1.事件的包含关系

定义 1.1 设A,B是随机试验E的事件，若事件A发生必然导致事件B的发生，则称事件A包含于事件B或事件B包含事件A，记作$A\subset B$.即集合A是集合B的**子集**(subset).

2.事件的相等

定义 1.2 设A,B是随机试验E的事件，若$A\subset B$且$B\subset A$，则称事件A与事件B**相等**(equality)，记作$A=B$.

3.事件的和

定义 1.3 事件$A\cup B=\{x\mid x\in A$或$x\in B\}$称为事件A与事件B的**和事**

件(sum event). 当且仅当 A,B 中至少有一个发生时，事件 $A\cup B$ 发生. $A\cup B$ 有时也记为 $A+B$. 类似地，称 $\bigcup\limits_{i=1}^{m}A_i$ 为 m 个事件 $A_1,A_2,\cdots,A_m$ 的和事件，称 $\bigcup\limits_{i=1}^{\infty}A_i$ 为可列事件 $A_1,A_2,\cdots$ 的和事件.

4. 事件的差

定义 1.4　设 A,B 是随机试验 E 的两个事件，称"事件 A 发生，而事件 B 不发生"这一事件为事件 A 与事件 B 的**差事件**(difference of events)，记作 $A-B$. 即集合 $A-B$ 为集合 A 与集合 B 的差集. 也就是，从集合 A 中去除集合 B 与集合 A 公共元素后，剩下元素的全体构成的集合.

5. 事件的积

定义 1.5　设 A,B 是随机试验 E 的两个事件，称"事件 A 与事件 B 同时发生"这一事件为事件 A 与事件 B 的**积事件**(product of events)，记作 AB 或 $A\cap B$. 即集合 $A\cap B$ 为集合 A 与集合 B 的交集. $\bigcap\limits_{i=1}^{m}A_i$ 称为 m 个事件 $A_1,A_2,\cdots,A_m$ 的积事件，$\bigcap\limits_{i=1}^{\infty}A_i$ 称为可列事件 $A_1,A_2,\cdots$ 的积事件. 假定积运算优先级高于和运算.

6. 事件的互斥(互不相容)

定义 1.6　设 A,B 为随机试验 E 的事件，若在同一次试验中，事件 A 与事件 B 不能同时发生，则称事件 A 与事件 B **互斥**(mutual exclusion)或互不相容. 显然，若事件 A 与事件 B 互斥，一定有 $AB=\varnothing$，即事件 A,B 的积事件为不可能事件.

7. 事件的对立

定义 1.7　设 A,B 为随机试验 E 的事件，若在同一次试验中，事件 A 与事件 B 互斥，同时事件 A 与事件 B 的和事件为必然事件，则称事件 A 与事件 B **对立**(oppose).

8. 事件运算所满足的运算律

定理 1.1　(1) 交换律 $A+B=B+A, AB=BA$.

(2) 结合律 $(A+B)+C=A+(B+C),(AB)C=A(BC)$.

(3) 分配律 $A(B+C)=AB+AC, A+BC=(A+B)(A+C)$.

(4) 对偶律 $\overline{A+B}=\overline{A}\,\overline{B},\overline{AB}=\overline{A}+\overline{B}$.

(5) 差律 $A\overline{B}=A-B=A-AB$.

例 1.1　一批产品中有合格品也有废品，从中有放回地抽取(将产品取出一件观察后放回)3 件产品，以 $A_i(i=1,2,3)$ 表示第 i 次抽到废品，试以事件的集合表示下列情况：

(1) 第 1 次和第 2 次抽取中，至少抽到 1 件废品；

(2) 只有第 1 次抽到废品；

(3)3 次都抽到废品.

解 (1)A_1+A_2 (2)$A_1\overline{A_2}\,\overline{A_3}$ (3)$A_1A_2A_3$.

1.2 概率的定义及性质

1.2.1 概率的公理化定义

1.频率

定义 1.8 在相同的条件下进行了 m 次试验，若事件 A 发生了 n 次，则 n 称为 m 次试验中 A 发生的频数. $f_m(A)=\dfrac{n}{m}$ 称为事件 A 发生的**频率**(frequency). 频率具有以下性质。

定理 1.2 (1) 对任何事件 A，有 $0\leqslant f_m(A)\leqslant 1$.

(2) 对必然事件 Ω，有 $f_m(\Omega)=1$.

(3) 对 k 个两两互不相容事件 $A_1,A_2,\cdots,A_k$：

$$f_m(A_1)+f_m(A_2)+\cdots+f_m(A_k)=f_m(\bigcup_{i=1}^{k}A_i)=\sum_{i=1}^{k}f_m(A_i) \qquad (1-1)$$

2.概率的公理化定义

定义 1.9 设 Ω 是随机试验 E 的样本空间，如果对于 E 中的任一事件 A，都对应一个实数 $P(A)$，且 $P(A)$ 满足：

(1)(非负性) 对于随机试验 E 的任一事件 A，有 $0\leqslant P(A)\leqslant 1$.

(2)(规范性) 对于随机试验 E 的必然事件 Ω，$P(\Omega)=1$.

(3)(可列可加性) 对于随机试验 E 的两两互不相容事件 $A_1,A_2,\cdots$：

$$P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i) \qquad (1-2)$$

该实数 $P(A)$ 称为事件 A 的**概率**(probability).

3.概率的其他几种典型公理化定义

(1) 概率的统计定义.

定义 1.10 当 $m\to\infty$ 时，m 次独立重复试验的频率 $f_m(A)=\dfrac{n}{m}$ 趋于稳定值 $P(A)$，称 $P(A)$ 为事件 A 的概率.

(2) 概率的古典定义.

定义 1.11 若全体试验结果是 m 个，且是等可能的，而有利于事件 A 的试验结果有 n 个，则规定 $\dfrac{n}{m}$ 为事件 A 的概率，记作 $P(A)$.

(3) 概率的几何定义.

定义 1.12　若随机试验E的可能结果有无限(不可列)个,每个基本事件发生的可能性相等,则当样本空间Ω与所求事件A都可以用几何量(长度L、面积S或体积V)来测度时,A发生的概率称为**几何概率**(geometric probability)$P(A)$. $P(A)=\dfrac{S(A)}{S(\Omega)}$,其中$S(A)$,$S(\Omega)$分别为$A$,$\Omega$的几何测度.

1.2.2　概率的性质

定理 1.3　(1) $P(\varnothing)=0$.

(2) 若$A_1,A_2,\cdots,A_k$是随机试验E的两两互不相容的事件,则有

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

此性质也称为概率的有限可加性.

(3) 若A,B是随机试验E的两个事件,且$A\subset B$,则有

$$P(B-A)=P(B)-P(A),P(B)\geqslant P(A)$$

(4) 若A为随机试验E的任一事件,$\overline{A}$是A的对立事件,则有$P(\overline{A})=1-P(A)$.

(5) 若A,B是随机试验E的任意两个事件,则

$$P(A+B)=P(A)+P(B)-P(AB) \tag{1-3}$$

此性质也称为**概率的加法公式**(plus formula of probability).特别地,如果事件A,B互不相容,即$AB=\varnothing$,则加法公式可表示$P(A+B)=P(A)+P(B)$

概率加法公式可以推广到多个事件的情形,一般地,对任意n个事件A_1,$A_2,\cdots,A_n$,有

$$P(A_1+A_2+\cdots+A_n)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)-\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n) \tag{1-4}$$

例 1.2　袋中有 5 个白球和 3 个黑球,从袋中任取 2 个球,求:

(1) 取得同色 2 个球的概率;

(2) 取得的 2 个球至少有 1 个是白球的概率.

解　(1) 从 8 个球中任取 2 个的取法总数为$C_8^2=28$种.取得 2 个同色球,可分为两种情况:2 个球皆为白色或 2 个球皆为黑色.这两种情况分别有C_5^2,C_3^2种取法,故取得 2 个同色球的取法为$C_5^2+C_3^2=13$种,故所求概率为$\dfrac{13}{18}\approx 0.464\ 3$.

此题也可以这样解，令 $A_1=\{$取得的 2 个球皆为白色$\}$，$A_2=\{$取得的 2 个球皆为黑色$\}$，$A_3=\{$取得的 2 个球同色$\}$. A_1，A_2 互不相容，$A_3=A_1+A_2$，而

$$P(A_1)=\frac{C_5^2}{C_8^2}=\frac{10}{28},\quad P(A_2)=\frac{C_3^2}{C_8^2}=\frac{3}{28}$$

故
$$P(A_3)=\frac{10}{28}+\frac{3}{28}=\frac{13}{28}\approx 0.464\,3$$

(2) 令 $A=\{$取得的 2 个球中至少有 1 个白球$\}$，则 $\overline{A}=\{$取得的 2 个球皆为黑球$\}$，故

$$P(A)=1-P(\overline{A})=1-\frac{C_3^2}{C_8^2}=1-\frac{3}{28}=\frac{25}{28}\approx 0.892\,9$$

在 Excel 中，

permut(number，number_chosen)函数用于计算排列 $P_{\text{number}}^{\text{number\_chosen}}$；

combin(number，number_chosen)函数用于计算组合 $C_{\text{number}}^{\text{number\_chosen}}$.

1.2.3 条件概率、全概率公式及事件间的独立性

定义 1.13 设 A，B 是试验 E 的两个事件，且 $P(B)\neq 0$，则在事件 B 发生的条件下 A 发生的概率，称为事件 A 在条件 B 下的条件概率(conditional probability)，$P(A\mid B)=\dfrac{P(AB)}{P(B)}$.

定理 1.4 设 $A_1,A_2,\cdots,A_n$ 是随机试验 E 的一组事件，$A_iA_j=\varnothing$，$P(A_i)\neq 0$，$i,j=1,2,\cdots,n$，$\bigcup\limits_{i=1}^{n}A_i=\Omega$，则对 E 的任一事件 B 有

$$P(B)=\sum_{i=1}^{n}P(B\mid A_i)P(A_i) \tag{1-5}$$

此公式称为**全概率公式**(total probability formula).

定理 1.5 设 $A_1,A_2,\cdots,A_n$ 随机试验 E 的一组事件，$A_iA_j=\varnothing$，$P(A_i)\neq 0$，$i,j=1,2,\cdots,n$，$\bigcup\limits_{i=1}^{n}A_i=\Omega$，$B$ 是 E 的任一事件且 $P(B)\neq 0$，则

$$P(A_i\mid B)=\frac{P(A_iB)}{P(B)}=\frac{P(B\mid A_i)P(A_i)}{\sum\limits_{j=1}^{n}P(B\mid A_j)P(A_j)} \tag{1-6}$$

此公式称为逆概率公式(**贝叶斯公式**).

定义 1.14 设 A，B 是任意两个事件，如果 $P(AB)=P(A)P(B)$，则称事件 A 与事件 B 相互独立，简称独立.

定义 1.15 设 A，B，C 是任 3 个随机事件，如果

$$P(AB)=P(A)P(B),\quad P(BC)=P(B)P(C),\quad P(AC)=P(A)P(C)$$

则称事件 A,B,C 两两独立.

定义 1.16　设 A,B,C 是任三个随机事件,如果事件 A,B,C 两两独立且

$$P(ABC)=P(A)P(B)P(C)$$

则称事件 A,B,C 相互独立.

例 1.3　袋中有同型号产品 10 件,其中 3 件是次品,7 件是正品. 作不放回抽取,每次取一件,取两次. 令 $B=\{$第一次取到次品$\}$,$A=\{$第二次取到次品$\}$,求 $P(A\mid B)$.

解法一　取样本空间 $\Omega=\{\omega_1,\omega_2,\cdots,\omega_{10}\}$,$\omega_1,\omega_2,\omega_3$ 是次品,$\omega_4,\omega_5,\cdots,\omega_{10}$ 是正品. 令 $\Omega^*=\{\omega_i,\omega_j,\omega_4,\cdots,\omega_{10}\}$,其中,$i,j$ 是 1,2,3 中两个不同的数. Ω^* 可看成在 B 出现条件下的样本空间,它含 2 件次品,7 件正品,所以 $P(A\mid B)=\dfrac{2}{9}$.

解法二　取样本空间 $\Omega=\{(x_1,x_2)\}$,x_1 表示第一次抽得的产品,x_2 表示第二次抽得的产品. 第一次有 10 件产品供抽取. 第二次有 9 件产品供抽取. 所以,Ω 含 10×9 个基本点. 第一次有 3 件次品供抽取,而在第一次取得次品后,第二次有 2 件次品供抽取,所以,AB 含 3×2 个基本点,B 表示第一次抽到次品时,对第二次抽到次品或正品没限制. 所以,B 含 3×9 个基本点,由定义得

$$P(A\mid B)=\frac{P(AB)}{P(B)}=\left(\frac{3\times2}{10\times9}\right)\Big/\left(\frac{3\times9}{10\times9}\right)=\frac{2}{9}$$

在 Excel 中,

product(number1,number2,…)函数用于计算连乘 number1×number2×….

例 1.4　某人有 n 个朋友,其地址各不相同. 他给每个朋友各写一封信,然后任意放入 n 个写好地址的信封中,每个信封内放一封信.

(1) 求没有一封信放对的概率;

(2) 求恰有 r 封信放对的概率($r\leqslant n$).

解　用 A_i 表示第 i 封信放对,$i=1,\cdots,n$,则 $\bigcap\limits_{i=1}^{n}\overline{A}_i$ 表示没有一封信放对,得

(1) 没有一封信放对的概率 $P(\bigcap\limits_{i=1}^{n}\overline{A}_i)=P(\overline{\bigcup\limits_{i=1}^{n}A_i})=1-P(\bigcup\limits_{i=1}^{n}A_i)$.

由上述定理 1.3 加法公式,有

$$P(\bigcup_{i=1}^{n}A_i)=\sum_{i=1}^{n}P(A_i)-\sum_{1\leqslant i<j\leqslant n}P(A_iA_j)+\sum_{1\leqslant i<j<k\leqslant n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)$$

注意到

$$P(A_i)=\frac{1}{n},\ P(A_iA_j)=P(A_i)P(A_j\mid A_i)=\frac{1}{n}\frac{1}{n-1}=\frac{(n-2)!}{n!}$$

$$P(A_iA_jA_k)=P(A_i)P(A_i\mid A_j)P(A_k\mid A_iA_j)=\frac{1}{n}\frac{1}{n-1}\frac{1}{n-2}=\frac{(n-3)!}{n!}$$

…………

$$P(A_1A_2\cdots A_n)=\frac{1}{n!}$$

代入得

$$P(\bigcup_{i=1}^{n}A_i)=C_n^1\frac{1}{n}-C_n^2\frac{(n-2)!}{n!}+C_n^3\frac{(n-3)!}{n!}-\cdots+(-1)^{n-1}\frac{1}{n!}=$$

$$\frac{1}{1!}-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n-1}\frac{1}{n!}$$

$$P(\bigcap_{i=1}^{n}\bar{A}_i)=\frac{1}{0!}-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\cdots+(-1)^n\frac{1}{n!}=\sum_{k=0}^{n}\frac{(-1)^k}{k!}$$

当 n 充分大时，近似为 $e^{-1}\approx 0.37$.

(2) 由以上讨论知，某指定的 r 封信放对的概率为 $\frac{(n-r)!}{n!}$，其余 $n-r$ 封信都未放对的概率为 $\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}$，而从 n 封信中指定 r 封信有 C_n^r 种方式，所以，恰有 r 封信放对的概率为 $C_n^r\frac{(n-r)!}{n!}\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}=\frac{1}{r!}\sum_{k=0}^{n-r}\frac{(-1)^k}{k!}$，当 n 充分大时，近似为 $\frac{1}{r!}e^{-1}$.

在 Excel 中，

fact(number)函数用于计算阶乘 number!.

例 1.5 设某群体中有 37.5%的人的血型为 A 型，20.9%的人的血型为 B 型，33.7%的人的血型为 O 型，7.9%的人的血型为 AB 型. 允许输血的血型配对见表 1.1(√表示允许输血，×表示不允许). 现从该群体中任取一人为输血者，再任取一人为受血者，求允许输血的概率.

表 1.1 允许输血的血型配对表

受血＼输血	A	B	AB	O
A	√	×	√	√
B	×	√	√	√
AB	√	√	√	√
O	×	×	×	√

解　在受血者为A型的条件下,允许输血的概率为1－20.9%;在受血者为B型的条件下,允许输血的概率为1－37.5%;在受血者为AB型的条件下,允许输血的概率为1;在受血者为O型的条件下,允许输血的概率为33.7%.又知受血者为A型的概率为37.5%,为B型的概率为20.9%,为AB型的概率为7.9%,为O型的概率为33.7%,所以,由全概率公式得

$$P(\text{允许输血})=\frac{37.5}{100}\times\left(1-\frac{20.9}{100}\right)+\frac{20.9}{100}\times\left(1-\frac{37.5}{100}\right)+\frac{7.9}{100}\times1+\frac{33.7}{100}\times\frac{33.7}{100}\approx0.6198$$

例1.6　根据临床记录,某种诊断癌症的试验效果如下:

令$A=\{\text{被诊者有癌症}\}$,$B=\{\text{被诊者对试验呈阳性反应}\}$,则$P(B\mid A)=P(\overline{B}\mid\overline{A})=0.95$,又知$P(A)=0.005$,求$P(A\mid B)$.

解　取样本空间$\Omega=\{\text{被诊者}\}$,$P(B\mid\overline{A})=1-P(\overline{B}\mid\overline{A})=0.05$,由贝叶斯公式得

$$P(A\mid B)=\frac{P(A)P(B\mid A)}{P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})}=0.087$$

1.3　一维随机变量及其概率分布

1.3.1　随机变量的定义

定义1.17　设Ω是随机试验E的样本空间,若对每一个样本点$\omega\in\Omega$,有一个实数$X(\omega)$与之对应,这样得到的一个定义在样本空间上的实单值函数$X=X(\omega)$,称为一个**随机变量**(random variable)X.

一、离散型随机变量及其分布

定义1.18　离散型随机变量(discrete variable)$X(\omega)$只可能取有限个或可列个值,一般地,记这些值为

$$x_1,x_2,\cdots,x_n,\cdots$$

离散型随机变量取各个值的概率,也就是$P(X=x_i)=p_i,i=1,2,\cdots$,称为随机变量$X$的分布律.

常见离散型随机变量的分布律:

1.两点分布(0－1分布)

最简单的随机试验是只有两种可能结果的试验,称之为贝努利试验.一般的,把两个试验结果分别看做是"成功"与"失败",用数值"1"和"0"表示,用随机变量