

统计学完全教程

〔美〕L.沃塞曼 著

张 波 刘中华
魏秋萍 代 金 译



科学出版社
www.sciencep.com

现代数学译丛 5

统计学完全教程

〔美〕L. 沃塞曼 著

张 波 刘中华 译
魏秋萍 代 金

科学出版社

北 京

图字: 01-2007-3533

内 容 简 介

由美国当代著名统计学家 L·沃塞曼所著的《统计学完全教程》是一本几乎包含了统计学领域全部知识的优秀教材. 本书除了介绍传统数理统计学的全部内容以外, 还包含了 Bootstrap 方法(自助法)、独立性推断、因果推断、图模型、非参数回归、正交函数光滑法、分类、统计学理论及数据挖掘等统计学领域的新方法和技术. 本书不但注重概率论与数理统计基本理论的阐述, 同时还强调数据分析能力的培养. 本书中含有大量的实例以帮助广大读者快速掌握使用 R 软件进行统计数据分析.

本书适用于统计学、数学、计算机科学、机器学习与数据挖掘等领域的高年级本科生、研究生, 对于相关领域的广大科研工作者和实际工作者来说也不失为一本有价值的参考书.

Translation from the English language edition:

All of Statistics by Larry Wasserman

Copyright © 2004 Springer-Verlag New York, Inc.

Springer is a part of Springer Science+Business Media

All Rights Reserved

图书在版编目(CIP)数据

统计学完全教程/(美)沃塞曼 著. 张波等译. —北京: 科学出版社, 2008
(现代数学译丛; 5)

ISBN 978-7-03-021705-9

I. 统… II. ①沃…②张… III. 统计学—教材 IV. C8

中国版本图书馆 CIP 数据核字(2008)第 056121 号

责任编辑: 陈玉琢 房 阳/责任校对: 曾 茹

责任印制: 赵德静/封面设计: 王 浩

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>

源海印刷有限责任公司 印刷

科学出版社编务公司排版制作

科学出版社发行 各地新华书店经销

*

2008 年 6 月第 一 版 开本: B5 (720 × 1000)

2008 年 6 月第一次印刷 印张: 22 3/4

印数: 1—3 000 字数: 423 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换(明辉))

译者前言

统计学是一门数据分析科学,它有着漫长的发展历程.值得一提的是,在 20 世纪 20 至 30 年代,数理统计学的基本理论框架形成了,继而得到了快速的发展,数理统计学更加系统化、数学化.但是统计学的主要任务仍然是分析数据.计算机技术的发展和广泛应用改变了统计学的学科结构和研究方法.1979 年斯坦福大学教授 Efron 提出的基于计算机的统计推断技术 Bootstrap 方法就是一个很典型的例子.

21 世纪统计学的教育是一个很值得思考和研究的重大课题.一方面我们继续注重统计学的基本理论素质的培养,另一方面强调提高数据分析的实际能力.这两个方面缺一不可,互相促进.但是,现存的国内统计学教材则无法满足这两个要求.数理统计学方面的教材虽然理论较严谨,但是忽视了统计学的背景和应用.而介绍数据分析的教材则较欠缺理论基础.

由美国当代著名统计学家拉里·沃塞曼所著的这本教材恰恰可以同时满足上述两个要求,也可以解决目前国内统计学教材存在的一些不足.拉里·沃塞曼是美国卡内基-梅隆大学统计学系教授,他还是 1999 年度“考普斯”总统奖获得者.正如书名一样,本书包含了统计学领域几乎全部的知识,除了传统的数理统计教材中的内容外,还包含了诸如非参数回归、自助法、分类等统计学领域的新方法和技术.我们对《统计学完全教程》一书进行了认真的阅读和研究,认为它是一本优秀的教材和参考书,将其翻译成中文介绍给我国的广大读者.

本书的第一个主要特点是其适用面广.作为教材,本书适用于数学、统计学、计算机科学的高年级本科生以及统计学、计算机科学的研究生.它也适用于即将从事统计工作而又需要补充数理统计背景知识的毕业生.读者可以根据自己的时间和需要,有选择地学习相关内容.

本书的第二个主要特点是取材面广.它包含了统计学领域几乎全部的知识.第一部分讲述了概率论的基本知识,而且与通常的概率论教材不同的是,该部分强调在统计学里常用到的概率知识,如随机变量的收敛性中的 Delta 方法.第二部分的统计推断则涵盖了点估计、假设检验、分布函数的估计和统计泛函、Bootstrap(自助法)方法、参数推断及贝叶斯推断和统计决策理论.而第三部分则介绍了统计模型和方法,既有常见的回归和多变量模型,也有因果推断、图模型、非参数模型、光滑方法、分类、模拟技术等统计学的前沿课题.

本书的第三个主要特点是既注重概率统计基本理论的讲述,又强调数据分析能力的培养.本书所有的基本概念和原理的讲述是清晰的,完整的.而同时本书具有大量的实际的例子,这些例子的原始数据可以在作者的个人主页上下载,并且附有相应的 R 程序. R 是统计学家最钟爱的统计分析软件之一,而且是一款免费的开源软件.广大读者通过实际的数据例子不但可以学到数据分析方法,而且还可以加深对、

统计学基本概念和方法的理解. 如果将统计理论和数据分析能力比作人的两条腿, 那么这本书无疑将教会学生如何用“两条腿走路”, 这与我们的统计教育目标是吻合的.

为了保持原书的风格和特色, 在翻译的过程中, 我们保留了原书的所有栏目, 尽可能地忠实于原著, 由于本书内容涵盖面很广, 并涉及很多统计学前沿的内容, 很多统计学词汇还没有严格的中文翻译. 在翻译过程中, 我们尽量参考现存的中文翻译, 对于没有相应中文翻译的专业词汇, 我们请教相关专家, 力求将本书翻译好. 由于时间紧迫, 加上我们水平有限, 译文中一定有不尽如人意之处, 敬请读者不吝指正.

阅读本书只需要具备微积分和线性代数的基本知识, 不需要概率论和数理统计的相关知识. 因此, 对于那些想尽快掌握概率统计基础知识的读者而言, 本书是一本很好的入门教材. 又由于其内容的完备性和前瞻性, 本书可作为统计学、数学、计算机科学、机器学习和数据挖掘领域的高年级本科生、研究生的教材. 对于想了解概率统计方法, 尤其是想了解统计学前沿的实际工作者, 本书也不失为一本有价值的参考书.

本书由代金翻译第 1~4 章, 张波翻译第 5~8 章, 魏秋萍翻译第 9~16 章, 刘中华翻译第 17~24 章, 全书由张波统检并负责校译.

感谢在本书翻译与校对过程中给予我们支持和帮助的同仁吴喜之教授、刘畅副教授、殷红博士和王星博士.

译者

2008 年 3 月

于中国人民大学统计学院

原 书 序

从字面含义上讲,书名“统计学完全教程”有些夸大其词,但从本书的内容上讲,使用此书名也并无不妥,因为它比一般的数理统计介绍书籍涉及的面要广泛的多.

本书是为那些希望快速掌握概率和统计知识的读者而编写的.它适合计算机科学、数学、统计学和其他相关学科的研究生或优秀本科生.本书包括许多最新的课题,如非参数曲线估计、自助法、分类等,这些课题通常都有相关的后续课程.学习本书时,读者需要了解一些有关积分和线性代数的知识,不需要概率和统计的相关知识.

统计、数据挖掘和机器学习三者都关注收集和分析数据.曾经一段时间,统计研究在统计部门进行,而数据挖掘和机器学习研究在计算机科学部门设置,统计学家认为计算机科学家是在重复劳动,而计算机科学家认为统计理论没有应用于他们的问题.

事过境迁,时代也发生了变化,现在的统计学家都已经意识到计算机科学家作出了卓越的贡献而计算机科学家也意识到统计理论和方法的普遍性.适应性强的数据挖掘算法通常比统计学家的思想更具有预见性,而形式化的统计理论又比计算机科学家意识到的更具有普遍性.

从事数据分析的学生或者立志开发新型方法进行数据分析的学生需要掌握良好的基本概率和数理统计基础.没有理解最基本的统计含义就去使用如神经网络、提升算法、支持向量机等高级的工具就如同在没有学会如何使用邦迪就去做脑部手术一样.

但是学生从哪儿能快速学习最基础的概率和统计知识呢?我的答案是无处不在,至少当我的计算机学同事问我“我要将我的学生送到哪儿去快速掌握现代统计的知识?”的时候我是如此回答的.典型的数理统计课程在枯燥乏味的内容(计数方法,二维积分等)上消耗了很多时间,而真正讲解前沿课题(自助法,曲线估计,图模型等)的时间却少之又少,所以我决定重新设计本科生在概率和数理统计上的课程,本书就是以此为宗旨进行编写的.本书的主要内容如下:

1. 本书适用于计算机科学的研究生和数学、统计学、计算机科学高年级的本科生,它也适用于即将从事统计工作而又需要补充数理统计背景知识的毕业生.
2. 本书包括了初级教程中没有的前沿课题,如非参数回归、自助法、密度估计和图模型.
3. 本书删掉了在统计推断中无足轻重的内容,如计数方法.
4. 本书旨在强调基本概念,力图避免枯燥的计算.
5. 本书在讲述参数统计推断之前先讲述非参数统计推断.

6. 本书摒弃了通常的“第一学期 = 概率”，“第二学期 = 统计”的教育模式。在以前的教育模式下，一些学生仅仅学习了前半部分有关概率的知识，而不懂得任何统计理论的知识，则可能产生难以弥补的过错。另外，当学生能够将概率应用到后面的统计中的时候，概率才能真正体现它的内在价值。本书随机过程是例外，它将在后面的章节里面讲述。

7. 本书前后节奏变化快，包括了很多内容。我的同事开玩笑说我在本书中几乎包括了所有的统计知识，因此采用了这一书名。本书学习起来比较费时费力，但我尽力使内容比较直观具体。尽管节奏变化很快，但仍保证内容有血有肉、易于理解。

8. 严格与清晰并非一个含义，本书内容尽量保证两者的平衡性，使它们达到有机的统一。为避免在一些细枝末节上陷入困境，很多结果没有给出具体的证明。每章结尾的文献评述为读者列出其他合适的参考书。

9. 在我的个人主页上有 R 语言编码程序，学生可以用这些程序来完成所有的计算，具体网页是 <http://www.stat.cmu.edu/~larry/all-of-statistics>

但是，本书并未讲述 R 及其常用的计算机语言。

本书第 I 部分讨论概率理论，它是表示不确定性的正式用语，是统计推断的基础，在概率论中我们研究的基本问题是

给定数据的生成过程，其输出结果具有什么样的性质？

本书第 II 部分包括统计推断及其相关内容、数据挖掘和机器学习，统计推断的基本问题可看成是概率论的反向思维：

给定输出结果，我们能得出关于数据生成过程什么样的性质？

以上思想如图 1 所示。

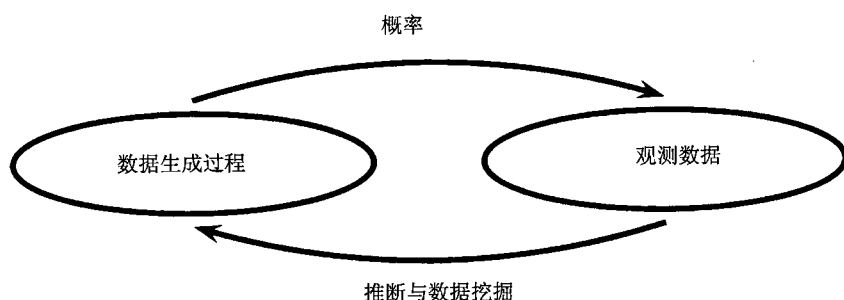


图 1 概率和统计推断的关系

预测、分类、聚类 and 估计都是统计推断的特殊情形。数据分析、机器学习和数据挖掘是统计推断在不同实践中的称谓，它依赖于具体使用的环境。

第 III 部分将第 II 部分的结论应用到具体问题中，如回归、图像模型、因果关系、密度估计、平滑、分类和模拟。第 III 部分也包括多章概率论的内容，具体内容

为随机过程,其中包括马尔可夫链.

本书的很多地方都引用了其他教材的内容,很多章节后面都有文献评述,一是为了向那些文献的作者表示由衷的感谢,其次是为了给读者提供有帮助的其他参考书. 在这里要特别提及的参考书是 (DeGroot and Schervish, 2002; Grimmett and Stirzaker, 1982), 本书从这两本参考书中引用了很多例子和练习.

花几年时间编写一本书,有些想法和课后练习的出处就失去了线索. 本书涉及的问题有一些是我自己构思的,有一些来源于平时的教育工作,还有一些摘自其他参考书. 如果未经作者的许可就擅自引用了他的内容,我希望没有冒犯对方. 就像我的同事 Mark Schervish 在他的书中 (Schervish, 1995) 写到:

“…… 每章后面的问题出自很多地方…… 有些我也不知道具体出自哪儿…… 如果我未经许可就使用了您的内容,请在此接受我的感谢.”

在这里我要感谢很多朋友,没有他们的帮助我不可能完成此书的编写. 首先要感谢使用本书早期版本并提供了很多反馈的广大学生. 要特别感谢 Liz Prather 和 Jennifer Bakal, 他们仔细阅读了这本书;要特别感谢 Rob Reeder, 他逐字逐句的阅读了这本书,并提供了无数宝贵的修改意见. 要特别感谢的还有 Chris Genovese, 他不仅对本书内容提供了很多创造性想法,还花了很多时间编写 Latex 程序,本书的版面设计归功于他的努力;由于本人所学有限,在一些格式上难免存在不足,请读者谅解. David Hand, Sam Roweis 和 David Scott 仔细阅读了本书并提供了大量改进的意见. John Lafferty 和 Peter Spirtes 也给了我很多启发. John Kimmel 对本书的编写过程作出了很大的贡献. 最后,我的妻子 Isabella Cerdinelli 给予我无限的关爱、支持和鼓励. 在此谨向他们致以深深的谢意.

L. 沃塞曼

宾夕法尼亚州 匹兹堡

2003 年 7 月

第二次印刷得益于下列人员的评价: Beth Ayers, Frederick Eberhardt, Luis Escobar, Warren Ewens, Pat Gray, John Lafferty, Valerie Ventura 和 Indrayana Rustandi.

统计/数据挖掘字典

统计学家和计算机科学家经常对相同的事情使用不同的用语, 如下这些术语将贯穿整本书的内容:

统计	计算机科学	含义
估计	学习	使用数据估计未知量
分类	有指导学习	由 X 预测离散的 Y
聚类	无指导学习	将数据分组
数据	训练样本	$(X_1, Y_1), \dots, (X_n, Y_n)$
协变量	特征	X
分类器	假设	协变量到结果的映射
假设	—	参数子集
置信区间	—	以给定频率包括空间 Θ 的未知量的一个区间
有向非循环图	贝叶斯网络	给定条件独立性下的多元分布
贝叶斯推断	贝叶斯推断	利用数据修正信度的统计方法
频率推断	—	保证频率行为的统计方法
大偏差界	PAC 学习	误差概率上的一致有界

目 录

译者前言

原书序

第 1 章	概 率	1
1.1	引言	1
1.2	样本空间和事件	1
1.3	概率	3
1.4	有限样本空间上的概率	4
1.5	独立事件	5
1.6	条件概率	7
1.7	贝叶斯理论	8
1.8	文献注释	9
1.9	附录	9
1.10	习题	9
第 2 章	随机变量	13
2.1	引言	13
2.2	分布函数和概率函数	14
2.3	一些重要的离散随机变量	18
2.4	一些重要的连续随机变量	20
2.5	二元分布	23
2.6	边际分布	25
2.7	独立随机变量	26
2.8	条件分布	27
2.9	多元分布与独立同分布 (IID) 样本	29
2.10	两个重要的多元分布	30
2.11	随机变量的变换	31
2.12	多个随机变量的变换	33
2.13	附录	34
2.14	习题	34
第 3 章	数学期望	37
3.1	随机变量的期望	37
3.2	期望的性质	39
3.3	方差和协方差	40
3.4	一些重要随机变量的期望和方差	41

3.5	条件期望	43
3.6	矩母函数	45
3.7	附录	46
3.8	习题	47
第 4 章	不等式	50
4.1	概率不等式	50
4.2	有关期望的不等式	52
4.3	文献注释	52
4.4	附录	53
4.5	习题	54
第 5 章	随机变量的收敛	55
5.1	引言	55
5.2	收敛的类型	55
5.3	大数定理	59
5.4	中心极限定理	59
5.5	Delta 方法	62
5.6	文献注释	63
5.7	附录	63
5.7.1	几乎必然收敛和 L_1 收敛	63
5.7.2	中心极限定理的证明	64
5.8	习题	65
第 6 章	模型、统计推断与学习	67
6.1	引言	67
6.2	参数与非参数模型	67
6.3	统计推断的基本概念	68
6.3.1	点估计	69
6.3.2	置信集	70
6.3.3	假设检验	72
6.4	文献注释	73
6.5	附录	73
6.6	习题	73
第 7 章	CDF 和统计泛函的估计	74
7.1	经验分布函数	74
7.2	统计泛函	76
7.3	文献注释	79
7.4	习题	80

第 8 章 Bootstrap 方法	81
8.1 随机模拟	81
8.2 Bootstrap 方差估计	82
8.3 Bootstrap 置信区间	83
8.4 文献注释	88
8.5 附录	88
8.5.1 刀切法 (Jackknife)	88
8.5.2 刀切法的百分位数置信区间	88
8.6 习题	89
第 9 章 参数推断	91
9.1 关注参数	91
9.2 矩估计	92
9.3 极大似然估计	93
9.4 极大似然估计的性质	96
9.5 极大似然估计的相合性	96
9.6 极大似然估计的同变性	98
9.7 渐近正态性	98
9.8 最优性	100
9.9 Delta 方法	101
9.10 多参数模型	102
9.11 参数 Bootstrap 方法	104
9.12 检验假设条件	104
9.13 附录	104
9.13.1 证明	104
9.13.2 充分性	106
9.13.3 指数族	109
9.13.4 计算极大似然估计	111
9.14 习题	114
第 10 章 假设检验和 p 值	117
10.1 Wald 检验	119
10.2 p 值	122
10.3 χ^2 分布	125
10.4 多项分布数据的 Pearson χ^2 检验	125
10.5 置换检验	127
10.6 似然比检验	129
10.7 多重检验	130

10.8	拟合优度检验	132
10.9	文献注释	133
10.10	附录	133
10.10.1	Neyman-Pearson 引理	133
10.10.2	t 检验	134
10.11	习题	134
第 11 章	贝叶斯推断	138
11.1	贝叶斯理论体系	138
11.2	贝叶斯方法	138
11.3	参数函数	141
11.4	随机模拟	142
11.5	贝叶斯过程的大样本属性	143
11.6	扁平先验、非正常先验和无信息的先验	143
11.7	多参数问题	144
11.8	贝叶斯检验	145
11.9	贝叶斯推断的优点和缺点	146
11.10	文献注释	150
11.11	附录	150
11.12	习题	150
第 12 章	统计决策理论	152
12.1	引言	152
12.2	比较风险函数	152
12.3	贝叶斯估计	155
12.4	最小最大规则	156
12.5	极大似然、最小最大和贝叶斯	158
12.6	容许性	159
12.7	Stein 悖论	161
12.8	文献注释	161
12.9	习题	161
第 13 章	线性回归和 Logistic 回归	163
13.1	简单线性回归	163
13.2	最小二乘和极大似然	166
13.3	最小二乘估计的性质	167
13.4	预测	168
13.5	多元回归	169
13.6	模型选择	171

13.7	Logistic 回归	174
13.8	文献注释	176
13.9	附录	176
13.10	习题	177
第 14 章	多变量模型	180
14.1	随机向量	180
14.2	相关系数的估计	182
14.3	多元正态分布	183
14.4	多项分布	183
14.5	文献注释	185
14.6	附录	185
14.7	习题	186
第 15 章	独立性推断	187
15.1	两个二值型变量	187
15.2	两个离散变量	190
15.3	两个连续变量	191
15.4	连续变量和离散变量	191
15.5	附录	192
15.6	习题	195
第 16 章	因果推断	196
16.1	反事实模型	196
16.2	超二值处理	200
16.3	观察研究和混淆	201
16.4	Simpson 悖论	202
16.5	文献注释	204
16.6	习题	204
第 17 章	有向图与条件独立性	205
17.1	引言	205
17.2	条件独立性	205
17.3	DAGs	206
17.4	概率与 DAGs	207
17.5	更多的独立性关系	208
17.6	DAGs 的估计	211
17.7	文献注释	212
17.8	附录	212
17.9	习题	215

第 18 章	无向图	218
18.1	无向图	218
18.2	概率与图	219
18.3	团与势	221
18.4	拟合图模型	222
18.5	文献注释	222
18.6	习题	222
第 19 章	对数线性模型	225
19.1	对数线性模型	225
19.2	图性对数线性模型	227
19.3	分层对数线性模型	229
19.4	模型生成元	230
19.5	拟合对数线性模型	231
19.6	文献注释	233
19.7	习题	233
第 20 章	非参数曲线估计	234
20.1	偏差 - 方差平衡	234
20.2	直方图	236
20.3	核密度估计	241
20.4	非参数回归	247
20.5	附录	251
20.6	文献注释	252
20.7	习题	252
第 21 章	正交函数光滑法	254
21.1	正交函数与 L_2 空间	254
21.2	密度估计	257
21.3	回归	261
21.4	小波	265
21.5	附录	270
21.6	文献注释	270
21.7	习题	270
第 22 章	分类	273
22.1	引言	273
22.2	错误率与贝叶斯分类器	274
22.3	高斯分类器与线性分类器	276
22.4	线性回归与 Logistic 回归	279

22.5	Logistic 回归与 LDA 之间的关系	281
22.6	密度估计与朴素贝叶斯	282
22.7	树	282
22.8	误差率评估与选择好的分类器	285
22.9	支持向量机	290
22.10	核方法	292
22.11	其他分类器	295
22.12	文献注释	297
22.13	习题	297
第 23 章	重温概率: 随机过程	299
23.1	引言	299
23.2	马尔可夫链	300
23.3	泊松过程	310
23.4	文献注释	313
23.5	习题	313
第 24 章	模拟方法	317
24.1	贝叶斯推断回顾	317
24.2	基本蒙特卡罗积分	317
24.3	重要抽样	321
24.4	MCMC 第一部分: Metropolis-Hastings 算法	324
24.5	MCMC 第二部分: 其他算法	327
24.6	文献注释	331
24.7	习题	331
参考文献		333
符号列表		337
名词索引		340

第 1 章 概 率

1.1 引 言

概率是描述不确定性的数学语言. 本章将介绍概率论的基本概念. 这里的介绍从样本空间入手, 样本空间指一切可能结果的集合.

1.2 样本空间和事件

样本空间 Ω 是某试验所有可能结果的集合. 样本空间 Ω 中的点 ω 称为**样本点**、**实现**或者**元素**. Ω 的子集称为**事件**.

1.1 例 假如将一枚硬币连续抛两次, 则 $\Omega = \{HH, HT, TH, TT\}$, 在抛第一次出现正面的事件是 $A = \{HH, HT\}$.

1.2 例 令 ω 表示某一物理量的测量结果, 如温度, 则 $\Omega = \mathbb{R} = (-\infty, +\infty)$. 因为温度是有下界的, 也许读者认为 $\Omega = \mathbb{R}$ 并不准确, 但将样本空间考虑得比实际范围大并没有什么坏处. 测量值大于 10 但小于等于 23 这一事件记为 $A = (10, 23]$.

1.3 例 假如永不停止地抛一枚硬币, 则样本空间就是无限集

$$\Omega = \{\omega = (\omega_1, \omega_2, \omega_3, \dots) : \omega_i \in \{H, T\}\}.$$

令 E 表示第一次正面朝上出现在第 3 次抛硬币这一事件, 则

$$E = \{(\omega_1, \omega_2, \omega_3, \dots) : \omega_1 = T, \omega_2 = T, \omega_3 = H, \omega_i \in \{H, T\}, i > 3\}.$$

给定事件 A , 令 $A^c = \{\omega \in \Omega : \omega \notin A\}$ 表示集合 A 的余集, 在非正式场合, A^c 可读作“非 A ”. Ω 的余集为空集 $\emptyset$.

A 与 B 的并集记为

$$A \cup B = \{\omega \in \Omega : \omega \in A \text{ 或 } \omega \in B \text{ 或 } \omega \in A \text{ 和 } B\}$$

并集可看成“ A 或者 B ”. 如果 $A_1, A_2, \dots$ 是一个集合序列, 那么

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \text{至少存在一个 } i \text{ 使得 } \omega \in A_i\}.$$

A 和 B 的交集定义为

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ 并且 } \omega \in B\}.$$