

 现代计算机科学与技术教材系列

算法设计与分析

(高级教程)

SUANFA SHEJI YU FENXI (GAOJI JIAOCHENG)



张德富 编著



国防工业出版社
National Defense Industry Press

现代计算机科学与技术教材系列

算法设计与分析

(高级教程)

张德富 编著



国防工业出版社

· 北京 ·

内 容 简 介

本书主要取材于反映当今计算机科学与技术学科中算法设计及分析发展潮流方面的内容。内容除包括国外一些比较成熟的算法技术,例如基本的随机算法以及近似算法,还包括一些最新的研究成果,例如基于近似和随机思想的混合算法:随机近似算法、在线算法、现代启发式算法等。本书包括大量的问题实例并给出了相应的求解方法。而工业应用领域的许多实际问题 and 疑难问题,都需要有效的求解算法,本书提供了大量的可供选择的解决途径。

本书可作为计算机科学系、数学系、管理科学等高级本科以及研究生课程的教材,也适合科研人员学习使用。

图书在版编目(CIP)数据

算法设计与分析高级教程/张德富编著. —北京:国防工业出版社,2007.2

(现代计算机科学与技术教材系列)

ISBN 7-118-04933-6

I. 算... II. 张... III. ①电子计算机—算法设计—高等学校—教材②电子计算机—算法分析—高等学校—教材 IV. TP301.6

中国版本图书馆 CIP 数据核字(2006)第 157769 号

※

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100044)

天利华印刷装订有限公司印刷

新华书店经售

*

开本 710×960 1/16 印张 14 1/4 字数 247 千字

2007 年 2 月第 1 版第 1 次印刷 印数 1—4000 册 定价 24.00 元

(本书如有印装错误,我社负责调换)

国防书店: (010)68428422

发行邮购: (010)68414474

发行传真: (010)68411535

发行业务: (010)68472764

《现代计算机科学与技术教材系列》

编审委员会

主 编	刘椿年	许满武		
执行副主编	陈国良	赵致琢		
编 委	王志英	康立山	刘大有	刘椿年
	许满武	宋方敏	林作铨	陈向群
	李师贤	陈国良	洪先龙	应明生
	陆汝占	傅育熙	赵致琢	唐常杰
	张明义	彭群生	朱 洪	周龙骧
	李立武	承恒达	李西宁	陈 湲
	阎瑞琪			

总 序

民众多好饮酒,中外概莫能外。酒馆和酿酒坊伴随饮酒客而起,人类对酒的喜爱造就了酒文化和一个庞大的产业。好酒能卖好价钱,能使文人诗兴大发,催生佳作,还能解人间百难。于是,酿天下名酒自然成为不少人的毕生追求。

怎样才能酿出好酒呢?国人的看法不尽相同。崇信洋酒的人主张引进国外的生产工艺,学习洋人的生产和经营理念,而喜欢国酒的人则主张走自己的路,但不排除借鉴国外先进的科学技术和管理经验。这样的争论或许永远不会终结,但外国人重视科学酿酒,这一点是值得我们学习和借鉴的。

计算机科学教育,如同酿酒工业的生产一样,科学办学迄今还只是部分学者的一种理想。与国内一样,国外的计算机科学教育并没有像他们的科学酿酒业一样,实现科学办学。也许科学办学要远比科学酿酒困难得多。譬如,怎么实现科学办学?甚至怎么推出一套科学的系列教材都是一篇大文章。

这套教材的创作始于教育部面向 21 世纪教育与教学改革 13-22 项目的研究。2000 年,在 13-22 项目研究工作即将完成之际,一些学者开始认识到面对计算机科学与技术的高速发展,我们亟需一套体现科学办学思想、反映内涵发展要求、服务教育与教学改革、参与构建学科人才培养科学体系的系列教材。强调系列教材是因为那时已经意识到计算机科学教育本质上是一项科学活动,但长期以来教师向学生传授科学技术知识的方式方法科学性不强。由于高等教育几百年来一直沿袭经验方式而非科学方式办学,大学教学的方式方法仍然还停留在古代作坊式的阶段,只不过今天使用的教学技术手段先进而已。在经验办学方式下,无论是研究型大学还是教学型大学,由于种种原因,教学活动的全过程存在着太多的漏洞和质量上的隐患。科学办学是对高等教育界传统的一个挑战,尽管在认识上,人们不难理解,科学办学是经验办学的最高形式,而经验办学应该成为科学办学的有益补充。

13-22 项目组积极探索,率先倡导科学办学理念,初步构建了一个体现科学办学思想,反映内涵发展要求的计算机科学一级学科人才培养科学理论体系,为学科专业教育探索新天地,走向科学办学和发展学科系列教材提供了一个认

知基础。

长期以来,学术界一直在探索计算机专业教育的规律。ACM 和 IEEE/CS 的专家小组在走访了全美 400 多位著名计算机科学家的基础上,以学科方法论作为切入点开展教学改革理论研究,于 1989 年发表了具有开创性意义的成果,尽管他们并未意识到自己的工作是以学科方法论的研究作为切入点,探讨内涵发展的道路。1990 年前后,在迷宫中探索行走的专家小组经大师和精英群的指点实际上已经摸到了走出迷宫的大门,却没有打开并进入一个崭新的天地。这一点从他们在 2000 年网上公布的 CC2001 报告最先删除了 CC1991 报告中有关学科方法论的内容便不难看出(注:后经中国人的提醒又补充写入)。

与此同时,中外教材建设也一直没有停止探索,国内外出版社先后出版了种类繁多的计算机类专业教材。这些教材中不乏精品和上乘之作,但难觅具有鲜明特色、真正一体化设计并且符合科学办学要求的系列教材。多数丛书和系列教材基本上还只是出版社对出自作者个人创作的教材通过冠名“丛书”或“系列”的方法结集出版以求强势效应。尽管如此,不少优秀作者和学者理所当然地进入了编审委员会的视线。西方发达国家在计算机科学学科的领先优势曾使许多人不自觉地将目光转向海外,试图从世界名牌大学使用的教材中去寻找蓝本。遗憾的是与国内一样,经验办学并没有使西方大学在教材建设方面摆脱“各自为政,各行其是”的阴影。此时,我们如梦初醒,毕竟科学办学是前无古人的一项创举。随着学科的不断深入,在迈向深蓝知识海洋的今天,外国人未必比中国人在科学办学方面占有更多天时地利的优势。不经意中的发现使我们惊喜和激动,同时深感责任重大且平添担忧:即使能够写出系列教材的一体化设计,我们是否真能确认这项改革的正确性?真能推出科学的系列教材?可是,除了实践和试验,我们别无捷径可循。

从 20 世纪 50 年代起,我国几代学者苦苦追赶了西方发达国家半个世纪,依靠引进、学习、跟踪、改进、创新的高新技术发展思维定式曾使我们付出了高昂的学费和沉重的代价。固然,在高新技术领域,依靠“引进”和“泊来”,取得了长足的发展和进步,填补了不少国内的“空白”,但在水准上始终与发达国家保持着一段距离,一种在行业内部看来时长时短、难于逾越的差距。这种差距主要表现在对高精尖学科的发展我们缺乏思想、概念、理论、方法、技术、制度、规范和设计的原始创新和发展模式的全面创新,研究工作总是跟在别人后面亦步亦趋。我们缺乏在发展中另辟蹊径,走自己道路的机制和氛围,迷信洋人,盲目追随西方学术发展道路的习惯思维方式几乎导致国人失去了创新的机能,这是一个国家和

民族发展高新技术学科和产业致命的硬伤。

高新技术领域竞争的成败,关键取决于人才与文化。现代科学技术的创新,已不单纯是一个学术问题,还是一个与文化、人文密切相关的问题。科学教育求真求是,技术教育求实求精,人文教育求灵求善,艺术教育求美求新。没有科学技术知识,人的认识和生活难免停留在原始社会,而没有人文精神和艺术的陶冶,科学技术的创新必然失去力量的源泉。可见,走自己的道路,发展中国的科技创新体系,在某种程度和意义上,成败的关键在于大学能否真正培养一大批高素质的人才。高等学校要实现培养大量高素质计算机专业人才培养的目标,需要在前进中不断地进行系统的、科学的总结和深刻的反省,需要对遇到的问题进行科学的分析和判断,作出正确的决策。

工欲善其事,必先利其器。倘若教师不能在思想上摆脱陈旧的思维定式,用先进的理念武装头脑,勇于探索前人没有走过的发展道路,那么,即使采用了世界一流大学的全套教材,恐怕也难于培养出世界一流的人才。中西文化、人文传统之间的差异之大,中外教育思想之间的差异之大使得中国教育的现代化决不是一个引进和模仿就可以轻易解决的问题。教师的职业不是贩卖知识。授业、传道、启蒙、解惑技能的高低,不仅取决于教师知识的广博和深厚,更重要的在于远见、卓识、探索、创新、敬业、求真的本领和身先垂范。

身处 21 世纪,面对国家的期望,处在科学技术发展浪潮之巅的计算机教师,任重道远。我们就像茫茫林海中的探险者,环顾苍翠的群山,犹如身陷迷宫一样。计算机科学教育,敢问路在何方?其实,我们的出路或许只有一条,那就是系统总结前人的经验和教训,设法努力登上山峰,居高眺望,探寻走出林海的希望之路。

2001 年初,《现代计算机科学与技术教材系列》编审委员会正式成立,一年后,计算机科学一级学科系列教材一体化设计研究报告的第一部分在《计算机科学》杂志第 6 期长篇发表。编审委员会为系列教材的出版制定了严格、详细的操作程序,并通过多种渠道,选择了第一批教材的作者。在体制创新方面本次教材出版设立了学术编审人,跟踪编审教材的创作内容,力求教材的尽善尽美。可以预期,《现代计算机科学与技术教材系列》将是一套基于计算机科学一级学科人才培养科学理论体系,体现内涵发展和科学办学要求,反映一体化设计的系列教材建设的一个尝试。然而,就像任何新生事物一样,她难免存在缺点和不足,我们诚恳地希望关心和使用本套系列教材的师生、读者,在使用中将批评或建议留下来,帮助改进教材建设工作,修正存在的不足之处。

一些学者对于一级学科人才培养科学理论体系的能行性表示怀疑。带着这个问题,在贵州大学和国内部分高等学校和一大批知名学者的支持下,从1999年夏天起在贵州大学连续举办了“计算机科学与技术高级研讨班”,向(博士)研究生和中青年教师陆续开设了研究生核心学位课程“高等计算机体系结构”、“并行算法设计基础”、“分布式算法设计基础”、“高等逻辑基础”和“形式语义学基础”,后来又进一步开设了本科生重点课程“算法设计与分析”、“数理逻辑基础”、“可计算性与计算复杂性”、“形式语言与自动机理论”等一系列课程。5年来,高级研讨班受到全国广泛关注、响应和支持,先后吸引了几十所大学三百多人次的师生参加听讲和学习,目前已经发展为由教育部批准资助、15所大学联合主办的高级研讨班。实践证明,高级研讨班为中国高等学校计算机科学教学改革和教育质量的提高正在发挥其独特的作用,并得到国内外一大批学者的充分肯定和好评。高级研讨班正在成为按照一级学科办学和教学改革的要求,对计算机科学系教师进行高起点、高标准、正规化研究生学位课程和本科重点课程培训的一个师资培训模式,有可能对未来计算机科学教育产生深远的影响。试想,如果高等学校的教师和培养的研究生普遍具有高级研讨班所开设的3门~4门课程的基础,不仅科学办学面临的主要困难迎刃而解,各大学科研学术队伍的素质也将得到显著提高。

一些学者对高起点研究生学位课程的必要性提出疑问:是否这些课程都要学习?我们认为,应该看到在高等教育界从来就存在着两种不同的教育观,一种是专才教育观,一种是通才教育观。持这两种教育观的人尽管都主张基础知识的重要性,但在对学以致用原则的理解和解释方面存在差异。一般地说,专才教育观主张在一定的基础上,通过深入钻研某一方向的学问,逐步扩展和加深自己的知识,缺什么基础补什么知识,学以致用,逐步成长为一个学科的专家。通才教育观则不同,它不主张在具备一定的基础后,就匆忙沿着某一方向钻研学问,单线独进,而是主张在一级学科的范围,通过尽可能系统地掌握从事本学科各个重要方向的研究所需要的基础知识,能够站在学科的各个至高点上,沿着学科的一个方向,以单线独进、多线并进或整体推进的观点,逐步扩展和加深一级学科的知识,融会贯通,学以致用,逐步成长为一个学科的专家。两种教育观都有其代表人物。迄今为止,高等教育中研究生教育主要以培养专才为主,专才教育观是主流。但是,两种不同的教育观各有其特点。一般地说,当一个学科的发展处于早期时,专才教育比较容易跟上学科的发展步伐,比较容易出成果,也比较容易迅速地达到较深的学术层面。而当一个学科的发展比较成熟,发展

速度比较平稳时,通才教育的优势就比较明显。因为,通才教育培养的人才可以在一级学科的范围内比较容易地向任何一个方向转向。特别是在胜任高难度重大创新人才的培养方面,在出综合性的大成果方面,在创立一套科学理论和开辟一个研究方向方面,通才教育的多种优势往往是专才教育所不具备的。当然,两种教育观谁优谁劣迄今并无定论,根据两种教育观的特点和现实情况,选择哪一种教育观实际仅反映了师生的一种选择策略。不过,实践告诉我们,尽管通才教育观的操作实现比较困难,但作为师资补充的来源,通才教育培养的人才更容易适应大学教学与科研的双重要求,理应更多地受到研究大学的青睐。在科学技术日益深化、高度分化又高度综合的今天,放眼未来,在高精尖学科中,通才教育观无疑有着更为宽广而美好的发展前景。

伴随着学科教学改革理论研究与实践探索的推进,社会热切地期待着一套与教学改革方案相配套的高质量系列教材问世。总结过去教材建设成功和失败的经验和教训,使我们清楚地认识到:教材建设必须建立在科学研究基础之上,按照科学的运作程序,动员在第一线从事科学研究、功底深厚、学有所长的优秀教师参与到教材的创作中来,才有可能推出高质量并符合学科发展要求的系列教材。我们的主张是:“让大学中的科学家来创作教材。”

冬去春来,年复一年。当我们终于从跟踪、学习、盲从西方大学教程的发展模式中走出时,感受到了一种从未有过的释然与激动,一种走自己的发展道路,独立自主的自豪与喜悦。这条道路虽然艰难,但前景光明。连续5年在贵州大学举办的全国计算机科学与技术高级研讨班的成功实践更进一步地坚定了我们内涵发展模式与科学办学之路的认识与追求。

今天,经过编审委员会、作者和出版社的共同努力,《现代计算机科学与技术教材系列》终于开始陆续出版发行。在新年即将到来的时候,我们怀着喜悦的心情,向祖国和人民,向计算机科学界献上一份完全由华夏学者按照他们对高等教育和计算机科学的理解与认识,倾力创作的新年贺礼,一份建立在科学研究基础之上的教育与教学改革最新成果。

新的世纪已经到来,愿《现代计算机科学与技术教材系列》的出版能够为新一代的莘莘学子攀登现代科学技术的高峰成就未来。

《现代计算机科学与技术教材系列》编审委员会
2006年10月

前 言

在工程技术、科学研究和经济管理的诸多领域,存在着大量的优化问题,特别是一些 NP 难问题,很难精确求解。当问题规模超过一定大小之后,即使超级计算机也无能为力。随着信息科学的发展,计算智能方面的问题日趋尖锐,成为实际应用的瓶颈,迫使计算机科学家和数学家寻找解决这类困难问题的新出路。本书介绍了三种有效的解决方案:随机算法、近似算法以及启发式算法。

本书深入浅出、通俗易懂地介绍了以下主要内容:

第 1 章简单介绍一些本书中需要用到的数学基础、问题的复杂性以及规划问题的一些预备知识。第 2 章重点介绍随机算法,其内容包括数值随机算法、Sherwood 算法、Las Vegas 算法和 Monte Carlo 算法的设计与分析。最后介绍如何扩大随机优势、随机复杂类的定义等内容。第 3 章重点介绍近似算法,其内容包括:基本定义,包括差界、相对性能界、多项式时间近似策略以及完全多项式时间近似策略。根据这些定义,可以衡量求解着色问题、调度问题、旅行商问题、覆盖问题、装箱问题、背包问题等的近似算法的性能。然后介绍了随机近似算法以及基于线性规划问题的近似算法及其应用。在 NP 难解性理论初步的基础上,进一步深入介绍近似的难度,最后介绍了在线算法。第 4 章重点介绍基于近似与随机思想的启发式算法的设计与分析,特别是现代启发式算法,例如模拟退火、禁忌搜索以及遗传算法等的应用实例。

本书写作过程中,得到了赵致琢教授关心,是他鼓励和支持我写作本书,并给出了许多宝贵的建议。本书内容曾在厦门大学计算机系、软件学院 2004 级和计算机系 2005 级研究生中试用,这些研究生为本书内容的完善提出了一些有益的建议。还有我的研究生刘艳娟、陈胜达、魏丽军以及陈青山等对本书进行了校对工作。本书的出版还得到了国防工业出版社辛再甫主任以及宋序一编辑的关心和支持。在此对他们一并表示感谢。我还要特别感谢我的妻子罗淀,对书稿中的文字进行了录入和校对,她还建议书不要写得太深奥复杂,这奠定了本书的风格。没有她的帮助,本书不可能按时完成。

X

限于水平和经验,书中难免有不妥和错误之处,还望广大读者指正,在此先表感谢。

谨以此书献给所有关心、鼓励和帮助过我的人们。

张德富

厦门大学计算机科学系

hustzdf@sina.com

2006年8月10日

目 录

第 1 章	预备知识	1
1.1	数学基础	1
1.1.1	常用的数学符号及公式	1
1.1.2	概率基础	1
1.2	问题的复杂性	5
1.2.1	最优化问题	5
1.2.2	计算复杂性理论	6
1.2.3	处理 NP 完全问题	8
1.2.4	算法性能的评估	9
1.3	规划问题	11
1.3.1	线性规划问题	12
1.3.2	整数规划问题	15
第 2 章	随机算法	17
2.1	基本概念	17
2.1.1	随机数发生器	17
2.1.2	随机算法	19
2.1.3	随机算法的复杂性分析	22
2.1.4	随机算法的应用	24
2.2	数值随机算法	26
2.2.1	Buffon 针	27
2.2.2	数值积分	29
2.3	Sherwood 算法	32
2.3.1	随机快速排序	33
2.3.2	随机选择	35
2.3.3	随机采样	38
2.4	Las Vegas 算法	41
2.4.1	基本概念	41
2.4.2	8 皇后问题	43

2.4.3	大数因子分解	46
2.5	Monte Carlo 算法	50
2.5.1	基本概念	50
2.5.2	主元素问题	51
2.5.3	最小切割问题	53
2.5.4	素数测试	58
2.5.5	模式匹配	62
2.5.6	扩大随机优势	64
2.6	随机复杂性	68
2.7	总结	70
第 3 章	近似算法	72
3.1	基本概念	72
3.1.1	基本概念	72
3.1.2	设计近似算法	76
3.2	调度问题	76
3.2.1	并行机器调度问题描述	77
3.2.2	$R(I)$ —近似算法	77
3.2.3	PTAS	81
3.3	旅行商问题	83
3.3.1	问题描述	83
3.3.2	R —近似算法	84
3.3.3	$R(I)$ —近似算法	89
3.4	覆盖问题	94
3.4.1	问题描述	94
3.4.2	R —近似算法	96
3.4.3	$R(I)$ —近似算法	97
3.5	Bin packing 问题	101
3.5.1	问题描述	102
3.5.2	R —近似算法	102
3.5.3	结论	107
3.6	背包问题	107
3.6.1	问题描述	108
3.6.2	动态规划	108
3.6.3	R —近似算法	110

3.6.4	PTAS	112
3.6.5	完全多项式时间近似策略	114
3.6.6	子集和问题的 FPTAS	119
3.7	随机近似算法	123
3.7.1	基本概念	123
3.7.2	可满足性问题(Satisfiability)	124
3.7.3	R—近似算法	125
3.8	基于线性规划的近似算法	127
3.8.1	理论基础	127
3.8.2	凑整技术的应用	129
3.8.3	混合近似算法	137
3.8.4	主对偶技术	139
3.9	近似的难度	142
3.9.1	绝对近似算法的消极结果	143
3.9.2	相对近似算法的消极结果	144
3.9.3	PTAS 的消极结果	145
3.9.4	FPTAS 的消极结果	146
3.10	在线算法	149
3.10.1	基本概念	149
3.10.2	在线算法的性能	150
3.10.3	随机在线算法	150
3.10.4	应用	152
3.11	总结	157
第 4 章	启发式算法	159
4.1	概述	159
4.1.1	启发式算法的分类及构造	159
4.1.2	常用启发式算法	160
4.1.3	启发式算法的性能分析	170
4.2	作业车间调度问题	171
4.2.1	问题描述	172
4.2.2	启发式算法	173
4.2.3	转换瓶颈算法	178
4.3	packing 问题	183
4.3.1	矩形 packing 问题的描述	183

4.3.2	启发式递归算法	184
4.3.3	GA + RP	189
4.3.4	圆形 packing 问题的描述	192
4.3.5	拟人退火算法	194
4.4	SAT 问题	201
4.4.1	SAT 问题描述及转换	201
4.4.2	拟人退火算法	202
4.4.3	改进的禁忌搜索算法	205
4.5	总结	210
4.5.1	启发式算法的特点	210
4.5.2	观点与思考	211
参考文献		212

第1章 预备知识

1.1 数学基础

1.1.1 常用的数学符号及公式

在分析一个算法的时间复杂度或者近似比时,经常要用到下面几种技术来估计它们。

1. 函数的底和顶

对任意给定的实数 x , 有

$$x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1$$

2. 幂和对数

对任意的 $|x| \leq 1$, 有

$$1 + x \leq e^x \leq 1 + x + x^2$$

对任意的 x , 有

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

对任意的 $x > -1$, 有

$$\frac{x}{1+x} \leq \ln(1+x) \leq x$$

3. 调和级数

$$\ln(n+1) \leq H(n) = \sum_{k=1}^n \frac{1}{k} \leq \ln n + 1$$

1.1.2 概率基础

在分析一个算法的性能时,通常要用到概率,下面介绍一下有关概率的知识。

为定义概率,需要样本空间,而样本空间是由一个个的基本事件所组成。每个基本事件可以看作一次试验的结果。一个离散随机变量 X 是一个函数: $S \rightarrow$

R , 其中 S 是一个有穷的或者是可数的无穷样本空间, 它对每个试验结果都赋予一个实数值。一般情况下, 可以将随机变量看成一个有随机输出的试验。随机变量值即为随机试验的观察输出。对于一个随机变量 X 和一个实数 x , 我们定义事件 $X = x$ 为 $\{s \in S: X(s) = x\}$, 因而该事件发生的概率可以定义为 $\Pr\{X = x\} = \sum_{\{s \in S: X(s) = x\}} \Pr\{s\}$ 。

如果样本空间有穷, 且每个基本事件 $s \in S$ 有概率

$$\Pr\{s\} = \frac{1}{|S|}$$

则得 S 上的一致概率分布。一般地, 随机地选择 S 中的元素就表示一致概率分布下的试验。特别地, 假定给定一个样本空间 S 和一个事件 s , 则关于事件 s 的二元随机变量 $I\{s\}$ 可以定义如下: 如果事件 s 发生, 则 $I\{s\} = 1$, 否则 $I\{s\} = 0$ 。

在有些情况下, 我们对某个试验的结果会有一些先验性的知识, 这就需要用到条件概率, 条件概率就是对先验知识的思想形式化的结果。比如, 我们猜 3 张牌, 在猜一次并且记住该牌的情况下, 下一次猜牌猜中的概率是多少。在假定事件 s_2 发生的条件下, 事件 s_1 发生的概率可以定义为

$$\Pr\{s_1 | s_2\} = \Pr\{s_1 \cap s_2\} / \Pr\{s_2\}$$

如果 n 个事件 $s_1, s_2, \dots, s_n$ 互相独立, 则有

$$\Pr\{s_1 \cap s_2 \cap \dots \cap s_n\} = \Pr\{s_1\} \Pr\{s_2\} \dots \Pr\{s_n\}$$

1. 随机变量的期望值

一个随机变量分布的最简单、最有用的值是该随机变量在样本空间中所有取值的平均, 即重复采样随机变量得到的值的平均。一个离散随机变量的期望值定义为

$$E[X] = \sum_x x \Pr\{X = x\}$$

引理 1.1.1 给定一个样本空间 S 和在 S 中的事件 s , 使 $X_s = I\{s\}$ 。则 $E[X_s] = \Pr\{s\}$ 。

证明: 根据二元随机变量和期望值的定义, 有

$$\begin{aligned} E[X_s] &= E[I\{s\}] = \\ &1 \cdot \Pr\{s\} + 0 \cdot \Pr\{s\} = \Pr\{s\} \end{aligned}$$

现在, 给出一个例子来简单说明上述概念的应用。

抛一个硬币, 其样本空间为 $S = \{H, T\}$, 其中 H 表示硬币正面向上, T 表示硬币正面向下。令 X_H 表示一个二元随机变量, 记录硬币正面向上的次数, Y