图书情报应用数学

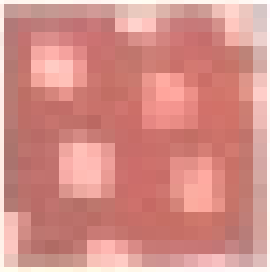
——知识组织、发现和利用中的数学方法

邹晓顺 王晓芬 邓珞华 编著

国家图書館出版社

图书情报应用数学

——统计描述、预测和决策中的数学方法



[redacted]

[redacted]

图书情报应用数学

——知识组织、发现和利用中的数学方法

邹晓顺　王晓芬　邓珞华　编著

國家圖書館出版社

图书在版编目(CIP)数据

图书情报应用数学:知识组织、发现和利用中的数学方法/邹晓顺,王晓芬,邓珞华编著. —北京:国家图书馆出版社,2012.7

ISBN 978-7-5013-4710-0

Ⅰ.①图… Ⅱ.①邹… ②王… ③邓… Ⅲ.①图书情报学:应用数学 Ⅳ.①G250-05

中国版本图书馆 CIP 数据核字(2012)第 257499 号

责任编辑:王 涛 李家儒

书名 图书情报应用数学——知识组织、发现和利用中的数学方法

著者 邹晓顺 王晓芬 邓珞华 编著

出版 国家图书馆出版社(原北京图书馆出版社)

(100034 北京市西城区文津街 7 号)

发行 010-66139745 66151313 66175620 66126153

66174391(传真) 66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 北京汉玉印刷有限公司

开本 787×1092(毫米) 1/16

印张 14.5

版次 2012 年 7 月第 1 版 2012 年 7 月第 1 次印刷

字数 310(千字)

书号 ISBN 978-7-5013-4710-0

定价 60.00 元

序

近几十年来，数学在各个领域的应用方兴未艾，取得了突出成果，充分显示了其广泛的渗透性和应用性。数学用形式语言描述研究对象，简化对象之间复杂的依存和相互影响关系，抽象具体的研究过程，精确地揭示对象之间的逻辑规律，以严格的逻辑推理保证结论的正确性。无论是作为思想还是作为工具，数学都显示出强大的生命力。因此，“一种科学只有当它达到能够应用数学的时候才算真正发展了”（马克思语）。今天，数学已成为分支繁多、体系严谨、应用面极广的庞大学科，在推动整个科学技术发展中扮演着举足轻重的角色。正如笛卡尔所说：“数学是人类知识活动留下来的最具威力的知识工具，是一些现象的根源。数学是不变的，是客观存在的，上帝必以数学法则建造宇宙。”

与自然科学和工程技术相比较，社会科学对数学的应用相对落后，这并不是说社会科学没有资格应用数学，而是社会科学的研究对象受人为因素的影响，较之自然科学和工程技术具有显著的不确定性、混沌性和不分明性，数学应用于社会科学的难度要大得多。但近几十年来，情况发生了根本变化。由于数学在经济学、管理学、社会学、心理学和行为科学等学科中的成功应用，大大提升了这些学科的科学特质，催生了许多新兴的研究领域，取得了前所未有的成果，为社会科学应用数学理论和方法提供了新的舞台和范例。

图书馆学情报学对数学的应用虽然不像经济学和管理学那样系统、全面和深入，但也达到了相当高的水平。以图书馆统计应用为标志，20 世纪三四十年代通过经验统计得到的，在图书馆学和情报学中具有奠基性的布拉德福定律、洛特卡定律、齐夫定律和文献的增长与老化定律先后引起了许多学者的关注，如维克利、费尔桑、莱姆库勒、曼德布罗特等。他们先后用概率论、信息论、函数论对这些经验定律给出了严格的理论推导，并将其统一到一个表达式中，揭示了文献情报流的规律，大大推动了图书馆学和情报学的定量化研究，使图书情报学科的理论和应用都上了一个新的台阶。在此基础上创立的文献计量学，使图书情报学科渐趋成熟，得以向其他学科输出研究方法。如果说数学理论和方法的应用在这一阶段仍然显得零散，但在计算机运用于图书情报工作之后，特别是网络兴起和普及应用之后，数学便广泛渗透到图书情报学科和图书情报工作的各个领域，涉及知识信息获取、组织、管理、发现和评价的各个方面，从系统到模型，从方法到原理，逐渐形成体系，其研究水平和应用深度大大提升了图书馆学和情报学的科学性、严谨性。我曾多次强调指出，图书馆学情报学要取得突破，在微观上需要解决两个关键问题：一是知识信息的表达和组织必须从物理层次的文献单元向认识层次的知识单元或情报单元转换；二是知识信息的计量必须从语法层次向语义和语用层次发展。其实，这两个

关键问题的突破也必然要仰仗数学的支撑。

20 世纪 80 年代初，我国图书情报学界的一些年轻学者在介绍和引进国外研究成果时，就意识到数学应用的重要性以及我国在定量化研究方面与发达国家存在的差距，自觉地在自己的研究和实际工作中应用数学方法。在文献情报流规律的揭示、图书情报服务统计、情报检索模型建立等不少方面都取得了较好的成果。本书作者之一邓珞华曾和孙清兰、范并思合作编写出版了《图书情报数学》，被一些高校图书情报专业用作教材，为推动我国图书馆学和情报学应用数学理论和数学方法起到了积极作用。90 年代以后，由于信息技术高度发展和广泛应用，全新的网络信息环境给图书馆学和情报学提出了许多重大课题，这些课题如果不采用数学方法进行深入研究就不能获得突破。以此为契机，我国图书情报界的理论研究和实际工作中开始了一个定量分析研究的繁荣时期。如今，当时的年轻学者都已进入花甲之年，看到我国的图书馆学情报学研究由于采用了定量分析手段而有了长足的进步，感到由衷的欣慰。

数学的应用取决于研究者自身的知识结构和相应的环境支持。改革开放以来，我国培养了大批硕士和博士，他们有比较合理的知识结构，有能力应用数学开展定量化研究，有不少学者也的确做得不错。但遗憾的是，一些学术期刊对此没有兴趣，不愿意刊载满是数学符号的论文，称其为玄学。当然，我们反对利用数学语言故弄玄虚，把简单的问题复杂化。但很难设想，一份完全没有数据、没有模型、没有推理论证的期刊会拥有很高的学术水平。在国内，图书情报界是较早应用数学开展研究的学科领域，但经济学界、管理学界，甚至公共管理和社会学界在近 10 多年来应用数学方面已经大大超越图书情报界，达到了较高水平。这些领域的学术期刊的强力推动和支持是重要原因。

在邓珞华等人 1983 年出版的《图书情报数学》基础上，充分考虑到网络时代知识获取、组织、发现与利用的内容和特征，以及数学应用的新进展，邹晓顺、王晓芬、邓珞华三位老师编写了《图书情报应用数学》。这是一件非常值得庆贺的事。该书在我国第一次比较全面、系统地总结了数学在图书馆学、情报学各个分支的应用，对推动我国图书馆学情报学的定量研究具有重要意义，对现代网络环境下的图书馆工作和情报工作也具有重要的参考价值。而国家图书馆出版社积极支持本书出版，显示了该社领导和编辑的远见卓识，不仅重视“图书馆人文”，也重视“图书馆数理”。

值得一提的是，本书的三位作者不是专职从事图书馆学情报学科研和教学的教师，而是长期供职于高校图书馆的实际工作者，他们完全是利用业余时间在从事研究和编写工作，这就更加难能可贵。邓珞华同志在图书馆工作 30 多年，笔耕不辍，一直坚持图书情报应用数学的研究，发表了不少有影响的文章；王晓芬同志是数学教授，又兼图书馆馆长，她的加盟使本书在数学理论和应用上有了更大的可信度；邹晓顺同志长期从事图书馆的系统管理和系

统设计，他撰写的章节使本书更具有鲜明的时代感。这一“黄金组合”保证了本书的成功写作。

由于图书情报应用数学涉及大量的数学分支和图书馆学情报学不同专业领域的知识，也没有同类著作可以参考，完成编写任务之艰巨可以想见。正因为如此，本书也必然存在不少有待改进之处。我相信读者会独具慧眼，在阅读和使用本书过程中与作者交流互动，不断锤炼，使本书的内容和结构更加完善。

《图书情报应用数学》付梓之际，作者邀我作序，写下上面的感想，以表达我对三位作者的祝贺和敬意。

马费成

2011 年 2 月于武昌珞珈山

前　言

大学期间，我和班上其他三位同学被图书馆学系选送到计算机科学系同时修图书馆学和计算机科学两个专业的课程，从此我与图书情报应用数学结下了不解之缘。1982 年我从武汉大学毕业，与之前结识的湖南大学图书馆学专业的学生范并思（后任华东师范大学信息管理系主任）开始筹划写一本关于数学在图书馆学情报学中应用的书。十分幸运的是，时任东北师范大学图书馆系主任的单行先生十分欣赏我们两个年轻人“初生牛犊不怕虎”的精神，派系里的数学教师孙清兰与我们合作，并把我们的书稿作为东北师范大学教材于 1983 年出版，定名为《图书情报数学》。

尽管《图书情报数学》在我国开创了数学在图书馆学情报学中系统应用的新阶段，对我国图书馆学情报学的发展起到一定的作用，但随着计算机和通信技术的发展及其在图书情报工作中的应用不断深入，27 年后再看这本书，虽然仍有闪光之处，但也有“落花流水春去也”的过时之感，尤其是图书馆自动化管理系统、数据库系统这两个数学应用十分重要的领域，由于当时条件所限，在书中显得十分单薄，此外该书体系上也有些凌乱。因此，很早以前，我就有重编该书的打算，但一直未能如愿：一是我担任武汉大学图书馆副馆长和湖北省高校图工委秘书长后，精力和时间有限，兴趣也有所转移；二是以我数学和计算机的那点功底，实在难以胜任重修这本书的重任。

2008 年，我结束了 33 年图书馆工作的历史，退休赋闲了，又重新燃起重修该书的愿望。当我谈起这件事时，我的两个图书馆界的朋友王晓芬（武汉体育学院图书馆馆长）和邹晓顺（武汉科技大学图书馆系统部主任）表示了浓厚的兴趣。前者是学数学出身，体育统计学的教授，又从事图书馆管理工作；后者是学物理出身，长期从事图书馆自动化管理系统的设计和管理。有他们二人的加盟，我想写出一本更科学、更实用、更富有时代气息的《图书情报应用数学》，应该不是一件难事了。

与《图书情报数学》相比，本书有两个特点：一是增加了网络环境下文献信息工作中的数学方法等方面的内容，尤其是图书馆自动化管理系统、知识组织和数据库系统等方面的内容，使本书的时代气息更浓，实用性更强；二是按照知识的组织、发现与利用划分章节，体系上更科学，针对性更强。关于本书的书名，有些争议。有些人认为许多图书馆学系情报学系都更名为信息管理系了，不如把书名改作《信息管理应用数学》之类的名称，但我总觉得“信息”这个词范围太广，容易产生语义混淆，图书馆学会和情报学会也没有更名为信息学会，所以还是采用《图书情报应用数学》这个题目更明确更醒目一些。

正如绪论中所说，定量分析是一门学科发展到高级阶段的标志，我们相信本书对于我国图书馆学和情报学以及图书情报工作的发展会起到一定的推动作用。一方面，它可以作为高校

图书情报和信息管理专业的教材和教学参考书;另一方面,它对于从事图书情报工作,尤其是管理工作和系统设计的实际工作人员,也有较大的参考价值。

由于本书涉及了大量的数学知识和图书馆学情报学方面的知识,而图书情报工作人员中懂得这些数学知识的人为数不多,数学工作者中懂得图书情报知识的人更是微乎其微,以至于该书的审稿都很困难,只能采取文责自负的办法,所以一般读者要看懂它还是有相当难度的。但随着计算机技术在图书情报工作中的普及和图书馆学情报学定量分析的需要,相信本书会有越来越多的"知音"。另外,由于图书情报应用数学还是一个正处在发展阶段的学科分支,很多问题还处在探讨阶段,难免存在疏漏甚至错误之处,还望读者不吝赐教,使之完善。

本书分工:邹晓顺负责第2、3、4、5章的撰写,王晓芬负责第9、10章的撰写,邓珞华负责第1、6、8章的撰写,中南财经政法大学图书馆的邓东宁负责第7章的撰写,最后由王晓芬负责数学方面的审查,邹晓顺、邓珞华统稿。此外,我国著名情报学专家、武汉大学信息管理学院原院长马费成教授百忙之中为本书作序,国家图书馆出版社的王涛、李家儒为本书的评审编辑工作付出了辛勤的劳动,大量的参考文献是本书的基础,尤其是范并思、孙清兰在本书的前身——《图书情报数学》一书中付出了大量心血,我们正是站在这些"肩膀"上攀登科学高峰的,在此对他们致以深深的谢意。

邓珞华

2011年11月于武昌珞珈山

目　录

1 绪论

1.1 图书情报应用数学的简短回顾

数学运用于人类的生产、生活，已有悠久的历史。从远古人的结绳记事，到古埃及人的几何丈量术，从古印度人的“0”的发明，到我国《周易》的排列算法，数学伴随着人类生产和生活中“数”和“形”的问题愈来愈受到重视。但直到公元前6世纪，这种知识还没有形成具有逻辑关系的体系，因而只能作为数学的萌芽载入史册。

从公元前5世纪到公元16世纪，逐步完备起来的算术、初等几何、初等代数等，以静止的数和形的简单组合为研究对象，形成我们现在所称的常量数学。

16世纪，“运动”的概念随着天文学和力学的发展进入自然科学的中心课题，这使得变量的概念应运而生。笛卡尔以力学的要求为背景，使几何内容的课题与代数形式的方法相结合，建立了解析几何学。17世纪下半叶，牛顿和莱布尼茨各自独立地建立了微积分。此后又有级数理论、微分方程、微分几何以及较晚一些的复变函数论的产生。与此同时，几何、数论、代数也都在继续发展着，画法几何学和射影几何学也相继产生。此外还开辟了或然数学的领域，建立了概率论。直到18世纪末的大约200年间，是以微积分为基本思想的变量数学的长足发展时期。

19世纪以来数学的发展在几何、代数、分析等方面都有了深刻的变化，一切可能的量，一切可以抽象出来的量以及这些量之间的关系都成为数学的研究对象。非欧几何、群论、集合论、拓扑学、突变函数论、泛函分析以及一些交叉学科，如代数几何、分析拓扑、大范围分析等，都十分明显地显现出这些特征。由于对数学基础的研究，又有了数理逻辑各个分支的建立和发展。

近几十年来，尤其是二战中运筹方法应用产生的影响，形成了优选学、规划论、对策论、排队论等运筹学科。由于研究通讯和自动控制系统的需要，几门介于数学和工程之间的学科——信息论、控制论、系统论又相继问世。电子计算机的发明大大促进了计算数学的发展，并形成了计算机科学的数学理论。数学和其他科学的相互渗透，又促使了物理数学、经济数学等边沿学科的产生。近20年，又出现以不分明的量为研究对象的模糊数学这一新的研究领域。今天，数学已成为分支繁多、体系严谨、应用面极广的庞大学科。

数学具有三大特征：一是抽象性，它撇开具体的研究对象，只把其中的量及其之间的关系抽象出来。像复数、函数、微积分、泛函、线性空间这样一些概念，其抽象程度大大超过了自然科学中的一般抽象，抽象到似乎与生活失去了一切联系，以至“凡夫俗子”除了感到莫名其妙以外什么也不能理解。二是精确性，首先它用数的形式保证任何两个对象在某一共同属性上比较的精确，其次它以严格的逻辑推理保证了结论的正确性。三是应用极其广泛，以至“一种科

学只有当它达到能够应用数学的时候才算真正发展了"(马克思语)。数学的各个分支在短短的二三十年间在图书馆学情报学与图书情报工作中迅速扩展蔓延开来,从无足轻重发展到举足轻重,正是数学强大威力的具体体现。

图书馆学与情报学运用数学手段的开始,是以图书馆统计的产生为标志的,但数学知识更广泛更深入的运用,则是在电子计算机运用于图书情报工作之后。19 世纪乃至 20 世纪 50 年代以前的图书馆中,图书资料的浩淼与人们对图书资料的特定需要之间的矛盾并不明显。人们利用目录索引查询图书资料,并不担心书库装不下书,目录柜装不下目录,也不担心目录卡片容不下著录项目,因此他们不必绞尽脑汁设计一个复杂系统,用最简短的符号把图书文献存储起来,用尽可能快的速度又准又全地把所需图书资料查找出来。这时人们对数学的兴趣,仅仅在于书的册次、读者的人次、工作人员的工作量、财政的预算决算等,而对于这样一些问题人们只需要具备一般的统计知识就足够了。因此在这一时期的图书馆学情报学中,统计应用即是数学应用的代名词。20 世纪初,图书文献的数量有了一定增长,文献流的研究提上了议事日程,人们对文献流研究的结果是发现了著名的布拉福德定律、齐夫定律、洛特卡定律以及文献增长规律、文献老化规律,这给情报学增加了量的概念和定量分析的方法,大大推动了情报学向纵深发展,至今这些定律仍是情报学的基本定律。但是,这些定律仍是用统计学手段发现的。虽然现在人们试图用比较高深的数学理论解释它们,但在当时,它们涉及的数学理论和数学方法却十分有限。只是到了 20 世纪 60 年代,电子计算机应用于图书情报界以后,数学方法才在更深的程度和更广的范围应用到图书馆学和情报学中来。

首先是计算机管理系统数学模型的建立。由于计算机管理系统比人工管理系统更模型化、规范化、条理化,人们必须用系统分析的方法,根据某种理论将图书情报管理系统抽象出一个个独立的部门(元、集合、点),并描述出这些部门之间的关系——函数关系、映射关系,主要是它们之间的信息交换关系,然后根据这些关系规定各部门之间的工作流程和工作走向,这些都涉及到图书情报管理系统的数学模型问题——采用什么样的模型使计算机管理起来更方便,更迅速,更经济。据统计,仅信息检索系统的数学模型就有:1968 年 G. Salton 的集合论模型、1976 年 A. Books 和 W. Cooper 的集合论模型、1976 年 T. Radecki 和 V. Tanani 的模糊集合模型、1977 年 Van Rijsbergen 和 Robertson 的概率集合模型、1977 年 Swets 的概率论和数理统计模型以及后来 G. Salton 提出的十分有创见的代数(矩阵—向量)模型,还有适合于理论分析的线性方程组模型。此外围绕检索系统的评价问题和文献标引问题,检索决策过程问题,人们也提出了不少数学模型,这使得高等数学各个分支蜂拥而入图书情报领域,打破了统计学手段在图书情报工作中一统天下的局面,开创了数学在图书馆学情报学中应用的繁荣时代。

其次是图书文献(包括二次文献、三次文献)本身的信息化、代码化问题也使得高等数学手段得以引进。机读型文献的信息化、代码化问题是与压缩计算机内存空间密切相关的。文献的外形从书本式发展到磁带式和缩微胶片式以后,目录的外形也从书本式、卡片式发展到机读型,在纸上不是大问题的文献符号的长度此时成了影响计算机应用的关键问题。机读型数据库的重要课题之一是如何压缩书目数据,使之代码化、通用化,将能够合并的书目信息尽量合

并,将字符长的书目信息用尽可能短的符号表示出来,并尽可能合理、充分地分配存储空间。这样做的结果之一,就是数学手段——如概率论与数理统计、排列组合论、数论、信息论——的引进。

再次是计算机系统的管理和评价问题。计算机系统的管理和评价,一是效率的管理和评价,二是效益的管理和评价,包括:计算机内部硬件模块与软件模块的联系和调配,费用的核算,效果检验指标的判定和检验方法,以及现在时兴的计算机网络安全的监控与防范等等。这些工作不采用高等数学的工具,如优选法、规划论、对策论、图论、排队论以及系统论、控制论等,也是难以深入下去的。

最后是由于计算机的快速准确使原来可以引进的但因工作量太大、时间太长不能投入实用的数学方法得以投入实用。如主题词表、分类词表的逻辑错误问题,国外早已能够用机器自动检查和纠正,我国也提出了树型图算法和关系矩阵算法进行自动检查和纠正,但无论哪种方法,不借助计算机都是难以实现的。

总之,尽管数学方法应用于图书馆学情报学还比较零乱,不太成熟,但从应用速度之快,应用范围之广这一趋势来看,已经显示出光明的前景。在短短的二三十年间,数学各大分支——微积分、线性代数、集合论、概率论与数理统计、数论、运筹学、信息论、系统论、控制论等——程度不同地运用到图书馆学情报学的各个分支,如基础理论、情报检索和情报检索语言、情报分析、图书馆自动化、图书馆管理、文献资源建设、甚至传统的目录学之中。在国外,这一势头更为明显,信手翻翻他们的图书情报专业书刊,公式、数据、图表比比皆是。然而在我国图书情报界,尽管近30年引进了不少计算机人才,但能自觉地、有意识地运用数学方法解决工作中和理论研究中的问题者却寥寥无几。究其原因,除了传统习惯的影响和人员知识结构的影响,主要是图书情报的管理系统不像自然界的宏观上和微观上的机械系统那样有比较明显的运动规律,容易模拟和运算,而是受人的主观因素以及其他随机因素的影响很大,很难构造出精确的模型。但这并不意味着数学方法就不能应用到图书馆学情报学中,更不意味着应该消极地对待图书情报应用数学的研究。正确的态度应该是:根据具体情况分别对待,对能够运用到图书馆学情报学中的数学方法应加紧其研究步伐;对暂不能模拟出数学模型的,不要牵强附会,同时要鼓励有志之士敢于攻关,因为从长远的观点——科学从定性分析的初级阶段向定量分析的高级阶段发展的观点——来看,数学方法在图书馆学情报学中的应用是一个必然的趋势。

用积极的态度对待图书情报应用数学的研究是有现实基础的。一方面,图书情报机构尽管有难以模拟的、包含太多随机因素和人为因素的系统,但也有比较容易模拟的、结构比较简单的系统,尤其是计算机管理系统,如检索系统、服务排队系统、数据库系统、词表系统等;另一方面,图书情报工作中的大量随机现象很多是服从规律性分布的,如单词的频率服从齐夫分布,文献在期刊中的分布以及馆藏图书利用率服从布拉德福分布,文献的增长服从普赖斯曲线,读者到馆率服从泊松分布,著者在文章中的分布服从洛特卡定律等,这些比较容易用数学模型模拟出来的系统以及这些呈规律性分布的随机现象是我们应用数学方法解决图书情报工作中的实际问题的基础。当然我们还应该看到图书情报机构中有些系统、有些过程由于受到

太多太复杂的人为因素和随机因素的影响，尤其受以上因素的综合影响，用数学方法模拟和计算都很困难。对这类问题我们也不必硬给它嵌进一个与实际误差太大的模型，或给它一个无法计算的公式。例如20世纪末讨论得很热烈的情报量的定义和计算问题，尽管对于情报流和情报系统的定量化描述具有十分重大的意义，但因为影响情报量的因素太多太复杂，因而至少在短期内无法给它一个能投入实用的定义和计算方法。对于这样的问题，我们不妨采取“放一放”的态度，避免咬文嚼字的文字游戏，同时鼓励人们在实际工作和理论研究中对该问题进行实质性的，特别是有实用价值的探讨。

1.2 数学与图书情报工作的客观联系

数学与图书情报客观世界有着不同一般的联系。

(1)图书情报工作中虽然也存在着数和形的现象，但这些数和形往往不是直接产生数学概念和公式的起源，解决图书情报工作中数和形的问题往往借助于数学中已有的概念、公式和方法，因此在借用这些概念、公式与方法时往往要根据图书情报工作的特殊环境和特殊条件对它们进行限定和修正。

例如，在定义情报量时，有人借用信息量的计算公式 $I = -\sum \log Pi$。但用于情报系统的情报量和用于通讯系统的情报量有所不同，那就是通讯系统中的信息往往处在一种固定的环境下，为机器所识别；而情报系统中的情报往往处在不同的环境下，为不同的人接收，它的情报量(价值量)随着接收条件的差异、接收者的需求和接受能力的不同而变化。因此，问题就转化为如何对信息量公式加修正系数了。

又例如，情报检索系统的向量—矩阵模型中提出的“概念空间”来源于高等数学“线性空间”概念，但与线性空间相比，概念空间受到更大的限制：第一，严格地说，它甚至不是一个线性空间；第二，概念空间中每一个向量各个元的值(加权值)局限在一个很小的范围，使得线性空间中的一些运算不能在概念空间中进行；第三，即使没有以上两点，概念空间也不是情报检索系统数和形的问题的精确反映。情报检索系统中的概念空间与高等代数中的线性空间有这样一些差距，主要是由于概念空间的向量中各元的取值范围太小，而文献与文献之间又不能简单相加、相减或相乘得出新的文献，这样在实数、整数或其他范围较宽的数域中可以进行的运算，受到了数域范围太小的局限，而且文献之间的运算很难找到一个完满的定义。

又例如，微积分中“极限”和“连续”的概念在图书情报工作中运用时也有加以说明的必要，即图书情报工作中许多函数都是在“假定是连续的”或“看成是连续的”条件下求导和积分的，但实际上，图书情报工作中许多函数表达式在其定义域内是不连续的。虽然在许多情况下作这样的假定不碍大局，但在有些情况下却不能作这样的假定，否则会得出不符实际的结论。我们至少应该明白，“假定的连续”和数学上严格定义的连续是有区别的。例如著名的文献增长规律的数学解析式为：

$$y = ae^{kt}\ (a>0, k>0)$$

它的函数图像如图 1 - 1。虽然 y 在数学定义里是连续的，但这里它却是离散的，因为文献不会有“几点几”篇这样的表示，y 显然只能取整数值。有人曾经试图用微分的方法求出 y 随 t 的增长速度，考察图像的走向。这样做虽然与数学上严格的求导条件相违背，假定了 y 的连续性，但并不影响我们从宏观上估计曲线的走向。但是，当 y 不是代表文献累计量，而是代表当年文献数量时，如果我们在 t 轴上截取一段时间[a,b]，对这段时间的文献量用积分方法求和，那就大错特错了。因为，由于 y 不连续，a、b 之间时间段的文献总量实际上只是 a、m_1、m_2…b 等年代文献量的总和，即 A、M_1、M_2…B 这些值的和，而积分 $\int_a^b y(t)\,dt$ 则是梯形 ABba 的面积，显然和上面求出来的解不是一回事，结果相距可能很大。因此我们不能为图像的连续画法所迷惑，要考察图像依据的数据是如何得来的，是否真的连续；如果是不连续的，要考虑是否可以假定连续。这使得我们可以用连续定义下的微积分方法而不影响问题的分析。

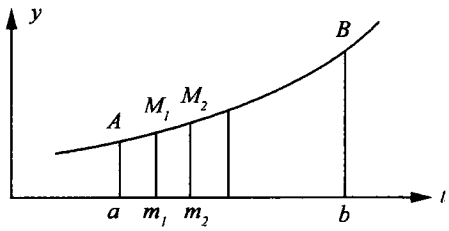


图 1 - 1　文献增长规律数学解析式的函数图

(2)图书情报管理系统的许多子系统不是简单的机械系统，而是受到多个复杂随机因素，尤其是人为因素影响的系统，在考虑应用数学方法(例如制定数学模型)时很难精确地模拟出所要应用的系统及其内部的函数关系，往往要忽略一些次要影响因素，尽量排除随机因素的影响。

数学应用在目前阶段，其用武之地主要还在机械系统中，当然这不是说对生命系统的模拟就无能为力了，尤其当这个系统充满大量的、规律性分布的随机事件时更非如此，例如排队论模型就能很好地模拟出图书馆的一些服务系统。但总的来讲，在生命系统中应用数学模拟方法还是比较困难的，尤其像图书情报管理系统这种机械和人的复合系统更是如此。20 世纪 80 年代，曾经有人提出了一个情报系统的微分方程组模型。他画出情报系统的一个动态流程图，然后用一组微分方程模拟了情报交流过程。在这个微分方程组模型中，提出了累积情报量、当用情报量、过时情报量、价值情报量等概念，后经他人修正，又提出反馈系数、浓缩系数、时滞系数、分配系数等概念。该模型对于从宏观上理解分析情报系统是有意义的，这些概念的提出于情报学理论也不无帮助，但制定一个数学模型绝不只是为了宏观地了解某个系统(事实上，一个动态流程图就可以达到这个目的)，而是要通过计算最终解决一个问题。就这一点来讲，该模型没有达到数学应用的目的，它无法进一步计算，因为这个方程组里的所有系数都涉及到大量复杂的随机因素，如经济因素、思想因素、甚至体制因素的影响，至少在目前阶段，这些因素

对情报系统的影响我们还无法测定出来(哪怕是粗略地测定出来)。但是有些图书情报管理系统,特别是一些具体的计算机管理系统、服务系统中,随机因素、人为因素的影响要小一些,我们可以先不考虑这些因素,在进行结果分析时再对这些因素加以考虑是可行的。例如设计计算机流通管理系统的评价分析子系统(或叫借阅统计子系统)时,要制定一系列的评价指标,如拒借率、流通率、借阅率等,这些指标是受多个方面因素影响的。在大多数情况下,它们能够反映和评价系统的工作情况,如拒借率反映图书管理水平,流通率反映图书的利用程度,借阅率反映读者对图书馆的兴趣等。但情况并不总是如此,如拒借率并不总是反映图书管理的水平,在某些情况下图书管理的水平提高,拒借率反而降低,或者维持在一定水平上;在某些情况下,一些与图书管理水平无关的因素(如购书经费的增减、馆舍面积的增减等)也引起拒借率的变化。那么当该评价分析子系统实施应用时,除了根据系统统计的拒借率、流通率、借阅率等指标分析流通工作外,还要结合当时当地的具体情况和具体条件用其他方法,如心理学、经济学的方法分析评价结果是否如实反映了流通动态,因为社会、经济等因素的影响一时无法测算,只好在设计时加以排除,在实施时再加以考虑。

数学应用于图书馆学情报学的这些特殊之处,使得图书情报工作者在应用数学方法进行研究时,必须对课题进行全面深入的调查,弄清本质的和非本质的部分,必然的和或然的部分,具有规律性分布和分布不规律(或暂时找不出分布规律)的部分,去粗取精,去伪存真,系统分析,综合考虑,才能提炼出能够较好反映系统结构和功能的数学模型、数学公式和计算方法,而且最后常常要用其他学科的方法分析验证计算结果。

当然,图书情报系统进行数学模型的提炼,并不是随心所欲、呼之即来的事,任何一种新的思想或方法的产生和应用,首先要有客观基础,即客观上存在着产生这种思想和方法的条件;其次要有客观需要,即科学的进步和技术的发展已经将应用这种思想和方法提到日程上来。数学应用于图书馆学情报学正是如此。如前所述,计算机在图书情报部门的应用要求用数学方法提出、分析和解决问题,但是仅有客观需要还不行,数学能够应用到图书馆学情报学中来,归根结底在于图书情报客观世界中存在着数学现象。下面把这些现象按图书馆学情报学的各个分支进行粗略的分类。

(1)检索语言(分类法、主题法、叙词法等)

检索语言中的数学现象是十分明显和有趣的。首先,任何一种检索语言都有着易于数学模拟的结构。当我们把每一个检索概念(类名、主题词、关键词)当做一个结点或线性空间中的一个点,把概念之间的关系(等级关系、并列关系、全等关系、参见关系等)当做一条条边或映射,就很容易用图论、矩阵、集合论的模型模拟出检索语言的结构。应用这个数学模型的最好的例证就是主题词表逻辑错误的自动检查。在编制主题词表的过程中,由于编制人员的专业知识面不同、理解不同、粗心大意等原因,难免出现各式各样的逻辑错误。有人统计,我国第一版《汉语主题词表》中,有大约10%的主题词下出现各式各样的错误(主要是逻辑错误)。在国外,早些年就解决了机器自动检查词表逻辑错误的问题。1982—1983年《情报学报》先后有作者提出了用树型结构节点扫描法和关系矩阵法检查主题词表逻辑错误的算法,其中树型结构