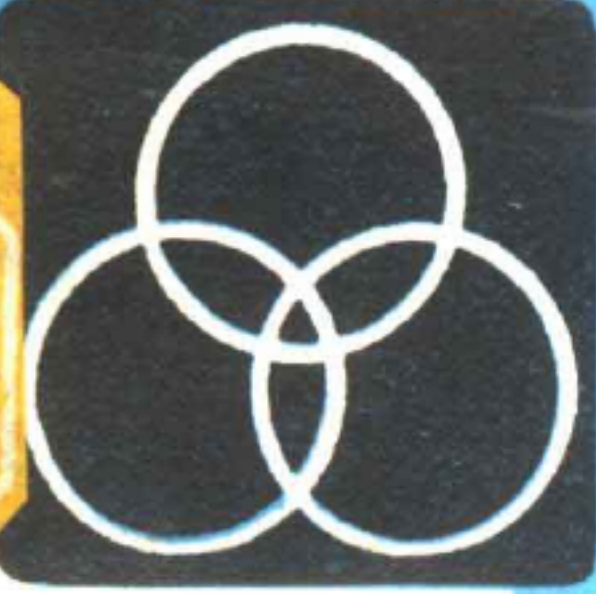


964499

G350
94344

情报计量学引论

[比] L·埃格希 R·鲁索



科学技术文献出版社

封面设计 崔文毅
责任编辑 向 明

科技新书目：278—117
ISBN 7-5023-1770-8/Z·295
定 价：9.60元

n
11
'9
17
5
2
0
8
6
4
2
0
8
5
3
1

钟莉妹 校

(京)新登字130号

内 容 简 介

情报计量学是近年来在文献计量学基础上发展起来的一门新兴学科,是当前国际图书、文献、情报研究范围内最活跃的领域之一。本书围绕情报计量学的基本原理和应用这一主题,进行了深入浅出的论述。全书共分4编:1.统计学;2.运筹学与图书馆管理;3.引文分析;4.情报计量学模型。本书采用了大量的公式、图表、应用实例,几乎概括了图书、文献、情报计量学方面的所有典型的公式、表达式、公理、定理、定律和图表,文字流畅、结构严谨、层次分明、易学易懂、使用方便。本书适用于作为大专院校图书、情报系师生的教材或教学参考书;同时对广大图书、文献、情报界从事定量分析和研究的工作人员也有很大参考价值。

INTRODUCTION TO INFORMETRICS

Quantitative Methods in Library,
Documentation and Information Science

Leo Egghe Ronald Rousseau

Elsevier Science Publishers

Amsterdam New York Oxford Tokyo

1990

情报计量学引论

[比] L. 埃格希 R. 鲁 索 著

田苍林 葛赵青 译

汤兆魁 钟莉妹 校

科学技术文献出版社出版

(北京复兴路15号 邮政编码100038)

中国科学技术情报研究所印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

850×1168毫米 32开本 13.125印张 353千字

1992年9月第1版 1992年9月第1次印刷

印数:1—2200册

科技新书目:278—117

ISBN 7-5023-1770-8/Z·295

定价:9.60元

译 者 序

情报计量学是近年来在文献计量学基础上发展起来的一门新兴学科。欧美及其它一些国家的图书、文献、情报院校也开设了此门课程，这本专著就是两位作者在比利时（及欧洲其它国家）从事情报计量学教学及研究工作的总结。

本书围绕情报计量学的基本理论和应用这一主题，进行了深入浅出的论述，其中有大量的应用实例、图表，使用极为方便。

目前，我国虽尚未正式出版有关情报计量学的专著，但可以预见，随着我国文献情报研究工作的进一步深入，情报计量学的理论和方法必将在我国图书、文献、情报界得到迅速而广泛的应用。

本书适合作为大学教材或教学参考书，对广大图书、文献和情报界从事图书、文献、情报定量研究的人员也有很大参考价值。

本书的概述、第 I、II 编由葛赵青翻译，序、第 III、IV 编、参考文献、附录由田苍林翻译。由汤兆魁、钟莉妹同志任审校。西安交通大学数学系讲师徐萍同志从数学角度审阅了部分稿件；徐琳、倪慧、宋乐同志为本书的出版亦做了很多工作，在此一并表示感谢。

我们欢迎广大读者对翻译中的缺点和不足之处提出批评。

译者

1991年10月于西安

序

1984年，我们两位著者进入文献计量学领域——现在我们宁可更为人们所接受的术语“情报计量学”。当时我们正积极致力于图书文献学高等教育大纲的研究工作，这个大纲是在弗莱米什校际理事会的倡议下由安特卫普大学组织的。不久表明，对情报计量学教材发生兴趣不仅在弗兰德斯，而且在瑞士也有。事实上，由于欧共体资助的伊拉兹马斯大纲取得的成果，利奥·埃格希也负责了阿姆斯特丹大学的情报计量学大纲的研究工作。

1986年前后，我们以埃格希的一部弗莱米什短期教程为基础，开始了本书的编写工作。现在情报计量学的领域已是如此广泛，以致目前还没有一部内容完整的入门性著作，因此我们试图在本书中编排尽可能多的专题。

在编写本书时，我们以文笔清晰、主题逻辑性强等特点，来引起非专家（和非数学家）的兴趣。在编写准备阶段，我们面临的文献分散在内容差别很大的各类期刊和图书中的难题。此外，正如人们所共知的，在这一领域工作的大部分科学家只把情报计量学家作为“第二选择”。他们都是作为图书馆员、物理学家、化学家、数学家、社会学家、心理学家或计算机科学家而接受教育的，并在其出版物中揭示了不同的背景。因此，要统一各种观点不是一件容易的事。

我们希望本书能对从事情报计量学课程教育的老师有所帮助，并希望它能引起图书情报学专业学生的兴趣，学生既可把它作为教材，也可作为参考读物。此外，我们希望从事实际工作的图书馆员也能发现它的用处，因为我们提供了很多简便易懂的、非数学的图书馆管理技巧。在科学政策领域工作的研究人员和学者也会对本书

产生兴趣，因为我们提供了许多最新资料。

本书分为四编，每编又由许多章、节、小节（必要的地方还有小小节）组成。为便于参照，我们采用了十进制编码，例如“Ⅰ.4.3.3”表示第Ⅰ编第4章第3节第3小节。方程式、表和图以同样方法编号，直到节。因此，“〔Ⅰ.5.18〕”表示第Ⅰ编第5章的第18个方程式。表和图的数字前有“表”、“图”字样。因此，“图Ⅱ.4.1”表示第Ⅱ编第4章的第一个图。表的标题注在表的上面，图的说明注在图的下面。按字母顺序排列的参考文献以下列方式给出：“姓名（年代）”。在参考文献目录中未出现的第一著者，有“见”参照；对于既作为第一著者又作为第二著者（不同论文的）的著者们，我们增加了“参见”参照。尽管书后所列的参考文献目录比较长，覆盖了情报计量学的许多方面，但这当然并不意味着它就是详尽无遗的。对于那些被遗漏了的任何著者，我们深表歉意。

感谢弗兰德斯文献与图书馆学高等教育大纲的奠基者H. D. L. 范弗里特教授。也非常感谢B. C. 布鲁克斯教授，是他鼓励我们写这本书，实际上，他是“情报计量学”一词的伟大倡议者。布鲁克斯教授也是埃格希在英国伦敦城市大学的博士指导老师，他的博士论文构成了第Ⅳ编的基础。第Ⅳ编是讨论情报计量学“定律”的。

真诚感谢与我们有许多联系的所有科学家。这里列出几位：A. 布克斯坦，T. 布朗，Q. L. 伯勒尔，W. 格兰泽尔，D. 克拉夫特，F. F. 莱姆库勒，I. K. 拉维钱德拉·拉奥，S. E. 罗伯逊，C. 萨尔顿，J. 塔格，R. 托多罗夫和A. F. J. 冯·拉安。

感谢我们所在单位：Limburgs Universitair Centrum, the Universitaire Instelling Antwerpen, Katholieke Industriële Hogeschool West-Vlaanderen给予的关心和支持。我们还要感谢比利时国家科学基金会在各种场合给予我们的支持。

感谢打字员雷恩德斯女士，她的工作产生了很好的打字照相版。

最后要说明，与埃尔塞维尔（Elsevier）科学出版社，特别是与海伦·冯·格尔德伦和苏珊·马索泰极其融洽的合作关系是非常令人愉快的。

著者欢迎对本书提出任何批评、指正，增补或其它任何形式的评论。最后，我们希望本书对从事各类情报工作的人均有裨益。

利奥·埃格希

罗纳德·鲁索

1989. 12

符号表

符 号	说 明
$[a, b[$	使 $a \leq x < b$ 的全部实数 x
$\log_{10} x$	x 的对数 (以 10 为底)
a^x	指数函数 $f(x) = a^x$
$\log_a x$	x 的对数 (以 a 为底)
x^a	幂函数 $f(x) = x^a$
$\sum_{i=1}^n x_i$	$x_1 + x_2 + x_3 + \cdots + x_n$
\bar{x}	平均值 (或加权平均值)
GM	几何平均值
HM	调和平均值
Md	中位数
Mo	模
σ	标准偏差
MD	平均偏差
Q_j, P_j	第 j 个四分位, 第 j 个百分位
V	变差系数
m_r, m'_r	r 阶矩
Ω	全域
$P(A)$	A 的概率
A^c	A 的余集
Φ	空集
$P(A B)$	条件概率
X	随机变量

$$\int_{x_1}^{x_2} f(x) dx, \int_{-\infty}^{+\infty} f(x) dx, \int_{-\infty}^{+\infty} f(x) dx$$

分别为 f 的有限积分及广义积分
累积分布

F

$$\frac{dF}{dx} = F'$$

F 的导数

$E(x)$

x 的平均值或期望值

$Var(x)$

x 的方差

$$\binom{n}{x}$$

$\frac{n!}{x!(n-x)!}$, 此处 $n! = n(n-1)\cdots 3\cdot 2\cdot 1$

π

$\pi = 3.1415927\cdots$

x^2

x 平方, x 的2次幂

Γ

Γ 函数

\square

证明结束

\bar{x}

样本均值

S^2

样本方差

H_0, H_1

虚假设, 备选假设

D

科莫格洛夫—斯米尔诺夫统计量

$O(k), E(k)$

分别为观测值和期望值

O_{11}, E_{11}

U

曼—惠特尼统计量

R

游程数

$S_{x, y}$

样本协方差

$Cov(x, y)$

x 和 y 的协方差

$R_{x, y}$

样本相关系数

$\rho(x, y)$

x 和 y 的相关系数

$\frac{\partial f}{\partial a}$

f 的偏导数

R_s

斯皮尔曼秩相关

\dagger	肯德尔 \dagger
$\text{sgn}(\bar{x})$	\bar{x} 的符号
$\#A$	A的基数=A中的元素数
$\hat{C}=(C_{ij})$	矩阵C
R^k	k维空间, $K=1, 2, 3, \dots$
$+9.11 \text{ E}-08$	$\frac{9.11}{10^8}$
d, d_i	距离, 度量, 不相似性
s	相似性
Arctg	反正切
D_i	距离或不相似矩阵
ESS	误差平方和
E	$E = \sum_{i=1}^{k-1} \text{ESS}_i$
C^t	矩阵C的换位
$::$	由定义相等
$\bar{N}_q, \bar{N}_s, \bar{N}$	分别在队列、服务和整个系统中的平均项数
ρ	利用率
λ, μ	分别为期望到达速率和期望服务速率
$\bar{T}_q, \bar{T}_s, \bar{T}$	分别为消耗在队列、服务和整个系统中的平均时间
m	服务员数量
$(A B M)$	排队模型
$(M M m)$	排队模型
R, R_t	某书在一年内被出借的平均次数
P_1, P_0	能借到(借不到)某书的概率
U, U_i	不满足度
(S_n)	随机过程
$P=(p_{ij})$	转移矩阵
T_m	上年借过m次的图书的借阅数量的随机变量
$N(m)$	$N(m)=E(T_m)$

T_m	$T_m = P(T_m = n)$
$P_r(t)$	一本图书将出借 r 次的概率
T	时间周期
$1/a$	比例系数
δ	$\delta = \frac{T}{T+a}$
β	由“死”书构成的那一部分藏书
$\hat{\beta}, \hat{\delta}, \dots$	β, δ, \dots 的最大似然估计量
\ln	$\ln = \log_e$
γ	欧拉数 $\gamma \approx 0.5772$
μ_A	隶属函数
$C(d)$	文献 d 中所有参考文献的集合
$C^{-1}(d)$	文献 d 的所有引文的集合
\forall	对每一个 (对所有的)
\hat{ZXS}	角 ZXS
\sin	正弦
Arccos	反余弦
JCR	《期刊引文报告》
IPF	效果系数
IMI	即时指数
CIT	引文数量
PUB	出版物数量
a	老化系数
$\vec{c} = (c_x, c_y)$	出版物中心
AI	活动指数
AAI	吸引指数
S_j	耶科特指数
S_e	萨尔顿余弦函数
IPP	情报生产过程
f	洛特卡函数

T	来源总数
r	等级, 秩
g	齐普夫函数或芒代尔布罗函数
h	帕累托函数
R	(一般) 莱姆库勒函数
A	项总数
r_0, y_0, k, p	布拉德福定律中的参数
SBS	成功倍增
φ	洛特卡函数
$\frac{\partial}{\partial T}$	偏导
B	β 函数 $B(j, m) = \frac{(j-1)!(m-1)!}{(j+m-1)!}$
λ	勒贝格测度
D.	分形维数或相似性维数
D	豪斯多夫-贝西科维特维数
(S, I, V)	一个IPP, 这里 $S=[0, T]$, $I=[0, A]$, V 为函数 $V: S \rightarrow I$
(I, S, U)	(S, I, V) 的对偶IPP
V^{-1}	V 的反函数
o	$o = u^{\perp}$, (一般) 布拉德福函数
ρ	$\rho = V^{\perp}$
K	布拉德福定律中的任意组布拉德福因子
r_d	格鲁斯偏移的横坐标
y_m	最富集源中的项数
$m(i)$	第 i 组中的最富集源中的项数
$[r_0]$	小于或等于 r_0 的最大整数
ζ	ζ 函数 ($\alpha > 1$) $\zeta(\alpha) = \sum_{j=1}^{\infty} \frac{1}{j^{\alpha}}$
$x = x(\theta)$	广义“80120法则”

$\alpha = \alpha(\theta)$	广义“普赖斯定律”
f	一般集中测度
π	排列
V	变分系数
G _a	加斯顿测度
A	阿利森修正平方系数
K	尤拉特性(函数)
J	辛普森指数
D	舒茨系数
C	普拉特测度
G	吉尼指数
Th	锡尔测度
L	对数方差
A(e)	阿金森指数
CON	CON指数
α	洛特卡 α
P(r)	广义普拉特测度
H	熵测度
(S(t), I(t), V _t)	取决于时间的IPP
P(t)	总体容量

目 次

概述	(1)
第 I 编 统计学	(5)
概述	(5)
I .1 描述统计学	(6)
I .1.1 表格	(6)
I .1.2 测度的标度	(9)
I .1.3 图形表示	(10)
I .1.4 中心趋势测度	(18)
I .1.5 离散测度	(22)
I .2 概率论基础	(26)
I .2.1 概率	(26)
I .2.2 分布函数	(28)
I .2.3 随机变量的特征值	(30)
I .2.4 举例	(31)
I .2.5 匣子占有问题	(35)
I .3 推理统计学：假设检验与显著性检验	(41)
I .3.1 抽样	(41)
I .3.2 假设检验综述	(42)
I .3.3 中心极限定理	(43)
I .3.4 平均值检验	(44)
I .3.5 χ^2 检验	(52)
I .3.6 科莫格洛夫-斯米尔诺夫检验	(58)
I .3.7 一些其它的非参数检验	(60)
I .3.8 回归与相关	(63)
I .4 抽样理论	(75)

I.4.1	传统的抽样规则	(75)
I.4.2	福斯勒抽样法	(80)
I.4.3	重叠	(89)
I.4.4	样本容量	(93)
I.5	多元统计学	(96)
I.5.1	多重回归与相关	(97)
I.5.2	主分量分析	(100)
I.5.3	多维标度	(106)
I.5.4	聚类分析	(113)
第 II 编	运筹学与图书馆管理	(125)
概述	(125)
II.1	规划问题	(126)
II.1.1	双变量线性规划问题的图解	(126)
II.1.2	线性规划问题和简化方法的形式说明	(129)
II.1.3	整数规划	(133)
II.1.4	运输和分配问题	(133)
II.1.5	举例	(137)
II.2	最短路算法	(141)
II.2.1	图论基础知识	(141)
II.2.2	迪克斯特拉最短路算法	(145)
II.2.3	迪克斯特拉算法的应用	(148)
II.2.4	在赋权图中寻找顶点对之间最短路长度的矩阵方法	(152)
II.2.5	旅行推销员问题	(154)
II.3	排队论	(156)
II.3.1	概述	(156)
II.3.2	(M M 1) 排队模型	(158)
II.3.3	(M M m) 排队模型	(161)
II.3.4	相互独立的服务人员的合并	(163)
II.4	图书流通干扰	(164)
II.4.1	一般情况: 概念	(164)
II.4.2	第一种特例: 完全拒借	(165)

II.4.3	第二种特例: 当图书不能立即得到时, 每一位潜在的 的借阅者都进行了预约登记	(167)
II.4.4	多复本	(169)
II.5	马尔可夫过程与莫尔斯模型	(172)
II.5.1	随机过程—马尔可夫过程	(172)
II.5.2	用于图书的莫尔斯马尔可夫模型	(175)
II.6	其它图书馆流通模型	(180)
II.6.1	用于图书馆借阅的布莱尔简单随机模型	(180)
II.6.2	更精细的模型	(186)
II.7	图书馆管理中的模糊集和试探法	(194)
II.7.1	模糊集合论	(194)
II.7.2	实例: 期刊装订决策	(196)
第III编	引文分析	(199)
概述	(199)
III.1	引文标引	(200)
III.1.1	参考文献与引文	(200)
III.1.2	引文标引原理	(201)
III.1.3	SCI、SSCI的说明与使用	(202)
III.1.4	A&HCI与SCI、SSCI和A&HCI的联机版	(203)
III.1.5	主题索引的缺陷与引文索引	(204)
III.2	引文与引用者的动机	(206)
III.2.1	引用者的动机问题	(206)
III.2.2	引用者的动机研究: 特雷斯·布鲁克斯所做的工作 (1985)	(210)
III.2.3	以引文分析为基础的几种假设和与使用引文数据有 关的几个问题	(211)
III.2.4	自引与合著	(216)
III.2.5	支持引文分析	(219)
III.3	引文网络与引文矩阵	(222)
III.3.1	引文图与引文矩阵概述	(222)

III.3.2	关于引文图的几条数学定律	(224)
III.3.3	由矩阵描述的出版物及引文过程	(227)
III.4	文献耦合与同引分析	(230)
III.4.1	文献耦合	(230)
III.4.2	同引: I	(235)
III.4.3	同引: II	(238)
III.4.4	引用关系分析	(247)
III.5	科学期刊的引文分析	(249)
III.5.1	《期刊引文报告》	(249)
III.5.2	以引文测度为基础的比较的可靠性	(254)
III.5.3	除了在JCR中所发表的引文之外的引文测度计划	(256)
III.6	老化	(260)
III.6.1	概述	(260)
III.6.2	与应用于科学文献类似的半衰期	(260)
III.6.3	老化速率与半衰期的测定	(261)
III.6.4	“真实的”与“表面的”——“同步的”与“历时 的”	(265)
III.7	科学政策的应用	(266)
III.7.1	概述	(266)
III.7.2	三种加权方法的比较	(266)
III.7.3	科学产品的动态统计	(272)
第四编	情报计量学模型	(277)
概述	(277)
IV.1	情报计量学模型的试探反射与历史实例	(278)
IV.1.1	一般途径	(278)
IV.1.2	情报生产过程: 源与项	(278)
IV.1.3	经验定律与相应的数学函数	(279)
IV.2	情报计量学定律的解释	(283)
IV.2.1	成功倍增原理	(283)
IV.2.2	布克斯坦的函数解析理论	(287)

IV.2.3	芒代尔布罗的组合分形理论	(292)
IV.3	IPP的形式理论、机理与对偶性	(300)
IV.3.1	IPP的定义	(300)
IV.3.2	IPP中的对偶性	(300)
IV.3.3	纯对偶性与经典情报计量学	(303)
IV.3.4	一般对偶性及其在洛特卡定律中的应用	(304)
IV.4	洛特卡定律 $f(j) = \frac{c}{j^\alpha}$, $j \in [1, p(A))$, $\alpha > 1$ 的等 价定律	(309)
IV.4.1	$\alpha = 2$ 的情况	(309)
IV.4.2	一般情况: $\alpha \neq 2$	(321)
IV.5	情报计量学逼近	(326)
IV.6	情报计量学定律的拟合方法	(328)
IV.6.1	布拉德福定律的拟合	(328)
IV.6.2	莱姆库勒函数 $R(r) = a \log(1 + br)$ 的拟合	(330)
IV.6.3	莱姆库勒函数第一部分的拟合	(334)
IV.6.4	广义莱姆库勒函数和洛特卡函数的拟合	(339)
IV.7	应用	(345)
IV.7.1	集中理论, 80/20法则, 普赖斯定律及集中测度的 有关问题	(345)
IV.7.2	数据库的压缩	(354)
IV.7.3	风格与作者	(354)
IV.7.4	计算机存贮与文本检索	(355)
IV.7.5	布拉德福定律与抽样	(356)
参考文献	(358)
附录	(394)