

张泽厚 主编

963649

C81
1237



统计信息理论与实践

中国统计出版社

统计信息理论与实践

主编 张泽厚

中国统计出版社

(京) 新登字041号

主 编：张泽厚

编写人员：张泽厚 孙云鹏 袁 卫 倪家勋 赵彦云

杨泽军 张 勇 王毓贤 刘海云

总 纂：张泽厚 袁 卫 杨泽军

统计信息理论与实践

TONGJI XINXI LILUN YU SHIJIAN

主编 张泽厚

中国统计出版社出版
新华书店北京发行所发行
京安印刷厂印刷

787×1092毫米 32开本 6.375印张 13万字
1992年11月第1版 1992年11月北京第1次印刷

印数 1—3000

ISBN 7-5037-0753-4 / C·466

定价：6.80元

目 录

一、信息的基本概念	(1)
(一) 信息概念的发展	(1)
(二) 信息分类	(9)
二、统计学的发展	(14)
(一) 古典统计学	(15)
(二) 近代统计学	(19)
(三) 现代统计学	(25)
三、统计学与信息论的关系	(29)
(一) 申农信息与费暄信息	(30)
(二) 信息论与数理统计学的关系	(38)
四、统计调查与概率信息决策	(44)
(一) 统计调查	(44)
(二) 概率和信息的基础知识	(51)
(三) 统计决策基础知识	(71)

五、统计信息分析方法	(77)
(一) 几种常用统计量	(78)
(二) 参数估计和假设检验	(80)
(三) 方差分析	(91)
(四) 回归分析与相关分析	(99)
六、统计信息是社会经济信息的主体	(116)
(一) 统计信息的主要特点和范围	(116)
(二) 社会经济信息的主要特点和范围	(125)
(三) 统计信息在社会经济信息中的主 体地位	(127)
(四) 统计信息在社会经济决策和调控 中的重要作用	(131)
七、强化统计整体功能，改善宏观经济管理	(137)
(一) 统计部门的职能作用	(137)
(二) 强化统计部门的职能作用是 经济改革形势的要求	(142)
(三) 强化统计部门的职能作用是改善 宏观管理的要求	(146)
八、加强统计信息与社会经济信息的管理	(150)
(一) 信息管理概念	(150)
(二) 国外统计信息与社会经济信息的 管理状况	(152)

(三) 我国统计信息系统的发展	(156)
(四) 加强我国统计信息系统建设 和社会经济信息管理	(159)
附录 1 熵与方差	(164)
附录 2 引自有关参考文献中的概念	(190)
后记	(198)

一、信息的基本概念

(一) 信息概念的发展

本书探讨的内容是统计信息在社会经济信息中的地位和作用。因而有必要先把信息的概念弄清。信息的概念是随着信息在社会发展中的作用和信息科学的发展而形成的，由于信息科学是一门新兴的学科，虽然近年来关于信息理论的研究有很大的进展，但是还未形成比较系统的信息科学理论，信息的概念也未最后形成。从目前公开出版的文献来看，信息的定义已不下数十种，但迄今为止尚没有一个普遍公认的定义。但从发展的过程看，人们对它的认识是在不断深化。从文字出现信息这个词到现在，信息的含义大致可划分为三个阶段。第一个阶段只是从语义学上产生信息这个词，把信息作为消息的同义语；第二阶段是以申农(Shannon)为代表的信息论的产生，我国把 Information^{*}一词译为信息，就给信息赋予了比较科学和精确的定义；第三阶段是广义信息科学的发展突破了申农信息论的范围，要求在更高的层次上来定义信息。

* 香港和台湾将 Information 译为“资讯”。

在人类漫长的历史发展过程中,信息本来是客观存在的,因为人们为了维持自身的生存和发展,就要不断地从它所处的外部环境中获得有关情况和消息,以便作出相应的行动以适应环境的变化。例如,原始的人类为了生存要用自己的感官来获取消息以寻找充饥的野果和食物,躲避凶猛野兽的进攻和避免各种灾害的侵袭,并且把这种消息在人群中相互传递,以便协同一致去采食猎物。这种消息实质上就是信息,只是那时候还没有信息这个名词而已,那时的传递信息也比较简单,只是利用人们的感觉器官,所以只能局限在同时同一地点面对面地传递信息。文字的产生使人类传递信息的技能产生了一个飞跃,使得信息可以记录下来横向远距离的传递和纵向由前人向后人传递信息。从古书的记载中我们知道我国在商周时代就建立了通信制度。人类每一个科学技术的发明也都与信息的传递密切相关。如指南针的发明使人们识别方向信息的能力增强,纸张和印刷的发明使人脑贮存信息的功能增强,望远镜和显微镜的发明使人类接受信息的能力有了扩展,因而信息对人类所起的作用也在日益加强。至于在我国文字中何时开始出现信息一词,目前尚无确切的考证。但从现有的资料发现,在一千多年前唐代诗人就曾有“梦断美人沉信息,目穿长路倚楼台”的诗句。在古典文学名著《水浒传》和《红楼梦》中也都有信息一词的出现。《水浒传》(120回本)的第56回“吴用使时迁盗甲,汤隆赚徐宁上山”中有这样一段描述:“这日徐宁正在家中纳闷,只听得有人扣门……汤隆见了徐宁纳头拜下,说道‘哥哥一向安乐?’徐宁答道:‘闻知舅舅归天去了,一者官身羁绊,二乃路途遥远,不能前来吊问。并不知兄弟信息,一向正在何处?今次自何而

来?’”《红楼梦》第16回“贾元春才选凤藻宫，秦鲸卿夭逝黄泉路”中有以下叙述：“……贾母又唤进赖大来细问端底，赖大禀道：‘奴才们只在外朝房伺候着，里头的信息一概不知’”。显然，以上的信息都是指音信和消息的含义。解放前出版的汉英新辞典中，信息还包括智力(Intelligence)的解释(见商务印书馆1918年李玉汶著的汉英新辞典)。从英文中相应的词“Information”来看，据韦勃氏辞典记载，约在14世纪将该词收入英语词典。那时信息论尚未产生，所以也只能是语义学上的含义。

关于信息的科学概念则是与信息论的产生相联系的。人类进入19世纪后，随着电的广泛应用，通信技术出现了新的飞跃，从通信的发展历史看有以下重大事件*：

- | | |
|------------|-------------------------|
| 1844年5月24日 | 美国发送了第一个实用的电报，从华盛顿至巴尔的摩 |
| 1858年8月5日 | 第一条穿越大西洋的电报电缆完成 |
| 1876年 | 贝尔宣布电话的发明并登记了专利 |
| 1878年1月28日 | 美国康涅狄克州纽黑文市第一个商业电话局成立 |
| 1895年 | 第一次无线电信号的传送 |
| 1901年 | 第一次穿越大西洋的无线电报 |
| 1904年 | 第一次照片电传 |
| 1915年1月25日 | 从纽约到旧金山横穿美国大陆的电 |

* 资料摘自科学百科辞典 Van Nostrand's Scientific Encyclopedia 第6版。

话通话

1920年8月20日 美国建立了第一个广播电台

由于信息传送技术的发展，如何提高通信效率的通信理论引起了人们的重视，一些人开始探索度量信息的科学方法。20年代美国的奈奎斯特和哈特莱曾就通信系统的传输能力提出度量信息的理论，40年代，美国贝尔电话实验室的工程师申农发表了《通信的数学理论》一文，进一步详细讨论了信息的度量和通信的有关理论问题，奠定了现代信息论的基础，他把信息定义为事物不确定性的减少，从而对传递的信息可加以计量，把信息论建立在科学的基础上，说明信息并不等同于消息，信息是消息的内容，而消息则是信息的形式。一条消息可能包含很多信息，也可能信息很少。同一个信息可以寓于不同形式的消息之中。信息的计量对信息论的形成起了奠基作用，正如物理学中的热学一样，虽然人们很早就有“冷”和“热”的感觉，但只有温度概念的提出和温度计的发明，找到确定冷热程度的方法和进行客观的计量以后，热学研究才成为一门科学。申农信息论在理论上是严谨的，对推动通信技术，提高通信效率发挥了重要的作用。

第二次世界大战以后，通信技术又有了进一步的发展，特别是电子计算机的产生起到了人脑功能延长和扩展的功能使信息在社会发展中的作用进一步增强，有人提出了社会在向信息社会发展，它有以下一些特征：

1. 信息传递的手段进一步多样化，效率进一步提高，信息的传播已能在短时期内冲破地域和国界辐射到全球的范围，从而大大促进了世界范围经济科学文化艺术的交流。例如1946年出现了转送信号的人造卫星；1951年美国建立了

第一个电视台；60年代发射了通信卫星；70年代将光导纤维用于电话系统。这些先进技术也很快在全世界传播。其他许多传递信息的新型技术（如报纸传真、激光通信、侦察卫星等）也得到了广泛应用。

2. 信息已渗透到各行各业，并日益发挥巨大作用，使得信息量急剧增加，知识更新的时间大大缩短，信息成为现代社会的重要资源，是一种无形的财富。做任何工作必须十分重视信息。对企业家来说，信息是影响生产力的重要因素，能否充分利用信息关系到企业的成败；对军事家来说，信息就是力量，它关系到战争的胜负；对领导者来说，信息就是主意和办法，是进行正确决策的基础；对教育工作者来说，信息就是知识，它关系到教育的质量。人们开始认识到信息与物质、能量相并立，是客观世界三位一体的组成部分之一。物质依靠能量才能运动，而物质能量要靠信息进行联系才能更好地发挥作用。

3. 一个围绕信息的生产、加工、传递和流动服务的产业正在形成。它包括信息设备制造业（如计算机制造及外部设备集成电路；通信、广播设备；其他电子信息设备等），信息传播报导业（如新闻、广播、出版、印刷、音像视听等），信息技术服务业（如信息处理服务、信息技术开发、软件服务、计算机销售维修等），信息流动服务业（如图书馆、情报所、邮政、电信、函授教育等）。而且这一产业在整个国民经济中的比重在日益增加。

由于上述的发展变化，对信息的研究已远远超出了申农信息论局限于通信工程的范围。一门广义的信息科学正在兴起，它是在信息论、控制论、计算机科学、仿生学、人工智

能和系统工程学的基础上发展起来的边缘科学，以信息为研究对象，以信息的运动规律和应用方法为主要内容，研究信息的取得、传递、存贮、变换、处理、利用和控制的一般规律。没有信息论对信息的定量描述，就不可能对信息的传输和处理进行严密的分析，也就不可能有信息科学。但信息科学比信息论研究的范围更为广阔，涉及的内容也更为深刻更为复杂。为此，申农的信息概念就显示出它的局限性。有人把信息分为语法、语义和语用三个层次，语法是指传递信息符号的语法结构，这是信息的形式方面；语义是指如何使传输的符号精确表达信息的含义；语用是指信宿收到信息含义如何影响它的行为问题。而申农的信息只解决了通信的技术问题，即语法问题，并没有涉及语义及语用问题，有人说，按照申农的信息概念，一个人知道自己的妻子生一男孩还是女孩，和知道掷一枚硬币出现正面还是反面其信息量相同（因为其不确定性相同，均为 $1/2$ ），显然是不恰当的，这主要是没有考虑语用的原因。早在 1953 年哥尔德曼在《信息论》一书中已指出，相同内容、相同比特数的信息对于不同的收信人有不同的价值，申农的信息概念只反映量的信息，没有考虑信息的价值。诺贝尔经济学奖获得者美国的肯尼斯·阿罗也认为“申农度量法——它在通信工程领域已经变得十分有用——对经济分析而言，并不是普遍适用的，因为它忽视了信息的价值。”*为了解决申农信息的这一局限，1968 年贝里斯和高艾斯曾提出了同时考虑信息的量和质的问题，试图建立“量—质”统一量度的信息概念，对每个信息 x_i ，引进

* 肯尼斯·阿罗：《信息经济学》第 159 页，北京经济学院出版社 1989 年版。

一个效用系数 U_i ，表示 x_i 出现对信宿的效用。令

$$U = \{U_1, U_2, \dots, U_n\} \quad U_i \geq 0$$

为信源的有效性分布，则信源的信息结构应同时考虑其不确定性和效用系数，表述如下：

$$S = \begin{pmatrix} a_1, a_2, \dots, a_n \\ u_1, u_2, \dots, u_i \\ P_1, P_2, \dots, P_i \end{pmatrix} = \begin{pmatrix} A \\ U \\ P \end{pmatrix}$$

有效信息定义为：

$$I(P, U) = - \sum_{i=1}^n U_i P_i \log P_i$$

即以效用系数加权修正的方法使在一定程度上反映了信息的价值，显然当 $U_1=U_2=\dots=U_n$ 时，则有效信息即为申农信息，可见申农信息是有效信息的一种特殊情况。当所有的 $U_i=0$ 时，不论信源中的平均信息量有多大，对信宿的价值为0。但这种方法对如何确定信源的有效性分布仍然是一个困难的问题。关于信息的语义问题卡尔纳普于1964年提出来，但研究表明，虽然定性地描述语义信息并非难事，但定量地描述语义信息却相当困难，至今没有多少实质性进展，不能与申农信息论相提并论。一种方法是仿效申农的思路，用逻辑概率代替统计概率，用状态描述代替可能事件，制定一种对消息的逻辑真实性和精确性作某种度量的方法*。

除了对申农的信息的计量方法加以补充以外，信息科学的发展也要求对信息有一个概括的定义。为此世界各国信息

*参见钟义信：《信息学漫谈》第97—100页，科学普及出版社1984年版。
本文转引自《系统科学原理》第164页，中国人民大学出版社1990年版。

科学的研究工作者也对信息下了各种各样的定义。这里值得一提的是美国著名的哲学家、数学家，控制论的创始人诺伯特·维纳(N. Wiener)曾给信息下过一个定义：“信息就是我们在适应外部世界，并使这种适应反作用于外部世界过程中同外部世界进行交换的内容的名称”^{*}。维纳的定义概括了信息的特征，但是没有解决信息的计量问题。此外，人同外部世界的交换不只是信息，也有物质和能量，目前普遍认为物质、能量和信息是三位一体存在于客观世界的，维纳的定义中混淆了信息与物质、能量之间的界限。我国的钟义信在概括了各种意见以后将信息定义为“事物运动状态和过程以及关于这种状态和过程的知识，它的作用在于消除观察者在相应认识上的不确定性，它的数值则以消除不确定性的大小或等效地以新增加的知识多少来度量。”^{**}我们认为这个定义既概括了信息的实际含义，也包括了申农的定义，目前看来还是比较全面的。

最后，在确定信息概念的时候，还必须注意到信息已渗透到其他学科的特点，而专门学科所关心的往往是信息现象的一个侧面或一个层次，所以在不同的学科中又有不同的信息定义，如生物学、计算机科学、管理学、统计学等对信息的解释各有特点，所以又形成信息的多定义。在词典中对信息的解释也不止一个，现以比较权威的英国韦勃氏大学辞典中对信息(Information)一词的解释摘录如下：^{***}

“1. 知识或智力的交流或接受。

^{*} N. 维纳《控制论与社会》，转引自钟义信《信息的科学》。

^{**} 《信息科学：它的内容方法和意义》北京邮电学院学报1984年3期。

^{***} 根据 Webster's Ninth New Collegiate Dictionary 翻译。

2. a. (1) 从调查、研究或教育得到的知识

(2) 智力、消息(新闻)

(3) 事实、数据

b. 在两个以上序列或某些东西(如DNA中的核苷酸或计算机程序中的二进位数)的排列的任一个中存在并传送的内在属性。

c. (1) 在一个通信系统或计算机中代表数据的一个信号或字符。

(2) 证明表示物质或精神经历的一个构造(如一个计划或理论)中的变化是正确的那些东西(如一消息、试验数据或一图象)。

d. 信量含量的一个数值测量。具体地讲,在完成一个试验结果中计量其不确定性的一个数字计量。

3. 告发某一个人的行为。

4. 由检察官作出的犯罪的正式起诉,区别于大陪审团作出的告发。”

从辞典的解释可以看到,除了对信息一般的语义解释外,还包括了计算机科学、生物学、法律学等学科的解释,其中也包括了申农信息的解释。在广义的信息概念基础上,研究社会经济信息与统计信息的关系,不得不涉及信息的子概念问题,而这方面问题的根本解决在于科学的信息分类。

(二) 信息分类

1. 信息分类的基本原则

信息分类的基本原则就是分类的客观依据。从目前看,

对信息分类客观依据的认识具有发展的相对性。虽然信息科学或信息理论发展迅速,但是,这种发展只是在狭义信息论及其应用技术方面比较成熟,而将狭义信息论向更大范围——人类活动整体上扩展却是很不成熟的,因而在信息及其分类的认识上带有一定程度的主观随意性。对目前关于信息分类的讨论作一简单归纳,有以下几种:第一,把信息按照其性质分为自然信息与物化信息(或称为社会信息),自然信息包括生物信息等,物化信息包括逻辑思维信息与形象情感思维信息。第二,把信息按照信源、信道、信容、信义、信宿可以分为不同的信息类型。第三,把信息按人类活动和社会活动的不同领域分为经济信息、社会信息、政治信息、军事信息、科技信息、文化信息、环境信息、外交信息等。第四,把信息按照信息的层次和计量方法分为物理信息、语法信息、语义信息、语用信息和模糊信息等。第五,把信息按照控制论闭环动态系统的需要分为必要信息、充足信息和多余信息。第六,把信息按照信息经济学关于最优政策形成过程三类时间因素分为连续信息、计算信息和累积信息……。由此可见,信息分类所应依据的原则,还有待进一步研究解决。从发展的观点,我们认为信息分类的主要客观依据是:

第一,以信息生成标志为依据。从人类活动整体出发,信息无论在认识客观方面,还是在组织活动方面,都有其生成的标志,例如,统计信息是运用统计方法认识数量方面而生成的信息,其生成标志就是统计认识方法。信息在信息理论出现以前就存在,但处于一种不自觉的发展过程,而现在则需要从过去信息发展的不自觉过程转化成自觉过程,以便提高人类认识客观和简化交流与组织活动的水平。正是出于

这样的目的，需要从信息生成的客观依据去认识信息整体及其结构。因此，信息生成标志是研究信息分类的首要依据。

第二，遵循系统的原则。信息论的产生和发展表明，认识信息的整体性以及开发信息的整体功能，提高信息的整体效率和加速信息的整体传输，都是依据系统原则认识与开发和提高的。即使将信息扩展到社会经济领域，这一原则也必须遵守，因为当今信息时代也同样是根据系统的信息认识客观规律及其决定关系。还应看到，信息从不自觉向自觉的发展过程，一个根本方向就是不断扩大信息系统和强化信息系统功能。因此，信息分类应为信息系统的设计和开发服务，遵循系统原则。

第三，满足国家管理的需要。这是从信息分类的目的性考虑的。目前，从广泛的信息作系统的归类研究，不能没有目的性。从最高层次国家管理的要求去研究信息整体，是非常重要的。满足国家管理需要并不等于按国家管理体制进行信息分类，而是体现国家管理职能的充分发挥，在一定意义上，信息系统职能的发挥为国家管理体制的改革提供客观依据。因此，信息分类兼顾国家管理需要，对提高国家管理的整体功能和效率是非常重要的。

2. 信息分类体系

根据上述基本分类原则，我们将信息分类整理归纳为图1。

图1反映了统计信息与社会经济信息、信息之间的关系。

(1) 社会经济信息的概念

社会经济信息应从它生成的客观基础上去认识，这就是