Report and Opinion

Websites:
http://www.sciencepub.net
http://www.sciencepub.net/report

Emails:
editor@sciencepub.net
reportopinion@gmail.com

MARSLAND PRESS
Multidisciplinary Academic Journal Publisher

个体集合和它的复杂程度
----《空中水文学初探》附录

张学文（新疆气象局）

Recommended: 王德奎 (Wang Dekui), 绵阳日报社, 绵阳, 四川 621000, 中国, y-tx@163.com

摘要：这里简单地点到了个体集合和它的复杂程度的应用，提供一组具有普遍意义又互相有联系的定义、概念、公式和原理，它们对于理解《空中水文学初探》书中的降水统计力学等内容有帮助

[张学文. **个体集合和它的复杂程度----《空中水文学初探》附录**.*Rep Opinion* 2023;15(1):9-19].ISSN1553-9873 (print);ISSN2375-7205(online).http://www.sciencepub.net/report. 02.doi:10.7537/marsroj150123.02.

关键词：个体、集合、标志值、特征值、复杂程度法

【0、引言】

现代科学之所以强大，除了它尊重“事实、测量、规律、推理”之外，还十分关注科学概念的提炼。提炼科学领域的基本概念具有原始创新意义。现代科学被划分为数千个分科，每个分科都有专属自己的若干专名词。它们在本学科内十分有用，但难以跨学科应用。

显然，能发现和提炼具有跨学科应用能力的定量概念，就对各个学科的发展和各个学科的统一具有十分重要的价值。

这里推荐的“个体”概念、“个体-标志值-集合”概念以及关于“个体-标志值-集合”的“复杂程度”概念等，都是可以横跨许多学科的通用性十分强的量化的概念。它们能把过去说不大清楚的“事”和“物”提炼为由含义清楚的“个体-标志值-集合”，并且送给你一些关于它们的“量”（如复杂程度）。而这些量化的对象就为研究、分析、发现新的定量规律做了准备。

“‘个体--标志值--集合’和它的复杂程度”是一组概念模型，它们不仅可以在各个学科（自然的、社会的）中找到自己的身影，而且也能派上新用场、充当新的思维工具、计量工具、推理工具。它是《组成论》里的概念和原理部分的发展、提炼。

这里要突出这样一些认识：

㈠把“个体”，以及“复杂”这两个词提升为社会和自然科学通用的、量化的科学名词（概念）。

㈡特意把以个体为元素，把每个个体的标志值都明确的集体称为“个体--标志值--集合”（简称个体集合或者个体集）.

㈢给复杂程度概念以定量的定义，最后把这些认识与信息熵、热力学第二定律等概念和原理再横向地串起来。

这里也谈及有关的原理和应用，但是重心是新概念的引入。

【1、把“个体”提炼为科学名词（科学概念）】

1. 个体

“5 个苹果，3 个梨子”，这里的每个苹果、每个梨子都是“个体”的特例。确实，小学生也知道“个体”二字的大概含义。生物个体、我国改革开放初期的“个体户”都是有关“个体”的著名例子。

我们几乎找不到一个领域不能使用个体这两个字。对这个使用领域非常宽的词，进行一番科学提炼，说不定可以惠及各个科学学科。

2. 个体概念的定义

把个体提升为一个科学名词，就要给它更清楚的定义。在文献的基础上，这里把个体定义为：个体是总体（宇宙、世界、群体）的某一部分，它可以独立存在、与外界有比较清楚的边界、在边界内有比较清楚的特征（特定质地、性质、结构、功能、目标……）。

3.个体的例子

每个电子、每个分子、每个细胞、每个生物体、每个人、每个城市、每个国家、每个星星、每个星系都符合个体的定义。所以过去在科学、生活和社会实践中，大家理解的各种“个体”都符合这里的个体的定义：有 10 张 100 元的钞票。每张钞票有清楚的边界，而且它们印有特定的图案文字，它们独立存在、使用，所以每张钞票就是一个个体。教室坐着 20 个学生，每个学生与其他学生都有清楚的分界，而且每个学生有特定内质，可以独立活动，自然应当承认每个学生是一个个体。

符合这个定义的个体不只这些类型。例如一个骰子，把它掷了 50 次，每次都与其他各次过程无关，应当承认每个过程与另外的过程有清楚的分界，每次过程都有确定的“点数”向上，每个掷骰子的过程都是一个独立过程。这种彼此独立的过程也可以算是“过程（事件）个体”。晚会上演唱了 10 个歌曲、我下了 5 盘象棋，这里的每个歌曲的演唱、每盘棋

都是一个过程（事件）个体。

英文里有 26 个字母，最近做了很多梦，大家提出了很多办法……这里的每个字母或者每个梦或者每个办法，都与抽象世界中的其他部分有清楚的边界，并且有特定的内容，而且它们独立存在。

这些抽象对象也符合个体的定义。它们算"抽象个体"。

很多文章里用到"我"，"我"就是个体的个例。

4. 著名的个体

100 元一张的人民币，是个体。一个人，是个体。电脑、手机、飞机、轮船、人造卫星都是著名的个体。统计学里的一次采样结果、一届奥运会、EXCEL 软件、一个歌曲、一次太空飞行都是著名的个体。

系统科学研究"系统"，而每个具体的"系统"也是相对意义下的个体。一个国家就是一个个体，多少人为建立一个国家而牺牲了自己；一个公司就是一个个体，多少人为建立一个公司而费尽心机，确实创立一个新个体（不是复制），一个在环境中得以稳定生存，以致发展的个体几乎都有一个不平凡的故事。

把"个体"作为科学名称来对待，我们就为研究一大类现象，找到了共同又精练的语言。"个体"应当成为科学领域的著名词汇。

5. 说明

1）个体的命名：

只要有需要，我们可以为每个个体专门取一个特定的名词代表这个相对完整的对象。每个人有姓名。这个姓名不是属于某人的某部分而是该完整对象。它体现着个体具有整体完整性。对已经发现的每个星星有命名、对每个城市、国家有命名都是例子。对每个个体临时给个号码，这也是对各个个体的命名。

2）个体的类别：

某些个体如果具有某种一致的特点，我们经常为该一批、一类个体取相同的名字。当我们说到 5 个苹果或者 7 个羊、9 个星星、3 个电子时，苹果、羊、星星、电子是一类个体的统称。

在该命名下的每个个体具有相同的地位。分别研究不同类别的个体，是科学得以分科的重要标准。

3）个体的边界：

既然"存在边界"是个体的共性，单独研究边界自然是个合格的科学课题。曹鸿兴著《系统周界的一般理论界壳论》开此研究之先河。

4）个体的内容：

边界把个体与外界环境分开，而个体内部是具有特定质地、性质、结构以至功能、目标。每个苹果、人、星星、国家都有内部特有的质地、性质、结构以至功能、目标，这些内容太丰富了，它们是各个具体科学分科的研究内容。我们在这里分析涉及"个体"的一般特点、统计量和规律，但是不是代替各个学科对它们的分门别类做研究。

6.个体的重要性质

1）个体的完整性：

个体这个名称是赋予该完整对象的，这个对象具有相对意义下的不可再分性。个体的存在体现着它的完整性。再分就失去了原来的定义和意义。我们不能说那里有 1/3 个人，不能把一张人民币撕开，一个分子再分就变成原子了。科学界经常用"量子"一词，量子就是强调研究对象是以完整个体的形态存在。可以说，量子就是个体。

2）个体的相对性：

我们所谓的个体其实仅是个相对概念。你可以把银河系这个完整整体看作是一个"个体"，并且称它为银河系。但是也可以在另外的层次上，把银河系看作是一个集体、总体，而它里面的各个恒星（如太阳系）才是其中的一个一个的个体。这一切取决于我们从什么层次上分析问题。在一定的场合我们把一个人看作是个体，而在必要时又以一个人体器官、一个细胞、一个生物大分子为个体。

所以个体概念有它的相对层次性。当代科学的一些分科就是在不同层次上研究不同层次意义下的个体。

3）个体存在的量化表示与可加性：

"存在"与否是涉及个体的基本问题。对这个问题可以用量化的办法表示："1"与"个"连在一起表示该个体的存在。如一个苹果、一个人等，这里的"一个"二字体现着"苹果"或者"人"的存在。如果该个体不存在，就用"0 个"表示，如我家有一个厨房、一个卫生间以及 0 个飞机和 0 个银行。这里的 1 和 0，不仅具有符号意义，而且具有"数"的意义，同类的独立存在的个体可以合并，其对应的 1，0 是可以做加法的（我们已经无形中引入了关于个体的运算了！）。

7.个体的量纲----个和量词

1）个体的存在与量化表示的深化

前面谈了个体的存在与否的量化表示，而这种表示还涉及了对个体的数量的计量单位。这个单位就是"个"。当我们说有 1 千克牛奶时，我们是以千克作为度量牛奶的单位的。类比地说，当我们说这里有 8 个苹果时，这里的"个"字就成为计量独立存在的苹果个体的数量单位了。从学术角度看，"个"就是（也是）一种"量纲"。它是关于个体的量的单位。所以当我们把个体提升为科学名词以后，"个"字就成为计量个体数量的单位（量纲）。

2）10 个人、100 个苹果以及 6.022×10^23 个氧原子

这里的"个"字的含义大家都明确。但是 6.022×10^23 个氧原子也就是 1 摩尔氧原子。这说明"个"这个新的单位原来与化学中用的摩尔有着本质上相同的意义，其区别仅在于"个"的 6.022×10^23 倍

是摩尔。这类似在天文学里用光年计量长度，而在生活中用米计量长度。请注意摩尔是 7 个基本科学单位之一，它在科学中有基础地位。

所以个体的单位“个”，早已经穿着摩尔的外衣进入了科学领域。这也提示我们：个体一词进入科学殿堂的前期准备工作已经就绪。

我们建议用小写的英文字符 ge（个）作为计量个体的单位（Ge 是化学元素锗）。

3）中文里有很多“量词”

3 匹马、2 架飞机、4 条鱼、5 只鸡都是例子。匹、架、条、只，这些词是所谓量词。可以认为它们是“个”这个一般单位在特殊场合的特殊运用。它们与摩尔类似，都是“个”的特例。在明确了个体概念以后，我们看到“个”是比摩尔的通用性更广的量纲。

【2、“个体--标志值--集合”】

1. 个体的标志

根据个体的定义，每个个体有特定的边界和内在特征，所以每个个体都根据其边界或者内部特征而具有若干个特征“标志”。

太阳系的每个行星都是一个个体。而这里的每个行星都具有自己的质量、体积、自转周期、它与太阳的距离等这些“特征标志”。

对于本班的同学，这里的每个同学都有特定的身高、体重、年龄等。身高、体重、年龄就是这些个体都具有的统一的特征标志，而质量、体积、自转周期、与太阳的距离是每个行星都具有的特征标志。

确实，我们都是通过各个特征来认识各个个体的，所以“个体具有特征标志”的几乎是不言自明。

2. 个体的标志值

对于每个个体，它不仅具有若干个特征标志，而且对每个标志，在确定的时刻（充分短的时间段落内），它都有具体的、确定的特征值。例如对本班的学生甲（个体），他身高 1.2 米，体重 24 千克，年龄 8 岁，这里的 1.2 米就是身高这个特征标志的标志值。24 千克，8 岁分别是体重和年龄的特征值。365 天自转一圈、半径 6370 千米分别是地球这个个体的自转角速度、半径这两个特征标志的标志值。

1）个体－标志－标志值举例：

a、一个学生----学生的体重----体重 37 千克；

b、一本书----书的价格----15 元；

c、一张选票----选举结果----赞成票；

d、一张人民币----人民币票面值----五元；

e、一升海水----海水的温度----23 度；

f、一张麻将牌----条、并、万等----柒条。

在上面中，列了一些个体的特征标志和特征值。这里的多数例子的特征值是有单位的具体数值量，这些与物理量十分类似。而选举的例子中，其标志值不是数值量而是“字符串”（文本）。这说明特征值也可以是含义明确的“字符串”，而不限于是有单位的“数”。

2）在统计学和概率论里人们用随机变量概念的标志

这里相当于哪里的“随机变量”，而标志值是随机变量在特定时刻或者特定的抽象试验的结局的取值。以后我们把“随机变量”与“标志”两个词混同使用。请注意这里谈的标志值都是针对该个体在特定时刻（特定场合）而言的。不同的时刻该个体可能有不同的标志值。

3. 同类个体

定义：具有某些相同的特征标志的个体称为同类个体。

同类个体往往有集体的名称（语言学里称为集体名词，以与特有名词相对）。行星、苹果、学生、劳动者、乘客、家用电器等这类个体的统称，都是它们的例子。

4. 个体--标志值--集合（个体集）

定义：由若干个（有限个）同类个体组成的集体里，如果就某一标志（或某些标志）而言，每个个体在确定时刻有确定的标志值，就把该集体称为“个体--标志值--集合”，简称为“个体集”或者“个体集合”。这里的 3 只羊的体重分别为 5、28、34 千克。每个羊作为个体而存在，体重是它们都具体的特征标志，而 5、28、34 千克是每只羊的标志值（这里的标志值是有单位的数量）。

它们就组成了一个个体--标志值--集合，即个体集。有 5 个球，3 红，2 黄，球是个体，颜色是它们共同具有的标志，红色、黄色是它们各自的标志值（这里的标志值不是数，是字符串）。所以这 5 球组成的集体是一个个体--标志值--集合，或者说是个明确的个体集。

已经知道本班 34 位同学的身高，这就是一个个体集。已经知道太阳系的各个行星的直径，也构成一个个体集。依然是这 34 位同学，知道每个学生的体重，也构成另外的一个“个体-标志值－集合”（个体集）。《组成论》里讲的广义集合就是这里的个体集。它是关于确定时刻的若干同类的个体以及各个个体的标志值的知识。

个体集内的每个个体具有的标志可能不只一个（如每个学生不仅具有身高，还具有体重、年龄……），但是为了集中研究特别关注的侧面，通常只分析它们的一个或者两个等少数标志，而暂时回避其它的标志。数学里有集合概念。这里的“个体--标志值--集合”既借用了集合的概念，又突出了以个体为元素的这个物理特点，它还涉及了标志值概念，它是集合概念向物理内容的靠拢。

5. 个体集的例子和类别

1）以下是六类个体集的特点

《组成论》提供了六类个体集（广义集合）的大

量例子，如个体集类型----个体的特点----标志值的特点----例子：

a、物质组成----物质名词（种）----物质名词（属）----三个苹果、两个梨；

b、时空场----空间、时间单元----空间或者时间的编号----一年有四季、礼堂座位编号；

c、运动----时间单元----物体在空间的位置----一辆运行中的公共汽车；

d、物理场----空间单元（有时也是质量）----物体在该处的特征值----一张有等高线的地图；

e、随机实验与概率----每次随机实验----实验的结果----掷一次骰子；

f、抽象事物----视问题而定----视问题而定----晚会上演唱了 5 个歌曲和 3 个小品、英文里有 26 个字母。

个体集的分布函数，对于每个明确的个体--标志值--集合（个体集），我们都可以提出这样一个问题：在确定的时刻，具有不同标志值的个体各有多少？既然每个个体在确定时刻的标志值是多少（是什么）是确定的（有时是已经知道的），自然该个体集内的各个标志值 x_i 和它对应的个体的数量 n_i 的关系也是确定（唯一的）。

如果已经知道每个同学的身高，自然知道本班的不同身高的同学各有多少，既然知道每只羊的体重，自然知道不同体重的羊各有多少。所以上面的问题必然有答案。这个答案确定了各个标志值与个体数量的对应关系，这个对应关系符合对函数的一般定义，我们把这个涉及个体集的特定函数称为该个体集的分布函数。这里的标志值是自变量 x_i，而各个标志值所对应的个体的数量 n_i 是函数值。

2）根据这个说明，前面的两个个体集的例子表示

这里的分布函数，如“三只羊（5 个球）组成的个体集的分布函数”可分别表示为：

a、三只羊的个体集----自变量（体重/千克）----5；28；34。

三只羊的个体集----函数值（羊的个数）----1；1；1。

b、5 个球的个体集----自变量（颜色）----红色；黄色；其它颜色。

5 个球的个体集----函数值（球的个数）----3；3；3。

3）在每个确定时刻，每个个体只能有唯一的标志值

因不能是多值，所以分布函数是单值函数。每个个体集在确定时刻有唯一的分布函数。它告诉我们标志值 x_i 的个体的数量 n_i。

a、用分布函数的表示这个函数就是：

自变量----x_1；$x_2 \ldots x_i \ldots x_p$。

函数值----n_1；$n_2 \ldots n_i \ldots n_p$。

b、用公式表示这个函数就是：

$$n=f(x) \qquad (1)$$

这里的函数二字的含义与数学、逻辑中的含义是一致的。

既然每个个体集都伴有唯一的单值的分布函数，我们也就认为：知道了分布函数也就确定了该个体集。

4）个体集的分布函数的例子

如与分布函数对应的问题----标志值----函数值意义：

a、不同身高学生各有多少（如三年级）----学生身高----学生数量；

b、不同的年产值的企业各有多少----企业年产值----企业数量；

c、不同人口数量的国家各有多少----国家的人口数----国家数量；

d、不同吨位的轮船各有多少----吨位----轮船数量；

e、不同质量的恒星各有多少（如银河系）----质量----恒星数量；

f、不同温度的日子各有多少（如 1 年中）----温度----天数；

g、不同海拔高度的面积各有多少（如中国）----拔海高度----面积；

h、不同随机变量 x 值出现概率是多少----随机变量 x----出现概率。

标志值可以是离散变量，也可以是连续变量，而函数可以是连续的也可以不连续的。上面最后一个例子把所有的概率分布函数都归入个体集模型内。但是这里在用词上与那里有一些差别。这里的连续型分布函数对应于那里的概率密度分布函数（那里的概率分布函数是概率密度分布函数的从变量下限开始到目前值的积分）。（旁白：我们讨论大家早就熟悉的个体二字，现在居然发现同类个体的集合必然存在着一个函数，好像我们走了一段熟路可居然看到了新景，有趣！）

【3、个体集的表示、运算和特征量】

在数学中一切进步都是引入表意符号后的反应（皮亚诺 G.Peano 语），让“个体”概念进入科学领域，还需要引入有关的“符号”以及量化、运算方法。

1. 字符多项式

我们一般地把有 p 项组成的 $a_1x_1+a_2x_2+\ldots+a_ix_i+\ldots+a_px_p$ 的式子称为字符多项式。在外型上，字符多项式类似初等代数里的多元一次多项式，但是它们的各个 a_i 和 x_i 都可以是有一定意义的字符串（不再单纯是数了）。而+号也需要针对需要去定义（说明）。由于它们经常不再是“数”，所以连写在一起，也（一般）不具有数的乘法的意义。

这里有 9 个水果：3 个苹果，2 个梨，4 个香蕉，可以用字符多项式写为（3 个）（苹果）+（2 个）（梨）+（4 个）（香蕉）。它对应的字符多项式里的字符串例子，见各个字符串在字符多项式一般公式中的含义如：符号 a_1—a_2—a_3；x_1—x_2—x_3：

含义—3 个—2 个—4 个；苹果—梨子—香蕉。

上面字符多项式里的"+"号具有"还有"的意义（不是强行把含义不同的东西做代数加法）。在算术运算里 3 个苹果不能与 2 个梨做加法，因为它们的单位不同。但是这里对加号有了另外的理解，于是 3 个（苹果）+2 个(梨)就是合格的表达式了。但是，至少可以这样认为，3 个（苹果）+2 个(梨)＝2 个(梨)+3 个（苹果）。在文献《字符多项式与表格数学》中，初步讨论了字符多项式和它的一些应用。那里还指出它可以表示各种表格。下面我们要用它表示个体集。

2.　个体集本身的符号表示

对于个体集本身，我们一般用被方括弧包起来的大写粗体的字母表示它（这与集合的表示类似）。如用[A]表示盘子里有 9 个水果：3 个苹果，2 个梨，4 个香蕉。[B]、[C]、[甲]等等符号，都可以表示某特定的个体集。

3.离散的分布函数的符号表示

鉴于经常出现分布函数的自变量不是过去常用的连续变量，而可以是离散的特征标志（如名字是苹果、梨、香蕉的水果，蓝色的、黄色的……），所以我们推荐用字符多项式去表示分布函数。例如前面的例子里，就用下式表示这个个体集的分布函数：

[A]=（3 个）（苹果）+（2 个）（梨）+（4 个）（香蕉）

这里[A]代表了一个个体集，而等号后面表示该个体集的分布函数。显然[B]=15 个（儿童）+16 个（成年人）+9 个（老年人）表示了不同年龄段的人各有多少，[B]就是一个分布函数明确的个体集。

这里把园括号省略了。[C]=14(70 分以下)+25（70-90 分）+10（90 分以上）表示了一次考试的全班成绩的个体集的分布函数，它说明 70 分以下的学生有 14 个，70 到 90 分的学生 25 个，高于 90 分的学生有 10 个。根据上面的说明，我们一般用方括弧包起来的大写字母表示个体集本身；而用字符多项式表示该个体集的分布函数。

这里的等号"＝"体现了前面的认识：知道了它的分布函数，也就等于知道了一个确定的个体集。而加号"+"的意义是"还有"、"还包括有"的意思。如果再借用数学里的求和符号∑，个体集的分布函数就可以一般地写为下面格式：

$$[A]=\sum n_i x_i \qquad (2)$$

这里的各个 x_i 是彼此不同的标志值（也称为变量）；而各个 n_i 是对应于该标志值的个体的数量（这里用符号 n 代替了 a（它包含计量单位"个"）。我们也用"系数"称呼它。

对于分布函数是连续变量的情况（标志值是连续变化的），它们对应的个体数量也是连续变化的情况（在个体数量十分大的情况下，才会出现化离散为连续的数学处理技巧问题），我们依然可以用数学里惯用的连续函数去表示它。

2.　个体集的运算

如果给某概念一个比较严的定义，而随后提不出有什么好处（定量的计算、新规律的发现……），这样的定义在科学上也就没有吸引力了。已知甲小学的不同年级各有多少学生，还知道乙小学不同年级各有多少学生。求两个学校合并以后的不同年级各有多少学生。

如果用个体集[甲]、[乙]分别表示两个学校的学生个体集，两个小学合并以后的个体集用[丙]表示，那么个体集[丙]就是[甲]、[乙]的"和"。这里的"和"是一种数学运算，这可以写为：

[丙]＝[甲]+ [乙]

如何对两个个体集做加法运算？其实，运用代数里的多项式加法（合并同类项）正合适。在这里已经看到定义的字符多项式的好处。

《组成论》里介绍了个体集的某些运算规则。它把过去熟悉的一些逻辑和代数运算规则，如加、减、乘、除等，引用到个体集中。通过这些运算可以得到含义明确的新的个体集。个体集不仅成为可以运算的对象，而且通过运算扩展了我们的知识。

3.　个体集的某些特征量（参数）

1）个体集的个体总数 N：

个体集存在着一些重要的特征量或者说重要参数。由于我们突出了个体的完整性（量子化）、离散化（现在时髦的称呼是"数字化"），所有个体集里所包括的个体的数量应当是个正整数。这个正整数自然称为该个体集的个体总数。这个小学校有 400 学生，中国有 14 亿人、那个盘子里有 7 个水果都是例子。根据定义，个体集的个体总数显然是个体集的分布函数多项式表示下的各个系数的和，即

$$N=\sum n_i \qquad (3)$$

2）个体集的标志值的平均值 m：

如果某一个体集的各个标志值 x_i 是物理量（包括单位）而不是字符串，那么按照统计学的一般做法，通过下面公式得到的数值显然应当称为该个体集的标志值的平均值

$$m=\sum (n_i x_i)/N \qquad (4)$$

公式中的 n_i 和 x_i 都是数，而且这里是真的在 n_i 和 x_i 进行普通的代数运算（相乘）。运算得到的 m 的量纲应当与 $x_i n_i$ 相同。甲班的学生-年龄个体集为：

[甲]＝3 个（12 岁）+15 个（13 岁）+11 个（14 岁）+1 个（15 岁）。那么甲班学生的平均年龄 m 显

然是：

$m=(3\times12+15\times13+14\times11+1\times15)/(12+15+14+1)=13.3$。

即平均年龄是 13.3 岁。

3）个体集内不同标志值占的百分比（权重）f_i：

它等于具有标志值 x_i 的个体数量 n_i 与总体内的个体总量 N 的比值。即： $f_i=n_i/N$ （5）

百分比本来就是大家熟悉的统计量，现在用到这里了。显然，各个标志值的 f_i 的合计值应当等于 1。在概率论语言中被称为归一性。

$$1=\sum f_i=\sum n_i/N \quad (6)$$

4）上面谈到的一些特征量的关系理解

可用“个体集----一般符号表示（或特例）----总计”来帮助理解。

a、标志值（区间）----$x_1+x_2+\ldots+x_i+\ldots$----从下界到上界；

b、个体数量----$n_1+n_2+\ldots+n_i+\ldots$----N；

c、比例---- $n_1/N+n_2/N+\ldots+n_i/N+\ldots$----N/N；

d、百分比----$f_1+f_2+\ldots+f_i+\ldots$----1；

（下面是个体集----特例----总计）

e、年龄（岁）----12+13+14+15+…----12--15；

f、学生（个）----3+15+11+1+…----30；

g、百分比----3/30+15/30+11/30+1/30+…----30/30。

5）定义：

百分比矢量 f 是个体集必然具有的一个矢量，它的各个分量就是具有特定的标志值 x_i 的个体的数量与个体集内个体总量的比值 n_i/N，即 $f_i(i=1,2,\ldots)$。如本班有 50 个同学，女生占 30 人，男生 20 人，则这个个体集的百分比矢量 f 的两个分量（顺序是女，男）是（0.6，0.4）。任何一个个体集的百分比矢量的各个分量代数和等于 1。

6）百分比的代数平均值和几何平均值

假设个体集内的每个个体用一个卡片代表它，而在卡片上根据它原来的标志值 x_i，写上该标志值对应的百分比的值 f_i，那么也可以认为百分比本身就是标志值了。即这个个体的标志值 x_i 就改用 x_i 占的百分比 f_i 来代替了。由于百分比这个新的标志值是数值，我们自然可以求这个个体集内各个个体的百分比的平均值。

而根据平均值的定义，$m=\sum(n_ix_i)/N$，注意到现在 f_i 代替 x_i（相当于对自变量的函数求平均值），而 $f_i=n_i/N$，自然有百分比的平均值公式：$m_f=\sum(n_i)\hat{}2/N\hat{}2=\sum(f_i)\hat{}2$ （7）

于是我们知道一个“个体集”内的不同的标志值占的百分比的平方和，就是各个个体的新标志（百分比）的平均值。按照统计学的语言，这种平均值应当称为代数平均值或者算术平均值。对于直角坐标系下的平面上的一个矢量，我们知道其分量的平方和是个有意义的量（开平方以后是该矢量的长度）。在 p 维空间中的一个矢量，其各个分量的平方和也是有意义的。对于个体集来说，它的百分比矢量的平方和等于百分比的平均值。另外对百分比还可以求其几何平均值。

根据几何平均值的定义，个体集内的百分比的几何平均值（mf)'应当是： $(m_f)'={}^N\sqrt{[\prod(f_i)^{n_i}]}$ （8）

7）个体集的复杂程度 C，其定义是：

$$C=[N(\log N)]-[\sum(n_i)(\log n_i)] \quad (9)$$

以上两个平均值都是个体集的重要特征量，它们与熵有特殊关系（见后面说明）。

【4、复杂程度概念】

1. “复杂”

小学生就知道“复杂”二字的含义。从科学的角度看，把“复杂”提炼为科学名词，就需要把它定量化。确实，如果“复杂”变成一个“量”，并方便地用到十分广泛的领域，就提炼了一个重要的科学概念。在系统科学里要研究“复杂”的系统，最近 30 年所谓复杂性研究又成为热门。在这些领域都标榜自己研究涉及“复杂”的问题，如果对“复杂”的定量计量问题都没有很好的解决，那么这个研究必然存在基本的弱点。因此，给复杂下个妥当的定义，并且在各种场合可以具体计算出来复杂程度是多少，这就是一件基础性的工作。

在前面交代了“个体-概念-集合”和分布函数概念，再引入“复杂程度”概念就十分方便了。

2. 用个体集的分布函数计算一个数（C）

对于由 N 个个体组成的一个个体集，根据其中每个标志值占有的个体数量 n_i，（i=1，2，…，p，显然 $p\le N$））都可以用下面的公式计算出一个数值来。我们用大写的 C 表示这个数：

$$C=[N(\log N)]-[\sum(n_i)(\log n_i)] \quad (9)$$

这就是前面給的公式（9）。这里的各个 n_i 就是各个标志值 x_i 所对应的个体的数量。或者说是分布函数的函数值。这里的 log 是取对数的意思（对数的底应当是大于 1 的实数）。如果对数是以 2 为底的，其复杂程度的单位就称为比特（bite），它与信息论里的单位是相同的。《组成论》把数值 C 称为该个体集的复杂程度。只要每个个体集的分布函数是知道的，就可以计算出该个体集的复杂程度 C。个体集可以用到物理、化学、生物、地理、天文以至社会科学，所以这个公式可以计算很多领域的研究对象（个体集）的复杂程度。

3. 计算复杂程度的例子

个体集[A]是指一个白球和两个黑球，求这个个体集的复杂程度：

[A]=1(白球)+2(黑球)。

C_A=3log3-1log1-2log2=2.75 比特（计算时对数以 2 为底，下同）。

我们可以把它理解为个体集内标志值的差别程度、可区分的状态的丰富程度或者状态的混乱程度。如果 3 个球都是一个颜色，那么公式变成了：C=3log3-3log3＝0。

即其复杂程度为零，它说明这是一个清一色的系统，其复杂程度为零（再简单不过，所以复杂程度的最小值是 0，它没有负值）。而且无论这里有多少个个体，其复杂程度都是 0。一个水库里有很多水分子，你可以说那里物质很多，但是该系统的内部的状态却是清一色的水，这十分单纯，所以状态的丰富（复杂）程度小到了最低值 0。

由 0，1，2，…，9 这 10 个数组成的个体集的复杂程度＝10log210=33.21928bite（注意 log1=0）。

由 26 个英文字母组成的个体集的复杂程度＝20log26=26×log26×3.321928=122.2 bite。一副扑克牌有 54 张，每张都不同，根据复杂程度公式，其复杂程度＝54log54=310.76 bite(比特)。钱夹里有 3 张 100 元、4 张 50 元、5 张 10 元的人民币，班里有 23 个女同学和 21 个男同学，这些个体集的复杂程度，不难根据公式计算。下面是求算复杂程度的更多的例子（题目），它们体现了很多领域都存在对应的复杂程度的计算。利用这些计算自然可以展开对应的分析研究。

确实，过去不少人围绕平均值做了很多文章；现在我们看到一切可以计算平均值的数据都可以拿来再计算另外一个重要的数----复杂程度 C，利用复杂程度再写文章的很多的机会已经来了。

a、已经知道不同年龄的中国人各有多少，于是可以求中国的人口年龄的复杂程度。

b、一个国家有 30 万公里的公路，已经知道不同等级的公路各有多少，求公路等级的复杂程度。

c、已经知道 500 强企业的资产，求企业的资产的复杂程度。

d、已经知道 10 万平方公里面积上受到了不同烈度的地震灾害的面积各有多少，求该地区地震烈度的复杂程度。

e、已经知道本省 12 万平方公里面积上不同雨量分别占了多少面积，求雨量分布的复杂程度。

f、已经知道本岛屿上不同物种的动物各有多少，求该岛上动物的复杂程度。

g、知道班上 57 位同学的考试成绩，求考试成绩的复杂程度。

h、运动会的 1 等门票有 200 张，2 等的 1000 张，3 等的 1200 张，求运动会门票的复杂程度。

4. 复杂程度的重要性质

根据复杂程度公式，注意到 N，以及各个 n_i 都是正整数，不难知道复杂程度不会出现负值。而当每个标志值仅占有一个个体时（各个个体的标志值都不相同时），复杂程度达到它的最大值，C=NlogN。当 N＝1 时，即这个个体集里只有一个个体，其复杂程度＝0。

个体集[A]与个体集[B]如果它们所包括的个体是同类个体，而且标志相同（标志值可以不同），当它们合并为一个个体集以后，新的个体集的复杂程度可以大于原个体集的复杂程度的"和"。如[A]是 1 个红球，[B]是 1 个白球，把它们合成一个由两个球组成的新的个体集，则新的个体集的复杂程度＝2log2，由于原个体集的复杂程度都是 0，所以个体集的合并（加法）体现了 0+0>0 的特点。

它对哲学家、政治家、系统科学研究者热心讨论的 1+1 大于 2 问题给了一个严格的数学、物理论证。在信息论的语言里，面对一个个体集进行一次（不是 2 次，或者多次）抽样的结局的不确定性 H 被表示为：

$$H=-\sum(p_i)(\log p_i) \quad (10)$$

注意到 $p_i=n_i/N$，有 C=NH，即信息熵 H 的 N 倍就是复杂程度 C。于是沿着现在的思路引入的复杂程度概念与信息论里的信息熵概念是成正比例的量（如果进行 N 次抽样，其结局的信息熵恰好等于复杂程度 C）。由于信息论中已经发现了信息熵的很多数学性质，借着这个关系，信息熵的很多知识也自动归入复杂程度的性质中。

《组成论》还指出，在热力学领域，复杂程度（乘玻尔兹曼常数以后）就是物理学中的热力学熵。于是我们看到复杂程度概念吸收了物理学的热力学熵、信息论的信息熵作为自己的特例，它同时把自己方便的用到一切个体集上。所以复杂程度概念的明朗化也是"熵概念"在新理解方式下的合理扩展（从热力学扩展到其它领域，这可以是概率领域，也可以是另外的领域）。

5. 复杂程度的物理意义

应当说规定前面的计算量 C 为复杂程度的核心理由，是它的计算结果符合大家在社会生活中对复杂一词的理解，这就使得用途十分广的"复杂"概念有了比较科学的量化途径。而各种场合对"复杂"二字的理解，自然就成为理解复杂程度公式的多种思路。

前面已经看到系统内（个体集）的不同（可分辨）的标志值（状态）越多，则复杂程度越大（相同的标志值的数量越多复杂程度越小）；个体数量（N）很大，则复杂程度大。所以数量 C 体现了该系统内部状态的多样性，而复杂程度是个中性词。

热力学关注物质分子运动的混乱、差异、丰富程度，并且用热力学熵表示它，所以你可以借助热力学熵理解复杂程度，也可以反过来，借助复杂程度理解热力学熵（目前流行的把热力学熵说成为无序程度是不准确的，因为"序"不能简单的用"复杂"来度量）。

在信息论里人们考虑抽样结局的不确定性，它也是理解复杂程度概念的一个角度。我们把它用到不同场合时，会逐步加深对这个重要的词汇的理解和运用。

【5、引言复杂程度的定律】

我们从"个体"概念的清晰化，引申出关于个体的集合概念，即个体-标志值-集合（个体集），随后提出了对个体集的分布函数概念的提炼和个体集的特征量的讨论，并且引出了复杂程度概念。

复杂程度概念与热力学熵概念、信息熵概念的联系固然是我们重要的收获，但是新概念的引入如果没有与新规律的揭露联系起来，人们可能就怀疑这样做是否值得。后面要说明复杂程度概念的引入是为了引入"复杂度定律"。

1. 最复杂原理

1） 个体集的随机性（不确定性）

掷一个骰子，它出现几点（那点向上）就有随机性，

从一副麻将牌中随便取一张牌，它究竟是什么牌也有随机性（不确定性）。玩麻将牌者要拿 13 张牌，这 13 张牌就构成了一个个体集（群体）。显然在没有拿到牌之前，这付牌究竟是什么也有随机性。

100 位顾客买了那些东西，他们的花费各是多少？在事先也有随机性。看来，某些个体集及其分布函数究竟是什么的随机性问题是值得研究的。而这也就引出了不同的个体集（如抓 13 张麻将牌）有不同的出现概率问题。

2）红绿灯模型

某人上班路上要经过 10 个有红绿灯路口。针对每次上班可以问：你遇上了几次红绿灯。本问题的答案显然有随机性。从个体集语言的角度看，"经过 10 个有红绿灯的路口"等价一个个体集有 10 个个体，"红灯"、"绿灯"对应两种标志值。而回答了红灯（或者绿灯）的出现次数也就等价于知道了分布函数（不同颜色的灯各有多少）。

如果遇到的红灯次数(概率)计算出来了，也就知道了不同的个体集（分布函数）的出现概率（机会）。本问题是概率论中的二项分布问题。设红绿灯的出现概率都是 0.5（相等），经过 10 次路口遇到 m 次红灯的概率 p(m)为：

$$p(m)=\{10!/[m!(10-m)!]\}(0.5)^{10} \quad (11)$$

3）下面是表示不同红绿灯的出现概率的例子：

如不同的个体集、分布函数、复杂程度的出现概率不同，其中第二行给出了不同的红绿灯次数 m 对应的不同的出现概率 p(m)，第三行是利用复杂程度公式（9）计算出来的对应的复杂程度值。

a、红灯次数 m---- m=0；m=1；m=2；m=3；m=4；m=5；m=6；m=7；m=8；m=9；m=10。

b、该事件出现的概率×10^24---1；10；45；120；210；252；210；120；45；10；1。

c、复杂程度(比特)--- 0；4.6；7.2；8.8；9.7；10；9.7；8.8；7.2；4.6；0。

以上说明：出现概率最高的个体集（事件）也是复杂程度最大的个体集（事件）。在概率的对数与复杂程度的坐标图（略）上它们是直线关系。

2．最复杂原理

如果个体的标志值不仅只是两个（红绿灯）而是 k 个可能值 x_1，x_2，…，x_i，…，x_k，而它们的出现的概率 p_1，p_2，…，p_i，…，p_k（也称为先验概率）可以彼此不相同、个体的总个数相当多（可以利用 Stirling 公式：lnN!=NlnN-N），利用概率论中的多项式分布公式，可以得到下面的关系：

$$\ln p = C + [\sum(n_i)(\ln p_i)] \quad (12)$$

n_i 是概率为 p 的标志值占有的个体数量。这个公式体现了不同的个体集的出现概率 p（大写的！）与其复杂程度 C 的关系。

它表示该个体集的出现概率的对数等于该个体集的复杂程度再加上"另外一项"。显然当概率达到最大值时，复杂程度与"另外一项"的合计值也达到最大值。概率最大的事情自然是在一次实验中最容易出现，所以最容易出现的个体集是复杂程度与"另外一项"的合计值最大的个体集。"另外一项"在本例中包括了各个结局的出现概率，它对应着一种限制、约束条件。在另外的场合它的外型可能不同。

但是它都包括着一些约束条件。考虑到有的个体集的出现还会附有其他的约束条件，我们把"另外一项"更含糊化为"在限制条件下"。

于是上面的公式、红绿灯的例子、概率公理和以后的例子，我们把："有随机性的客观事物（个体集）都自动使自己内部状态的复杂程度在限制条件下达到最大值"称为最复杂原理。最复杂原理的正确性体现于多次实践中它经常是对的，而不是每次必然正确，它与"正方形的面积等于边长的平方"之类的确定性规律在品格上是不同的。

3．定性的例子

你在街上会遇到很多人，有买东西的、上学的、做生意的、出差的、看病的等等。把遇到的人看成是一个个体集，根据最复杂原理这个个体集内的人（个体）的活动目标（标志值）会自动最复杂化。

从商店出来的人都买了相同商品的事件的出现概率就非常小的，最容易出现的情况是买什么东西的人都有。所以最容易出现的也正是复杂程度最大的。你把一个玻璃杯摔碎了，碎玻璃的大小都相同？根据最复杂原理，它们是尽量地不相同，碎玻璃的大小尽量复杂化。

仔细想想，符合最复杂原理的事物几乎是司空见惯。

4．定量的例子----斩乱麻问题

最复杂原理不仅在生活的事例早已司空见惯，而且也可以推导出很多定量规律。利用分布函数可以计算该个体集的复杂程度。对此也可以反过来思考：对于有随机性的个体集，根据最复杂原理（复杂程度最大），能否反算出分布函数？这个问题非常有吸引力。

但是，在离散变量的情况要利用求函数的极值的技巧反求自变量（分布函数）。在连续变量的情况下就要利用变分技术，去反算未知函数（连续型的分布函数）。“拉哥朗日乘子方法”经常帮助我们达到目标。现以“斩乱麻问题”为例说明之。

把长为 L 的绳子随机地切割成充分多的 N 段，问不同长度的线段各有多少？这就是斩乱麻问题。把切碎的绳子看作是个体集，这就是利用随机性从理论推测其分布函数。切割的随机性造成了线段长短不齐，它体现了事物的复杂性。以 f(x)这个分布函数表示不同长度的线段占的百分比（权重），那么它对应的复杂程度 C 为：

$$C=-\int[Nf(x)]\,[\ln f(x)]\,(dx) \qquad (13)$$

切割当然不是烧掉，切割后的线头总长度应当等于 L，即

$$L=\int[Nf(x)]\,(dx) \qquad (14)$$

根据最复杂原理，C 应当在一定的约束条件下达到最大值。而上式就是一个约束，另外，百分比的合计值应当等于 1 也是一个约束条件：

$$1=\int f(x)\,(dx) \qquad (15)$$

现在的问题是：在式（14）、（15）的约束条件下使 C 最大的分布函数 f(x)应当是什么？根据求泛函数极值的拉哥朗日方法，构造一个新的函数 F:

$$F=-\int[f(x)\ln f(x)]dx+C1[\int xf(x)dx-(L/N)]+C2[\int f(x)dx-1] \qquad (16)$$

这里的 C_2，C_2 是待定常数。很显然，复杂程度 C 达到最大值，F 也达到了最大值。在 F 达到最大值（体现最复杂原理，也考虑了本问题的特殊约束条件）时它对未知函数 f(x)的变分应当等于 0。利用 F 对 f 的变分为 0，我们得到：

$$f(x)=\exp(-1+C_1Cx+C_2)。$$

利用（14）和（15）消去未知数 C_2，C_2 解得：$f(x)=(N/L)\exp(-Nx/L)$。

注意到 L/N 的含义是线头的平均长度，以 a 表示它(也是常数)，有：

$$f(x)=(1/a)\exp[-(x/a)] \qquad (17)$$

它是一个负指数函数。即一堆斩乱麻中，不同长度的线头占的百分比（分布函数），应当是负指数函数（相对密度分布函数）。它显示长度 x 短（小）的线头多而长线头很少。

根据相对密度分布函数的定义，线头长度在 x→x+Δx 范围的有 Nf(x)Δx 段。如 L=200m，切成 N=20000 段（平均值 a=1cm），可以计算出长度在 3→4 cm 范围的线头有 604 段，占总量的 3% 。

我们可以在电脑上做一个数值模拟实验以得到一个斩乱麻的样本。对比显示最复杂原理得到的理论分布与模拟实验结果是一致的。这说明最复杂原理与实际相符合。例子中虽然利用了复杂程度最大，但是没有计算复杂程度的最大值究竟是多少？原因是我们主要兴趣是利用复杂程度最大去推算理论的分布函数。

只要某个体集具有随机性，就可以利用最复杂原理配合不同的约束条件得到不同的理论分布函数。客观存在的有随机性的个体集很多，利用最复杂原理求分布函数的一般方法有广泛的应用和价值。

5．熵原理与最复杂原理

我们从不同的“个体--标志值--集合”的出现概率入手，利用了概率最大对应于复杂程度最大，从而得到了复杂程度最大的个体集也就是出现概率高的个体集。我们把它称为最复杂原理。由于已经说明热力学熵是复杂程度的特例，所以联系着热力学熵的热力学第二定律（熵原理，熵最大原理）也是最复杂原理的特例。

1）复杂度定律

复杂程度是物质自身天然具有的属性，它和质量、能量一样地真实。关于物质的质量、物质的能量已经有了质量守恒、能量守恒定律，与之对应也应当存在一个关于物质的复杂程度的变化规律，现在暂称为“复杂度定律”。限于篇幅这里仅指出某些观点：

a.、个体以及个体集概念有层次性，复杂程度也具有层次性。不同层次（形态）的复杂程度都客观存在。这类似物质的能量有化学能和核能。不同形态的能量可以互相转化，不同层次（形态）的复杂程度也可以互相转化。

b、有限的物质仅具有有限的质量和能量，它也仅具有有限的复杂程度。“物质可以无限分隔”的观点等价于有限物质的复杂程度为无穷大，这个观点使复杂程度概念失去物质性，所以它是错误的。

c、建议把目前流行的世界的三元观（物质+能量+信息）修改为物质属性的三元观（质量+能量+复杂程度）。

d、存在着质量变换机构（如木器厂）、能量变换机构（如水利发电厂）和信息变换机构（如电视机），但是使输出的质量、能量、信息大于输出的变换机构是不存在的。

e、爱因斯坦的质量能量公式可能要扩大为质量、能量、复杂程度之间的定量互相转化关系，而其总量具有保守性。写为公式，对于孤立系统，有

$$\Delta m+\Delta E/v^2+k\Delta C=0 \qquad (18)$$

这里的 v 是光速、C 是复杂程度，m、E 分别是质量和能量。而 k 是待定常数。

f、我们暂且把上面的认识连同最复杂原理合称为复杂度定律。

2）小结

根据个体集（客观事物）的分布函数，利用公式（9）可以计算该广义集合的内部状态的复杂程度。具有随机性的广义集合，它的复杂程度常常是在一定约束条件下的最大值。利用这个最复杂原理可以推导很多事物（广义集合）的理论分布函数。

复杂程度是客观物质具有的基本物理量，它与质量、能量一样的真实，在一定的意义下它与信息熵成正比例。关于客观事物的复杂程度的一般规律与最复杂原理一并称为“复杂度定律”。

【6、这些概念和定律的应用】

1. 概念的应用

我们把“个体”二字提炼为科学名词。为了研究同类的个体（如学生）以及其中每个个体都具有的标志值（如学生身高），我们在个体、标志值的概念基础上又综合了“个体--标志值--集合”概念（即个体集、个体集合、广义集合概念）。这些努力使数学界受到重视的集合概念有了更具体一些的物理内容，从而可以在数学以外的自然科学、社会科学中得到广泛应用。大家不难在前面的举例和自己研究的领域找到大量的“个体”、“个体集”的例子。

这样做不是简单的套概念，做“概念”游戏。其重要的回报就是沿着这个思路自然地引入了它们所对应的分布函数。于是我们无形中把零散的数据（这经常体现为大量的资料）归结为一个函数。

这自然为引入数学工具和数学处理做了概念和技术准备。科学家经常把平凡的资料归结为一个公式，我们自己的资料那么多，也借用这里的思路归结为函数、公式，难道不好吗？确实，把数据归结为分布函数（公式）就为后续的理论论证做了技术准备。它可能使你从资料的守财奴变成指点江山的理论家。应当承认在过去，不同科学领域的学者在驾驭数据的能力上是有比较大的区别的。“个体－标志值－集合”概念的把握和分布函数概念的把握，会使这种差别缩小。

复杂程度是个类似平均值的统计量，过去多少学者围绕平均值做过大量的工作，今天，还利用相同的资料数据，又可以统计出了该对象（实为个体集）的复杂程度。自然，我们围绕复杂程度这个量还可以做大量的文章。具体应用的例子在《组成论》书里找到更多。

2. 与统计概率的应用领域一样的宽

不理解什么是黑洞、波粒二象性、量子力学的学者很多。而这也说明“黑洞、波粒二象性、量子力学”之类的概念的应用领域有限。反之，统计概率知识几乎是横扫各个学科。在统计与概率的知识系统中，概率分布知识，不同的概率分布函数的知识占据重要地位。

目前统计学对经常用的那些分布函数为什么符合资料数据的实际，没有给出统一的理论说明。我们想说明统计与概率论中的经常用到的十多种分布函数是个体集合的分布函数的特例。

而这些不同的统计或者概率分布函数，都是最复杂原理或者说熵最大原理在不同约束条件下的体现。即利用最复杂原理配合不同的约束条件，从而得到不同的概率（统计）分布函数。这就为概率分布函数的理论成因，给了理论性的指点与说明。这不仅是统计学的进步，也是最复杂原理（熵原理）的重要应用。这说明最复杂原理的应用领域应当与统计、概率的应用领域一样的宽。

3. 物理学里的例子

物理学是熵概念和熵原理的发源地。在熵概念和热力学第二定律的基础上所形成的统计力学、物理化学知识，是物理学和化学知识的重要组成部分。而我们从日常生活中的“个体”概念入手引入了“个体集”概念直到分布函数、复杂程度，再到最复杂原理，这固然没有借用物理学的这些专门知识，但是我们也说明物理学里的熵概念与复杂程度概念是一致的，熵原理是高概率事件容易出现这个浅显道理的体现。所以熵概念在物理学中的应用和熵原理在物理学中的应用都是个体集、分布函数、复杂程度概念和最复杂原理的应用的重要体现。

希望物理学理解我们的思路是问题的一个侧面；看到物理学的成就辉煌，借着这里介绍的思路把大家试图利用熵概念和熵原理的热情，变成在物理学以外的学科领域的平行应用是问题的另外一个侧面。我们相信用这里的思路参照物理学的成就，提高其他学科的水平是大有作为的。多年来，我们用复杂程度代替熵，用最复杂原理代替熵原理在思考、推理问题上一直是顺畅的。

4. 信息论

信息论是十分重要的一门科学。信息熵概念是信息论中的基础概念。而我们已经说明信息熵与复杂程度概念是互相成正比例的量。

所以信息熵概念、信息概念的理解、信息量的很多公式，都是可以沿着个体集合--复杂程度的思路推演出来。

话也可以反过来说：信息论里的概念、公式等等，都已经是关于个体集合和它的复杂程度的知识和应用。

5. 熵气象学

作者的专业是气象，正是在多年对气象难题的探索中，使我领会和形成了这里介绍的思路。而这个思路在气象学中的应用，使我们在 20 世纪 90 年代初都归入“熵气象学”的名下。

【7、结束语】

《熵气象学》一书，就是这个努力的阶段性成果。它让我们看到分布函数概念，不仅把气象概率分布都包括了进去，而且还引出了很多又很重要的大气分布函数。这些分布函数的理论，说明常可以用最大熵原理（最复杂原理）给予说明。

上面简单地点到了个体集合和它的复杂程度的应用，其实这些也仅是开始。它们在更广泛领域的应用正等待着我们大家继续努力挖掘。

参考文献

[1]张学文、周少祥，空中水文学初探，气象出版社 2010 年 3 月；
[2]张学文，马力.斩乱麻问题，数理统计与应用概率，1997 年，12 月，12（4）：315-321；
[3]张学文，马力.熵气象学，北京：气象出版社，1992；
[4]张学文，字符多项式与表格数学，计算机工程与应用，2002 年，38 卷增刊的 124-126，128 页；
[5]张学文，组成论，中国科技大学出版社，2003 年 12 月；
[6]叶眺新，蝴蝶效应莫比乌斯圈太极图病毒全息----读《病毒博物馆》说多极与全球化，Academ Arena，August 25, 2022；
[7]王德奎，三旋理论初探，四川科学技术出版社，2002 年 5 月；
[8]孔少峰、王德奎，求衡论----庞加莱猜想应用，四川科技出版社，2007 年 9 月；
[9]钱金，重元素的量子色动化学----自然科学与社会科学全息交叉探索（1），Academ Arena，June 25, 2022。

11/8/2022