

# 选择模型的理论扩展分析<sup>\*</sup>

尉伟峰<sup>†</sup>

(北京交通大学电子信息工程学院, 北京 100044)  
(2009 年 3 月 17 日收到, 2009 年 3 月 29 日收到修改稿)

选择模型为解释指数律、幂律的成因提供了基于偏好的概率分析方法. 通过对选择试验的研究, 在视野规律和选项分布之间导出了一对变换, 称其为偏序变换. 对偏序变换连续化, 导出了对应的连续变换. 仿照离散分布, 定义了连续分布的比较函数, 并沟通了离散和连续分布的伴随关系. 通过选择模型, 对幂律访问现象进行了仿真. 选择理论对于复杂巨系统和社会行为的研究具有较重要的借鉴作用, 可扩展统计物理学和概率论的理论基础.

关键词: 选择模型, 幂律现象, 偏序变换, 比较函数

PACC: 0250, 0500

## 1. 引 言

幂律现象在现实世界中广泛存在, 在人类社会、物理系统和生物领域中十分普遍, 对应的幂律分布曲线具有长尾现象, 显示出事务的多样性共存<sup>[1-17]</sup>. 随着对网络和复杂巨系统的深入研究, 发现了很多著名的幂律现象, 如节点度分布和 Web 访问律等, 引起了整个学界的关注. 而选择机理解释幂律和指数律现象<sup>[18, 19]</sup>时, 不再依赖随机过程和动力系统, 而主要考虑偏好特性、随机视野和确定优选等基本因素, 从而得到概率论解释. 其中偏好序列代表一元性, 随机视野代表多元性, 因此人们在满足物理条件和相应需求下, 选择公认的最好选项作为选定结果. 在此基础上, 通过改变视野中选项数的分布规律(视野规律), 可将模型推广到一般情形, 为一类分布找到基于加法原理的数理依据; 因此研究相应的偏序试验和概率理论具有很强的学术意义.

## 2. 理论的扩展

选择理论可以更加深入的扩展, 基于统一选择试验, 可以导出一类概率变换, 并且通过比较函数研究离散和连续分布间的伴随现象.

### 2.1. 统一选择试验

通过对两个基本选择模型和相应扩展研

究<sup>[18, 19]</sup>, 可将其融合为统一的选择试验, 步骤如下.

1) 列出选项集合, 以偏好序列中的次序标识选项, 按偏好从高到底排序, 依次为 1 到  $N$  选项.

2) 对于任意一次选择, 按视野规律  $q_l$  随机生成选项个数  $l$ , 其可取值在 1 到  $N$  之间.

3) 随机或按物理量大小逐次选择选项, 选项不重复, 以概率  $p_s$  确定是否能够选入视野, 依次进行, 直到视野中的选项个数为  $l$  时停止; 由于选项常是有限的, 当  $l$  较大时, 可能出现选到最后一个选项都未选满  $l$  个, 此时也算此视野有效.

4) 从视野中选定  $n$  个偏好最高的选项, 如果视野个数  $l$  小于  $n$  个, 则将所有选项选出; 如果选项本身具有排它特性, 非此即彼, 则  $n = 1$ ;  $n$  也可以是随机的, 步骤 3) 和 4) 等价于按概率  $p_s$  逐步挑选选项, 同时进行比较, 保留  $n$  个最高偏好选项, 直到比较的选项总数达  $l$  个或选项全部挑完停止.

5) 重复试验, 统计每个选项被选中的次数, 并归一化为概率分布, 称其为选项分布.

通过统一选择试验, 当改变视野规律  $q_l$ , 选入视野的概率  $p_s$  以及选定选项个数  $n$  等参数, 就可以完成各类概率分布的生成机理. 假设步骤 3) 中每个选项被选择的先后次序等概, 通常地, 若视野规律是较平的或者单调的, 其试验结果往往呈现幂律特征; 若其在  $l$  较大处有单峰, 常呈现指数律特征, 如图 1 所示. 此试验中, 当没有视野规律, 直接对每个选项

<sup>\*</sup> 国家自然科学基金(批准号: 60772043, 60672069), 国家重点基础研究发展计划(批准号: 2007CB307101)资助的课题.

<sup>†</sup> E-mail: weiwefeng@sjtu.edu.cn

判断是否选入视野,则此模型退化为模型 1,此时  $p_s$  和模型 1 中  $p$  的意义一致,而若  $p_s = 1$  时,选项不经过筛选选入视野,等价于门限前的选项全部选取,则

退化为模型 2,需要指出的是此试验中概率  $p_s$  仅用来筛选选项,而模型 2 中的  $p$  可控制门限规律或视野规律<sup>[18,19]</sup>.

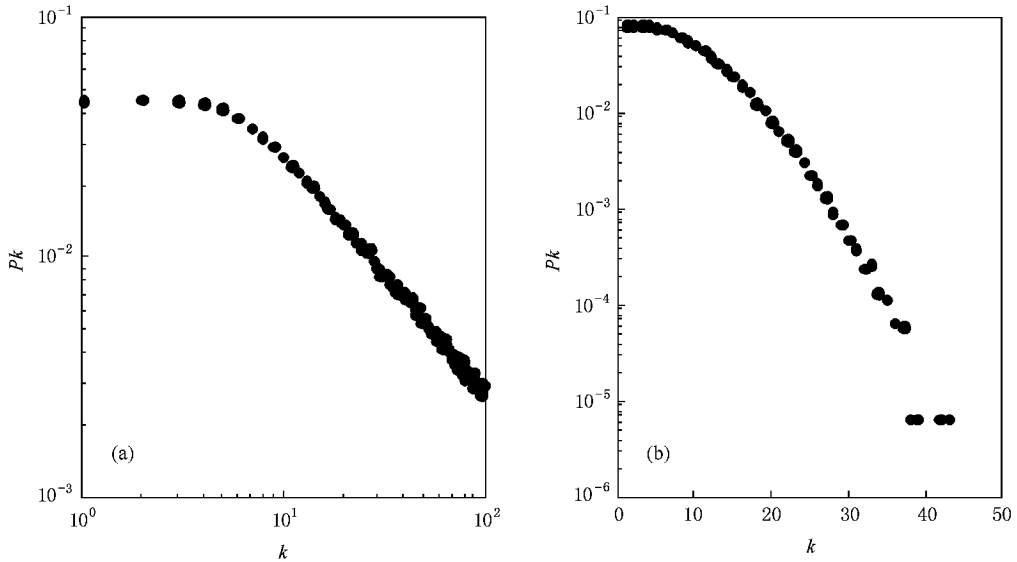


图 1 统一选择试验的仿真图(选项个数  $N = 100$ , 筛选概率  $p_s = 0.5$ , 选定个数  $n = 3$ , 试验次数 50000) (a)对数图,反比视野规律  $q_l \sim l^{-1} (1 \leq l \leq 50)$ ,  $l$  大于 50 概率为零,当  $k$  较大时分布为直线,幂指数为 1 (b)纵轴对数化图,固定视野规律  $q_l \sim \alpha (l - 25)$ ,当  $k$  较大时分布为直线

当选择选项的次序等概, $N$  较大, $p_s$  较大且与  $q_l$  无关时,概率  $p_s$  影响试验结果较小(其意义将在后述中说明),统计选项分布时可不计,各选项在视野中情况类同,易证明分布是单调的( $n > 1$  时头部  $n$  点是平的).在研究幂律现象时,由实验与推导<sup>[18]</sup>可知, $n$  较小时分布幂指数常与  $n = 1$  时相同,甚至  $n$  随机取的情况下也如此;且通常理论分析取  $n = 1$  即可,其他情况可做类似处理,得到分布如下:

$$P_k \sim \sum_{l=1}^{N-k+1} \frac{C_{N-k}^{l-1}}{C_N^l} \cdot q_l \quad (k = 1, 2, \dots, N). \quad (1)$$

### 2.2. 偏序概率变换

在复杂巨系统内,由于关系到人们自身在视野中的选项个数倾向,有些视野规律并不容易测得,只能依概率按经济学上类似效益函数作假设.而选项分布统计规律往往是较易获得的,通过(1)式,容易推断选择模型下的视野规律.给出  $P_k$ ,解(1)式构成的方程组可得(推导过程见附录 A)

$$q_l \sim C_N^l \cdot \sum_{k=1}^l (-1)^{-k} C_{l-1}^{k-1} P_{N-k+1} \quad (l = 1, 2, \dots, N), \quad (2)$$

可见(1)式和(2)式形成一对变换,它沟通了视野规律和选项分布的对应关系,可以相互导出,称其为偏序概率变换.偏序变换中,若要获得  $P_k \sim \frac{1}{k-2}$  的视野规律,只需将此分布代入(2)式,即得  $q_l \sim \frac{1}{l \cdot (N-l) \cdot (N-1-l)}$ ,类似地,变换可推导出众多恒等式<sup>[19]</sup>,并提供生成机理和物理意义.而且此变换可导出更多的类似变换,比如模型 1 也可对应变换,在(1)式和(2)式中用  $C_{N-k}^l$  代入  $q_l$ ,即得如下变换,称为几何概率变换;其和 Boustrophedon 变换<sup>[20]</sup>类似

$$P_k \sim \sum_{l=1}^{N-k+1} C_{N-k}^{l-1} \cdot q_l$$

$$q_l \sim \sum_{k=1}^l (-1)^{-k} C_{l-1}^{k-1} P_{N-k+1}. \quad (3)$$

### 2.3. 变换连续逼近

(1)式中视野规律是离散的,可连续化近似如下(推导过程见附录 B):

$$P_k \sim \int_0^1 y e^{-ky} \cdot q_y dy \quad (k = 1, 2, \dots, \infty). \quad (4)$$

改变连续视野规律  $q_y$ , 即可得到相应的  $P_k$ ; 其意义是将模型中的选项用  $(0, 1)$  的数据来表示, 视野规律可在  $(0, 1)$  上定义并满足逼近条件, 按小数据高偏好进行选择试验, 并以数据大小次序排出  $k$ , 对应相应的选择分布, 当数据越来越密集, 视野规律趋向为连续的, 同时选项数趋向无穷. 特别地, 若  $q_y \sim y^r (r \geq -1)$  逼近过程中满足条件  $q_l \sim q_{l/N}$  (见附录 B), 当  $k$  较大时, 可得近似幂律分布, 此积分运算较组合运算容易地多, 如  $q_y \sim 1$  由分部积分法易得  $P_k \sim \frac{1 - (k+1)e^{-k}}{k^2}$ . 若视野规律为连续冲击函数, 即  $q_y \sim \delta(y - \lambda) (0 < \lambda < 1)$  则  $P_k \sim e^{-\lambda k}$  为几何分布.

如(4)式, 将(1)式中  $k, l$  分别推广为实数  $x, y$ , 则有概率密度函数如下:

$$p(x) \sim \int_0^\infty y e^{-xy} \cdot q(y) dy \quad (0 < x, y < \infty), \quad (5)$$

式中可以将  $yq(y)$  看作一个整体, 则  $p(x)$  即为对  $yq(y)$  拉普拉斯变换的象函数, 从而可得反变换为

$$q(y) \sim \frac{1}{yi} \cdot \int_L e^{xy} \cdot p(x) dx, \quad (6)$$

其中  $L$  为复变域内的无界纵向直线<sup>[21]</sup>; 称(5)式和(6)式为连续偏序变换. 当将  $q(y) \sim y^r$  代入(5)式作相应拉普拉斯变换, 可得  $p(x) \sim x^{-(r+2)}$ ; 若对应的视野规律为  $q_l \sim (l+r) \cdot (l+r-1) \cdot \dots \cdot (l+1)$ , 并令  $N \rightarrow \infty$ , 代入(1)式可得<sup>[19]</sup>,  $P_k \sim B(r+2, k)$ ; 由此可见, 前者为 Zipf 分布, 后者为 Yule 分布, 两分布之间存在对应关系. 通过(5)式, 改变  $q(y)$ , 对照拉普拉斯变换表, 可导出很多密度函数; 简单地, 若  $q(y) \sim a^y = e^{y \ln a}$ , 则可求得  $p(x) \sim \frac{1}{(x - \ln a)^2} (x > \ln a)$ , 可见  $q(y)$  当为指数规律时, 也可以导出近似 Zipf 分布, 但此时门限规律已在实数域相对独立, 不再对应离散门限规律. 这样模型就扩展到连续形态, 并提供相应物理意义和构造方法.

对于几何分布和指数分布的对应关系, 同样可以通过偏序变换来分析. 由文献<sup>[19]</sup>可知, 当视野规律为单点分布  $q_l = \delta(l - m)$  时, 代入(1)式, 得到固限分布  $P_k = \frac{C_{N-k}^{m-1}}{C_N^m} (k = 1, 2, \dots, N - m + 1)$ , 当  $N$  充分大时, 其近似为几何分布  $P_k \sim e^{-\frac{m-1}{N}k}$  (推导过程见附录 C); 对应地, 将连续冲击函数  $q(y) = \delta(y - \lambda)$  代入(5)式, 可得概率密度满足指数分布  $p(x) =$

$\lambda e^{-\lambda x} (x > 0)$ ; 可见几何分布和指数分布之间也存在基本对应关系; 虽然这是较明显的, 但在理论证据上, 还需要更多分析, 以期待应用于其他分布. 一般地,  $q_l$  可表示为  $q_l \sim \sum_{j=1}^N q_j \cdot \delta(j - l)$ , 通过(1)式和上述分析可知, 选项分布可以由固限分布叠加导出, 而固限分布又可近似为几何分布, 故选项分布可以近似由几何分布的叠加构成, 这和排队论和随机过程中常见连续分布通过指数分布构成类似. 此外, 离散冲击函数代入(1)式与连续冲击函数代入(4)式所得结果类同, 说明单点分布和连续冲击具有对应关系. 由此可见, 离散分布和连续分布间存在很紧密的对应关系.

对(3)式的连续近似较为困难, 所以用一个特例说明另一种连续化近似. 由(1)式知,  $P_k \sim \sum_{i=1}^{N-k+1} \frac{C_{N-k}^{i-1}}{C_N^i} \cdot 1 \sim \frac{N^2}{k(k+1)}$ , 可得  $P_{k/N} \sim \left(\frac{N}{k}\right)^2$ , 当  $N$  充分大, 得到  $p(x) \sim x^{-2} (0 < x < 1)$ ; 由(3)式  $P_k \sim \sum_{i=1}^{N-k+1} C_{N-k}^{i-1} \cdot 1 \sim 2^{-k}$ , 定义  $P_{k/N} \sim \sqrt[N]{P_k}$ , 则  $P_{k/N} = 2^{-k/N}$ , 当  $N$  充分大, 得到  $p(x) \sim 2^{-x} = e^{-(\ln 2) \cdot x} (0 < x < 1)$ ; 两特定模型做简单运算, 在区间  $(0, 1)$  上连续化, 可见也是具有上述对应关系. 值得指出的是 Zipf 特性不会随着空间的线性压缩或膨胀而变形. 另外, 泊松和高斯分布间的对应可通过对二项分布的极限近似导出<sup>[21]</sup>; 所以这种对应关系较为普遍, 分布之间存在紧密的构造联系.

### 2.4. 分布比较函数

对于离散分布, 概率比较函数(也称为概率比函数)定义为<sup>[18]</sup>

$$\psi(k) \triangleq \frac{P_k}{P_{k+1}}. \quad (7)$$

考虑到正态分布是连续核心分布, 将其离散化后得到  $P_k \sim e^{-\frac{k^2}{2\sigma^2}}$ , 比较函数  $\psi(k) = e^{\frac{2k+1}{2\sigma^2}}$ , 此为指数形态. 设想连续分布在离散化后的比较函数可约表示为  $\psi(k) = e^{\psi(k)}$ ; 于是可知  $e^{\psi(k)[(k+1)-k]} = \frac{P_k}{P_{k+1}}$ , 对照地有  $e^{\psi(x)[(x+\Delta x)-x]} = \frac{p(x)}{p(x+\Delta x)}$ , 其中  $p(x)$  为概率密度函数, 当  $\Delta x \rightarrow 0$  时, 定义连续分布比较函数为

$$\varphi(x) \triangleq -\frac{d \ln p(x)}{dx}. \quad (8)$$

基于(7)式和(8)式, 若离散分布和连续分布的

比较函数类型一致,称互为伴随分布;常用的离散分布在比较函数上对应着常用的连续分布(见附录 D). Zipf 分布或 Pareto 分布可指连续分布,而幂律分布常指  $P_k \sim k^{-r}$  的离散分布;由(8)式,Zipf 分布的比较函数是双曲线型的,而由(7)式,幂律分布的比较函数却不是,而比较函数是双曲线并且可近似为幂律分布的最典型的是 Yule 分布,故 Yule 分布伴随 Zipf 分布.所以在复杂巨系统中,特别网络测量出的幂律曲线不像物理界中那样(如引力场、电场强度等连续的平方反比律),其头部往往偏低,而几何和指数分布间却不会出现类似偏差,在伴随分布意义上,这些都是正常的.给出比较函数,即可得相应的分布(推导过程见附录 E),其沟通了分布之间的关联.

## 2.5. 选择模型应用

在类似博客站点(blogs)入站链接的幂律现象上<sup>[22]</sup>地域特征可以通过选择试验来刻画.各用户节点(sites)了解各站点的次序是不同的,假设以网速为参照物理量为例,网速大的易先考虑.假设用户数和网速间的关系满足近似反比关系(指数为 1 的

幂律);若用节点到站点的距离来表示网速,若距离减小,网速反比增大,则用户数和距离成正比关系;这相当于在平面上以站点为中心的区域随机撒点,随着距离增大,周长增大,相应用户数呈正比递增.当然,此中距离也可理解为其他物理量.

一般地,用户链接博客站点具有保持性,考虑每个节点只选定一个站点的情形.当按距离依次考查站点,并且按一定的选择概率  $p_s$  来判断此站点能否选入视野,通过公认的偏好序列(可理解为站点服务排名),优选出站点.若视野规律满足反比分布( $q_i \sim i^{-1}$ ),节点到站点间距离远近等可能(只高维空间可实现)则入站链接分布就会近似呈现出幂律现象( $P_k \sim k^{-1}$ )<sup>[19]</sup>;为了方便作图,仅在平面上仿真示意.如图 2 所示,当  $p_s$  较小时,用户社团(同一站点节点集)相对均散;而当  $p_s$  增大时,用户社团的集中程度变高,形成鲜明的势力范围.可见当站点很多时,概率  $p_s$  虽影响分布幂指数较小,但决定社团的聚合程度.

无标度网络中,若每个节点对各自所取连接

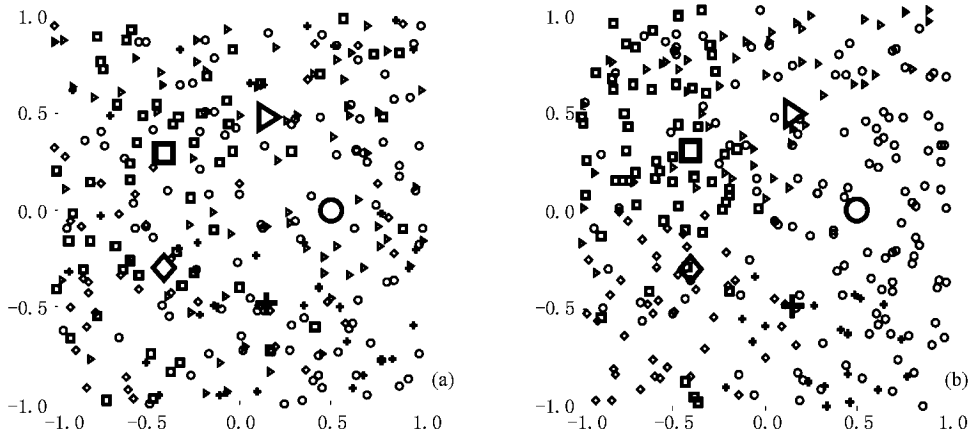


图 2 博客站点的用户社团仿真图(平面上随机撒点,点数为 300;大图形中心对应站点,分布在直径为 1 的圆上,总数为 5,圆形站点访问偏好最大,按逆时针逐个降低;用户节点选定站点,即用相应站点图形标识)(a)  $p_s = 0.5$  时仿真图 (b)  $p_s = 1.0$  时仿真图

数进行选择,那么各选项分别为 1 度点、2 度点等,此时选项具有排它特性,往往大多数节点没有大的资源供别人连接,或者总希望能够“搭便车”(free rider),只满足需求而不做贡献,于是选低度点作为高偏好选项.那么通过选择模型,当视野规律较为单调,可知度数和节点数分布(节点度分布)为幂律分布,假设视野规律是均匀的,那么有  $P_d \sim d^{-2}$ ,其中

$d$  为度数,  $P_d$  为  $d$  度节点个数.按此规律连接各节点,即可得到近似幂律图;在选择与其他节点连接时,同样可以运用选择模型来处理,比如按地理位置来考虑节点选入连接视野的次序等.若各点非常相似,视野规律相差无几,通过拥有单峰性的单点分布和二项分布(密集时可对应高斯分布),那么度分布将满足近似几何分布,其为指数律,在各节点类似的

P2P 网络中容易出现.

在文件分块的 P2P 传播系统中,类似文件传播模型<sup>[18]</sup>某节点拥有文件,并扩散其分块,整个文件将由不同分块树到达各节点.若偏好弱,任节点随机地在有相应分块的节点处下载,那么分块由随机树到达各节点,节点间形成均匀交互特性.若在报复机理(Tit-for-Tat)或网速择优情况下,存在节点间的偏好,则分块由偏序树到达各节点,运用选择理论,不同的视野规律和选择次序可导出不同的节点间交互特性.

### 3. 结 论

基于偏好的选择试验是一种值得深入研究的分布生成机理.若选项进入视野的次序是非等概的,其也可以生成非单调的分布.一般情况下,可以得到偏序概率变换和对应的连续化变换,离散和连续分布间能够通过比较函数关联.因此选择理论具有重要的数学内涵,需进一步分析.在网络研究中,选择模型也可引导建模和仿真,具有较强的实用性和拓展能力.

选择模型的完善将为统计物理学和概率论提供更丰富的分析方法,有助于揭示自然和人为现象背后的一般性成因,能对群体性行为的研究提供较重要的借鉴和启发.

衷心感谢陈常嘉教授、郭海燕硕士对本文工作提出的建议.

### 附录 A (2) 式的推导过程

由(1)式,设定归一化系数  $b$ ,可得  $P_k = b \cdot \sum_{l=1}^{N-k+1} \frac{C_{N-k}^{l-1}}{C_N^l} \cdot q_l$ ,展开即得方程组如下:

$$P_N = b \cdot \sum_{l=1}^1 \frac{C_0^{l-1}}{C_N^l} \cdot q_l = b \cdot C_0^0 \cdot \frac{q_1}{C_N^1},$$

$$P_{N-1} = b \cdot \sum_{l=1}^2 \frac{C_1^{l-1}}{C_N^l} \cdot q_l = b \cdot \left( C_1^0 \cdot \frac{q_1}{C_N^1} + C_1^1 \cdot \frac{q_2}{C_N^2} \right),$$

...

$$P_1 = b \cdot \sum_{l=1}^N \frac{C_{N-1}^{l-1}}{C_N^l} \cdot q_l = b \cdot \left( C_{N-1}^0 \cdot \frac{q_1}{C_N^1} + C_{N-1}^1 \cdot \frac{q_2}{C_N^2} + \dots + C_{N-1}^{N-1} \cdot \frac{q_N}{C_N^N} \right).$$

令  $P = (P_N, P_{N-1}, \dots, P_1)^T, q = \left( \frac{q_1}{C_N^1}, \frac{q_2}{C_N^2}, \dots, \frac{q_N}{C_N^N} \right)^T$ ,则方程组可表示为  $M \cdot q = P$ ,其中  $M$  为下三角系数矩阵,  $|M| = C_0^0 C_1^1 \dots C_{N-1}^{N-1} = 1$ ;由克莱姆法则、代数余子式和下三角矩阵行列式性质易得  $q_l = \frac{C_N^l}{b} \cdot \sum_{k=1}^l (-1)^{-k} C_{l-1}^{k-1} P_{N-k+1}$ ,归一化后略去系数,即得(2)式.

### 附录 B (4) 式的推导过程(此处 $N \rightarrow \infty$ 可理解为 $N$ 充分大)

由(1)式,可以写成组合展开形式,并多乘一个因子  $1/N$ ,可得如下:

$$P_k \sim \frac{P_k}{N} \sim \sum_{l=1}^{N-k+1} l \cdot \frac{(N-k) \cdot (N-k-1) \cdot \dots \cdot (N-k-l+2)}{N \cdot (N-1) \cdot \dots \cdot (N-l+1)} \cdot q_l \cdot \frac{1}{N} \quad (k, l = 1, 2, \dots, N),$$

其中

$$\frac{(N-k) \cdot (N-k-1) \cdot \dots \cdot (N-k-l+2)}{N \cdot (N-1) \cdot \dots \cdot (N-l+1)} = \left(1 - \frac{k}{N}\right) \cdot \left(1 - \frac{k-1}{N}\right) \cdot \dots \cdot \left(1 - \frac{k-l+2}{N-l+2}\right) \cdot \left(\frac{1}{N-l+1}\right),$$

于是可知

$$\lim_{N \rightarrow \infty} \frac{(N-k) \cdot (N-k-1) \cdot \dots \cdot (N-k-l+2)}{N \cdot (N-1) \cdot \dots \cdot (N-l+1)} \approx \lim_{N \rightarrow \infty} \left( \left(1 - \frac{k}{N}\right)^N \right)^{\frac{l-1}{N}} \cdot \frac{1}{N} = \lim_{N \rightarrow \infty} (e^{-k})^{\frac{l-1}{N}} \cdot \frac{1}{N},$$

于是有

$$P_k \sim \lim_{N \rightarrow \infty} \sum_{l=1}^{N-k+1} l \cdot \frac{(N-k) \cdot (N-k-1) \cdot \dots \cdot (N-k-l+2)}{N \cdot (N-1) \cdot \dots \cdot (N-l+1)} \cdot q_l = \lim_{N \rightarrow \infty} \sum_{l=1}^N \frac{l}{N} \cdot (e^{-k})^{l-1} \cdot q_l \cdot \frac{1}{N},$$

若满足条件  $q_l \sim q_{UN}$ (比如  $q_l \sim \frac{1}{l} \sim \frac{1}{(l/N)} \sim q_{UN}$ ),但若  $q_l \sim 2^l$ ,则  $q_{UN} \sim 2^{UN} = \sqrt{2^U} \sim \sqrt{q_l}$ ,不满足  $q_l \sim q_{UN}$ ,令  $y = \frac{l}{N} \in (0, 1]$ ,当  $N \rightarrow \infty, y \in (0, 1)$  连续,即得(4)式如下:

$$P_k \sim \int_0^1 y e^{-ky} \cdot q_y dy \quad (k = 1, 2, \dots, \infty).$$

若令  $k$  连续化为  $x, x \in (0, \infty)$ ,即得

$$\mu(x) \sim \int_0^1 y e^{-xy} \cdot q_y dy \quad (0 < x < \infty, 0 < y < 1).$$

### 附录 C 固限分布逼近几何分布

固限分布可表示为(式中  $\delta(l-m)$  为仅  $m$  处有值的离散冲击函数)

$$P_k = \sum_{l=1}^{N-k+1} \frac{C_{N-k}^{l-1}}{C_N^l} \cdot \delta(l-m) \quad (k = 1, 2, \dots, N-m+1).$$

由附录 B 的推导, 当  $N$  较大时, 可得近似如下:

$$\begin{aligned} P_k &\approx \sum_{l=1}^{N-k+1} \frac{l}{N} \cdot e^{-\frac{(l-1)k}{N}} \cdot \delta(l-m) \\ &= \frac{m}{N} e^{-\frac{m-1}{N}k} \sim e^{-\frac{m-1}{N}k} \\ &\quad (k = 1, 2, \dots, N-m+1). \end{aligned}$$

## 附录 D 伴随分布 (仅列出最基本的三类)

线型: 泊松分布  $P_k = \frac{\lambda^k e^{-\lambda}}{k!}$ ,  $\psi(k) = \frac{k+1}{\lambda}$ ; 正态分布

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}, \varphi(x) = \frac{x}{\sigma^2}.$$

常数型: 几何分布  $P_k = p(1-p)^{k-1}$ ,  $\psi(k) = \frac{1}{1-p}$ ; 指数

分布  $p(x) = \lambda e^{-\lambda x}$ ,  $\varphi(x) = \lambda$ .

双曲线型: Yule 分布  $P_k = rB(r+1, k)$ ,  $\psi(k) = 1 +$

$$\frac{r+1}{k} \quad \text{Zipf 分布 } p(x) = \frac{x^{-(\rho+1)}}{\zeta(\rho+1)}, \varphi(x) = \frac{\rho+1}{x}.$$

## 附录 E 由比较函数导出分布

离散时, 由(7)式可得

$$\begin{aligned} P_k &= \frac{P_{k-1}}{\psi(k-1)} = \frac{P_{k-2}}{\psi(k-1) \cdot \psi(k-2)} = \dots \\ &= P_1 \cdot \prod_{l=1}^{k-1} \frac{1}{\psi(l)} \sim \prod_{l=1}^{k-1} \frac{1}{\psi(l)} = e^{-\sum_{l=1}^{k-1} \ln \psi(l)}, \end{aligned}$$

连续时, 由(8)式可得  $p(x) = C \cdot e^{-\int_0^x \varphi(l) dl} \sim e^{-\int_0^x \varphi(l) dl}$ .

- [ 1 ] Zipf G K 1949 *Human Behavior and the Principle of Least Effort* (Cambridge: Addison-Wesley) p1
- [ 2 ] Albert R, Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
- [ 3 ] Vázquez A, Satorras R P, Vespignani A 2002 *Phys. Rev. E* **65** 66130
- [ 4 ] Yan D, Qi G N 2006 *Acta Phys. Sin.* **55** 3799 (in Chinese) [ 闫栋、祁国宁 2006 物理学报 **55** 3799 ]
- [ 5 ] Zhao W, He H S, Lin Z Y, Yang K Q 2006 *Acta Phys. Sin.* **55** 3906 (in Chinese) [ 赵伟、何红生、林中材、杨孔庆 2006 物理学报 **55** 3906 ]
- [ 6 ] Guo J L, Wang L N 2007 *Acta Phys. Sin.* **56** 5635 (in Chinese) [ 郭进利、汪丽娜 2007 物理学报 **56** 5635 ]
- [ 7 ] Ouyang M, Fei Q, Yu M H 2008 *Acta Phys. Sin.* **57** 6763 (in Chinese) [ 欧阳敏、费奇、余明晖 2008 物理学报 **57** 6763 ]
- [ 8 ] Huang Z X, Wang X R, Zhu H 2004 *Chin. Phys.* **13** 273
- [ 9 ] Li Y, Liu Y, Shan X M, Ren Y J 2005 *Chin. Phys.* **14** 2153
- [ 10 ] Pei W D, Chen Z Q, Yuan Z Z 2008 *Chin. Phys.* **B 17** 373
- [ 11 ] Chen Y Z, Fu C H, Chang H, Li N, He D R 2008 *Chin. Phys.* **B 17** 3580

- [ 12 ] Albert R, Barabási A L 1999 *Sci.* **286** 509
- [ 13 ] Barabási A L, Albert A 1999 *Phys. A* **272** 173
- [ 14 ] Bak P 2001 *How Nature Works* (Wuhan: Central China Normal University Press) p10 (in Chinese) [ 帕巴克 2001 大自然如何工作 (武汉: 华中师范大学出版社) 第 10 页 ]
- [ 15 ] Carlson J M, Doyle J 2000 *Phys. Rev. Lett.* **84** 2529
- [ 16 ] Reed W J, Hughes B D 2002 *Phys. Rev. E* **66** 067103
- [ 17 ] Hao B L 2001 *Physics* **30** 466 (in Chinese) [ 郝伯林 2001 物理 **30** 466 ]
- [ 18 ] Wei W F 2009 *Acta Phys. Sin.* **58** 2127 (in Chinese) [ 尉伟峰 2009 物理学报 **58** 2127 ]
- [ 19 ] Wei W F 2009 *Acta Phys. Sin.* **58** 6696 (in Chinese) [ 尉伟峰 2009 物理学报 **58** 6696 ]
- [ 20 ] <http://mathworld.wolfram.com/BoustrophedonTransform.html>
- [ 21 ] Team of Mathematic Handbook 1979 *Mathematic Handbook* (Beijing: Higher Education Press) p553—786 (in Chinese) [ 数学手册编写组 1979 数学手册 (北京: 高等教育出版社) 第 553—786 页 ]
- [ 22 ] <http://kottke.org/03/02/screw-power-law>

# Analysis on theory of selective model<sup>\*</sup>

Wei Wei-Feng<sup>†</sup>

( School of Electronics and Information Engineering , Beijing Jiaotong University , Beijing 100044 , China )

( Received 17 March 2009 ; revised manuscript received 29 March 2009 )

## Abstract

Selective model provided a way to explain the generation of index law and power law on the basis of preferential probability analysis. A transformation formula between the selective scope rule and selective distribution was deduced through the research on preferential trials. It was named as preferential transformation , which offered a mean to study the rule of monotone distributions. A corresponding continuous transformation can be deduced by the preferential transformation for analyzing probability density function. Compared with the discrete distribution , we defined a comparison function for continuous distribution , and pointed out the relationship between the discrete distribution and continuous distribution. Through selective model , we simulated and analyzed the power law phenomena. Selective theory was important to the research on huge complex system and social behavior. It could also expand the foundation of statistical physics and probability theory.

**Keywords** : selective model , power law phenomena , preferential transformation , comparison function

**PACC** : 0250 , 0500

---

<sup>\*</sup> Project supported by the National Natural Science Foundation of China ( Grant Nos. 60772043 , 60672069 ) , and the State Key Development Program for Basic Research of China ( Grant No. 2007CB307101 ) .

<sup>†</sup> E-mail : norax@sina.com