几何增长系统中的分布及其演化

Kim Chol-jun*Department of Astronomy, Faculty of Physics,* 金日成 *University, DPR Korea*邮政编码：+850电子邮箱：
cj.kim@ryongnamsan.edu.kp

2023年2月28日

arXiv:2302.13781v1 [physics.soc-ph] 24 Feb 2023

摘要

最近，我们提出了几何增长系统理论。本文表明，该理论无需引入精细的经济学模型，仅通过统计方法即可解释幂律分布的一些现象，包括经典的人口统计学和经济学现象，以及新出现的大流行病现象。分布的小规模部分具有凸性，这是该理论的一个特点，而幂律分布中不存在这一特点。我们发现，几何增长系统的分布在系统演化过程中可能趋于平坦，从而使系统内规模的相对比例增加。该系统可以充当一种逆向机器，将参数空间中的扩散转化为规模分布中的集中。

关键词：幂律；企业规模分布；新冠肺炎大流行；城市人口；谱硬化；

JEL分类代码：C11;O1

重要性：大多数看似呈现幂律分布的经济学系统，都是通过吉布拉特模型（即几何增长系统）进行分析的，而该模型似乎会产生对数正态分布。我们已证明，如果考虑参数之间的相关性，该系统可以产生渐近幂律。本文表明，只要系统的扩散（或方差增量）与系统的增长相伴，该系统就可能导致谱硬化。

1 引言

首先，我们解释相关问题和一些术语。一个系统由成员组成，我们将每个成员称为一个项目。项目具有一种可测量属性，称为规模。城市人口和企业规模可以视为规模，而城市和企业则代表国家这一系统中的项目。幂律，即齐夫定律或帕累托分布，指出某一项目出现的概率与该项目规模的某个幂成反比：$p(Z) = \frac{M}{Z^{\gamma}}$，其中 $p(Z)$ 表示规模为x的项目的频率，γ 表示幂指数，M 表示归一化常数。

从历史上看，Pareto（1896）指出，收入分布遵循幂律。Estoup（1916）和 Zipf（1932）在小说中观察到词频遵循幂律，而 Auerbach（1913）和 Zipf（1949）则指出，城市人口规模也遵循这一规律。许多种类繁多的事物都呈现幂律分布，关于这一现象的综述可参阅许多文献（例如 Mitzenmacher，2004；Newman，2005）。事实上，作者曾对宇宙射线谱感兴趣，它似乎是典型的幂律分布。Salpeter（1955年）发现，恒星质量分布遵循幂律。

针对幂律分布，人们提出了若干生成模型，可将其分为几类。第一类模型基于优先连接或“富者愈富”过程（Barabasi & Albert，1999年；Simon，1955年；Yule，1924年）。第二类模型追求尺度不变性，这是幂律分布的一个特性（Bak, Tang & Wiesenfeld，1987年；Sneppen等人，1995年）。第三类模型从需求优化出发（Mandelbrot，1953年）。此外，还有一些复合模型根据对参数的基本分布所作的特定假设，推导出幂律分布（Gabaix，1999年；Gibrat，1931年；Miller，1957年；Reed & Jorgensen，2004年）。这些模型展现了生成幂律分布的多种可能性。然而，它们都建立在特定假设之上。尽管逻辑始于假设，但假设最好具有普遍性；逻辑也最好能涵盖数据中更广泛的规模范围。

2 几何增长系统中分布的形式体系

最近，我们基于统计上最具合理性的假设，即参数分布的正态性，提出了一种几何增长系统（GGS）理论（Chol-jun，2022年）。如果系统中每个项目的规模都按几何方式或成比例增长，我们就称该系统为几何增长系统。几何增长系统可以建模为

$$Z = (1+\alpha)^t Z_0, \tag{1}$$

其中， Z 是系统中一个项目的规模， α 是增长率（下文简称增长）， t 表示增长年龄[1]，而 Z_0 是该项目的初始规模。

我们可以假设 α 和 t都服从正态分布，这在统计上是最合理的。这里，我们可以引入 α 与 t 之间的相关性 R，而不失一般性，因为即使 α 和 t的配置完全任意，也可以设定这种相关性。[2] 如果相关性为正（$R>0$），那么上限处的对数尺寸（$Y=\log Z$）可以近似为

$$Y_3 = (Ax_3 + B)^2 + C, \tag{2}$$

而如果相关性为零或负（$R \leqslant 0$），则可近似为[3]

$$Y_4 = Fx_4 + G, \tag{3}$$

其中， x_3， x_4是服从标准正态分布 $N(0,\ 1)$并且，当 μ_t， σ_t， μ_α， σ_α， μ_i和 σ_i分别表示 t， α 以及 $Y_0=\log\ Z_0$ 的均值和标准差时，这些参数如下所示:

$$A = a, \qquad B = \frac{\sqrt{4a^2b^2+d^2}}{2a}, \qquad C = c - \mathtt{sgn}(R)\frac{d^2}{4a^2}, \tag{4}$$

$$F = \mathtt{sgn}(R)(a^2+b^2)+c, \qquad G = \sqrt{2a^4+4a^2b^2+d^2}, \tag{5}$$

其中

$$a = \sqrt{\sigma_\alpha\sigma_t|R|}, \qquad b = \frac{|\mu_\alpha\sigma_t + \mathtt{sgn}(R)\mu_t\sigma_\alpha|}{2\sqrt{\sigma_\alpha\sigma_t}}, \qquad c = \mu_i + \mu_t\mu_\alpha - \mathtt{sgn}(R)b^2,$$
$$d = \sqrt{\sigma_i^2 + (\mu_t^2\sigma_\alpha^2 + \mu_\alpha^2\sigma_t^2)(1-|R|) + \sigma_t^2\sigma_\alpha^2(1-R^2)}, \tag{6}$$

并且 $\mathtt{sgn}(R)$ 表示 R 的符号，而 $|\cdot|$ 表示绝对值。

因此，如果 $R>0$，对数尺寸表现为 χ^2 变量；而在 $R \leqslant 0$的情况下，则表现为正态变量。我们可以推导出两种情况下规模的概率密度函数（PDF）：

$$p_{Z_3}(Z) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\sqrt{(\log Z - C)}-B}{A}\right)^2\right]}{2\sqrt{2\pi}AZ\sqrt{(\log Z - C)}} \qquad \text{for } R>0, \tag{7}$$

$$p_{Z_4}(Z) = \frac{1}{\sqrt{2\pi}FZ}\exp\left[-\frac{(\log Z - G)^2}{2F^2}\right] \qquad \text{for } R \leqslant 0. \tag{8}$$

我们将 $p_{Z_3}(Z)$ 自由度为 1 的对数完全平方卡方（χ）分布称为对数-CSχ_1分布（简称 log-CS），而 $p_{Z_4}(Z)$ 则是众所周知的对数正态分布。有趣的是，log-CS 分布的渐近指数，即其概率密度函数在双对数尺度图上的渐近斜率，会趋于一个常数:

$$\gamma_{3\infty} = \lim_{Y\to+\infty(Z\to+\infty)\ \text{in}\ R>0} \frac{d(\log(p_{Z_3}(Z)))}{d(\log Z)} = -\left(1+\frac{1}{2A^2}\right) = -\left(1+\frac{1}{2\sigma_\alpha\sigma_t R}\right), \tag{9}$$

[1]在 Chol-jun（2022年）中，年龄表示为 n。[2] Chol-jun（2022年）对新冠肺炎大流行的研究显示出一种系统性相关性：较晚爆发大流行病的国家，其增长相对较低，即呈现正相关，这可能是因为它们能够获得预警或进行准备。Gabaix & Ioannides (2004)指出，在某些年代，大城市增长更快；而在另一些年代，小城市增长更快。这意味着年龄（假设较老的城市规模更大）与增长之间的相关性符号会偶然发生反转。[3]F 和 G 与 Chol-jun（2022年）中的情况互换。

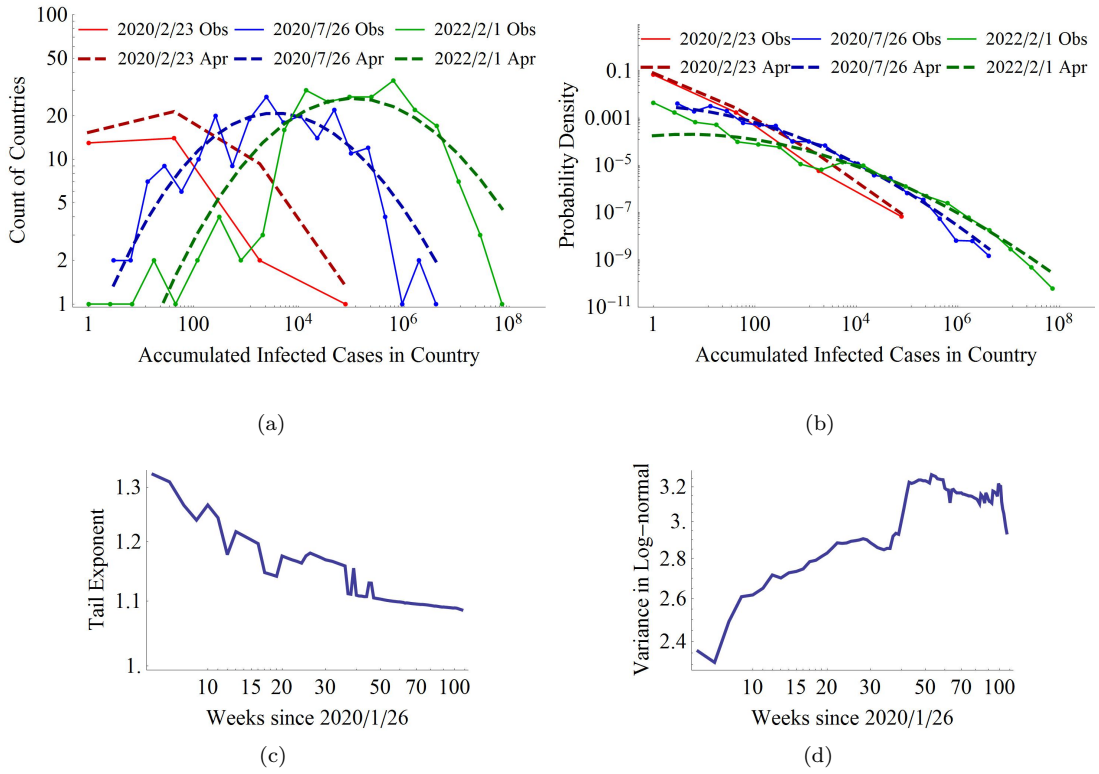


图1：各国新冠肺炎大流行累计感染病例的分布及其演化。传播的几个阶段中的分布，以及（a）计数直方图和（b）概率密度中的对数-CSχ_1 近似。（c）概率分布的尾指数（见（b））和（d）对数正态近似中的方差的演化。

这表明 log-CS 分布具有渐近幂律$p(Z) = \frac{M}{Z^{\gamma_{3\infty}}}$ 行为。特别是，渐近指数仅取决于年龄和增长的方差。顺便说一下，该指数为负，因此我们通常只考虑其绝对值。

对于对数正态分布，局部斜率由方差（更准确地说，是标准差） F决定，而方差又取决于参数的均值和方差。

3 新冠肺炎疫情统计及其演化

大流行病的传播可视为几何增长系统的典型例子。尽管存在季节性起伏、封锁措施、疫苗接种以及变异株的出现，自疫情暴发以来的两年间，大流行病的传播仍在加速。大流行病在国家之间传播的早期阶段，各国感染病例的幂律分布已有报道（Beare & Toda，2020年；Blasius，2020年）。然而，一旦国家间的传播趋于饱和，该分布就应偏离幂律分布。

Chol-jun（2022年）指出，2021年5月各国累计感染人数的分布可以用对数-CS极好地近似。请注意，这种近似并非通过最佳拟合得出，而是根据大流行病的演变历程推导出来的。事实上，年龄和增长的分布与正态分布相似，而且二者的相关性为正。[2] 图1展示了大流行病不同阶段的观测结果与对数-CS近似之间的一致性：例如，2020年2月下旬（早期阶段）、2020年7月下旬（国家间传播饱和阶段）以及2022年2月上旬（奥密克戎变异株传播阶段）。[4] 特别是在2020年7月，观测曲线围绕对数-CS近似曲线起伏。2022年2月上旬的分布似乎因奥密克戎变异株前所未有的快速传播而变得扭曲。

在上述计数直方图和概率直方图中，我们几乎无法察觉概率斜率的变化

[4]COVID-19 https://ourworldindata.org/ 传播数据可通过例如 Our World in Data 网站获取。 coronavirus/

曲线。如果我们t 尝试使用最大似然（ML）估计幂律指数（Newman，2005 ）

$$\hat{\gamma} = 1 + N \left(\sum_{i=1}^{N} \log \frac{z_i}{z_{\min}} \right)^{-1}, \tag{10}$$

其中 z_i 表示数据项目的规模， N 表示项目数量，我们应设定 $z_{\min}$，即允许的最小规模或截断尺度。不过，分布并非在整个规模范围内都遵循幂律，只在大规模的尾部区域遵循幂律。我们可以将 $z_{\min}$ 设为概率直方图（图1(b)）中的众数（概率最大的规模），并据此确定尾指数。图1(c)显示，尽管局部有所上升，尾指数在过去这段时间里似乎总体呈下降趋势。

我们还可以用对数正态分布的方差作为尾指数的另一种替代指标。事实上，分布也可以用对数正态分布（式 8）近似。该分布没有渐近指数，其局部斜率由方差 F （式 5）决定。我们也可以通过最大似然法估计该方差：

$$\hat{F}^2 = \frac{1}{N} \sum_{i=1}^{N} (\log z_i - \hat{G})^2, \qquad \hat{G} = \frac{1}{N} \sum_{i=1}^{N} \log z_i. \tag{11}$$

与上述尾指数评估方法相比，这种方法的优点是无需选择最佳截断尺度，即 $z_{\min}$ 。在对数-CS 分布和对数正态分布中，规模越大，斜率越大，因此即使分布本身保持不变，规模越大，评估得到的尾指数也可能越大，反之亦然。方差越大，局部斜率或尾指数越小。在新冠肺炎大流行的演化过程中，方差似乎在增大（图1(d)），这与前面提到的尾指数下降相一致。

4 城市与国家人口统计

城市人口分布是幂律的经典实例。几个世纪以来的人口增长呈现指数型或几何型曲线，但有时会达到饱和状态，因此可用逻辑斯蒂函数表示。因此，我们可以通过几何增长系统（GGS）来表达人口的演化。

首先，我们使用 stellarium-0.21.1 中的数据分析了一个国家内部的城市人口。[5] 在分析美国的城市人口时，Gabaix（1999年）指出，最大城市的人口服从幂律分布，且齐普夫指数几乎为 1。我们得到美国的齐普夫指数为 1.29（图 2(a)），这一数值不同于 Gabaix（1999年）所给出的 1；例如，伊拉克的齐普夫指数为 0.639。我们使用多种近似方法对美国城市人口进行了最佳拟合分析：对数-CS、对数正态分布和幂律分布（图 2(b)）。我们通过马尔可夫链蒙特卡罗（MCMC）方法推断最佳拟合参数，尤其采用了 Metropolis-Hastings算法（Hastings，1970年；Metropolis 等，1953年）。 R^2 对数-CS 的最佳拟合参数评估为： $R^2 = 0.9956$ ，幂律分布为 $R^2 = 0.9905$ ，对数正态近似为 $R^2 = 0.9833$ 。因此，在考察国家内部的城市人口时，我们可以优先选择对数-CS，因为它是最接近的近似。Ioannides与Skouras（2013年）声称，美国大多数城市服从对数正态分布，但其上尾部分以及因此占大多数的人口服从幂律分布。事实上，对数-CS 分布与对数正态分布几乎无法区分，只有在无穷远处例外（Chol-jun，2022年）。

接下来，我们利用《世界人口展望》（WPP）2015 数据集[6] 及其近似分布，对全球各国家和地区的人口分布进行最佳拟合分析（图 2(c)）。根据 MCMC方法推断出的最佳拟合参数，评估 R^2 ： 对数-CS为 $R^2 = 0.9851$ ，幂律为$R^2 = 0.9380$ ，对数正态近似为 $R^2 = 0.9920$ 。因此，在这种情况下，我们可以认为对数正态分布是最接近的近似。

另一方面，我们尝试应用 GSS 方法。我们预期年龄与增长之间存在相关性[2]。为此，我们查阅了《朝鲜大百科全书》中各国家的年龄信息[7]。我们以首个政权出现的时间作为国家的起始时期，例如首个王朝或城邦出现的时间、独立时间等。然而，对于许多非洲和美洲国家，我们应该考虑到

[5]我们使用开源天文软件 stellarium-0.21.1 中编制的、作为全球观测地点的城市数据集，该软件可在 Stellarium GitHub网页 https://github.com/Stellarium/stellarium/releases/ 获取。该数据集涵盖了城市及其位置、人口和其他信息，这些信息收集于至2019年之间。由于较低人口数据存在不连续性，我们将城市限定为人口超过20,000的城市。~24 000 2006

[6]数据可在《世界人口展望》的网站 https://population.un.org/wpp/ 获取。[7]《朝鲜大百科全书》，平壤，百科全书出版社，2001年

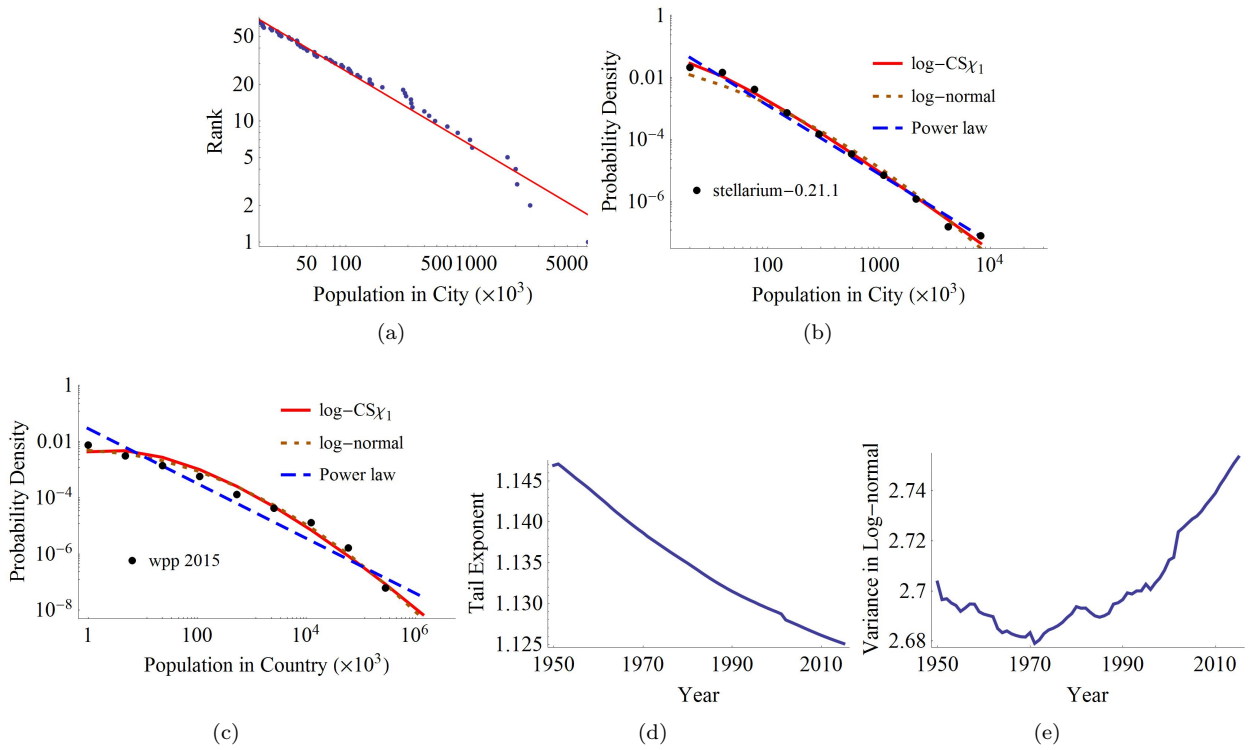


图2：城市和国家的人口分布及其演化。（a）美国城市的 log（排名）与 log（人口）关系图。线性回归得到 log（排名） = 6.1946 - 0.6387 log（人口）。（b）美国城市人口分布的最佳拟合。对数-CS似乎最为接近。（c）2015年世界各国和地区的人口。（d）世界各国和地区人口的对数正态近似中尾指数和（e）方差的演化。

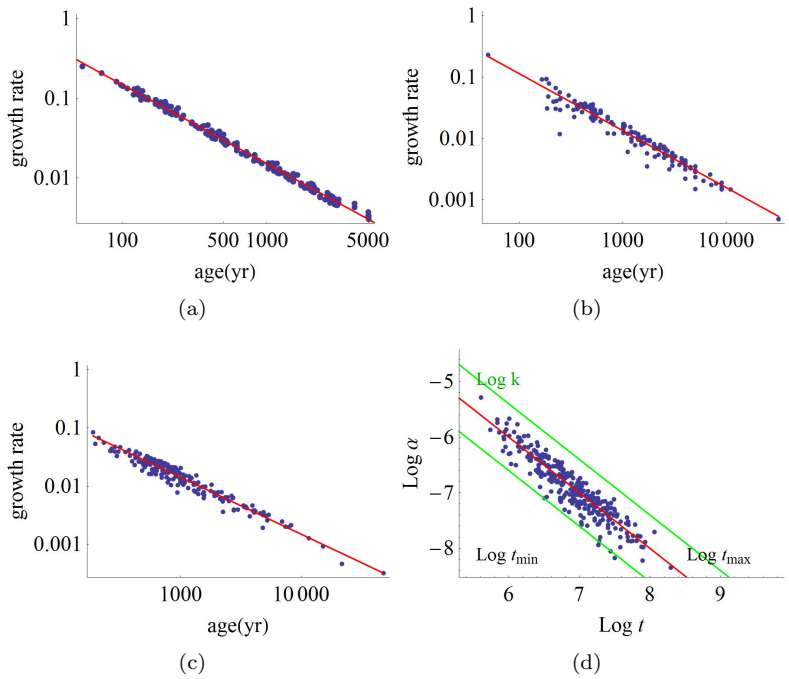


图3：假设极端初始条件时，全球国家和地区的人口年龄与增长之间的关系。(a)“行政年龄”与相应增长之间的关系。在双对数尺度下，斜率为 -0.9943，年龄与增长之间的相关性为 $R = -0.6191$。(b)对于“居住年龄”，斜率为 -0.9281，相关性为 $R = -0.3599$。(c)对于“外推年龄”，斜率为 -0.9771，相关性为 $R = -0.3911$。(d)规模差异形成一条条带，对应负相关。

殖民地的建立极大地改变了这些国家的人口构成。我们担心，在提取典型日期时，可能会将许多国家和地区缺失或失真的真实历史官方记录当作真实记录。我们对各国的增长率作如下评估：假设人口源自一对亚当和夏娃。Chol-jun（2022年）将这一近似方法应用于新冠肺炎大流行。令人惊讶的是，增长率与年龄呈现精确的反向关系，而且其指数几乎为1（图3(a)）。这使年龄与增长之间呈负相关，因此我们当然可以预期，世界人口应服从对数正态分布。

此外，我们还从 Britanica[8] 中提取了部落首次居住或移民的日期。这类年龄可称为“居住年龄”，比前述“行政年龄”古老得多。不过，仍然得到了负相关（图3(b)）。我们还考虑了另一种年龄，即根据人口的当前增长外推得出的年龄：将年龄的起点设定为初始人口同样为一对人。我们称这种年龄为“外推年龄”。图3(c)展示了根据 WPP数据集1950-2015年间的数据外推得到的年龄与增长之间的关系。同样，二者呈负相关。对于任何一种年龄，年龄与增长之间的反向关系都依然成立，且其指数接近1。

我们分析这一事实。由式（1）可推导出

$$\log z - \log z_0 = t \log(1+\alpha) \approx t \cdot \alpha. \tag{12}$$

如果 $t\alpha = b$ 和 b 被确定在 k 倍的范围内，那么 t 或 α 也被确定在 k 倍的范围内，并且在 t 与 α 的双对数图中会出现一条宽度为$\log k$ 的条带（图3(d)）。这条条带的斜率必为 -1，因此 α 与 t 之间呈负相关。要获得正相关，这条条带必须覆盖数据集中 t 的全部范围。因此，必须满足 $k \geqslant \sqrt{\frac{t_{\max}}{t_{\min}}}$，我们可以由公式（12）改写得：

$$\frac{(\log z - \log z_0)_{\max}}{(\log z - \log z_0)_{\min}} \geqslant \frac{t_{\max}}{t_{\min}}, \tag{13}$$

[8]《大英百科全书终极版》。参考套装。芝加哥：《大英百科全书》，2014年。

其中， $t_{\max}$ 和 $t_{\min}$ 可以是数据集的适当极值，例如 3σ 区域的极值。这可能是 α 与 t之间呈正相关的必要条件。总而言之，我们可以提出一个定理。

定理 1. 只有当式（13）满足条件时，几何增长系统中的增长与年龄之间才可能存在正相关。

由该定理我们可以推导出另一个结论：

推论 1. 如果规模最小的新项目在几何增长系统中持续产生，则该系统应可由对数正态分布近似。另一方面，在停止创建规模最小的新项目后，该系统可以由对数-*CS*近似。

对于全世界的国家，如果我们取初始条件 $z_0 = 2$ （一对）并认为目前各国人口的最大值和最小值分别为 $z_{\max} = 10^9$、 $z_{\min} = 10^3$ 和 $t_{\max} = 5000$ 和 $t_{\min} = 50$，那么式（ 13 ）右侧得到 100，而左侧只有 3。然而，如果我们考虑一个非恒定的初始条件（在实际情况下），左侧的值可能会大得多，从而可以进行对数-CS近似。对于新冠肺炎大流行的情况， $t_{\max} - t_{\min} \approx 120$ 和 $z_{\min} = 1$、 $z_0 = 1$，因此我们可以在国家之间的传播达到饱和后不久应用对数-CS近似。注意，式（13）并非充分条件。

我们考察世界人口分布的演化。WPP 2015数据集提供了1950—2015年各国家和地区的人口数据。如前所述，分布上端的斜率可以通过两种方法评估：尾指数和对数正态近似中的方差。图2(d)和2(e)显示，斜率变平的趋势与新冠肺炎大流行期间的情况相似。这意味着年龄和增长等参数的方差不断增大，导致规模分布的方差增大，尾指数则下降。

一国内城市人口的尾指数也会发生演化。引用以往研究，Gabaix & Ioannides (2004)指出，1900年至1990年间，美国的尾指数有所下降，这意味着人口集中度提高。Gonz´ales-Val（2010）指出，只要城市截断数量维持在10,000，尾指数就会随时间单调下降。有趣的是，观察1970—2007年中东欧国家的城市动态数据（Necula 等人，2010），我们发现，在大多数国家，指数与人口本身几乎呈负相关：人口增加时，指数下降，反之亦然（图4）。中世纪欧洲城市的指数似乎在1500年后开始下降，之后欧洲城市才开始出现齐夫定律（Bairoch、Batou & Ch`evre，1987；Gonz´ales-Val，2019；Dittmar，2011）。

考虑到各国在财富、规模和地理位置方面存在很大差异，Pinto、Lopes 与 Machado (2012) 声称，财富水平较高的国家往往表现出较高的指数值，而大多数非洲国家则呈现出较小的指数值。在我们的模拟中，伊拉克和中国等历史较悠久的亚洲国家似乎具有最小的指数。然而，他们的推理并不明显：如果这一推理是正确的，那么随着世界经济的发展，全球指数应当不断上升，但指数实际上必然是在下降。我们的方法可以给出一个清晰的解释：较年轻的国家可能具有较大的指数，反之亦然。事实上，最发达的国家大多较为年轻，而较老的国家往往处于欠发达状态，因此，较富裕的国家可能表现出较高的指数；不过，这只是偶然现象，并非必然如此。

Gabaix & Ioannides (2004) 将城市化与经济因素联系起来，例如经济一体化和国际贸易。那么，为什么尽管经济不断进步，一些国家的指数仍在上升？Necula 等（2010）提出，政治因素决定城市化进程。我们可以在考虑所有经济因素、政治因素或其他因素之前，或将它们纳入考量后，进行统计分析。例如，指数取决于参数的方差。

5 企业规模统计

幂律分布广泛出现在经济学和金融现象中（例如，参见 Farmer 与 Geanakoplos，2008）。很久以前，人们就指出，企业规模的幂律分布可以通过多种属性来测量（Ijiri 与 Simon，1977；Zipf，1949）。利用1997年经济普查，Axtell（2001）展示了以雇员人数和收入衡量的企业规模幂律分布。企业规模是一种可以适合按几何方式增长的量。事实上，我们通常用比例性而非可加性来评估企业的增长。

我们分析了 Axtell（2001）的数据，其中明确列出了以雇员人数表示的企业规模数值数据（即就业规模）。事实上，该分布在小规模部分呈凸形，相较于纯幂律模型，这更支持对数-CS模型。Giovanni，Levchenko

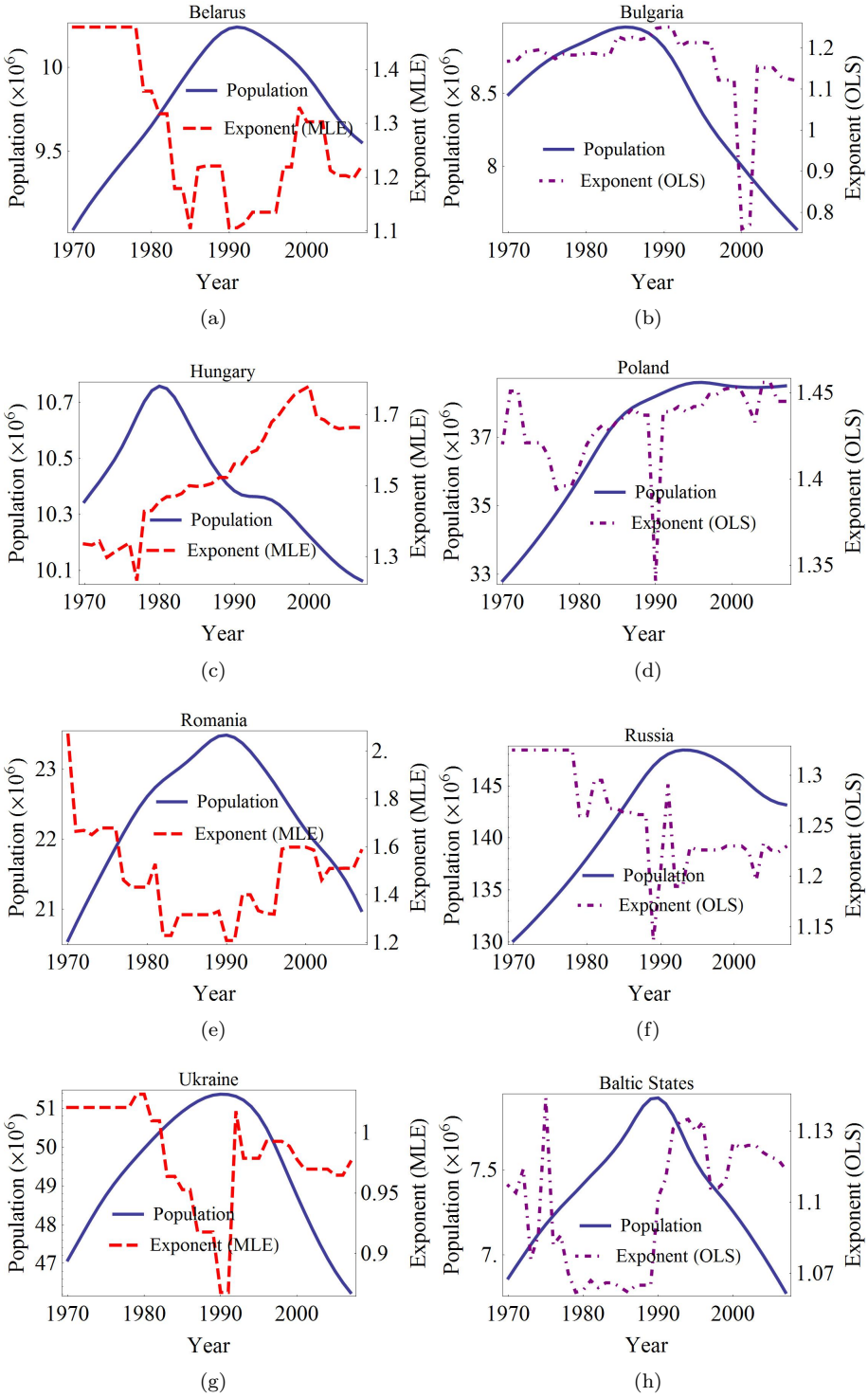


图4：中东欧国家人口与尾指数的演化。(a) 白俄罗斯，(b) 保加利亚，(c) 匈牙利，(d) 波兰，(e) 罗马尼亚，(f) 俄罗斯，(g) 乌克兰，(h) 波罗的海国家（爱沙尼亚、拉脱维亚和立陶宛）。数据摘自 Necula 等（2010），其中指数采用两种方法计算：用于线性回归的普通最小二乘法（OLS）和最大似然估计（MLE）。对于某些国家，两种指数估计方法中的一种显示出骤升骤降，因此我们将其忽略。除保加利亚和波兰外，对于大多数国家，我们可以观察到人口与尾指数在演化过程中呈反向关系。

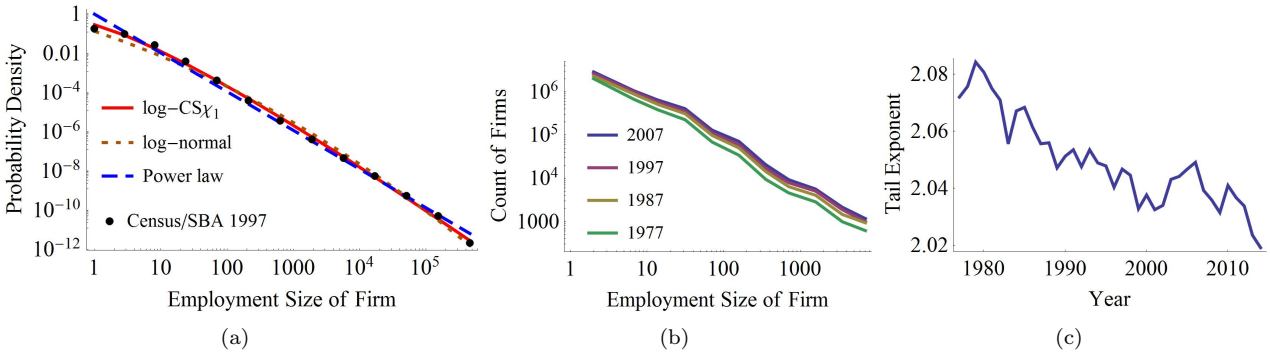


图5：企业就业规模的分布及尾指数的演化。(a) 人口普查局/小企业管理局1997年数据集中企业就业规模分布的对数-CS$_{\chi_1}$、对数正态分布和幂律拟合。BDS 1977—2014年数据集中企业就业规模的 (b) 直方图和 (c) 尾指数的演化。尾指数不是通过 (b) 这样的计数直方图，而是通过 (a) 这样的概率直方图进行线性回归估算的。

Ranci`ere（2010）分析了法国企业，得到了类似的凸形分布曲线。和 Axtell 一样，我们忽略规模为 0 的企业。我们使用 MCMC方法，对各种近似模型进行了最佳拟合（图5(a)）。幂律拟合得到的 R^2 与 Axtell 得到的相同： $R^2 = 0.9932$。对数-CS拟合得到的值更大： $R^2 = 0.9987$。我们应当预期，该数据集可以用对数正态分布近似：$R^2 = 0.9952$。这表明，对数-CS可能是最接近真实数据集的模型。

如前所述，根据增长与年龄之间的相关性，几何增长系统可以分别用对数-CS或对数正态分布进行建模。人口普查局商业动态统计和纵向商业数据库提供的按企业年龄划分的就业动态数据（1987年—2005年）显示，如果能够存续，年轻企业的就业增长率高于老年企业（Haltiwanger、Jarmin & Miranda，2009年、2010年）。这可能是因为，较老或规模较大的企业由于市场限制，其增长似乎已经饱和，而这一限制并不影响年轻企业；因此，尽管年轻企业的企业退出率更高，后者仍表现出更高的增长率。Navaretti、Castellani & Pieri（2012）分析了EFIGE调查的数据，该调查涵盖了2001年至2008年期间的法国、意大利和西班牙企业。结果显示，年轻企业在短期内（例如1年）和长期内（即其现有年龄阶段）都更有可能实现高增长率。因此，我们可以预期企业年龄与企业增长之间存在负相关。这应当会导致采用对数正态分布来拟合该分布，尽管实际数据集似乎更接近对数-CS。这可能源于年龄和增长并不服从正态分布。

我们追踪了该分布的演化。我们考察了1977—2014年的人口普查局商业动态统计（BDS）数据。[9] 尽管数据表明，与较高规模部分相比，较低规模部分的指数有所降低，这可能意味着分布具有凸性（图5(b)），但由于最高规模区间的上限不适合代表无穷大，我们可以通过线性回归估计尾指数，并排除最低和最高规模区间。图5(c)清楚地表明，指数随时间推移而下降。因此，我们可以看到尾指数正向更低的方向演化，即分布正在变平。正如前文所述，我们的方法可以给出一个解释：年龄或增长的方差增大，从而使规模的方差增加，规模分布趋于平坦。

6 结论与讨论

本文考虑了具有大流行病、人口统计学和经济现象的几何增长系统（GSS）中分布的一些特殊性质。

首先，该分布在规模较小的部分呈现凸性。这并不奇怪，因为它仅代表众数（最可能出现的规模）；这种情况在几乎所有分布中都很常见，但幂律分布中没有。事实上，对数-CS还具有额外的凹性和奇异性。在上述近似中，对数-CS分布或对数正态分布（也可能二者兼有）占主导地位，而非幂律分布。这意味着，GSS可以恰当地解释人口统计学、大流行病和经济学分布。事实上，在大多数情况下，对数-CS分布与对数正态分布之间的差异并不大，也不重要。重要的是，这两种分布都代表GSS的分布，而且二者通常都呈现凸性。然而，如果分布的

[9]数据可在 美国小企业管理局网站上获取 https://www.sba.gov/sites/default/files/advocacy/

参数偏离正态分布，这些分布的形态可能会发生变化。对于对数-CS分布或对数正态分布，不会出现具有特定指数的纯幂律分布所产生的发散问题。

如果新项目不断以较小的规模诞生并持续蓬勃增长，这种凸性就会减弱，分布似乎也会更加接近幂律。然而，一旦系统中的项目数量达到饱和，随着系统演化，低规模项目的数量便会减少，除非它们脱离整体、始终不再增长；与此同时，低规模部分会逐渐出现一种滚降。在这一时期的早期阶段，几何增长系统的分布应仅用对数正态分布来描述；而在项目数量饱和很久之后，则可以采用对数-CS拟合。

其次，随着年龄、增长等参数发生扩散，系统演化过程中尾指数降低，分布趋于平坦，这通常称为谱硬化。如前所述，分布的斜率完全或部分取决于参数的方差。在许多系统中，例如布朗运动，参数的方差会随时间增长。经济活动的多样性和经济增长的方差也会随时间加速扩大。热力学第二定律只表明，物质应通过扩散而散开。然而，物质却在宇宙中聚集并凝结。尽管从物理角度看，这可以用引力等因素来解释，但从统计角度看，几何增长系统可以充当一种逆向机器，将参数空间中的扩散转化为规模分布的集中。

分布变平反过来意味着，系统中规模最大与最小的项目之间的相对比例扩大。这或许可以解释城市化、垄断化等现象。城市化早在古代就已出现，例如古罗马。城市化不仅可能由经济和政治原因引发的迁移造成，也可能源于增长本身的随机性，例如出生率（或死亡率）不同，或由上述所有因素共同作用。如果将这一性质应用于呈幂律分布、且具有几何增长趋势的财富分配，那么只要钱能生钱，我们就可以预期，“富者愈富”的过程和经济体制中的垄断化会自然加剧。

有趣的是，集中度可能与扩散并存，甚至由扩散驱动。GSS 中的增长可能导致参数空间中的扩散，进而促使物质空间中的集中化。“富者愈富”现象并不意味着系统中的排名应当固定，也就是说，最富有或规模最大者并不能理所当然地一直稳居第一。排名可能由增长率决定，而增长率的多样性会随时间变化。正如我们所见，最富有的职业或规模最大的城市会随着时代更迭而交替。

我们可以在许多其他现象中发现几何增长系统的特性。我们希望我们的方法有助于分析这些问题。

利益冲突

作者没有需要披露的利益冲突。

数据可用性

本文所用数据　　可通过所示网站地址获取，也可联系作者获取。　　.

参考文献

Auerbach, F.，《人口集中定律》，《Petermanns 地理通讯》LIX（1913年），73—76；Axtell, R.，《美国企业规模的齐夫分布》，《科学》293，1818-1820页，2001年；Bairoch, P.、J. Batou和P. Ch`evre, 1988年，《欧洲城市人口》。日内瓦：Droz；Bak, P.、Tang, C.、Wiesenfeld, K.，《自组织临界性：对1/f噪声的解释》，《物理评论快报》59(4):381，1987年；Barabasi, A.、Albert, R.，《随机网络中标度的出现》，《科学》286，509-512（1999年）；Beare, B. K.、Toda, A. A.，《COVID-19 病例分布中幂律的出现》，《物理学 D》412:132649，arXiv:2004.12772v2 [physics.soc-ph]（2020年）；Blasius, B.，《确诊 COVID-19 病例数量的幂律分布》，《混沌》30，093123，https://doi.org/10.1063/5.0013031，arXiv:2004.00940v2 [q-bio.PE]（2020年）；Chol-jun, K.，《几何增长系统中的幂律分布：COVID-19 大流行病的统计》，《混沌》32，013111（2022年），doi: 10.1063/5.0068220

Dittmar, J.，2011年，《城市、市场与增长：齐夫定律的出现》，工作论文。伦敦政治经济学院，内部印行本；Estoup, J. B.，《速记符号》，巴黎：法国速记学院，1916年；Farmer J. D.、Geanakoplos J.，2008年，《经济学及其他领域中的幂律》，《经济学年度评论》1(1)，1；Gabaix, X.，《城市齐夫定律：一种解释》，《经济学季刊》114(3)，739-767（1999年）；Gabaix, X.、Ioannides, Y. M.，2004年，《城市规模分布的演变》，载于《城市与区域经济学手册》第4卷，第53章，J. V. Henderson和J. F. Thisse编，2341-2378页。阿姆斯特丹：爱思唯尔科学出版社；Gibrat, R.，《经济不平等》，Libraire du Recueil Siray，巴黎（1931年）；di Giovanni, J.、Levchenko, A. A.、Ranci`ere, R.，（2010年），《企业规模与贸易开放度中的幂律：测度与含义》，国际货币基金组织工作论文，WP/10/109；Gonz´ales-Val, R.，2010年，《从长期视角看美国城市规模分布的演变（1900-2000年）》，《区域科学杂志》50(5)，952-972；Gonz´ales-Val, R.，2019年，《欧洲历史城市增长（1300-1800年）》，《区域科学论文集》98(2)，1115-1136；Haltiwanger, J. C.、Jarmin, R. S.、Miranda, J.，《企业动态统计简报：高增长企业、失败企业或年轻企业》，使用美国人口普查局企业动态统计数据的系列报告之四，2009年；Haltiwanger, J. C.、Jarmin, R. S.、Miranda, J.，《谁创造就业岗位：小企业、大企业还是年轻企业》，美国国家经济研究局工作论文16300，马萨诸塞州剑桥02138，2010年；Hastings, W. K.，（1970年）《使用马尔可夫链的蒙特卡罗抽样方法及其应用》，《生物统计学》52(1):97-109；Ijiri, Y.、Simon, H. A.，《偏态分布与企业规模》，《数理经济学研究》。北荷兰出版社，阿姆斯特丹，1977年；Ioannides, Y.、Skouras, S.，2013年，《美国城市规模分布：稳健地服从帕累托分布，但仅在尾部如此》，《城市经济学杂志》73，18-29；Mandelbrot, B.，1953年，《语言统计结构的信息论》（马萨诸塞州沃本：Butterworth）；Mitzenmacher, M.，《幂律和对数正态分布生成模型简史》，《互联网数学》1，226-251（2004年）；Metropolis, N.、Rosenbluth, A. W.、Rosenbluth, M. N.、Teller, A. H.和Teller, E.，1953年，《利用快速计算机进行状态方程计算》，《化学物理学杂志》21(6):1087-1092；Miller, G. A.，《间歇性沉默的一些影响》，1957年，《美国心理学杂志》70，311-314；Navaretti, G. B.、Castellani, D.、Pieri, F.，《年龄对企业规模动态的影响：学习效应还是增长意愿？》，EFIGE工作论文60，2012年；Necula, C.、Ibragimov, M.、Valetka, V.、Bobeica, G.、Radu, A-N.、Mukhamedkhanova, K.、Radyna, A.，2010年，《转型经济体中的城市规模分布动态：一项跨国研究》，全球发展网络工作论文；Newman, M. E. J.，《幂律、帕累托分布与齐夫定律》，《当代物理学》第46卷，323-351页，http://arxiv.org/abs/cond-mat/0412004v3（2005年）；Pareto, V.，《政治经济学教程》。瑞士日内瓦，Droz（1896年）；Pinto, C. M. A.、Lopes, A. M.、Machado, J. A. T.，《现实生活现象中的幂律综述》，《非线性科学与数值模拟通讯》17(9)，47:3558-78，doi:/10.1016/j.cnsns.2012.01.013；Reed, W.、J.、Jorgensen, M.，《双帕累托对数正态分布——一种新的规模分布参数模型》，《统计通讯：理论与方法》33(8):1733-1753，DOI: 10.1081/STA-120037438（2004年）；Salpeter, E. E.，《光度函数与恒星演化》，《天体物理学杂志》121，161-167（1955年）；Simon, H. A.，《一类偏态分布函数》，《生物统计学》42(3/4)，425-440，DOI:10.2307/2333389（1955年）；Sneppen, K.、Bak, P.、Flyvbjerg, H.、Jenssen, M. H.，《作为自组织临界现象的演化》，《美国国家科学院院刊》92，5209-5213（1995年）；Yule, G.，《基于J. C. Willis博士（F.R.S.）结论的进化的数学理论》，《皇家学会哲学汇刊》，伦敦，213，21-87（1924年）

Zipf, G., 语言中的选择性研究与相对频率原理。马萨诸塞州剑桥：哈佛大学出版社（1932年）Zipf, G., 人类行为与最小努力原则。马萨诸塞州剑桥：Addison-Wesley（1949年）