连续型幂律分布的参数估计

陈月萍，陈庆华

（福建师范大学 数学与计算机学院，福州　350007）

摘　要：文章探索连续型幂律分布的参数估计，主要研究参数的极大似然估计和有效估计. 给出了参数的一个渐近有效估计，并且证明了不存在参数的有效无偏估计。

关键词：幂律分布；极大似然估计；有效估计
doi：10.3969/j.issn.1000－5757.2012.03.113
中图分类号：O21　　文献标志码：A　　文章编号：1000-5757（2012）03-0113-04

1 引言

幂律关系广泛存在于物理学、生物学、地球与行星科学、经济与金融学、计算机科学、人口统计学和社会科学中.[1] 例如，城市人口、地震大小、[2] 行星间碎片大小的分布、[3] 月球表面上月坑直径的分布、[4] 战争规模的分布[5] 和个人财富的分布[6] 均服从幂律关系.

近年来，科学家们在复杂网络的研究过程中亦发现了许多大规模网络的度分布都是幂律分布，呈现无标度特征. 虽然这些网络在结构及功能上是如此的千变万化，相差迥异. 如万维网网页的入度分布、出度分布都具有幂律尾部；[7] 1999 年，Faloutos 等人在两个层次研究了因特网，其度分布都遵循幂律分布；[8] 1998 年，Redner 研究了由科学信息协会按目录列入的 783339 篇论文的引文分布，统计结果表明一篇论文被引用 k 次的概率服从幂律分布。[9]

目前许多学者关注幂律是如何形成的. 近几十年来，为了解释幂律分布形成的原因，科学家们提出了几种不同的机制，包括增长与优先连接、[10] 自组织临界、[11] HOT 理论[12] 等. 对于一些随机过程，如“随机行走”模型可以解释物种寿命所呈现的幂律分布.[13]

传统上，数理统计学研究的概率分布基本上都是轻尾部的，而幂律分布是重尾部的，具有本质区别；特别地，离散型幂律分布的统计分析是比较困难的.

2005 年，Newman 给出了连续型幂律分布参数的极大似然估计.[1] 本文在此基础上计算了极大似然估计的期望和方差，进而给出了参数的一个渐近有效估计，并且证明了不存在参数的有效无偏估计。

2 极大似然估计

定义 1　设一个连续型随机变量 X 的概率密度函数具有幂律形式：

$$p(x;\alpha)=\frac{\alpha-1}{x_{\min}}\left(\frac{x}{x_{\min}}\right)^{-\alpha},x\geq x_{\min}>0,\alpha>1 \quad (1)$$

则称随机变量 X 服从幂律分布，其中 $x_{\min}$ 是 X 的最小值，α 是唯一的分布参数.

Newman 在文献[1]中给出了未知参数 α 的极大似然估计，即有如下定理 1：

定理 1[1]：设总体 X 服从连续型幂律分布，具有概率密度函数 $p(x;\alpha)$，$X_1,X_2,\cdots,X_n$ 为取自该总体的随机样本，则未知参数 α 的极大似然估计为

$$\hat{\alpha}=1+n\left[\sum_{i=1}^{n}\ln\frac{X_i}{x_{\min}}\right]^{-1} \quad (2)$$

证明：设 $x_1,x_2,\cdots,x_n$ 为样本观测值，则似然函数为

*收稿日期：2011-11-07
基金项目：福建省自然科学基金资助项目（Z0511020）
作者简介：陈月萍（1987—），女，福建泉州人，硕士研究生，研究方向：随机图与复杂网络；
陈庆华（1962—），男，福建莆田人，教授，博士，研究方向：应用统计学。

$$L = L(\alpha; x_1, \cdots x_n)$$
$$= \prod_{i=1}^{n} p(x_i; \alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{x_{\min}} (\frac{x_i}{x_{\min}})^{-\alpha}$$

两边取对数

$$\ln L = \sum_{i=1}^{n} [\ln(\alpha - 1) - \ln x_{\min} - \alpha \ln \frac{x_i}{x_{\min}}]$$
$$= n\ln(\alpha - 1) - n\ln x_{\min} - \alpha \sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}$$

则似然方程为

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha - 1} - \sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}} = 0$$

解得

$$\hat{\alpha} = 1 + n [\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}]^{-1}$$

而且

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = - \frac{n}{(\alpha - 1)^2} < 0$$

所以，$\hat{\alpha}$ 为未知参数 α 的极大似然估计.

定理 2：极大似然估计 $\hat{\alpha}$ 的数学期望和方差分别为:

$$E(\hat{\alpha}) = 1 + n \frac{\alpha - 1}{n - 1} = 1 + \frac{n}{n - 1}(\alpha - 1) \tag{3}$$

$$D(\hat{\alpha}) = n^2 \frac{(\alpha - 1)^2}{(n - 1)^2 (n - 2)}$$
$$= \frac{n^2}{(n - 1)^2 (n - 2)} (\alpha - 1)^2 \tag{4}$$

证明：$\hat{\alpha}$ 的期望和方差分别为

$$E(\hat{\alpha}) = 1 + nE [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-1},$$

$$D(\hat{\alpha}) = n^2 D [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-1}$$

令 $Y_i = \ln \frac{X_i}{x_{\min}}$，$i = 1, 2, \cdots, n$，则 $Y_i > 0$ (a.e)，并有

$$P(Y_i = \ln \frac{X_i}{x_{\min}} < y) = P(x_{\min} < X_i < x_{\min} e^y)$$
$$= \int_{x_{\min}}^{x_{\min} e^y} p(x; \alpha) dx$$
$$= -(\frac{x}{x_{\min}})^{-\alpha + 1} \Big|_{x_{\min}}^{x_{\min} e^y}$$
$$= 1 - e^{-(\alpha - 1) y}$$

所以，$Y_i = \ln \frac{X_i}{x_{\min}}$ 的密度函数为

$$p_Y(y; \alpha) = \begin{cases} (\alpha - 1) e^{-(\alpha - 1) y} & y > 0 \\ 0 & y \leqslant 0 \end{cases}$$

即 $Y_i = \ln \frac{X_i}{x_{\min}}$ ($i = 1, 2, \cdots, n$) 服从参数为 $\alpha - 1$ 的指数分布，且 $Y_1, Y_2, \cdots Y_n$ 相互独立，则

$Z = \sum_{i=1}^{n} Y_i$ 服从参数为 $(n, \alpha - 1)$ 的 Gamma 分布 $\Gamma(n, \alpha - 1)$。

因此

$$E [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-1} = \int_0^{+\infty} \frac{1}{z} \frac{(\alpha - 1)^n}{\Gamma(n)} z^{n-1} e^{-(\alpha - 1) z} dz$$
$$= \frac{\alpha - 1}{\Gamma(n)} \int_0^{+\infty} (\alpha - 1)^{n-1} z^{(n-1)-1} e^{-(\alpha - 1) z} dz$$
$$= \frac{(\alpha - 1) \Gamma(n - 1)}{\Gamma(n)}$$
$$= \frac{\alpha - 1}{n - 1}$$

$$E [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-2} = \int_0^{+\infty} \frac{1}{z^2} \frac{(\alpha - 1)^n}{\Gamma(n)} z^{n-1} e^{-(\alpha - 1) z} dz$$
$$= \frac{(\alpha - 1)^2}{\Gamma(n)} \int_0^{+\infty} (\alpha - 1)^{n-2} z^{(n-2)-1} e^{-(\alpha - 1) z} dz$$
$$= \frac{(\alpha - 1)^2 \Gamma(n - 2)}{\Gamma(n)}$$
$$= \frac{(\alpha - 1)^2}{(n - 1)(n - 2)}$$

$$D [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-1} = E [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-2} - \{E [\sum_{i=1}^{n} \ln \frac{X_i}{x_{\min}}]^{-1}\}^2$$
$$= \frac{(\alpha - 1)^2}{(n - 1)(n - 2)} - \frac{(\alpha - 1)^2}{(n - 1)^2}$$
$$= \frac{(\alpha - 1)^2}{(n - 1)^2 (n - 2)}$$

从而

$$E(\hat{\alpha}) = 1 + n \frac{\alpha - 1}{n - 1} = 1 + \frac{n}{n - 1}(\alpha - 1)$$

$$D(\hat{\alpha}) = n^2 \frac{(\alpha - 1)^2}{(n - 1)^2 (n - 2)}$$
$$= \frac{n^2}{(n - 1)^2 (n - 2)} (\alpha - 1)^2.$$

3 有效估计

定理 3：对于连续型幂律分布族 $\{p(x, \alpha) : \alpha > 1\}$，存在参数 α 的 Cramer - Rao 下界.

证明：将(1)式改写成

$$p(x;\alpha)=\frac{\alpha-1}{x_{\min}^{1-\alpha}}\exp\{-\alpha\ln x\}I_{\{x\geqslant x_{\min}>0\}},\alpha>1 \quad (5)$$

可取 $c(\alpha)=\frac{\alpha-1}{x_{\min}^{1-\alpha}}$,$c_1(\alpha)=-\alpha$,$T_1(x)=\ln x$,$h(x)=I_{\{x\geqslant x_{\min}>0\}}$,并且其支撑:$\{x:p(x;\alpha)>0\}=[x_{\min},+\infty)$ 不依赖于未知参数 α,则连续型幂律分布族是指数型分布族,[14] 而指数族是 *Cramer - Rao* 正则族,[14] 所以存在参数 α 的 C - R 下界.

定理 4:对于连续型幂律分布族 $\{p(x,\alpha):\alpha>1\}$,不存在参数 α 的有效无偏估计.

证明:设 $X=(X_1,X_2,\cdots,X_n)$ 为其随机样本,则样本 X 的联合概率密度函数为

$$\prod_{i=1}^{n}p(x_i;\alpha)=\exp\{-\alpha\sum_{i=1}^{n}\ln\frac{x_i}{x_{\min}}+n\ln\frac{\alpha-1}{x_{\min}}\} \quad (6)$$

令 $T(x)=-\sum_{i=1}^{n}\ln\frac{x_i}{x_{\min}}$,$c(\alpha)=n\ln\frac{\alpha-1}{x_{\min}}$,$d(x)=0$,则由上述定理 2 的证明过程可知,$T(X)=-Z$,且 $Z\sim\Gamma(n,\alpha-1)$,所以

$$E_\alpha T(X)=-E(Z)=-\frac{n}{\alpha-1}$$

显然,不存在常数 a,b,使得 $\alpha=a\times E_\alpha T(X)+b$ 成立,根据参考文献[14]中的定理 2.8,不存在参数 α 的有效无偏估计.

定理 5:设总体 X 服从连续型幂律分布,具有概率密度函数 $p(x;\alpha)$,$X_1,X_2,\cdots,X_n$ 为取自该总体的随机样本,则统计量

$$\tilde{\alpha}=(n-1)(\sum_{i=1}^{n}ln\frac{X_i}{X_{min}})^{-1}+1 \quad (7)$$

为未知参数 α 的一个渐近有效估计.

证明:将(1)式取对数后得

$$\ln p(x;\alpha)=\ln(\alpha-1)-\ln x_{\min}-\alpha\ln\frac{x}{x_{\min}}$$

则上式对 α 的二阶偏导数为

$$\frac{\partial}{\partial\alpha}\ln p(x;\alpha)=\frac{1}{\alpha-1}-\ln\frac{x}{x_{\min}}$$

$$\frac{\partial^2\ln p(x;\alpha)}{\partial\alpha^2}=-\frac{1}{(\alpha-1)^2}$$

根据信息量的性质可知,信息量

$$I(\alpha)=-E\left[\frac{\partial^2\ln p(X;\alpha)}{\partial\alpha^2}\right]=\frac{1}{(\alpha-1)^2}$$

从而参数 α 的 Cramer - Rao 下界为:

$$\frac{1}{nI(\alpha)}=\frac{(\alpha-1)^2}{n}$$

现在选取统计量

$$\tilde{\alpha}=(n-1)(\sum_{i=1}^{n}\ln\frac{X_i}{x_{\min}})^{-1}+1$$

作为参数 α 的一个估计.由定理 2,估计量 $\tilde{\alpha}$ 的期望和方差分别为

$$\begin{aligned}E(\tilde{\alpha})&=(n-1)E\left[\sum_{i=1}^{n}\ln\frac{X_i}{x_{\min}}\right]^{-1}+1\\&=(n-1)\frac{\alpha-1}{n-1}+1=\alpha\end{aligned}$$

$$\begin{aligned}D(\tilde{\alpha})&=(n-1)^2D\left[\sum_{i=1}^{n}\ln\frac{X_i}{x_{\min}}\right]^{-1}\\&=(n-1)^2\frac{(\alpha-1)^2}{(n-1)^2(n-2)}\\&=\frac{(\alpha-1)^2}{n-2}\end{aligned}$$

因此,$\tilde{\alpha}$ 是参数 α 的一个无偏估计量.但是,其方差大于 Cramer - Rao 下界,即

$$D(\tilde{\alpha})=\frac{(\alpha-1)^2}{n-2}>\frac{(\alpha-1)^2}{n}=\frac{1}{nI(\alpha)}$$

所以 $\tilde{\alpha}$ 不是参数 α 的一个有效估计,其有效率为

$$e=\frac{\frac{1}{nI(\alpha)}}{D(\tilde{\alpha})}=\frac{\frac{(\alpha-1)^2}{n}}{\frac{(\alpha-1)^2}{n-2}}=\frac{n-2}{n}<1$$

当 $n\to+\infty$ 时,有效率 $e\to1$,所以是 $\tilde{\alpha}$ 参数 α 的一个渐近有效估计.

4 结论与讨论

本文主要研究了连续型幂律分布的参数估计,给出了参数的极大似然估计和一个渐近有效估计.我们将进一步研究参数的假设检验,以及离散型幂律分布的统计分析.

参考文献:

[1] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law [J]. Contemporary Physics, 2005, 46(5).

[2] B. Gutenberg and R. F. Richter, Frequency of earth-quakes in california [J]. Bulle - tin of the Seismological Society of America, 1944, (34): 185 - 188.

[3] E. T. Lu and R. J. Hamilton, Avalanches of the distribution of solar Flares [J]. A - strophysical Journal, 1991, (380): 89 - 92.

[4] G. Neukum and B. A. lvanov, Crater size distributions and impact probabilities on Earth from lunar, terrestialplanet,

and asteroid cratering data [Z] . In T. Geh - rels(ed.) , Hazards Due to Comets and Asteroids, University of Arizona Press, Tucson, AZ, 1994, 359 - 416.

[5] D. C. Roberts and D. L. Turcotte, Fractality and selforganized criticality of w - Ars [J] . Fractals, 1998, (6) : 351 - 357.

[6] V. Pareto, Cours d´Economie Politique, Droz, Geneva, 1896.

[7] R. Albert, H. Jeong and A. - L. Barabási, Diameter of the world - wide web [J] . Nature, 1999, (401) : 130 - 131.

[8] M. Faloutsos, P. Faloutsos and C. Faloutsos, On power - law relationships of the internet topolopy [J] . Computer Communications Review, 1999, (29) : 251 - 262.

[9] S. Redner, How popular is your paper? An empirical study of the citation dist - ribution [J] . Eur. Phy. J. B, 1998, (4) : 131 - 134.

[10] Albert R, A. - L. Barabási, Science, 1999, (286) : 509.

[11] Bak P, Tang C, Wiesenfeld K. Phys. Rev. Lett, 1987, (59) : 381.

[12] Carls on J M, Doyle J, Phys. Rev. E, 1999, (60) : 1412.

[13] Sneppen K, Bak P, Flyvbjerg Hetal, Proc Natl Acad Sci USA, 1995, (92) : 5209.

[14] 茆诗松，王静龙，濮晓龙．高等数理统计(第二版) [M] ．北京: 高等教育出版社，2006.

[15] 严士健，刘秀芳．测度与概率 [M] ．北京: 北京师范大学出版社，2003.

Parameter Estimations for Continuous Power - law Distributions

CHEN Yue-ping, CHEN Qing-hua

(School of Mathematics and Computer, Fujian Normal University, Fuzhou 350007, China)

Abstract: This paper explores the parameter estimations of continuous power - law distributions. Maximum likelihood estimation and effective estimation of the parameter are researched. An asymptotically effective estimation is given, and it is proved that the effectively unbiased estimation is inexistent.

Key words: power - law distributions; maximum likelihood estimation; effective estimation

(责任编辑: 刘春林　责任校对: 林　子)

(上接第 112 页) 表多篇科研学术论文，体现出较强的专业素质和创新能力。

参考文献:

[1] 胡寿松．自动控制原理(第五版) [M] ．北京: 科学出版社，2008.

[2] 马德秀．寻找人才培养模式突破 致力培养创新人才 [J] ．中国高等教育，2006，(11) : 19 - 21.

[3] 任长松．探究式学习中的科学观教育: 18 条内涵 [J] ．教育理论与实践，2005，(5) : 38 - 42.

[4] 冯爱祥，左信．“自动控制原理”课程实施研究型教学模式的探讨 [J] ．电气电子教学学报，2009，(5) : 12 - 13 .

[5] [美] 戴尔蒙德 著，黄小苹 译．课程与课程体系的设计和评价实用指南 [M] ．浙江: 浙江大学出版社，2006.

Research on Inquiry Teaching of Automatic Control Theory Course

FU Qiang

(College of Science and Technology, Ningbo University, Ningbo Zhejiang 315212, China)

Abstract: Based on the automatic control theory course, this paper presents a series of inquiry teaching measures: advancing students´ inquiry willingness and innovation consciousness through project - driven method, cultivating students´ critical thinking skills by interactive learning, promoting students´ participation and initiative with the advanced teaching means and reasonable evaluation system. Practice shows that inquiry teaching can help students to improve their ability of innovation and application in the process of learning professional knowledge and better meet the needs of society for applied innovative talents demand.

Key words: automatic control theory; inquiry teaching; project - driven; interactive learning; evaluation system

(责任编辑: 刘春林　责任校对: 林　子)