

基于粒子群优化算法的洛伦兹曲线参数估计及基尼系数的计算

张 转¹, 常安定¹, 岳莎莎², 王晓晨¹

(长安大学 1. 理学院; 2. 经济与管理学院, 西安 710064)

摘要:根据已有洛伦兹曲线模型,应用粒子群优化算法对洛伦兹曲线参数进行估计。由拟合的洛伦兹曲线计算 2002—2011 年我国城镇和农村的基尼系数,再利用“分组加权法”得出我国总体的基尼系数。结果表明:粒子群优化算法可以有效地估计洛伦兹曲线中的参数;由该算法的估计参数所确定的洛伦兹曲线估算 2002—2011 年基尼系数,其变化趋势与其它预测相符。从而为洛伦兹曲线提供了可靠、有效的参数估计方法。

关键词:粒子群优化算法;洛伦兹曲线;基尼系数

中图分类号: F224 **文献标志码:** A **文章编号:** 1671-1807(2014)02-0083-05

居民收入分配关系到广大人民群众的生活水平,基尼系数(Gini coefficient)是衡量一个经济体收入分配状况重要经济指标,洛伦兹(Lorenz curve)曲线是用来刻画收入分配不平等程度

经济指标。图 1 中所示的直线 $y = x$ 为绝对平等线,曲线即为洛伦兹曲线。

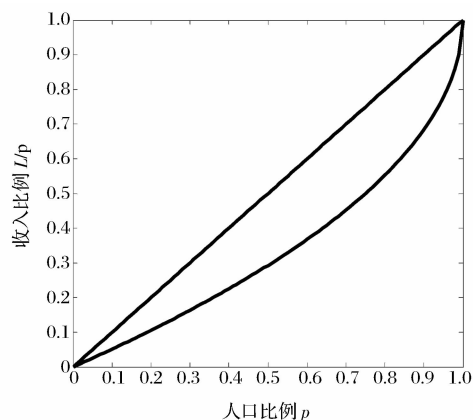


图 1 洛伦兹曲线

洛伦兹曲线与基尼系数之间的关系式为:

$$G = \frac{S_A}{S_A + S_B} = 2S_A = 1 - 2S_B \quad \text{其中 } S_A \text{ 为绝对}$$

平均线与洛伦兹曲线间的几何面积, S_B 为洛伦兹曲线下方的曲三角形面积,对 S_B 的计算大致可分为离

散法和连续法两种^[1],其中连续法是用一个函数 $L(p)$ 来拟合洛伦兹曲线,则 $S_B = \int_0^1 L(p) dp$ 。由此基尼系数可定义为:

$$G = 1 - 2 \int_0^1 L(p) dp \quad (1)$$

对于洛伦兹曲线方程,国内外学者如 Bas-mann^[2]、Sarabia^[3]、Chotikapanich^[4]已提出了各种形式的函数来拟合洛伦兹曲线,已有的这些模型存在诸多问题。例如:拟合效果欠佳,有些模型甚至不满足洛伦兹曲线的条件。本文采用张奎、王原提出的 Sarabia 洛伦兹曲线的推广模型^[5],数值实验说明此模型拟合效果远优于前人提出的模型。而对模型中参数的估计大都采用传统的方法,如:密度函数的 Kernel 估计法、无约束非线性最小二乘估计法以及回归分析等。现代智能优化算法在经济、工程各个领域中有广泛的应用。隶属于智能优化算法的粒子群优化算法对复杂非线性问题具有较强的寻优能力以及简单通用等显著特点,此算法概念清晰,容易实现,同时又有深刻的智能背景。鉴于此,本文对模型中的参数尝试采用粒子群优化算法进行估计,进而分别确立城镇与农村的洛伦兹曲线。由该算法估计参数确定的洛伦兹曲线计算 2002—2011 年我国城镇和农村的基尼系数,再利用“分组加权法”^[6]得出总体

收稿日期: 2013-12-06

基金项目: 陕西省软科学项目(2013KRM46)

作者简介: 张转(1988—),女,陕西咸阳人,长安大学理学院,在读硕士,研究方向:最优化理论与方法;通讯作者:常安定(1964—),男,陕西大荔人,长安大学理学院硕导,教授,博士,研究方向:水文地质的数学方法。

的基尼系数。

1 洛伦兹曲线方程与基尼系数

1.1 洛伦兹曲线方程

$L(p, \tau)$ 是定义在 $[0, 1]$ 区间上、取值于 $[0, 1]$ 区间上的函数,若满足以下特征

$$L(0, \tau) = 0, L(1, \tau) = 1, L'(p, \tau) \geq 0, L''(p, \tau) \leq 0$$
 (2)

则 $L(p, \tau)$ 被称为洛伦兹曲线方程^[7], τ 为洛伦兹曲线方程中的参数。文献[8]中已提出很多种洛伦兹曲线模型,大都是一次函数、二次函数、幂函数和指数函数。本文采用张奎,王原提出的对 Sarabia 洛伦兹曲线的推广模型:

$$L(p) = p^\alpha [1 - (1 - p^\gamma)^\beta]^\eta$$
 (3)

其中 $\alpha \geq 0, \gamma \geq 0, \beta \in (0, 1], \eta \geq 1, \alpha + \gamma \geq 1$ 均为参数。 p 为收入群体的累计人口比例, $0 \leq p \leq 1$; L 为该群体拥有的总收入比例, $0 \leq L \leq 1$ 。

1.2 基尼系数

首先根据 2002—2011 年我国农村和城市分组数据拟合出历年农村和城市的洛伦兹曲线 $L_1(p)$ 和 $L_2(p)$, 由此计算出各年的农村和城镇的基尼系数 G_1 和 G_2 , 再利用 Sundrum 提出的“分组加权法”, 计算总体基尼系数。总体基尼系数计算公式为:

$$G = P_1^2 \frac{m_1}{m} G_1 + P_2^2 \frac{m_2}{m} G_2 + P_1 P_2 \left| \frac{m_2 - m_1}{m} \right|$$
 (4)

式(4)中 m 为全体居民的人均收入, m_1 为农村居民人均纯收入, m_2 为城镇居民人均总收入, P_1 和 P_2 分别为当年农村和城市的人口比例。“利用此式计算总体的居民收入基尼系数时的假设条件是‘城乡居民收入分布不重叠’, 但我国全国和各省级区域城镇和农村收入分布大多存在重叠现象”^[9], 收入分组不重叠这一难题并没有得到解决, 因此本文仍用式(4)计算总体的居民收入基尼系数。

2 粒子群优化算法(PSO)及算法步骤

2.1 PSO 算法原理

粒子群优化(Particle Swarm Optimization)算法^[10]简称 PSO, 算法中以每个粒子的位置表示待优化参数的解, 初始时随机生成一个具有 n 个粒子的种群, 每个粒子的维数均为 d 。由待优化问题的目标函数确定的适应度值评价每个粒子的性能, 每次迭代都保留适应度值较优的粒子的位置。 $p_i(t)$ 是粒子 i 本身到目前为止寻到的最优解, 称为个体最优解。 $p_g(t)$ 是整个粒子群迄今搜索的最优解, 称为全局最优解。第 $t+1$ 次迭代计算, 粒子 i 根据下列规则来更新自己的速度和位置:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 (p_i(t) - x_i(t)) + c_2 r_2 (p_g(t) - x_i(t))$$
 (5)

$$x_i(t+1) = x_i(t) + r v_i(t+1)$$
 (6)

其中 ω 为惯性权重, c_1 、 c_2 为学习因子, r_1 、 r_2 为 0~1 之间的随机数, r 为约束因子。

2.2 PSO 算法基本步骤

步骤 1: 初始化粒子种群、个体最优解以及全局最优解;

步骤 2: 根据式(5)和式(6), 更新每个粒子的位置和速度;

步骤 3: 计算每个粒子的适应度值。本文取适应度函数为拟合值和实际值的均方误差(MSE, mean squared error) 即:

$$MSE = \frac{1}{d} \sum_{i=1}^d [L(p_i) - L_i]^2$$
 (7)

步骤 4: 将每个粒子的适应度值与个体最优值进行比较, 如果较优则更新当前的个体最优解;

步骤 5: 将每个粒子适应值与全局最优值进行比较, 如果较优则更新当前的全局最优解;

步骤 6: 如果达到预先设定的最大迭代次数或者目标函数的精度要求, 则停止计算; 否则返回步骤 2。

3 实例分析

表 1 农村居民人均纯收入分布情况

年份	低收入户 20%	中低收入户 20%	中等收入户 20%	中高收入户 20%	高收入户 20%
2002	1 551.790	2 288.340	3 025.170	4 075.600	7 567.220
2003	1 573.400	2 327.990	3 123.120	4 219.520	7 999.280
2004	1 778.850	2 667.290	3 535.270	4 784.950	8 889.910
2005	2 090.022	3 024.444	4 022.665	5 453.554	10 210.560
2 006	2 244.693	3 248.596	4 346.766	6 002.890	11 066.200
2007	2 554.566	3 718.387	5 042.252	6 797.532	12 926.910
2008	3 072.258	4 264.095	5 764.925	7 930.940	14 895.390
2009	3 151.620	4 431.380	6 057.320	8 487.970	16 006.510

年份	低收入户 20%	中低收入户 20%	中等收入户 20%	中高收入户 20%	高收入户 20%
2010	3 566.190	5 101.710	6 986.420	9 702.070	18 327.390
2011	4 420.540	6 168.250	8 462.410	11 685.510	22 278.150

表 2 城镇居民人均总收入分布情况

年份	最低收入户 10%	较低收入户 10%	中等偏下收入户 20%	中等收入户 20%	中等偏上收入户 20%	较高收入户 10%	最高收入户 10%
2002	2 527.68	3 833.01	5 209.18	7 061.37	9 437.99	12 555.07	20 208.43
2003	2 762.43	4 209.16	5 705.67	7 753.86	10 463.66	14 076.07	23 483.95
2004	3 084.83	4 697.62	6 423.89	8 746.65	11 870.79	16 156.02	27 506.23
2005	3 377.68	5 202.12	7 177.05	9 886.96	13 596.66	18 687.74	31 237.52
2006	3 871.37	5 946.10	8 103.73	11 052.05	15 199.70	20 699.63	34 834.39
2007	4 604.09	6 992.55	9 568.02	12 978.61	17 684.55	24 106.62	40 019.22
2008	5 203.83	7 916.53	10 974.63	15 054.73	20 784.19	28 518.85	47 422.40
2009	5 950.68	8 956.81	12 345.17	16 858.36	23 050.76	31 171.69	51 349.57
2010	6 703.70	10 247.04	13 970.99	18 920.72	25 497.81	34 254.64	56 435.17
2011	7 819.44	11 751.28	15 880.67	21 439.70	29 058.92	39 215.49	64 460.67

表 1 和表 2 是根据历年国家统计局公布的“人民生活”相关数据整理分析所得。应用 PSO 算法对 (p_i, L_i) 这组数据进行估计,使拟合误差达到最小时所确定的 $L(p)$ 方程中的参数即为所求的最优解。PSO 算法在解决不同问题时参数设定均不同,针对本文问题经过多组参数的调试,结果表明将参数设定^[11]为: $c_1 = 1.7592$, $c_2 = 1.7592$, $w = 0.725$, $r = 0.5$, 种群规模 50,最大迭代次数 100 为较佳。选

取此组参数可使算法快速收敛、粒子震荡轨迹的幅度随时间不断减小,并且能够快速搜索到全局最优解,避免了过度探索。以下以我国 2002 年农村居民人均纯收入和城镇居民人均总收入的分组数据为例对洛伦兹曲线中的参数进行估计,并计算其对应的基尼系数。

3.1 估计参数

表 3 2002 年农村居民按等级人口与收入分布情况

累计人口比重	0.2	0.4	0.6	0.8	1.0
累计收入比重实际值	0.083 8	0.207 5	0.370 9	0.591 1	1.000 0
累计收入比重估计值(LM)	0.081 3	0.208 2	0.373 1	0.589 4	1.000 0
累计收入比重估计值(PSO)	0.081 1	0.208 0	0.373 4	0.591 0	1.000 0

表 4 2002 年城镇居民按等级人口与收入分布情况

累计人口比重	0.1	0.2	0.4	0.6	0.8	0.9	1.0
累计收入比重实际值	0.041 6	0.104 6	0.190 2	0.306 3	0.461 4	0.667 8	1.000 0
累计收入比重估计值(LM)	0.037 5	0.081 7	0.185 9	0.317 2	0.498 1	0.632 7	1.000 0
累计收入比重估计值(PSO)	0.038 2	0.082 1	0.185 7	0.317 0	0.498 5	0.633 6	1.000 0

表 5 LM、PSO 算法绝对误差比较(农村)

LM 绝对误差	0.002 5	0.000 7	0.002 2	0.001 7	0.000 0
PSO 绝对误差	0.002 7	0.000 5	0.002 5	0.000 1	0.000 0

表 6 LM、PSO 算法绝对误差比较(城镇)

LM 绝对误差	0.004 1	0.022 9	0.004 3	0.010 9	0.036 7	0.035 1	0.000 0
PSO 绝对误差	0.003 4	0.022 5	0.004 5	0.010 7	0.037 1	0.034 2	0.000 0

表 3 和表 4 中的第 4、5 行数据都是通过 Levenberg-Marquardt 无约束最小二乘算法和 PSO 算法

分别估计出洛伦兹曲线方程中的参数,再根据此方程计算得到 $L(p)$ 的估计值。表 5、表 6 是两算法对

$L(p)$ 估计值的绝对误差比较。表 5 中第 2、4 组数据,表 6 中第 1、2、4、6 组数据的 PSO 比 LM 算法的绝对误差小。显然,对于 2002 年数据洛伦兹曲线方程中的参数估计,PSO 比 LM 算法较优。利用 PSO 算法的估计参数所确定的 2002 年农村拟合洛伦兹曲线方程为:

$$L_1(p) = p^{0.4820} [1 - (1 - p^{0.7362})0.6101]1.0911$$

(8)

城镇的拟合洛伦兹曲线方程为:

$$L_2(p) = p^{0.2085} [1 - (1 - p^{0.8035})0.4386]1.0722$$

(9)

再根据式(1),计算农村和城镇的基尼系数分别为 $G_1 = 0.3243$ 、 $G_2 = 0.4091$ 。由已求得农村和城镇的洛伦兹曲线和基尼系数,根据式(4)和表 8 中 2002 年相关数据计算我国当年总体的基尼系数为 $G = 0.4606$ 。

3.2 计算基尼系数

类似于计算 2002 年基尼系数的过程,估计 2003—2011 年我国农村和城镇的洛伦兹曲线中的参数及基尼系数的计算结果如表 7 和表 8。

表 7 PSO 估计 2003—2011 年农村洛伦兹曲线中的参数及基尼系数

年份	MSE10 ⁻⁶	α	γ	β	η	G_1
2003	3.442 2	0.288 3	0.926 6	0.612 9	1.071 5	0.333 1
2004	3.435 5	0.783 9	0.434 9	0.533 5	1.078 5	0.329 3
2005	3.355 7	0.386 2	0.707 9	0.606 7	1.199 1	0.327 4
2006	7.913 0	0.472 1	0.753 7	0.593 2	1.066 0	0.330 5
2007	3.635 1	0.266 8	0.937 2	0.611 8	1.076 6	0.334 0
2008	7.186 7	0.685 6	0.418 1	0.532 5	1.210 9	0.333 0
2009	13.173	0.413 5	0.827 0	0.582 7	1.042 9	0.340 0
2010	9.094 5	0.528 0	0.641 3	0.574 9	1.141 9	0.340 7
2011	8.510 1	0.290 8	0.757 6	0.612 0	1.256 1	0.339 2

表 8 PSO 估计 2003—2011 年城镇洛伦兹曲线中的参数及基尼系数

年份	MSE10 ⁻⁴	α	γ	β	η	G_2
2003	5.006 7	0.345 7	0.748 7	0.417 8	1.016 7	0.416 8
2004	4.974 8	0.541 7	0.475 3	0.377 3	1.126 2	0.438 8
2005	5.529 3	0.831 3	0.307 2	0.363 8	1.184 0	0.452 8
2006	5.213 6	0.674 1	0.356 3	0.392 0	1.243 9	0.433 7
2007	5.369 0	0.440 5	0.645 4	0.385 5	1.025 9	0.436 8
2008	5.144 2	0.398 2	0.717 0	0.392 3	1.021 6	0.444 2
2009	4.868 1	0.240 7	0.782 8	0.404 2	1.038 4	0.427 2
2010	5.200 3	0.255 9	0.750 5	0.418 6	1.052 1	0.411 1
2011	5.178 2	0.642 3	0.596 4	0.411 2	1.068 7	0.454 8

应用上述方法计算的我国 2003—2011 年居民的基尼系数如表 9 中最后一列所示,我国在 2000 年公布的基尼系数为 0.412,根据本文的估计 2009 年我国总体基尼系数是 0.5003 与 0.5 相差 0.003,2011

年的基尼系数为 0.4891,距 0.5 也非常接近。由此可知本文所估计的基尼系数的变化趋势与大多数学者的预测相符。

表 9 2003—2011 年全国居民的相关数据

年份	P_1	P_2	m_1 /元	m_2 /元	m /元	G_1	G_2	G
2002	0.609 1	0.390 9	2 475.6	8 177.40	4 704.400 0	0.324 3	0.409 1	0.460 6
2003	0.594 7	0.405 3	2 622.2	9 061.22	5 231.949 6	0.333 1	0.416 8	0.474 3
2004	0.582 4	0.417 6	2 936.40	10 128.51	5 939.825 8	0.329 3	0.438 8	0.480 2
2005	0.570 1	0.429 9	3 254.90	11 320.77	6 722.417 2	0.327 4	0.452 8	0.486 5
2006	0.556 6	0.443 4	3 587.00	12 719.19	7 596.054 2	0.330 5	0.433 7	0.487 8
2007	0.541 1	0.458 9	4 140.36	14 908.61	8 979.630 1	0.334 0	0.436 8	0.495 6

年份	P_1	P_2	$m_1/元$	$m_2/元$	$m/元$	G_1	G_2	G
2008	0.530 1	0.469 9	4 760.62	17 067.78	10 382.813 0	0.333 0	0.444 2	0.499 4
2009	0.516 6	0.483 4	5 153.17	18 858.09	11 538.339 8	0.340 0	0.427 2	0.500 3
2010	0.500 5	0.499 5	5 919.01	21 033.42	13 469.658 0	0.340 7	0.411 1	0.478 2
2011	0.487 3	0.512 7	6 977.29	23 979.20	15 694.169 0	0.339 2	0.454 8	0.489 1

4 总结

本文旨在提出一种估计洛伦兹曲线参数的新方法——粒子群优化算法。根据已有的洛伦兹曲线方程,应用粒子群优化算法对其参数进行估计,进而计算基尼系数。实例结果表明:经粒子群优化算法的估计参数所确定的洛伦兹曲线的拟合效果较好,而且由该洛伦兹曲线所估算 2002—2011 年的基尼系数,其变化趋势与多数学者的预测相符。PSO 算法在参数估计时区别于传统的参数估计方法,具有较高的精度和较快的计算速度,而且操作简单。从而为估计洛伦兹曲线中的参数提供了一种有效、可靠、快捷的计算方法。

参考文献

[1] 田卫民. 省域居民收入基尼系数测算及其变动趋势分析[J]. 经济科学,2012(2):48—59.

[2] BASMANN R L,HAYES K J,SLOTTJE D J,JOHNSON J D. A general functional form for approximating the Lorenz curve[J]. Journal of Econometrics,1990(43):77—90.

[3] SARABIA J,CASTILLO E,SLOTTJE D J. An ordered family of Lorenz curves[J]. Journal of Econometrics,1999,(91):

43—60.

[4] DUANGKAMON CHOTIKAPANICH,WILLIAM E GRIF-FITHS. Estimating Lorenz Curves Using a Dirichlet Distribution[J]. Journal of Business & Economic Statistics,2002,20(6):290—295.

[5] 张奎,王原君. Sarabia 洛伦兹曲线模型的推广[J]. 应用数学,2010(3):501—507.

[6] R M SUNDRUM. Income Distribution in Less Developed Countries[M]. New York:Chapman and Hall, 1990.

[7] WANG Z X,Y-K NG,R SMYTH. A general method for creating Lorenz curves[J]. The Review of Income and Wealth, 2011(57):563—587.

[8] SCHADER M,SCHMID F. Fitting parametric Lorenz curves to grouped income distribution-A critical note[J]. Empirical Economics,1994(19):361—370.

[9] 洪兴建,李金昌. 如何正确测算我国居民收入基尼系数[J]. 经济研究,2005(4):54.

[10] 向长城,殷礼胜. 现代智能算法理论及应用[M]. 北京:科学出版社,2009:179—208.

[11] TRELEA I C. The particle swarm optimization algorithm: convergence analysis and parameter selection[J]. Information Processing Letters,2003,85(6):317—325.

Based on Particle Swarm Optimization Algorithm to Estimate Parameters of Lorenz Curve and Calculation of the Gini Coefficient

ZHANG Zhuan¹, CHANG An-ding¹, YUE Sha-sha², WANG Xiao-chen¹

(1. College of Science; 2. School of Economics and Management,Changan University,Xi'an 710064,China)

Abstract: According to established Lorenz curve equation,this paper uses particle swarm optimization algorithm to estimate parameters of the equation; Then it calculates the Gini coefficients of Chinese urban and rural from 2002 to 2011 by the fitted Lorenz curve. After that it computes Chinese overall Gini coefficient by weighting method. The final conclusion show that particle swarm optimization algorithm can effectively estimate parameters of Lorenz curve equation; And the tendency of the Gini coefficients from 2002 to 2011 corresponds to the opinion of most scholars. So it provides a reliable and effective method for parameters estimation of Lorenz curve.

Key words: Particle swarm optimization algorithm;Lorenz curve;Gini coefficient