

基于伽马模型的中等收入定位与人口度量模型研究

徐思军, 张晓敏*, 商金耀, 王立洪

(宁波大学 理学院, 浙江 宁波 315211)

摘要: 主要运用 Gamma 分布模型刻画收入分配结构及描述洛伦兹曲线, 较好地反映了社会收入分配的公平程度. 利用随机模拟和最大似然方法求得 Gamma 分布的参数估计, 同时对收入空间法和人口空间法进行了改进, 并将改进的方法对实际数据进行处理, 得到了针对每组数据的中等收入区间.

关键词: 洛伦兹曲线; Gamma 分布; 随机模拟; 最大似然估计

中图分类号: O213.9

文献标志码: A

文章编号: 1001-5132 (2015) 03-0053-04

近年, 我国居民生活水平明显提高, 中等收入群体的社会责任被赋予了新的内涵, 其界定标准也随之变化. 现有中等收入指标体系非常广泛, 统计方法和标准有 20 多种, 不同标准之间差距较大, 得出的结果差别很大. 因此, 新形势下科学地界定中等收入群体, 合理把握划分标准, 对扩大中等收入群体比重和构建和谐社会具有重要的现实意义. 文献[1]利用经验分布拟合分组数据直接形成收入分配的近似分布, 但太依赖数据, 可推广性弱. 文献[2]讨论了“收入空间法”和“人口空间法”的缺陷, 但未给出测算中等收入人口比例大小和改进的方法. 因此, 笔者在 Gamma 分布模型基础上刻画了收入分配, 确定中等收入区间. 结合数据改进了“收入空间法”和“人口空间法”缺陷, 得到了比较好的确定中等收入的方法.

1 Gamma 分布模型

一般情况下, 我国收入分配从统计上来说服从某个分布, 而 Gamma 分布是常见的收入分配分布, 它仅用 2 个参数就可刻画多种形态的收入分配曲线. 收入分配经验表明^[1], 收入分配曲线一般正偏, 只需参数取得适当就符合 Gamma 分布的图形.

1.1 模型建立

设收入的随机变量为 x , 若它的密度函数为:

$$f(x) = \frac{\theta e^{-\theta x} (\theta x)^{\alpha-1}}{\Gamma(\alpha)}, x \geq 0, \alpha > 0, \theta > 0,$$

则为服从参数 (α, θ) 的 Gamma 分布, 记为 $\Gamma(\alpha, \theta)$.

$\Gamma(\alpha, \theta)$ 的分布函数为:

$$F(x; \alpha, \theta) = P(X \leq x) = \int_0^x \frac{\theta e^{-\theta t} (\theta t)^{\alpha-1}}{\Gamma(\alpha)} dt,$$

其数学期望与方差分别为:

$$E(x) = \alpha / \theta, \text{Var}(x) = \alpha / \theta^2.$$

由于收入均值 μ 已知, 而 $\mu = E(x) = \alpha / \theta$, 故 $\theta = \alpha / \mu$, 从而基于 Gamma 分布模型的中等收入定位与人口度量模型研究可简化为单参数模型.

当收入 x 服从 Gamma 分布时, 洛伦兹曲线函数为:

$$L(p) = \frac{1}{\mu} \int_0^x t f(t) dt = \frac{\alpha}{\mu \theta} \int_0^x \frac{\theta^{\alpha+1}}{\Gamma(\alpha+1)} t^{\alpha+1-1} e^{-\theta t} dt =$$

$$\frac{\alpha}{\mu \theta} F(x; \alpha+1, \theta) = F(F^{-1}(p; \alpha, \theta); \alpha+1, \theta),$$

$$\theta = \alpha / \mu, p = F(x; \alpha, \theta).$$

证明上式时需符合以下条件: (1)定义在 $[0, 1]$ 区间上; (2)取值于 $[0, 1]$ 区间; (3) $L(0, \tau) = 0, L(1, \tau) = 1, L'(p, \tau) \geq 0, L''(p, \tau) \leq 0$ (τ 为参数向量).

对于当收入 x , 服从 Gamma 分布的 $L(p) = F(F^{-1}(p; \alpha, \theta); \alpha+1, \theta)$. 显然满足条件(1)和(2); 当 $p=0$ 时, $x=0$, 故 $L(0; \alpha, \theta) = 0$; 当 $p=1$ 时, $x=$

收稿日期: 2014-10-08.

宁波大学学报(理工版)网址: <http://journal.lg.nbu.edu.cn/>

基金项目: 国家自然科学基金(11271210); 宁波大学研究生专业选修课程建设项目(ZDKC2013009).

第一作者: 徐思军(1989-), 男, 安徽芜湖人, 在读硕士研究生, 主要研究方向: 精算科学与金融数学. E-mail: 804487007@qq.com

*通信作者: 张晓敏(1977-), 男, 湖北鄂州人, 博士/副教授, 主要研究方向: 精算科学与金融数学. E-mail: zhangxiaomin@nbu.edu.cn

$+\infty$, 故 $L(1; \alpha, \theta) = 1$, 由 $L'(p) = x / \mu$ 和 $L''(p; \alpha, \theta) = 1 / \mu f(x)$, 因此,

$$L'(p; \alpha, \theta) = \frac{\theta x}{\alpha} \geq 0,$$

$$L''(p; \alpha, \theta) = \frac{\theta}{\alpha f(x)} = \Gamma\left(\frac{1}{\theta}\right) / (\alpha(\theta x)^{\alpha-1} e^{-\theta x}) \geq 0.$$

所以, 由 Gamma 分布导出的 $L(p)$ 是满足以上 3 个条件的洛伦兹曲线。

1.2 Gamma 模型与经典模型的比较

文献[3-5]中的 2 种代表性模型使用非线性最小二乘法对表 1 进行拟合, 并用均方误差 MSE 衡量模型对实际数据拟合精度, 其中,

$$\text{MSE}(L, \hat{L}) = \frac{1}{n} \sum_{i=1}^n (L_i - \hat{L}_i)^2,$$

如果 MSE 值越小, 说明模型拟合越精确; 每个模型对应的 MSE 取值及模型的参数取值见表 1。

为直观看拟合精度, 也可分别绘制此 2 个模型及 Gamma 模型的拟合图像。

综合 MSE 和参数个数等指标来看, Gamma 模型具有良好的拟合精度, 且是常见分布, 可用来很好地刻画社会收入的洛伦兹曲线形状。

2 中等收入区间

2.1 收入空间法的改进

为克服中等收入区间取法的任意性, 利用 Gamma 模型以确定中等收入区间。利用随机模拟在 $(0, 1)$ 上均匀地产生 n 个随机数, 对任意数判断属于哪个 $[p_i, p_{j+1}]$, 在其对应的 $[x_j, x_{j+1}]$ 上产生 1 个随机数, 得到 n 个随机数, 它较好地仿真模拟了实际收入。由于考虑到特高收入人群较少, 对人群用指数分布模拟。为检验随机模拟的数据是否符合 Gamma 模型, 对所得数据进行 $K-S$ 检验, 若 P -Value 值 > 0.05 就认为模拟的数据来自某个 Gamma 分布, 这也说明 Gamma 模型的优良性。

对于模拟的数据, 设处于中间收入人口的比例是 60%, 运行程序找出 1 个最短且覆盖了 60% 人口的收入区间, 即为中等收入区间。

该优化模型的公式为:

$$\min(x_h - x_l),$$

$$\text{s.t.} \begin{cases} p_h - p_l = F(x_h) - F(x_l) = 60\%, \\ 0 < x_l < m < x_h. \end{cases}$$

为得到确定的区间, 需进行 m 次模拟, 将得到的区间求均值, 即为 1 个确定的中等收入区间。此外, 还可以利用最大似然估计或矩估计得到原参数 α 的估计值, 从而得到符合总体收入的 Gamma 分布, 再对此模型求中等收入区间, 则该区间是 1 个确定能反映大多数人收入状况的动态区间。

对 Gamma 分布参数 α , 利用矩估计 $\bar{x} = \alpha / \theta$, $\delta^2 = \alpha / \theta^2$, 得到 α 的估计值 $\hat{\alpha}$, 对于 100 次随机模拟得到的 $\hat{\alpha}$ 求平均值, 则认为这是 α 的真实值。由此, 得到原模型中的参数 $\alpha = 1.579 467$ 。

同样方法, 对于已知的 Gamma 模型求出其中等收入区间 $[997.987, 7 468.057]$ 。此区间和随机模拟求得的区间存在差异, 因为一个是由抽象的模型得到的光滑曲线, 而另一个是由仿真样本数据得到的结果, 并且人为设定了中等收入起始点, 所以差异是确实存在的。

利用最大似然的方法求模型的参数 α , 通过运行 R 软件程序, 得到 $\alpha = 1.597 656$, 中等收入区间为 $[1 057.745, 7 535.209]$ 。

2.2 人口空间法的改进

对于一个确定的收入分布, 可利用最大似然估计得到符合 Gamma 模型的参数 α , 从而得到分布函数 $F(x)$ 。则 $\frac{1}{0.4} \int_0^{0.4} F^{-1}(q) dq$ 表示收入最低的 40% 人口所拥有的收入期望, 而 $\frac{1}{0.4} \int_{0.6}^1 F^{-1}(q) dq$ 表示收入最高的 40% 人口所拥有的收入期望。令:

$$x_l = \frac{1}{0.4} \int_0^{0.4} F^{-1}(q) dq, \quad x_h = \frac{1}{0.4} \int_{0.6}^1 F^{-1}(q) dq,$$

$$p_l = F^{-1}(x_l), \quad p_h = F^{-1}(x_h),$$

则区间 $[x_l, x_h]$ 为中等收入区间, 而 x_l, x_h 对应的 p_l, p_h 值即为中等收入人口的范围。

对表 1 数据利用上述方法可得 $p_l = 0.196 935 1$, $p_h = 0.849 240 7$, 从而可确定相应中等收入区间。

表 1 洛伦兹曲线模型拟合精度对比

模型	模型表达式	约束条件	参数 α 取值	MSE
Gamma 模型	$L(p) = F(F^{-1}(p, \alpha, \theta); \alpha + 1, \theta)$	$\alpha > 0$	$\alpha = 1.657 8$	0.000 044
模型 1 Ortega(1991) ^[4]	$L(p) = [1 - (1 - p)^\alpha]^{1/\beta}$	$\alpha > 0, 0 < \beta \leq 1$	$\alpha = 0.781 9, \beta = 0.567 2$	0.000 018
模型 2 Wang (2011) ^[4]	$L(p) = p(\alpha + (1 - \alpha)p)$	$0 \leq \alpha \leq 1$	$\alpha = 0.000 0$	0.002 170

2.3 两极分化指标的定义

设有收入分配是 $[0, 4]$ 和 $[1, 3]$ 上的均匀分布,这时的中位收入分别是 $m = 2$ 和 $m = 2$.

为比较不同收入分配,需做单位变换,即将收入分配区间均除以中位收入.则 $[0, 4]$ 上的收入分布变为 $[0, 2]$ 上的均匀分布,而图1中的 S_1 和 S_2 分别表示高低收入人口比例.同理 $[1, 3]$ 变为 $[0.5, 1.5]$ 上的均匀分布.

从图1和图2可见,这2个分布的中等收入人口比例是相等的,且均为60%.但收入分配显然不同.说明不进行单位化,直接通过数字分析是偏离实际的.因此,用 S_1/S_2 (记为 R)的比值定义为1个刻画两极分化程度指标,显然 R 越大,两极分化越大^[2].该指标反映到图上就是图2的收入分配两极分化程度低于图1的收入分配两极分化程度.

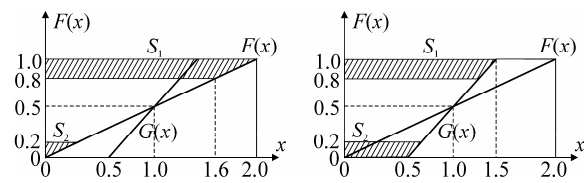


图1 收入单位化1

图2 收入单位化2

图1中 $S_1 = 0.36$, $S_2 = 0.04$, 则 $R = 9$; 图2中 $S_1 = 0.28$, $S_2 = 0.12$, 则 $R = 2.33$. 针对Gamma模型, 该两极分化比值 R 为:

$$R = S_1 / S_2 = \int_{0.8}^1 F^{-1}(q) dq / \int_0^{0.2} F^{-1}(q) dq.$$

2.4 改进后确定中等收入的应用

2.4.1 改进的收入空间法的应用

针对表2~表5, 利用改进的收入空间法可求中等收入区间. 但表4和表5中的数据有些特殊, 以表4为例: 极端低收入特别少, 而由随机模拟性质可知, 模拟出来收入较低的人比真实情况多, 使得模拟数据出现偏差. 对于这种情况, 对该区间进行模拟修正, 以 $\text{Beat}(a, b)$ 分布模拟, 通过对不同参数值作对比, 知 a, b 分别取5, 1时, 比较接近该收入分配的真实情况. 记产生最后区间的随机数的分布为平移指数分布, 均值为 λ , 平移后, 以0为起点, 在随机模拟过程中考虑到均值相等, 则:

$$x_1 \cdot \frac{a}{a+b} f_1 + \sum_{i=1}^9 \frac{x_i + x_{i+1}}{2} \cdot f_i + (\lambda + x_{10}) f_{10} = \mu,$$

其中, μ 表示表2~表5中给出的平均值以确定 λ .

因此通过以上方法可得出表2的结果, 而最终

计算后得到的结果数据见表5.

表3是表2~表5给出的均值与随机模拟的随

表2 A和B地区不同年份的模拟结果

地区	理论	年份一			
		Wight	Left	Right	α
A改进	随机模拟	4 639.6	1 840.2	6 460.8	2.509 4
	伽马模拟	6 192.4	2 931.6	9 116.7	
	伽马理论	6 166.2	3 182.3	9 347.2	
B改进	随机模拟	9 417.3	8 362.9	17 739.2	4.553 6
	伽马模拟	16 167.3	10 671.3	26 806.4	
	伽马理论	15 472.5	12 177.2	27 642.1	

地区	理论	年份二			
		Wight	Left	Right	α
A改进	随机模拟	6 649.7	2 559.5	9 210.7	2.450 4
	伽马模拟	8 774.0	4 036.8	12 810.4	
	伽马理论	8 749.0	4 314.3	13 059.6	
B改进	随机模拟	13 602.8	11 038.0	24 557.7	4.540 9
	伽马模拟	20 399.0	15 822.5	36 202.1	
	伽马理论	20 164.8	16 196.4	36 353.6	

表3 各年份理论平均值与随机值

地区	年份一(mean)				
	理论值	模拟值			
A	6 281.3	6 208.8	6 367.7	6 367.8	6 114.5
B	16 938.5	17 051.8	16 953.9	17 000.4	16 832.1

地区	年份二(mean)				
	理论值	模拟值			
A	8 890.2	8 885.5	8 876.15	8 905.8	8 927.6
B	22 228.5	22 218.7	22 262.1	22 277.7	22 211.3

表4 A和B地区各年份中位数

地区	年份一	年份二
A	5 254.75	7 461.83
B	14 422.27	19 814.62

表5 A和B地区各年份中等收入范围

地区	年份一		
	Width	Left	Right
A	0.879 32	0.350 20	1.229 52
B	0.650 13	0.579 86	1.229 99
地区	年份二		
	Width	Left	Right
A	0.891 74	0.343 01	1.234 75
B	0.682 31	0.557 06	1.239 37

机值,对比后可发现,随机模拟得到的均值在均值附近波动,表明随机模拟的方法是可信的。

便于比较,统一用 Gamma 分布得到的中等收入区间作为各地区各年份的中等收入区间。由于地区状况不同,必须将其单位化,即把得到的区间的两端点数值除以中位数。

由于数据是分区间给出的,故无法确定中位数,但可认为 $P=0.05$ 处的点作为中位数,而表 4 中数据即为各地区各年份的中位数。

表 5 表明,对同地区不同年份进行对比知, A、B 两地区在年份一、二的中等收入区间长度无明显变化,那么 2 个地区的收入分配格局无明显变化。同一年份, B 地区的整体收入明显比 A 地区的高,且 B 地区的中等收入区间长度较 A 地区短,所以 B 地区的收入分配结构比 A 地区合理。

2.4.2 改进的人口空间法的应用

根据改进的人口空间法,可求 p_1 和 p_2 以及中

表 6 A 和 B 地区各年份人口空间法与改进后对比

地区	年份一(改进方法)			
	Mid	p_1	p_2	R
A	0.615 3	0.189 1	0.846 4	5.788 4
B	0.637 5	0.182 2	0.843 3	3.610 1
地区	年份二(改进方法)			
	Mid	p_1	p_2	R
A	0.614 1	0.189 5	0.846 5	5.919 6
B	0.637 4	0.182 2	0.843 4	3.616 8

等收入人口百分比 Mid 和两极分化指标 R , 并通过这些数据对表 2~表 5 进行定量定性的分析。

表 6 表明,对于同一地区而言,从年份一到二 R 值的差距不大,那么可知收入分配无明显差异,虽然从年份一到二的收入均有所上涨。但同一年份 B 地区的 R 值明显小于 A 地区,即 B 地区的分配收入构造优于 A 地区,且 B 地区的收入也明显高于 A 地区。

参考文献:

- [1] Chotikapanich D, Rao D S P, Tang K K. Estimating income inequality in China using grouped data and the generalized Beta distribution[J]. The Review of Income and Wealth, 2007, 53(1):127-147.
- [2] Foster J E, Wolfson M C. Polarization and the decline of the middle class: Canada and the U.S.[J]. Journal of Economic Inequality, 2009(8):247-273.
- [3] WANG Zuxiang, Smyth R, Ng Y K. A new ordered family of Lorenz curves with an application to measuring income inequality and poverty in rural China[J]. China Economic Review, 2009, 20:218-235.
- [4] WANG Zuxiang, Ng Y K, Smyth R. A general method for creating Lorenz curves[J]. The Review of Income and Wealth, 2011, 57(3):561-582.
- [5] WANG Zuxiang, Smyth R. A hybrid method for creating Lorenz curves with an application to measuring world income inequality[J]. [2014-05-08]. <http://www.buseco.monash.edu.au/eco/research/papers/2013/index.html>.

Middle-income Positioning and Population Metrics Based on Gamma Model

XU Si-jun, ZHANG Xiao-min*, SHANG Jin-yao, WANG Li-hong

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: In this paper, the Gamma distribution model is mainly used to characterize the income distribution structure and describe the Lorenz curve, which well reflects the fairness of the distribution of social income. The stochastic simulation and maximum likelihood estimation method are utilized to obtain Gamma distribution, and to improve both the income space method and the population space method. The presented methods are applied to process the real data and the middle-income range is obtained in accordance with each data group.

Key words: Lorenz curve; Gamma distribution; stochastic simulation; maximum likelihood estimation

(责任编辑 章践立)