

参赛密码 _____
(由组委会填写)



第十届华为杯全国研究生数学建模竞赛

题 目 中等收入定位与人口度量模型研究

摘 要:

分配公平是群众广泛关注的社会经济话题。其中中等收入人口比重是反映收入分配格局的重要指标,本文针对中等收入定位与人口度量模型研究,文中所做主要创新工作与得到的结论如下:

第一,总结已有的十种洛伦兹曲线拟合模型,并比较了它们的拟合精度,得到拟合精度最高的前三个模型为 Wang.ZX(2009)、Wang.ZX(2011)、Ortega(1991),提出的拟合模型具体表达式 5.1.1 节。提出一种有五个参数的洛伦兹曲线拟合模型 $L(p) = (1 - (1 - p)^{\beta_1})^{\alpha} (1 - \omega(1 - p)^{\beta_2})^{\gamma}$, 并证明该模型符合洛伦兹曲线拟合条件,并给出了参数范围,该模型的拟合精度与现有模型的最高精度相比高出 5 倍左右。

第二,对于问题二。首先,讨论洛伦兹曲线上的特殊点的性质,研究了均匀分布、指数分布这两种特殊的收入分配分布的洛伦兹曲线的特性,据此指出了收入空间法与人口空间法缺陷的问题所在。其次,研究了洛伦兹曲线形态与性质的关系,主要为:洛伦兹曲线的凹凸程度与基尼系数,洛伦兹曲线的不对称程度与贫富分化结构。基于以上的研究,改进后收入空间法与人口空间法,主要有以下优点:中等收入的收入区间半径与人口区间半径的选取,主要考虑了洛伦兹曲线的凹凸程度与不对称程度,这种方法克服题目中所提到的收入区间选取的随意性;同时克服题中所提出的不适合均匀分布的缺陷。

第三,根据题目中的数据,得到了 A, B 地区两年的洛伦兹曲线与基尼系数。根据问题二改进的人口空间法,计算了 A、B 地区两年的中等收入人口区间与累积收入区间,然后根据洛伦兹曲线方程与密度函数、分布函数关系,得到了 A, B 地区两年的分布函数与密度函数曲线,从而得到中等收入人口区间与收入区间。对比 A, B 两地区两年密度曲线与分布曲线,主要有如下结论:相对 B,

A 地区贫富分化程度大，中等收入人口区间窄，中等收入的收入区间小，积累财富少。从时间来看，A 地区的经济有向坏的发展趋势。B 反之。并证明了问题二改进方法的有效性。

第四，提出了一种基于洛伦兹曲线中等收入半径模型。该模型以洛伦兹曲线极点（洛伦兹曲线一阶导数等于 1 的点）为圆心，连接极点与完全不平等线折点（坐标为(1, 0)）线段长度 s 的正相关函数为半径，画圆。把圆与洛伦兹曲线的交点确定为中等收入的范围，用线段长度 s 衡量该社会经济的稳定程度，线段 s 长度越长，中等收入人口所占比重越大，社会经济越稳定。这样解决了不能同时确定收入区间与人口区间的缺陷，并赋予了明确的经济意义。在此基础上，还构造了基于曲线凹凸程度与不对称程度的评价指标，并对曲线的未来走势做出了预测。本节最后，还分析了极点的不同运动演化情况代表了不同的社会经济发展趋势。

文中所改进的中等收入定位模型，较好地解决了人口空间法与收入空间法的缺陷；所建立的基于洛伦兹曲线中等收入半径模型较好的解决了不能同时确定人口区间与收入区间的问题，为科研人员提供了一种新的思路。

最后，基于文中所做的工作与结论，为政府决策人员提供了一点有价值的决策方案与建议。

关键词：中等收入定位，洛伦兹曲线，基尼系数，贫富分化，收入分配

目录

1 问题重述	4
2 问题分析	5
2.1 全文建模流程	5
2.2 问题一	5
2.3 问题二	6
2.4 问题三	7
2.5 问题四	7
3 模型假设	7
4 符号说明	7
5 模型建立与求解	8
5.1 问题一——洛伦兹曲线拟合模型	8
5.1.1 现有洛伦兹曲线拟合模型	8
5.1.2 本文洛伦兹曲线拟合模型的提出与证明	9
5.2 问题二——改进的收入空间法和人口空间法	12
5.2.1 对洛伦兹曲线上的几个特殊点的认识	12
5.2.2 两种特殊的收入分配分布	13
5.2.3 关于洛伦兹曲线形态与性质的讨论	15
5.2.4 改进的收入空间法改进的人口空间法	17
5.3 问题三——中等收入人口的定量描述与变化趋势	21
5.3.1 中等收入人口的定量描述与变化趋势	21
5.3.2 比较两个地区的中等收入人口、收入变化情况	24
5.4 问题四——中等收入定位与人口度量模型	25
5.4.1 收入空间法与人口空间法是一对矛盾的方法	25
5.4.2 中等收入定位于人口度量模型	26
5.5 模型补充——对中等收入定位于人口度量模型的进一步讨论	29
5.5.1 解决两条相切的洛伦兹曲线的人口度量问题	29
5.5.2 关于样本分布方差问题	30
5.6 对政府决策部门的建议	30
6 全文评价与创新点	30
7 参考文献	32

1 问题重述

居民收入分配关系到广大民众的生活水平，分配公平程度是广泛关注的话题。其中中等收入人口比重是反映收入分配格局的重要指标，这一人口比重越大，意味着收入分配结构越合理，称之为“橄榄型”收入分配格局。在这种收入分配格局下，收入差距不大，社会消费旺盛，人民生活水平高，社会稳定。一般经济发达国家都具有这种分配格局。我国处于经济转型期，收入分配格局处于重要的调整期，“橄榄型”收入分配格局正处于形成阶段。因此，监控收入分配格局的变化是经济社会发展的重要课题。

一个国家的收入分配可以用统计分布表示，收入分配的密度函数为 $f(x)$ ，对应的分布函数为 $F(x)$ ，则 $p = F(x)$ 表示收入低于或等于 x 的人口比例。记收入低于或等于 x 的人口群体拥有收入占总收入的比例为 $L(p)$ ，称之为收入分配的洛伦兹曲线。

经济学界采用洛伦兹曲线模型 $L(p, \tau)$ 拟合统计年鉴上发布分组数据，作为近似的洛伦兹曲线来进行收入分配分析，从而确定相应的统计密度与分布的估计。

经济理论界考虑取收入落在中位收入 m 的一个范围内的人口为中等收入人口，视这种方法为“收入空间法”。取其中收入属于 (x_l, x_h) 中的人口为中等收入人口，这时中等收入人口比例 M 显然等于 $F(x_h) - F(x_l)$ 。由于这种方法中 x_l 与 x_h 的取法具有任意性，收入区间是变化的，使纵向比较各年的中等收入人口时出现困难^[1]。

另一种方法为“人口空间法”，即选择 $F(m) = 1/2$ 邻近的一个范围为中等收入人口，例如取范围 $p_1 = 20\%$ 到 $p_2 = 80\%$ ，当然，按定义，中等收入人口比例已经取定为 60%。再用此 60% 的人口所拥有的收入占总收入的比例来描述中等收入人口的状态，此时中等收入人口的收入范围 $[x_l, x_h]$ 容易算得^[1]。

中等收入人口的多少与两极分化(polarization)的程度有关，所谓两极分化，用密度函数表示时，例如严重右偏且厚尾，也即中间部分空洞化。两极分化与收入不平等(inequality)是不同的概念，有学者对这两个概念进行了准确阐述^[2]。文献^[2]建立了一种指数，说明两极分化的大小或严重程度，该指数扩大意味着两极分化严重了，这时表示中等收入人口缩小了。反之若该指数缩小了，则意味着中等收入人口扩大了。但该文献并没有给出测算中等收入人口比例大小的方法。

为此，需要研究中等收入定位与人口度量问题，根据题目给出的分组数据，用数学模型研究下列问题：

一、构造满足条件的洛伦兹曲线拟合模型 $L(p, \tau)$ ，使得能很好的拟合分组数据、反映经济规律。在现有参考文献中找出至少 10 种模型，与提出的模型进行比较。通过比较拟合精度的好坏，说明所建立的洛伦兹曲线拟合模型不差。

二、研究可否改进已有的收入空间法，这时需要研究确定中等收入的范围、中等收入人口的范围的科学方法，以克服中等收入区间取法的任意性；研究可否改进已有的人口空间法，例如研究在各年中 p_1 与 p_2 取不同的值时，纵向比较各年中等收入人口与收入的变动的方法。

三、利用所给 A、B 两个地区前后两个不同年份的收入分配分组数据，研究：(1)对各地区、各年份的中等收入的数量(或范围)、中等收入人口的数量或范围进行定量描述，说明中等收入人口的变化趋势；(2)比较两个地区的中等收入人口、

收入等变化情况。

四、除二题中所述方法外，提出中等收入人口的定义、原理及经济学意义，并提出与之相应的中等收入人口的测算方法、模型或指数，说明其经济学意义。要求所提出的方法应满足普遍的价值判断或价值取向，也应反映经济规律。

2 问题分析

2.1 全文建模流程

本文围绕中等收入定位与人口度量这一问题，通过对洛伦兹曲线的讨论与研究，建立满足一定精度的洛伦兹曲线的拟合模型。利用洛伦兹曲线的拟合模型改进收入空间法与人口空间法，以克服收入空间法区间选取随意性问题及克服人口空间法无法解决均匀分布两极分化问题。对 A、B 两个地区的中等收入人口进行定量与变化描述，最后建立符合一定经济规律的中等人口收入模型，并建立中等收入人口的测算方法等。全文的建模流程如图 1 所示。

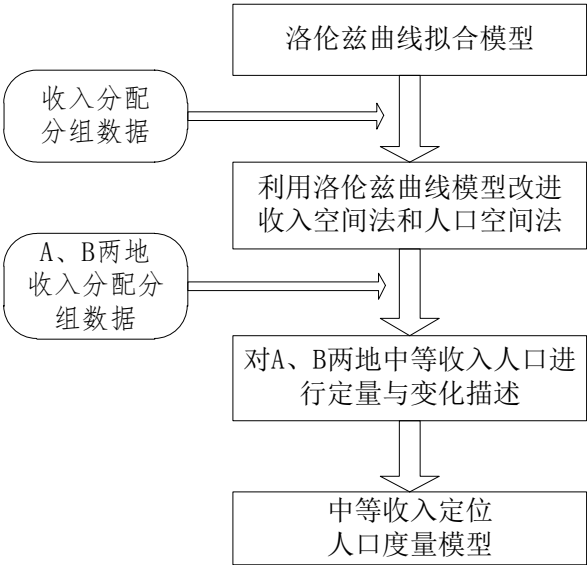


图 1 全文建模流程

2.2 问题一

根据题目的要求，问题一主要是构造洛伦兹曲线拟合模型 $L(p, \tau)$ 。 $L(p, \tau)$ 是定义在 $[0,1]$ 区间上、取值于 $[0,1]$ 区间的函数，应满足如下所给条件

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

在满足条件的情况下，确定其中参数向量 τ 的估计值 $\hat{\tau}$ ，并计算所构造的模型是否满足拟合精度条件，然后用 $L(p, \hat{\tau}) = \hat{L}(p)$ 作为近似的洛伦兹曲线来进行收入分配分析。问题二建模流程与方法如图 2 所示。

关于现有洛伦兹曲线拟合模型的研究，前人学者做了许多工作。文献 2 对此进行了详细的总结，并提出了一系列的方法洛伦兹曲线拟合模型的研究方法，尤其是两个洛伦兹曲线指数的乘积，为构造洛伦兹曲线模型提供了思路，文中还提出了拟合精度较高的洛伦兹曲线模型。本文也可据此思路构造满足条件的洛伦兹曲线模型。

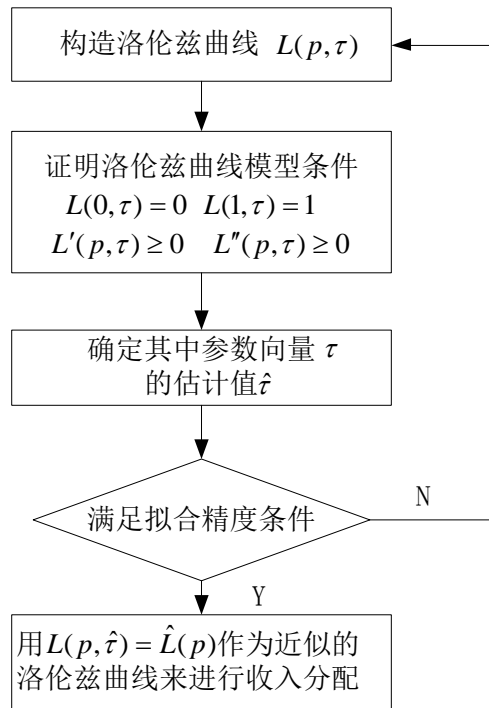


图 2 问题一建模流程与方法

2.3 问题二

问题二主要集中在中等收入人口区间的界定（有哪些人是中等收入）与中等收入区间的界定（中等收入人口的收入是多少）这两个问题。目前，确定中等收入的方法主要有如题目中的“人口空间法”与“收入空间法”。但这两种方法都有一定的缺陷：收入空间法无法克服区间选取的随意性，人口空间法无法克服两级分化对结果的影响。

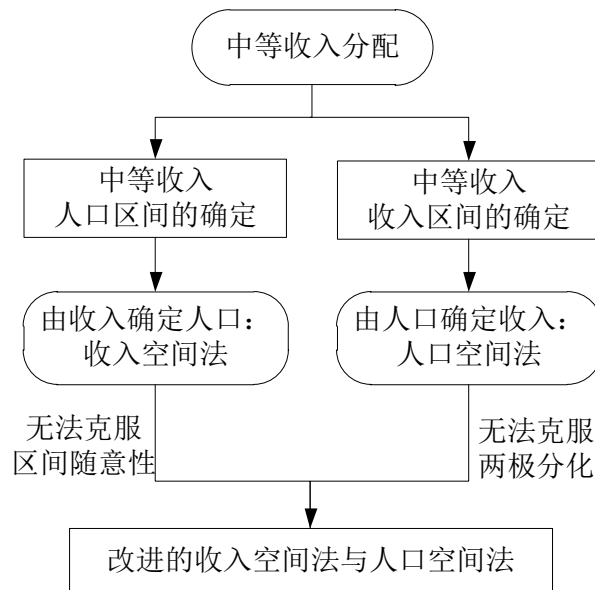


图 3 问题二建模流程

因此，本节所做的工作主要在洛伦兹曲线特性的基础上，克服人口空间法与

收入空间法的缺陷，并给出符合经济学规律的数学模型计算方法。对于区间选取的随意性这一问题，应朝着提出符合经济意义的具体特定指标放行进行；而对于人口空间法无法克服两级分化对结果的影响主要是考虑到收入分布样本方差等指标的影响。

2.4 问题三

基于问题二中建立的模型，利用数据分析 A, B 地区连续两年的中等收入区间变化情况。分析对比优化后的人口空间法与收入空间法，得出 A, B 地区中等收入区间变化的结论。

2.5 问题四

针对同一样本，人口空间法与收入空间法不能同时确定中等收入与中等收入人群。需要建立一个新的中等收入区间模型，以致于更好地定义、划分中等收入区间。然后再根据所建立的模型，分析器经济意义的发展趋势。

3 模型假设

在数学建模的过程中，在不影响模型意义与计算精度的前体下，为了使模型简单明确，并且计算建立了如下假设：

- 1) 把国民的收入分配结果分为三类：穷人、中等收入人口、富人。
- 2) 中等收入半径越大，社会越稳定。

4 符号说明

文章中所用主要符号与意义说明如表 1 所示。

表 1 符号说明

符号	名称与意义	备注
$f(x)$	收入分配的密度函数	
x	收入	
$F(x)$	分布函数	
p	收入低于或等于 x 的人口比例	$p = F(x)$
m	中位数	
μ	平均收入	
$L(p)$	洛伦兹曲线：收入低于或等于 x 的人口群体拥有收入占总收入的比例	$L(p) = \frac{1}{\mu} \int_0^x tf(t)dt \quad p = F(x)$

5 模型建立与求解

5.1 问题一——洛伦兹曲线拟合模型

5.1.1 现有洛伦兹曲线拟合模型

洛伦兹曲线作为衡量收入分配均衡程度的一种重要的图形指标,由于国家政府部门公布的收入分配分组数据,这种数据形式无法直接得到收入密度函数,而是洛伦兹曲线上的点。经济学界通常采用的洛伦兹曲线模型 $L(p, \tau)$ 拟合上述数据。因此构造较好精度的洛伦兹曲线显得尤为重要,许多学者对此作了大量的研究,如表 2 所示,列出了十种洛伦兹曲线拟合模型^{[3][4][5]}。

表 2 已有的洛伦兹曲线模型

洛伦兹拟合模型	参数	作者与时间
$L(p)_1 = (1 - (1 - p)^{\beta_1})^\alpha (1 - (1 - p)^{\beta_2})^\beta$	$\beta_1, \beta_2 \in (0, 1],$ $\alpha \geq 0, \beta \geq 0, \alpha + \beta \geq 1$	Wang.ZX (2009)
$L(p)_2 = p^\gamma - ap^\alpha (1 - p)^\beta$	文献中没有提到	Schader
$L(p)_3 = (1 - (1 - p)^\alpha)^{\frac{1}{\beta}}$	$0 < \alpha \leq 1, 0 < \beta \leq 1$	Rasche (1980)
$L(p)_4 = p^\alpha (1 - (1 - p)^\alpha)^\beta$	$0 < \alpha, 0 < \beta < 1$	Wang.ZX (2011)
$L(p)_5 = \frac{e^{\lambda p} - 1}{e^\lambda - 1}$	$0 < \lambda$	不详
$L(p)_6 = p - ap^\alpha (1 - p)^\beta$	$0 < \alpha, 0 < \beta, 0 < a$	Kakwani (1986)
$L(p)_7 = p^\alpha (1 - (1 - p)^\beta)$	$0 \leq \alpha, 0 < \beta \leq 1$	Ortega (1991)
$L(p)_8 = p^\alpha (1 - \omega(1 - p)^\beta)$	$0 < \beta \leq 1, 0 < \omega < 1, \alpha \geq 1$	Wang.ZX (2009)
$L(p)_9 = pA^{p-1}$	$0 < A$	Gupta (1984)
$L(p)_{10} = p^\delta e^{-\eta(1-p)}$	$1 < \delta < 2, 0 < \eta$	Kakwani(1973)

对于洛伦兹曲线拟合模型拟合精度的好坏常采用以下三种标准来度量。

均方误差(MSE, mean squared error):

$$\frac{1}{n} \sum_{i=1}^n [L(p_i, \hat{\tau}) - L_i]^2$$

平均绝对误差(MAE, mean absolute error):

$$\frac{1}{n} \sum_{i=1}^n |L(p_i, \hat{\tau}) - L_i|$$

最大绝对误差(MAS, maximum absolute error):

$$\max_{1 \leq i \leq n} |L(p_i, \hat{\tau}) - L_i|$$

采用上述十个洛伦兹曲线拟合模型,对题目中所给的数据进行拟合,得到了十条拟合的洛伦兹曲线,并且得到了拟合精度,如表 3 所示。

表 3 拟合精度

	<i>MSE</i>	<i>MAE</i>	<i>MAS</i>		<i>MSE</i>	<i>MAE</i>	<i>MAS</i>
L_1	1.8209e-5	0.003770	0.007969	L_6	9.5768e-6	0.007597	0.020999
L_2	9.9845e-6	0.002807	0.004580	L_7	9.3830e-6	0.002741	0.0055232
L_3	1.8209e-5	0.003770	0.007969	L_8	3.2312e-6	0.001552	0.00323
L_4	3.3735e-6	0.001560	0.003641	L_9	1.6766e-4	0.010164	0.02559
L_5	2.7383e-4	0.014086	0.03588	L_{10}	1.2762e-4	0.009263	0.022637

注：表中 L_1 、 $L_2 \cdots L_{10}$ 分别对应上述表中 10 种洛伦兹曲线拟合模型。

对上述三个拟合精度标准指标进行排序，结果如下：

MSE: $L_5 > L_9 > L_{10} > L_1 > L_3 > L_2 > L_6 > L_7 > L_4 > L_8$

MAE: $L_5 > L_9 > L_{10} > L_6 > L_1 > L_3 > L_2 > L_7 > L_4 > L_8$

MAS: $L_5 > L_9 > L_{10} > L_6 > L_1 > L_3 > L_7 > L_2 > L_4 > L_8$

从拟合精度指标的排序情况来看，洛伦兹曲线拟合精度最高的前三个为 Wang.ZX (2009) 提出的 $L_8^{[5]}$ 、Wang.ZX(2011) L_4 、Ortega(1991) L_7 ：

$$L(p)_8 = p^\alpha (1 - \omega(1-p)^\beta) \quad (2)$$

$$L(p)_4 = p^\alpha (1 - (1-p)^\alpha)^\beta \quad (3)$$

$$L(p)_7 = p^\alpha (1 - (1-p)^\beta) \quad (4)$$

5.1.2 本文洛伦兹曲线拟合模型的提出与证明

根据文献^[2]，当 $f(p), g(p)$ 为洛伦兹曲线时，两者的指数乘积在参数满足一定的条件下，均满足洛伦兹曲线模型条件，即

$$L(p) = f(p)^\alpha g(p)^\beta$$

据此，提出有五个参数的洛伦兹曲线模型，参数向量 τ

$$L(p) = (1 - (1-p)^{\beta_1})^\alpha (1 - \omega(1-p)^{\beta_2})^\gamma \quad (5)$$

定理 1: 当 $\alpha \geq 1, \gamma \geq 1, \beta_1 \in (0, 1], \beta_2 \in (0, 1], 0 < \omega < 1$ 时，(1) 式定义的 $L(p)$ 满足洛伦兹曲线的条件。若 $\alpha \geq 1, \gamma \geq 1, \beta_1 \in (0, 1], \beta_2 \in (0, 1], \omega = 1$ 时，(1) 式中式定义的 $L(p)$ 满足洛伦兹曲线的条件。

证明 1: 因为 $1 - \omega(1-p)^{\beta_2}$ 不满足洛伦兹曲线的条件，所以不能采用文献^[2]中通过组合已有洛伦兹曲线来构造新洛伦兹曲线的方法。需要通过定义，满足洛伦兹曲线拟合条件来证明定理 1。

当 $\alpha \geq 1, \gamma \geq 1, \beta_1 \in (0, 1], \beta_2 \in (0, 1], 0 < \omega < 1$ 时，

1) 当 $p = 0$ 时

$$L(0) = (1 - (1-p)^{\beta_1})^\alpha (1 - \omega(1-p)^{\beta_2})^\gamma = (1 - (1-0)^{\beta_1})^\alpha (1 - \omega(1-0)^{\beta_2})^\gamma = 0 \quad (6)$$

2) 当 $p = 1$ 时

$$L(1) = (1 - (1-1)^{\beta_1})^\alpha (1 - \omega(1-1)^{\beta_2})^\gamma = (1 - 0^{\beta_1})^\alpha (1 - 0^{\beta_2})^\gamma = 1 \quad (7)$$

3) 证明 $L'(p, \tau) \geq 0$

对 $L(p)$ 取对数，有

$$\ln L(p) = \alpha \ln(1 - (1-p)^{\beta_1}) + \gamma \ln(1 - \omega(1-p)^{\beta_2}) \quad (8)$$

对上式两边进行求导，得：

$$\frac{L'(p)}{L(p)} = \alpha\beta_1 \frac{(1-p)^{\beta_1-1}}{1-(1-p)^{\beta_1}} + \gamma\omega\beta_2 \frac{(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}} \quad (9)$$

进而得到，

$$L'(p) = [\alpha\beta_1 \frac{(1-p)^{\beta_1-1}}{1-(1-p)^{\beta_1}} + \gamma\omega\beta_2 \frac{(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}}]L(p) \quad (10)$$

令 $L'(p) = G(p)L(p)$ ，因为 $L(p) \geq 0$ ，所以要令 $L'(p, \tau) \geq 0$ ，只需 $G(p) \geq 0$ ，即：

$$\alpha\beta_1 \frac{(1-p)^{\beta_1-1}}{1-(1-p)^{\beta_1}} + \gamma\omega\beta_2 \frac{(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}} \geq 0 \quad (11)$$

易知，当 $\alpha \geq 0, \gamma \geq 0, \beta_1, \beta_2 \geq 0, \omega \in (0,1)$ 时，所以上式成立。

4) 证明 $L''(p, \tau) \geq 0$

$$\begin{aligned} L''(p) &= G'(p)L(p) + G(p)L'(p) = G'(p)L(p) + G(p)(G(p)L(p)) \\ &= [G'(p) + G^2(p)]L(p) \end{aligned} \quad (12)$$

其中，

$$G'(p) = [\alpha\beta_1 \frac{(1-p)^{\beta_1-1}}{1-(1-p)^{\beta_1}} + \gamma\omega\beta_2 \frac{(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}}]' = [x(p) + y(p)]' \quad (13)$$

因 $x(p), y(p)$ 具有相似的结构，现讨论 $y(p)$ 的求导情况。

$$y'(p) = \frac{\gamma\omega\beta_2(1-\beta_2)(1-p)^{\beta_2-2}[1-\omega(1-p)^{\beta_2}] - \gamma\omega^2\beta_2^2(1-p)^{2\beta_2-2}}{[1-\omega(1-p)^{\beta_2}]^2} \quad (14)$$

$$G^2(p) = [\alpha\beta_1 \frac{(1-p)^{\beta_1-1}}{1-(1-p)^{\beta_1}} + \gamma\omega\beta_2 \frac{(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}}]^2 \quad (15)$$

所以，当 $\gamma\omega\beta_2 \frac{(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}}$ 的平方项大于 $\frac{\gamma\omega^2\beta_2^2(1-p)^{2\beta_2-2}}{[1-\omega(1-p)^{\beta_2}]^2}$ ($\alpha\beta_1 \frac{(1-p)^{\beta_1-1}}{1-(1-p)^{\beta_1}}$ 的平方项大于 $\frac{\alpha\beta_1^2(1-p)^{2\beta_1-2}}{[1-\omega(1-p)^{\beta_1}]^2}$) 时，即能保证 $G'(p) + G^2(p) \geq 0$ ，即 $L''(p) \geq 0$ 。即：

$$\begin{aligned} &[\frac{\gamma\omega\beta_2(1-p)^{\beta_2-1}}{1-\omega(1-p)^{\beta_2}}]^2 - \frac{\gamma\omega^2\beta_2^2(1-p)^{2\beta_2-2}}{[1-\omega(1-p)^{\beta_2}]^2} \geq 0 \\ &\frac{\gamma^2\omega^2\beta_2^2(1-p)^{2\beta_2-2} - \gamma\omega^2\beta_2^2(1-p)^{2\beta_2-2}}{[1-\omega(1-p)^{\beta_2}]^2} \geq 0 \end{aligned} \quad (16)$$

于是，有：

$$\omega^2\beta_2^2(1-p)^{2\beta_2-2}(\gamma^2 - \gamma) \geq 0 \quad (17)$$

所以：

$$\gamma \geq 1 \quad (18)$$

同理， $\alpha \geq 1$ 。

所以，当 $\alpha \geq 1, \gamma \geq 1$ 时， $L''(p) \geq 0$ 。

综上所述，当 $\alpha \geq 1, \gamma \geq 1, \beta_1, \beta_2 \in (0,1], \omega \in (0,1)$ 时， $L(p)$ 满足洛伦兹曲线条件。

当 $\alpha \geq 1, \gamma \geq 1, \beta_1 \in (0,1], \beta_2 \in (0,1], \omega = 1$ 时，证明见 Wang. ZX(2009) 文献^[2]。

根据所给的数据与提出了洛伦兹曲线拟合模型，其中参数向量 τ 的估计值 $\hat{\tau}$ 得到曲线如图 4 所示。

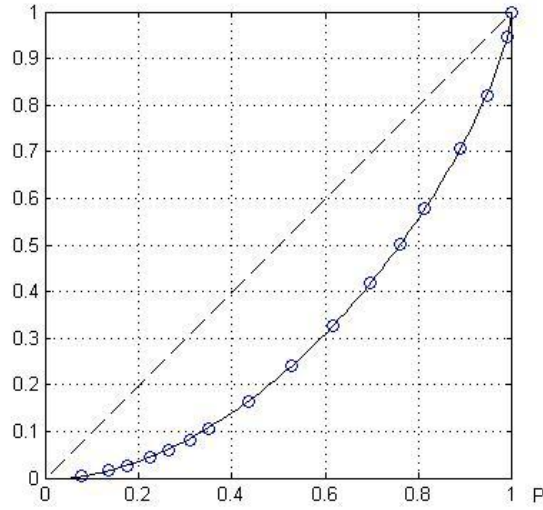


图 4 洛伦兹曲线 $L(p)$

根据题中，洛伦兹曲线模型拟合精度与上述十种模型最高拟合精度的 L_8 对比，其拟合精度指标如表 4 所示。

表 4 本文模型 L 与 Wang.ZX L_8 拟合精度

	MSE	MAE	MAS
Wang.ZX L_8	3.23E-06	0.001553	0.00323
本文模型 L	1.84526E-07	0.000388	0.000777

为了进一步度量洛伦兹曲线拟合模型的拟合精度，假设拟合残差为：

$$L(p_i, \hat{\tau}) - L_i$$

据此，得到本文的洛伦兹曲线拟合模型与 Wang.ZX L_8 拟合残差精度对比，其拟合精度指标对比如图 5、图 6 所示。

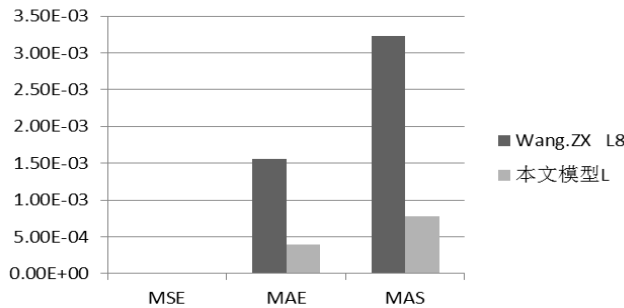


图 5 拟合精度指标对比

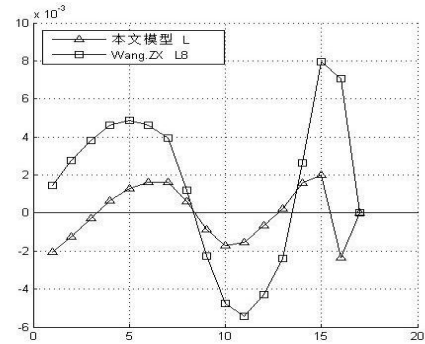


图 6 拟合残差

从上图中可以看出，本文提出的洛伦兹曲线拟合模型相对于 Wang.ZX L_8 ，各类指标均提高了 5-6 倍，有很高的精度。

综上所述，本节得到以下结论：

1. 由定理 1 的证明，得到了本文提出的洛伦兹曲线模型 L ，满足洛伦兹拟合模型条件，符合曲线基本要求；
2. 本文的洛伦兹曲线模型 L 相对于当前较为拟合精度非常好的模型 L_8 ，其拟合精度指标有较大幅度的提高，表明模型符合拟合精度条件。

5.2 问题二——改进的收入空间法和人口空间法

现有中等收入分配格局方法主要是收入空间法与人口空间法，前文已经提到了这两种方法本身就具有一定的局限性，不能从经济发展的角度去准确度量中等收入的人口区间与收入区间。

因此，对应的中等收入的人口区间与收入区间应从经济发展的角度去选取，这就需结合以下两点综合考虑：

- 1) 地区或国家的收入分配与两极分化情况；
- 2) 经济水平发展情况。

5.2.1 对洛伦兹曲线上的几个特殊点的认识

洛伦兹曲线对描述国家或地区财富分配信息提供了一种便利的图形方法，得到了广泛应用。研究中等收入分配格局，应对洛伦兹曲线有充分的认识。下面将详细说明洛伦兹曲线上的关键点：极点 N、50%财富累积点 Z、50%累积人口对应点 M。

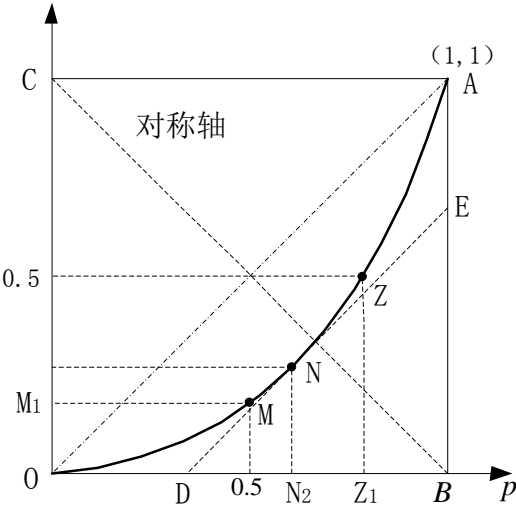


图 7 洛伦兹曲线上的特殊点

1) 极点 N

定义 1: 洛伦兹曲线方程为 $L(p)$ ，其一阶导数为 $L'(p)$ ，对于 $p \in (0, 1)$ 总存在一个 p 使得 $L'(p)=1$ ，把这个 p 对应的点称之为极点。

根据洛伦兹曲线的特性，极点 N 主要有如下性质：

- a、极点偏离完全平等线 OA 的程度表明洛伦兹曲线的弯曲程度；
- b、极点在对称轴的上方或下方，表明社会的贫富分配结构。

2) 50%财富累积点 Z

在洛伦兹曲线是坐标为 $(Z_1, 0.5)$ 的点，对于一个非完全平等的社会， Z_1 一定在横坐标 0.5 的右方，表明一定是超过 50% 的低收入人拥有 50% 的财富。

3) 50%累积人口对应点 M

在洛伦兹曲线上选取 M 点 $(0.5, M_1)$ ，M 点将人群分为相等的两部分，即 M 点为人口区间的中点（中位数）。而 M 点的纵坐标即为在收入中位数以下人群所累积的总财富。

5.2.2 两种特殊的收入分配分布

在改进现有收入空间法和人口空间法之前，先分析一下两种特殊的收入分配分布情况，一种是收入分配均匀分布，一种是收入分配指数分布。

1) 收入分配均匀分布

设某地区（或国家）收入分布在区间 $[a, b]$ 上为均匀分布，概率密度函数为：

$$f(x) = \frac{1}{b-a}, \quad x \in [a, b] \quad (19)$$

分布函数为：

$$F(x) = \frac{x-a}{b-a}, \quad x \in [a, b] \quad (20)$$

均值为：

$$\mu = \frac{a+b}{2} \quad (21)$$

洛伦兹曲线为：

$$L(p) = \frac{1}{\mu} \int_0^p F(x)^{-1} dx = \frac{2}{a+b} \int_0^p [(b-a)x+a] dx = \frac{b-a}{b+a} p^2 + \frac{2a}{b+a} p \quad (22)$$

洛伦兹曲线的图形为一条过原点的抛物线。基尼系数为：

$$G = 1 - 2 \int_0^1 L(p) dp = \frac{b-a}{3(a+b)} \quad (23)$$

接下来求解均与分布的洛伦兹曲线上的极点，求 $L(p)$ 的一阶导数：

$$L'(p) = 2 \frac{b-a}{b+a} p + \frac{2a}{b+a} \quad (24)$$

令 $L'(p)=1$ ，解得

$$p = 0.5 \quad (25)$$

结果表明，若收入分配分布是均与分布，其洛伦兹曲线的极点的横坐标一定为 0.5。其经济意义表现为，在平均财富收入值以下的人口与平均财富收入值以上的人口数量相等。

现在进一步讨论题目中所举出的不适合人口空间法的反例。当收入分配均匀分布区间分别为 $[10000, 30000]$ 和 $[0, 40000]$ 时，如按照人口空间法来计算，发现中间 60%的人口拥有的总收入仍然是 60%，这与经济直观不符合。从图 8 中来看，两个不同的样本中 20%至 80%的人口，所占有的累积财富值有很大的差距。由式(23)均匀分布基尼系数公式得到两种分布样本的基尼系数分别为 0.16665 和 0.3333，由于基尼系数表示贫富分布差距与两极分化程度的度量。

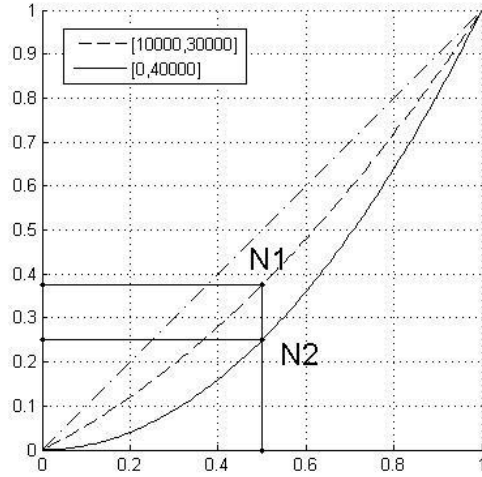


图 8 均匀分布的洛伦兹曲线

综上所述，人口空间法出现这种缺陷的主要原因与结论如下：

1. 收入均匀分布的跨度增大，尽管均匀程度没有变化（这种均匀程度不表示平等程度），但是贫富分化程度加大了，表现在基尼系数的增加。
2. 收入均匀分布的跨度增大，在数学统计上表现为方差增大，因此在考虑样本中等收入分配区间时，应综合考虑样本方差的影响。

2) 指数分布

某地区（或国家）收入分布在区间 $[a, b]$ 上为指数分布，设概率密度函数为：

$$f(x) = \lambda e^{-\lambda x}, \quad \lambda > 0 \quad (26)$$

分布函数为

$$F(x) = 1 - e^{-\lambda x}, \quad \lambda > 0 \quad (27)$$

均值为

$$\mu = \frac{1}{\lambda} \quad (28)$$

洛伦兹曲线为：

$$L(p) = \frac{1}{\mu} \int_0^p F(x)^{-1} dx = \lambda \int_0^p \frac{\ln(1-x)}{-\lambda} dx = (1-p)\ln(1-p) + p \quad (29)$$

洛伦兹曲线的图形为一条过 $(1, 0)$ 点的对数曲线，这条曲线与 λ 的取值没有关系。其基尼系数为：

$$G = 1 - 2 \int_0^1 L(p) dp = 0.5 \quad (30)$$

接下来求解指数分布的洛伦兹曲线上的极点，求 $L(p)$ 的一阶导数，令 $L'(p)=1$ ，解得：

$$p = 1 - e^{-1} \quad (31)$$

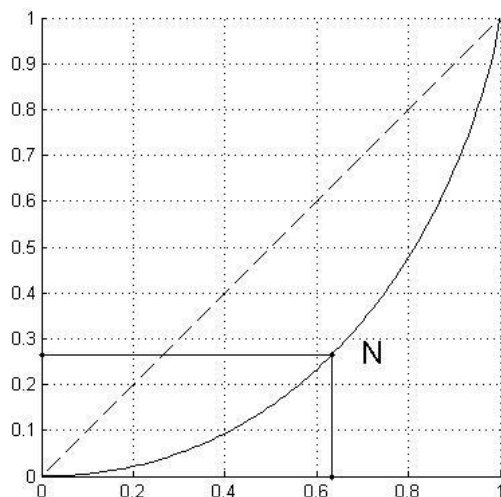


图 9 指数分布的洛伦兹曲线

5.2.3 关于洛伦兹曲线形态与性质的讨论

前面小节讨论了洛伦兹曲线上的几个特殊点，还给出了两种的特殊分布的洛伦兹曲线，并依此再进一步讨论现行人口空间法与收入空间的缺陷。

为了更加理解洛伦兹曲线，下面的洛伦兹曲线的形态与性质进行几点讨论。

1) 洛伦兹曲线的凹凸程度

基尼系数是 20 世纪初意大利学者科拉多 基尼根据劳伦茨曲线所定义的反映社会财富两极分化的指标^[6]。当基尼系数越大时，社会稳定程度越小，中等收入人群也越少。

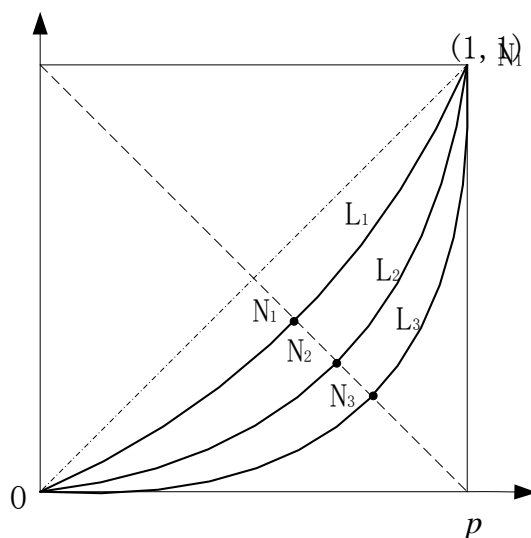


图 10 洛伦兹曲线的凹凸程度

由图 10 可知，当洛伦兹曲线的凹凸程度越大时，与完全平等曲线围成的面积越大，根据基尼系数的定义大小为 $G_3 > G_2 > G_1$ ，即社会两极分化程度逐渐减小。

2) 洛伦兹曲线的不对称程度——洛伦兹不对称系数

如图 11 所示, **Weiner** 与 **Solbrig** 发现不同分布的洛伦兹曲线可能拥有相同的基尼系数^[7]。在此基础上 **Christian Damgaard** 和 **Jacob Weiner** 提出, 洛伦兹曲线的不对称性有助于刻画两极分化的特性, 并提出了洛伦兹不对称系数^[8]。

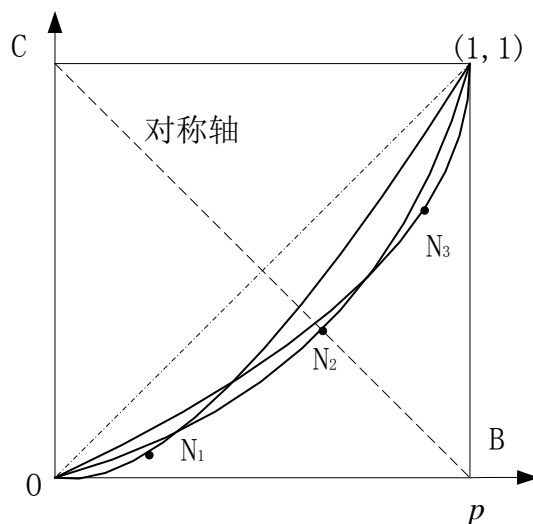


图 11 洛伦兹曲线不对称程度

由图 11 可知, N_1 , N_2 , N_3 拥有相同大小的基尼系数, 说明这三条洛伦兹曲线的凹凸程度一样。但是明显可知, 这三条洛伦兹曲线所对应的贫富分化程度是不同的。

Christian Damgaard 和 **Jacob Weiner** 证明了当 N 点偏离对称轴的程度 (不对称程度) 越大时, 社会贫富分化程度越大, 因此中等收入半径 r 也应该越小。

3) 两条交叉的洛伦兹曲线——引起贫富分化差距的一个原因

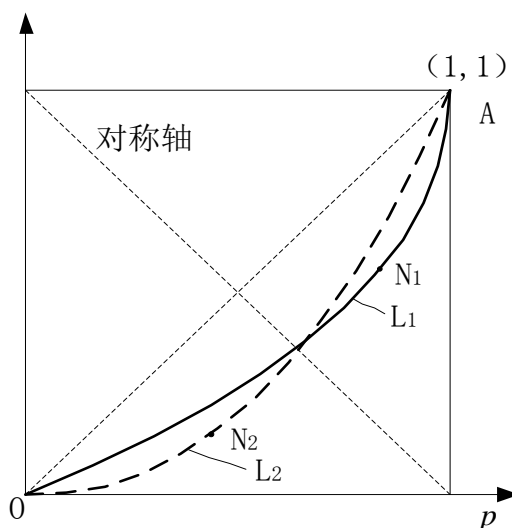


图 12 两条交叉的洛伦兹曲线

如图所示, 有两条交叉的洛伦兹曲线, 且这两种情况下基尼系数相等。对于 L_1 , 其极点 N_1 位于对称轴的上方, 表示贫富分化差距引起的主要原因是少部分的富人拥有的财富过多; 对于 L_2 , 其极点 N_2 位于对称轴的下方, 表示贫富分化差距引起的主要原因是大部分的穷人拥有的财富太少。

4) 两条相切的洛伦兹曲线

从图 1 中可以看到，两条相切的洛伦兹曲线，且其切点为极点 N 。这两种情况的基尼系数显然 $G_4 > G_3$ ，表明两种情况下的社会贫富分化程度不同，但是其洛伦兹曲线的凹凸程度与不对称程度均相同。

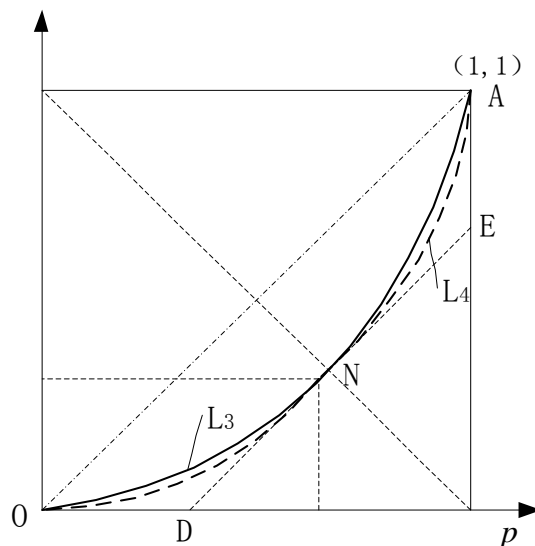


图 13 两条相切的洛伦兹曲线

5.2.4 改进的收入空间法改进的人口空间法

1) 中等收入的人口区间与收入区间

人口空间法是将 M 的横坐标作为基准点，以半径 r_p 的区域定义为中等收入人口区间，其洛伦兹曲线在纵坐标上对应的累积收入区间。收入空间法是将 M 的纵坐标作为基准点，以半径 r_i 的区域定义为中等收入收入区间。

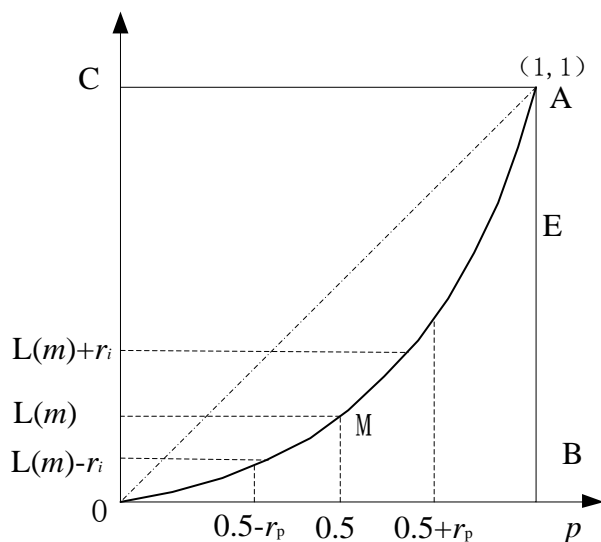


图 14 人口区间与收入区间

（注：这里的收入区间在洛伦兹曲线上表示出来时以总财富的形式，通过密度函数 $f(x)$ 与洛伦兹曲线的关系，可以得到中等收入收入区间 $[x_l, x_h]$ 。）

已有的人口空间法，实际上是定义人群中一个特定的区间内为中等收入人口区间。选取 r_p 为 0.3，即 [0.2, 0.8] 作为中等收入人口区间，即表示 20%-80% 的人群即中等收入人群。而中等收入人群所在的累计收入区间为 $[L(0.5-r_p), L(0.5+r_p)]$ 。

而已有的收入空间法，实际上是定义在某一个收入区间为中等收入，而收入区间内的人即为中等收入人群。选取的中等收入区间为 $[L_1, L_2]$ （也可以为 $[x_l, x_h]$ ，根据所给样本的数据形式选择），此区间所对应的洛伦兹曲线在横坐标上的投影即为中等收入人口区间。

人口空间法在中等收入半径 r_p 的选取上具有随意性，洛伦兹曲线的形态对中等收入人口的影响，因此对模型的结果有较大的影响。而收入空间法也有同样的缺陷。

因此，对于这两种方法的改进主要是在寻找区间半径上，并且区间半径的确定必须与洛伦兹曲线的凹凸程度与不对称程度联系起来。

2) 中等收入半径 r 的模型建立

先说明一种特殊的情况。以人口空间法为例，对于一个理想的社会（完全平等），也就是洛伦兹曲线达到了完全平等曲线时，整个社会都是中等收入，此时 r 的半径取得最大值 0.5。而对于一个极端社会（完全不平等），当洛伦兹曲线趋近于完全不平等折线时，整个社会两极分化程度最大，此时 r 的半径取得最小 0，即整个社会没有中等收入人口。

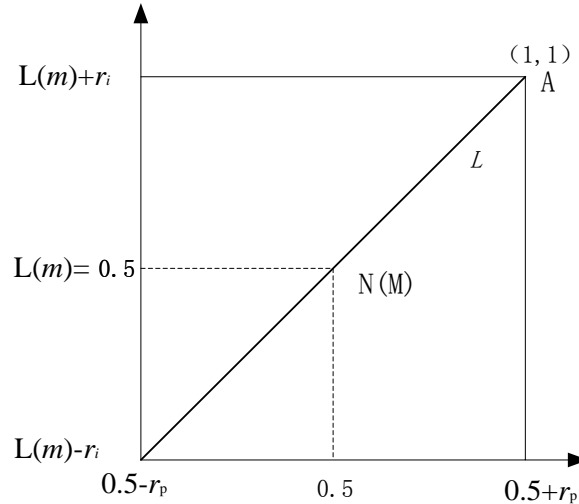


图 15 完全平等情况下的 r

对于一般的情况，考虑到收入空间的基准点在纵轴上位置具有不确定性，因此可推得 r_i 的取值范围为 $[0, \max(L(0.5), 1-L(0.5))]$ ；对于 r_p ，当且仅当洛伦兹曲线趋近于完全平等曲线时取得最大值 0.5，趋近于完全不平等线时取最小值 0。因此 r_i 与 r_p 的取值范围为：

$$\begin{cases} r_p \in [0, 0.5] \\ r_i \in [0, \max(L(0.5), 1-L(0.5))] \end{cases} \quad (32)$$

于是收入区间半径与人口区间半径 r_i 与 r_p 的都需满足条件应为：

- a、当洛伦兹曲线逐渐接近于完全平等线时， r_i 与 r_p 等于 0.5；
- b、当洛伦兹曲线逐渐接近于完全不平等线时， r_i 与 r_p 等于 0。

下面讨论一般的情况下满足上述条件的 r_i 的具体求法。

根据 5.2.3 小节的洛伦兹曲线形态与性质的讨论结论，将 r_i 与洛伦兹曲线的凹凸程度联系起来，而将 r_p 与洛伦兹曲线的不对称程度联系起来。

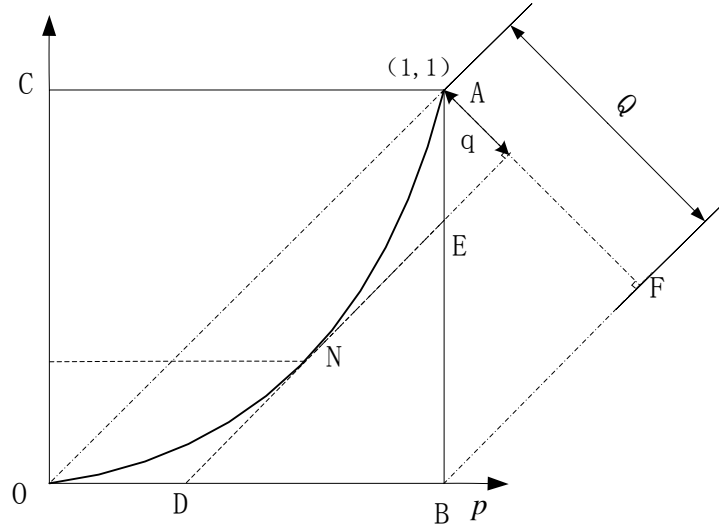


图 16 收入区间半径 r_i 与洛伦兹曲线的凹凸程度

如图 11 所示，过 A (1, 1) 点做 OA 的垂线 AF，垂足为 A；过 B (1,0) 点做 AF 的垂线 BF，垂足为 F。设 A 点与 N 点垂足的距离为 q ，AF 为 Q ，我们定义收入区间半径 r_i 为：

$$r_i = 0.5 \cdot \left(1 - \frac{q}{Q}\right) = 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \quad (33)$$

由上式可知，当洛伦兹曲线逐渐接近于完全平等线时， q 趋近于 0，此时

$$r_i = 0.5 \cdot \left(1 - \frac{0}{\sqrt{2}/2}\right) = 0.5 \quad (34)$$

当洛伦兹曲线逐渐接近于完全不平等线时， q 趋近于 $\frac{\sqrt{2}}{2}$ ，此时，

$$r_i = 0.5 \cdot \left(1 - \frac{\sqrt{2}/2}{\sqrt{2}/2}\right) = 0.5(1-1) = 0 \quad (35)$$

因此可知，公式(33)合理地满足了中等收入半径的条件。

基尼系数是对洛伦兹曲线的凹凸程度的一种描述。而文中提出的收入区间半径 r_i 的大小对凹凸程度一种很好的描述。因此， r_i 是基尼系数的一种体现。

由于实际的洛伦兹曲线形状为凹形，所以在收入空间方法中，M 点的纵坐标会在纵轴上 0.5 左右做大幅度的变化。直接使用 r 作为中等收入半径，经过洛伦兹曲线有可能会超过坐标轴 (0, 1) 的上下限。

而随着洛伦兹曲线的凹凸程度越大，即基尼系数越大时，M 点的纵坐标偏离 0.5 的距离也就越大。

因此，利用基尼系数对收入的收入区间半径 r_i 做适当的优化，优化后的收入空间半径为：

$$r_i = \begin{cases} 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \cdot \frac{(G_{i\max} - G_i)}{G_{i\max}}, & 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \cdot \frac{(G_{i\max} - G_i)}{G_{i\max}} < 0.5 \\ 0.5 & , \quad 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \cdot \frac{(G_{i\max} - G_i)}{G_{i\max}} \geq 0.5 \end{cases} \quad (36)$$

式中， G_i 为样本基尼系数， $G_{i\max}$ 选取为同类样本中该年度基尼系数最大的值。

下面讨论一般的情况下满足上述条件的 r_p 的具体求法。

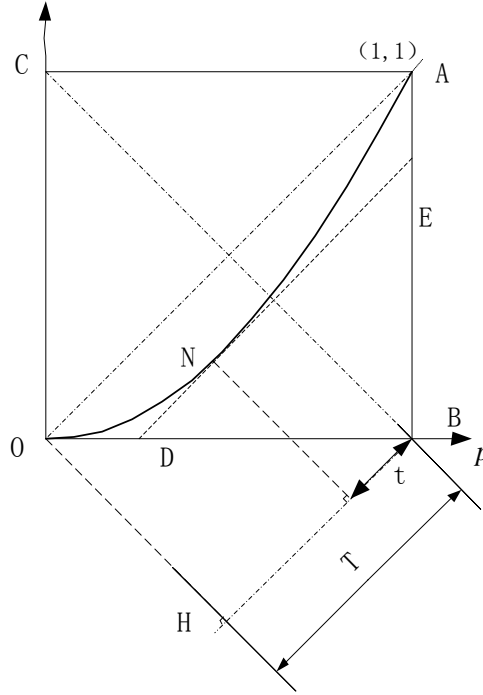


图 17 人口区间半径 r_p 与洛伦兹曲线的不对称程度

过 B 点做 OA 的平行线 BH；分别过 N 点，O 点做 BH 的垂线，设 B 点与 N 点垂足的距离为 t ，B 点与 O 点垂足的距离为 T ，则：

$$r_p = 0.5 \cdot \left(1 - \frac{t}{T}\right) = 0.5 \cdot \left(1 - \frac{t}{\sqrt{2}/2}\right) \quad (37)$$

由上式可知，当洛伦兹曲线逐渐接近于对称轴时时， q 趋近于 0，此时

$$r_p = 0.5 \cdot \left(1 - \frac{0}{\sqrt{2}/2}\right) = 0.5 \quad (38)$$

当洛伦兹曲线逐渐接近于原点 O 或者 B 点时， q 趋近于 $\frac{\sqrt{2}}{2}$ ，此时，

$$r_p = 0.5 \cdot \left(1 - \frac{\sqrt{2}/2}{\sqrt{2}/2}\right) = 0.5(1-1) = 0 \quad (39)$$

因此可知，公式(39)合理地满足了中等收入半径的特点。

3) 对收入空间法与人口空间法的改进

基于上一节的所得到的人口区间半径与收入区间半径，对收入(人口)空间方

法的改进行如下改进：

针对收入空间法，中等收入收入区间为：

$$[L(m)-r_i, L(m)+r_i] \quad (40)$$

其中，

$$r_i = \begin{cases} 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \cdot \frac{(G_{i\max} - G_i)}{G_{i\max}}, & 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \cdot \frac{(G_{i\max} - G_i)}{G_{i\max}} < 0.5 \\ 0.5, & 0.5 \cdot \left(1 - \frac{q}{\sqrt{2}/2}\right) \cdot \frac{(G_{i\max} - G_i)}{G_{i\max}} \geq 0.5 \end{cases} \quad (41)$$

因此，中等收入收入区间对应应在洛伦兹曲线上在横轴的投影为中等收入人口区间，即

$$[L^{-1}(L(m)-r), L^{-1}(L(m)+r_i)] \quad (42)$$

改进优点：中等收入收入区间半径 r_p 的选取时主要是考虑了洛伦兹曲线的凹凸程度，并且会随着经济的发展情况而发生变化。并且这种方法克服题目中所提到的收入区间选取的随意性。

针对人口空间法，中等收入人口区间为：

$$p \in [0.5 - r_p, 0.5 + r_p] \quad (43)$$

其中，

$$r_p = 0.5 \cdot \left(1 - \frac{\sqrt{2}/2}{\sqrt{2}/2}\right) = 0.5(1-1) = 0$$

因此，中等收入人口区间对应应在洛伦兹曲线上在纵轴的投影为中等收入收入区间，即

$$[L(0.5 - r_p), F(0.5 + r_p)] \quad (44)$$

改进优点：中等收入人口区间半径 r_i 的选取主要是考虑了洛伦兹曲线的走势，因而更具有经济意义，而不是随意选取的区间范围。并且这种方法克服题目中所提出的不适合均匀分布的缺陷。

4) 对收入空间法与人口空间法的改进的进一步说明

a、 本节中提到的改进方法是基于洛伦兹曲线进行的，主要是用指标描述了洛伦兹曲线的形态与走势，即凹凸程度与不对称程度。

b、 对于中等收入收入区间，方法中是用洛伦兹曲线纵坐标财富累积区间来表示，这个区间可以通过洛伦兹曲线方程与密度函数关系得到。

5.3 问题三——中等收入人口的定量描述与变化趋势

本节以 5.2 节提出的改进的人口空间法，对题目中所给出的数据进行中等人口收入的定量描述与变化趋势，并比较 A, B 两个地区中等收入人口的变化情况。

5.3.1 中等收入人口的定量描述与变化趋势

根据题目中所给的数据，首先得到了 A, B 两个地区两个年份的洛伦兹曲线，并标记出了曲线上极点 N 的位置关系。从洛伦兹曲线的图相信心中可以明

显看出 A，B 两个地区两个年份的贫富分化程度与收入分配格局的情况。

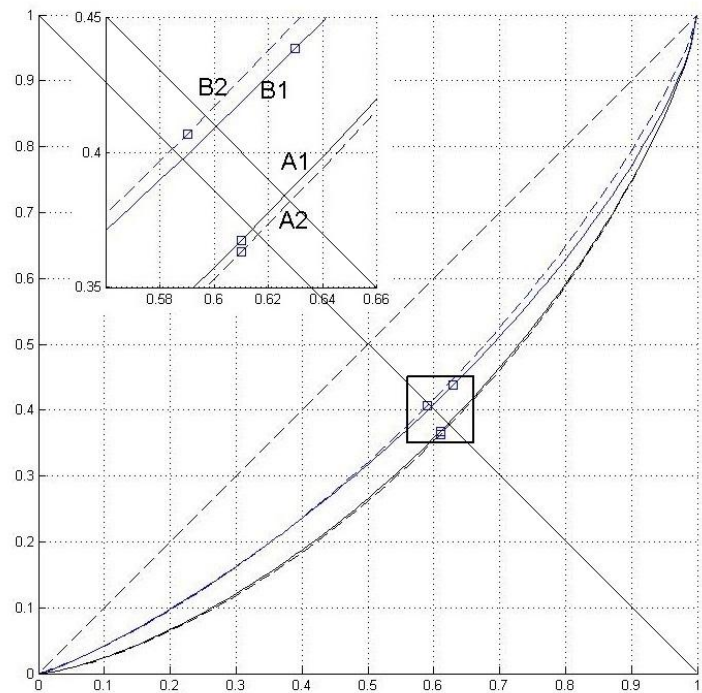


图 18 A、B 地区两个年份的洛伦兹曲线

（注：A1、A2、B1、B2 分别表示 A 地区年份一、年份二，B 地区年份一、年份二的数据）

同时得到了 A1、A2、B1、B2 四个基尼系数如表 5 所示，对比如图 19 所示。

表 5 A、B 地区两个年份基尼系数

	A1	A2	B1	B2
Gini 系数	0.3414	0.3465	0.27069	0.25581

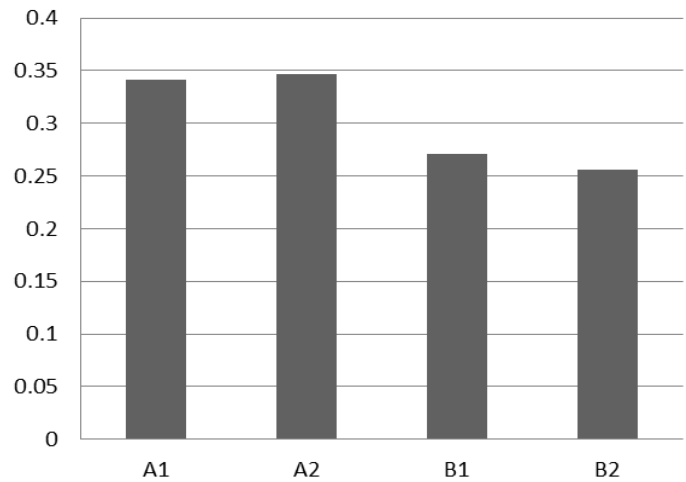


图 19 A、B 地区两个年份的基尼系数

从图 19 中可以看出，A 地区两个年份的基尼系数都相对较大（最大 0.34），并且随着年份的增长有增大的趋势；根据联合国发布的指标 A 地区可能为贫富分化程度适中的发展中国家或地区。

B 地区两个年份的基尼系数都相对较小（最大为 0.27），并且随着年份的增长有减小的趋势。根据联合国发布的指标 B 地区可能为贫富分化程度不大的发

达国家或地区。

根据改进的人口空间法，得到 A, B 两个地区两个年份的中等收入的人口区间与收入区间，其中人中范围表示人口右区间与左区间的差，表示人口跨度，累积收入表示累积收入上区间与下区间的差，表示财富累积。如表 6 所示。

表 6 A, B 两个地区两个年份的中等收入的人口区间与累积收入区间

	人口 左区间	人口 右区间	累积收入 下区间	累积收入 上区间	人口 范围	累积 收入
A1	0.1298	0.8702	0.0356	0.6973	74.04%	66.17%
A2	0.1335	0.8665	0.0358	0.689	73.30%	65.32%
B1	0.1235	0.8765	0.0544	0.7349	75.30%	68.05%
B2	0.0929	0.9071	0.0381	0.8061	81.42%	76.80%

从上述表中可以看出 A1、A2、B1、B2 的人口区间范围，累积收入区间范围，由于 A 地区的基尼系数又变大的趋势，所以 A 地区的人口范围与累积收入范围都有变小的趋势，说明经济情况在向坏的方向发展。

由于 B 地区的基尼系数又变小的趋势，所以 B 地区的人口范围与累积收入范围都有变小的趋势，说明经济情况在向好的方向发展。

根据 $L(p)$ 与 $f(x)$ 具有关系

$$L'(p) = \frac{x}{\mu} \quad (45)$$

$$f(x) = \frac{1}{\mu L''(p)} \quad (46)$$

得到了 A, B 两个地区两个年份的中等收入的数量与范围、中等收入人口的数量与范围的定量描述，其具体数值如表 7 所示，对比如图 20-23 所示。

表 7 A, B 两个地区两个年份的中等收入人口区间与收入区

	X_l	X_h	p_l	p_2
A1	2339.73	10141.91	12.98%	87.02%
A2	3354.45	14166.5	13.35%	86.65%
B1	8883.501	8883.501	12.35%	87.65%
B2	10670.87	35288.09	9.29%	90.71%

(注：表中 $[X_l, X_h]$ 表示收入区间， $[p_l, p_2]$ 人口区间)

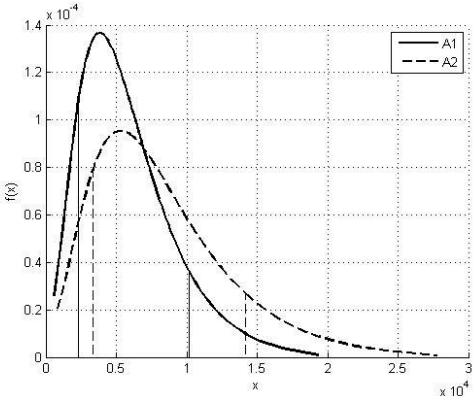


图 20 A 地区两个年份密度曲线

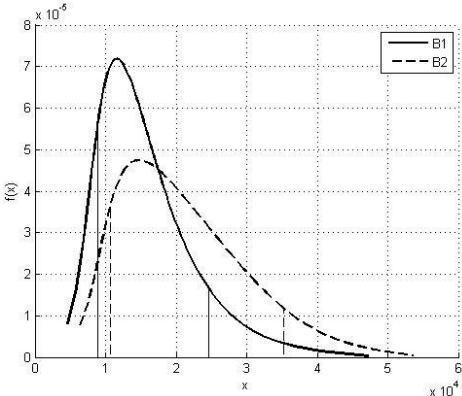


图 21 B 地区两个年份密度曲线

(注：上述两个图中的两条实线或虚线之间的范围表示中等收入收入区间，下同)

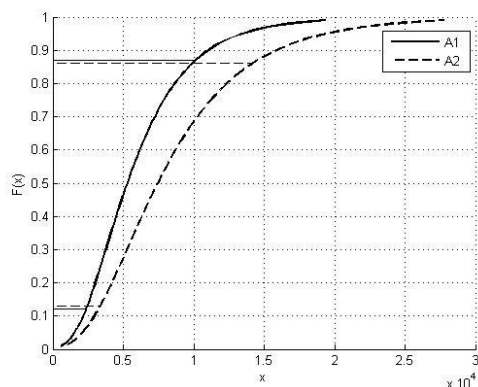


图 22 A 地区两个年份分布曲线

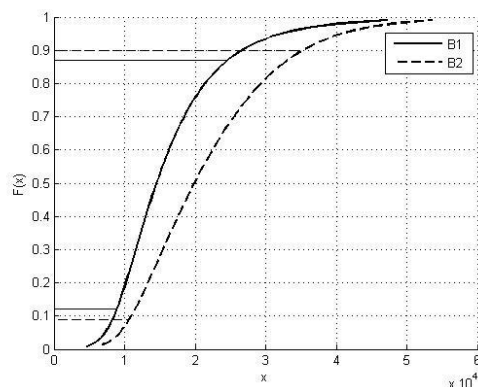


图 23 B 地区两个年份分布曲线

(注: 上述两个图中的两条实线或虚线之间的范围表示中等收入人口区间, 下同)

从表中可以看出, 对于 A 地区年份一中等收入的数量为[2339.73, 10141.91], 中等收入人口范围为[12%, 87%], 结合其基尼系数的大小, 表明其在中等收入定位在一个合理范围内。

对于 B 地区年份一中等收入的数量为[8883.501, 8883.501], 中等收入人口范围为[12.35%, 87.65%], 其收入基数较大, 可能是 B 地区经济发达。又因其基尼系数较小, 因此, 可能是 B 地区经济发达且财富分配状况优秀的地区或国家。

上述图表充分表现了 A, B 两个地区两个年份的中等收入的人口区间与收入区间及其变化趋势, 也进一步表明问题二中所提出的改进人口空间法的有效性。

5.3.2 比较两个地区的中等收入人口、收入变化情况

根据题目要求需要比较年份一、年份二 A、B 两个地区中等收入的人口区间与收入区间的变化情况。

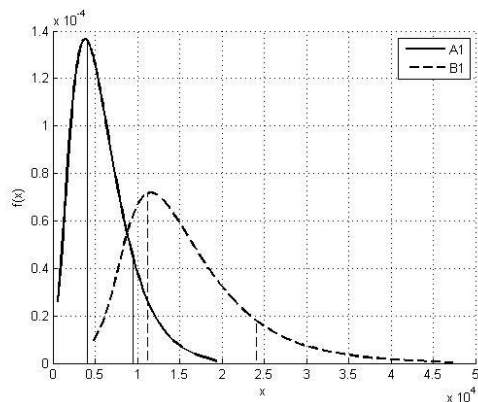


图 24 A、B 地区年份一密度曲线

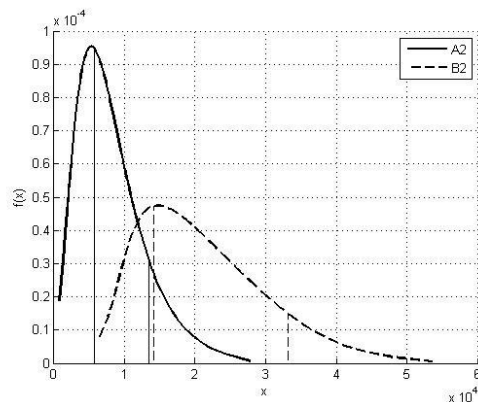


图 25 A、B 地区年份二密度曲线

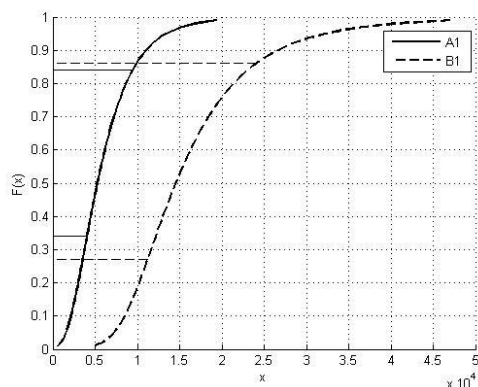


图 26 A、B 地区年份一分布曲线

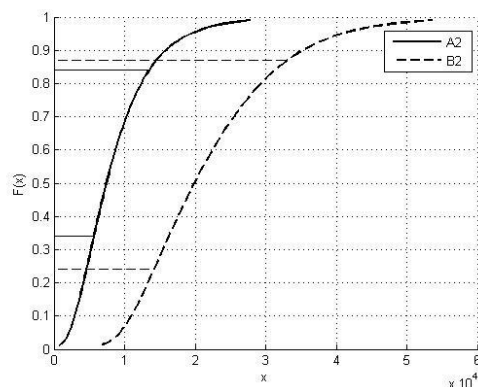


图 27 A、B 地区年份二分布曲线

从图 1 与图 2 中可以看出，对于年份一，从密度曲线来看，A 地区的图形表现为尖且拥挤，B 地区的图形表现为钝且宽松，说明 A 地区的收入分配情况较 B 地区更为集中，因此 A 地区的收入分配情况较 B 地区差。从横向来看，B 地区两条虚线所处的位置相对于 A 地区整体靠右，表明 B 地区的经济发展水平角 A 地区差。从中等收入人口来看，A 地区的人口区间与 B 地区的人口区间范围几乎相等略小，但整体收入水平较地，且掌握的社会财富较少。

5.4 问题四——中等收入定位与人口度量模型

二题中所述人口空间法与收入空间法，尽管有一些缺陷，但在目前情况来看还是倍大多数的经济学家采用，同时也题出了很多质疑[2]。改进后的人口空间法语收入空间法在一定程度上克服了原有方法的缺陷，但限于这两种方法本身的缺陷，不能做到很满意的中等收入人口定于。

本节的主要工作是提出一种能同时克服原有人口空间法与收入空间法的缺点，并希望能得有明确的经济意义。

5.4.1 收入空间法与人口空间法是一对矛盾的方法

从讨论问题二中可以知道收入空间法与人口空间法是一对矛盾的方法，因为这两种方法都不能同时解决收入区间与人口区间的问题。如图 1 所示

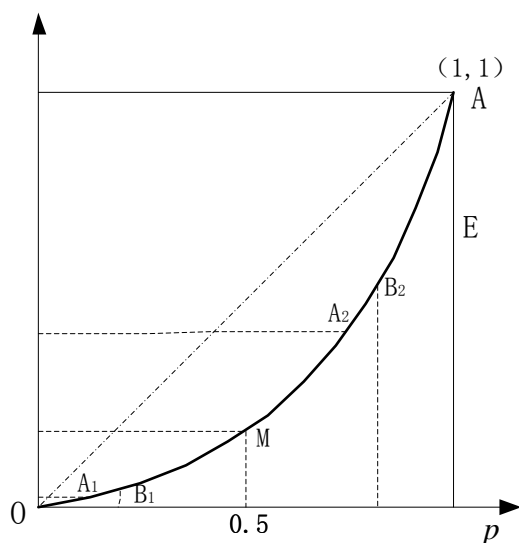


图 28 收入空间法与人口空间法确定同一样本

由图可知，对于同一样本，收入空间法与人口空间法在确定同一样本时，会得到两个不同的中等收入区间。并且，这两种方法都是通过确定一个区间，然后再根据密度函数或分布函数得到另外一个区间。

因此，我们需要解决这个矛盾问题，就是要同时得到收入区间与人口区间。

5.4.2 中等收入定位于人口度量模型

1) 几点假设与说明

由上文收入空间法与人口空间法是一对矛盾的方法，不能同时得到收入区间与人口区间，为此，我们通过极点来解决这个问题。

为了使模型更加准确与移动，我们提出以下假设：

a、中等收入半径 r 越大，社会越稳定。

中等收入群体一个社会主要的劳动力，同时也是主要的消费市场^[9]。因此，中等收入群体在社会的稳定和发展过程中扮演重要的角色^[10]。当洛伦兹曲线趋近于完全平等线时， r 逐渐增大，此时中等收入区间的人群增加，社会两级分化趋势减弱。

所以，当中等收入 r 越大时，整个社会经济越稳定。

b、点 N 与点 $B(1,0)$ 距离 s 越大时，社会越经济不稳定。

当某个社会的洛伦兹曲线趋近于完全不平等线时，点 N 趋近于点 B 。因此，令 s 为两点之间的距离，即

$$s = \sqrt{(x_n - 1)^2 + y_n^2}$$

$$s \in [0, \frac{\sqrt{2}}{2}] \quad (47)$$

因此采用 s 来衡量该社会经济的稳定程度。

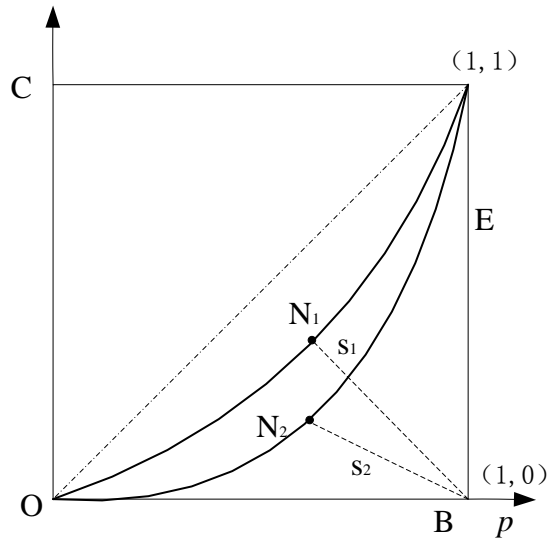


图 29 不同 s 的洛伦兹曲线

由图(29)可知，当 s 越大时，社会经济越稳定，当 s 越小时，社会经济越容易动乱。

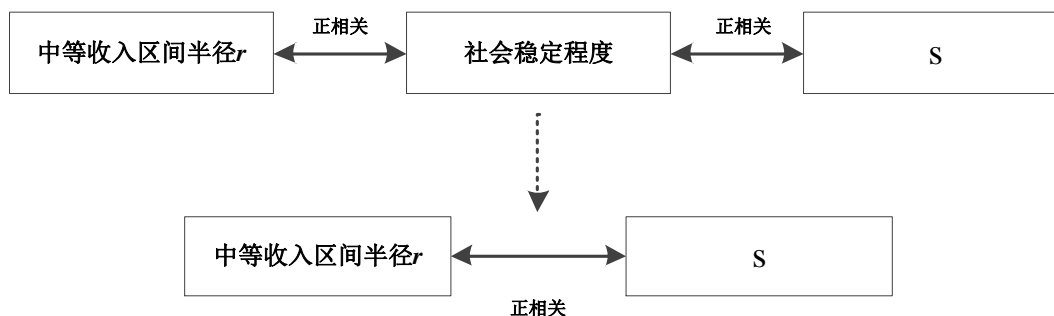


图 30 r 与 s 的关系

结合图(30)，可得到中等收入半径 r 与 s 关系：当 s 越大时， r 越大；当 s 越小时， r 越小。

2) 模型的建立

原有的人口空间法，与收入空间法都是在单独人口或者收入的条件下刻画中等收入。第二问所建立的模型，将原有的人口与收入因素，与洛伦兹曲线的特征相结合，优化了原有的人口与收入的中等收入模型。

但是，优化后的模型仍然面临如下的问题：

(1)中等收入半径 r 收到中位收入 M 点的位置影响，特别是优化后的收入空间法。

(2)人口空间法与收入空间法的结果有较大的差异，

因此，我们提出一种基于洛伦兹曲线的中等收入半径 r 的模型。

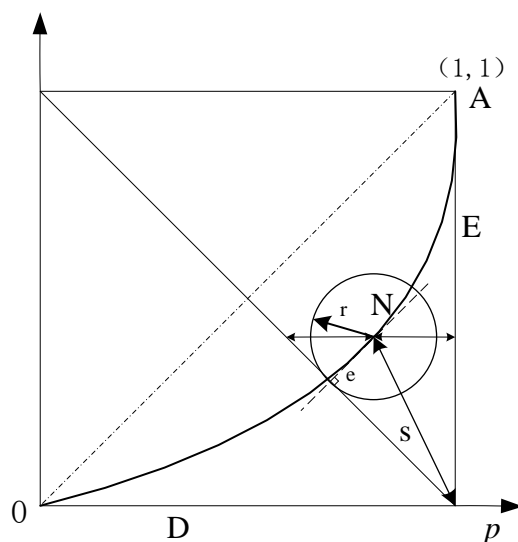


图 31 基于洛伦兹曲线的中等收入半径模型

由上节可知， r 与 s 成负相关，因此建立 r 模型如下：

$$r = j(s) \quad (48)$$

其中函数 $j(s)$ 满足以下条件：

$$\begin{aligned} j(0) &= 0; \\ j(1) &= \sqrt{2}/2; \\ j'(s) &\geq 0; \end{aligned} \quad (49)$$

以 N 为圆心，以 r 为半径做圆。与洛伦兹曲线交与 U, V 两点，将洛伦兹曲

线上的弧UV段定义为中等收入，其对应的经济意义社会经济稳定弧，显然符合直观经济规律。其在横坐标对应的区间为中等收入人口区间，其纵坐标对应的区间为中等收入收入区间。

$$\begin{cases} (p - p_N)^2 + [L(p) - L(p_N)]^2 = r^2 \\ L(p) = \frac{1}{\mu} \int_0^x t f(t) dt \end{cases} \quad (50)$$

以前文中提到的洛伦兹曲线凹凸程度，与不对称程度作为指标建立的基础。用 s 与其最大值的比作为刻画凹凸程度的指标。

由 N 点向对称轴作垂线段 e ,

$$e \in [0, \sqrt{2}/2]$$

同理，用 e 与其最大值的比作为刻画不对称程度的指标。

联合式(51)(53),得

所以指数 k 的范围为 $[0,1]$ 。且当指数 k 越大时,说明社会贫富分化程度越小,中等收入人群越大;且当指数 k 越小时,说明社会贫富分化程度越大,中等收入人群越小。

对于关于社会经济发展趋势的讨论。 N 点 N_0 趋近 N_1 过程有三种变化趋势，一种是在对称轴的上方趋近，一种是在对称轴的下方趋近，一种是在对称轴的上、下方交替趋近。其中两种情况如图 32.33 所示。

图 32 N 点在对称轴上方趋近

图 33 N 点在对称轴下方趋近

以从完全平等状态发展到完全不平等状态，即 N 点由 N_0 趋近 N_1 过程为例。依据 N 点的轨迹曲线在对称轴上下为依据，该过程可以分为两类，如图 32,33 所示。

当 N 点轨迹在对称轴上方时，说明该社会的贫富分化的主要原因是富人减少^[11]。此时，达到平均收入的人增多，中等收入的人群增加。因此中等收入的区间整体向高收入一方偏移。当 N 点轨迹在对称轴下方时，说明该社会的贫富分化的主要原因是贫穷人口增加。此时，达到平均收入的人减少，中等收入人群减少。因此中等收入区间整体向低收入一方偏移。

通过观察 N 点的位置，可以得出现阶段社会贫富分化的原因人群，能够为政府宏观调控提供参考。而根据以往 N 点的轨迹，可以对下一步社会的发展进行一定的预测。

5.5 模型补充——对中等收入定位于人口度量模型的进一步讨论

文中所改进的中等收入定位模型，较好地解决了人口空间法与收入空间法的缺陷；所建立的基于洛伦兹曲线中等收入半径模型较好的解决了不能同时确定人口区间与收入区间的问题。

但是，所建立的模型，还是有一些问题与缺陷，这里拿出进一步讨论。

5.5.1 解决两条相切的洛伦兹曲线的人口度量问题

对于两条相切且切点为位于极点的问题，问题二所建立的模型不能很好的解决。因为，此时极点 N 在两个方向的投影距离相等。为此，对相切模型做进一步补充：

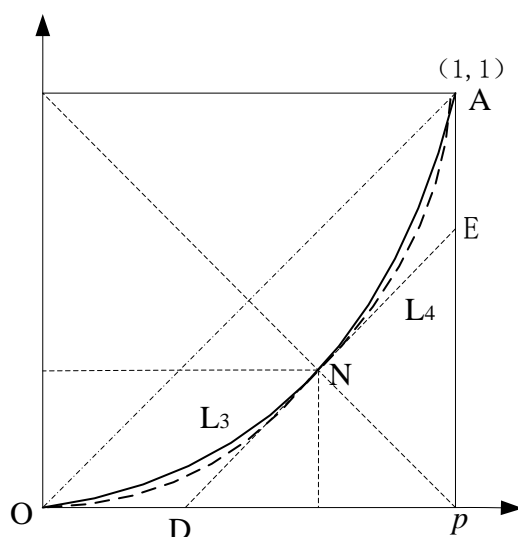


图 34 两条相交的洛伦兹曲线

由上图可知， L_4 斜率变化程度小于 L_3 ，且曲线整体更接近与 N 点的切线。因此，可知 L_4 所具有的中等收入区间应大于 L_3 。

本文次用采用洛伦兹曲线与完全平等线围成的面积 ONA 与梯形面积 $ODEA$ 之比作为衡量洛伦兹曲线整体与 N 点切线的接近程度。

而洛伦兹曲线与完全平等曲线所围成的面积为：

$$S_{ONA} = 0.5 - \int_0^1 pL(p) dp \quad (55)$$

洛伦兹曲线与 N 点切线，坐标轴所围成的面积为：

$$S_{ODEA} = 0.5(2 + L(x_N) - x_N)(x_N - L(x_N)) \quad (56)$$

可得：

$$k_3 = \frac{S_{ONA}}{S_{ODEA}} \quad (57)$$

将 k_3 定义为饱和度，反应的是洛伦兹曲线的相对饱和度。将此饱和度指标与中等收入半径联系起来，可解决两洛伦兹曲线相切模型。

5.5.2 关于样本分布方差问题

考虑收入分布的统计量对收入结构的影响，方差越小说明人口收入越集中，收入格局越均匀、合理。通过本文模型利用 A 地区数据计算得到第一年的中等收入区间半径为 0.37，第二年的中等收入区间半径为 0.36，第一年收入方差小，其中等收入区间略大于第二年，因此推断方差对收入格局有影响，但对其的影响程度不大。

5.6 对政府决策部门的建议

根据前文所做的研究与结论，对政府决策部门提出如下建议：

1)不仅要提高低收入者的收入水平，更重要的是提高低收入者的收入比重。提高低收入者的收入水平非常重要，可以保障低收入者的基本生活，维持社会稳定。不过，从提高中等收入群体的比重角度来看，仅仅提高低收入者的收入水平是不够的，关键还要提高低收入者的收入比重。

2)通过调节过高收入，扩大中等收入比重的方式，使洛伦兹曲线整体左移，降低社会的收入不平等程度。政府可以通过对高收入者的高税收等手段，调节过高收入者的收入比重，避免财富的过度集中，这样将改善的收入分配不平等程度。

3)对于两级分化社会，应关注低收入者的收入转移，尽可能的使高收入者的财富转移到低收入者，而不是使收入者的财富转移到高收入者。

6 全文评价与创新点

第二问模型的优缺点：

- (1) 解决了半径确定的随意性，建立一种可以得出具体半径参考值的模型。
- (2) 该模型基于洛伦兹曲线的几何特性，直观并且具有一定经济学意义。
- (3) 该模型不能对两条相切条件下的洛伦兹曲线的收入分配问题。

第四问模型的优缺点：

- (1) 将中等收入区间定义为，以 N 点为圆心，以 r 为半径的圆。具有一定的

创新性，并消除了对 M 点的依赖性。模型具有更好的适应性。

(2) 从社会经济学上，抽象地建立了中等收入区间半径 r ，与 N 点到 B 点距离 s 的关系。并以此确立了 r 的数学模型。

(2) 模型能够对给出较合理的社会当前贫富分化的原因，并对未来走势做出一定预测，能够为政府部门提供决策建议。

创新点：

1) 文中所改进的中等收入定位模型，较好地解决了人口空间法与收入空间法的缺陷；

2) 所建立的基于洛伦兹曲线中等收入半径模型较好的解决了不能同时确定人口区间与收入区间的问题。

7 参考文献

- [1] Wikipedia, Lorenz curve , https://en.wikipedia.org/wiki/Lorenz_curve . 2013.9.24.
- [2] Wang, Z., Y. Ng and R. Smyth, A General Method For Creating Lorenz Curves. *Review of Income and Wealth*, 2011. 57(3): p. 561--582
- [3] Schader, M. and F. Schmid, Fitting parametric Lorenz curves to grouped income distributions—a critical note. *Empirical Economics*, 1994. 19(3): p. 361-370
- [4] Wang, Z. and R. Smyth, A hybrid method for creating Lorenz Curves with an application to measuring world income inequality. 2013
- [5] 王祖祥, 张奎与孟勇, 中国基尼系数的估算研究. *经济评论*, 2009(3): 第 14-21 页.
- [6] Wikipedia, Gini coefficient, https://en.wikipedia.org/wiki/Gini_coefficient. 2013.9.24
- [7] Weiner, J. and O.T. Solbrig, The meaning and measurement of size hierarchies in plant populations. *Oecologia*, 1984. 61(3): p. 334-336.
- [8] Damgaard, C. and J. Weiner, Describing inequality in plant size or fecundity. *Ecology*, 2000. 81(4): p. 1139-1142
- [9] Foster, J.E. and M.C. Wolfson, Polarization and the Decline of the Middle Class: Canada and the US. *The Journal of Economic Inequality*, 2010. 8(2): p. 247--273.
- [10] Thurow, L.C.: The disappearance of the middle class. *New York Times*, F.3, February (1984)
- [11] Wikipedia, Lorenz asymmetry coefficient ,https://en.wikipedia.org/wiki/Lorenz_asymmetry_coefficient,2013.9.24