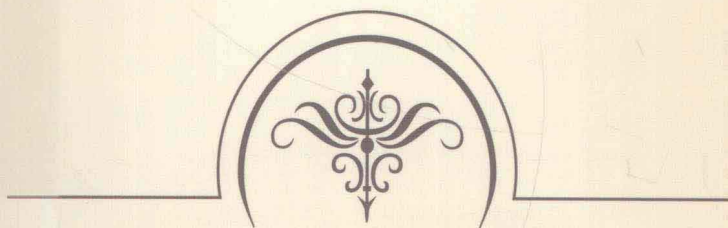


国家自然科学基金
资助项目
(70671093)

JINI XISHU YU
SHOURU FENBU YANJIU

基尼系数 与 收入分布研究



胡祖光 等 著



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

国家自然科学基金
资助项目
(70671093)

JINI XISHU YU
SHOURU FENBU YANJ

基尼系数 与 收入分布研究



胡祖光 等 著

 浙江工商大学出版社

图书在版编目(CIP)数据

基尼系数与收入分布研究/ 胡祖光等著. —杭州: 浙江
工商大学出版社, 2010. 12

ISBN 978-7-81140-218-6

I. ①基… II. ①胡… III. ①收入分配—经济理
论—研究 IV. ①F014. 4

中国版本图书馆 CIP 数据核字(2010)第 238516 号

基尼系数与收入分布研究

胡祖光 等 著

责任编辑 任晓燕
责任校对 张振华
封面设计 刘 韵
责任印制 汪 俊
出版发行 浙江工商大学出版社
(杭州市教工路 198 号 邮政编码 310012)
(E-mail: zjgsupress@163. com)
(网址: <http://www.zjgsupress.com>)
电话: 0571-88904980, 88831806(传真)

排 版 杭州兴邦电子印务有限公司
印 刷 杭州杭新印务有限公司
开 本 889mm×1194mm 1/32
印 张 6. 25
字 数 150 千字
版 次 2010 年 12 月第 1 版 2010 年 12 月第 1 次印刷
书 号 ISBN 978-7-81140-218-6
定 价 20. 00 元

版权所有 翻印必究 印装差错 负责调换

浙江工商大学出版社营销部邮购电话 0571-88804227

前 言

在中国这样一个具有“不患寡而患不均，不患贫而患不安”传统的国度里，基尼系数的理论研究和实际测算具有特别重要的意义。收入差距的不断扩大会产生严重的不良后果，特别是增加社会不安定因素，增加改革的阻力和难度，从而影响经济增长和社会发展。怎样使经济增长与发展过程更加公平，使增长成果能够更广泛地被分享，已经成为制订发展战略所关注的重点。如果以基尼系数达到0.4为警戒点，我们需要时时关注：本国或本地区的基尼系数是接近还是超过了这个警戒点。正因为如此，基尼系数不仅是理论研究的一个重要问题，也一直是各级政府关注的一个实际问题。

本书具有下面三个独到的创新之处。

创新之一是论证了基尼系数的理论最佳值，提出计算基尼系数的简易但又相当精确的

计算公式。仅仅利用统计年鉴上的城乡分列的居民收入数据，就可以计算城乡合一的基尼系数，这使我们得以及时监控基尼系数的大小变化，避免出现收入分配过于悬殊的现象。

创新之二是给出多拐点倒U型收入分配不平等曲线。传统计量模型刻画的倒U型曲线没有拐点或者仅有一个拐点，难以测度经济转型时期中国倒U型曲线存在多个拐点的情况。提出利用非参数局部多项式估计方法，测度经济增长与收入分配不平等的多拐点倒U型曲线，从而为系统认识中国的收入分配不平等发展状况提供可选择的测度方法。研究发现中国的收入分配不平等曲线存在三个拐点：经历了凹性的快速上升到凸性的减速上升，然后又到凹性的快速上升最后到凸性的减速上升“过山车”模式。

创新之三是把基尼系数融入到收入分布的研究。已有的关于收入分配不平等的理论研究，主要关注基尼系数不平等指标的测算本身，而本书从收入分布变化着手，采用核密度估计、累积分布函数、洛伦兹曲线等方法刻画我国收入分布的动态演进，并采用Kolmogorov-Smirnov非参数方法检验经验分布函数的拟合效果和基尼系数计算结果的精确度。经验研究结果表明，N.C.Kakwani经验函数和胡祖光（2004）简易公式具有较高的基尼系数估计精度，广义Logistic分布函数是一个可行的选择。进一步地，利用半参数分析框架结合反事实研究方法，考察中国收入分布函数的变化成因，讨论了家庭类型、养老金、家庭的社会属性、就业情况以及长期经济回报对于中国收入分布的影响。这种研究为改变基尼系数居高不下的现实提供了一种可选择的政策策略。

本书内容框架如下：第一章提出基尼系数的理论最佳值及简易公式，并给出实际案例；第二章为大学生消费基尼系数及其对KNN聚类的影响；第三章为混合基尼系数和收入分配格局；第四章为收入差距测算和分析；第五章提出倒U型多拐点

的非参数测度方法；第六章系统介绍收入分布的动态演进及统计检验；第七章进一步利用半参数分析方法探讨收入分布变化成因；第八章为城市收入差距扩大成因的分析。

本书是浙江工商大学胡祖光教授主持的国家自然科学基金项目“基于大规模数据调查的基尼系数理论研究”（批准号：70671093）的研究成果。参加本书写作的有浙江工商大学胡祖光教授、许冰教授、崔杰研究员、顾文涛副教授、章上峰博士、曾菊英博士以及硕士研究生杨舒、夏鹏、李熠和谭纯平等。本书中存在的不足之处，敬请读者批评指正！

胡祖光

2010年10月于浙江工商大学

目 录

第一章 基尼系数的理论最佳值及简易公式 /1

一、引言 /2

二、基尼系数理论最佳值为1/3的数学证明 /4

三、计算基尼系数的简易公式—— $g=P_5-P_1$ /9

四、两个测算案例 /14

第二章 大学生消费与KNN聚类 /25

一、背景介绍 /25

二、研究设计 /26

三、数据分析 /26

四、基于KNN聚类Gini系数对大学生消费的动态影响 /34

第三章 混合基尼系数和收入分配格局 /41

一、数据的收集及处理 /41

二、模型的拟合 /44

三、居民收入分配格局 /50

第四章 收入差距测算和分析 /55

一、收入差距的测算 /55

二、收入差距的地区分析 /67

第五章 倒U型测度方法 /74

- 一、文献回顾与问题的提出 /74
- 二、倒U测度模型与数据说明 /77
- 三、经验结果分析 /85
- 四、结论 /89

第六章 收入分布的演进及检验 /91

- 一、引言 /91
- 二、数据说明与统计描述 /93
- 三、城乡收入分布动态演进 /96
- 四、拟合优度检验 /102
- 五、基尼系数估计精度比较 /107
- 六、结论 /110

第七章 收入分布变化的半参数分析 /112

- 一、引言 /112
- 二、数据描述 /114
- 三、度量收入的不平等性 /115
- 四、结论 /129

第八章 城市收入差距扩大的成因分析 /133

- 一、引言 /134
- 二、影响收入分布的主要因素 /140
- 三、实证模型的构建 /143
- 四、中国城市间收入分布变化的实证分析 /151

参考文献 /171

第一章

基尼系数的理论最佳值及简易公式

众所周知,要计算一个国家或地区的基尼系数,必须将该国家或地区的所有居民按收入从低到高进行排列,并将排列后的所有居民均分为若干个组(至少分 5 个组),并计算出每个组的收入占该国家或地区的总收入的百分比数据,然后才可以进行基尼系数的计算。但是,在《中国统计年鉴》和各省的统计年鉴中,有关城镇和农村居民收入数据都是单列的。这样,根据《中国统计年鉴》和各省的统计年鉴中的数据,我们只能分别计算城镇居民收入分布的基尼系数和农村居民收入分布的基尼系数;而难以直接测算城乡合一的基尼系数。

本章提出的理论将改变原来根据《中国统计年鉴》和各省的统计年鉴中的数据只能分别计算城镇居民收入分布的基尼系数和农村居民收入分布的基尼系数的状况,使得利用《中国统计年鉴》和各省的统计年鉴中的数据就可以直接测算中国或各省的城乡合一的基尼系数。

本章证明以下两个结论:

第一,基尼系数的理论最佳值为 $1/3$ 。

第二,基尼系数等于收入分组五分法中最高收入组与最低收入组各自所占的收入比重之差。即:

$$g = P_5 - P_1$$

本章还根据世界银行利用收入分组五分法所公布的 34 个国家的统计数据用上述简易公式对基尼系数进行了重新计算,结果发现所计算出来的各国基尼系数与世界银行公布的基尼系数极其接近。

本章最后给出了两个测算案例:一个是我国城乡合一的基尼系数的测算,另一个是基于浙江省的统计数据的基尼系数的测算。

一、引言

国内学术界关于基尼系数的分析和研究,通常集中在以收入分配差距为中心内容的成因、走势以及与此相关的评价等方面。赵人伟等(1999)根据两次大规模的抽样调查,对 1988 年、1992 年和 1995 年的全国基尼系数包括城镇和农村的基尼系数进行了计算;李实等(1998)通过个人可支配收入与国家统计局使用的人均收入这两者的主要不同之处的比较分析,对反映我国 1995 年城镇和农村居民收入差距的基尼系数进行了新的估算,认为城乡之间的收入差距呈不同的变动轨迹;王学力(2000)曾将世界银行、国家统计局、中国社科院、中国人民大学等分别测度的基尼系数进行归纳分析,并在此基础上就我国收入差距的现状和问题展开探讨。诚然,以上分析对我国改革开放以来收入分配变动格局以及存在的问题,的确是做出了具有说服力的解释,但由于运用收入分组五分法来计算基尼系数,难以在统计资料不全的情况下准确测度城乡合一的基尼系数,^①并且计算

① 例如,依据《中国统计年鉴》的数据,我们只能以收入分组五分法来分别测度城镇或农村的基尼系数,而不能测度出城乡合一的基尼系数。从这个方面考察,国内学者有关农村或城乡合一基尼系数的测算是值得探究的。

程序相当纷繁;因此,我们有必要在测度基尼系数时探索出一种能够合理回避信息不完全并且测算起来比较简捷但又依然精确的计算方法。

据笔者所掌握的资料,有关基尼系数理论最佳值和基尼系数简易计算公式的分析文献,迄今尚不多见。比较有代表性的文献是 Cowell(1995)和 Sen(1997)针对基尼系数优良属性的深入而广泛的研究。在这些文献中,虽然他们刻画了基尼系数在描述收入分布不均等方面的机理特征,但无论是在理论论述还是在实际测度上,都未曾对基尼系数的理论最佳值和简易计算方法进行过探讨。世界银行所公布的有关各国的基尼系数,通常也是按照收入分组五分法并以传统计算方法计算得出的。

本章所论述的基尼系数的理论最佳值,是关乎收入分配在何种状态下才有利于经济运行的一个规范性概念,是从生产和消费相适应或从全社会消费品梯度更新角度来评判“理想收入分配”的一个数值界定。我们将基尼系数的理论最佳值视为一个规范性概念,一方面是因为现实经济中的收入分配客观上存在着这样的最佳值,尽管它会随着生产、消费和分配总体格局的变化而变化;另一方面则是基于不同收入阶层的消费传导而要求政府不断调整收入分配政策的考虑。这是因为,社会收入分配的终极结果会构成若干个级差显著的收入阶层,不同阶层的消费行为和消费倾向存在着明显的差异;基尼系数理论最佳值的确定可以使政府对某一时期收入分配格局有一个比较准确的基准数,从而有利于收入分配政策的宏观调控。

本章用严格的数学方法证明推导出基尼系数的简易计算公式,而且,简易计算公式的精确度几乎与精确公式相同。简易计算公式的理论价值在于它解决了计算城乡合一基尼系数的难题,使研究者不需要通过繁杂的计算就能够测度出某一时期城乡合一的收入分配的基尼系数。这对政府及时监控本国或本地区收入分配状况具有非常重要的意义。

二、基尼系数理论最佳值为 $1/3$ 的数学证明

收入分配不均等状况是市场经济的一种常态,它会影响经济增长对贫困减少的效果。事实证明,经济增长并不总是能够缩小不同收入阶层的差距。近期世界银行一份研究报告(Dollar 和 Kraay, 2001)表明,经济的整体增长有可能在促使贫困人口收入增长的同时进一步扩大不同收入阶层的差距;林伯强(2003)从低于或等于一定收入水平的狭义贫困线定义出发,通过对表征收入不均等状况的基尼系数的分析,研究了中国 1985—2001 年经济增长与贫困减少之间的关系,认为如果平均收入的提高朝着有利于贫困人口的方向发展,社会贫困就会在很大程度上减少;反之贫困不但不会减少反而会增多。实践证明,收入分布平均或悬殊都不是最佳状态。收入分布过于悬殊的弊病可以说是人所共知;至于收入分布过于平均,一般人只知道为了追求收入分布平均而牺牲效率是不好的。但我们将指出,即使是在不牺牲效率下的收入绝对平均,也不是一件好事。这可以从生产与消费相适应以及消费品更新的角度来说明。

如果一国或一个地区的收入分布平均化趋势十分明显,通常会导致一种可称之为居民消费的“排浪”现象。即当一种新的消费品问世时,你要是买不起,我也买不起;你若是买得起,我也一定买得起,不存在无支付能力的收入阶层。当该类消费品成为一种时尚或成为消费偏好的主要对象时,以追逐这类消费品为目标的“排浪”现象就会出现。此时,生产该类消费品的厂家就会频频告急;但等到“排浪”过去后,该类消费品往往又乏人问津。我国 20 世纪 80 年代就曾经出现过这种景况。显然,这是一种收入分布过于平均所导致的对社会生产有着明显不利影响的情况。另一方面,由于社会需求变化要求消费品不断更新,收入分布平均化的“排浪”现象会致使下一轮的“排浪”需求排斥上一轮的热点消费品。于是,原先被需求所刺激而生产出来的大

量消费品只能积压在仓库中,居民家中的二手货也不易处理,从而出现根植于收入分布平均化而导致的社会生产和消费的低效率的情况。

收入平均化有损于效率这一事实,实际上是要求社会的收入分布必须呈阶梯状。阶梯状的收入分布,首先,可以维系一种让失去时尚或偏好的、价格看跌的耐用消费品由较高收入阶层转入次级收入阶层的消费秩序。例如,居民月收入可分为能买得起百元级商品(手表、自行车、MP4、普通手机等)、千元级商品(照相机、VCD机、彩电、高档手机、低档手提电脑等)、万元级商品(摩托车、高档手提电脑等)、10万元级商品(汽车等)和百万元级商品(住宅、别墅等)等多种。这就避免了出现“排浪式”消费对生产的冲击;其次,当一个阶层的居民要想更新消费品时,会有下一个阶层的居民来接手。这样不仅不会产生旧货的浪费现象,还能使相对较为贫穷的阶层居民能够买到价廉物美的物品(如美国的二手汽车特别便宜,凡是到过美国的人都有深刻印象)。那么,现在要问:如此最优的消费秩序是否反映在基尼系数上存在着某一特定的最优数值规定呢?或者说某一时期的收入分布是否存在着一个理论最佳值呢?关于这一问题,回答是肯定的。

假设一个社会共有 $m+1$ 个人,其中最低者的收入为 a 元,按收入从低到高排序,后者收入比前者依次高出一个货币单位,收入排列依次分别是 $a+1, a+2, a+3, \dots, a+n$ 。根据基尼系数大小与洛伦兹曲线的相关性,在洛伦兹曲线单调递增且凸性特征明显的情况下(Kakwani, 1980),则图 1—1 成立。

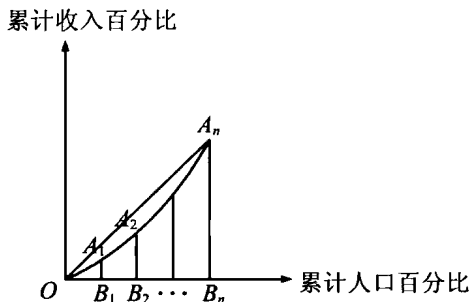


图 1—1 洛伦兹曲线和基尼系数公式

图 1—1 中, 曲线 $OA_1A_2A_n$ 是洛伦兹曲线, $OB_n = A_nB_n = 1$, 记 S_1 为拟弓形 $OA_1A_2A_n$ 的面积, S_2 为曲边三角形 $OA_1A_2A_nB_n$ 的面积, S 为三角形 OA_nB_n 的面积, 显然 $S = 0.5$, 由基尼系数的定义:

$$G = \frac{S_1}{S} = \frac{S_1}{0.5} = 2S_1 = 2(S - S_2) = 1 - 2S_2$$

为了求曲边三角形的面积 S_2 , 把 OB_n 分成 n 份, 每份长度为 $1/n$, 并从 $B_1, B_2, \dots, B_n$ 作横轴的垂线 $B_1A_1, B_2A_2, \dots, B_nA_n$, 显然这些垂线表示收入的累计百分比, 记 $P_1, P_2, \dots, P_n$, 分别表示 n 个阶层的收入在总收入中的比例, 则有:

$$A_1B_1 = P_1$$

$$A_2B_2 = P_1 + P_2$$

.....

$$A_nB_n = P_1 + P_2 + \dots + P_n = 100\%$$

这样, 曲边三角形 $OA_1A_2A_nB_n$ 被分割为高为 $1/n$ 的曲边梯形(其中第一个是小曲边三角形 OA_1B_1)之和, 其面积:

$$\begin{aligned} S_2 &= \frac{1}{2} A_1B_1 \times \frac{1}{n} + \frac{1}{2} (A_1B_1 + A_2B_2) \times \frac{1}{n} + \frac{1}{2} (A_2B_2 + A_3B_3) \times \frac{1}{n} + \dots + \frac{1}{2} (A_{n-1}B_{n-1} + A_nB_n) \times \frac{1}{n} \\ &= \frac{1}{2n} (2A_1B_1 + 2A_2B_2 + \dots + 2A_{n-1}B_{n-1} + A_nB_n) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2n} [2P_1 + 2(P_1 + P_2) + \cdots + 2(P_1 + P_2 + \cdots + P_{n-1}) + (P_1 + P_2 + \cdots + P_{n-1} + P_n)] \\
 &= \frac{1}{2n} [2(n-1)P_1 + 2(n-2)P_2 + \cdots + 2P_{n-1} + (P_1 + P_2 + \cdots + P_{n-1} + P_n)] \\
 &= \frac{1}{2n} [(2n-1)P_1 + (2n-3)P_2 + \cdots + 3P_{n-1} + P_n] \\
 \therefore G &= 1 - 2S_2 = 1 - 2 \cdot \frac{1}{2n} [(2n-1)P_1 + (2n-3)P_2 + \cdots + 3P_{n-1} + P_n] \\
 &= 1 - \frac{(2n-1)P_1 + (2n-3)P_2 + \cdots + 3P_{n-1} + P_n}{n} \\
 &= 1 - \frac{[n+(n-1)]P_1 + [n+(n-3)]P_2 + [n+(n-5)]P_3 + \cdots + [n+(3-n)]P_{n-1} + [n+(1-n)]P_n}{n} \\
 &= 1 - \frac{(nP_1 + nP_2 + \cdots + nP_n) + [(n-1)P_1 + (n-3)P_2 + (n-5)P_3 + \cdots + (3-n)P_{n-1} + (1-n)P_n]}{n} \\
 &= 1 - \frac{n \times 1 + [(n-1)P_1 + (n-3)P_2 + (n-5)P_3 + \cdots + (3-n)P_{n-1} + (1-n)P_n]}{n} \\
 &= \frac{(n-1)P_1 + (n-3)P_2 + (n-5)P_3 + \cdots + (3-n)P_{n-1} + (1-n)P_n}{n} \\
 &= \frac{(1-n)P_1 + (3-n)P_2 + (5-n)P_3 + \cdots + (n-3)P_{n-1} + (n-1)P_n}{n} \\
 &= \frac{1}{n} [(P_2 - P_1) + (P_3 - P_1) + (P_3 - P_2) + (P_4 - P_1) + (P_4 - P_2) + (P_4 - P_3) + \cdots + \\
 &\quad (P_{n-1} - P_1) + (P_{n-1} - P_2) + (P_{n-1} - P_3) + \cdots + (P_{n-1} - P_{n-2}) + (P_n - P_1) + (P_n - P_2) \\
 &\quad + (P_n - P_3) + \cdots + (P_n - P_{n-2}) + (P_n - P_{n-1})]
 \end{aligned}$$

上式就是基尼系数的计算公式,其经济意义是明确的:要计算基尼系数,就必须计算全社会任何两个人(或收入组)之间的收入比率之差,即“算尽人间不平”。

把上式用求和符号表示有:

$$\begin{aligned}
 G &= \frac{1}{n} \sum_{i=2}^n \sum_{1=j<i} (P_i - P_j) \\
 &= \frac{1}{n} \sum \sum \left(\frac{I_i}{I} - \frac{I_j}{I} \right) \\
 &= \frac{1}{nI} \sum \sum (I_i - I_j)
 \end{aligned}$$

上式中 I_i 表示第 i 个人(或组)的收入, I 则表示全社会所有成员的收入之和。当全社会 $m+1$ 个成员的收入呈严格的阶梯状时(为了使下标更便于记忆,我们假定全社会有 $n=m+1$ 成员),有:

$$I_0 = a$$

$$I_1 = a + 1$$

$$I_2 = a + 2$$

$$\dots\dots$$

$$I_m = a + m$$

$$\begin{aligned}\therefore I &= \sum I_i \\ &= (m+1)a + \frac{m(m+1)}{2} \\ &= \frac{(m+1)(m+2a)}{2}\end{aligned}$$

$$\begin{aligned}G &= \frac{1}{(m+1)I} \sum_{i=1}^m \sum_{0 \leq j < i} (I_i - I_j) \\ &= \frac{1}{(m+1)I} \{1 + (2+1) + (3+2+1) + [m + (m-1) + \dots + 1]\} \\ &= \frac{1}{(m+1)I} \left[1 + 3 + 6 + \dots + \frac{m(m+1)}{2} \right]\end{aligned}$$

上式括号内为二阶等差级数,其首项 $A_1 = 1$,各项的一阶差分为 $2, 3, 4, \dots$,二阶差分为 $1, 1, 1, \dots$,故一阶差分首项为 $R_1 = 2$,二阶差分首项为 $R_2 = 1$,由高阶等差级数前 n 项的求和公式可得:

$$\begin{aligned}G &= \frac{1}{(m+1)I} (C_m^1 A_1 + C_m^2 R_1 + C_m^3 R_2) \\ &= \frac{1}{(m+1) \cdot \frac{(m+1)(m+2a)}{2}} \left[m \times 1 + \frac{m(m-1)}{2} \times 2 + \frac{m(m-1)(m-2)}{6} \times 1 \right] \\ &= \frac{2m}{(m+1)^2(m+2a)} \cdot \left[1 + m - 1 + \frac{(m-1)(m-2)}{6} \right] \\ &= \frac{2m}{(m+1)^2(m+2a)} \cdot \frac{6m + (m-1)(m-2)}{6} \\ &= \frac{m(m+1)(m+2)}{3(m+1)^2(m+2a)} \\ &= \frac{m(m+2)}{3(m+1)(m+2a)}\end{aligned}$$

按照我们对社会成员收入由低到高序列的分析设定,由于

最穷的成员收入 a 是一个常数,因此,当社会成员数趋向于无穷时,上式中 G 值便趋向于 $1/3(0.3333)$ 。笔者以为, 0.3333 可视为基尼系数的理论最佳值。这是因为,以上严格的数学推导不仅是以社会成员收入分布呈阶梯状为基础,更重要的则是在这一数学推导结论中蕴涵着生产消费相适应以及消费品梯度更新对社会收入分配的规定性要求的思想。

三、计算基尼系数的简易公式—— $g = P_5 - P_1$

《中国统计年鉴》中城镇收入与农村收入的统计是分列的,而按照通常的收入分组五分法来计算基尼系数,就必须把全国的居民收入从高到低分成人数相等的五个组,但由于在《中国统计年鉴》中只有城镇居民的收入分组五分法数据,而没有城乡合一的收入分组五分法数据,这使我们利用通常的计算方法测算城乡合一基尼系数时遇到极大的困难。但是,我们发现,可以只根据全国最富的和最穷的 20% 人口的收入比重这两个特征值来测算城乡合一基尼系数,换句话说,基尼系数就是收入分组五分法中最高收入组与最低收入组各自所占的收入比重之差。下面给出“基尼系数就是这两个收入比重之差”的数学证明。

沿袭国内外学者利用收入分组五分法数据计算基尼系数的思路,并根据前文分析基尼系数理论最佳值的计算公式,取 $n=5$,我们有以下收入分组五分法背景下的数学推导:

$$\begin{aligned} G &= \frac{1}{n} \sum_{i=2}^5 \sum_{j<i} (P_i - P_j) \\ &= \frac{1}{5} [(P_2 - P_1) + (P_3 - P_1) + (P_3 - P_2) + (P_4 - P_1) \\ &\quad + (P_4 - P_2) + (P_4 - P_3) + (P_5 - P_1) + (P_5 - P_2) + \\ &\quad (P_5 - P_3) + (P_5 - P_4)] \end{aligned}$$

把上式中正的项与负的项分别归并,有:

$$G = \frac{1}{5} (P_2 + 2P_3 + 3P_4 + 4P_5 - 4P_1 - 3P_2 - 2P_3 - P_4)$$