

基尼系数理论最佳值及其简易计算公式研究^{*}

胡祖光

(浙江工商大学 310035)

内容提要:本文针对《中国统计年鉴》有关城镇和农村居民收入数据单列而难以直接测算城乡合一基尼系数的情况,证明了:在收入五分法下基尼系数的计算方法,可以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来简易计算。文章以严格的数学推导证明出基尼系数的理论最佳值为三分之一,并用推导出的简易公式对世界银行利用收入五分法所公布的32个国家的统计数据进行了重新计算,结果发现,以笔者的简易公式所计算出的基尼系数与世界银行公布的基尼系数极其接近。本文关于基尼系数的简易计算公式及其理论最佳值的分析,为中国城乡合一的基尼系数的测算及分析提供了一种新的思路。

关键词:基尼系数 理论最佳值 收入五分法 简易计算公式

一、引 论

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,第9、49页)根据两次大规模的抽样调查,对1988、1992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过探讨。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,的确是作出了具有说服力的解释。但由于运用收入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,并且计算程序相当纷繁;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,通常也是按照收入五分法并以传统计算方法计算得出的。

本文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

^{*} 作者感谢匿名审稿人提出的宝贵意见,但对文中可能存在的错误负责。

例如,依据《中国统计年鉴》的数据,我们只能以收入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深究的。

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优劣的基准数,从而有利于收入分配政策的宏观调控。

本文所推导的基尼系数的简易计算公式,则是笔者根据《中国统计年鉴》中城镇居民收入分配五分法统计资料和农村居民的收入分层资料,在数据不完全时测算城乡合一基尼系数的一种理论分析尝试。在《中国统计年鉴》中,由于城镇收入统计与农村收入统计是分列的,这便决定了无法直接用国家统计局公布的数据来测算城乡一体的基尼系数。从基尼系数的通行计算方法来看,人们一般是利用划定的分组收入数据以不同收入组占总收入比重而代入公式来进行测算的。显然,当统计年鉴上不列出城乡合一收入而以单独的城乡分组数据出现时,人们便不能根据通行的计算方法来测算了。鉴于此,我们考虑采用一种迂回的测算方法。据国家统计局公布的数据,中国农村人口占 60%,而其中最穷的 33%人口收入在《中国统计年鉴》上可以查到,这 33%最穷的农村人口恰好占全国人口的 20% ($60\% \times 33\% = 20\%$);同样,城镇人口占全国人口的 40%,而其中最富的 50%人口的收入在《中国统计年鉴》的城镇家计调查数据中也可以查到,这 50%的人口即全国最富的 20%人口 ($40\% \times 50\% = 20\%$)。笔者认为,目前我国最富的和最穷的 20%人口的收入比重显示了测算基尼系数的特征值。

笔者用严格的数学方法证明了:我们可以利用上述这两个特征值来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不仅在于它解决了计算城乡合一基尼系数的难题,更重要的是它使政府在较短的时间内不需要通过纷繁的计算就能够测度出某一时期收入分配的走势。

二、基尼系数理论最佳值及其数学推导

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,研究了我国 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;反之则贫困不但不减少反而会增多。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,从而出现了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,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、自行车等)、千元级商品(VCD机、彩电等)、万元级商品(摩托车、手提电脑等)、10万元级商品(汽车等)和百万元级商品(大套住宅、别墅等)等多种。这就避免了出现“排浪式”的消费对生产的冲击;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,以下的数学推导可谓是一种分析尝试。

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 $a+1, a+2, a+3, \dots, a+m$ 。根据基尼系数大小与洛仑茨曲线的相关性,在洛仑茨曲线单调递增且凸性特征明显的情景下(Kakwani, 1980),有图1成立。

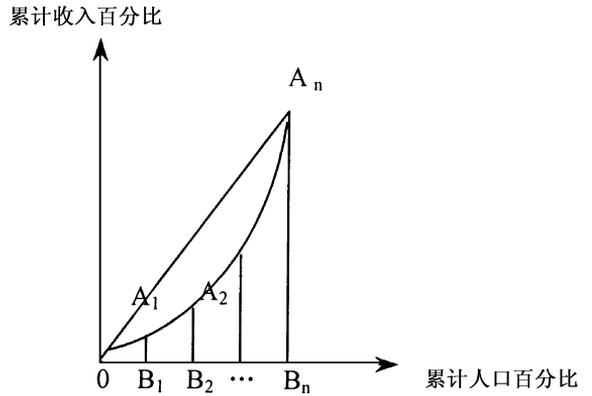


图1

图1中,曲线 $OA_1A_2A_n$ 是洛仑茨曲线, $OB_n = A_nB_n = 1$, 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 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:

$$\begin{aligned} A_1B_1 &= P_1 \\ A_2B_2 &= P_1 + P_2 \\ &\dots\dots \\ A_nB_n &= P_1 + P_2 + \dots + P_n = 100\% \end{aligned}$$

这样, 曲边三角形 $OA_1A_2A_nB_n$ 被分割成高为 $1/n$ 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:

$$\begin{aligned} S_2 &= \frac{1}{2} A_1B_1 \times \frac{1}{n} + \frac{1}{2} (A_1B_1 + A_2B_2) \times \frac{1}{n} + \frac{1}{2} (A_2B_2 + A_3B_3) \frac{1}{n} + \dots + \frac{1}{2} (A_{n-1}B_{n-1} + A_nB_n) \times \frac{1}{n} \\ &= \frac{1}{2n} (2A_1B_1 + 2A_2B_2 \dots + 2A_{n-1}B_{n-1} + A_nB_n) \\ &= \frac{1}{2n} [2P_1 + 2(P_1 + P_2) + \dots + 2(P_1 + P_2 + \dots + P_{n-1}) + (P_1 + P_2 + \dots + P_{n-1} + P_n)] \end{aligned}$$

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

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:

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

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,有(为了使下标更便于记忆,我们假定全社会有 $n = m+1$ 个成员)

$$\begin{aligned}
 I_0 &= a \\
 I_1 &= a + 1 \\
 I_2 &= a + 2 \\
 &\dots \dots \dots \\
 I_m &= a + m \\
 I &= \sum I_i = (m+1)a + \frac{m(m+1)}{2} = \frac{(m+1)(m+2a)}{2} \\
 G &= \frac{1}{(m+1)I} \sum_{i=1}^m \sum_{j<i}^m (I_i - I_j)
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{(m+1)I} \{1 + (2+1) + (3+2+1) + [m + (m-1) + \dots + 1]\} \\
 &= \frac{1}{(m+1)I} [1 + 3 + 6 + \dots + \frac{m(m+1)}{2}]
 \end{aligned}$$

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

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

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大即 $m \rightarrow \infty$ 时, G 值便趋向于三分之一(0.33)。笔者以为,0.33 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、基尼系数的简易公式 —— $g = P_5 - P_1$

如上所述,《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五分法数据,而没有城乡合一的收入五分法数据,这便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困难。但是,我们可以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,其具体方法就是以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来简易计算基尼系数。为此,我们有必要探寻“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,我们有以下收入五分法背景下的数学推导:

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

把上式中正的项与负的项分别归并,有

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

上式可称之为收入五分法下的基尼系数的精确计算公式。为使计算简单化,我们将五大收入组称为最低收入组、中下收入组、中等收入组、中上收入组和最高收入组,并进一步假定五大组收入占总收入的 5 个比重 P_1 、 P_2 、 P_3 、 P_4 、 P_5 呈等差数列(记公差为 D),则有:

$$P_2 = P_1 + D$$

$$P_3 = P_1 + 2D$$

$$P_4 = P_1 + 3D$$

$$P_5 = P_1 + 4D$$

代入(1)式,得

$$\begin{aligned}
 G &= \frac{4(P_5 - P_1) + 2(P_4 - P_2)}{5} \\
 &= \frac{4(P_1 + 4D - P_1) + 2(P_1 + 3D - P_1 - D)}{5} \\
 &= \frac{4 \times 4D + 2 \times 2D}{5} \\
 &= 4D \\
 &= P_1 + 4D - P_1 \\
 &= P_5 - P_1
 \end{aligned}$$

诚然,由于现实中五大收入组的收入之差并不一定存在同一的公差,因此,我们以五大收入组的收入比重呈等差数列的假定所推导的基尼系数计算公式,只能是一个简易的近似公式(下面我们将看到,这是一个相当精确的简易公式);为强调上述公式只是近似公式,在以下的分析中,我们以 g 代替 G 并将基尼系数的近似计算公式表示为:

$$g = P_5 - P_1 \tag{2}$$

其意义是:基尼系数近似地等于五分法中收入最高的那组人的收入百分比与收入最低的那组人的收入百分比之差。这一公式十分有用。因为,通常我们见到的报道是:某国居民收入的收入分配十分不公平,收入最高的 20% 的人们的收入已占全部收入的 55%,而收入最低的 20% 的穷人的收入只占全部收入的 4%,等等。根据上述简易公式,我们马上可以算出,这个国家的收入分配的基尼系数为

$$g = 55\% - 4\% = 0.51$$

而且,对于不懂基尼系数的人来说,我们也可以简单地告诉他们:基尼系数大致就是 20% 的最高收入者的收入比重与 20% 的最低收入者的收入比重之差。

上述计算基尼系数的简易公式,简易是简易了,但精确性如何呢?对照世界银行以收入分组五分法测算的 34 个国家的基尼系数,我们以简易计算公式进行复算的比较结果如表 1。

表 1 世界银行公布的 34 个国家的基尼系数及本文用简易公式计算的基尼系数对比 (%)

国家	年份	基尼系数(%, 世行数据)	最低收入 的 20%	第二个 20%	第三个 20%	第四个 20%	最高收入 的 20%	基尼系数(%, 简易公式计算)
澳大利亚	1994	35.2	5.9	12.0	17.2	23.6	41.3	35.4
孟加拉国	1995	33.6	8.7	12.0	15.7	20.8	42.8	34.1
巴西	1997	59.1	2.6	5.7	10.3	18.5	63.0	60.4
保加利亚	1997	26.4	10.1	13.9	17.4	21.9	36.8	26.7
加拿大	1994	31.5	7.5	12.9	17.2	23.0	39.3	31.8
智利	1996	57.5	3.4	6.3	10.5	17.9	62.0	58.6
中国	1998	40.3	5.9	10.2	15.1	22.2	46.6	40.7
丹麦	1992	24.7	9.6	14.9	18.3	22.7	34.5	24.7
法国	1995	32.7	7.2	12.6	17.2	22.8	40.2	33.0
匈牙利	1998	24.4	10.0	14.7	18.3	22.7	34.4	24.4
印度	1997	37.8	8.1	11.6	15.0	19.3	46.1	38.0
印度尼西亚	1999	31.7	9.0	12.5	16.1	21.3	41.1	32.1
以色列	1992	35.5	6.9	11.4	16.3	22.9	42.5	35.6
意大利	1995	27.3	8.7	14.0	18.1	22.9	36.3	27.6
德国	1994	30.0	8.2	13.2	17.5	22.7	38.5	30.3
日本	1993	24.9	10.6	14.2	17.6	22.0	35.7	25.1
韩国	1993	31.6	7.5	12.9	17.4	22.9	39.3	31.8
马来西亚	1997	49.2	4.4	8.1	12.9	20.3	54.3	49.9
墨西哥	1996	51.9	4.0	7.6	12.2	19.6	56.7	52.7
荷兰	1994	32.6	7.3	12.7	17.2	22.8	40.1	32.8
尼日利亚	1996	50.6	4.4	8.2	12.5	19.3	55.7	51.3
挪威	1995	25.8	9.7	14.3	17.9	22.2	35.8	26.1
巴基斯坦	1996	31.2	9.5	12.9	16.0	20.5	41.1	31.6
菲律宾	1997	46.2	5.4	8.8	13.2	20.3	52.3	46.9
波兰	1998	31.6	7.8	12.8	17.1	22.6	39.7	31.9
俄罗斯	1998	48.7	4.4	8.6	13.3	20.1	53.7	49.3
斯里兰卡	1995	34.4	8.0	11.8	15.8	21.5	42.8	34.8
泰国	1998	41.4	6.4	9.8	14.2	21.2	48.4	42.0
西班牙	1990	32.5	7.5	12.6	17.0	22.6	40.3	32.8
英国	1991	36.1	6.6	11.5	16.3	22.7	43.0	36.3
美国	1997	40.8	5.2	10.5	15.6	22.4	46.4	41.2
委内瑞拉	1997	48.8	4.1	8.3	13.2	20.7	53.7	49.6
越南	1998	36.1	8.0	11.4	15.2	20.9	44.5	36.5
赞比亚	1998	52.6	3.3	7.6	12.5	20.0	56.6	53.3

资料来源: The World Bank, World Development Indicators 2001, 其中最后一列基尼系数的数值系笔者用简易公式自己计算。

表 1 说明,我们以简易公式得到的基尼系数与世界银行给出的基尼系数数值相当接近。但在表 1 中,有一个基本事实值得注意,那就是所有国家的 5 个收入组所占的收入比重并不呈等差数列,而只是前 3 个收入组呈现出近似的等差数列排列,自第 4 个收入组开始,其收入比重就开始“起飞”了,到最高收入组则是“一飞冲天”。这种景况似乎在一定程度上否定了我们以等差数列为假定的数学推导,但是,以简易公式所计算的结果与世界银行数据十分接近这一事实,又在很大程度上支持我们所推导的简易公式。对此,有必要作一些说明。

首先,我们摈弃原来“五个收入组的收入比重呈现等差数列排列”的假定,而代之以以下描述:

$$\begin{aligned} P_2 &= P_1 + D \\ P_3 &= P_1 + 2D + s \\ P_4 &= P_1 + 3D + m \\ P_5 &= P_1 + 4D + B \end{aligned}$$

上式中 s 是比较小的数 (small), B 是较大的数 (BIG), m 是比 s 略大的介于 s 与 B 中间的数 (middle)。引入 s 表示 P_1 、 P_2 、 P_3 三个收入组的收入比重只是大致地呈现等差数列排列;引入 m 表示从 P_4 组起收入在等差值 $P_1 + 3D$ 的基础上开始“起飞”;引入 B 表示 P_5 组的收入在等差数值 $P_1 + 4D$ 的基础上已一飞冲天。在上述表述之下,由五分法的基尼系数的精确计算公式(1)可得

$$\begin{aligned} G &= \frac{4(P_5 - P_1) - 2(P_4 - P_2)}{5} \\ &= \frac{4[(P_1 + 4D + B) - P_1] + 2[(P_1 + 3D + m) - (P_1 + D)]}{5} \\ &= \frac{4(4D + B) + 2(2D + m)}{5} \\ &= \frac{20D + 4B + 2m}{5} \\ &= \frac{20D + 5B - B + 2m}{5} \\ &= 4D + B - \frac{B - 2m}{5} \\ &= (P_1 + 4D + B) - P_1 - \frac{B - 2m}{5} \\ &= P_5 - P_1 - \frac{B - 2m}{5} \\ &= g - \frac{B - 2m}{5} \\ g &= G + \frac{B - 2m}{5} \end{aligned}$$

上式表示,用近似的基尼系数计算公式 $g = P_5 - P_1$ 计算所得的数值 g 要比用精确的基尼系数计算公式计算所得的数值 G 大 $(B - 2m)/5$ 。此项中 B 与 $2m$ 相比大不了多少,再被分母 5 除一下就更小,故此项可以忽略不计。

检验 2 列独立随机变量数据有无显著差别要用统计检验的方法。但由于此处 2 列基尼系数不独立,是相互相关的(都来自 5 个收入组数据),因此,一般数理统计书上的显著性检验公式不能用。进一步的讨论已经超出了本文范围。

四、一个测算案例:2002年中国城乡合一的基尼系数

现在,我们根据《中国统计年鉴2003》中有关2002年中国城镇和农村居民收入的分类统计资料,运用简易计算公式来测算城乡合一的基尼系数。

从表2可以看出,在2002年,中国农村33.33%居民的人均纯收入在1650元以下,而同年城镇居民困难户的人均收入为2063.95元。因此,我们有理由认为中国收入最低的20%的人群在农村(由于2002年农村人口占全国人口的60%, $60\% \times 33.33\% = 20\%$),另外,中国城镇较为富裕的人群无疑也是全国较为富裕的人群。按照前面我们推导出的基尼系数的简易公式

$$g = P_5 - P_1$$

我们可以根据《中国统计年鉴》中的城镇调查数据计算富裕的 P_5 而根据农村调查数据计算贫穷的 P_1 。这样,我们就解决了《中国统计年鉴》中由于城乡数据分列而不能计算基尼系数的难题。

具体计算方法如下:由于2002年中国城镇人口占总人口的40%,因此,城镇居民中较富的50%正好占全国人口的20%($40\% \times 50\% = 20\%$)。不算出城镇居民抽样调查中较为富裕的一半人在2002年的人均年收入为11859元。至于农村最穷的33.33%的人(他们就是全国最穷的20%的居民)的人均年收入,我们可以通过《中国统计年鉴2003》中2002年农村最贫穷家庭的收入分布数据获取。从表2的数据我们可以算出这一最穷收入组的人均收入是876.00元。

另外,根据全国城镇与农村居民收入调查的数据可以推算出2002年全国城乡居民的人均年收入为5340.13元。有了最低、最高收入组的人均收入以及全国的人均收入数据,就可以测算出这两个收入组的收入分别占全国总收入的百分比:

$$P_1 = 3.28\%$$

$$P_5 = 44.41\%$$

$$\text{故, } g = P_5 - P_1 = 44.41\% - 3.28\% = 0.4113$$

用类似的方法还可以算出1999年中国城乡合一的基尼系数是0.3983。也就是说,从1999年到2002年,中国的基尼系数在3年间从0.3983增加到0.4113。

我们以简易计算公式所测算出的中国2002年和1999年的城乡合一基尼系数,与一些学者有关相近年份基尼系数之测算结果有一定的差距。例如,赵涛(2000)认为,20世纪90年代下半期中国的基尼系数出现迅速扩大的趋势,1996年为0.42,1997年提高到0.43,1998年则提高至0.46;马敏娜(2001)曾依据历年的《中国统计年鉴》测算出1996年的基尼系数为0.424,1997年为0.425,1998年为0.445。不过,对国内学者计算基尼系数所依据的资料来源持怀疑态度的也不乏其人,王学力(2000)就认为,利用国家统计局颁布的数据,只能分别计算城镇内部和农村内部收入差距的基

表2 2002年中国农村最贫穷家庭的收入分布

人均年收入分组	在农村总收入中的比重(%)	累计比重(%)
100元以下	0.40	0.40
100元—200元	0.19	0.59
200元—300元	0.28	0.87
300元—400元	0.50	1.37
400元—500元	0.79	2.16
500元—600元	1.25	3.41
600元—800元	3.62	7.03
800元—1000元	4.98	12.01
1000元—1200元	6.16	18.17
1200元—1300元	3.25	21.42
1300元—1500元	6.89	28.31
1500元—1650元	5.02	33.33

资料来源:《中国统计年鉴2003》,表中最后1行的数据是笔者通过插值得到的。

尼系数,而不能直接计算全国城乡合一的基尼系数。当然,如果不使用《中国统计年鉴》上的城乡分组单列的数据而自行进行家计调查,是可以计算中国全国城乡合一的基尼系数的。如李强(1995)援引了中国人民大学社会调查中心 1994 年在全国范围的收入调查资料,并据此计算出全国城乡合一的基尼系数为 0.445;中国社会科学院经济研究所也进行过两次大规模的抽样调查,计算出 1995 年全国的基尼系数值为 0.452(李实,赵人伟,1999)。

在笔者看来,国家统计局在《中国统计年鉴》上公布的城乡分列的数据不足以用于计算全国城乡合一的基尼系数这一问题现在已经被解决了:因为我们已经有了基尼系数的简易而又相当精确的计算公式。我们以简易计算公式所计算的 1999 年的城乡合一的基尼系数 0.3983 与以国际惯例计算的 0.39 十分接近。至于这一数值是否偏小,则是统计数据的问题。这已经超出了我们的讨论范围。

参考文献

- 李强,1995:《我国社会各阶层收入差距分析》,《科技导报》第 11 期。
 李实等,1998:《中国经济转型与收入分配变动》,《经济研究》第 4 期。
 李实、赵人伟,1999:《中国居民收入分配再研究》,《经济研究》第 4 期。
 林伯强,2003:《中国的经济增长、贫困减少与政策选择》,《经济研究》第 12 期。
 马敏娜,2001:《我国居民收入差距扩大对消费需求的影响》,《当代经济研究》第 1 期。
 王学力,2000:《个人收入差距的现状、问题和对策》,《改革》第 6 期。
 赵人伟等,1999:《中国居民收入分配再研究》,中国财政经济出版社。
 赵涛,2000:《收入分配不公的状况亟待扭转》,《经济研究参考》第 54 期。
 Cowell, F., 1995, *Measuring Inequality* (Second Edition), Prentice Hall and Harvester Wheatsheaf, London, Sen, A., 1997, *On Economic Inequality*, Clarendon Press, Oxford.
 Dollar D., and Kraay A., 2001, "Growth Is God for the Poor", Mimeo, World Bank, Washington D. C., Policy Research Working Paper, 2587, 11—49.
 Kakwani, Nanak C., 1980, "Income Inequality and Poverty Methods of Estimation and Policy Applications", Oxford University Press.

A Study of the Best Theoretical Value of Gini Coefficient and Its Concise Calculation Formula

Hu Zuguang

(Zhejiang Gongshang University, China)

Abstract: The paper pointed out the difficulty of calculating China's Gini coefficient because the income distribution data in the "China Statistics Yearbook" is given separately under the title of "Urban" and "rural". The paper got a concise formula for calculating Gini coefficient: Gini Coefficient equals the difference between the income percentage of the highest income group and the income percentage of the lowest income group in the quinquartite income. The paper proved with strict mathematics that the best theoretical value of Gini coefficient is 0.3333, or a third. The paper calculated Gini coefficient of 32 countries by using the concise formula and the data published by the World Bank. A comparison between the calculated Gini coefficients and the published Gini coefficients by the World Bank shows that the results are very close to each other. The analysis of the best theoretical value and the concise calculation formula of Gini coefficient in the paper provided a simple and quick way to calculate Gini coefficient of China as a whole.

Key Words: Gini Coefficient; Best Theoretical Value; Quinquartite Method of Income; Concise Calculation Formula

JEL Classification: D630, D300

(责任编辑:郑 健)(校对:林)

前国务院总理朱镕基在九届人大四次会议举行的记者招待会上曾说过“据 1999 年的调查,中国 1999 年按国际惯例计算的基尼系数是 0.39”。