

案例1

基尼系数基于概率意义上的解释及其估计

1、背景知识

基尼系数是意大利经济学家基尼于1922 年根据洛伦茨曲线提出的定量测定收入分配差异程度的指标. 该系数的经济含义是: 在全部居民收入中用于不平均分配的百分比. 这个系数用来刻画贫富差异及测度区域差异, 基尼系数越大, 收入分配越是趋向不平等, 区域差异也越大. 刻画随机变量取值的散布程度, 最常用的是方差和标准差, 它们在使用上的一个不便之处在于其值与量度单位有关. 统计上的变异系数可以免除上述缺点. 基尼系数与变异系数起同样的作用, 但基尼系数有一个重要的优点: 其值介于0 与1 之间, 因而赋予一种直观的标示意义. 基尼系数与变异系数, 虽同属刻画变异, 二者之间并无同步关系. 例如有两个人群A、B, 在人群A, 所有的人具有相同财富, 在人群B, 全部财富集中于一人之手. 从通常的观点看, 这两个人群中财富变异都小, 但基尼系数在人群A 为0, 而在人群B 则为1.

2、案例分析

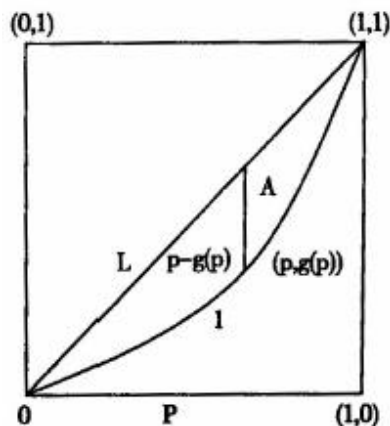
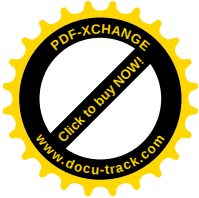
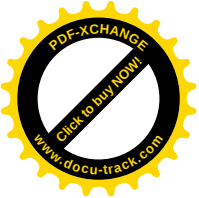
(1) 基尼系数的概率定义

假设一个总体有 M 人, 以 X 表示此人群中单个人所拥有的财富, 不妨把 X 的概率分布视为连续的, 即有一个概率密度函数 $f(x)$, 其相应的分布函数记为 $F(x)$, 则总体的人均拥有财富为 $m = \int_0^{\infty} xf(x)dx$, 而财富总量为 $N = mM$, 固定 $p, (0 \leq p \leq 1)$, 找 x_p , 使 $F(x_p) = p$, 在统计上把 x_p 称为其分布 F 的 P 分位数, 这部分人拥有的财富总量为 $N_p = M \int_0^{x_p} xf(x)dx$, 由此算出, 这部分人所拥有的财富在该群总财富中所占份额, 为 $g(p) = \frac{N_p}{N} = \frac{1}{m} \int_0^{x_p} xf(x)dx$, 显然有 $g(0) = 0, g(1) = 1$, 直观上用以看明白, 当 $0 \leq p \leq 1$ 时, 有 $0 \leq g(p) \leq p$.

因为低收入者所占财富份额, 显然超不过其在人口数中所占份额. 当 p 由0 变化到1 时, 点 $(p, g(p))$ 描绘出一条曲线, 即图1中的曲线 l , 由式(1) 知 l 必位于图1中正方形对角线 L 的下方, $p - g(p)$ 反映出低收入者财富所占比例, 与在财富绝对平均时应有的比例之间的差距, 差距愈大, 表示财富分配愈不平均, 而曲线 l 就愈远离对角线 L , 故可以把 L 与 l 之间所夹出的这一块的面积 A , 作为衡量财富在该人群中分布不均的一个综合指标. 由于 A 处在正方形的对角线的下方, 基尼取比值面积 A/L 下方的正方形的面积作为指标, 此即基尼系数, 由于 L 下方的正方形面积为 $1/2$, 故有 $G = \text{基尼系数} = 2A$, 这一定义的好处是使基尼系数在0 到1 之间.

(2) 基尼系数的点估计

由于在估计基尼系数 G 这样的问题中, 一般会碰到很大的样本量 n , 但样本多以分组的形式给出, 即只给出财富在某区间内的人数是多少, 而不知各人财富的确值, 这时, 估计公式要相应地作一点调整.



设样本值处在的区间为 $[a, b]$, 分成 n 个长为 $\frac{b-a}{n}$ 的等距组 (a_{i-1}, a_i) , $i=1, 2, \dots, n$ 其中 $a_i = a_0 + \frac{i(b-a)}{n}$, $a_0 = a, a_n = b$, 第 i 组区间 (a_{i-1}, a_i) 内的样本数为 n_i , 都算作取值于区间中点 $a + \frac{i(b-a)}{2n}$ 处, 对这样的分组样本, 平均值 m 的估计使用 $m^* = \sum_{i=1}^n (a + \frac{i(b-a)}{2n}) \frac{n_i}{n}$, 不致产生多大偏差, 利用数据直接估计出曲线 $g(p)$ (即图1 中的曲线 I) , 注意到:

在 a_i 点处, $p = P(X \leq a_i)$ 的估值为 $p_i = \frac{1}{n} \sum_{k=1}^i n_k$, 而积分 $\int_0^{a_i} xf(x)dx$ 的估值为

$$t_i = \sum_{j=1}^i (a + \frac{j(b-a)}{2n}) \frac{n_j}{n}$$

由此得到 $g(p)$ 的估值 $g_i = \frac{t_i}{m^*}$,

这样, 我们得到图1 中的曲线 I 上的一些点 $(p_i, g_i), 1 \leq i \leq n$, 可以用某种光滑化的方法, 利用这些点(近似地) 决定曲线 I , 或直接利用数值积分的方法, 以决定图1 正方形内曲线 I 下方那一块面积的值, 并由此决定基尼系数 G .