Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística

.

PICME - Programa de Iniciação Científica e Mestrado

Índice de Gini e distribuição de Pareto

Aluno: Vinícius Litvinoff Justus

Orientador: Prof. Dr. Mauricio Enrique Zevallos Herencia

Órgão financiador: CNPq

Palavras chave: Desigualdade; Gini; Econometria.

Campinas
2021

Resumo: o coeficiente de Gini é um índice de dispersão relativa amplamente utilizado para medir desigualdade de renda e patrimonial. Segundo Gastwirth, "a cauda superior da distribuição de renda é geralmente aproximada pela distribuição de Pareto"[1] (2016, pág. 4, tradução nossa). Apesar das limitações do método e da crítica de alguns autores,"[...] tornou-se geralmente aceito, principalmente com base em evidências empíricas e não em qualquer fundamento teórico, que a maioria das distribuições de renda realmente exibia o comportamento da cauda de Pareto" [2] (Arnold, 2014, pág. 1, tradução nossa). Tendo em vista a ampla utilização desta distribuição para o estudo da distribuição de renda, este trabalho visa estudar se a distribuição de Pareto é capaz de aproximar bem toda a distribuição; além disso, pretende-se estudar as propriedades da distribuição, verificando se o valor do coeficiente do Gini é bem predito, além de se considerar o índice de Palma, medida de desigualdade em ascensão hoje.

Palavras chave: Desigualdade; Gini; Econometria.

1 Introdução

Dizemos que uma variável aleatória X segue uma distribuição de Pareto com parâmetros (α, m) se a sua função densidade de probabilidade (fdp) é dada por:

$$f(x) = \frac{\alpha m^{\alpha}}{x^{\alpha+1}} I_{[m,\infty)}(x) \tag{1}$$

Perceba que o suporte da distribuição depende de m: X assume valores inferiores a m com probabilidade nula.

Sejam G e L, respectivamente, o índice de Gini e a curva de Lorenz de uma variável aleatória. Possivelmente, um dos resultados mais importates para este trabalho é o fato de que o valor destes indicadores associado a uma distribuição de Pareto depende apenas de α, isto é, eles não dependem de m:

$$G(X) = \frac{1}{2\alpha - 1} \tag{2}$$

$$L(p) = 1 - (1-p)^{1-\frac{1}{\alpha}} \tag{3}$$

Embora isto não elimine a importância de se conhecer m e, portanto, conhecer toda a distribuição, isto traz a vantagem de que, para a finalidade de analisar exclusivamente as medidas clássicas de desigualdade, basta estimar um parâmetro. Também é possível demonstrar que o índice de Palma, outro indicador de desigualdade econômica, também depende apenas de α.

Esta Seção explorará a questão de como estimar α a partir de dados agregados.

1.1 Estimação de α

Sabemos (Rytgaard, 1990) que o estimador de máxima verossimilhança para α é dado por:

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} ln(X_i / min_j(x_j))} \tag{4}$$

[1] "The upper tail of the income distribution is often approximated by a Pareto distribution [...]"

[2] "[...] it became, chiefly on the basis of empirical evidencerather than on any theoretical grounds, generally accepted that most income distributions did indeedexhibit Paretian tail behavior"

No entanto, na maioria das situações com o qual nos deparamos dentro desta área, não temos acesso a toda amostra $(x_1, x_2, ..., x_n)$ de rendas individuais ou a uma estatística suficiente para α, o que torna necessário construir novos estimadores que sejam funções das quantidades presentes nos bancos de dados.

Exceto caso especificado em contrário, este resumo trabalhará apenas com exemplos usando decis; no entanto, os procedimentos com outros números de quantis são análogos. Sejam $p_1, p_2, ..., p_{10}$ as proporções de renda em cada decil[3] - isto é, $0 \leq p_i \leq 1$ para todo $i = 1, 2, .., 10$ e $\sum_{i=1}^{10} p_i = 1$. No decorrer da iniciação científica, foi criado o seguinte estimador para α:

$$A = \frac{1}{1 - log_{0.1}(p_{10})} \tag{5}$$

Seja $Q(P)$ o quantil P, para $0 \leq P \leq 1$. Foi deduzido que:

$$Q(P) = \frac{m}{(1-P)^{\frac{1}{\alpha}}} \tag{6}$$

Também foi deduzido que a proporção de renda entre os quantis Q(P) e Q(P') é:

$$(1-P)^{\frac{\alpha-1}{\alpha}} - (1-P')^{\frac{\alpha-1}{\alpha}} \tag{7}$$

Assim, uma vez que estimamos α, podemos estimar $p_1, p_2, ..., p_{10}$ pelo princípio plug-in e, deste modo, podemos comparar os quantis "empíricos"com os quantis estimados, o que permite mostrar a efetividade do método proposto.

O estimador A foi construído sobre a ideia de que, se uma população segue uma distribuição de Pareto (α, m), então a proporção de renda entre $Q(0.9)$ e $Q(1)$ é igual a $(0.1)^{\frac{\alpha-1}{\alpha}}$, o que permite encontrar o valor $\hat{\alpha}$ que satisfaz $(0.1)^{\frac{\hat{\alpha}-1}{\hat{\alpha}}} = p_{10}$. Por análogo argumento, é possível construir um estimador a partir dos 10% inferiores ao invés dos 10% superiores:

$$A_2 = \frac{1}{1 - log_{0.9}(1 - p_1)} \tag{8}$$

Aigner e Goldberger (1970) apresentam diversos métodos para estimar α a partir do uso de dados agregados. Nos concentraremos especificamente no estimador de máxima verossimilhança, pois ele apresenta - junto com o estimador de mínimos quadrados generalizados - a menor variância assintótica (Aigner e Goldberger, 1970, pág. 721).

O estimador de máxima verossimilhança para dados agregados (a_l) é dado pela solução da equação:

$$\left[\sum_{t=0}^{T-1} f_t \frac{x_{t+1}^{-a} x_{t+1}^{*} - x_t^{-a} x_t^{*}}{x_t^{-a} - x_{t+1}^{-a}} \right] - f_t x_t^{*} = 0, \tag{9}$$

onde $x_0 = 1 < x_1 < x_2 < ... < x_{T+1} = \infty$ são os quantis, f_i é a proporção de pessoas no i-ésimo intervalo e $x_t^{*} = ln(x_t)$. O valor de a deve ser obtido numericamente.

Uma limitação importante é o fato de que a expressão para encontrar o estimador a_l assume que $x_0 = 1$, o que, em outras palavras, implica $m = 1$. Portanto, o método não pode ser aplicado para outros valores de m, exceto caso ele passe por alguma modificação que não foi encontrada na literatura ou descoberta durante o projeto.

[3] Informalmente, a definição correta de quantil é uma "Linha divisória"com certas propriedades; como abuso de linguagem, este termo será usado para se referir a todo o "conteúdo"entre as "Linhas divisórias", isto é, todas as observações entre um decil e o decil subsequente.

Tabela 1: Estimadores

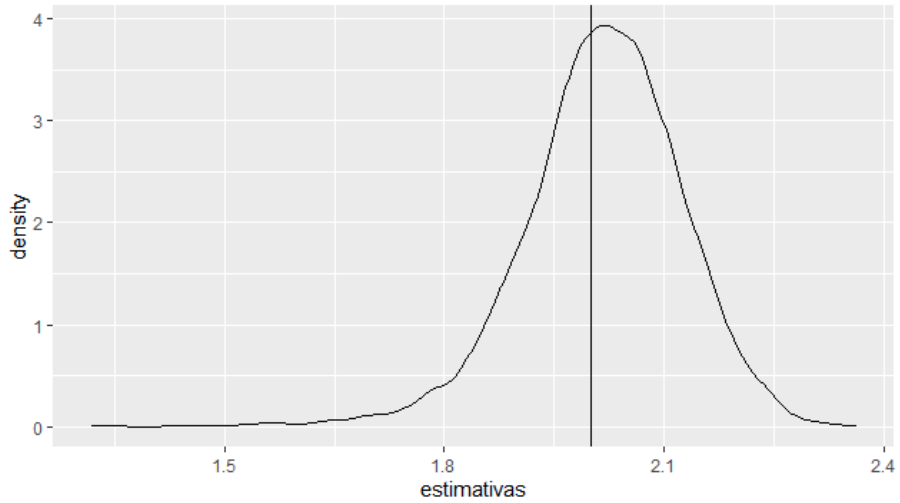
Quantidade	A1	A2	al
1° quartil	1.953	1.967	1.969
Mediana	2.021	2.010	2.001
3° quartil	2.088	2.051	2.033
Média	2.016	2.005	2.001

1.2 Simulação: desempenho dos estimadores

Seja X uma variável aleatória de Pareto com parâmetros $(\alpha = 2, m = 1)$. Foram feitas 10000 simuações para o estimador proposto no relatório anterior (A) e para o estimador de máxima verossimilhança para dados agregados (a_l), todos com uma amostra $n = 2000$.

O estimador A apresentou desvio padrão 0.1109471 e vício 0.016; o estimador A_2 apresentou desvio padrão 0.06980441 e vício 0.005; o estimador A apresentou desvio padrão 0.04766818 e vício 0.001. Os quartis e a média dos estimadores podem ser vistos na Tabela 1.

Figura 1: Distribuição do estimador A para $\alpha = 2, m = 1, n = 2000$.



O estimador de máxima verossimilhança para dados agregados apresentou menos desvio padrão e menor viés do que os outros dois estimadores. Por outro lado, uma possível vantagem do método proposto é a existência de uma solução analítica para $\hat{\alpha}$.

2 Referências

AIGNER, Dennis J.; GOLDBERGER, Arthur S. Estimation of Pareto's law from grouped observations. Journal of the American Statistical Association, v. 65, n. 330, p. 712-723, 1970.

ARNOLD, Barry C. Pareto distribution. Wiley StatsRef: Statistics Reference Online, p. 1-10, 2014.

BAKLIZI, Ayman. Estimation of the Pareto scale parameter based on grouped data. Journal of Interdisciplinary Mathematics, v. 5, n. 2, p. 177-182, 2002.

Figura 2: Distribuição do estimador A_2 para $\alpha = 2, m = 1, n = 2000$.

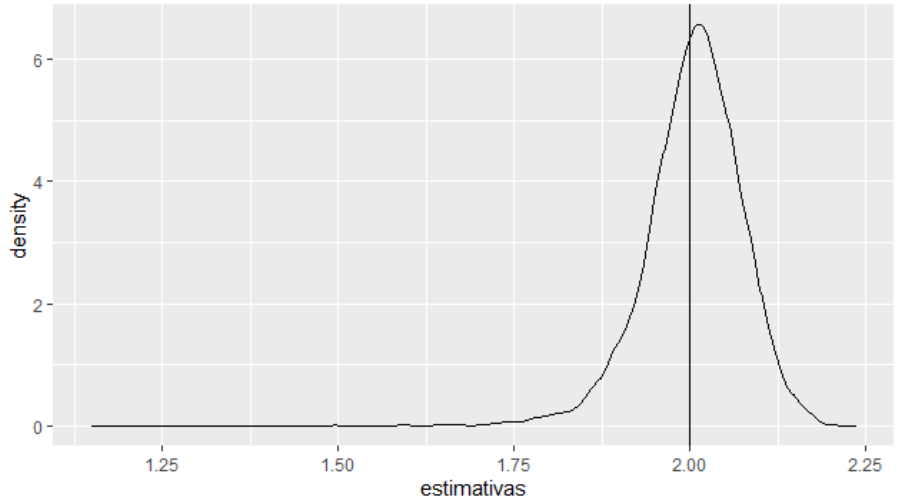
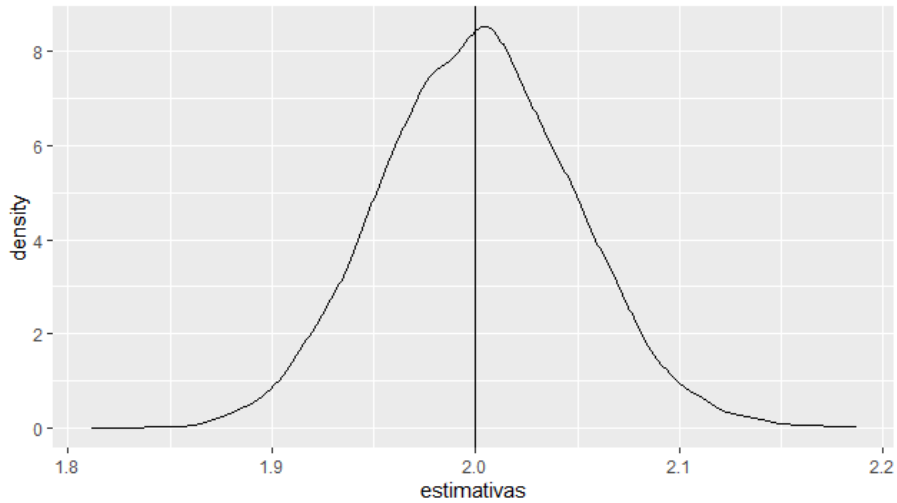


Figura 3: Distribuição do estimador a_l para $\alpha = 2, m = 1, n = 2000$.



GASTWIRTH, Joseph L. Is the Gini index of inequality overly sensitive to changes in the middle of the income distribution?. Statistics and Public Policy, v. 4, n. 1, p. 1-11, 2017.

GASTWIRTH, Joseph L. Measures of economic inequality focusing on the status of the lower and middle income groups. Statistics and Public Policy, v. 3, n. 1, p. 1-9, 2016.

HOFFMANN, Rodolfo; BOTASSIO, D. C.; JESUS, J. G. Distribuição de renda: medidas de desigualdade, pobreza, concentração, segregação e polarização. São Paulo: Editora da Universidade de São Paulo, 2019.

RYTGAARD, Mette. Estimation in the Pareto distribution. ASTIN Bulletin: The Journal of the IAA, v. 20, n. 2, p. 201-216, 1990.