Models of Income Distributions for Knowledge Discovery

Alfred Ultsch, Stefan Schnabel, Michael C. Thrun

Databionics Research Group, University of Marburg, Germany

Descriptions of income distributions using a single distribution, like Lognormal or Gamma are often quite poor in describing the tails of the distribution [1]. This led to separate models for the upper vs. lower parts of income distributions [2]. For example [3-5] describe the high-income region with the Pareto power laws. Other authors model the low to medium income region using Exponential [6], Lognormal [7] or Gamma distributions [8, 9]. The high income range is often modeled using the cumulative distribution function (cdf) [10], whereas the low to medium income regions are modeled using the probability density function (pdf) [11]. Usually no systematic limits between low, medium and high income are defined [3]. A goal for a valid and suitable model for income distributions is to derive a theory of the mechanisms which operate in a society (Computational Social Science) and explains the observed distribution [12]. Here a model for income distributions as a mixture of components is proposed. The model is derived using the Pareto Density Estimation (PDE) [14] for an estimation of the pdf. PDE has been designed in particular to identify groups/classes in a dataset [13]. Precise limits for the classes can be calculated using the theorem of Bayes. Our model suggests that there are different groups/classes in a society, which contribute to the total distribution of income in their own way. The approach is demonstrated on several real world data sets including actual income data from Germany.

1. Dagum, C., New Model of personal Income-Distribution-Specification and Estimation, Economie appliquée, 30(3), pp 413-437, 1977.
2. Richmond, P., et al., A Review of empirical Studies and Models of Income Distributions in Society. Wiley, Berlin, 2006.
3. Chatterjee, A., B.K. Chakrabarti, and R.B. Stinchcombe, Master Equation for a kinetic Model of a Trading Market and its analytic Solution. Phys Rev E Stat Nonlin Soft Matter Phys. 72(2 Pt 2, p. 026126, 2005.
4. Levy, M. and S. Solomon, New Evidence for the power-law Distribution of Wealth, Physica A: Statistical Mechanics and its Applications, 242(1), p. 90-94, 1997.
5. Drăgulescu, A. and V.M. Yakovenko, Exponential and power-law Probability Distributions of Wealth and Income in the United Kingdom and the United States. Physica A: Statistical Mechanics and its Applications, 299(1), pp 213-221, 2001.
6. Chakrabarti, A.S. and B.K. Chakrabarti, Statistical theories of income and wealth distribution. Economics: The Open-Access, Open-Assessment E-Journal. 4, p. 4, 2010.
7. Clementi, F. and M. Gallegati, Pareto's law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States, in Econophysics of wealth distributions, Springer, New York, pp 3-14, 2005.
8. Ferrero, J.C., The statistical Distribution of Money and the Rate of Money Transference. Physica A: Statistical Mechanics and its Applications, 341, p. 575-585, 2004.
9. Scafetta, N., S. Picozzi, and B.J. West, An out-of-equilibrium Model of the Distributions of Wealth. Quantitative Finance. 4(3): pp 353-364, 2004.
10. Di Matteo, T., T. Aste, and S. Hyde, Exchanges in complex Networks: Income and Wealth Distributions, arXiv preprint, 2003.
11. Drăgulescu, A. and V.M. Yakovenko, Evidence for the exponential Distribution of Income in the USA. The European Physical Journal B-Condensed Matter and Complex Systems, 20(4), pp 585-589, 2001.
12. Cioffi-Revilla, C., Introduction to Computational Social Science: Principles and Applications. Springer, Berlin, 2013.
13. Ultsch, A., Pareto Density Estimation: A Density Estimation for Knowledge Discovery, in Innovations in classification, data science, and information systems, Springer, New York, pp 91-100, 2005.