Estimating Small Area Income Distributions and Income Statistics via the Inequality Process (IP)

John Angle
angle@inequalityprocess.org
Inequality Process Institute
www.inequalityprocess.org
Post Office Box 215
Lafayette Hill, Pennsylvania 19444-0215

and

Kenneth C. Land
kland@soc.duke.edu
Center for Population Health and Aging
Post Office Box 90088
Duke University
Durham, North Carolina 27708-0088

1. Small Area Estimation of Income via "Borrowing Strength"

The phrase "borrowing strength" is used in empirical Bayes estimation of small area statistics to refer to the use of information on the statistic to be estimated in adjoining, nearby, or enveloping areas (Carlin and Louis 2009; Fay and Herriot 1979). If "borrowing strength" works, it must do so because the process generating the unknown value of the statistic in the small area of interest is the same process generating its known value in adjoining, nearby, or enveloping areas, where the values of the parameters of the process may be similar to those in the small area of interest. In nearly all cases of "borrowed strength" estimates, the common process generating the statistic to be estimated is unknown. If the process were known, its direct use might yield better estimates than those of linear "borrowed strength" models that indirectly and partially capture its similar effects in adjoining, nearby, or enveloping areas. A scientific law, if applicable, is the most concentrated form of "strength" [information] for small area estimation. In Bayesian terminology, the existence of a scientific law leads to a highly concentrated prior distribution for model parameters and reduces uncertainty in their estimation very substantially.

1.1 The Inequality Process (IP) as a Source of "Strength" in Estimating the Distribution of Earned Income Distribution and Associated Statistics in a Small Area

The present paper uses a parametric model of the process generating income statistics and income distributions to do small area estimation of income statistics and income distribution. This model is what is known as the Inequality Process (IP) (Angle, 1983-2007), a stochastic particle dynamics system at the most abstract level. The Macro Model of the Inequality Process is a gamma probability density function (pdf) approximation to its stationary distribution, expressed in terms of the parameters of the Inequality Process (see the Appendix for a mathematical representation of the model). The Inequality Process has a wide scope of explained income and wealth phenomena. See Table 1.

Table 1. Published Applications/Explanandum of the Inequality Process (IP)

1. Why the gamma family of probability density functions (pdfs) is a useful approximation to wage income distributions conditioned on education and why the unconditional distribution of wage income has a right tail whose heaviness approximates that of a Pareto pdf;
2. How the unconditional distribution of annual wage income appears to be gamma distributed although the gamma distribution is not closed under mixture, i.e., aggregation by area; [Mixture is a different operation from convolution.]
3. The shapes of the distribution of wage incomes of workers by level of education, why this sequence of shapes changes little over decades, and why it is similar to the sequence of shapes of the unconditional distribution of wage income over the course of techno-cultural evolution;
4. The dynamics of the distribution of wage income conditioned on education as a function of the unconditional mean of wage income and the distribution of education in the labor force;
5. Why the distribution of wage income is different from the distribution of income from tangible assets;
6. Why the IP's parameters estimated from certain statistics of the wage incomes of individual workers in longitudinal data on annual wage incomes are ordered as predicted by the IP's meta-theory and approximate estimates of the same parameters from the fit of the IP's stationary distribution (the IP's Macro Model) to the distribution of wage income conditioned on education;
7. The Kuznets pattern in the Gini concentration ratio of earned income during the industrialization of an agrarian economy;
8. In an elaboration of the basic IP, if a particle in a coalition of particles has a probability different from 50% of winning a competitive encounter with a particle not in the coalition, this modified IP can reproduce features of the joint distribution of income to African-Americans and other Americans: a) the % minority effect on discrimination (the larger the minority, the more severe discrimination on a per capita basis); b) the relationships among: i) % of a U.S. state's population that is non-white; ii) median non-African-American male earnings in a U.S. state; iii) the Gini concentration of non-African-American male earnings in a U.S. state; and iv) the ratio of African-American male to non-African-American male median earnings in a U.S. state. This version of the IP (Angle, 1992) appeared sufficiently significant to the late Prof. John Hope Franklin to be included in the Inventory of the John Hope Franklin Papers, 1889-1998 and undated, bulk 1970-1998, Rare Book, Manuscripts, and Special Collections Library, Duke University Libraries [http://library.duke.edu/digitalcollections/rbmscl/franklinjohnhope/inv/] .

1.1.1 *The Inequality Process as a Success of Econophysics*

Angle (1990, 2000) places the Inequality Process (IP) as a mathematical object in the class of stochastic models known as particle systems. Angle (1990) compares the Inequality Process to the classic particle system of statistical mechanics, the Kinetic Theory of Gases (KTG). The KTG is the mechanical basis of thermodynamics. The KTG explains the heat of a volume of mass as the kinetic energy of the molecules of the mass and the temperature of the mass as a function of mean kinetic energy of the molecules.

The Inequality Process dates from 1983. Physicists began independently investigating the properties of a particle system model that is a hybrid of the IP and the KTG in 2000 (Chakraborti and Chakrabarti, 2000). Partial rediscovery of a few of the properties of the IP published by Angle in the 1980's excited physicists attempting to apply classes of mathematical models, such as the particle system, used in statistical mechanics to economic phenomena, an interdisciplinary field, econophysics.

At an international 2005 meeting called to celebrate five years progress on the Chakraborti and Chakrabarti model, Thomas Lux, chair of the Economics Department of Kiel University, Germany, told the conference that the five years of progress had been anticipated in the literature on the Inequality Process (IP). The most recent IP paper he cited was published nine years earlier, nine years during which the IP's published explanandum continued to expand. In the abstract of the paper in the conference's proceedings volume, Prof. Lux writes ".... We point out that recent models of wealth condensation which draw their inspiration from molecular dynamics have, in fact, reinvented a process introduced quite some time ago by Angle (1986) in the sociological literature. ....". (Lux, 2005: 51).

Reviewing the history of particle system models of income distribution, Yakovenko and Rosser (2009) write on page 3 of "Colloquium: Statistical Mechanics of Money, Wealth, and Income". [on-line at http://arxiv.org/abs/0905.1518] about particle system models of income: "Actually, this approach was pioneered by the sociologist John Angle (1986, 1992, 1993, 1996, 2002) already in the 1980s. However, his work was largely unknown until it was brought to the attention of econophysicists by the economist Thomas Lux (2005). Now, Angle's work is widely cited in econophysics literature (Angle, 2006)."

2. Applying the Macro Model of the Inequality Process (IP) to the Estimation of the Earned Income Distribution and Income Statistics in a Small Area

The American Community Survey's PUMS (public use microdata sample) provides records of individuals in households surveyed in the American Community Survey (ACS), a project of the U.S. Bureau of the Census to collect detailed information on U.S. communities and make them available to the public. The ACS' PUMS has observations on the earned income of individuals. The ACS' PUMS files also contain geographic information: the state, PUMA (public use microdata area), in some cases the identification of sub-PUMA areas as well. PUMA's are geographic areas, often corresponding to a city or a county, or a group of counties, containing at least 100,000 residents (U.S. Bureau of the Census, 2009: 3). If not coextensive with a single county, PUMA's must be a combination of contiguous counties or census tracts. A state's PUMA's are a disjoint covering of its area. An example of a PUMA in the state of Pennsylvania is PUMA #42170, co-extensive with Montgomery County, Pennsylvania, a suburban county just to the west of Philadelphia. Summary ACS data are available on earned income by gender for county subdivisions with at least 65,000 residents. Lower Merion Township is an example of a county subdivision in Montgomery County, Pennsylvania. The ACS' PUMS file can be used to estimate the distribution of earned income in every PUMA, both the unconditional distribution and the distribution conditioned on any variable in the ACS' questionnaire, for example, the distribution of earned income conditioned on education and gender.

2.1 An Application of the Inequality Process to the Estimation of Earned Income for Small Areas

In areas with fewer residents than the minimum of 100,000 for a PUMA but with at least 65,000 residents, the following three types of summary statistics based on ACS data are published by the U.S. Bureau of the Census:

1) a tabulated distribution (histogram, if relative frequencies are estimated and plotted) of the number of men and women earners in a set of earned income bins (ranges of annual earned income) up to the topcode; the published income bins have different widths (the maximum binned income minus the minimum binned income);

2) the mean and the median of earned income of men and women in #1), the tabulated distribution;

3) the mean and the median earned income of men and women, 25 years of age or older, by level of education.

These summary statistics on county subdivisions within a PUMA permit an evaluation of the ability of the Inequality Process to estimate the distribution of earned income within small areas. The empirical target to be explained in this test is the distribution of earned income by gender in a county subdivision. Its Inequality Process estimate is based on a) the distribution of earned income conditioned on education, gender, and age in the PUMA, b) the distribution of the employed by gender, age, and level of education in the county subdivision, c) the mean and median of earned income in the county subdivision, and d) the median of earned income by age and gender in the county subdivision.

This empirical evaluation will be performed on PUMA's and subcounty divisions in and around Philadelphia, Pennsylvania and the Raleigh-Durham area in North Carolina. To show that there is more "strength" in the Inequality Process than in a standard empirical Bayes estimate, the county subdivision's earned income distribution will be estimated directly from a) and b). c) and d) can be incorporated into the Inequality Process' estimate but there is no obvious way to incorporate that information into an empirical Bayes estimate.

Appendix. The Macro Model of the Inequality Process (IP)

The macro model of the Inequality Process (IP) is a gamma probability density function (pdf), $f_{\psi t}(x)$, a model of the wage income, x, of workers at the same level of education, the ψ^{th} at time step t . The IP's macro model in the ω_ψ equivalence class is:

$$f_{\psi t}(x) \equiv \frac{\lambda_{\psi t}^{\alpha_\psi}}{\Gamma(\alpha_\psi)} x^{\alpha_\psi - 1} e^{-\lambda_{\psi t} x} \tag{1}$$

or in terms of the IP's parameter in the ω_ψ equivalence class:

$$f_{\psi t}(x) = \exp\left(\left(\frac{1-\omega_\psi}{\omega_\psi}\right) \ln\left(\frac{1-\omega_\psi}{\tilde{\omega}_t \mu_t}\right) - \ln \Gamma\left(\frac{1-\omega_\psi}{\omega_\psi}\right) + \left(\frac{1-2\omega_\psi}{\omega_\psi}\right) \ln(x) - \left(\frac{1-\omega_\psi}{\tilde{\omega}_t \mu_t}\right) x \right) \tag{2}$$

where:

α_ψ = the shape parameter of the gamma pdf that approximates the distribution of wealth, x, in the IP's ω_ψ equivalence class, intended to model the wage income distribution of workers at the ψ^{th} level of education regardless of time;

$$\alpha_\psi > 0$$
$$\alpha_\psi \approx \frac{1-\omega_\psi}{\omega_\psi} \tag{3}$$

and:

$\lambda_{\psi t}$ = the scale parameter of the gamma pdf that approximates the distribution of wealth, x, in the IP's ω_ψ equivalence class, intended to model the wage income distribution of workers at level ψ of education in a labor force with a given unconditional mean of wage income and a given harmonic mean of ω_ψ's at time step t;

$$\lambda_{\psi t} > 0$$

$$\lambda_{\psi t} \approx \frac{(1-\omega_{\psi})\left(\frac{w_{1t}}{\omega_1} + \ldots \ldots + \frac{w_{\psi t}}{\omega_{\psi}} + \ldots \ldots + \frac{w_{\Psi t}}{\omega_{\Psi}}\right)}{\mu_t}$$

$$\lambda_{\psi t} \approx \frac{1-\omega_{\psi}}{\tilde{\omega}_t \ \mu_t} \tag{4}$$

where:

μ_t = unconditional mean of wage income at time step t

$\tilde{\omega}_t$ = harmonic mean of the ω_{ψ}'s at time step t

$w_{\psi t}$ = the proportion of particles in the ω_{ψ} equivalence class at time step t; assumption that the $w_{\psi t}$ of all distinct ω_{ψ} are small and no particular $w_{\psi t}/\omega_{\psi}$ ratio dominates $\tilde{\omega}_t$.

and μ_t and the $w_{\psi t}$'s are exogenous and the sole source of change in a population of particles where Ψ ω equivalence classes are distinguished. Consequently, the dynamics of (1), the IP's Macro Model, are exogenous, that is, driven by the product ($\tilde{\omega}_t\ \mu_t$) and expressed as a scale transformation, i.e., via $\lambda_{\psi t}$.

References

Angle, J. 1983. "The surplus theory of social stratification and the size distribution of Personal Wealth." ***1983 Proceedings of the American Statistical Association,*** Social Statistics Section**.** Pp. 395-400. Alexandria, VA: American Statistical Association.

_____, 1986. "The surplus theory of social stratification and the size distribution of Personal Wealth." ***Social Forces*** 65:293-326.

_____. 1990. "A stochastic interacting particle system model of the size distribution of wealth and income." ***1990 Proceedings of the American Statistical Association,*** Social Statistics Section. Pp. 279-284. Alexandria, VA: American Statistical Association.

_____. 1992. "The Inequality Process and the distribution of income to blacks and whites". ***Journal of Mathematical Sociology*** 17:77-98.

_____. 1993. "Deriving the size distribution of personal wealth from 'the rich get richer, the poor get poorer'". ***Journal of Mathematical Sociology*** 18:27-46.

_____. 1996. "How the gamma law of income distribution appears invariant under aggregation". ***Journal of Mathematical Sociology***. 21:325-358.

_____. 2000. "The binary interacting particle system (bips) underlying the maxentropic derivation of the gamma law of income distribution". ***2000 Proceedings of the American Statistical Association,*** Social Statistics Section. Pp. 270-275. Alexandria, VA: American Statistical Association.

_____. 2006 "The Inequality Process as a wealth maximizing algorithm". ***Physica A: Statistical Mechanics and Its Applications*** 367:388-414 (DOI: 10.1016/j.physa.2005.11.017).

_____. 2007. "The Inequality Process is an evolutionary process". ***The Constructal Theory of Social Dynamics***. Adrian Bejan and Gilbert Merkx, eds. (Proceedings of the Conference on the Constructal Theory of Social Dynamics, Duke University, April 2006). New York: Springer.

_____. 2007. "The Macro Model of the Inequality Process and The Surging Relative Frequency of Large Wage Incomes". Pp. 171-196 in A. Chatterjee and B.K. Chakrabarti, (eds.), ***The Econophysics of Markets and Networks*** (Proceedings of the Econophys-Kolkata III Conference, March 2007 Milan: Springer.

Carlin, B. P. and T. A. Louis. 2009. **Bayesian Methods for Data Analysis**. Third Edition. New York: CRC Press.

Chakraborti, A., B.K. Chakrabarti. 2000. "Statistical mechanics of money: How saving propensity affects its distribution". **European Physics Journal B**: 17: 167-

Fay, R. E., and R. A. Herriot. 1979. "Estimates of income for small places: An application of James-Stein procedures to census data." **Journal of the American Statistical Association** 74:269-277.

Lux, Thomas. 2005. "Emergent statistical wealth distributions in simple monetary exchange models: a critical review". Pp. 51-60 in A. Chatterjee, S. Yarlagadda, and B.K. Chakrabarti, (eds.), **Econophysics of Wealth Distributions**, (the proceedings volume of the International Workshop on the Econophysics of Wealth Distributions, March, 2005, Kolkata, India.

U.S. Bureau of the Census. 2009. "A Compass for Understanding and Using American Community Survey Data: What PUMS Data Users Need to Know".