# JOURNAL OF INCOME DISTRIBUTION

AN INTERNATIONAL QUARTERLY

NEW SERIES

TORONTO YORK UNIVERSITY

VOLUME 18, NUMBER 2
JUNE 2009

# JOURNAL OF INCOME DISTRIBUTION



The purpose of the *Journal of Income Distribution* is to foster scholarly research internationally into all aspects of income distribution. The *Journal* supports theoretical, empirical, and technical studies pertaining to income distribution research and its methodologies. It also promotes analysis of the complex, subtle, and subjective principles which enter into questions of policy.

Communications and permission requests are to be sent to *Journal of Income Distribution*, Professor O.F. Hamouda, Editor, Glendon College, York University, York Hall 365, 2275 Bayview Avenue, Toronto, Ontario, Canada M4N 3M6 (telephone: 416-487-6778). Information is also available from the publisher, through the same address. All copy is subject to publisher's approval.

Volume 18, Number 2 (June 2009)

ISSN: 0926-6437
Printed in Canada, by Thistle Printing Limited



*Ad libros publications inc.* is the publisher of the *Journal of Income Distribution*. Business communications, subscription orders, and change-of-address requests (indication of old address as well as new address requested) are to be sent to *Ad libros publications inc., Journal of Income Distribution,* 235 Third Street, Cambridge, MA 02142. or by e-mail to jid@glendon.yorku.ca Advertising rates and information are available at either of the above addresses. Copyright is held by the Journal of Income Distribution. The periodical's mailing cost is paid in US or Canadian postage.

Subscription rates: Individuals: $204/1 year, $276/2 yrs; $316/3 yrs; Institutions: $412/1 year, $708/ 2 yrs; $940/3 yrs. For all subscriptions outside the US or Canada, add $50 per year for postage. Payment is to be made in US funds or equivalent. New Series current & back copies: Individuals $60; Institutions $95

# Maximum Entropy and the Entropy of Mixing for Income Distributions

G.R. Mohtashami Borzadaran
Ferdowsi University of Mashhad* [1]

Zahra Behdani
Statistical Research and Training Center, Tehran†

Over the last 100 years, a large number of distributions has been proposed for the modeling of size phenomena, notably the size distribution of personal incomes. The most widely known of these models are the Pareto, log-normal, generalized log-normal, Generalized Gamma, generalized Beta of the first and of the second kind, the Dagum, and the Singh-Madala distributions. They are discussed as a group in this note, as general forms of income distributions. Several well-known models are derived from them as sub-families with interesting applications in economics. The behaviour of their entropy is what is here under study. Maximum entropy formalism chooses certain forms of entropy and derives an exponential family of distributions under certain constraints. Finding constraints that income distributions have maximum entropy is another direction of this note. In economics and social statistics, the size distribution of income is the basis of concentration on the Lorenz curve. The difference between the tail of the Lorenz function and the Lorenz function itself determines the entropy of mixing. In the final section of this note, theoretical properties of well-known income distributions are also derived in view of the entropy of mixing.

*Keywords*: Maximum entropy, Lorenz curve, Lorenz order, income distributions, Generalized Beta distributions
*JEL Classifications*: D31, C02

**Introduction**

In June 1905, a paper written by Otto Lorenz truly revolutionized the economics and statistics of studying income distributions. Personal income distributions had

* *Address for Correspondence*: Department of Statistics and Ordered and Spatial Data Center of Excellence, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: gmb1334@yahoo.com

† *Address for Correspondence*: Department of Statistics, The University of Birjand, Birjand, Iran and Statistical Research and Training Center (SRTC), Tehran, Iran. E-mail: Zbehdani@yahoo.com

originated with Pareto's formulation of income distributions. Income and wealth are important in economics, and in this note we concentrate on some properties of income distributions (see for example Kleiber and Kotz, 2003). Empirical statisticians have studied income distributions as different notions, and in different regional, age, and educational groups, sometimes fitting them to analytical functions or suitable mixtures. Information theory is usually considered a product of the period of the Second World War. An original paper by Shannon was published in 1948, and new branch of science was born. The concept of entropy became 'expected information in distributions', and here we will concentrate on its connection to income distributions. Maximum entropy formalism specifically chooses a form of entropy and derives an exponential family of distributions by constraint. From Kagan, Linnik, and Rao (1973), any Maximization Entropy Probability Distribution (MEPD) is derived in a class of income distributions subject to certain constraints. In econophysics, Yakovenko (2003) has discussed aspects of the statistical mechanics of money, income and wealth. The size distribution of income has, however, in economics and social statistics, made the Lorenz curve the basis of concentration. While Lavenda (2005; 2006a) obtained statistical results deriving from mean entropy and the geometric entropy of mixing, respectively. By Lavenda (2006b) they also examined the entropy of mixing and its links with Lorenz curves, in connection with the limited applicability of Lorenz ordering.

In this note, a brief history and review of income distributions, and their characterizations by Shannon (1948), regarding entropy and univariate maximum entropy, are presented. For certain income distributions, there exist characterizations based on the maximization of entropy subject to certain constraints. The entropy of mixing can be obtained directly from the parent income distribution of extreme value theory, noting that the Lorenz curve that has a closed form implies a nice entropy of mixing. For others, which are not flexible enough for modeling, we can find the Lorenz curve and entropy of mixing of them numerically.

**Income Distributions**

In the inequality literature, a substantial number of formula are used to model various aspects of income distribution (see, for example, Kleiber and Kotz (2003)). Some of the most important ones include the following:

- The classic **Pareto** distribution, ParI, defined in terms of

$$F(x) = 1-(\frac{x}{x_0})^{-\alpha}, x > x_0 > 0, \tag{1}$$

where $\alpha > 0$ is shown as $X \sim \text{ParI}(x_0, \alpha)$. For a random sample of this distribution, we have $X_{(1)} \sim \text{ParI}((x_0, n\alpha)$ and $X_{(n)} \sim Stopa$ distribution. The second Pareto model, ParII, possesses the cummulative distribution function (cdf)

$$F(x) = 1 - (1 + \frac{x-\mu}{x_0})^{-\alpha}, x > \mu, \tag{2}$$

where $x_0, \alpha > 0$ is often referred to as $X \sim \text{Par(II)}(x_0, \alpha)$. Generalization of ParI, by introducing a power transformation of the distribution, is given by

$$F(x) = [1 - (\frac{x}{x_0})^{-\alpha}]^{\theta}, x > x_0 > 0, \tag{3}$$

as a **Stopa** distribution, where $\alpha, \theta > 0$ and $\theta = 1$ implies ParI.

- The distribution arising from the distribution of $X = e^Y$, where $Y \sim N(\mu, \sigma^2)$ is a **log-normal** distribution ($X \sim LN(\mu, \sigma^2)$), an income distribution with a cdf

$$F(x) = \Phi(\frac{\ln x - \mu}{\sigma}), x > 0, \tag{4}$$

where $\Phi$ denotes the cdf of the standard normal distribution. Also, $Y^r \sim LN(r\mu, r^2\sigma^2)$. The log-normal can be fitted to various income data. Let $Y$ be a generalized error distribution with a probability distribution function (pdf).

$$f(y) = \frac{1}{2r^{\frac{1}{r}}\sigma_r\Gamma(1+\frac{1}{r})} e^{-\frac{1}{2\sigma_r^r}|y-\mu|^r}, y \in \Re, \tag{5}$$

is, as an income distribution, called a **three parameter generalized log-normal**, where $\mu \in \Re, \sigma_r = (E|Y-\mu|^r)^{\frac{1}{r}}, r > 0$. The distribution can be considered a generalization of the log-normal distribution, which is obtained for $r = 2$, and also a generalization of the log-Laplace distribution, which is obtained for $r = 1$.

- The pdf of the **Generalized Gamma** (GG):

$$f(y/\alpha, \tau, \lambda) = \frac{\tau}{\lambda^{\alpha\tau}\Gamma(\alpha)} y^{\alpha\tau - 1} e^{-(\frac{y}{\lambda})^\tau}, y \geq 0, \alpha, \tau, \lambda > 0. \tag{6}$$

It includes several well-known models as sub-families of GG. So, $(\alpha = \tau = 1), (\tau = 1)$, and $(\alpha = 1)$ are the exponential, the Gamma and the Weibull distributions, respectively. Let $(\tau = 2)$, and we obtain a sub-family of Equation 6, which is known as the Generalized Normal distribution, GN$(2\alpha, \lambda)$, and $\alpha \to \infty$ leads us to a log-normal distribution. The GN sub-family includes the half-normal $(\alpha = \frac{1}{2})$, the Reyleigh $(\alpha = 1)$, the Maxwell-Boltzman $(\alpha = \frac{3}{2})$ and the Chi $(\alpha = \frac{k}{2}, k = 1, 2, \ldots)$ distributions.

- The pdf of **Generalized Beta, type II** [GBII$(a,b.p,q)$]:

$$f(x)=\frac{ax^{ap-1}}{b^{ap}B(p,q)}[1+(\frac{x}{b})^a]^{-(p+q)}, x>0, \tag{7}$$

where $a,b,p,q>0$ and $B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, such that the $\Gamma(.)$ is the Gamma function. The Sing-Maddala ($p=1$), the Beta of the second type ($a=1$), the Dagum ($q=1$), the Lomax (ParII) ($p=1, a=1$), the log-logistic($p=q=1$), and the inverse Lomax ($a=q=1$) distributions are all special cases of GBII$(a,b.p,q)$. The GG distribution emerges as a limiting case of GBII when $b=q^{\frac{1}{a}}\beta$ and $q\to\infty$, so, the Gamma and the Weibull distributions are also special cases of GBII. Also, if we put $1-(\frac{x}{b})^a$ in place of $1+(\frac{x}{b})^a$ in Equation 7, for $0<x<b$, then GBII becomes GBI$(a,b.p,q)$.

- The pdf of the form

$$f(x)=kx^{-a}[1-g(\frac{x}{b})^{\lambda}]^{q-1}[1-h(\frac{x}{c})^{\mu}]^{r}, 0<x<b, \tag{8}$$

which is called the **nine-parameter Generalized Beta distribution.** We apply specialized versions of the above, by taking values of $a,g,\lambda,b,q,c,h,\mu,r$, such as ($g=0$, and $q=\infty$), ($h=0$, and $r=\infty$), ($g=0, q=\infty, h=0$, and $r=\infty$), ($g=0$, and $h=1$), and ($g=1$, and $h=0$). The general case is not easy to work out. Also, ($g=0, h=1, a=1-a_2, r=1-a_3, c=a_2a_3$, and $e^{-\mu}=a_1$) gives the Singh-Maddala distribution of the parameters $a_1, a_2$, and $a_3$.

**Entropy and Maximum Entropy for Income Distributions**

Shannon entropy or, here, simply the entropy of $f$ relative to $\mu$, is defined by

$$H(f,\mu)=-\int_{\Omega} f\ln f d\mu, \text{ (with } f\ln f=0, \text{ if } f=0), \tag{9}$$

and is assumed to be defined for cases in which $f\ln f$ is integrable. Upon noting that $X$ is a random variable (rv) with pdf $f$, we then refer to $H$ as the entropy of $X$ and denote it also by the notation $H_X$ in place of $H(f,\mu)$. In the case where $\mu$ is a version of a counting measure, Equation 9 leads us to a specialized version of entropy of $X$, introduced by Shannon (1948) as $H_X=-\sum_{i=1}^{n} p_i\ln p_i$, where $p_i\geq 0$ and $\sum_{i=1}^{n} p_i=1$. The following theorem has given the general class of distributions for which Shannon entropy is maximized (Kagan, Linnik, and Rao, 1973).

**Theorem 1** *Let $X$ be a rv with density $f$ (with respect to (wrt) measure $\mu$) such that $f(x)>0$, for $x\in(a,b]$, and 0 elsewhere. Let further $h_1,h_2,...,h_n,..$ be integrable functions on $(a,b]$ satisfying the conditions $\int_{(a,b]} h_i(x)f(x)d\mu(x)=\lambda_i$ and*

*$i = 1, 2, \ldots$, with $\lambda_1, \lambda_2, \lambda_3, \ldots$ as constants. Then, maximum entropy is attained by the distribution with density $f$ of the form $f(x) = \exp\{c_0 + c_1 h_1(x) + c_2 h_2(x) + \ldots\}$, whenever there exist $c_0, c_1, \ldots$ so that the constraint is met.*

- The entropy of Equation 3 is calculated easily, and when $E(\ln \frac{X}{x_0}) = \frac{1}{\alpha}[\psi(\theta + 1) - \psi(1)]$, and $E(\ln[1 - (\frac{X}{x_0})^{-\alpha}]) = -\frac{1}{\theta}$, and $\psi(.) = \frac{\Gamma'(.)}{\Gamma(.)}$, for $x > 0$, the MEPD is obtained. For the case $\theta = 1$, ParI will be the MEPD. The entropy of log-normal distribution is equal to the entropy of normal distribution when the mean is equal to zero.
- Let $E(\ln X) = \ln b + \frac{1}{a}[\psi(p) - \psi(q)]$, $E(\ln[1 + (\frac{X}{b})^a]) = [\psi(p+q) - \psi(q)]$, and $\psi(.) = \frac{\Gamma'(.)}{\Gamma(.)}$, for $x > 0$. Then, under some conditions specified in the above theorem, Equation 7 maximizes entropy. $(p = 1)$, $(q = 1)$, and $(p = 1, q = 1)$ imply that BurrXII, BurrIII, and log-logistic distributions, respectively, are also MEPD. For a nine-parameter generalized Beta distribution and $\mathrm{GBI}(a, b, p, q)$, we can have similar characterizations.
- The maximization of second-order entropy was achieved by Kemp (1975), paralleling Kagan, Linnik, and Rao (1973). The probability distributions that are obtained by the maximum second-order entropy with the perscribed $E(X^{-k})$ are Pareto distributions.
- The entropy of Equation 6 is $H_X = \ln \lambda + \ln \Gamma(\alpha) + \alpha - \ln \tau + (\frac{1}{\tau} - \alpha)\psi(\alpha)$, namely, increasing in $\alpha$ and increasing in $\tau$ when $\alpha < 1.5$. It gives the expression of the entropy of the distributions that can be derived from GG distributions.
- For non-negative, continuous random variable $Y$, let $E(\ln Y) = \frac{1}{\lambda} + \frac{1}{\tau}\psi(\alpha)$ and $E(Y^\tau) = \frac{\lambda^\tau \Gamma(\alpha\tau + 1)}{\Gamma(\alpha)}$, then, the GG distribution is MEPD. $\tau = 1$ implies that the Pearson type-V distribution is MEPD, when the geometric and harmonic means are as prescribed. The inverse Gaussian distribution is MEPD, when the arithmetic, geometric and harmonic means are as prescribed.

### Entropy of Mixing in view of Lorenz Order

The Lorenz curve for general distribution, supported on a non-negative half line with finite and positive first moments, is $L(u) = \frac{1}{E(X)} \int_0^u F^{-1}(t)dt$, and $u \in [0, 1]$, where $F^{-1}(x) = \sup\{x/F(x) \leq t, t \in [0, 1]\}$, $E(X) = \int_0^1 F^{-1}(t)dt$, and $L$ is an increasing, convex, and continuous function on $[0, 1]$, with $L(0) = 0$ and $L(1) = 1$. The difference between the tail of the Lorenz function and the Lorenz function itself determines the entropy of mixing. This is used as a measure of dispersion of the distributions. So, $S(p) = \frac{1 - L(1-p) - L(p)}{1 - 2^{-D}}$, where $D$ is given in terms of the shape parameter of the income distribution. The Lorenz partial order, $\leq_L$ on Ł (Ł denotes the set of all non-negative random variables with finite and positive means), is defined by

$$X \leq_L Y \Longleftrightarrow L_X(u) \geq L_Y(u), \forall u \in [0,1].$$

Here, we are interested in the Lorenz ordering and the entropy of mixing order. In the cases in which the Lorenz curves intersect, we need to use the generalized Lorenz order. It is defined in terms of a generalized Lorenz curve, $GL_X(u)$, where

$$GL_X(p) = E(X)L_X(p) = \int_0^p F_X^{-1}(u)du, p \in [0,1].$$

Generalized Lorenz curves are non-decreasing, continuous, and convex, with $GL_X(0) = 0$ and $GL_X(1) = E(X)$. A distribution is uniquely determined by its generalized Lorenz curve. Hence, also,

$$S(p) = \frac{1 - \frac{GL(1-p)}{E(X)} - \frac{GL(p)}{E(X)}}{1 - 2^{-D}}.$$

If, however, $L_X(u) \geq L_Y(u), \forall u \in [0,1]$, then, $1 - L_X(1-u) \geq 1 - L_Y(1-u)$. Hence it implies that $S_X(u) \geq S_Y(u)$, so the Lorenz order leads to the order of $S(.)$. The Lorenz curve is only a partial order, so what does one do if two Lorenz curves intersect? The most widely used alternative to the Lorenz order is the generalized Lorenz order, of Shorrocks (1983). The generalized Lorenz partial order, $\leq_{GL}$ on Ł, is defined by:

$$X \leq_{GL} Y \Longleftrightarrow E(X)L_X(u) \leq E(Y)L_Y(u), \forall u \in [0,1].$$

The Generalized Lorenz order can also imply ordering due to $S(.)$.

- Most of the results that are achieved for distributions in view of the Lorenz order or the generalized Lorenz order can be applied for ordering due to $S(.)$. For example, consider rv $X_1$ and $X_2$, where $X_i \sim$GB2$(a_i, b_i, p_i, q_i), i = 1, 2$. Wilfling (1990, 1996) found that $X_1 \geq_L X_2 \Rightarrow S_{X_1}(.) \geq S_{X_2}(.), \Rightarrow a_1p_1 \leq a_2p_2, a_1q_1 \leq a_2q_2$, where $(a_1 \leq a_2, p_1 \leq p_2, q_1 \leq q_2)$. A complete characterization of the Lorenz ordering within the generalized Beta family of distributions includes some of the results due to Lorenz ordering, generalized Lorenz ordering, and ordering due to $S(.)$ for a big class of distributions.
- The Zenga curve provides a partial ordering between random variables. The Zenga partial order, $\leq_Z$ on Ł, is defined by $X \leq_Z Y \Longleftrightarrow Z_X(u) \geq Z_Y(u), \forall u \in [0,1]$, that can be connected to the ordering of $S(.)$ through its connection with Lorenz ordering.

- The Gini index, which derives from the Lorenz curve, is the most well-known and broadly used measure of inequality in the economics literature. The Gini index is given as twice the area between the Lorenz curve and equality line, $G = 1 - 2\int_0^1 L(u)du$. The Gini index has attractive theoretical and statistical properties that other inequality measures do not, which explains why it is used by most researchers.

- The extended Gini uses a parameter to emphasize various parts of the distribution. As a measure of inequality, the entropy of mixing is comparable with the Gini index. The Gini index is a global criterion of inequality; the entropy of mixing is a local one. Yet connection between them is achievable.

- For exponential distributions, the entropy of mixing is nested in those of the Pareto and the power-function entropy of mixing. The entropy of mixing of exponential distributions will be smaller than those of the Pareto and power-function distributions.

- The difference between the tail of the Lorenz function and the Lorenz function itself determines the entropy of mixing. It is comparable to the Gini index and a familiar concept in information theory.

- Entropy of mixing for families whose Lorenz cuvres are not an explicit function can be found through numerical arguments.

**Conclusions**

In this note, we reviewed and produced results for income distributions in view of entropy and maximum entropy. Also, we located the entropy of mixing used in the process of Lorenz functions. The classic Pareto and log-normal distributions do fit nicely as models into entropy formulation, but the more complicated distributions of income, which have been proposed, are fitted purely numerically. To extend analysis to a multivariate set-up for the many noted income distribution families, in view of entropy, maximum entropy, and entropy of mixing, could be one of the directions for the future of this work.

## Notes

1 *Acknowledgements*: This research was in part supported by a grant from STRC.

## References

Kagan, A.M., Y.V. Linnik, and C.R. Rao. 1973. *Characterization Problems in Mathematical Statistics*. New York, NY: Wiley.

Kemp, A.W. 1975. "Characterization based on second-order entropy", in *Statistical Distributions in Scientific Work*, Vol. 3, eds G.P. Patil, S. Kotz, and J.K. Ord. New York, NY: Springer Publishing Co, pp. 313-319.
Kleiber, C. and S. Kotz. 2003. *Statistical size distributions in economics and actuarial sciences*. New York, NY: John Wiley.
Lavenda, B.H. 2005. "Mean entropies", *Open Systems and Information Dynamics* 12: 289-302.
Lavenda, B.H. 2006a. "Geometric entropies of mixing", *Open Systems and Information Dynamics* 13: 91-96.
Lavenda, B.H. 2006b. "Entropies of mixing and the Lorenz order", *Open Systems and Information Dynamics* 13: 75-90.
Shannon, C.E. 1948. "A mathematical theory of communication", *Bell System Technical Journal* 27: 379-423.
Shorrocks, A.F. 1983. "Ranking income distributions", *Economica* 50: 3-17.
Wilfling, B. 1996. "A sufficient condition for Lorenz ordering", *The Indian Journal of Statistics* 58, B., 1: 62-69.
Wilfling, B. 1990. "Concepts of Lorenz-ordering with a special regard to the Lorenz ordering of order statistics", Diploma Thesis, Fakultät Statistik, Universität Dortmund, Germany.
Yakovenko, V.M. 2003. "Research in Econphysics", <Cond-math.stat-mech, [arxiv.cond-math/0302270v2]>