MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

A Note on Inequality Measures for Mixtures of Double Pareto-Lognormal Distributions

William Griffiths, Duangkamon Chotikapanich
and Gholamreza Hajargasht

October 2021

Working Paper 14/21

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

A Note on Inequality Measures for Mixtures of Double Pareto-Lognormal Distributions

William Griffiths
University of Melbourne

Duangkamon Chotikapanich*
Monash University

Gholamreza Hajargasht
University of Queensland

9 August, 2021

Abstract

Formulas are derived for the Gini, Theil and Pietra coefficients for a population-weighted mixture of double Pareto-lognormal distributions; they are applied to South America for two years. Results are also provided for the special case Pareto-lognormal and lognormal distributions. The formulas are useful for measuring regional or global inequality in large-scale projects that utilise double Pareto-lognormal distributions or their special cases.

Keywords: Gini coefficient; Theil index; Pietra index.

*Corresponding author
Duangkamon Chotikapanich
Department of Econometrics and Business Statistics
Monash University
Vic 3045
Australia
Duangkamon.chotikapanich@monash.edu
Phone +61 3 9903 2791

A Note on Inequality Measures for Mixtures of Double Pareto-Lognormal Distributions

1. Introduction

A vast number of distributions have been suggested in the literature for modelling income. An appreciation of the wide variety of distributions and their characteristics can be obtained from many sources; two comprehensive ones are Kleiber and Kotz (2003) and Chotikapanich (2008). We are concerned with two distributions that have received less attention than most, but which are nevertheless worthy of consideration: the double Pareto-lognormal distribution (dPLN) and its special case, the Pareto-lognormal distribution (PLN). The PLN, introduced by Colombi (1990), combined two features often considered desirable by income distribution modellers: a general lognormal shape and Pareto behaviour in the right tail. It was extended to the dPLN by Reed (2002, 2003) and Reed and Jorgenson (2004) to capture Pareto behaviour in both tails. They illustrated the relevance of the dPLN in a variety of applications including income distributions, size of human settlements, particle size, oil-field size, and stock price returns. Giesen et al. (2010) used it for modelling city size. Toda (2012) argued that, conditional on education and experience, income is distributed as a double Pareto distribution, and, unconditionally, it is dPLN. He also advocates using the dPLN to model a cross-section distribution for consumption (Toda, 2017). Concerned that the assumption of a constant elasticity inverse demand function relies on the possibly unrealistic assumption of a Pareto income distribution, Fabinger and Weyl (2018) explore the dPLN as an alternative. Hajargasht and Griffiths (2013) described how to estimate the PLN and dPLN using grouped data and provided expressions for calculating inequality and poverty measures from their parameters.

Our focus is on the use of the PLN and dPLN in large-scale projects designed to investigate regional or global income distributions, and regional or global inequality. In such studies, it is often convenient to estimate parametric income distributions for the component countries and to use a population-weighted mixture of those distributions to study the combined distribution and its characteristics. Inequality measures from the mixture distribution have the advantage of reflecting both within-country and between-country inequality. An example is the study by Chotikapanich et al., (2012) who estimated component-country beta-2 distributions to study global inequality. With their general lognormal shape and Pareto tail behaviour, the PLN and dPLN are also good candidates for estimating the component-country distributions. In a study of 10 separate countries, Hajargasht and Griffiths (2013) compared results from estimating the PLN, dPLN and generalized beta distributions. There was no clear ranking in terms of goodness-of-fit; the preferred distribution depended on what country was chosen and the goodness-of-fit measure that was used. When pursuing estimates of regional and global inequality using large scale projects involving an uncertain choice of the best parametric distribution, it is prudent to investigate the sensitivity of inequality estimates to alternative distributional assumptions. The purpose of this note is to provide expressions that can be used to compute regional or global inequality for a mixture distribution whose components are PLN or dPLN. Formulas are provided for the Gini, Theil and Pietra coefficients as functions of the parameters of the PLN and dPLN components. For completeness, we also report expressions for the special-case lognormal (LN) distribution.

Section 2 is devoted to specification of the density and distribution functions for the dPLN and PLN. Expressions for the Gini, Theil and Pietra coefficients as functions of the parameters are provided in Sections 3, 4, and 5, respectively. A proof for the Gini coefficient is given in a supplementary material. In Section 6 we calculate inequality for South America as an illustrative example, comparing estimates from the PLN, dPLN and LN distributions.

2. The double Pareto-lognormal distribution

A dPLN random variable is distributed as the product of independent lognormal and double Pareto random variables. Its density function for a random variable $y > 0$ is

$$f(y)=\frac{\alpha\beta}{(\alpha+\beta)y}\left[\frac{1}{y^{\alpha}}\exp\left\{\frac{\alpha^2\sigma^2}{2}+\alpha m\right\}\Phi\left(\frac{\ln y-m-\alpha\sigma^2}{\sigma}\right)\right.$$
$$\left.+y^{\beta}\exp\left\{\frac{\beta^2\sigma^2}{2}-\beta m\right\}\Phi\left(\frac{-\ln y+m-\beta\sigma^2}{\sigma}\right)\right] \quad (1)$$

with parameters $\alpha>0, \beta>0, -\infty<m<\infty,$ and $\sigma>0;$ $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (cdf). For $\alpha>1$, its mean is given by

$$\mu=\frac{\alpha\beta}{(\alpha-1)(\beta+1)}\exp\left\{\frac{\sigma^2}{2}+m\right\} \quad (2)$$

Its cdf is obtained by setting $k=0$ in equation (3) that follows. Its first moment distribution function, defined as $F^{(1)}(y)=\int_0^y t f(t)dt\big/\mu$, is obtained by setting $k=1$ in equation (3); $\alpha>1$ is required for its existence.

$$F^{(k)}(y)=\Phi\left(\frac{\ln y-m-k\sigma^2}{\sigma}\right)-\frac{\beta+k}{\alpha+\beta}y^{k-\alpha}\exp\left\{\frac{\sigma^2}{2}(\alpha^2-k^2)+m(\alpha-k)\right\}\Phi\left(\frac{\ln y-m-\alpha\sigma^2}{\sigma}\right)$$
$$-\frac{\alpha-k}{\alpha+\beta}y^{\beta+k}\exp\left\{\frac{\sigma^2}{2}(\beta^2-k^2)-m(\beta+k)\right\}\Phi\left(\frac{-\ln y+m-\beta\sigma^2}{\sigma}\right) \quad (3)$$

Equations (1)-(3) represent the essential features of the distribution necessary for deriving the inequality measures for its mixtures. More details of its properties can be found in Reed and Jorgensen (2004) and Hajargasht and Griffiths (2013).

The PLN distribution is a special case of the dPLN, obtained by letting $\beta\to\infty.$ Its density function is given by

$$f(y)=\frac{\alpha}{y^{\alpha+1}}\exp\left\{\frac{\alpha^2\sigma^2}{2}+\alpha m\right\}\Phi\left(\frac{\ln y-m-\alpha\sigma^2}{\sigma}\right) \quad (4)$$

Its cdf and first moment distribution function are obtained by setting $k=0$ and $k=1$, respectively, in the following equation.

$$F^{(k)}(y)=\Phi\left(\frac{\ln y-m-k\sigma^2}{\sigma}\right)-y^{k-\alpha}\exp\left\{\frac{\sigma^2}{2}(\alpha^2-k^2)+m(\alpha-k)\right\}\Phi\left(\frac{\ln y-m-\alpha\sigma^2}{\sigma}\right) \quad (5)$$

For $\alpha>1$, its mean is

$$\mu=\frac{\alpha}{\alpha-1}\exp\left\{\frac{\sigma^2}{2}+m\right\} \quad (6)$$

The familiar lognormal distribution is obtained by letting $\alpha\to\infty$.

Now, suppose we have *n* distributions indexed by *i* (and, later, also by *j*), and that the population share of the *i*-th country is λ_i. The density and distribution functions of their population-weighted mixture are given by

$$f(y)=\sum_{i=1}^{n}\lambda_i f_i(y) \qquad F(y)=\sum_{i=1}^{n}\lambda_i F_i(y) \qquad \text{and} \qquad F^{(1)}(y)=\frac{1}{\mu}\sum_{i=1}^{n}\lambda_i\mu_i F_i^{(1)}(y)$$

where we now denote $\mu=\sum_{i=1}^{n}\lambda_i\mu_i$ as the mean of the mixture distribution, and μ_i as the mean of the *i*-th component. Similarly, the parameters of the *i*-th component will be indexed as $(\alpha_i,\beta_i,m_i,\sigma_i)$.

3. Gini coefficient

A general expression for the Gini coefficient is

$$G=-1+\frac{2}{\mu}\int_0^\infty y\,f(y)F(y)dy \quad (7)$$

For the dPLN mixture, we show that it is given by

$$G^{\text{dPLN}}=-1+\frac{2}{\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\left(P_{ij}+q_{ij}Q_{ij}-r_{ij}R_{ij}\right) \quad (8)$$

where

$$P_{ij} = \mu_i \Phi\left(\frac{m_i - m_j + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right) \tag{9}$$

$$Q_{ij} = \mu_j \exp\left\{\alpha_i (m_i - m_j - \sigma_j^2) + \frac{\alpha_i^2}{2}(\sigma_i^2 + \sigma_j^2)\right\} \Phi\left(\frac{m_j - m_i - \alpha_i(\sigma_i^2 + \sigma_j^2) + \sigma_j^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right) \tag{10}$$

$$R_{ij} = \mu_i \exp\left\{\beta_j (m_i - m_j + \sigma_i^2) + \frac{\beta_j^2}{2}(\sigma_i^2 + \sigma_j^2)\right\} \Phi\left(\frac{m_j - m_i - \beta_j(\sigma_i^2 + \sigma_j^2) - \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right) \tag{11}$$

$$\begin{aligned} q_{ij} = {} & \frac{\alpha_i \beta_i (\alpha_j - 1)(\beta_j + 1)}{\alpha_j \beta_j (\alpha_i + \beta_i)(\alpha_j + \beta_j)} \left[\frac{\alpha_j + \beta_j}{\alpha_i - 1} - \frac{\beta_j}{\alpha_i + \alpha_j - 1} + \frac{\alpha_j}{\alpha_i - \beta_j - 1}\right] \\ & + \frac{\beta_i (\alpha_j - 1)(\beta_j + 1)}{(\alpha_i + \beta_i)(\alpha_i - \beta_j - 1)(\alpha_i + \alpha_j - 1)} \end{aligned} \tag{12}$$

and

$$\begin{aligned} r_{ij} = {} & \frac{\alpha_j (\alpha_i - 1)(\beta_i + 1)}{(\alpha_j + \beta_j)(\alpha_i - \beta_j - 1)(\beta_i + \beta_j + 1)} \\ & + \frac{\alpha_j \beta_j (\alpha_i - 1)(\beta_i + 1)}{\alpha_i \beta_i (\alpha_i + \beta_i)(\alpha_j + \beta_j)} \left[\frac{\alpha_i + \beta_i}{\beta_j + 1} + \frac{\beta_i}{\alpha_i - \beta_j - 1} + \frac{\alpha_i}{\beta_i + \beta_j + 1}\right] \end{aligned} \tag{13}$$

Proof: See the supplementary material.

For the PLN distribution this expression reduces to

$$G^{\mathrm{PLN}} = -1 + \frac{2}{\mu} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j (P_{ij} + e_{ij} Q_{ij}) \tag{14}$$

where

$$e_{ij} = \frac{\alpha_j - 1}{(\alpha_i - 1)(\alpha_i + \alpha_j - 1)} \tag{15}$$

For the lognormal distribution, we have

$$G^{\mathrm{LN}} = -1 + \frac{2}{\mu} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j P_{ij} \tag{16}$$

4. Theil indices

Theil (1967) suggested two possible measures of inequality. To derive them, we use Lemma 1 from Sarabia et al. (2017); it provides a convenient device for obtaining the expectations $E\left[\ln y\right]$ and $E\left[y\ln y\right]$. Denoting the Theil indices by T_0 and T_1, we can show that

$$\begin{aligned} T_0^{\text{dPLN}} &= \int_0^\infty \ln\left(\frac{\mu}{y}\right) f(y)dy = \ln\mu - \sum_{i=1}^n \lambda_i \int_0^\infty \ln y\, f_i(y)dy \\ &= \ln\mu - \sum_{i=1}^n \lambda_i \left(m_i + 1/\alpha_i - 1/\beta_i\right) \end{aligned} \tag{17}$$

$$T_0^{\text{PLN}} = \ln\mu - \sum_{i=1}^n \lambda_i \left(m_i + 1/\alpha_i\right) \tag{18}$$

$$T_0^{\text{LN}} = \ln\mu - \sum_{i=1}^n \lambda_i m_i \tag{19}$$

$$\begin{aligned} T_1^{\text{dPLN}} &= \int_0^\infty \left(\frac{y}{\mu}\right)\ln\left(\frac{y}{\mu}\right) f(y)dy = \frac{1}{\mu}\sum_{i=1}^n \lambda_i \int_0^\infty y\ln y\, f_i(y)dy - \ln\mu \\ &= \frac{1}{\mu}\sum_{i=1}^n \lambda_i \mu_i \left(m_i + \sigma_i^2 + \frac{1}{\alpha_i - 1} - \frac{1}{\beta_i + 1}\right) - \ln\mu \end{aligned} \tag{20}$$

$$T_1^{\text{PLN}} = \frac{1}{\mu}\sum_{i=1}^n \lambda_i \mu_i \left(m_i + \sigma_i^2 + \frac{1}{\alpha_i - 1}\right) - \ln\mu \tag{21}$$

$$T_1^{\text{LN}} = \frac{1}{\mu}\sum_{i=1}^n \lambda_i \mu_i \left(m_i + \sigma_i^2\right) - \ln\mu \tag{22}$$

5. Pietra index

The Pietra index which is equal to the maximum distance between the Lorenz curve and the equality line is given by

$$\begin{aligned} P &= \frac{1}{2\mu}\int_0^\infty \left|y-\mu\right| f(y)dy = \int_0^\mu f(y)dy - \frac{1}{\mu}\int_0^\mu y f(y)dy \\ &= F(\mu) - F^{(1)}(\mu) \end{aligned} \tag{23}$$

The Pietra index can be written as the difference between the cdf and the first moment distribution function, both evaluated at μ. Replacing y with μ in the equations for these functions specified in equation (3), subtracting, and simplifying, yields

$$\begin{aligned} P^{\text{dPLN}} &= \sum_{i=1}^{n} \lambda_i \Phi\left(\frac{\ln\mu - m_i}{\sigma_i}\right) - \frac{1}{\mu}\sum_{i=1}^{n}\lambda_i\mu_i\Phi\left(\frac{\ln\mu - m_i - \sigma_i^2}{\sigma_i}\right) \\ &+ \sum_{i=1}^{n}\frac{\lambda_i\beta_i}{(\alpha_i+\beta_i)(\alpha_i-1)\mu^{\alpha_i}}\exp\left\{\frac{\alpha_i^2\sigma_i^2}{2}+\alpha_i m_i\right\}\Phi\left(\frac{\ln\mu - m_i - \alpha_i\sigma_i^2}{\sigma_i}\right) \\ &- \sum_{i=1}^{n}\frac{\lambda_i\alpha_i\mu^{\beta_i}}{(\alpha_i+\beta_i)(\beta_i+1)}\exp\left\{\frac{\beta_i^2\sigma_i^2}{2}-\beta_i m_i\right\}\Phi\left(\frac{-\ln\mu + m_i - \beta_i\sigma_i^2}{\sigma_i}\right) \end{aligned} \tag{24}$$

For the PLN distribution, this expression simplifies to

$$\begin{aligned} P^{\text{PLN}} &= \sum_{i=1}^{n} \lambda_i \Phi\left(\frac{\ln\mu - m_i}{\sigma_i}\right) - \frac{1}{\mu}\sum_{i=1}^{n}\lambda_i\mu_i\Phi\left(\frac{\ln\mu - m_i - \sigma_i^2}{\sigma_i}\right) \\ &+ \sum_{i=1}^{n}\frac{\lambda_i}{(\alpha_i-1)\mu^{\alpha_i}}\exp\left\{\frac{\alpha_i^2\sigma_i^2}{2}+\alpha_i m_i\right\}\Phi\left(\frac{\ln\mu - m_i - \alpha_i\sigma_i^2}{\sigma_i}\right) \end{aligned} \tag{25}$$

For the LN distribution, it simplifies further to

$$P^{\text{LN}} = \sum_{i=1}^{n} \lambda_i \Phi\left(\frac{\ln\mu - m_i}{\sigma_i}\right) - \frac{1}{\mu}\sum_{i=1}^{n}\lambda_i\mu_i\Phi\left(\frac{\ln\mu - m_i - \sigma_i^2}{\sigma_i}\right) \tag{26}$$

6. Example

To illustrate computations from the above results, we present estimates of inequality for South America and its component countries for the years 2006 and 2012. We have omitted Guyana, Suriname and French Guiana whose data were unavailable or incomplete. Their combined population is less than 0.4% of the total population of South America. Formulas for the inequality measures for the PLN and dPLN components can be found in Hajargasht and Griffiths (2013); those for the lognormal distribution can be found in Kleiber and Kotz (2003). The parameter estimates are drawn from a subset of distributions estimated as a part of a larger project on the construction of a global panel of income distributions.[1] The data were expressed in terms of common currency units using the purchasing power parities from the UQICD website.[2] Table 1 contains the population proportions that make up the weights in the mixture and the annual mean incomes estimated using parameter estimates from each of the distributions. Brazil is the dominant country, comprising half of the total population of South America. Chile and Uruguay have the highest per capita incomes. Most of the mean income estimates are robust with respect to the choice of distribution; Chile is an obvious exception.

Inequality estimates for each of the countries and their South American mixture are presented in Table 2. We note the following:

1. The income shares used to estimate the distributions for Argentina and Venezuela were the same in both years, leading to identical inequality estimates in both years.[3]
2. Examining the sensitivity of the inequality estimates with respect to choice of distribution, we find the results are mixed. Some estimates are relative consistent, for example Colombia and Ecuador, whereas others, such as Chile and Uruguay show much more variation.
3. One might expect the tails of the PLN and dPLN distributions to lead to higher inequality estimates than those obtained using the lognormal distribution. While that is often the case, it is not universally true. A larger lognormal estimate for σ can compensate for the existence of the Pareto tails.
4. All distributions and all inequality measures suggest there has been a reduction in inequality from 2006 to 2012, in all countries except Argentina, Venezuela, Brazil and

[1] The parameter estimates are available from the authors on request.

[2] uqicd.economics.uq.edu.au.

[3] Data limitations meant that some income shares had to be interpolated or extrapolated using data from available years. This led to the identical income shares for Argentina and Venezuela.

Paraguay. As explained in point (1), the inequality estimates were the same in both years for Argentina and Venezuela. For Brazil, the PLN distribution suggested an increase in inequality; for Paraguay, the PLN and dPLN distributions suggested an increase, and the lognormal distribution, a decrease.

5. Except for the PLN estimates for Theil-1, all distributions and all inequality measures suggest inequality in South America as a whole has declined from 2006 to 2012.

References

Chotikapanich, D., W.E. Griffiths, G. Hajargasht, W. Karunarathne & D.S. Prasada Rao (2018) Using the GB2 income distribution. *Econometrics* 6, 1-24.

Chotikapanich, D., W.E. Griffiths, W., D.S. Prasada Rao & V. Valencia (2012) Global income distributions and inequality, 1993 and 2000: incorporating country-level inequality modelled with beta distributions. *Review of Economics and Statistics* 94, 52-73.

Colombi, R. (1990) A new model of income distribution: The Pareto-lognormal distribution. In C. Dagum & M. Zenga (Eds.), *Income and Wealth Distribution, Inequality and Poverty*. Springer, 18-32.

Fabinger, M. and E.G. Weyl (2018) Functional forms for tractable economic models and the cost structure of international trade. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2194855

Gupta, S. & S. Pillai (1965) On linear functions of ordered correlated normal random variables. *Biometrika* 52, 367-379.

Giesen, K., A. Zimmermann & J. Suedekum (2010) The size distribution across all cities – double Pareto lognormal strikes. *Journal of Urban Economics* 68, 129-137.

Hajargasht, G. & W.E. Griffiths (2013) Pareto-lognormal distributions: inequality, poverty, and estimation from grouped income data. *Economic Modelling* 33, 593-604.

Kleiber, C. & S. Kotz (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*. John Wiley & Sons.

Reed, W.J. (2002) On the rank-size distribution for human settlements. *Journal of Regional Science* 42, 1-17.

Reed, W.J. (2003) The Pareto law of incomes – an explanation and extension. *Physica A* 319, 469-486.

Reed, W.J. & M. Jorgensen (2004) The double Pareto-lognormal distribution – a new parametric model for size distributions. *Communications in Statistics Theory and Methods* 33, 1733-1753.

Sarabia, J.M., V. Jordá & L. Remuzgo (2017) The Theil indices in parametric families of income distributions—a short review. *Review of Income and Wealth* 63, 867–80.

Theil, H. (1967) *Economics and Information Theory*. North Holland.

Toda, A.A. (2012) The double power law in income distribution: explanations and evidence. *Journal of Economic Behaviour & Organization* 84, 364-381.

Toda, A.A. (2017) A note on the size distribution of consumption: more double Pareto than lognormal. *Macroeconomic Dynamics* 21, 1508-1518.

Table 1 Population weights and mean incomes

		Argentina	Bolivia	Brazil	Chile	Colombia	Ecuador	Paraguay	Peru	Uruguay	Venezuela	Overall
Pop (%)												
2006		10.49	2.51	50.18	4.38	11.53	3.77	1.58	7.50	0.89	7.18	100
2012		10.48	2.61	49.89	4.36	11.56	3.89	1.61	7.40	0.85	7.37	100
Means												
2006	dPLN	7219	2962	5322	9794	4813	4712	2780	4514	8236	6205	5565
	PLN	7291	2708	5363	9145	4934	4724	2785	4597	7750	5810	5546
	LN	7168	2720	5552	7918	5053	4997	2971	4571	7333	5863	5599
2012	dPLN	10673	3921	8862	15118	7128	6724	4797	7462	14679	10019	8878
	PLN	10770	3804	9582	15182	7369	6810	4716	7350	13555	11019	9333
	LN	10588	3793	9240	13164	7423	6801	4584	7340	13536	9472	8944

Table 2 Inequality measures for South America and its component countries

		Argentina	Bolivia	Brazil	Chile	Colombia	Ecuador	Paraguay	Peru	Uruguay	Venezuela	Overall
Gini												
2006	dPLN	0.470	0.642	0.528	0.583	0.513	0.489	0.491	0.510	0.5158	0.508	0.520
	PLN	0.466	0.606	0.529	0.553	0.524	0.489	0.490	0.517	0.487	0.540	0.531
	LN	0.454	0.605	0.539	0.470	0.526	0.507	0.511	0.513	0.451	0.457	0.523
2012	dPLN	0.470	0.532	0.506	0.539	0.511	0.465	0.508	0.469	0.464	0.508	0.499
	PLN	0.466	0.523	0.538	0.541	0.515	0.469	0.494	0.461	0.419	0.540	0.529
	LN	0.454	0.522	0.518	0.456	0.522	0.468	0.476	0.460	0.418	0.457	0.506
Theil0												
2006	dPLN	0.383	0.797	0.515	0.602	0.480	0.431	0.434	0.476	0.462	0.457	0.521
	PLN	0.381	0.726	0.520	0.538	0.501	0.432	0.434	0.490	0.413	0.546	0.525
	LN	0.364	0.724	0.544	0.395	0.513	0.470	0.478	0.483	0.359	0.369	0.514
2012	dPLN	0.383	0.525	0.466	0.506	0.477	0.384	0.453	0.392	0.371	0.457	0.477
	PLN	0.381	0.505	0.527	0.510	0.487	0.391	0.433	0.378	0.305	0.546	0.513
	LN	0.364	0.503	0.494	0.368	0.504	0.390	0.405	0.376	0.303	0.369	0.473
Theil1												
2006	dPLN	0.445	1.029	0.519	1.009	0.486	0.434	0.438	0.479	0.606	0.522	0.571
	PLN	0.403	0.729	0.522	0.806	0.525	0.434	0.436	0.500	0.487	0.547	0.547
	LN	0.364	0.724	0.544	0.395	0.513	0.470	0.478	0.483	0.359	0.369	0.503
2012	dPLN	0.445	0.528	0.470	0.773	0.480	0.388	0.553	0.395	0.440	0.522	0.513
	PLN	0.403	0.507	0.580	0.783	0.488	0.393	0.474	0.379	0.306	0.547	0.564
	LN	0.364	0.503	0.494	0.368	0.504	0.390	0.405	0.376	0.303	0.369	0.467
Pietra												
2006	dPLN	0.340	0.481	0.388	0.433	0.376	0.357	0.359	0.373	0.375	0.370	0.387
	PLN	0.339	0.453	0.390	0.407	0.385	0.358	0.359	0.380	0.354	0.399	0.390
	LN	0.330	0.453	0.398	0.343	0.387	0.372	0.375	0.377	0.328	0.333	0.385
2012	dPLN	0.340	0.391	0.371	0.396	0.375	0.338	0.370	0.341	0.336	0.370	0.373
	PLN	0.339	0.385	0.396	0.398	0.378	0.342	0.361	0.336	0.304	0.399	0.388
	LN	0.330	0.384	0.381	0.332	0.384	0.341	0.347	0.336	0.303	0.333	0.371

Supplementary Material for

A Note on Inequality Measures for Mixtures of Double Pareto-Lognormal Distributions

William Griffiths*
University of Melbourne

Duangkamon Chotikapanich
Monash University

Gholamreza Hajargasht
University of Queensland

Derivation of Gini coefficient for a mixture of double Pareto-lognormal distributions

Supplementary Material for

A Note on Inequality Measures for Mixtures of Double Pareto-Lognormal Distributions

Derivation of Gini coefficient for a mixture of

Deriving the Gini coefficient for the mixture of double-Pareto lognormal distributions involves integrals of products of normal cumulative distribution functions. There are three distinct cases; the results for these cases are provided in the following lemma.

***Lemma**:*

We assume $\sigma_1 > 0, \sigma_2 > 0$ and define $B_i = \exp\left\{ab_i + a^2\sigma_i^2/2\right\}, i = 1, 2.$ Then,

(i) For $a < 0$,

$$\int_0^\infty y^{a-1}\Phi\left(\frac{\ln y - b_1}{\sigma_1}\right)\Phi\left(\frac{\ln y - b_2}{\sigma_2}\right)dy = -\frac{1}{a}\left[B_1\Phi\left(\frac{b_1 - b_2 + a\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right) + B_2\Phi\left(\frac{b_2 - b_1 + a\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)\right]$$

(ii) For $-\infty < a < \infty$,

$$\int_0^\infty y^{a-1}\Phi\left(\frac{\ln y - b_1}{\sigma_1}\right)\Phi\left(\frac{-\ln y + b_2}{\sigma_2}\right)dy = -\frac{1}{a}\left[B_1\Phi\left(\frac{b_2 - b_1 - a\sigma_1^2}{\sqrt{\sigma_i^2 + \sigma_2^2}}\right) - B_2\Phi\left(\frac{b_2 - b_1 + a\sigma_2^2}{\sqrt{\sigma_i^2 + \sigma_2^2}}\right)\right]$$

(iii) For $a > 0$,

$$\int_0^\infty y^{a-1}\Phi\left(\frac{-\ln y + b_1}{\sigma_1}\right)\Phi\left(\frac{-\ln y + b_2}{\sigma_2}\right)dy = \frac{1}{a}\left[B_1\Phi\left(\frac{b_2 - b_1 - a\sigma_1^2}{\sqrt{\sigma_i^2 + \sigma_2^2}}\right) + B_2\Phi\left(\frac{b_2 - b_1 - a\sigma_2^2}{\sqrt{\sigma_i^2 + \sigma_2^2}}\right)\right]$$

Proof:

We provide the proof for (i). Proofs for (ii) and (iii) follow similar lines and can be obtained from the authors upon request.

Let $u = (\ln y - b_1)/\sigma_1$, so that $\ln y = b_1 + \sigma_1 u,\ \ y = \exp\left\{b_1 + \sigma_1 u\right\},\ \ dy = \sigma_1 \exp\left\{b_1 + \sigma_1 u\right\} du,$ and

$$\frac{\ln y - b_2}{\sigma_2} = \frac{\delta + \sigma_1 u}{\sigma_2}$$

where $\delta = b_1 - b_2$. Then,

$$\begin{aligned} I &= \int_0^\infty y^{a-1}\Phi\left(\frac{\ln y - b_1}{\sigma_1}\right)\Phi\left(\frac{\ln y - b_2}{\sigma_2}\right)dy \\ &= \sigma_1 \exp\left\{ab_1\right\}\int_{-\infty}^\infty \exp\left\{a\sigma_1 u\right\}\Phi(u)\,\Phi\left(\frac{\delta + \sigma_1 u}{\sigma_2}\right)du \\ &= \sigma_1 \exp\left\{ab_1\right\} I_0 \end{aligned}$$

To evaluate I_0, we use integration by parts. Let

$$X = \Phi(u)\,\Phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)$$

and $dW = \exp\{a\sigma_1 u\}\,du$. Then,

$$I_0 = XW\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} W\,dX$$

Consider

$$XW = \frac{\Phi(u)\Phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)}{\exp\{-a\sigma_1 u\}/a\sigma_1}$$

Given $a < 0$, it is clear that

$$\lim_{u\to\infty} XW = 0$$

For $u \to -\infty$, the limits of the numerator and denominator of XW both approach zero. Using l'Hôpital's rule, differentiating will lead to terms in the numerator which involve the normal density function; the term in u in the denominator will be unchanged. The normal density function contains terms such as $\exp\{-u^2/2\}$ which approach zero as $u \to -\infty$. Moreover, they approach zero at a faster rate than that for the term in the denominator. Thus,

$$\lim_{u\to-\infty} XW = 0$$

and

$$\begin{aligned}
I_0 &= -\int_0^{\infty} W\,dX \\
&= -\frac{1}{a\sigma_1}\int_{-\infty}^{\infty} \exp\{a\sigma_1 u\}\left[\Phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)\phi(u) + \Phi(u)\phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)\frac{\sigma_1}{\sigma_2}\right]du \\
&= -\frac{1}{a\sigma_1}\left[A_1 + A_2\right]
\end{aligned}$$

where

$$A_1 = \int_{-\infty}^{\infty} \exp\{a\sigma_1 u\}\,\Phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)\phi(u)\,du$$

$$A_2 = \int_{-\infty}^{\infty} \exp\{a\sigma_1 u\}\Phi(u)\phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)\frac{\sigma_1}{\sigma_2}du$$

Consider A_1, and focus on the terms in the exponent of $\exp\{a\sigma_1 u\}\phi(u)$, namely

$$a\sigma_1 u - \frac{u^2}{2} = -\frac{1}{2}(u - a\sigma_1)^2 + \frac{a\sigma_1^2}{2}$$

Then,

$$A_1 = \exp\left\{\frac{a^2\sigma_1^2}{2}\right\}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(u-a\sigma_1)^2\right\}\Phi\left(\frac{\delta+\sigma_1 u}{\sigma_2}\right)du$$

Let $v = u - a\sigma_1$, so that $u = v + a\sigma_1$ and $du = dv$. Then,

$$A_1 = \exp\left\{\frac{a^2\sigma_1^2}{2}\right\}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{v^2}{2}\right\}\Phi\left(\frac{\delta+a\sigma_1^2+\sigma_1 v}{\sigma_2}\right)dv$$

$$= \exp\left\{\frac{a^2\sigma_1^2}{2}\right\}\Phi\left(\frac{\delta+a\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)$$

The last equality follows from the following result provided by Gupta and Pillai (1965).

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\}\Phi(cz+d)dz = \Phi\left(\frac{d}{\sqrt{1+c^2}}\right)$$

Now, consider A_2, and focus on the terms in the exponent of $\exp\{a\sigma_1 u\}\phi\left[(\delta+\sigma_1 u)/\sigma_2\right]$, namely

$$a\sigma_1 u - \frac{1}{2\sigma_2^2}(\delta+\sigma_1 u)^2 = -\frac{\sigma_1^2}{2\sigma_2^2}\left(u + \frac{\delta - a\sigma_2^2}{\sigma_1}\right)^2 + \frac{a^2\sigma_2^2}{2} - a\delta$$

Then,

$$A_2 = \exp\left\{\frac{a^2\sigma_2^2}{2} - a\delta\right\}\frac{\sigma_1}{\sigma_2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\sigma_1^2}{2\sigma_2^2}\left(u+\frac{\delta-a\sigma_2^2}{\sigma_1}\right)^2\right\}\Phi(u)du$$

Let

$$v = \frac{\sigma_1}{\sigma_2}\left(u + \frac{\delta - a\sigma_2^2}{\sigma_1}\right)$$

so that

$$u = \frac{\sigma_2}{\sigma_1}v - \frac{\delta - a\sigma_2^2}{\sigma_1}$$

and

$$du = \frac{\sigma_2}{\sigma_1} dv$$

Then,

$$A_2 = \exp\left\{\frac{a^2\sigma_2^2}{2} - a\delta\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{v^2}{2}\right\} \Phi\left(\frac{-\delta + a\sigma_2^2 + \sigma_2 v}{\sigma_1}\right) dv$$

Using again the result from Gupta and Pillai (1965),

$$A_2 = \exp\left\{\frac{a^2\sigma_2^2}{2} - a\delta\right\} \Phi\left(\frac{-\delta + a\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$$

Collecting all the terms, we have

$$\begin{aligned}
I &= \sigma_1 \exp\{ab_1\} I_0 \\
&= \sigma_1 \exp\{ab_1\} \left[-\frac{1}{a\sigma_1}(A_1 + A_2)\right] \\
&= -\frac{1}{a}\left[\exp\left\{ab_1 + \frac{a^2\sigma_1^2}{2}\right\}\Phi\left(\frac{\delta + a\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) + \exp\left\{ab_1 + \frac{a^2\sigma_2^2}{2} - a\delta\right\}\Phi\left(\frac{-\delta + a\sigma_2^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\right] \\
&= -\frac{1}{a}\left[\exp\left\{\frac{a^2\sigma_1^2}{2} + ab_1\right\}\Phi\left(\frac{b_1 - b_2 + a\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right) + \exp\left\{\frac{a^2\sigma_2^2}{2} + ab_2\right\}\Phi\left(\frac{b_2 - b_1 + a\sigma_2^2}{\sqrt{\sigma_1^2+\sigma_2^2}}\right)\right]
\end{aligned}$$

This completes the proof of the lemma.

Derivation of the Gini coefficient:

The Gini coefficient for the mixture is given by

$$\begin{aligned}
G &= -1 + \frac{2}{\mu}\int_0^{\infty} yf(y)F(y)dy \\
&= -1 + \frac{2}{\mu}\int_0^{\infty} y\sum_{i=1}^{n}\lambda_i f_i(y)\sum_{j=1}^{n}\lambda_j F_j(y)dy \\
&= -1 + \frac{2}{\mu}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j\int_0^{\infty} yf_i(y)F_j(y)dy
\end{aligned}$$

Let

$$I = \int_0^\infty y f_i(y) F_j(y) dy$$

$$= \int_0^\infty y\left(\frac{\alpha_i\beta_i}{(\alpha_i+\beta_i)y}\right)\left[\frac{1}{y^{\alpha_i}}\exp\left\{\frac{\alpha_i^2\sigma_i^2}{2}+\alpha_i m_i\right\}\Phi\left(\frac{\ln y - m_i - \alpha_i\sigma_i^2}{\sigma_i}\right)\right.$$
$$\left. + y^{\beta_i}\exp\left\{\frac{\beta_i^2\sigma_i^2}{2}-\beta_i m_i\right\}\Phi\left(\frac{-\ln y + m_i - \beta_i\sigma_i^2}{\sigma_i}\right)\right]$$
$$\times\left[\Phi\left(\frac{\ln y - m_j}{\sigma_j}\right) - \left(\frac{\beta_j}{\alpha_j+\beta_j}\right)\frac{1}{y^{\alpha_j}}\exp\left\{\frac{\alpha_j^2\sigma_j^2}{2}+\alpha_j m_j\right\}\Phi\left(\frac{\ln y - m_j - \alpha_j\sigma_j^2}{\sigma_j}\right)\right.$$
$$\left. - \left(\frac{\alpha_j}{\alpha_j+\beta_j}\right) y^{\beta_j}\exp\left\{\frac{\beta_j^2\sigma_j^2}{2}-\beta_j m_j\right\}\Phi\left(\frac{-\ln y + m_j - \beta_j\sigma_j^2}{\sigma_j}\right)\right] dy$$

Let $E_i = \exp\left\{\alpha_i m_i + \alpha_i^2\sigma_i^2/2\right\}$ and $G_i = \exp\left\{-\beta_i m_i + \beta_i^2\sigma_i^2/2\right\}$.

Then,

$$I = \left(\frac{\alpha_i\beta_i}{\alpha_i+\beta_i}\right)\left[E_i I_1 - \left(\frac{\beta_j}{\alpha_j+\beta_j}\right)E_i E_j I_2 - \left(\frac{\alpha_j}{\alpha_j+\beta_j}\right)E_i G_j I_3 + G_i I_4 - \left(\frac{\beta_j}{\alpha_j+\beta_j}\right)G_i E_j I_5\right.$$
$$\left. - \left(\frac{\alpha_j}{\alpha_j+\beta_j}\right)G_i G_j I_6\right]$$

where, from the lemma,

$$I_1 = \int_0^\infty y^{-\alpha_i}\Phi\left(\frac{\ln y - m_i - \alpha_i\sigma_i^2}{\sigma_i}\right)\Phi\left(\frac{\ln y - m_j}{\sigma_j}\right)dy$$
$$= \frac{1}{\alpha_i - 1}\left[\exp\left\{m_i(1-\alpha_i) + \frac{\sigma_i^2}{2}\left(1-\alpha_i^2\right)\right\}\Phi\left(\frac{m_i - m_j + \sigma_i^2}{\sqrt{\sigma_i^2+\sigma_j^2}}\right)\right.$$
$$\left. + \exp\left\{m_j(1-\alpha_i) + \frac{\sigma_j^2}{2}\left(1-\alpha_i\right)^2\right\}\Phi\left(\frac{m_j - m_i - \alpha_i\sigma_i^2 + (1-\alpha_i)\sigma_j^2}{\sqrt{\sigma_i^2+\sigma_j^2}}\right)\right]$$

$$I_2 = \int_0^\infty y^{-\alpha_i-\alpha_j}\Phi\left(\frac{\ln y - m_i - \alpha_i\sigma_i^2}{\sigma_i}\right)\Phi\left(\frac{\ln y - m_j - \alpha_j\sigma_j^2}{\sigma_j}\right)dy$$
$$= \frac{1}{\alpha_i + \alpha_j - 1}\left[\exp\left\{\frac{\sigma_i^2}{2}\left(1 - 2\alpha_j - \alpha_i^2 + \alpha_j^2\right) + m_i(1-\alpha_i-\alpha_j)\right\}\Phi\left(\frac{m_i - m_j + (1-\alpha_j)\sigma_i^2 - \alpha_j\sigma_j^2}{\sqrt{\sigma_i^2+\sigma_j^2}}\right)\right.$$
$$\left. + \exp\left\{\frac{\sigma_j^2}{2}\left(1 - 2\alpha_i - \alpha_j^2 + \alpha_i^2\right) + m_j(1-\alpha_i-\alpha_j)\right\}\Phi\left(\frac{m_j - m_i - \alpha_i\sigma_i^2 + (1-\alpha_i)\sigma_j^2}{\sqrt{\sigma_i^2+\sigma_j^2}}\right)\right]$$

$$I_3 = \int_0^{\infty} y^{\beta_j - \alpha_i} \Phi\left(\frac{\ln y - m_i - \alpha_i \sigma_i^2}{\sigma_i}\right) \Phi\left(\frac{-\ln y + m_j - \beta_j \sigma_j^2}{\sigma_j}\right) dy$$

$$= \frac{1}{\alpha_i - \beta_j - 1}\left[\exp\left\{\frac{\sigma_i^2}{2}\left(\beta_j^2 - \alpha_i^2 + 1 + 2\beta_j\right) + m_i\left(\beta_j - \alpha_i + 1\right)\right\}\Phi\left(\frac{m_j - m_i - \beta_j\left(\sigma_i^2 + \sigma_j^2\right) - \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right.$$

$$\left. - \exp\left\{\frac{\sigma_j^2}{2}\left(\alpha_i^2 - \beta_j^2 + 1 - 2\alpha_i\right) + m_j(\beta_j - \alpha_i + 1)\right\}\Phi\left(\frac{m_j - m_i - \alpha_i\left(\sigma_i^2 + \sigma_j^2\right) + \sigma_j^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right]$$

$$I_4 = \int_0^{\infty} y^{\beta_i} \Phi\left(\frac{\ln y - m_j}{\sigma_j}\right) \Phi\left(\frac{-\ln y + m_i - \beta_i \sigma_i^2}{\sigma_i}\right) dy$$

$$= \frac{1}{\beta_i + 1}\left[\exp\left\{m_i(1 + \beta_i) + \frac{\sigma_i^2}{2}\left(1 - \beta_i^2\right)\right\}\Phi\left(\frac{m_i - m_j + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right.$$

$$\left. - \exp\left\{m_j(1 + \beta_i) + \frac{\sigma_j^2}{2}(1 + \beta_i)^2 \Phi\left(\frac{m_i - m_j - \beta_i \sigma_i^2 - (1 + \beta_i)\sigma_j^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right\}\right]$$

$$I_5 = \int_0^{\infty} y^{\beta_i - \alpha_j} \Phi\left(\frac{\ln y - m_j - \alpha_j \sigma_j^2}{\sigma_j}\right) \Phi\left(\frac{-\ln y + m_i - \beta_i \sigma_i^2}{\sigma_i}\right) dy$$

$$= \frac{1}{\alpha_j - \beta_i - 1}\left[\exp\left\{m_j(\beta_i - \alpha_j + 1) + \frac{\sigma_j^2}{2}\left(\beta_i^2 - \alpha_j^2 + 2\beta_i + 1\right)\right\}\Phi\left(\frac{m_i - m_j - \beta_i\left(\sigma_i^2 + \sigma_j^2\right) - \sigma_j^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right.$$

$$\left. - \exp\left\{m_i(\beta_i - \alpha_j + 1) + \frac{\sigma_i^2}{2}\left(\alpha_j^2 - \beta_i^2 - 2\alpha_j + 1\right)\right\}\Phi\left(\frac{m_i - m_j - \alpha_j\left(\sigma_i^2 + \sigma_j^2\right) + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right]$$

and

$$I_6 = \int_0^{\infty} y^{\beta_i + \beta_j} \Phi\left(\frac{-\ln y + m_i - \beta_i \sigma_i^2}{\sigma_i}\right) \Phi\left(\frac{-\ln y + m_j - \beta_j \sigma_j^2}{\sigma_j}\right) dy$$

$$= \frac{1}{\beta_i + \beta_j + 1}\left[\exp\left\{m_i\left(\beta_i + \beta_j + 1\right) + \frac{\sigma_i^2}{2}\left(\beta_j^2 - \beta_i^2 + 2\beta_j + 1\right)\right\}\Phi\left(\frac{m_j - m_i - \beta_j\left(\sigma_i^2 + \sigma_j^2\right) - \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right.$$

$$\left. + \exp\left\{m_j\left(\beta_i + \beta_j + 1\right) + \frac{\sigma_j^2}{2}\left(\beta_i^2 - \beta_j^2 + 2\beta_i + 1\right)\right\}\Phi\left(\frac{m_i - m_j - \beta_i\left(\sigma_i^2 + \sigma_j^2\right) - \sigma_j^2}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right)\right]$$

Recognizing that the means of the distributions are given by

$$\mu_i = \frac{\alpha_i \beta_i}{(\alpha_i - 1)(\beta_i + 1)} \exp\left\{\frac{\sigma_i^2}{2} + m_i\right\}$$

and simplify each of the terms in I, reveals three distinct terms which we define as follows

$$P_{ij} = \mu_i \Phi\left(\frac{m_i - m_j + \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right)$$

$$Q_{ij} = \mu_j \exp\left\{ \alpha_i (m_i - m_j - \sigma_j^2) + \frac{\alpha_i^2}{2}(\sigma_i^2 + \sigma_j^2) \right\} \Phi\left(\frac{m_j - m_i - \alpha_i(\sigma_i^2 + \sigma_j^2) + \sigma_j^2}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right)$$

$$R_{ij} = \mu_i \exp\left\{ \beta_j (m_i - m_j + \sigma_i^2) + \frac{\beta_j^2}{2}(\sigma_i^2 + \sigma_j^2) \right\} \Phi\left(\frac{m_j - m_i - \beta_j(\sigma_i^2 + \sigma_j^2) - \sigma_i^2}{\sqrt{\sigma_i^2 + \sigma_j^2}} \right)$$

Collecting all the terms yields

$$I = P_{ij} + d_{ij} Q_{ij} + g_{ij} Q_{ji} - h_{ij} R_{ij} - k_{ij} R_{ji}$$

where

$$d_{ij} = \frac{\alpha_i \beta_i (\alpha_j - 1)(\beta_j + 1)}{\alpha_j \beta_j (\alpha_i + \beta_i)(\alpha_j + \beta_j)} \left[\frac{\alpha_j + \beta_j}{\alpha_i - 1} - \frac{\beta_j}{\alpha_i + \alpha_j - 1} + \frac{\alpha_j}{\alpha_i - \beta_j - 1} \right]$$

$$g_{ij} = \frac{\beta_j (\alpha_i - 1)(\beta_i + 1)}{(\alpha_j + \beta_j)(\alpha_j - \beta_i - 1)(\alpha_i + \alpha_j - 1)}$$

$$h_{ij} = \frac{\alpha_j (\alpha_i - 1)(\beta_i + 1)}{(\alpha_j + \beta_j)(\alpha_i - \beta_j - 1)(\beta_i + \beta_j + 1)}$$

and

$$k_{ij} = \frac{\alpha_i \beta_i (\alpha_j - 1)(\beta_j + 1)}{\alpha_j \beta_j (\alpha_i + \beta_i)(\alpha_j + \beta_j)} \left[\frac{\alpha_j + \beta_j}{\beta_i + 1} + \frac{\beta_j}{\alpha_j - \beta_i - 1} + \frac{\alpha_j}{\beta_i + \beta_j + 1} \right]$$

Then, the Gini coefficient can be written as

$$\begin{aligned} G &= -1 + \frac{2}{\mu} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j I \\ &= -1 + \frac{2}{\mu} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \left(P_{ij} + q_{ij} Q_{ij} - r_{ij} R_{ij} \right) \end{aligned}$$

where $q_{ij} = d_{ij} + g_{ji}$ and $r_{ij} = h_{ij} + k_{ji}$.

That completes the derivation of the Gini coefficients.