1D 2014 NNN+ 1

Exploring Lognormal Incomes

Milo Schield
Augsburg College
Editor: www.StatLit.org
US Rep: International Statistical Literacy Project

www.StatLit.org/
pdf/2014-Schield-Explore-LogNormal-Incomes-Slides.pdf
XLS/Create-LogNormal-Incomes-Excel2013.xlsx

1D 2014 NNN+ 2

Log-Normal Distributions

A Log-Normal distribution is generated from a normal with mu = Ln(Median) and sigma = Sqrt[2*Ln(Mean/Median)].

The lognormal is always positive and right-skewed.

Examples:

- Incomes (bottom 97%), assets, size of cities
- Weight and blood pressure of humans (by gender)

Benefit:

- calculate the share of total income held by the top X%
- calculate share of total income held by the 'above-average'
- explore effects of change in mean-median ratio.

1D 2014 NNN+ 3

Log-Normal Distributions

"In many ways, it [the Log-Normal] has remained the Cinderella of distributions, the interest of writers in the learned journals being curiously sporadic and that of the authors of statistical test-books but faintly aroused."

"We … state our belief that the lognormal is as fundamental a distribution in statistics as is the normal, despite the stigma of the derivative nature of its name."

Aitchison and Brown (1957). P 1.

1D 2014 NNN+ 4

Lognormal and Excel

Use Excel to focus on the model and the results.

Excel has two Log-Normal functions:

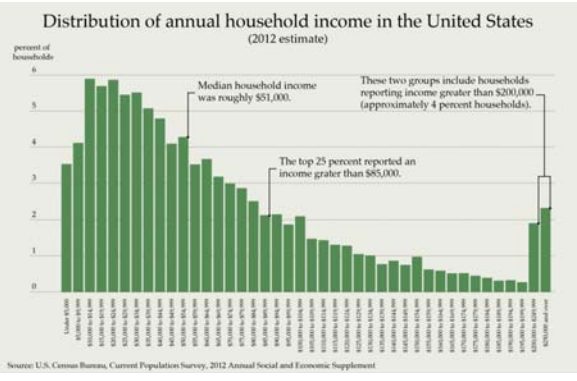
- Standard: =LOGNORM.DIST(X, mu, sigma, k)
  k=0 for PDF; k=1 for CDF.
- Inverse: =LOGNORM.INV(X, mu, sigma)

Use Standard to calculate/graph the PDF and CDF.

Use Inverse to find cutoffs: quartiles, to 1%, etc.

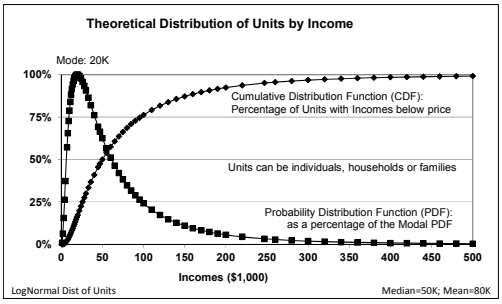
Use Excel to create graphs that show comparisons.

1D 2014 NNN+ 5



1D 2014 NNN+ 6

Log-Normal Distribution of Units



1D 2014 NNN+ 7

Paired Distributions

For anything that is distributed by X, there are always two distributions:

1. Distribution of subjects by X
2. Distribution of total X by X.

Sometime we ignore the 2nd: height or weight.

Sometimes we care about the 2nd: income or assets.

Surprise: If the 1st is lognormal, so is the 2nd.

1D 2014 NNN+ 8

Distribution of Households and Total Income by Income

Suppose the distribution of households by income is log-normal with normal parameters mu# and sigma#.

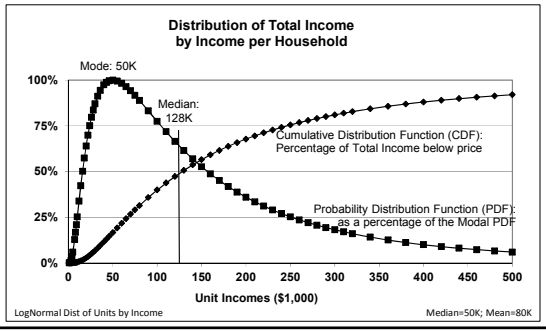
Then the distribution of total income by amount has a log-normal distribution with these parameters: mu$ = mu# + sigma#^2; sigma$ = sigma#.

See Aitchison and Brown (1963) p. 158.

Special thanks to Mohammod Irfan (Denver University) for his help on this topic.

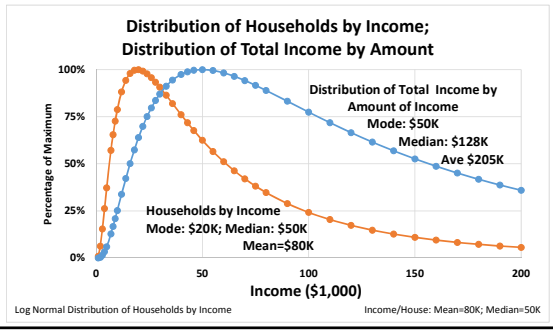
1D 2014 NNN+ 9

Distribution of Total Income



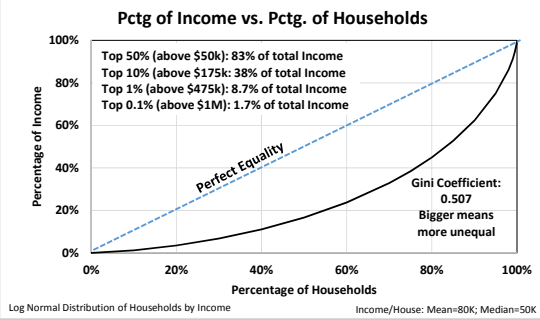
1D 2014 NNN+ 10

Distribution of Households and Total Income



1D 2014 NNN+ 11

Lorenz Curve and Gini Coefficient

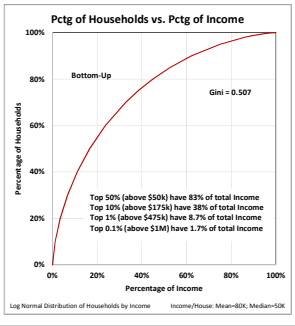


1D 2014 NNN+ 12

Champagne-Glass Distribution

The Gini coefficient is determined by the Mean#/Median# ratio.

The bigger this ratio the bigger the Gini coefficient and the greater the economic inequality.



1D 2014 NNN+ 13

Log-Normal Balance Conjecture

Conjecture: If household (HH) income is distributed log-normally and X% of households have below-average incomes, then X% of all income is earned by HH with above-average incomes.

Example: If 60% of HH have below-average incomes, then 60% of total income is earned by HH having above-average incomes.

Evidence using Excel spreadsheet:
Suppose Mean# = 50K and Median# = 80K.

- 68.61%: Percentage of HH having below-average income
- 68.61%: Percentage of total income that is associated with HH having above-average incomes. QED

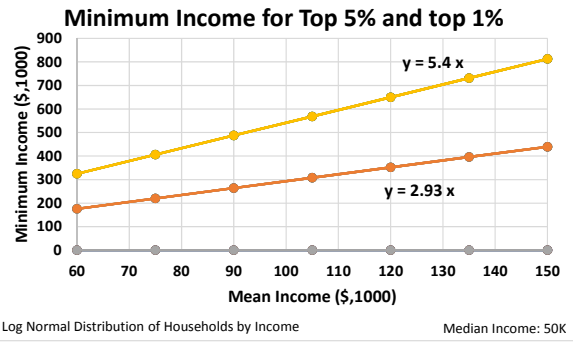
1D 2014 NNN+ 14

As Mean-Median Ratio ↑ Rich get Richer (relatively)

Log-normal distribution. Median HH income: $50K.

	Top 5%		Top 1%		
Mean#	Min$	%Income	Min$	%Income	Gini
55	103	11%	138	2.9%	0.24
60	135	15%	204	4.2%	0.33
65	165	18%	270	5.5%	0.39
70	193	20%	337	6.6%	0.44
75	220	23%	406	7.7%	0.48
80	246	25%	477	8.7%	0.51
85	272	27%	549	9.7%	0.53
90	298	29%	623	10.7%	0.56

1D 2014 NNN+ 15



1D 2014 NNN+ 16

Which parameters best model US household incomes?

US Median Income (Table 691*)

- $46,089 in 1970; $50,303 in 2008

Share of Total Income by Top 5% (Table 693*)

- 16.6% in 1970; 21.5% in 2008

Best log-normal fits:

- 1970 Median 46K, Mean 53K: Ratio = 1.15
- 2008 *Median 50K, Mean 73K*; Ratio = 1.46

* 2011 US Statistical Abstract (2008 dollars).

1D 2014 NNN+ 17

Conclusion

Using the LogNormal distributions provides a principled way students can explore a plausible distribution of incomes.

Allows students to explore the difference between part and whole when using percentage grammar.

1D 2014 NNN+ 18

Bibliography

Aitchison J and JAC Brown (1957). *The Log-normal Distribution*. Cambridge (UK): Cambridge University Press. Searchable copy at Google Books: http://books.google.com/books?id=Kus8AAAAIAAJ

Cobham, Alex and Andy Sumner (2014). Is inequality all about the tails?: The Palma measure of income inequality. Significance. Volume 11 Issue 1. www.significancemagazine.org/details/magazine/5871201/Is-inequality-all-about-the-tails-The-Palma-measure-of-income-inequality.html

Limpert, E., W.A. Stahel and M. Abbt (2001). Log-normal Distributions across the Sciences: Keys and Clues. *Bioscience* 51, No 5, May 2001, 342-352. Copy at http://stat.ethz.ch/~stahel/lognormal/bioscience.pdf

Schield, Milo (2013) Creating a Log-Normal Distribution using Excel 2013. www.statlit.org/pdf/Create-LogNormal-Excel2013-Demo-6up.pdf

Stahel, Werner (2014). Website: http://stat.ethz.ch/~stahel

Univ. Denver (2014). Using the LogNormal Distribution. Copy at http://www.du.edu/ifs/help/understand/economy/poverty/lognormal.html

Wikipedia. LogNormal Distribution.

Exploring Lognormal Incomes

Milo Schield
Augsburg College
Editor: www.StatLit.org
US Rep: International Statistical Literacy Project

www.StatLit.org/
pdf/2014-Schield-Explore-LogNormal-Incomes-Slides.pdf
XLS/Create-LogNormal-Incomes-Excel2013.xlsx

Log-Normal Distributions

A Log-Normal distribution is generated from a normal with mu = Ln(Median) and sigma = Sqrt[2*Ln(Mean/Median)].

The lognormal is always positive and right-skewed.

Examples:

- Incomes (bottom 97%), assets, size of cities
- Weight and blood pressure of humans (by gender)

Benefit:

- calculate the share of total income held by the top X%
- calculate share of total income held by the ‘above-average’
- explore effects of change in mean-median ratio.

Log-Normal Distributions

"In many ways, it [the Log-Normal] has remained the Cinderella of distributions, the interest of writers in the learned journals being curiously sporadic and that of the authors of statistical test-books but faintly aroused."

"We … state our belief that the lognormal is as fundamental a distribution in statistics as is the normal, despite the stigma of the derivative nature of its name."

Aitchison and Brown (1957). P 1.

Lognormal and Excel

Use Excel to focus on the model and the results.

Excel has two Log-Normal functions:

- Standard: =LOGNORM.DIST(X, mu, sigma, k)
  k=0 for PDF; k=1 for CDF.
- Inverse: =LOGNORM.INV(X, mu, sigma)

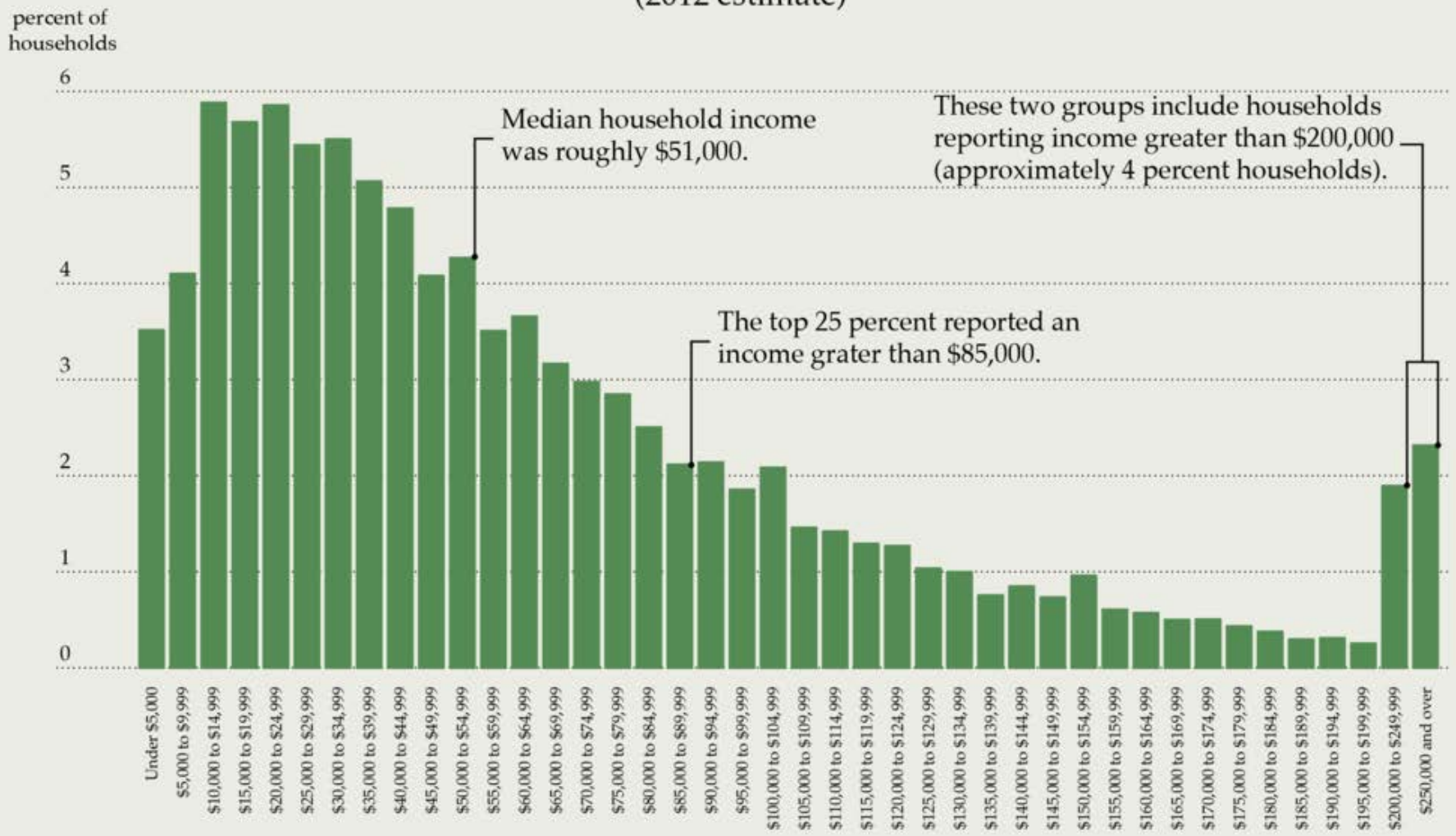
Use Standard to calculate/graph the PDF and CDF.

Use Inverse to find cutoffs: quartiles, to 1%, etc.

Use Excel to create graphs that show comparisons.

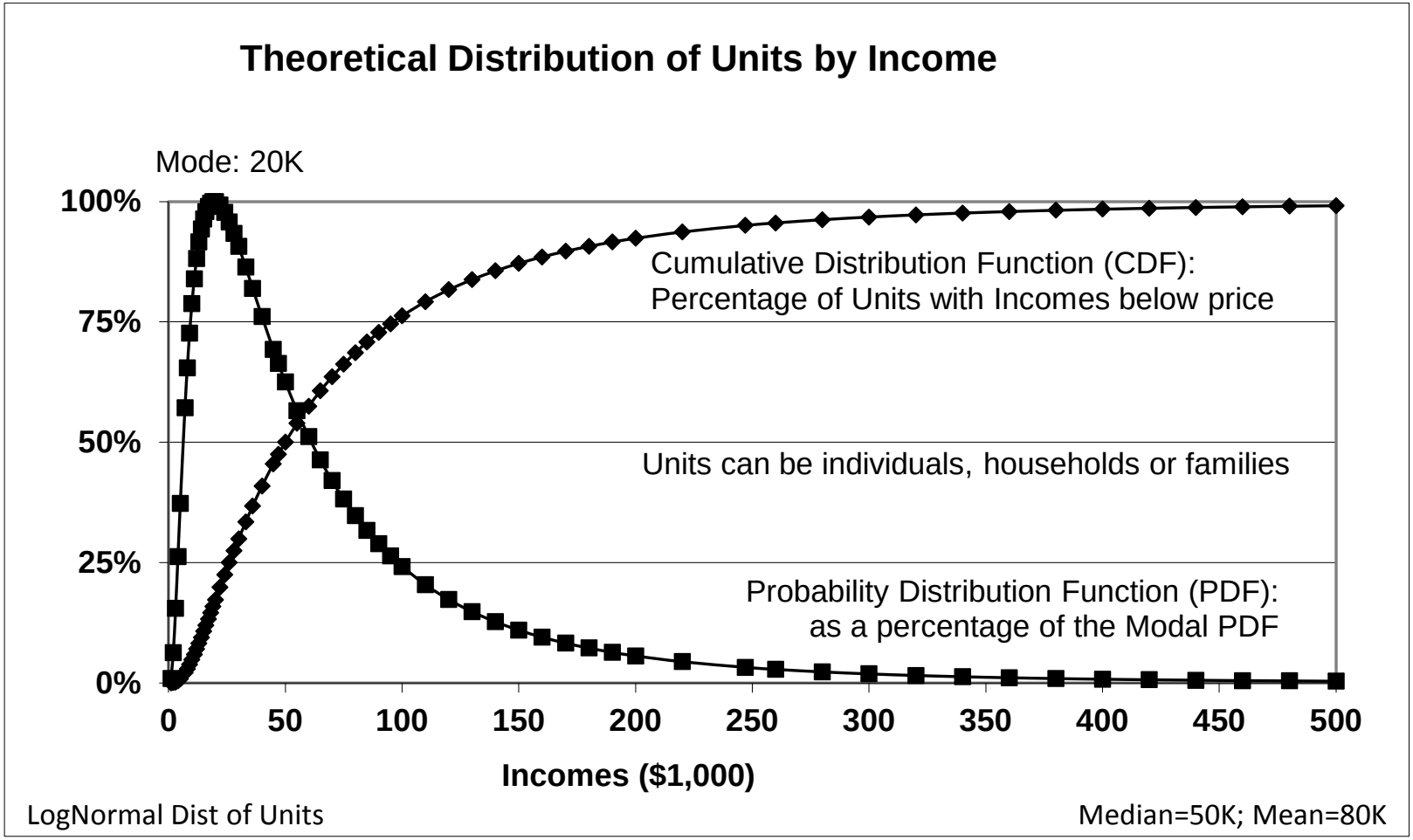
Distribution of annual household income in the United States

(2012 estimate)



Source: U.S. Census Bureau, Current Population Survey, 2012 Annual Social and Economic Supplement

Log-Normal Distribution of Units



Paired Distributions

For anything that is distributed by X, there are always two distributions:

1. Distribution of subjects by X
2. Distribution of total X by X.

Sometime we ignore the 2nd: height or weight.

Sometimes we care about the 2nd: income or assets.

Surprise: If the 1st is lognormal, so is the 2nd.

Distribution of Households and Total Income by Income

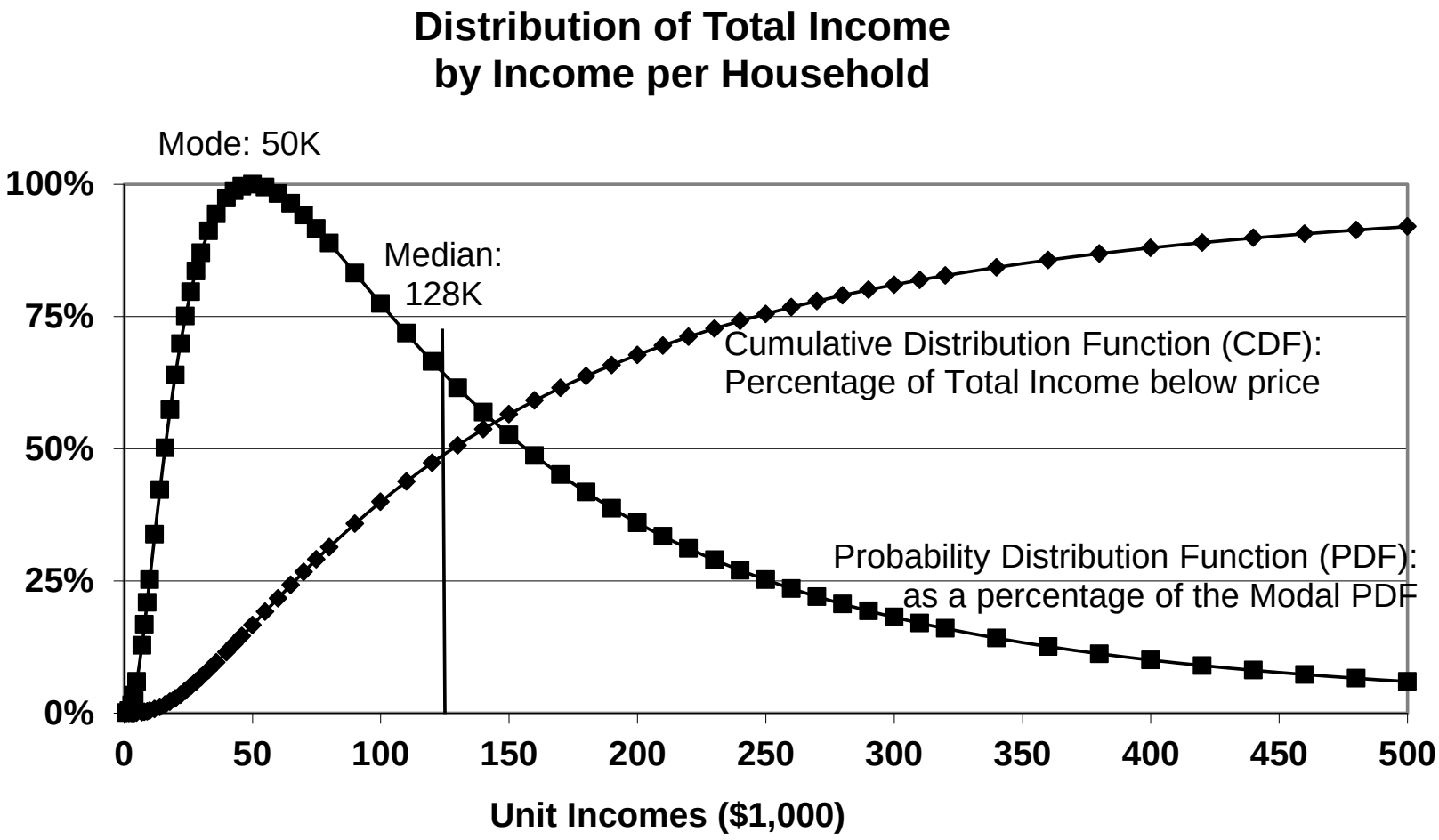
Suppose the distribution of households by income is log-normal with normal parameters mu# and sigma#.

Then the distribution of total income by amount has a log-normal distribution with these parameters:
mu$ = mu# + sigma#^2; sigma$ = sigma#.

See Aitchison and Brown (1963) p. 158.

Special thanks to Mohammod Irfan (Denver University) for his help on this topic.

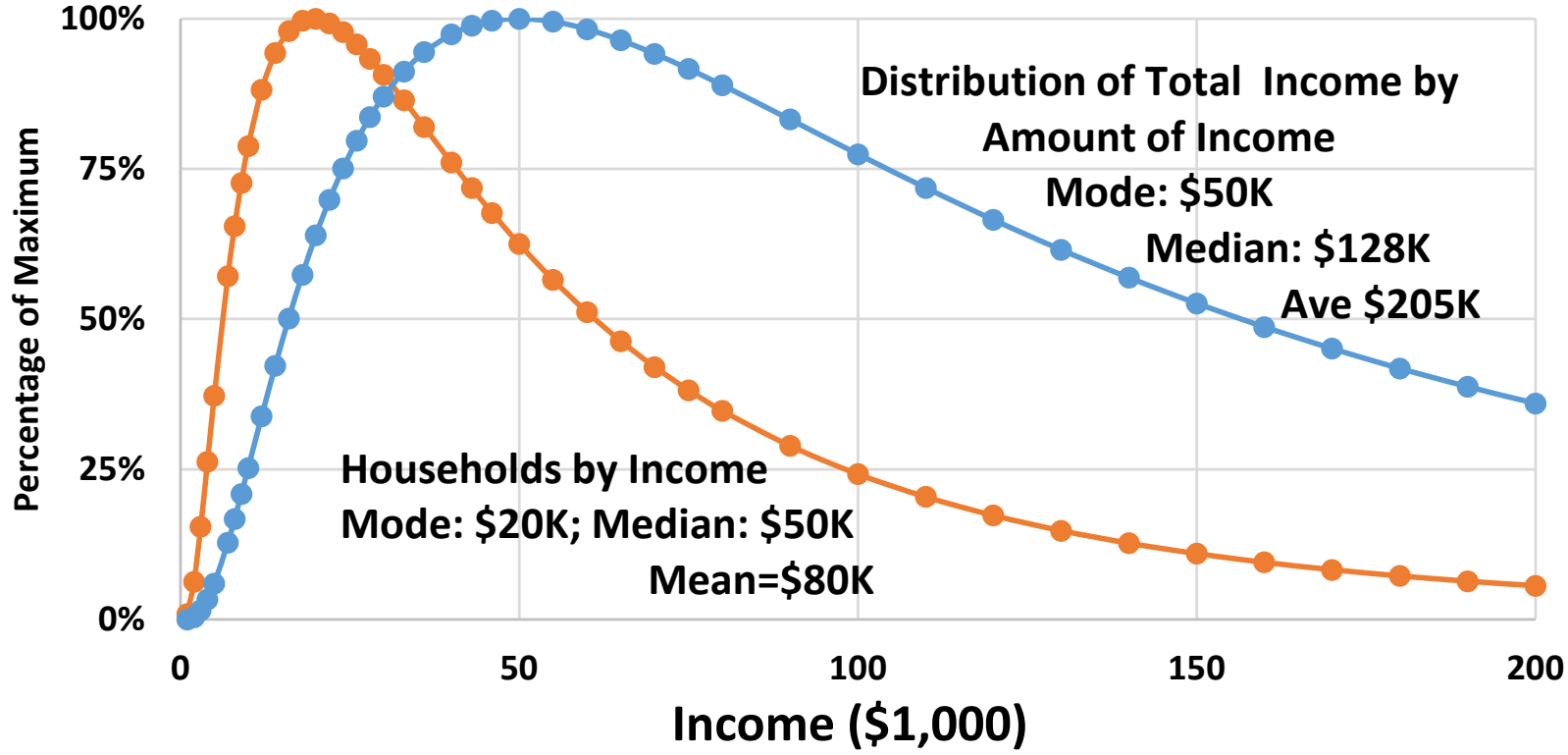
Distribution of Total Income



LogNormal Dist of Units by Income

Median=50K; Mean=80K

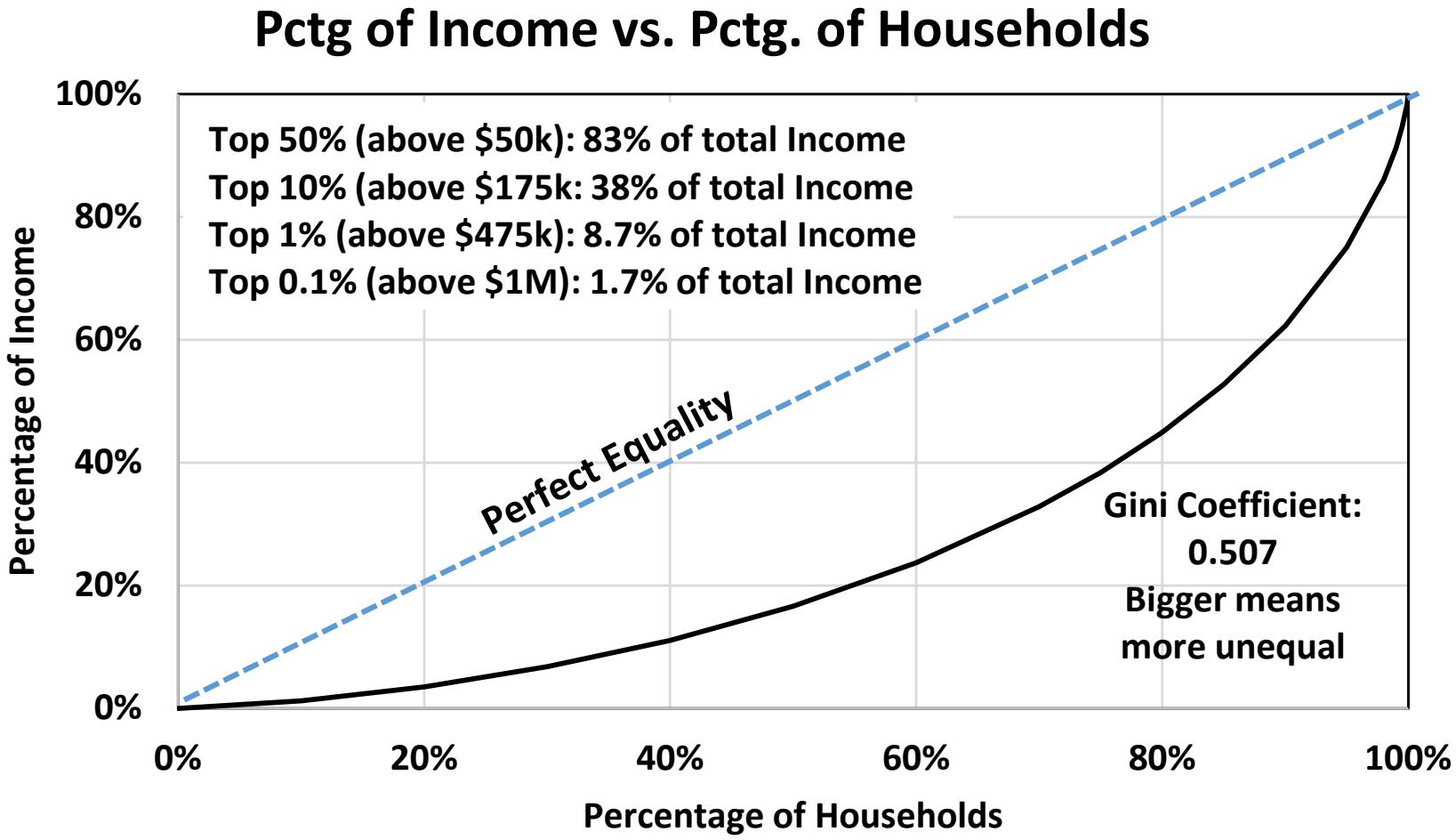
Distribution of Households and Total Income



Log Normal Distribution of Households by Income

Income/House: Mean=80K; Median=50K

Lorenz Curve and Gini Coefficient



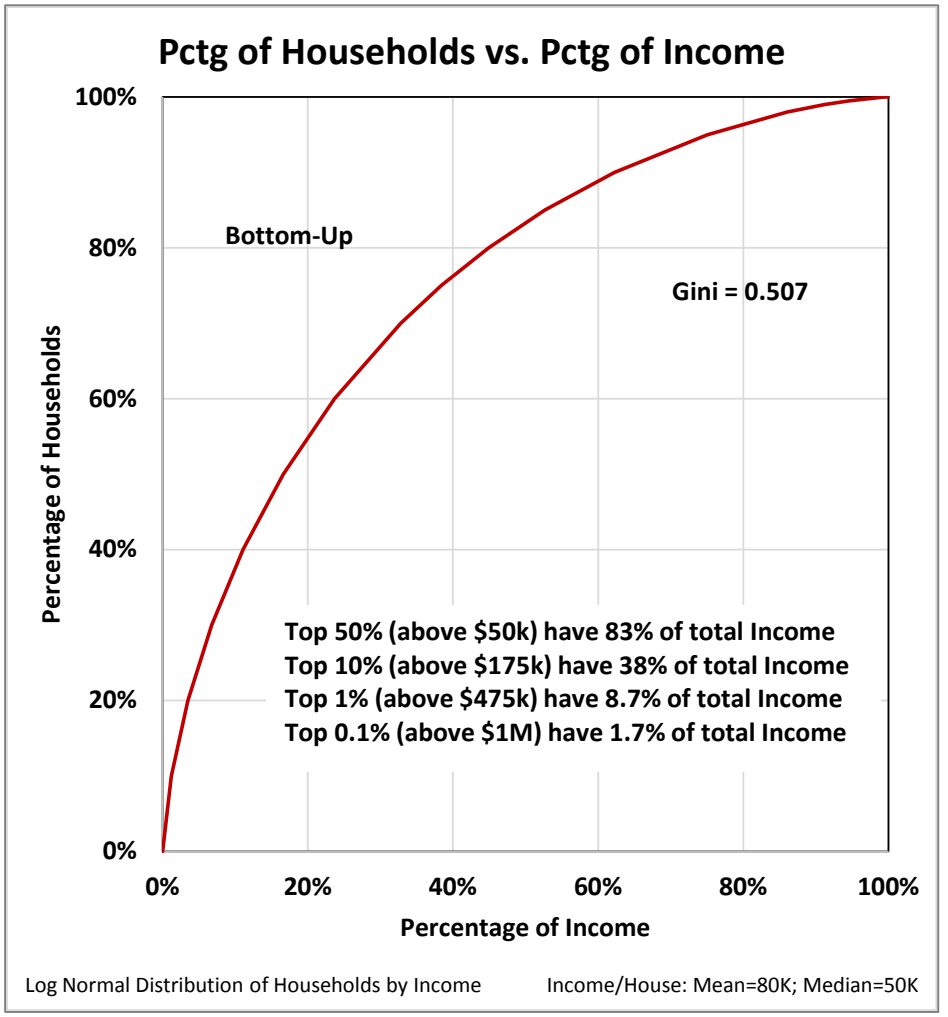
Log Normal Distribution of Households by Income

Income/House: Mean=80K; Median=50K

Champagne-Glass Distribution

The Gini coefficient is determined by the Mean#/Median# ratio.

The bigger this ratio the bigger the Gini coefficient and the greater the economic inequality.



Log-Normal Balance Conjecture

Conjecture: If household (HH) income is distributed log-normally and X% of households have below-average incomes, then X% of all income is earned by HH with above-average incomes.

Example: If 60% of HH have below-average incomes, then 60% of total income is earned by HH having above-average incomes.

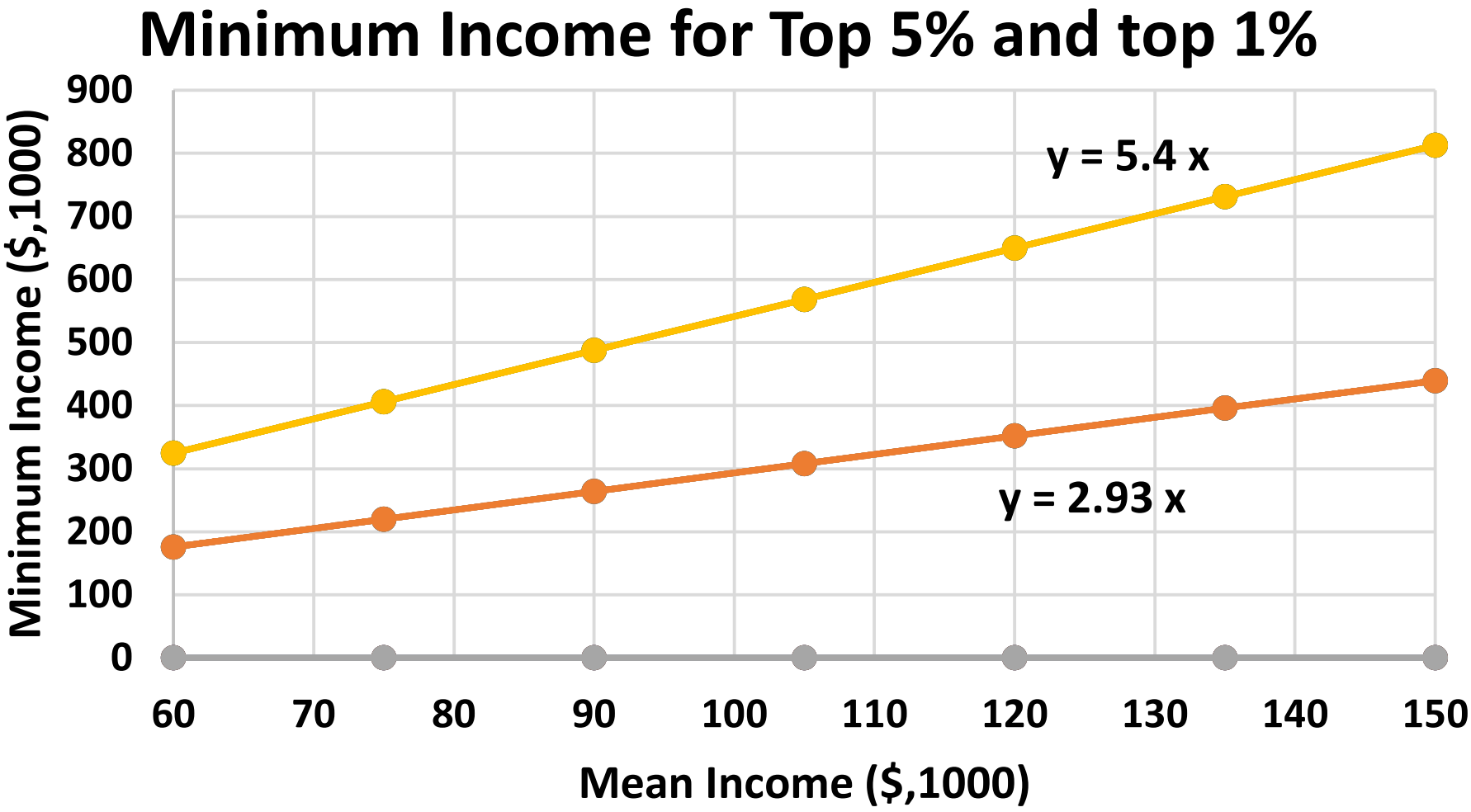
Evidence using Excel spreadsheet:
Suppose Mean# = 50K and Median# = 80K.

- 68.61%: Percentage of HH having below-average income
- 68.61%: Percentage of total income that is associated with HH having above-average incomes. QED

As Mean-Median Ratio ↑ Rich get Richer (relatively)

Log-normal distribution. Median HH income: $50K.

	Top 5%		Top 1%		
Mean#	Min$	%Income	Min$	%Income	Gini
55	103	11%	138	2.9%	0.24
60	135	15%	204	4.2%	0.33
65	165	18%	270	5.5%	0.39
70	193	20%	337	6.6%	0.44
75	220	23%	406	7.7%	0.48
80	246	25%	477	8.7%	0.51
85	272	27%	549	9.7%	0.53
90	298	29%	623	10.7%	0.56



Log Normal Distribution of Households by Income

Median Income: 50K

Which parameters best model US household incomes?

US Median Income (Table 691*)

- $46,089 in 1970; $50,303 in 2008

Share of Total Income by Top 5% (Table 693*)

- 16.6% in 1970; 21.5% in 2008

Best log-normal fits:

- 1970 Median 46K, Mean 53K: Ratio = 1.15
- 2008 *Median 50K*, *Mean 73K*; Ratio = 1.46

* 2011 US Statistical Abstract (2008 dollars).

Conclusion

Using the LogNormal distributions provides a principled way students can explore a plausible distribution of incomes.

Allows students to explore the difference between part and whole when using percentage grammar.

Bibliography

Aitchison J and JAC Brown (1957). *The Log-normal Distribution*. Cambridge (UK): Cambridge University Press. Searchable copy at Google Books: http://books.google.com/books?id=Kus8AAAAIAAJ

Cobham, Alex and Andy Sumner (2014). Is inequality all about the tails?: The Palma measure of income inequality. Significance. Volume 11 Issue 1. www.significancemagazine.org/details/magazine/5871201/Is-inequality-all-about-the-tails-The-Palma-measure-of-income-inequality.html

Limpert, E., W.A. Stahel and M. Abbt (2001). Log-normal Distributions across the Sciences: Keys and Clues. *Bioscience* 51, No 5, May 2001, 342-352. Copy at http://stat.ethz.ch/~stahel/lognormal/bioscience.pdf

Schield, Milo (2013) Creating a Log-Normal Distribution using Excel 2013. www.statlit.org/pdf/Create-LogNormal-Excel2013-Demo-6up.pdf

Stahel, Werner (2014). Website: http://stat.ethz.ch/~stahel

Univ. Denver (2014). Using the LogNormal Distribution. Copy at http://www.du.edu/ifs/help/understand/economy/poverty/lognormal.html

Wikipedia. LogNormal Distribution.