Human Development Index: PNG progress and a simulated interpretation

Peter K Anderson

Abstract

The Human Development Index attempts to measure human well-being and its development over time in multiple countries across the world. Relative values of this index seem to possess an undesirable inherent stability with little indication of the removal of inequality. Monte Carlo simulation is used to explore possible causes of this stability. Where collected historical data can be best-fitted to a particular theoretical distribution, some of the inherent properties of the data can be revealed. Per-capita Gross National Income data is at least visibly consistent with a Lognormal probability distribution suggesting that poverty may be the result of multiplicatively interdependent factors. Thus there may be a certain inevitability that, without special intervention, the rich will become richer and the poor, poorer.

Key words:

Human Development Index, Gross National Income, probability distribution, frequency distribution, cumulative frequency distribution, lognormal distribution, law of proportionate effect, Monte Carlo Simulation.

Introduction

Well known probability or frequency distributions arising from those used in statistics model the behavior of random variables whose characteristics are known. These variables arise from various real world situations. When a particular distribution can be fitted to a set of empirical data, the distribution is commonly used to make predictions about probable future behavior of the system generating the data. However, the fitting can also be used to suggest assumptions about the origin or causes of the empirical data based on knowledge of characteristics of the variable giving rise to a particular distribution (e.g. Hahn & Shapiro, 1967).

After exploring the origin of the lognormal distribution using Monte Carlo simulation, this paper reviews some of the data recorded in the Human Development Reports (HDR) developed over the past two decades. It notes the relative progress of PNG and its near neighbors on the Human Development Index (HDI). The perceived lack of progress relative to more developed countries in the same region leads to the examination of one of the several factors, the per capita Gross National Income (GNI), from the perspective of its empirical data fit to the lognormal distribution. The assumption is made that if empirical random data from an entity can be fitted to a particular distribution, hypotheses may be established concerning the underlying natural or other causes of the behavior of the entity.

Human Development Index (HDI)

The HDI is a composite statistic intended to be a holistic measure of human wellbeing calculated from data collected annually by the United Nations Development Program (UNDP) for each country in the world where data is available. The information compiled includes data on aspects of human and economic life such as life expectancy, achieved educational levels, reduced maternal mortality rates, measures of poverty and health, all as indicators of standard of living. These measures of human well-being are combined with per capita GNI, a quantitative measure of national economic growth, to produce the HDI, a ranking index ranging from approximately 0.3 (the low human development group) to nearly 1 (the very high human development group) for advanced countries. As data is collected annually, changing levels of estimated human development or wellbeing can be tracked for the 186 countries for which data is available.

The world map of Human Development Index (Figure 1) in 2013 (The Human Development Index: Wikipedia and based on HDR (2013), Table 1, p 144) identifies a general disparity in HDI values on a world map. The North (darker colours) South (lighter colours cutting a diagonal swathe from left to right) division is apparent. Australia and New Zealand provide an interesting anomaly, being "high human development" countries in the far south and their relative geographical locations support the comparisons made in the paper.

Very High
High
Medium
Low
Data unavailable

Figure 1 World map by quartiles of Human Development Index in 2013 (The Human Development Index: Wikipedia) showing the North (darker colours) South (lighter colours cutting a diagonal swathe from left to right) division and based on HDR (2013), Table 1, p 144.

The limitations of HDI, an index from easily measured quantities, as a measure of the quality of human life are readily acknowledged. "…. human well-being and freedom, and their connection with fairness and justice in the world, cannot be reduced simply to the measurement of GDP and its growth rate" (UNDP, p 24). Thus there is a need to avoid a reductionist approach which would equate human wellbeing completely with these easily measured indicators. Despite this acknowledged limitation, this paper assumes that the HDI data is still useful and proceeds to make best use of its availability.

In 2012, Papua New Guinea (PNG) ranked 156 out of the 186 ranked countries and is classified as a country of "low human development" (Human Development Report, 2013, Table, p 144).

Neighboring Solomon Islands (SI) was ranked 143, but still within the same low human development group. These rankings can be compared with those of Australia (rank 2) and New Zealand (rank 6), other near neighbors and sources of overseas aid for PNG who are ranked in the "very high development" group on the HDI. The disparity between these countries could hardly be much greater. PNG has, however, shown some limited improvement in HDI (Figure 2 and Table 1) with its HDI ranking growing from 0.324 (1980) to 0.466 (2012). Despite this upward trend, there has been a downward trend in growth rate (Figure 3) as measured over consecutive 10 year periods and as indicated by the decreasing slope of the plotted lines from 1980 to 2012.

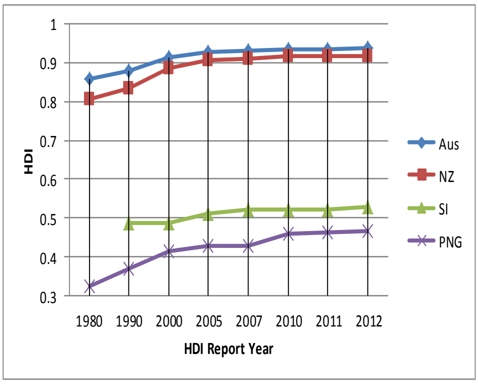

Figure 2 HDI Growth curves compared between selected countries in the Pacific region show little change in relative positions over time.

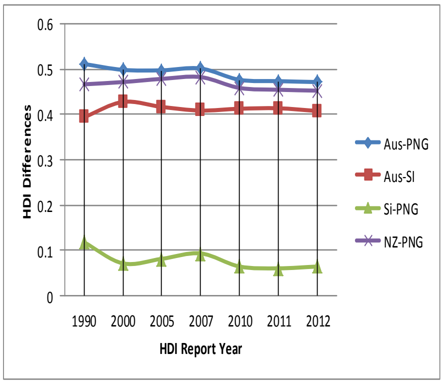

Figure 3 HDI differences compared as in Figure 1 showing little change in differences despite decades of overseas aid.

.

	1980	**1990**	**2000**	**2005**	**2007**	**2010**	**2011**	**2012**
Aus.	0.857	0.880	0.914	0.927	0.931	0.935	0.936	0.938
NZ	0.807	0.835	0.887	0.908	0.912	0.917	0.918	0.919
SI			0.486	0.510	0.522	0.522	0.526	0.530
PNG	0.324	0.368	0.415	0.429	0.429	0.458	0.462	0.466

Table 1 Growth in HDI values for selected neighbouring countries in the Pacific showing progressive relative development (HDR, 2013, Table 2 p148)**.** Whilst there have been changes in the method of HDI calculation over the years, the data presented here has been recalculated according to the most recent method.

HDI differences between these counties are quite stable (relatively flat plotted lines in Figure 3 and data in Table 2) showing little evidence of reduction of HDI disparity countries classified with “low human development” and their higher ranking neighbours despite decades of aid from the latter. This is here interpreted as suggesting that there might be other factors operating to produce these apparently stable disparities.

	1990	**2000**	**2005**	**2007**	**2010**	**2011**	**2012**
Aus-PNG	0.512	0.499	0.498	0.502	0.477	0.474	0.472
Aus-SI	0.394	0.428	0.417	0.409	0.413	0.414	0.408
SI-PNG	0.118	0.071	0.081	0.093	0.064	0.060	0.064
NZ-PNG	0.467	0.472	0.479	0.483	0.459	0.455	0.453

Table 2 Differences in HDI values for selected neighbouring countries in the Pacific showing only very small convergence of HDI values between neighbouring Pacific Island countries (calculated from data supplied in HDR, 2013, Table 2 p148).

Possible factors influencing HDI

The hypothesis of this paper is motivated by the way in which factors affecting HDI appear to be compounded as suggested in the HDR (Human Development Report 2013). The report notes that:

> “Environmental threats …. and natural disasters affect everyone, but they hurt poor countries and poor communities the most” (HDR Overview, p6).

It is further noted that:

> “Although low HDI countries contribute least to global change, they are likely to endure the greatest loss in annual rainfall and sharpest increase in its variability … with dire consequences for agricultural production and livelihoods” (HDR Overview, p6)

as a result of this change. These observations are consistent with the well-known observation that "the rich get richer and the poor get poorer" seemingly as a quite natural consequence of being where they are.

These perceptions suggest that causative factors of HDI values may be multiplicative meaning that the value of a human development variable at any time is proportionate to its value at a previous period of time. Thus a negative impact on a national economy will hurt poor counties more than those that are wealthy. If causative factors combine in such a multiplicative manner, the lognormal distribution suggests itself as a possible statistical model to fit the empirical data listed in the HDR.

Lognormal Distribution

When a random variable is the total effect of a large number of qualitatively different interacting factors, such that the influence of one factor is proportional to the magnitude of the other factors, the variable displays a lognormal distribution (Aitchison & Brown, 1969; Crow & Shimizu, 1988). This is in contrast to the well-known normal distribution in which the randomly varying contributing factors are independent and simply add together without interaction. With the lognormal distribution, the contributing factors are known to multiply rather than add together.

As an example of interacting factors, consider a variable x as the time for human recovery after a medical operation (cf. Lawrence, 1988). Influencing factors might be seriousness of the operation (SO), age of patient (AP) and state of health (SoH) of the patient. The effect of AP is reasonably dependent on SO (e.g. being greater for more serious operations) or on SoH and so on. Such more elementary variables, therefore, combine their influence in a multiplicative, rather than an additive way (as noted with the normal distribution).

Thus, if T_0 is the recovery time for a patient after an average operation:

$$T_1 = T_0 + \varepsilon_1 T_0 = T_0 (1 + \varepsilon_1)$$

where ε_1 is a random proportion of T_0 for the effect of SO;

$$T_2 = T_1 + \varepsilon_2 T_1 = T_1 (1 + \varepsilon_2) = T_0 (1 + \varepsilon_1)(1 + \varepsilon_2)$$

and where ε_2 involves the effect of AP. Similarly, we can write:

$$T_3 = T_2(1 + \varepsilon_3) = T_0\,(1 + \varepsilon 1)\,(1 + \varepsilon_2)\,T_2(1 + \varepsilon_3)$$

indicating the multiplicative effect of the factors influencing the time of recovery after an operation.

In general the multiplicative effect can be represented as:

$$T_j = T_{j-1}(1 + \varepsilon_j) \text{ or } T_j - T_{j-1} = \varepsilon_j T_{j-1} \qquad (1)$$

which is a recurrence relationship where epsilon ε_j is a random proportion of $T_{j-1,}$ the index j is an integer ranging from 1 to n, and T_j is a variable (recovery time in this example) resulting from n multiplicative effects. This embodies what is known as the *law of proportionate effect:* the change in the value of a variable at any step of the process is a random proportion of the previous value of the variable (Aitchison & Brown, 1969: 22) working back through previous steps in a first order recurrence sequence.

Variables resulting from such multiplicative effects of many small, qualitatively different, elementary variables may be transformed into normal random variables with the natural logarithm, ln(x), function (in which multiplicative effects become additive) and ln(x) is distributed as $N(\mu,\sigma^2)$ where N denotes a normal distribution with mean μ and variance σ^2 . The form of the function:

$$\text{where } z = (\ln(x) - \mu)/\sigma \qquad (2)$$

has a shape characterised by positive skewing, a peak near zero, a lower bound on the x axis, and the mode and median score falling below the mean. The parameters μ and σ are, respectively, the mean and variance of the normal distribution which would be obtained by considering the natural log of the X variable values (ln x). For the lognormal distribution the corresponding parameters are: expected value: $\exp(\mu+0.5\sigma^2)$, variance: $(\exp(\sigma^2-1)\exp(2\mu+\sigma^2)$, mode: $\exp(\mu-\sigma^2)$ and median: $\exp \mu$.

The effect of these parameters is firstly to explain the positive skewing given that the expected value, mode and median are all different and so separated. Secondly they allow considerable variation in possible patterns of data that the lognormal function (2) can fit. The parameter σ functions as a scale parameter (Figure 4, where μ is kept constant) and μ as a position parameter (Figure 5, where σ is kept constant). This suggests that there is a strong possibility that some form of the lognormal function may be found to fit empirical data characterised by a lower limit of zero and typically small rather than large values.

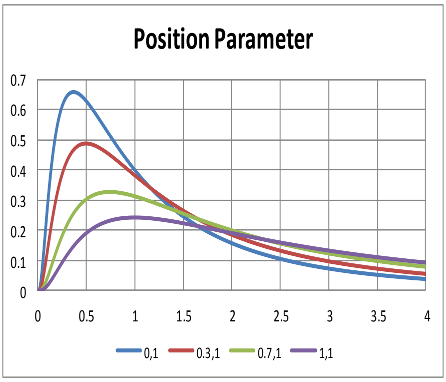

Figure 4 Variations in Lognormal distributions as the scale parameter σ varies (= 0.1, 0.2, 0.3 & 0.7) with position parameter μ (=0) constant.

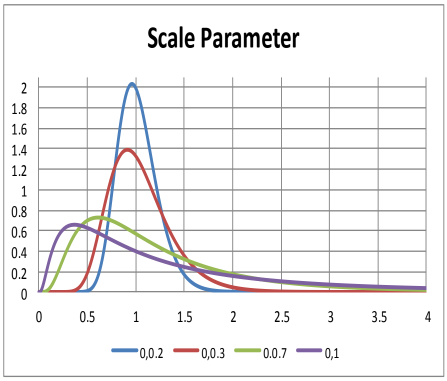

Figure 5 Variations in Lognormal distributions as the position parameter μ (= 0, 0.3, 0.7, 1) varies with scale parameter σ (=1) constant.

A modeling example

For purposes of modeling of the origin of lognormal distributions, a Monte-Carlo method (Manno, 1999) of simulating such a distribution using random variables was used with both a spreadsheet (Excel, 2010) and the R platform for data analysis (Kabacoff, 2011). The simulations considered 5000 theoretical income earners, with initial wealth I_0 ($1000), being rewarded with 30 periodic incomes, each of which was a proportion of the income from the previous period.

The total accumulated wealth for each earner, from the law of proportionate effect (see (1) above), is given by:

$$I_n = I_0(1+r_1)(1+r_2)........(1+r_n), \qquad (3)$$

for n periods of income earning. For the spreadsheet simulation the random proportion value r_i was generated with the RANDBETWEEN function as in the following:

$$X_j = X_{j-1}*(1 + \text{RANDBETWEEN}(1,10)/10).$$

The effect of this function as displayed here is to generate an r_i value uniformly distributed between 0.1 and 1. The final result (in column 31) was then divided by an appropriate power of 10 to produce a number between3 and 5 digits. The effect of this simulation was to produce a characteristic lognormal distribution (Figures 6 & 7) with large positive skewing and a preponderance of small values. The strong positive skew shows how initially equal wealth units become separated with time as a result of purely random effects.

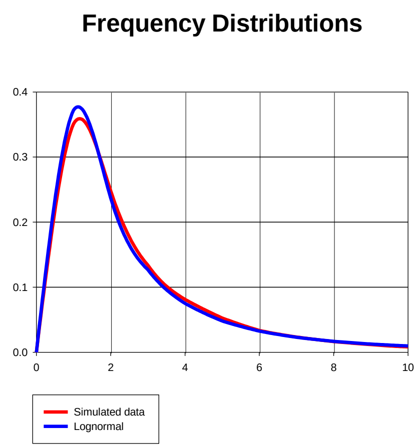

Figure 6 Simulated frequency wealth data for a theoretical set of income earners where yearly income is a random fraction of the previous year's income.

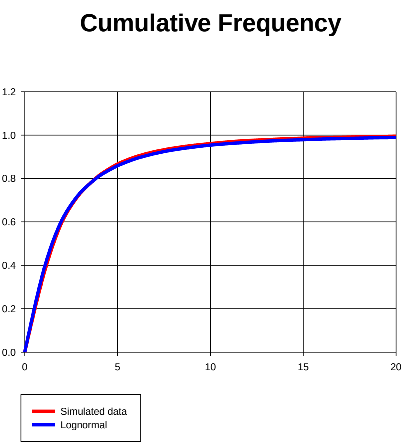

Figure 7 Cumulative frequency data resulting from the simulation as described for Figure 6.

The simulated data (red) appears to quite closely fit the corresponding lognormal theoretical distribution, a closeness to be explored later in the paper.

A second simulation was carried out using R programming (Kabacoff, 2011), an open source scripting language. A script (see Appendix: R Source Code) was used to generate random incomes but this time the proportion variable (r_i) was drawn from a standard normal distribution (rather than the evenly distributed random distribution used with the spreadsheet simulation). Because this variable can take positive and negative values, all the standard increments were multiplied by a factor of 3% before addition to prevent negative incomes. Such a factor could conceivably correspond to common interest rates, a base rate at which money could accrue.

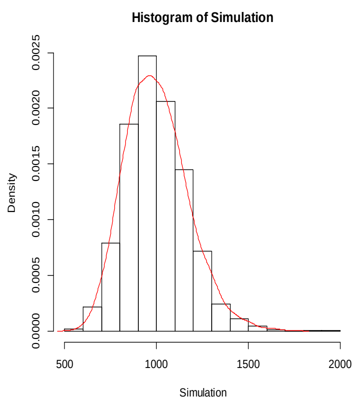

Figure 8 Histogram of data simulated using a standard normal distribution to model the fraction of the previous year's income. The red line shows a best fit lognormal frequency distribution.

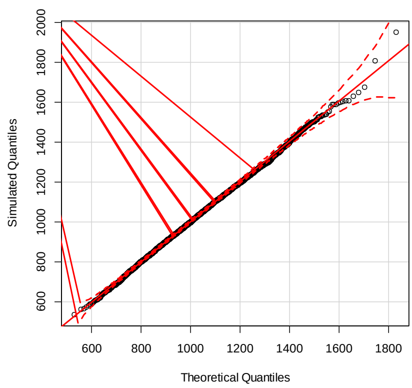

Figure 9 qqPlot of quantiles for simulated data (y axis) and theoretical distribution (x axis) falling with most points lying between the red 95% confidence interval lines.

The R script also generated a frequency histogram for 5000 income earners after 30 income earning periods (Figure 8) with a best fitting lognormal curve shown as an apparently well-fitting overlay. Also confirming a lognormal fit to the data is the

Quantile - Quantile plot (qq Plot in Figure 9) used to determine if two data sets come from populations with a common distribution.

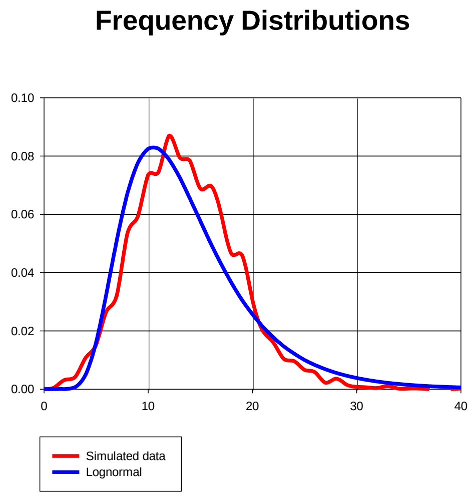

Figure 10 Comparative frequency distributions, simulated and theoretical, from a 5000 run simulation using data generated using the R script.

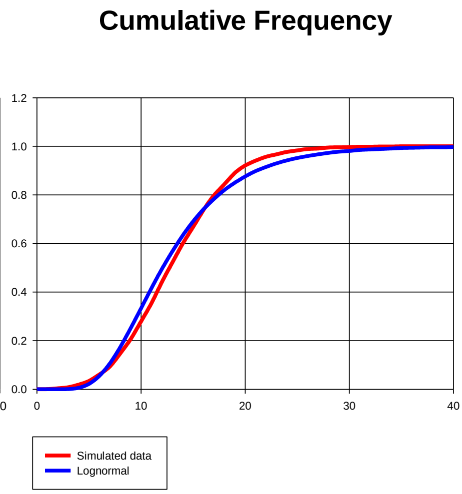

Figure 11 Comparative cumulative frequency distributions, simulated and theoretical, from a 5000 run simulation.

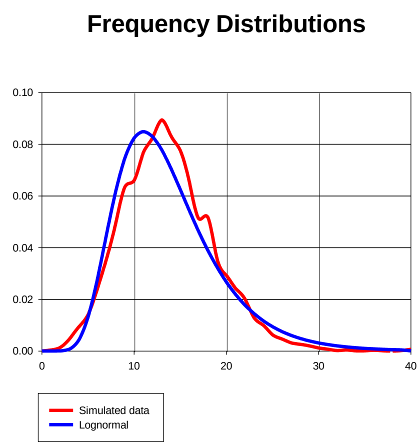

Figure 12 Comparative frequency distributions, simulated and theoretical, from a 10000 run simulation using data generated using the R script.

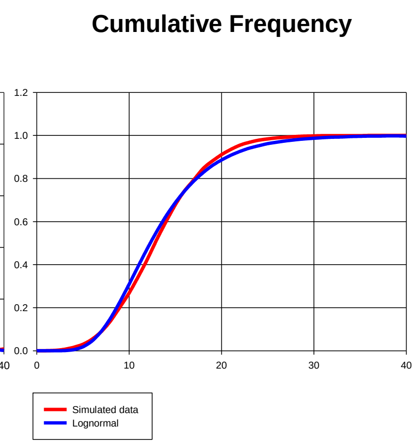

Figure 13 Comparative cumulative frequency distributions, simulated and theoretical, from a 10000 run simulation.

The two data sets are ordered and equivalent positions (quantiles or equally probable positions) are matched. If they do come from the same distribution, plotted points should fall on the 45° reference line which they clearly do in this simulation with most points lying between the 95% confidence lines shown in red.

Further graphic displays (Figures 10 to 13, using the Input Analyser display tool from Arena simulation software (Kelton et al., 2010) show the relation between simulated data (red lines) and corresponding theoretical lognormal distributions (blue lines). For reasons which are not presently clear, these graphs show a somewhat poorer closeness of fit than do those obtained from the R script (Figure 8), although running the simulation for 10000 cases (Figures 11 & 12) does show a visible improvement on the simulation run for 5000 cases only (Figures 10 & 11).

However simplistic this modeling as a simple random process (Aitchison & Brown, 1969: 116) may appear (e.g. income earners do not usually possess equal initial wealth), there does appear to be supportive visual evidence that such a dynamic as modeled here is active as a discussion of results of the per capita Gross National Incomes in the HDR would also suggest.

Modeling per capita GNI

World per capita GNI data for 1995 and 2011 are available (HDR, 2013) for consideration as lognormal distributions. Some summary data (Table 3) shows the scale of variation in the

	GNI Data 1995	**GNI Data 2011**
Average	$6949.58	$12700.93
St. Dev.	$7264.12	$13794.91
N	174	181
Aus.	$19632	$34548
NZ	$17627	$24818
SI	$2230	$2581
PNG	$2500	$2500

Table 3 GNI data for 1995 and 2011 are compared for all counties for which data was available and for comparison between the countries previously discussed.

Countries discussed earlier to show disparities and relative locations of developing countries.

The GNI data sets also show the lognormal characteristic of a positively skewed distribution (Figures 14 & 15 for 1995 data and Figures 16 & 17 for 2011 data) consistent with outcomes resulting from multiplicative effects discussed previously. Actual values (red lines) from most of the 186 countries which have received a HDI ranking and for which GNI data was available, were sorted into 40 intervals chosen for optimum histogram display (using the previously mentioned Input Analyser utility). The total numbers of scores are shown in Table 3. Best fitting theoretical lognormal functions (blue lines) to the empirical data provide visible indication of goodness of fit. Both frequency functions (Figures 14 and 16) and cumulative frequency functions (Figures 15 and 17) provide reasonably confirming visibility tests for the claim of lognormal fitting to the GNI data.

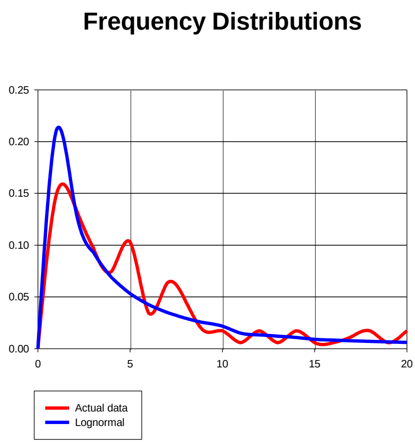

Figure 14 Frequency distribution of 1995 per capita GNI data with the red curve indicating the empirical data and blue the closest fit lognormal curve.

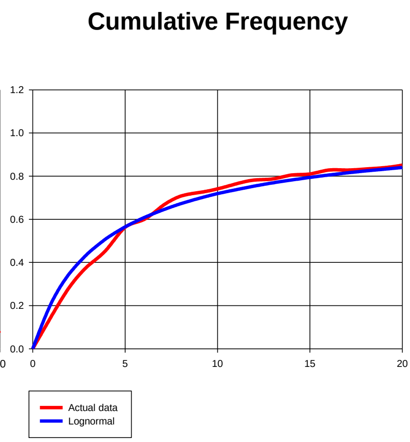

Figure 15 Corresponding cumulative frequency distribution of 1995 per capita GNI data.

Comparison of the two sets of data (1995 & 2011), at least for the frequency (probability) functions, tends to suggest, at least from visibility, an improved fit for the 2011 data.

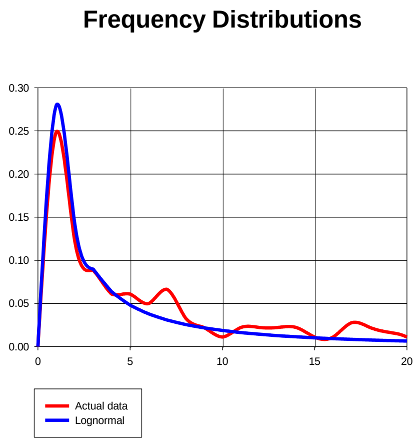

Figure 16 Frequency distribution of 2011 per capita GNI data with the red curve indicating the empirical data and blue the closest fit lognormal curve.

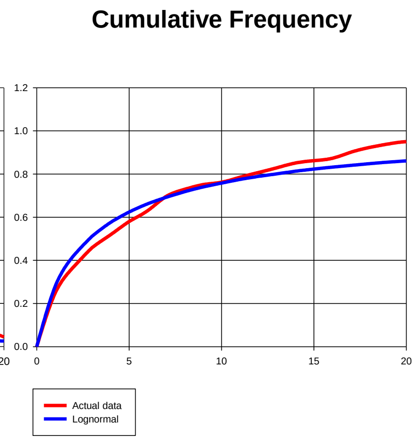

Figure 17 Corresponding cumulative frequency distribution of 2011 per capita GNI data.

Whilst the visibility tests provided so far might be reasonably convincing, statistical tests are also available for more objective confirmation of any possible claims which might be made for these distributions.

Other candidate distributions

Simulation		1995		2011	
Function	Sq. Error	Function	Sq Error	Function	Sq Error
Lognormal	0.000671	Weibull	0.00546	Beta	0.00236
Weibull	0.00252	Gamma	0.00576	Weibull	0.00395
Gamma	0.00323	Erlang	0.00742	Lognormal	0.00435
Erlang	0.00325	Lognormal	0.00926	Gamma	0.00446
Beta	0.0174	Beta	0.0126	Erlang	0.0146

Table 4 Various asymmetric distributions fitted to simulated and HDI data with an error term (mean square error) indicating the closeness of fit**.** The results for the simulation data come from the 5000 run spreadsheet generated cases.

It needs to be acknowledged that there are numerous other statistical distributions which model positively skewed data such as the HDI data discussed in this paper. Relative degree of fittings of the data to candidate distributions can be estimated with a mean square error term (Sq. Error in Table 4, with error terms generated by Input Analyser referred to previously)

Clearly, whilst the simulated data is best fitted with the lognormal distribution, there are other distribution functions which provide better fits to the HDI 1995 and 2011 empirical data than the lognormal despite the positive indications of the "visibility tests" referred to above**.** Thus, whilst the lognormal distribution may not provide the best fit, further tests can be applied to determine if the data is at least consistent with that distribution.

Statistical tests for Goodness of Fit

In the statistical tests conducted here, the following notation is used:

Ho: (Null Hypothesis) the data follow the lognormal distribution
H_A: (Alternative Hypothesis) the data do not follow the lognormal distribution

We assume the null hypothesis (Ho) that the pairs of data sets (simulated & empirical) come from the same theoretical distribution (lognormal) and then apply statistical tests such as the Kolmogorov-Smirnov (KS) Test and the Chi-Squared ($\chi 2$) goodness of fit test which are conveniently available. These tests are used to determine if sample data are consistent with a specified distribution function, in this case empirical data with the lognormal distribution.

The p-values obtained from these tests provide an estimate of the probability of obtaining a test statistic (a measure of the difference between the empirical data and the fitted distribution) as extreme as that obtained, under the null hypothesis (Ho) that all data sets (in this case the best fitted lognormal distribution and, in turn, each of the two empirical data sets) come from the same distribution.

It is clear from the results provided (Table 5) that p-values are small compared with minimum values of p (> 0.1) required for the null hypothesis to be supported. Thus the null hypothesis is consistently rejected and the tests cannot support the null hypothesis that any of these sets of data are best described by the lognormal distribution to a required confidence level.

	Simulated data		**Empirical (1995)**		**Empirical (2011)**	
Test	**Test statistic**	**p-value**	**Test statistic**	**p-value**	**Test statistic**	**p-value**
K-S	0.0227	0.0123	0.0941	0.0892	0.0976	0.0622
Rejection level		strong		Low		Low
Data points	5000		174		181	
$\chi 2$	32.1	0.0324	19.1	<0.005	21.3	<0.005
DoF	19		6		6	
Rejection level		strong		Very strong		Very strong

Table 5 Results of statistical tests **(**Kolmogorov-Smirnov (KS) and Chi-Squared ($\chi 2$)) used to determine if data can be confirmed as being well modeled by the lognormal distribution**.** The results for the simulation data come from the 5000 run spreadsheet generated cases.

Thus the hypothesis of this paper that the empirical data follows the lognormal distribution and that multiplicative effects (law of proportionate effects discussed earlier) are a factor in worldwide HDI data is left to be supported only by the “visibility tests” and the general observations (HDR Overview p 6 quoted above) of the unequal effects of adverse conditions on poor countries. It should also be noted that the tests applied above may use criteria that are too conservative (as suggested by the strong support of the “visibility tests” referred to earlier) and so may hide the appearance of real effects.

Conclusion

Probability or frequency distributions exhibit intimate relationships which make explicit their properties, underlying assumptions and the nature of the causes which produce such distributions in real systems. Knowing the physical and other characteristics of an entity exhibiting random behavior, a suitable choice of distribution function may be made to model that behavior. If empirical data on an entity such as HDI data can be fitted to a particular distribution, hypotheses may be established concerning the underlying causes of the behavior of the entity. An application of this has been made by suggesting possible hypotheses concerning causes of empirically determined HDI data by attempting to fit that data to a lognormal distribution.

Attempt at this data fitting was motivated by observations in the HDR (2013) consistent with the operation of multiplicative factors in determining relative HDI values across many countries for which data is available. Whilst visibility data (Figures 6 to 18) seemed to support the hypothesis of the paper, the more objective statistical tests did not.

Further questions to be explored using the process described in this paper could include investigating possible prevailing factors influencing GNI, and how might they be taken account of in a simulation. For example, consideration could be given to possible effects on the GNI of various deterministic factors such as whether a country is landlocked, experiences high levels of corruption, or is Muslim.

Finally, the world community is still left with the problem of a highly skewed distribution of measures of human wellbeing with large disparities between even neighboring countries. Without redistribution of wealth by taxation, which is possible within a nation, perhaps the nations of the world need to take seriously the notion of the periodic application of international debt relief, to overcome the inexorable effects of what at least has the appearance of a lognormal process.

References

Aitchison, J., & Brown, J.A.C. (1969). *The Log-normal Distribution*. UK: Cambridge Uni. Press.

Crow, E.L., & Shimizu, K., (1988). *Log-normal Distribution, Theory and applications*. New York: Mariel Dekker.

Hahn, G.J., & Shapiro, S.S., (1967). *Statistical Models in Engineering.* New York: John Wiley.

Human Development Report 2013: The Rise of the South: Human Progress in a Diverse World, (2013). UNDP: New York, Retrieved 27 May 2014 from http:// hdr.undp.org.

Kabacoff, R.L., (2011). *R in Action*, US:Manning.

Kelton, W.D., Sadowski, R.P., & Swets, N.B., (2010). *Simulation with Arena (5th Edn)*. Boston, Ma: McGraw- Hill.

Law, A.M., & Kelton, W.D. (1992). *Simulation Modeling and Analysis*. New York: McGraw-Hill.

Lawrence, R.J. (1988). The Lognormal as Event-Time Distribution. In Crow, E.L., & Shimizu,K., (2008). *Log-normal Distribution, Theory and applications* (pp. 211-228). New York: Mariel Dekker.

Manno, I., (1999). *Introduction to the Monte-Carlo Method*, Hungary:Budapest.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill.

Acknowledgements

Special thanks are due to Dr R King, formerly of the University of Western Sydney, who provided a critical review of the text as well as assisting with the R source code. However, all errors of fact or quality of expression must remain with the author.

Author

Prof. Peter K Anderson PhD
Head, Department of Information Systems
Divine Word University
Email: panderson@dwu.ac.pg

Dr Peter K Anderson is Professor and foundation head of the Department of Information Systems at DWU where he specialises in data communications. He holds a PhD in thermodynamic modeling from the University of Queensland. His research interests include exploration of the implication of lognormal distributions.

Glossary

GNI	Per capita Gross National Income
HDR	Human Development Report
HDI	Human Development Index
KS	Kolmogorov-Smirnov Test
PNG	Papua New Guinea
SI	Solomon Islands
UNDP	United Nations Development Project
$\chi 2$	Chi-Squared Goodness of Fit Test

R Source Code

```
# Simulation of 30 increments added over time to a capital of $1000
library(distr) library(MASS)
library(car)
n <- 30;         # no of increments N
<- 5000;       # no of simulations
mult.fac <- 0.03 # needed to
prevent negative incomes

# storage matrix for random normal variates, mean = 0, sd = 1
stoch.incr <- matrix(rnorm(N*n, 0,
1),nrow = N, ncol = n)

# storage matrix for generated incomes - to be overwritten
I <- matrix(0, nrow = N, ncol = n +
1)
I[ ,1] <- 1000 # initial income
$1000 in col 1 of I

# set the random generator so the results are reproducible
set.seed(1271)

#  simulation
 for(i in 1:n){
            I[ ,i + 1] = I[ ,i]* (1 +
mult.fac*stoch.incr[ ,i])
           }

#Result is a 5000 row 31 column matrix
I.final <- I[ ,n+1]# column 31 is
the final 5000 incomes
hist(I.final) # shows the histogram
of simulated data

# results for I.final written to file
setwd("E:/Res_2014/SimulatedDat
aR/Run5000")
filename <-
"E:/Res_2014/SimulatedDataR/Ru
n5000/Rsim.csv"
write.table(I.final, file = filename,
sep = ",")

# find the mean and sd of a lognormal curve fitting # the simulated data: respectively meanlog,sdlog
lnorm.fit <-
fitdistr(I.final,"lognormal")
meanlog <-
lnorm.fit$estimate["meanlog"]  #
6.895
sdlog   <-
lnorm.fit$estimate["sdlog"]    #
0.1664

# 1. Comparison with a lognormal distribution: generate 10000 random lognormal variates from a distribution of mean = meanlog, sd = sdlog
lnrv = rlnorm(5000,meanlog,sdlog)
```

combine histogram and density lines

```
hist(I.final, prob=T)  # various options for histogram not used
lines(density(lnrv), col="red")
```

 # QQplot of simulated data and theoretical lognormal

```
qqPlot(I.final,dist=
"lnorm",meanlog=lnorm.fit$estimate["meanlog"],
sdlog=lnorm.fit$estimate["sdlog"],
xlab=”Theoretical Quantiles”,
ylab=”Simulated Quantiles”)
```