International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: 6.614
Available Online at www.journalijcar.org
Volume 8; Issue 06 (E); June 2019; Page No.19266-19271
DOI: http://dx.doi.org/10.24327/ijcar.2019.19271.3707

Research Article

HEAVY TAILS AND THE FAILURE OF THE CENTRAL LIMIT THEOREM

Thomas B. Fowler

Adjunct Professor of Engineering George Mason University Fairfax, VA USA

A R T I C L E I N F O

Article History:
Received 06th March, 2019
Received in revised form 14th April, 2019
Accepted 23rd May, 2019
Published online 28th June, 2019

Key words:

Heavy tails, Central Limit theorem, Random variables, Financial modeling

A B S T R A C T

Heavy-tailed distributions have become very important in fields as diverse as telecommunications and economics. They often occur in situations where one would expect that the Central Limit Theorem should apply. This article investigates why the Central Limit Theorem fails, and shows one mechanism by which heavy-tailed behavior can arise. This is by addition of what the article defines as "hypercorrelated" random variables. The article also shows that heavy-tailed behavior cannot arise due to addition of linearly related random variables. The failure of the Central Limit Theorem to be applicable in many areas where it has traditionally been assumed to apply has important real-world consequences, especially in finance and financial modeling.

INTRODUCTION

The Central Limit Theorem is one of the most profound and important results in mathematics. It tells us that under fairly general conditions, the net result of many actions or events can reliably be described by a normal distribution. Some common examples are:

- Height of male or female freshman students at a college (result of the combined action of many independent genetic and environmental factors).
- Time to make a certain long trip by automobile (result of weather, road conditions, accidents, and other independent factors)
- Electricity consumption in a city (result of a large number of independent consumers)

Other cases which would seem to fit (but in fact do not) are

- Height of ocean waves (result of the sum of many different forces)
- Stock market prices (result of many trades by independent investors)
- Length of Internet messages (result of many different activities by disparate users)

The Central Limit Theorem is used to estimate the distribution of some quantity when it is the result of many small contributions from various sources. The requirements for the Central Limit Theory to be applicable are:[i]

1. Variables summed must be independent.
2. All variables must have finite mean and variance
3. No variable can make an excessively large contribution to the sum

Despite its success in many instances, the Central Limit Theorem is known to fail as a descriptive tool in cases where its applicability seems assured. In many of these cases, we find that we are dealing with heavy-tailed distributions, and decisions made on the basis of normal distribution characteristics often lead to catastrophic results. Perhaps the most famous recent example is the meltdown of a hedge fund, Long Term Capital Management, in 1998. The fund assumed that certain currency movements would be governed by the normal distribution on account of the (presumed) applicability of the Central Limit Theorem. In reality, a set of movements occurred which, according to the fund managers, was a "10 sigma event", triggering huge leveraged losses which nearly destabilized the global financial system.[ii] Of course, 10 sigma events do not occur, since their probability is 7.62×10^{-24}. Indeed, if the event in question is measured on a daily basis, which includes settlement of most market-based accounts, the event would occur once every 1.31×10^{23} days. This corresponds to about 3.6×10^{20} years—roughly 10 orders of magnitude longer than the age of universe. What happened was that a heavy-tailed distribution, not a normal distribution, governed the phenomenon.

Another case of great interest in the telecommunications arena is the distribution of the file size of Internet messages. Indeed, many parameters associated with Internet traffic are governed by heavy-tailed distributions, including page requests/site, reading time per page, and session duration.[iii] Heavy-tailed distributions cause serious problems in the design of IP-based

Corresponding author:* **Thomas B. Fowler
Adjunct Professor of Engineering George Mason University Fairfax, VA USA

systems, because the usual queuing theory methods employed to determine the required size of network components to meet performance specifications break down when variance and higher-order moments cannot be calculated. It has been necessary to develop special techniques, such as the Transform Approximation Method (TAM) in order to cope with this problem.[iv,v] The TAM utilizes the fact that one does not need actual infinite times and infinite variances, but can utilize finite approximations to get useful results.

What goes wrong, and why? Which of the three conditions is the problem? How can heavy tails emerge from sums of small-contribution random variables? We must examine the three conditions to determine which is likely to fail in practical cases, and how such a failure would manifest itself. Clearly the second condition is not of much interest because if the mean and variance of one or more components do not exist, we already have a heavy tail to all intents and purposes. Nor is the third likely to be of great interest either, because we want the heavy tails to emerge from the sum of many small contributions, not one big one. So the first condition would seem to be an ideal candidate, because correlated variables presumably could have large tails. Presumably feedback mechanisms of some type give rise to the larger-than-expected probabilities of events. However, it is not obvious that this is the cause, because adding random variables with the distributions that are just multiples of each other (100% correlated) just yields a similar distribution with different mean and variance, but not a heavy tail. Thus more is needed than simple correlation. Specifically, what is needed is some type of nonlinear feedback mechanism which can yield large changes. Thus the relationship among the variables being summed would include this kind of nonlinear feedback, and that would ultimately generate the heavy tails.

Heavy-tailed Distribution and Infinite Variance

The size of files sent over the Internet has been determined to have what is known as a "heavy-tailed" probability distribution.[vi] That is, the probability of a given file length occurring falls off very slowly with increasing file length, unlike ordinary distributions, where the rate of fall-off is much faster. Technically, given a probability distribution function *f(x)*, if for large *x* values, its cumulative distribution function *F*(*x*) has the property that its complementary distribution

$$1 - F(x) \approx \kappa_1 x^{-\beta}$$

where $\kappa_1 > 0$ and. $\beta \in [0,2]$, then the distribution function is said to be *heavy-tailed* because it falls off very slowly with increasing values of *x*.[vii,viii] In turn, this has an important consequence. Setting β=2 and differentiating above equation,

$$f(x) = 2\kappa_1 x^{-3}$$

Since $Var(x) = E(x^2) - [E(x)]^2$ and $[E(x)]^2$ is fixed, it follows that the variance is determined by

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) = \int_{-\infty}^{\infty} \frac{2\kappa_1}{x} dx = \infty$$

That is, the Variance is Infinite

The physical significance of infinite variance can be readily understood. Recalling that variance measures central tendency, for any *finite* variance, values are known to be clustered around a central measure, the mean. Increasing the coarseness of the horizontal scale will result in the distribution appearing more "peaked," i.e., clustered around the mean. However, in the case of *infinite* variance, there is no such clustering, and regard-less of the scale on which measurements are made, there is no change in their central tendency — in effect, all scales look the same.[ix] Physically, this means that it is difficult or impossible to put limits on the values of random variable that one may observe. Such values can become arbitrarily large in absolute value with a much higher frequency than is the case with better-behaved distributions such as the normal distribution.

The simplest heavy-tailed distribution (and the most famous in packet network analysis) is the *Pareto distribution,* [7] which has the general form

$$F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}, \quad \alpha, k > 0 \quad x \geq k$$

$$p(x) = \frac{dF(x)}{dx} = \alpha k^{\alpha} x^{-\alpha-1}$$

As α decreases, the "heavy-tail" effect increases. For $\alpha < 1$, the variance becomes infinite. This distribution is a type of negative power law distribution, similar to the one discussed above.

Certifying infinite variance in practical cases is very difficult. In many cases, infinite variance is a much stronger condition than necessary; often a very large variance is enough to cause significant financial, queuing, or other problems. For such cases, the ratio of the standard deviation to the mean can become quite large. A common way to identify heavy tails is to look for a falloff of the density function that takes the form of a power law (straight line on a logarithmic plot). This can be easily recognized (see Figure 1).

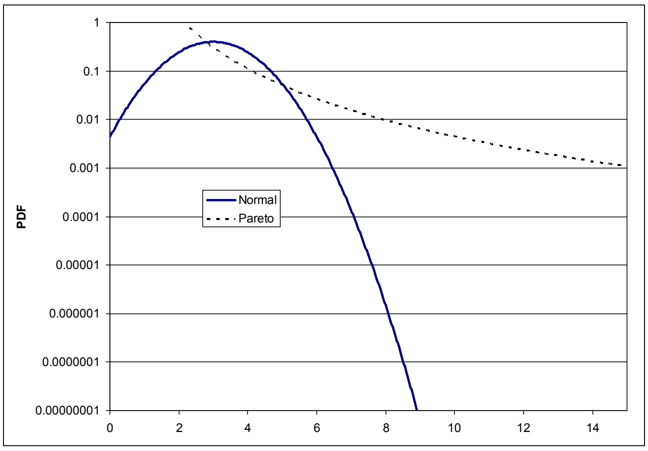


Figure 1 Comparison of normal and Pareto distribution with respect to falloff for large values of *x*. Note the approximately linear power law behavior of the Pareto distribution for larger values along the abscissa.

Failure of the Central Limit Theorem

To understand how the failure of Central Limit Theorem manifests itself, and why more is required than simple correlation of random variables added, consider the following two cases:

Random variables *X*, *Y* are completely correlated, i.e., □ = 1. In that case, $Y = aX + b$. A typical data set might be:

X	1	2	3	4
Y	2	4	6	8

Here $Y = 2X$.

Random variables *X*, *Y* which are independent, but which have similar distributions:

X: (μ,σ^2)
Y: $(2\mu, 4\sigma^2)$

In this case, though mean values are related as $\mu_Y = 2\mu_X$, correlation $\rho = 0$.

Now observe what happens when these are added. Let $Z = X + Y$. Then for the two cases:

$\mu_Z = \mu_X + \mu_Y$ since expected value is always additive. Thus $\mu_Z = \mu_X + 2\mu_X = 3\mu_X$

For the variance we have

$$Z^2 = (X+Y)^2 = X^2 + 2XY + Y^2$$

$$\begin{aligned} E(Z^2) &= E(X^2) + E(2XY) + E(Y^2) \\ &= E(X^2) + E(2X \cdot 2X) + E([2X]^2) \\ &= E(X^2) + E(4X^2) + E(4X^2) \\ &= E(9X^2) = 9E(X^2) \end{aligned}$$

Then Var (Z) can be calculated as

$$\begin{aligned} Var(Z) &= E(Z^2) - [E(Z)]^2 \\ &= 9E(X^2) - [3E(X)]^2 \\ &= 9E(X^2) - 9[E(X)]^2 \\ &= 9Var(X) \end{aligned}$$

$$\mu_Z = \mu_X + \mu_Y$$
$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$$

For the above case this becomes

$$\mu_Z = \mu_X + \mu_Y = \mu_X + 2\mu_X = 3\mu_X$$
$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 = \sigma_X^2 + 4\sigma_X^2 = 5\sigma_X^2$$

Note that in the calculation for variance of the correlated variables, the middle term $E(2XY)$ does not drop out as it would for independent variables, since in general for correlated variables $E(XY) \neq E(X)E(Y)$. In the case of independent variables we have

$$\begin{aligned} V(X+Y) &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= E(X^2 + 2XY + Y^2) - [E(X) + E(Y)]^2 \\ &= E(X^2) + 2E(X)E(Y) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \\ &= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 \\ &= V(X) + V(Y) \end{aligned}$$

In general case, for *n* random variables $X_1 \ldots X_n$, where $X_i = a_i X_1$, $a_i > 0$ for all a_i and $Z = \sum_{i=1}^{n} X_i$, we have

$$E(Z) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} E(a_i X_1) = \sum_{i=1}^{n} a_i E(X_1) = E(X_1)\sum_{i=1}^{n} a_i$$

$$V(Z) = \left(\sum_{i=1}^{n} a_i\right)^2 V(X_1) \text{ or } \sigma_Z = \sigma_{X_1} \sum_{i=1}^{n} a_i$$

In effect, this says that the *standard deviation scales at the same rate as the mean*. Thus the *shape* of the distribution is unchanged, and specifically, it does not exhibit the flattening required for a heavy tail.

But if the random variables are independent, $E(Z)$ is the same but

$$V^*(Z) = \sum_{i=1}^{n} a_i^2 V(X_1) < V(Z)$$

That is, standard deviation scales as $\sqrt{\sum_{i=1}^{n} a_i^2}$, which is *less* than the rate at which the mean scales. This implies that the shape of the distribution is being compressed—exactly the opposite behavior to that required for heavy tails. Naturally such a result is just what we would expect, since the Central Limit Theorem governs this case.

In general, comparing these, we have

$$\frac{V(Z)}{V^*(Z)} = \frac{\left(\sum_{i=1}^{n} a_i\right)^2}{\sum_{i=1}^{n} a_i^2} > 1$$

This is true for variables regardless of their distribution, so it shows that *heavy tails cannot arise by linear combinations of correlated random variables*. The new summed distribution has the same shape as the old, only scaled up. When converted to a pdf, it will not have a heavy tail, defined as variance tending to ever larger values in the pdf. Obviously, adding independent random variables will not lead to heavy tails because the standard deviation scales at a smaller rate than the mean, so the distribution tends to contract. The two situations are illustrated in Figures 2 to 4.

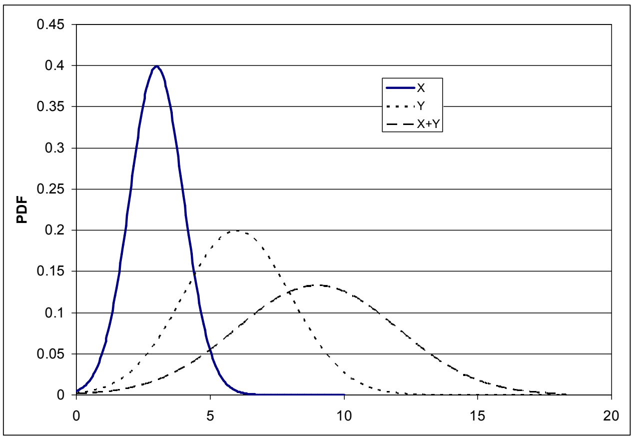


Figure 2 Two correlated random variables and their sum

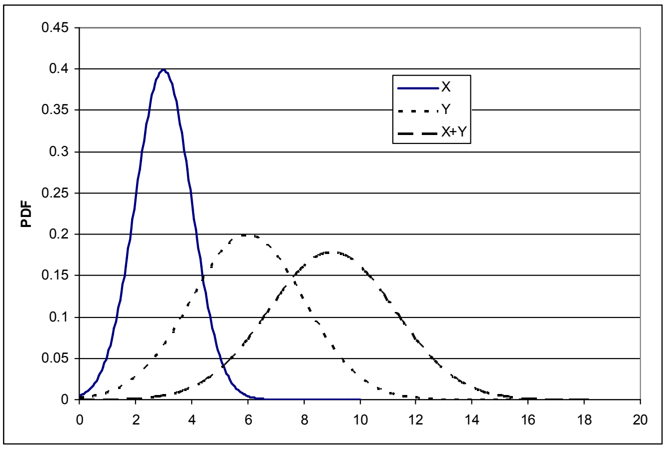


Figure 3 Two independent random variables and their sum

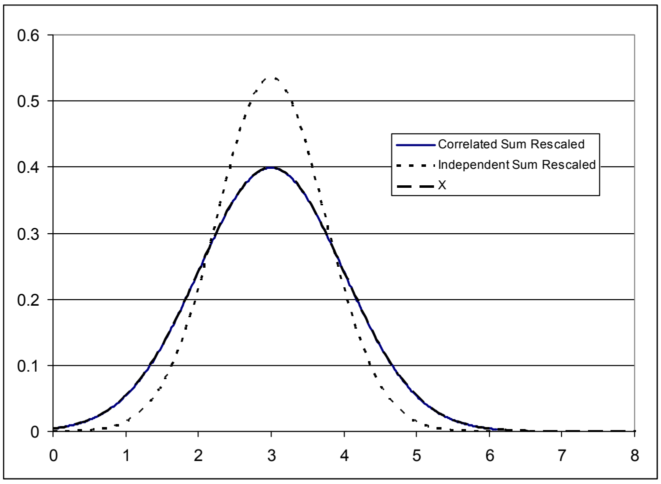


Figure 4 Rescaled sums compared to original random variable *X*. Note that the correlated sum and *X* overlap.

Hypercorrelation

However, the fact that the variance of a sum of correlated random variables is greater than the corresponding sum for independent random variables suggests the next step. If normal correlation cannot create heavy tails, but only replicate the shape of the original distribution, perhaps some type of *amplified* correlation can do so. One way in which an "amplification" can occur in correlated distributions is if a power-law relationship exists for the variables. So we shall consider relationships of the form $Y = aX^n$, and we will define ρ' as the standard correlation coefficient ρ for the natural log of variables *Y*, *X* where the values of *Y* and *X* have been fitted using an exponential regression, so that *a* and *n* are known. That is, ρ' will measure the correlation for $\ln Y = \ln a + n \ln X$. This is equivalent to finding ρ for *W* and *Z* where $W = a' + b'Z$. Then the *hypercorrelation* between *X* and *Y* is given by $\rho^* = n\rho'$. Thus if ρ' is 0, there is no hypercorrelation. Conversely, if ρ' is 1, there is hypercorrelation of *n*.

It is clear that *n* measures the strength of the amplification effect, i.e., how much faster *Y* increases than *X*. In real-world situations, we would want to assume a distribution for *a* and a distribution for *n*, and then determine under what circumstances the heavy-tailed behavior would arise. This could be done by fitting the tail of the summed hypercorrelated random variables to a Pareto distribution.

Specifically, let us assume that we have a large number of hypercorrelated variables Y_i, such that $Y_i = a_i X^{n_i}$, each of which has an associated hypercorrelation ρ_i^*. We should be able to show that if a large number of Y_i's are summed, $\sum_{i=1}^{k} \rho_i^* / k$ increases, and standard deviation./mean continues to increase, and thus heavy tails emerge since the sum of the Y_i's is bounded on the lower end.

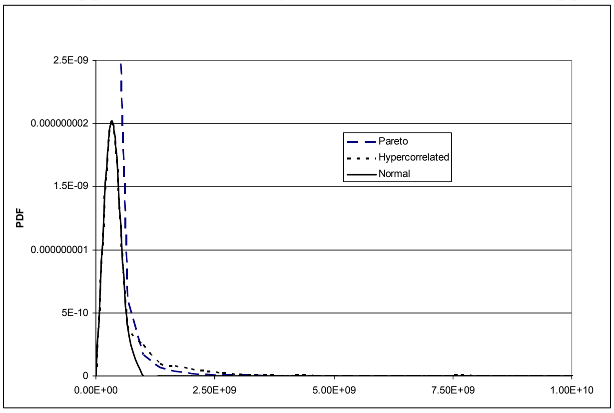
Simulation

Efforts to verify the creation of heavy-tailed distributions by means of hypercorrelation have thus far been done by means of numerical simulation. In these experiments, hypercorrelated random variables are generated, and then added. The resulting distributions are then analyzed for heavy-tailed behavior. Such behavior is quite obvious on logarithmic plots, especially when compared to usual result of adding random variables, namely a normal distribution. The procedure is as follows:

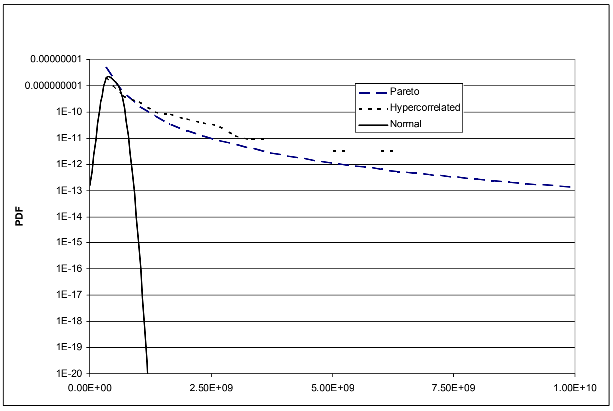
1. Assume a functional form of $Y_{ij} = a_i X^{n_j}$
2. Generate 5 normally distributed random values for the a_i, and for each of those, 10 normally distributed random values for the n_j.
3. Generate a normally distributed set of values for *X*.
4. Calculate the Y_{ij} terms and add, $Y_i = \sum_{j=1}^{10} Y_{ij}$
5. Calculate the distribution of the sum $Z = \sum_{i=1}^{5} Y_i$
6. Fit a normal and a Pareto distribution to the distribution for *Z*
7. Record Pareto parameters and average hypercorrelation
8. Repeat for new set of *a, n* values with different mean, standard deviation.

Some typical graphs resulting from step 6 are shown in Figures 5 and 6 on both linear and log scales. These graphs illustrate clearly the heavy-tailed behavior of the summed hypercorrelated variables, as compared to a normal distribution. Note that in Figure 5(a), the normal and hypercorrelated distributions look almost identical until well past the mean; this could easily fool an observer into thinking that the process in question was really normally distributed when in fact it has a heavy tail. The difference, of course, is immediately obvious when the distribution is plotted on a logarithmic scale. A smaller value of *n*, which corresponds to less hypercorrelation and presumably less heavy-tailed behavior, gives correspondingly a higher value for the Pareto constant α. Note also in Figure 6 that the degree of heavy-tailed behavior is much lower than in Figure 5, as illustrated by the decreased divergence between the hypercorrelated (and Pareto) curves and the normal curve. This result is expected, since the exponent *n* in the runs for Figure 5 is 6, whereas it is

2 for the runs of Figure 6. As the exponent n approaches 1 from the upper side, the heavy-tailed behavior will disappear.

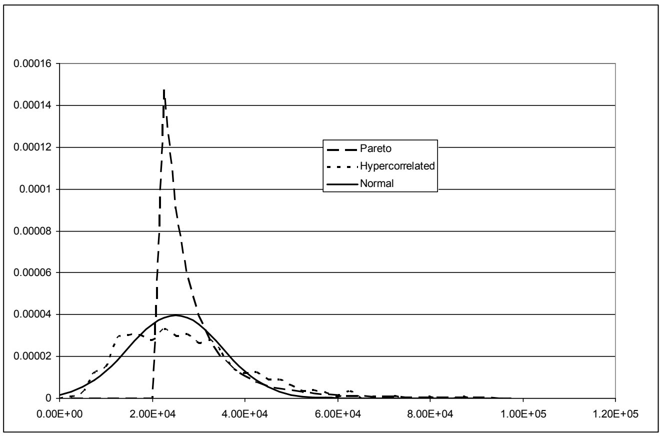


(a) Linear scale

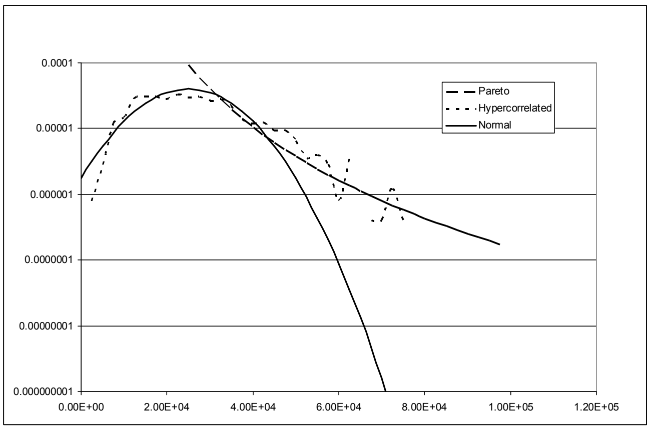


(b) Log scale

Figure 5 Summed hypercorrelated variables. Distribution of a: : μ=10, σ=3. Distribution of n: μ=6, σ=2. Average hypercorrelation value: 5.923; Pareto constant α=2.1.



(a)



(b)

Figure 6 Summed hypercorrelated variables. Distribution of a: μ=10, σ=3 Distribution of n: μ=3, σ=1. Average hypercorrelation value: 1.868; Pareto constant α=3.6.

A summary plot of many runs, given in Figure 7, shows the relationship between average hypercorrelation value and Pareto constant α. As expected, the relationship is inverse but well-defined, as indicated by the regression line.

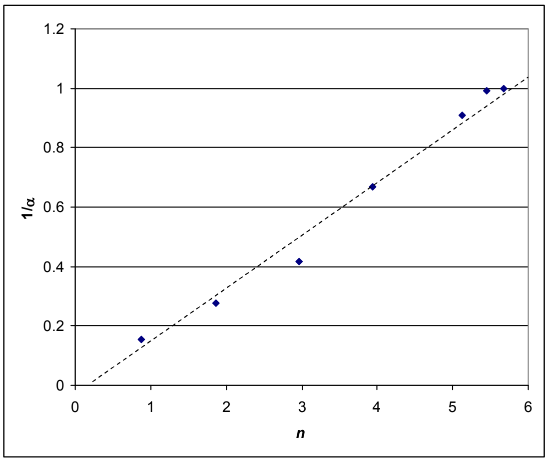


Figure 7 Relationship between hypercorrelation n and Pareto constant □ based on simulation runs.

Similarly, the ratio of standard deviation to mean varies from 1.39 to 18.31 over the range of n. These results furnish a demonstration of the emergence of heavy-tailed behavior by means of hypercorrelation. Naturally, no claim is made that *all* instances of heavy tailed distributions arise in this way, only that this is one way in which the Central Limit Theorem can fail, resulting in heavy-tailed distributions.

How Heavy Tails Might Arise Through Hypercorrelation

It is useful to consider how heavy tails due to hypercorrelation might arise. First, consider the case of investors, whose behavior is known to be subject to crowd influences ("if everyone is doing it, I should too!"). So Investor 1 says, "I'm buying 10 shares of company *A*." Investor 2 says, "I see what investor 1 is doing, and I like it, so I'm going to beat him and buy 100 shares of company *A*." Investor 3 says, "Everybody else is doing it, so I'm going to beat them and buy 1,000 shares of company *A*!" Thus if the number of shares bought is increasing by a factor, so that we have N, N^2, N^3, etc., all added together, then we have a hypergeometric distribution sum and a heavy tail condition. Heavy-tailed behavior can be particularly dangerous in today's financial environment, where the amount of derivatives in existence is estimated at about $500 trillion—about 10 times the world's annual GDP. If hypercorrelated linkages exists in some of the derivatives—and very little is known of how they would behave in a crisis situation—then the results could be catastrophic. Since hypercorrelated behavior can mimic ordinary normal behavior under many circumstances, it might lie there undiscovered.

In the case of Internet traffic, the psychology of users and its impact on traffic has not been well-studied, and the heavy-tailed nature of the traffic may have multiple sources. But some ideas come to mind. If success breeds a desire for more of the same, for instance, if successful retrieval of one web page leads to requests for larger numbers by the same user, or pages with more information, a type of hypercorrelation will emerge. As an example, let us consider how most commercial sales web sites are setup. If a user is seeking information about a product that may be of interest, he would first arrive at a summary page which gives highlights of the product. Such a page is typically designed to load fast. Then, the user might want to drill down a bit, and ask for a more detailed specifications page. Later, he might ask for a product brochure download, or a page with photos or images of some type. The ratio of file size here could easily be 1:5:25, or something on that order. This would explain at least in part the heavy-tailed character of traffic, page requests/site, reading time/page, and session duration.

Conclusion and Future Work

Heavy-tailed probability density functions can arise through summations of hypercorrelated variables, which are variables of the general form $Y = aX^n$. The larger the value of *n*, the greater the degree of heavy-tailed behavior. Distributions of such sums, $Z = \sum_i Y_i$, can often mimic normal distributions for certain values *Z*. Therefore application of the Central Limit Theorem must be carefully monitored to ensure that hypercorrelation is not present. Heavy-tailed behavior can be made more apparent by plotting the sums on a logarithmic scale. This research only deals with the case of heavy-tailed distributions arising from sums of random variables; heavy tails may also originate in other ways. Future work will concentrate on a more theoretical understanding of the emergence of heavy tails from hypercorrelated random variables. It will also look at empirical reasons for the emergence of heavy-tailed behavior, to better understand the circumstances under which such behavior can be expected.

Referance

1. Paul Meyer, *Introductory Probability and Statistical Applications*, 2nd edition, Reading, MA: Addison-Wesley, 1970, p. 251.
2. http://www.investopedia.com/terms/l/longtermcapital.asp.
3. Thomas Fowler, "A Short Tutorial on Fractals and Internet Traffic," *The Telecommunications Review*, 1999.
4. Martin Fischer, Carl Harris, "A Method for Analyzing Congestion In Pareto and Related Queues", *The Telecommunications Review*, 1999.
5. Martin Fischer, Donald Gross, Denise Masi, John Shortle, "Analyzing the Waiting Time Process in Internet Queueing Systems With the Transform Approximation Method", *The Telecommunications Review*, 2001, p. 21-31.
6. Crovella, M. and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Transactions on Networking*, 5(6); pp. 835-846, December 1997
7. Willinger, W., and V. Paxson, "Where Mathematics Meets the Internet," *Notices of the American Mathematical Society*, Vol. 45, No.8, pp. 961-970, August 1998.
8. Crovella, M. and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes,"
9. *IEEE/ACM Transactions on Networking*, 5(6); pp. 835-846, December 1997.
10. Thomas Fowler, "A Short Tutorial on Fractals and Internet Traffic," *The Telecommunications Review*, 1999.

How to cite this article:

Thomas B. Fowler (2019) 'Heavy Tails and the Failure of the Central Limit Theorem', *International Journal of Current Advanced Research,* 08(06), pp. 19266-19271. DOI: http://dx.doi.org/10.24327/ijcar.2019.19271.3707

****