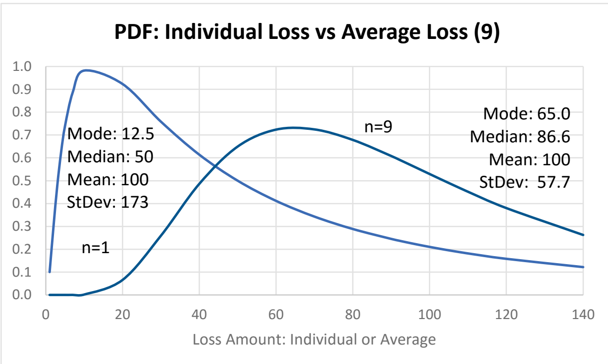
Averaging Log-Normal Losses
Different Sample Sizes

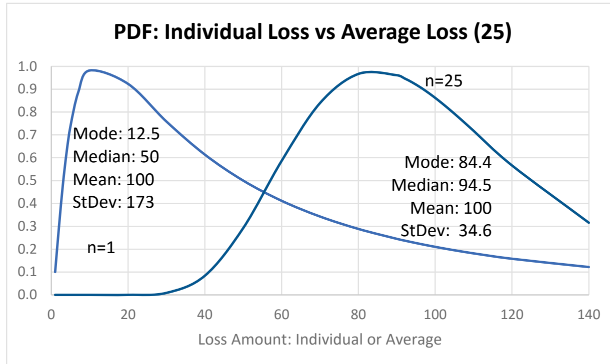
Individuals vs. Aggregate Average Insurance Losses

Suppose that individual losses have a log-normal distribution with a median of 50 and a mean of 100. The most likely loss (mode) is 12.5. The standard deviation around the mean is a massive 173 – due to the long right tail. If risk adverse individuals wanted to set aside savings to cover risks up to the 95th percentile, they would need to set aside almost 2.5 times their expected loss.

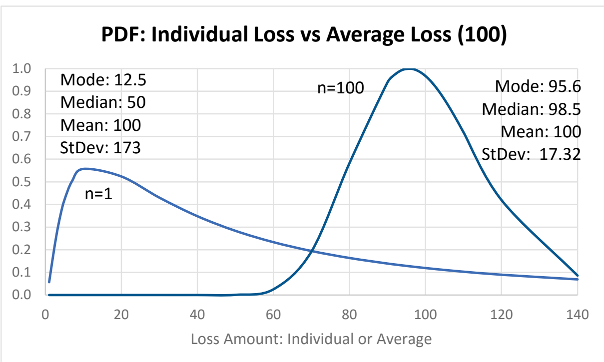
Now suppose an insurance company sells policies agreeing to cover all such losses. Their average loss (mean) is always 100.0. The big difference is in the spread – the standard deviation.

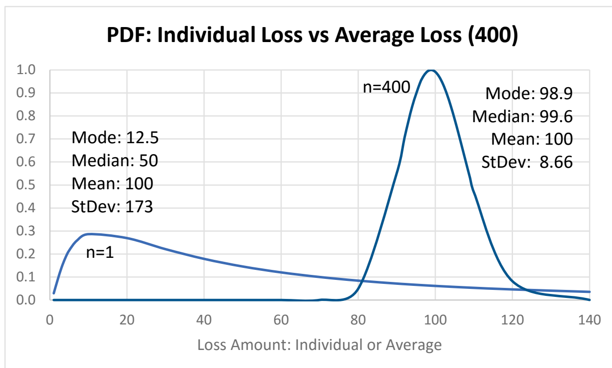
The distribution of individual and average losses (#) for n=9 and n = 25



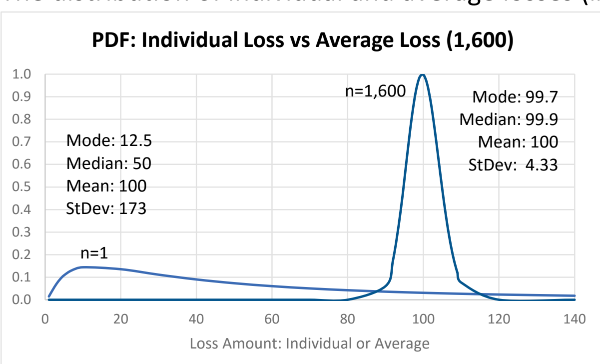


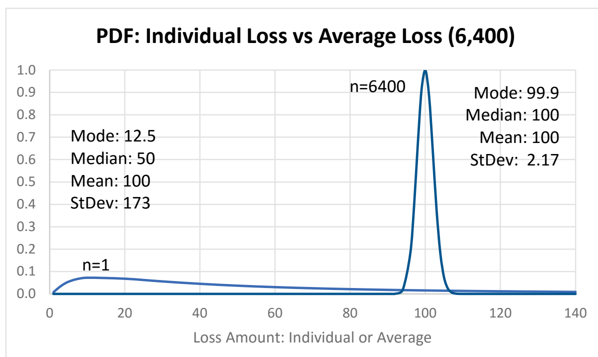
The distribution of individual and average losses (#) for n=100 and n = 400:





The distribution of individual and average losses (#) for n=1,600 and n = 6,400:





Averaging Log-Normal Losses
Different Sample Sizes

BENEFIT: As the number of losses increases, the standard deviation decreases. This is what enables insurance companies to sell individuals a policy with a small premium above the cost of expected losses. That premium includes the selling commission and the company's reserve for losses in excess of expected. To see how excess loss reserve varies with the number of losses, consider this table.

Data Summary:

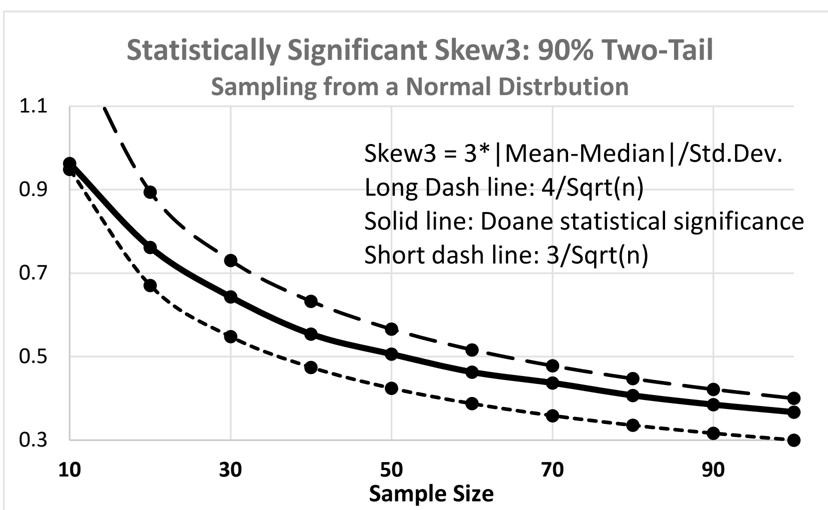
A	B	C	D	E	F	G	H
# Sample	Mode	Median	Mean	Std.Error	Skew3	Sigma/Sqrt(n)	Coef.Var
1	12.5	50	100	173	0.866	173.2	1.73
9	65.0	86.6	100	57.74	0.696	57.74	0.577
25	84.4	94.5	100	34.64	0.477	34.64	0.346
100	95.7	98.5	100	17.32	0.254	17.32	0.173
400	98.9	99.7	100	8.66	0.129	8.66	0.087
1,600	99.7	99.9	100	4.33	0.065	4.33	0.043
6,400	99.9	100	100	2.17	0.032	2.17	0.022

Note that as the sample size increases: (1) the mode increases and approaches the median, (2) the median increases and approaches the mean, and (3) the skew decreases and approaches zero.

Statisticians see the log-normal slowly turning into a symmetric (normal) distribution. Statisticians see this process obeying the Central Limit Theorem since the Standard Error always equals Sigma / Sqrt(n).

Note that the skew (labelled Skew3) is defined as the 3*(mean-median)/standard deviation. Statisticians include the factor of three to align it with a more technical definition of skew.

Using simulation, Doane and Seward (2011) determined the minimum value for Skew3 to be statistically significant when sampling from a Normal distribution. From their exact values, there are two simple conditions: one necessary; the other sufficient. If the population is Normal, a Skew3 of 3/sqrt(n) is necessary (4/sqrt(n) is sufficient) for a 10% level of significance with two tails for 10 < n < 100.



Source: Schield (2016). www.StatLit.org/pdf/2016-Schield-Statisticsl-Significance-Skew-from-Normal.pdf

Doane, David and Lori Seward (2011). Source:
http://www.amstat.org/publications/jse/v19n2/doane.pdf

ANALYSIS OF SKEW3:

On that basis, none of the Skew3 values shown above are statistically significant at the 10% level. This seems counter-intuitive since the log-normal is by its nature skewed right. Sample size in this context means *n* random values from the underlying distribution – not the entire distribution.

So for n=9, imagine sampling nine times, getting nine data values such that they had a mean of 100, a median of 86.6 and a standard deviation of 57.74. Here are nine data values that fit that condition:

20, 50, 60, 70, 86.6, 120, 140, 160 and 197.5.

In looking at just these nine data values, could one reject the null hypothesis that they were obtained from a normal distribution with a mean of 100 and a standard deviation of 57.74? Probably not. Note that we are not choosing between two competing hypotheses: a normal distribution vs. a log-normal distribution.

LOSS RESERVES

The loss reserves of the insurance company can be divided into two parts: the expected (the mean amount per loss) and the excess (coverage between the expected and some percentile).

The value of these excess loss reserves cannot be obtained from the distribution of average losses. The distribution of average losses is a distribution of paid losses – a count. It is not the distribution of the total loss amount – a dollar amount.

But, it is quite obvious that whatever the amount of the excess loss reserve, it decreases as the standard error of the distribution of losses decreases – which occurs as the sample size increases.

A plausible estimate of that amount – as a fraction of the expected losses -- might be the coefficient of variation – the standard error of the sampling distribution divided by the mean (the expected value).

Source: http://www.statlit.org/pdf/2016-Schield-Sum-Identical-LogNormals-FW-Excel2013-Demo.pdf