Properties of Probability Models: Part Three

What they forgot to tell you about the Lognormals

Donald J. Wheeler

The best analysis is the simplest analysis that provides the needed insight. Of course this requires enough knowledge to strike a balance between the needed simplicity and unnecessary complexity.   In parts one and two of this series we looked at the properties of Weibull and Gamma probability models and discovered some unexpected characteristics for each of these families. Here we shall examine some basic properties of the family of Lognormal models.

As we have seen, Weibull and Gamma distributions have an elongated tail, with the elongation increasing with the skewness. Yet contrary to expectation, this elongation does not increase the area in the tail. In fact just the opposite, it is the area in the central portion, the area within one standard deviation of the mean, that increases with increasing skewness. Here we consider if this property of Weibulls and Gammas also holds true for Lognormal probability models.

THE REQUIREMENTS FOR A LOGNORMAL VARIABLE

Lognormal distributions should not be used as often as they are.

While many people have gotten used to using a lognormal model whenever their histogram shows an elongated upper tail, there are certain stringent assumptions that are inherent in the use of a lognormal distribution. To explain this we need to go back to one of the foundations of modern statistics, Laplace's Central Limit Theorem.

In 1810 Laplace proved that when a variable, Y, is the *sum* of many different cause-and-effect relationships it will tend to have a normal distribution. This ability to characterize the distribution of a sum is the foundation of many statistical techniques. It also explains why our data so often have some sort of mound-shaped histogram since most observations can be thought of as the sum of many different sources of variation beginning with the effects of raw materials, continuing with the effects of production and operations, and finishing with effects of the measurement process.

Now consider what would happen if a variable, X, happened to be the *product of* many different cause-and-effect relationships. That is, the value for X is the product of the values for A, B, C, etc:

$$X = A \ x \ B \ x \ C \ x \ D \ x \ E \ x \dots$$

Then the logarithm of X, $Y = \ln(X)$ would be the sum of the logarithms of these different cause-and-effect relationships:

$$Y = \ln(X) = \ln(A) + \ln(B) + \ln(C) + \ln(D) + \ln(E) + \dots$$

and as a consequence of Laplace's theorem, Y will be approximately normally distributed.

Because of this bit of mathematics, the *assumption* that a variable X is lognormally distributed

places some very stringent requirements upon *X*. Specifically, the assumption that *X* is lognormally distributed requires that *every* source of variation that affects *X* must operate in a multiplicative manner. This includes measurement errors, sources of variation in the production of the items measured, and variation in the raw materials used. The effects of all of these causes of variation must operate multiplicatively to produce an observed value for *X* in order for *X* to be a lognormal random variable. If the value for *X* is not the *product* of *all* of these cause-and-effect relationships, if one or more of the cause-and-effect relationships are additive rather than being multiplicative in nature, then *X* may not be lognormally distributed. In my experience this requirement for lognormal distributions is rarely satisfied, in fact, it is rarely even mentioned, yet it is part and parcel of the mathematics behind all lognormal distributions.

Since using models without regard for their mathematical foundations is not good form, you should be very careful about using a lognormal probability model. Any use of a Lognormal model must be justifiable in terms of the actual physical process that generates the observations. Having a skewed histogram is not an adequate basis for using a lognormal probability model. (After all, skewed histograms are often the result of a process that is moving around.) Nevertheless, because various software packages encourage their users to try to fit Lognormal models to all sorts of data, we will consider the properties of the family of lognormal distributions.

THE LOGNORMAL FAMILY OF DISTRIBUTIONS

While software facilitates the use of Lognormal distributions, the following formulas are given here in the interest of clarity. If *X* is a lognormally distributed random variable with parameters alpha and beta, then $Y = \ln(X)$ will be a normally distributed variable with:

$$MEAN(Y) = \ln(\alpha) \quad \text{and} \quad SD(Y) = \beta.$$

The mean for the lognormal variable *X* will be:

$$MEAN(X) = \alpha \, \exp(0.5\,\beta^2)$$

while the variance parameter for the lognormal variable *X* will be:

$$VAR(X) = \alpha^2\,[\exp(2\,\beta^2) - \exp(\beta^2)]$$

And the cumulative distribution function of *X*, *F(x)*, will be found by evaluating the cumulative standard normal distribution *F(z)* at the point:

$$z = \frac{\ln(x) - \ln(\alpha)}{\beta}$$

While the value for the alpha parameter defines both the median and the scale for the distribution of *X*, it is the value for the beta parameter that defines the shape of the distribution of *X*. The skewness and kurtosis of the Lognormal distribution will increase as beta increases. Figure 1 shows the standardized versions of the lognormal distribution for beta values of 0.25, 0.50, 0.75, 1.00 and 1.25. While all lognormal distributions are said to be mound-shaped, Figure 1 shows that the distinction between J-shaped and mound-shaped blurs for large values of beta.

To compare different Lognormal distributions Figure 2 uses 18 different models with beta values ranging from 0.10 to 1.50. For each model we have the skewness and kurtosis, the areas

within fixed-width central intervals (encompassing one, two, and three standard deviations on either side of the mean), and the z-score for the 99.9th percentile of the model.
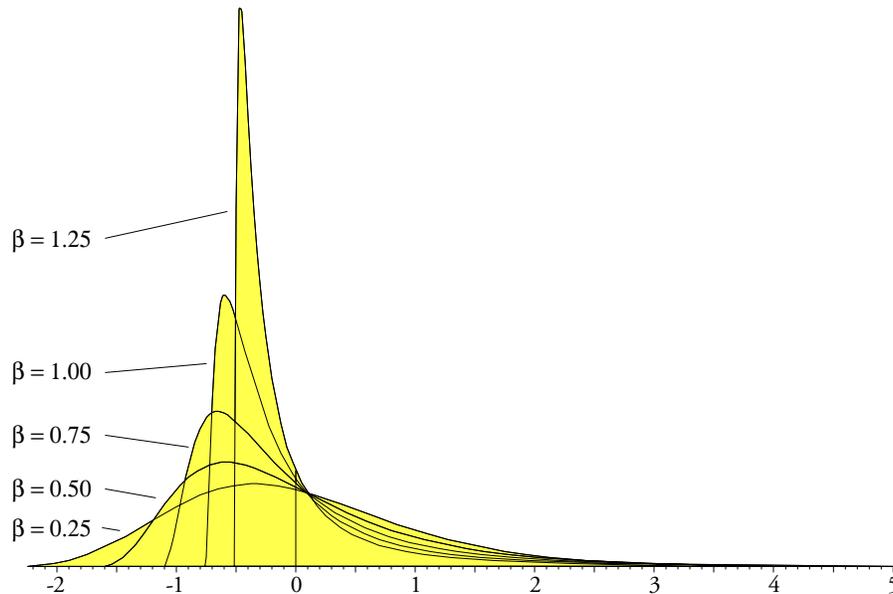

Figure 1: Five Standardized Lognormal Distributions

Lognormal Parameter Beta	Skewness	Kurtosis	Fixed-Width Central Intervals			Most Extreme ppt z-score
			Area Within One SD	Area Within Two SD	Area Within Three SD	
0.10	0.30	3.16	0.686	0.955	0.996	3.54
0.20	0.61	3.68	0.694	0.957	0.993	4.05
0.25	0.78	4.10	0.701	0.958	0.992	4.33
0.30	0.95	4.64	0.709	0.958	0.990	4.61
0.40	1.32	6.26	0.732	0.957	0.987	5.23
0.50	1.75	8.90	0.764	0.956	0.985	5.89
0.60	2.26	13.3	0.805	0.955	0.983	6.58
0.70	2.89	20.8	0.854	0.956	0.982	7.30
0.75	3.26	26.5	0.877	0.957	0.981	7.67
0.80	3.69	34.4	0.891	0.958	0.981	8.03
0.90	4.75	60.4	0.900	0.960	0.981	8.74
1.00	6.18	114	0.909	0.963	0.982	9.41
1.10	8.21	233	0.919	0.966	0.983	10.00
1.20	11.2	518	0.927	0.969	0.984	10.50
1.25	13.1	800	0.932	0.971	0.985	10.70
1.30	15.6	1266	0.936	0.972	0.985	10.87
1.40	22.5	3404	0.944	0.976	0.987	11.09
1.50	33.5	10078	0.952	0.979	0.988	11.14

Figure 2: Characteristics for Various Lognormal Models

The z-scores in the last column of Figure 2 would seem to validate the idea that increasing skewness corresponds to elongated tails. As the skewness gets larger the z-score for the most

extreme part per thousand also increases. This may be seen in Figure 3 which plots the skewness versus the z-scores for the most extreme part per thousand. So once again skewness is directly related to elongation, as is commonly thought. But what about the weight of the tails?
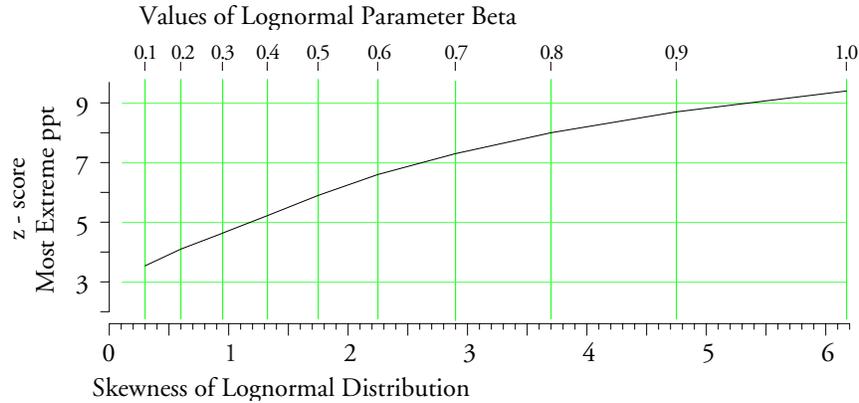

Figure 3: Skewness and Elongation for Lognormal Models

Figure 4 plots the areas for the fixed-width central intervals against the skewness of models from Figure 2. The bottom curve of Figure 4 ($k = 1$) shows that the areas found *within* one standard deviation of the mean of a Lognormal distribution increase with increasing skewness. Since the tails of a probability model are traditionally defined as those regions that are more than one standard deviation away from the mean, the bottom curve of Figure 4 shows us that the areas in the tails must *decrease* with increasing skewness. This contradicts the common notion about skewness and a heavy tail.
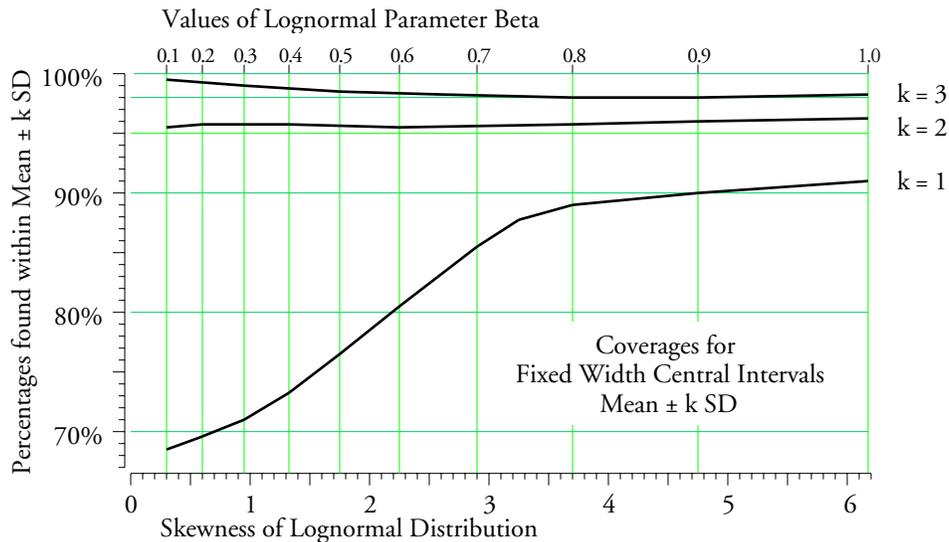

Figure 4: Coverages for Fixed-Width Central Intervals for Lognormal Distributions

So while the infinitesimal areas under the extreme tails will move further away from the mean with increasing skewness, the classically defined tails do not get heavier. Rather they actually get much lighter with increasing skewness. To move the outer few parts per thousand

further away from the mean you have to compensate by moving a much larger percentage closer to the mean. This compensation is unavoidable and inevitable. To stretch the long tail you have to pack an ever increasing proportion into the center of the distribution!
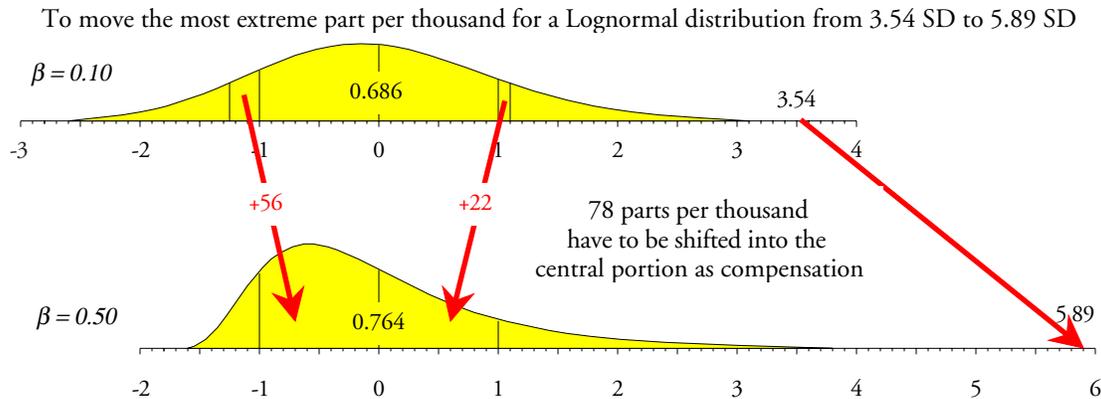
Figure 5: How the Tails Get Lighter with Skewness for Lognormal Distributions

So while skewness is associated with one tail being elongated, that elongation does not result in a heavier tail, but rather in a lighter tail. Increasing skewness is rather like squeezing toothpaste up to the top of the tube: while concentrating the bulk at one end, little bits get left behind and are squeezed down toward the other end. As these little bits become more isolated from the bulk the "tail" becomes elongated.

However, once again, there are a couple of surprises about this whole process. The first of these is the middle curve of Figure 4 ($k = 2$) which shows the areas within the fixed-width, two-standard-deviation central intervals. The flatness of this curve shows that, regardless of the skewness, the areas within two standard deviations of the mean of a Lognormal stay around 96 percent or more.

In statistics classes students are taught that having approximately 95% within two standard deviations of the mean is a property of the normal distribution. In parts one and two of this series we found that this property also applied to the families of Weibull and Gamma models. Here we see that this property also applies to the Lognormal distributions. Until the beta parameter gets larger than 1.00, Lognormal distributions will have approximately 95.5 percent to 96 percent within two standard deviations of the mean.

The second unexpected characteristic of the Lognormals is seen in the top curve of Figure 4 ($k = 3$) which shows the areas within the fixed-width, three-standard-deviation central intervals. While these areas drop slightly at first with increasing skewness for the Lognormals, they stabilize around 98 percent before beginning to climb back up. This means that a fixed-width, three-standard-deviation central interval for a Lognormal distribution will always contain approximately 98 percent or more of that distribution.

So if you think your data are modeled by a Lognormal distribution, then even without any specific knowledge as to which Lognormal distribution is appropriate, you can safely say that 98% or more will fall within three standard deviations of the mean, and that approximately 96% or more will fall within two standard deviations of the mean. *Fitting a particular Lognormal*

probability model to your data will not change either of these statements to any practical extent.
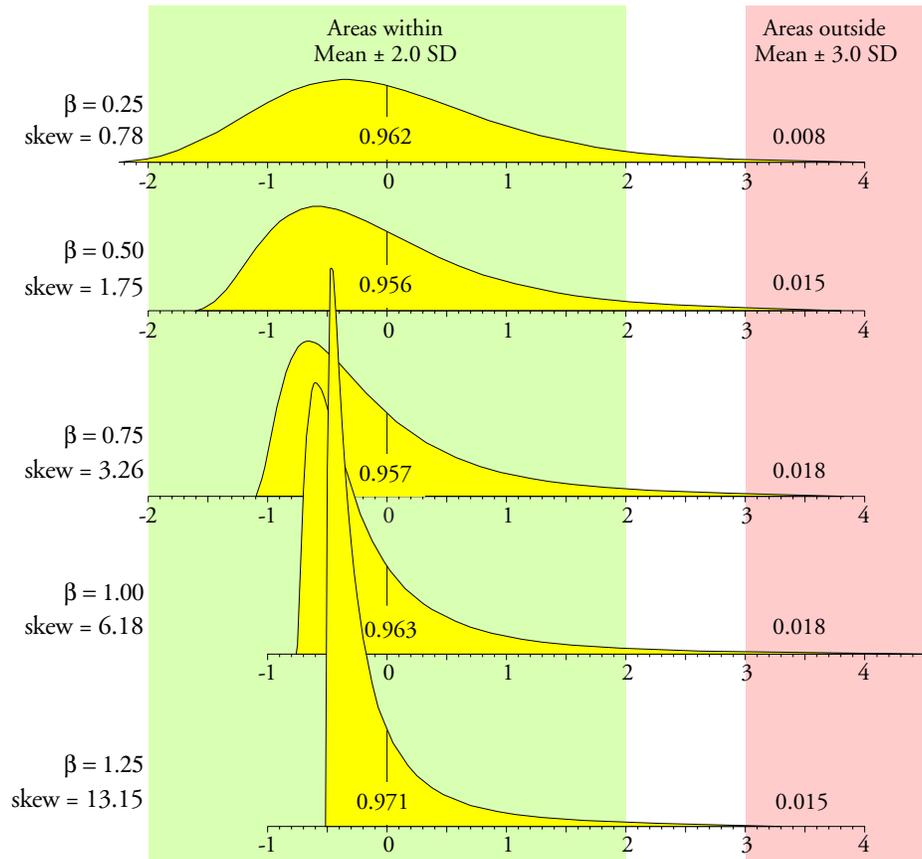

Figure 6: Common Characteristics of Lognormal Distributions

For many purposes these two results will be all you need to know about your Lognormal model. Without ever actually fitting a Lognormal probability model to your data, you can filter out either 95% or 98% of the probable noise using generic, fixed-width central intervals.

WHAT GETS STRETCHED?

If the tail gets both elongated and thinner at the same time, something has to get stretched. To visualize how skewness works for Lognormal models we can compare the widths of various fixed-coverage central intervals. These fixed-coverage central intervals will be symmetrical intervals of the form:

$$\text{MEAN}(X) \quad \pm \quad Z \;\; \text{SD}(X)$$

While this looks like the formula for the earlier fixed-width intervals, the difference is in what we are holding constant and what we are comparing. Earlier we held the widths fixed and compared the areas covered by the intervals. Here we hold the coverages fixed and compare the widths of the intervals. These widths are characterized by the z-scores in Figure 7. For example, a lognormal model with a beta parameter of 0.50 will have 92 percent of its area within 1.49 standard deviations of the mean, and it will have 95 percent of its area within 1.89 standard

deviations of the mean.

Lognormal Model			Fixed Coverages						
Beta	Skew	Kurt	0.92	0.95	0.96	0.98	0.99	0.995	0.999
0.10	0.30	3.16	1.74	1.95	2.05	2.34	2.62	2.90	3.54
0.20	0.61	3.68	1.71	1.92	2.03	2.39	2.78	3.17	4.05
0.25	0.78	4.10	1.67	1.91	2.03	2.44	2.89	3.33	4.33
0.30	0.95	4.64	1.64	1.89	2.03	2.51	3.00	3.49	4.61
0.40	1.32	6.26	1.57	1.88	2.07	2.64	3.22	3.81	5.23
0.50	1.75	8.90	1.49	1.89	2.10	2.75	3.42	4.13	5.89
0.60	2.26	13.3	1.43	1.89	2.11	2.83	3.60	4.43	6.58
0.70	2.89	20.8	1.38	1.86	2.10	2.89	3.76	4.72	7.30
0.75	3.26	26.5	1.34	1.83	2.08	2.90	3.82	4.85	7.66
0.80	3.69	34.4	1.31	1.80	2.06	2.91	3.88	4.97	8.03
0.90	4.75	60.4	1.22	1.73	1.99	2.90	3.95	5.17	8.74
1.00	6.18	114	1.12	1.64	1.90	2.85	3.98	5.32	9.41
1.10	8.21	233	1.02	1.52	1.79	2.76	3.95	5.40	10.00
1.20	11.2	518	0.91	1.40	1.66	2.63	3.87	5.41	10.50
1.25	13.2	800	0.85	1.33	1.59	2.56	3.81	5.39	10.71
1.30	15.6	1266	0.80	1.26	1.51	2.47	3.73	5.34	10.87
1.40	22.5	3404	0.68	1.12	1.36	2.29	3.54	5.19	11.09
1.50	33.5	10078	0.58	0.97	1.20	2.08	3.31	4.97	11.14

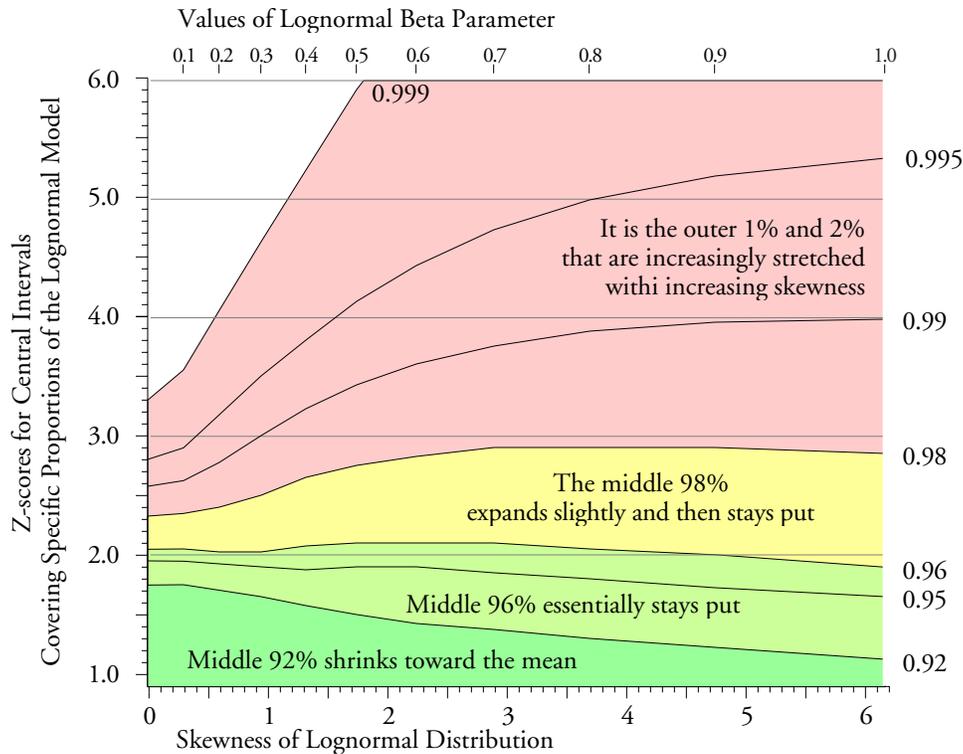**Figure 7: Widths of Fixed-Coverage Central Intervals for Lognormal Models**

Figure 8: Widths of Fixed-Coverage Central Intervals for Lognormals

Figure 8 shows the values in each column of Figure 7 plotted against skewness. The bottom curve shows that the middle 92 percent of a Lognormal will shrink with increasing skewness. The 95 percent and 96 percent curves remain remarkably flat in the neighborhood of 2.0 standard deviations until the increasing mass in the center of the distribution eventually begins to pull these curves down. The 98 percent curve initially grows, and then plateaus just below 3.0 standard deviations before it too begins to drop down with very high skewness.

The spread of the top three curves shows that for the Lognormal models it is primarily the outermost two percent that gets stretched into the extreme upper tail. While 920 parts per thousand are moving toward the mean, and while another 60 parts per thousand get slightly shifted outward and then stabilize, it is primarily the outer 20 parts per thousand that bear the brunt of the stretching and elongation that goes with increasing skewness.

THE BENEFITS OF FITTING A LOGNORMAL DISTRIBUTION

So if the context for the data provides a convincing rationale for using a Lognormal model, what do you gain by fitting a model to your data? The value for the beta parameter may be estimated from the data either by finding the standard deviation of the logarithms of the original data, or by using some other function of the average and standard deviation of the original data. This estimate will, in turn determine the shape of the specific Lognormal model you fit to your data. Since the average and standard deviation statistics will be more dependent upon the middle 95% of the data than the outer one or two percent, you will end up primarily using the middle portion of the data to choose your Lognormal model. Since the tails of a Lognormal model become lighter with increasing skewness, you will end up making a much stronger statement about how much of the area is within one standard deviation of the mean than about the size of the elongated tail. Fitting a Lognormal model is not so much about the tails as it is about how much of the model is found within one standard deviation of the mean. So, while we generally think of fitting a model as matching the elongated tail of a histogram, the reality is quite different.

Once you have chosen a specific Lognormal model, you can then use that model to extrapolate out into the extreme tail (where you are unlikely to have any data) to compute critical values that correspond to infinitesimal areas under the curve. However, as may be seen in Figure 3, even small errors in estimating the parameter beta can have a large impact upon the critical values computed for the infinitesimal areas under the extreme tail of your Lognormal model. As a result, the critical values you compute for the upper one or two percent of your Lognormal model will have virtually no contact with reality. Such computations will always be more of an artifact of the model used than a characteristic of either the data or the process that produced the data. To understand the problems attached to this extrapolation from the region where we have data to the region where we have no data see "Why We Keep Having 100-Year Floods," **QDD**, June 2013, and "The Parts Per Million Problem" **QDD** May 2015.

INDUSTRIAL DATA ANALYSIS

What impact does all this have on how we analyze data? To answer this it will be helpful to contrast the traditional statistical approach with the approach pioneered by Walter Shewhart.

The statistical approach uses fixed-coverage intervals for the analysis of experimental data.

In some cases these fixed-coverage intervals are not centered on the mean, but rather involve fixed coverages for the tail areas, but this is still analogous to the fixed-coverage central intervals used above. Fixed coverages are used because experiments are designed and conducted to detect specific signals, and we want the analysis to detect these signals in spite of the noise present in the data. The complexity and cost of most experiments will justify a fair amount of complexity in the analysis. By using fixed coverages statisticians can fine-tune just how much of the noise is being filtered out. This fine-tuning is important because additional data are not generally going to be available and we need to get the most out of the limited amount of experimental data. Thus, the complexity and cost of most experiments will justify a fair amount of complexity in the analysis. Moreover, to avoid missing real signals within the experimental data, it is traditional to filter out only 95 percent of the probable noise.
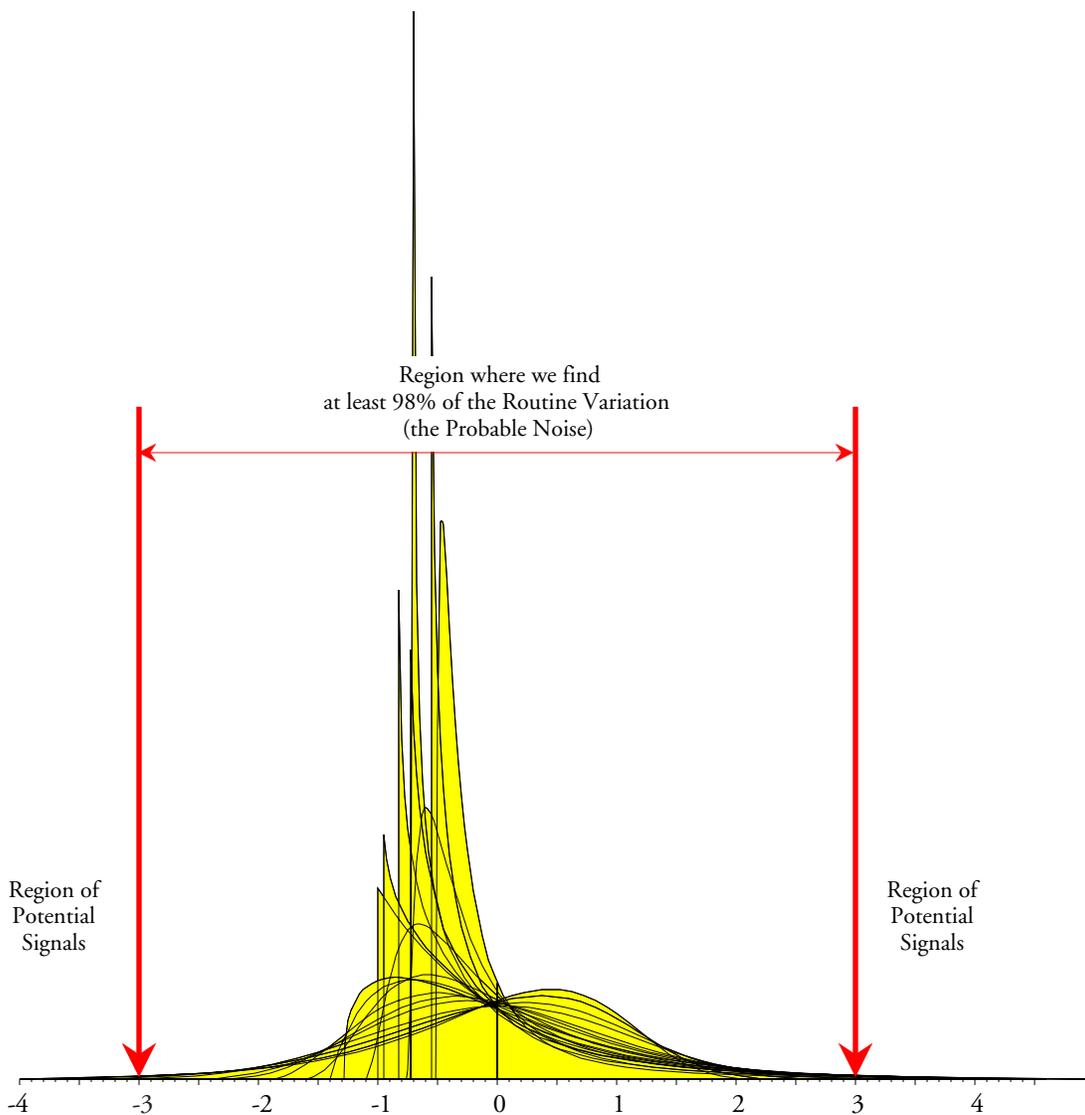

Figure 9: How Three-Sigma Limits Work with Weibulls, Gammas, and Lognormals

Shewhart's approach was created for the continuing analysis of observational data that are the by-product of operations. To this end Shewhart used a fixed-width interval rather than a fixed-coverage interval. His argument was that we will never have enough data to ever fully specify a particular probability model for the original data. Moreover, since additional data will typically be available, we do not need to fine-tune our analysis—the exact value of the coverage is no longer critical. As long as the analysis is reasonably conservative it will allow us to find those signals that are large enough to be of economic importance without getting too many false alarms. So, for the real-time analysis of observational data Shewhart chose to use a fixed-width, three-sigma central interval. As we have seen, such an interval will routinely filter upwards of 98 percent of the probable noise.

What we have discovered here is that Shewhart's simple, generic, three-sigma limits centered on the average will provide a conservative analysis for *any* and *every* data set that might logically be considered to be modeled by a Weibull, Gamma, or Lognormal distribution. As the reader may easily verify, it also works for any other mound-shaped or reasonable J-shaped probability model you might choose.

When we analyze observational data from our operations we only need limits that are conservative enough to keep false alarms from getting in our way as we look for the exceptional variation that is associated with process changes. Three-sigma limits are sufficient to do this because the difference between exceptional variation and routine variation is dramatic enough that we do not need a very sharp axe to separate the two.

This is why finding exact critical values for a specific probability model is not a prerequisite for using a process behavior chart. Once you filter out at least 98 percent of the probable noise, anything left over is a potential signal. As Bill Scherkenback said, "The purpose for collecting data is to take action." And Shewhart's three-sigma limits tell you when you need to take action and when to refrain from taking action. We only need watershed points to separate the bulk of the probable noise from the region of potential signals. Effort spent in fine-tuning the width of the limits to achieve a specific coverage is effort wasted.