Properties of Probability Models: Part One

What they forgot to tell you about Weibull distributions

Donald J. Wheeler

Some commonly held ideas about skewed probability models are incorrect. These incorrect ideas are one source of complexity and confusion regarding the analysis of data. By examining the basic properties of skewed distributions this article can help you to greater clarity of thought and may even simplify your next data analysis.

How would you characterize a skewed distribution? When asked this question most will answer, "A skewed distribution is one that has a heavy, elongated tail." This idea is expressed by saying that a distribution becomes more heavy-tailed as its skewness and kurtosis increase. To examine these ideas we shall use a popular family of skewed distributions, the Weibulls.

THE WEIBULL FAMILY OF DISTRIBUTIONS

Weibull distributions are widely used in reliability theory and are generally found in most statistical software packages. This makes these distributions easy to use without having to work with complicated equations. The following equations are included here in the interest of clarity. The Weibull distributions depend upon two parameters: alpha, α, and beta, β. The cumulative distribution function for the Weibull family has the form:

$$F(x) \; = \; 1 \; - \; exp\,\{\,-\,[\,\tfrac{x}{\beta}\,]^{\alpha}\,\} \qquad \text{for } x \, \geq \, 0,\, \alpha \, > 0,\; \beta \, > 0.$$

The mean of a Weibull distribution is:

$$Mean(X) \; = \; \beta \; \Gamma\,[\,(\alpha+1)/\alpha\,]$$

and the variance of a Weibull distribution is:

$$Variance(X) \; = \; \{\,\beta^{\,2}\;\Gamma\,[\,(\alpha+2)/\alpha\,]\,\} \; - \; [\;Mean(X)\;]^{2}$$

where $\Gamma[A]$ denotes the gamma function (for $A > 0$):

$$\Gamma[A] \; = \; \int_{0}^{\infty} x^{A-1}\;e^{-x}\;dx$$

The Weibull alpha parameter determines the shape of the distribution while the beta parameter determines the scale. Since we will consider the Weibulls in standardized form, where the distribution is shifted to have a mean of zero and stretched or shrunk to have a standard deviation of 1.00, the value for the beta parameter will not matter. Changing the value of β will not affect any of the results noted here. Thus, the horizontal scale for the probability models shown in the figures that follow are in standard deviation units.

When the alpha parameter is 1.00 or less the Weibull model will be J-shaped. When the alpha parameter is between 1.00 and 3.60 the Weibull distributions will be mound-shaped with positive

skewness. When the parameter α has the value 3.60 the Weibull model will have a skewness that is very near to zero. When the parameter α is greater than 3.60 the Weibull models will be mound-shaped and negatively skewed. Since the shape of these negatively skewed Weibull models essentially stops changing as the alpha parameter increases beyond 10.0 the negatively skewed Weibulls are of little practical interest. This freezing in shape can be seen by comparing the standardized distributions for α = 15 and α = 800 in Figure 1.
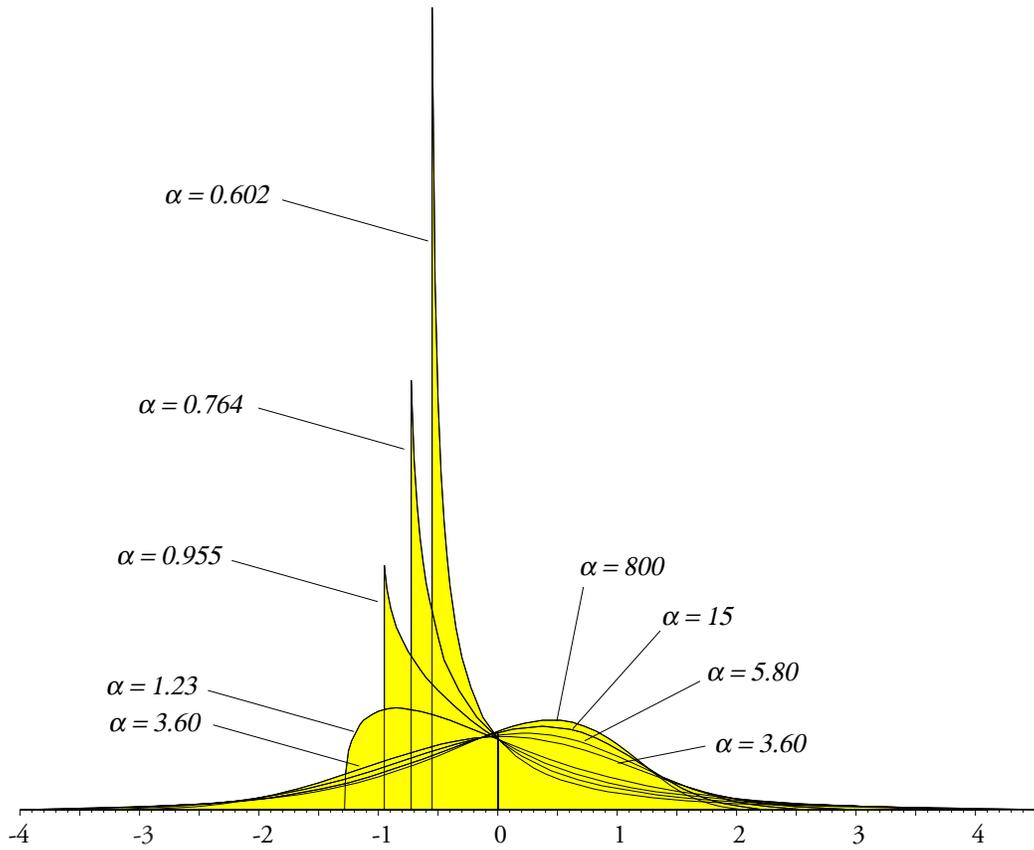

Figure 1: Eight Standardized Weibull Distributions

So what is changing as you select different Weibull probability models? To answer this question Table 1 considers twenty-three different Weibull models. 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 most extreme part per thousand of the model.

The z-scores in the last column of Table 1 would seem to validate the idea that increasing skewness corresponds to elongated tails. As the skewness gets larger in magnitude the z-score for the most extreme part per thousand also increases in magnitude. This may be seen in Figure 2 which plots the skewness versus the z-scores for the most extreme part per thousand. So skewness is directly related to elongation, as is commonly thought. But what about the weight of the tails?

Table 1: Characteristics for Various Weibull Models

| Weibull Parameter Alpha	Skewness	Kurtosis	Fixed-Width Central Intervals			Most Extreme ppt z-score
			Area Within One SD	Area Within Two SD	Area Within Three SD	
800	−1.14	11.70	0.723	0.957	0.988	−4.92
67	−1.05	5.00	0.718	0.957	0.989	−4.77
15	−0.79	4.00	0.702	0.957	0.991	−4.24
5.8	−0.35	3.00	0.681	0.956	0.997	−3.36
3.6	0.00	2.72	0.670	0.958	0.999	2.91
2.245	0.48	3.01	0.670	0.965	0.996	3.55
1.615	0.95	4.00	0.691	0.957	0.991	4.25
1.37	1.24	5.00	0.717	0.953	0.988	4.71
1.226	1.47	6.00	0.747	0.952	0.985	5.09
1.128	1.67	7.00	0.780	0.951	0.984	5.40
1.056	1.84	8.00	0.819	0.950	0.983	5.67
1.00	2.00	9.00	0.865	0.950	0.982	5.91
0.955	2.14	10.00	0.868	0.950	0.981	6.12
0.9173	2.28	11.00	0.870	0.950	0.980	6.32
0.8858	2.40	12.00	0.873	0.950	0.980	6.49
0.8349	2.63	14.00	0.877	0.951	0.979	6.81
0.7636	3.03	18.00	0.884	0.952	0.979	7.32
0.6958	3.53	24.01	0.891	0.954	0.978	7.90
0.6019	4.57	40.00	0.904	0.957	0.979	8.88
0.5443	5.55	59.99	0.913	0.960	0.979	9.60
0.50	6.12	87.72	0.922	0.963	0.980	10.22
0.40	11.35	290.6	0.942	0.972	0.984	11.70
0.30	28.33	2346.	0.967	0.983	0.990	12.35
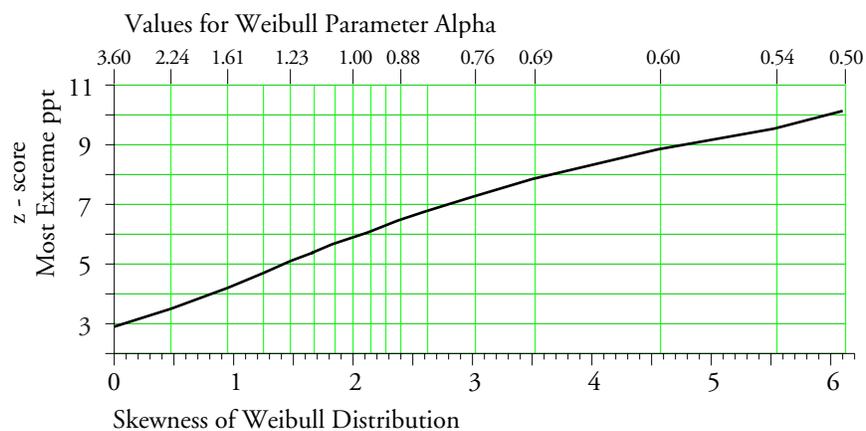
Figure 2: Skewness and Elongation for Weibull Models

Figure 3 plots the areas for the fixed-width central intervals against the skewness of models from Figure 2. The bottom curve of Figure 3 shows that the areas found *within* one standard deviation of the mean of a Weibull 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 3 shows us that the areas in the tails must *decrease* with increasing skewness. This contradicts the common notion about skewness and a heavy tail.
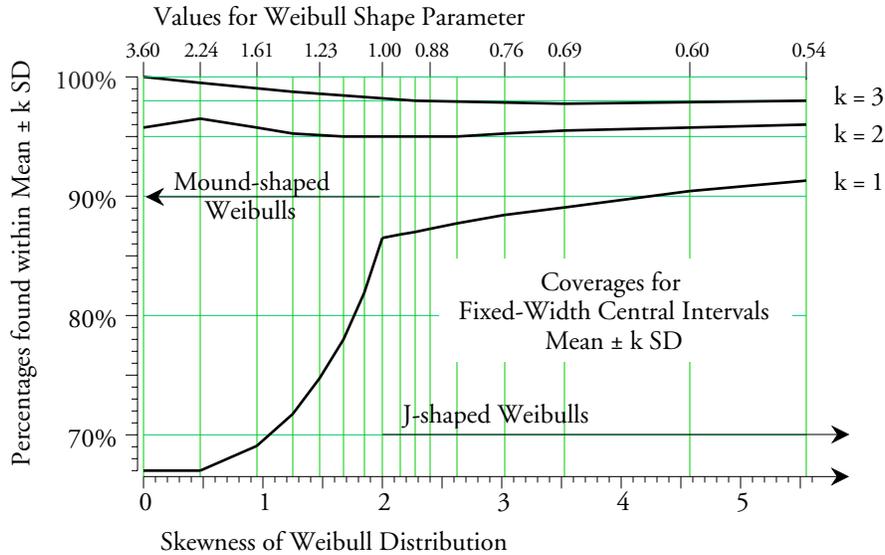

Figure 3: How the Coverages Vary with Skewness for Weibull Distributions

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!
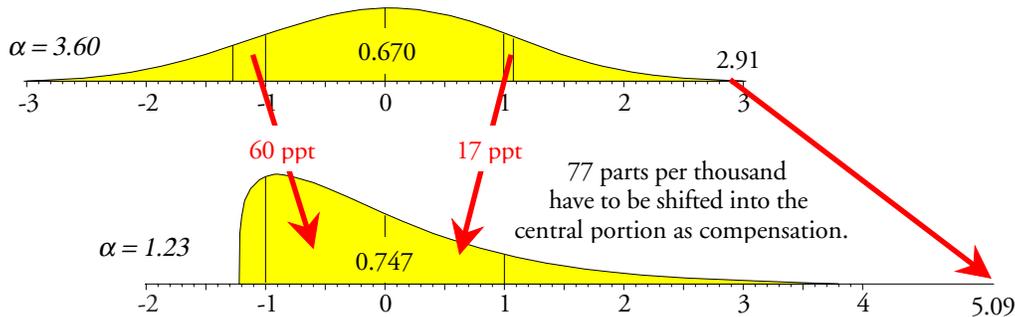

Figure 4: How the Tails Get Lighter with Skewness for Weibull Distributions

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, there are a couple of surprises about this whole process. The first of these is the middle curve of Figure 4 which shows the areas within the fixed-width, two-standard-deviation central intervals. The flatness of this curve shows that the areas within two standard deviations of the mean of a Weibull stay around 95 percent to 96 percent *regardless of the skewness.*

In statistics classes students are taught that having approximately 95% within two standard deviations of the mean is a property of the normal distribution. While this is true, the fact that this property also applies to most of the Weibull distributions is unexpected. From the negatively skewed Weibulls, through the positively skewed mound-shaped Weibulls, and for all but the most extreme of the J-shaped Weibulls there will be approximately 95 percent to 96 percent within two standard deviations of the mean.
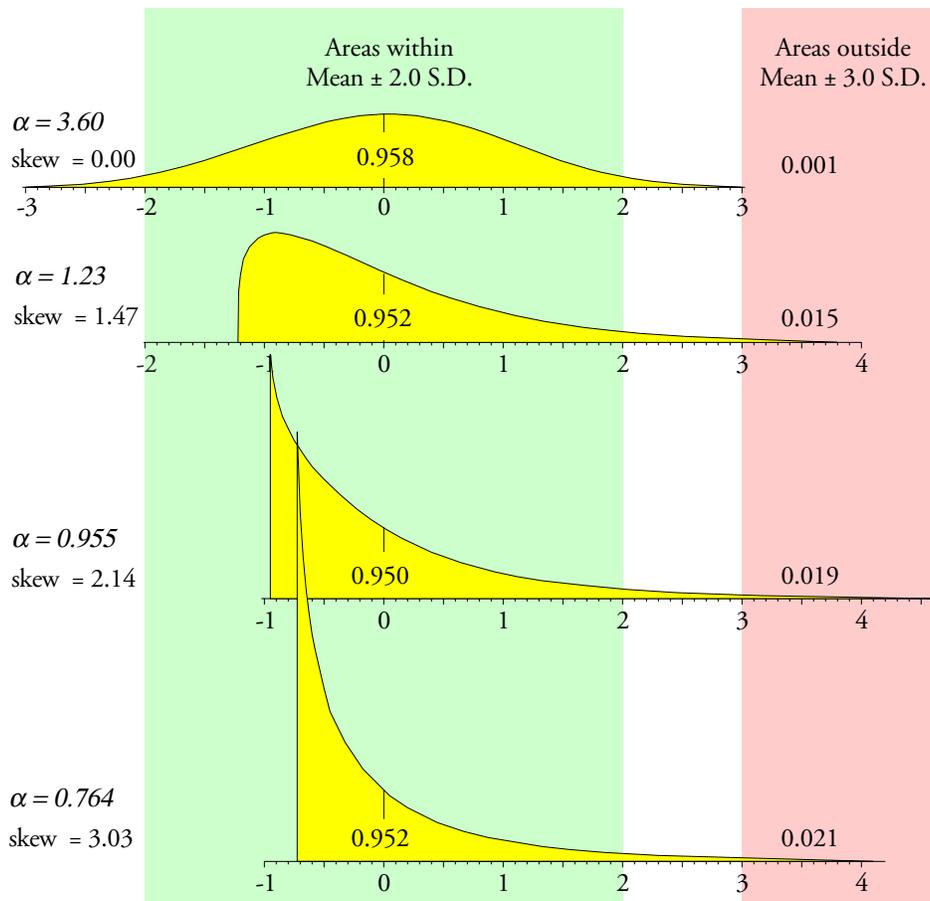

Figure 5: What Weibull Distributions Have in Common

The second unexpected characteristic of the Weibulls is seen in the top curve of Figure 3 which shows the areas within the fixed-width, three-standard-deviation central intervals. While these areas do drop slightly at first, they stabilize for the J-shaped Weibulls at about 98 percent. This means that a fixed-width, three-standard-deviation central interval for a Weibull distribution will always contain approximately 98 percent or more of that distribution.

So if you think your data are modeled by a Weibull distribution, then even without any specific knowledge as to which Weibull distribution is appropriate, you can safely say that approximately 98% or more will fall within three standard deviations of the mean, and that approximately 95% or more will fall within two standard deviations of the mean. *Fitting a particular Weibull probability model to your data will not change either of these statements to any practical extent.*

For many purposes these two results will be all you need to know about your Weibull model. Without ever actually fitting a Weibull 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 Weibull 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. With the fixed-width intervals we compared the areas covered. With the fixed-coverage intervals we compare the widths of the intervals. These widths are characterized by the z-scores in Table 2. For example, a Weibull model with an alpha parameter of 1.000 will have 92 percent of its area within 1.52 standard deviations of the mean, and it will have 99 percent of its area within 3.61 standard deviations of the mean.

Table 2: Widths of Fixed-Coverage Central Intervals for Weibull Models

Weibull Model			Coverages						
Alpha	Skew	Kurt	0.92	0.95	0.975	0.98	0.99	0.995	0.999
3.600	0.00	2.72	1.75	1.94	2.18	2.25	2.44	2.61	2.94
2.245	0.48	3	1.69	1.87	2.16	2.28	2.61	2.91	3.55
1.615	0.95	4	1.56	1.89	2.37	2.52	2.95	3.36	4.25
1.370	1.24	5	1.56	1.94	2.48	2.65	3.16	3.65	4.71
1.226	1.47	6	1.56	1.97	2.56	2.75	3.31	3.86	5.09
1.128	1.67	7	1.55	1.99	2.62	2.82	3.43	4.03	5.40
1.056	1.84	8	1.54	1.99	2.66	2.87	3.52	4.18	5.67
1.000	2.00	9	1.52	2.00	2.69	2.91	3.61	4.30	5.91
0.955	2.14	10	1.51	2.00	2.71	2.95	3.67	4.41	6.12
0.9173	2.28	11	1.50	1.99	2.73	2.98	3.73	4.50	6.32
0.8858	2.40	12	1.49	1.99	2.75	3.00	3.78	4.59	6.49
0.8349	2.63	14	1.46	1.98	2.77	3.04	3.87	4.73	6.81
0.7636	3.03	18	1.41	1.95	2.80	3.08	3.99	4.95	7.32
0.6958	3.53	24	1.34	1.90	2.80	3.11	4.11	5.18	7.90
0.6019	4.57	40	1.21	1.79	2.76	3.10	4.25	5.51	8.88
0.5443	5.55	60	1.09	1.68	2.69	3.05	4.30	5.71	9.60

Each column in Table 2 is plotted against the skewness in Figure 6. The bottom curve shows that the middle 92 percent of a Weibull will shrink with increasing skewness. The 95 percent

fixed-coverage intervals are remarkably stable until the increasing mass near the mean eventually begins to pull this curve down. The 98 percent fixed-coverage intervals initially grow, and then they plateau near three standard deviations.
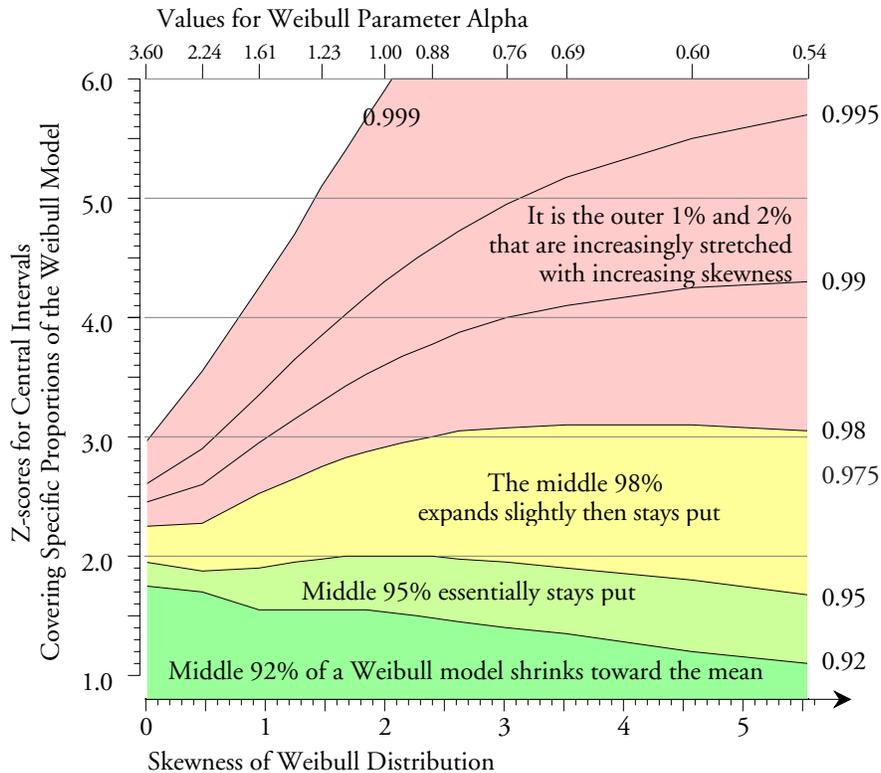

Figure 6: Widths of Fixed-Coverage Central Intervals for Weibull Models

The spread of the top three curves shows that for the Weibull 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 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 WEIBULL DISTRIBUTION

So what do you gain by fitting a Weibull model to your data? The value for the alpha parameter may be estimated from the average and standard deviation statistics, and this will, in turn determine the shape of the specific Weibull model you fit to your data. Since these statistics will be more dependent upon the middle 95% of the data than the outer one percent or less, you will end up primarily using the middle portion of the data to choose a Weibull model. Since the tails of a Weibull 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 Weibull distribution 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 a specific Weibull 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 alpha can have a large impact upon the critical values computed for the infinitesimal areas under the extreme tail of your Weibull model. As a result, the critical values you compute for the upper one or two percent of your Weibull 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? It helps to provide some perspective on how and why there are two distinctly different approaches to data analysis. For clarity call these the statistical approach and Shewhart's approach.

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. 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 they 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.

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.
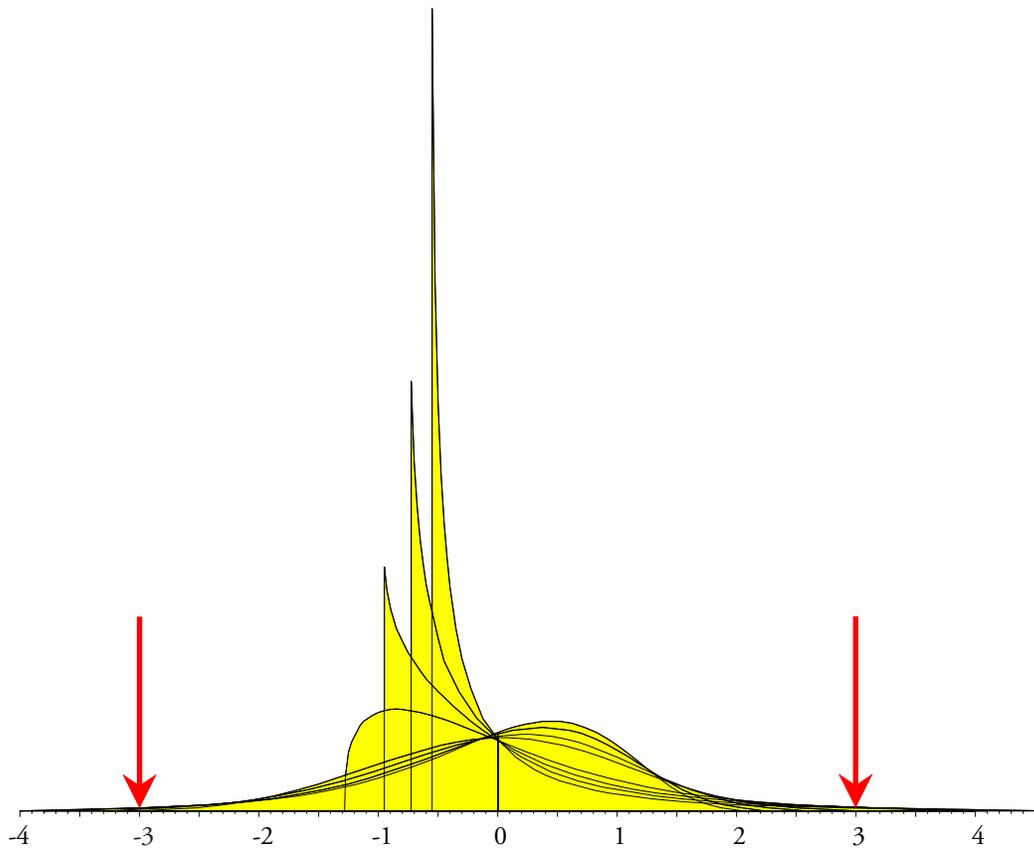
Figure 7: How Three-Sigma Limits Work with Weibull Distributions

What we have discovered here is that Shewhart's simple, generic, three-sigma limits will provide a conservative analysis for *any* and *every* data set that might logically be considered to be modeled by a Weibull distribution. 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.