Properties of Probability Models: Part Two

What they forgot to tell you about the Gammas

Donald J. Wheeler

Clear thinking and simplicity of analysis require concise, clear, and correct notions about probability models and how to use them. Last month we looked at the properties of the Weibull probability models and discovered that some ideas about skewed distributions are incorrect. Here we shall examine the basic properties of the family of Gamma models.

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. Last month, for Weibull models at least, we discovered that as the tail was elongated it grew lighter, not heavier. Does this happen with other families of probability models? Here we consider the Gamma models.

THE GAMMA FAMILY OF DISTRIBUTIONS

Gamma distributions are widely used in all aras of statistics, and are found in most statistical software. Since software facilitates our use of the Gamma models, the following formulas are given in the interest of clarity. Gamma models depend upon two parameters, once again denoted by alpha, α, and beta, β. The probability density function for the Gamma family has the form:

$$f(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-x/\beta} \qquad \text{for } x > 0,\ \alpha > 0, \text{ and } \beta > 0$$

where the symbol $\Gamma(\alpha)$ denotes the gamma function (for $\alpha > 0$):

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}\, e^{-x}\, dx$$

The mean and variance for a Gamma distribution are:

$$Mean = \alpha\,\beta \qquad Variance = \alpha\,\beta^2$$

The alpha parameter determines the shape of the Gamma model. When the value for alpha is 1.00 or less the Gamma distributions will be J-shaped. As the value for alpha increases above 1.00 the Gamma distributions become mound-shaped and as the value for alpha gets large the Gammas approach the normal distribution. Since we consider these distributions in standardized form the value for the beta parameter will not affect any of the following results. Six standardized Gamma distributions are shown in Figure 1.

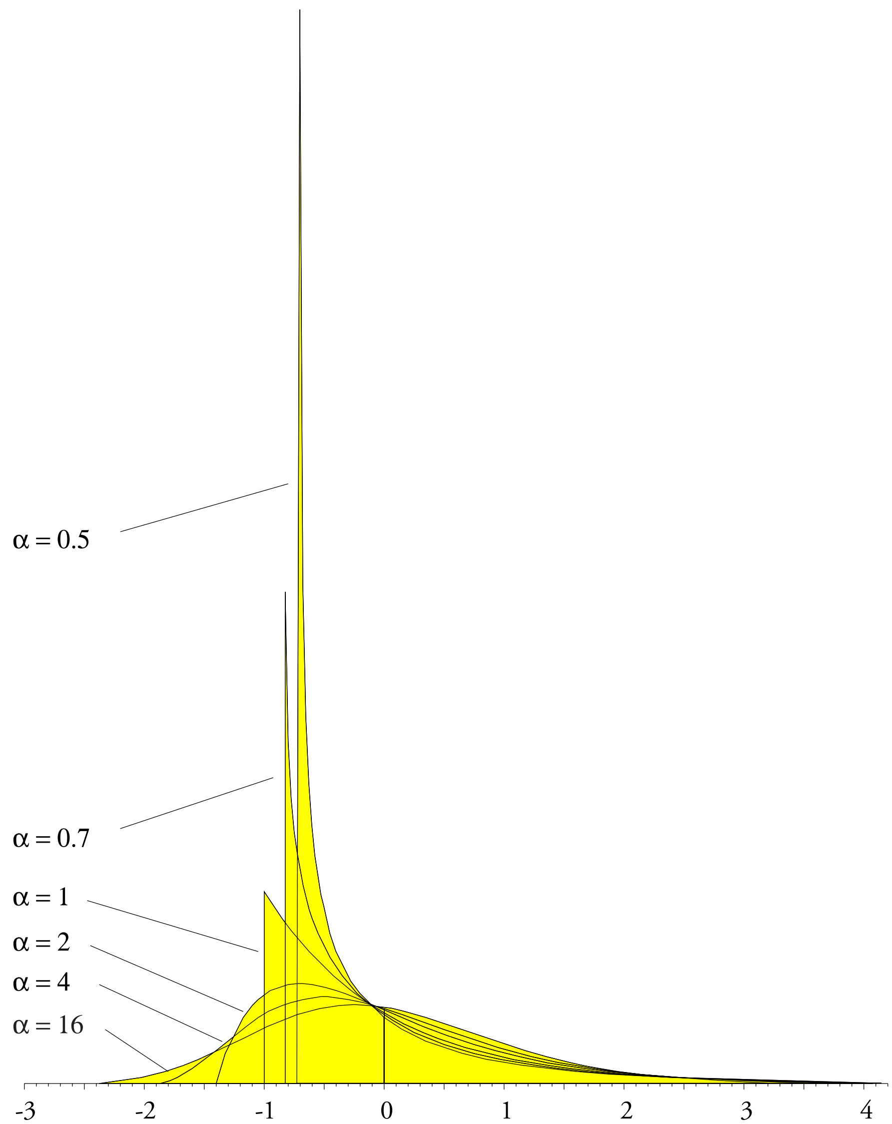

Figure 1: Six Standardized Gamma Distributions

So what is changing as you select different Gamma probability models? To answer this question Table 1 considers nineteen different Gamma 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 99.9th percentile 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 the z-score for the most extreme part per thousand also increases. 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 Gamma Models

			Fixed-Width Central Intervals			
Gamma Parameter Alpha	Skewness	Kurtosis	Area Within One SD	Area Within Two SD	Area Within Three SD	Most Extreme ppt z-score
64	0.25	3.09	0.684	0.955	0.997	3.45
26	0.39	3.23	0.686	0.956	0.996	3.65
16	0.50	3.38	0.688	0.957	0.995	3.81
10	0.63	3.60	0.691	0.959	0.993	4.00
7	0.76	3.86	0.695	0.959	0.992	4.18
4	1.00	4.50	0.706	0.958	0.990	4.53
3	1.15	5.00	0.715	0.956	0.988	4.75
2	1.41	6.00	0.738	0.953	0.986	5.11
1.50	1.63	7.00	0.766	0.952	0.984	5.42
1.25	1.79	7.80	0.796	0.951	0.983	5.63
1.00	2.00	9.00	0.865	0.950	0.982	5.91
0.80	2.24	10.5	0.869	0.950	0.980	6.21
0.70	2.39	11.57	0.872	0.949	0.980	6.41
0.60	2.58	13.00	0.875	0.949	0.979	6.65
0.50	2.83	15.00	0.880	0.950	0.978	6.95
0.40	3.16	18.00	0.886	0.950	0.977	7.34
0.30	3.65	23.00	0.894	0.952	0.976	7.88
0.20	4.47	33.00	0.907	0.955	0.976	8.72
0.16	5.00	40.50	0.915	0.957	0.976	9.22

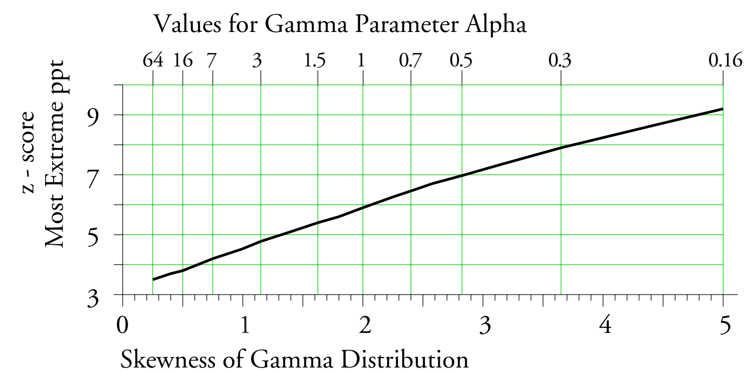

Figure 2: Skewness and Elongation for Gamma Models

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 ($k = 1$) shows that the areas found *within* one standard deviation of the mean of a Gamma 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.

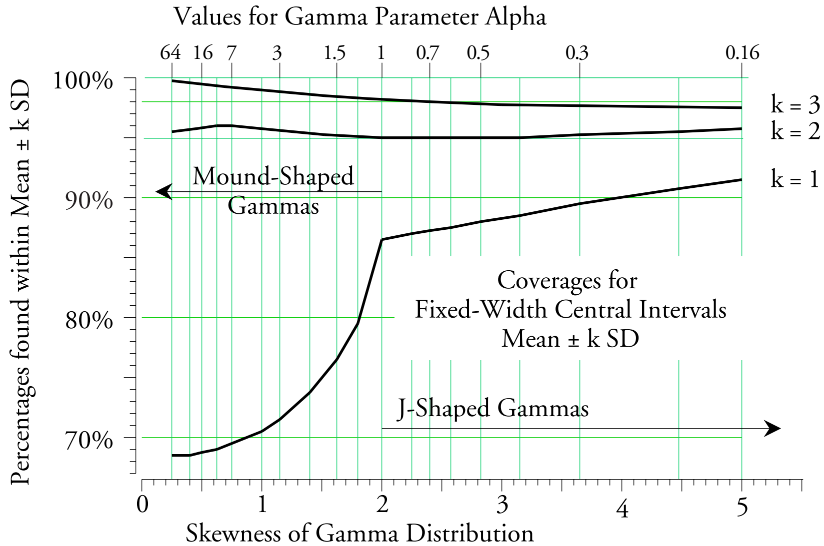

Figure 3: How the Coverages Vary with Skewness for Gamma Distributions

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!

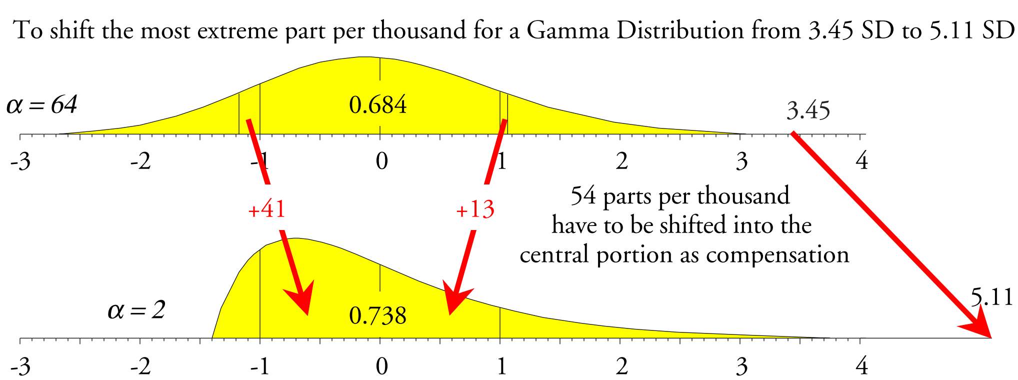

Figure 4: How the Tails Get Lighter with Skewness for Gamma Distributions

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 3 (k = 2) 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 Gamma 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. Last month we found that this was a property of the family of Weibull models. Here we see that this property also applies to the Gamma distributions! Beginning with the mound-shaped Gammas and continuing through the J-shaped Gammas there will be approximately 95 percent to 96 percent within two standard deviations of the mean.

The second unexpected characteristic of the Gammas is seen in the top curve of Figure 3 (k = 3) which shows the areas within the fixed-width, three-standard-deviation central intervals. While the area within three standard deviations of the mean does drop slightly at first, it stabilizes for the J-shaped Gammas at about 97.5 percent. This means that a fixed-width, three-standard-deviation central interval for a Gamma distribution will always contain at least 97.5 percent of that distribution.

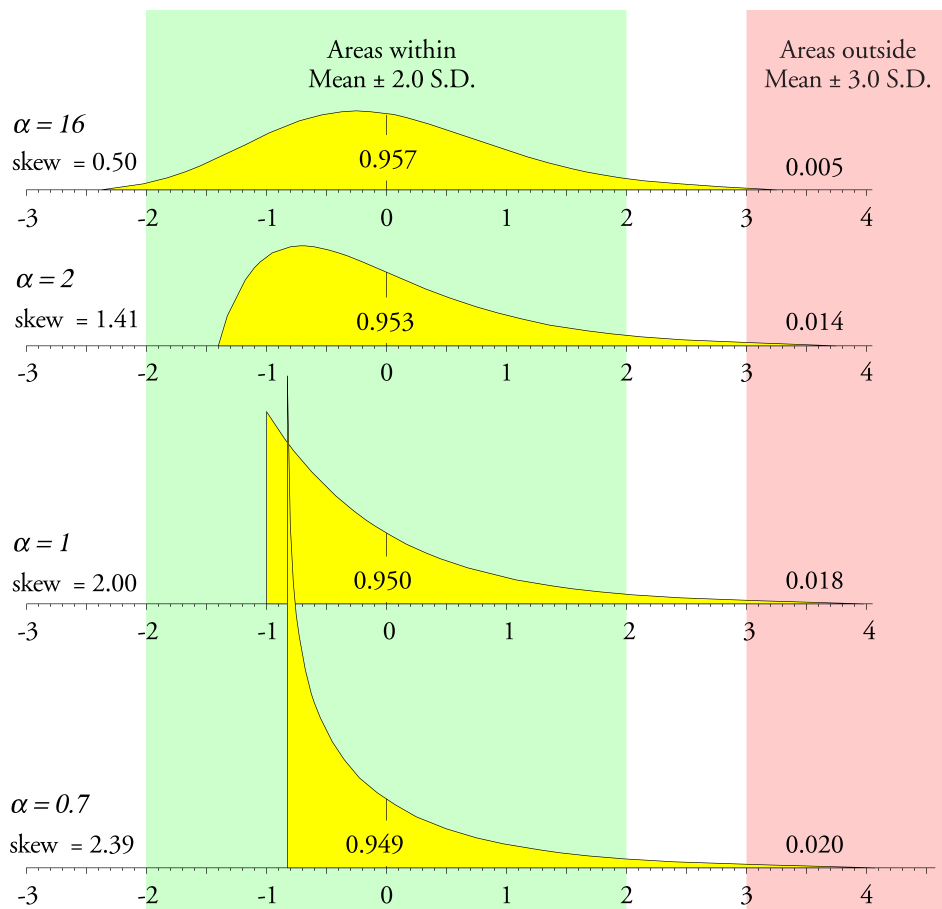

Figure 5: What Gamma Distributions Have in Common

So if you think your data are modelled by a Gamma distribution, then even without any specific knowledge as to which of the Gamma distributions is appropriate, you can safely say that

97.5% 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 Gamma 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 Gamma model. Without ever actually fitting a Gamma 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 Gamma models we can compare the widths of various fixed-coverage central intervals. These fixed-coverage central intervals will be symmetrical intervals of the form:

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

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. When we hold the widths fixed we compare the areas covered by the intervals. When we hold the coverages fixed we compare the widths of the interrvals. These widths are characterized by the z-scores in Table 2. For example, a Gamma model with an alpha parameter of 1.25 will have 92 percent of its area within 1.53 standard deviations of the mean, and it will have 99 percent of its area within 3.49 standard deviations of the mean.

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

Gamma Model			Fixed Coverages						
Alpha	Skew	Kurt	0.92	0.95	0.975	0.98	0.99	0.995	0.999
64	0.25	3.09	1.74	1.95	2.24	2.33	2.60	2.86	3.44
26	0.39	3.23	1.73	1.94	2.24	2.44	2.64	2.95	3.64
16	0.50	3.38	1.72	1.93	2.25	2.35	2.69	3.04	3.79
10	0.63	3.60	1.70	1.91	2.26	2.38	2.78	3.16	4.00
7	0.76	3.86	1.67	1.89	2.29	2.43	2.86	3.27	4.18
4	1.00	4.50	1.61	1.88	2.38	2.54	3.02	3.49	4.53
3	1.15	5.00	1.57	1.90	2.44	2.61	3.12	3.62	4.75
2	1.41	6.00	1.53	1.94	2.53	2.71	3.28	3.84	5.11
1.5	1.63	7.00	1.53	1.97	2.59	2.79	3.41	4.02	5.42
1.25	1.79	7.80	1.53	1.98	2.64	2.84	3.49	4.14	5.63
1.0	2.00	9.00	1.52	2.00	2.69	2.91	3.61	4.30	5.91
0.8	2.24	10.5	1.51	2.01	2.74	2.98	3.72	4.47	6.21
0.7	2.39	11.57	1.50	2.01	2.77	3.02	3.80	4.58	6.41
0.6	2.58	13.00	1.49	2.01	2.81	3.07	3.88	4.71	6.65
0.5	2.83	15.00	1.46	2.01	2.85	3.12	3.98	4.86	6.95
0.4	3.16	18.00	1.42	2.00	2.89	3.18	4.11	5.07	7.34
0.3	3.65	23.00	1.34	1.96	2.92	3.24	4.27	5.33	7.88
0.2	4.47	33.00	1.20	1.86	2.93	3.30	4.48	5.71	8.72
0.16	5.00	40.50	1.09	1.77	2.91	3.30	4.57	5.92	9.22

Figure 6 shows the values in each column of Table 2 plotted against skewness. The bottom curve shows that the middle 92 percent of a Gamma 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 97.5 percent fixed-coverage intervals initially grow until they plateau near three standard deviations.

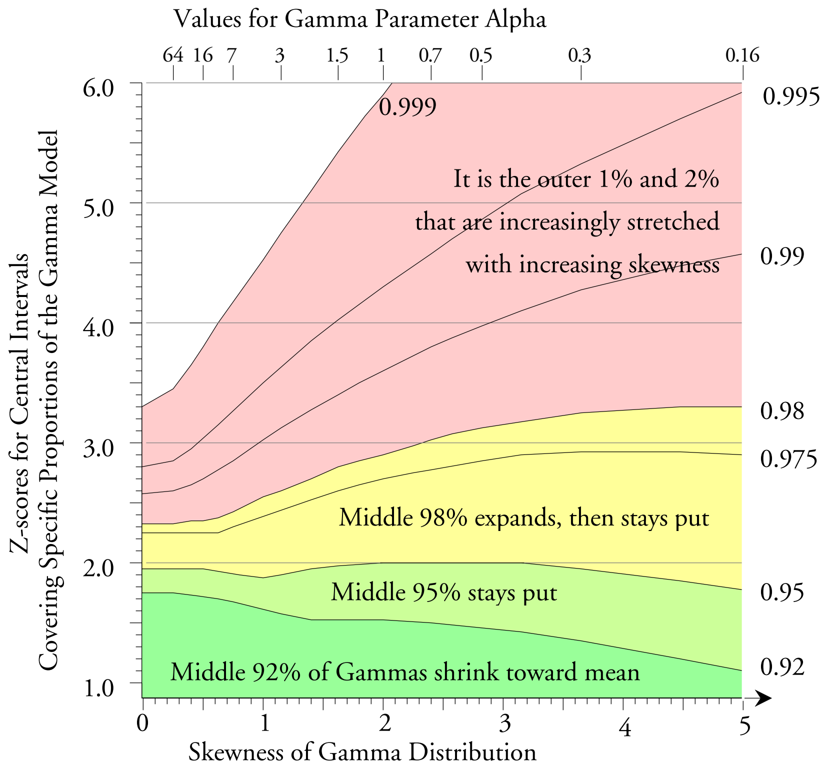

Figure 8: Widths of Fixed-Coverage Central Intervals for Gamma Models

The spread of the top three curves shows that for the Gamma 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 perts 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 GAMMA DISTRIBUTION

So what do you gain by fitting a Gamma model to your data? The value for the alpha parameter may be estimated from the average and standard deviation statistics, and this estmate will, in turn determine the shape of the specific Gamma 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 Gamma model. Since the tails of a Gamma 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 Gamma 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 Gamma model, you can then use the 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 Figures 3 and 8, even small errors in estmating the parameter alpha can have a large impact upon the critical values computed for the infinitesimal areas under the extreme tail of your Gamma model. As a result, the critical values you compute for the upper one or two percent of your Gamma model will have virtually no contact with reality. Such computations will always be more of an artifact of the model used than a charactistic of either the data or the process that produced the data.

To illustrate this point I generated 5000 data sets of 100 values each using an exponential distribution (which is a Gamma with an alpha parameter of 1.000). For each data set I estimated the value of alpha. These estimates ranged from 0.495 to 2.103. From Figure 3 we can see that this range of values for alpha will result in Gamma models that have their most extreme part per thousand anywhere in the range from 5 SD to 7 SD above the mean. Thus, the uncertainty in your estimate of the alpha parameter will create large uncertainties in the location of the infinitesimal areas under the extreme tail. Consequently, any extreme tail critical values you compute will be more of an artifact of your model than a characteristic of your data.

INDUSTRIAL DATA ANALYSIS

What impact does all this have on how we analyze data? It turns out that 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 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.

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.

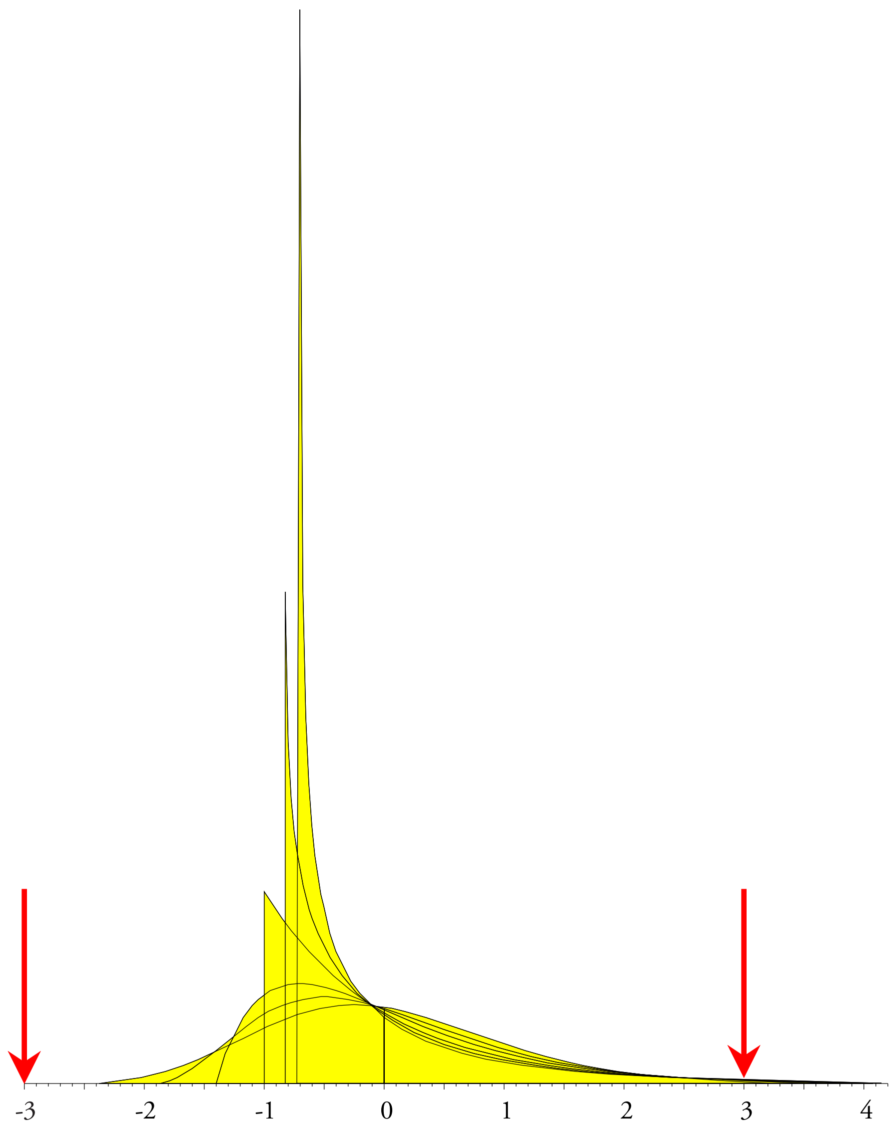

Figure 7: How Three-Sigma Limits Work with Gamma Distributions

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 Gamma distribution. Last month we discovered that Shewhart's simple, generic, three-sigma limits also 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 approximately 98 percent or more of the probable noise, anything left over is a potential signal.