Global Journal of Researches in Engineering
General Engineering
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 2249-4596 Print ISSN:0975-5861

Process Capability Analysis using Curve Fitting Methods

By John J. Flaig & Fred Khorasani
University of Engineering and Technology (BUET), United States

Abstract - This paper offers a new approach to process capability measurement based on techniques for estimating the fraction nonconforming in the tails of the observed process distribution. The methodology proposed can provide significantly more accurate results and applicability to a much larger class of process distributions than the standard approaches of calculating capability indices based on making parametric assumptions or using standard curve fitting techniques.

The tail probability estimation approach to capability analysis can significantly enhance the practitioners understanding of the processes they are attempting to develop and/or control because it provides a more robust and accurate measure of capability. This clearer insight into process performance becomes increasingly important as the allowable fraction nonconforming levels decline.

Keywords : process capability, fraction nonconforming estimation.

GJRE-J Classification : FOR Code: 010301

PROCESS CAPABILITY ANALYSIS USING CURVE FITTING METHODS

Strictly as per the compliance and regulations of :

Process Capability Analysis using Curve Fitting Methods

John J. Flaig [α] & Fred Khorasani [σ]

Abstract - This paper offers a new approach to process capability measurement based on techniques for estimating the fraction nonconforming in the tails of the observed process distribution. The methodology proposed can provide significantly more accurate results and applicability to a much larger class of process distributions than the standard approaches of calculating capability indices based on making parametric assumptions or using standard curve fitting techniques.

The tail probability estimation approach to capability analysis can significantly enhance the practitioners understanding of the processes they are attempting to develop and/or control because it provides a more robust and accurate measure of capability. This clearer insight into process performance becomes increasingly important as the allowable fraction nonconforming levels decline.

Keywords : process capability, fraction nonconforming estimation.

I. History

It is critical in process development and ongoing monitoring to have an understanding of how capable the process is of meeting requirements. These requirements reflect internal and external demands that are expressed in terms of specifications. Historically, assessment of process capability using the indices such as Cp and Cpk became popular in the early 1980's. Engineers used these indices to determine if a process should be released to production (i.e., during qualification) and customers demanded that suppliers provide them as measures of their process performance. Clearly, important decisions were based on these indices. Then, around 1990, questions began to be raised about their validity in industrial applications where the assumptions underlying the calculation of the indices were often not met. Numerous papers have been published that discuss the shortcomings of capability indices [Gunter, 1989 and 1991][Somerville, 1997]. However, we still find today that capability indices are the primary tool for assessing and communicating the process capability.

Author α : Applied Technology 1237 Clark Way San Jose, California, 95125, U.S.A. www.e-AT-USA.com Copyright © 2013 by John J. Flaig. E-mail : johnflaig@yahoo.com

II. Introduction

If process capability can be defined as the ability of a process to produce products or services that meet the specified requirements [ASQC, 1983] [Duncan, 1986], the question then becomes; how can this ability be measured? A reasonable approach might be to try to estimate the probability that the product or service falls within the acceptance region defined by the specifications. There are three common methods for generating this estimate:

1. Empirical: Based on sampling the process to determine the number of conforming items divided by the total number of items sampled. This is the relative frequency approach to capability assessment and in the limit it would provide a true measure of capability assuming that the process is stable. Unfortunately, many real world processes are not stable.
2. Parametric: Based on the assumption that the observed values come from some theoretical distribution. This top-down approach is the classical method used by many practitioners to assess process capability [Somerville, 1997]. The parametric assumption might be given credibility because the nature of the process may "a priori" give rise to the theoretical distribution or it might be supported by goodness-of-fit tests. This approach has two risks, the first is the assumption of stability and the second is the subjective nature of the assumed distribution.
3. Modeling: Based on curve-fitting techniques such as polynomial regression or Johnson curves. This is a bottom-up approach [Pyzdek, 1992] [Farnum, 1996]. This approach also assumes process stability. However, model selection is less subjective because it is based upon the limited set of choices typically offered by the computer program. The problem is that the limited set may not include a "good" fitting distribution.

There is also another and more common approach to measuring and communicating the assessment of process capability. This methodology involves generating so-called capability indices. These indices are generally just functions of the processes descriptive statistics and specifications limits. Based on our experience using the fraction conforming as an estimator of process capability offers several advantages over the use of capability indices because it is more intuitive. That is, most people have an intuitive understanding of what percent nonconforming means (i.e., high yield implies the process is capable and low yield implies that it is incapable) whereas there is no such intuitive understanding for an abstract numeric capability index. In addition, there are so many capability indices (in excess of one hundred) that it is difficult to recall the merits of each.

III. Alternate Methods

There are a number of different approaches to estimating the fraction nonconforming. For example, Pyzdek and Farnum discussed using Johnson curves to estimate the fraction nonconforming [Pyzdek, 1992] [Farnum, 1996]. Other researchers have expressed concerns with the curve fitting approach because of accuracy issues [Wheeler, 1995]. The author's agree with Wheeler that curve fitting methods will typically not be able to resolve nonconformance rates down to low levels unless there is a relatively large amount of data available.

If the process distribution and specification limits are reasonably well structured (i.e., the process distribution is mostly within the specification limits), then the problem of determining process capability becomes one of estimating tail probabilities. A major criticism of the curve fitting approach is that a single function is probably not sufficient to fit the observed data in the tails and in the middle of the distribution simultaneously. This follows from the observation that least squares regression analysis will tend to fit the bulk of the data (i.e., the central mass) and miss-fit the limited amount of data in the tails. Because of this, attempts to fit parametric distributions such as Normal, Johnson, or Weibull to mound shaped empirical data sets will give rise to tail fit errors. This problem is further complicated because real world processes are generally dynamic – meaning that the data may not be coming from a single or static distribution generator.

Our proposed approach to fitting the process distribution differs from the classical curve fitting methodology in two ways:

1. The distribution is divided into three parts (left, middle, and right) and the tails are fit separately.
2. A well-known and very flexible modeling approach is used to fit the tails of the distribution so that the left and right tails are approximated by unique functions.

The first point focuses attention and statistical techniques where they should be -- on the tail probabilities and not the bulk of the distribution. Johnson, Kotz and Pearn proposed a somewhat similar analysis approach [Johnson, 2006]. However, they divided the process distribution in half, which is an improvement but still has the central mass fitting issue.

The second point allows the practitioner to fit the observed data in a realistic way. For example, there are distributions where the observed data is increasing and then decreasing in the tail, so the fitting function should have this property. The classical approach of assuming a Normal distribution (which goes to zero in the tail) is clearly unrealistic. Bounded or truncated distributions offer another example, where the standard approaches do not work very well. For example, fitting a Johnson curve to a bounded distribution gives rise to a function (SB type) that goes to zero in the tail whereas the bounded function may have no tail area (e.g., if the LSL is less than the lower bound).

IV. Analytic Methodology

Techniques from reliability analysis will be used to fit various functions to the tail distributions of data drawn at random from known distributions [Tobias, 1995]. The fitted curve results will be compared with the true results from the actual distribution and contrasted with the results of using the classical assumption of normality.

The first class of distributions to be considered are the bounded type (i.e., the domain (t) of the function is bounded on one or both sides and the range (y) does not go to zero on at least one side). A triangular distribution defined by, $y = -2t+2$ on the interval [0, 1] will be used in this example. This function was selected because it offers a challenging test of the classical normality assumption and its ability to yield a realistic assessment of process capability.

The analysis is carried out as follows and displayed in Table 1 and Figure 1:

1. One thousand data points are generated at random from the triangular (Tri) distribution
2. The data is sorted smallest to largest
3. The first one hundred (left tail) and last one hundred values (right tail) are selected
4. Normal (Nor), Johnson (Jon), and Weibull (Wei) distributions are fitted to the tail values
5. The PDF and CDF functions for each distribution are generated and graphed
6. Several estimates of forecast accuracy are generated so that the results can be compared.

Table 1 : Tail Probability Analysis

Left Tail

t	CDF Tri	PDF Tri	CDF Nor	PDF Nor	ERR Nor	ABS Err	CDF Wei	PDF Wei	ERR Wei	ABS Err	CDF Jon	PDF Jon	ERR Jon	ABS Jon
0.00	1,000,000	2.000	993,862	0.527	6,138	6,138	1,000,000	-	0	0	975,008	14.582	24,992	24,992
0.01	980,100	1.980	986,112	1.076	-6,012	6,012	981,574	2.049	-1,474	1,474	769,536	22.893	210,564	210,564
0.02	960,400	1.960	971,070	2.005	-10,670	10,670	960,321	2.182	79	79	537,568	23.250	422,832	422,832
0.03	940,900	1.940	944,428	3.406	-3,528	3,528	938,173	2.240	2,727	2,727	306,798	22.778	634,102	634,102
0.04	921,600	1.920	901,371	5.276	20,229	20,229	915,630	2.265	5,970	5,970	89,640	19.542	831,960	831,960
0.05	902,500	1.900	837,867	7.453	64,633	64,633	892,946	2.270	9,554	9,554	#NUM!	#NUM!	#NUM!	#NUM!
0.06	883,600	1.880	752,397	9.601	131,203	131,203	870,280	2.262	13,320	13,320	#NUM!	#NUM!	#NUM!	#NUM!
0.07	864,900	1.860	647,425	11.278	217,475	217,475	847,738	2.245	17,162	17,162	#NUM!	#NUM!	#NUM!	#NUM!
0.08	846,400	1.840	529,775	12.081	316,625	316,625	825,396	2.222	21,004	21,004	#NUM!	#NUM!	#NUM!	#NUM!
0.09	828,100	1.820	409,446	11.801	418,654	418,654	803,312	2.194	24,788	24,788	#NUM!	#NUM!	#NUM!	#NUM!
0.10	810,000	1.800	297,140	10.513	512,860	512,860	781,527	2.162	28,473	28,473	#NUM!	#NUM!	#NUM!	#NUM!
Mean					151,601	155,275			11,055	11,323			#NUM!	#NUM!
Sigma					188,423				10,633				#NUM!	

Right Tail

t	CDF Tri	PDF Tri	CDF Nor	PDF Nor	ERR Nor	ABS Err	CDF Wei	PDF Wei	ERR Wei	ABS Err	CDF Jon	PDF Jon	ERR Jon	ABS Jon
0.90	10,000	0.200	8,179	0.146	1,821	1,821	9,259	0.155	741	741	100,794	1.895	-90,794	90,794
0.91	8,100	0.180	6,832	0.124	1,268	1,268	7,815	0.134	285	285	83,112	1.644	-75,012	75,012
0.92	6,400	0.160	5,685	0.106	715	715	6,568	0.116	-168	168	67,835	1.414	-61,435	61,435
0.93	4,900	0.140	4,712	0.089	188	188	5,497	0.099	-597	597	54,762	1.204	-49,862	49,862
0.94	3,600	0.120	3,891	0.075	-291	291	4,581	0.085	-981	981	43,687	1.014	-40,087	40,087
0.95	2,500	0.100	3,200	0.063	-700	700	3,801	0.072	-1,301	1,301	34,405	0.845	-31,905	31,905
0.96	1,600	0.080	2,622	0.053	-1,022	1,022	3,140	0.061	-1,540	1,540	26,717	0.696	-25,117	25,117
0.97	900	0.060	2,139	0.044	-1,239	1,239	2,582	0.051	-1,682	1,682	20,429	0.565	-19,529	19,529
0.98	400	0.040	1,739	0.036	-1,339	1,339	2,114	0.043	-1,714	1,714	15,356	0.452	-14,956	14,956
0.99	100	0.020	1,408	0.030	-1,308	1,308	1,722	0.036	-1,622	1,622	11,326	0.356	-11,226	11,226
1.00	0	0.000	1,135	0.025	-1,135	1,135	1,397	0.030	-1,397	1,397	8,177	0.276	-8,177	8,177
Mean					-277	1,002			-907	1,093			-38,918	38,918
Sigma					1,121				857				27,429	

Total Distribution

t	CDF Tri	PDF Tri	CDF Nor	PDF Nor	ERR Nor	ABS Err	CDF Wei	PDF Wei	ERR Wei	ABS Err	CDF Jon	PDF Jon	ERR Jon	ABS Jon
0.00	1,000,000	2.000	927,658	0.599	72,342	72,342	1,000,000	-	0	0	979,148	0.893	20,852	20,852
0.10	810,000	1.800	847,031	1.027	-37,031	37,031	803,882	2.030	6,118	6,118	816,764	2.060	-6,764	6,764
0.20	640,000	1.600	722,069	1.458	-82,069	82,069	614,642	1.730	25,358	25,358	608,420	1.995	31,580	31,580
0.30	490,000	1.400	561,281	1.714	-71,281	71,281	459,333	1.378	30,667	30,667	428,443	1.587	61,557	61,557
0.40	360,000	1.200	389,522	1.668	-29,522	29,522	337,850	1.060	22,150	22,150	291,089	1.169	68,911	68,911
0.50	250,000	1.000	237,197	1.343	12,803	12,803	245,434	0.798	4,566	4,566	191,981	0.827	58,019	58,019
0.60	160,000	0.800	125,045	0.895	34,955	34,955	176,481	0.590	-16,481	16,481	122,894	0.568	37,106	37,106
0.70	90,000	0.600	56,493	0.494	33,507	33,507	125,792	0.431	-35,792	35,792	76,073	0.379	13,927	13,927
0.80	40,000	0.400	21,707	0.226	18,293	18,293	88,975	0.311	-48,975	48,975	45,239	0.246	-5,239	5,239
0.90	10,000	0.200	7,054	0.085	2,946	2,946	62,503	0.223	-52,503	52,503	25,595	0.153	-15,595	15,595
1.00	0	0.000	1,931	0.027	-1,931	1,931	43,635	0.158	-43,635	43,635	13,589	0.091	-13,589	13,589
Mean					-4,272	36,062			-9,866	26,022			22,797	30,285
Sigma					46,848				31,102				31,131	

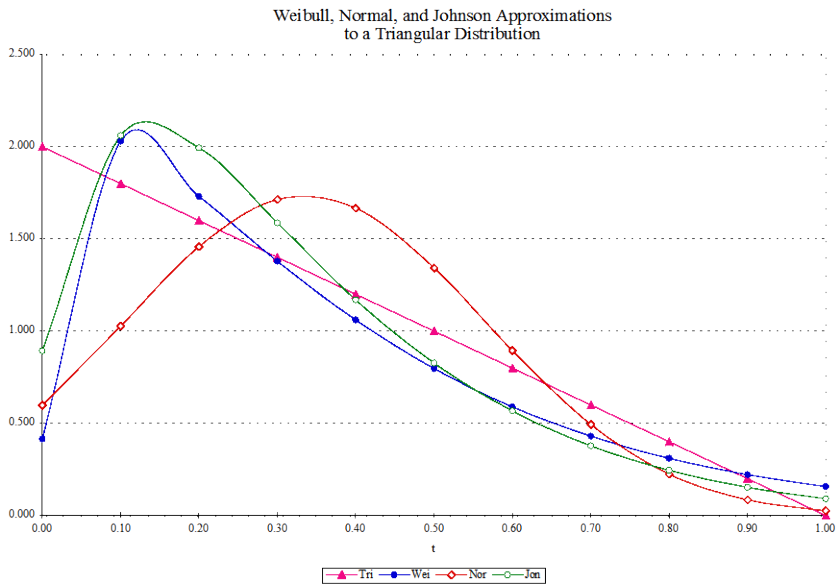

Figure 1 : Approximating a Triangular Distribution

V. Analysis

It can be seen from Table 2 that the CDF errors for Weibull and Normal were about equal in the right tail and both were significantly better than Johnson. For the left tail the Weibull error is significantly less than Normal or Johnson. Thus, the Weibull estimates more accurately reflect the true tail probabilities than does the Normal or Johnson curve for this triangular distribution.

Table 2 : Errors in Estimating Probabilities (DPM)

	Left Tail (0, .1)	Right Tail (.9, 1)	Total Distribution
Weibull	11,323 *	1,093	26,022
Normal	155,275	1,002	36,062
Johnson	Very Large	38,918	30,285

* Mean absolute deviation of the True CDF from the Estimated CDF measured in defectives per million (DPM).

The second type of distribution to be considered is the unbounded type. The data for this distribution arose as part of a real world study at a semiconductor equipment manufacturer. Five hundred and sixty nine readings were taken and the distribution formed by this data is displayed in Figure 2:

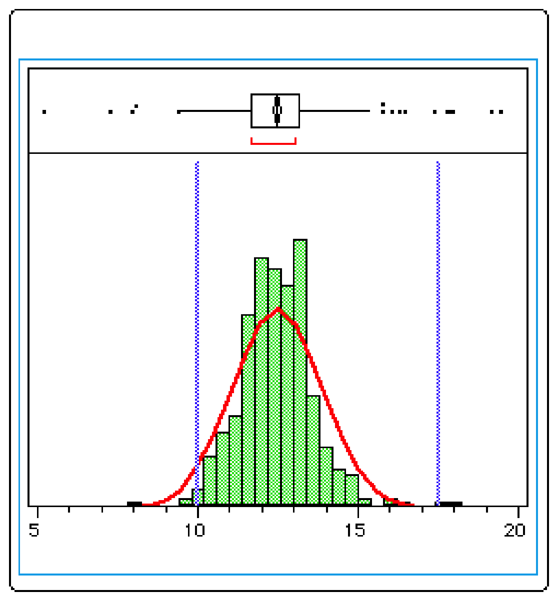

Figure 2 : Real Process Histogram and Normal Distribution

The distribution has a mean of 12.482 and a standard deviation of 1.395; it is roughly symmetrical and highly peaked (as indicated by a kurtosis of 7.5). This distribution is also non-Normal as can be seen by comparing it to the superimposed Normal distribution and this is confirmed by the Shapiro-Wilk's normality test statistic of .932.

If the USL = 17.5 and LSL = 10, then the fraction nonconforming can be estimated based of the various distribution assumptions. This analysis is given below:

Observed

The amount of product falling outside of specification limits based on the observed data is given below:

Percentage of units above the USL =	1.23%	12,302 DPM
Percentage of units below the LSL =	1.41%	14,060 DPM
Total percent nonconforming =	2.64%	26,362 DPM

Normal

Assuming normality, the amount of material falling outside the specification limits is given below:

Percentage of units above the USL =	0.02%	237 DPM
Percentage of units below the LSL =	3.98%	39,754 DPM
Total percent nonconforming =	4.00%	39,990 DPM

Johnson

Using a Johnson curve to approximate the observed distribution we have:

Percentage of units above the USL =	0.04%	385 DPM
Percentage of units below the LSL =	2.31%	23,053 DPM
Total percent nonconforming =	2.34%	23,437 DPM

Weibull

Using a Weibull curve to approximate the observed distribution we have:

Percentage of units above the USL =	1.14%	11,367 DPM
Percentage of units below the LSL =	1.43%	14,261 DPM
Total percent nonconforming =	2.56%	25,628 DPM

It can be seen from Table 3 that the Weibull fraction nonconforming matched the observed values better than the Normal or Johnson in both the left and right tails. Thus, the Weibull estimates more accurately reflect the observed tail probabilities than does the Normal or Johnson curve for this empirical distribution.

Table 3 : Tail Probabilities (DPM)

	Left Tail	Right Tail
Observed	14,060	12,302
Weibull	14,261	11,367
Johnson	23,053	385
Normal	39,754	237

The most disconcerting part of this study is the realization that many practitioners are currently basing their process capability analysis and conclusions on the assumption of normality, which can be seen, in this example, to yield very unrealistic results.

VI. Summary

The Weibull tail fitting approach to capability analysis has been shown to offer good accuracy in estimating the fraction nonconforming when compared with two other common fitting distributions in the examples tested. The use of capability indices for measuring process capability seem weak because they offer limited intuitive communication ability and they do not map one-to-one into an accurate estimate of the fraction nonconforming which is what management is interested in knowing. The standard curve fitting approach is handicapped by the attempt to force a single function to fit the entire distribution (which may be a mixture of several distributions) when only the tails are generally of interest in capability analysis.

The merits of this new approach are:

1. It attempts to estimate an intuitively reasonable measure of process capability (i.e., the fraction conforming which is one minus the fraction nonconforming).
2. It separates the data distribution into three parts (left tail, middle, and right tail) so that the analysis can be focused where it should be -- on the tails. This approach results in significantly increased accuracy in estimating the tail probabilities.
3. Using the Weibull curve offers significantly greater flexibility than Normal or Johnson curves when applied to the tails. This increased flexibility

translates into a greater ability to mimic the observed data distribution, which resulting in more accurate tail probability estimates.

The probability density functions (pdf's) used in this paper are listed below:

Normal

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2}, -\infty < \mu < \infty, \ \sigma > 0, -\infty < t < \infty$$

Johnson

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty, -\infty < t < \infty$$

$$z = \gamma + \eta k_i(t, \lambda, \varepsilon)$$

$$k_1(t, \lambda, \varepsilon) = \sinh^{-1}\left(\frac{t-\varepsilon}{\lambda}\right)$$ Unbounded (SU type)

$$k_2(t, \lambda, \varepsilon) = \ln\left(\frac{t-\varepsilon}{\lambda+\varepsilon-t}\right)$$ Bounded (SB type)

$$k_3(t, \lambda, \varepsilon) = \ln\left(\frac{t-\varepsilon}{\lambda}\right)$$ Lognormal (SL type)

Weibull

$$f(t) = \frac{\eta}{\sigma}\left(\frac{t-\mu}{\sigma}\right)^{\eta-1} e^{-\left(\frac{t-\mu}{\sigma}\right)^{\eta}}, -\infty < \mu < \infty, \ \eta > 0, \ \sigma > 0, t \geq \mu$$

References Références Referencias

1. American Society for Quality Control, Statistics Division: Glossary and Tables for Statistical Quality Control, 2nd ed., ASQC, Milwaukee, WI.
2. Duncan, A. J. (1986). Quality Control and Industrial Statistics, 5-th ed., Irwin, Homewood, IL.
3. Farnum, N. R. (1996). Using Johnson Curves to Describe Non-normal Process Data, Quality Engineering, Marcel Dekker, Vol. 9, No. 2, pp. 329-336.
4. Flaig, J. J. (1999). Process Capability Sensitivity Analysis. Quality Engineering, Marcel Dekker, Vol. 11, No. 4, pp. 587-592.
5. Gunter, B. (1989). The Use and Abuse of Cpk. Quality Progress, January through September.
6. Gunter, B. (1991). Process Capability Studies. Quality Progress, Vol. 24, No. 8, pp. 123-132.
7. Pearn, W. L. and Kotz S. (2006). Encyclopedia and Handbook of Process Capability Indices, World Scientific Publishing Company, Vol. 12, pp. 305-307.
8. Pyzdek, T. (1992). Process Capability Analysis Using Personal Computers, Quality Engineering, Marcel Dekker, Vol. 4, No. 3, pp. 419-440.
9. Somerville, S. E. and Montgomery, D. C. (1997). Process Capability Indices and Non-normal Distributions, Quality Engineering, Marcel Dekker, Vol. 9, No. 2, pp. 305-316.
10. Tobias, P. A. and Trindade, D. C. (1995). Applied Reliability, Van Nostrand Reinhold.
11. Wheeler, D. J. (1995). Advanced Topics in Statistical Process Control, Statistical Process Controls, Inc.

Global Journals Inc. (US) Guidelines Handbook 2013

www.GlobalJournals.org