Process Capability Indices for Non-Normal Data

MARTIN KOVÁŘÍK, LIBOR SARGA
Department of Statistics ands Quantitative Methods
Tomas Bata University in Zlín
Mostní 5139, 760 01 Zlín
CZECH REPUBLIC
m1kovarik@fame.utb.cz, sarga@fame.utb.cz

Abstract: - When probability distribution of a process characteristic is non-normal, C_p and C_{pk} indices calculated using conventional methods often lead to erroneous interpretation of process capability. Various methods have been proposed for surrogate process capability indices (PCIs) under non-normality but few literature sources offer their comprehensive evaluation and comparison, in particular whether they adequately capture process capability under mild and severe departures from normality, and what is the best method to compute true capability under each of these circumstances. We overview 9 methods as to their performance when handling PCI non-normality. The comparison is carried out by simulating log-normal and data and the results presented using box plots. We show performance to be dependent on the ability to capture tail behavior of the underlying distribution.

Key-Words: - process, capability, index, distribution, method, transformation, asymmetric

1 Introduction

Process capability indices (PCIs) are widely used to determine whether a given process is capable to output products in a given tolerance band. C_p a C_{pk}, ranking among the most popular, are defined as:

$$C_p = \frac{tolerance\ width}{process\ width}, \quad (1)$$

$$C_{pk} = \min\{C_{pu}, C_{pl}\} = \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right\}, \quad (2)$$

where USL and LSL denote upper and lower specification limits, respectively, and C_{pk} the minimum of (C_{pu}, C_{pl}). Due to unknown mean process value μ and variance σ^2, they are usually estimated from sample counterparts, $\bar{x}$ and s^2. English and Taylor analyzed non-normality in PCIs and posited C_{pk} is more sensitive to deviations than C_p. Kotz and Johnson tested properties of the indices, e.g., Clements' method, Johnson-Kotz-Pearn method and "distribution-independent" PCIs under non-normal distributions as well. Index by Wright [19] integrates skewness correction factor in denominator of C_{pmk} proposed by Johnson. Choi and Bai [5] suggested a weighted variance heuristic method to correct PCIs for skewness by considering standard deviation above and beyond mean individually. Some authors proposed new-generation PCIs based on assumptions about the population distribution. Pearn and Kotz [16] tied their capability indices to Pearson distribution. Johnson et al. [11] substituted 6σ in the numerator of equation (2) by 6θ where θ was selected so that „capability" was not substantially influenced by the shape of the distribution. Castagliola presented computational methods for non-normal PCIs by estimating ratio of unsatisfactory products using Burr distribution. Vännman introduced $C_p(u, v)$ index family with (u, v) comprising many other indices as special cases. Deleryd investigated suitable u's and v's for skewed process distributions. $C_p(1, 1)$, equivalent to C_{pmk}, is recommended as the best fit for treating non-normality in PCIs. Even though it is accepted C_p and C_{pk} are not suitable for process capability with non-normal distributions and the previously-mentioned methods constitute viable substitutes, a complex treatment is missing. Experts are interested in which methods are mutually comparable, which are sensitive to the normality assumption, and which are suitable for use with non-normal distributions.

We will investigate 9 methods with respect to process capability by comparing C_{pu} obtained from simulations with a target C_{pu}. In case the data are non-normally distributed, it will be necessary to obtain estimates as $\hat{L}_{0.00135}$, $\hat{U}_{0.99865}$, and $\hat{Me}_{0.5}$ quantiles; we will evaluate some methods.

Sometimes, the shape of the distribution is known a priori or based on knowledge of physical process properties. It is then necessary to use goodness of fit test to determine whether the model correctly represents the data, estimates distribution parameters, and computes the quantiles for log-normal, Weibull, exponential distributions, etc. Some literature sources list equations for parameter and quantile estimations which can be also inputted to software. Fig. 1 depicts how a PCI is obtained for a non-normally distributed quality attribute [16,17,18].

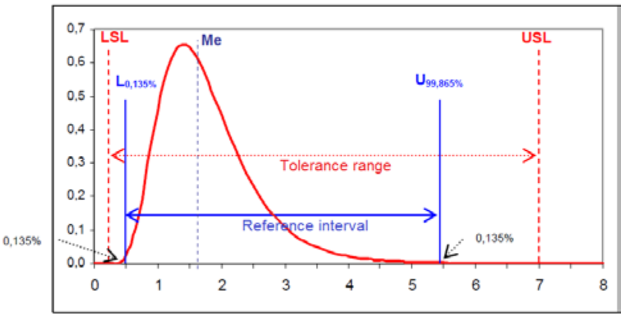

Fig. 1 PCI for a non-normally distributed quality attribute [8]

2 Surrogate PCIs for Non-Normal Distributions

Here we will present methods to compute PCIs for non-normal data distributions.

2.1 Probability Graph

A popular approach to compute PCIs is to use normal distribution graph which allows to investigate the normality assumption. As with normal distribution where process width spans the 0.135 and 99.865 percentile range, a surrogate PCI can be obtained from suitable probability distribution:

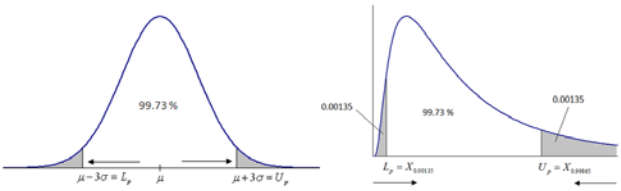

Fig. 2 Normal distribution: interval spanning 99.73% of values, and non-normal distribution: interval spanning 99.73% of values [Own work]

$$C_P = \frac{USL - LSL}{Upper\ 0.135\% - lower\ 0.135\%\ centiles} = \frac{USL - LSL}{U_P - L_P}, \quad (3)$$

where L_p and U_p are 0.135 and 99.865 percentiles, respectively, trivially obtained in statistical software such as Minitab, NCSS, XLStatistics, etc. Due to median being preferred as asymmetric distributions' location parameter, C_{pu} and C_{pl} are defined as

$$C_{pu} = \frac{USL - median}{x_{0.99865} - median}, \quad (4)$$

$$C_{pl} = \frac{median - LSL}{median - x_{0.99865}}, \quad (5)$$

Because the method is graphical, it can produce errors if statistical software is eschewed in favor of a probability paper where x axis is linear and y probabilistic, conforming to a selected probability distribution. N observations are diagramed according to $(x(i))$ with the respective counterparts of the empirical distribution function $((i)\ /\ N)$. Interpolated line is the best fit for a function of quality attribute's expected probability distribution. Quantiles corresponding to (0,00135; 0,50; 0,99865) can be inferred and instated to the performance indicators. However, we repeat the method can be erroneous if statistical software is not involved [19,20].

2.2 Distribution-Independent Tolerance Intervals

Chan et al. [4] devised a method utilizing distribution-independent tolerance intervals to compute C_p a C_{pk} for non-normally distributed processes:

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{USL - LSL}{\frac{3}{2}(4\sigma)} = \frac{USL - LSL}{3(2\sigma)}. \quad (6)$$

This implies

$$C_p = \frac{USL - LSL}{\omega} = \frac{USL - LSL}{\frac{3}{2}\omega_2} = \frac{USL - LSL}{3\omega_3}, \quad (7)$$

$$C_{pk} = \frac{\min[(USL - \mu, \mu - LSL)]}{\omega/2}, \quad (8)$$

where ω is tolerance interval's width. If 99% reliability is required, the interval ensures 75% coverage of population statistics, the method is thus the only one giving different result even if the underlying distribution is normal. In that case, 99.73% of the quality attribute's values lie within ±3s tolerance bounds, C_p = 1 means 99.72% of values lie within the bound. In equation (6), 6 in denominator is generally dependent on sample size out of which s is estimated, and on quality attribute's distribution. The non-normality limitation can be mitigated by selecting suitable percentage of items within the tolerance bounds, e.g., for 99%, 5.15 may be used, a constant valid for many probability distributions with skewness from 0 to 3.111 and kurtosis from 1 to 5.997 [5,10].

2.3 Weighted Variance Method

Choi and Bai [5] proposed a heuristic weighted variance method for PCI correction according to population skewness. Let P_x be probability a variable X is lower or equal to the population mean

$$P_x = \frac{1}{n}\sum_{i=1}^{n} I\left(\overline{X} - X_i\right), \tag{9}$$

where $I(x)=1$ when $x>0$, and $I(x)=0$ when $x<0$. PCI is then defined as [5]

$$C_p = \frac{USL - LSL}{6\omega W_x}, \tag{10}$$

where $W_x = \sqrt{1+|1-2P_x|}$. Therefore

$$C_{pu} = \frac{USL-\mu}{2\sqrt{2P_x\sigma}}, \tag{11}$$

$$C_{pl} = \frac{\mu - LSL}{3\sqrt{2(1-P_x)\sigma}}. \tag{12}$$

2.4 Clements' Method

Clements [6] supplanted 6σ in equation (6) with $U_p - L_p$:

$$C_p = \frac{USL - LSL}{U_p - L_p}, \tag{13}$$

where U_p is the 99.865 percentile. For C_{pk}, μ is estimated by median M and 3σ by $U_p - M$ and $M - U_p$. We get

$$C_{pk} = \min\left[\frac{USL-M}{U_p - M}, \frac{M-LSL}{M-L_p}\right]. \tag{14}$$

Clements' method employs traditional estimates for skewness and kurtosis based on the third and the fourth central moments both of which are unreliable for small sample sizes. The PCIs were devised for a class of *Pearson distributions*, i.e., *beta*, *gamma*, and *Student* [6,16].

<u>Procedure for calculating PCIs:</u>

a) tolerance bounds *USL*, *LSL*,

b) $\overline{x}$, s, Sk (skewness), Ku (kurtosis),

c) standardized quantiles L'_p, U'_p (for selected p and known Sk, Ku can be obtained from tables [6] or in software),

d) standardized median M‘ from tables [6],

e) L_p, U_p: $L_p = \overline{x} - s\cdot L'_p$, $U_p = \overline{x} + s\cdot U'_p$,

f) M: $M = \overline{x} + s\cdot M'$

g) C_p, C_{pk} from equations (13) and (14).

Pearn and Kotz [16] expanded the method with C_{pm}, C_{pm*}, C_{pmk} defined as:

$$C'_{pm} = \frac{USL-LSL}{6\sqrt{\left(\frac{U_p - L_p}{6}\right)^2 + (M-T)^2}}, \tag{15}$$

$$C'_{pm*} = \frac{\min(USL-T, T-LSL)}{3\sqrt{\left(\frac{Up+Lp}{6}\right)^2 + +(M-T)^2}}, \tag{16}$$

$$C'_{pmk} = \min\left[\begin{array}{c}\dfrac{USL-M}{3\sqrt{\left(\frac{U_p - M}{3}\right)^2 + +(M-T)^2}}, \\ \dfrac{M-LSL}{3\sqrt{\left(\frac{M-L_p}{3}\right)^2 + (M-T)^2}}\end{array}\right]. \tag{17}$$

2.5 Burr Percentile Method

Burr [2,3] proposed Burr XII distribution to compute the random variable X's percentiles. Its probability density function, y, is of the form

$$f(y\,|\,c,k) = \begin{cases} \dfrac{cky^{c-1}}{(1+y^c)^{k+1}} & when\ y\ge 0, c\le 1, k\ge 1, \\ 0 & when\ y<0, \end{cases} \tag{18}$$

where c and k represent skewness and kurtosis, respectively.

Liu and Chen introduced modified Clements' method with percentiles from Burr XII distribution instead of from the Pearson's curve.

1. Estimate sample mean $\overline{x}$ and standard deviation s. Skewness s_3 and kurtosis s_4 from empirical sample size are calculated as:

$$s_3 = \frac{n}{(n-1)(n-2)}\sum\left(\frac{x_j - \overline{x}}{s}\right)^3, \tag{19}$$

and

$$s_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^4 - \\ - \frac{3(n-1)^2}{(n-2)(n-3)}. \qquad (20)$$

2. Compute standardized skewness and kurtosis moments α_3 and α_4, respectively:

$$\alpha_3 = \frac{(n-2)}{\sqrt{n(n-1)}} s_3, \qquad (21)$$

and

$$\alpha_4 = \frac{(n-2)(n-3)}{(n^2-1)} s_4 + 3\frac{(n-1)}{(n+1)}. \qquad (22)$$

Standardized kurtosis is known as excess; if it equals 3, the distribution is normal (i.e., mesokurtic). Negative values indicate platykurtic (broader) distribution with higher standard deviation, positive values leptokurtic (slender) distribution with lower standard deviation [1,2,3].

3. Use α_3 and α_4 to select suitable *c* and *k*, then calculate Burr XII distribution standardized value

$$Z = \frac{Y - \mu}{\sigma}, \qquad (23)$$

where *Y* is the selected Burr value, μ and σ mean and standard deviation, respectively. All central moment characteristics are in tables [2,3] which also supply standardized 0.00135, 0.5, and 0.99865 percentiles corresponding to $Z_{0.00135}$, $Z_{0.5}$, and $Z_{0.99865}$. We get them by comparing

$$\frac{X - \bar{x}}{s} = \frac{Y - \mu}{\sigma}.$$

4. We now have estimates for lower, medium, and upper percentiles:

$$L_p = \bar{x} + s\ Z_{0.00135},$$

$$M = \bar{x} + s\ Z_{0.5},$$

$$U_p = \bar{x} + s\ Z_{0.99865}.$$

5. Compute the PCI from equations for the Clements' method.

Other procedures can be used to estimate Burr distribution's parameters apart from the third and the fourth central moments, e.g., maximum likelihood method, probability weighted moments, etc. Moment characteristics are the most usual due to their simplicity and speed of computation [2,3].

2.6 Transformation Techniques

1. Variance stabilization requires a transformation $y = g(x)$ with constant variance $\sigma^2(y)$. In case variance of x is a function of the type $\sigma^2(x) = f_1(x)$, $\sigma^2(y)$ is computed as

$$\sigma^2(y) \approx \left[\frac{dg(x)}{dx} \right]^2, f_1(x) = C, \qquad (24)$$

where C is a constant; $g(x)$ is then obtained by solving a differential equation

$$g(x) \approx C \int \frac{dx}{\sqrt{f_1(x)}}. \qquad (25)$$

Multiple methods and tools were devised for constant relative error $\delta(x)$, e.g., $\sigma^2(x)$ is given by $f_1(x) = \delta^2(x)$, $x^2 =$constant. After instating, we get $g(x) = \ln x$. Here, logarithmic transformation is optimal together with geometric mean. When $\sigma^2(x) = f_1(x)$ exhibits a power relation, $g(x)$ will be of power type as well. For normal distribution, mean is independent on variance, i.e., variance-stabilizing transformation will ensure convergence to normality.

2. Sample distribution symmetrization is possible using simple power transformation

$$y = g(x) = \begin{cases} x^{\lambda} & for\ \lambda > 0 \\ \ln x & for\ \lambda = 0 \\ -x^{-\lambda} & for\ \lambda < 0. \end{cases} \qquad (26)$$

Scale is not preserved, is discontinuous due to λ, and suitable only for positive values. Optimal $\hat{\lambda}$ estimate is found by minimizing the suitable asymmetry characteristics. Skewness defined in the following equation is robust and can be used in place of $\hat{g}_1(y)$:

$$\hat{g}_{1R}(y) = \frac{(\bar{y}_{0.75} - \bar{y}_{0.50}) - (\bar{y}_{0.50} - \bar{y}_{0.25})}{\bar{y}_{0.75} - \bar{y}_{0.25}}. \qquad (27)$$

For a symmetric distribution, $\hat{g}_1(y) = 0$, as well as $\hat{g}_1(y)$ and $\hat{g}_{1R}(y)$. Optimal $\hat{\lambda}$ is found from a rankit plot; its quantiles will lie approximately on a line.

3. **Hine-Hines selection graph** (x axis: $\tilde{x}_{0.5} / x_{1-P_i}$, y axis: $\tilde{x}_{P_i} / \tilde{x}_{0.5}$)

A selection graph as a diagnostic tool for finding optimal λ value.

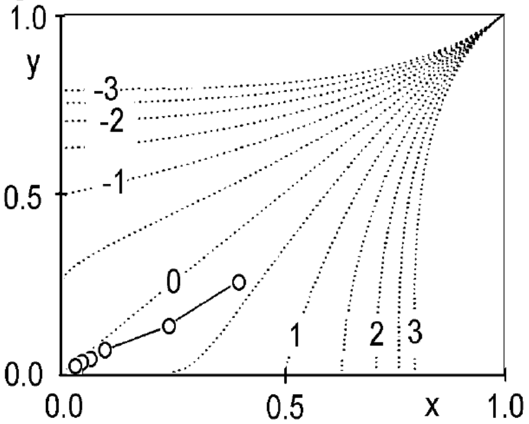

Fig. 3 Selection graph for a sample from log-normal distribution [14]

The graph originates from asymmetry requirements of quantiles around median

$$\left(\frac{\tilde{x}_{P_i}}{\tilde{x}_{0.5}}\right)+\left(\frac{\tilde{x}_{0.5}}{\tilde{x}_{1-P_i}}\right)^{-\lambda}=2, \tag{28}$$

where for each ordinal probability, values of $P_i = 2^{-i}$, $i = 2,3$ are usually assigned. To compare experimental and theoretical observation for selected λ, $y^{\lambda} + x^{-\lambda} = 2$ for $0 \le x \le 1$ and $0 \le y \le 1$ is plotted into the graph:

a) for $\lambda = 0$, the solution is a line $y = x$,

b) for $\lambda < 0$, the solution is a ratio $y = \left(2 - x^{-\lambda}\right)^{1/\lambda}$,

c) for $\lambda > 0$, the solution is a ratio $x = \left(2 - y^{\lambda}\right)^{-1/\lambda}$.

Judging from the positions of experimental points on theoretical curves in the selection graph, λ can be inferred and quality of transformation assessed in varying distances from median. To converge the sample distribution to normal, Box-Cox transformation is a viable method due to skewness and kurtosis [14].

2.6.1 Box-Cox Transformation

Disadvantages of a simple power transformation (discontinuity around zero and non-aligned transformed scales) are mitigated by Box-Cox transformation $X^{(\lambda)}$, a linear variation of a simple power transformation $X_p^{(\lambda)}$. Box-Cox class of polynomial transformations is of the form

$$X^{(\lambda)} = \begin{cases} \dfrac{X^{\lambda}-1}{\lambda} & \text{for } \lambda \neq 0 \\ \ln X & \text{for } \lambda = 0. \end{cases} \tag{29}$$

To determine its quality and optimal λ, quantile dispersion graphs (QDG) and quantile graphs (Q-Q) can be plotted, with normality tests after power transformation more suitable. A well-known Shapiro-Wilk test is equivalent to testing tangent's significance in a Q-Q plot, i.e., linearity can be analyzed as well. The Box-Cox family of transformations depends on λ estimated by Maximum Likelihood Estimate (MLE) or Least-Squares Estimate (LSE). For $\lambda = 1$, additive measurement model is suitable while for $\lambda = 0$, a multiplicative one should be preferred. First, λ is chosen from a selected range, and the following statistic computed

$$L_{\max} = -\frac{1}{1}\ln\hat{\sigma}^2 + \ln J(\lambda, X) = \\ = \frac{1}{2}\ln\hat{\sigma}^2 + (\lambda-1)\sum_{i=1}^{n}\ln X_i, \tag{30}$$

where

$$J(\lambda, X) = \prod_{i=1}^{n}\frac{\partial W_i}{\partial X_i} = \prod_{i=1}^{n} X_i^{\lambda-1}\ \forall\lambda,$$

so that $\ln J(\lambda, X) = (\lambda-1)\sum_{i=1}^{n}\ln X_i$. The $\hat{\sigma}^2$ estimate for fixed λ is $\hat{\sigma}^2 = S(\lambda)/n$ where $S(\lambda)$ is a residual sum of squares in analysis of variance $X(\lambda)$. After computing $L_{\max}(\lambda)$ for several λ in the range, $L_{\max}(\lambda)$ is plotted against λ. MLE for λ is computed from the λ maximizing $L_{\max}(\lambda)$. With optimal λ^*, every X in the specification bounds is transformed to a normal value with the help of equation (29) [4,18].

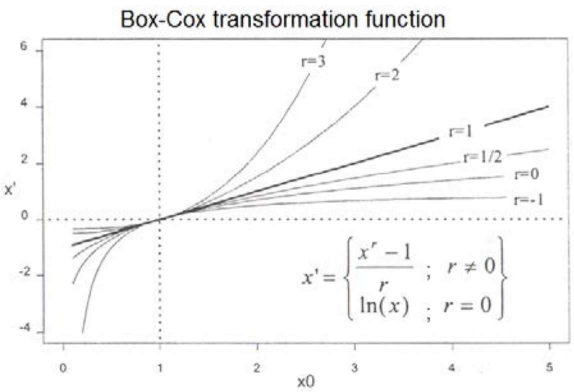

Fig. 4 Shape of Box-Cox transformation for r = -1, 0, 0.5, 1, 2, 3 [13]

Box-Cox transformation estimates λ by minimizing standard deviation of the normalized transformed variable; software is capable to evaluate many of them. *y* is a transformed value of the *x* quality attribute [18].

λ denotes the function's shape and allows to find *F(x)* satisfying the maximum normality or symmetry condition. Different shapes of the Box-Cox transformation function (specified as *r*) for various λ are depicted in Fig. 4. The aim is to find *r* ensuring maximum symmetry "measured" by skewness, or (better) maximum normality "measured" by likelihood, after transformation. Maximum symmetry (zero skewness condition) or maximum likelihood condition is computed iteratively [15].

2.6.2 Backward Transformation After Box-Cox and Power Transformation

If a suitable transformation leading to approximate normality is found, $\bar{y}, s^2(y)$, and confidence intervals $\bar{y} \pm t_{1-\alpha/2}(n-1)s(y)/\sqrt{n}$ can be found and statistical tests performed. However, we fist need to transform the statistics back to their original forms.

1. **Incorrect approach** simply backward transforms $\bar{x}_R = g^{-1}(\bar{y})$. For a simple power case, backward transformation leads to mean of the form

$$\bar{x}_R = \bar{x}_\lambda = \left[\frac{\sum_{i=1}^{n} x_i^\lambda}{n}\right]^{1/\lambda}. \tag{31}$$

For $\lambda = 0$, $\ln x$ is used instead of x^λ and e^x instead of $x^{1/\lambda}$. $\bar{x}_R = \bar{x}_{-1}$ represents harmonic, $\bar{x}_R = \bar{x}_0$ geometric, $\bar{x}_R = \bar{x}_1$ arithmetic, and $\bar{x}_R = \bar{x}_2$ quadratic mean, respectively. The approach leads to biased results.

2. **Correct (more exact) approach** is based on Taylor expansion $y = g(x)$ around $\bar{y}$. For untransformed $\bar{x}_R$, an approximate relation is

$$\bar{x}_R \approx g^{-1}\left[\bar{y} - \frac{1}{2}\frac{d^2 g(x)}{dx^2}\left(\frac{dg(x)}{dx}\right)^{-2} \cdot s^2(y)\right] \tag{32}$$

for mean and

$$s^2(x_R) \approx \left(\frac{dg(x)}{dx}\right)^{-2} \cdot s^2(y) \tag{33}$$

for variance. Each derivation is computed at $x = \bar{x}_R$. For $100(1-\alpha)\%$ confidence interval for mean of original data, it holds that $x_R - I_L \le \mu \le x_R + I_U$ where

$$I_L = g^{-1}\left(\bar{y} + G - t_{1-\alpha/2}(n-1)\frac{s(y)}{\sqrt{n}}\right), \tag{34}$$

$$I_U = g^{-1}\left(\bar{y} + G - t_{1-\alpha/2}(n-1)\frac{s(y)}{\sqrt{n}}\right), \tag{35}$$

$$G = -0.5\frac{d^2 g(x)}{dx^2}\left(\frac{dg(x)}{dx}\right)^{-2} \cdot s^2(y). \tag{36}$$

$t_{1-\alpha/2}(n-1)$ denotes $100(1-\alpha/2)\%$ Student quantile with $(n-1)$ degrees of freedom. With $y = g(x)$ and $\bar{y}$, $s^2(y)$, it is trivial to compute $\bar{x}_R$ and $s^2(x_R)$:

a) For a special case of $\lambda = 0$, i.e., logarithmic transformation of the type $g(x) = \ln x$:

$\bar{x}_R \approx \exp\left[\bar{y} + 0.5 s_2(y)\right]$, $s^2(x_R) \approx \bar{x}_R^2 s^2(y)$.

b) For $\lambda \ne 0$, $\bar{x}_R$ is the following quadratic equation's root:

$$\bar{x}_{R,1,2} = \begin{bmatrix} 0.5(1+\lambda\bar{y}) \pm \\ \pm 0.5\sqrt{1 + 2\lambda(\bar{y} + s^2(y)) + (\lambda^2 - 2s^2(y))} \end{bmatrix}^{1/\lambda}. \tag{37}$$

$\bar{x}_{R,i}$ closer to $\bar{x}_{0.5} = g^{-1}(\bar{y}_{0.5})$ is taken as an estimate of $\bar{x}_R$. With known untransformed mean, variance can be computed as $s^2(x) = \bar{x}_R^{-2\lambda+2} \cdot s^2(y)$ [14].

2.6.3 Johnson Transformation

When observations are non-normally distributed, Johnson transformation can be used; the data will then be *N (0,1)*-normally distributed. It's of the form

$$\begin{gathered} z = \gamma + \eta\tau(x;\varepsilon,\lambda), \\ \eta > 0, -\infty < \gamma < \infty, \lambda > 0, -\infty < \varepsilon < \infty, \end{gathered} \tag{38}$$

where *z* is a standardized normal random variable and *x* a variable fitted using Johnson transformation. Four parameters to be estimated are γ, η, ε, and λ with τ an arbitrary function taking one of the following three forms. Johnson transformation therefore selects one type depending on whether the random variable is "log-normal", "bounded", or "unbounded" [9,11,12].

Log-normal System (S_L)

$$\tau_1(x;\varepsilon,\lambda)=\log\left(\frac{x-\varepsilon}{\lambda}\right), x\ge\varepsilon \tag{39}$$

Titled Johnson log-normal S_L distribution, it is tied to a log-normal system. Parameter estimates are

$$\hat{\eta}=1.645\left[\left(\frac{x_{0.95}-x_{0.5}}{x_{0.5}-x_{0.05}}\right)\right]^{-1}, \tag{40}$$

$$\hat{\gamma}^*=\hat{\eta}\log\left(\frac{1-\exp(-1.645/\hat{\eta})}{x_{0.5}-x_{0.05}}\right), \tag{41}$$

$$\hat{\varepsilon}=x_{0.5}=\exp\left(-\hat{\gamma}^*/\hat{\eta}\right), \tag{42}$$

where 100α% percentile is computed as $\alpha(n+1)$ th value from n observations. If necessary, linear interpolation can be used to get the desired percentile [11,12].

Unbounded System (S_U)

$$\tau_2(x;\varepsilon,\lambda)=\sinh^{-1}\left(\frac{x-\varepsilon}{\lambda}\right), -\infty<x<\infty. \tag{43}$$

The curves are unbounded and the system consists of both Student's and normal distributions among others. Hahn and Shapiro supplied a table for $\hat{\gamma}$ and $\hat{\eta}$ based on given skewness and kurtosis for filtering.

Bounded System (S_B)

$$\tau_3(x;\varepsilon,\lambda)=\log\left(\frac{x-\varepsilon}{\lambda+\varepsilon-x}\right), \varepsilon\le x\le\varepsilon+\lambda. \tag{44}$$

The system includes bounded distributions, i.e., gamma and beta. Because each may be lower-bounded ε or upper-bounded $(\varepsilon+\lambda)$,, the following can occur:

- *Fluctuation range is known*: both end points are specified, the parameters are obtained by

$$\hat{\eta}=\frac{z_{1-\alpha'}-z_\alpha}{\log\left(\frac{(x_{1-\alpha'}-\varepsilon)(\varepsilon+\lambda-x_\alpha)}{(x_\alpha-\varepsilon)(\varepsilon+\lambda-x_{1-\alpha'})}\right)}, \tag{45}$$

$$\hat{\gamma}=z_{1-\alpha'}-\hat{\eta}\log\left(\frac{x_{1-\alpha'}-\varepsilon}{\varepsilon+\lambda-x_{1-\alpha'}}\right). \tag{46}$$

- *One known end point*: additional equation (given by respective medians) is required to supplement $\hat{\eta}$ and $\hat{\gamma}$; given as

$$\hat{\lambda}=(x_{0.5}-\varepsilon)\left\{\begin{array}{l}(x_{0.5}-\varepsilon)(x_\alpha-\varepsilon)+(x_{0.5}-\varepsilon)\\(x_{1-\alpha}-\varepsilon)-2(x_\alpha-\varepsilon)(x_{1-\alpha}-\varepsilon)\end{array}\right\}\times \tag{47}$$
$$\times\left\{(x_{0.5}-\varepsilon)^2-(x_\alpha-\varepsilon)(x_{1-\alpha}-\varepsilon)\right\}^{-1}.$$

- *No end point is known*: four percentiles must be compared to those from normal distribution. The equation for i = 1, 2, 3, 4:

$$z_i=\hat{\gamma}+\hat{\eta}\log\left(\frac{x_i-\hat{\varepsilon}}{\hat{\varepsilon}+\hat{\lambda}-x_i}\right). \tag{48}$$

Being non-linear, it is solved numerically.

Hill et al. devised an algorithm to assign first four central moments of X to the above-mentioned distribution types. PCIs are quantified from equations (1) and (2) [11,12].

2.7 Wright's Process Capability Index C_s

Wright proposed a process capability index C_s incorporating skewness correction factor in the numerator of C_{pmk}. It is of the form

$$C_s=\frac{\min(USL-\mu,\mu-LSL)}{3\sqrt{\sigma^2+(\mu-T)^2+|\mu_3/\sigma|}}= \tag{49}$$
$$=\frac{\min(USL-\mu,\mu-LSL)}{3\sqrt{\sigma^2+|\mu 3/\sigma|}},$$

where $T=\mu$ and μ_3 is the third central moment [19].

Some methods are widely employed in industrial applications, such as probability graphs and Clements' method. Conversely, Box-Cox transformation is relatively unknown among experts. If the distribution is normal, all methods except for the distribution-independent one should provide identical results to traditional C_p and C_{pk} as per equations (1) and (2). However, as different statistics and procedures are used for parameter estimates, there will be some variability in C_p and C_{pk} with that of C_p in particular decreasing with increasing sample size, e.g., $\chi^2_{1-\alpha}$ distribution assuming normality. Variability of C_{pk} can be substantial for all sample sizes as it depends on process variability caused by mean shifts and other changes.

3 Case Study

3.1 Data Comparison Metric

Different test metrics may lead to different results, accentuating that it is imperative to set a suitable performance comparison measure. Researchers use various ones to address the PCI non-normality problem. English and Taylor fixed C_p and C_{pk} to 1.0 for all their simulations investigating PCI robustness under non-normality. The basis of their comparison was a $\hat{C}_p$ and $\hat{C}_{pk}$ ratio (established from the simulation) higher than 1.0 in case of non-normal distribution. Deleryd focused on a ratio of unsatisfactory items to determine C_p. Estimated $C_p(u,v,)$'s bias and variance were compared with a target C_p. As direct relation between PCI and defects exists, choice of index also sets acceptable percentage of detects. Rivera et al. [17] tweaked population's upper specification bounds to get true number of defects and respective C_{pk}. C_{pk} from simulations of transformed data were compared with target values of C_{pk}.

In practice, PCIs are frequently used to monitor performance and compare processes. Even though such application are not recommended when normality is untested, "substitute" PCIs for non-normal data should be compatible with PCIs computed under normality when defect ratios are approximately equal. This explains solution proposed by Rivera et al: the defects ratio is set a priori using suitable specification limits and the PCIs are then compared with the target value, leading to analyzing one-sided tolerance where C_{pu}, a one-sided tolerance limit for PCI, is used as a comparison measure. C_{pu} is computed as

$$Ratio\ of\ unsatistifactory\ items = \Phi(-3C_{pu}), \tag{50}$$

Ratio of unsatisfactory items for asymmetric two-sided tolerance:

$$\begin{aligned} &Ratio\ of\ unsatistifactory\ items = \\ &= \Phi(-3C_{pl}) + \Phi(-3C_{pu}). \end{aligned} \tag{51}$$

In the study, $C_{pu} = 1,\ 1.5$ and 1.667 and was derived from respective USL for log-normal and Weibull distributions with identical ratio of unsatisfactory items as targets. USL's serve to estimate C_{pu} which are then compared with the target values. The method whose C_{pu} sample mean estimate exhibits the least deviation with the smallest variance as measured by dispersion or standard deviation, is the best. Graphical representation depicting the two characteristics is a simple box and whisker plot.

3.1.1 Test Distribution

To investigate effects of non-normal data on PCIs, log-normal and Weibull distributions were selected as it is known their parameters can represent fractional, medium and high deviations from normality. They are also known for different properties near tails which substantially influence process capability. Skewness and kurtosis for both are summarized in Table 1, distribution functions are given respectively as

$$F(x;\mu,\sigma) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right),\ x \geq 0, \tag{52}$$

$$F(x;\eta,\sigma) = 1 - \exp\left[-\left(\frac{x}{\sigma}\right)^n\right],\ x \geq 0. \tag{53}$$

Log-normal distribution		Weibull distribution	
μ	σ^2	Shape parameter	Scale parameter
0.0	0.1	1.0	1.0
0.0	0.3	2.0	1.0
0.0	0.5	4.0	1.0
1.0	0.5	0.5	1.0

Tab. 1 Parameters used in performance comparison simulations [Own work]

Fig. 5 depicts shapes of log-normal and Weibull distributions with parameters from Tab. 1.

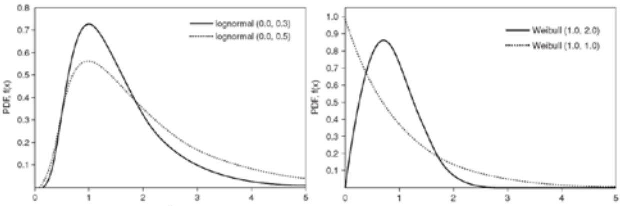

Fig. 5 Probability density functions for log-normal and Weibull distributions with parameters from Table 1 [18], modified

3.2 Monte Carlo Simulations

A series of simulations was performed with sample sizes n = 50, 75, 100, 125 and 150, target C_{pu} were set to 0.1, 1.33, 1.5, 1.667 and 1.8 using log-normal and Weibull distributions. We only list results for n = 100 and C_{pu}'s 1.0, 1.5 and 1.667. In each run, statistics such as $\bar{x}$, s, upper and lower 0.135 percentiles aas well as median were obtained from random variables generated from both distributions. For the transformation methods, they are basically obtained from the transformed values. C_{pu}'s were subsequently determined using the 9 methods presented above with the statistics for each USL target [4,7,18].

Log-normal distribution			
μ	σ^2	Skewness $\sqrt{\beta_1}$	Kurtosis β_2
0.0	0.5	2,9388	21,5073
0.0	0.3	1,9814	10,7057
0.0	0.1	1,0070	4,8558
1.0	0.5	2,9388	21,5073
Weibull distribution			
Shape parameter	Scale parameter	Skewness $\sqrt{\beta_1}$	Kurtosis β_2
1,0	1,0	2,0000	9,0000
2,0	1,0	0,6311	3,2451
4,0	1,0	-0,0872	2,7478
0,5	1,0	2,0000	9,0000

Tab. 2 Skewness and kurtosis for the investigated values [Own work]

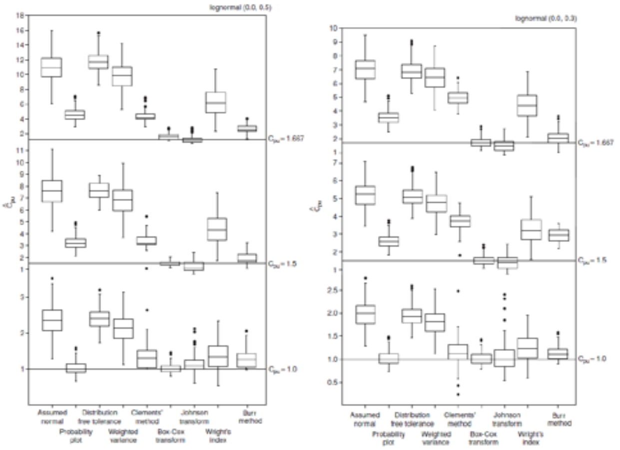

Fig. 6 Probability density functions for log-normal and Weibull distributions with parameters from Tab. 1 [18], modified

Each iteration was run 100 times to obtain mean $\overline{C}_{pu}$ from $\hat{C}_{pu}$. To investigate which method is the best for handling non-normality, box plots where $\hat{C}_{pu}$'s will be plotted with target C_{pu}'s for both distributions will be generated. The plots depict several statistics pertaining to C_{pu}, e.g., mean, variance, outliers and extreme values. Box plots which could capture the target C_{pu}, mean will intersect the horizontal line on the target level [18].

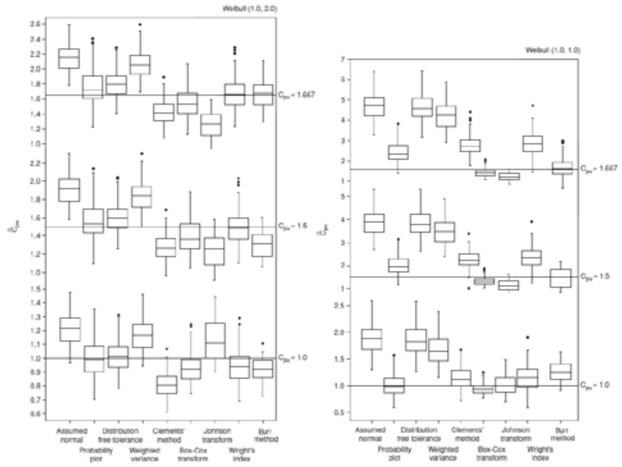

Fig. 9 $\hat{C}_{pu}$ box plots (100 observations each) for log-normal distribution (μ= 0, σ^2 = 0.5, 0.3), target C_{pu} = 1.0, 1.5, 1.667 [Own work]

4 Discussion and Conclusion

Generally, the methods can be put into two categories. Probability graph, distribution-independent tolerance intervals, weighted variance method, and Wright's process capability index can be classified as non-transformation methods. The second category comprises transformation techniques: Clements', Burr's, Box-Cox's, and Johnson's. In the performance comparison part, two measures are considered: accuracy and correctness with respect to sample size. For the former, we will look at difference between simulated mean $\overline{C}_{pu}$ and target C_{pu}, for the latter, we will consider variance or range of simulated means. The lower the variance or range, the better the performance. The results show transformation methods are better for both distributions, with two exceptions. First: Clements' method does not work as well as the rest. For Weibull distribution, its performance is lower than that of probability graph method. Second: box plots show probability graph method is the only one which outperforms transformation techniques for Weibull distribution with σ = 1.0 and η = 2.0. For log-normal distribution, transformation techniques supply consistently better $\overline{C}_{pu}$'s closer to the true C_{pu}'s, all other inflate the results. From a practical point of view, non-transformation methods are easier to compute but were proven insufficient for quantifying process capability except where population distribution is approximately normal.

Differences in $\hat{C}_{pu}$'s are mostly higher compared to transformation techniques and their use is not recommended for distributions significantly deviating from normal. When one method fails, transformed data are required which converge them closer to normal distribution. For further computations, they need to be backward transformed. Transformation class' performance is fairly consistent with respect to correctness and accuracy; however, they are sensitive to sample size as evident for $\overline{C}_{pu}$ where n = 50 and 150: the differences are 33%, 11% and 26% for Clements', Box-Cox, and Johnson transformation methods, respectively, which is more than 10% difference compared to non-transformation ones. To summarize, transformation techniques exhibit lower $\hat{C}_{pu}$ variability as well as higher differences between $\hat{C}_{pu}$ computed from small and large sample sizes which cannot be ascribed only to inherent statistical variability. This hints at them being favorable for smaller sample sizes. Possible reasons include:

1. They usually involve conversion from distribution's scale of range to shape. This may induce unwanted process shifts influencing $\hat{C}_{pu}$, especially for smaller sample sizes.
2. For monitoring capability, their success depends on the ability to detect behavior near tails via UCL percentiles.

Transformation techniques should be used to monitor process capability with $n \geq 100$. Nevertheless, Box-Cox's dominance is apparent from the graphs, particularly for log-normal distribution which is to be expected since it provides exact transformations to natural logarithms. In case of log-normal distribution, this results in convergence to normality. Comparing Box-Cox's and Johnson's methods, the former is more accurate as it requires maximizing logarithmic likelihood function to get precise λ while the latter operates on estimates of the first four central moments. Probability graph method is recommended for processes slightly deviating from normality, it also can detect anomalies such as bimodality and outliers. Another finding is that Burr percentile method is comparable to Box-Cox's transformation. Their performance for small sample sizes, specifically 30 sample sizes with n = 50, was analyzed to determine which one is better for samples frequently encountered in practice. The results are depicted in Fig. 8.

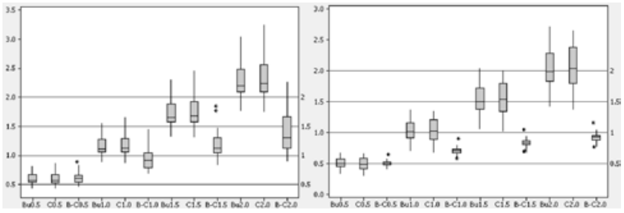

Fig. 8 Box plots (100 observations each) for Weibull distribution (σ = 0.0, 1.0, η = 1.0, 2.0), target Cpu = 1.0, 1.5, 1.667 [Own work]

The graph shows Burr percentile method generally allows for better process capability estimates in small, non-normal sample sizes. Monte Carlo simulations corroborated traditional PCIs produce biased process capability results in case of non-normal distributions as they firmly rely on the normality assumption.

Acknowledgements

This contribution was written within the framework of the GA ČR (Czech Science Foundation) grant-maintained project: Reg. No. 407/12/0821, Creating a Czech Instrument for Measuring Academic Tacit Knowledge, and with the financial support of GA ČR.
The authors are also thankful to the Operational Programme Education for Competitiveness co-funded by the European Social Fund (ESF) and national budget of the Czech Republic for the grant No. CZ.1.07/2.3.00/20.0147 - "Human Resources Development in the field of Measurement and Management of Companies, Clusters and Regions Performance", which provided financial support for this research.

References:

[1] E. Abouheif, A method for testing the assumption of phylogenetic independence in comparative data, Evol. Ecol. Res. 1 (1999) 895-909.
[2] I. W. Burr, Cumulative frequency functions, Ann. Math. Stat. 13:2 (1942) 215-232.
[3] I. W. Burr, Parameters for a general system of distribution to match a grid of α_3 a α_4, Commun. Stat. 2:1 (1973) 1-21.
[4] L. K. Chan, S. W. Cheng, F. A. Spiring, A graphical technique for process capability, Trans. ASQC Annu. Qual. Congr. 42 (1988) 268–275.
[5] I. S. Choi, D. S. Bai, Process capability indices for skewed populations, Proc. 20th Int. Conf. Comp. Ind. Eng. (1996) 1211-1214.

[6] J. A. Clements, Process capability indices for non-normal calculations, Qual. Prog. 22 (1989) 49-55.

[7] M. Dlouhý, J. Fábry, M. Kuncová, T. Hladík, Simulace podnikových procesů, Brno, Computer Press, 2007.

[8] F. Fabian, V. Horálek, J, Křepela, J. Michálek, V. Chmelík, J. Chodounský, J. Král, Statistické metody řízení jakosti, Praha, ČSJ, 2007.

[9] I. D. Hill, R. Hill, R. L. Holder, Algorithm AS99: Fitting Johnson curves by moments, J. R. Stat. Soc. 25:2 (1976) 180-189.

[10] K. Ishikawa, What Is Total Quality Control? The Japanese Way, New York, Prentice Hall, 1988.

[11] N. L. Johnson, S. Kotz, W. L. Pearn, Flexible process capability indices, Pak. J. Stat. 10 (1992) 23-31.

[12] V. E. Kane, Process capability indices, J. Qual. Technol. 18 (1986) 41-52.

[13] K. Kupka, Statistické řízení jakosti, Pardubice, TriloByte, 2001.

[14] M. Meloun, J. Militký, Statistická analýza experimentálních dat, second ed., Praha, Academia, 2004.

[15] M. Meloun, J. Militký, Kompendium statistického zpracování dat, second ed., Praha, Academia, 2006.

[16] W. Pearn, S. Kotz, Application of Clements' method for calculating second- and third-generation process capability indices for non-normal Pearsonian populations, Qual. Eng. 7:1 (1994) 139-145.

[17] L. A. R. Rivera, N. F. Hubele, F. D. Lawrence, Cpk index estimation using data transformation, Comput. Ind. Eng. 29:1-4 (1995) 55-58.

[18] L. C. Tang, S. E. Than, B. W. Ang, A graphical approach to obtaining confidence limits of Cpk. Qual. Reliab. Eng. Int. 13:6 (1997) 337-346.

[19] P. A. Wright, A process capability index sensitive to skewness, J. Stat. Comput. Sim. 52:3 (1995) 195-203.

[20] K. Yang, B. El-Haik, Design For Six Sigma: A Roadmap for Product Development, New York, McGraw-Hill, 2003.