Int J Adv Manuf Technol (2007) 31: 957–968
DOI 10.1007/s00170-005-0279-3

ORIGINAL ARTICLE

S. H. Chung · W. L. Pearn · Y. S. Yang

A comparison of two methods for transforming non-normal manufacturing data

Received: 5 February 2004 / Accepted: 5 May 2005 / Published online: 22 March 2006

Abstract Many statistical methods applied to manufacturing quality control and operations management have been under the assumption that the process characteristic investigated is normally distributed. If the process characteristic is not normally distributed, a popular approach is to transform the non-normal data into a normal one. In this paper, we consider the Box-Cox transformation, and compare the transformation power using two different parameter estimation methods, including the maximum likelihood estimator (MLE) and the method of percentiles (MOP). The performance comparison is based on the pass rate under the Shapiro-Wilk normality test. The results show that, in general, the MOP has better pass rate, while the MLE has smaller power variation for most cases investigated. For small sample size (n=5, 10) both methods perform equally well. For large sample size, the MOP is recommended due to its simplicity and significantly higher pass rate.

Keywords Box-Cox transformation · Non-normal distribution · Parameter estimation

1 Introduction

For most industrial applications, normality is assumed due to the advantage of analytical convenience and existing effective statistical methods. But, for many engineering operations such as locating pins or automatic sensors, the manufacturing data is often truncated or appears to be non-normal. Pezdek [9] gave a non-normal data example and perform process performance analysis. Pezdek [9] demonstrated how the non-normal characteristic would significantly impact the data analysis result and the conclusion, thus conveying incorrect process information. If the process characteristic is not normally distributed, a popular approach is to transform the non-normal data into a normal one. Box and Cox [1] modified the family of power transformation introduced by Tukey [13] which, in its simple form, consists of transformations T_λ: $y \rightarrow y^\lambda$ defined as

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda} (\lambda \neq 0) \\ \ln y (\lambda = 0) \end{cases} \quad (1.1)$$

The transformation in Eq. 1.1 is defined for y>0. It is hoped that for some value of λ, the data transformed for a non-normal one can be fitted to a normal distribution. Box and Cox [1] considered the maximum likelihood method and the Bayesian approach for estimating the parameter λ. Draper and Cox [2] derived an analytical expression for the accuracy of maximum likelihood estimate of λ. Hinkley [5] proposed an analytical procedure to estimate the transformation parameter based on order statistics. Hinkley [5] further proposed a simple method for choosing appropriate power transformation for non-normal data. Hernandes and Johnson [4] introduced an approach, by minimizing the Kullback-Leibler information, to optimize the normal transformation; but their procedure for selecting the best value of λ uses the information function, and is therefore restricted to cases with large sample size n. Taylor [12] used the skewness coefficient as a measure of the process symmetry to estimate the transformation power. Taylor [12] showed that the use of skewness coefficient could significantly improve the transformation power for some special cases. Han [3] investigated a non-parametric approach, and proposed an estimator for the transformation based on Kendall's rank correlation. Due to the complexity in computing the maximum likelihood estimate of an extended Box-Cox model, Ogwang [8] proposed a simple algorithm which accommodates the classical model in Box and Cox [1]. An illustrative example is presented in Ogwang [8] to demonstrate good performance of the algorithm.

In this paper we consider the Box-Cox transformation [1] and compare the transformation power using two

S. H. Chung (✉) · W. L. Pearn · Y. S. Yang
Department of Industrial Engineering & Management,
National Chiao Tung University,
Taiwan, Republic of China
e-mail: shchung@mail.nctu.edu.tw

different parameter estimation methods, including the maximum likelihood estimator (MLE) and the method of percentiles (MOP). The performance comparison is based on the pass rate under the Shapiro-Wilk normality test [11].

2 Box-Cox transformation based on MLE

Suppose that $Y_1,\ldots,Y_n$ are continuous nonnegative independent and identically distributed random variables. A restriction on the variables to be positive is necessary for the transformation in Eq. 1.1 to be valid. Consider the power transformation in Eq. 1.1 and its extension simplified from Han [3]:

$$y_i^{(\lambda)} = \beta + e_i, i = 1, 2, \cdots n, \tag{2.1}$$

where β is a constant and e_i is the error term. Assume that the errors are independent and approximately normally distributed with mean zero and variance σ^2. The Jacobian of transformation from e_i to y_i is $y_i^{\lambda-1}$, and the log-likelihood of the observed sample, $y_1, y_2,\ldots,y_n$, is given by

$$\begin{aligned} L(\lambda) &= -(n/2)\ln(2\pi) - (n/2)\ln\sigma^2 \\ &\quad - \left(2\sigma^2\right)^{-1}\sum_{i=1}^{n}\left[y_i^{(\lambda)} - \beta\right]^2 + (\lambda-1)\sum_{i=1}^{n}\ln y_i \end{aligned} \tag{2.2}$$

The maximum likelihood estimator can be obtained by maximizing the log–likelihood function of Eq. 2.2 with respect to β, σ^2 and λ. Dividing by its geometric mean scales, $\dot{y} = \exp\left(n^{-1}\sum\ln y_i\right)$, the Jacobian term of Eq. 2.2 becomes zero. Schlesselman [10] noted that the constant β in the model is essential to the scale invariance property of the Box-Cox transformation. This implies that we can obtain the scaled version maximum likelihood estimate of the parameter λ for the Box-Cox transformation, and then retrieve λ of the original un-scaled transformation. Correspondingly, dividing both sides of Eq. 2.1 by $\dot{y}^{\lambda}$ with some manipulations we obtain the following,

$$y_i^{*(\lambda)} = \beta^* + e_i^*, i = 1, 2, \cdots, n \tag{2.3}$$

where $y_i^* = y_i\big/\dot{y}^{\lambda}$, $\beta^* = \left[\beta - \dot{y}^{(\lambda)}\right]\big/\dot{y}^{\lambda}$, *and* $e_i^* = e_i\big/\dot{y}^{\lambda}$. The log-likelihood of the scaled sample $y_1^*, \cdots, y_n^*$ is given by

$$\begin{aligned} L^*(\lambda) &= -(n/2)\ln(2\pi) - (n/2)\ln\sigma^{*2} \\ &\quad - \left(2\sigma^{*2}\right)^{-1}\sum_{i=1}^{n}\left[y_i^{*(\lambda)} - \beta^*\right]^2 \end{aligned} \tag{2.4}$$

where σ^{*2} is the variance of e_i^* . The maximum likelihood estimators of the parameters in Eq. 2.3 can be derived by maximizing Eq. 2.4 with respect to β, σ^{*2} and λ. Taking partial differentiation with respect to β^*, σ^{*2} and λ, and setting the partial derivatives to zero, we obtain the following normal equations:

$$\sum_{i=1}^{n} e_i^* = 0 \tag{2.5}$$

$$\sigma^{*2} = \sum_{i=1}^{n} e_i^{*2}\Big/ n \tag{2.6}$$

$$\left(-1/\lambda^2\sigma^{*2}\right)\sum_{i=1}^{n} e_i^*\left[\lambda y_i^{*\lambda}\ln y_i - \left(y_i^{*\lambda} - 1\right)\right] = 0 \tag{2.7}$$

From Eq. 2.5, we solve for the maximum likelihood estimator of β^* to obtain

$$\widehat{\beta}^* = \overline{y}^{*(\lambda)} \tag{2.8}$$

By plugging Eq. 2.8 in Eq. 2.6, the maximum likelihood estimator of σ^{*2} is

$$\widehat{\sigma}^{*2} = \sum_{i=1}^{n}\left(y_i^{*(\lambda)} - \overline{y}^{*(\lambda)}\right)^2\Big/ n \tag{2.9}$$

Multiplying both sides of Eq. 2.7 by $-\lambda\sigma^{*2}$, and setting $\widehat{e}_i^* = y_i^{*(\lambda)} - \overline{y}^{*(\lambda)}$, we obtain

$$\begin{aligned} &\sum_{i=1}^{n}\widehat{e}_i^*\left(y_i^{*\lambda}\ln y_i - y_i^{*(\lambda)}\right) \\ &\qquad = \sum_{i=1}^{n}\widehat{e}_i^*\left(y_i^{*\lambda}\ln y_i - \widehat{\beta}^* - \widehat{e}_i^*\right) = 0 \end{aligned} \tag{2.10}$$

With Eq. 2.5 and $\widehat{\beta}^*\sum_{i=1}^{n}\widehat{e}_i^* = 0$, then Eq. 2.7 can be rewritten as

$$\sum_{i=1}^{n}\widehat{e}_i^*\left(y_i^{*\lambda}\ln y_i - \widehat{e}_i^*\right) = \sum_{i=1}^{n}\widehat{e}_i^*\left(f(\lambda) - \widehat{e}_i^*\right) = 0 \tag{2.11}$$

Ogwnag [8] suggested using the first-order Taylor expansion to approximate f(λ) around λ_0, to avoid tedious computation, which provides a unique solution that is as good as that of using higher order ones. Therefore, Eq. 2.11 can be rewritten and solved for λ as:

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} \widehat{e}_i^{*2} + \sum_{i=1}^{n} \widehat{e}_i^{*}[\lambda_0 f'(\lambda_0) - f(\lambda_0)]}{\sum_{i=1}^{n} \widehat{e}_i^{*} f'(\lambda_0)} \tag{2.12}$$

Based on the above estimation, we develop the following procedure, which is simple to implement for practitioners, to calculate the power of the Box-Cox transformation using the maximum likelihood estimator:

Step 1
: Choose a number as an initial guess λ_0 of λ for a given random sample.

Step 2
: Transform the original sample using Eq. 1.1 with parameter λ_0, then calculate $\widehat{\beta}^*$ and $\widehat{\sigma}^{*2}$ based on Eqs. 2.8 and 2.9, respectively.

Step 3
: Use Eq. 2.12 to solve for a current value λ_c of λ.

Step 4
: Check whether the difference between λ_c and λ_0 is less than a predetermined precision level. If not, reset λ_0 to λ_c and iterate between step 1 and step 3 until the difference between λ_c and λ_0 is smaller than the predetermined precision level.

Step 5
: Use the λ_c obtained in step 4 as the optimal value for $\widehat{\lambda}$. Apply the Shapiro-Wilk test to check the normality of the transformed sample.

3 Box-Cox transformation based on percentile estimator

Hinkley [5] proposed an alternative method, which does not require as much calculation as that required for using the maximum likelihood method. We note that the method discussed in Hinkley [5] is also sensitive to outliers. Assume that the sample $Y_1,\ldots,Y_n$ can be described by Eq. 2.1 with the common distribution function $F(y)$. The percentile η_p is defined as $F(\eta_p)=p(0<p<1)$. If there is a λ such that the transformed data $y^{(\lambda)}$ defined in Eqs. 1.1 and 2.1 follow a normal distribution, then the η_p and η_{1-p} percentiles will be symmetric to the median. This property suggests using the order statistics of the random sample, which is corresponding to the tail probabilities p and $1-p$ for some p. As pointed out by Hinkley [5], $y^{(\lambda)}$ is seldom perfectly symmetrical for most λ. But, we anticipate that there might be a value of λ making the transformed data nearly symmetric. Therefore, we look for the power transformation in Eq. 1.1 for which

$$\eta_{0.5}^{\lambda} - \eta_p^{\lambda} = \eta_{1-p}^{\lambda} - \eta_{0.5}^{\lambda} \tag{3.1}$$

Note that $Y_{(1)},\ldots,Y_{(n)}$ are the ordered values and $\widetilde{Y}$ is the median of $Y_1,\ldots,Y_n$. Clearly, except the trivial solution, $\widetilde{Y} = Y_{(r)} = Y_{(n-r+1)}$, other solutions of λ must be found by solving the following Eq. 3.2:

$$\widetilde{Y}^{\lambda} - Y_{(r)}^{\lambda} = Y_{(n-r+1)}^{\lambda} - \widetilde{Y}^{\lambda}, \quad where\, r = [np] \tag{3.2}$$

Equation 3.2 provides exact solutions for λ. Another solution to Eq. 3.2 is $\lambda=0$. This occurs when

$$\widetilde{Y} \Big/ Y_{(r)} = Y_{(n-r+1)} \Big/ \widetilde{Y} \tag{3.3}$$

which is the condition for sample percentiles of ln y symmetric about the median. If $\lambda \neq 0$, we could easily rewrite Eq. 3.2 as the form given below:

$$\left(Y_{(r)} \Big/ \widetilde{Y}\right)^{\lambda} + \left(Y_{(n-r+1)} \Big/ \widetilde{Y}\right)^{\lambda} = 2 \tag{3.4}$$

Hinkley [5] showed that there must exist one nonzero solution to Eq. 3.4, and suggested the use of multiple values of p, say $p_1<\ldots<p_m<1/2$, to obtain multiple equations of Eq. 3.4, and then summing them to obtain a reasonably efficient estimator of λ. The resulting equation is given below:

$$\sum_{j=1}^{m} c_j \left[\left(\frac{Y_{(r_j)}}{\widetilde{Y}}\right)^{\lambda} + \left(\frac{Y_{(n-r_j+1)}}{\widetilde{Y}}\right)^{\lambda}\right] = 2\sum_{j=1}^{m} c_j \tag{3.5}$$

where $r_j=[np_j]$ and $c_1,\ldots,c_m$ are arbitrary weights. Hinkley [5] showed that using equal weights on c_j, with $p_j \leq 0.05$ and $m=3$, which average out the asymmetry characteristic of $y^{(\lambda)}$, increases the precision of the transformation. We may solve Eq. 3.5 for a good approximated λ. However, if the following condition is met, then the solution $\widehat{\lambda} = 0$ will be chosen.

$$\sum_{j=1}^{m} c_j \ln\left(Y_{(r_j)} Y_{(n-r_j+1)}\right) = 2\sum_{j=1}^{m} c_j \ln \widetilde{Y} \tag{3.6}$$

Hinkley [6] proposed a similar method for choosing a symmetrizing transformation based on the asymmetry degree of the sample, which is measured by

$$d = (sample\ mean - sample\ median) \big/ sample\ scale. \tag{3.7}$$

If the underlying distribution is symmetric, then the mean and the median must be identical. Thus, the sample data drawn from such distribution should reflect such property, and a good estimate of λ should minimize the value of d.

Base on the above argument with setting λ to $-2 \leq \lambda \leq 2$ (as recommended by Tukey [13]), a step-by-step procedure for computing the power of Box-Cox transformation based on MOP may be presented as follows:

Step 1
: Choose –2 as an initial guess λ_0 of λ for a given random sample.

Step 2
: Transform the original sample by taking the power λ_0 and then find the sample mean, $\overline{Y}^{(\lambda)}$, sample median, $\widetilde{Y}^{(\lambda)}$, and sample inter-quartile range, r, for the transformed random sample, $Y_1^{(\lambda)}, Y_2^{(\lambda)}, \cdots, Y_n^{(\lambda)}$.

Step 3
: Calculate d defined in Eq. 3.7 using the inter-quartile range as the sample scale.

Step 4
: Check whether d is less than a predetermined precision level. If not, iterate steps 1–3 by increasing the magnitude of λ by 0.05 as new λ_0, till the difference between λ_0 and λ_c is smaller than the predetermined precision level.

Step 5
: Use the λ derived from step 4 as the optimal estimated $\widehat{\lambda}$. Employ the Shapiro-Wilk test to check the normality of the transformed sample.

4 Implementation and application

To illustrate the Box-Cox transformation using the two estimators, we consider the data given in Table 1 collected from a forging manufacturing process making a specific type of piston rings for automotive engines. Before starting data analysis, the normality assumption must be checked. Figure 1 gives a standard normal plot and the standardized original data, which does not appear to be normal. The original data is then transformed to one that is likely to be normal so that data analysis can be performed.

After estimating λ by MOP and MLE, the original data is then transformed using Eq. 1.1 with powers of 0.25 and 0, respectively. The transformed data of the piston rings for automotive engines, using the two estimators, MLE and MOP, are given in Tables 2 and 3. In Fig. 2, the curve marked by crosses is the normal plot for the transformed data using MOP, and the curve marked by triangles is one for the transformed data using MLE. The near normality for both sets of the transformed data indicates that the transformation is effective; the normal assumption required

Table 1 Thirty piston ring diameters for automotive engines

0.32	0.47	0.52	0.59	0.77
0.81	0.81	0.90	0.96	1.18
1.20	1.20	1.31	1.35	1.43
1.51	1.62	1.74	1.87	1.89
1.95	2.05	2.10	2.20	2.48
2.81	3.00	3.09	3.37	4.75

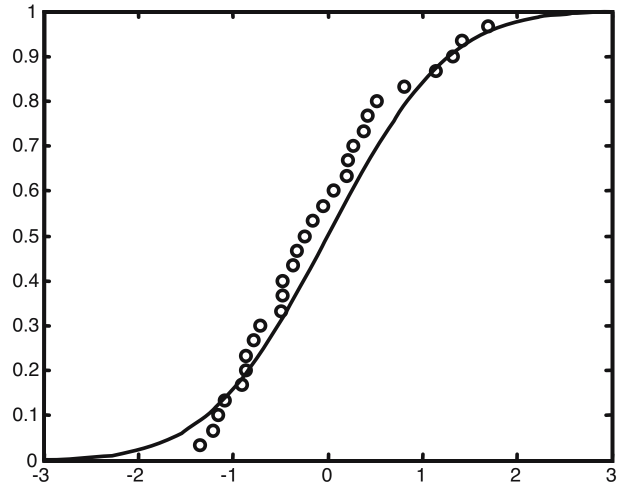


Fig. 1 Normal plot of standardized original data indicated by *circles*

for applying the statistical method may be satisfactory, and the manufacturing quality analysis could be performed.

5 Comparison of the transformation power

To compare the transformation power using the two estimators (MLE and MOP), a set of distributions widely applied to engineering applications modelling different process characteristics are selected. Those distributions can be grouped into three categories: (1) negatively skewed, and the beta (5, 1) distribution; (2) positively skewed, including the gamma distributions, F distributions, and the lognormal distributions; and (3) the symmetrical distributions, including the beta (0.5, 0.5) distribution, uniform distributions with flat kurtosis, and the T distributions with sharp kurtosis. To obtain useful information, 1,000 samples of size 5, 10, 25, 50, 100, and 500 are generated for each distribution in the simulation. For each sample, we use the optimal $\widehat{\lambda}$ obtained using the two estimators for the

Table 2 The thirty diameters of the piston rings transformed by MOP

2.0085	2.3120	2.3967	2.5057	2.7470
2.7947	2.7947	2.8960	2.9594	3.1690
3.1865	3.1865	3.2794	3.3116	3.3742
3.4341	3.5127	3.5941	3.6776	3.6900
3.7268	3.7863	3.8152	3.8715	4.0196
4.1789	4.2643	4.3033	4.4196	4.9052

Table 3 The thirty diameters of the piston rings transformed by MLE

–1.1394	–0.7550	–0.6539	–0.5276	–0.2614
–0.2107	–0.2107	–0.1054	–0.0408	0.1655
0.1823	0.1823	0.2700	0.3001	0.3577
0.4121	0.4824	0.5539	0.6259	0.6366
0.6678	0.7178	0.7419	0.7885	0.9083
1.0332	1.0986	1.1282	1.2149	1.5581

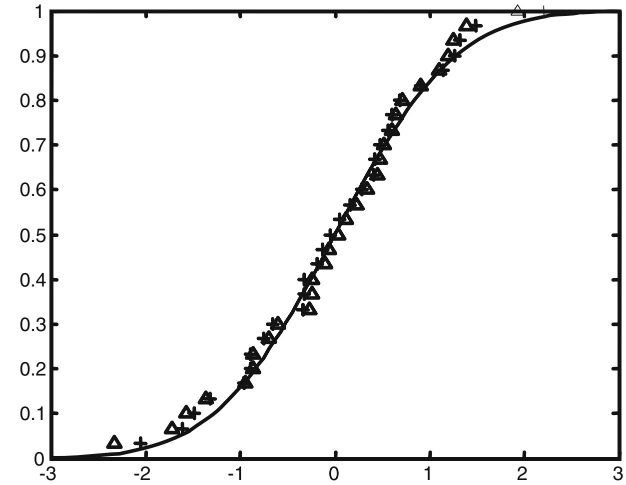


Fig. 2 Normal plots of standardized transformed data indicated by *crosses* for MOP and *triangles* for MLE

transformation. To test the normality of the transformed data, the Shapiro-Wilk test is employed. Madansky [7] showed that the Shapiro-Wilk test, in general, is more powerful than other goodness-of-fit tests. Sample means and variances of the estimated optimal powers, $\widehat{\lambda}$, are calculated, and the pass rates of the transformed data using the Shapiro-Wilk test are then computed. A flow chart illustrating the simulation procedure is presented in Fig. 3.

5.1 Negatively skewed distributions

We first consider the beta distributions with heavy left tails. Suppose a random sample of size 100 is generated from beta (10, 2) (Fig. 4), with the histograms of the transformed samples by MOP and MLE in Figs. 5 and 6, respectively. In correcting the asymmetry, MOP seems to be more effective than the MLE. Figure 7 gives the pass rates of beta (5, 1) for various sample sizes. The MOP outperforms MLE in

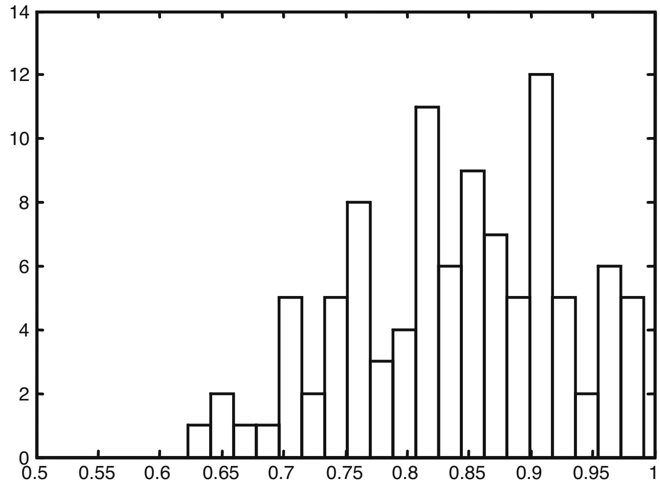


Fig. 4 A negatively-skewed random sample of size 100 drawn from beta (10, 2)

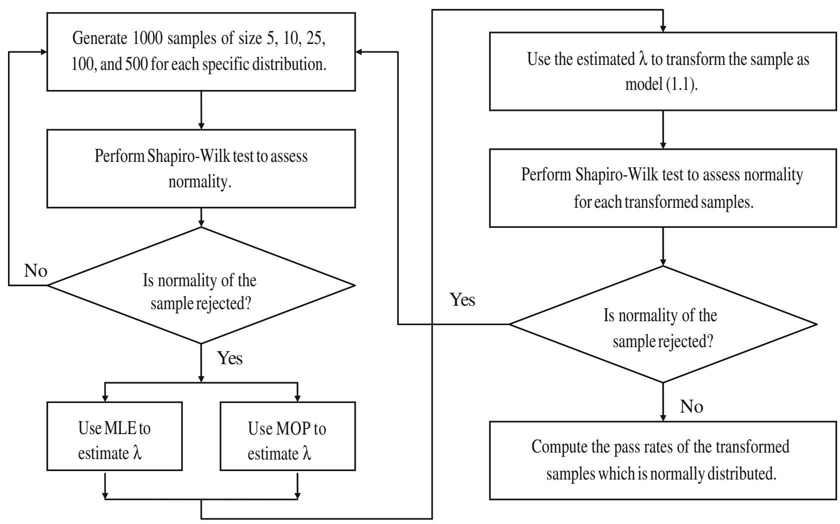
pass rates except for n=5, 500. In Fig. 8, we note that the transformation is rather effective using the MOP with powers between 1.5 and 2. While using the MLE results in much the same value of $\widehat{\lambda}$ near zero, the variation of $\widehat{\lambda}$ using the MOP is significantly greater than that of using the MLE, as seen in Fig. 9.

5.2 Positively skewed distributions

Next, we consider the positively skewed distributions including the gamma family, F, some beta distributions, and lognormal distributions. A random sample drawn from beta (2, 10) is presented to show the transformation in Figs. 10, 11, 12. Both MOP and MLE seem to be effective in centralization and lightening tails.

Among these positively skewed distributions considered, exponential (1), Weibull (5, 1), and χ^2 (1) are highly positively skewed. According to Fig. 13, MOP has higher pass rates than MLE in most cases for different sample

Fig. 3 Flow chart for simulation procedure



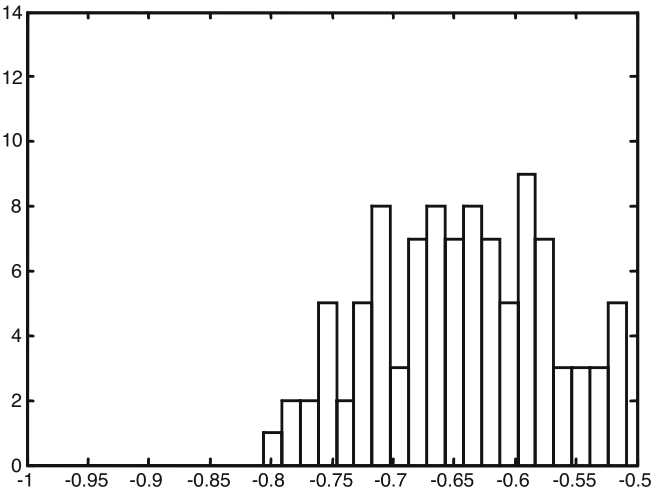


Fig. 5 Histogram of the transformed sample by MOP, $\widehat{\lambda} = 2$

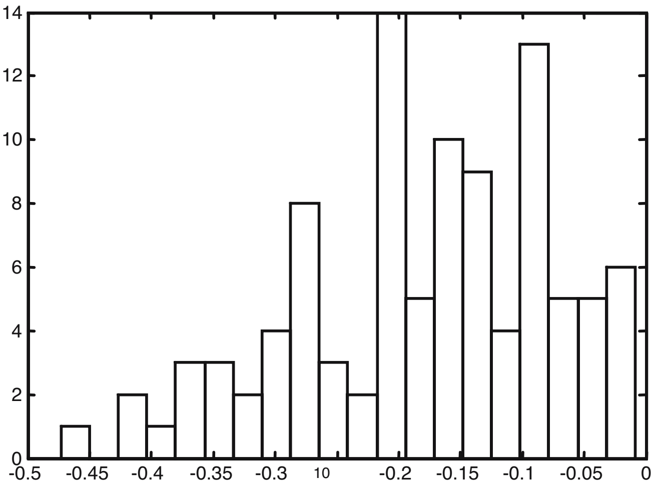


Fig. 6 Histogram of the transformed sample by MLE, $\widehat{\lambda} = 0$

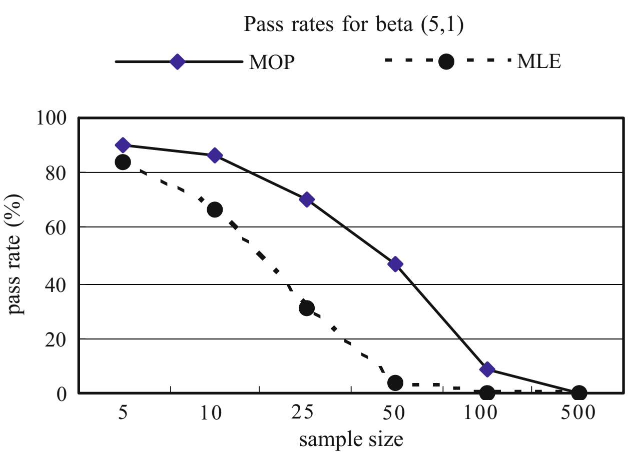


Fig. 7 Pass rates for beta (5, 1) random sample

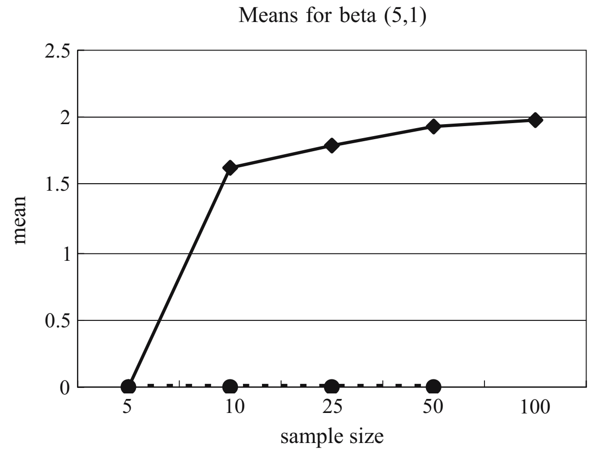


Fig. 8 Means for $\widehat{\lambda}$ values of beta (5, 1) random sample

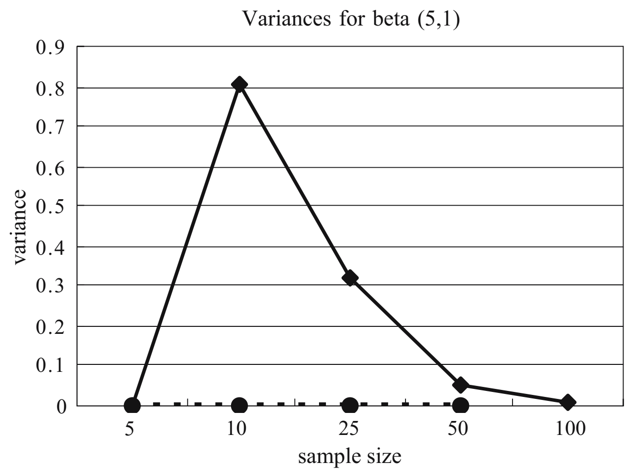


Fig. 9 Variances for $\widehat{\lambda}$ values of beta (5, 1) random sample

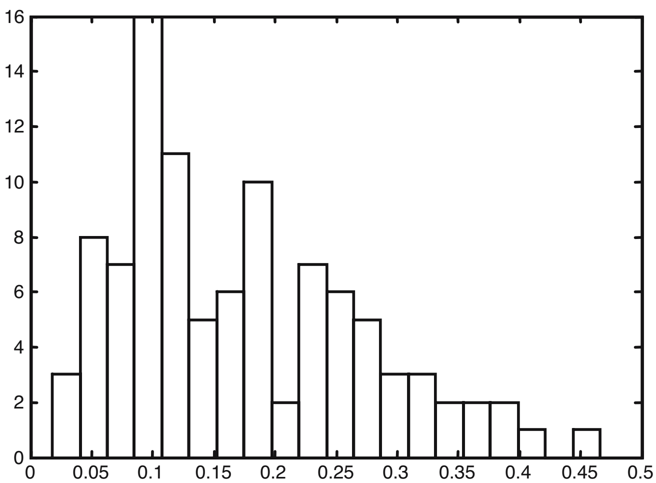


Fig. 10 A positive-skewed random sample of size 100 drawn from beta (2,10)

sizes. MOP provides nearly constant pass rates. For MLE the larger the sample size, the smaller the pass rate. While n is equal to 5 or 10 as shown in Fig. 14, the best choice of $\widehat{\lambda}$ is zero no matter which method was employed. MLE suggests the use of square root as the power of the transformation for exponential and chi-square distribu-

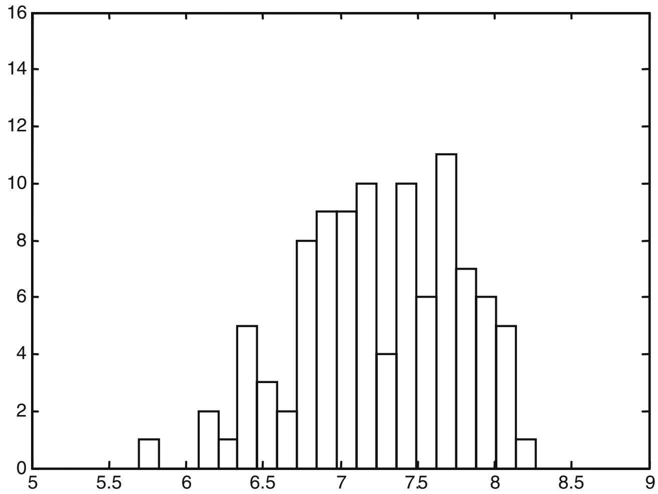


Fig. 11 The sample transformed from Fig. 10 by MOP with $\widehat{\lambda} = 0.1$

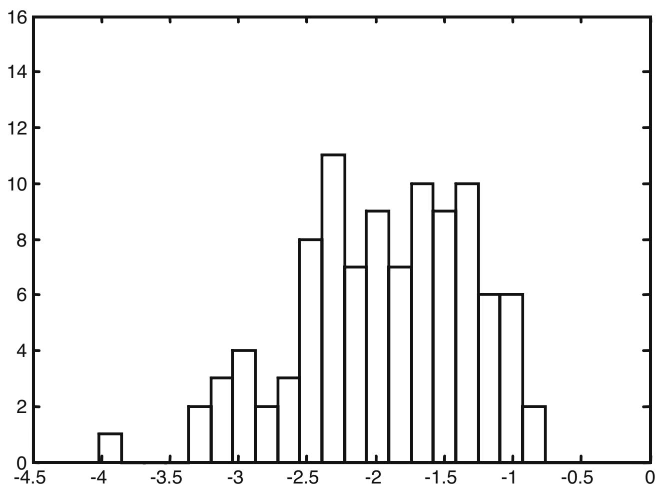


Fig. 12 The sample transformed from Fig. 10 by MLE with $\widehat{\lambda} = 0$

tions, and zero for the Weilbull distribution; whereas MOP advises the values of $\widehat{\lambda}$ as 0.25 and 1.25, respectively. The variances of estimated $\widehat{\lambda}$ by both methods are consistent for most distributions with the exception of the Weilbull distribution due to its extreme skewness (Fig. 15).

For the F distributions, the pass rates of normality tests are shown in Fig. 16, where it seems that there is a uniform pass rate except when the sample size is as large as 500. No matter what method was employed, zero is the best choice for λ as the power parameter when the sample size is as small as five. Figure 17 also gives the estimated values of $\widehat{\lambda}$ to transform the random samples drawn from F distributions into normal ones. MOP suggests negative values of powers for mildly skewed F distributions but MLE gives positive ones. Once again, MOP is more inefficient because of the larger variances of $\widehat{\lambda}$ than MLE as shown in Fig. 18.

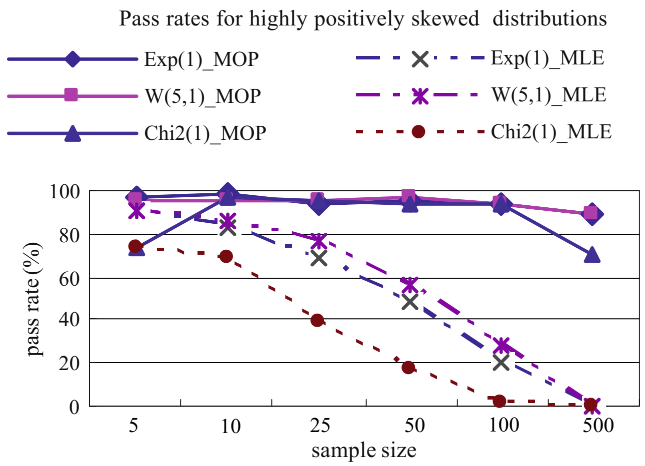


Fig. 13 Pass rates for highly positively skewed random samples

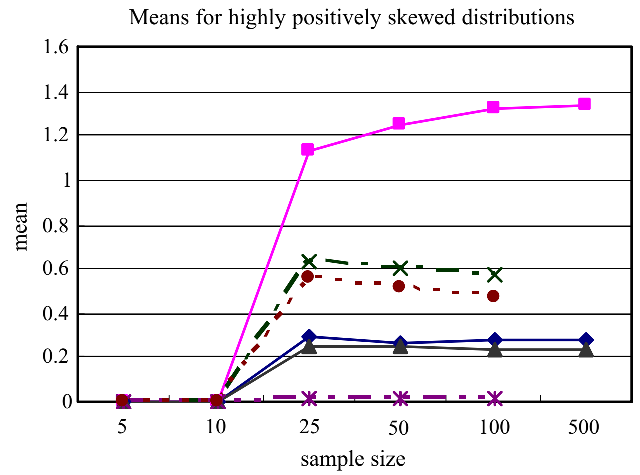


Fig. 14 Means for $\widehat{\lambda}$ values of highly positively skewed random samples

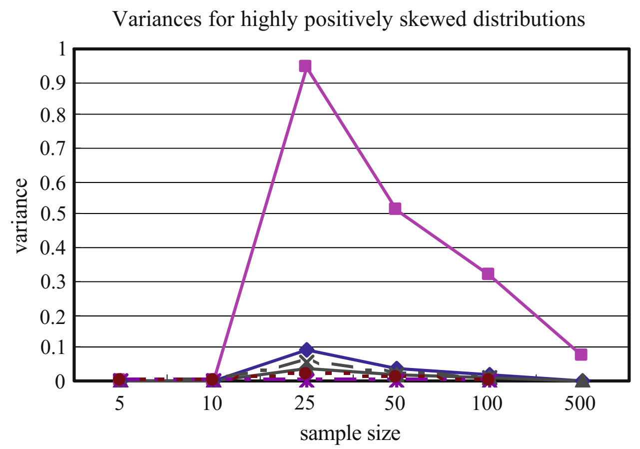


Fig. 15 Variances for $\widehat{\lambda}$ values of highly positively skewed random samples

For lognormal distributions, both methods produce much the same results for the lognormal distributions investigated here, as seen in Figs. 19, 20, 21.

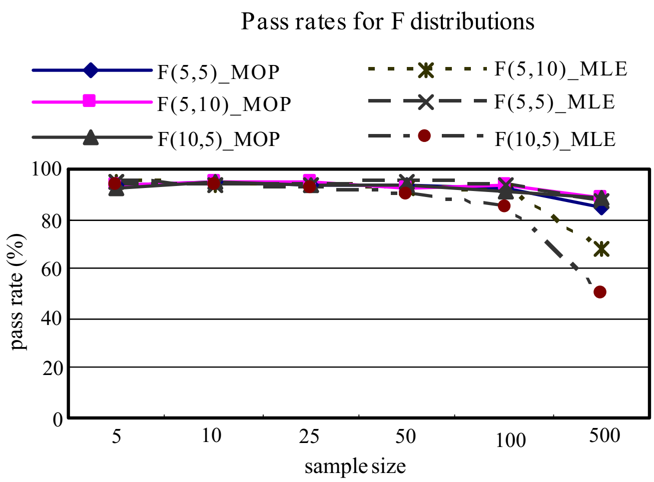


Fig. 16 Pass rates for random samples of F distribution

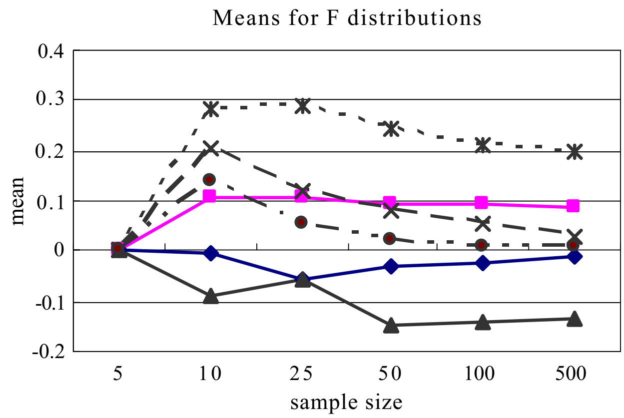


Fig. 17 Means for $\widehat{\lambda}$ values of random samples drawn from F distributions

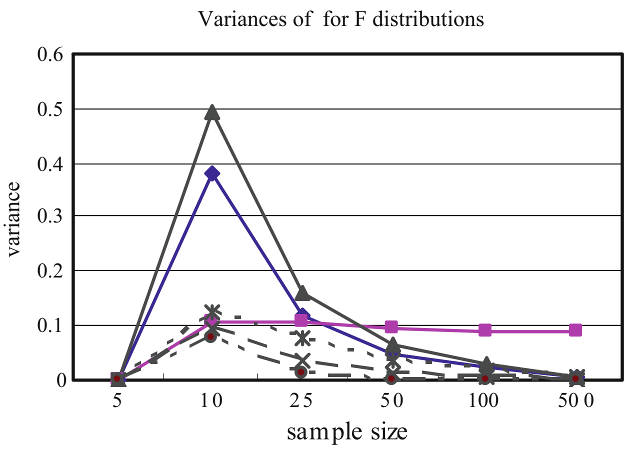


Fig. 18 Variances for $\widehat{\lambda}$ values of random samples drawn from F distributions

The exceptional case is the lognormal distribution with small-scale parameter, 0.1. For near symmetrical lognormal distributions, MLE is superior to MOP in pass rates of transformation. Figures 22, 23, 24 summarize the simulation results for gamma, chi-square, and Rayleigh distributions. For chi-square distributions, the larger the degree of freedom, the higher the pass rates, especially while the sample size is very large. MOP surpasses MLE in pass rates more as sample sizes increase. Zero is the only choice of transformation power estimated by MLE, while the preponderance of $\widehat{\lambda}$ values obtained by MOP indicate transformations using powers between 0.2 and 0.3. The trends of Rayleigh and gamma distributions are the same as chi-square distributions.

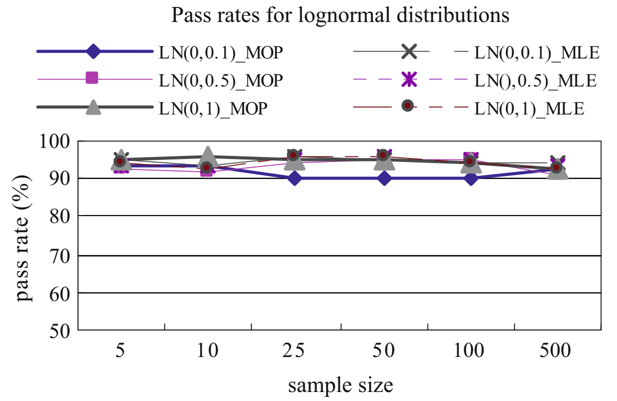


Fig. 19 Pass rates for random samples of lognormal distributions

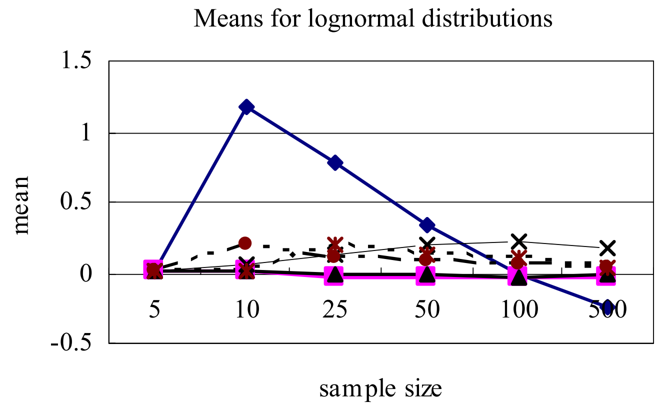


Fig. 20 Means for $\widehat{\lambda}$ values of random samples drawn from lognormal distributions

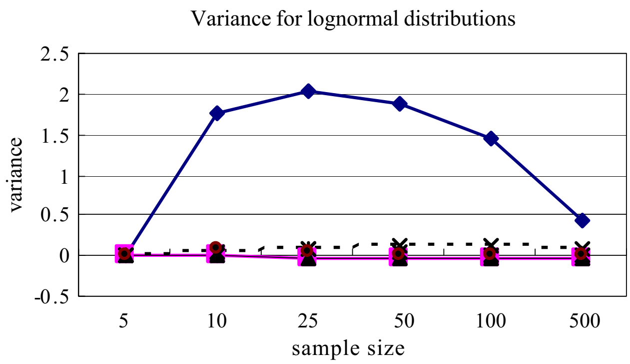


Fig. 21 Variances for $\widehat{\lambda}$ values of random samples drawn from lognormal distributions

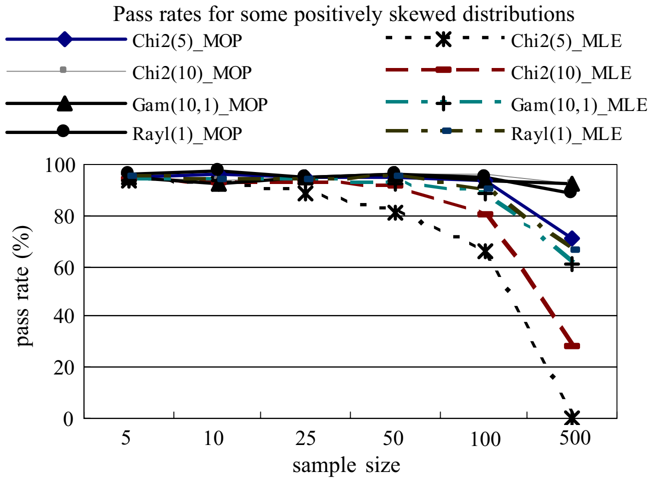


Fig. 22 Pass rates for random samples of positively skewed distributions

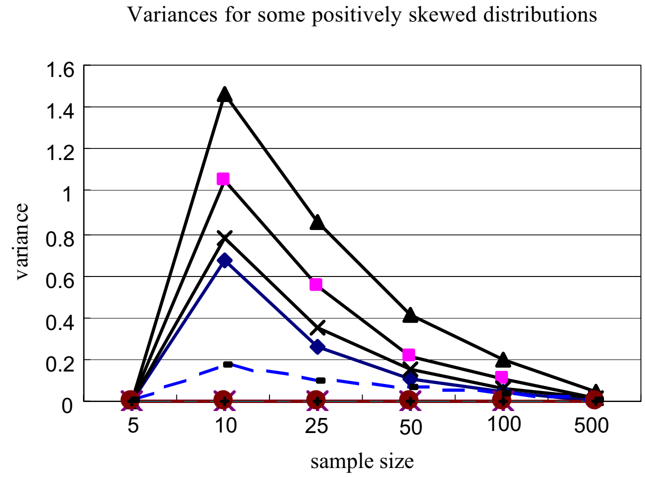


Fig. 24 Variances for $\hat{\lambda}$ values of random samples drawn from positively skewed distributions

5.3 Symmetric distributions

Four different symmetric distributions are included in our investigation. We consider the beta distribution with parameters (0.5, 0.5) as a symmetric one with concave heavy tails. In contrast to the beta (0.5, 0.5) distribution, the T distribution may represent symmetric leptokurtic ones. In addtion, uniform distributions with constant symmetry and Rayleigh distributions with parameters 10 and 20, which are near symmetrical, are considered. Figures 25, 26, 27 demonstrate the transformation for platykurtic beta (0.5, 0.5), while Figs. 28, 29, 30 depict leptokurtic beta (40, 40).

Since the distribution shown in Fig. 25 appears to have a very heavy tail, neither MOP nor MLE perform well in normal transformation. Nevertheless, not only MOP but also MLE works well for leptokurtic beta (40, 40) as shown in Figs. 28, 29, 30.

For beta (0.5, 0.5) and uniform distributions, Figs. 31 and 34 show the same pattern due to samples drawn from heavy tailed distributions. The pass rates of MOP are greater than MLE between moderate sample sizes from n=10 to n=100. The pass rates decrease as sample size increases no matter what method is applied. Figures 32 and 35 depict that the suggested transformation power for uniform and beta distributions of all sample sizes by MLE is zero, which is the same as MOP when the sample size is no greater than ten. If using MOP, $\hat{\lambda}$ should be greater than 0.8 and less than 1 when the sample size is greater than or equal to 25. Referring to Figs. 33 and 36, the variances decrease as sample size increases from 25, because the large sample size can increase the precision of the estimation.

For leptokurtic distribution, 1,000 samples were generated from each T distribution with different degrees of freedom. MOP and MLE seem to perform very equally in pass rates. As can be seen from Fig. 37, higher kurtosis of the distribution results in lower pass rate. The estimated values of λ obtained using MLE (referring to Fig. 38) are zero, while the estimated powers by MOP are between 1.6 and 1.8 for median sample sizes, but drop steeply to almost square-root as sample size goes up to 500. Unlike platy distributions, as shown in Fig. 39, the variances increase with the increase of sample size.

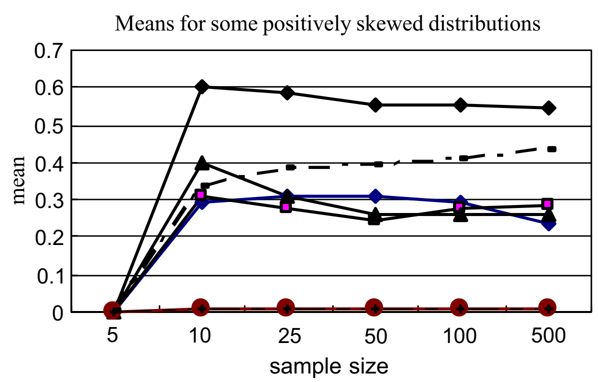


Fig. 23 Means for $\hat{\lambda}$ values of random samples drawn from positively skewed distributions

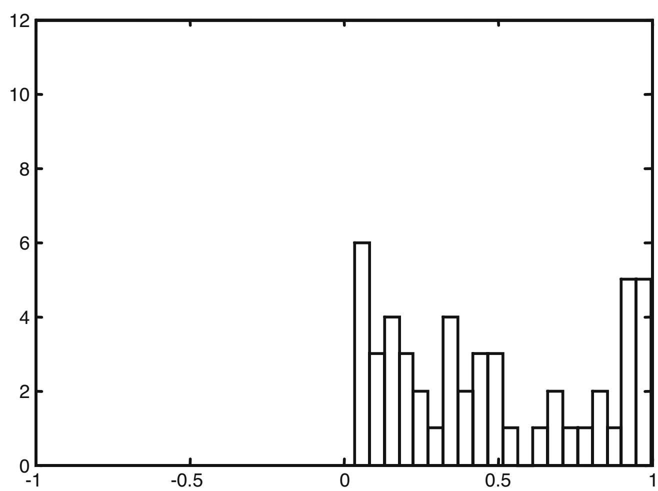


Fig. 25 A platykurtic and symmetric random sample with heavy tails of size 100 drawn from beta (0.5, 0.5)

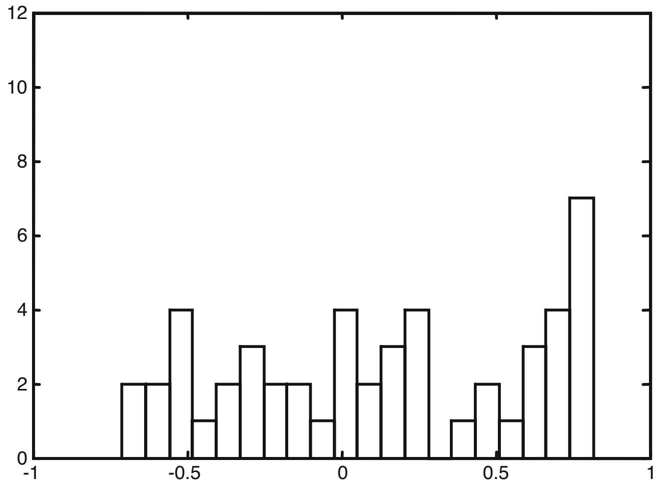


Fig. 26 The sample transformed from Fig. 25 by MOP with $\widehat{\lambda} = 0.55$

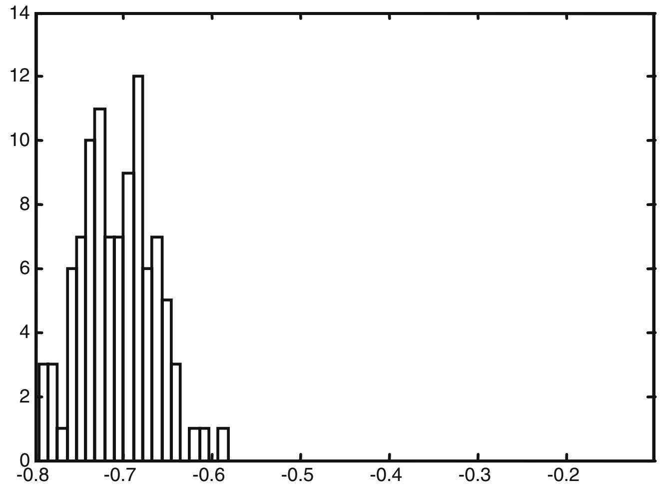


Fig. 29 The sample transformed from Fig. 28 by MOP with $\widehat{\lambda} = 1.35$

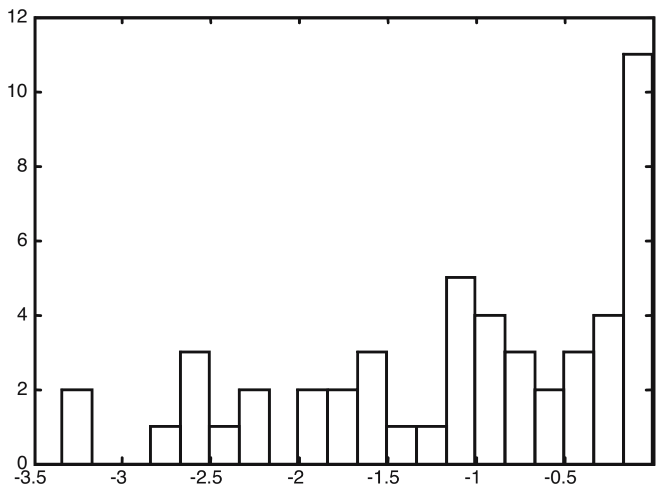


Fig. 27 The sample transformed from Fig. 25 by MLE with $\widehat{\lambda} = 0$

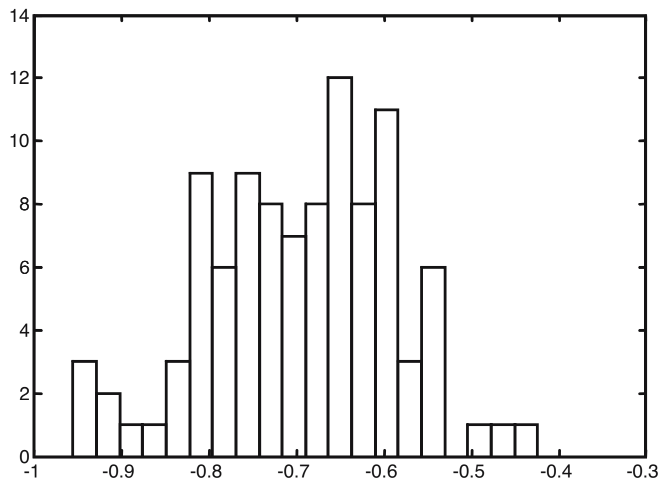


Fig. 30 The sample transformed from Fig. 28 by MLE with $\widehat{\lambda} = 0$

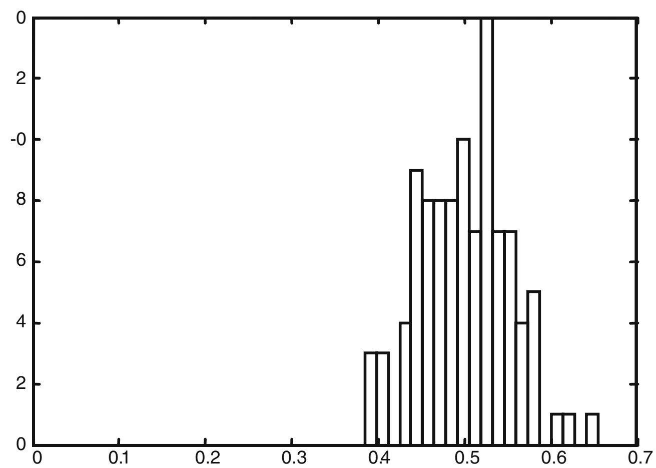


Fig. 28 A leptokurtic random sample of size 100 drawn from beta (40, 40)

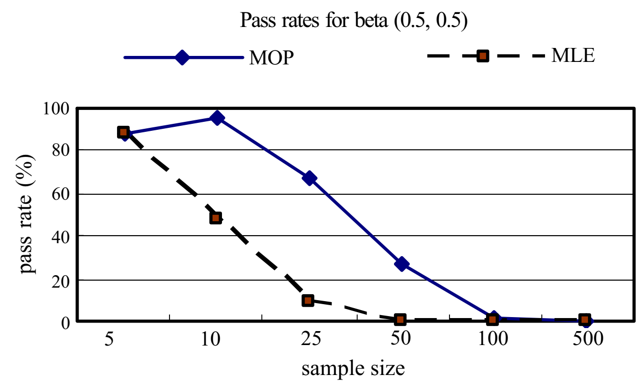


Fig. 31 Pass rates for random samples of beta (0.5, 0.5)

Since the Rayleigh distributions considered here are very near symmetric, their behaviors are similar to T distributions in pass rates and means of the estimated powers, as shown in Figs. 40 and 41. Only variances do not follow the pattern of T distributions as shown in Fig. 42.

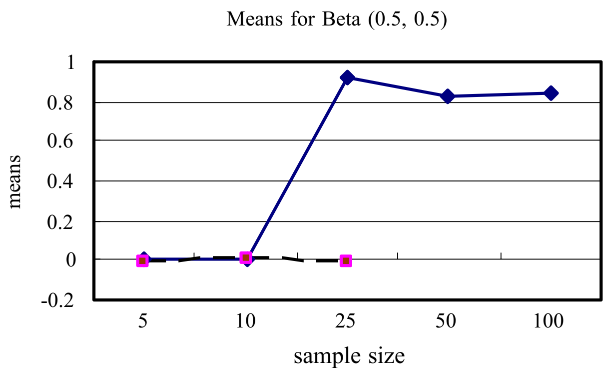


Fig. 32 Means for $\hat{\lambda}$ values of random samples drawn from beta (0.5, 0.5)

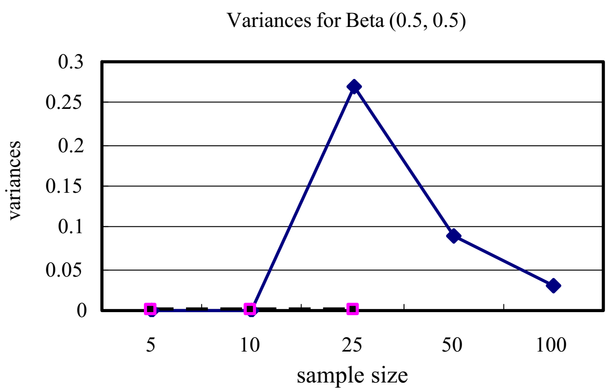


Fig. 33 Variances for $\hat{\lambda}$ values of random samples drawn from beta (0.5, 0.5)

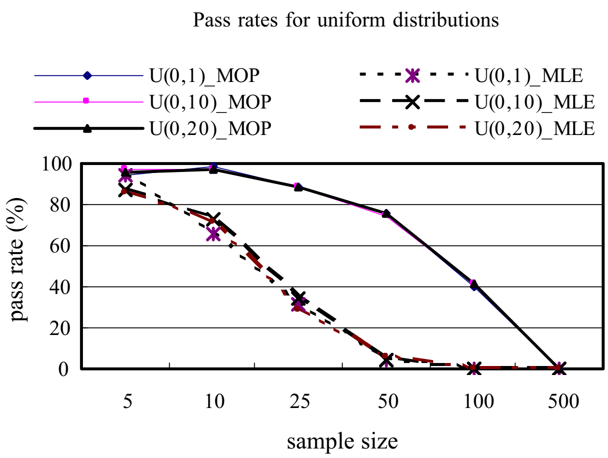


Fig. 34 Pass rates for random samples of uniform distributions

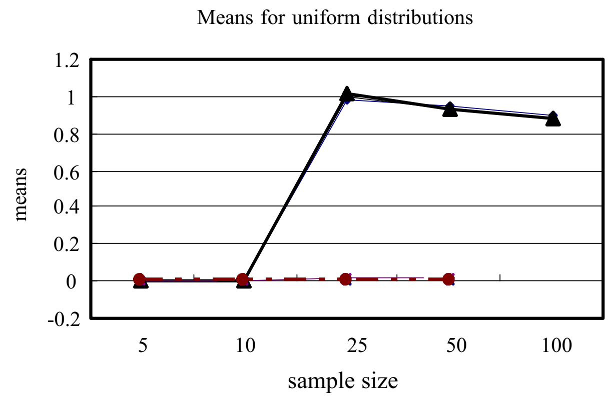


Fig. 35 Means for $\hat{\lambda}$ values of random samples drawn from uniform distributions

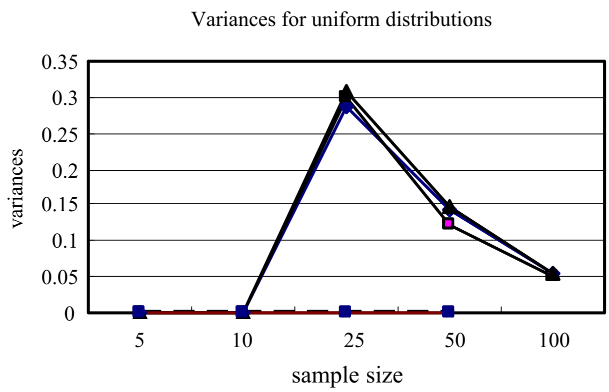


Fig. 36 Variances for $\hat{\lambda}$ values of random samples drawn from uniform distributions

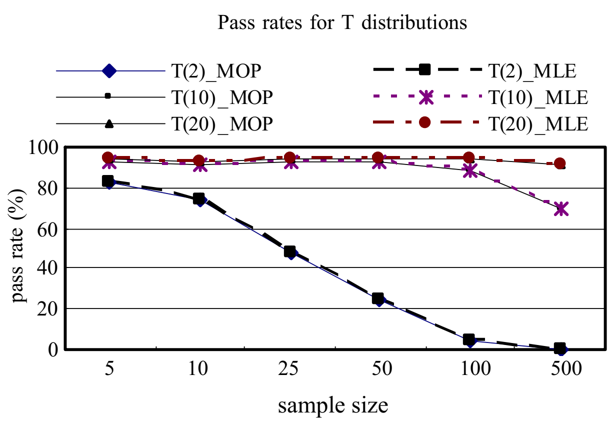


Fig. 37 Pass rates for random samples of T distributions

6 Conclusions and recommendations

In the practice of manufacturing quality control and operations management, many statistical methods applications require that the process characteristic of interest is normally distributed. If the data collected is not normally distributed, using normal-base techniques may result in incorrect conclusions. For such situations, the most popular approach is to transform the non-normal data into a normal one. In this paper, we consider the Box-Cox transformation, and compare the transformation power using two different parameter estimation methods, including the maximum likelihood estimator (MLE) and the method of percentiles (MOP). The performance comparison is based on the pass rate under the Shapiro-Wilk normality test. We note that small and median samples generally result in higher pass rates than those of large samples. The results also show that in general the MOP has better pass rate, while the MLE has smaller power variation for most cases investigated. For small sample size (n=5, 10), both

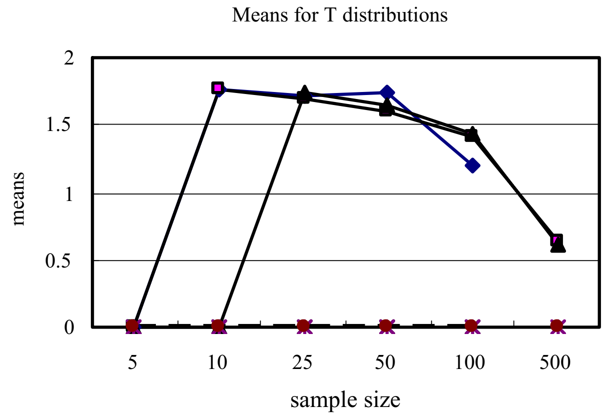


Fig. 38 Means for $\widehat{\lambda}$ values of random samples drawn from T distributions

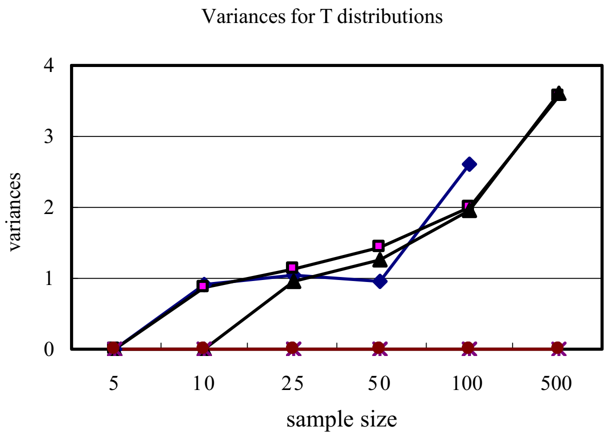


Fig. 39 Variances for $\widehat{\lambda}$ values of random samples drawn from T distributions

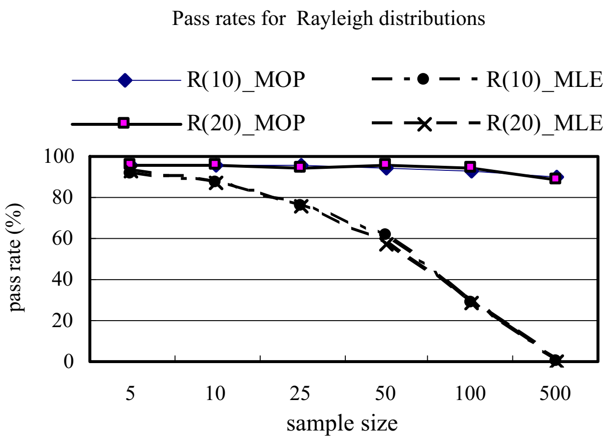


Fig. 40 Pass rates for random samples of Rayleigh distributions

methods perform equally well. For large sample size, the MOP is recommended due to its simplicity and significantly higher pass rate.

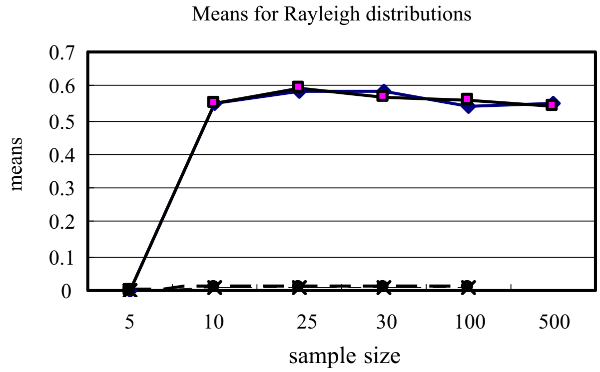


Fig. 41 Means for $\widehat{\lambda}$ values of random samples drawn from Rayleigh distributions

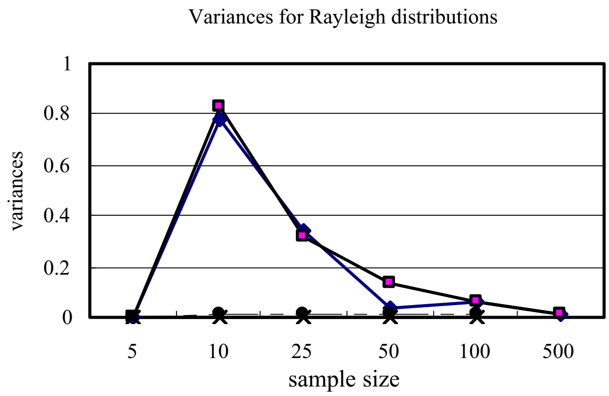


Fig. 42 Variances for $\widehat{\lambda}$ values of random samples drawn from Rayleigh distributions

References

1. Box GEP, Cox DR (1964) An analysis of transformations. J Roy Stat Soc B 26:211–252
2. Draper NR, Cox DR (1969) On distributions and their transformations to normality. J Roy Stat Soc B 31:472–476
3. Han AK (1987) A non-parametric analysis of transformation. J Econometrics 35:191–209
4. Hernandes F, Johnson RA (1980) The large sample behavior of transformation to normality. J Am Stat Assoc 75:855–861
5. Hinkley DV (1975) On power transformations to symmetry. Biometrika 62(1):101–111
6. Hinkley DV (1976) On quick choice of power transformation. Appl Stat 26(1):67–69
7. Madansky A (1988) Prescriptions for working statisticians. Springer, New York
8. Ogwang T, Gouranga Rao UL (1997) A simple algorithm for estimating Box-Cox models. Statistician 46(3):399–409
9. Pyzdek T (1995) Why normal distributions aren't [all that normal]. Qual Eng 7(4):767–777
10. Schlesselman J (1971) Power families. A note on the Box-Cox transformation. J Roy Stat Soc B 33:307–371
11. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality. Biometrika 52(3–4):591–611
12. Taylor JMG (1985) Power transformations to symmetry. Biometrika 72(1):145–152
13. Tukey JW (1957) The comparative anatomy of transformations. Ann Math Stat 28:602–632