

Full Paper

Bootstrap confidence intervals for the process capability index under half-logistic distribution

Wararit Panichkitkosolkul^{1,*} and Luckhana Saothayanun²

¹ Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Phatum Thani, 12121, Thailand

² Department of Applied Statistics, Faculty of Science, University of the Thai Chamber of Commerce, Bangkok, 10400, Thailand

* Corresponding author, e-mail: wararit@mathstat.sci.tu.ac.th

Received: 21 August 2011 / Accepted: 26 July 2012 / Published: 27 July 2012

Abstract: This study concerns the construction of bootstrap confidence intervals for the process capability index in the case of half-logistic distribution. The bootstrap confidence intervals applied consist of standard bootstrap confidence interval, percentile bootstrap confidence interval and bias-corrected percentile bootstrap confidence interval. Using Monte Carlo simulations, the estimated coverage probabilities and average widths of bootstrap confidence intervals are compared, with results showing that the estimated coverage probabilities of the standard bootstrap confidence interval get closer to the nominal confidence level than those of the other bootstrap confidence intervals for all situations.

Keywords: bootstrap confidence interval, process capability index, half-logistic distribution.

INTRODUCTION

Balakrishnan [1] introduced the half-logistic distribution as the distribution of the absolute logistic random variable – that is, if Y is a logistic random variable, then $X = |Y|$ has a half-logistic distribution. The probability density function ($f(x)$) and the cumulative distribution function ($F(x)$) are given by

$$f(x) = \frac{2 \exp\{-(x - \mu)/\sigma\}}{\sigma [1 + \exp\{-(x - \mu)/\sigma\}]^2}, \quad (1)$$

and

$$F(x) = \frac{1 - \exp\{-(x - \mu)/\sigma\}}{[1 + \exp\{-(x - \mu)/\sigma\}]}, \quad x > \mu, \quad \sigma > 0, \quad (2)$$

where μ and σ are the location and the scale parameters respectively. The graph of the probability density function for half-logistic distribution is shown in Figure 1. The mean and the variance of X are defined as

$$E(X) = \mu + \sigma \ln(4) \quad \text{and} \quad Var(X) = \sigma^2 \left[\frac{\pi^2}{3} - (\ln(4))^2 \right].$$

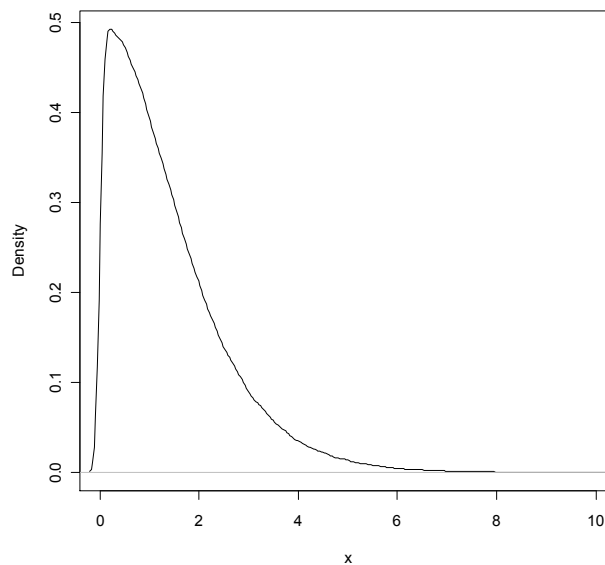


Figure 1. Probability density function for half-logistic distribution with $\mu = 0$ and $\sigma = 1$

The half-logistic distribution has been widely used for many applications. For example, Balakrishnan [1] has suggested the usage of this distribution as a possible lifetime model with an increasing hazard rate. In addition, Balakrishnan and Chan [2] have shown that the failure times of air conditioning equipment in a Boeing 720 airplane fit the half-logistic distribution quite well. This distribution was also applied to environmental and sports records data [3]. In recent papers, several authors have applied the half-logistic distribution under progressive Type-II censoring [4-6]. As mentioned above, it is known that the half-logistic distribution is an increasing failure rates model with considerable importance in quality control and reliability studies [7-9].

In product quality control, the process capability index (PCI) has been widely adopted as a useful tool. Several process capability indices have been proposed to numerically measure whether a process is capable of manufacturing products that meet customer requirements or specifications [10]. Even though there are many process capability indices, the two most commonly used indices are C_p and C_{pk} [11-12]. The more popular process capability index C_{pk} can be defined as follows [11]:

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

where USL and LSL denote the upper and lower specification limits of the process respectively, σ is the process standard deviation, and μ is the process mean. As the process standard deviation and the process mean are unknown, they must be estimated from the sample data $\{X_1, \dots, X_n\}$. The sample mean \bar{X} ; $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ and the sample standard deviation S ; $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$ are used to estimate the unknown parameters μ and σ respectively in Eq.(3). The estimator of the process capability index C_{pk} therefore is

$$\tilde{C}_{pk} = \min \left\{ \frac{USL - \bar{X}}{3S}, \frac{\bar{X} - LSL}{3S} \right\}.$$

However, the underlying process distribution is non-normal in some situations. Hence it may be a skewed distribution. To deal with these situations, Clements [13] proposed a new method for computing the estimator of the process capability index C_{pk} when the process distribution is non-normal. This estimator is defined as

$$\tilde{C}_{pk} = \min \left\{ \frac{USL - M}{U_p - M}, \frac{M - LSL}{M - L_p} \right\}, \quad (4)$$

where U_p, L_p and M denote the 99.865th, 0.135th and 50th percentiles of the distribution respectively. The advantage of \tilde{C}_{pk} shown in Eq.(4) is that it can be applied to any distribution. Kantam et al. [8] discussed the relationship between \tilde{C}_{pk} and the probability of a product falling outside the specification limits. When X has a half-logistic distribution, this probability is given by

$$P_i = P(X \leq LSL) + P(X \geq USL) = 1 + F(LSL) - F(USL),$$

where $F(\cdot)$ is the cumulative distribution function of a half-logistic distribution shown in Eq.(2). In the case of standard half-logistic distribution, i.e. $\mu = 0$, $\sigma = 1$ in Eq.(1), the values of U_p, L_p and M are 7.300122639, 0.002700002 and $\ln(3) \approx 1.098612289$ respectively. On the other hand, if a scale parameter σ is introduced and known, i.e. $\mu = 0$, $\sigma \neq 1$ in Eq.(1), the optimal estimator of C_{pk} is given by [8]:

$$\tilde{C}'_{pk} = \min \left\{ \frac{USL - \sigma M}{\sigma(U_p - M)}, \frac{\sigma M - LSL}{\sigma(M - L_p)} \right\}.$$

In practice, the scale parameter σ is unknown. Therefore, the unknown σ must be estimated by its estimator. In this paper the method of moments is used for calculating this estimator. The estimator of C_{pk} for a half-logistic distribution therefore is

$$\hat{C}_{pk} = \min \left\{ \frac{USL - \hat{\sigma} M}{\hat{\sigma}(U_p - M)}, \frac{\hat{\sigma} M - LSL}{\hat{\sigma}(M - L_p)} \right\}. \quad (5)$$

where $\hat{\sigma}$ is the method of moments estimator of σ given by $\hat{\sigma} = \bar{X} / \ln(4)$. The properties of the estimator of C_{pk} with mean squared error (MSE) and absolute of bias (|Bias|) are considered. Using Monte Carlo simulations, the MSE and |Bias| are plotted in Figure 2. If σ is fixed and $n \rightarrow \infty$,

$MSE(\hat{C}_{pk}) \rightarrow 0$ and $|Bias(\hat{C}_{pk})| \rightarrow 0$. Therefore, the estimator of C_{pk} given in Eq.(5) is an approximate estimator in terms of MSE and $|Bias|$.

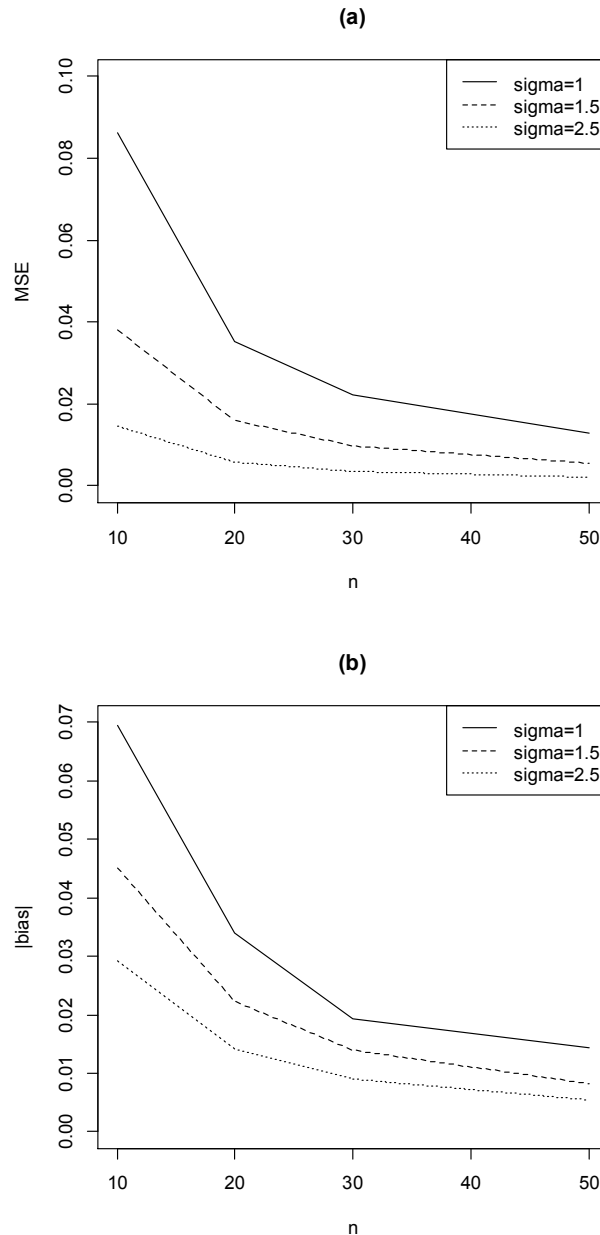


Figure 2. (a) MSE of the estimator of C_{pk} and (b) $|Bias|$ of the estimator of C_{pk} when $\mu = 0$ and $\sigma = 1, 1.5, 2.5$

BOOTSTRAP CONFIDENCE INTERVALS

The bootstrap is a computer-based and resampling method for assigning measures of accuracy to statistical estimates [14]. The advantage of the bootstrap method is that it is a simple approach for estimating biases, standard errors, confidence intervals and so forth for complicated estimators. Furthermore, the distribution of the variable of interest is not mathematically estimated, but rather empirically developed on the characteristics of the distribution of original data [15]. There are many types of bootstrap methods for constructing confidence intervals that have been

introduced, e.g. the standard bootstrap method (SB), the percentile bootstrap method (PB) and the bias-corrected percentile bootstrap method (BCPB) [14].

For a sequence of independent and identically distributed random variables, the bootstrap procedure can be defined as follows [14]. Let the random variables $\{X_{n,j}^*, 1 \leq j \leq m\}$ be the results from sampling m times with replacement from n observations X_1, \dots, X_n . The random variables $\{X_{n,j}^*, 1 \leq j \leq m\}$ are called the bootstrap samples from the original data X_1, \dots, X_n . The construction of confidence intervals of the process capability index C_{pk} using bootstrap techniques are described in what follows.

Standard Bootstrap (SB) Confidence Interval

Let X_b^* , where $1 \leq b \leq B$, be the b^{th} bootstrap sample and let X_1^*, \dots, X_B^* be the B bootstrap samples. The b^{th} bootstrap estimator of C_{pk} is computed by [14]:

$$\hat{C}_{pk}^{*(b)} = \min \left\{ \frac{USL - \hat{\sigma}^{*(b)}M}{\hat{\sigma}^{*(b)}(U_p - M)}, \frac{\hat{\sigma}^{*(b)}M - LSL}{\hat{\sigma}^{*(b)}(M - L_p)} \right\},$$

where

$$\hat{\sigma}^{*(b)} = \bar{X}^{*(b)} / \ln(4), \text{ and } \bar{X}^{*(b)} = m^{-1} \sum_{j=1}^m X_{n,j}^*.$$

Thus, the standard bootstrap $(1 - \alpha)100\%$ confidence interval is

$$CI_{SB} = \left(\bar{C}_{pk}^* - Z_{1-\alpha/2} S_c^*, \bar{C}_{pk}^* + Z_{1-\alpha/2} S_c^* \right), \tag{6}$$

where $Z_{1-\alpha/2}$ is a $(1 - \alpha / 2)^{th}$ quantile of the standard normal distribution $\bar{C}_{pk}^* = B^{-1} \sum_{i=1}^B \hat{C}_{pk}^{*(i)}$ and

$$S_c^* = \sqrt{\frac{1}{B-1} \sum_{i=1}^B \left(\hat{C}_{pk}^{*(i)} - \bar{C}_{pk}^* \right)^2}.$$

Percentile Bootstrap (PB) Confidence Interval

The percentile bootstrap $(1 - \alpha)100\%$ confidence interval is given by [14]:

$$CI_{PB} = \left(\hat{C}_{pk}^* \left(\frac{\alpha}{2} \right), \hat{C}_{pk}^* \left(\left(1 - \frac{\alpha}{2} \right) B \right) \right), \tag{7}$$

where $\hat{C}_{pk(r)}^*$ is the r^{th} ordered value on the list of the B bootstrap estimator of C_{pk} .

Bias-Corrected Percentile Bootstrap (BCPB) Confidence Interval

The bootstrap distributions obtained using only a sample of the complete bootstrap distribution may be shifted higher or lower than would be expected. Therefore, this approach has been introduced in order to correct for the potential bias. Firstly, using the ordered distribution of \hat{C}_{pk}^* , compute the probability $P_0 = P(\hat{C}_{pk}^* \leq \hat{C}_{pk})$. Then, $Z_0 = \Phi^{-1}(P_0)$. Therefore, the percentile of the ordered distribution $G^*(\hat{C}_{pk}^*)$, $P_L = \Phi(2Z_0 - Z_{1-\alpha/2})$ and $P_U = \Phi(2Z_0 + Z_{1-\alpha/2})$ are obtained,

where $\Phi(\cdot)$ is the standard normal cumulative function. Finally, the bias-corrected percentile bootstrap $(1-\alpha)100\%$ confidence interval is defined as follows [14]:

$$CI_{BCPB} = \left(\hat{C}_{pk(P_L B)}^*, \hat{C}_{pk(P_U B)}^* \right), \quad (8)$$

where $\hat{C}_{pk(r)}^*$ is the r^{th} ordered value on the list of the B bootstrap estimator of C_{pk} .

To study the different confidence intervals, their estimated coverage probabilities and average widths are considered. For each of the methods considered, a $(1-\alpha)100\%$ confidence interval denoted by (L, U) is obtained (based on $M = 10,000$ replicates). The estimated coverage probability and the average width are given by [16]:

$$\text{Coverage Probability} = \frac{\#(L \leq C_{pk} \leq U)}{M},$$

and

$$\text{Average Width} = \frac{\sum_{i=1}^M (U_i - L_i)}{M}.$$

In the following section, the simulation results are presented in order to evaluate the performance of the confidence intervals CI_{SB} , CI_{PB} , and CI_{BCPB} based on their estimated coverage probabilities and average widths.

SIMULATION RESULTS

In the following, the bootstrap confidence intervals CI_{SB} , CI_{PB} and CI_{BCPB} are compared via Monte Carlo simulation. Using R statistical software [17-19], the data sets are generated from a half-logistic distribution given by Eq.(1) with $\mu = 0$, $\sigma = 1.0, 1.5$ and 2.5 . The scope of the simulation is set under the sample sizes $n = 10, 20, 30, 50$ and 100 , and LSL and USL are taken as $LSL = 1$ and $USL = 29$. To obtain the estimated coverage probabilities and average widths, the 90% and 95% confidence levels are computed by drawing 1,000 bootstrap samples of sizes $m = n$. The simulation results are summarised in Tables 1-2. As expected, the results show that the estimated coverage probabilities for all confidence intervals get closer to the nominal confidence level with increasing sample sizes n . Likewise, the average widths of all confidence intervals get shorter when n increases. This is intuitive in nature because as n increases, it is possible to estimate the scale parameter σ more accurately. A more interesting result is that the estimated coverage probabilities of the CI_{SB} get closer to the nominal confidence level than those of CI_{PB} and CI_{BCPB} . For example, the estimated coverage probabilities attained by the CI_{SB} , CI_{PB} , and CI_{BCPB} are 0.9451, 0.9335 and 0.9341 respectively, for $n = 50$ and $\sigma = 1.0$. Consequently, the average widths of CI_{SB} are longer than those of CI_{PB} and CI_{BCPB} for all situations.

Table 1. Estimated coverage probabilities and average widths of 90% bootstrap confidence intervals of the process capability index

n	σ	Coverage probability			Average width		
		SB	PB	BCPB	SB	PB	BCPB
10	1.0	0.8926	0.8250	0.8310	0.9829	0.9098	0.8730
	1.5	0.8892	0.8206	0.8222	0.6524	0.6052	0.5813
	2.5	0.9102	0.8335	0.8057	0.3937	0.3638	0.3421
20	1.0	0.8992	0.8634	0.8650	0.6150	0.5962	0.5819
	1.5	0.8942	0.8630	0.8645	0.4095	0.3969	0.3873
	2.5	0.8906	0.8555	0.8524	0.2471	0.2395	0.2331
30	1.0	0.8958	0.8713	0.8734	0.4853	0.4754	0.4674
	1.5	0.8890	0.8669	0.8674	0.3248	0.3182	0.3128
	2.5	0.8934	0.8695	0.8705	0.1947	0.1908	0.1876
50	1.0	0.9015	0.8863	0.8861	0.3675	0.3629	0.3592
	1.5	0.8932	0.8781	0.8789	0.2463	0.2432	0.2407
	2.5	0.8932	0.8730	0.8778	0.1467	0.1449	0.1434
100	1.0	0.9003	0.8930	0.8922	0.2565	0.2546	0.2532
	1.5	0.8983	0.8905	0.8900	0.1709	0.1696	0.1687
	2.5	0.9007	0.8932	0.8928	0.1024	0.1017	0.1012

Table 2. Estimated coverage probabilities and average widths of 95% bootstrap confidence intervals of the process capability index

n	σ	Coverage probability			Average width		
		SB	PB	BCPB	SB	PB	BCPB
10	1.0	0.9357	0.8757	0.8809	1.1651	1.1234	1.0788
	1.5	0.9428	0.8806	0.8818	0.7817	0.7526	0.7234
	2.5	0.9508	0.8881	0.8602	0.4689	0.4495	0.4222
20	1.0	0.9438	0.9102	0.9121	0.7325	0.7219	0.7049
	1.5	0.9386	0.9078	0.9091	0.4878	0.4806	0.4693
	2.5	0.9442	0.9170	0.9143	0.2919	0.2874	0.2799
30	1.0	0.9448	0.9247	0.9256	0.5829	0.5768	0.5672
	1.5	0.9458	0.9215	0.9238	0.3859	0.3819	0.3756
	2.5	0.9424	0.9220	0.9212	0.2321	0.2296	0.2257
50	1.0	0.9451	0.9335	0.9341	0.4396	0.4364	0.4318
	1.5	0.9493	0.9358	0.9373	0.2930	0.2907	0.2876
	2.5	0.9475	0.9327	0.9331	0.1758	0.1744	0.1726
100	1.0	0.9470	0.9393	0.9377	0.3055	0.3038	0.3021
	1.5	0.9509	0.9409	0.9428	0.2037	0.2025	0.2014
	2.5	0.9509	0.9419	0.9428	0.1222	0.1215	0.1209

ILLUSTRATIVE EXAMPLE

To illustrate the bootstrap confidence intervals of the process capability index developed in Section 2, a simulated example is presented. The random samples of sizes $n = 20$ are generated from the half-logistic distribution with $\mu = 0$ and $\sigma = 1$. In this case, we set $LSL = 1$, $USL = 29$, and the true process capability index, $C_{pk} = -1/3$. The random samples generated are:

0.04 0.14 0.19 0.20 0.23 0.44 0.75 0.81 0.88 1.07
 1.07 1.09 1.29 1.50 1.62 1.83 1.91 3.56 5.04 5.15.

In addition, the density plot of the generated samples is displayed in Figure 3. Assuming the half-logistic distribution for the corresponding random samples, three bootstrap confidence intervals of the process capability index with a confidence level of 95% are constructed, and they are shown in Table 3. The value of the true C_{pk} lies in the bootstrap confidence intervals. Additionally, the widths of the confidence intervals are similar to the simulation results.

Table 3. Bootstrap confidence intervals and their widths for the process capability index

Method	Confidence interval	Width
SB	(-0.3608 , 0.5187)	0.8795
PB	(-0.4499 , 0.4111)	0.8610
BCPB	(-0.4448 , 0.4129)	0.8577

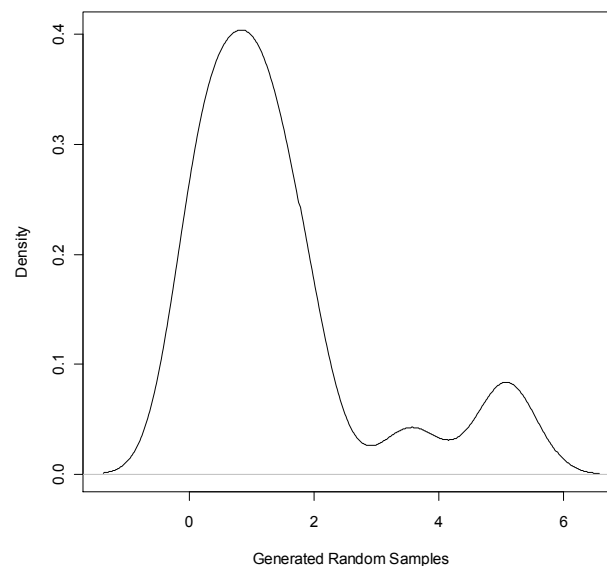


Figure 3. Density plot of generated random samples

CONCLUSIONS

The bootstrap confidence intervals of the process capability index for half-logistic distribution have been proposed. The following were considered: the standard bootstrap confidence interval, the percentile bootstrap confidence interval and the bias-corrected percentile bootstrap

confidence interval. By means of Monte Carlo experiments, the performance of the bootstrap prediction intervals was compared by considering their coverage probabilities and average widths. Based on simulation study, the standard bootstrap confidence interval achieved better coverage probability than the other confidence intervals. Thus, the standard bootstrap confidence interval is more appropriate than its counterparts in this setting.

ACKNOWLEDGEMENTS

The authors are very appreciative of the valuable insights and comments provided by the three anonymous referees, which led to many improvements to the paper. The authors also acknowledge the excellent suggestions given by Professor Dr. R.R.L. Kantam, Professor Dr. David E. Giles and Associate Professor Dr. Punzeyer Patthanangkoor.

REFERENCES

1. N. Balakrishnan, "Order statistics from the half logistic distribution", *J. Stat. Comput. Simul.*, **1985**, 20, 287-309.
2. N. Balakrishnan and P. S. Chan, "Estimation for the scale half logistic distribution under Type II censoring", *Comput. Stat. Data Anal.*, **1992**, 13, 123-141.
3. A. K. Mbah and C. P. Tsokos, "Record values from half logistics and inverse Weibull probability distribution functions", *Neural Parallel Sci. Comput.*, **2008**, 16, 73-92.
4. S. B. Kang, Y. S. Cho and J. T. Han, "Estimation for the half logistic distribution under progressive Type-II censoring", *Commun. Korean Stat. Soc.*, **2008**, 15, 815-823.
5. N. Balakrishnan and H. M. Saleh, "Relations for moments of progressively Type-II censored order statistics from half-logistic distribution with applications to inference", *Comput. Stat. Data Anal.*, **2011**, 55, 2775-2792.
6. D. H. Jang, J. Park and C. Kim, "Estimation of the scale parameter of the half-logistic distribution with multiply type II censored sample", *J. Korean Stat. Soc.*, **2011**, 40, 291-301.
7. R. R. L. Kantam and K. Rosaiah, "Half logistic distribution in acceptance sampling based on life tests", *IAPQR Trans.*, **1998**, 23, 117-125.
8. R. R. L. Kantam, K. Rosaiah and R. Subba Rao, "Estimation of process capability index for half logistic distribution", *Int. Trans. Math. Sci. Comput.*, **2010**, 3, 61-66.
9. R. R. L. Kantam, K. Rosaiah and M. S. R. Anjaneyula, "Estimation of reliability in multicomponent stress-strength model: Half logistic distribution", *IAPQR Trans.*, **2000**, 25, 43-52.
10. S. Niwitpong and J. Sappakitkamjorn, "Biased estimation of the capability indices C_p , C_{pk} and C_{pm} using bootstrap and jackknife methods", *Asia Pac. Manag. Rev.*, **2006**, 10, 1-7.
11. V. E. Kane, "Process capability indices", *J. Qual. Technol.*, **1986**, 18, 41-52.
12. J. Zhang, "Conditional confidence intervals of process capability indices following rejection of preliminary tests", Ph.D. Thesis, **2010**, The University of Texas at Arlington, USA.
13. J. A. Clements, "Process capability calculations for non-normal distributions", *Qual. Prog.*, **1989**, 22, 95-100.

14. B. Efron and R. J. Tibshirani, "An Introduction to the Bootstrap", Chapman and Hall, New York, **1993**.
15. N. Péladeau and Y. Lacouture, "SIMSTAT: Bootstrap computer simulation and statistical program for IBM personal computers", *Behav. Res. Methods*, **1993**, 25, 410-413.
16. W. Panichkitkosolkul, "A simulation comparison of new confidence intervals for the coefficient of variation of a Poisson distribution", *Silpakorn U. Sci. Technol. J.*, **2010**, 4, 14-20.
17. R. Ihaka and R. Gentleman, "R: A language for data analysis and graphics", *J. Comput. Graph. Statist.*, **1996**, 5, 299-314.
18. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R: A Programming Environment for Data Analysis and Graphics, Version 2.15.1", R Core Team, **1999-2012** (Online at: <http://cran.r-project.org/doc/manuals/R-intro.pdf>).
19. The R Core Team, "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, **1999-2010** (Online at: <http://cran.r-project.org/doc/manuals/refman.pdf>).