Applied Mechanics and Materials Vol. 404 (2013) pp 520-525

The Nonparametric Confidence Interval for the Process Capability Index C^*_{pmk}

Shu-fei Wu*

Department of Statistics, Tamkang University, Taiwan, R.O.C.

100665@mail.tku.edu.tw

Keywords: Process Capability Index, Bootstrap Method, Jackknife Method

Abstract. The process capability index C^*_{pmk} which is a generalization of C_{pmk} is defined by the use of the idea of Chan et al. [1] for asymmetric tolerance. In this paper, we proposed a Jackknife confidence interval for and compare its coverage probability with the other three Efron and Tibshirani's [2] bootstrap interval estimate techniques. The simulation results show that the Jackknife method has higher chance of reaching the nominal confidence coefficient for all cases considered in this paper. Therefore this method is recommended for used. One numerical example to demonstrate the construction of confidence interval for the process capability index is also given in this paper.

Introduction

The index C_p only measures the process variation without considering the process centering. The index C_{pk} take the process variation and process centering into account, but not considering the process targeting to the preset target. The index C_{pm} takes the process variation and the process targeting to the preset target into account. Combing the factors considered by indices C_{pk} and C_{pm}, Pearn [3] developed the index C_{pmk}. For asymmetric tolerance($T \neq m$), a simulation comparison study for estimating the process capability index C^*_{pm} is done in Wu [4]. Making use of the idea of Chan et al. [1], the process capability index C^*_{pmk} , a generalization index of C_{pmk} defined as $C^*_{pmk} = \dfrac{\min(D_L - |T-\mu|, D_U - |T-\mu|)}{3\sqrt{\sigma^2 + (\mu - T)^2}} = \dfrac{D^* - |T-\mu|}{\sqrt{\sigma^2 + (\mu - T)^2}}$, where *USL* and *LSL* are the upper and lower specification limits preset by the process engineers, μ is the process mean σ is the process standard deviation, m=(*USL*+*LSL*)/2 is the midpoint of specification limits and d=(*USL*-*LSL*)/2 is the half length of the specification interval, $D_L = T - LSL$, $D_U = USL - T$ and $D^* = \min(D_L, D_U)/3$. Replacing parameters μ and σ^2 by sample mean $\bar{X}$ and sample variance S^2 respectively, then we have the natural estimator as $\hat{C}^*_{pmk} = \dfrac{D^* - |T-\mu|}{\sqrt{S^2 + (\bar{X} - T)^2}}$.

The process must be stable in order to produce the reliable estimates of μ and σ^2.

Since the distribution of $\hat{C}^*_{pmk}$ is quite complicated under normal assumption, Franklin and Wasserman [5] make used of the three Bootstrap confidence interval techniques proposed by Efron and Tibshirani [2] to construct the confidence intervals for C_{pk}. The advantage of Eforn's three interval techniques is nonparametric or free of distribution assumptions of X . In this paper, we proposed a nonparametric Jackknife confidence interval for the index C^*_{pmk} and compare its coverage probabilities with the other three Efron and Tibshirani's methods by simulation study. According the simulation results, the third Efron and Tibshirani's method (BCPB method) always has the highest coverage probability than the other two methods and it can also reach the normal confidence coefficient for some cases when simulation sample is coming from the normal distribution. The simulation results also show that BCPB method has the highest chance of having highest coverage probability than the other three nonparametric methods and this method is recommended for the interval estimation of process capability indices C^*_{pmk} for simulation sample coming from normal or a heavily skewed distribution. For symmetric tolerance ($T = m$), the index C^*_{pmk} is reduced to C_{pmk}. Therefore, all results for the index C^*_{pmk} are applicable for the index C_{pmk}. At last, one real life example is given to demonstrate the construction of confidence interval for the index C^*_{pmk}.

Introduction of four methods

The Bootstrap method was introduced by Efren [6]. Let $X_{1,\ldots}X_n$ be the original random sample from a process with distribution *F*. A Bootstrap sample is one of size *n* drawn (with replacement) from the original sample and is denoted by $X^*_{1,\ldots}X^*_n$. There are a total of n^n such possible samples. Let *B* be the number of Bootstrap samples and B is taken to be 1000 throughout this paper. Let $\bar{X}^*(i)$ and $S^{2*}(i)$ be the sample mean and sample variance based on the *i*th Bootstrap sample. The first three Bootstrap interval estimate methods proposed by Efron and Tibshiani(1986) and the fourth Jackknife are introduced as follows :

Standard bootstrap confidence interval (SB). First, calculate the natural estimator of C^*_{pmk} given by $\hat{C}^*_{pmk}(i) = \frac{D^* - |T - \bar{X}^*(i)|}{\sqrt{S^{2*}(i) + (\bar{X}^*(i) - T)^2}}$ based on the *i*th Bootstrap sample, *i*=1, …, B. Then calculate the sample average of the Bootstrap estimates $\hat{C}^*_{pmk}(\cdot) = \frac{1}{B}\sum_{=1}^{B}\hat{C}^*_{pmk}(i)$ and the sample standard deviation of Bootstrap estimates $S_{\hat{C}^*_{pmk}} = \sqrt{\frac{1}{B}\sum_{=1}^{B}[\hat{C}^*_{pmk}(i) - \hat{C}^*_{pmk}(\cdot)]^2}$.

Then the $(1-\alpha)100\%$ confidence interval for C^*_{pmk} is $(\hat{C}^*_{pmk} \pm Z_{\alpha/2} S_{\hat{C}^*_{pmk}})$, where $Z_{\alpha/2}$ is the right tail $\alpha/2$ percentile of a standard normal random variable Z. If a 95% confidence interval is desired, then $Z_{\alpha/2} = 1.96$. If a 97.5% lower confidence interval of the index is desired, the lower confidence limit can be easily obtained by simply selecting the lower value of the two-sided confidence interval.

Percentile bootstrap confidence interval (PB). Let $\hat{C}^*_{pmk}(1) \le \hat{C}^*_{pmk}(2) \le \cdots \le \hat{C}^*_{pmk}(B)$ be the sorted Bootstrap estimates. Then $\hat{C}^*_{pmk}(B*\alpha/2)$ and $\hat{C}^*_{pmk}(B*(1-\alpha/2))$ are the $\alpha/2$ and (1-α/2) percentile points of the distribution of $\hat{C}^*_{pmk}(i)$. The $(1-\alpha)100\%$ approximate confidence interval for C^*_{pmk} is given by $(\hat{C}^*_{pmk}(B*\alpha/2), \hat{C}^*_{pmk}(B*(1-\alpha/2)))$.

Biased corrected percentile bootstrap confidence interval(BCPB). Since the Bootstrap distribution may be a biased distribution, the third method was developed to correct for this potential bias. For example, if $\hat{C}^*_{pmk}$ is 1.63 and in the order values of $\hat{C}^*_{pmk}(i)$ we have $\hat{C}^*_{pmk}(412) = 1.61$ and $\hat{C}^*_{pmk}(423) = 1.66$, then $p_0 = P(\hat{C}^*_{pmk} \le 1.63) = 412/1000 = .412$. Calculate $Z_0 = \phi^{-1}(p_0) = \phi^{-1}(.412) = -.222$, where ϕ^{-1} is the inverse of the distribution function standard normal random variable Z. Then calculate $P_L = \phi(2Z_0 - Z_{\alpha/2})$ and $P_U = \phi(2Z_0 + Z_{\alpha/2})$, where ϕ is the cdf of a standard normal variable Z. Then the $(1-\alpha)100\%$ approximate confidence interval for C^*_{pmk} is given by $[\hat{C}^*_{pmk} = ([P_L \times B]+1), \hat{C}^*_{pmk} = ([P_U \times B]+1)]$, where [x] denotes the largest integer being less than or equal to *x*.

Jackknife method. Quenoulli [7] originally introduced the Jackknife as method of reducing the bias of an estimator of a serial correlation coefficient. We employ his method as follows: Let $\hat{\theta} = \hat{C}^*_{pmk}$ denote the natural estimator of $\theta = C^*_{pmk}$ based on the complete sample. Eliminating the first observation, we make use of the remaining $n-1$ observations to calculate the first natural estimator of C^*_{pmk} and denoted by $\hat{\theta}_{(1)}$ .Similarly, eliminating the second observation, we can have the second natural estimator of C^*_{pmk} and denoted by $\hat{\theta}_{(2)}$ based on the remaining n-1 observations. Repeat the same procedure, we can have n natural estimators denoted by $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(n)}$ based on the subsample of size $n-1$. The ith perseudo value is defined as $\hat{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$, i=1,2,……,n. The Quenoulli's estimator is the mean of the $\hat{\theta}_i, \hat{\bar{\theta}}$. The Jackknife estimator of standard error is $S_{\hat{\bar{\theta}}} = \sqrt{\dfrac{\sum_{i=1}^{n}(\hat{\theta}_i - \hat{\bar{\theta}})^2}{k(k-1)}}$. Turkey[8] suggested that the statistic $t = \dfrac{\hat{\bar{\theta}} - \theta}{S_{\hat{\bar{\theta}}}}$ should be distributed approximately as Student's t with $k-1$ degrees of freedom. Then the $(1-\alpha)100\%$ approximate confidence interval for C^*_{pmk} is given by $[\hat{\bar{\theta}} \pm t_{\alpha/2}(n-1)S_{\hat{\bar{\theta}}}]$, where $t_{\alpha/2}(n-1)$ is the right tail $\alpha/2$ percentile of a student's t distribution.

Simulation Comparisons

For simulation studies, the upper limit and lower limit of the process are set to be USL=60 and LSL=40 and then midpoint of the specification limits is m=50. All simulation studies are accomplished by Fortran IMSL[9] subroutines for n=10(10)40(20)60 from a normal distribution with combinations of (μ, σ^2)=(50,4),(50,9),(52,4),(52,9). The target values are set to T=55 under asymmetric tolerance, then the corresponding true index values are C^*_{pmk}=(.309,.286,.462,.393). With 1000 simulation runs, the percentage of times the actual index contained in the intervals of four methods out of 1000 is calculated and the average length of the 95% confidence intervals is also computed. All simulation results are listed in Tables 1 for normal process and a highly skewed process, where a highly skewed process has the same structure of means and variances and is created by simulating a Chi-square distribution with 4 degrees of freedom and suitably scaling and shifting the distribution. The frequency of coverage is a Binomial event with p=.95 and n=1000. Thus a 95% confidence interval surrounding the expected coverage frequency .95 would have a bound of $\pm 1.96\sqrt{(.95)(.05)/1000} = \pm .0135$. The frequency of coverage significantly different from the expected value of .95 are marked by an asterisk (*) in Tables 1. From Table 1, the coverage probabilities increase and the average lengths decrease when the sample size n increases for most cases. Four methods have better performance for the normal process than the heavily skewed process because the heavily skewed process is always the most difficult process to deal with. The optimal methods based on the highest coverage probability in order to reach the nominal confidence coefficient for different situations are listed in Table 2. From Table 2, BCPB method has the highest chance of having highest coverage probability than the other three nonparametric methods and this method is recommended for the interval estimation of process capability indices C^*_{pmk}.

Table 1: Th

	n		
$(\mu,\sigma^2)=(50,4)$	n=20		n
	Coverage	Length	C
SB	0.913*	1.257	0.
PB	0.913*	1.255	0.
BCPB	0.939	1.308	0.
Jackknife	0.871*	1.396	0.
$(\mu,\sigma^2)=(50,9)$			
SB	0.908*	0.980	0.
PB	0.895*	0.977	0.
BCPB	0.910*	1.001	0.
Jackknife	0.834*	1.131	0.
$(\mu,\sigma^2)=(52,4)$			
SB	0.947	0.873	0.
PB	0.936	0.862	0.
BCPB	0.941	0.855	0.
Jackknife	0.950	0.931	0.
$(\mu,\sigma^2)=(52,9)$			
SB	0.924*	0.900	0.
PB	0.922*	0.886	0.
BCPB	0.930*	0.881	0.
Jackknife	0.938	0.988	0.

Table 2: The

Numerical Example

The example 5-1 in Montgom
confidence interval estimates
example, the inside diameter

Table 1: The coverage probability and average length with T=55

	normal process						heavily skewed process					
$(\mu,\sigma^2)=(50,4)$	n=20		n=40		n=60		n=20		n=40		n=60	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
SB	0.913*	1.257	0.918*	0.887	0.921*	0.718	0.871*	0.183	0.897*	0.136	0.911*	0.112
PB	0.913*	1.255	0.912*	0.885	0.906*	0.717	0.873*	0.179	0.908*	0.133	0.919*	0.110
BCPB	0.939	1.308	0.939	0.912	0.924*	0.730	0.878*	0.186	0.911*	0.137	0.921*	0.113
Jackknife	0.871*	1.396	0.876*	0.973	0.869*	0.784	0.882*	0.196	0.901*	0.140	0.910*	0.114
$(\mu,\sigma^2)=(50,9)$												
SB	0.908*	0.980	0.922*	0.680	0.918*	0.554	0.864*	0.320	0.912*	0.229	0.902*	0.187
PB	0.895*	0.977	0.886*	0.680	0.873*	0.553	0.869*	0.312	0.918*	0.225	0.915*	0.184
BCPB	0.910*	1.001	0.913*	0.682	0.907*	0.551	0.879*	0.325	0.925*	0.231	0.926*	0.189
Jackknife	0.834*	1.131	0.848*	0.773	0.858*	0.625	0.868*	0.348	0.915*	0.239	0.901*	0.193
$(\mu,\sigma^2)=(52,4)$												
SB	0.947	0.873	0.939	0.607	0.954	0.493	0.872*	0.389	0.920*	0.283	0.906*	0.229
PB	0.936	0.862	0.934*	0.602	0.947	0.489	0.886*	0.380	0.921*	0.279	0.917*	0.226
BCPB	0.941	0.855	0.940	0.601	0.958	0.489	0.894*	0.392	0.929*	0.285	0.922*	0.229
Jackknife	0.950	0.931	0.946	0.627	0.958	0.504	0.881*	0.421	0.919*	0.295	0.906*	0.235
$(\mu,\sigma^2)=(52,9)$												
SB	0.924*	0.900	0.945	0.641	0.954	0.515	0.872*	0.518	0.914*	0.389	0.921*	0.327
PB	0.922*	0.886	0.943	0.636	0.954	0.511	0.896*	0.506	0.922*	0.383	0.922*	0.323
BCPB	0.930*	0.881	0.948	0.634	0.957	0.511	0.902*	0.517	0.926*	0.390	0.928*	0.327
Jackknife	0.938	0.988	0.946	0.669	0.953	0.529	0.878*	0.579	0.916*	0.413	0.928*	0.341

Table 2: The optimal method among four nonparametric methods

	T=55	
(μ,σ^2)	Normal	Heavily Shewed
(50,4)	BCPB	Jackknife
(50,9)	SB	BCPB
(52,4)	Jackknife	BCPB
(52,9)	Jackknife	BCPB

Numerical Example

The example 5-1 in Montgomery [10] is used to demonstrate the construction of 90% and 95% confidence interval estimates and the 95% and 97.5% lower confidence limit of C^*_{pmk}. In that example, the inside diameter measurement data of the 125 Piston rings for an automotive engine produced by a forging process is recorded in Table 5-1. The sample mean and the sample variance are obtained as 74.001176 and .010199. The upper limit, lower limit of the specification interval are given by 74.041972 and 73.96038 respectively and thus the midpoint is m=74.001176. The target for asymmetric tolerance and is given by 74.003. The natural point estimates of the corresponding index is $\hat{C}^*_{pmk} = 1.078$ for asymmetric tolerance. Their confidence interval estimates or the lower confidence limits are presented in Table 3. Usually, if a process with $C^*_{pmk} > 1$, then it can be considered to be a capable process. From Table 3, we can conclude that this Piston rings process is incapable for asymmetric tolerance ($T = 74.003 \neq m$).

Table 3: The 90% and 95% confidence intervals (length) or the 95% and 97.5% lower confidence bound for C_{pmk} with T=74.001176 (symmetric tolerance) and for C^*_{pmk} with T=74.003 (asymmetric tolerance).

T=74.003 $\hat{C}^*_{pmk}$=1.078	90% confidence intervals (length) 95% lower confidence bound	T=74.003 $\hat{C}^*_{pmk}$=1.078	95% confidence intervals (length) 97.5% lower confidence bound
SB	(0.867,1.289)(0.422) (0.867, ∞)	SB	(0.831,1.325)(0.494) (0.831, ∞)
PB	(0.871,1.290)(0.418) (0.871, ∞)	PB	(0.837,1.332)(0.495) (0.837, ∞)
BCPB	(0.858,1.281)(0.423) (0.858, ∞)	BCPB	(0.828,1.326)(0.498) (0.858, ∞)
Jackknife	(0.850,1.295)(0.445) (0.850, ∞)	Jackknife	(0.806,1.326)(0.520) (0.806, ∞)

Conclusion

In estimating any process capability index that confidence intervals estimates should be used instead of the simple point estimates. The nonparametric confidence intervals estimates can protect the user from the error of calculating confidence intervals based on an assumed normal process if the process is a distinctly non normal process. Among four methods, the BCPB method is recommended for use. A computer program is provided by authors to obtain the interval estimates of the index by four methods for symmetric or asymmetric tolerance.

Acknowledgment

The author wish to thank an associate editor and referees for their careful reading and valuable suggestions so that the article is more readable and applicable. The author's research was supported by The National Science Council NSC101-2118-M-032-003- in Taiwan, ROC.

Corresponding Author

Shu-Fei Wu, 100665@mail.tku.edu.tw, 88620953305675

References

[1] L.K. Chan, S.W. Cheng and F. A: Spiring, A new measure of process capability: C_{pm}. Journal of Quality Techology, Vol. 20 (1988), p. 162-175.

[2] B. Efron and R. J. Tibshirani: Bootstrap Method for standard errors, confidence Intervals, and Other Measures of Statistical Accuracy. Statistic Science, Vol. 1 (1986), p. 54-77.

[3] W. L. Pearn, G. H. Lin and K. S. Chen: Distributional and Inferential Properties of the process accuracy and process precision indices. Communications in Statistics-Theory and Methods, Vol. 27 (1998), p. 985-1000

[4] S. F. Wu: A Simulation Comp with Asymmetric Tolerances. Vol. 20 (2009), p. 243-253

[5] L. A. Franklin and G. Was introduction. Communications 231-242

[6] B. Efron: Bootstrap methods: (1979), p. 1-26

[7] M. H. Quenouille: Approxima Statistical Society, B 11 (1949)

[8] J. W. Tukey: Bias and Confider 29 (1958), p. 614

[9] Microsoft Developer Studio F

[10] D. C. Montgomery: Introduct (2001)

[4] S. F. Wu: A Simulation Comparison Study for Estimating the Process Capability Index C^*_{pm} with Asymmetric Tolerances. International Journal of Information and Management Sciences, Vol. 20 (2009), p. 243-253

[5] L. A. Franklin and G. Wasserman: Bootstrap confidence interval estimates of C_{pk}: An introduction. Communications in Statistics-Simulation and Computations, Vol.20 (1991), p. 231-242

[6] B. Efron: Bootstrap methods: Another look at the Jackknife. *The annals of statistics*, Vol. 7 (1979), p. 1-26

[7] M. H. Quenouille: Approximate tests for the correlation in time series. Journal of the Royal Statistical Society, B 11 (1949), p. 68-84

[8] J. W. Tukey: Bias and Confidence in not quite large samples. Annals of Mathematical Statistics, 29 (1958), p. 614

[9] Microsoft Developer Studio Fortran Powerstage 4.0 and IMSL, 1995, Microsoft Corporation

[10] D. C. Montgomery: Introduction to statistic quality control. John Wiley & Sons, New York, NY. (2001)