QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2002; **18**: 111–116 (DOI: 10.1002/qre.450)

ON THE DISTRIBUTION OF THE ESTIMATED PROCESS YIELD INDEX S_{pk}

J. C. LEE[1]*, H. N. HUNG[1], W. L. PEARN[2] AND T. L. KUENG[1]

[1] *Institute of Statistics, National Chiao Tung University, 1001 TA Hsueh Road, Hsinchu, Taiwan 30050, Republic of China*
[2] *Department of Industrial Engineering and Mangement, National Chiao Tung University, 1001 TA Hsueh Road, Hsinchu, Taiwan 30050, Republic of China*

SUMMARY

This paper considers an asymptotic distribution for an estimate $\hat{S}_{pk}$ of the process yield index S_{pk} proposed by Boyles (1994). The asymptotic distribution of $\hat{S}_{pk}$ is useful in statistical inferences for S_{pk}. An illustrative example is given for hypothesis testing and for interval estimation on the yield index S_{pk}.

KEY WORDS: central limit theorem; illustrative example; normal distribution

1. INTRODUCTION

Process capability indices, establishing the relationship between the actual process performance and the manufacturing specifications, have been the focus of recent research in quality assurance and capability analysis. Those capability indices, quantifying process potential and process performance, are essential to any successful quality improvement activities and quality program implementation. Some basic capability indices that have been widely used in the manufacturing industry include C_p, C_a and C_{pk}. These indices are explicitly defined as follows [1–3]:

$$C_p = \frac{USL - LSL}{6\sigma} \qquad C_a = 1 - \frac{|\mu - m|}{d}$$

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

where USL and LSL are the upper and the lower specification limits, respectively, μ is the process mean, σ is the process standard deviation, $m = (USL + LSL)/2$ is the midpoint of the specification interval and $d = (USL - LSL)/2$ is half the length of specification interval. We will focus on the situation in which the specification interval is two-sided with the target value T at m, which is most common in practice.

The index C_p measures the overall process variation relative to the specification tolerance and, therefore, only reflects process potential (or process precision). The index C_a measures the degree of process centering, which alerts the user if the process mean deviates from its target value. Therefore, the index C_a only reflects process accuracy. The index C_{pk} takes into account the magnitudes of process variation as well as the degree of process centering, which measures process performance based on yield (proportion of conformities). For a normally distributed process with a fixed value of C_{pk}, the bounds on process yield are given by

$$2\Phi(3C_{pk}) - 1 \le \%\text{Yield} < \Phi(3C_{pk})$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$, the standard normal distribution. For example, if $C_{pk} = 1.00$, then it guarantees that the %Yield will be no less than 99.73%, or no greater than 2700 ppm (parts per million) of nonconformities. We note that the index C_{pk} only provides an approximate rather than an exact measure of the process yield.

To obtain an exact measure, Boyles [4] considered a yield index, referred to as S_{pk}, for processes with normal distributions. The index S_{pk} is defined as:

$$S_{pk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{USL - \mu}{\sigma}\right) + \frac{1}{2}\Phi\left(\frac{\mu - LSL}{\sigma}\right)\right\} \tag{1.1}$$

It can be written as

$$S_{pk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{1 - C_{dr}}{C_{dp}}\right) + \frac{1}{2}\Phi\left(\frac{1 + C_{dr}}{C_{dp}}\right)\right\} \tag{1.2}$$

*Correspondence to: J. C. Lee, Institute of Statistics, National Chiao Tung University, 1001 TA Hsueh Road, Hsinchu, Taiwan 30050, Republic of China.

Received 27 June 2001
Revised 10 November 2001

Table 1. Various S_{pk} values and the corresponding process yield

S_{pk}	Process yield
1.00	0.997 300 204
1.24	0.999 800 777
1.33	0.999 933 927
1.50	0.999 993 205
1.67	0.999 999 456
2.00	0.999 999 998

where Φ^{-1} is the inverse function of Φ, $C_{dr} = (\mu - m)/d$ and $C_{dp} = \sigma/d$. For a process with $S_{pk} = c$, we can obtain %Yield $= 2\Phi(3c) - 1$. Obviously, there is a one-to-one relationship between S_{pk} and the process yield. Thus, the yield index S_{pk} provides an exact measure of the process yield. For normal processes, the expected number of non-conformities corresponding to a capable process with $S_{pk} = 1.00$ is 2700 ppm, a satisfactory process with $S_{pk} = 1.33$ is 63 ppm, an excellent process with $S_{pk} = 1.67$ is 0.6 ppm and a super process with $S_{pk} = 2.00$ is 0.002 ppm, as summarized in Table 1.

2. AN ESTIMATOR AND ITS DISTRIBUTION

Let $X_1, \ldots, X_n$ be a random sample from the normal process, a natural estimator of S_{pk} is

$$\hat{S}_{pk} = \frac{1}{3}\Phi^{-1}\left\{\frac{1}{2}\Phi\left(\frac{1-\hat{C}_{dr}}{\hat{C}_{dp}}\right)+\frac{1}{2}\Phi\left(\frac{1+\hat{C}_{dr}}{\hat{C}_{dp}}\right)\right\} \tag{2.1}$$

where $\hat{C}_{dr}$ and $\hat{C}_{dp}$ are estimates of C_{dr} and C_{dp}, respectively and are defined as

$$\hat{C}_{dr} = (\bar{X} - m)/d$$
$$\hat{C}_{dp} = S/d \tag{2.2}$$

with $\bar{X} = 1/n\sum_{i=1}^{n} X_i$ and $S^2 = 1/(n-1)\sum_{i=1}^{n}(X_i - \bar{X})^2$. The distribution of $\hat{S}_{pk}$ is non-trivial as it is a complex function of the statistics $\bar{X}$ and S^2. However, a useful approximate distribution of $\hat{S}_{pk}$ can be furnished by considering the following asymptotic expansion of $\hat{S}_{pk}$. Let $Z = \sqrt{n}(\bar{X} - \mu)$, $Y = \sqrt{n}(S^2 - \sigma^2)$, then Z and Y are independent and since the first two moments of $\bar{X}$ and S^2 exist, by the Central Limit Theorem they converge to $N(0, \sigma^2)$ and $N(0, 2\sigma^2)$, respectively, as n goes to infinity. Consequently, $\hat{S}_{pk}$ can be expressed as

$$\hat{S}_{pk} = S_{pk} + \frac{1}{6\sqrt{n}}(\phi(3S_{pk}))^{-1}W + \mathrm{O_p}(n^{-1}) \tag{2.3}$$

where

$$W = -\frac{d}{2\sigma^3}Y\left[(1+C_{dr})\phi\left(\frac{1+C_{dr}}{C_{dp}}\right) + (1-C_{dr})\phi\left(\frac{1-C_{dr}}{C_{dp}}\right)\right] - \frac{1}{dC_{dp}}Z\left[\phi\left(\frac{1-C_{dr}}{C_{dp}}\right) - \phi\left(\frac{1+C_{dr}}{C_{dp}}\right)\right] \tag{2.4}$$

which is normally distributed with a mean of zero and a variance of $a^2 + b^2$,

$$a = \frac{d}{\sqrt{2}\sigma}\left\{(1-C_{dr})\phi\left(\frac{1-C_{dr}}{C_{dp}}\right) + (1+C_{dr})\phi\left(\frac{1+C_{dr}}{C_{dp}}\right)\right\} \tag{2.5}$$
$$b = \phi\left(\frac{1-C_{dr}}{C_{dp}}\right) - \phi\left(\frac{1+C_{dr}}{C_{dp}}\right)$$

and ϕ is the probability density function of the standard normal distribution. It is noted that the asymptotic expansion of $\hat{S}_{pk}$, as given in (2.3), indicates that asymptotically $\hat{S}_{pk}$ is normally distributed with mean S_{pk} and variance $(a^2 + b^2)/36n(\phi(3S_{pk}))^2$. In (2.3), C_{dr} and C_{dp} appear in the asymptotic expression of $\hat{S}_{pk}$ as a consequence of the Taylor expansion used in the asymptotic expansion of $\hat{S}_{pk}$ around the true values C_{dr} and C_{dp}. In practice, $\hat{C}_{dr}$ and $\hat{C}_{dp}$ are used instead because they will converge to C_{dr} and C_{dp}, respectively. Also, the remaining terms $\mathrm{O_p}(n^{-1})$ represent the error of the expansion having a leading term of order n^{-1} in probability.

The first-order approximation of $\hat{S}_{pk}$, as given in (2.3), can produce an adequate approximate distribution for a large enough sample size. Figures 1 and 2 depict approximate and exact distributions, obtained by simulations, with a sample size of n = 100, 200, 300, 400, 500, 1000. It is clear that as the sample size n reaches 1000, the approximate and exact distributions are almost indistinguishable. In fact, even with $n = 100$ the approximation is quite reasonable for practical purposes.

3. INFERENCE BASED ON $\hat{S}_{pk}$

From (2.3) and (2.4), it is clear that $\hat{S}_{pk}$ is an asymptotically unbiased estimator of S_{pk}. Also, thanks to the asymptotic distribution of $\hat{S}_{pk}$ given in (2.3)–(2.5), hypothesis testing and a confidence interval for

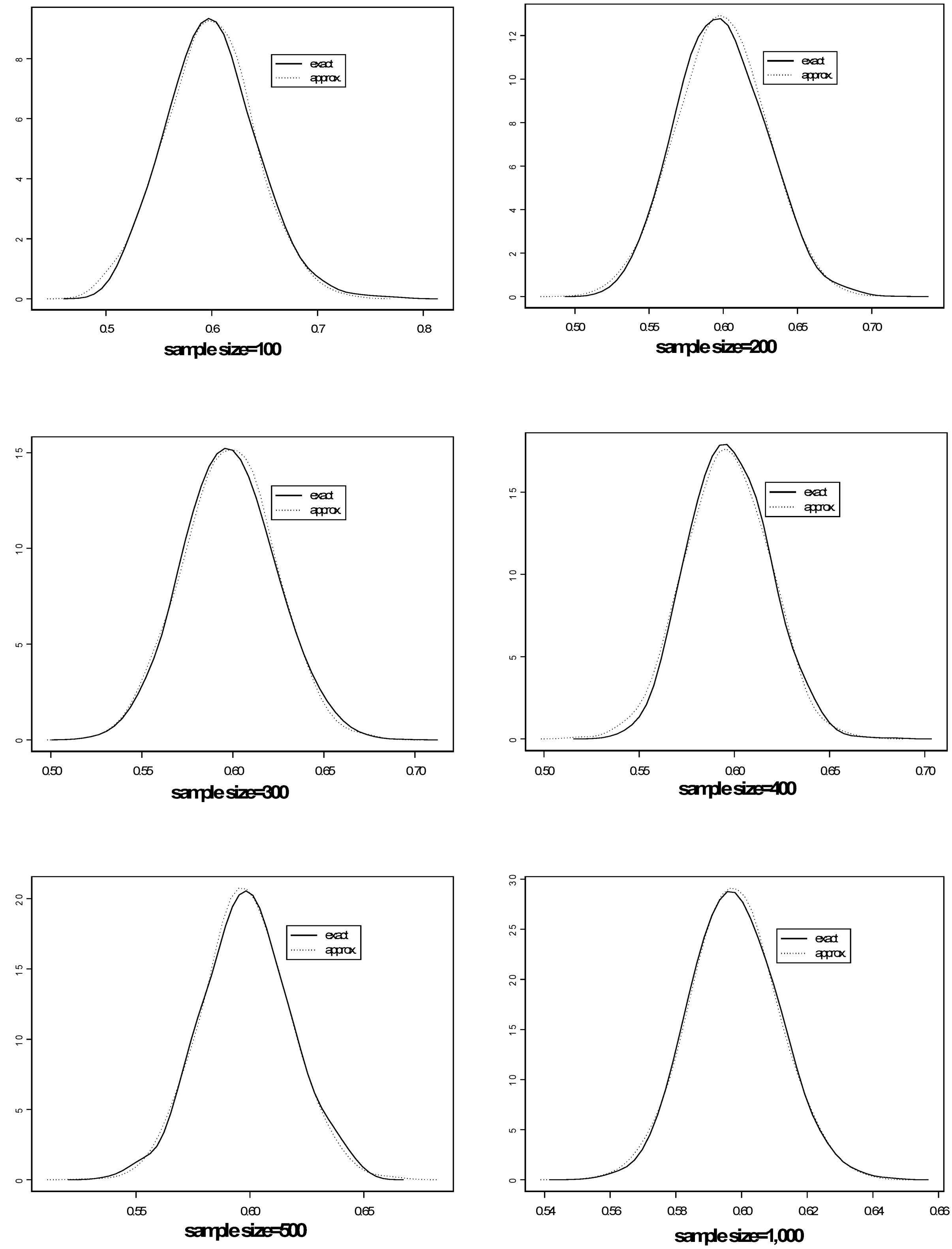

Figure 1. Comparison of approximate and exact densities via simulations

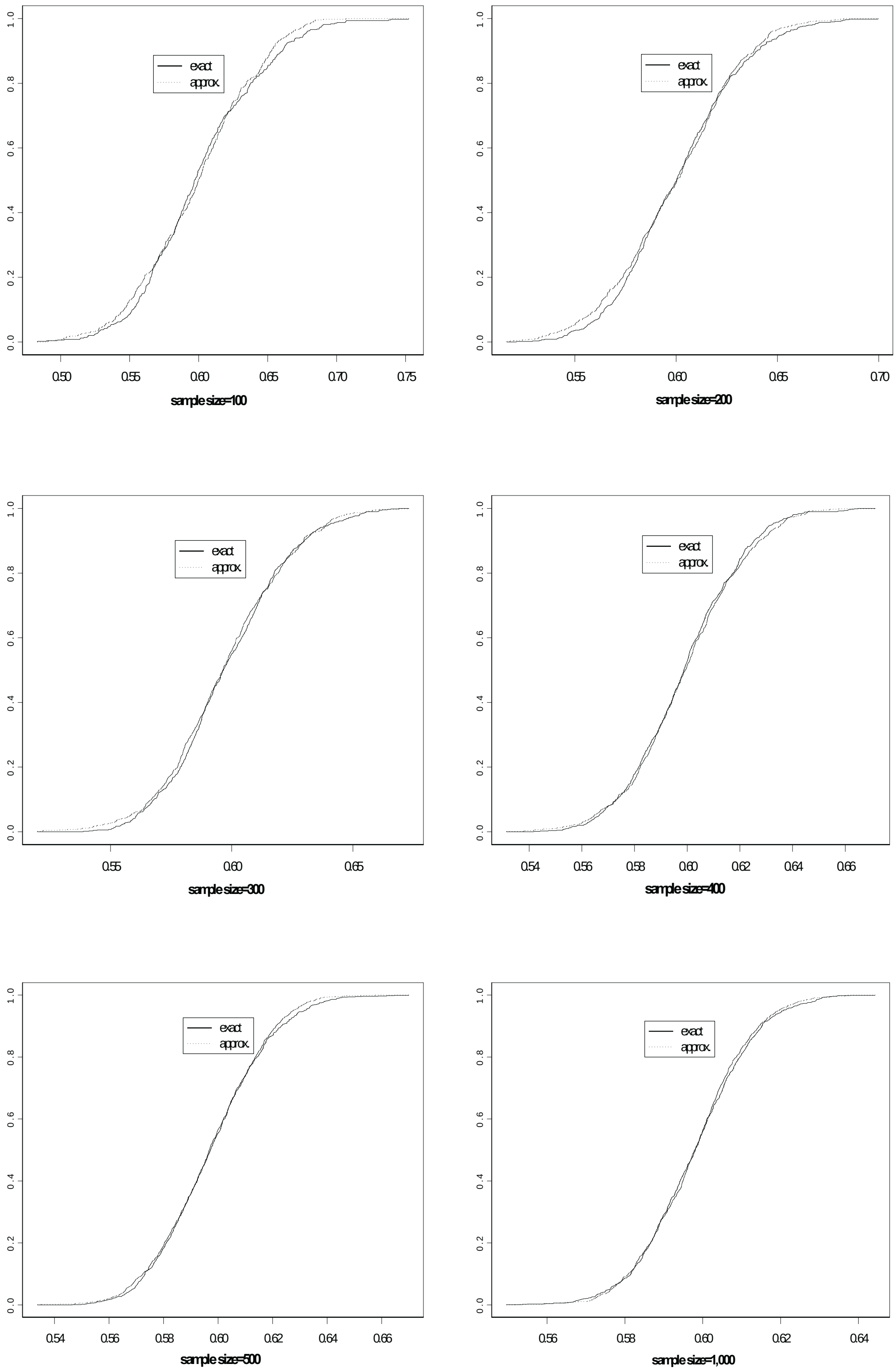

Figure 2. Comparison of approximate and exact cumulative distribution functions via simulations

S_{pk} can be constructed. For example, the following null hypothesis

$$H_0 : S_{pk} \leq c, \quad \text{a specified value} \tag{3.1}$$

versus alternative hypothesis

$$H_1 : S_{pk} > c \tag{3.2}$$

can be executed by considering the testing statistic

$$T = \frac{6(\hat{S}_{pk} - c)\sqrt{n}\phi(3\hat{S}_{pk})}{\sqrt{\hat{a}^2 + \hat{b}^2}} \tag{3.3}$$

where $\hat{a}$ and $\hat{b}$ are estimates of a and b, with C_{dr}, C_{dp} and σ replaced by $\hat{C}_{dr}$, $\hat{C}_{dp}$ and S, respectively. The null hypothesis H_0 is rejected at α level if $T > z_\alpha$, where z_α is the upper $100\alpha\%$ point of the standard normal distribution.

An approximate $1 - \alpha$ confidence interval for S_{pk} is

$$\left(\hat{S}_{pk} - \frac{\sqrt{\hat{a}^2 + \hat{b}^2}}{6\sqrt{n}\phi(3\hat{S}_{pk})} z_{\alpha/2},\ \hat{S}_{pk} + \frac{\sqrt{\hat{a}^2 + \hat{b}^2}}{6\sqrt{n}\phi(3\hat{S}_{pk})} z_{\alpha/2}\right) \tag{3.4}$$

An illustrative example is given in the next section.

4. AN ILLUSTRATIVE EXAMPLE

Consider the following example taken from a supplier manufacturing high-end audio speaker drivers including 3-inch tweeters, 3-inch full-range drivers, 5-inch mid-range drivers, 6.5-inch woofers and 8-inch, 10-inch and 12-inch subwoofers. A standard woofer driver consists of components including an edge, cone, dustcap, spider (damper), voice coil, lead wire, frame, magnet, front plate and back plate. The edge (on the top) and the spider (on the bottom) are glued onto the frame to hold the cone for the piston movement and the dustcap (glued onto the cone to cover the top of the voice coil) decouples the noise from the musical signals. One characteristic, which reflects the bass performance, musical image, clarity and cleanness of the sound, transparence and compliance (excursion movement) of the mid-range, full-range or subwoofer driver units, is F_0 (the free-air resonance frequency). Some key factors, which determine F_0 values, include the hardness, thickness, weight of the damper, weight of the edges and the weight of the cone. Typical ranges of F_0 values are 25–40 Hz for subwoofers, 40–60 Hz for woofers, 50–100 Hz for full-range, and 500–5000 Hz for mid-ranges.

One particular model of the 3-inch full-range drivers, designed particularly for the central and background channels of home-theater applications, has used the specially designed Pulux edge, Pulux dustcap and PP-mica cone. This model of 3-inch driver requires the F_0 value to be 80 Hz with ± 10 Hz tolerance. The production specification limits for this particular model of drivers are therefore set to $(LSL, T, USL) = (70, 80, 90)$ for F_0. The quality requirement was predefined as $S_{pk} \geq 1.00$ (equivalent to $USL - LSL = 6\sigma$). A total of 100 samples of data were collected from the factory, which are displayed in Table 2.

Table 2. F_0 measures for the 3-inch drivers

81	80	82	79	78	76	78	78	76	81
83	78	81	85	81	78	79	79	80	82
79	79	82	78	82	80	75	85	80	80
80	75	81	78	82	84	76	78	80	79
82	82	78	78	82	78	82	80	82	83
81	78	83	81	82	79	80	79	81	82
79	80	82	77	81	80	81	81	75	76
83	86	82	79	82	85	80	80	77	75
78	85	81	79	81	83	78	78	80	80
79	76	77	74	85	83	76	80	75	82

Thus, statistical inferences on the index, S_{pk}, such as hypothesis testing and interval estimation can be considered. For testing the null hypothesis H_0 as given in (3.1) with $c = 1$ against the alternative hypothesis as given in (3.2), the testing statistic T, as given in (3.3), yields 3.1389. Since $3.1389 > z_{0.05} = 1.96$, the null hypothesis H_0 is rejected at $\alpha = 0.05$. We may conclude that the process satisfies the capability requirement $S_{pk} \geq 1.00$. Moreover, an approximate 95% confidence interval for S_{pk} is easily obtained from (3.4) as

$$(1.1078, 1.4664)$$

which is consistent with the hypothesis testing result.

REFERENCES

1. Kane VE. Process capability indices. *Journal of Quality Technology* 1986; **18**(1):41–52.
2. Pearn WL, Kotz S, Johnson NL. Distributional and inferential properties of process capability indices. *Journal of Quality Technology* 1992; **24**(4):216–231.
3. Pearn WL, Lin GH, Chen KS. Distributional and inferential properties of the process accuracy and process precision indices. *Communications in Statistics: Theory and Methods* 1998; **27**(4):985–1000.
4. Boyles RA. Process capability with asymmetric tolerances. *Communications in Statistics: Computation and Simulation* 1994; **23**(3):615–643.

Authors' biographies:

Jack C. Lee is a Professor at the Institute of Statistics, National Chiao Tung University, Taiwan. He received a BA in Business from the National Taiwan University, an MA in Economics from the University of Rochester and a PhD in Statistics from SUNY at Buffalo.

H. N. Hung is an Associate Professor at the Institute of Statistics, National Chiao Tung University, Taiwan. He received a BS in Mathematics from the National Taiwan University, an MS in Mathematics from National Tsing Hua University and a PhD in Statistics from the University of Chicago.

W. L. Pearn recieved his MS degree in Statistics and PhD degree in operations Research from the University of Maryland at College Park, USA. He worked for Bell Laboratories as a quality engineer before he joined the National Chiao Tung University. Currently, he is a Professor in the Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan.

T. L. Kueng is a Senior Engineer at the Behavior Design Corporation, Taiwan. He received a BA in Business from the National Taiwan University and an MS in Statistics from the National Chiao Tung University, Taiwan.

