axioms

MDPI

Article

Modified C_p and C_{pk} Indices Based on Left-Truncated Data

Yimin Yin [1], Bin Yan [1,*] and Pengfei Liu [2,*]

[1] School of Mathematics and Statistics, Hunan First Normal University, Changsha 410205, China; yinyimin@hnfnu.edu.cn
[2] College of Advanced Interdisciplinary Studies, National University of Defense Technology, Changsha 410073, China
* Correspondence: byan068@hnfnu.edu.cn (B.Y.); liupengfei08a@nudt.edu.cn (P.L.)

Abstract

The process capability indices C_p and C_{pk} are commonly used in industry to evaluate process capability, but they usually require that quality data follow a normal distribution. However, in the actual supply–demand relationship, some suppliers artificially eliminate products that do not meet the inspection requirements in order to make buyers accept their products, and these truncated sample data have a more significant impact on process capability evaluation. Based on the left-truncated sample, two modified process capability indices, C_p^T and C_{pk}^T, are proposed, and bootstrap confidence interval estimation methods are established for each of them. Extensive simulation experiments are conducted on the modified indices by varying the sample size and truncation location parameters, and the results are compared with those of traditional methods. The comparison reveals that the new methods outperform the traditional ones across a range of sample sizes and truncation locations. Finally, a real example is used to validate the usefulness of the new method in guiding production management.

Keywords: process capability index; left truncation; bootstrap; production management

MSC: 62P30



Academic Editor: Hari Mohan Srivastava

Received: 1 August 2025
Revised: 10 September 2025
Accepted: 13 September 2025
Published: 16 September 2025

Citation: Yin, Y.; Yan, B.; Liu, P. Modified C_p and C_{pk} Indices Based on Left-Truncated Data. *Axioms* **2025**, *14*, 699. https://doi.org/10.3390/axioms14090699

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The process capability index (PCI) is a quality management tool for evaluating whether an industrial process meets given standards. The prerequisites for using PCIs are that the process must be under control and that the quality data used for evaluation follow a normal distribution. However, in practical applications, the assumption of independent normality may be violated—for example, the data may be non-normally distributed [1,2], or contain autocorrelation [3–5], etc. In recent years, some manufacturing companies have had data fraud incidents, such as Mitsubishi, which admitted that the company had quality data fraud incidents in order to meet customer demand standards from 2015 to 2017. The most common form of data fraud in industrial production activities is data truncation or bad data deletion, which not only destroys trust in the supply chain but also misleads decisions and choices, and then leads to larger-scale economic losses.

The truncated normal distribution is a commonly observed distribution in industrial production, where some factories produce products that must be scrapped or reworked if their specifications are not met. When product specifications follow a normal distribution, the remaining qualified products follow a truncated normal distribution [6]. In addition, in the supply chain of industrial products, some salesmen will select products that meet

the specifications for testing by the demander in order to become potential suppliers of a factory, and these test samples also follow a truncated normal distribution [7]. Since the process capability index is one of the most useful tools for evaluating true product quality, it has attracted increasing attention from quality experts and academics as the market becomes more competitive and manufacturing enterprises set higher quality standards.

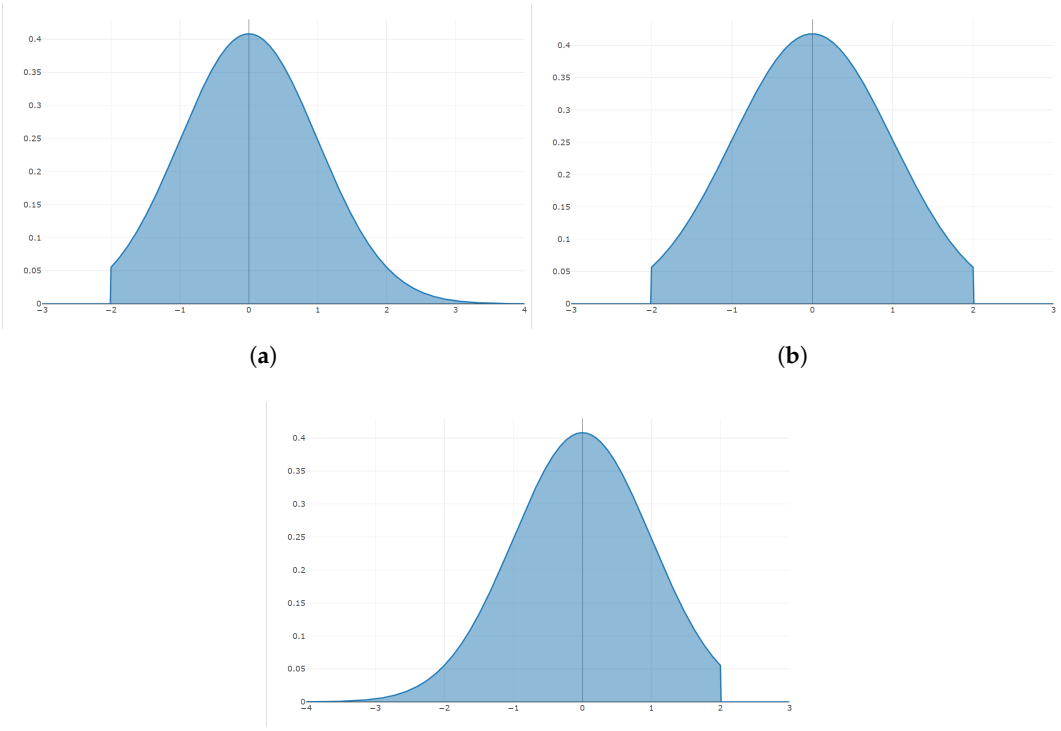
The earliest research on process capability indices for truncated data dates back to the end of the last century, when Alan et al. [8] used the Johnson transform method to transform truncated data into an approximate normal distribution before estimating process capability index values. This is a more stringent method on the data constraints, and the accuracy of the estimation results is low. Since the turn of the century, researchers have conducted numerous studies on the impact of truncated data on process capability. For instance, English & Taylor [9] initially investigated the effect of the exponentially truncated normal distribution on the accuracy of process capability evaluation using the Monte Carlo method while studying process capability robustness, while Pearn et al. [6] studied the accuracy of process capability for the double-truncated normal distribution under multi-parameter transformation conditions. One of the main issues in current research is the analysis of the process capability of truncated data of supply products. Some researchers attempt to derive the true distribution of the truncated samples using maximum likelihood estimation from a theoretical point of view, and then infer the true capability of the process based on the true distribution once more. Related studies are shown in [7,10].

Compared with the precision simulation and point estimation of truncated data, the interval estimation of the truncated sample process capability index has not made great breakthroughs because of its complicated distribution function. Based on a double-censored normal distribution, Yock [11] has made a preliminary study on the interval estimation of the process capability index, but has not formed a complete theory. However, researchers have used bootstrap methods to study the interval estimation of process capability indices for complex distributions and have achieved fruitful results, such as [12–14]. This provides a basis for interval estimation of the process capability index for truncated samples.

Comprehensive results of previous research on the process capability index for the truncated normal distribution are primarily focused on the double-truncated normal distribution, whereas actual quality management activities often involve quality data following a left-truncated normal distribution, such as the hardness and tensile strength of metal products, etc. To the best of our knowledge, the evaluation of the left-truncated normal distribution of quality data has not been reported in the literature. Therefore, this paper focuses on the study of the left-truncated normal distribution case for point estimation of process capability and constructs its bootstrap confidence interval. This paper is organized as follows: Section 1 reviews the research literature on the process capability of truncated distributions; Section 2 introduces the estimation methods of the expectation and variance of the left-truncated normal distribution; Section 3 establishes two process capability indices based on the left-truncated distribution; Section 4 presents the bootstrap interval estimation method for the proposed process capability indices; Section 6 describes the Monte Carlo simulation of the new and comparative methods; and Section 7 provides the conclusions.

2. Singly-Truncated Normal Distribution Data

Truncated samples are commonly seen in areas such as biomedical, aerospace, materials engineering, automotive manufacturing, and electrical and electronic engineering [15–18]. There are three types of truncated normal distributions—double-truncated, left-truncated, and right-truncated; see Figure 1.



(**a**) (**b**)

(**c**)

Figure 1. Truncation normal distribution: (**a**) Left-truncated; (**b**) double-truncated; (**c**) right-truncated.

2.1. Statistical Characteristics of Left-Truncated Normal Distribution Data

Let $f(x;\theta_1,\theta_2,\cdots,\theta_p)$ and $F(x;\theta_1,\theta_2,\cdots,\theta_p)$ represent the probability density function (pdf) and cumulative distribution function (cdf) of the unconstrained distribution with parameters $\theta_1,\theta_2,\cdots,\theta_p$, and abbreviate them to $f(x)$, $F(x)$, respectively. Then, the *pdf* of X, given the restriction that it is truncated at $x=a$, is

$$f(x)=\begin{cases}\frac{f(x)}{1-F(a)} & a\le x<\infty\\ 0 & otherwise\end{cases}. \tag{1}$$

Assuming that X is a quality characteristic and follows a normal distribution, we denote it as $N(\mu,\sigma)$. Then, the pdf of the truncated normal distribution is

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}(1-F(a))}e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \qquad a\le x<\infty \tag{2}$$

Let ξ be the standardized point of truncation.

$$\xi=\frac{a-\mu}{\sigma}, \tag{3}$$

Then, the k-th moment of the truncated distribution at the point of truncation is

$$\mu_k=\frac{1}{\sigma\sqrt{2\pi}(1-F(a))}\int_a^{\infty}(x-a)^k e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx. \tag{4}$$

Let $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ be the *pdf* of the standard normal distribution, and $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ be the *cdf* of the standard normal distribution. Then, as a function of ξ, $\mu_k(\xi)$ can be rewritten as

$$\mu_k(\xi) = \frac{\sigma^k}{1 - \Phi(\xi)} \int_{\xi}^{\infty} (t - \xi)^k \phi(t) dt. \tag{5}$$

Here, it can be easily verified that $\Phi(\xi) = F(a)$.

Let the k-th moment about 0 of the truncated standard normal distribution be defined as follows in relation to the original complete distribution:

$$\alpha_k(\xi) = \frac{1}{1 - \Phi(\xi)} \int_{\xi}^{\infty} t^k \phi(t) dt. \tag{6}$$

When $k = 1$ and then 2, from Equation (6), we can obtain

$$\alpha_1(\xi) = \frac{\phi(\xi)}{1 - \Phi(\xi)}, \qquad \alpha_2(\xi) = 1 + \frac{\xi \phi(\xi)}{1 - \Phi(\xi)} \tag{7}$$

For simplicity of the symbol, let

$$Q(\xi) = \frac{\phi(\xi)}{1 - \Phi(\xi)}. \tag{8}$$

Here, the inverse Mills ratio function of the standard normal distribution is also referred to in this context as $Q(\xi)$. Thus, Equation (7) follows

$$\alpha_1(\xi) = Q(\xi), \qquad \alpha_2(\xi) = 1 + \xi Q(\xi). \tag{9}$$

Let $k = 0, 1$, and 2 in Equation (5), we get

$$\begin{aligned} &\mu_0(\xi) = 1, \\ &\mu_1(\xi) = \sigma(Q(\xi) - \xi), \\ &\mu_2(\xi) = \sigma^2[1 - \xi(Q(\xi) - \xi)]. \end{aligned} \tag{10}$$

By combining Equations (5), (6), (9), and (10), the variance of the truncated samples X and the expected value of $X - a$ can be calculated as

$$\begin{aligned} &Var(X) = \mu_2(\xi) - (\mu_1(\xi))^2 = \sigma^2[1 - Q(\xi)(Q(\xi) - \xi)], \\ &E(X - a) = \sigma(Q(\xi) - \xi). \end{aligned} \tag{11}$$

2.2. Moment Estimators for Left-Truncated Normal Distribution Samples

Suppose $X = \{X_1, X_2, \cdots, X_n\}$ is a sample set from a truncated normal distribution with n observations. A fundamental assumption in this study is that the observations are independently and identically distributed (i.i.d.). From the definition of moment estimation, Cohen [19] introduced the moment estimators for singly-truncated normal distribution samples as

$$\bar{X} = a + \mu_1(\xi), \qquad S_n^2 = \mu_2(\xi) - (\mu_1(\xi))^2, \tag{12}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, $S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

Substituting Equation (12) into Equation (11), we have

$$\begin{aligned} &S_n^2 = \sigma^2[1 - Q(\xi)(Q(\xi) - \xi)], \\ &\bar{X} - a = \sigma(Q(\xi) - \xi). \end{aligned} \tag{13}$$

From Equation (3), μ follows

$$\mu = a - \sigma\xi. \tag{14}$$

To estimate μ in Equation (14), we should estimate σ and ξ first. Eliminating the two equations in Equation (13), we get

$$\frac{S_n^2}{(\bar{X}-a)^2} = \frac{1 - Q(\xi)(Q(\xi)-\xi)}{(Q(\xi)-\xi)^2} = \alpha(\xi), \tag{15}$$

and from the second equation of Equation (13), we have

$$\sigma = \frac{\bar{X}-a}{Q(\xi)-\xi}. \tag{16}$$

Finally, the estimators of σ^2 and μ can be derived as [19]:

$$\begin{aligned} \hat{\sigma}^2 &= S_n^2 + \hat{\theta}(\bar{X}-a)^2, \\ \hat{\mu} &= \bar{X} - \hat{\theta}(\bar{X}-a). \end{aligned} \tag{17}$$

where

$$\hat{\theta} = \frac{Q(\hat{\xi})}{Q(\hat{\xi})-\hat{\xi}}, \tag{18}$$

and $\hat{\xi}$ can be solved using Equation (15)

Estimating $\hat{\xi}$ by Equation (15) is a complicated task. A table of auxiliary functions for the estimator $\theta(\alpha)$ was presented in [19], which facilitates its use in the following application. If an equation-solving approach is employed, it can be implemented programmatically, as demonstrated in the pseudocode of Algorithm 1 below.

Algorithm 1 Calculate the $\hat{\xi}$ function

```
function F(s)
    Define Q(x) ← φ(x)/(1−Φ(x))                ▷ Inverse Mills ratio
    Define f1(x, s) ← (1−Q(x)(Q(x)−x))/(Q(x)−x)^2 − s
    r_root ← uniroot(f1, [−6, 6], s)           ▷ Find root in [−6, 6]
    ξ̂ ← Q(r_root)/(Q(r_root)−r_root)
    return ξ̂
end function
```

3. C_p and C_{pk} Indices Based on Singly-Truncated Normal Distribution Data

The process capability index is a key tool in statistical process control (SPC). It is a technique that is frequently used in modern industrial production to control and improve process capability. The process capability index provides a numerical quality standard for the ability of a process to produce a product that satisfies the factory's pre-defined quality requirements, allowing the production department to improve a less capable process and raise the quality level. In the practical process capability analysis, two things must be confirmed. First, the data must be in a control state, and second, they must be examined to see if the data follow a normal distribution. If the data follow the normal distribution, the famous C_p and C_{pk} indices can be used.

3.1. Classical C_p and C_{pk} Indices

The C_p index is suitable for cases where the process mean μ coincides with the target line, i.e., when the process does not shift. However, in actual manufacturing processes,

the process usually fluctuates, and the mean value may shift from the target line; thus, an adjusted index, namely C_{pk}, is used. Table 1 provides the process capability evaluation reference for different C_p and C_{pk} values.

Table 1. CPI evaluation criteria [20].

C_p	C_{pk}	Level	Evaluation Results
$C_p < 0.67$	$C_{pk} < 0.57$	V	Process capability is severely inadequate
$0.67 \leq C_p < 1.00$	$0.57 \leq C_{pk} < 0.92$	IV	Insufficient process capability
$1.00 \leq C_p < 1.33$	$0.92 \leq C_{pk} < 1.27$	III	Process capability is generally adequate
$1.33 \leq C_p < 1.67$	$1.27 \leq C_{pk} < 1.63$	II	Full process capability
$C_p \leq 1.67$	$C_{pk} \leq 1.63$	I	Excessive process capability

3.1.1. C_p Index

A C_p index was defined as

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma}. \tag{19}$$

where USL denotes upper specification limits, LSL denotes lower specification limits, and σ is the process standard deviation. The higher the C_p value, the more sufficient the process capability; conversely, the lower the C_p value, the less sufficient the process capability.

When a quality process is stationary and follows a Gaussian distribution, the estimator of the C_p index is derived as

$$\hat{C}_p = \frac{\text{USL} - \text{LSL}}{6S}, \tag{20}$$

where $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ is the standard deviation of the observations.

3.1.2. C_{pk} Index

The C_{pk} index was proposed to address the issue of process mean deviation from the target value, and it is defined as

$$C_{pk} = \min\left\{\frac{\text{USL} - \mu}{3\sigma}, \frac{\mu - \text{LSL}}{3\sigma}\right\}. \tag{21}$$

The estimator of C_{pk} is

$$\hat{C}_{pk} = \min\left\{\frac{\text{USL} - \mu}{3S}, \frac{\mu - \text{LSL}}{3S}\right\}. \tag{22}$$

3.2. Modified C_p and C_{pk} Indices

Assume that X_T denotes a set of quality characteristics with n observations drawn from a stationary process and following a truncated normal distribution. Then, the original expectation μ_T and σ_T^2 can be estimated using Equation (17):

$$\begin{aligned} \hat{\sigma}_T^2 &= S_T^2 + \hat{\theta}(\bar{X}_T - a)^2, \\ \hat{\mu}_T &= \bar{X}_T - \hat{\theta}(\bar{X}_T - a). \end{aligned} \tag{23}$$

where $\bar{X}_T = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_T^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_T)^2$, and $\hat{\theta}$ is calculated using Equation (18).

The modified C_p and C_{pk} indices based on the left-truncated normal distribution are defined as

$$\hat{C}_p^T = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}_T}, \tag{24}$$

and

$$\hat{C}_{pk}^{T} = \min\left\{\frac{\mathrm{USL}-\mu}{3\hat{\sigma}_T}, \frac{\mu-\mathrm{LSL}}{3\hat{\sigma}_T}\right\}. \tag{25}$$

where $\hat{\sigma}_T = \sqrt{S_T^2 + \hat{\theta}(\bar{X}_T - a)^2}$.

4. Confidence Interval Estimation of the Modified C_p and C_{pk} Indices

The modified C_p and C_{pk} indices may have significant inaccuracy due to the randomness of the sample collection; thus, the confidence interval of the two modified indices should also be investigated. It is very complicated to derive the distribution of $\hat{C}_p^T$ and $\hat{C}_{pk}^T$ indices directly, but Guevara [21] proposed a simpler Monte Carlo approach for computing the process capability indices' confidence intervals.

The bootstrap method was first proposed by Efron. Due to its simplicity and ease of use, among other things, it is frequently used in statistical inference and in estimating the distributions of unknown statistics. It is a traditional computer-intensive nonparametric method that is highly dependent on computers. It does not require knowing the sample distribution in advance. The fundamental principle of the bootstrap method is to create a new sample set by randomly selecting an equal number of samples from the observed sample set, and then computing the statistics for the new sample set. The aforementioned procedure is repeated a total of B times, resulting in a cluster of statistics. From the cluster, the empirical distribution of the statistics can be obtained, and point estimation, interval estimation, and hypothesis testing can be conducted using this distribution. The fundamental tenets of the bootstrap approach can be summed up as Table 2:

Table 2. Bootstrap parameter estimation process

Theoretical Distribution		Experience Distribution		Bootstrap Sample		Bootstrap Parameter Estimation
F	$\rightarrow$	$\hat{F} = \{X_i\}$	$\rightarrow$	$X_i^{(1)}$	$\rightarrow$	$\hat{\theta}^{(1)}$
				$X_i^{(2)}$	$\rightarrow$	$\hat{\theta}^{(2)}$
				$X_i^{(3)}$	$\rightarrow$	$\hat{\theta}^{(3)}$
				$\vdots$		$\vdots$
				$X_i^{(B)}$	$\rightarrow$	$\hat{\theta}^{(B)}$

4.1. Bootstrap Confidence Interval of C_p and C_{pk} Indices

Assuming that $X = \{X_1, X_2, \cdots, X_n\}$ is the quality set with n observations, from which n samples are drawn back a total of B times, and $X^{(i)} = \{X_1^{(i)}, X_2^{(i)}, \cdots, X_n^{(i)}\}, i = 1, 2, \cdots, B$, denotes the set of samples drawn for the i-th time. The statistic $\theta^{(i)}, i = 1, 2, \cdots, B$, for each sample is the point estimate of the statistic θ; thus, the estimated standard deviation of the statistic θ is s. The bootstrap confidence intervals of the statistic $\hat{\theta}$ are defined as follows: The standard deviation of the estimate is given by $\hat{\theta} = \bar{\theta} = \frac{1}{B}\sum_{i=1}^{B}\theta^{(i)}$; therefore, the bootstrap confidence intervals for the statistic $\hat{\theta}$ can be defined as in Equation (26).

$$\bar{\theta} \pm z_{\alpha/2}\hat{\sigma}_\theta, \tag{26}$$

where $\hat{\sigma}_\theta$ is the bootstrap variance of θ.

After resampling B times, we obtain two new datasets, $\hat{C}_p^T = \left\{ \left(\hat{C}_p^T\right)^{(1)}, \left(\hat{C}_p^T\right)^{(2)}, \cdots, \left(\hat{C}_p^T\right)^{(B)} \right\}$ and $\hat{C}_{pk}^T = \left\{ \left(\hat{C}_{pk}^T\right)^{(1)}, \left(\hat{C}_{pk}^T\right)^{(2)}, \cdots, \left(\hat{C}_{pk}^T\right)^{(B)} \right\}$; here, $(\hat{C}_p^T)^{(i)}$ is the estimate of i-th sampling, and so is $(C_{pk}^T)^{(i)}$.

$$\left(\hat{C}_p^T\right)^{(i)} = \frac{USL - LSL}{6\hat{\sigma}_T^{(i)}}, \tag{27}$$

and

$$\left(\hat{C}_{pk}^T\right)^{(i)} = \min\left\{ \frac{USL - \mu}{3\hat{\sigma}_T^{(i)}}, \frac{\mu - LSL}{3\hat{\sigma}_T^{(i)}} \right\}, \tag{28}$$

where $\hat{\sigma}_T^{(i)}$ is the variance estimation of the i-th sampling set.

Thus, the confidence intervals of $\hat{C}_p^T$ and $\hat{C}_{pk}^T$ are defined as

$$\begin{aligned} \hat{C}_p^T: &\quad \left[\bar{C}_p^T - z_{\alpha/2}V\left(\hat{C}_p^T\right), \bar{C}_p^T + z_{\alpha/2}V\left(\hat{C}_p^T\right)\right], \\ \hat{C}_{pk}^T: &\quad \left[\bar{C}_{pk}^T - z_{\alpha/2}V\left(\hat{C}_{pk}^T\right), \bar{C}_{pk}^T + z_{\alpha/2}V\left(\hat{C}_{pk}^T\right)\right]. \end{aligned} \tag{29}$$

where $z_{\alpha/2}$ is the α-quantile of the standard normal distribution, $\bar{C}_p^T$ and $\bar{C}_{pk}^T$ are the means of $(\hat{C}_p^T)^{(i)}$ and $(\hat{C}_{pk}^T)^{(i)}$. Because both $\hat{C}_p^T$ and $\hat{C}_{pk}^T$ asymptotically follow a normal distribution, their variances can be estimated using traditional variance estimation methods, as

$$\begin{aligned} V\left(\hat{C}_p^T\right) &= \frac{1}{B-1}\sum_{i=1}^{B}\left(\left(\hat{C}_p^T\right)^{(i)} - \bar{C}_p^T\right)^2, \\ V\left(\hat{C}_{pk}^T\right) &= \frac{1}{B-1}\sum_{i=1}^{B}\left(\left(\hat{C}_{pk}^T\right)^{(i)} - \bar{C}_{pk}^T\right)^2. \end{aligned} \tag{30}$$

4.2. Bootstrap Confidence Interval Estimation Procedure

To obtain the confidence interval of the investigated indices, the bootstrap method is used, and the procedure is simply presented as follows:

Step 1. Collect the quality data for a certain character;

Step 2. Use a histogram to check whether the data follow a left-truncated normal distribution; otherwise, recollect the quality data or estimate process capability using the traditional estimator.

Step 3. When the observations follow a left-truncated normal distribution, calculate $\hat{\sigma}_T^2$ and $\hat{\mu}_T$ using Equation (23);

Step 4. Calculate the confidence interval for the process capability index based on the following scenarios: (1) normal distribution: calculate $\hat{C}_p$, $\hat{C}_{pk}$,$V(\hat{C}_p)$,$V\left(\hat{C}_{pk}\right)$, respectively, and then compute the process capability under the normal distribution as follows:

$$\left[\bar{C}_p - z_{\alpha/2}V(\hat{C}_p), \bar{C}_p + z_{\alpha/2}V(\hat{C}_p)\right], \tag{31}$$

and

$$\left[\bar{C}_{pk} - z_{\alpha/2}V\left(\hat{C}_{pk}\right), \bar{C}_{pk} + z_{\alpha/2}V\left(\hat{C}_{pk}\right)\right], \tag{32}$$

respectively; (2) truncated normal distribution: calculate $\bar{C}_p^T$, $\bar{C}_{pk}^T$,$V\left(\hat{C}_p^T\right)$,$V\left(\hat{C}_{pk}^T\right)$ using Equation (24), Equation (25), and Equation (26), respectively, and then the process capability under normal distribution is calculated as follows:

$$\left[\bar{C}_p^T - z_{\alpha/2}V\left(\hat{C}_p^T\right), \bar{C}_p^T + z_{\alpha/2}V\left(\hat{C}_p^T\right)\right], \tag{33}$$

and

$$\left[\bar{C}_{pk}^T - z_{\alpha/2}V\left(\hat{C}_{pk}^T\right), \bar{C}_{pk}^T + z_{\alpha/2}V\left(\hat{C}_{pk}^T\right)\right], \tag{34}$$

respectively.

The above bootstrap confidence interval calculation steps were summarized in Figure 2.

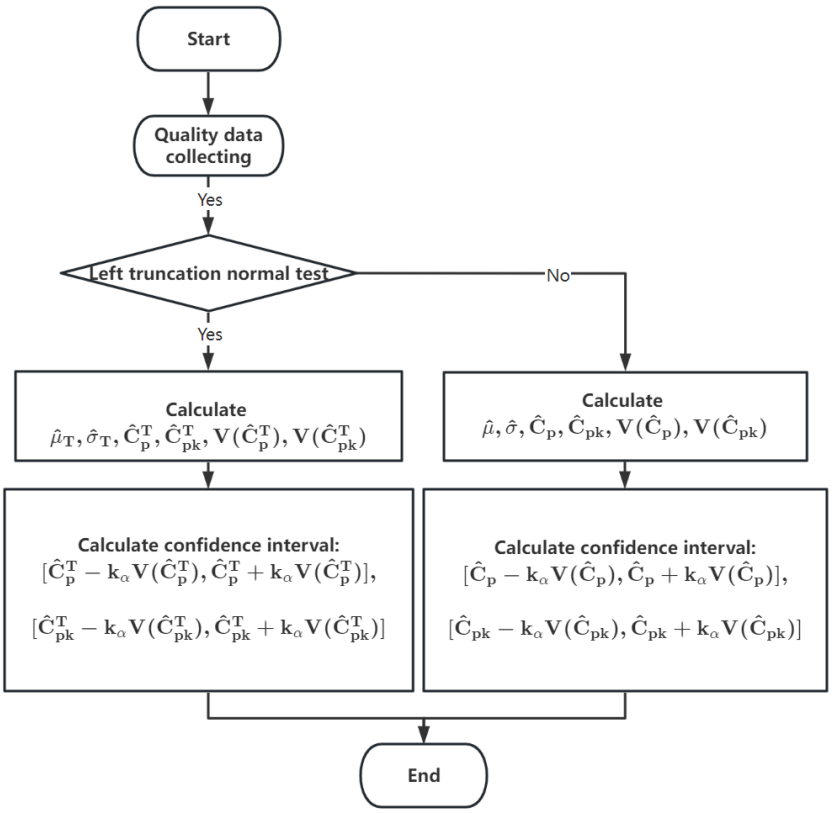


Figure 2. Bootstrap confidence interval estimation procedure.

4.3. Computational Complexity of the Bootstrap Procedure

The scalability of the bootstrap method with respect to sample size (n) and resampling iterations (B) can be characterized as follows:

(1) Computational complexity with respect to sample size (n).

For each bootstrap iteration, the core operations include the following:

Resampling with replacement: Generate a bootstrap sample of size n from the original data, which requires $O(n)$ time (linear in sample size).

Statistic recalculation: Compute the modified indices for the resampled data. Since the calculation of these indices involves the mean, standard deviation, and quantile estimates—all of which are $O(n)$ operations—the total complexity per iteration is dominated by O(n).

(2) Computational complexity with respect to resampling iterations (B).

The bootstrap procedure repeats the above resampling and statistic-calculation steps B times. Thus, the overall computational complexity scales linearly with B, resulting in a total time complexity of $O(B \times n)$ for the entire bootstrap process.

5. Simulations and Numerical Analysis

5.1. Performance of $\hat{C}_p^T$ and $\hat{C}_{pk}^T$ Indices

The following simulation experiments were conducted on the new indices $\hat{C}_p^T$ and $\hat{C}_{pk}^T$, and the results were compared with the true process capability values C_p and C_{pk} in order to determine whether the new indices based on truncated sample theory can evaluate the process capability of the observed samples more accurately. The comparison also includes the $\hat{C}_p$ and $\hat{C}_{pk}$ indices without taking truncation information into account, in order to more clearly show the efficacy of the suggested approach. Here, $\hat{C}_p^T$ and $\hat{C}_{pk}^T$ are calculated in Formulas (24) and (25); C_p and C_{pk} are defined in Formulas (19) and (21); and the estimating formulas for $\hat{C}_p$ and $\hat{C}_{pk}$ are presented in Equations (20) and (22).

The comparison experiment was conducted from three aspects. First, by controlling the number of samples and the expectation and variance of the sample as a whole, and varying the position of the left-truncated point, we explored the impact of different cutoff point positions on the estimation of C_p and C_{pk}. Second, by controlling the position of the left-truncated point and the expectation and variance of the sample as a whole, and letting the number of samples grow from small to large, we studied the impact of different sample sizes on the accuracy of the estimation results.

Since any normal distribution can be converted to a normal distribution, without loss of generality, the first two sets of simulation experiments use standard normal distribution observations, i.e., the population $X \sim N(0,1)$. The upper and lower specification lines are designed as $USL = 3$, $LSL = -3$, respectively. Thus, the true process capability value of the sample is as follows:

$$\begin{aligned} C_p &= \frac{USL - LSL}{6\sigma} = \frac{3-(-3)}{6\times 1} = 1, \\ C_{pk} &= \min\left\{\frac{USL-\mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right\} = \min\left\{\frac{3-0}{3\times 1}, \frac{0-(-3)}{3\times 1}\right\} = 1. \end{aligned} \tag{35}$$

All experiments were conducted on a PC using R software. All simulations were performed 10,000 times, with the average value being used as the final result in order to make the experimental results representative, given the random nature of sample collection.

Experiment 1. Set the observation sample size as $n = 100$, and let the value of the truncation point a change from -3 to 3 in intervals of 0.1. Then, generate n random truncated samples using the rnormTrunc() function in the EnvStats package with parameter a as the truncation point, and estimate the values of $\hat{C}_p$, $\hat{C}_{pk}$, $\hat{C}_p^T$, and $\hat{C}_{pk}^T$ using Equations (20), (22), (24), and (25), respectively. The results are shown in Figures 3 and 4.

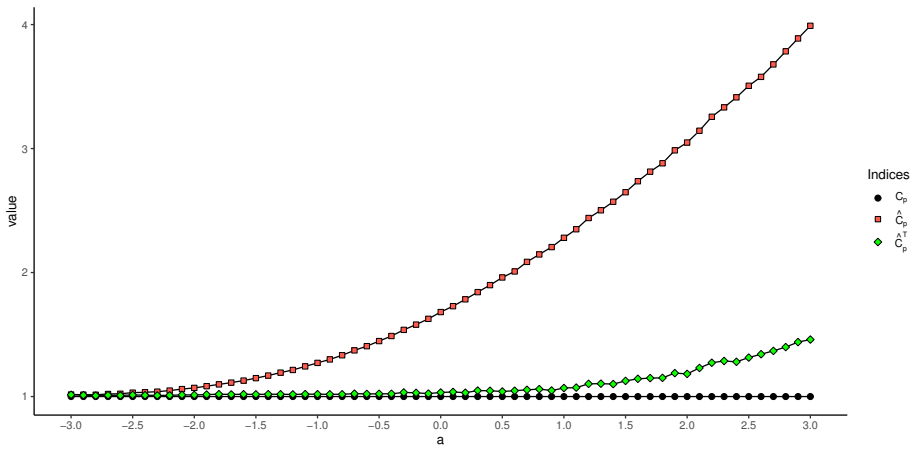


Figure 3. Values of C_p, $\hat{C}_p$, and $\hat{C}_p^T$ based on different truncation values.

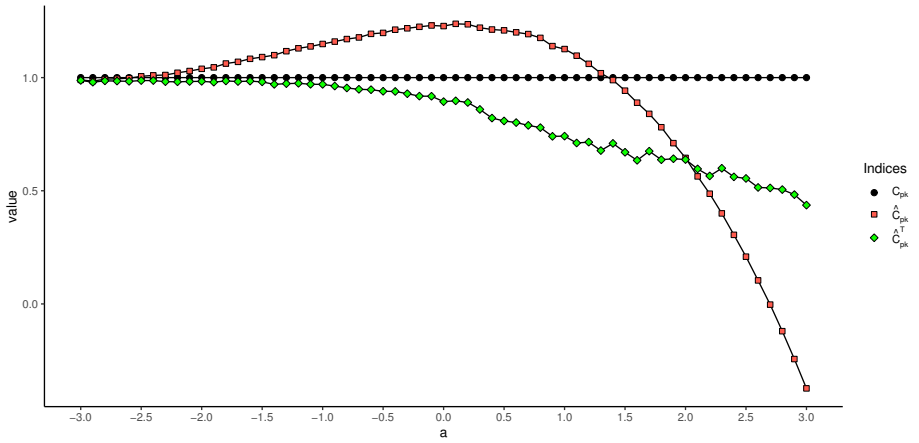


Figure 4. Values of C_{pK}, $\hat{C}_{pK}$, and $\hat{C}^T_{pK}$ based on different truncation values.

As can be seen in Figures 3 and 4, the difference between $\hat{C}_p$ and C_p, as well as the difference between $\hat{C}_{pk}$ and C_{pk}, increases with the growing value of the truncation point, i.e., if the truncation information of the sample is not taken into account when calculating the process capability index for truncated samples, the final estimated result will be higher than the true value. For the newly proposed estimation method, the results of $\hat{C}^T_p$ and $\hat{C}^T_p$ slowly increase and drop, respectively. Nevertheless, both $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ outperform $\hat{C}_p$ and $\hat{C}_{pk}$. Moreover, on the left side of the symmetry axis $x = 0$, the difference between the estimates of $\hat{C}_p$ and $\hat{C}_{pk}$ and the true values of C_p and C_{pk} is very small, which implies that the proposed two process capability indices have a good chance of being applied when the truncation point is on the left side of the symmetry axis.

Experiment 2. Another issue we are concerned about is how the accuracy of the proposed index estimation varies with changes in the sample size. Understanding this issue will assist users in determining the appropriate sample size for practical applications. In this experiment, we compare the results of different estimations with the true values. To assess the sensitivity of the proposed method to sample size, the number of samples was increased progressively from 50 to 1000, and the truncation point was set to -2, -1, 0, 1, and 2, respectively. The simulated results are shown in Tables 3 and 4.

Tables 3 and 4 show that, when the cutoff value is controlled, the estimation results for $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ converge to the true value $C_P = 1$, whereas $\hat{C}_p$ and $\hat{C}_{pk}$ also converge to the true value with the increase in the sample size. However, there is always a significant difference between the estimation results $\hat{C}_p$ and $\hat{C}_{pk}$, and the true values of C_p and C_{pk}. This phenomenon can be further seen in Figures 5 and 6. When the truncation value is set to $a = -1$, as can be seen, the $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ indices, which account for truncation information, outperform $\hat{C}_p$ and $\hat{C}_{pk}$ in terms of estimating effect. Additionally, Tables 3 and 4 show that when the truncated value increases, the estimation results for the C_p index deteriorate; when the sample size is greater than 100, the estimation results for the newly proposed $\hat{C}^T_{pk}$ index are much closer to the true value. In addition, it can also be seen from Tables 3 and 4 that, for the estimation of the C_p index, the results of the $\hat{C}_p$ index estimates become worse as the truncation value increases. However, the proposed $\hat{C}^T_p$ index does not differ much from the true values when the sample size is greater than 100. For the performance of the C_{pk} index estimator, the performance of $\hat{C}^T_{pk}$ still outperforms that of $\hat{C}_{pk}$.

Table 3. $\hat{C}_p$ and $\hat{C}_p^T$ values under different sample sizes and truncation points.

n	C_p	$a=-2$		$a=-1$		$a=0$		$a=1$		$a=2$	
		$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$
50	1.000	1.035	0.982	1.159	0.962	1.248	0.884	1.152	0.709	1.143	0.739
100		1.068	1.011	1.270	1.015	1.682	1.028	2.287	1.085	2.273	1.047
150		1.064	1.005	1.269	1.017	1.661	1.007	2.258	1.038	2.277	1.049
200		1.067	1.008	1.266	1.008	1.667	1.011	2.260	1.023	2.252	1.017
250		1.063	1.003	1.262	1.004	1.669	1.016	2.256	1.018	2.252	1.012
300		1.064	1.004	1.263	1.004	1.665	1.010	2.257	1.024	2.260	1.029
350		1.064	1.003	1.263	1.005	1.666	1.012	2.259	1.015	2.247	1.008
400		1.063	1.003	1.262	1.004	1.663	1.008	2.246	1.010	2.248	1.006
450		1.064	1.003	1.260	1.001	1.666	1.014	2.252	1.012	2.246	1.007
500		1.063	1.002	1.262	1.003	1.664	1.006	2.248	1.008	2.250	1.009
550		1.064	1.002	1.260	1.001	1.662	1.007	2.248	1.005	2.251	1.009
600		1.064	1.002	1.263	1.004	1.661	1.004	2.242	1.002	2.244	1.007
650		1.063	1.002	1.260	1.001	1.659	0.998	2.248	1.012	2.247	1.009
700		1.064	1.002	1.260	1.000	1.660	1.003	2.246	1.007	2.245	1.010
750		1.063	1.002	1.262	1.003	1.660	1.003	2.245	1.004	2.246	1.010
800		1.064	1.002	1.261	1.001	1.663	1.005	2.239	0.996	2.247	1.008
850		1.064	1.002	1.262	1.001	1.661	1.003	2.245	1.004	2.245	1.006
900		1.063	1.001	1.260	1.001	1.661	1.003	2.243	1.001	2.247	1.006
950		1.064	1.001	1.261	1.001	1.662	1.003	2.247	1.001	2.246	1.006
1000		1.062	1.000	1.261	1.001	1.660	1.000	2.243	1.000	2.246	1.002

Table 4. $\hat{C}_{pk}$ and $\hat{C}_{pk}^T$ values under different sample sizes and truncation points.

n	C_p	$a=-2$		$a=-1$		$a=0$		$a=1$		$a=2$	
		$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$	$\hat{C}_p$	$\hat{C}_p^T$
50	1.000	1.035	0.982	1.159	0.962	1.248	0.884	1.152	0.709	1.143	0.739
100		1.036	0.982	1.151	0.965	1.239	0.902	1.125	0.742	1.120	0.718
150		1.037	0.982	1.147	0.979	1.219	0.902	1.110	0.771	1.122	0.772
200		1.043	0.988	1.146	0.975	1.225	0.922	1.113	0.783	1.107	0.768
250		1.039	0.984	1.142	0.975	1.226	0.938	1.111	0.806	1.108	0.786
300		1.042	0.987	1.143	0.976	1.222	0.939	1.111	0.828	1.112	0.839
350		1.041	0.988	1.142	0.979	1.223	0.946	1.113	0.830	1.105	0.827
400		1.042	0.988	1.142	0.980	1.221	0.944	1.104	0.840	1.106	0.833
450		1.044	0.989	1.140	0.978	1.223	0.957	1.108	0.859	1.105	0.842
500		1.042	0.989	1.140	0.982	1.223	0.948	1.106	0.863	1.107	0.868
550		1.043	0.990	1.139	0.981	1.220	0.952	1.106	0.858	1.108	0.867
600		1.042	0.990	1.142	0.985	1.220	0.952	1.102	0.862	1.103	0.873
650		1.043	0.991	1.139	0.981	1.218	0.948	1.105	0.879	1.105	0.874
700		1.043	0.991	1.140	0.982	1.218	0.955	1.104	0.886	1.104	0.883
750		1.044	0.991	1.141	0.985	1.218	0.957	1.104	0.882	1.104	0.894
800		1.043	0.992	1.141	0.984	1.221	0.961	1.101	0.879	1.105	0.891
850		1.044	0.993	1.141	0.985	1.219	0.960	1.104	0.891	1.104	0.894
900		1.043	0.992	1.139	0.984	1.220	0.961	1.103	0.891	1.105	0.901
950		1.044	0.993	1.140	0.986	1.221	0.961	1.105	0.905	1.105	0.897
1000		1.042	0.991	1.140	0.986	1.219	0.961	1.103	0.893	1.105	0.897

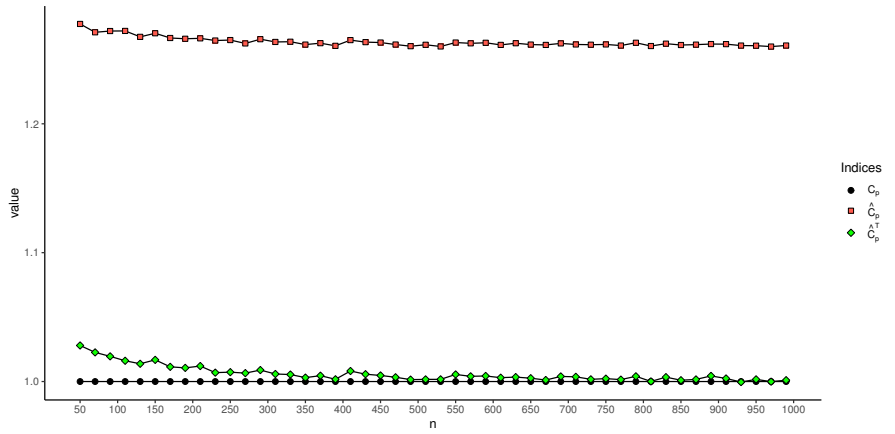


Figure 5. Values of C_p, $\hat{C}_p$, $\hat{C}_p^T$ based on different sample sizes with truncation value $a=-1$.

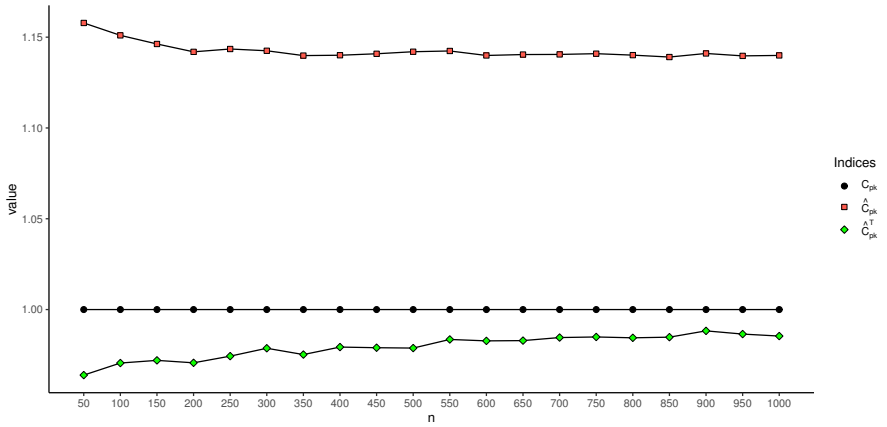


Figure 6. Values of C_{pK}, $\hat{C}_{pK}$, and $\hat{C}^T_{pK}$ based on different sample sizes with truncation value $a = -1$.

Experiment 3. To further investigate the absolute bias and root mean square error (RMSE) values of $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ under varying sample sizes, censoring rates, and mean shifts, Experiment 3 was designed. The absolute biases of the modified indices are presented in Figures 7 and 8, while the RMSE results are shown in Figures 9 and 10.

Analysis of the data in Figures 7 and 8 reveals certain patterns in the absolute biases of $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ under different sample sizes, censoring levels (indicated by the censoring point), and mean shifts. When the sample size is 50, under various censoring points, the absolute biases of both $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ generally decrease as the mean shift increases. Moreover, a larger censoring point (i.e., a lower degree of censoring) corresponds to greater absolute biases under the same mean shift. Similar trends are observed for a sample size of 100. Compared to the case with a sample size of 50, the absolute biases tend to decrease in some scenarios. When the sample size increases to 200, these patterns become more pronounced, and the absolute biases are further reduced in certain cases compared to those with a sample size of 100.

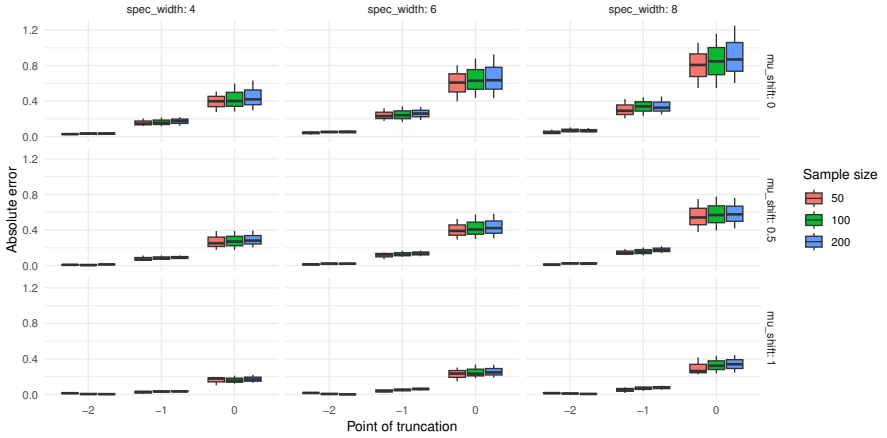


Figure 7. The absolute deviation of $\hat{C}^T_p$.

Overall, as the sample size increases, the absolute biases of $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ tend to decrease under the same censoring level and mean shift. For a fixed sample size, a lower degree of censoring (i.e., larger censoring point) generally leads to higher absolute bias. In addition, as the mean shift increases, the absolute biases of $\hat{C}^T_p$ and $\hat{C}^T_{pk}$ mostly exhibit a declining trend. These findings indicate that sample size, censoring level, and mean shift all influence the absolute biases of $\hat{C}^T_p$ and $\hat{C}^T_{pk}$. In practical applications, these patterns can be utilized to optimize relevant operations or improve prediction accuracy.

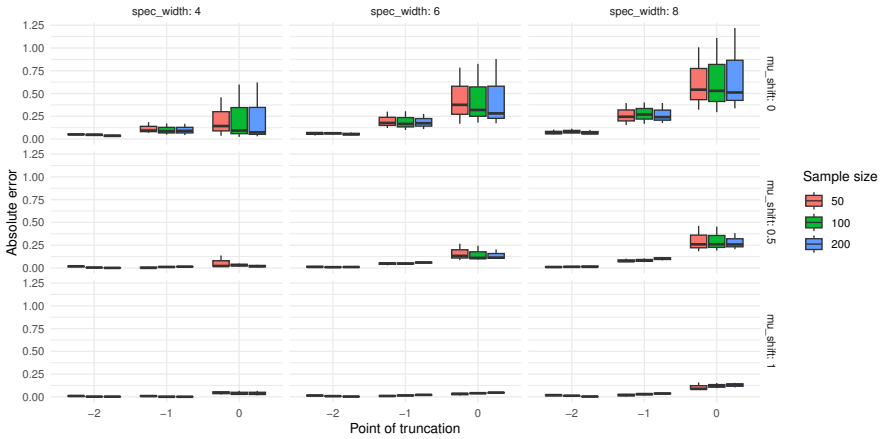


Figure 8. The absolute deviation of $\hat{C}_{pk}^{T}$.

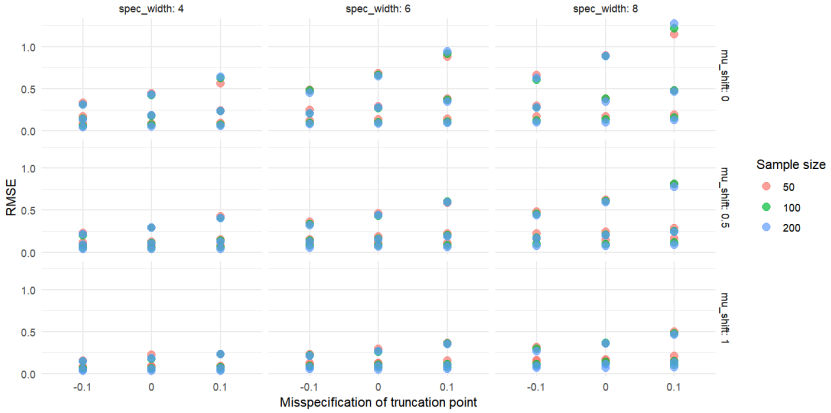


Figure 9. The RMSE of $\hat{C}_{p}^{T}$.

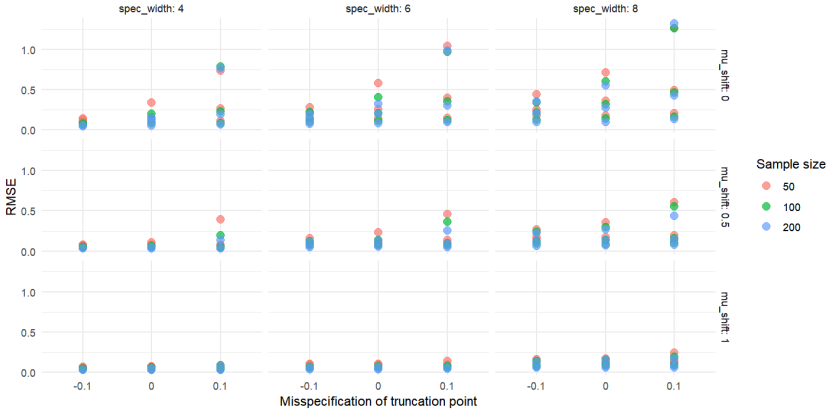


Figure 10. The RMSE of $\hat{C}_{pk}^{T}$.

As shown in Figures 9 and 10, from the perspective of sample size, when the censoring level and mean shift are fixed, the RMSE values of both $\hat{C}_{p}^{T}$ and $\hat{C}_{pk}^{T}$ generally exhibit a decreasing trend as the sample size increases from 50 to 200. For example, at a censoring point of -2 and a mean shift of 0, the RMSE of cpt decreases from 0.359639 (sample size 50) to 0.315073 (sample size 100), and further to 0.283292 (sample size 200); similarly, the RMSE of cpkt decreases from 0.375543 (sample size 50) to 0.329441 (sample size 100), and then to 0.287575 (sample size 200). These results indicate that larger sample sizes may contribute to reducing the RMSE values of $\hat{C}_{p}^{T}$ and $\hat{C}_{pk}^{T}$.

Regarding the censoring level, under the same sample size and mean shift, a larger censoring point (i.e., lower degree of censoring) tends to correspond to higher RMSE values

for both $\hat{C}_p^T$ and $\hat{C}_{pk}^T$. For instance, at a sample size of 100 and a mean shift of 0, as the censoring point changes from -2 to 0, the RMSE of cpt increases from 0.315073 to 0.808805, and that of cpkt rises from 0.329441 to 0.688743.

In terms of the effect of mean shift, when the sample size and censoring level are fixed, the RMSE values of $\hat{C}_p^T$ and $\hat{C}_{pk}^T$ mostly show a decreasing trend as the mean shift increases. Taking a sample size of 50 and a censoring point of -1 as an example, the RMSE of cpt decreases from 0.536742 (mean shift 0) to 0.363781 (mean shift 1), while the RMSE of cpkt drops from 0.520519 (mean shift 0) to 0.303010 (mean shift 1).

5.2. Bootstrap Confidence Interval for C_p and C_{pk} Indices

One of the hotspots that researchers and quality managers are concerned about is the impact of shortened samples on process capability. According to Section 4.1, the truncation position can be found in a variety of places, and the sample size or truncation value will affect how accurately the process capability indices are estimated. The following simulation studies examine the interval estimation of the C_p and C_{pk} indices under different parameter situations; the interval estimation of the C_p and C_{pk} indices with different sample sizes and different truncation point locations is simulated below. This is done in order to further analyze the effect of samples containing truncation information on the interval estimation of the process capability indices.

The sample size n for the simulations that follow was changed from 50, 100, and 200 to 500, and the truncation values $a = -3, -2, -1, 0, 1, 2, 3$ were selected with $USL = 3$ and $LSL = -3$. Using R software, random samples were created, the procedure was repeated 10,000 times, and the random seed was set to 123. The 95% SB confidence intervals of the C_p and C_{pk} indices calculated by the traditional and proposed methods are shown in Tables 5 and 6.

Table 5. Simulation results of the SB confidence interval for the C_p index under various process parameters with $\alpha = 0.05$ and true $C_p = 1$.

n	a	$V(\hat{C}_p)$	$V\left(\hat{C}_p^T\right)$	$L_{\hat{C}_p}$	$U_{\hat{C}_p}$	$L_{\hat{C}_p^T}$	$U_{\hat{C}_p^T}$
50	-3	0.104	0.110	1.014	1.026	1.016	1.030
	-2	0.110	0.133	1.074	1.088	1.022	1.038
	-1	0.135	0.185	1.278	1.294	1.031	1.054
	0	0.210	0.285	1.685	1.711	1.046	1.081
	1	0.326	0.426	2.305	2.345	1.135	1.188
	2	0.476	0.561	3.140	3.199	1.363	1.431
	3	0.625	0.743	4.092	4.169	1.712	1.805
100	-3	0.071	0.075	1.012	1.021	1.010	1.019
	-2	0.069	0.084	1.065	1.073	1.007	1.018
	-1	0.093	0.128	1.266	1.278	1.011	1.027
	0	0.149	0.211	1.668	1.687	1.020	1.045
	1	0.221	0.308	2.259	2.286	1.047	1.085
	2	0.319	0.423	3.056	3.095	1.188	1.241
	3	0.437	0.550	3.940	3.995	1.417	1.485
200	-3	0.050	0.052	1.009	1.015	1.005	1.011
	-2	0.050	0.061	1.064	1.070	1.004	1.011
	-1	0.066	0.089	1.263	1.270	1.006	1.016
	0	0.098	0.140	1.663	1.675	1.008	1.025
	1	0.158	0.232	2.251	2.270	1.009	1.039
	2	0.219	0.324	2.991	3.017	1.059	1.099
	3	0.301	0.400	3.847	3.885	1.207	1.256
500	-3	0.031	0.033	1.007	1.011	1.001	1.005
	-2	0.032	0.039	1.063	1.067	1.002	1.006
	-1	0.040	0.056	1.260	1.265	1.001	1.008
	0	0.065	0.092	1.660	1.668	1.001	1.012
	1	0.101	0.151	2.243	2.255	1.001	1.019
	2	0.154	0.233	2.964	2.983	1.007	1.036
	3	0.191	0.291	3.790	3.814	1.066	1.103

In Tables 5 and 6, $L_{\hat{\theta}}$ and $U_{\hat{\theta}}$ denote the lower and upper confidence intervals of the estimated parameter θ. As can be seen from Table 5, the upper and lower confidence intervals of both $\hat{C}_p$ and $\hat{C}_p^T$ are larger than the true value $C_p = 1$, which indicates that when the truncation information exists, both $\hat{C}_p$ and $\hat{C}_p^T$ are overestimated. But this overestimation is inversely proportional to the change in sample size. Moreover, the $\hat{C}_p^T$ index is overestimated to a much smaller extent than the $\hat{C}_p$ index, and this tendency is reflected more clearly with the rightward shift of the truncation point. For example, when $n = 50$ and $a = 3$, the confidence interval of $\hat{C}_p$ is $[1.014, 1.026]$, and the confidence interval of $\hat{C}_p^T$ is $[1.016, 1.030]$. At this point, the length of the confidence interval of $\hat{C}_p$ is 0.012, which is smaller than the length of the confidence interval of $\hat{C}_p^T$, 0.014, and it is closer to the true C_p value. However, when the sample size increases, the advantage of the $\hat{C}_p$ confidence interval estimation results is quickly lost. For example, when $n = 500$, $a = -3$, $\hat{C}_p^T$'s confidence interval $[1.001, 1.005]$ is obviously better than $\hat{C}_p$'s confidence interval $[1.007, 1.011]$, even though the lengths of their confidence intervals are the same.

If the sample is controlled so that the truncation position is gradually moved from -3 to 3, the confidence interval estimation advantage of the $\hat{C}_p^T$ index becomes more evident than that of the $\hat{C}_p$ index. For example, under the condition of a sample size of 50, when $a = 3$, the confidence interval of the $\hat{C}_p^T$ index, $[1.712, 1.805]$, is far better than that of the $\hat{C}_p$ index $[4.092, 4.169]$.

Table 6. Simulation results of the SB confidence interval for the C_{pk} index for various process parameters with $\alpha = 0.05$, true $C_{pk} = 1$.

n	a	$V(\hat{C}_{pk})$	$V(\hat{C}_{pk}^T)$	$L_{\hat{C}_{pk}}$	$U_{\hat{C}_{pk}}$	$L_{\hat{C}_{pk}^T}$	$U_{\hat{C}_{pk}^T}$
50	-3	0.102	0.108	0.975	0.988	0.977	0.990
	-2	0.110	0.134	1.034	1.048	0.979	0.996
	-1	0.151	0.202	1.157	1.176	0.959	0.984
	0	0.191	0.363	1.238	1.262	0.862	0.907
	1	0.200	0.515	1.132	1.157	0.702	0.765
	2	0.146	0.424	0.658	0.676	0.614	0.666
	3	0.034	0.346	-0.387	-0.383	0.326	0.369
100	-3	0.074	0.078	0.984	0.994	0.982	0.991
	-2	0.074	0.088	1.033	1.043	0.978	0.988
	-1	0.103	0.138	1.145	1.158	0.963	0.980
	0	0.134	0.261	1.224	1.241	0.885	0.917
	1	0.135	0.450	1.108	1.125	0.720	0.775
	2	0.097	0.459	0.639	0.651	0.582	0.637
	3	0.024	0.355	-0.375	-0.372	0.413	0.458
200	-3	0.051	0.053	0.990	0.996	0.985	0.992
	-2	0.055	0.062	1.038	1.045	0.983	0.991
	-1	0.075	0.095	1.141	1.150	0.970	0.982
	0	0.088	0.160	1.220	1.231	0.918	0.938
	1	0.095	0.355	1.107	1.119	0.762	0.806
	2	0.065	0.480	0.625	0.633	0.571	0.631
	3	0.016	0.378	-0.366	-0.364	0.472	0.520
500	-3	0.032	0.033	0.995	0.999	0.989	0.993
	-2	0.035	0.040	1.042	1.047	0.988	0.993
	-1	0.044	0.060	1.139	1.144	0.980	0.987
	0	0.058	0.108	1.219	1.226	0.942	0.955
	1	0.060	0.207	1.102	1.110	0.846	0.872
	2	0.045	0.403	0.619	0.625	0.658	0.708
	3	0.011	0.427	-0.360	-0.358	0.515	0.567

The same experiment becomes a little more complicated for the C_{pk} index. First of all, in the estimation of the C_{pk} index, $\hat{C}_{pk}^{T}$ will always be underestimated, while $\hat{C}_{pk}$ is overestimated when the truncation point is located on the left side of the symmetry axis. When the truncation point is moved to the right side of the symmetry axis, the estimation of the $\hat{C}_{pk}$ quickly becomes underestimated. This phenomenon can be seen in Figure 5 and Table 6.

Overall, the $\hat{C}_{pk}$ index estimate without considering truncation information may outperform the $\hat{C}_{pk}^{T}$ index only when the truncation point is located within $(1, 2)$. In the rest of the cases, the $\hat{C}_{pk}^{T}$ index, which considers sample truncation information, outperforms the $\hat{C}_{pk}$ index in confidence interval estimation.

Combining the results from Tables 5 and 6, it is clear that the proposed new indices $\hat{C}_{p}^{T}$ and $\hat{C}_{pk}^{T}$ outperform conventional estimation techniques in the majority of cases without taking into account sample truncation information in the confidence interval estimation. This shows that in the application of actual industrial production, we cannot ignore the existence of truncation information, especially for samples sent for testing in order to pass inspection. It is particularly noteworthy that both of the newly proposed indices show surprising accuracy when the truncation position of the sample is on the left side of the symmetry axis, for both point and interval estimation.

Building on the experiments above, we now discuss the bootstrap interval coverage of the true value under varying sample sizes, censoring rates, and mean shifts. The results are shown in Figures 5 and 6

As shown in Figures 11 and 12, from the perspective of sample size, when the censoring level and mean shift are held constant, the bootstrap interval coverage of the true value of both $\hat{C}_{p}^{T}$ and $\hat{C}_{pk}^{T}$ generally exhibit a decreasing trend as the sample size increases from 50 to 200. For instance, at a censoring point of -2 and a mean shift of 0, the bootstrap interval coverage of the true value of $\hat{C}_{p}^{T}$ decreases from 0.36 (sample size 50) to 0.32 (sample size 100), and further to 0.28 (sample size 200). Similarly, the bootstrap interval coverage of the true value of $\hat{C}_{pk}^{T}$ declines from 0.38 (sample size 50) to 0.33 (sample size 100), and then to 0.29 (sample size 200). These results suggest that larger sample sizes may help reduce the RMSE values of $\hat{C}_{p}^{T}$ and $\hat{C}_{pk}^{T}$.

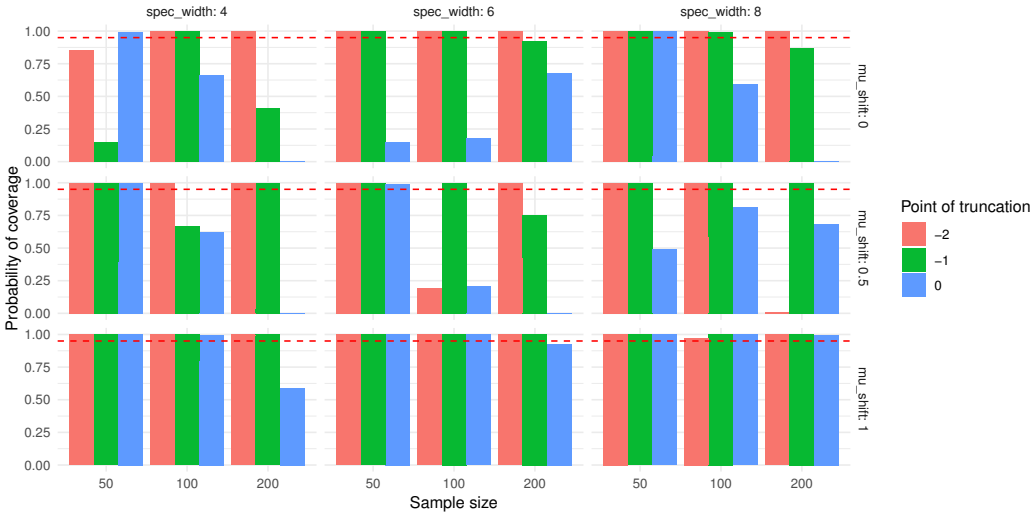


Figure 11. The bootstrap interval coverage of the true value of $\hat{C}_{p}^{T}$. The red dashed line means a probability of 0.95.

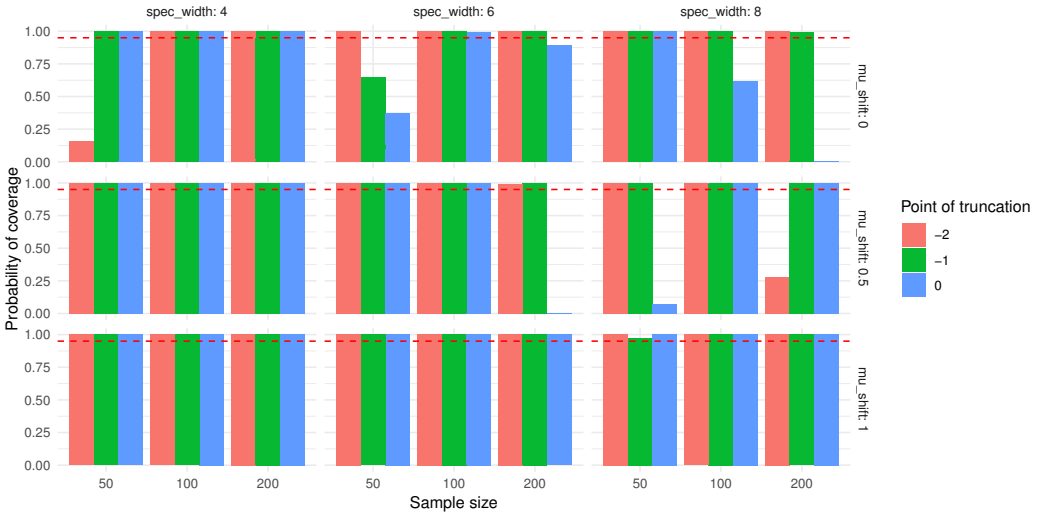


Figure 12. The bootstrap interval coverage of the true value of $\hat{C}_{pk}^{T}$.. The red dashed line means a probability of 0.95.

With respect to the censoring level, under the same sample size and mean shift, a larger censoring point (i.e., lower degree of censoring) tends to correspond to higher RMSE values for both $\hat{C}_{p}^{T}$ and $\hat{C}_{pk}^{T}$. For example, at a sample size of 100 and a mean shift of 0, as the censoring point increases from -2 to 0, the RMSE of cpt rises from 0.32 to 0.81, and that of $\hat{C}_{pk}^{T}$ increases from 0.33 to 0.69.

Regarding the effect of mean shift, when the sample size and censoring level are fixed, the RMSE values of cpt and $\hat{C}_{pk}^{T}$ mostly show a declining trend with increasing mean shift. Taking a sample size of 50 and a censoring point of -1 as an example, the RMSE of cpt decreases from 0.54 (mean shift 0) to 0.36 (mean shift 1), while the RMSE of $\hat{C}_{pk}^{T}$ drops from 0.52 (mean shift 0) to 0.30 (mean shift 1).

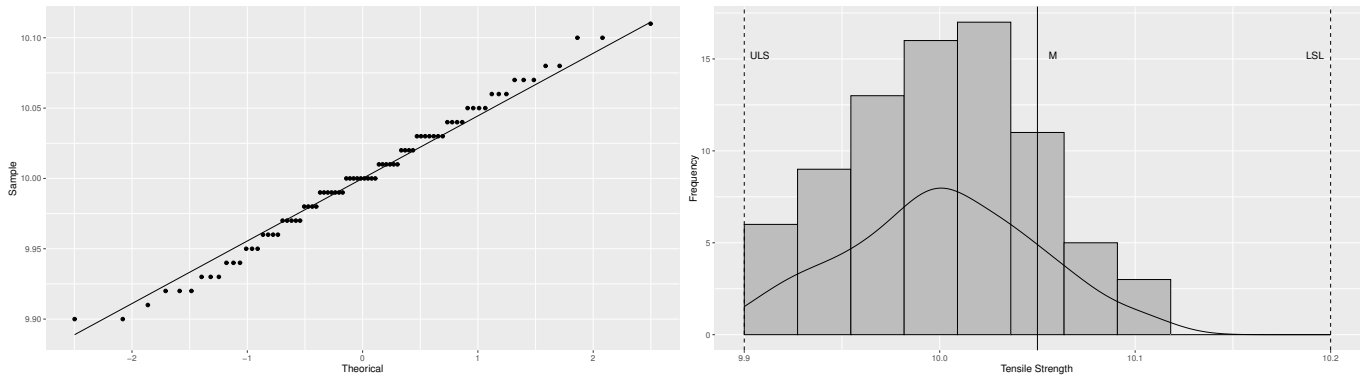
6. Numerical Example

To illustrate the proposed method, this study uses information from a supplier who provided a new type of insulating material to an electronics company in Changsha. A central concern of the electronics company was the tensile strength of the flame-retardant material, which was required to be within the range of 10.05 ± 0.15. The test data provided by the supplier contained a total of 80 samples, and the characteristics of the tensile strength values are shown in Table 7.

According to the manufacturer's demands, we know that $USL = 10.20$, $LSL = 9.90$, and the target value is $M = 10.05$. Firstly, the Q-Q plot method is used to test the data for normal distribution, and the test results are shown in Figure 13a. From the test results in Figure 13a, it can be seen that the data are basically distributed around the qq-line, indicating that the test data follow an approximately normal distribution.

Table 7. Tensile strength data for an electronics company from a supplier.

10.07	10.03	9.97	10.06	9.99	10.00	9.96	10.08	10.01	9.95	10.07	10.01
9.96	9.98	9.93	10.03	10.00	10.03	10.00	10.01	9.94	10.10	10.04	10.10
10.05	9.99	9.90	10.06	9.92	10.04	9.97	9.99	10.03	10.11	10.07	10.05
10.03	9.94	10.08	10.00	10.00	10.00	10.01	10.06	9.96	10.05	9.92	9.96
10.00	10.00	9.90	10.02	10.05	10.00	9.93	9.97	9.93	9.91	9.99	9.99
9.99	10.01	10.03	10.03	10.02	9.92	10.02	9.95	9.97	10.04	9.98	9.94
9.95	9.97	9.98	10.02	10.04	9.98	9.99	10.01				



(**a**) (**b**)

Figure 13. Test for dataset: (**a**) Q-Q plot of the test data. (**b**) Histogram of the test data.

We create a histogram of the quality data in order to better examine the distribution pattern. The data for this quality characteristic clearly follow a left-truncated normal distribution, as shown in Figure 13. The truncation value was empirically set to 9.90 based on the inspection process.

6.1. Hypotheses Text for the Truncation Threshold

- Null hypothesis H_0: Data are not left-truncated (complete data).
- Alternative hypothesis H_1: Data are left-truncated.

6.2. Test Statistic

Let $c = \min\{X_i\}$ be the observed minimum value. Under H_0, estimate distribution parameters:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{(sample mean)} \tag{36}$$

$$\hat{\sigma} = S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad \text{(sample standard deviation)} \tag{37}$$

Calculate the probability below c:

$$p = P(X < c \mid \hat{\mu}, \hat{\sigma}) = \Phi\left(\frac{c - \hat{\mu}}{\hat{\sigma}}\right) \tag{38}$$

where Φ is the standard normal cumulative distribution function.

Let Y be the actual number of observations below c. Since c is the minimum value, $Y = 0$. Under H_0:

$$Y \sim \text{Binomial}(n, p) \tag{39}$$

Therefore,

$$P(Y = 0) = (1 - p)^n \tag{40}$$

When p is small and n is large, we can use the Poisson approximation:

$$Y \sim \text{Poisson}(\lambda), \quad \lambda = np \tag{41}$$

$$P(Y = 0) \approx e^{-\lambda} \tag{42}$$

6.3. Decision Rule

Given the significance level $\alpha = 0.05$:

$$\text{If } P(Y = 0) < \alpha \Rightarrow \text{reject } H_0 \text{ (evidence for left truncation)} \tag{43}$$

$$\text{Otherwise} \Rightarrow \text{fail to reject } H_0 \tag{44}$$

If H_0 is rejected, estimate the truncation point as $c = \min\{X_i\}$.

6.4. Test Procedure

1. Calculate sample statistics: $n, c = \min\{X_i\}, \hat{\mu}, \hat{\sigma}$.
2. Compute z-score: $z = \dfrac{c - \hat{\mu}}{\hat{\sigma}}$.
3. Calculate $p = \Phi(z)$.
4. Compute the binomial test p-value: $P_{\text{binom}} = (1 - p)^n$.
5. Compute the Poisson approximation p-value: $P_{\text{pois}} = e^{-np}$.
6. Compare the p-value with α and make a decision.

6.5. Result

For the given tensile strength data ($n = 80$), let $c = 9.90$, following the truncation threshold test processing, we get $P_{\text{pois}} = 0.016$ Since $P_{\text{binom}} < 0.05$, we accept H_0, and suggest the data follow a left-truncated distribution.

6.6. Results

Point and interval estimation of the process capability index for this quality process is performed below. First, the expectation and variance of the quality data are calculated using the conventional method without considering the truncation information and the newly proposed method, respectively, and the results are obtained as follows:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 10.0005,$$
$$\hat{\mu}_T = \bar{X} - \hat{\theta}(\bar{X} - a) = 9.996.$$

Thus, the corresponding standard deviations are as follows:

$$\hat{\sigma} = \bar{X} = S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = 0.0401,$$
$$\hat{\sigma}_T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \hat{\theta}(\bar{X} - a)^2 = 0.0526.$$

The process capability index values calculated by conventional and modified methods can be obtained as follows:

$$\hat{C}_p = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}} = 1.022,$$
$$\hat{C}_p^T = \frac{\text{USL} - \text{LSL}}{6\hat{\sigma}_T} = 0.9509,$$
$$\hat{C}_{pk} = \min\left\{\frac{\text{USL} - \hat{\mu}}{3\hat{\sigma}}, \frac{\hat{\mu} - \text{LSL}}{3\hat{\sigma}}\right\} = 0.6853,$$
$$\hat{C}_{pk}^T = \min\left\{\frac{\text{USL} - \hat{\mu}_T}{3\hat{\sigma}_T}, \frac{\hat{\mu}_T - \text{LSL}}{3\hat{\sigma}_T}\right\} = 0.6115.$$

According to Table 1, it can be seen that the value of process capability C_p calculated using the conventional method belongs to the III level, i.e., the process capability is realized as general, which means that the technical management capability of this product is more reluctant and should be improved to the level II, while the value of the C_p^T index of the revised method is 0.9509, which belongs to the IV level, which represents the insufficiency of the process capability, and should be diagnosed for the production process, and necessary measures should be taken to improve production. In terms of the performance of the C_{pk} index, the results calculated by both the conventional and revised methods show that the process capability of the product is seriously inadequate, and that the production line needs to be shut down and reorganized.

Based on the above conclusions, we constructed the corresponding confidence intervals for the process capability index to analyze the reliability of the conclusions. The 95% confidence intervals were calculated using the conventional method described in Equations (31) and (32) as follows:

$$\begin{aligned}&\left[\bar{C}_p - z_{\alpha/2}V\left(\hat{C}_p\right), \bar{C}_p + z_{\alpha/2}V\left(\hat{C}_p\right)\right] = [1.036, 1.039],\\&\left[\bar{C}_{pk} - z_{\alpha/2}V\left(\hat{C}_{pk}\right), \bar{C}_{pk} + z_{\alpha/2}V\left(\hat{C}_{pk}\right)\right] = [0.695, 0.697].\end{aligned}$$

The 95% confidence intervals were calculated using the modified method described in Equations (33) and (34) as follows:

$$\begin{aligned}&\left[\bar{C}_p^T - z_{\alpha/2}V\left(\hat{C}_p^T\right), \bar{C}_p^T + z_{\alpha/2}V\left(\hat{C}_p^T\right)\right] = [0.964, 0.967],\\&\left[\bar{C}_{pk}^T - z_{\alpha/2}V\left(\hat{C}_{pk}^T\right), \bar{C}_{pk}^T + z_{\alpha/2}V\left(\hat{C}_{pk}^T\right)\right] = [0.619, 0.623].\end{aligned}$$

According to the results of the above confidence interval estimation, the traditional method's C_p index result always indicates that it is generally adequate, while the improved method indicates that the process capability is insufficient, which feeds back to the actual activity, i.e., according to the result of the traditional C_p index, the power company can accept the product, while according to the new method, the company should reject the product.

7. Conclusions

This study modifies two process capability evaluation indices, C_p and C_{pk}, based on the truncation theory of statistical estimation for left-truncated samples, and then constructs the bootstrap confidence intervals for these two modified indices. To test the performance of the modified indices, we first performed Monte Carlo simulations of the point estimates of the proposed indices by setting different parameters and comparing them with the traditional estimation methods under the same conditions. The results demonstrate that the modified indices outperform the traditional methods in determining the process capability of left-truncated samples under various parameter conditions. Then, the accuracy and stability of the new technique are further validated by simulated tests using bootstrap confidence interval estimation.

Author Contributions: Conceptualization, Y.Y. and B.Y.; methodology, B.Y. and Y.Y.; software, B.Y. and P.L.; validation, B.Y., Y.Y., and P.L.; formal analysis, Y.Y.; resources, P.L.; data curation, B.Y.; writing—original draft preparation, B.Y.; writing—review and editing, Y.Y. and P.L.; visualization, B.Y.; supervision, P.L.; funding acquisition, Y.Y., B.Y., and P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (No. 12204540), the Project of the Social Science Popularization Base in Hunan Province (No. XJK22ZDJD35), and the Excellent Youth Project of the Hunan Provincial Education Department (No. 24B0865).

Data Availability Statement: The experimental datasets were randomly generated using R4.3.2 software, and the real datasets used in the test are available in the main text.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Borucka, A.; Kozłowski, E.; Antosz, K.; Parczewski, R. A New Approach to Production Process Capability Assessment for Non-Normal Data. *Appl. Sci.* **2023**, *13*, 6721. [CrossRef]
2. Orosz, Á.; Varbanov, P.S.; Klemeš, J.J.; Friedler, F. Process synthesis considering sustainability for both normal and non-normal operations: P-graph approach. *J. Clean. Prod.* **2023**, *414*, 137696.
3. Banihashemi, A.; Fallah Nezhad, M.S.; Amiri, A. Developing process-yield-based acceptance sampling plans for AR (1) auto-correlated process. *Commun. Stat.-Simul. Comput.* **2021**, *52*, 4230–4251. [CrossRef]
4. Song, S.; Bai, Z.; Wei, H.; Xiao, Y. Copula-based methods for global sensitivity analysis with correlated random variables and stochastic processes under incomplete probability information. *Aerosp. Sci. Technol.* **2022**, *129*, 107811. [CrossRef]
5. Chakraborty, A.K.; Chatterjee, M. *Handbook of Multivariate Process Capability Indices*; CRC Press: Boca Raton, FL, USA, 2021.
6. Pearn, W.L.; Hung, H.N.; Peng, N.F.; Huang, C.Y. Testing process precision for truncated normal distributions. *Microelectron. Reliab.* **2007**, *47*, 2275–2281. [CrossRef]
7. Yang, J.; Meng, F.; Huang, S.; Cui, Y. Process capability analysis for manufacturing processes based on the truncated data from supplier products. *Int. J. Prod. Res.* **2020**, *58*, 6235–6251. [CrossRef]
8. Polansky, A.M.; Chou, Y.M.; Mason, R.L. Estimating Process Capability Indices for a Trltncated Distribution. *Qual. Eng.* **1998**, *11*, 257–265. [CrossRef]
9. English, J.R.; Taylor, G.D. Process capability analysis—A robustness study. *Int. J. Prod. Res.* **1993**, *31*, 1621–1635. [CrossRef]
10. Khadse, K.G.; Khadse, A.K. Assessing supplier's process capability using truncated normal distribution data. *J. Univ. Shanghai Sci. Technol.* **2020**, *52*, 82–96.
11. Lai, Y.W.; Chew, E.P. Gauge capability assessment for high-yield manufacturing processes with truncated distribution. *Qual. Eng.* **2000**, *13*, 203–210. [CrossRef]
12. Tong, L.; Chen, J. Bootstrap confidence interval of the difference between two process capability indices. *Int. J. Adv. Manuf. Technol.* **2003**, *21*, 249–256. [CrossRef]
13. Park, C.; Dey, S.; Ouyang, L.; Byun, J.H.; Leeds, M. Improved bootstrap confidence intervals for the process capability index Cpk. *Commun. -Stat.-Simul. Comput.* **2020**, *49*, 2583–2603. [CrossRef]
14. Besseris, G.J. Evaluation of robust scale estimators for modified Weibull process capability indices and their bootstrap confidence intervals. *Comput. Ind. Eng.* **2018**, *128*, 135–149. [CrossRef]
15. Fields, E.; Osorio, C.; Zhou, T. A Data-Driven Method for Reconstructing a Distribution from a Truncated Sample with an Application to Inferring Car-Sharing Demand. *Transp. Sci.* **2021**, *55*, 1–22. [CrossRef]
16. Nie, G.Q.; Zhou, X.Y. Empirical Bayes test problem for the parameter of two-side truncated distribution families: In the case of NA samples. *Breast Cancer Res.* **2014**, *41*, 134–139.
17. Gu, K.; Jia, X.Z.; You, H.L.; Liang, T. The yield estimation of semiconductor products based on truncated samples. *Int. J. Metrol. Qual. Eng.* **2014**, *4*, 215–220. [CrossRef]
18. Kong, X.F.; He, Z.; Che, J.G. A judgement study on process capability of suppliers truncated treatment. *Xitong Gongcheng Lilun Shijian/Syst. Eng. Theory Pract.* **2008**, *28*, 75–80.
19. Cohen, A.C. *Truncated and Censored Samples: Theory and Applications*; CRC Press: Boca Raton, FL, USA, 1991.
20. Kane, V.E. Process capability indices. *J. Qual. Technol.* **1986**, *18*, 41–52. [CrossRef]
21. Guevara, R.D.; Vargas, J.A. Comparison of process capability indices under autocorrelated data. *Rev. Colomb. Estadíst.* **2007**, *30*, 301–316.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.