The yield estimation of semiconductor products based on truncated samples

K. Gu[*], X.Z. Jia, H.L. You, and T. Liang

School of Microelectronics, Xidian University, Xi'an 710071, P.R. China

Abstract. Product yield reflects the potential product quality and reliability, which means that high yield corresponds to good quality and high reliability. Yet consumers usually couldn't know the actual yield of the products they purchase. Generally, the products that consumers get from suppliers are all eligible. Since the quality characteristic of the eligible products is covered by the specifications, then the observations of quality characteristic follow truncated normal distribution. In the light of maximum likelihood estimation, this paper proposes an algorithm for calculating the parameters of full Gaussian distribution before truncation based on truncated data and estimating product yield. The confidence interval of the yield result is derived, and the effect of sample size on the precision of the calculation result is also analyzed. Finally, the effectiveness of this algorithm is verified by an actual instance.

Keywords: Truncated normal distribution; product quality; product yield; maximum likelihood estimate; fisher information matrix

1 Introduction

As semiconductor technologies continue to evolve, users' demands for even higher product quality and reliability are increasing. So product yield draws more and more attention. In manufacturing system, product yield level depicts product quality performance to a certain extent. The six sigma handbook [1] has ever pointed out that there is an obviously positive correlation among yield, quality and reliability, which means that the more nonconforming products during manufacturing, i.e. the lower product yield, the lower quality grade and poorer reliability of the eligible products; in contrast, when the amount of nonconforming products is reduced, the failure rate of eligible products also falls, i.e. the quality of products is better. Therefore, when purchasing goods the consumers can take the product yield as a sign of product quality and reliability.

However, in practice, users usually can't know the actual product yield of manufacturers. Furthermore, manufacturers have removed the nonconforming products prior to delivery. That is to say the quality characteristic data of the eligible products are truncated. Typically, quality characteristic is normally distributed. Hence, the quality characteristic of eligible products has a truncated normal distribution. To evaluate the product quality, some users calculate product yield by taking the truncated data as the complete sample. Nonetheless, this method will overestimate the product yield of suppliers. In addition, some

researchers propose to transform the truncated sample by data transformation [2,3], and then calculate process capability index [4–6] by the coded data. Actually, such a transformation will alert the distribution of the original data and cause a non-negligible calculation error [7].

How to judge the quality of products supplied by manufacturers based on truncated samples is an important task not only for the semiconductor industry but also for all the manufacturing industry. The paper introduces a method to estimate product yield by truncated normal data; then the confidence interval of the yield and the requirement for minimum sample size for this method are analyzed. Finally, detailed steps of calculating product yield is presented with actual test data of a semiconductor product, and the effectiveness of this algorithm is also verified.

2 Estimation of product yield based on truncated samples

Suppose that X is the quality characteristic, LSL and USL are the lower and upper specification limits, respectively. The truncated sample, $\{x_1, x_2, \ldots, x_n\}$ is a random sample from a truncated normal distribution, and the complete sample has a normal distribution. Let the normal population have a mean of μ and a standard deviation of σ. According to the six sigma design concept, the P value of $\pm P\sigma$ corresponding to the specification interval can be regarded as the judgment of production

* Correspondence: kkgukai@vip.qq.com

level and product quality. Then $USL - LSL = 2h = 2P\sigma$, and the midpoint of the specification interval, which is denoted as T_0, is $(LSL + USL)/2$. In actual production, μ and T_0 usually don't coincide. Assume that μT_0 is $\delta\sigma$, for the given specification limits, LSL and USL, product yield η can be calculated as [8]:

$$\eta = \Phi(\delta + P) - \Phi(\delta - P) \tag{1}$$

where Φ is the cumulative distribution function of the standard normal distribution. This formula shows that the product yield can be obtained as long as the P-value of $P\sigma$ design level and the δ-value, the degree of deviation between the complete sample mean μ and the midpoint of the specification interval T_0, are estimated correctly by the truncated normal sample.

In this section, a common parameter estimation technique, maximum likelihood estimation, is applied to truncated normal sample. The probability density function (p.d.f.) of the truncated normal distribution is expressed as [9]:

$$f(x; \mu, \sigma) = \begin{cases} \frac{\phi(\frac{x-\mu}{\sigma})}{\sigma\left[\Phi(\frac{USL-\mu}{\sigma}) - \Phi(\frac{LSL-\mu}{\sigma})\right]}, & LSL \leqslant x \leqslant USL \\ 0 & \text{otherwise} \end{cases}$$

where φ is the probability density function of the standard normal distribution. Converting the above formula to be the function of P and δ gives the following expression for p.d.f. of the truncated normal distribution:

$$f(x; P, \delta) = \begin{cases} \frac{P\phi[(Px - PT_0 - \delta h)/h]}{h[\Phi(\delta+P) - \Phi(\delta-P)]}, & LSL \leqslant x \leqslant USL \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Then the log-likelihood function is:

$$\text{Ln}L = \sum_{i=1}^{n} \log f(x_i; P, \delta). \tag{3}$$

The maximum likelihood estimators of P and δ are found by solving the following partial differential equations:

$$\begin{cases} \frac{\partial \text{Ln}L}{\partial P}\Big|_{P=\hat{P}} = 0 \\ \frac{\partial \text{Ln}L}{\partial \delta}\Big|_{\delta=\hat{\delta}} = 0. \end{cases} \tag{4}$$

As shown in equation (4), there is no explicit expression for P and δ. Hence, the estimators of P and δ should be calculated by newton-raphson iteration method (this method will not be elaborated in this paper. Please refer to [10] for detailed steps.).

3 The relationship between confident interval and sample size

The previous section introduces a method to calculate product yield based on maximum likelihood estimation. Actually, the calculation precision is affected by the sample size. In this section, the relationship among the estimator of yield, the confidence interval of the maximum likelihood estimator and the sample size will be discussed.

3.1 Analysis of confident interval

In order to obtain the confidence interval of the maximum likelihood estimator the concepts of Fisher information and Fisher information matrix should be introduced. The Fisher information is a way of measuring the amount of information that an observable random variable Y carries about an unknown parameter θ upon which the probability of Y depends. The probability function for Y, which is also the likelihood function for θ, is a function $f(Y; \theta)$. The partial derivative with respect to θ of the natural logarithm of the likelihood function is called the score. Under certain regularity conditions [11], it can be shown that the first moment of the score is 0. The second moment is called the Fisher information:

$$I(\theta) = E\left[\left(\frac{\partial^2}{\partial^2\theta} \log f(X;\theta)\right)\Big|\theta\right]$$
$$= \int \left(\frac{\partial^2}{\partial^2\theta} \log f(X;\theta)\right) f(X;\theta)\right) dX.$$

If $\log f(Y;\theta)$ is twice differentiable with respect to θ, then the Fisher information may also be written as [12]:

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} \log f(X;\theta)\Big|\theta\right]. \tag{5}$$

When there are v parameters, so that $\boldsymbol{\theta}$ is a $v \times 1$ vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_v)^{\text{T}}$, then the Fisher information takes the form of a $v \times v$ matrix, the Fisher Information Matrix (FIM), with typical element:

$$(I(\theta))_{i,j} = E\left[\left(\frac{\partial}{\partial\theta_i} \log f(X;\theta)\right)\left(\frac{\partial}{\partial\theta_j} \log f(X;\theta)\right)\Big|\theta\right].$$

The FIM is a $v \times v$ positive semidefinite symmetric matrix. Under certain regularity conditions, the Fisher Information Matrix may also be written as:

$$(I(\theta))_{i,j} = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j} \log f(X;\theta)\Big|\theta\right]. \tag{6}$$

It has been proved that the maximum likelihood estimator has asymptotically normal distribution [13]. That is to say if the true value of θ is θ_0 and the maximum likelihood estimator of θ is $\hat{\theta}$, then

$$\hat{\theta} - \theta_0 \sim N\left(0, \frac{1}{nI_1(\theta_0)}\right). \tag{7}$$

where $I_1(\theta_0)$ is the Fisher Information Matrix of sample size 1. For a given sample of size n from a distribution with the unknown parameter θ, $I_1(\theta_0)$ can be computed using the estimated value $\hat{\theta}$. According to equation (7), it can be further argued that the distribution of $\hat{\theta} - \theta_0$ asymptotically follows normal distribution:

$$\hat{\theta} - \theta_0 \sim N\left(0, \frac{1}{nI_1(\hat{\theta})}\right).$$

Since the standard normal distribution is symmetric about 0, then

$$P\left\{-\frac{Z(1-\alpha/2)}{\sqrt{nI_1(\hat{\theta})}} \leqslant \theta_0 - \hat{\theta} \leqslant \frac{Z(1-\alpha/2)}{\sqrt{nI_1(\hat{\theta})}}\right\} \approx 1-\alpha.$$

where Z_α is the quantile of the standard normal distribution. Manipulation of the inequalities yields the following approximate $100(1-\alpha)\%$ confidence interval

$$\hat{\theta} - \frac{Z(1-\alpha/2)}{\sqrt{nI_1(\hat{\theta})}} \leqslant \theta_0 \leqslant \hat{\theta} + \frac{Z(1-\alpha/2)}{\sqrt{nI_1(\hat{\theta})}}. \tag{8}$$

In summary, as long as the type of distribution is known, distribution parameter can be estimated by maximum likelihood estimation, and then FIM may be calculated by equation (5) or equation (6). Lastly, the confidence interval of the maximum likelihood estimator can be obtained by equation (8).

For truncated normal sample, the unknown distribution parameter vector is $\boldsymbol{\theta} = (P, \delta)^{\mathrm{T}}$, and the log-likelihood function is:

$$\log f(X; P, \delta) = -\frac{1}{2}\ln(2\pi) - \frac{(PX - PT_0 - \delta h)^2}{2h^2}$$
$$- \log[\Phi(\delta + P) - \Phi(\delta - P)]$$
$$+ \log P - \log h.$$

For the convenience of calculating the following intermediate functions are defined:

$$y(P, \delta) = \Phi(\delta + P) - \Phi(\delta - P)$$
$$a(P, \delta) = \frac{\partial y}{\partial P} = \phi(\delta + P) + \phi(\delta - P)$$
$$b(P, \delta) = \frac{\partial y}{\partial \delta} = \phi(\delta + P) - \phi(\delta - P).$$

The second derivatives with respect to P and δ of the log-likelihood function are:

$$\frac{\partial^2}{\partial^2 P}\log f(X; P, \delta) = -\frac{1}{P^2} - \frac{(X-T_0)^2}{h^2}$$
$$- \frac{y(P,\delta)\frac{\partial}{\partial p}a(P,\delta) - a^2(P,\delta)}{y^2(P,\delta)}$$

$$\frac{\partial^2}{\partial^2 \delta}\log f(X; P, \delta) = -1 - \frac{y(P,\delta)\frac{\partial}{\partial \delta}b(P,\delta) - b^2(P,\delta)}{y^2(P,\delta)}$$

$$\frac{\partial^2}{\partial P \partial \delta}\log f(X; P, \delta) = \frac{\partial^2}{\partial \delta \partial P}\log f(X; P, \delta) = \frac{X-T_0}{h}$$
$$- \frac{y(P,\delta)\frac{\partial}{\partial \delta}a(P,\delta) - a(P,\delta)b(P,\delta)}{y^2(P,\delta)}.$$

Let $I(P, \delta)$ be the FIM of truncated normal distribution. According to equation (6), the elements of FIM are:

$$(I(P,\delta))_{11} = -E\left[\frac{\partial^2}{\partial P^2}\log f(X; P, \delta)\Big| P, \delta\right]$$
$$= \frac{1}{P^2} + \frac{E(X-T_0)^2}{h^2}$$
$$+ \frac{y(P,\delta)\frac{\partial}{\partial p}a(P,\delta) - a^2(P,\delta)}{y^2(P,\delta)} \tag{9}$$

$$(I(P,\delta))_{22} = -E\left[\frac{\partial^2}{\partial \delta^2}\log f(X; P, \delta)\Big| P, \delta\right]$$
$$= 1 + \frac{y(P,\delta)\frac{\partial}{\partial \delta}b(P,\delta) - b^2(P,\delta)}{y^2(P,\delta)} \tag{10}$$

$$(I(P,\delta))_{12} = (I(P,\delta))_{21} = -E\left[\frac{\partial^2}{\partial P \partial \delta}\log f(X; P, \delta)\Big| P, \delta\right]$$
$$= -\frac{E(X-T_0)}{h}$$
$$+ \frac{y(P,\delta)\frac{\partial}{\partial \delta}a(P,\delta) - a(P,\delta)b(P,\delta)}{y^2(P,\delta)}. \tag{11}$$

On the basis of statistical theory [1], when X is a truncated normal random variable, the expectation and variance of X are:

$$E(X) = T_0 + \frac{\delta h}{P} + \frac{hb(P,\delta)}{Py(P,\delta)}$$
$$\mathrm{Var}(X) = \left\{1 - \frac{Pa(P,\delta) + \delta b(P,\delta)}{y(P,\delta)} - \left[\frac{b(P,\delta)}{y(P,\delta)}\right]^2\right\}\frac{h^2}{P^2}.$$

Then, it can be concluded that:

$$E(X-T_0) = E(X) - T_0 = \frac{\delta h}{P} + \frac{hb(P,\delta)}{Py(P,\delta)} \tag{12}$$

$$E(X-T_0)^2 = E\left[X - E(X) + E(X) - T_0\right]^2$$
$$= \mathrm{Var}(X) + [E(X) - T_0]^2$$
$$= \mathrm{Var}(X) + \left[\frac{\delta h}{P} + \frac{hb(P,\delta)}{Py(P,\delta)}\right]^2. \tag{13}$$

Substituting equations (12) and (13) into equations (9) and (11) reveals that the elements of the FIM of truncated normal distribution are:

$$(I(P,\delta))_{11} = \frac{1}{P^2} + \left\{1 - \frac{Pa(P,\delta) + \delta b(P,\delta)}{y(P,\delta)} - \left[\frac{b(P,\delta)}{y(P,\delta)}\right]^2\right\}\frac{1}{P^2}$$
$$+ \left[\frac{\delta}{P} + \frac{b(P,\delta)}{Py(P,\delta)}\right]^2 + \frac{y(P,\delta)\frac{\partial}{\partial p}a(P,\delta) - a^2(P,\delta)}{y^2(P,\delta)}$$

$$(I(P,\delta))_{22} = 1 + \frac{y(P,\delta)\frac{\partial}{\partial\delta}b(P,\delta) - b^2(P,\delta)}{y^2(P,\delta)}$$

$$(I(P,\delta))_{12} = (I(P,\delta))_{21} = -\frac{\delta}{P} - \frac{b(P,\delta)}{Py(P,\delta)}$$

$$+ \frac{y(P,\delta)\frac{\partial}{\partial\delta}a(P,\delta) - a(P,\delta)b(P,\delta)}{y^2(P,\delta)}.$$

Summing up the above, if $\hat{P}$ and $\hat{\delta}$ are the maximum likelihood estimators of P and δ, respectively, $\hat{\boldsymbol{\theta}} = (\hat{P}, \hat{\delta})^T$ asymptotically follows normal distribution $N(\boldsymbol{\theta}_0, \frac{1}{nI_1(\hat{\boldsymbol{\theta}})})$. According to equation (8), the 95% confidence interval of P is:

$$\left(\hat{P} - Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{\boldsymbol{\theta}}))_{1,1}}}, \hat{P} + Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{\boldsymbol{\theta}}))_{1,1}}} \right) \tag{14}$$

the 95% confidence interval of δ is:

$$\left(\hat{\delta} - Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{\boldsymbol{\theta}}))_{2,2}}}, \hat{\delta} + Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{\boldsymbol{\theta}}))_{2,2}}} \right) \tag{15}$$

P-value reflects the concentration of the quality characteristic, and δ-value reflects the deviation between the mean of the complete data and the midpoint of the specification interval. Based on equation (1), it is obvious that the product yield increases as P-value and decreases as δ-value. Hence, the 95% confidence interval of product yield η is:

$$\left(\Phi\left(\hat{\delta} + Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{2,2}}} + \hat{P} \right.\right.$$

$$\left. -Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{1,1}}} \right)$$

$$- \Phi\left(\hat{\delta} + Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{2,2}}} - \hat{P} \right.$$

$$\left. +Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{1,1}}} \right),$$

$$\Phi\left(\hat{\delta} - Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{2,2}}} + \hat{P} \right.$$

$$\left. +Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{1,1}}} \right)$$

$$- \Phi\left(\hat{\delta} - Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{2,2}}} - \hat{P} \right.$$

$$\left.\left. -Z_{0.05/2}\sqrt{\frac{1}{n(I_1(\hat{P},\hat{\delta}))_{1,1}}} \right)\right). \tag{16}$$
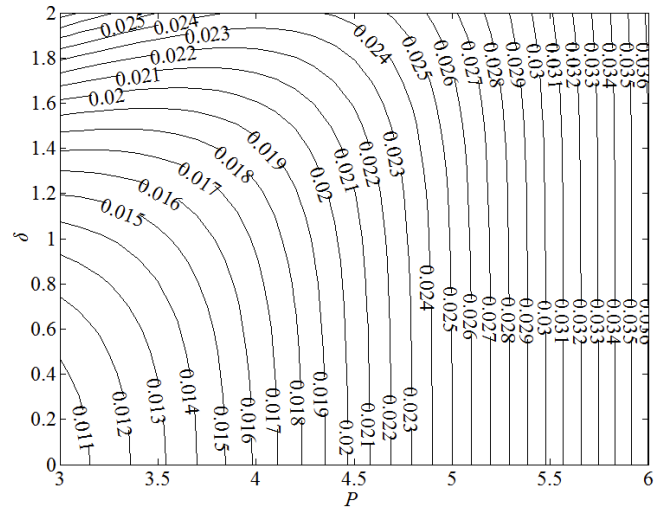

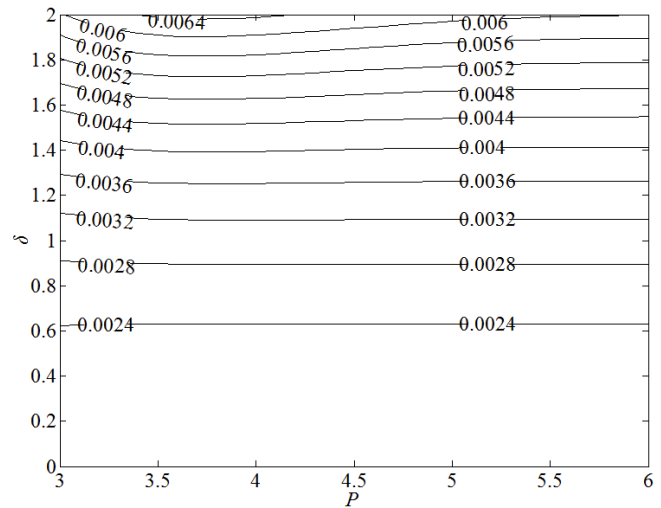**Fig. 1.** MSE of the maximum likelihood estimator of P.

**Fig. 2.** MSE of the maximum likelihood estimator of δ.

3.2 The effect of sample size on precision

Central limit theorem is the foundation of maximum likelihood estimation, so the accuracy of estimators is closely related to the sample size. The larger the sample size, the smaller the Mean Squared Error (MSE) of maximum likelihood estimation, and the narrower the confidence interval. Then the precision of estimation is higher. This study finds that the MSE of the maximum likelihood estimators of P and δ is influenced not only by the sample size but also by the values of P and δ. For a given sample size, $n = 500$, the relationship among P, δ and the MSE of their maximum likelihood estimators is analyzed as shown in Figures 1 and 2.

The following conclusions can be drawn from Figure 1: the MSE of the maximum likelihood estimator of P increases as P-value increases, and when P-value is large enough, say larger than 4.7, the MSE has little to do with δ-value; the MSE of maximum likelihood estimator of P decreases as P-value decreases, and when P-value reduces

Table 1. The relationship among P, δ and product yield, and the minimum sample sizes for the confidence interval not being larger than $\pm10\%$ of P and δ.

δ	$P = 2.0$		$P = 2.5$		$P = 3.0$		$P = 3.5$		$P = 4.0$	
	Yield	Sample size	Yield	Sample size	Yield	Sample size	Yield	Sample size	Yield	Sample size
0.0	0.9545	470	0.9876	285	0.9973	222	0.9995	201	0.9999	194
0.5	0.9270	517	0.9759	313	0.9936	237	0.9986	208	0.9998	197
1.0	0.8400	674	0.933	405	0.9772	288	0.9938	232	0.9986	207
1.5	0.6912	996	0.8413	593	0.9332	392	0.9772	286	0.9938	232
2.0	0.5000	1585	0.6915	940	0.8413	587	0.9332	392	0.9772	286
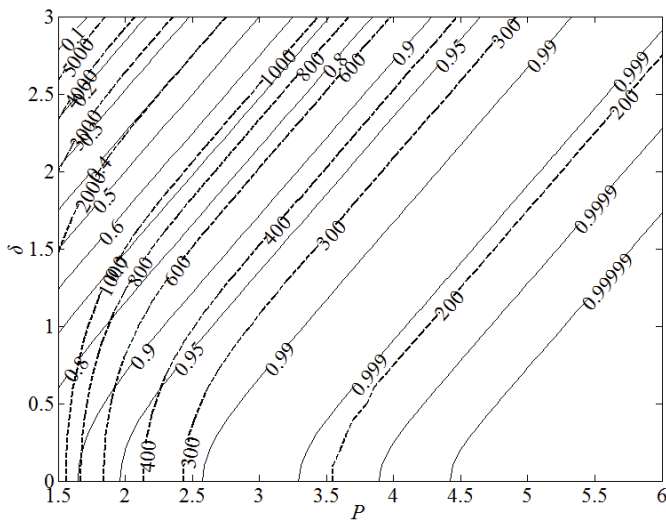
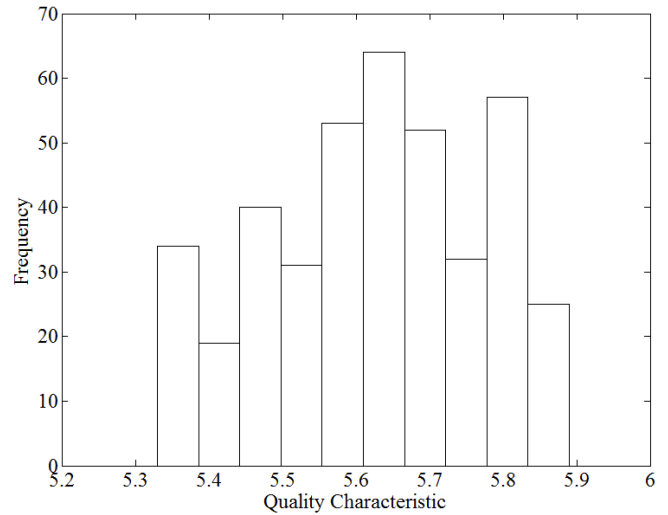

Fig. 3. The contours of yield and the minimum sample size for the confidence interval not being larger than $\pm10\%$ of P and δ (solid line: product yield; dash line: minimum sample size).

Fig. 4. Histogram of the 407 test data.

to a certain extent, δ-value will affect the MSE dramatically. Figure 2 reveals that the MSE of the maximum likelihood estimator of δ increases as δ-value increases, whereas P-value almost has no effect on the MSE. As both P-value and δ-value determine the product yield, in order to ensure that the maximum likelihood estimators have the same precision, different sample sizes should be adopted at different yield levels as shown in Table 1. The minimum sample sizes are also listed for the confidence interval not being larger than $\pm10\%$ of P and δ. The contours in Figure 3 describe the yield results at different P-value and δ-value and the minimum sample size for the confidence interval not being larger than $\pm10\%$.

It can be concluded from Table 1 and Figure 3 that the lower the product yield, the larger the sample size needed to ensure the same accuracy of the estimators.

According to the above analysis, this paper proposes the following steps to calculate product yield based on truncated data:

(1) Choose 200 to 300 eligible products randomly, test and record the observations.
(2) Calculate the maximum likelihood estimators of P and δ by the method introduced in Section 2, and compute product yield using equation (1).

(3) If the yield result is larger than 95%, this means that the sample size is proper and the yield result is believable.
(4) If the yield result is smaller than 95%, to improve the accuracy of estimation, the required minimum sample size has to be obtained from equations (14) and (15) and then users should increase the sample size to meet the requirement and recalculate the product yield; an alternative method is just using the 95% confidence interval of the product yield as the judgment of product quality.

The authors have developed a software package on the basis of the above steps. Users can easily evaluate the quality of products supplied by manufacturers with the help of this package.

4 Illustrative instance

The quality characteristic of a specific semiconductor is denoted as X, and the upper and lower specification limits are known to be 5.9 and 5.3, respectively. 407 eligible parts supplied by manufacturer are tested. The histogram of the test data is shown in Figure 4.

According to the steps for computing yield presented in Section 4, the truncated data are analyzed by the authors'

software package. The yield of this product is calculated as 82.63%, and the 95% confidence interval of the yield is (73.56%, 91.36%).

In the light of the information provided by the manufacturer, the total number of this batch of products is 462. Since there are 407 eligible products, then the actual yield is approximately 88.1%. It is obvious that there is excellent agreement between the actual yield and the confidence interval calculated in this paper. This result shows that the conclusion of the yield computed by the proposed method is in line with the actual situation.

As shown in Figure 3, if the yield value is between 80% and 90%, a minimum of 600 to 800 samples is required to ensure the calculation precision. Hence, in practical application, if the sample size of the truncated data could be increased, the calculation accuracy of product yield can be further improved.

5 Conclusion

In order to scientifically and correctly evaluate and control the product quality and reliability provided by suppliers, the distribution parameters of complete data should be estimated accurately according to truncated samples at first, and then the product yield should be calculated. This study is founded on statistical theory. The distribution parameters of complete samples are estimated by truncated data, that has a truncated normal distribution, and then the product yield of suppliers can be calculated. The authors have also presented the confidence intervals of the results and discussed the effect of sample size on calculation precision. The case study shows that the statistical properties of the sample prior to truncation and the product yield can be restored scientifically by the method proposed in this paper.

In the paper, the algorithms are applied to the case that the quality characteristic of products is normally distributed and has two-side specification limits, which means that a normal sample of the quality characteristic is two-sided truncated. However, the algorithm for estimating parameters of complete samples and the derivation of confidence intervals can also apply to the case of one-sided truncation.

References

1. T. Pyzdek, P. Keller, *The Six Sigma Handbook*, 3rd edn. (McGraw-Hill Professional, New York, 2010)
2. P.H. Mo, Corruption and economic growth, J. Comp. Econom. **29**, 66–79 (2001)
3. W.L. Pearn et al., Testing process precision for truncated normal distributions, Microelectron. Reliab. **47**, 2275–2281 (2007)
4. V.E. Kane, Process capability indices, J. Qual. Technol. **18**, 41–52 (1986)
5. S. Kotz, N.L. Johnson, Process capability indices – A review, 1992–2000, J. Qual. Technol. **34**, 1–19 (2002)
6. C.W. Wu, W.L. Pearn, K. Samuel, An overview of theory and practice on process capability indices for quality assurance, Int. J. Prod. Econ. **117**, 338–359 (2009)
7. S. Rose-Ackerman, *Redesigning the State to Fight Corruption: Transparency, Competition and Privatization* (Public Policy for Private Sector, World Bank, Washington, DC, 1996)
8. M.-T. Chao, D.K.J. Lin, Another Look at the Process Capability Index, Qual. Reliab. Eng. Int. **22**, 153–163 (2006)
9. L. Norman Johnson, S. Kotz, N. Balakrishnan, *Continuous Univariate Distributions*, 2nd edn. (Wiley-Interscience, New York, 1994), Vol. 1
10. J.-F. Bonnans, J.C. Gilbert, C. Lemarechal, *Numerical optimization: Theoretical and practical aspects*, 2nd edn. (Springer-Verlag, Berlin, 2006)
11. F.Y. Edgeworth, On the Probable Errors of Frequency-Constants, J. Roy. Stat. Soc. **71**, 499–512 (1908)
12. J.W. Pratt, F.Y. Edgeworth, R.A. Fisher, On the Efficiency of Maximum Likelihood Estimation, Ann. Stat. **4**, 501–514 (1976)
13. W.K. Newey, D. McFadden, *Handbook of Econometrics* (Elsevier Science, Amsterdam, 1994), Vol. 4