Distribution of the Coefficient of Variation of the Continuous Sample in the Electronic Testing of the Raw Silk Size

Jian-Mei Xu*, Lun Bai*

School of Material Science, Soochow University, Suzhou 215021, Jiangsu Province, China

Abstract: As we know, the coefficient of variation (CV) is one of the main quality indexes in the electronic testing for raw silk. Therefore, to study the classification of this index, it's nessary to get the distribution and other statistical information of this variable. In our previous study, we have gotten its distribution when the population is normal and the sampling is random. However, in the real testing, the CV value is obtained by computing a series of continuous silk size tested by the electronic machine. Due to the continuity of the silk filament, the series of raw silk size can be regarded as a continuous sample. The characteristics of the size from the continuous sample will be different from that of the random sample. This paper analyzes the results of the continuous sampling tests of the CV of the raw silk size, and deduces the distribution of CV of the raw silk size and its characteristics in the case of continuous sampling test, which provides a theoretical basis for the development of electronic testing standard.

Keywords: random sampling, continuous sampling, distribution of the coefficient of variation, equivalent sample

1. Introduction

As we know, the electronic testing system has been widely used in the textile industries of chemical fiber, cotton and wool for its objective testing results and high efficiency. However, in the silk industry only some enterpries in the silk consuming countries are using the electric testing method. There is no global use and universal standards in the electronic testing and classification. However, the demand for silk is global; the international trade of raw silk demands industry standards for raw silk quality to enable the buyers to purchase raw silk at internationally accepted grades. Therefore, it is of great importance for us to make series of researches on the electronic testing and classificaiton.

In the electronic testing system, coefficient of variation (CV) is an important quality index as it embodies the relative variations of the raw silk size. Here the size means the linear density of the fiber. In the testing, we sample a bobbin from a silk lot, then sample a long segment of raw silk from the bobbin [1], and test its size in a certain testing length, thus get a size series, from which we can get CV of the raw silk size of the long segment. Repeating this test several times [1], we get the average value of CV of the raw silk size, thus we can use the average value to estimate the CV of the raw silk size of this silk lot.

To scientifically and rationally make the raw silk testing standard that takes the size coefficient of variation as the quality index, we need to find the sampling distribution of the size coefficient of variation and its statistic characteristics of the continuous sample in the electronic testing of raw silk size. In our previous research [2], we have found the distribution of the CV of the random sample from the normal population. In the raw silk sampling test, we can assume that the size of a whole lot of raw silk has normal distribution [3]; and this assumption is based on the real testing results that had been done by many reseachers [4]. In the traditional testing, sample of the raw silk size is a random sample. However, in the electronic testing the sample is a long segment of raw silk, the test is also a continuous process, so what we get is not a random sample, but a continuous sample. Here "continuous" means the data of the raw silk size tested by the machine is one after another. At this time due to the continuous change of the raw silk size, the

* Corresponding author's email: *xujianmei@suda.edu.cn; *bailun@suda.edu.cn

data of the sample have correlation with each other. It is very difficult to get the sampling distribution of the CV when the sample is continuous. By analyzing the statistical character of the CV of the continuous sample, this paper get a random sampling distribution of the CV that is equivalent to a continuous sampling distribution of the CV, thus make the CV of the continuous sample take an equivalent expression of the random sample. Therefore, we can get the distribution of the CV of the continuous sample, as well as its mean and variance. By simulation, the feasibility and the accuracy of the equivalence are verified.

2. Continuous Sampling Method and Random Sampling Method

Assume that we are testing the size of a lot of raw silk, the testing length is meters, then we can get a size series $\{x_1, x_2, x_3, \cdots x_N\}$ of this lot, we may regard the series as a normal population $N(\mu, \sigma^2)$ whose samples have correlation with each other, and μ , $\sigma^2$are the mean and variance of the raw silk size respectively[3].

In the electronic raw silk testing, we sample a long silk segment from this lot of raw silk, and test its size every l meter continuously, and this kind of sampling method is regarded as the continuous sampling method. Assume that n data of the raw silk size are obtained by this method, and the series of data can be expressed as $\{x_1, x_2, x_3, \cdots x_n\}$, then x_i is a variable of the raw silk size of a certain length, thus the expected value and variance of the raw silk size can be written respectively as

$$E(x_i) = \mu, \quad D(x_i) = \sigma^2 .$$

As the raw silk size is tested continuously, the series of the data have a periodic change, from thick to thin, and then from thin to thick, thus change in cycles. Therefore, the individual datum of the series is not independent, and there are certain correlations among them. In the statistics, this correlation is usually expressed by the autocorrelation coefficient

$$\rho_s = \frac{E[(x_i - \mu)(x_{i+s} - \mu)]}{\sigma^2},$$

where s is the potential difference.

As the autocorrelation coefficient of the raw silk size attenuates gradually, and the attenuation process takes on a cosine change, an autocorrelation coefficient model [5] of the raw silk size is established, i.e.

$$\rho_s = e^{-\lambda s} \cos\frac{2\pi s}{T}, \quad (s = 0, 1, 2, \cdots\cdots)$$

where parameter λ is the attenuation rate, which expresses the attenuating speed of the autocorrelation coefficient; T is the mimetic period, which embodies the change period of the raw silk size. In this paper, the values of λ and T are all based on the fact that the testing length is l meters long.

The average and variance of the raw silk size of the continuous sample are as follows

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 ;$$

as n in the electronic testing is always very large, there will have no big error, when the variance of the raw silk size is written as

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \tag{1}$$

Coefficient of variation of the raw silk size of the sample is

$$c = s / \bar{x} . \tag{2}$$

Repeat this test for N times; we will get the average of the sample $\bar{c} = \frac{1}{N}\sum_{i=1}^{N} c_i$, which can be regarded as the estimation of the size coefficient of variation of this lot of raw silk. When , this value will be approximately equal to the CV of the population,

$$E(c) = \sigma / \mu \tag{3}$$

On the other hand, random sampling is that we get n samples from the size series $\{x_1, x_2, x_3, \cdots x_N\}$

randomly, i.e., every sample is sampled independently. Moreover, the expected value of the average of the random sample also equals to the mean of the population.

3. Analytical Results of the Size Variance of the Continuous Sample of the Raw Silk

As the mean of the average value of the continuous sample and that of the random sample are equal to the mean of the population, the statistical character of the variance of the continuous sample is to be analyzed. And from Eq.(1) the expected value of the sample variance of the raw silk size can be written as

$$D(x_i) = E[s^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu + \mu - \bar{x}\right)^2\right]$$
$$= E\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right] - E\left[(\bar{x} - \mu)^2\right] \tag{4}$$

According to the equation that shows the relation between the size variances of two cases of different testing length in reference [4], we can get the variance of the average of the sample

$$E\left[(\bar{x} - \mu)^2\right] = \frac{\sigma^2}{n}\left[1 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i\right]. \tag{5}$$

When the sample size n is very big, by equation (5) and (4), we can get

$$E\left[s^2\right] = \sigma^2\left\{1 - \frac{1}{n}\left[1 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i\right]\right\}. \tag{6}$$

Then the expected value of the standard deviation of the continuous sample approximately is

$$E\left[s\right] = \sigma\sqrt{1 - \frac{1}{n}\left[1 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i\right]}\ . \tag{7}$$

4. The Variance and the CV of the Random Sample Equivalent to the Continuous Sample

Eq. (6) and (7) are the means of the variance and the standard deviation of the raw silk size of the continuous sample. Due to the continuity of the sampling, the mean of the variance of the sample is not the size variance of the population, but has a difference with the variance of the population, and the value of the difference is correlated with the continuity of the raw silk size.

Here, we will regard (7) as mean of the standard deviation of random sample, thus we can use the mode of the random sampling to express the continuous sampling. And we can use the sampling distribution result of the CV of the random sample to analyze the problem of the continuous sample. Thus Eq.(6) and Eq.(7) can be defined as equivalent variance and equivalent standard deviation, and

$$r = \sqrt{1 - \frac{1}{n}\left[1 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i\right]} \tag{8}$$

is the equivalent factor, r shows the change of the standard deviation of the raw silk size due to the continuous sampling, this change has correlation with the continuity of the raw silk size.

Similarly, if we ignore the error produced when the expected value of the population is substituted for the mean of the sample (in fact, when the testing times is big enough, this approximation will not cause big error), we can get the CV of the continuous sampling

$$E(c) = E\left(\frac{s}{\bar{x}}\right) \approx \frac{E(s)}{\mu}$$
$$= \frac{\sigma}{\mu}\sqrt{1 - \frac{1}{n}\left[1 + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i\right]}\ . \tag{9}$$

According to the idea of equivalent variance and equivalent standard deviation, we can change the CV of the continuous sample into equivalent CV of the random sample. Here the equivalence in fact means that the sample getting from the correlated normal population by continuous and ordered sampling is equivalent to the sample from the normal population by random sampling. The random sample getting from this normal population $N(\mu,\sigma^2)$ by continuous and ordered sampling is equivalent to the sample from the normal population

$$N\left(\mu,\sigma^2\left\{1-\frac{1}{n}\left[1+\frac{2}{n}\sum_{i=1}^{n-1}(n-i)\rho_i\right]\right\}\right)$$

by random sampling. The random sample getting from this normal population is regarded as equivalent sample. Thus by such equivalent sample we can get the distribution of the CV and its statistical character of the continuous sample from the obtained theory we have made in our previous study [2].

5. Distribution of the CV of the Continuous Sample and its Characteristics

In our previous study, we have found the CV distribution of the random sample, which is the theorem 1 in reference [2].

Theorem 1 If $\overline{X}$ and S are the mean and standard variance of a random sample $x_1, x_2, \text{L L } x_n$ of size taken from a normal population, then the probability $N(\mu,\sigma^2)$ density function of the coefficient of variation $CV = S/\overline{X}$ of the sample is

$$f(z)=\begin{cases} kz^{n-1}\int_0^{+\infty} x^n e^{-\frac{n}{2\sigma^2}\left[z^2x^2+(x-\mu)^2\right]}dx, & z>0 \\ -kz^{n-1}\int_{-\infty}^{0} x^n e^{-\frac{n}{2\sigma^2}\left[z^2x^2+(x-\mu)^2\right]}dx, & z\le 0, \end{cases}$$

where

$$k=\frac{2(n/2)^{(n+1)/2}}{\sigma^{n-1}\Gamma(n/2)\Gamma(1/2)}.$$

Using the result of the equivalent sample, from theorem 1, the sampling distribution of CV and its characteristics in the actual raw silk testing can be obtained.

Lemma 1 Let the size of raw silk in a bobbin tested every l meters has normal distribution $N(\mu,\sigma^2)$, sample a long silk segment for L meters from the bobbin, and test the long segment every l meters long, let $n = L/l$. According to the concept of equivalent sample, the long silk segment can be regarded as a sample that is sampled from a normal populatio $N\left(\mu,\sigma^2\left\{1-\frac{1}{n}\left[1+\frac{2}{n}\sum_{i=1}^{n-1}(n-i)\, e^{-\lambda i}\cos\frac{2\pi i}{T}\right]\right\}\right)$, its sample size is ; then the density function of the CV of the sample, namely, the density function of the size coefficient of variation of the long silk segment will be

$$f(z)=\begin{cases} \dfrac{2\left(\frac{n}{2}\right)^{\frac{n+1}{2}}}{\sigma_l{}^{n+1}\Gamma(n/2)\Gamma(1/2)} z^{n-1}\int_0^{+\infty} x^n e^{-\frac{n}{2\sigma_l^2}\left[z^2x^2+(x-\mu)^2\right]}dx & z>0 \\ -\dfrac{2\left(\frac{n}{2}\right)^{\frac{k+1}{2}}}{\sigma^{n+1}\Gamma(n/2)\Gamma(1/2)} z^{n-1}\int^{0} x^n e^{-\frac{n}{2\sigma_l^2}\left[z^2x^2+(x-\mu)^2\right]}dx & z\le 0 \end{cases} \tag{10}$$

where

$$\sigma_l=\sigma\sqrt{1-\frac{1}{n}\left[1+\frac{2}{n}\sum_{i=1}^{n-1}(n-i)\, e^{-\lambda i}\cos\frac{2\pi i}{T}\right]}.$$

Thereby, we can also get the mean and variance of the CV in the actual raw silk testing respectively

$$E(z)=r\,c, \tag{11}$$

$$D(z)=\frac{r^2c^2}{2n}\left(1+2r^2c^2\right), \tag{12}$$

where

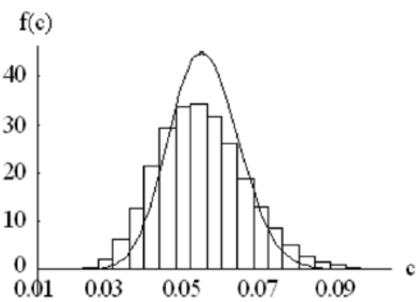
$$c=\sigma/\mu,$$

$$r=\sqrt{1-\frac{1}{n}\left[1+\frac{2}{n}\sum_{i=1}^{n-1}(n-i)\, e^{-\lambda i}\cos\frac{2\pi i}{T}\right]}.$$

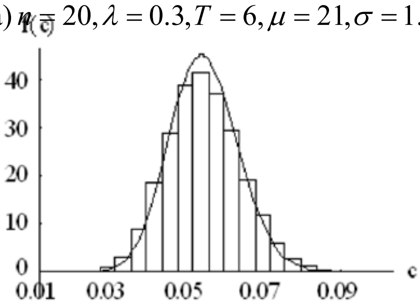
6. Error of Equivalent Sample

According to the definition of the equivalent sample, we can get that the mean of the CV of the continuous sample and that of the random sample are approximately equal. Therefore, the error caused by the equivalence also has relation with the variance of the CV of the sample. Here the value of the variance also has correlation with the continuity of the continuous sampling. By simulation, the extent of the error can be found out.

Figure 1 is the distribution of the CV of the continuous sample and that of the equivalent random sample. In the figure, the histogram is the continuous sampling distribution of the CV from



(a) $n = 20, \lambda = 0.3, T = 6, \mu = 21, \sigma = 1.2$



(b) $n = 20, \lambda = 0.7, T = 6, \mu = 21, \sigma = 1.2$

Figure 1 Distributions of the CV of the continuous sample and the equivalent random sample

The normal population $N(\mu, \sigma^2)$ whose samples have correlation with each other, and the data of the histogram are gotten by simulation, while the real line is the random sampling distribution of the CV, which is drawn according to (10) directly.

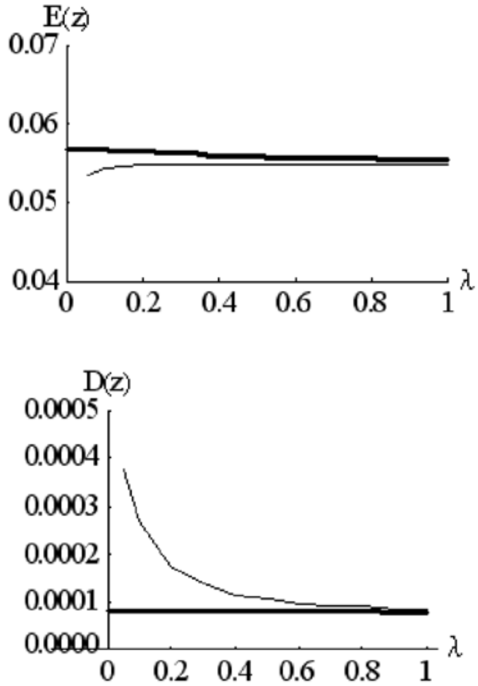
Figure 2 to Figure 6 are the curves showing the changes of the mean and variance of the CV with the change of the parameters of the size continuity, and the two curves in each figure are gotten from the two fore-mentioned sampling methods. In the figures, the thin lines are the curves of the characteristics of the CV of the continuous sample from the correlated normal population $N(\mu, \sigma^2)$, and these curves changing with each parameter, are obtained by simulation. The thick lines are the curves of the characteristics of the CV of the random sample, changing with each parameter, and these curves are drawn according to (11) (12) directly. In each figure, we analyze the effects of the sample size n, attenuation rate λ, the mimetic period T, standard deviation of raw silk size σ and the mean of the raw silk size μ on the equivalent sample error.

6.1 Attenuation Rate and the Equivalent Sample Error

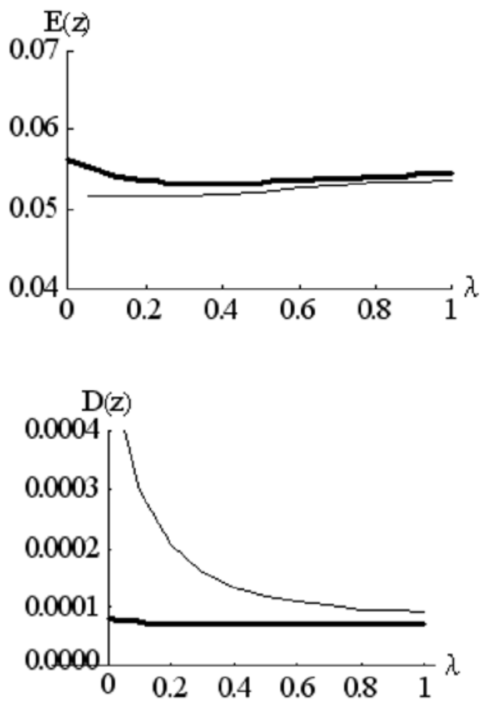
From the distributions of the CV of the continuous sample and the random sample when the attenuation rate is different and other parameters are same in Figure 1, we can see that the two kinds of distributions are approximately uniform, especially when the attenuation rate is big.

From Figure 2, we can find that, with the increase of the attenuation rate, the mean and variance of the CV of the continuous sample and the equivalent random sample will be closer. It indicates that the distributions of the two kinds of samples will be more similar, and that the equivalent error will decrease with the increase of the attenuation rate. The reason is that the correlation between the raw silk sizes will be weaker with the increase of the attenuation rate, and the continuous sampling will be more similar to the random sampling. Therefore, when the attenuation rate is comparatively bigger, the error of the equivalent transform will be smaller.

Figure 3 shows the characteristic curves of the CV of the continuous sample and the equivalent random sample changing with the mimetic period when other parameters take the same value. In the figure, with the increase of the mimetic period, the difference of the mean and variance of the CV between the two sampling method will be bigger; but the change is very small. The reason is that the correlation between the raw silk sizes will be stronger with the increase of the mimetic period, and at this time the error produces in the substitution of the random sample for the continuous sample will be relatively bigger, but the effect of mimetic period is rather less than that of the attenuation rate, therefore, the effect of the change of the mimetic period on the equivalent error is generally ignored.



(a) $n = 20, T = 6, \mu = 21, \sigma = 1.2$



(b) $n = 20, T = 16, \mu = 21, \sigma = 1.2$

Figure 2 Effect of the attenuation rate on the error of the equivalent sample

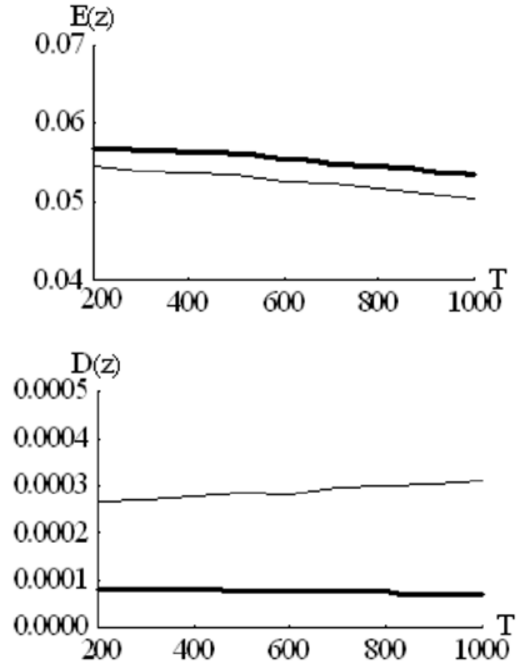


Figure 3 Effect of the mimetic period on the equivalent error
$n = 20, \lambda = 0.1, \mu = 21, \sigma = 1.2$

6.3 Sample Size and the Equivalent Sample Error

Figure 4 shows the characteristic curves of the CV of the continuous sample and the equivalent random sample changing with sample size n when the other parameters take the same value. In the figure, with the increase of the sample size, the mean and variance of the CV of the continuous sample and the equivalent random sample will be closer, which indicates that the equivalent error will decrease with the increase of the sample size.

6.4 Size Variance of the Population and the Equivalent Sample Eerror

Figure 5 shows the characteristic curves of the CV of the continuous sample and the equivalent random sample changing with size deviation of the population σ when the other parameters are same. In the figure, with the increase of the size deviation, the difference between the means of the CV of the continuous sample and the equivalent random sample basically remains invariable, while the difference between the variance curves of the CV of the two kinds of sampling methods has a tendency of increasing gradually, but the increasing extent is not big.

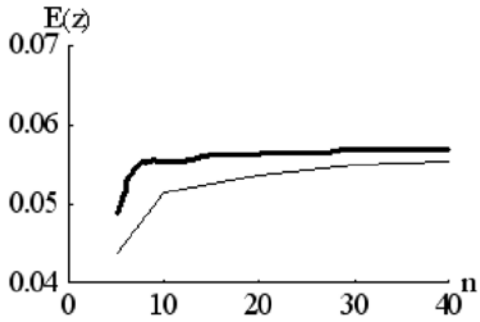


Figure 4 Effect of the sample size on the equivalent error

$T=8, \lambda=0.1, \mu=21, \sigma=1.2$

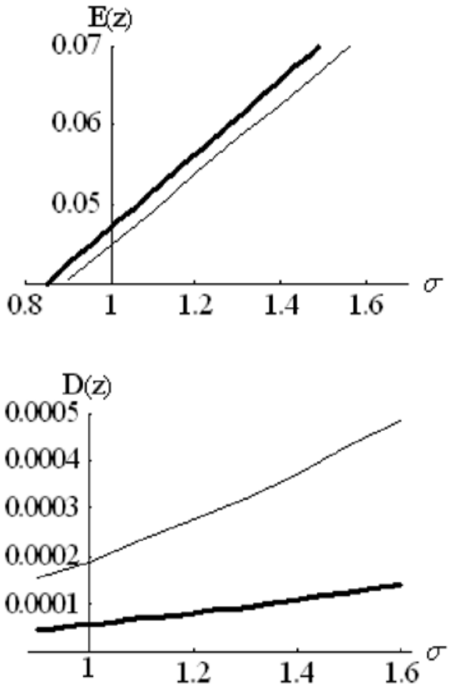


Figure 5 Effect of the size variance of the population on the equivalent error

$k=20, T=8, \lambda=0.1, \mu=21$

6.5 Size Mean of the Population and the Equivalent Sample Error

Figure 6 shows the characteristic curves of the CV of the continuous sample and the equivalent random sample changing with size mean of the population μ when other parameters are same. In the figure, with the increase of the size mean, the difference between the means of the CV of the continuous sample and the equivalent random sample has no obvious change, while the variance curves of the CV of the two kinds of sampling methods become more and more near, but the changing extent is not big.

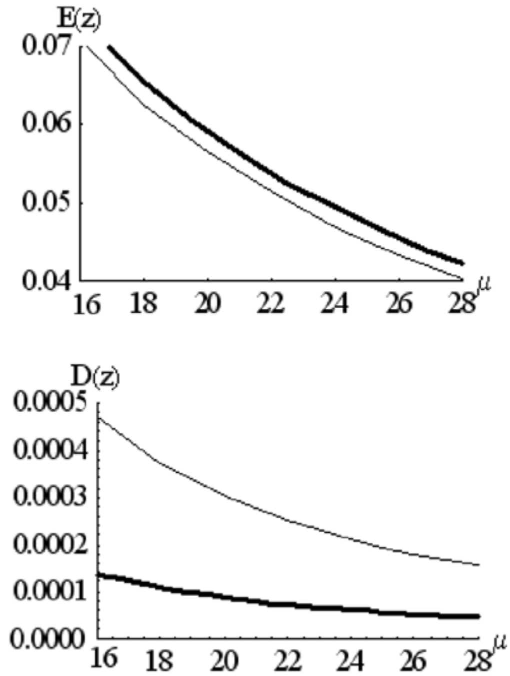


Figure 6 Effect of the size mean of the population on the equivalent error

$k=20, T=8, \lambda=0.1, \sigma=1.2$

From the above analyses, we can concluded that the error caused by using the method of equivalent sample to analyze the CV in the continuous sampling is basically very small in the general electronic testing of the raw silk size, and this method is feasible in application. Especially when the sample size and the attenuation rate are comparably bigger, the effect of the equivalent sample analysis will be better.

7. Conclusion

By comparison of the difference between the random sample and the continuous sample from the correlated normal population, the paper proposes a method to obtain the distribution and the characteristics of the CV of the sampling silk segment in the electronic testing of raw silk size; and this method is realized by equivalent sampling based on the obtained random sampling distribution theory of CV. By simulation, the error caused by the equivalence is verified to be very small. When the sample size and the attenuation rate are comparatively bigger, the effect of the use of equivalent sample will be better.

The research is supported by the doctoral special foundation of the high education (series number: 200400285088)

References:

[1] Xu JM, Zhou Y, Sun QL, Di CX, Gu H, Bai L. Research on sample capacity in the electronic test of raw silk size. Silk Monthly 2006; 9:21-23.

[2] Xu JM. Sampling distribution of the coefficient of variation when the population is normal. 2008. Information on http://www.paper.edu.cn

[3] Suzhou Institute of Silk Textile Technology, Silk reeling book, Peking, Chinese Textile Publication, 1994. p.147.

[4] Shimazaki A. Introduction to the control-engineering, Farming & Forestry Statistical Association, Japan, 1973. p.193-196.

[5] Bai L. Studies on the relation between the continuous property of raw silk size and the size deviation (continuation), Suzhou Institute of Silk Textile Technology 1990; 10:1-11.