

Application of truncated distribution for *tsumugi* yarn width

Abstract

Hand spun silk yarn made from floss silk is referred to as *tsumugi* yarn. It gives a unique appearance on the products, *tsumugi*, which are considered to be an important type of fabrics for their aesthetics. The essential factor of this visual effect is originated by the thickness variation of *tsumugi* yarn, however the detail of the variation is unknown. In this paper, as a fundamental study of *tsumugi* yarn, its width distribution is discussed. To model the width distribution, Polya-Eggenberger distribution is introduced. Moreover some modifications are applied for the distribution to describe the yarn situation. That is to say, it is impossible to be thinner than a level of thickness for the yarn to maintain a shape of a thread, on the other hand *tsumugi* yarn is not limited to contain some thicker parts in itself. Additionally, it is not avoidable to involve errors and fluctuations in the actual data. Hence the distribution is truncated and convoluted. By using an image scanner, the width of *tsumugi* yarn is measured. With the measured data, the parameters of the width distribution are estimated using the maximum likelihood method. The results shows that the width distribution of *tsumugi* yarn is asymmetric and well summarized by the distribution.

Key words

truncated distribution, *tsumugi* yarn, convolution, Polya-Eggenberger distribution, maximum likelihood method

The origin of *tsumugi* is utilization of inferior cocoons which cannot provide quality raw silk. By degumming and spreading the cocoons, floss silk is produced. Then the floss is processed by hand into *tsumugi* yarn which is very irregular in thickness due to the entanglement of the fibers in the floss silk. The yarn is raw material of *tsumugi*, a silk fabric with unique irregularity on its surface. Although it began as a secondary product, people became to find beauty in it. Nowadays, *tsumugi* is considered to be an important type of fabrics. Thus several techniques have been developed to introduce the effect into other kind of textiles. Although, it is easy to produce irregular and/or uneven yarns, it is difficult to generate intended effect on the produced textiles. In many cases, uneven yarns result in faults and/or defects on the products. It is surely that the characteristics of *tsumugi* which is valued by people is originated by the thickness variation of *tsumugi* yarn, however the detail of the variation is unknown. There must be a law and/or a theorem of the unevenness to give favorable impression by the yarn used in the textile products. It will be helpful to find the law to elevate the producer's technique and the consumer's satisfaction.

In order to analyze the yarns with thickness variation, some researchers have tried to extract characteristic parts of the yarns. For example, with slub yarns, the slubs correspond to characteristic parts. The length and distance of the slubs have been investigated [1, 2]. Also the influence of the characteristic parts on the fabrics has been studied using doupion silk [3-5]. However, to talk about *tsumugi* yarn, it is difficult to determine the boundaries of the characteristic parts due to the irregular and continuous variation of the yarn thickness.

In this paper, *tsumugi* yarn width distribution which is indispensable to investigate the visual effect of the yarn is discussed. To maintain the shape of a thread, it is impossible to be thinner than a certain level of thickness for the yarn. On contrary to this, a *tsumugi* yarn is not limited to include thicker parts. In order to describe this asymmetric situation, we introduce the idea of truncated distribution. In the previous work, we have proposed the truncated normal distribution convoluted with normal distribution. The likelihood equations for the distribution are derived as well [6]. In this paper, truncated Polya-Eggenberger distribution (TPED) is discussed.

Theoretical

Asymmetric nature of tsumugi yarn

In the textile area, liner density, the weight of the material in unit length, is used to indicate thickness. In order to measure the liner density of a yarn, a long segment of the yarn is sampled and weighed, then the liner density of the sampled yarn is calculated based on the segment length. Suppose we have lots of data obtained with this method. They are expected to follow a normal distribution approximately. The reason is as follows: within a sample, we can assume many short intervals. The weight corresponding to a short interval may be influenced by neighboring intervals whereas it can be considered to be independent from distant ones. So we can consider many of intervals are independent each other in a long sample. Weighing the sample at one time is equal to the averaging the weights corresponding to the short intervals. The averaged values obtained from independent data are expected to follow a normal distribution by the central limit theorem.

In order to investigate *tsumugi* yarn thickness which varies abruptly within short length, it is required to sample in very short length. In this case, we also can assume minute intervals in the sample. However, the minute intervals are not distant each other, hence we should consider that they are not independent. Therefore we cannot apply the central limit theorem for this time. Consequently, it makes difficult to assume normal distribution for the measured data with very short interval.

Additionally, it is impossible to be thinner than a certain level of thickness for a yarn to maintain the shape of thread. Contrary to this, it is not limited to involve thicker parts for *tsumugi* yarn in itself. That is to say, the limitation factors of *tsumugi* yarn thickness are different between thinner and thicker side. Therefore an asymmetric distribution will be required to model *tsumugi* yarn thickness variation. In this purpose, we introduce TPED in the following.

Polya-Eggenberger distribution

Suppose short intervals along a line which will correspond to a yarn. We will determine the thickness of the yarn corresponding to each interval by ball drawing from an urn. The urn contains NP black balls and $N(1 - P)$ red balls where $0 < P < 1$ and N is the total number of balls. If you draw a black ball from the urn, you place a fiber in the interval. Otherwise, if you have a red one, you have to pass. In an interval, the trial mentioned above is repeated M times, and the drawn ball is returned to the urn after each trial. The number of fiber, namely the thickness of the yarn, in each interval follows the Poisson distribution when P is very small and M is very large. The distribution is well known as the idealized model of yarn irregularity.

Now, we extend the model. In this time, at the stage of returning the ball, we add some more same color balls in the urn. Thus, the probability of the same color increases in the next trial. This model is also well known as the Polya's urn. In our case, if a fiber placed in an interval, the probability of the additional fiber increases. We consider that such behavior of the Polya's urn resembles the fiber disposition in *tsumugi* yarn. In the raw material, floss silk, the fibers cross over each other. As the *tsumugi* yarn is produced from floss silk by hand directly, a fiber will catch in another fiber at the cross point. Thus a fiber will lead another one during the yarn formation process. If the effect is dominant, the number of fibers in an interval follows the Polya-Eggenberger distribution:

$$p_o(t) = \frac{\lambda^t}{t!} (1 + \rho)^{-\frac{\lambda}{\rho} - t} \prod_{i=0}^{t-1} \left(1 + \frac{i\rho}{\lambda}\right) \quad (t = 0, 1, 2, \dots) \quad (1)$$

where λ is the mean of the distribution, and ρ is the coefficient of spread which characterizes this distribution.

Truncation

As mentioned before, to maintain the shape of thread, the yarn is needed to be thicker than a level of thickness or it breaks off. In order to reflect the situation on our model, we truncate the lower side of the distribution as follows:

$$p(t) = \frac{1}{1 - p_o(0) - p_o(1) - \dots - p_o(a - 1)} p_o(t) \quad (t = a, a + 1, a + 2, \dots) \quad (2)$$

where a is the truncation point. While we assume TPED, the majority of data need to be described

by the distribution before truncated. Therefore the truncation point should takes a small value in comparison with the mean of the original distribution. So we introduce a truncation coefficient ($0 < k < 1$) and restrict the truncation point to satisfy the following equation.

$$a = k\lambda \quad (3)$$

Convolution

If the yarn thickness follows exactly TPED defined in Equation 2, the data less than the truncation point never appear. However, it is not avoidable to involve errors during actual measurement. Furthermore, the truncation point may fluctuate along the yarn. Thus in this paper, we assume the observed value x is formed by the sum of the error ε and the variable t which ruled by TPED as follows:

$$x = t + \varepsilon \quad (4)$$

Additionally, we can assume the error follows a normal distribution. Ordinary, it is a continuous distribution, however TPED is a discrete distribution. In order to define the error convoluted TPED, the discrete error is approximated as follows:

$$q(\varepsilon) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{\varepsilon^2}{2s^2}} \quad (\varepsilon = \dots, -2, -1, 0, 1, 2, \dots) \quad (5)$$

where s is the standard deviation of the distribution.

The observed value x follows the convoluted distribution of Equations 2 and 5:

$$r(x) = \sum_{u=a}^{\infty} p(u)q(x-u) \quad (6)$$

Maximum likelihood estimation

The logarithmic likelihood function of the convoluted TPED for given data, (x_1, x_2, \dots, x_n) is

$$\log L(\lambda, \rho, k, s) = \sum_{i=1}^n \log r(x_i)$$

(7)

where n is the number of data.

Generally, in the maximum likelihood method, the partial differentiations of Equation 7 are derived to obtain the likelihood equations. Then the equations are solved to estimate the parameters. However, it seemed to be difficult to prove our problem in this way. Hence in this paper, we employ the solver of computer software, and search the combination of the parameters which maximize the Equation 7.

Other distributions

In this paper, we have selected two distributions as the representation of asymmetric distributions to compare with TPED. One is Pearson type V distribution:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}} \quad (8)$$

where α is the shape parameter and β is the scale parameter. The distribution is also called as inverse gamma distribution.

Another is log normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \quad (9)$$

This distribution also has two parameters named μ and σ .

Information criteria

The maximized likelihood value of each distribution obtained from the data is required to select the most suited one. However, when the number of parameters is different each other, the fact also should be concerned. For this purpose, we introduce Akaike's information criterion (AIC) defined as follows:

$$AIC = -2 \log L + 2\psi \quad (10)$$

where L is the maximized likelihood value and ψ is the number of parameters. Among the candidates, if a distribution model takes the minimum value of AIC, the model is selected as the most suited for the data.

Experimental

Materials

A *tsumugi* yarn (129 tex) was used as an experimental material. The total yarn length was 480 m.

Image processing

In order to inspect *tsumugi* yarn's width which shows quick change along yarn length, we need to measure the width of the yarn in detail. An image scanner (CanoScan LiDE 210, Canon) was employed for yarn image capturing. The resolution was 600 dpi that corresponds to 42.3 μm length and width of a pixel. At one time, 20 cm length of yarn was captured in 8 bit gray scale mode. We captured 10 m length of the yarn into 50 images successively. These images formed a unit of data. In this way, we obtained 48 units of data which correspond to 480 m length of the yarn.

In order to extract yarn part, the captured gray images were converted into binary images. The threshold level was selected by Otsu's method [7, 8]. The number of pixels corresponding to yarn was counted to determine the width of the yarn in each column of pixels as shown in Figure 1. The image processing was performed with Scilab on a personal computer. In this way, we obtained 236,200 width data in each unit.

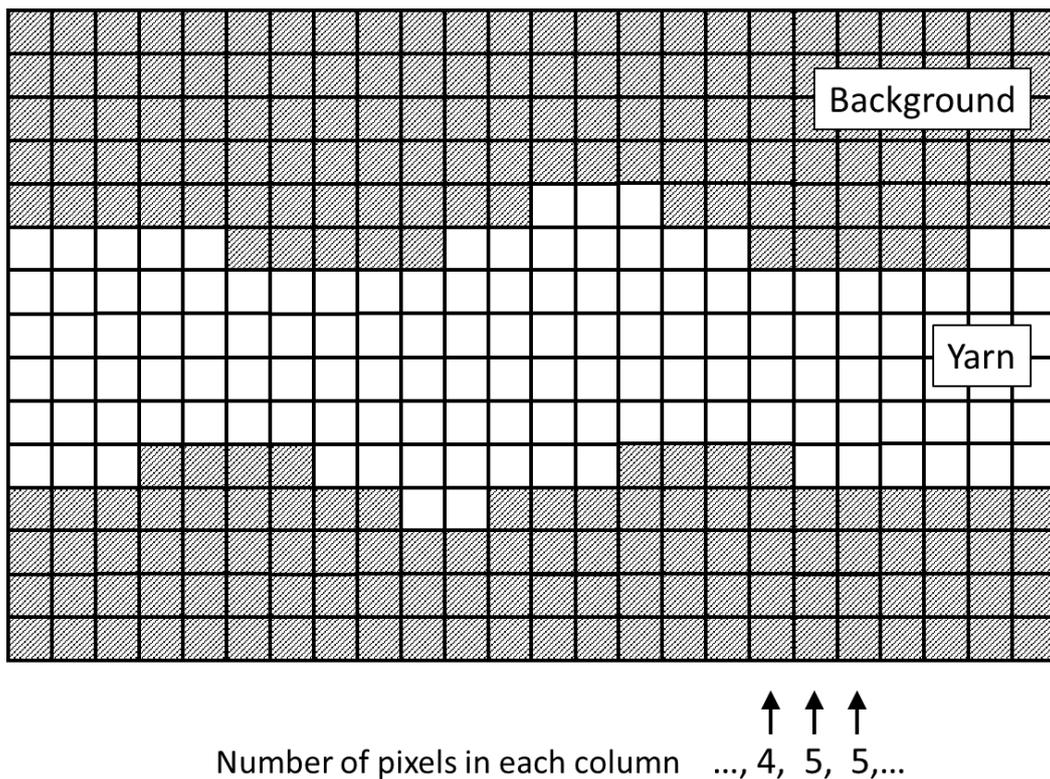


Figure 1. Schematic illustration of yarn width measuring.

Estimation of parameters

The histograms of yarn width distribution were plotted. In order to estimate the parameters of the convoluted TPED in the maximum likelihood method, the solver of Microsoft Excel was employed.

Thereafter the truncation point was calculated by Equation 3.

In the same way, the parameters of Pearson type V distribution and log normal distribution were estimated. The maximized likelihood values and AICs of the distributions were also calculated simultaneously.

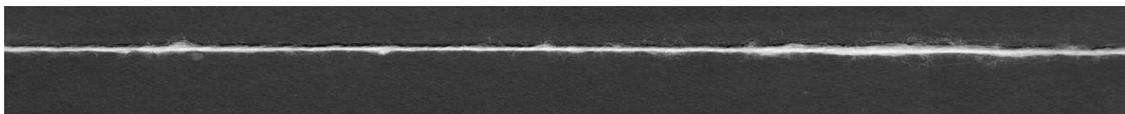
Scatter plots of estimated parameters

In order to verify the dependency among parameters of TPED, scatter plots of the estimated parameters were prepared and discussed.

Results and discussion

Captured images

Figure 2 shows gray and binary images of *tsumugi* yarn. Each image is corresponding to 20 cm length of *tsumugi* yarn. However 20 cm is very short interval for general yarn, *tsumugi* yarn width changes abruptly within the interval as shown in the figure. These characteristics are considered to be important for the product's surface feature. Hence we need to inspect *tsumugi* yarn in minute intervals. By comparing these images, it is considered that the yarn parts are appropriately extracted by Otsu's method.



a) Gray image.



b) Binary image.

Figure 2. Images of *tsumugi* yarn.

Estimated parameters

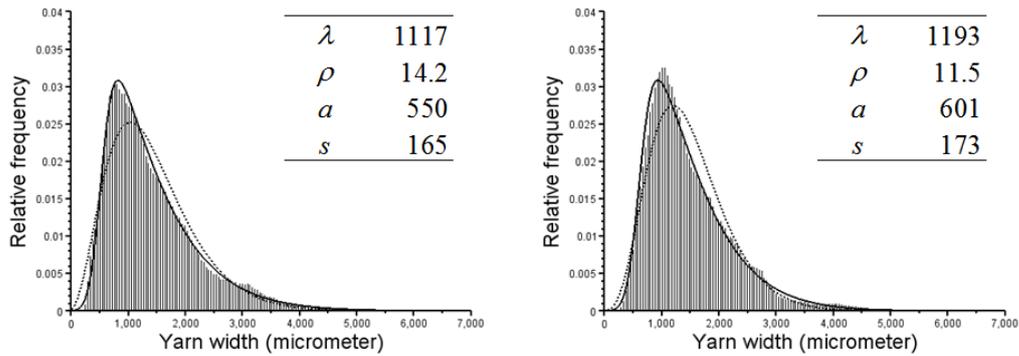
Figure 3 shows the examples of *tsumugi* yarn width distribution. The histograms in the figure are corresponding to measured data. They are asymmetric and not truncated vertically. This implies that some errors and/or fluctuations are involved in the data.

By using the maximum likelihood method, the parameters of the convoluted TPED is estimated. The truncation point is obtained by Equation 3 from the estimated parameters.

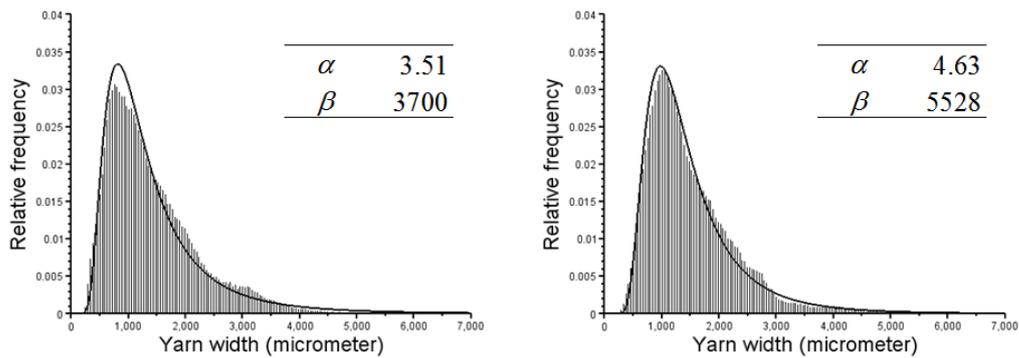
The curves in Figure 3(a) shows the estimated TPED. The original Polya-Eggenberger distribution is also asymmetric as shown in the dotted lines in the graphs. However, the original distributions are not mature to describe *tsumugi* yarn width distribution. Therefore truncation and convolution are needed

to be applied to this distribution for our purpose. The results show that the convoluted TPED well summarizes the distribution of *tsumugi* yarn width.

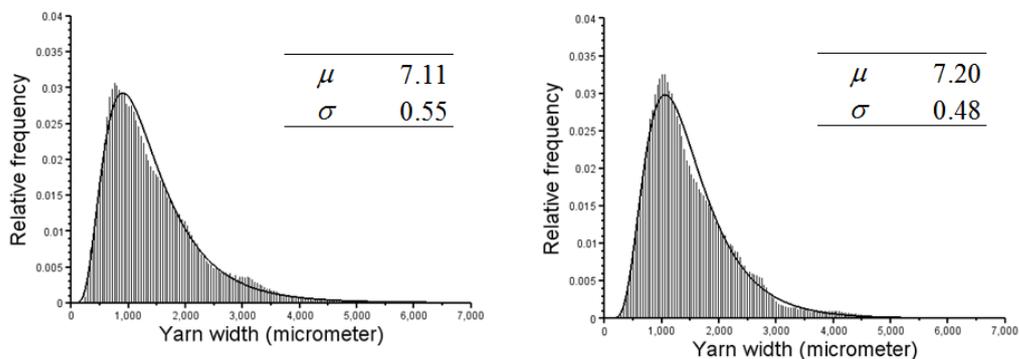
Figures 3(b) and 3(c) show the estimated results of Pearson type V distribution and log normal distribution, respectively. They also show good agreement with the measured data.



a) Truncated Polya-Eggenberger distribution



b) Pearson type V distribution



c) Log-normal distribution

Figure 3. Histograms and estimated curves of *tsumugi* yarn width distribution.

Information criteria

Table 1 shows the calculated results of AIC for each distribution. Among them, the value of AIC corresponding to TPED is the smallest. In the same way, we have inspected 48 units of data and same results have been obtained. Therefore we have selected TPED as the most suitable distribution for our data.

Table 1. Calculated results of AIC ($\times 10^6$).

Data unit number	7	16
TPED	1.97	1.95
Pearson type V	3.75	3.72
log normal	3.58	3.55

Scatter plots of the parameters

Figure 4 shows the scatter plots of the estimated parameters of TPED. The truncation point (a) tend to increase, as the increase of the mean before truncation (λ). However, it is not easy to estimate the former from the latter, as the coefficient of determination is not so large. Similarly, the standard deviation of the convoluted error (s) shows weak relation with truncation point (a). The other combinations of estimated parameters are also inspected, however the coefficients of determination corresponding to other combinations are smaller than the above. Therefore it seems to be difficult to convert this model into simpler one and we need to estimate all of the parameters from the data with this model.

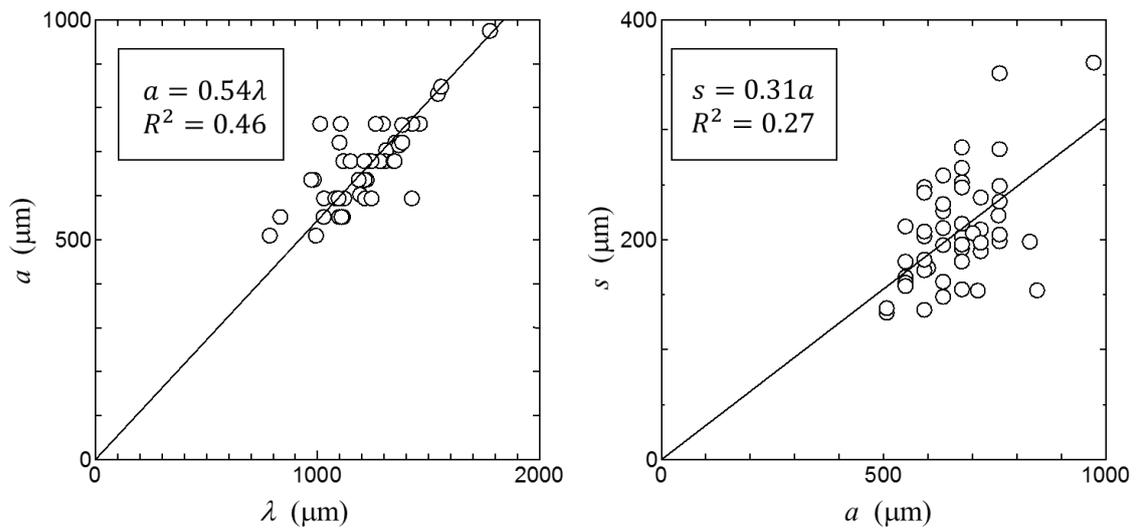


Figure 4. Scatter plots of estimated parameters.

Conclusions

In order to model *tsumugi* yarn width distribution, TPED is introduced. To estimate parameters of the distributions from the observed data, the convoluted TPED is derived. Actual *tsumugi* yarn width is measured with an image scanner. The obtained data are well summarized by the convoluted TPED estimated by the maximum likelihood method. Hence we conclude as follows: 1) the *tsumugi* yarn width is mainly ruled by the Polya-Eggenberger distribution, 2) the strict limitation in the thinner side of the yarn width distorts the distribution, 3) the measuring errors and/or the fluctuations of the truncation point also affect the measured data. By considering these three factors, we are able to understand the *tsumugi* yarn width distribution.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Liu JH, Xie ZF, Gao WD and Jiang HX. Automatic determination of slub yarn geometrical parameters based on an amended similarity-based clustering method. *Textile Res J* 2010; 80: 1075-1082.
2. Pan RR, Gao WD, Liu JH and Wang HB. Recognition the parameters of slub-yarn based on image analysis. *J Engineered Fibers and Fabrics* 2011; 6: 25-30.
3. Nishioka T, Shimazaki A, Watanabe K and Hyoudou M. Characteristics of the knot count in the arbitrary fixed area of the shantung cloth. *J Seric Sci Jpn* 1989; 58: 474-482.
4. Nishioka T, Miura M, Shimazaki A and Morikawa H. Characteristics of knot pattern appearing in the cloth woven from doupion silk. *J Seric Sci Jpn* 1992; 61: 361-369.
5. Takabayashi C, Tsuboi H, Nakamura K, Nakajima K, Teramoto H and Miyasaka T. Measurement of the dupion silk slugs in the silk reeling process and the simulation of slugs occurrence states on the fabric. *J Silk Sci Tech Jpn* 2001; 10: 23-30.
6. Sakaguchi A, Yasuhara K, Matsumoto Y and Kimura H. Truncated normal distribution investigation of *tsumugi* yarn-width. *J Silk Sci Tech Jpn* 2014; 22: 133-138.
7. Otsu N. An automatic threshold selection method based on discriminant and least squares criteria. *Trans Inst Electronics and Communication Engineers of Jpn* 1980; J63-D: 349-356.
8. Sakaguchi A, Matsumoto Y, Toriumi K and Kim H. Novel observation technique for contact condition of fabric. *Textile Res J* 2002; 72: 221-226.