2007 Beltwide Cotton Conferences, New Orleans, Louisiana, January 9-12, 2007

TEXTILE FIBER-LENGTH PROPERTIES OF SEED COTTON - PART 1: TEXTILE MATHEMATICAL PROPERTIES OF THE NORMAL LENGTH DISTRIBUTION

Kearny Q. Robert
Southern Regional Research Center, USDA-ARS
New Orleans, LA
Zhide Fang
Tumulesh Solanky
University of New Orleans
New Orleans, LA

Abstract

The industrially relevant problem investigated herein was the relation of the traditional textile quality-related properties of fiber length distribution (such as UHML or Uniformity Index) to the two parameters of a Gaussian distribution (mean and standard deviation), when the fiber length distribution happens to be Gaussian. In particular, the form of the reference distribution (or seed "paragon") for cotton fiber length has been proposed recently to be Gaussian by mass instead of the more traditionally assumed Gaussian by number. Such normal distributions are believed to be intrinsic to natural fibers, as well as to synthetic staple fibers produced by random long-gage cutting or other random-breakage processes.

Introduction

Practical fiber length distributions (*e.g.,* textile fiber length) are characterized by a variety of mathematical properties of length that are defined in terms of the entire length distribution. Most of these properties can be calculated based on either mass or number weighting. Some examples are:

- Mean Fiber Length = ML
- Standard Deviation of Fiber Length = SD
- Coefficient of Variation of Length = CV = SD/ML
- Upper Quartile Length = UQL = Length exceeded by 25% of fibers
- Lower Quartile Length = LQL = Length exceeded by 75% of fibers
- Short Fiber Content = SFC = Fraction of total distribution having length less than 0.5 inch.

These properties are trivial to calculate for the Gaussian case. The normalized Gaussian distribution is defined mathematically by two parameters: the mean and the standard deviation, which are referred to herein as alpha and sigma, respectively. The properties listed above are easily expressed in terms of alpha and sigma, as long as the frequency weighting is the same (number or mass) as the property weighting. In other words, the UQL by mass of a normal length distribution of mass frequency (*i.e.*, a Gaussian by mass) is easy to calculate. The UQL is simply the value of length, $\boldsymbol{x}$, corresponding to the value of the cumulative normal probability integral, $\mathbf{P}(\boldsymbol{x})$, where $\mathbf{P}(\boldsymbol{x}) = 0.75$. The value of normal deviate can be readily obtained from a handbook, scientific calculator, or spreadsheet software, so that the value of UQL is found to be:

$$(1) \qquad UQL = ML + (0.67449...)SD \; .$$

On the other hand, there are other textile mathematical properties that may be of equal or greater importance, but which are combinations of mass and number integrals. For instance, the calculated UQL by number of a length distribution that is Gaussian (symmetric) by mass is not as obvious. Two crucial properties for cotton fiber length are herein defined as:

- Upper Half Mean Length = UHML = Mean by number of the longer half by mass
- Uniformity Index = UI = ML/UHML = Ratio of Mean Length by mass to UHML.

Objectives & Approach

The problem that needed to be addressed was to determine analytically these last two quantities (UHML and UI) for Gaussian paragons in terms of alpha and sigma. The approach was to evaluate the integral expressions for UHML by use of lognormal approximations over ranges of values where the normal and lognormal distributions are practically indistinguishable.

Methods & Definitions

Frequency Distributions of Length

Robert [1991] inferred from experiment that each partly-broken cotton fiber length distribution could be associated with a theoretically unbroken reference distribution, which he referred to as the "paragon" distribution. Robert and Blanchard [1997], Robert *et al.* [2000], and Robert [2000b] showed quantitatively how fiber length distributions are transformed by experimental breakage from longer to shorter functions. Robert [1991] hypothesized that the useful form of the cotton fiber paragon is Gaussian by mass (and not by number). Recently, Robert [2005a] offered experimental evidence for the mass-based Gaussian paragon based on single-fiber studies of fiber length of cotton on the seed. Therefore, consider the case wherein the unbroken mass probability density function (p.d.f.) of fiber length is Gaussian.

For a population of fibers, define $n(x)$ to be the frequency distribution of number of fibers per unit interval of length at length x. Furthermore, define $m(x)$ to be the corresponding distribution of frequency of mass (*i.e.*, weight) of fibers per unit length at the same length x in the same population. These are referred to as the differential distributions:

(2) $\quad n(x)\,dx \;=\;$ number of fibers having length within the increment between x and $(x + dx)$, and

(3) $\quad m(x)\,dx \;=\;$ mass of fibers having length within the increment between x and $(x + dx)$.

The corresponding p.d.f. for number or mass are defined as $N(x)$ and $M(x)$, respectively, which can be obtained by normalizing the number and mass frequency distributions:

(4) $$N(x) = n(x) \Big/ \int_0^\infty n(y)\,dy = N_0\, n(x), \text{ and}$$

(5) $$M(x) = m(x) \Big/ \int_0^\infty m(y)\,dy = M_0\, m(x),$$

where the quantities N_0 and M_0 are the normalization constants for number and mass, respectively. The normalized distributions (p.d.f.'s) have the property that the cumulative probability area under the curve is equal to unity:

(6) $$\int_0^\infty N(y)\,dy = 1 \quad \text{, and}$$

(7) $$\int_0^\infty M(y)\,dy = 1 .$$

Gaussian Distributions

The Gaussian distribution (also known as the normal distribution or the standard curve of error) is identified by the Central Limit Theorem as the form of the distribution of the mean values of a property of sub-samples (*e.g.*, of fibers) drawn at random from a large population. This is true mathematically, regardless of the form of the distribution of the universe population. The Gaussian frequency distribution, $\phi(x)$, is a second-order exponential function that is symmetrical and extends from negative infinity to positive infinity. It is defined mathematically by two parameters: the centroid value and the dispersion parameter, which are referred to herein respectively as α and σ:

(8) $$\phi(x) = \exp\left(-(x-\alpha)^2 / 2\sigma^2\right).$$

The normalization constant, Φ_0 , is defined by:

(9) $$\Phi_0 = \left[\int_{-\infty}^{\infty} \phi(y)\, dy\right]^{-1} = \left[\int_{-\infty}^{\infty} \exp\left(-(y-\alpha)^2/2\sigma^2\right) dy\right]^{-1} = \left[\sqrt{2\pi}\,\sigma\right]^{-1},$$

so that the normalized Gaussian p.d.f. is:

(10) $$\Phi(x) = \Phi_0\, \phi(x) = \exp\left(-(x-\alpha)^2/2\sigma^2\right) / \left(\sqrt{2\pi}\,\sigma\right),$$

for the range of x from minus infinity to plus infinity. It can be shown that the Mean Length of a Gaussian length distribution is equal to α, and the Standard Deviation is equal to σ, and furthermore that the Median Length, *L50*, is equal to α.

Log-Gaussian Distributions

Practical fiber length distributions, however, are constrained by the absence of negative values. Utilization of the Gaussian frequency for a bounded variable can be accomplished by the application of the distribution to the logarithm of the variable. This formulation, known as the log-normal distribution, is defined for the quasi-bounded range of length: $0 < x < \infty$, and has a log-normal frequency, $\theta(x)$, defined in terms of the Mode, a, and the Geometric Standard Deviation (*GSD*), γ, and its natural log, s, such that:

(11) $$s = \ln(GSD) \equiv \ln\gamma, \text{ and:}$$

(12) $$w = \ln(x) - \ln(a) = \ln(x/a), \text{ so that:}$$

(13) $$\theta(w)\, dw = \exp\left(-w^2/2s^2\right) dw.$$

For the lognormal distribution, the Mode, a, is equal to the value of the Median, *L50*. It follows that:

(14) $$dw = x^{-1}\, dx, \text{ and:}$$

(15) $$\theta(w)\, dw \propto x^{-1} \exp\left(-(\ln(x/a))^2/2s^2\right) dx,$$

so that the normalized p.d.f. can be written as:

(16) $$\Theta(w)\, dw = \Theta_0\, \theta(w)\, dw = \left(\sqrt{2\pi}\, s\right)^{-1} \theta(w)\, dw.$$

The Gaussian distributions can be meaningful in the science of fiber length when the value of α is sufficiently larger than σ, so that the unrealistic negative values of x are negligible. In that case, the lognormal and normal forms have approximately the same quality of fit to experimental data, and in fact are practically indistinguishable. To illustrate this equivalence of normal and lognormal distributions when the distribution is "tight" and well-removed from zero, set the value of the lognormal $\gamma << 1$. The value of lognormal dispersion, s, will be much closer to zero than to one, so the range of meaningful values of x is $x \approx a$, and the values $x/a \approx 1$. Using the approximation $ln(1+\Delta) \approx \Delta$ yields:

(17) $$w = \ln(x/a) \cong \ln\left(1-((x/a)-1)\right) = \ln\left(1-((x-a)/a)\right) = (x-a)/a.$$

Substituting into eq. 15 we find that:

(18) $$\theta(w) \cong \exp\left(-(x-a)^2/2a^2s^2\right),$$

which by its similarity to eq. 8 is recognizable as the form of a normal distribution having median value $a = \alpha$ and dispersion parameter given by:

(19) $$\sigma = s\,a\,.$$

In that case, the cumulative values of the normal Gaussian p.d.f. are:

(20) $$(\Phi]_{x=-\infty}^{x=0} \cong 0\,,$$ and: (21) $$(\Phi]_{x=0}^{x=-\infty} \cong 1\,,$$

so that for the bounded range $x > 0$:

(22) $$\Phi(x) \cong \exp\left(-(x-\alpha)^2 / 2\sigma^2\right) / \left(\sqrt{2\pi}\,\sigma\right), \text{ (for } 0 < x < \infty \text{ if } \alpha \gg \sigma\text{)}.$$

A useful rule of thumb is a "six-sigma" criterion, *i.e.*, that the equivalence of normal and lognormal distributions is assured when $\alpha > 6\sigma$, or in other words, when the centroid is removed from zero by six or more standard deviations.

Derivation of Results

Relations between Mass and Number Distribution

If the mass and number distributions are related by an assumption of uniform linear mass density of the fibers, then the number frequency at length x is proportional to the mass frequency at length x divided by x:

(23) $$n(x) \propto m(x)/x\,.$$

Similarly, the mass frequency at length x is proportional to the product of x and the value of the number frequency at length x:

(24) $$m(x) \propto x\,n(x)\,.$$

It follows from applying eqs. 23 and 24 to eqs. 4 and 5 that the corresponding p.d.f.'s are given by:

(25) $$N(x) = M(x) \Big/ x\int_0^\infty \left(M(y)/y\right)dy\,,$$ and

(26) $$M(x) = x\,N(x) \Big/ \int_0^\infty y\,N(y)dy = xN(x)/\mu_n\,.$$

The last two equations are especially important because they link the forms of the number-based and mass-based p.d.f.'s. The significance of that will become evident in the following derivations.

Upper Half Mean Length

The cotton fiber UHML is defined as the average length by number of the longer half of fibers by mass.
If the mass-mean length of the p.d.f. is:

(27) $$\mu_{\mathrm{m}} = \frac{\int_0^\infty x\,m(x)\,dx}{\int_0^\infty m(x)\,dx} = \alpha\,,$$

then the fiber UHML is given by:

(28) $$UHML = \mu_n \text{ of the } UH_m = \frac{\int_\alpha^\infty x\,n(x)\,dx}{\int_\alpha^\infty n(x)\,dx}$$

If the mass distribution is approximately Gaussian in the coordinate x with $\alpha \gg \sigma$ (and, therefore, indistinguishably log-Gaussian) then the mass distribution can be written as:

(29) $$m(x)\,dx \propto x^{-1}\exp\left(-(\ln x - \ln\alpha)^2 / 2s^2\right)dx\,.$$

The corresponding number distribution is:

(30) $$n(x)\,dx \quad \propto \quad x^{-2}\exp\left(-(\ln x-\ln\alpha)^2/2s^2\right)dx \;,$$

for x from 0 to ∞ if $\alpha \gg \sigma$). The fiber UHML, therefore, is equal to:

(31) $$UHML \;=\; \frac{\int_{\alpha}^{\infty} e^{\left(\frac{-(\ln x-\ln\alpha)^2}{2s^2}\right)} x^{-1}\,dx}{\int_{\alpha}^{\infty}\left(e^{\left(\frac{-(\ln x-\ln\alpha)^2}{2s^2}\right)}\right) x^{-2}\,dx} \;=\; \frac{I_1}{I_2}\,.$$

The numerator, I_1 , is

(32) $$I_1 \;=\; \int_{\alpha}^{\infty} e^{\left(\frac{-(\ln x-\ln\alpha)^2}{2s^2}\right)} x^{-1}\,dx \;=\; \frac{1}{2}\int_{-\infty}^{\infty} e^{\left(\frac{-w^2}{2s^2}\right)} dw\,,$$

which is related to the normalization constant for one-half of a Gaussian, so that the result is:

(33) $$I_1 \;=\; \sqrt{\pi/2}\;s\,.$$

The denominator, I_2, is more complicated:

(34) $$I_2 \;=\; \int_{\alpha}^{\infty}\exp\left(-(\ln x-\ln\alpha)^2/2s^2\right)\,x^{-2}\,dx$$

To evaluate this integral, use the GSD to define a coordinate transform:

(35) $$z \;=\; \left(\ln x-\ln\alpha+s^2\right)/s\,.$$

This is equivalent to:

(36) $$I_2 \;=\; (s/\alpha)\exp\left(-s^2/2\right)\int_{s}^{\infty}\exp\left(-z^2/2\right)ds$$, which can be written as:

(37) $$I_2 \;=\; (s/\alpha)\,\exp\left(-s^2/2\right)\sqrt{2\pi}\,\left[1-P(s)\right]$$, where the quantity P(s) is the cumulative normal probability integral for the normal deviate s. Now we can calculate the ratio of the two integrals as:

(38) $$\frac{I_1}{I_2} \;=\; \sqrt{\frac{\pi}{2}}\;s\,(\alpha/s)\,\exp\left(s^2/2\right)\frac{1}{\sqrt{2\pi}}\,\left[1-P(s)\right]^{-1}\,.$$

Therefore, the value of *UHML* is given finally by the expression:

(39) $$\boxed{UHML \;=\; \frac{\alpha}{2}\,\exp\left(s^2/2\right)\left[1-P(s)\right]^{-1}}\,.$$

The corresponding value of Uniformity Index is:

(40) $$\boxed{UI \;=\; 2\,\exp\left(-s^2/2\right)\left[1-P(s)\right]}\,.$$

Discussion

These last two equations are precise to several significant figures for a length distribution that is normal by mass, has a mean value equal to α, and a dispersion factor of $s = \sigma/\alpha$ less than about 1/6. The significance for breeders and fiber scientists is that there may only be two parameters (α and σ) that need to be tracked for cotton fiber on the seed. These values, however, are not the mean length and standard deviation of the ginned lint. Almost any ginning technique (including pulling by hand) and any length measurement technique (particularly AFIS) will break a substantial number of fibers, resulting in measured ML and SD that are not representative of the native values of α and σ on the seed. Unfortunately, no techniques are known for reasonably fast measurement of the cotton fiber length distribution on the seed.

Conclusion

Quantitative expressions were derived for the UHML and UI of a fiber length distribution in terms of the ML and SD of the distribution. These expressions equations are precise to several significant figures for any length distribution that has a lognormal mass p.d.f., or a normal mass p.d.f. with a ML greater than about six times the SD.

Acknowledgment

• The present work was performed while the first-author was stationed on temporary duty on the campus of the University of New Orleans before returning to SRRC after Hurricane Katrina; the cooperation of the staff of the Department of Mathematics at UNO is deeply appreciated.

Note

• The Southern Regional Research Center in New Orleans, LA, is one of the facilities of the Mid South Area, Agricultural Research Service, U.S. Department of Agriculture.

References

Robert, K. Q. 1991. Determining cotton short-fiber content from the shape of the length distribution of the longer fibers. pp. 880-886. In Proc. Beltwide Cotton Conf., San Antonio, TX. 6-10 Jan. 1991. Natl. Cotton Counc. Am., Memphis, TN.

Robert, K. Q., and Blanchard, L. J. 1997. Cotton cleanability, part i: modeling fiber breakage. Textile Research Journal 67:417-427.

Robert, K. Q., Price, J. B., and Cui, X. 2000. Cotton cleanability, part ii: effect of simple random breakage on fiber length distribution, Textile Research Journal 70(2):108-115.

Robert, K. Q. 2005a. The original cotton fiber length distribution. p. 2673-2686. In Proc. Beltwide Cotton Conf., New Orleans, LA 4-7 Jan. 2005. Natl. Cotton Counc. Am., Memphis, TN.

Robert, K. Q. 2005b. Cotton fiber breakage and its relation to length distribution, short fiber, and uniformity. p. 3074-3086. In Proc. Beltwide Cotton Conf., New Orleans, LA 4-7 Jan. 2005. Natl. Cotton Counc. Am., Memphis, TN.