Tables And Characteristics of the Standardized Lognormal Distribution

Nick T. Thomopoulos, Stuart Graduate School of Business, Illinois Institute of Technology, 565 W. Adams St., Chicago, IL 60661, Phone: 312-906-6536, Fax: 312-906-6549

Arvid C. Johnson, Graduate School of Business & Information Systems, Dominican University, 7900 W. Division St., River Forest, IL 60305, Phone: 708-524-6465, Fax: 708-524-6939

Proceedings of the Decision Sciences Institute, 2003, pp. 1031-1036

Abstract: While lognormal distributions have demonstrated great utility in a number of applications related to decision sciences, practitioners find few – if any – tables of its cumulative distribution function available to support their work. This paper describes a "standardized" form of the lognormal distribution and a methodology by which tables of its cumulative distribution function can be generated. This paper, then, provides these reference tables and illustrates their use.

INTRODUCTION

A lognormally-distributed random variable is a random variable whose logarithm is normally-distributed. Johnson, *et. al.*, (1994) note that some practitioners maintain "that the lognormal distribution is as fundamental as the normal distribution" and that the lognormal distribution has found applications in fields including the physical sciences, life sciences, social sciences, and engineering. Aitchison and Brown (1957) is the classic reference on the lognormal distribution, and the compendium edited by Crow and Shimizu (1988) provides a more recent update.

Despite the lognormal distribution's utility, practitioners find few – if any – tables of its cumulative distribution function available to support their work. Moshman (1953) has published selected upper and lower percentile points (0.5%, 1%, 2.5%, 5%, and 10%) as a function of the shape parameter. Similarly, Broadbent (1956) provides upper and lower 1% and 5% values as a function of the coefficient of variation. It is the objective of this paper to provide practitioners with more comprehensive tables of the cumulative distribution function of the lognormal distribution. Additionally, we will illustrate a methodology by which "critical values" of a "standardized" lognormal distribution may be calculated directly.

THE LOGNORMAL DISTRIBUTION

Consider a lognormally-distributed variable x with mean μ_x and standard deviation σ_x– denoted $LN(\mu_x, \sigma_x^2)$. The variable y, where

$$y = ln(x) , (1)$$

is normally-distributed with mean μ_y and standard deviation σ_y and is denoted $N(\mu_y, \sigma_y^2)$. The probability density function $f(x)$ of the lognormal distribution is given by

$$1 \neg \ln x - \mu_y)^{2} - \square\ \square 1 2 \square\ \sigma_y \square \sqrt{2\pi}\, e^{=\Downarrow} (2)\ x s_{y} = f \quad (x)_{2p}$$

as noted, for example, in Hines and Montgomery (1990). As can be seen from Figure 1, the distribution is skewed with a longer tail to the right of the mean.

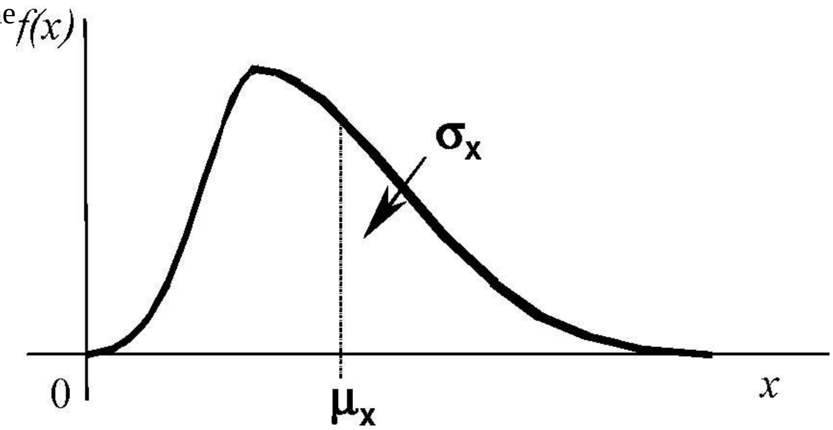

Figure 1 - Probability density function of the lognormal distribution.

Aitchison and Brown (1957) note that, when μ_y and σ_y are known for *y*, the corresponding mean and variance for *x* can be found from the following:

$$\mu_x = e^{\,m + \sigma\, \frac{1}{2}\, {}_{y}{}^{2}{}^{y}} \quad (3)$$

$$s = e^{\,2m + \sigma_y s\ {}^{22}_{2y\ \ y}} \left(e - 1 \right) \quad (4)$$

$$x$$

Similarly, using Equation (1), when μ_x and σ_x are known for *x*, the corresponding mean and variance for *y* can be determined from the following:

$$2$$

$$\mu$$

$$\mu_y = \ln \frac{\mu_x^2}{\sqrt{\mu_x^2+\sigma_x^2}} \quad (5)$$

$$\sigma_y^2 = \ln\left(1+\frac{\sigma_x^2}{\mu_x^2}\right) \quad (6)$$

Johnson, *et. al*., (1994) note various ways in which the lognormal distribution has been standardized. In this paper, we study the properties of the "standardized lognormal distribution" that arises when the mean of its normal counterpart is zero – i.e., $\mu_y = 0$ so that y is $N(0, \sigma_y^2)$. For this case, the mean and variance of x become

$$m_x = e^{\frac{1}{2}s_y^2} \quad (7)$$

$$s_x^2 = e^{s_y^2}\left(e^{s_y^2} - 1\right) \quad (8)$$

In the event that the mean of *y* is not equal to zero, the random variable can be transformed into standardized form *y'* as follows:

$$y' = y - \mu_y \quad (9)$$

We use this standardized lognormal distribution form to develop tables for the cumulative distribution function *F(x)* in the next section.

TABLES OF THE CUMULATIVE DISTRIBUTION FUNCTION

In this section, we present a methodology by which "critical values" of the standardized lognormal distribution may be calculated. We, then, introduce reference tables for use by practitioners.

The "critical value" associated with α^{th} percentile point of the standardized lognormal distribution is denoted at x_α and corresponds to the value of the lognormally-distributed random variable *x* at which the cumulative distribution function *F(x)* is equal to α.

$$F(x_a) = P(x \le x_a) = \alpha \quad (10)$$

Let Z_α denote the value of the standardized *normal* variate associated with the α^{th} percentile of that distribution. Then, x_α can be calculated as follows:

$$x_a = e^{m_y + Z_a s_y} \quad (11)$$

In this way, tables of the critical values of the lognormal variable *x* can be generated for general use. Such tables are provided in Tables 1, 2, and 3 of this paper. All of the tables are generated for the case in which the counterpart normal variable *y* is $N(0, \sigma_y^2)$. The tables are of the form $F(x_\alpha) = P(x \le x_\alpha) = \alpha$, where the particular values of α are as follows: 0.01(0.01)0.05, 0.05(0.05)0.95, and 0.95(0.01)0.99.

The tables list the values of x_α for $F(x_\alpha) = \alpha$, and the tables are arranged by σ_y as follows:

- Table 1: $\sigma_y = 0.1(0.1)1.0$

- Table 2: σ_y = 1.0(0.1)2.0
- Table 3: σ_y = 2.0(0.1)3.0

y: N(my, s y 2)	my s y	0 0.10	0 0.20	0 0.30	0 0.40	0 0.50	0 0.60	0 0.70	0 0.80	0 0.90	0 1.00
x: LN(m x, s x 2)	mx s x Cx	1.005 0.101 0.100	1.020 0.206 0.202	1.046 0.321 0.307	1.083 0.451 0.417	1.133 0.604 0.533	1.197 0.788 0.658	1.278 1.016 0.795	1.377 1.304 0.947	1.499 1.675 1.117	1.649 2.161 1.311
F(x)		Values of xa for F(x a) = a									
0.01		0.792	0.628	0.498	0.394	0.312	0.248	0.196	0.156	0.123	0.098
0.02		0.814	0.663	0.540	0.440	0.358	0.292	0.237	0.193	0.157	0.128
0.03		0.829	0.686	0.569	0.471	0.390	0.324	0.268	0.222	0.184	0.152
0.04		0.839	0.705	0.591	0.496	0.417	0.350	0.294	0.246	0.207	0.174
0.05		0.848	0.720	0.611	0.518	0.439	0.373	0.316	0.268	0.228	0.193
0.10		0.880	0.774	0.681	0.599	0.527	0.464	0.408	0.359	0.316	0.278
0.15		0.902	0.813	0.733	0.661	0.596	0.537	0.484	0.436	0.393	0.355
0.20		0.919	0.845	0.777	0.714	0.657	0.604	0.555	0.510	0.469	0.431
0.25		0.935	0.874	0.817	0.764	0.714	0.667	0.624	0.583	0.545	0.509
0.30		0.949	0.900	0.854	0.811	0.769	0.730	0.693	0.657	0.624	0.592
0.35		0.962	0.926	0.891	0.857	0.825	0.794	0.764	0.735	0.707	0.680
0.40		0.975	0.951	0.927	0.904	0.881	0.859	0.837	0.817	0.796	0.776
0.45		0.988	0.975	0.963	0.951	0.939	0.927	0.916	0.904	0.893	0.882
0.50		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.55		1.013	1.025	1.038	1.052	1.065	1.078	1.092	1.106	1.120	1.134
0.60		1.026	1.052	1.079	1.107	1.135	1.164	1.194	1.225	1.256	1.288
0.[illegible]		1.039	1.080	1.123	1.167	1.212	1.260	1.310	1.361	1.415	1.47[illegible]

Table 1 – Critical Values of x_α for σ_y from 0.1 to 1.0 ($\mu_y = 0$).

Table 2 – Critical Values of x_α for σ_y from 1.0 to 2.0 ($\mu_y = 0$).

Table 3 – Critical Values of x_α for σ_y from 2.0 to 3.0 ($\mu_y = 0$).

The tables also list the mean, standard deviation, and coefficient of variation c_x of the associated lognormally-distributed random variable *x*. To use these tables, the lognormal variable needs to be scaled so that its average is less than or equal to 90, which corresponds to the largest mean value listed on Table 3.

In reviewing the three tables, some comments on the shape of the lognormal variable can be made. (Recall that *x* is the lognormal variable, and *y* is the associated normal variable with mean zero and standard deviation σ_y.)

y: N(my, s y 2)	my s y	0 0.10	0 0.20	0 0.30	0 0.40	0 0.50	0 0.60	0 0.70	0 0.80	0 0.90	0 1.00
x: LN(m x, s x 2)	mx s x Cx	1.005 0.101 0.100	1.020 0.206 0.202	1.046 0.321 0.307	1.083 0.451 0.417	1.133 0.604 0.533	1.197 0.788 0.658	1.278 1.016 0.795	1.377 1.304 0.947	1.499 1.675 1.117	1.649 2.161 1.311
F(		Values of xa for F(x a) = a									

y: N(my, s y 2)	my s y	0 0.10	0 0.20	0 0.30	0 0.40	0 0.50	0 0.60	0 0.70	0 0.80	0 0.90	0 1.00
x: LN(m x, s x 2)	mx s x Cx	1.005 0.101 0.100	1.020 0.206 0.202	1.046 0.321 0.307	1.083 0.451 0.417	1.133 0.604 0.533	1.197 0.788 0.658	1.278 1.016 0.795	1.377 1.304 0.947	1.499 1.675 1.117	1.649 2.161 1.311
F(x)		Values of xa for F(x a) = a									
0.01		0.792	0.628	0.498	0.394	0.312	0.248	0.196	0.156	0.123	0.098
0.02		0.814	0.663	0.540	0.440	0.358	0.292	0.237	0.193	0.157	0.128
0.03		0.829	0.686	0.569	0.471	0.390	0.324	0.268	0.222	0.184	0.152
0.04		0.839	0.705	0.591	0.496	0.417	0.350	0.294	0.246	0.207	0.174
0.05		0.848	0.720	0.611	0.518	0.439	0.373	0.316	0.268	0.228	0.193
0.10		0.880	0.774	0.681	0.599	0.527	0.464	0.408	0.359	0.316	0.278
0.15		0.902	0.813	0.733	0.661	0.596	0.537	0.484	0.436	0.393	0.355
0.20		0.919	0.845	0.777	0.714	0.657	0.604	0.555	0.510	0.469	0.431
0.25		0.935	0.874	0.817	0.764	0.714	0.667	0.624	0.583	0.545	0.509
0.30		0.949	0.900	0.854	0.811	0.769	0.730	0.693	0.657	0.624	0.592
0.35		0.962	0.926	0.891	0.857	0.825	0.794	0.764	0.735	0.707	0.680
0.40		0.975	0.951	0.927	0.904	0.881	0.859	0.837	0.817	0.796	0.776
0.45		0.988	0.975	0.963	0.951	0.939	0.927	0.916	0.904	0.893	0.882
0.50		1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.55		1.013	1.025	1.038	1.052	1.065	1.078	1.092	1.106	1.120	1.134
0.60		1.026	1.052	1.079	1.107	1.135	1.164	1.194	1.225	1.256	1.288
0.[illegible]		1.039	1.080	1.123	1.167	1.212	1.260	1.310	1.361	1.415	1.47[illegible]

In all situations, the median of *x* (lognormal) occurs at *x* = 1 – so 50% of all observations of *x* lie between 0 and 1 and 50% above 1. When σ_y is small, *x* is shaped similarly to a normal distribution. This is seen in Table 1 when $\sigma_y = 0.10$ since the range of x values below $x = 1$ is almost the same as the range of x values above 1. Also, the coefficient of variation of x is 0.10, which is much like a normal distribution.

As σ_y increases, the mode, which is given by

$$\text{mode} = e^{-\sigma_y^2}, \quad (12)$$

starts to approach zero, and the shape is more skewed to the right. For example, note that, when $\sigma_y = 1.00$, $c_x = 1.31$ which indicates that the distribution is highly skewed to the right – i.e., while 50% of the observations range between 0 to 1, 50% range from 1 to 6 or higher. In all situations, the probability density of x starts out small, increases to a peak at the mode ($x < 1$), and then spreads out to the right to large numbers indeed. Recall that the probability density of the exponential distribution ($c_x = 1$) is large at $x = 0$ and continually drops as x increases. For a lognormally-distributed random variable x with $\sigma_y = 0.8$, $c_x = 0.95$ – which is similar to an exponential except that the probability density of x starts out small at $x = 0$, quickly rises to the mode, and then drops for all remaining values of x.

In the next section, we use two examples to illustrate how practitioners may use the tables.

USE OF THE TABLES

Example 1

Consider the lognormal x where it is desired to find the middle 90 percent range of values when y is $N(0, 1.4^2)$. Table 2 applies, and the entries show $x_{0.05} = 0.10$ for $F(x) = 0.05$ and $x_{0.95} = 10.01$ for $F(x) = 0.95$. Thus, the 90% range for x is between 0.10 and 10.01. Further, x has a mean of 2.66, a standard deviation of 6.58, and a coefficient of variation of 2.47.

Example 2

Suppose a sample of size n from a lognormal variable x' is observed. For each x', the natural log transformation is taken as $y' = \ln(x')$, and y' is, thereby, normal. This gives y_i' ($i = 1$ to n) that yield estimates for the mean and variance of y'. For convenience, the estimated parameters of y' are denoted as $\mu_{y'}$ and $\sigma_{y'}$. To apply the tables of this paper, it is necessary to convert the variable y' to y – where y is also normal but has a mean of zero. This can be done by utilizing Equation 9.

$$y = y' - \mu_{y'} \quad (13)$$

So now $\mu_y = 0$ and $\sigma_y = \sigma_{y'}$, and y is normal and consistent with the table entries. Assume that the sample gives $\mu_{y'} = 4$ and $\sigma_{y'} = 1.4$. Then,

$$y = y' - \mu_{y'} = y' - 4 \quad (14)$$

is also normal with a mean of zero and a standard deviation of $\sigma_y = 1.4$ – i.e., y is $N(0, 1.4^2)$. The 90 percent middle range on x is obtained in the following way:

1. For $y = N(0, 1.4^2)$, $x_{0.05} = 0.10$ and $x_{0.95} = 10.01$, as in Example 1.
2. Rearranging Equation 13, gives $y' = y + \mu_{y'}$. Similarly, Equation 1 gives $x = e^y$. These can be combined as follows:

$$x' = e^{y'} = e^{y} e^{m_{y'}} = x e^{\mu_{y'}} \quad (15)$$

3. Applying Equation 15 to the results from step 1, we find:

$$x'_{0.05} = x_{0.05} e^{4} = 0.10 * 54.59 = 5.46$$

$$x'_{0.95} = x_{0.95} e^{4} = 10.01 * 54.59 = 546$$

Thus, the 90% range for *x'* is between 5.46 and 546. Similarly, the mean of *x'* becomes 145.25.

CONCLUSION

This paper has illustrated a methodology by which "critical values" of a "standardized" lognormal distribution (one in which the counterpart normal distribution has a mean of zero) may be calculated directly. This methodology was, then, utilized to develop tables of the cumulative distribution function for use by practitioners. Finally, the use of these tables was illustrated.

REFERENCES

Aitchison, J., and J.A.C. Brown (1957), *The Lognormal Distribution*, Cambridge University Press, Cambridge.

Broadbent, S.R. (1956), "Lognormal Approximation to Products and Quotients," *Biometrika*, **43**, 404-417.

Crow, E.L., and K. Shimizu (1988), *Lognormal Distributions: Theory and Applications*, Marcel Dekker, New York.

Hines, W.W., and D.C. Montgomery (1990), *Probability and Statistics in Engineering and Management Science*, 3rd Edition, John Wiley & Sons, New York.

Johnson, N.L., N. Balakrishnan, and S. Kotz (1994), *Continuous Univariate Distributions 1*, 2nd Edition, John Wiley & Sons, New York.

Moshman, J.E. (1953), "Critical Values of the Log-normal Distribution," *Journal of the American Statistical Association*, **48**, 600-609.