BROADLEAF CAPITAL INTERNATIONAL PTY LTD
ABN 24 054 021 117

23 Bettowynd Road
Pymble
NSW 2073
Australia
www.Broadleaf.com.au

Tel: +61 2 9488 8477
Mobile: +61 419 433 184
Fax: + 61 2 9488 9685
Cooper@Broadleaf.com.au

Specialists in Strategic, Enterprise and Project Risk Management

LOGNORMAL DISTRIBUTION SUMMARY

Introduction

I was stuck recently in a distant part of Papua New Guinea without reference sources. I had a lognormal distribution defined in terms of its mean and 95-percentile values, and I needed help in determining its standard deviation. Many people from the RISKANAL list responded to my request (see the list below; many thanks, all of you) with a wealth of specific and general information. I have summarised the main points here.

Definition

A random variable X is said to follow a lognormal distribution if the random variable Y = log(X) is normally distributed, N(mu, sigma ^ 2). A lognormal distribution is defined by a density function of

f(y) = EXP(- ((LOG(y) – mu) ^ 2) / (2 * sigma ^ 2)) / (y * sigma * SQR(2 * pi)), for y > 0.

Lognormal distributions are typically specified in one of two ways throughout the literature. One way is to specify the mean and standard deviation of the underlying normal distribution (mu and sigma) as described above. The other way is to specify the distribution using the mean of the lognormal distribution itself and a term called the “error factor.” The error factor for a lognormal distribution is defined as the ratio of the 95^{th} percentile to the median, or, equivalently, the ratio of the median to the 5^{th} percentile. Physically, its square represents the width of a 90% confidence interval with respect to the median. The mathematical relationship between the input mean and error factor, and the parameters of the underlying normal distribution (mu and sigma) is shown by the following relations:

sigma = LOG(error factor) / 1.645
mu = LOG(mean) – (sigma ^ 2 / 2)

When the mean and error factor are used as input for the lognormal distribution, both input parameters must be positive, and the error factor must be greater than one. If mu and sigma are specified, there is no restriction on mu, but sigma must be positive.

Formulae

Two parameters are generally sufficient to define a lognormal distribution.

The majority (but not all) of the formulae listed below are taken from a freeware program called LOGNORM4 for uniquely determining the parameters of lognormal distributions from minimal information (e.g. a mean and a median), and for manipulating and generating lognormal distributions. This was written by Daniel J. Strom and is available as freeware from the author at the Risk Analysis and Health Protection Group, MSIN K3-56, Battelle Pacific Northwest Laboratories, PO Box 999,

Richland, Washington 99352-0999 USA; phone (509) 375-2626; fax (509) 375-2019; Internet daniel.j.strom@pnl.gov.

It was sent to me by Robert Lee. Many of the other contributions are reflected in this set of formulae.

"You are asked to choose values or pairs of values that uniquely determine a lognormal distribution. From these, the distribution parameters [mean, median, mode, geometric standard deviation (GSD) mu (= ln(median)), sigma (=ln(GSD)), standard deviation (SD), coefficient of variation (CV), variance, skewness, and kurtosis] are calculated. You can then use the resultant distribution

1) to determine percentiles, quantiles, or z-values for values you supply, or
2) to determine values for percentiles, quantiles, or z-values you supply.

For example, if you supply the geometric mean (that is, the median) and the GSD of a lognormal distribution, this program calculates the arithmetic mean (the same as the average and the expectation value)."

```
sigma    =  SQR(2 * LOG(mean / median))
         =  SQR(2 * LOG(mean / mode) / 3)
         =  SQR(LOG(median / mode))
         =  LOG(value1 / median) / z1
         =  SQR(LOG(CV ^ 2 + 1))
         =  LOG(GSD)
         =  LOG(value1 / value2) / (z1 - z2)
         =  (LOG(value1) – LOG(value2)) / (z1 - z2)
         =  SQR(mu - LOG(mode))
         =  LOG(error factor) / 1.645
```

[Note: value1 = EXP(mu + z1 * sigma), where z1 is derived from the standard normal cdf]

```
(SQR(EXP(sigma^2)-1) * EXP(sigma^2/2)) - SD/median  =  0
(SD ^ 2 / value1 ^ 2) – (EXP(-2 * z1 * sigma + sigma ^ 2) * (EXP(sigma ^ 2) - 1))  =  0
LOG(mean) - LOG(value1) - 0.5 * (sigma ^ 2) + (z1 * sigma)  =  0
```

[Note: These equations must be solved numerically. They may yield 1 or 3 values of sigma for z1 > 1]

```
If mean < value, then      sigma = z1 + SQR(z1 ^ 2 + 2 * LOG(mean / value))
If mean > value, then      sigma = z1 - SQR(z1 ^ 2 + 2 * LOG(mean / value))
If mode > value, then      sigma = (-z1 + SQR(z1 ^ 2 - 4 * LOG(mode / value))) / 2
If mode < value, then      sigma = (-z1 - SQR(z1 ^ 2 - 4 * LOG(mode / value))) / 2
```

```
GSD      =  EXP(sigma)
median   =  mean * EXP(-sigma ^ 2 / 2)
         =  mode * EXP(sigma ^ 2)
         =  EXP(mu)
mean     =  median * EXP(sigma ^ 2 / 2)
mu       =  LOG(median)
         =  LOG(mean) – (sigma ^ 2 / 2)
         =  LOG(mode) + sigma ^ 2
         =  (z2 * LOG(value1) - z1 * LOG(value2)) / (z2 - z1)
         =  LOG(value1) + LOG(value2 / value1) * (0 - z1) / (z2 - z1)
         =  LOG(value1) - sigma * z1
         =  LOG(mode) + LOG(CV ^ 2 + 1)
mode     =  EXP(mu - sigma ^ 2)
         =  median * EXP(-sigma ^ 2)
mode ^ 2 * SD ^ 2 - median ^ 4 + median ^ 3 * mode = 0   [Note: must be solved numerically]
```

CV	=	SQR(EXP(sigma ^ 2) - 1)
	=	SD / mean
SD	=	CV * mean
	=	mean * SQR(EXP(sigma ^ 2) - 1)
	=	EXP(mu + (sigma ^ 2) / 2) * SQR(EXP(sigma ^ 2) - 1)
variance	=	SD ^ 2
	=	EXP(2 * mu + sigma ^ 2) * (EXP(sigma ^ 2) - 1)
	=	(mean ^ 2) * (EXP(sigma^2) – 1)
skewness	=	CV^3 + 3 * CV
kurtosis	=	CV^8 + 6 * CV^6 + 15 * CV^4 + 16 * CV^2
Zmode	=	- sigma
Zmean	=	sigma / 2
Zmedian	=	0
value	=	EXP(mu + zvalue * sigma)
Zvalue	=	(LOG(value) - mu) / sigma
jth moment	=	EXP(j * mu + ½ * (j^2) * sigma^2)

Combining Independent Lognormal Distributions

"Again, assuming independence of each factor, the probability distributions can now be combined. This is particularly simple if each distribution can be treated as approximately lognormal. In such instances, the final distribution is lognormal with the logarithmic standard deviation given by the square root of the sum of squares of the individual geometric standard deviations. If the distributions are far from lognormal, Monte Carlo methods can be used to combine them."

The product of two (independent) L-variates is also a L-variate such that if x1 is L(mu1; sigma1^2) and x2 is L(mu2; sigma2^2), then x1*x2 is L(mu1+mu2; sigma1^2+sigma2^2)

If x is L(mu; sigma^2) and b and c are constants, where c>0, (say c= exp(a)), then c*x^a is L(a+b*mu; b^2*sigma^2)

Manual Calculation

"If you have the mean (median) and 95-percentile, you can plot those values on log-probability paper, connect them with a straight line, and use that line to calculate the geometric standard deviation.

"The geometric standard deviation is the 84.13% value divided by the 50% value, which equals the 50% value divided by the 15.87% value. Provided that the distribution is log-normal or at least a close approximation.

"Additionally, 95% of all values will lie between the (geometric mean)/(sigma-g)^2 and (sigma-g)^2/(geometric mean).

"Also, for a lognormal distribution, 95% of the observations will lie BELOW exp(mu + 1.65sigma), where mu is the mean of the log of the original data and sigma is the standard deviation of the log values."

Excel Calculation

"If you have Excel, you can use the loginv function and GoalSeek to find the GSD (specify the GSD by reference to another cell, and have GoalSeek optimize the cell value for the specified 95th percentile)."

Contributors

The following people contributed information used in this summary. Many thanks, and please excuse me if you were not quoted verbatim – I was trying to simplify and condense.

Clark Carrington	RiskyLogic@compuserve.com
Jerry Falo	Gerald.Falo@APG.AMEDD.ARMY.MIL
Bill Huber	whuber@quantdec.com
Jim Irish	Jim.Irish@uts.edu.au
Yongsung Joo	yj12@cornell.edu
Robert Lee	RLee@GOLDER.com
Fred Leverenz	leverenz@BATTELLE.ORG
Ray Martin	Ray_Martin@AltaVista.net
Joseph Minarick	JOSEPH.W.MINARICK.III@cpmx.saic.com
Sim Ooi	oois@mel.egisconsult.com.au
Peter Smit	pj.smit@pers.vhall.nl
Greg Wyss	gdwyss@sandia.gov

References

Aitchison, J. and Brown, J.A.C. The Lognormal Distribution. Cambridge: Cambridge University Press, 1957.

Strom, D.J. Determining Parameters for Lognormal Distributions from Minimal Information. Submitted to Risk Analysis; 1993.

Gregory D. Wyss and Kelly H. Jorgensen, A User's Guide to LHS: Sandia's Latin Hypercube Sampling Software, Risk Assessment and Systems Modeling Department, Sandia National Laboratories, PO Box 5800, Albuquerque, NM 87185-0747, SAND98-0210, February 1998.

Additional Below-Detection-Limit References

I received a number of contributions relating to estimation, including estimation of distributions with missing data (e.g. data below a detection threshold). I have included the references here for information. They were compiled by DJ Strom and HR Pritchard, rev. July 28, 1992.

Atwood CL, Blackwood LG, Harris GA, Loehr CA. Recommended methods for statistical analysis of data containing less-than-detectable measurements. Springfield, VA: National Technical Information Service; Rev. 1; EGG-SARE--9247-Rev.1; 1991

Finkel, A.M. A Simple Formula for Calculating the "Mass Density" of a Lognormally Distributed Characteristic: Applications to Risk Analysis. Risk Analysis 10(2):291-301; 1990

Gilbert RO, Kinnison RR. Statistical Methods for Estimating the Mean and Variance from Radionuclide Data Sets Containing Negative, Unreported or Less-Than Values. Health Phys. 40:377-390, 1981.

Helsel DR. Less than obvious: statistical treatment of data below the detection limit. Environ. Sci. Technol. 24(12):1766-1774, 1990.

Hertzler CL, Atwood CL, Harris GA. Current Methods of Handling Less-Than-Detectable Measurements and Detection Limits in Statistical Analysis of Environmental Data. Idaho Falls, ID: Idaho National Engineering Laboratory; EGG-SARE--8609; 1989. See Atwood et al. 1991.

Hornung R, Reed LD. Estimation of Average concentration in the Presence of Nondetectable Values. Appl. Occup. Environ. Hyg. 5(1):46-51, 1990.

Newman MC, Dixon PM. UNCENSOR: A program to estimate means and standard deviations for data sets with below detection limit observations. American Environmental Laboratory, April, 1990.

Newman MC, Dixon PM, Looney BB, Pinder III JE. Estimating Mean and Variance for Environmental Samples with Below Detection Limit Observations. Water Resources Bulletin 25(4):905-916, 1989.

Nielson KK, Rogers VC. Statistical Estimation of Analytical Data Distributions and Censored Measurements. Analytical Chemistry 61:2719-2724, 1989.

Rappaport SM, Selvin S. A Method for Estimating the Mean from a Lognormal Distribution. Am. Ins. Hyg. Assoc. J. 48(4):374-389, 1987.

Selvin S, Rappaport S, Spear R, Schulman J, Francis M. A Note on the Assessment of Exposure Using One-Sided Tolerance Limits. Am. Ind. Hyg. Assoc. J. 48(2):89-93, 1987.

Strom DJ. Estimating Individual and Collective Doses to Groups With 'Less than Detectable' Doses: A Method for Use in Epidemiologic Studies. Health Phys. 51(4):437-445, 1986.

Strom DJ. LOGNORML. (Earlier version: LPROBIT). Oak Ridge, TN: Radiation Shielding Information Center Code No. PSR-370; RSIC Newsletter 325:5, December, 1991. [Note: this version is obsolete.]

Taylor NA. Estimation of dose received when dosemeter results are recorded below a threshold level. J. Radiol. Prot. 11(3) 191-198, 1991.

Troyer GL, Jones RA, Jensen L. The Utility of Reporting Negative Counting Values. Radioactivity and Radiochemistry 2(2):48-56, 1991.

Waite DA. Interpretation of Environmental Radioactivity Measurements. In: CRC Handbook of Environmental Radiation, AW Klement, ed. Boca Raton: Chemical Rubber Company, 1982.

Waters MA, Selvin S, Rappaport SM. A Measure of Goodness-of-Fit for the Lognormal Model Applied to Occupational Exposures. Am. Ind. Hyg. Assoc. J. 52(11):493-502, 1991.

Compiled by:
Dr Dale F Cooper
Cooper@Broadleaf.com.au

June 1999