Everything Is Lognormal… or Is It?

Daniel J. Strom

Pacific Northwest National Laboratory

Richland, Washington USA

+1 509 375 2626 strom@pnl.gov

Paper TAM-F.11

52nd Annual Meeting of the Health Physics Society, Portland, Oregon

Tuesday, July 10, 2007

Work supported by the U.S. Department of Energy under Contract No. DE-AC05-76RLO 1830

Battelle

Pacific Northwest National Laboratory

Introduction

- Variability
  - Occupational and environmental radiation monitoring data are usually lognormally-distributed (see, e.g., UNSCEAR 1977 Annex E)
    - external dosimetry
    - bioassay
    - dose rate
    - air samples
    - surface contamination
- Uncertainty
  - When used in support of compensation decisions, each reconstructed dose must be characterized by a quantitative expression of uncertainty in the form of a distribution with associated parameters
- In many cases, variability is the largest contributor to uncertainty
  - Example: Time-weighted average exposures to airborne radioactivity in the workplace (Davis & Strom, MPM-A.6)

Problems

- Paucity of data
- Data are summarized by parameters such as minimum, average, and maximum values
- Left-censoring

Determining a lognormal distribution from minimal information

1	the mean and median (or their natural logs)
2	the mean and mode (or their natural logs)
3	the median and mode (or their natural logs)
4	the median (or its natural log) and the *GSD* or sigma = ln(*GSD*)
5	the mean (or its natural log) and the *GSD* or sigma = ln(*GSD*)
6	the mode (or its natural log) and the *GSD* or sigma = ln(*GSD*)
7	a value and its percentile OR fractile OR std norm deviate and *GSD* or sigma=ln(*GSD*)
8	the median and a value with its percentile OR fractile OR std normal deviate
9	the mean and a value with its percentile OR fractile OR std normal deviate
10	the mode and a value with its percentile OR fractile OR std normal deviate
11	the median and [arithmetic] standard deviation OR coefficient of variation
12	the mean and [arithmetic] standard deviation OR coefficient of variation
13	the mode and [arithmetic] standard deviation OR coefficient of variation
14	a value and its percentile OR fractile OR std norm deviate and [arithmetic] *SD* or *CV*
15	a pair of values and their percentiles OR fractiles OR std normal deviates
16	minimum, maximum, and mean values

LOGNORM4

- The first 15 can be done by LOGNORM4.EXE
  - (Alt-Enter for full screen)
- Strom and Stansbury (2000)

Using Minimum, Mean, and Maximum Values without Number of Observations

- if the x_{min} and x_{max} values are symmetric about the geometric mean x_{50}, then the 3 values uniquely determine a lognormal distribution
- $f_{min} = 1 - f_{max}$, so that $-z_{min} = z_{max}$

$$\mu = \frac{\ln x_{min} + \ln x_{max}}{2} \text{ or}$$

$$x_{50} = \sqrt{x_{min} x_{max}}$$

$$\sigma^2 = 2 \ln \bar{x} - \ln x_{min} - \ln x_{max}$$

$$\sigma = \sqrt{2 \ln \bar{x} - \ln x_{min} - \ln x_{max}}$$

Warning! May give lousy or even impossible results if one of the extremes is an outlier.

$$z_{min} = \frac{\ln x_{min} - \mu}{\sigma}$$

$$\text{Check: } n \approx \frac{1}{2 f_{min}} ?$$

Censored Individual Observations

- Sometimes values are reported as “less-than” some number or as zero
  - This is referred to as **left-censoring**
- One cannot take the logarithm of zero or a less-than value
- Simple averaging of natural logs won’t work

The Lognormal Fitting Utility

- Strom (2007, 2007a)
- Consider this data set: 18, <2, 5, <2, 2, 3, <2, 8
- Lognormal Fitting Utility

Combining Multiple Uncertainties

- Uncertainty distributions due to *n* different causes can be multiplied:

$$U = \prod_{i=1}^{n} U_i$$

- Lognormals have a convenient analytical property:

$$\Lambda\left(\mu_1, \sigma_1^2\right)\Lambda\left(\mu_2, \sigma_2^2\right) = \Lambda\left(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2\right)$$

$$\Lambda = \prod_{i=1}^{n} \Lambda_i = \Lambda\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Sums of Lognormally Distributed Variables

$$y = \sum_i x_i \text{, where } x_i \text{ are sampled from lognormals}$$

$$\bar{y} = \sum_i \bar{x}_i$$

$$y_{50} > \sum_i x_{50,\,i}$$

- The distribution of the sum *y* is
  - very skewed
  - not even remotely Normal
  - not too different from a lognormal
  - "…the central-limit theorem is, at best, a weak theorem for the case of the lognormal. …the distribution of the sum of log-normal variates for many cases of interest is very accurately represented by a lognormal" (Mitchell 1968)

Conclusions

- No, not everything is lognormal, but many quantities of interest are
- Lognormal distributions have useful properties
  - Simple, analytical products (not true for normal distributions)
  - Significant immunity to the central limit theorem for sums
- Both LOGNORM4 and the Lognormal Fitting Utility can be downloaded at http://qecc.pnl.gov/LOGNORM4.htm

References

Aitchison J and JAC Brown. 1981. *The Lognormal Distribution.* Cambridge University Press, Cambridge.

Mitchell RL. 1968. "Permanence of the Log-Normal Distribution." Journal of the Optical Society of America 58(9):1267-1272.

Strom DJ. 1994. *LOGNORM4: Determining Parameters of Lognormal Distributions from Minimal Information.* http://qecc.pnl.gov/LOGNORM4.htm, Pacific Northwest National Laboratory, Richland, Washington.

Strom DJ. 2006. *Lognormal Fitting Utility Program*. http://qecc.pnl.gov/Lognormal_Fitting_Utility.htm, Battelle, Richland, Washington.

Strom DJ. 2007. *Default Assumptions and Methods for Atomic Weapons Employer Dose Reconstruction*. Battelle TIB-5000 rev. 0; http://www.cdc.gov/niosh/ocas/awedocs.html, NIOSH, Cincinnati, Ohio.

Strom DJ. 2007a. "Statistical Issues in Applied Health Physics." Chapter 6 in *Radiological Assessment: A Handbook for Detection, Identification, and Evaluation*, ed. KA Higley. Medical Physics Publishers, Madison, Wisconsin.

Strom DJ and PS Stansbury. 2000. "Determining Parameters of Lognormal Distributions from Minimal Information." American Industrial Hygiene Association Journal 61(Nov-Dec):877-880.