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
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

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>the mean and median (or their natural logs)</td>
</tr>
<tr>
<td>2</td>
<td>the mean and mode (or their natural logs)</td>
</tr>
<tr>
<td>3</td>
<td>the median and mode (or their natural logs)</td>
</tr>
<tr>
<td>4</td>
<td>the median (or its natural log) and the $GSD$ or sigma = ln($GSD$)</td>
</tr>
<tr>
<td>5</td>
<td>the mean (or its natural log) and the $GSD$ or sigma = ln($GSD$)</td>
</tr>
<tr>
<td>6</td>
<td>the mode (or its natural log) and the $GSD$ or sigma = ln($GSD$)</td>
</tr>
<tr>
<td>7</td>
<td>a value and its percentile OR fractile OR std norm deviate and $GSD$ or sigma=ln($GSD$)</td>
</tr>
<tr>
<td>8</td>
<td>the median and a value with its percentile OR fractile OR std normal deviate</td>
</tr>
<tr>
<td>9</td>
<td>the mean and a value with its percentile OR fractile OR std normal deviate</td>
</tr>
<tr>
<td>10</td>
<td>the mode and a value with its percentile OR fractile OR std normal deviate</td>
</tr>
<tr>
<td>11</td>
<td>the median and [arithmetic] standard deviation OR coefficient of variation</td>
</tr>
<tr>
<td>12</td>
<td>the mean and [arithmetic] standard deviation OR coefficient of variation</td>
</tr>
<tr>
<td>13</td>
<td>the mode and [arithmetic] standard deviation OR coefficient of variation</td>
</tr>
<tr>
<td>14</td>
<td>a value and its percentile OR fractile OR std norm deviate and [arithmetic] $SD$ or $CV$</td>
</tr>
<tr>
<td>15</td>
<td>a pair of values and their percentiles OR fractiles OR std normal deviates</td>
</tr>
<tr>
<td>16</td>
<td>minimum, maximum, and mean values</td>
</tr>
</tbody>
</table>
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 $x_{\text{min}}$ and $x_{\text{max}}$ values are symmetric about the geometric mean $x_{50}$, then the 3 values uniquely determine a lognormal distribution.
- $f_{\text{min}} = 1 - f_{\text{max}}$, so that $-z_{\text{min}} = z_{\text{max}}$

$$
\mu = \frac{\ln x_{\text{min}} + \ln x_{\text{max}}}{2} \quad \text{or} \quad x_{50} = \sqrt{x_{\text{min}}x_{\text{max}}}
$$

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

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

$$
z_{\text{min}} = \frac{\ln x_{\text{min}} - \mu}{\sigma} \quad \text{Check: } n \approx \frac{1}{2f_{\text{min}}}?
$$

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

\[
\Lambda(\mu_1, \sigma_1^2)\Lambda(\mu_2, \sigma_2^2) = \Lambda(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)
\]

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


