Handout 2. The Log-Normal Distribution and Particle Size Presentation

<u>Introduction</u>

Often when a measured quantity takes on different values at different times, it is convenient to describe its behavior using a probability density function, also called a frequency distribution. For example, the daily values of the 8-hour average concentration of a chemical in one worker's breathing zone could be described by a distribution function for the N values of concentration, C_i, where i = 1,2…..N, as follows:

$\Delta n(C_i)$ = number of observations between C_i and $(C_i + \Delta C_i)$, where ΔC_i is a small increment relative to C_i. Figure 2.1 shows a histogram of 20 years' hypothetical data on 8 hour average concentrations of substance X in the breathing zones of 8 workers. Further, it is useful to define:

$\frac{\Delta n(C_i)}{N}$ = the fraction of all observations in the above range. Figure 2.2 shows the same data plotted as the fraction in each value interval. When two or more distributions are compared, it is most useful to make the comparison using:

$$f_n(C_i) = \frac{\Delta n(C_i)}{N \Delta C_i} \qquad 2.1$$

The function *f* is the probability density function, or frequency distribution, for C_i, and represents the fraction of the total observations in each interval, divided by the interval width. The subscript *n* on the left-hand side denotes that this is the distribution by count. Figure 2.3 shows the concentration data again, plotted as the frequency distribution and smoothed to obliterate the histogram bars. There are often other ways of enumerating the values in each interval than by count, as will be demonstrated shortly.

Rearrangement of equation 2.1 shows that the number fraction in any interval is the product of the frequency distribution value for that interval divided by the interval width:

$$\therefore \frac{\Delta n(C_i)}{N} = f_n(C_i)\Delta C_i$$

In figure 2.3, if a histogram were used instead of a smooth curve, it would be evident that the height of each histogram bar is the value of f_n, while the area of each bar is the fraction contained in that interval. Therefore the area under the entire distribution is 1.00.

If the data are normally distributed, the Gaussian PDF applies:

$$f_n(C_i) = \frac{1}{(\sqrt{2\pi})\sigma} \exp\left[\frac{-(C_i - \overline{C})^2}{2\sigma^2}\right] \qquad 2.2$$

where $\overline{C}$ is the mean, and σ is the standard deviation. The data in figures 2.1-2.3 are not normally distributed.

Regardless of the shape of the distribution, the mean is defined as:

$$meanC = \overline{C} = \sum_i C_i \frac{\Delta n(C_i)}{N} = \sum_i C_i f_n(C_i)\Delta C_i \qquad 2.3$$

The cumulative fraction is the fraction of the observations that are less than or equal to C_J:

$$F_n(C_J) = \sum_{i=1}^{J} \frac{\Delta n(C_i)}{N} = \sum_{i=1}^{J} f_n(C_i)\Delta C_i \qquad 2.4$$

The frequency distribution f_n and the cumulative distribution F_n are equivalent ways of displaying the same data. Depending on what use is to be made of the measurement data, one of these representations may be more useful than the other.

Particle Size Distribution Data

Particles present in contaminated air are rarely all the same size, and a convenient way of dealing with this fact is to describe the frequency distribution of particle sizes in an air sample. Extending the above definitions, by number:

$$f_n(d_i) = \frac{\Delta n(d_i)}{N\Delta d_i} \qquad 2.5$$

Figure 2.4 shows a typical particle size distribution enumerated by number (count) of particles in each size interval. The mean diameter by count is shown as is the standard deviation of the distribution. Note that since this distribution is not normal, the interpretation of the mean and standard deviation of the size distribution is ambiguous.

If the particles are enumerated by mass:

$$f_m(d_i) = \frac{\Delta m(d_i)}{M\Delta d_i} \qquad 2.6$$

The latter is the mass fraction contained in the diameter interval divided by the width of the interval and is denoted using the subscript *m*.

A very frequent observation for airborne particles in many settings is that the distribution of particle sizes is normal if the size is plotted as the logarithm of diameter. Figure 2.5 shows the same data as Figure 2.4, but plotted using the logarithms of the diameters (the x-axis is distorted into a logarithmic scale to reflect this). Notice that now the distribution is a symmetric bell-shaped curve. Such a distribution is called "log-normal."

The Log-Normal probability density function:

$$f_n(d_i) = \frac{\Delta n(d_i)}{N\Delta(\log_{10} d_i)} = \frac{1}{\left(\sqrt{2\pi}\right)\log_{10}\sigma_g}\exp\left[-\frac{\left(\log_{10} d_i - \log_{10}\bar{d}\right)^2}{2\left(\log_{10}\sigma_g\right)^2}\right] \qquad 2.7$$

and

$$f_m(d_i) = \frac{\Delta m(d_i)}{M\Delta(\log_{10} d_i)} = \frac{1}{\left(\sqrt{2\pi}\right)\log_{10}\sigma_g}\exp\left[-\frac{\left(\log_{10} d_i - \log_{10} MMD\right)^2}{2\left(\log_{10}\sigma_g\right)^2}\right] \qquad 2.8$$

In these expressions, $\bar{d}$ is the count mean diameter, σ_g is the geometric standard deviation, and MMD is the mass mean diameter. An important property of any normal or log-normal distribution is that the count mean diameter and the count median diameter are equal; likewise, the mass mean diameter and the mass median diameter are equal.

Therefore the count median diameter, CMD, can be calculated as

$$\log_{10} CMD = \sum_i \frac{\Delta n(d_i)}{N}\log_{10} d_i = \sum_i f_n(d_i)\Delta(\log_{10} d_i)\log_{10} d_i \qquad 2.9$$

but a simpler way is to use the following definitions:

The count median diameter CMD is the value of d at which $F_n(d_i) = 0.5$. In figure 2.5, the count median diameter of the distribution is given. When the particles are enumerated by mass, the same data again produce a log-normal distribution, but with a different location on the diameter axis, as shown in Figure 2.6.

The mass median diameter MMD is the value of d at which $F_m(d_i) = 0.5$, and it is not the same as the count median diameter for these particles.

In general, the geometric standard deviation of the distribution requires laborious calculation. However, if the distribution is log-normal, or even approximately so, the geometric standard deviation σ_g is given by :

$$\log_{10}\sigma_g = \log_{10} d(0.84) - \log_{10} d(0.50) = \log_{10}\left[\frac{d(0.84)}{d(0.50)}\right] \qquad 2.10$$

$$\Rightarrow \sigma_g = \frac{d(0.84)}{d(0.50)} \qquad 2.11$$

A curious finding is that the geometric standard deviation for these particles is the same, regardless of how the particles are enumerated. This is a universal property of the geometric standard deviation. It should also be noted that the geometric standard deviation has no units, since it is the ratio of two diameter percentile values, and the smallest possible value for the geometric standard deviation is 1.00. This will arise when all particles have the same diameter: the rare "monodisperse" particle size distribution.

The cumulative version of the (count-enumerated) particle size distribution is shown in Figure 2.7 for the same data.. One very useful feature of this plot is that the median diameter can be read directly from the graph without need for calculation: it is the diameter corresponding to the fraction value of 0.5. As noted above, if the distribution is log-normal, the determination of the geometric standard deviation is also much simplified. To determine if this condition is met, the cumulative fraction can be plotted on a probability, or probit scale. This is shown in Figure 2.8 for the same data, again for particle counts. A log-normal distribution will appear as a straight line when plotted in this form. Because the diameters are plotted on a logarithmic scale, this is referred to as a "log-probability" or "log-probit" plot. The probit scale is constructed so that zero corresponds to a fraction of 0.5, and each probit unit above or below this is one standard deviation. Therefore the probit value +1 corresponds to a fraction of 0.84, since in a normal distribution one standard deviation above the mean corresponds to this fraction (probit value +2 corresponds to a fraction of 0.975, probit value -1 corresponds to a fraction of 0.16, etc.). Use of this representation to determine the median diameter and any other percentiles is illustrated in Figure 2.9.

<u>"Hatch-Choate" relationships</u>:

To simplify some of the calculations that would otherwise be time-consuming, advantage can be taken of several properties of the log-normal distribution. For particles meeting this criterion, the MMD and the mode diameter can be calculated simply, once the CMD and geometric standard deviation are known:

$$MMD = CMD \exp\left[3\left(\ln \sigma_g\right)^2\right] \quad 2.12$$

$$d_{mode} = CMD \exp\left[-\left(\ln \sigma_g\right)^2\right] \quad 2.13$$

For example, a log-normal size distribution with CMD = 1.00 µm and σ_g = 2.00, will yield:

The mode diameter = 0.62 µm
The MMD = 4.23 µm
The count mean diameter = 1.27 µm

Note that mode diameter<count median diameter<mass median diameter. These properties of the particle size distribution will always have this relationship, unless the geometric standard deviation is 1.00, where all diameters including the mode and each of the medians are the same.