Chemometrics

MA4605

Week 3. Lecture 5. The Log-normal distribution

September 19, 2011

The Log-normal distribution

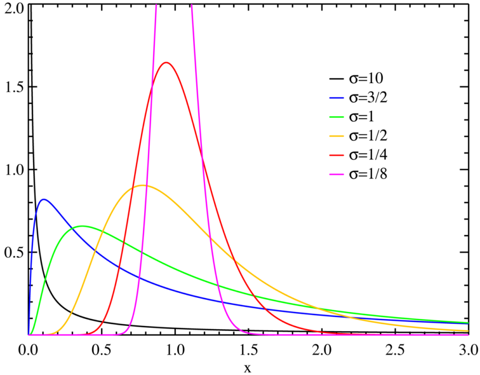
Measurements made on each of a number of specimens can lead to distributions that are not Normal.
In particular the **log-normal** distribution is often encountered.

- A random variable is said to follow a lognormal distribution if its *logarithm* transformation follows a Normal distribution
- **X** is a lognormal variable if **Y=log(X)** is Normal.

Example 1: The antibody concentration in human blood sera.
Example 2: The particle size of the droplets formed by the nebulisers used in flame spectroscopy.
Example 3: The particle size distributions in atmospheric aerosols.
Example 4: Equipment failure rates.

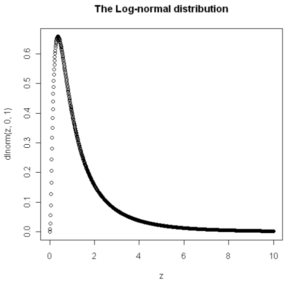
The frequency distribution is approximated by an asymmetric curve, skewed to the right.

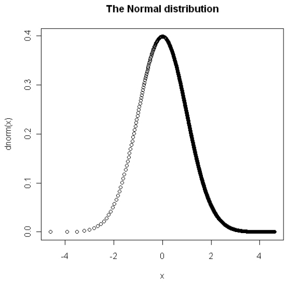
Log-normal density



The log transformation of a LN variable

The log transformation of a log normal variable Z, leads to a normally distributed variable X.





Confidence Intervals for data that come from a log-normal distribution are calculated based on the *logarithm* of the measurements.

- Calculate the limits of the CI for the mean of the transformed data
- Take the exponential value of these limits to obtain the CI for the geometric mean of the original measurements.

The confidence interval obtained is not for the *arithmetic mean* but for the **geometric mean**!

- The geometric mean of the original distribution is given by $\sqrt[n]{x_1 x_2 \ldots x_3}$

A sample of values for the antibody concentration in human blood serum for 8 healthy adults 2.15 1.13 2.04 1.45 1.35 1.09 0.99 2.07

- Calculate the 95%CI for the geometric mean, assuming that the antibody concentration is log-normal.
- Apply the *log* function to these values.
  In *R* the **log** computes by default natural logarithms.
- The log transformed values are: 0.7654 0.1222 0.7129 0.3715 0.3001 0.0861 -0.0100 0.7275.

A sample of values for the antibody concentration in human blood serum for 8 healthy adults 2.15 1.13 2.04 1.45 1.35 1.09 0.99 2.07

- Calculate the 95%CI for the geometric mean, assuming that the antibody concentration is log-normal.
- Apply the *log* function to these values.
  In *R* the **log** computes by default natural logarithms.
- The log transformed values are: 0.7654 0.1222 0.7129 0.3715 0.3001 0.0861 -0.0100 0.7275.

The mean of the logged values is 0.3844974.
The standard deviation of the logged values is 0.3142897.

Example 2.10.1

A sample of values for the antibody concentration in human blood serum for 8 healthy adults 2.15 1.13 2.04 1.45 1.35 1.09 0.99 2.07

- Calculate the 95%CI for the geometric mean, assuming that the antibody concentration is log-normal.
- Apply the *log* function to these values.
  In *R* the **log** computes by default natural logarithms.
- The log transformed values are: 0.7654 0.1222 0.7129 0.3715 0.3001 0.0861 -0.0100 0.7275.

The mean of the logged values is 0.3844974.
The standard deviation of the logged values is 0.3142897.
The 95% CI of the mean of the logged values is
$0.3844974 \pm 2.364624 \frac{0.3142897}{\sqrt{8}} = [0.1217447, 0.6472502]$.

A sample of values for the antibody concentration in human blood serum for 8 healthy adults 2.15 1.13 2.04 1.45 1.35 1.09 0.99 2.07

- Calculate the 95%CI for the geometric mean, assuming that the antibody concentration is log-normal.
- Apply the *log* function to these values.
  In *R* the **log** computes by default natural logarithms.
- The log transformed values are: 0.7654 0.1222 0.7129 0.3715 0.3001 0.0861 -0.0100 0.7275.

The mean of the logged values is 0.3844974.
The standard deviation of the logged values is 0.3142897.
The 95% CI of the mean of the logged values is
$0.3844974 \pm 2.364624 \frac{0.3142897}{\sqrt{8}} = [0.1217447, 0.6472502]$.
The 95% CI of the geometric mean of the initial values is
$[exp(0.1217447), exp(0.6472502)] = [1.129466, 1.910281]$.

In experimental work, the quantity to be determined is often from a **combination** of observable quantities.
Even simple operations involve several stages, each subject to errors.
The final calculations are combinations(sum, difference, product, power) of measured quantities and the variance of the final results is affected by errors of all quantities involved.

Propagation of the error in the linear combination of independent measurements

If the final value **y** is obtained as the linear combination of measured variables a,b,c,... with standard deviations σ_a, σ_b , σ_c,..., where k, k_a, k_b, k_c,... are known constants:

$y = k + k_a \cdot a + k_b \cdot b + k_c \cdot c + \ldots$

Then the variance of y, $\sigma_y^2 = (k_a\sigma_a)^2 + (k_b\sigma_b)^2 + (k_c\sigma_c)^2 + \ldots$

The standard deviation of y,

$\sigma_y = \sqrt{(k_a\sigma_a)^2 + (k_b\sigma_b)^2 + (k_c\sigma_c)^2 + \ldots}$

Example 2.11.1

In a titration the initial reading in a burette is a=3.51 ml and the final reading is b=15.67 ml, both with a standard deviation $\sigma_a = \sigma_b = 0.02$ml.
The volume of titrant used is y=$k_a a$+$k_b b$= $k_b + k_a(-1)$ = $15.67 - 3.51$=12.16
The standard deviation y is $\sigma_y = \sqrt{(k_a\sigma_a)^2 + (k_b\sigma_b)^2}$
$\sigma_y = \sqrt{(-1 \cdot \sigma_a)^2 + (1 \cdot \sigma_b)^2} = \sqrt{0.02^2 + 0.02^2} = 0.028$
Standard deviation of final quantity y is $>$ standard deviation of linear components a or b.

Propagation of the error in multiplicative expressions

If the final value **y** is obtained as a product $y = \frac{k \cdot a \cdot b}{c \cdot d}$
where a,b,c,d are independent measured quantities with standard deviations σ_a, σ_b , σ_c, σ_d, and k is a known constants.
Then the **relative standard deviation** of y

$$\frac{\sigma_y}{y} = \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 + \left(\frac{\sigma_c}{c}\right)^2 + \left(\frac{\sigma_d}{d}\right)^2}$$

Example 2.11.2

The quantum yield of fluorescence, y, is calculated from the expression $y = \frac{I_f}{k \cdot c \cdot l \cdot I_0 \cdot \epsilon}$.
where the quantities involved are defined below with an estimate of their standard deviation between brackets:

I_0= incident light intensity (0.5)
I_f= fluorescence intensity (2)
ϵ= molar absorptivity (1)
c= concentration (0.2)
l= path-length (0.2)
k = constant

Then the **relative standard deviation** of y
$\frac{\sigma_y}{y} = \sqrt{2^2 + 0.2^2 + 0.2^2 + (0.5)^2 + 1^2}$=2.3 %.

Propagation of the error in other functions

If the final value **y** is a general function of **x**, then the standard deviations of **x** and **y** are related by:

$$\sigma_y = |\sigma_x \frac{dy}{dx}|$$

Example 2.11.3

The absorbance, A, of a solution is given by A=-log$T$ where T is the transmittance. If the measured value of T is 0.501 with a standard deviation σ_T= 0.001, calculate A and its standard deviation.

A= -log$T$=-log 0.501= 0.6911492

From the previous slide $y = A$ and $x = T$.

$$\sigma_A = |\sigma_T \frac{dA}{dT}| = |0.001 \cdot (-\frac{1}{T})| = |0.001 \cdot (-\frac{1}{0.501})| = .002$$