The Lognormal Distribution

In general, the most important property of the **lognormal process is that it represents a product of independent random variables**. (Class Handout on Lognormal Processes)

> "A lognormal process is one in which the random variable of interest results from the product of many independent random variables multiplied together"

Often when natural processes are concerned the lognormal process is applicable. Some examples of common phenomena that can be represented by the lognormal distribution are:

- The occurrence of alien visitations (the rate varies with the changing population of cows on the planet and increases by a proportion of the bovine population)
- leaf growth (the area of the leaf increases by some random proportion)
- yearly population growth (the growth rate is a random variable because the growth rate varies in response to annual fluctuations in economic, health and social conditions)
- interest on a savings account (compounded daily by a varying national interest rate and the amount increases by some proportion of the initial amount)
- fixed amount of tracer pollutant in a pond some days later (the flow of the water through the pond varies from hour to hour; therefore the dilution factor (proportion) varies by some proportion of the initial concentration)

The concept I'm attempting to convey is **change by a proportion that can vary.**
The important formulas related to the lognormal are; (all equations can be found on the class handout, "Lognormal Processes")

Probability Density Function (PDF)

$$f_x(x) = \begin{cases} \dfrac{1}{x s_{\ln x}\sqrt{2p}}\, e^{-0.5\left(\frac{\ln x - m_{\ln x}}{s_{\ln x}}\right)^2} & -\infty < m < \infty \\ 0 & o/w \end{cases}$$

Expected Value

$$a = E(X) = m_x = e^{m_{\ln x} + \frac{s_{\ln x}^2}{2}}$$

Variance

$$b^2 = V(X) = s_x^2 = e^{2m_{\ln x} + s_{\ln x}^2}\left(e^{s_{\ln x}^2} - 1\right)$$

note: these show the relationship between the normal and lognormal values.

Problem Statement:

? What are the Lognormal distribution parameters?

If we are given α & β, (these are the μ_x & σ_x) we must re-write the above equations in terms of μ_{lnx} & σ_{lnx}, which are the parameters of the Lognormal Distribution.

$$m_{\ln x} = \ln m_x - \frac{1}{2} s^2_{\ln x}$$

$$s_{\ln x} = \sqrt{\ln\left(1 + \frac{s^2_x}{m^2_{xl}}\right)}$$

note: the lognormal standard deviation, σ_{lnx}, must be computed first

It is given that μ_x = 6.267 & σ^2_x = 10.73

But we want:

X~Lognormal(μ_{lnx}, σ_{lnx})

Therefore, apply the equations:

$$s_{\ln x} = \sqrt{\ln\left(1 + \frac{s^2_x}{m^2_{xl}}\right)} = \sqrt{\ln\left(1 + \frac{10.73}{6.267^2}\right)} = \mathbf{0.4915 = \sigma_{lnx}}$$

$$m_{\ln x} = \ln m_x - \frac{1}{2} s^2_{\ln x} = \ln 6.267 - \frac{1}{2} \cdot 0.4915^2 = \mathbf{1.7145 = \mu_{lnx}}$$

X~Lognormal(1.7145,0.4915)

??What are the 1, 10, 50, 90, 99th percentiles of the Lognormal?

The x_p percentile represents the value of x at which p% of the population is below. Mathematically,

$$P(X \le x_p) = p$$

There are several ways to determine the value of x_p corresponding to a given percentile value for the lognormal distribution.

Method 1: (Use the Standard Normal Tables)

As shown in class:

- Standardization of Lnx using the Z-tables:

$$z_p = \frac{\ln x_p - m_{\ln x}}{s_{\ln x}}$$

- Rearranging to solve in terms of x_p:

$$x_p = e^{z_p \cdot s_{\ln x} + m_{LNX}}$$

HINT You had better have this equation on your cheat sheet!!

- Now we find the z_p value corresponding to the percentile that we are interested in and back calculate for x_p

For example, say we want the 37th percentile, we'd look up 0.37 on the Z-table and find that the z having this probability is –0.33. Plugging that value and the Lognormal parameters into $x_p = e^{z_p \cdot s_{\ln x} + m_{LNX}}$ we'd get the x_p value

$$x_{0.37} = e^{(-0.33*0.4915+1.7145)} = 4.72$$

For the answers to our specific problem refer to Table 1[1]

Method 2- Use EXCEL

Using the Excel spreadsheet can be pretty quick and easy too. What you have to is use the CDF function, (which is all Excel has for the lognormal anyway- the PDF is not an intrinsic function. If you want to graph the PDF you must enter it in manually. To learn more about this, check out the section on the PDF graph)

Steps:

1) You'll enter into a cell the CDF function: =LOGNORMDIST(x,$\mu_{\ln x}$,$\sigma_{\ln x}$)
   For example: =LOGNORMDIST($A1,$B$3,$B$4), where $A1 is the column of x values, B3, & B4 are the lognormal parameters,$\mu_{\ln x}$ & $\sigma_{\ln x}$ repsectively

2) Then choose (under "Tools") Goal Seek.
   - The dialog box will have "Set Cell" -that cell should be where you have the CDF function.
   - Enter in the decimal value of the percentile of interest in the "To Value" cell
   - The "By Changing" cell should reference the cell where you want the x, $A1, value to appear. Also, this cell should be the cell referenced in your CDF equation.

And as if by magic, you receive your answer-beautiful!-Actually, it's Numerical Analysis wizardry at work (iteration after iteration)! Consult a numerical analysis text for further details if you have a burning curiosity about the specific algorithm that Excel uses.

***To note: You may get a response of "no solution" if you haven't entered in an initial guess or there's a crazy value in the cell- try "1" in each cell and it should work fine.

---

[1] Located under Method 2

The solutions I got were slightly different than those achieved manually. That's probably just round-off or truncation error in one of the algorithms. Table 1 lists the values achieved by both methods.

Table 1- Percentile Values calculated by Mathematical and Numerical Techniques

Percentile	z_p			x_p		Manual value	Excel value
1st	$z_{0.01}$	=	-2.33	$x_{0.01}$	=	1.767	1.7675
10th	$z_{0.1}$	=	-1.29	$x_{0.1}$	=	2.946	2.9582
50th	$z_{0.5}$	=	0	$x_{0.5}$	=	5.554	5.5531
90th	$z_{0.9}$	=	1.29	$x_{0.9}$	=	10.470	10.4000
99th	$z_{0.99}$	=	2.33	$x_{0.99}$	=	17.455	17.2087

Method 3- Use Integration

Integrate the PDF but whoa-why when there are other less painful ways? Well, unfortunately, sometimes the CDF and tables are not available and we must do this. (Hey and why else did we take 3 semesters of Calculus?)

$$F_x(x) = \int_{-\infty}^{x} \frac{1}{u s_{\ln x} \sqrt{2p}} e^{-0.5\left(\frac{\ln u - m_{\ln x}}{s_{\ln x}}\right)^2} du$$

This function can be integrated mathematically by using integration by parts OR numerically by your friend the calculator.

Graphs of the PDF and CDF of the Lognormal

Figure 1 is the graph of the probability density function of the lognormal.

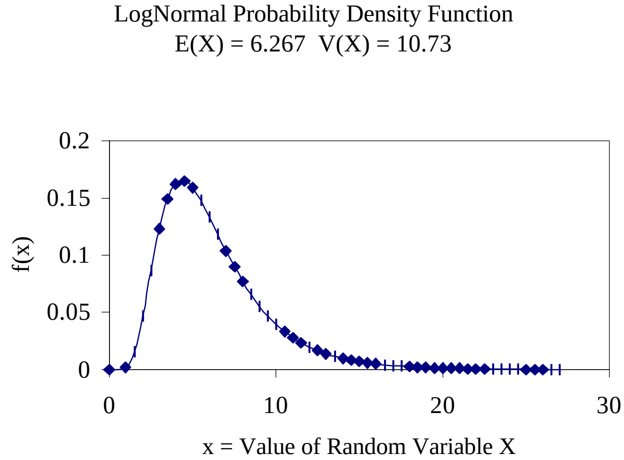

Figure 1- PDF of the Lognormal Distribution

Unlike the symmetry observed in the normal distribution, the lognormal is characterized by its skewidness[2] (that means it's skewed), a single mode and a long tail to the right.

The Lognormal Cumulative Density Function is illustrated in Figure 2. Notice that the 50th percentile is NOT E(X) = 6.267. Rather it is 5.55. This is so because of the skewidness of the lognormal. (important concept)

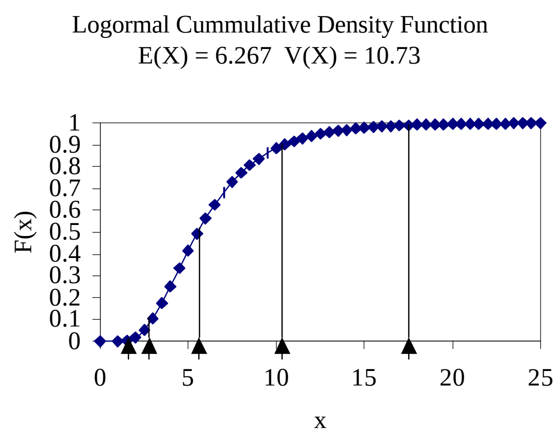

Figure 2- Cumulative Density Function-Lognormal Distribution.
The arrows (left to right) indicate the 1st, 10th, 50th, 90th , and 99th percentiles.
Refer to Table 1 for numerical values

[2] Some words may have been created by Geppie

Lastly, Figure 3 illustrates the relative shape and location of the pdf's of the lognormal and the normal distributions.

PDF of the Normal and Lognormal Distributions
E(X) = 6.267 V(X) = 10.73