LOGNORMAL DISTRIBUTIONS I

A *lognormal* random variable is a random variable Y that only takes on positive values and has the property that the random variable X = logY is normal. Here "log" refers to "natural log." So equivalently: $Y = e^X$ for some normal random variable X.

Would it matter if we used log base 10?

Exercise: Using what you know about the shape of a normal distribution and about the exponential function, figure out the general shape of a lognormal distribution.

Corollary: If Y is log normal and the mean of logY is μ, then e^{μ} is the _____________ of Y.

"The Multiplicative Central Limit Theorem": Suppose $Y_1, \ldots, Y_n$ are independent random variables and $Y = Y_1 \ldots Y_n$ is their product. Then

logY = ____

So if n is large enough, logY is ____ ,

so Y is ____.

Examples:

1. During a four-year period in the early 1990's, the pesticide heptachlor was used to control ants on pineapples in Hawaii. Heptachlor breaks down into harmless constituents after a suitable length of time, but in this case, clippings from the pineapple plants were used as supplemental cattle feed before the necessary time for the heptachlor to break down. The amount of heptachlor varied and could not be measured exactly in any case, so must be considered a random variable, say Y_1 (units ppm = parts per million, or mg heptachlor per kg clippings). The clippings were mixed with regular feed in variable amounts, so this rate is another random variable Y_2 (can be considered unitless, since it's a rate in the same units, or as a percent, or in kg heptachlor per kg total feed). The cattle ate variable amounts of the contaminated feed, giving another random variable Y_3 (units kg per day). The cattle were processed as beef for human consumption. The amount of beef an adult human eats per day is another random variable Y_4 (units mg/day). Adult body weight is another random variable Y_5 (units kg). The amount of fat in beef is also a random variable Y_6 (unitless, or percent, or mg fat per mg beef). The incremental lifetime risk of cancer from exposure to the heptachlor is proportional to $Y_1Y_2Y_3Y_4(1/Y_5)Y_6$ (the constants of proportionality, such as the cancer potency factor, make the units come out right), and these random variables are independent, so we expect that the incremental lifetime risk of cancer from the exposure is approximately lognormally distributed.

2. Many biological processes produce lognormal distributions. For example, in cell division, one cell division more will multiply the number of cells by 2; one cell division less will divide it by half. How might this explain why "exceptionally large" fruits and vegetables seem to be reported every year?

3. A modified Galton board can illustrate how a lognormal distribution can arise from a process where at every stage, there is a random choice of either moving to the right by a factor of c or to the left by a factor of 1/c. See a simulated modified Galton board at http://www.inf.ethz.ch/personal/gut/lognormal/

4. Examples where data appear to be lognormal include:

- Concentration of elements in the earth's crust.
- Latent periods (time from infection to first symptoms) of infectious diseases.
- Survival times after cancer diagnosis.
- Size of crystals in ice cream.

Problems:

1. Find a formula for the pdf of a lognormal random variable Y with logY = X normal with parameters μ and σ. (Remember that we are using natural logs.) [Hints: a. Remember that Y only takes on positive values. b. To find the formula for the pdf for positive values, start by finding the cdf $F_Y(y)$ of Y in terms of the cdf $F_X(x)$ of X.]

2. Use your answer to Problem 1 to find the mode of Y – that is, the place where the graph of the pdf of Y is highest.