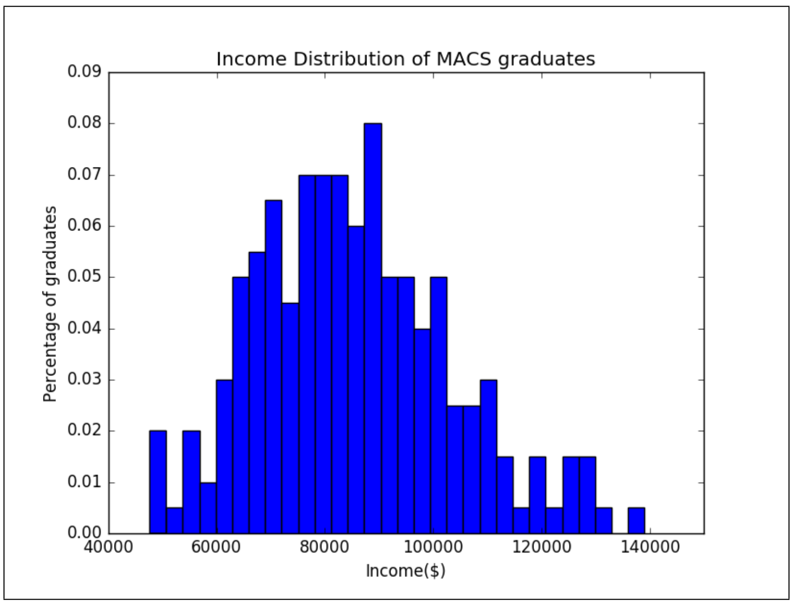
Problem Set #2

MACS 30100, Dr. Evans
Yinxian Zhang

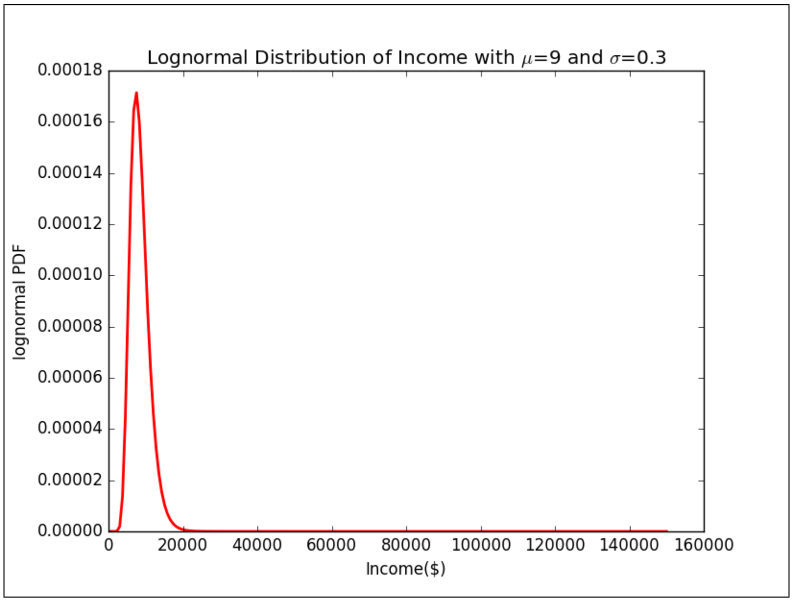
Problem 1 Lognormal Distribution of Some Income Data
Part (a).
The histogram of percentages of the annual income of MACS graduates is plotted in Figure 1. The histogram has 30 bins that are weighted to represent the percentage of income observations in that bin.

Figure 1: Income Distribution



Part (b).
Given the lognormal distribution with $\mu = 9$ and $\sigma = 0.3$, and given the annual income of the MACS students, the value of the log likelihood function for this parameterization and this data is -8298.637. The lognormal PDF for 0 <= x <= 150,000 is plotted in Figure 2.

Figure 2: Lognormal PDF



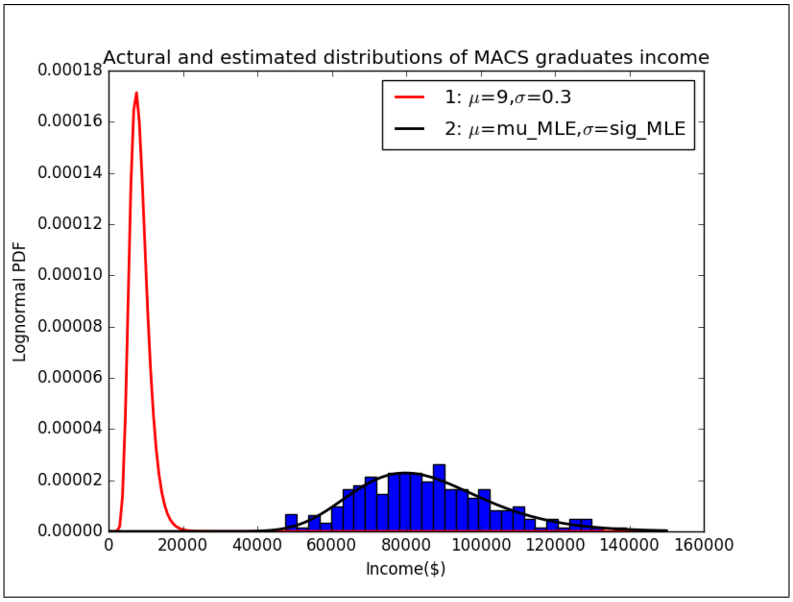
Part (c).

The maximum likelihood estimates for $\mu_{(MLE)}$ and $\sigma_{(MLE)}$ of the annual income assuming a lognormal distribution are 11.334 and 0.215, respectively. The value of the log likelihood function is -2238.938. The variance-covariance matrix is:

$$\begin{bmatrix} 0.000239 & 0.000011 \\ 0.000011 & 0.000131 \end{bmatrix}$$

The PDF with ML estimates is plotted against the PDF from Part(b) and the histogram of the real income data, and is shown in Figure 3.

Figure 3: Lognormal PDF and histogram distribution of income



Part (d).
The likelihood ratio rest shows that the probability that the data in `income.txt` came from the distribution in part(b) is 0.00, meaning that the annual income of MACS graduates is very unlikely to comply to the lognormal distribution with a μ of 9 and σ of 0.3.

Part (e).
Based on the maximum likelihood estimates, the probability that I will earn more than \$100,000 is 20.17%; and the probability that I will earn less than \$75,000 is 30.66%.

Problem 2 Linear regression and MLE
Part (a).
The maximum likelihood estimate of the model gives us the value of the likelihood function of 876.865, and we can get the following parameter estimates:
$\beta_0 = 0.252$
$\beta_1 = 0.013$
$\beta_2 = 0.401$
$\beta_3 =$ -0.00999
$\sigma^2 = 0.0000091$ ($\sigma = 0.00302$)
The varian-covariance matrix is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Part (b).

The likelihood ratio test shows that the probability that β_0 is 1 and σ^2 is 0.01 is very low with a value of 0.00, indicating that it is very unlikely that age, number of children and avergae winter temperature have no effect on the number of sick day.