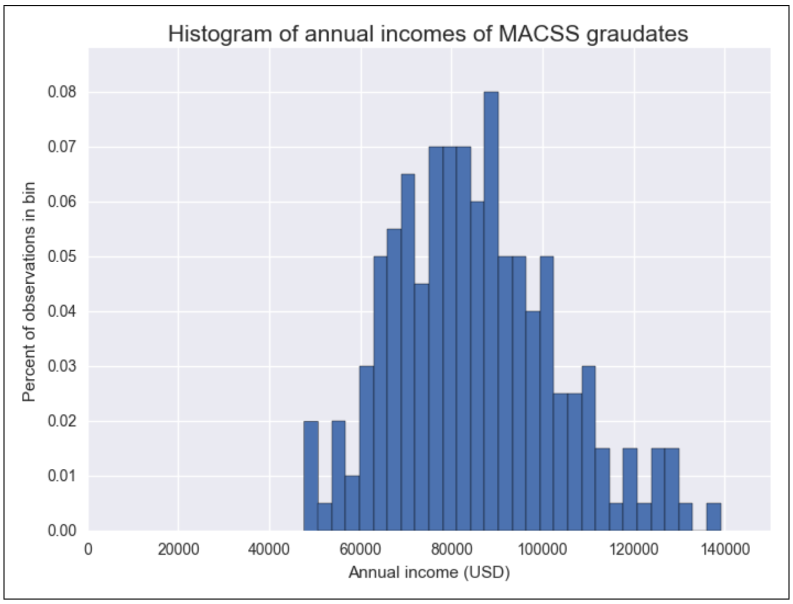
Problem Set #2

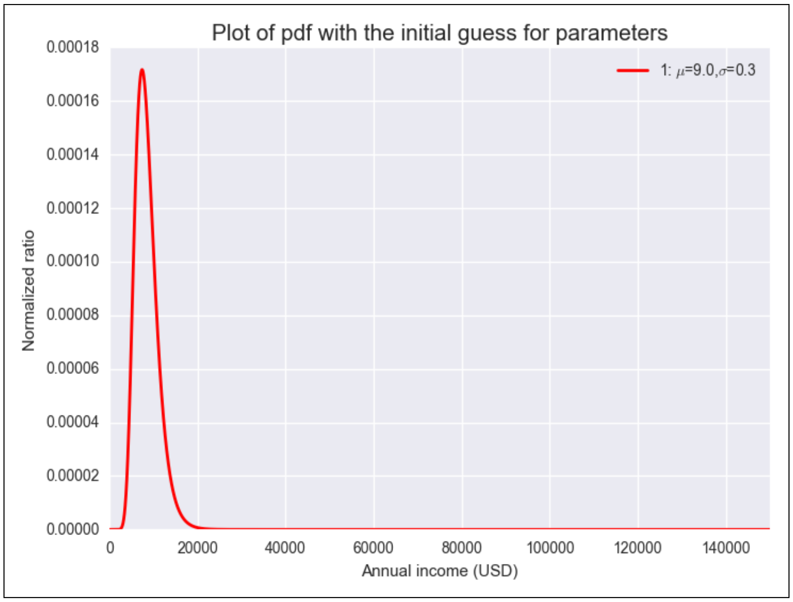
MACS 30100, Dr. Evans
Bobae Kang

1. **Some income data, lognormal distribution, and hypothesis testing.**

   (a) Plot a histogram of percentages of the income.txt data with 30 bins. Make sure that the bins are weighted such that the height of each bin represents the percent of the income observations in that bin. In other words, all the bin heights should sum to 1. Make sure your plot has correct x-axis and y-axis labels as well as a plot title.

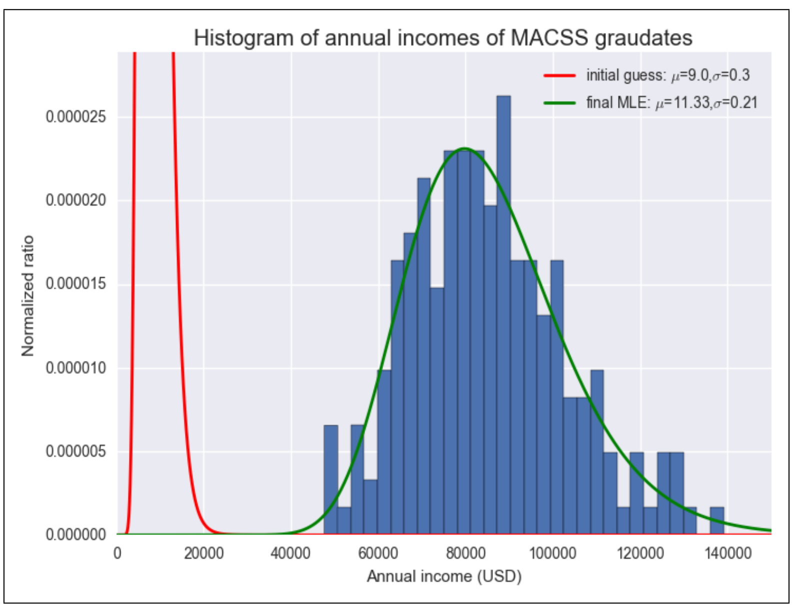


   (b) Plot the lognormal PDF $f(x|\mu = 9.0, \sigma = 0.3)$ for $0 \leq x \leq 150{,}000$. What is the value of the log likelihood value for this parameterization of the distribution and given this data?



Log-likelihood value = -8298.63695601

(c) Estimate the parameters of the lognormal distribution by maximum likelihood and plot its PDF against the PDF from part (b) and the histogram from part (a). Plot the estimated PDF for $0 \leq x \leq 150{,}000$. Report the ML estimates for μ and σ, the value of the likelihood function, and the variance-covariance matrix.



MLE estimate for μ = 11.331440374

MLE estimate for σ = 0.211674587081

Log-likelihood value = -2239.5347439980096

VCV = [[2.47898671e-04 2.46830454e-06]
[2.46830454e-06 1.12322196e-04]]

(d) Perform a likelihood ratio test to determine the probability that the data in ‘incomes.txt‘ came from the distribution in part (b).

χ^2 of H_0 with 2 degrees of freedom p-value = 0.0

(e) With your estimated distribution of incomes for Chicago MACSS students from part (c), you now have a model for what your own income might look like when you graduate. Using that estimated model from part (c), What is the probability that you will earn more than $100,000? What is the probability that you will earn less than $75,000?

The probability that you will earn more than $100,000 is: 19.56%
The probability that you will earn less than $75,000 is: 30.79%

2. **Linear regression and MLE**

(a) Estimate the parameters of the model $(\beta_0, \beta_1, \beta_2, \beta_3, \sigma^2)$ by maximum likelihood using the fact that each error term ε_i is distributed normally $N(0, \sigma^2)$. Report your estimates, the value of the log likelihood function, and the estimated variance covariance matrix of the estimates.

MLE estimate for β_0 = 0.251644293661
MLE estimate for β_1 = 0.0129334154112
MLE estimate for β_2 = 0.400501865882
MLE estimate for β_3= -0.00999168487451
MLE estimate for σ^2 = 9.10892518541e-06
Log-likelihood value = 876.865053329
VCV = [[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.]]

(b) Use a likelihood ratio test to determine the probability that $\beta_0 = 1.0$, σ^2 = 0.01 and $\beta_1, \beta_2, \beta_3 = 0$. That is, what is the likelihood that age, number of children, and average winter temperature have no effect on the number of sick days?

The χ^2 of H_0: age, number of childeren, and average winter temperature have no effect on the number of sick days, with 5 degrees of freedom, is: 0.