Problem Set #2

MACS 40200, Dr. Evans
Dongping Zhang

Problem 1: Descriptive Statistics and Density Histograms

Part (a). The descriptive statistics for the *Households Health Expenditure* dataset are:

- mean = 720.28
- median = 172.21
- maximum = 227967.25
- minimum = 0.01
- variance = 15783543.67
- standard deviation = 3972.66

Figure 1: Households Health Expenditure (Raw)

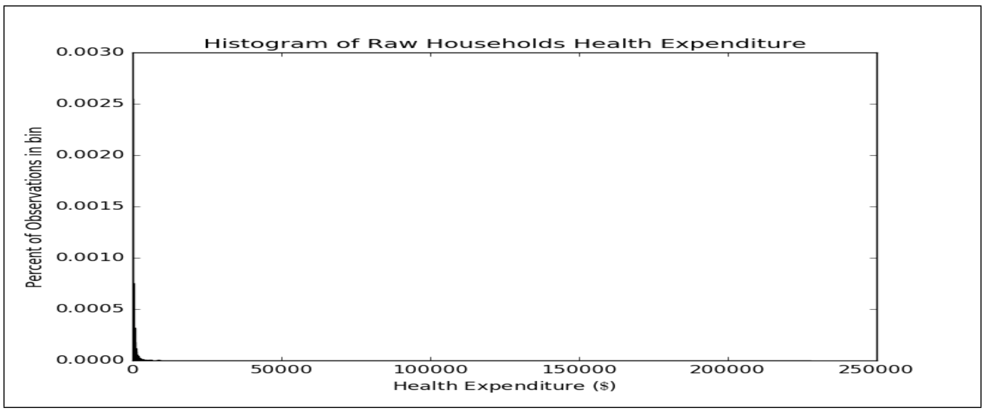
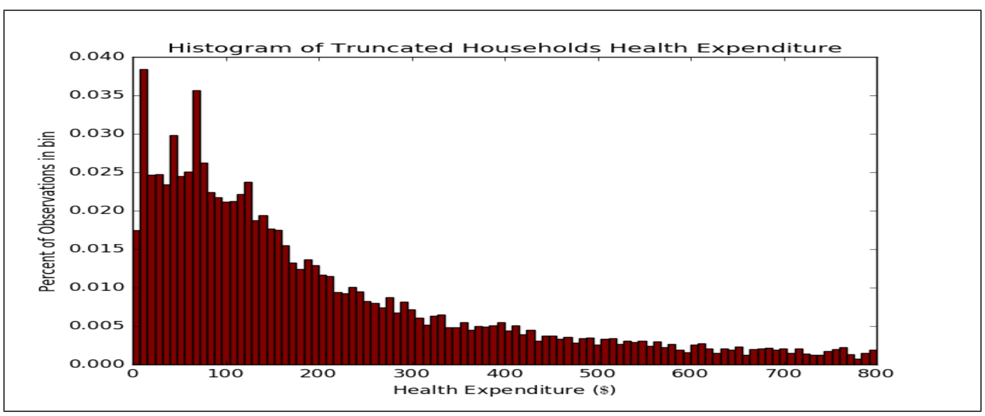


Figure 2: Households Health Expenditure (Truncated)



The sum of probabilities represented by all bins in **Figure 2** is: 0.849420849421, which is not 1 because it only displays a fraction of the health expenditure data.

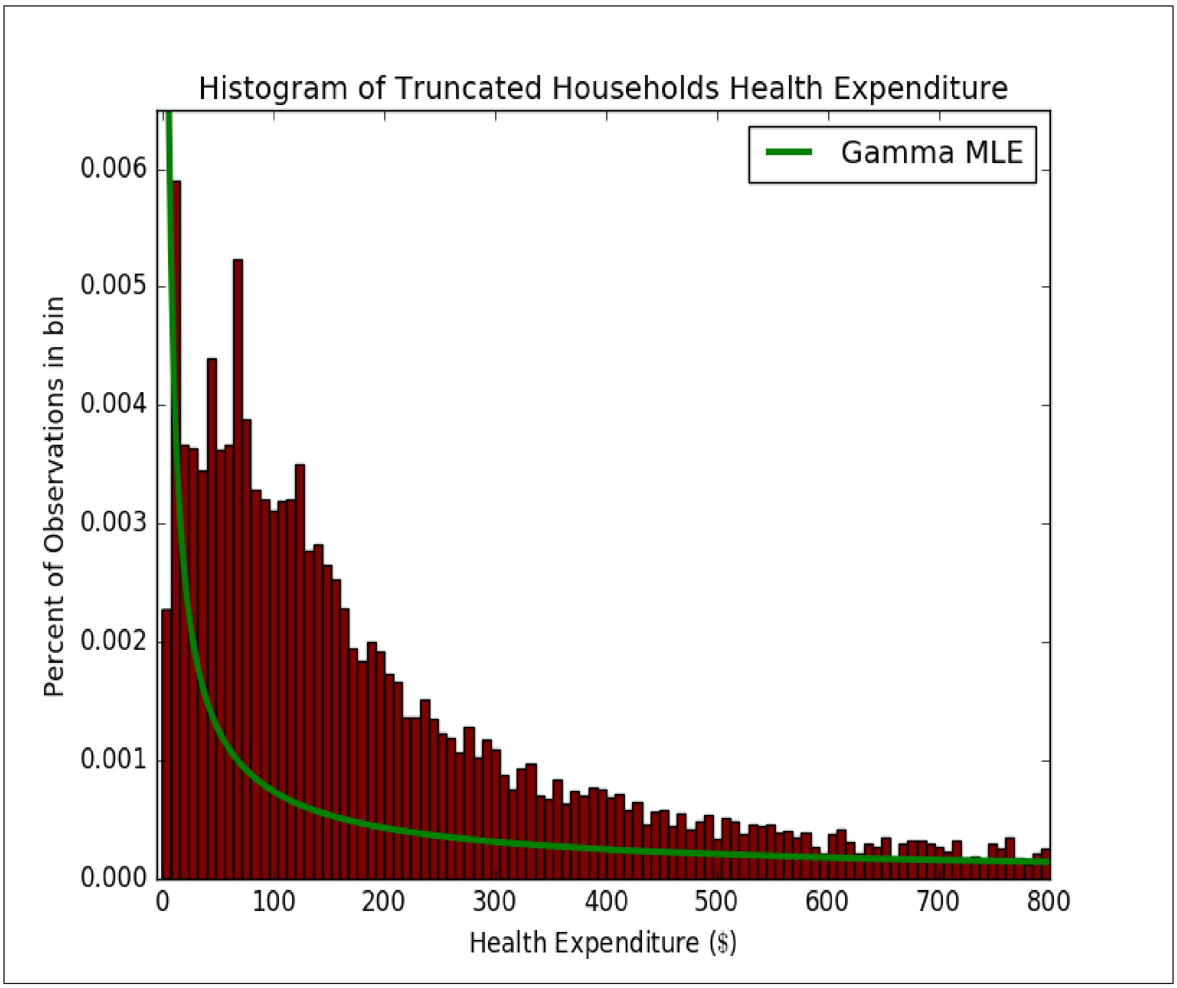
Comparing with the two histograms, the second one might be preferred because it displays a clear and informative distribution for later MLE analysis. In addition,

Figure 1 has shown that most of the data points are located near 0, and the number of those large health expenditure data points are infinitesimal and could be considered to be outliers. Thus, the second histogram plotting with truncated Health Expenditure data would be preferred.

Part (b). After using Maximum Likelihood Estimation to fit a Gamma GA(x; α, β) distribution on the individual observation data, I have obtained:

- $\hat{\alpha}_{MLE}$ = 0.225120237974
- $\hat{\beta}_{MLE}$ = 20397.7875956
- $\mathcal{L}(\hat{\theta})$ = -81932.5529069

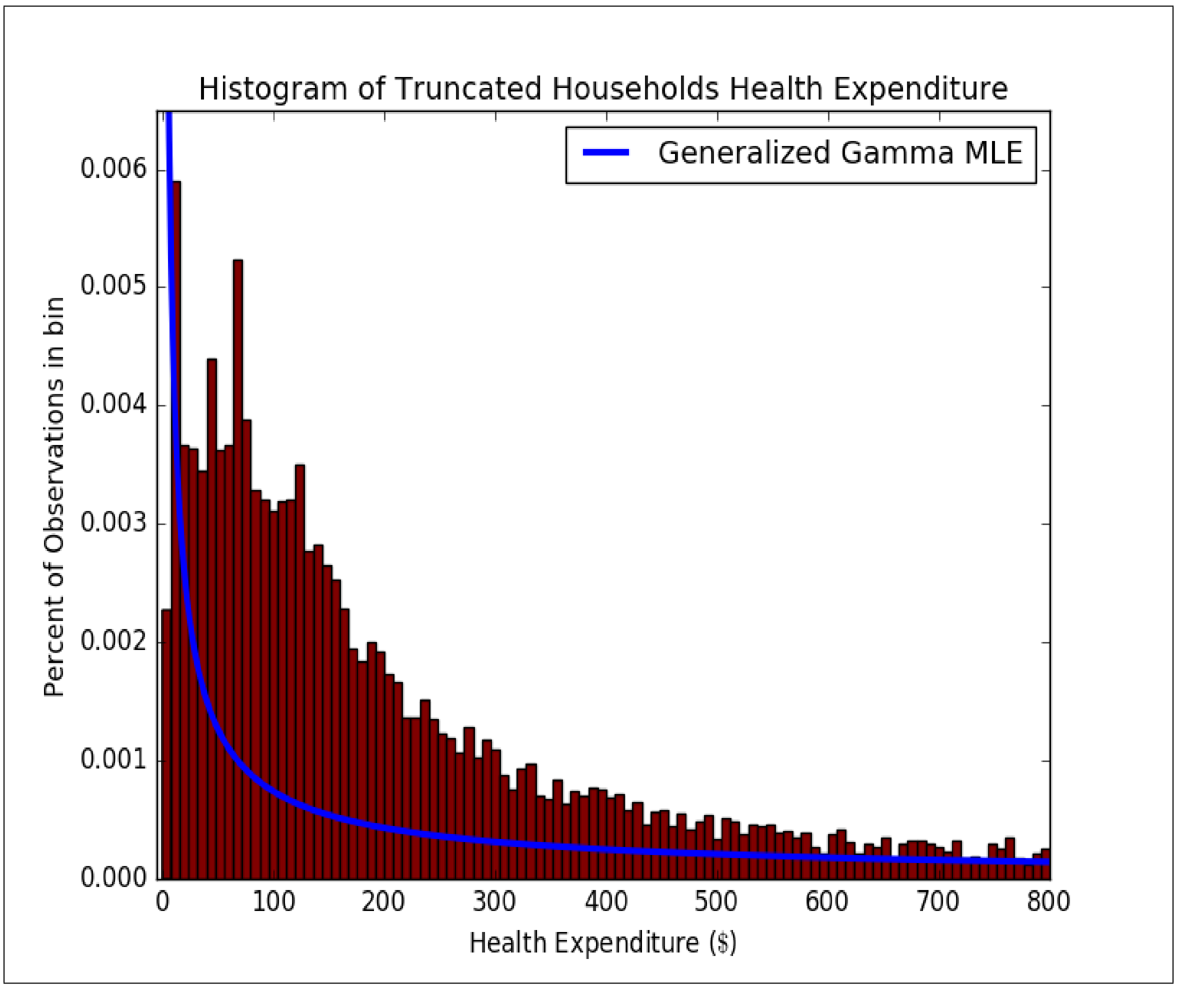
Figure 3: Households Health Expenditure (Truncated) Density Histogram Overlayed on a Gamma



Part (c). After using Maximum Likelihood Estimation to fit a Generalized Gamma GG(x; α, β, m) distribution on the individual observation data, I have obtained:

- $\hat{\alpha}_{MLE}$ = 0.225127191128
- $\hat{\beta}_{MLE}$ = 20397.787543
- m = 1.00773265469
- $\mathcal{L}(\hat{\theta})$ = -81932.496017

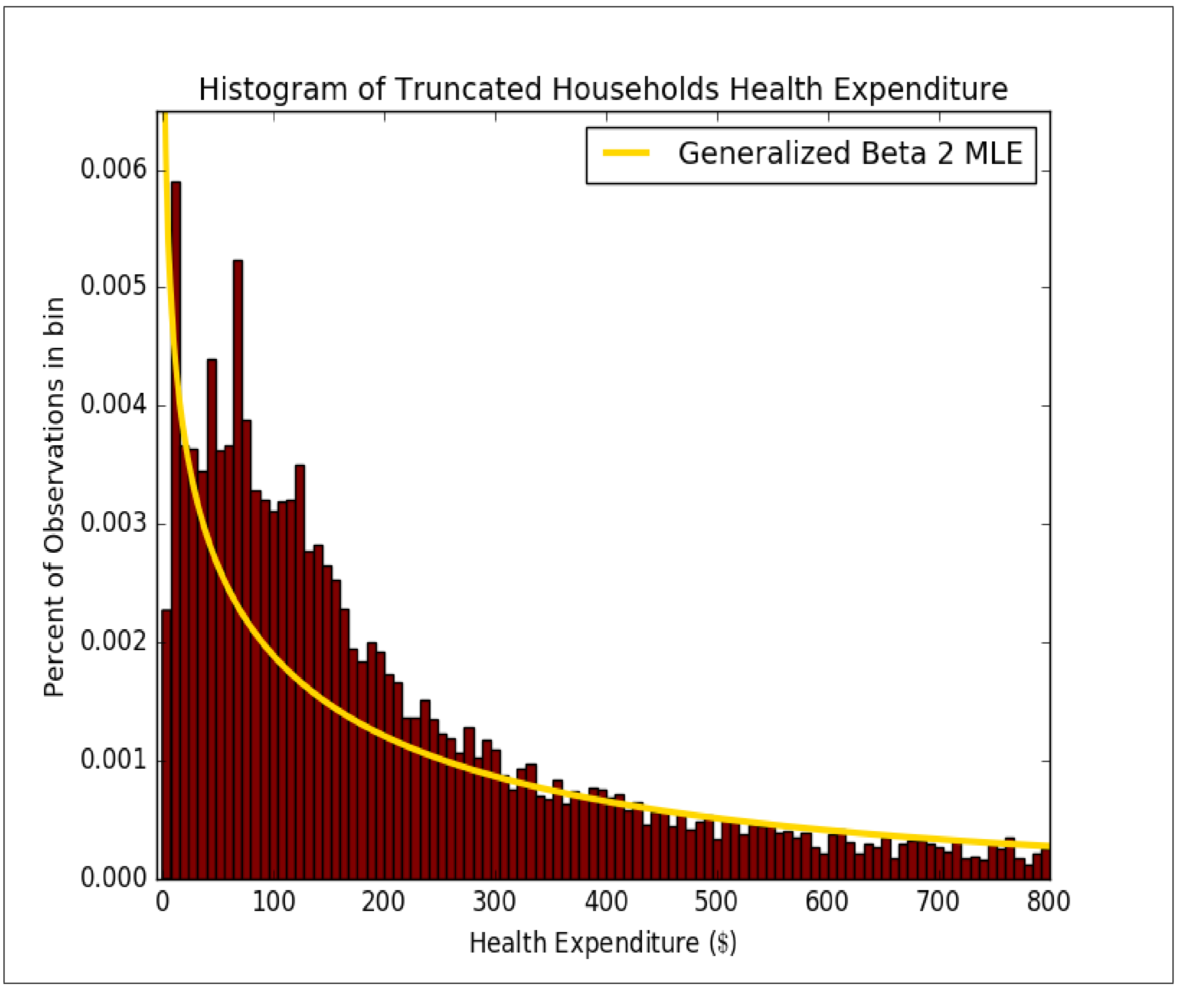
Figure 4: Households Health Expenditure (Truncated) Density Histogram Overlayed on a Generalized Gamma



Part (d). After using Maximum Likelihood Estimation to fit a Generalized Beta 2 GB2(x; a, b, p, q) distribution on the individual observation data, I have obtained:

- a = 0.610079339645
- b = 19344749.2778
- p = 1.32437879958
- q = 1000.08538436
- $\mathcal{L}(\hat{\theta})$ = -76126.9422141

Figure 5: Households Health Expenditure (Truncated) Density Histogram Overlayed on a Generalized Beta 2



Part (e). After performing Likelihood Ratio Test for each of the estimated in part (b) and (c), respectively against the GB2 specification in part (d), I have obtained the following test statistics and p-value:

- The Likelihood Ratio test statistics of the estimated Gamma distribution against the Generalized Beta 2 distribution is: -11611.2213855
- The p-value corresponds to the test statistics following a $\chi^2(4)$: 1.0

- The Likelihood Ratio test statistics of the estimated Generalized Gamma distribution against the Generalized Beta 2 distribution is: -11611.1076057
- The p-value corresponds to the test statistics following a $\chi^2(4)$: 1.0

Part (f). After approximating the probability of observing data points greater than \$1,000 under the estimated Generalized Beta 2 distribution, GB2(x; $\hat{a}_{MLE}, \hat{b}_{MLE}, \hat{p}_{MLE}, \hat{q}_{MLE}$), and the Gamma distribution, GA(x; $\hat{\alpha}_{MLE}, \hat{\beta}_{MLE}$), I have obtained:

- The estimated probability of observing a monthly health care expense of more than \$1,000 following the estimated Generalized Beta 2 distribution is: 0.1465991500932743
- The estimated probability of observing a monthly health care expense of more than \$1,000 following the estimated Gama distribution is: 0.4487256712610217