

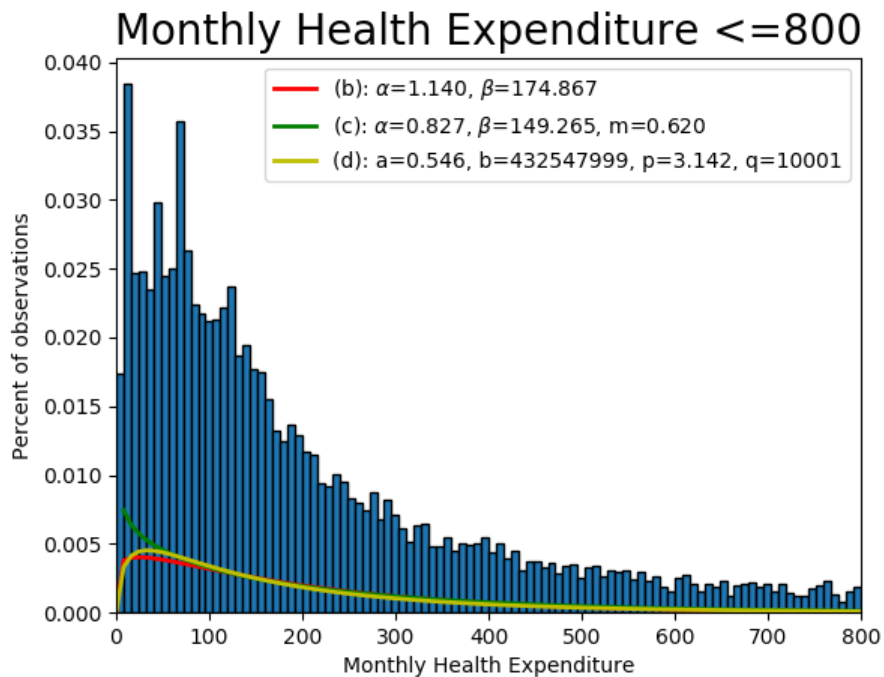
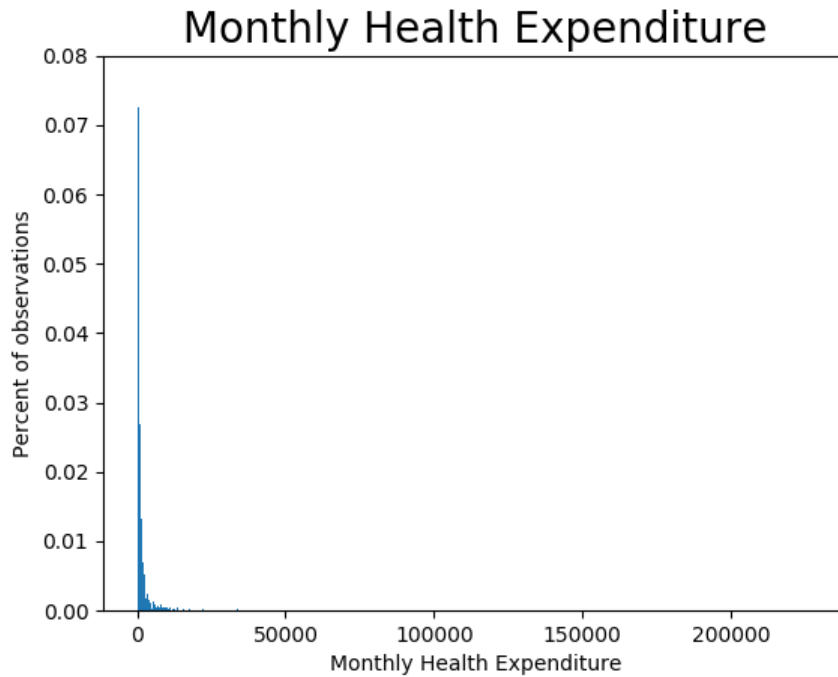
Problem Set 2

MACS 40200, Dr. Evans

Nobuyuki Bambi Furuta

Problem 1

(a) Mean: 720.278, Median: 172.210, Max: 227967.25, Min:0.01, Standard Deviation: 3972.664



The second histogram is preferred because the mean is around 720 and most of the data are concentrated below mean. In addition, it takes away outliers such as 227967.

(b) $\hat{\alpha} = 1.13976691235$, $\hat{\beta} = 174.866836155$, $\ln L = -56732.5986796$

The estimated gamma distribution is depicted in the graph above with the red line.

(c) $\hat{\alpha} = 0.826528658566$, $\hat{\beta} = 149.264584308$, $\hat{m} = 0.618997875459$, $\ln L = -56732.5986796$

The estimated generalized gamma distribution is depicted in the graph above with the green line.

(d) $\hat{a} = 0.545045409007$, $\hat{b} = 432528268.017$, $\hat{p} = 3.16725263428$, $\hat{q} = 9999.89896562$, $\ln L = -56800.0038081$

The estimated generalized beta 2 distribution is depicted in the graph above with the yellow line.

(e) p-value with the gamma distribution: 1.0

p-value with the generalized gamma distribution: 1.0

(f) Using the generalized beta 2 distribution, the probability that one has a monthly health care claim for more than \$1,000 is 0.01182126643710546.

With the gamma distribution, it is 0.00457132368760782.

Please see the attached Jupyter Notebook for the detail and codes used. I suppose that my calculation for the generalized gamma distribution is wrong, because it provided the weird estimated line.