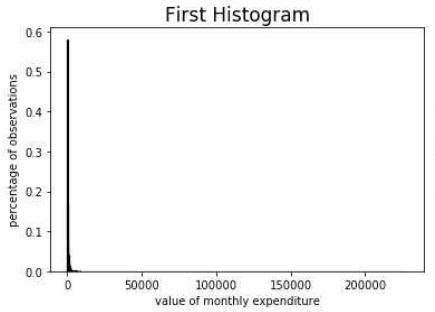
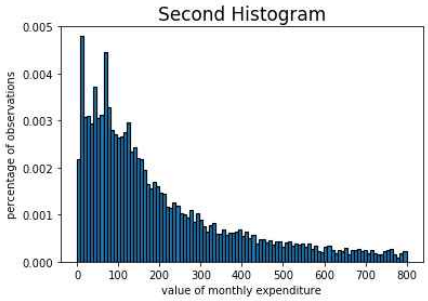
OSM Boot Camp, Economics, Week 4

Eun-Seok Lee

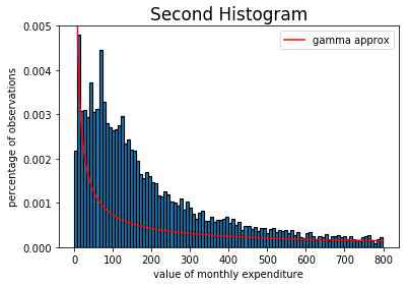
1. (a) **In second histogram, I could get an intuition of distribution regarding data in a reason that I can compare each height of bars.**

Mean = 720.277975327
Median 172.21
Max = 227967.25
Min = 0.01
Std = 3972.66375639

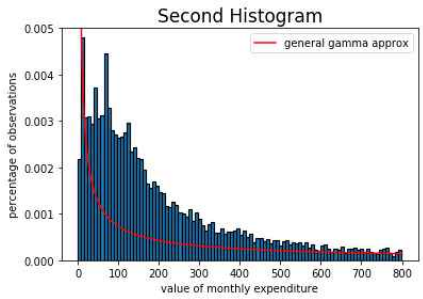




(b) **Definitely, its fitting is not good.**
alpha0_MLE= 0.221755308612 beta0_MLE= 21911.0646993
log_likelihood= -82076.4516057

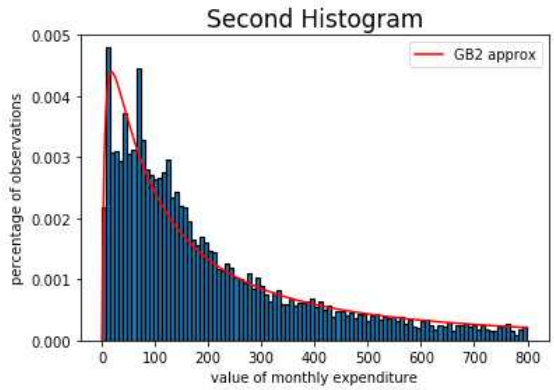


(c) **Generalized gamma approximation is also not appropriate for the distribution.**
alpha0_MLE= 0.222263480874 beta0_MLE= 21913.9457803 m_MLE= 0.997681780879
log_likelihood= -82076.745568



(d) **The fitting is now plausible at least from the judge by appearance. I also found that GB2 would become very similar to the shape of Generalized Gamma distribution when q goes to infinity. I used the PDF function that I wrote directly, so it is missing some points such as very large numbers in denominator. When computer recognize the large values as infinite, I corrected it to 1e+10. However, in built-in function, it handles more appropriately as you can see in (b) and (c); there was no error such as (d).**

a_MLE= 0.11404108258 b_MLE= 21913.9465417 p_MLE= 54.1132240339 q_MLE= 92.8968411452
log_likelihood= -74862.3422261



(e) **The result p values are 1.0, and it is plausible as you can compare graphs above.**
chi squared of H0 with 4 degrees of freedom p-value (gamma) = 1.0
chi squared of H0 with 4 degrees of freedom p-value (general gamma) = 1.0

(f) **I got the 13.4162% (Monte Carlo). and its difference from the ratio of original dataset is very small.**

Probability more than $1000(GB2) = 0.134161652373
Difference between Original Dataset = 0.0115512370794
Probability more than $1000(GA) = 0.455020433064
Difference between Original Dataset = 0.332410017771

2. In this problem, "a" should be 0.42 theoretically after the calculation with models. Also, mu should be around 9.5 ~ 10. I obtained very similar results between (a) and (b). I reported the result like below.

However, when I restored the equations with the parameters I obtained, its each element of the vector was far from zero even though its smaller than other value of parameters. (I arbitrarily set the parameter and checked whether it is close to zero but it was not.)

By the way, I think the values below are the best fit for data.

(a)
alpha_MLE= 0.457509740793 rho_MLE= 0.720493225849 mu_MLE= 9.52281143774 sigma_MLE= 0.0919960686003

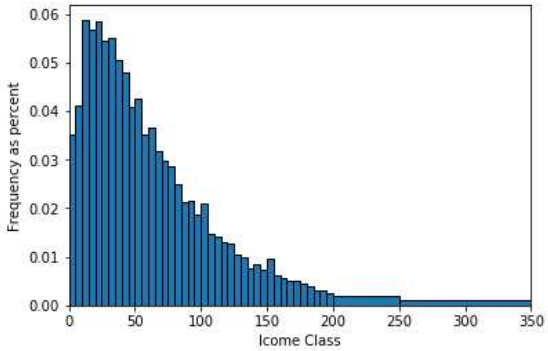
log_likelihood= 96.7069080647
VCV(MLE) = [[1.51156913e+01 -2.36913700e+01 -2.15416226e+02 -8.12357241e-01]
[-2.36913700e+01 4.02509812e+01 3.37273701e+02 8.95233561e-01]
[-2.15416226e+02 3.37273701e+02 3.06997571e+03 1.16211943e+01]
[-8.12357241e-01 8.95233561e-01 1.16211943e+01 9.02239012e-02]]

(b)
alpha_MLE= 0.457492063757 rho_MLE= 0.720504293511 mu_MLE= 9.37072990581 sigma_MLE= 0.0919959782314

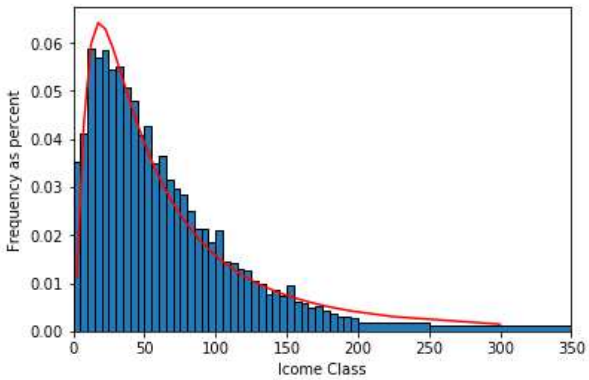
log_likelihood= 96.7069080414
VCV(MLE) = [[1.28817433e+00 -5.32358432e+00 -2.33234473e+01 -3.17381590e-01]
[-5.32358432e+00 2.59337351e+01 9.64073491e+01 1.67159559e+00]
[-2.33234473e+01 9.64073491e+01 4.22291108e+02 5.74832082e+00]
[-3.17381590e-01 1.67159559e+00 5.74832082e+00 1.11440783e-01]]

(c) Probability more than "r > 1"----->> [0.68268563]

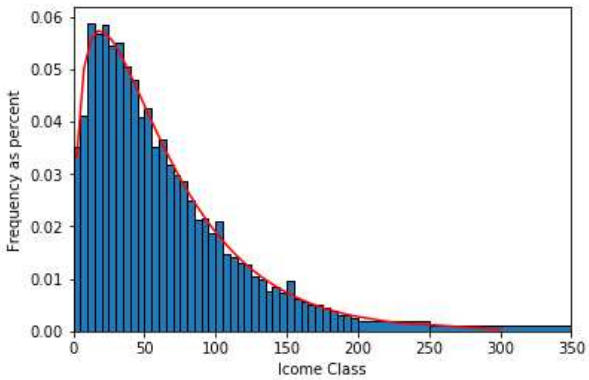
3. (a)



(b) mu_GMM= 3.9329187021 sigma_GMM= 1.01697916188 critic = 0.0302126130654

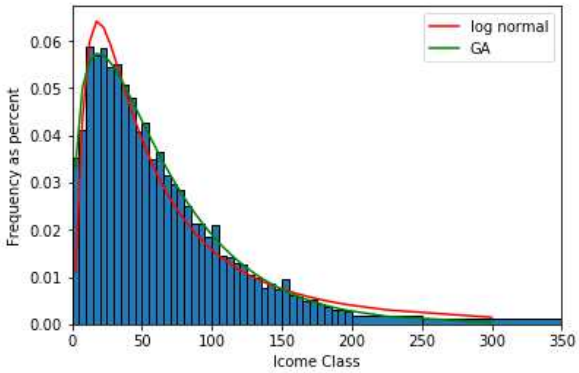


(c) alpha_GMM= 1.41639418788 beta_GMM= 44937.6193075 critic = 0.960093780566

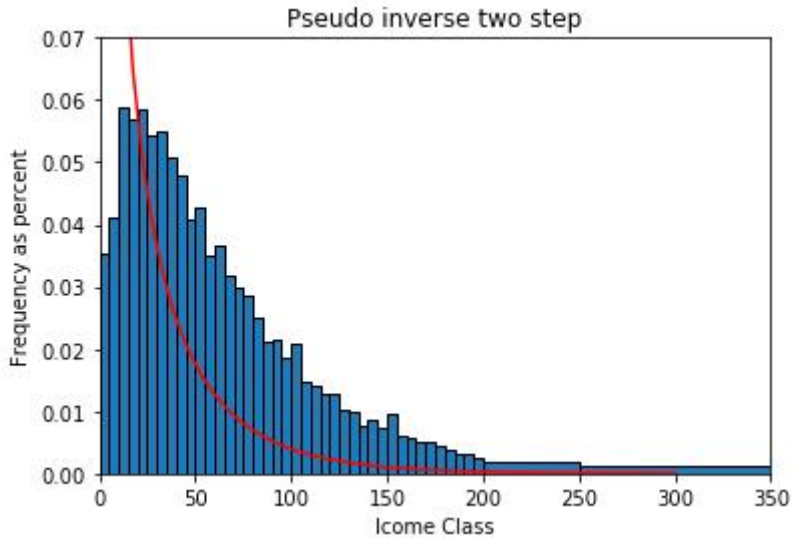


(d) **I could not decide which one has better fitting for the data by only the information of the picture. since GMM does not depend on prior model so much, I expect that, from large samples, each distribution would be very similar. However, I found some problems when I calculate GA distribution. When I worked on the scaled income divided by 1000, GMM could not find optimal converged values very well. GMM decided all the initial values are already converged values. When I worked on the original income, they could find the good fitting curve.**

alpha_GMM= 4.61265115176 beta_GMM= 20000.0009378 critic = 0.959104333249



(e) **In two-step estimation, the fit was worse. Also estimate of alpha changes a lot while the estimate of beta does not change so much.**



4. **Actually, it changes so much depending on initial values.**

'''

Initial values = alpha = 0.42, beta = 0.96, rho = 0, mu = 9.9

alpha_GMM= 0.419999995898 beta_GMM= 0.989999995 rho_GMM= 1.59258810433e-13 mu_GMM= 9.90000000003

'''

'''

Initial values = alpha = 0.42, beta = 0.97, rho = 0, mu = 9

alpha_GMM= 0.465655761447 beta_GMM= 0.98999999978 rho_GMM= 0.0131213684588 mu_GMM= 9.21730600318

'''