ZHAO Shengchun UMID 6773 4538

Choose Option 2:

What is the impact of non-normality on the t-test?

How large does the sample size need to be for the central limit theorem to overcome extreme non-normality?

Choose several distributions (including at least one mixture distribution) and multiple sample sizes to examine.

Project Proposal:

Analysis the impact of non-normality on the t-test:

I am going to use MC simulation to calculate the Type I error or Type II error of one-sample-t-test and two-sample-t-test to find out the robustness of t test to non-normal distributed data.

The data I am going to use not only based on some commonly seen data distributions (gamma, beta, chi), but also I am going to quote the **generalized lambda distribution** proposed by John S. Ramberg, Edward J. Dudewicz, Pandu R Tadikamalla, and Edward F. Mykytka(the article link:https://www.researchgate.net/publication/245001135_A_Probability_Distribution_and_Its_Uses_in_Fitting_Data), it could generate different type of data distributions (symmetric, extreme unsymmetric, etc.) by adjusting the four lambda values (scale, shape…). In the meantime, I will also control the sample size, variance equal or not, and compare the results' differences, then draw my conclusions.

Analysis the sample size needs to be for the central limit theorem to overcome extreme non-normality:

In this question, I will use similar data distributions mentioned in *Analysis the impact of non-normality on the t-test*. Also, I will use different sampling methods (simple sampling, Metropolis-Hastings Algorithm sampling, Gibbs sampling method (*in Gibbs, use mixture distributions to analysis*)) to compare whether the sampling methods will influence the sample size that is needed for CLT to overcome extreme non-normality distributed data. The MC method will be applied to determine the sample size to overcome the non-normality, and I will also check whether the MH algorithm or Gibbs method is appropriate to draw samples on a given data distribution.