ZIP : A modification over Poisson Distribution

Name : Shantanu Nayek Roll no. : 452

Supervisor : Prof. Surupa Chakraborty

Given a count data in hand and if the count seems to be unbounded, statisticians are more inclined towards fitting a Poisson Model blindfold in most cases. But this may sometimes lead to erroneous conclusions regarding analysis or other statistical works based on the data. We know that for a Poisson Model, the mean and variance for a sample are very close in general, but for the data it may not be. There may be another problem like if there is a value (say 0 or 1 or 2 or any integer) which has a very high frequency in the sample, then an inflation may be present at the probability curve at that particular value. Now let us suppose for example, we have a data in hand which is zero-inflated, it is not a wise job to fit a Poisson Distribution in general. We can think of some modifications to the Poisson distribution before fitting. One of such modifications is Zero Inflated Poisson Distribution.

Diane Lambert's Zero-inflated Poisson model concerns a random event containing excess zero-count data in unit time. The ZIP distribution with parameters π and λ, is denoted by ZIP (π, λ). The parameter π gives the extra probability thrust at the value 0; when it vanishes, ZIP (π, λ) reduces to Poisson(λ). The zero-inflated Poisson (ZIP) model mixes two zero generating processes. The first process generates zeros. The second process is governed by a Poisson distribution that generates counts, some of which may be zero.

In this study, data on number of seminars/webinars attended by college students during a certain span was collected. Since it is an unbounded count data, initially we showed interest to fit a Poisson Model for analysis of the data. But presence of excess zeroes, led to thinking of modification of the Poisson Model. Various measures were taken for justification of fitting a ZIP over a Poisson Model. Some of the justifications are, testing for the parameter π , standard errors for the two models, AIC values, etc. On giving preference to the ZIP model, our next point of interest is to estimate the parameters of the model by Method of Maximum Likelihood and Method of Moments. Moreover, we wish to give a measure of efficacy of the two estimation methods by comparing the standard errors of the Maximum Likelihood Estimator and Moment Estimator for π and λ.