NEW DISCRETE DISTRIBUTON FOR ZERO-INFLATED COUNT DATA

Peer Bilal Ahmad[1] Mohammad Kafeel Wani[2]

•

[1,2]Department of Mathematical Sciences, Islamic University of Science & Technology, Kashmir, INDIA, 192122

[1,*]bilalahmadpz@gmail.com , [2]wanimk5@gmail.com

Abstract

Over-dispersed models are commonly utilized when the variation is more than what the model actually predicts. Since one of the reasons for over-dispersion is the large number of zeros, we employ zero-inflated models instead of more traditional ones to handle this observed occurrence. We present a zero-inflated version of a discrete distribution that was developed in 2021 in our research. Significant statistical characteristics of the suggested model have been identified, such as moments, the over-dispersion feature, generating functions, and related measures, among others. We have carried the parametric estimation using the maximum likelihood estimate. Maximum likelihood estimates are checked for usefulness in a simulation exercise. We evaluated the applicability of our developed model using three real-world data sets,

Keywords: Over-dispersion, Zero-inflation, Discrete distribution, Simulation, Goodness-of-fit, Testing of hypothesis.

I. Introduction

To perform statistical analysis, statisticians use one of several methods, and these methods are the building blocks of statistical models. Mathematical representations of observable data are provided by statistical models. We choose statistical modeling of data for the purpose of understanding a wide range of random events across disciplines. Its applications are not limited to mathematical and statistical studies; rather, they permeate a wide variety of fields of study. Count data plays an important role in almost every scientific study, no matter how big or small. This data is used to draw inferences in relation to the population from which it is collected but, typically this data exhibits more variation than what is predicted by our hypothesized model. More precisely, this observable fact is called as over-dispersion (variance goes beyond mean). One cause of over-dispersion in count data is the presence of many more zeros than predicted by a statistical model. This phenomenon of finding excessive number of zeros is referred to as zero-inflation, and in order to model such dilemma, we use zero-inflated models rather than the more often used standard models.

Over-dispersion in count data due to zero-inflation is common, thus researchers are always developing new ideas and methods to shed light on this phenomenon. In order to deal with an excessive amount of zeros in count data, Lambert developed a new model called as zero-inflated Poisson (ZIP) regression model [7]. She used this model to investigate manufacturing flaws and