Properties and Applications of the Zero-Inflated Generalized Poisson (ZIGP) Regression Model

Count data:

Count Data refers to numerical data that represents the number of occurrences of an event within a fixed period, space, or category. It consists of non-negative integer values (0, 1, 2, …) and is typically modelled using discrete probability distributions like the Poisson, Negative Binomial, or Binomial distributions. Statistical Models for Count Data: Poisson Regression (when mean ≈ variance) Negative Binomial Regression (for overdispersed data) Zero-Inflated Models (when excessive zeros are present) Count data is widely used in epidemiology, ecology, finance, and various fields requiring frequency analysis.

Poisson fit:

The Poisson distribution assumes that the mean (μ) equals the variance (σ^2), i.e., $E(Y) = Var(Y)$. However, in many real-world datasets, the variance is often greater than the mean, a phenomenon known as overdispersion. When overdispersion is present, the Poisson model may not be appropriate, leading to underestimated standard errors and misleading inferences. To handle this, we can use, Negative Binomial Distribution. An extension of the Poisson distribution with an additional dispersion parameter allows $Var(Y) > E(Y)$ by introducing gamma-distributed heterogeneity in the Poisson mean. Generalized Poisson Distribution is another alternative that introduces an additional parameter to adjust for overdispersion. Can model both overdispersion ($Var(Y) > E(Y)$) and underdispersion ($Var(Y) < E(Y)$). These models provide better goodness-of-fit compared to the standard Poisson when the data exhibit overdispersion.

In many real-world datasets, count data often contain a large number of zeros, which standard Poisson or Negative Binomial models may fail to handle properly. This phenomenon is known as zero inflation. Models for Zero-Inflated Data Zero-Inflated Poisson (ZIP) – Extends the Poisson model by introducing a separate probability for extra zeros. Zero-Inflated Negative Binomial (ZINB) – Similar to ZIP but accounts for overdispersion as well. Hurdle Model – Separates the zero and nonzero counts into two different processes.

The Zero-Inflated Poisson (ZIP) model and the Zero-Inflated Generalized Poisson (ZIGP) model are used for count data with excess zeros, where traditional Poisson or Negative Binomial models fail to provide a good fit.

1**. Zero-Inflated Poisson (ZIP) Model**: Assumes that zeros arise from two sources, Structural zeros (always zero due to some underlying process), Random zeros (generated from a Poisson process). Uses a mixture model combining a Bernoulli (zero/nonzero) process and a Poisson distribution.

2. Zero-Inflated Generalized Poisson (ZIGP) Model: Extends the ZIP model by incorporating a dispersion parameter to handle overdispersion or underdispersion in the data. More flexible than ZIP, as it can model scenarios where variance ≠ mean, making it suitable for real-world datasets with varying dispersion levels. Applications of ZIP and ZIGP models are widely used in ecology, healthcare, finance, and insurance, where datasets often contain excess zeros (e.g., disease counts, accident reports, or species abundance data).

Here we use the Fisher Information Matrix (FIM) for the **Zero-Inflated Generalized Poisson (ZIGP) regression model** which is crucial for parameter estimation, inference, and hypothesis testing. It quantifies the amount of information the observed data provides about the parameters.

Finally, we can model the relation between Y and X variables, when modelling count data with excess zeros, two commonly used regression models are the **Zero-Inflated Poisson (ZIP) model** and the **Zero-Inflated Generalized Poisson (ZIGP) model**. These models account for overdispersion and excess zeros in the response variable Y, given predictor variables X.