A Flexible Zero-Inflated Count Regression Model

Eric S. Roemmele , Derek S. Young

Department of Statistics, University of Kentucky

JSM Presentation
August 2, 2018

Outline of Topics

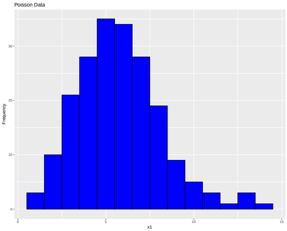
1. Introduction to Zero-Inflated Models
2. Semiparametric Extension to ZI
3. Hypothesis Testing
4. Real Data Example : Meth Lab Seizures
5. Future Directions

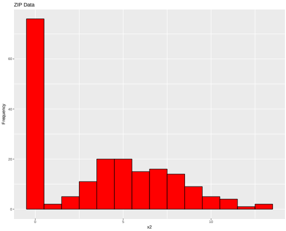
Outline of Topics

1. Introduction to Zero-Inflated Models
2. Semiparametric Extension to ZI
3. Hypothesis Testing
4. Real Data Example : Meth Lab Seizures
5. Future Directions

Motivation

- Suppose$\{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$ is observed, where the response Y_i is a discrete random variable, and $\boldsymbol{X}_i$ are covariates.
- Typically, we would model such a process by Poisson or Negative Binomial Regression.
- Sometimes, we observe *zero-inflation* - the observed amount of zeros significantly exceed model assumptions
- Employ Zero-Inflated models - define a latent process which says zeros come from two states - a degenerate and random state (i.e. zero comes from a count distribution)





ZI-Regression Model Definition

Definition

Let the discrete random variable Y_i be a count variable of interest and $(\boldsymbol{X}_i, \boldsymbol{Z}_i)$ be vectors of predictor variables measured for each subject i, $i = 1, \ldots, n$. Let $p\,(y|\mu, \theta)$ be a pmf function with mean μ and dispersion/scale/heterogeneity parameter θ. The **ZI pmf** is given by

$$f(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i, \mu, \theta) = \pi_i I\{y_i = 0\} + (1 - \pi_i)p(y|\mu_i, \theta) \tag{1}$$

where $0 \leq \pi_i \leq 1$. Parameterizing the count distribution in terms of its mean, μ_i, we relate this quantity to the predictor vector as

$$\log(\mu_i) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta},$$

while the mixing proportions can be modeled as

$$\text{logit}(\pi_i) = \boldsymbol{z}_i^{\mathrm{T}}\boldsymbol{\alpha},$$

where the predictors in $\boldsymbol{z}_i$ may be uncoupled from those predictors in $\boldsymbol{x}_i$.

ZI-Reg Continued

- Did zero come from the degenerate or count component?
- Example : $Y_i :=$ number of visits to a physician in a year
    - A patient could have zero visits because they were never ill, and thus had no reason to visit a doctor (strategic zero).
    - Or, maybe the patient was ill, but didn't have insurance, or followed alternative medicine (incidental zero).
- The most common choices for $p(\cdot)$ are the Poisson and Negative Binomial.
    - NB - $p(y|\mu,\theta) = \frac{\Gamma(\theta+y)}{y!\Gamma(\theta)}\left(\frac{\mu}{\theta+\mu}\right)^y\left(\frac{\theta}{\theta+\mu}\right)^\theta$
    - The expectation is $\mu + \mu^2\theta$
    - The NB is commonly used to characterize *over-dispersion* - the variability is increasing with the mean.
- For this presentation, we'll focus on the ZIP pmf

Optimization

- The observed log-likelihood for ZIP Regression is

$$\ell(\boldsymbol{\beta},\boldsymbol{\alpha}) = \sum_{y_i=0} \log(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\alpha} + \exp(-e^{\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}})) + \sum_{y_i>0} (y_i \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})) \\ - \sum_{i=1}^{n} \log(1+\exp(\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{\alpha})) - \sum_{i=1}^{n} \log(y_i!)$$

- Could always use any gradient based method (Newton-Raphson, IRLS, etc.).
- We'll employ the EM Algorithm (Dempster, Laird , Rubin 1997)

EM Algorithm

EM Algorithm for ZIP

Define

$$R_i = \begin{cases} 1 & Y_i \text{ is from the zero state} \\ 0 & Y_i \text{ is from the Poisson state} \end{cases}$$

Then, the log likelihood for the completed data $(\boldsymbol{Y}, \boldsymbol{R})$ is

$$\begin{aligned} \ell_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) &\propto \sum_{i=1}^{n} \Big(r_i \boldsymbol{w}^{\mathrm{T}} \boldsymbol{\alpha} - \log(1 + \exp(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{\alpha})) \Big) + \sum_{i=1}^{n} (1 - r_i) \Big(y_i \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta} - \exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}) \Big) \\ &= \ell_C(\boldsymbol{\alpha}) + \ell_C(\boldsymbol{\beta}) \end{aligned}$$

Iterate from $k = 1, \ldots$ til convergence

1. **E-Step** Update posterior memberships

$$r_i^{(k+1)} = \mathbb{P}(R_i = 1 | y_i, \boldsymbol{\beta}^{(k)}, \boldsymbol{\alpha}^{(k)})$$

2. **M-Step** Maximize ℓ_C above by

    1. Maximize $\ell_C(\boldsymbol{\beta})$, which is equivalent to running a Poisson regression of y_i on $\boldsymbol{x}_i$ with weights $1 - r_i^{(k+1)}$. Call this $\boldsymbol{\beta}^{(k+1)}$

    2. Maximize $\ell_C(\boldsymbol{\alpha})$, which is equivalent to running logistic regression of r_i on $\boldsymbol{w}_i$. Call this $\boldsymbol{\alpha}^{(k+1)}$.

Outline of Topics

1. Introduction to Zero-Inflated Models

2. Semiparametric Extension to ZI

3. Hypothesis Testing

4. Real Data Example : Meth Lab Seizures

5. Future Directions

Semiparametric Extension

- The (linear) ZIP and ZINB models are great at modeling zero-inflated data, and while count data is commonly heteroskedastic, the assumption of globally linear mixing proportions can often be too strong.
- Thus, we propose a semiparametric extension to the ZIP model.
    - Same set-up as before, but we now let $\pi(\boldsymbol{z})$ be an arbitrary function of continuous covariates.
    - For simplicity, let $\boldsymbol{z}$ be one dimensional, although we can extend to higher dimensions.
    - If the dimension of $\boldsymbol{z}$ is high (above 2 or 3), then one needs to be cognizant of the curse of dimensionality (Bellman 1957)

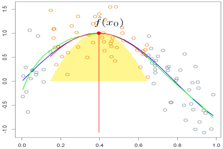
Local Likelihoods

- To learn $\pi(\cdot)$ we'll use local or smoothed likelihood method (Loader 1999)
- At a fixed point z_0 in the range of z, define the smoothed likelihood at z_0 as

$$\mathcal{L}^s_{z_0}(\theta_0) = \sum_{i=1}^{n} w_i \log(f(y_i|\theta_0)) \tag{2}$$

  where $w_i = h^{-1}K(\frac{z_i - z_0}{h})$ where $K(\cdot)$ is a kernel function and h is the bandwidth.

- h controls the size of the neighborhood around z_0 - think of bin-width on histogram
- Then, we estimate $\widehat{\theta}(z_0) = \underset{\theta_0}{\text{argmax}}\ \mathcal{L}^s_{z_0}(\theta_0)$



Estimation of the Semiparametric ZIP

- The challenge in estimating our model is that in addition to the mixture structure, the model contains both (global) parametric and local components.
- Therefore, we propose a three-step "back-fitting" algorithm, which alternates between local and global estimation.
- An "EM-like" algorithm is proposed for each step.
- Define $\boldsymbol{\theta}(z_0) = (\pi(z_0), \boldsymbol{\beta}(z_0))$.

Backfitting Procedure

Backfitting Procedure

1. Initial Local Step - For the observed $\mathcal{Z} = \{z_1, \ldots, z_n\}$, maximize the local-likelihood at each $z_i \in \mathcal{Z}$

$$\mathcal{L}_1^S(\boldsymbol{\theta}(z_i)) = \sum_{j=1}^{n} K_h(z_j - z_i) \log f(y_i | \boldsymbol{x}_i, z_i, \boldsymbol{\theta}(z_j)) \tag{3}$$

   To do this use an "EM" Algorithm analogous to the parametric EM Algorithm. Call these estimates $\widetilde{\pi}(z_j)$ and $\widetilde{\boldsymbol{\beta}}(z_j)$.

2. Global Step - Given the mixing proportions estimates $\widetilde{\pi}(z_i)$ for $i = 1, \ldots, n$, perform a global estimation of $\boldsymbol{\beta}$ by maximizing

$$\mathcal{L}_2(\boldsymbol{\beta}) = \sum_{i=1}^{n} \log f(y_i | \boldsymbol{\beta}, \widetilde{\pi}(z_i), \boldsymbol{x}_i, z_i) \tag{4}$$

   As before, an EM algorithm is implemented for estimation. Call this estimate $\widehat{\boldsymbol{\beta}}$.

3. Final Local Step - Given the global estimate of $\boldsymbol{\beta}$, update the mixing proportions at each z_i by maximizing

$$\mathcal{L}_3^S(\pi(z_i)) = \sum_{j=1}^{n} K_h(z_j - z_i) \log f(y_i | \pi(z_i), \widehat{\boldsymbol{\beta}}, \boldsymbol{x}_i, z_i) \tag{5}$$

   Again, this is done by an "EM" Algorithm. Call this estimate $\widehat{\pi}(z_i)$

Asymptotic Properties of Estimators

Asymptotic Properties of Estimators (Huang and Yao 2012)

1. Let $\boldsymbol{\theta}(z) = (\pi(z), \boldsymbol{\beta}(z))$. Assume $\sqrt{nh} \to \infty$ and $h \to 0$.
   Then, the estimator at step 1

$$\sqrt{nh}\{\widetilde{\boldsymbol{\theta}}(z) - \boldsymbol{\theta}(z) - b(z)h^2 + o(h^2)\} \xrightarrow{L} N(0, g^{-1}(z)\mathcal{I}^{-1}(z)v) \tag{6}$$

2. The estimator of $\boldsymbol{\beta}$ at Step 2

$$\sqrt{n}\{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\} \xrightarrow{L} N(0, B^{-1}\Sigma B^{-1}) \tag{7}$$

3. Finally, the final estimator of the mixing proportions $\pi(z)$

$$\sqrt{nh}\{\widehat{\pi}(z) - \pi(z) - \widehat{b}(z)h^2 + o(h^2)\} \xrightarrow{D} N(0, g^{-1}(z)\mathcal{I}_{\pi}^{-1}(z)v) \tag{8}$$

It can be shown that the asymptotic bias and variance of $\widehat{\pi}(z)$ is smaller than $\widetilde{\pi}(z)$.

Ascent Properties

- The Classic EM Algorithm posses the *ascent property* - the objective function increases at each iteration.
- Can't claim the overall likelihood increases at each iteration, but weaker ascent properties can be established.

Ascent Property (Huang and Yao 2012)

1. **Asymptotic Ascent** For the "EM" Algorithm in step one, if $nh \to \infty$ and $h \to 0$, it follows

$$\liminf_{n\to\infty}\; n^{-1}\left[\mathcal{L}_1(\boldsymbol{\theta}^{(k+1)}(z)) - \mathcal{L}_1(\boldsymbol{\theta}^{(k)}(z))\right] \geq 0$$

   in probability.

2. The ascent property in Step 2 follows immediately from the theory of ordinary EM Algorithms
3. For the EM Algorithm in Step 3, the local likelihood will be monotonically increasing at any z; that is, $\mathcal{L}_3(\pi^{(k+1)}(z)) \geq \mathcal{L}_3(\pi^{(k)}(z))$

Outline of Topics

1. Introduction to Zero-Inflated Models
2. Semiparametric Extension to ZI
3. Hypothesis Testing
4. Real Data Example : Meth Lab Seizures
5. Future Directions

Generalized Likelihood Ratio Test

- We are interested in testing

$$H_0 : \pi(z) \in \mathcal{M}_\alpha$$
$$H_1 : \pi(z) \notin \mathcal{M}_\alpha$$

  where $\mathcal{M}_\alpha$ is a parametric family of models.

- Fan (1999) argued that the LRT is still a good test provided that the nonparametric quantity is replaced with a good estimator.

$$\lambda(h) = \ell(H_1) - \ell(H_0) \tag{9}$$

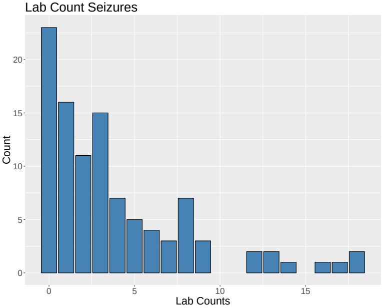
- Furthermore, the limiting distribution should be free of any nuisance parameters ($\boldsymbol{\beta}$ and the true value $\pi(z)$), and should be χ^2.
- However, there is no well-defined degrees of freedom in H_1, so having a closed form rejection region is not feasible.
- But, we can employ a parametric bootstrap to estimate the limiting distribution, and then obtain a bootstrap p-value.

Outline of Topics

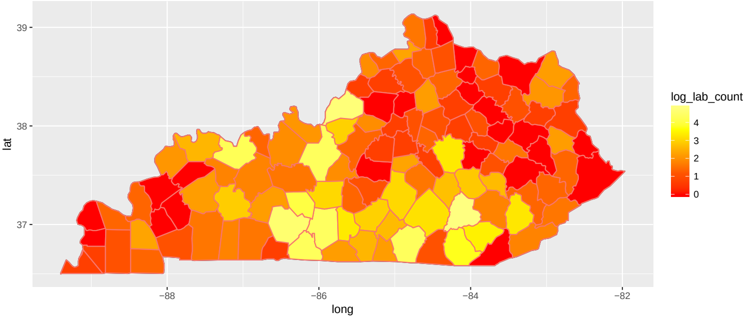
1. Introduction to Zero-Inflated Models
2. Semiparametric Extension to ZI
3. Hypothesis Testing
4. Real Data Example : Meth Lab Seizures
5. Future Directions

Meth Lab

- The data set consists of the number of clandestine meth lab seizures in each county of Kentucky, Louisiana, and Illinois in the year 2011. The main predictor of interest is the amount of pseudophedrine (PSE) sold per 100 people. In total, there are 286 counties across the three states.
- Predictors :
  - Poisson Component : State, PSE
  - Zero State : PSE
- 98 out of 286 counties were zero ($\approx 34\%$). Most of the data (67%) in Louisiana are zero.



Heat Map for KY



Estimated Model

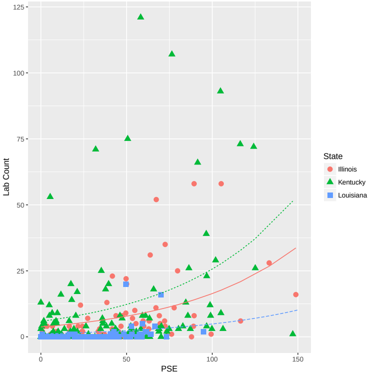


Figure: Fitted Values

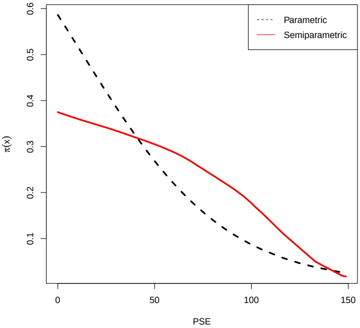


Figure: Mixing Proportions

Meth Analysis

Model	Log-Like
Semi-ZIP	-1843.352
ZIP	-1845.496

The LRT statistic is $\lambda_n = 2.14$, with the bootstrap p-value of .14 on $B = 100$ bootstrap iterations.

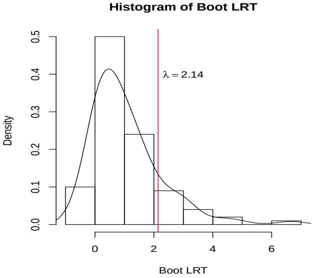


Figure: Wilks Phenomenon

Outline of Topics

1. Introduction to Zero-Inflated Models
2. Semiparametric Extension to ZI
3. Hypothesis Testing
4. Real Data Example : Meth Lab Seizures
5. Future Directions

Conclusions & Future Directions

- Zero-Inflated Models are a great way to account for excessive zeros in the response.
- The Semi-ZIP model provides flexibility in modeling the amount of zero inflation, and can be a confirmation of the parametric model.
- **Future Directions**
  - Spatial Model
    - As we can see from the heat map, neighboring counties tend to have similar counts
    - Account for auto-correlation by a Conditionally Autoregressive Model (Cressie 1991)
    - Usually done through Bayesian methodology