Zero Inflated Poisson Regression, Lambart, Feb 92

- Model for count data with excess zeros.
- prob. p - the only possible observation is zero
  $1-p$ - a poisson(λ) random variable is observed.

eg When manufacturing equipment is properly aligned, defects may be nearly impossible. But when it is misaligned, defects may occur according to a Poisson (λ) distribution.

p = prob of perfect state
λ = mean no. of defects in imperfect state

may / may not be related to each other **
both can depend on covariates *.

MLE's are approximately normal for ZIP regression models in large samples.

CI can be constructed either by
- inverting LRT or using ⟵ better
- approximating normality of MLEs

- Poisson mean $= \lambda \Rightarrow ne^{-\lambda}$ items with no defects (given n items) (approx)

* $\log(\lambda)$ = linear f^n of covariates
$\log\left(\frac{p}{1-p}\right)$ = linear f^n of covariates

** eg when p is a decreasing f^n of λ, (eg $p = \frac{1}{1+\lambda^{\tau}}$)
the prob of perfect state and mean of imperfect state improve or deteriorate together

Mathematical Model

responses $\overline{Y} = (Y_1, \ldots Y_n)^T$ are independent and

$$Y_i \sim 0 \quad \text{with probability } p_i$$
$$\sim \text{Poisson}(\lambda_j) \quad \text{with probability } 1-p_i$$

$$\Rightarrow \quad Y_i = 0 \quad \text{with prob. } p_i + \underbrace{(1-p_i)e^{-\lambda_i}}_{k=0 \text{ in poisson}(\lambda)}$$
$$= k \quad \text{" "} \quad (1-p_i)\frac{e^{-\lambda_i}\lambda_i^k}{k!}$$
$$k = 1, 2, 3, \ldots$$

Parameters $\overline{\lambda} = (\lambda_1, \ldots \lambda_n)^T$

$\overline{P} = (p_1, \ldots, p_n)^T$ satisfy

$$\log(\overline{\lambda}) = \overline{B}\overline{\beta}$$
$$\text{logit}(\overline{p}) = \overline{G}\overline{\gamma}$$ where $\overline{B}$ and $\overline{G}$ are covariate matrices.

If the same covariates affect $\overline{p}$ and $\overline{\lambda}$ (i.e. $\overline{B} = \overline{G}$) we can reduce the number of parameters (from twice compared to Poisson regression) by thinking of $\overline{p}$ as a function of $\overline{\lambda}$.

Assume that the function is known upto a constant
$\Rightarrow$ The parameters needed are nearly halved.

TODO — Computation Analysis ML / Least squares
$O(K^3)$

Parameterize wrt the constant

$$\log(\lambda) = \bar{B}\bar{\beta}$$
$$\text{logit}(\bar{p}) = -\tau \bar{B}\bar{\beta}$$

ZIP(τ)

- Logit link for $\bar{p}$ is symmetric around 0.5.
- $\log(-\log \bar{p}) = \tau \bar{B}\bar{\beta}$ log-log link

$\log(\log(1-\bar{p})) = -\tau \bar{B}\bar{\beta}$ Complementary log log link

$\log(-\log \bar{p}) = \bar{B}\bar{\beta} + \log \tau$ Additive log log link

$\text{logit}(\bar{p}) = \log \alpha - \tau \bar{B}\bar{\beta}$ linear logit link

$\log(-\log \bar{p}) = \log \alpha + \tau \bar{B}\bar{\beta}$ linear log log link

Equivalently

- $p_i = (1+\lambda_i^{\tau})^{-1}$ Logit
- $p_i = \exp(-\lambda_i^{\tau})$ LL

$p_i = 1 - \exp(-\lambda_i^{-\tau})$ CLL

for $\tau > 0$, $\lambda_i \uparrow \Rightarrow p_i \downarrow$

if $\tau \to \infty$ $p_i \to 0$ for fixed λ_i

As $\tau \to -\infty$ $p_i \to 1$ } except additive log-log link ($\tau < 0$ not allowed) ($\tau \to 0$, $p_i \to 1$)

$\tau < 0$, $p_i \uparrow \Rightarrow \lambda_i \uparrow$ (Poisson mean $\uparrow$ as excess zeros become more likely)

$\Rightarrow$ improve fraction of perfect items leads to more defects in imperfect items.

Maximum likelihood estimation -

Number of parameters that can be estimated depends on richness of the data.

→ If the observed information matrix is non singular

Case 1 - $\bar{\lambda}$ and $\bar{p}$ are unrelated. (ZIP regression)

Log-likelihood (log, logit links)

$$L(\bar{\gamma}, \beta \mid \bar{y}) = \sum_{y_i=0} \log P(y=y_i) + \sum_{y_i>0} \log P(y=y_i)$$

Now

$$\sum_{y_i=0} \log y_i = \sum_{y_i=0} \log\left(p_i + (1-p_i)e^{-\lambda_i}\right)$$

$$= \sum_{y_i=0} \log\left[(1-p_i)\left(\frac{p_i}{1-p_i} + \exp(-\lambda_i)\right)\right]$$

$$= \sum_{y_i=0} \log\left[e^{\bar{G}_i\bar{\gamma}} + \exp\left(-e^{\bar{B}_i\bar{\beta}}\right)\right] + \underbrace{\sum_{y_i=0} \log(1-p_i)}$$

$$\underbrace{-\sum_{y_i=0} \log\left(1+e^{\bar{G}_i\bar{\gamma}}\right)}_{(1)}$$

$$\frac{p_i}{1-p_i} = \exp(\bar{G}_i\bar{\gamma})$$

$$\Rightarrow p_i = \exp(\bar{G}_i\bar{\gamma}) - p_i \exp(\bar{G}_i\bar{\gamma})$$

$$\Rightarrow p_i = \frac{\exp(\bar{G}_i\bar{\gamma})}{1+\exp(\bar{G}_i\bar{\gamma})}$$

$$\Rightarrow 1-p_i = \frac{1}{1+\exp(\bar{G}_i\bar{\gamma})} \quad -(2)$$

and

$$\sum_{y_i>0} \log P(y=y_i) = \sum_{y_i>0} \log(1-p_i) + \sum_{y_i>0} \log\left(\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)$$

$$= -\sum_{y_i>0} \log\left(1+e^{\bar{G}_i\bar{\gamma}}\right) + \sum_{y_i>0} \log(y_i!) - \sum_{y_i>0} \lambda_i + \sum_{y_i>0} y_i \log \lambda_i$$

$$= -\sum_{y_i>0} \log\left(1+e^{\bar{G}_i\bar{\gamma}}\right) - \sum_{y_i>0}\log(y_i!) - \sum_{y_i>0} e^{\bar{B}_i\bar{\beta}} + \sum_{y_i>0} y_i \bar{B}_i\bar{\beta} \quad (2)$$

Adding (1) and (2)

$$\Rightarrow L(\bar{\gamma},\bar{\beta}\mid\bar{y}) = \sum_{y_i=0} \log\left[e^{\bar{G}_i\bar{\gamma}} + \exp\left(-e^{\bar{B}_i\bar{\beta}}\right)\right]$$

$$- \sum_{i=1}^{n} \log\left(1+e^{\bar{G}_i\bar{\gamma}}\right) - \sum_{y_i>0}\log(y_i!)$$

$$+ \sum_{y_i>0}\left(y_i\bar{B}_i\bar{\beta} - e^{\bar{B}_i\bar{\beta}}\right)$$

$\bar{G}_i$ and $\bar{B}_i$ are i^{th} rows of $\bar{G}$ and $\bar{B}$ respectivly

$\Rightarrow$ sum of exponentials inside log complicates the maximization of log-lik.

Assume latent variable Z_i denotes whether Y_i is from perfect zero state ($Z_i=1$) or from Poisson distribution ($Z_i=0$).

Then log-likelihood of complete data $(\bar{y},\bar{Z})$

$$L_c(\bar{\gamma},\bar{\beta}\mid\bar{y},\bar{Z}) = \sum_{i=1}^{n}\log\left(f(Z_i\mid\bar{\gamma})\right) + \sum_{i=1}^{n}\log\left(f(y_i\mid Z_i,\bar{\beta})\right)$$

$$P(Z_i=1) = P_i = f(\bar{\gamma})$$

$$P(Y=y_i) = P(Z_i=1)\, I_{y_i=0} \times P(y_i\mid y_i=0) + P(Z_i=0)\, I_{y_i} = \text{Poiss}(\lambda_i) \times P(y_i\mid \text{poisson})$$

Marginalize over Z_i

$$= \sum_{Z_i=0,1} P(Z_i=1)\underbrace{I_{y_i=0}}_{Z_i} + \underbrace{P(Z_i=0)}_{1-P(Z_i=1)}(1-Z_i)\times \underbrace{\left(\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)}_{f(\bar{\beta})}$$

$$f(y_i\mid\bar{\gamma},\bar{\beta}) = f(Z_i\mid\bar{\gamma})\, f(y_i\mid Z_i,\bar{\beta})$$

$$P(Z_i = 1) = p_i$$

$$P(Z_i = 0) = 1 - p_i$$

$$P(y = y_i \mid Z = z_i) = \begin{cases} 1_{\{y_i = 0\}} & z_i = 1 \\ \dfrac{e^{-\lambda_i}\lambda^{y_i}}{y_i!} & z_i = 0 \end{cases}$$

Law of TP

$$P(y = y_i) = \sum_{z_i} P(y = y_i \mid Z = z_i)\, P(z_i)$$

$$= 1_{\{y_i = 0\}} P(Z_i = 1) + \frac{e^{-\lambda_i}\lambda^{y_i}}{y_i!} P(Z_i = 0)$$

$$= 1_{\{y_i = 0\}} p_i + \frac{e^{-\lambda_i}\lambda^{y_i}}{y_i!}(1 - p_i)$$

$$L_c(\bar{\gamma}, \bar{\beta} \mid \bar{y}, \bar{z}) = \sum_{i=1}^{n} \log\Big[\big(\underbrace{P(y = y_i \mid Z = z_i)}_{P(y = y_i \mid Z = z_i, \bar{\beta})}\, \underbrace{P(z_i)}_{P(z_i \mid \bar{\gamma})}\Big]$$

given z_i $P(y)$ depends on $\bar{\beta}$

z_i depends on $p \to \bar{\gamma}$

$$= \sum_{i=1}^{n} \log\left(P(Z = z_i \mid \bar{\gamma})\right) + \sum_{i=1}^{n} \log\left(P(y = y_i \mid z_i, \bar{\beta})\right)$$

Now

$$\log\left(P(Z = z_i \mid \bar{\gamma})\right) = \begin{cases} \log p_i & z_i = 1 \\ \log(1 - p_i) & z_i = 0 \end{cases}$$

$$= z_i \log p_i + (1 - z_i)\log(1 - p_i)$$

$$= z_i \log \frac{\exp(G_i \bar{\gamma})}{1 + \exp(G_i \bar{\gamma})} - (1 - z_i)\log\left[1 + \exp(\bar{G}_i \bar{\gamma})\right]$$

$$= z_i \bar{G}_i \bar{\gamma} - z_i \log\left[1 + \exp(\bar{G}_i \bar{\gamma})\right] - \log\left[1 + \exp(\bar{G}_i \bar{\gamma})\right] + z_i \log\left[1 + \exp(\bar{G}_i \bar{\gamma})\right]$$

$$= z_i \bar{G}_i \bar{\gamma} - \log\left[1 + e^{\bar{G}_i \bar{\gamma}}\right]$$

Also,

$$\log\left(P(y = y_i \mid z_i, \bar{\beta})\right) = \begin{cases} \log\left(1_{\{y_i = 0\}}\right) & z_i = 1 \\ \log\left(\frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}\right) & z_i = 0 \end{cases}$$

0 or $-\infty$

Not useful for maximizing

$$= (1 - z_i)\left(y_i \log \lambda_i - \lambda_i - \log(y_i!)\right)$$

$$= (1 - z_i)\left(y_i \bar{B}_i \bar{\beta} - e^{\bar{B}_i \bar{\beta}}\right) - (1 - z_i)\log(y_i!)$$

$$\Rightarrow L_c(\bar{\gamma}, \beta \mid \bar{y}, \bar{z}) = \sum_{i=1}^{n}\left(z_i G_i \bar{\gamma} - \log\left[1 + e^{\bar{G}_i \bar{\gamma}}\right]\right) + \qquad (4)$$

$$\sum_{i=1}^{n} (1 - z_i)\left(y_i \bar{B}_i \bar{\beta} - e^{\bar{B}_i \bar{\beta}}\right) + \sum_{i=1}^{n} (1 - z_i)\log(y_i!)$$

$L_c(\bar{\gamma} \mid \bar{y}, \bar{z})$

$L_c(\bar{\beta} \mid \bar{y}, \bar{z})$

Now $L_c(\bar{\gamma}\mid\bar{y},\bar{z})$ and $L_c(\bar{\beta}\mid\bar{y},\bar{z})$ can be maximized separately.

<u>EM Algorithm for log-likelihood maximization</u> -

<u>E step</u> - Estimate expectation of z_i under current estimates of $(\bar{\gamma},\bar{\beta})$

$z_i^{(k)}$ = Posterior mean of z_i under current estimates $\bar{\gamma}^{(k)}, \bar{\beta}^{(k)}$.

$$= P(\text{perfect state}\mid y_i, \bar{\gamma}^{(k)}, \bar{\beta}^{(k)})$$

$$= \frac{\overbrace{P[y_i\mid \text{perfect state}]\,P[\text{perfect state}]}^{P[y_i \cap \text{perf. state}]}}{\underbrace{P[y_i\mid\text{perfect state}]\,P(\text{perfect state}) + P[y_i\mid\text{poisson}]\,P(\text{Poisson})}_{P(y_i)}}$$

$$= \begin{cases} 0 & \text{if } y_i = 1,2,3,\ldots \\ \dfrac{1\cdot p}{1\cdot p + (1-p)\cdot e^{-\lambda}} & \text{if } y_i = 0 \end{cases}$$

$$= \frac{1}{1+\frac{1-p}{p}e^{-\lambda_i}} = \frac{1}{1+e^{-\bar{G}_i\bar{\gamma}}\cdot e^{-\exp(\bar{B}_i\bar{\beta})}}$$

$$= \left(1+e^{-\bar{G}_i\bar{\gamma}-\exp(\bar{B}_i\bar{\beta})}\right)^{-1}$$

<u>M step for $\bar{\beta}$</u> - Find $\bar{\beta}^{(k+1)}$ by maximizing $L_c(\bar{\beta}\mid\bar{y},\bar{z}^{(k)})$

- Weighted log-linear regression (Poisson) with weights $1-\bar{z}^{(k)}$

M step for $\bar{\gamma}$ -

Maximize $L_c(\bar{\gamma}|\bar{y}, \bar{Z}^{(k)}) = \sum_{y_i=0} Z_i^{(k)} \bar{G}_i \bar{\gamma} - \sum_{y_i=0} Z_i^{(k)} \log(1+e^{\bar{G}_i\bar{\gamma}}) -$

$$\sum_{i=1}^{n} (1-Z_i^{(k)}) \log(1+e^{\bar{G}_i\bar{\gamma}})$$

$$(\because \; Z_i \bar{G}_i \bar{\gamma} = 0 \text{ for } y_i > 0) \quad (5)$$

(from (4))

let n_0 of y_i's are 0 s.t.

$$\bar{y}_*^T = (\underbrace{y_1, \ldots, y_n}_{\text{Non zero}}, \underbrace{y_{i_1}, \ldots, y_{i_{n_0}}}_{0}) \text{ and}$$

$$G_*^T = (\bar{G}_1^T, \ldots, \bar{G}_n^T, \bar{G}_{i_1}^T, \ldots, \bar{G}_{i_{n_0}}^T)$$

$$P_*^T = (P_1, \ldots, P_n, P_{i_1}, \ldots, P_{i_{n_0}})$$

Define diagonal matrix $\bar{W}^{(k)}$ with diagonal $\bar{w}^{(k)}$

$$\bar{W}^{k} = (1-Z_1^{(k)}, \ldots, 1-Z_n^{(k)}, Z_{i_1}^{(k)}, \ldots, Z_{i_{n_0}}^{(k)})^T$$

In this notation,

Removes $i = n+1$ to $n+n_0$ $\because y_i^* = 0$

Removes entries if $y_i = 0$ in 1 to n $Z_i = 1 \Rightarrow 1-Z_i = 0$

$$? \quad L_c(\bar{\gamma}|\bar{y}, \bar{Z}_p^{(k)}) = \underbrace{\sum_{i=1}^{n+n_0} y_{*i} w_i^{(k)} \bar{G}_{*i} \bar{\gamma}}_{\text{sum over non zero } y_i \text{ entries } y_i>0} - \sum_{i=1}^{n+n_0} w_i^{(k)} \log(1+e^{\bar{G}_{*i}\bar{\gamma}})$$

sum over all non-zero entries (y_i) in 1 to n and zero from $n+1$ to $n+n_0$ $\equiv$ sum over all n entries in original data $\bar{y}$

$\bar{y}_* \bar{W} \bar{G}_*^T \bar{\gamma} - \bar{W} \log(1+\bar{G}_*^T \bar{\gamma})$

$G^T W \bar{y}_* - \bar{W} \frac{e^{G_*^T \gamma} \cdot G^T}{1+G_*\gamma}$

Gradient is $\bar{G}_*^T \bar{W}^{(k)} (\bar{y}_* - \bar{P}_*) = 0$

Hessian / Neg. of Information Matrix is $-\bar{G}_*^T \bar{W}^{(k)} \bar{Q}_* \bar{G}_*$

where $\bar{Q}_*$ is diagonal matrix with $\bar{P}_*(1-\bar{P}_*)$ on the diagonal.

$\frac{e^{G\gamma}}{1+e^{G\gamma}}$

$\frac{e^{G\gamma} \cdot G(1+e^{G\gamma}) - e^{G\gamma} e^{G\gamma} G}{(1+e^{G\gamma})^2}$

$= \frac{e^{G\gamma} G(1+e^{G\gamma} - e^{G\gamma})}{(1+e^{G\gamma}) \; e^{G\gamma}} = P G (1-P) G$

- These functions are identical to those for a weighted logistic regression with

response $\bar{q}_*$

Covariate matrix $\bar{G}_*$

prior weights $\bar{W}^{(k)}$

Thus $\bar{\gamma}^{(k)}$ can be found by weighted logistic regression.

<u>Convergence and Initialization</u>

Use MLE for positive Poisson log-likelihood as initial guess for $\bar{\beta}$.

$$L_+(\bar{\beta} \mid \bar{y}_+) = \sum_{y_i>0} \left(y_i \bar{B}_i \beta - e^{\bar{B}_i \bar{\beta}} \right) - \sum_{y_i>0} \log\left(1 - e^{-\exp(\bar{B}_i \bar{\beta})}\right) - \sum_{y_i>0} \log(y_i!) \quad (\text{from}(2))$$

Gradient

$$\bar{B}_+^T \left(\bar{y}_+ - \frac{e^{\bar{B}_+ \bar{\beta}}}{1 - \exp(-e^{\bar{B}_+ \bar{\beta}})} \right) = 0$$

Hessian $\quad \bar{B}_+^T \bar{D} \bar{B}_+$

$+ \rightarrow$ only elements or rows corresponding to positive y_i's are used.

$\bar{D}$ - diagonal matrix with diagonal

$$\frac{e^{\bar{B}_+ \beta}\left(1 - e^{-e^{\bar{B}_+ \beta}}\left(1 + e^{\bar{B}_+ \bar{\beta}}\right)\right)}{\left(1 - \exp(-e^{-\bar{B}_+ \bar{\beta}})\right)^2}$$

Gradient eqn solved by iteratively reweighted least squares.

- $\log(1+\exp(a))$ should be computed carefully in tails.

$\bar{\gamma}$ guess found to be unimportant

If $\bar{\gamma}$ includes an intercept, estimate intercept as $\log \frac{\hat{p}_0}{1-\hat{p}_0}$

$$\hat{p}_0 = \frac{\#(y_i=0) - \sum \exp(-e^{\bar{B}_i \beta^{(0)}})}{n}$$

average probability of an excess 0.

<u>Case 2</u> $\bar{p}$ a function of $\bar{\lambda}$: ZIP(τ) model

$$L(\bar{\beta}, \tau / \bar{y}) = \sum_{y_i=0} \log(e^{-\tau \bar{B}_i \beta} + \exp(-e^{\bar{B}_i \bar{\beta}}))$$

$$+ \sum_{y_i>0} (y_i \bar{B}_i \bar{\beta} - e^{\bar{B}_i \bar{\beta}})$$

$$- \sum_{i=1}^{n} \log(1 + e^{-\tau \bar{B}_i \bar{\beta}})$$

EM is not useful here. The author suggests Newton's method with initial guess $\bar{\beta}^{(0)} = \hat{\beta}_u$

$\tau^{(0)} = -\text{median}\left(\frac{\hat{\bar{\gamma}}_u}{\hat{\bar{\beta}}_u}\right)$ where $(\hat{\bar{\gamma}}_u, \hat{\bar{\beta}}_u)$ are ZIP MLEs

exclude the intercepts of $\hat{\gamma}_u$ and $\hat{\bar{\beta}}_u$

If results are not satisfactory,

- first maximize over β for few choices of fixed $\tau^{(0)}$
- Then starting at $(\hat{\beta}(\tau_0), \tau_0)$ with the largest log-likelihood.

Standard Errors and CIs -

In large samples, the MLE's $(\hat{\bar{\gamma}}, \hat{\bar{\beta}})$ for ZIP regression and $(\hat{\bar{\beta}}, \hat{\tau})$ for ZIP(τ) regression are approximately normal with means (γ, β) and (β, τ) and variances equal to the inverse of observed information matrices $\bar{I}^{-1}$ and $\bar{I}_{\tau}^{-1}$ respectively.

Thus for large enough samples, the MLEs and regular functions of MLEs, such as prob of defect and mean number of defects are nearly unbiased.

LR confidence Intervals - More difficult to compute but are often more trustworthy.

For β_1 in ZIP regression, 2 sided $(1-\alpha)100\%$ CI

- compute set of β_1 for which

$$2\left[L(\hat{\bar{\gamma}}, \hat{\bar{\beta}}) - \max_{\bar{\gamma}, \bar{\beta}_{-1}} L\left((\beta_1, \bar{\beta}_{-1}), \bar{\gamma}\right)\right] < \chi^2_{\alpha,1}$$

vector $\bar{\beta}$ without its first element β_1

upper α quantile of a χ^2 dis with 1 df

** Van den Broek (1995) developed a score test (zero. test) for comparing a standard Poisson model with ZIP model. (assuming p does not dep. on covariates)

** ZIP model vs ZINB score test also developed by Ridout (2001)

** Vuong test for ZIP vs Hurdle model comparison