PS5841

Data Science in Finance & Insurance

Poisson Regression

Yubo Wang

Autumn 2022

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Poisson Distribution

- For modeling event counts $Y \sim Po(\lambda)$

$$\Pr(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \ldots$$

- Mean & Variance

$$E(Y) = Var(Y) = \lambda$$

- Sum $Y_1 \sim Po(\lambda_1), Y_2 \sim Po(\lambda_2)$, independent

$$Y_1 + Y_2 \sim Po(\lambda_1 + \lambda_2)$$

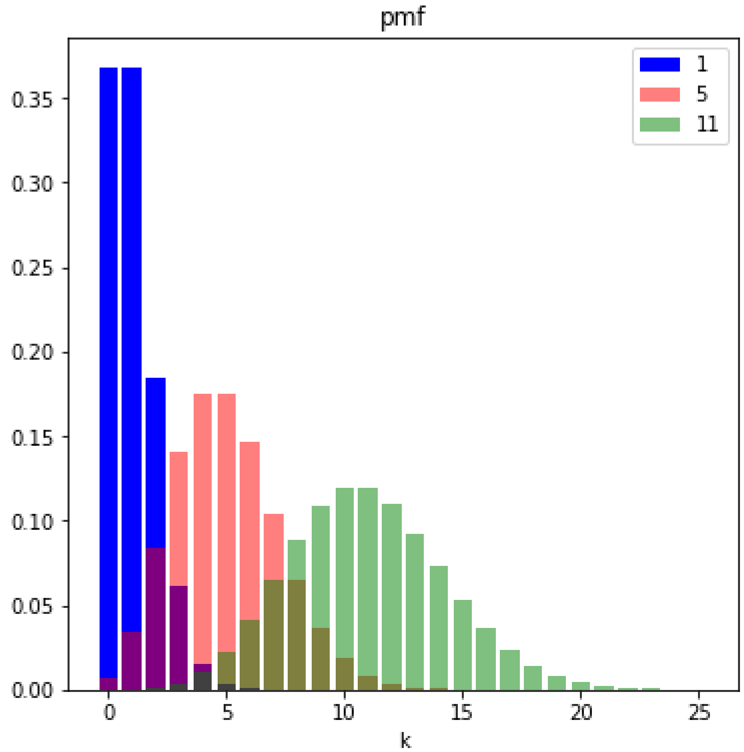
- Waiting time

$$\Pr(T > t) = e^{-\lambda t}$$

Poisson Distribution

- For modeling event counts

$$\Pr(Y = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \ldots$$



Poisson Regression

$$E(Y_i) = \mu_i = n_i \theta_i$$

exposure

$$\log(\mu_i) = \log(n_i) + \boldsymbol{x}_i^T \boldsymbol{\beta}$$

offset

Pearson Residuals

- Pearson residuals

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}}$$

$$X^2 = \sum_{i=1}^{N} r_i^2$$

- Pearson chi-squared statistic

$$X^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(N - p)$$

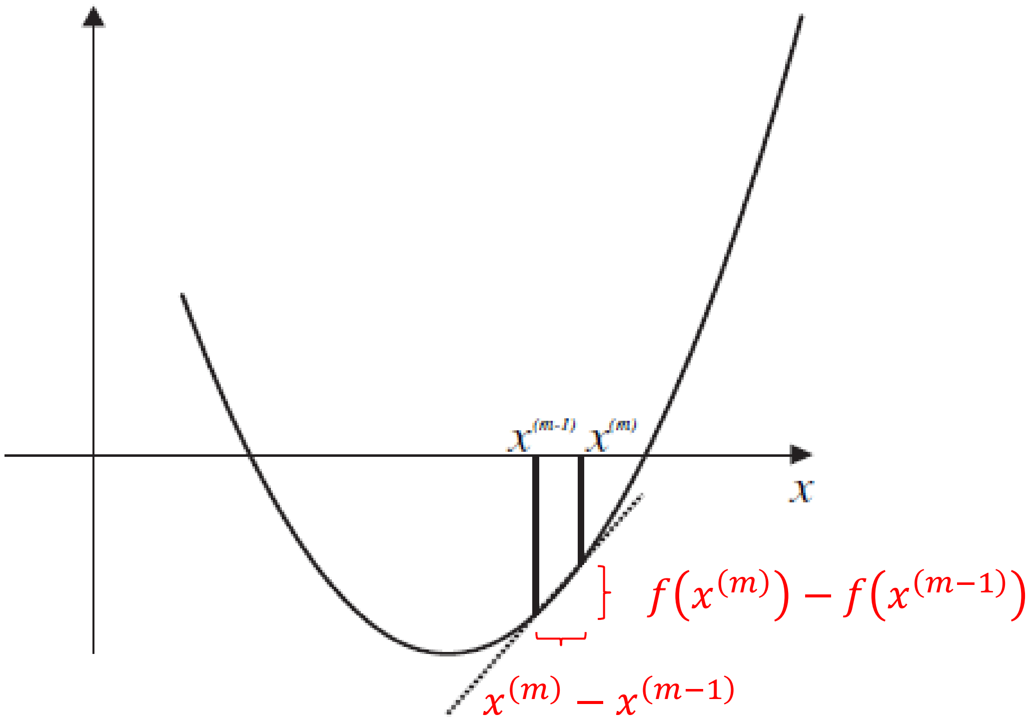
Pseudo R^2

$$pseudo\ R^2 = \frac{l(\widehat{\boldsymbol{\beta}}_{min}) - l(\widehat{\boldsymbol{\beta}})}{l(\widehat{\boldsymbol{\beta}}_{min})}$$

where the minimum model contains only the intercept.

Newton Raphson

Find a root of a function



Root: $f\left(x^{(m)}\right) = 0$ for some m.

$$f'\left(x^{(m-1)}\right) = \frac{f\left(x^{(m)}\right) - f\left(x^{(m-1)}\right)}{x^{(m)} - x^{(m-1)}}$$

$$x^{(m)} = x^{(m-1)} - \frac{f\left(x^{(m-1)}\right)}{f'(x^{(m-1)})}$$

Newton Raphson

- $f: \mathcal{R} \to \mathcal{R}$

$$x^{(r+1)} = x^{(r)} - \frac{f\left(x^{(r)}\right)}{f'\left(x^{(r)}\right)}$$

- $f: \mathcal{R}^p \to \mathcal{R}^q$

$$\boldsymbol{x}^{(r+1)} = \boldsymbol{x}^{(r)} - \left[J\left(\boldsymbol{x}^{(r)}\right)\right]^{-1} f\left(\boldsymbol{x}^{(r)}\right)$$

$$J = \frac{\partial f}{\partial \boldsymbol{x}}, \qquad J_{ij} = \frac{\partial f_i}{\partial x_j}$$

Example: MLE

- To maximize the log-likelihood function (or minimize $-1 \times$ log-likelihood function)

$$l(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{n} l_i(\boldsymbol{\beta}; y_i)$$

- Updating

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - [J(\boldsymbol{\beta}^{(r)})]^{-1} \boldsymbol{\nabla} l(\boldsymbol{\beta}^{(r)})$$

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[H(\boldsymbol{\beta}^{(r)})\right]^{-1} \left(\frac{\partial l(\boldsymbol{\beta}^{(r)})}{\partial \boldsymbol{\beta}}\right)$$

$$J = \frac{\partial}{\partial \boldsymbol{\beta}} \left(\frac{\partial l}{\partial \boldsymbol{\beta}^T}\right), \qquad J_{ij} = \frac{\partial}{\partial \beta_i} \left(\frac{\partial l}{\partial \beta_j}\right) = \frac{\partial}{\partial \beta_i} \left(\frac{\partial \sum_k l_k}{\partial \beta_j}\right)$$

$$H = \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}, \qquad H_{ij} = \frac{\partial^2 l}{\partial \beta_i \partial \beta_j} = \frac{\partial^2 \sum_k l_k}{\partial \beta_i \partial \beta_j}$$

Poisson Regression MLE (1)

- Likelihood

$$L(\boldsymbol{\beta};\boldsymbol{y}) = \prod_{i=1}^{n} \frac{(\mu_i)^{y_i} e^{-\mu_i}}{y_i!}, \qquad \mu_i = e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}$$

$$l(\boldsymbol{\beta};\boldsymbol{y}) = \sum_{i=1}^{n} y_i \boldsymbol{x}_i^T\boldsymbol{\beta} - e^{\boldsymbol{x}_i^T\boldsymbol{\beta}} - \log y_i!$$

- Normal equations

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{n} (y_i - \mu_i) x_{ij} = 0, \qquad x_{i0} = 1$$

Poisson Regression MLE (2)

- If $\mu_i = e^{\beta_0+\beta_1 x_i}$

$$l(\boldsymbol{\beta};\boldsymbol{y}) = \sum_{i=1}^{n} y_i(\beta_0+\beta_1 x_i) - e^{\beta_0+\beta_1 x_i} - \log y_i!$$

- Normal equations

$$\frac{\partial l}{\partial \beta_0} = \sum_{i=1}^{n}(y_i - e^{\beta_0+\beta_1 x_i}) \rightarrow \frac{\partial^2 l}{\partial \beta_0^2} = -\sum_{i=1}^{n} e^{\beta_0+\beta_1 x_i}$$

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{n}\left(y_i - e^{\beta_0+\beta_1 x_i}\right)x_i \rightarrow \frac{\partial^2 l}{\partial \beta_1^2} = -\sum_{i=1}^{n} x_i^2 e^{\beta_0+\beta_1 x_i}$$

$$\frac{\partial^2 l}{\partial \beta_1 \partial \beta_0} = \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1} = -\sum_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i}$$

- Hessian

$$H = \begin{bmatrix} -\sum_{i=1}^{n} e^{\beta_0+\beta_1 x_i} & -\sum_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i} \\ -\sum_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i} & -\sum_{i=1}^{n} x_i^2 e^{\beta_0+\beta_1 x_i} \end{bmatrix}$$

Poisson Regression MLE (3)

update	b0	b1	log-likelihood	gain in log-likelihood
0	0.15	0.04	-88.700971	
1	0.125189501	0.112141212	-74.20676404	14.4942
2	0.214265247	0.087113309	-60.85135835	13.3554
3	0.301197211	0.077049722	-60.25317102	0.5982
4	0.307842724	0.076359801	-60.25116153	0.0020
MLE	0.307866396	0.076357325	-60.25116151	0.0000

Example: LSE

- To minimize the loss function

$$R(\boldsymbol{\beta}) = \sum_{i=1}^{n} R_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i(\boldsymbol{\beta}))^2$$

- Updating

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - [J(\boldsymbol{\beta}^{(r)})]^{-1} \nabla R(\boldsymbol{\beta}^{(r)})$$

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \left[H(\boldsymbol{\beta}^{(r)})\right]^{-1} \left(\frac{\partial R(\boldsymbol{\beta}^{(r)})}{\partial \boldsymbol{\beta}}\right)$$

$$J = \frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{\partial R}{\partial \boldsymbol{\beta}^T}\right), \qquad J_{ij} = \frac{\partial}{\partial \beta_i}\left(\frac{\partial \sum_k R_k}{\partial \beta_j}\right)$$

$$H = \frac{\partial^2 R}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}, \qquad H_{ij} = \frac{\partial^2 \sum_k R_k}{\partial \beta_i \partial \beta_j}$$

Example

- To minimize the loss function

$$R(\boldsymbol{\beta}) = \sum_{i=1}^{n} R_i(\boldsymbol{\beta}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2})^2$$

- $\boldsymbol{\nabla} R(\boldsymbol{\beta}) = \begin{pmatrix} \sum_i e_i \\ \sum_i e_i x_{i1} \\ \sum_i e_i x_{i2} \end{pmatrix}$

- $J(\boldsymbol{\beta}) = \begin{pmatrix} X_0^T X_0 & X_0^T X_1 & X_0^T X_2 \\ X_1^T X_0 & X_1^T X_1 & X_1^T X_2 \\ X_2^T X_0 & X_2^T X_1 & X_2^T X_2 \end{pmatrix}$

That was