Maximum Likelihood Estimation (MLE) for Poisson Regression

Model Setup

Poisson regression models count data $y_i \in \mathbb{N}_0$ (i.e., non-negative integers). The target variable is modeled as:

$$y_i \sim \text{Poisson}(\lambda_i), \quad \text{where} \quad \lambda_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$$

Here, $\boldsymbol{\beta} \in \mathbb{R}^d$ is the parameter vector, and the exponential ensures $\lambda_i > 0$.

Likelihood Function

The probability mass function of the Poisson distribution is:

$$P(y_i \mid \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}$$

Substituting $\lambda_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$:

$$P(y_i \mid \mathbf{x}_i; \boldsymbol{\beta}) = \frac{e^{-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \exp(\mathbf{x}_i^\top \boldsymbol{\beta} \cdot y_i)}{y_i!}$$

The likelihood over n i.i.d. samples is:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{-\exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \exp(\mathbf{x}_i^\top \boldsymbol{\beta} \cdot y_i)}{y_i!}$$

Log-Likelihood

Taking the log of the likelihood:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i(\mathbf{x}_i^\top \boldsymbol{\beta}) - \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) - \log(y_i!) \right]$$

Gradient of the Log-Likelihood

Let $\lambda_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$. The gradient is:

$$\frac{\partial \ell}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} (y_i - \lambda_i)\mathbf{x}_i$$

Vectorized form:

$$\nabla_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}) = X^\top (\mathbf{y} - \boldsymbol{\lambda})$$

Where:

- $X \in \mathbb{R}^{n \times d}$ is the feature matrix,
- $\mathbf{y} \in \mathbb{R}^n$ is the vector of observed counts,
- $\boldsymbol{\lambda} \in \mathbb{R}^n$ is the vector of expected counts $\lambda_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$

MLE Optimization

No closed-form solution exists for $\boldsymbol{\beta}$, so we use iterative methods:

- **Gradient Ascent**:
$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} + \eta \cdot \nabla_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta})$$
- **Newton-Raphson**
- **Quasi-Newton** methods (e.g., BFGS)

Deriving Poisson Regression from the Exponential Family

The Poisson distribution belongs to the exponential family:

$$P(y \mid \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right)$$

For the Poisson distribution (with $\phi = 1$), we have:

$$\theta = \log \lambda, \quad b(\theta) = e^\theta, \quad c(y) = -\log(y!)$$

In the GLM framework:

- The mean $\mu = \mathbb{E}[y] = \lambda = \exp(\mathbf{x}^\top \boldsymbol{\beta})$
- The canonical link function is the log: $g(\mu) = \log(\mu)$
- The linear predictor is $\theta = \mathbf{x}^\top \boldsymbol{\beta}$

Summary

- Poisson regression models count data using a log-linear model.
- The likelihood is based on the Poisson distribution.
- MLE seeks to maximize the log-likelihood.
- The gradient can be derived analytically.
- No closed-form solution; optimization is done numerically.
- Poisson regression is a GLM with log link and Poisson noise.