UNIVERSITY OF WATERLOO

ACTSC 632 Spring 2023

Data Science with Actuarial Applications

Week 7

Xintong Li

Department of Statistics and Actuarial Science

Last Week

- Y is a key ratio: claim frequency or claim severity
- X is a vector of rating factors, modeled as categorical variables
- Exponential Dispersion Models (EDM)
- Multiplicative models (logarithmic link function)
- Claim frequency: Poisson regression
- Claim severity: Gamma regression

Today

- Estimating the parameters of a GLM from:
- Writing the log-likelihood function
- Taking the derivative of the log-likelihood function
- Solving for the maximum likelihood estimators (MLE)

2.6 Parameter Estimation (MLE)

Taking derivative using the chain rule

Evaluating each partial derivative

Remark: satured model

Corresponding GLM overfits. However, it is useful in the definition of deviance.

Multiplicative Poisson frequency model

Multiplicative gamma severity model

Summary of Theoretical Results

- **Goal:** estimate the parameters of a GLM using MLE
- The $(r+1)$ equations:

$$\sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{v(\mu_i)\, g'(\mu_i)} x_{ij} = 0.$$

- For multiplicative Poisson frequency model:

$$\sum_{i=1}^{n} w_i (y_i - \mu_i) x_{ij} = 0.$$

- For multiplicative gamma severity model:

$$\sum_{i=1}^{n} w_i \frac{y_i - \mu_i}{\mu_i} x_{ij} = 0.$$

Example: Moped dataset

- **Goal:** Use everything we learned to build a GLM for the moped dataset.
- **Tasks:**
  1. Load the data
  2. Explore the data
  3. Set the base tariff cells
  4. Fit the frequency model
  5. Fit the severity model

Recap from Last Time

- Estimating the parameters of a GLM using MLE
- Example in R: Moped dataset

Today

- Hypothesis testing for GLMs
- Deviance statistic: likelihood ratio test
- Pearson's chi-squared test
- Estimating the dispersion parameter ϕ
- Testing hierarchical models
- Confidence intervals based on Fisher's information

2.7 GLM Model Building

Deviance Statistic from Likelihood Ratio Test

Deviance Statistic from Likelihood Ratio Test

Null hypothesis:

$$H_0 : \text{fitted GLM model}$$

Test Statistic:

$$D^* = 2\{l(y_1, \ldots, y_n; , \theta, \phi) - l(\hat{\mu}_1, \ldots, \hat{\mu}_n; , \theta, \phi)\}.$$

Under general conditions, the deviance statistic D^* is approximately chi-squared distributed with $n - p$ degrees of freedom, where

- n is the number of observations,
- and $p = r + 1$ is the number of parameters in the model.

Compare the deviance statistic with $c_{1-\alpha}$,

- $c_{1-\alpha}$ is the $1 - \alpha$ quantile of the chi-squared distribution with $n - p$ degrees of freedom.
- Reject if $D^* > c_{1-\alpha}$, accept otherwise.

Using p-values, reject if p-value $< \alpha$, accept otherwise.

Pearson's Goodness-of-Fit Test

Null hypothesis: H_0 : fitted GLM model

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\text{Var}(Y_i)} = \frac{1}{\phi} \sum_{i=1}^{n} w_i \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)}.$$

X^2 is approximately chi-squared distributed with $n - (r + 1)$ degrees of freedom.
Compare the Pearson test statistic with $c_{1-\alpha}$,

- $c_{1-\alpha}$ is the $1 - \alpha$ quantile of the chi-squared distribution with $n - p$ degrees of freedom.
- Reject if $X^2 > c_{1-\alpha}$, accept otherwise.

Using p-values, reject if p-value $< \alpha$, accept otherwise.

Estimation of ϕ

Testing Hierarchical Models

Consider two models $M_s \subset M_t$:

$$H_0 : \text{ data comes from } M_s$$
$$H_\alpha : \text{ data comes from the more complicated } M_t$$

Test statistic:

$$D^*(\mathbf{y}, \hat{\mu}^{(s)}) - D^*(\mathbf{y}, \hat{\mu}^{(t)}).$$

- M_s and M_t are required to be from the same EDM family.
- $D^*(\mathbf{y}, \hat{\mu}^{(s)}) - D^*(\mathbf{y}, \hat{\mu}^{(t)})$ is approximately chi-squared distributed with $p_t - p_s$ degrees of freedom.
- Using p-values, reject if p-value $< \alpha$, accept otherwise.

Confidence intervals based on Fisher's information