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

Last Week

- Y is a key ratio: claim frequency or claim severity
- X is a vector of rating factors, modeled as categorical variables

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)

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)

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

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:

Today

- Estimating the parameters of a GLM from:
- Writing the log-likelihood function

Today

- Estimating the parameters of a GLM from:
- Writing the log-likelihood function
- Taking the derivative of the log-likelihood function

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: saturated 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

Summary of Theoretical Results

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

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

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:

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

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.

Example: Moped dataset

- **Goal:** Use everything we learned to build a GLM for the moped dataset.
- **Tasks:**
  1. Load the data

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

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

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

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

Recap from Last Time

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

Today

- Hypothesis testing for GLMs

Today

- Hypothesis testing for GLMs
- Deviance statistic: likelihood ratio test

Today

- Hypothesis testing for GLMs
- Deviance statistic: likelihood ratio test
- Pearson’s chi-squared test

Today

- Hypothesis testing for GLMs
- Deviance statistic: likelihood ratio test
- Pearson's chi-squared test
- Estimating the dispersion parameter ϕ

Today

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

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:

$$\mathrm{H}_0 : \text{fitted GLM model}$$

Deviance Statistic from Likelihood Ratio Test

Null hypothesis:

$$\mathrm{H}_0 : \text{fitted GLM model}$$

Test Statistic:

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

Deviance Statistic from Likelihood Ratio Test

Null hypothesis:

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

Test Statistic:

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

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.

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.

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

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

X^2 is approximately chi-squared distributed with $n - (r + 1)$ degrees of freedom.

Pearson's Goodness-of-Fit Test

Null hypothesis: H_0 : fitted GLM model

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

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.

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

- **Method 1**: Matching moments with the Pearson statistic

$$\hat{\phi} = \frac{1}{n-(r+1)} \sum_{i=1}^{n} w_i \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)}.$$

where we set $X^2 = n-(r+1)$ by matching the expected value of X^2

$$\mathbb{E}[X^2] = n-(r+1).$$

Estimation of ϕ

- **Method 1**: Matching moments with the Pearson statistic

$$\hat{\phi} = \frac{1}{n - (r+1)} \sum_{i=1}^{n} w_i \frac{(y_i - \hat{\mu}_i)^2}{\nu(\hat{\mu}_i)}.$$

where we set $X^2 = n - (r+1)$ by matching the expected value of X^2

$$\mathbb{E}[X^2] = n - (r+1).$$

- **Method 2**: Matching moments with the deviance statistic

$$\hat{\phi}_D = \frac{D}{n - (r+1)}.$$

where we set $D^* = n - (r+1)$ by matching the expected value of D^*

Testing Hierarchical Models

Consider two models $M_s \subset M_t$:

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

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:

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

Testing Hierarchical Models

Consider two models $M_s \subset M_t$:

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

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.

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.

Testing Hierarchical Models

Consider two models $M_s \subset M_t$:

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

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

To get the confidence intervals, we need the Fisher's information matrix:

$$[\mathrm{I}]_{jk} = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k}\right], \qquad j, k = 0, 1, \ldots, r.$$

Confidence intervals based on Fisher's information

To get the confidence intervals, we need the Fisher's information matrix:

$$[I]_{jk} = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k}\right], \qquad j, k = 0, 1, \ldots, r.$$

It can be shown that $I = X^T D X$, where D is a diagonal matrix with diagonal elements

$$d_i = \frac{w_i}{\phi(\hat{\mu}_i)(g'(\hat{\mu}_i)^2)}.$$

Confidence intervals based on Fisher's information

To get the confidence intervals, we need the Fisher's information matrix:

$$[I]_{jk} = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \beta_j \partial \beta_k}\right], \qquad j, k = 0, 1, \ldots, r.$$

It can be shown that $I = X^T D X$, where D is a diagonal matrix with diagonal elements

$$d_i = \frac{w_i}{\phi(\hat{\mu}_i)(g'(\hat{\mu}_i)^2)}.$$

Asymptotically, we have $\hat{\beta} \sim N(\beta, I^{-1})$. Thus we can find the confidence intervals for β_j:

$$\hat{\beta}_j \pm z_{1-\alpha/2}\sqrt{[I^{-1}]_{jj}}.$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

Confidence intervals based on Fisher's information

In practice, we also want to find the confidence intervals for μ_i.

Confidence intervals based on Fisher's information

In practice, we also want to find the confidence intervals for μ_i.

We do this by finding the confidence intervals for η_i:

$$\hat{\eta}_i \pm z_{1-\alpha/2}\sqrt{[I_\eta^{-1}]_{jj}}.$$

where $I_\eta^{-1} = (x_{i0}, x_{i1}, \ldots x_{ir}) I^{-1} (x_{i0}, x_{i1}, \ldots x_{ir})^T$ is the variance of $\hat{\eta}_i$.

Confidence intervals based on Fisher's information

In practice, we also want to find the confidence intervals for μ_i.

We do this by finding the confidence intervals for η_i:

$$\hat{\eta}_i \pm z_{1-\alpha/2}\sqrt{[I_\eta^{-1}]_{jj}}.$$

where $I_\eta^{-1} = (x_{i0}, x_{i1}, \dots x_{ir}) I^{-1} (x_{i0}, x_{i1}, \dots x_{ir})^T$ is the variance of $\hat{\eta}_i$.

Then we can find the confidence intervals for μ_i:

$$\left(e^{\hat{\eta}_i - z_{1-\alpha/2}\sqrt{[I_\eta^{-1}]_{jj}}}, e^{\hat{\eta}_i + z_{1-\alpha/2}\sqrt{[I_\eta^{-1}]_{jj}}}\right).$$

Example: Moped dataset

- **Goal:** Test our built GLM for the moped dataset.

Example: Moped dataset

- **Goal:** Test our built GLM for the moped dataset.
- **Tasks:**
  1. Test the frequency model using deviance test using its p-value.
  2. Test the severity model using deviance test using its p-value.
  3. Test the frequency model using Pearson's chi-squared test using its p-value.
  4. Test the severity model using Pearson's chi-squared test using its p-value.
  5. Estimate ϕ for the severity model.
  6. Combine zone 5, 6, 7 into zone 4. Fit a new frequency model with the new zone 4.
  7. Test the new frequency model against the previous one.
  8. Repeat step 6 and 7 for the severity model.
  9. Find the 95% confidence intervals for the parameters and the relativities of the frequency model.
  10. Repeat step 9 for the severity model.