STAT347: Generalized Linear Models
Lecture 11

Winter, 2024
Jingshu Wang

Today’s topics:

- Quasi-likelihood
- Estimating equations and the Sandwich estimator

Quasi-likelihood method

- Using the NB GLM instead of Poisson GLM / Beta-binomial GLM instead of a binomial GLM
  - Replace with a more complicated parametric distribution allowing an extra dispersion parameter in the variance of data
  - Hard to check whether the more complicated parametric distribution is the correct model or not

- We can provide a more general solution: the quasi-likelihood method
  - No parametric distributional assumption needed on the response
  - Only require the correct specification of a mean-variance relationship
  - We do not have a likelihood for the data, but we can still have an estimating equation to estimate the parameters and perform statistical inference (even when the mean-variance relationship is incorrectly specified)

Quasi-likelihood method

Remind the the score equation for the exponential family distributed data is:

$$\frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{\mathrm{Var}(y_i)} \frac{1}{g'(\mu_i)} = 0$$

- These score equations only involve $E(y_i) = \mu_i$ and $\mathrm{Var}(y_i)$.
- Quasi-likelihood: we replace $\mathrm{Var}(y_i)$ by some other mean-variance relationship that we believe can better fit the data.
- Typically, the mean-variance relationship can involves another unknown dispersion parameter.
- Here, we DO NOT assume any other aspects of the distribution of y_i besides mean and variance.

Common forms of mean-variance relationship

- Proportional: $a(\mu_i, \phi) = \phi v^\star(\mu_i)$.
  - counts: assume $a(\mu_i, \phi) = \phi\mu_i$
  - grouped Binary data: $a(\mu_i, \phi) = \phi\mu_i(n_i - \mu_i)/n_i$

- For counts we can also assume $a(\mu_i, \phi) = \mu_i + \phi\mu_i^2$ as in the Negative-Binomial distribution

- For grouped Binary data we can also assume $a(\mu_i, \phi) = \mu_i(n_i - \mu_i)(1 + (n_i - 1)\phi)$ as in the Beta-Binomial distribution

How to estimate with quasi-likelihood

- Plug in the mean-variance relationship into the following "score equation" (we now call it the estimating equation) for β

$$\varphi_{1j}(\beta,\phi) = \frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{a(\mu_i,\phi)} \frac{1}{g'(\mu_i)} = 0$$

  - For proportional mean-variance relationship, ϕ will be canceled
  - For other mean-variance relationship, the estimating equation becomes a function for both β and ϕ
- We need another estimating equation for estimating ϕ
  - Use the following moment condition to build an estimating equation for ϕ :

$$\varphi_2(\beta,\phi) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{a(\mu_i,\phi)} - (n-p) = 0$$

How to estimate with quasi-likelihood

When $a(\mu_i, \phi) = \phi v^\star(\mu_i)$, we can get $\hat{\beta}$ thus $\hat{\mu}_i$ first without knowing ϕ. Then define

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\phi v^\star(\hat{\mu}_i)}$$

We can solve ϕ by solving $X^2 = n - p$ (we use $n - p$ instead of n to correct for the degree of freedom in the estimated $\hat{\mu}_i$), which is

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{v^\star(\hat{\mu}_i)}$$

How to estimate with quasi-likelihood

For other forms of $a(\mu, \phi)$, we need to solve ϕ and β simultaneously from equations

$$\varphi_{1j}(\beta, \phi) = \frac{\partial L}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{a(\mu_i, \phi)} \frac{1}{g'(\mu_i)} = 0 \tag{1}$$

$$\varphi_2(\beta, \phi) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{a(\mu_i, \phi)} - (n - p) = 0 \tag{2}$$

– $\mathbb{E}[\varphi_{1j}(\beta, \phi)] = 0$ and $\mathbb{E}[\varphi_2(\beta, \phi)]/n \to 0$. Solutions $\hat{\beta}$ and $\hat{\phi}$ are called Z-estimators. Under proper regularity conditions, we can show that both $\hat{\beta}$ and $\hat{\phi}$ are consistent.

Properties of the estimates

- The proportional mean-variance relationship is the easiest for the computation of $\hat{\beta}$ as ϕ cancels and does not affect solving the score equations for β.

- $\mathrm{Var}(\hat{\beta})$ is affected by ϕ for any of the above mean-variance relationships.

- Including ϕ helps to get a correct uncertainty quantification of $\hat{\beta}$.

Statistical inference for quasi-likelihood estimator

- How to estimate the variance of $\hat{\beta}$ from the quasi-likelihood equations?

- And what if we do not even know the true form of the mean-variance relationship?

Estimating equations

- The equations (2) is one type of estimating equations. In general, the estimating equations for parameters θ (here $\theta = (\beta, \phi)$ or $\theta = \beta$) have the form:

$$u(\theta) = \sum_i u_i(\theta) = 0$$

  Denote the solution of these equations as $\hat{\theta}$ and the true θ as θ_0.

  - Consistency: roughly speaking, when p is small, if $E(u(\theta_0)) \to 0$ when $n \to \infty$, then we can have $\hat{\theta} \to \theta_0$ (with some additional conditions).
  - Variance of $\hat{\theta}$. Under consistency, we can estimate the asymptotic variance of $\hat{\theta}$ by first-order Taylor expansion (see later).

Estimating equations

- The score equations

$$u(\beta) = \sum_i \frac{(y_i - \mu_i)x_{ij}}{v^\star(\mu_i)} \frac{1}{g'(\mu_i)} = 0$$

are valid estimating equations ($\mathbb{E}[u(\beta_0)] = 0$) as long as as the link function is correct. The response y_i does not need to follow the assumed exponential family distribution and $v^\star(\mu_i)$ does not need to be the correct form of variance.

- Even the simple $\sum_i (y_i - \mu_i)x_{ij} = 0$ are always valid estimating equations. The problem is that $\mathrm{sd}(\hat{\beta})$ may be large if samples have unequal variances.

Sandwich estimator

Let's now calculate the asymptotic variance of $\hat{\theta}$ for

$$\mu(\hat{\theta}) = 0$$

By first-order Taylor expansion, we have

$$0 = u(\hat{\theta}) \approx u(\theta_0) + \dot{u}(\theta_0)(\hat{\theta} - \theta_0)$$

Thus, we have

$$\hat{\theta} - \theta_0 \approx -\dot{u}(\theta_0)^{-1} u(\theta_0)$$

Roughly speaking, we have

- Law of large numbers:

$$\frac{1}{n}\dot{u}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\dot{u}_i(\theta_0) \to E\left(\frac{1}{n}\sum_{i=1}^{n}\dot{u}_i(\theta_0)\right) = A$$

- CLT:

$$\frac{1}{\sqrt{n}}u(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_i(\theta_0) \approx N(0, V)$$

Thus

$$\text{Var}(\hat{\theta}) \approx A^{-1}VA^{-T}/n$$

In practice, we can estimate A and V by

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^{n}\dot{u}_i(\hat{\theta})$$

and

$$\widehat{V} = \frac{1}{n}\sum_{i}u_i(\hat{\theta})u_i(\hat{\theta})^T$$

Different from before when we work on the score equations (more parametric-free)

Comments

$$\text{Var}(\hat{\theta}) \approx A^{-1} V A^{-T}/n$$

In practice, we can estimate A and V by

$$\widehat{A} = \frac{1}{n}\sum_{i=1}^{n} \dot{u}_i(\hat{\theta})$$

and

$$\widehat{V} = \frac{1}{n}\sum_{i} u_i(\hat{\theta}) u_i(\hat{\theta})^T$$

- We use the sample variance to approximate V without knowing the distribution of the data
- The Sandwich estimator provides an estimate of the variance of $\hat{\beta}$ even when model assumption is violated.

Revisit the horseshoe crab data

- Check Example7 R notebook