Lecture 12. Quasi-likelihood

Nan Ye

School of Mathematics and Physics
University of Queensland

Looking Back: Course Overview

Generalized linear models (GLMs)

- Building blocks

  systematic and random components, exponential familes
- Prediction and parameter estimation
- Specific models for different types of data

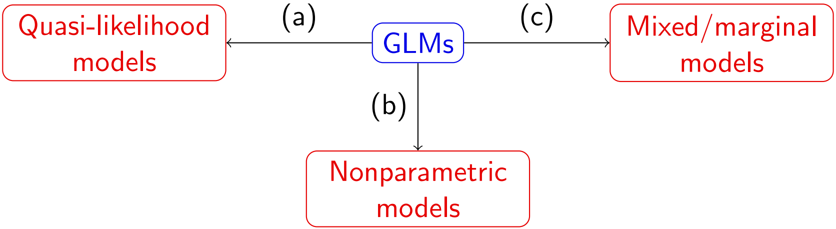
  continuous response, binary response, count response...
- Modelling process and model diagnostics

Extensions of GLMs

- Quasi-likelihood models
- Nonparametric models
- Mixed models and marginal models

Time series

Extending GLMs



(a) Relax assumption on the random component.
(b) Relax assumption on the systematic component.
(c) Relax assumption on the data (independence).

Recall

Gamma regression

- When Y is a non-negative continuous random variable, we can choose the systematic and random components as follows.

$$\text{(systematic)} \quad \mathbb{E}(Y \mid \mathbf{x}) = \exp(\beta^\top \mathbf{x})$$
$$\text{(random)} \quad Y \mid \mathbf{x} \text{ is Gamma distributed.}$$

- We further assume the variance of the Gamma distribution is μ^2/ν (ν treated as known), thus

$$Y \mid \mathbf{x} \sim \Gamma(\mu = \exp(\beta^\top \mathbf{x}), \text{var} = \mu^2/\nu),$$

where $\Gamma(\mu = a, \text{var} = b)$ denotes a Gamma distribution with mean a and variance b.

We have seen how to estimate β for Gamma regression. How do we estimate the dispersion parameter $\phi = 1/\nu$?

Poisson regression

- Poisson regression requires data variance to be the same as mean, but this is seldom the case in real data.
- Overdispersion: variance in data is larger than expected based on the model.
- Underdisperson: variance in data is smaller than expected based on the model.
- For count data, we used quasi Poisson regression to allow both overdisperson and underdispersion.

  How is the quasi-Poisson model defined? How are the parameters estimated?

This Lecture

- Estimation of dispersion parameter
- Quasi-likelihood: derivation and parameter estimation

Estimation of Dispersion Parameter

Recall: Fisher scoring for Gamma regression

- Consider the Gamma regression model

$$Y \mid \mathbf{x} \sim \Gamma(\mu = \exp(\beta^\top \mathbf{x}), \mathrm{var} = \mu^2/\nu),$$

- Let $\mu_i = \exp(\mathbf{x}_i^\top \beta)$, then gradient and Fisher information are

$$\nabla \ell(\beta) = \sum_i \frac{\nu(y_i - \mu_i)}{\mu_i} \mathbf{x}_i, \qquad I(\beta) = \sum_i \nu \mathbf{x}_i^\top \mathbf{x}_i,$$

- Fisher scoring updates β to

$$\beta' = \beta + I(\beta)^{-1} \nabla \ell(\beta).$$

Update of β does not depend on the dispersion parameter $\phi = 1/\nu$!

Moment estimator for the dispersion parameter

- We first estimate β with Fisher scoring.
- Recall: if a GLM model with $\text{var}(Y) = \phi V(\mu)$ is correct, then

$$\frac{X^2}{\phi} = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\phi V(\hat{\mu}_i)} \sim \chi^2_{n-p}$$

  where X^2 is the generalized Pearson statistic, n is the number of examples, and p is the number of parameters in β.
- That is, we have $\mathbb{E}(X^2/\phi) = n - p$.
- The gives us the moment estimator

$$\hat{\phi} = \frac{X^2}{n-p} = \frac{1}{n-p} \sum_i \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

The formula can be used for any GLM with unknown ϕ!

Example

For Gamma regression, $\text{var}(Y) = \phi\mu^2$, so $V(\mu) = \mu^2$.

```
> fit.gam.inv = glm(time ~ lot * log(conc), data=clot,
 family=Gamma)
(Dispersion parameter for Gamma family taken to be 0.002129707)
> mu = predict(fit.gam.inv, type='response')
> sum((fit.gam.inv$y - mu)**2 / mu**2) / (length(mu) -
 length(coef(fit.gam.inv)))
[1] 0.002129692
```

Our estimate is consistent with the `summary` function.

Quasi-Likelihood

Recall: Fisher scoring for GLM

- Let $\mu_i = \mathbb{E}(Y_i \mid \mathbf{x}_i, \beta) = g(\mathbf{x}_i^\top \beta)$ and $V_i = \mathrm{var}(Y_i \mid \mathbf{x}_i, \beta)$.
- The gradient, or score function, is

$$\nabla \ell(\beta) = \sum_i \frac{y_i - \mu_i}{g'(\mu_i) V_i} \mathbf{x}_i.$$

- The Fisher information is

$$I(\beta) = \sum_i \frac{1}{g'(\mu_i)^2 V_i} \mathbf{x}_i \mathbf{x}_i^\top.$$

- Fisher scoring updates β to

$$\beta' = \beta + I^{-1}(\beta) \nabla \ell(\beta).$$

- Fisher scoring for GLM can thus be written as

$$\beta' = \beta + \left(\sum_i \frac{1}{g'(\mu_i)^2 V_i} \mathbf{x}_i \mathbf{x}_i^\top \right)^{-1} \left(\sum_i \frac{y_i - \mu_i}{g'(\mu_i) V_i} \mathbf{x}_i \right).$$

- We just need to know the link function g and the variances V_i's.
- In particular, if we know $V_i = \phi V(\mu_i)$, then the update does not depend on ϕ.
- Thus we can determine β even if ϕ is unknown.

Quasi-model via Fisher scoring

- A GLM has the following structure

$$\begin{aligned}\text{(systematic)} \quad & \mu = \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}), \\ \text{(random)} \quad & Y \mid \mathbf{x} \text{ follows an exponential family distribution.}\end{aligned}$$

- A *quasi-model* relaxes the assumption on the random component

$$\begin{aligned}\text{(systematic)} \quad & \mu = \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}), \\ \text{(random)} \quad & \operatorname{var}(Y \mid \mathbf{x}) = \phi V(\mu),\end{aligned}$$

where ϕ is a dispersion parameter, $V(\mu)$ is a variance function, and β is determined using Fisher scoring!

Hi, I'm Quasimodo.

Quasi-model via quasi-likelihood

- A quasi-model relaxes the assumption on the random component

$$\begin{aligned} \text{(systematic)} \quad & \mu = \mathbb{E}(Y \mid \mathbf{x}) = h(\beta^\top \mathbf{x}), \\ \text{(random)} \quad & \mathrm{var}(Y \mid \mathbf{x}) = \phi V(\mu), \end{aligned}$$

  where ϕ is a dispersion parameter, $V(\mu)$ is a variance function, and β is determined by maximizing quasi-likelihood!
- Quasi-likelihood is a surrogate for the log-likelihood of the mean parameter μ given an observation y, when we only know $\mathrm{var}(Y \mid \mathbf{x}) = \phi V(\mu)$.

Construction of quasi-likelihood

- Recall: a score function $\ell(\mu)$ satisfies

$$\mathbb{E}(\ell) = 0,$$
$$\mathrm{var}(\ell) = -\mathbb{E}(\ell').$$

- Define $S(\mu) = \frac{Y-\mu}{\phi V(\mu)}$, then $S(\mu)$ is similar to a score function:

$$\mathbb{E}(S) = 0,$$
$$\mathrm{var}(S) = -\mathbb{E}\, S' = \frac{1}{\phi V(\mu)}.$$

- $S(\mu)$ is thus called a quasi-score function.

- The usual log-likelihood is an integral of the score function.
- By analogy, the quasi-likelihood (quasi log-likelihood) is

$$Q(\mu; y) = \int_y^\mu \frac{y - t}{\phi V(t)} dt.$$

Quasi-likelihood for some variance functions

$V(\mu)$	$Q(\mu; y)$	distribution	constraint
1	$-(y-\mu)^2/2$	normal	-
μ	$y \ln \mu - \mu$	Poisson	$\mu > 0, y \geq 0$
μ^2	$-y/\mu - \ln \mu$	Gamma	$\mu > 0, y \geq 0$
μ^3	$-y/(2\mu^2) + 1/\mu$	inverse Gaussian	$\mu > 0, y \geq 0$
μ^m	$\mu^{-m}\left(\frac{\mu y}{1-m} - \frac{\mu^2}{2-m}\right)$	-	$\mu > 0, m \neq 0, 1, 2$
$\mu(1-\mu)$	$y \ln \frac{\mu}{1-\mu} + \ln(1-\mu)$	binomial	$\mu \in (0,1), 0 \leq y \leq 1$
$\mu^2(1-\mu^2)$	$(2y-1) \ln \frac{\mu}{1-\mu} - \frac{y}{\mu} - \frac{1-y}{1-\mu}$	-	$\mu \in (0,1), 0 \leq y \leq 1$
$\mu + \mu^2/k$	$y \ln \frac{\mu}{k+\mu} + k \ln \frac{k}{k+\mu}$	negative binomial	$\mu > 0, y \geq 0$

Parameter estimation for quasi-model

- In a quasi-model, μ is a function of β, and the quasi-likelihood is also a function of β

$$Q(\beta) = \sum_i Q(\mu_i(\beta); y_i)$$

- The Fisher scoring update for Q is given by

$$\begin{aligned}\beta' &= \beta + (-\mathbb{E}\,\nabla^2 Q(\beta))^{-1}\,\nabla Q(\beta)\\ &= \beta + \left(\sum_i \frac{1}{g'(\mu_i)^2 \phi V(\mu_i)} \mathbf{x}_i \mathbf{x}_i^\top\right)^{-1} \left(\sum_i \frac{y_i - \mu_i}{g'(\mu_i) \phi V(\mu_i)} \mathbf{x}_i\right).\end{aligned}$$

  The update is independent of ϕ.
- ϕ is estimated as $\hat{\phi} = \frac{X^2}{n-p}$ after β is estimated.

Recall: quasi-Poisson regression

- Quasi-Poisson regression model introduces an additional dispersion paramemeter ϕ.
- It replaces the original model variance V_i on $\mathbf{x}_i$ by ϕV_i.
- $\phi > 1$ is used to accommodate overdispersion relative to the original model.
- $\phi < 1$ is used to accommodate underdispersion relative to the original model.
- ϕ is usually estimated separately after estimating β.

Estimating ϕ in quasi-Poisson regression

```
> fit.qpo <- glm(Days ~ Sex + Age + Eth + Lrn, data=quine,
 family=quasipoisson)
(Dispersion parameter for quasipoisson family taken to be
 13.16691)
> mu = predict(fit.qpo, type='response')
> sum((fit.qpo$y - mu)**2 / mu) / (length(mu) -
 length(coef(fit.qpo)))
[1] 13.16684
```

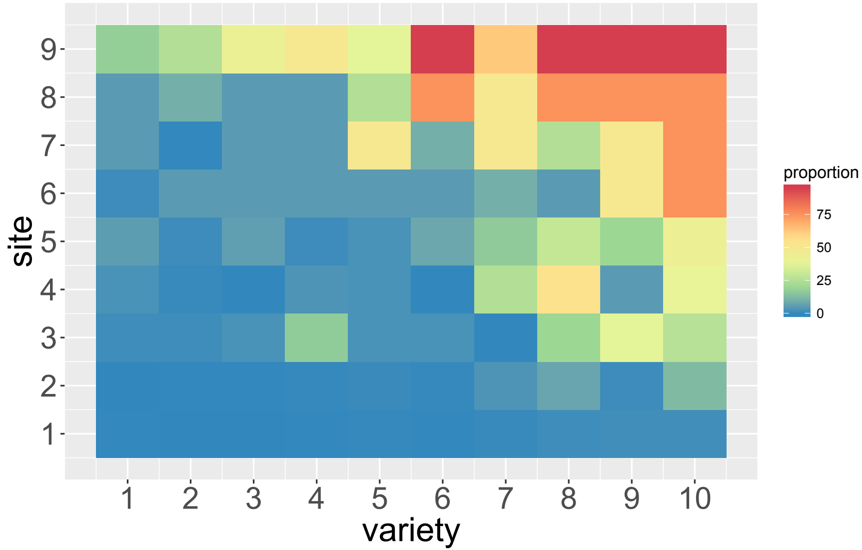
Example

Data

Site	Variety										Mean
	1	2	3	4	5	6	7	8	9	10	
1	0.05	0.00	0.00	0.10	0.25	0.05	0.50	1.30	1.50	1.50	0.52
2	0.00	0.05	0.05	0.30	0.75	0.30	3.00	7.50	1.00	12.70	2.56
3	1.25	1.25	2.50	16.60	2.50	2.50	0.00	20.00	37.50	26.25	11.03
4	2.50	0.50	0.01	3.00	2.50	0.01	25.00	55.00	5.00	40.00	13.35
5	5.50	1.00	6.00	1.10	2.50	8.00	16.50	29.50	20.00	43.50	13.36
6	1.00	5.00	5.00	5.00	5.00	5.00	10.00	5.00	50.00	75.00	16.60
7	5.00	0.10	5.00	5.00	50.00	10.00	50.00	25.00	50.00	75.00	27.51
8	5.00	10.00	5.00	5.00	25.00	75.00	50.00	75.00	75.00	75.00	40.00
9	17.50	25.00	42.50	50.00	37.50	95.00	62.50	95.00	95.00	95.00	61.50
Mean	4.20	4.77	7.34	9.57	14.00	21.76	24.17	34.81	37.22	49.33	20.72

- Incidence of leaf blotch on 10 varieties of barley grown at 9 sites.
- The response is the percentage leaf area affected.

Heatmap for the data



```
> fit.qbin = glm(proportions ~ as.factor(site) +
  as.factor(variety), family = quasibinomial)
```

- A binomial model satisfies $\text{var}(Y) = \mu(1-\mu)$.
- A quasibinomial model assumes that $\text{var}(Y) = \phi\mu(1-\mu)$, where ϕ is the dispersion parameter.
- The probability of having leaf blotch for variety j at site i has the form

$$p_{ij} = \frac{\exp(b + \alpha_i + \beta_j)}{1 + \exp(b + \alpha_i + \beta_j)}$$

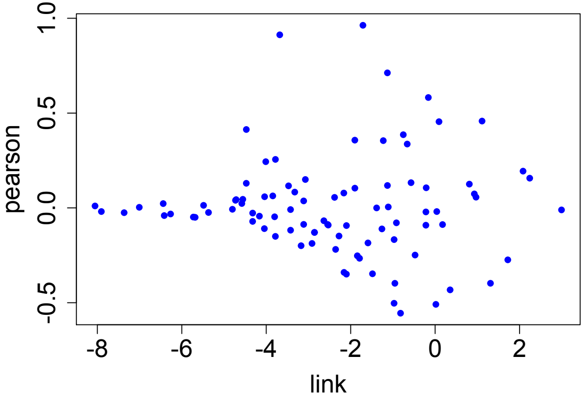
```
> summary(fit.qbin)
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          -8.0546     1.4219  -5.665 2.84e-07 ***
as.factor(site)2      1.6391     1.4433   1.136 0.259870
as.factor(site)3      3.3265     1.3492   2.466 0.016066 *
as.factor(site)4      3.5822     1.3444   2.664 0.009510 **
as.factor(site)5      3.5831     1.3444   2.665 0.009493 **
as.factor(site)6      3.8933     1.3402   2.905 0.004875 **
as.factor(site)7      4.7300     1.3348   3.544 0.000697 ***
as.factor(site)8      5.5227     1.3346   4.138 9.38e-05 ***
as.factor(site)9      6.7946     1.3407   5.068 3.00e-06 ***
```

```
as.factor(variety)2    0.1501     0.7237   0.207 0.836289
as.factor(variety)3    0.6895     0.6724   1.025 0.308587
as.factor(variety)4    1.0482     0.6494   1.614 0.110910
as.factor(variety)5    1.6147     0.6257   2.581 0.011895 *
as.factor(variety)6    2.3712     0.6090   3.893 0.000219 ***
as.factor(variety)7    2.5705     0.6065   4.238 6.58e-05 ***
as.factor(variety)8    3.3420     0.6015   5.556 4.39e-07 ***
as.factor(variety)9    3.5000     0.6013   5.820 1.51e-07 ***
as.factor(variety)10   4.2530     0.6042   7.039 9.38e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for quasibinomial family taken to be 0.08877)
```

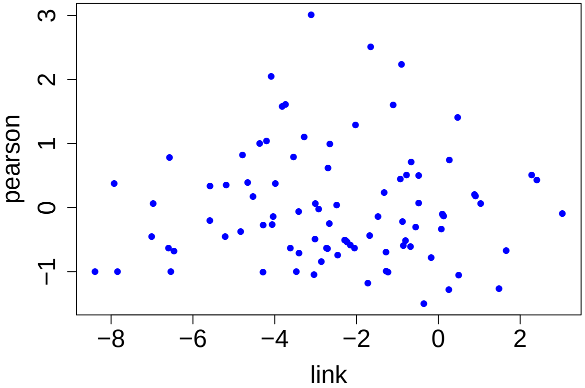
- We can see that both α_i and β_j are increasing as i, j increase.
- This is consistent with the trend in data.

Pearson residual plot



- The residuals are more or less symmetrically distributed around 0.
- Thus the mean function appears to be a good fit.
- However, the residuals are very close to 0 at both ends, and this suggests that the variance function is not good.

Pearson residual plot with $V(\mu) = \mu^2(1-\mu)^2$



- The residual plot is better than that with $V(\mu) = \mu(1-\mu)$.
- The variance function $V(\mu) = \mu^2(1-\mu)^2$ better fits the data than $V(\mu) = \mu(1-\mu)$.

What You Need to Know

- Moment estimator of the dispersion parameter: $\hat{\phi} = X^2/(n - p)$.
- Quasi-likelihood
  - Derivation
  - Estimation of β using Fisher scoring
  - Estimation of ϕ using moment matching