Quasi-likelihood Estimation

Li, Hyland, Yabor, Gueli, Yao

May 2, 2021

Sections

1. From Likelihood to Quasi-likelihood by Bridget Hyland
2. Derivation by Kai Li
3. Properties of the QL Estimator and Variance by Peng Fei Yao
4. QL for Poisson and Binomial GLMs by Chad Gueli
5. Examples by Vinny Yabor

Notation

As in the book:

- $Y \in \mathbb{R}^n$ is a random vector with independent components from identical distributions in the natural exponential family.
- $\boldsymbol{y} \in \mathbb{R}^n$ is the vector of realized values of Y with explanatory variables $X \in \mathbb{R}^{n \times p}$
- $E[Y] = \boldsymbol{\mu} \in \mathbb{R}^n$
- We assume that Y is linked to X by g and linear systematic component $\boldsymbol{\eta} = X\boldsymbol{\beta} \in \mathbb{R}^n$, where $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of coefficients.

Then the GLM is $g(\boldsymbol{\mu}) = \boldsymbol{\eta} = X\boldsymbol{\beta}$.

Back to Basics, Just ML

To obtain maximum likelihood estimates:

1. Assume a specific distribution for Y
2. Use distribution to determine $\text{Var}(Y) \in \mathbb{R}^{n \times n}$
   - $\text{Var}(Y)$ is a diagonal matrix with the variance for the ith observation as the ith component
3. The ML estimate $\hat{\beta}$ will satisfy

$$\boldsymbol{u}(\boldsymbol{\beta}) = X^T \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \text{Var}(Y)^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = 0. \tag{1}$$

- $\partial \boldsymbol{\mu} / \partial \boldsymbol{\eta} \in \mathbb{R}^{n \times n}$ is the Jacobian of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\eta}$, and is diagonal because $\partial \mu_i / \partial \eta_j = 0$ for $i \neq j$.
- $\boldsymbol{u}$ is called the score function and is dependent on $\boldsymbol{\beta}$ through $\boldsymbol{\eta}$.

From ML to QL

- ML estimate depends on distribution only through mean and variance.
- Instead, we can assume that there exists a distribution in the exponential family with mean μ and variance $v(\mu)$. This is quasi-likelihood estimation.
- Under this assumption, the QL estimate is the ML estimate for the unspecified distribution.

Likelihood I

As before Y_i are iid from an exponential dispersion distribution; i.e.

$$f(y_i \mid \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right). \tag{2}$$

Denote

$$L_i = \log f(y_i \mid \theta_i, \phi) = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi), \tag{3}$$

and get

$$\frac{\partial L_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} \qquad \frac{\partial^2 L_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{a(\phi)}. \tag{4}$$

Likelihood II

Fisher Information Theorem 1: For a pdf f dependent on θ,

$$E\left(\frac{\partial \log f_\theta(x)}{\partial \theta}\right) = \int \frac{\partial \log f_\theta(x)}{\partial \theta} f_\theta(x)\, dx = 0. \tag{5}$$

By this theorem,

$$0 = E\left(\frac{\partial L_i}{\partial \theta_i}\right) = E\left(\frac{Y_i - b'(\theta_i)}{a(\phi)}\right) = \frac{EY_i - b'(\theta_i)}{a(\phi)} = \frac{\mu_i - b'(\theta_i)}{a(\phi)}, \tag{6}$$

implying $\mu_i = b'(\theta_i)$ and

$$\frac{\partial \mu_i}{\partial \theta_i} = b''(\theta_i). \tag{7}$$

Likelihood III

Fisher Information Theorem 2: For a pdf f dependent on θ,

$$-E\left(\frac{\partial^2 \log f_\theta(x)}{\partial \theta^2}\right) = E\left(\left[\frac{\partial \log f_\theta(x)}{\partial \theta}\right]^2\right). \tag{8}$$

Due to this theorem, and because $\mu_i = b'(\theta_i)$

$$\frac{b''(\theta_i)}{a(\phi)} = -E\left(\frac{\partial^2 L_i}{\partial \theta_i^2}\right) = E\left(\left[\frac{\partial L_i}{\partial \theta_i}\right]^2\right) \tag{9}$$

$$= E\left(\left[\frac{Y_i - b'(\theta_i)}{a(\phi)}\right]^2\right) \tag{10}$$

$$= \frac{E(Y_i - \mu_i)^2}{a(\phi)^2} = \frac{\mathrm{Var}(Y_i)}{a(\phi)^2}, \tag{11}$$

therefore $b''(\theta_i) = \mathrm{Var}(Y_i)/a(\phi)$.

Likelihood IV

Put the preceding together to get

$$\frac{\mathsf{Var}(Y_i)}{a(\phi)} = b''(\theta_i) = \frac{\partial \mu_i}{\partial \theta_i} \tag{12}$$

or equivalently $\partial\theta_i/\partial\mu_i = \mathsf{Var}(Y_i)/a(\phi)$. Then

$$\mathcal{L} = \log\left(\prod_i f(y_i \mid \theta_i, \phi)\right) = \sum_i L_i \tag{13}$$

gives

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_i \frac{\partial L_i}{\partial \beta_j} = \sum_i \frac{\partial L_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \sum_i \frac{Y_i - \mu_i}{\mathrm{Var}(Y_i)}\frac{\partial \mu_i}{\partial \eta_i} x_{ij}. \tag{14}$$

Quasi-score Function

Noting that $\boldsymbol{\eta} = X\boldsymbol{\beta}$, we have

$$X^T \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} = \left(\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \right)^T X \right)^T = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} X \right)^T = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\beta}} \right)^T = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right)^T. \tag{15}$$

Using this substitution and $\mathrm{Var}(Y) = v(\boldsymbol{\mu})$ yields the quasi score function

$$\boldsymbol{u}(\beta) = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right)^T v(\boldsymbol{\mu})^{-1}(\boldsymbol{y} - \boldsymbol{\mu}). \tag{16}$$

QL Estimator Properties

- The quasi-score function is an unbiased estimating function
- If the mean and variance function are specified correctly, then $\hat{\beta}$ is
  - asymptotically efficient among estimators that are locally linear in $\boldsymbol{y}$
  - asymptotically normal with variance

$$V = \left[\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right)^T v(\boldsymbol{\mu})^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right]^{-1}, \tag{17}$$

  where $\partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$ is the Jacobian of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\beta}$.
- The QL estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$ even if the variance function is misspecified, as long as the specification is correct for the link function and linear predator.

Sandwich Covariance Adjustment

For misspecified variance functions, the Sandwich Adjustment

$$V\left[\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right)^T v(\boldsymbol{\mu})^{-1}\mathrm{Var}(Y)v(\boldsymbol{\mu})^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right]V \tag{18}$$

may be used to approximate the true variance. The Wald approximation with $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$ and $\widehat{\mathrm{Var}}(Y) = \mathrm{Diag}[\boldsymbol{y} - \hat{\boldsymbol{\mu}}]^2$ has been shown to be asymptotically accurate.
Assuming $\mathrm{Var}(Y) = v(\boldsymbol{\mu})$ gives

$$V\left[\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right)^T v(\boldsymbol{\mu})^{-1}v(\boldsymbol{\mu})v(\boldsymbol{\mu})^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}\right]V = VV^{-1}V = V \tag{19}$$

Overdispersion in Poisson GLMs

- For count data with unfixed margins, we can use quasi-likelihood with $v(\boldsymbol{\mu}) = \text{Diag}[\phi\boldsymbol{\mu}]$
  - $\phi = 1$ for regular Poisson distribution
  - $\phi > 1$ for over-dispersed data
- The estimated variance for $\hat{\beta}$ is

$$\left[\left(\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\beta}}\right)^T \text{Diag}[\phi\boldsymbol{\mu}]^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\beta}}\right]^{-1} = \phi\left[\left(\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\beta}}\right)^T \text{Diag}[\boldsymbol{\mu}]^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\boldsymbol{\beta}}\right]^{-1} \quad (20)$$

which is ϕ times the variance of the coefficients for the regular Poisson.

Estimating ϕ

Let X^2 be the Pearson Chi-squared statistic for the regular Poisson model, and $X^2/\phi = \sum_i (y_i - \hat{\mu}_i)^2/\phi\hat{\mu}_i$ be the X^2 stat for the generalized model. If one of the following two requirements holds

1. X^2/ϕ has an approximate χ^2 distribution
2. $\boldsymbol{\mu}$ is approximately linear in $\boldsymbol{\beta}$ with $v(\hat{\boldsymbol{\mu}})$ close to $v(\boldsymbol{\mu})$

then $E(X^2/\phi) \approx n - p$, implying that $E(X^2/[n-p]) \approx \phi$.

$$\hat{\phi} \approx X^2/(n-p) \tag{21}$$

Overdispersion in Binomial GLMs

Overdispersion in the binomial case works the same as in the Poisson case.

- $v(\pi_i) = \phi\pi_i(1-\pi_i)/n_i = \phi\text{Var}(Y_i)$
- Estimate ϕ by $\hat{\phi} = X^2/(n-p)$
- Estimate the variance of $\hat{\beta}$ by multiplying the variance of the standard model by $\hat{\phi}$

Horse Shoe Crab Dispersion Test

```
glm1 <- glm(sat ~ width, family=poisson, data=crabs)
check_overdispersion(glm1)

## # Overdispersion test
##
##        dispersion ratio =   3.182
##   Pearson's Chi-Squared = 544.157
##                 p-value = < 0.001
```

The `check_overdispersion` function from the `performance` package may be used to test for overdispersion.

Computing SE

```
sa <- tapply(crabs$sat, crabs$width, sum)
mu <- tapply(predict(glm1, type="response"),
             crabs$width, sum)
(chi_square <- sum((sa-mu)^2 / mu))
(phi <- chi_square/(66-2))
(SE_quasi <- sqrt(phi)*summary(glm1)$coefficients[2, 2])
```

```
## [1] 174
## [1] 2.72
## [1] 0.0329
```

QL Poisson Regression

```
glm1_quasi <- glm(sat ~ width, data=crabs,
                  family=quasipoisson)

confint(glm1)
confint(glm1_quasi)

##               2.5 % 97.5 %
## (Intercept) -4.366 -2.241
## width        0.125  0.203
##                2.5 % 97.5 %
## (Intercept) -5.1965 -1.405
## width        0.0936  0.233
```

Teratology Example

```
glm2 <- glm(cbind(R, N-R) ~ -1+grp, data=ter,
            family=binomial)

(pred <- unique(predict(glm2, type="response")))
(SE <- sqrt(pred*(1-pred)/tapply(ter$N, ter$grp, sum)))
(phi <- sum(resid(glm2, type="pearson")^2)/(58-4))
(SE_adj <- sqrt(phi)*SE)

## [1] 0.7584 0.1017 0.0345 0.0481
##      1      2      3      4
## 0.0237 0.0278 0.0240 0.0210
## [1] 2.86
##      1      2      3      4
## 0.0401 0.0471 0.0406 0.0355
```

QL For Binomial Model

```
glm2_quasi <- glm(cbind(R, N-R) ~ -1+grp, data=ter,
                  family=quasibinomial)

confint(glm2)
confint(glm2_quasi)

##        2.5 % 97.5 %
## grp1  0.896   1.40
## grp2 -2.829  -1.62
## grp3 -5.141  -2.17
## grp4 -4.027  -2.19
##       2.5 % 97.5 %
## grp1  0.73   1.59
## grp2 -3.35  -1.28
## grp3 -7.08  -1.56
## grp4 -4.97  -1.74
```