Quasilikelihood and overdispersion

Quasilikelihood

Recall: IWLS only requires X, g (the link function), and v (the variance function) to fit a glm. X and g describe the mean and v describes the variance.

Once fitted, we also need an estimate of $a(\phi) = \phi$ in order to estimate $\operatorname{Var}\hat{\boldsymbol{\beta}} = (X^T\hat{\Sigma}^{-1}X)^{-1}$, namely

$$\hat{\phi} = \frac{X^2}{n-p} = \frac{1}{n-p}\sum_i \frac{(Y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)} \quad (*).$$

Consequently, we can consider the following generalisation of the glm:

$$\begin{aligned} g(\mu_i) &= \mathbf{x}_i^T\boldsymbol{\beta} \\ \mathbb{E}Y_i &= \mu_i \\ \operatorname{Var}Y_i &= \phi v(\mu_i). \end{aligned}$$

That is, we do not need to restrict ourselves to exponential families. We can use IWLS to fit the model, and estimate ϕ using $(*)$ as before.

Quasilikelihood

The problem is, if we do not have that Y_i is an exponential family, then we do not know that IWLS is equivalent to maximum likelihood, so we do not have that $\hat{\beta}$ is asymptotically normal and all the other benefits of MLE.

Quasilikelihood is a general method for model fitting and inference that works when we do not have a complete likelihood. As such it is more widely applicable than maximum likelihood, however in general it provides less efficient estimators (larger variance than MLE).

For the quasilikelihood, we only need the mean μ and variance $\phi v(\mu)$.

Quasilikelihood

If $\mathbb{E}\, Y_i = \mu_i$ and $\mathrm{Var}\, Y_i = \phi v(\mu_i)$, define the score

$$u_i = \frac{Y_i - \mu_i}{\phi v(\mu_i)}.$$

We have

$$\mathbb{E}\, u_i = 0, \quad \mathrm{Var}\, u_i = \frac{1}{\phi v(\mu_i)},$$

$$\begin{aligned} -\mathbb{E}\frac{\partial u_i}{\partial \mu_i} &= -\mathbb{E}\frac{-\phi v(\mu_i) - (Y_i - \mu_i)\phi v'(\mu_i)}{(\phi v(\mu_i))^2} \\ &= \frac{1}{\phi v(\mu_i)} = \mathbb{E}\, u_i^2 \end{aligned}$$

These properties are all shared by $\frac{\partial \log L}{\partial \mu}$ and it is from these that many of the properties of MLE are derived. This suggests that we can use $\int u d\mu$ like a log-likelihood.

Quasilikelihood

Define

$$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi v(t)} dt \quad (\leq 0),$$

then $\frac{\partial Q_i}{\partial \mu_i} = u_i$ and the quasi log-likelihood is

$$Q = \sum_{i=1}^{n} Q_i.$$

It can be shown that maximizing the quasi (log) likelihood gives a consistent estimator for β. We estimate ϕ using $X^2/(n-p)$ as before.

Quasideviance

We can also form a quasideviance. The quasi (log) likelihood for the saturated model is clearly 0, so we get the scaled quasideviance:

$$\begin{aligned} \frac{D_Q}{\phi} &= -2\left[\sum Q_i - \sum Q_i^s\right] \\ &= -2\sum Q_i \end{aligned}$$

and the quasideviance is

$$D_Q = -2\sum \int_{y_i}^{\mu_i} \frac{y_i - t}{v(t)} dt.$$

Quasideviance

The difference of two scaled quasideviances, for nested models, approximately follows χ^2_s, where s is the difference in the number of parameters, and

$$
\begin{aligned}
F_Q &= \frac{\left[D_Q(\text{reduced model}) - D_Q(\text{full model})\right]/s}{\hat{\phi}} \\
&\approx F_{s,n-p},
\end{aligned}
$$

where $n - p$ is the degrees of freedom for $\hat{\phi}$.

Quasi Binomial and Quasi Poisson

Our main application of quasilikelihood theory is to generalise the binomial and Poisson regression models to allow for overdispersion. That is, count data where $\mathrm{Var}\, Y_i$ is larger than it should be, according to a binomial or Poisson model.

For a binomial observation, if there is dependence between the trials, or if the success probability changes from trial to trial, then we can get overdispersion. Similarly, for a Poisson process dependence between events or a non-constant rate can produce overdispersion.

Quasi Binomial

$$\mathrm{Var}\, Y = \phi m p(1-p) = \phi \mu(m-\mu)/m.$$

So $v(\mu) = \mu(m-\mu)/m$ just as for the binomial. The difference is that we no longer fix $\phi = 1$.

$$Q_i = \frac{1}{\phi}\left[y_i \log \frac{\mu_i}{1-\mu_i} + \log(1-\mu_i) \right]$$

which is the same as the binomial log likelihood, except for the factor $\frac{1}{\phi}$.

Thus, IWLS will give that $\boldsymbol{\beta}$ which maximises the quasi (log) likelihood. Because IWLS does not depend on ϕ, we can fit a quasi binomial by pretending it is a regular binomial regression model.

Quasi Poisson

$$\mathrm{Var}\, Y = \phi\mu.$$

So $v(\mu) = \mu$ just as for the Poisson, but now we no longer require $\phi = 1$.

$$Q_i = \frac{1}{\phi}\left[y_i \log \mu_i - \mu_i\right]$$

which is the same as the Poisson log likelihood, except for the factor $\frac{1}{\phi}$.

So, as for the quasi binomial, we can fit a quasi Poisson using IWLS by just pretending it is a regular Poisson regression model.

Quasi Binomial and Quasi Poisson

It also follows that for the quasi binomial and quasi Poisson, the deviance is the same as for the binomial and Poisson models applied to the same data. However, for the quasi models, the scaled deviance and the deviance are different.

Thus, for quasi binomial/Poisson models, we can't use the deviance directly to test for model adequacy, and when comparing models we can't just compare deviance and use a χ^2 test, instead we have to scale by s (the difference in df) and $\hat{\phi}$, and use an F test. In R, this can be done using the anova command (drop1 and step do not work).

Overdispersion

If you ignore overdispersion when fitting a binomial/poisson regression, your estimation is unaffected, but your inference changes. With overdispersion, your F statistic is reduced , making model comparison less significant in general (so you may end up with fewer significant variables in the model).

With overdispersion, our estimate of Σ also increases, so we get larger CI for our parameter estimates.

Example

See troutegg.pdf.