Overdispersion and quasilikelihood

Learning goals

Be able to explain what overdispersion is.

Understand quasilikelihood methods:

- Be able to define a quasi log-likelihood.
- Be able to compute quasi log-likelihood, scaled quasideviance, and quasideviance.
- Be able to perform a test for model selection between nested models
- Be able to estimate parameters and perform inference for quasi binomial/Poisson regressions.

Overdispersion

$Y_1, Y_2 \cdots Y_n$ since count data $\Rightarrow$ $Y_i \sim \text{Poisson}(\lambda)$

$E(Y_i) = \text{var}(Y_i) = \lambda$

based on data $100 = \bar{Y}$, $S^2 = 300$ (in theory should be close to 100)

one way handle overdispersion ① $\Rightarrow$ maybe fit $Y_i \sim NB(r,p)$ $E(Y_i) = \frac{rp}{1-p} = \mu$ $\text{Var}(Y) = \frac{rp}{(1-p)^2} = \frac{\mu}{1-p}$ $\Rightarrow$ overdispersed

$Y_i \sim NB(r,p)$ $\Leftrightarrow$ $Y_i \sim \text{Poisson}(\lambda_i)$ where $\lambda_i \sim \text{Gamma}(k,\theta)$.

Data are overdispersed when the actual $\text{Var}\, Y_i$ is larger than it should be, according to a model for Y_i (e.g., Poisson and binomial distributions).

variance is completed determined by mean

② $Y_i \sim \text{Bin}(m, p_i)$ where $p_i \sim \text{Beta}(\ )$

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.

Quasilikelihood is one way to deal with overdispersion. Our main application of quasilikelihood theory is to generalise the binomial and Poisson regression models to allow for overdispersion.

linear model
$E(Y_i) = x\beta$. $var(Y_i) = \sigma^2$
$g(\mu_i) = \mu_i$ $v(\mu) = 1$

Quasilikelihood

So far we have been fitting models using maximum likelihood. This has meant assuming that there is a probability distribution for the data.

Suppose that we are able to specify the link function (i.e., how the mean response is affected by several predictors) and the variance function (i.e., how the variability of the response changes with the average response) of the data, but that we do not have a strong idea about the appropriate distributional form for the response variable.

The IWLS algorithm requires only the link (g) and variance (v) functions to fit a glm and it assumes:

$$\begin{aligned} g(\mu_i) &= x_i^T \beta \\ \mathbb{E} Y_i &= \mu_i \\ \mathrm{Var}\, Y_i &= \phi v(\mu_i). \end{aligned}$$

The IWLS algorithm does not require distributional assumptions for the response variable.

Quasilikelihood

However, many inference procedures still need a likelihood. Can we propose a suitable substitute for a likelihood that can be computed without assuming a distribution?

Quasilikelihood is a general method for model fitting and inference that works when we do not have a likelihood. As such it is more widely applicable than maximum likelihood, however in general it provides less efficient estimators (larger variance than MLE). So if you have information about the distribution, you are advised to use it.

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

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

$E[\frac{\partial \log L}{\partial \mu}] = 0$

$E[\frac{\partial \log L}{\partial \mu^2}] = -E[(\frac{\partial \log L}{\partial \mu})^2]$

$u_i = \frac{\partial \log L}{\partial \mu_i}$ $\quad E[u_i] = 0$ $\quad E[\frac{\partial u_i}{\partial \mu_i}] = -E[u_i^2]$

$\log L = \int u_i \, \partial \mu_i$

We have

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

since $Eu_i = 0$ $\quad Eu_i^2 = \mathrm{var}\, u_i$

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

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

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

make sure $Q_i < 0$ (since Q_i is a substitude for log-likelihood)

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

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

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

The usual asymptotic properties expected of maximum likelihood estimators also hold for quasi-likelihood-based estimators.

Quasilikelihood is more widely applicable than maximum likelihood, however in general it provides less efficient estimators (larger variance than MLE). So if you have information about the distribution, you are advised to use it.

Quasideviance

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

for saturated model

$\hat{\mu}_i = y_i$

$Q = \int_{y_i}^{y_i} \frac{y_i - t}{\phi v(t)} dt = 0.$

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

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

and the quasideviance is

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

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

Scaled quasideviance for model selection (nested models)

If model A is nested within model B, under the null hypothesis that model A is correct we have

$$
\begin{aligned}
F_Q &= \frac{(D_Q^A - D_Q^B)/s}{\hat{\phi}} \\
&\approx F_{s,n-p},
\end{aligned}
$$

→ # of parameters between two models.

where D_Q^A and D_Q^B denote quasideviances for models A and B, respectively, we have n observations, model A has $p - s$ parameters, and model B has p parameters.

- $\hat{\phi} = X^2/(n-p)$, where X^2 (Pearson's chi-squared) is calculated using model B.

Quasi Poisson

$Y \sim \text{Poisson}(\lambda)$

$E[Y] = \lambda = \mu \quad var(Y) = \lambda = \underbrace{\mu = v(\mu)}_{\text{identity function}}. \quad \phi = 1.$

$var(Y) = v(\mu)\phi = \mu\phi.$

$$\text{Var } Y = \phi\mu.$$

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

$Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi \, \underbrace{v(t)}_{=t}} dt$

$= \frac{1}{\phi} \int_{y_i}^{\mu_i} \frac{y_i}{t} - 1 \, dt$

$= \frac{1}{\phi} [y_i \log t - t]_{y_i}^{\mu_i}$

$= \frac{1}{\phi} [\underbrace{(y_i \log \mu_i - \mu_i)}_{A} - \underbrace{(y_i \log y_i - y_i)}_{B}]$

since $g(\mu_i) = x_i\beta$. β only appear in A, B is constant

ignore constant part.

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

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

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

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

Quasi binomial

Binomial $Y_i \sim Bin(m, p)$ $E(Y) = mp = \mu$

$Var(Y) = mp(1-p) = \mu(1-\frac{\mu}{m})$.

$v(\mu) = \mu(1-\frac{\mu}{m})$

Exercise!

for Quasi binomial

$E(Y) = mp = m$

$var(Y) = v(\mu)\phi = \phi\mu(\frac{m-\mu}{m})$.

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

$= \frac{1}{\phi}\int_{y_i}^{\mu_i} \frac{y_i - t}{t(1-\frac{t}{m})} dt$

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

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

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

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

So, as for the quasi Poisson, we can fit a quasi binomial using IWLS by just pretending it is a regular binomial 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 binomial/Poisson models, the scaled deviance and the deviance are different. (since $\phi \neq 1$)

Thus, for quasi binomial/Poisson models, we can't use the deviance 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 in binomial/poisson regression

you will underestimate CI & false discovery (recognise significant but in truth insignificant).

If you ignore overdispersion when fitting a binomial/poisson regression, your estimation is unaffected, but your inference changes.

$F = \frac{D_\theta^A - D_\theta^B / s}{\phi \uparrow} \downarrow$

F go down → less significant

With proper modelling of 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).
- Your estimate of Σ also increases, so you get larger CI for your parameter estimates.

var ↑

$\Sigma_{ii} = g'(\mu_i)^2 v(\mu_i) \phi > 1.$ (since $\phi \neq 1$)

$var(\hat{\beta}) \uparrow = (X^T \Sigma^{-1} X)^{-1} \uparrow \propto \Sigma \uparrow$

Example

See troutegg.pdf.

Learning goals

Be able to explain what overdispersion is.

Understand quasilikelihood methods:

- Be able to define a quasi log-likelihood.
- Be able to compute quasi log-likelihood, scaled quasideviance, and quasideviance.
- Be able to perform a test for model selection between nested models.
- Be able to estimate parameters and perform inference for quasi binomial/Poisson regressions.