# STAT 770 Oct. 19 Lecture 14
# Quasilikelihood and Overdispersion

Reading and Topics for this lecture: Secs. 4.4.1-4.4.5, '`quasi`' `family` within `glm` in `R` , also Ch 13 through 13.1.2, Sec. 14.3-14.4

**(1)** Idea of Overdispersion – Important!

**(2)** Specific Models for Overdispersion, e.g. Negative Binomial

**(3)** Quasilikelihood estension of GLM

**(4)** `R` Code for Quasilikelihood fits

**(5)** Summarizing Predictive Power of Logistic Regression (Sec. 6.3)

# Extra Heterogeneity among Observations

A key assumption required for validity of GLM methods is homogeneity and independence (*iid individual observations*).

What if observations in groups (i.e., **clusters**) share a random effect, i.e., observations are dependent among themselves ?

Examples:

(a) **Beta-Binomial** $Y_i \sim \text{Binom}(n_i, p_i), \; p_i \sim \text{Beta}(\alpha, \beta)$,

(b) **Negative-Binomial** $Y_i \sim \text{Poisson}(\lambda_i), \; \lambda_i \sim \text{Gamma}(k, k/\mu)$

$$\begin{aligned} \text{Var}(Y_i) &= E(\text{Var}(Y_i \,|\, \lambda_i)) + \text{Var}(E(Y_i \,|\, \lambda_i)) = E(\lambda_i) + \text{Var}(\lambda_i) \\ &= \mu + \mu^2/k \qquad \text{over-dispersion} \end{aligned}$$

# NegBin Model

(material covered in detail in Sec. 14.4)

In example (b): Negative Binomial $\Big(k, \mu/(\mu+k)\Big)$ distribution

$$P(Y_i = m) \;=\; \int_0^\infty \frac{(k/\mu)^k}{(k-1)!}\, x^{k-1}\, e^{-xk/\mu}\, e^{-x}\, \frac{x^m}{m!}\, dx \;=$$

$$\frac{(k/\mu)^k}{\Gamma(k)\,\Gamma(m+1)} \int_0^\infty x^{k+m-1}\, e^{-x(1+k/\mu)}\, dx$$

$$= \; \frac{\Gamma(m+k)}{\Gamma(k)\,\Gamma(m+1)}\,(k/\mu)^k\,(1+k/\mu)^{-(m+k)}$$

Randomness in the Poisson parameter has the effect of changing the link! More generally, need estimation strategy to recognize overdispersion.

# Quasilikelihood Extension of GLM

Assume instead of exponential family, $Y_i \sim h(y,\phi) e^{(\theta y - c(\theta))/a(\phi)}$

$\phi$ is a new dispersion parameter

Check again (same as before) that $c'(\theta) = E(Y_i)$. For variances:

$$0 = \frac{\partial}{\partial\theta} \int (y - c'(\theta)) f(y,\theta)\, dy = \int \left[ \frac{(y - c'(\theta))^2}{a(\phi)} - c''(\theta) \right] f(y,\theta)\, dy$$

which implies: $\text{Var}(Y) = a(\phi) c''(\theta)$

So score estimating equation (for $\theta$) becomes

$$\mathbf{0} = \sum_{i=1}^n \nabla_\beta \left[ (\theta_i Y_i - c(\theta_i))/a(\phi) \right] = \sum_{i=1}^n X_i \frac{Y_i - \mu_i}{g'(\mu_i)\, a(\phi)\, v(\mu_i)}$$

same with factor $1/a(\phi)$ and $J = \mathbf{X}^{tr} (W/a(\phi))\, \mathbf{X}$

# Dispersion Parameter

$a(\phi)$ generally found not from likelihood but from a moment expression like

$$n^{-1} \sum_{i=1}^{n} \frac{(Y_i - \widehat{\mu}_i)^2}{v(\widehat{\mu}_i)} = n^{-1} \sum_{i=1}^{n} \frac{(Y_i - g^{-1}(X_i^{tr}\,\widehat{\beta})^2}{v \circ g^{-1}(X_i^{tr}\,\widehat{\beta})}$$

to indicate overdispersion when this estimate is clearly $> 1$

# R Script with Illustrations of Methods

(i) Quasilikelihood example (simulated), `RscriptLec14.RLog`

(ii) Quasilikelihood example as in book with Crab mating, in script `GLMsScript.RLog`

(iii) Predictive Power of Logistic Regression Models

- % reduction in sum of squared errors
- % reduction in deviance
- ROC plot and AUC

# Predictive Power of Logistic Regressions

Continue with this topic later on, using

– Binning of predicted values to make sure that the approximately correct proportions of $Y = 1$ cases are seen within each bin

– Cross-validated estimates of proportion of correct predictions

– Comparison with ‘Machine Learning’ algorithms
for binary prediction