Overdispersion and Distributions

Nydjie Grimes & Jon Moyer

9/28/2017

Overdispersion

- In possion model

$$y_i \sim poisson\ (\lambda_i)$$
$$E(y_i) = (\lambda_i)$$
$$V(y_i) = (\lambda_i)$$

- GLM 'estimation notation'

$$\mu_i = E(y_i)$$
$$V(\mu_i) = \phi\mu_i, \phi > 1$$
$$= \phi V(\mu_i)$$

Note: $\phi > 1$ implies overdispersion in your model

- Likelihood equations

$$\sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{v(\mu_i)}(\frac{\partial \mu_i}{\partial \eta_i}) = 0$$

for j= 1,...,p, this depends on the distribution of y through μ, $v(\mu)$

if add ϕ

$$\phi \sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{v(\mu_i)}$$

Having $\frac{1}{\phi}$ in these equations doesn't change MLE of β's (it ends up dropping out of the equation)

However:

$$cov(\hat{\beta}) = (X^T W X)^{-1} = \phi cov_{GLM}(\hat{\beta})$$

Note: cov_{GLM} with no dispersion

W is a matrix with diagonal w_i

$$w_i = \frac{(\frac{\partial \mu_i}{\partial \eta_i})^2}{v(y_i)} = \frac{(\frac{\partial \mu_i}{\partial \eta_i})^2}{\phi \mu_i}$$

When is Overdispersion Necessary Ex: 4.7.4

Start with standardized residual from non-OD model

$$Z_i = \frac{y_i - (\hat{y}_i)}{sd(\hat{y}_i)} = \frac{y_i - \mu_i}{\sqrt{\mu_i}} \sim N(0, 1)$$

$$\sum_{i=1}^{N} z_i^2 \sim \chi^2_{n-k}$$

Note: k is the number of parameters

$\hat{\phi} = \frac{\sum z_i^2}{n-k}$ summarizes overdispersion in data compared to model, if $\hat{\phi} > 1$ then we should use overdispersion

R code is glm(. . . ., family = "quassipoisson" or "quasibinomial")

Back-of-envelope

$$SE_{overdispersion}(\hat{\beta}) = SE_{overdispersion}(\hat{\beta}) * \hat{\phi}1/2$$

We can estimate the sampling distribution of $\hat{\beta}'s$:

$cov(\hat{\beta}) = (X^T W X)^{-1}$

1. Confidence interval
$$\hat{\beta} + 2 * SE(\hat{\beta})$$
   Under certain conditions, $SE(\hat{\beta})$ achieves the lowest possible variance for an unbiased estimator.The conditions include: independence of observations, random sample from population of interest, large enough sample size, and the model is true.
2. Hypothesis Test
$$H_0 : \beta = \beta_0$$
   p-value is when we assume H_0 is true and calculate the probability of observing $\hat{p}$ or $Pr(\hat{\beta}|\beta_0$ is true)

Likelihood Bayesian Analysis

- No fixed "true" β
- Posterior distribution is same as our likelihood

  posterior = likelihood (prior)