Regression for Overdispersed Binomial Data
`bod.regn()`

1 Overview

This function permits joint estimation of a mean function and a variance function for counts of successes/failures.

2 Model

A logistic regression is used to model the mean response rate. Let Y_i denote the number of successes out of N_i trials. The mean model is:

$$\begin{aligned} E[Y_i \mid N_i] &= N_i p_i \\ \text{logit}(p_i) &= X_i \beta \quad . \end{aligned}$$

The variance of $Y_i \mid N_i$ is specified by a second regression model. Let τ_i represent a parameter that specifies the variance, $V[Y_i \mid N_i] = f(N_i, p_i, \tau_i)$. We consider (3) choices for the functional form of the variance:

1. **"scale"**
   $V[Y_i \mid N_i] = N_i p_i (1 - p_i) \tau_i$.

2. **"correlation"**
   $V[Y_i \mid N_i] = N_i p_i (1 - p_i) \{1 + (N_i - 1)\tau_i\}$.

3. **"variance"**
   $V[Y_i \mid N_i] = N_i p_i (1 - p_i) + N_i (N_i - 1)\tau_i$.

Given the form of the variance, a second regression is specified as:

$$\log(\tau_i) = Z_i \alpha \quad .$$

3 Estimation

Estimation of (β, α) is obtained through extended quasi-likelihood. We solve for a root of the pair of estimating equations:

$$\begin{aligned}
U_1(\beta, \alpha) &= \sum_{i=1}^{K} \frac{\partial \mu_i}{\partial \beta}^T \nu_i(\beta, \alpha)^{-1}(Y_i - \mu_i(\beta)) \\
U_2(\beta, \alpha) &= \sum_{i=1}^{K} \frac{\partial \nu_i}{\partial \alpha}^T W_{2i}(\beta, \alpha)\{(Y_i - \mu_i(\beta))^2 - \nu_i(\beta, \alpha)\}
\end{aligned}$$

where $\mu_i = N_i p_i = E[Y_i \mid N_i]$, $\nu_i = V[Y_i \mid N_i]$, and W_{2i}^{-1} approximates the variance of $(Y_i - \mu_i)^2$. Given α, U_1 is the quasi-score function for β and hence an optimal estimating equation for β. Similarly, for a fixed β, U_2 is an optimal estimating equation for α if W_{2i}^{-1} is the variance of the squared residuals, $(Y_i - \mu_i)^2$.

The estimating equations are solved using Gauss-Seidel iteration. A model based estimate of the variance of the estimator $\hat{\beta}$ is avaiable given $\hat{\alpha}$. Without further model assumptions (third and forth moment models for Y_i), only an empirical estimate of the variance of $(\hat{\beta}, \hat{\alpha})$ is available. Note: given interest we could implement a higher moment model based on the beta-binomial model.

4 Usage

Call:

```
fit <- bod.regn( y, n, x, z, beta, alpha, dispersion=‘‘variance’’ )
```

Arguments:

- **y** : vector of counts of successes.
- **n** : vector of counts of number of trials.
- **x** : mean regression design matrix.
- **z** : variance regression design matrix.
- **beta** : initial estimate of mean regression parameter.
- **alpha** : initial estimate of variance regression parameter.

dispersion : variance form. Choices are "scale", "correlation", and "variance".

Return:
The function returns a list with the following components:

dispersion: dispersion model used.

beta: estimate of mean regression parameter.

alpha: estimate of variance regression parameter.

cov.model: model based variance estimate for $\hat{\beta}$.

cov.empirical: sandwich based variance estimate for $(\hat{\beta}, \hat{\alpha})$.