intermediate generalized linear models

Ben Bolker

overdispersion

overdispersion

- more variance than expected based on statistical model
- e.g. variance > mean for Poisson
- in general leads to *overconfidence*
  - overly narrow confidence intervals
  - too-small p-values
  - inflated type I error

methods

- quasi-likelihood models
- compounded distributions
- observation-level random effects

quasi-likelihood

- quantify excess variance
- e.g. ϕ=`sum(residuals(m,type="pearson")^2)/df.residual(m)`
- multiply estimated standard errors by $\sqrt{phi}$
- recompute Z/t statistics, p values
- `family=quasipoisson` or `family=quasibinomial` does this automatically
- no likelihood/AIC available

compounded distributions

- instead of Poisson/binomial/etc., use a compounded distribution
- Gamma + Poisson = negative binomial (e.g. `MASS::glmer.nb`)
- Beta + binomial = beta-binomial (e.g. `glmmTMB`)

observation-level random effects

- use mixed models; add a Normal deviate to each observation (on the link-function/linear predictor scale)
- e.g. logit-Normal-binomial, or log-Normal-Poisson

offsets

-

complete separation

- Wald approximation breaks down (*Hauck-Donner effect*)

zero-inflation

alternative families and links

Gamma

complementary log-log

beyond the exponential family

beta regression

- GLMs require counts (denominators), e.g. 40% = 4/10
- what if data don't have obvious denominators
- e.g. cover scores, activity budgets
- *Beta distribution*

negative binomial regression