Introduction to modelling count data in glmmTMB
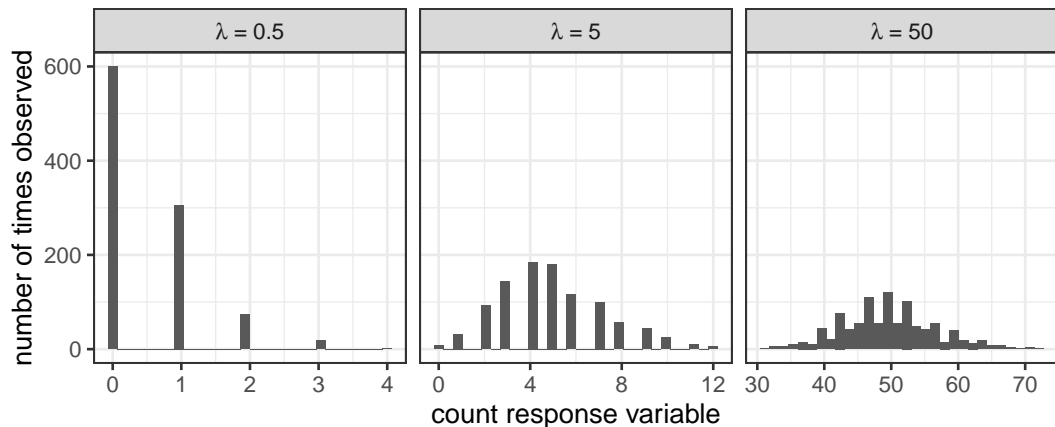
Mollie Brooks

June 2023

This video covers

- several distributions useful for count data

- overdispersion relative to the Poisson

- offsets

Poisson

- variance = mean
- poisson(link = "log")

Negative binomial distributions

- variance $>=$ mean

- `family = nbinom2(link = "log")`

  - Poisson distributions with gamma distributed λs

  - variance is $\mu(1 + \mu/\theta)$

  - larger θ corresponds to a **lower** variance

- `family = nbinom1(link = "log")`

  - variance is $\mu(1 + \alpha)$

  - Hardin and Hilbe (2007)

see `code_counts.R`

Conway-Maxwell-Poisson distribution

- fits over- or underdispersed distributions

- `family = compois(link = "log")`

- typically, CMP GLMs have been parameterized by the mode

- new methods for mean-parameterization (Huang 2017)

- mean-parameterization implemented in `glmmTMB`

- slower than other distributions

Offsets

- allow response rate to be relative
- account for unequal sampling effort per observation
- e.g. jellyfish per 1000 liters of water filtered

$N_i \sim Nbinom2(\mu_i, \phi)$, log link

$\mu_i = e^{\hat{\beta}_0 + \hat{\beta}_d * depth_i + \hat{\beta}_o * oxygen_i + \hat{\beta}_s * salinity_i + log(volume\_filtered_i)}$

$\mu_i = e^{\hat{\beta}_0 + \hat{\beta}_d * depth_i + \hat{\beta}_o * oxygen_i + \hat{\beta}_s * salinity_i} * volume\_filtered_i$

$\mu_i / volume\_filtered_i = e^{\hat{\beta}_0 + \hat{\beta}_d * depth_i + \hat{\beta}_o * oxygen_i + \hat{\beta}_s * salinity_i}$

Recap

- Poisson has strict variance=mean

- negative binomial handles more variance

- Conway-Maxwell-Poisson allows more or less variance

- DHARMa to test dispersion

- predictions on the response scale

- offsets