Ch1\_Wolf\_modelDistribution

Erin Tattersall

May 8, 2018

Modelling detections rates between November 2015 and April 2018 for wolves.  
Deciding on most appropriate distribution for response variable.

Proportion of zeros in data

sum(det$Wolf==0, na.rm = TRUE)/nrow(det)

## [1] 0.9105556

91% of the data is zeroes –> very likely zero-inflated.

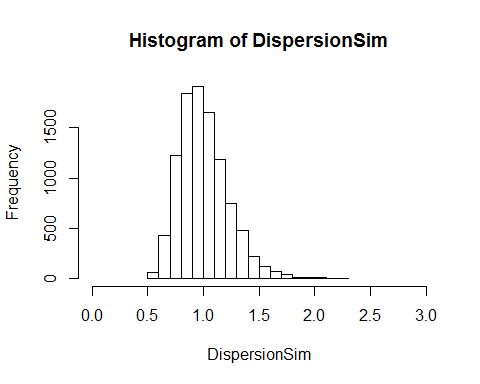
### Fitting a basic poisson GLM and checking overdispersion, using global model (doesn’t yet include SnowDays)

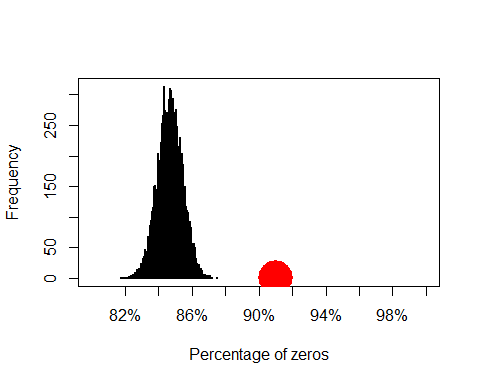
glm1 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays, data = det, family = poisson)  
# Residuals and overdispersion  
E1 <- resid(glm1, type="pearson")  
N <- nrow(det)  
p <- length(coef(glm1))  
sum(E1^2)/(N-p)

## [1] 2.321066

Dispersion value of 2.30 indicates overdispersion

Simulating data to check probability of calculated dispersion, given a Poisson GLM.

 Histogram indicates that an overdispersion statistic of 2.30 is outside the likely distribution of dispersion statistics for Poisson distributed response variables, suggesting that Wolf data is likely overdispersed.

Comparing the proportion of zeros in data to simulated zeros from model shows that Wolf data has more zeroes than would be expected in a Poisson GLM  This presents a case for using zero-inflated models, which can be verified with model selection of GLMMs ## Model selection: choosing model form and distribution

Comparing the same GLMM (including random effects of Site and Month) modeled as a poisson, nb, ZIP and ZINB (with nbinom1 and nbinom2 differing in how variance changes with mean) yields:

## Model comparisons of distributions and zero-inflation  
glm1 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays + (1| Site) + (1|Month), data = det, family = poisson)  
glm2 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom1(link= "log")) #Warning 'matrix not positive definite'  
glm3 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom2(link = "log"))  
glm4 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = poisson)  
glm5 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays +(1| Site) + (1|Month), zi = ~1, data = det, family = nbinom1(link= "log")) #Warning 'matrix not positive definite'  
glm6 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = nbinom2(link= "log"))

## dLogLik dAIC df weight  
## Nbinom2 48.8 0.0 12 0.6641  
## ZINB2 49.1 1.4 13 0.3273  
## ZINB1 45.4 8.7 13 0.0084  
## ZIP 40.0 17.5 12 <0.001  
## Nbinom1 36.7 24.0 12 <0.001  
## Poisson 0.0 95.5 11 <0.001

Nbinom1 models showed warnings (see above), but output includes parameter estimates, so models are included. However, Nbinom2 distributions perform better

## Family: nbinom2 ( log )  
## Formula:   
## Wolf ~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays +   
## (1 | Site) + (1 | Month)  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 1228.5 1294.2 -602.2 1204.5 1758   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 1.2393 1.1132   
## Month (Intercept) 0.3166 0.5627   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom2 family (): 0.618   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.79438 1.37084 -4.956 7.18e-07 \*\*\*  
## TreatmentHumanUse 1.48865 0.57078 2.608 0.00911 \*\*   
## TreatmentNatRegen 0.01069 0.59613 0.018 0.98569   
## TreatmentSPP 0.59072 0.50680 1.166 0.24378   
## LineWidth 0.01596 0.14258 0.112 0.91087   
## LD1250 0.37324 0.35616 1.048 0.29466   
## VegHt 0.10352 0.21104 0.491 0.62376   
## low500 -1.80276 0.96601 -1.866 0.06201 .   
## ActiveDays 0.15266 0.01884 8.102 5.41e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ZINB model output:

## Family: nbinom2 ( log )  
## Formula:   
## Wolf ~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays +   
## (1 | Site) + (1 | Month)  
## Zero inflation: ~1  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 1229.9 1301.1 -601.9 1203.9 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 1.2278 1.1081   
## Month (Intercept) 0.3223 0.5677   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom2 family (): 1.08   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.50970 1.39453 -4.668 3.04e-06 \*\*\*  
## TreatmentHumanUse 1.47934 0.56698 2.609 0.00908 \*\*   
## TreatmentNatRegen 0.01158 0.59232 0.020 0.98440   
## TreatmentSPP 0.57854 0.50391 1.148 0.25093   
## LineWidth 0.01286 0.14177 0.091 0.92771   
## LD1250 0.38093 0.35417 1.076 0.28213   
## VegHt 0.10729 0.20956 0.512 0.60868   
## low500 -1.79041 0.95977 -1.865 0.06212 .   
## ActiveDays 0.15086 0.01855 8.133 4.20e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1.344 1.148 -1.17 0.242

Active Days affects the probability of observing a zero in data –> should be included in ZI model. It could be argued that it should NOT be included in conditional. I will test both

glm7 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 + ActiveDays +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom2(link= "log"))  
glm8 <- glmmTMB(Wolf~ Treatment + LineWidth + LD1250 + VegHt + low500 +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom2(link= "log"))

## dLogLik dAIC df weight  
## ZINB2-AD1 51.2 0.0 13 0.7339  
## Nbinom2 48.8 2.8 12 0.1767  
## ZINB2 49.1 4.3 13 0.0871  
## ZINB1 45.4 11.6 13 0.0022  
## ZIP 40.0 20.4 12 <0.001  
## Nbinom1 36.7 26.9 12 <0.001  
## Poisson 0.0 98.4 11 <0.001

## Family: nbinom2 ( log )  
## Formula:   
## Wolf ~ Treatment + LineWidth + LD1250 + VegHt + low500 + (1 |   
## Site) + (1 | Month)  
## Zero inflation: ~ActiveDays  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 1225.6 1296.8 -599.8 1199.6 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 1.1837 1.0880   
## Month (Intercept) 0.2638 0.5136   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom2 family (): 0.636   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.19304 1.22050 -1.797 0.0724 .  
## TreatmentHumanUse 1.34242 0.55807 2.406 0.0162 \*  
## TreatmentNatRegen -0.11789 0.58438 -0.202 0.8401   
## TreatmentSPP 0.52631 0.49656 1.060 0.2892   
## LineWidth 0.02425 0.13993 0.173 0.8624   
## LD1250 0.37148 0.34974 1.062 0.2882   
## VegHt 0.07564 0.20764 0.364 0.7156   
## low500 -1.76958 0.94706 -1.868 0.0617 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 13.425 12.212 1.099 0.272  
## ActiveDays -1.477 1.548 -0.954 0.340

Model with ActiveDays in Conditional and ZI produced estimates, but not std. errors. Excluded. Including ActiveDays in the ZI model results in more logLikelihood of model and more AIC weight. For Wolf models, I will therefore use zero inflated GLMMs with a nbinom2 distribution (where variance changes quadratically with the mean), including ActiveDays in the ZI model.  
# Model hypotheses  
## Finding random structure

## dLogLik dAIC df weight  
## r2 51.5 0.0 16 1   
## rSite 41.3 18.3 15 <0.001  
## rMonth 12.7 75.5 15 <0.001  
## r0 0.0 98.9 14 <0.001

Continue modelling with 2 random effects