Ch1\_Moose\_modelDistribution

Erin Tattersall

May 8, 2018

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

Proportion of zeros in data

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

## [1] 0.9355556

94% of the data is zeroes –> very likely zero-inflated.

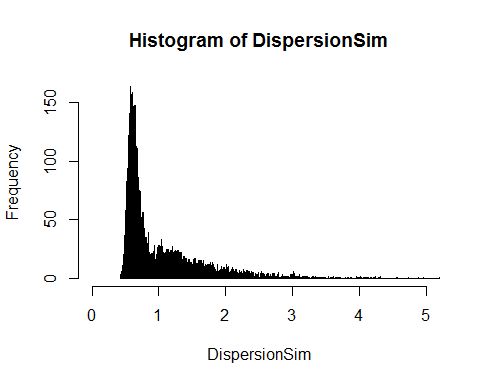
### Fitting a basic poisson GLM and checking overdispersion, using global model (doesn’t yet include SnowDays)

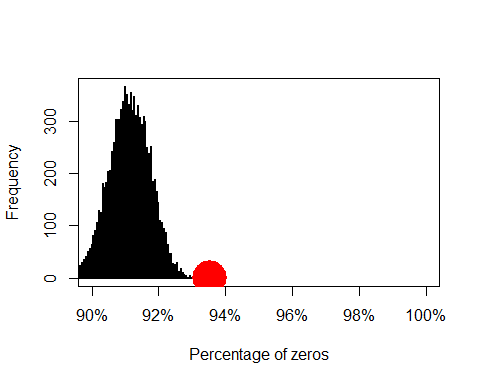
glm1 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + ActiveDays, data = det, family = poisson)  
# Residuals and overdispersion  
E1 <- resid(glm1, type="pearson")  
N <- nrow(det)  
p <- length(coef(glm1))  
sum(E1^2)/(N-p)

## [1] 1.02837

Dispersion value of 1.03 indicates data is not necessarily overdispersed

Simulating data to check probability of calculated dispersion.

 Histogram indicates that an overdispersion statistic of 1.03 is well within the likely distribution of dispersion statistics for Poisson distributed response variables, suggesting that Moose data is not necessarily overdispersed.

Comparing the proportion of zeros in data to simulated zeros from model shows that Moose 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(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays + (1| Site) + (1|Month), data = det, family = poisson)  
glm2 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom1(link= "log"))  
glm3 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom2(link = "log"))  
glm4 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = poisson)  
glm5 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays +(1| Site) + (1|Month), zi = ~1, data = det, family = nbinom1(link= "log")) ## warning 'matrix not positive definite'  
glm6 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = nbinom2(link= "log"))

## dLogLik dAIC df weight  
## Nbinom2 23.4 0.0 13 0.6678  
## ZINB2 23.4 2.0 14 0.2457  
## Nbinom1 20.7 5.2 13 0.0484  
## ZINB1 21.3 6.1 14 0.0311  
## ZIP 18.8 9.1 13 0.0071  
## Poisson 0.0 44.7 12 <0.001

ZINB1 displayed a warning when modeled, but parameter estimates + s.e. obtained, so included in model selection.  
Nbinom2 model’s summary output is:

## Family: nbinom2 ( log )  
## Formula:   
## Moose ~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km +   
## ActiveDays + (1 | Site) + (1 | Month)  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 894.8 966.0 -434.4 868.8 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.7509 0.8665   
## Month (Intercept) 0.5398 0.7347   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom2 family (): 0.532   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -9.93847 1.80702 -5.500 3.8e-08 \*\*\*  
## TreatmentHumanUse 0.58802 0.54417 1.081 0.2799   
## TreatmentNatRegen -0.29384 0.51236 -0.574 0.5663   
## TreatmentSPP 0.02923 0.46262 0.063 0.9496   
## LineWidth 0.26335 0.13397 1.966 0.0493 \*   
## LD1250 -0.19604 0.29398 -0.667 0.5049   
## VegHt 0.43951 0.18613 2.361 0.0182 \*   
## low250 -0.55050 0.69884 -0.788 0.4309   
## Dist2Water\_km -0.12976 0.12491 -1.039 0.2989   
## ActiveDays 0.20229 0.04621 4.377 1.2e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ZINB model output:

## Family: nbinom2 ( log )  
## Formula:   
## Moose ~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km +   
## ActiveDays + (1 | Site) + (1 | Month)  
## Zero inflation: ~1  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 896.8 973.5 -434.4 868.8 1756   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.7508 0.8665   
## Month (Intercept) 0.5398 0.7347   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom2 family (): 0.532   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -9.93842 1.80701 -5.500 3.8e-08 \*\*\*  
## TreatmentHumanUse 0.58802 0.54417 1.081 0.2799   
## TreatmentNatRegen -0.29384 0.51236 -0.574 0.5663   
## TreatmentSPP 0.02924 0.46262 0.063 0.9496   
## LineWidth 0.26335 0.13397 1.966 0.0493 \*   
## LD1250 -0.19604 0.29398 -0.667 0.5049   
## VegHt 0.43951 0.18613 2.361 0.0182 \*   
## low250 -0.55051 0.69884 -0.788 0.4308   
## Dist2Water\_km -0.12976 0.12491 -1.039 0.2989   
## ActiveDays 0.20228 0.04621 4.378 1.2e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -16.91 4725.20 -0.004 0.997

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(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km + ActiveDays +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom2(link= "log"))  
glm8 <- glmmTMB(Moose~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom2(link= "log"))  
summary(glm8)

## Family: nbinom2 ( log )  
## Formula:   
## Moose ~ Treatment + LineWidth + LD1250 + VegHt + low250 + Dist2Water\_km +   
## (1 | Site) + (1 | Month)  
## Zero inflation: ~ActiveDays  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 896.2 972.9 -434.1 868.2 1756   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.7475 0.8646   
## Month (Intercept) 0.5388 0.7340   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom2 family (): 0.568   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.79550 1.12902 -3.362 0.000774 \*\*\*  
## TreatmentHumanUse 0.59257 0.54348 1.090 0.275568   
## TreatmentNatRegen -0.28314 0.51191 -0.553 0.580191   
## TreatmentSPP 0.03399 0.46168 0.074 0.941311   
## LineWidth 0.26162 0.13378 1.956 0.050510 .   
## LD1250 -0.19838 0.29340 -0.676 0.498956   
## VegHt 0.44424 0.18759 2.368 0.017880 \*   
## low250 -0.52283 0.69824 -0.749 0.453987   
## Dist2Water\_km -0.13455 0.12479 -1.078 0.280955   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 10.0827 5.5119 1.829 0.0674 .  
## ActiveDays -0.4554 0.2678 -1.701 0.0890 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## dLogLik dAIC df weight  
## Nbinom2 23.4 0.0 13 0.4467  
## ZINB2-AD1 23.6 1.5 14 0.2136  
## ZINB2 23.4 2.0 14 0.1643  
## ZINB2-AD2 24.0 2.7 15 0.1175  
## Nbinom1 20.7 5.2 13 0.0324  
## ZINB1 21.3 6.1 14 0.0208  
## ZIP 18.8 9.1 13 0.0048  
## Poisson 0.0 44.7 12 <0.001

Nbinom2 without zero-inflation still comes out as top model. However, there is evidence that Moose detection data is zero-inflated, justifying including zero-inflation a priori. Further, examining the output for ZINB2-AD1 shows that ActiveDays in the ZI model has a small but near significant effect on detecting structured zeros (and that there is a high chance of observing a structured zero in data regardless).  
I will therefore model moose data as zero-inflated nbinom2 with ActiveDays in the ZI model.

# Model hypotheses  
## Finding random structure

## dLogLik dAIC df weight  
## r2 17.8 0.0 17 1   
## rSite 9.4 14.9 16 <0.001  
## rMonth 7.1 19.5 16 <0.001  
## r0 0.0 31.6 15 <0.001

Continue modelling with 2 random effects