Ch1\_Caribou\_modelDistribution

Erin Tattersall

April 26, 2018

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

Proportion of zeros in data

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

## [1] 0.9544444

95% of the data is zeroes –> very likely zero-inflated.

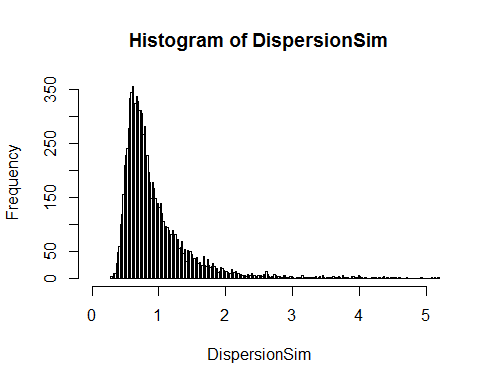
### Fitting a basic poisson GLM and checking overdispersion, using global model (doesn’t yet include SnowDays)

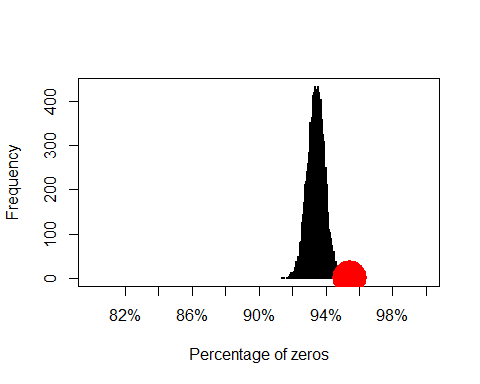
glm1 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays, data = det, family = poisson)  
# Residuals and overdispersion  
E1 <- resid(glm1, type="pearson")  
N <- nrow(det)  
p <- length(coef(glm1))  
sum(E1^2)/(N-p)

## [1] 0.8899986

Dispersion value of 0.89 indicates underdispersion, which would correspond to zero-inflation

Simulating data to check probability of calculated dispersion.

 Histogram indicates that an overdispersion statistic of 0.89 is well within the likely distribution of dispersion statistics for Poisson distributed response variables, suggesting that Caribou data is not necessarily overdispersed.

Comparing the proportion of zeros in data to simulated zeros from model shows that Caribou 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(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays + (1| Site) + (1|Month), data = det, family = poisson)  
glm2 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom1(link= "log"))  
glm3 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays + (1| Site) + (1|Month), data = det, family = nbinom2(link = "log"))  
glm4 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = poisson)  
glm5 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays +(1| Site) + (1|Month), zi = ~1, data = det, family = nbinom1(link= "log"))  
glm6 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays + (1| Site) + (1|Month), zi = ~1, data = det, family = nbinom2(link= "log"))

## dLogLik dAIC df weight  
## Nbinom1 23.9 0.0 12 0.7131  
## ZINB1 23.9 2.0 13 0.2623  
## Nbinom2 20.2 7.4 12 0.0177  
## ZINB2 20.2 9.4 13 0.0065  
## ZIP 16.2 15.5 12 <0.001  
## Poisson 0.0 45.8 11 <0.001

Where the Nbinom1 model’s summary output is:

## Family: nbinom1 ( log )  
## Formula:   
## Caribou ~ Treatment + LineWidth + LD1750 + VegHt + low1750 +   
## ActiveDays + (1 | Site) + (1 | Month)  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 634.3 700.1 -305.2 610.3 1758   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.6644 0.8151   
## Month (Intercept) 0.7980 0.8933   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom1 family (): 0.931   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -12.70557 2.18230 -5.822 5.81e-09 \*\*\*  
## TreatmentHumanUse -0.48517 0.67563 -0.718 0.473   
## TreatmentNatRegen -1.30174 1.17484 -1.108 0.268   
## TreatmentSPP 0.58697 0.51189 1.147 0.252   
## LineWidth 0.18267 0.15741 1.160 0.246   
## LD1750 -0.81014 0.50870 -1.593 0.111   
## VegHt -0.56600 0.45574 -1.242 0.214   
## low1750 8.94381 1.88716 4.739 2.14e-06 \*\*\*  
## ActiveDays 0.13723 0.02778 4.940 7.80e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ZINB model output:

## Family: nbinom1 ( log )  
## Formula:   
## Caribou ~ Treatment + LineWidth + LD1750 + VegHt + low1750 +   
## ActiveDays + (1 | Site) + (1 | Month)  
## Zero inflation: ~1  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 636.3 707.5 -305.2 610.3 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.6644 0.8151   
## Month (Intercept) 0.7980 0.8933   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom1 family (): 0.931   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -12.70563 2.18231 -5.822 5.81e-09 \*\*\*  
## TreatmentHumanUse -0.48516 0.67563 -0.718 0.473   
## TreatmentNatRegen -1.30172 1.17483 -1.108 0.268   
## TreatmentSPP 0.58697 0.51189 1.147 0.252   
## LineWidth 0.18267 0.15741 1.160 0.246   
## LD1750 -0.81014 0.50870 -1.593 0.111   
## VegHt -0.56599 0.45574 -1.242 0.214   
## low1750 8.94387 1.88718 4.739 2.14e-06 \*\*\*  
## ActiveDays 0.13723 0.02778 4.940 7.80e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -17.71 5298.55 -0.003 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(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 + ActiveDays +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom1(link= "log"))  
glm8 <- glmmTMB(Caribou~ Treatment + LineWidth + LD1750 + VegHt + low1750 +(1| Site) + (1|Month), zi = ~ActiveDays, data = det, family = nbinom1(link= "log"))

## dLogLik dAIC df weight  
## ZINB1-AD1 26.0 0.0 13 0.6916  
## Nbinom1 23.9 2.3 12 0.2200  
## ZINB1 23.9 4.3 13 0.0809  
## Nbinom2 20.2 9.7 12 0.0055  
## ZINB2 20.2 11.7 13 0.0020  
## ZIP 16.2 17.8 12 <0.001  
## Poisson 0.0 48.1 11 <0.001

## Family: nbinom1 ( log )  
## Formula:   
## Caribou ~ Treatment + LineWidth + LD1750 + VegHt + low1750 +   
## (1 | Site) + (1 | Month)  
## Zero inflation: ~ActiveDays  
## Data: det  
##   
## AIC BIC logLik deviance df.resid   
## 632.0 703.3 -303.0 606.0 1757   
##   
## Random effects:  
##   
## Conditional model:  
## Groups Name Variance Std.Dev.  
## Site (Intercept) 0.6592 0.8119   
## Month (Intercept) 0.8051 0.8973   
## Number of obs: 1770, groups: Site, 59; Month, 12  
##   
## Overdispersion parameter for nbinom1 family (): 0.966   
##   
## Conditional model:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -8.4133 1.9762 -4.257 2.07e-05 \*\*\*  
## TreatmentHumanUse -0.4827 0.6709 -0.720 0.472   
## TreatmentNatRegen -1.3811 1.1748 -1.176 0.240   
## TreatmentSPP 0.5434 0.5100 1.065 0.287   
## LineWidth 0.1770 0.1566 1.130 0.258   
## LD1750 -0.8362 0.5094 -1.642 0.101   
## VegHt -0.5388 0.4489 -1.200 0.230   
## low1750 8.8133 1.8779 4.693 2.69e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Zero-inflation model:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 8.3695 9.1929 0.910 0.363  
## ActiveDays -0.9041 1.2113 -0.746 0.455

The ZINB including ActiveDays in conditional and ZI did not produce std. errors for parameter estimates, and was therefore excluded.  
Including ActiveDays in the ZI model results in more logLikelihood of model and more AIC weight. For caribou models, I will therefore use zero inflated GLMMs with a nbinom1 distribution (where variance changes linearly with the mean), including ActiveDays in the ZI model.  
# Model hypotheses  
## Finding random structure

## dLogLik dAIC df weight  
## r2 25.5 0.0 16 0.9939  
## rMonth 19.4 10.2 15 0.0061  
## rSite 5.7 37.6 15 <0.001  
## r0 0.0 47.0 14 <0.001

Continue modelling with 2 random effects