Nonlinear, Non-normal models

Lecture 05.3: Determining your Distribution

Lauren Sullivan

Module: Linear, Nonlinear, and Mixed Effects Models

Readings

Required for class:

- NA

Optional:

- Sullivan et al. 2016 *PLOS ONE*
- Bolker, B. *Ecological Models and Data in R - Ebook version*
- Brooks-Bartlett, J. *Probability concepts explained: Maximum likelihood estimation*

What distribution to use?

Sometimes you are in a situation where you have your data, but you do not know what distribution fits your dependent variables (Y's) and thus you don't know what approach to take for the rest of your analysis.

You need a method for determining which distribution fits your data most appropriately.

Maximum Likelihood

To determine what distribution best fits your data, you need to use maximum likelihood estimation, which allows you to determine the *maximum likelihood estimates* (MLE) of the parameters of the potential distributions you are testing. These parameter estimates define the distribution so it **MOST LIKELY** fits the data.

NOTE: We try to maximize the log of the likelihood. But note that in most cases people *minimize* the negative log-likelihood instead of maximizing the positive log-likelihood.

Data

Restored prairie experiment to understand how the presence/absence of hervivores alter the movement of the native Partridge Pea (*Chamaecrista fasciculata*) away from source locations.

Chamaecrista fasciculata

Figure 1: From: Sullivan et al. 2016 PLOS ONE

Experimental Design

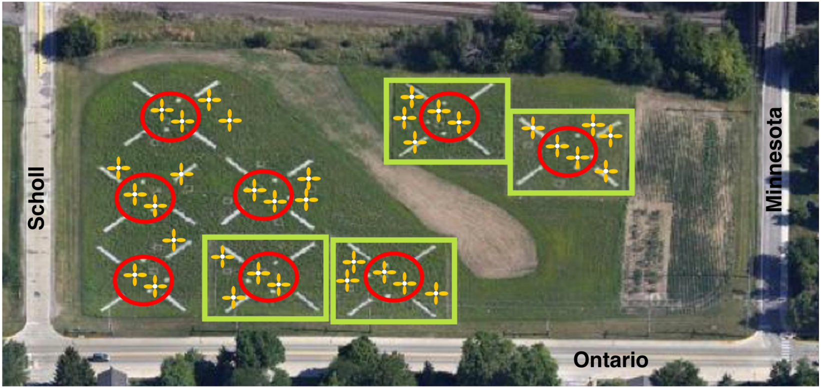
Have experimental units that either allow/deny herbivores access to the restored prairies.

Experimental Design

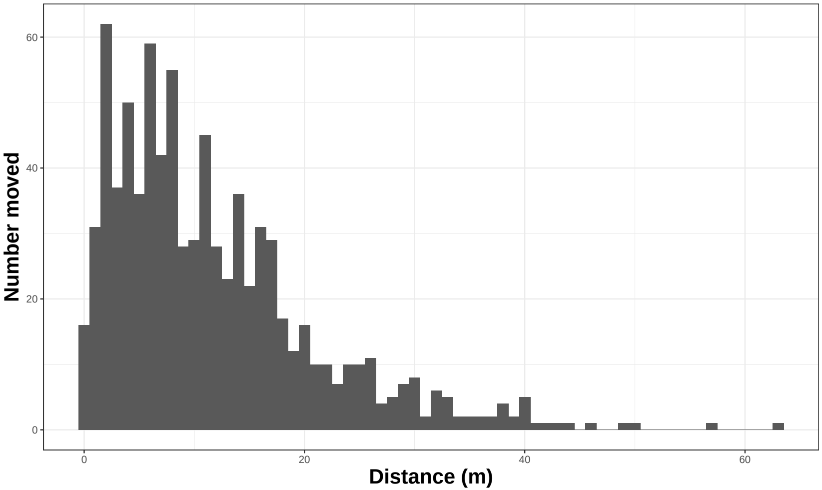
Plant Partridge Pea in known locations.

Experimental Design

Measure how far away from the sources the new plants move and determine the shape of this distribution.



Graph the Y's to see what shape they take.



Fit different distributions with MLE - Normal

```
fit.norm.disc <- mle2(chafas$distance_.1bins~dnorm(mean=mu, sd=s),
                      start=list(mu=mean(chafas$distance_.1bins),
                                 s=sd(chafas$distance_.1bins)),
                      data=chafas)
summary(fit.norm.disc)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = chafas$distance_.1bins ~ dnorm(mean = mu, sd = s),
##     start = list(mu = mean(chafas$distance_.1bins), s = sd(chafas$distance_.
##     data = chafas)
##
## Coefficients:
##    Estimate Std. Error z value     Pr(z)
## mu 11.74002    0.32745  35.853 < 2.2e-16 ***
## s   9.41667    0.23154  40.669 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 6055.989
```

Fit different distributions with MLE - Exponential

```
fit.exp.disc<- mle2(chafas$distance_.1bins~dexp(rate=rate),
                    start=list(rate=11), data= chafas)
summary(fit.exp.disc)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = chafas$distance_.1bins ~ dexp(rate = rate),
##     start = list(rate = 11), data = chafas)
##
## Coefficients:
##       Estimate Std. Error z value     Pr(z)
## rate 0.0851962  0.0029626  28.758 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 5727.808
```

Fit different distributions with MLE - Poisson

```r
fit.pois.disc<-mle2(chafas$distance_.1bins~dpois(lambda=l),
                    start=list(l=mean(chafas$distance_.1bins)),
                    data= chafas)
summary(fit.pois.disc)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = chafas$distance_.1bins ~ dpois(lambda = l),
##     start = list(l = mean(chafas$distance_.1bins)), data = chafas)
##
## Coefficients:
##   Estimate Std. Error z value     Pr(z)
## l 11.74002    0.11915  98.534 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 9031.689
```

Fit different distributions with MLE - Negative Binomial

```
fit.nbinom.disc<-mle2(chafas$distance_.1bins~dnbinom(mu=m, size=s),
                      start=list(m=10, s=.1),data= chafas)
summary(fit.nbinom.disc)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = chafas$distance_.1bins ~ dnbinom(mu = m, size = s),
##     start = list(m = 10, s = 0.1), data = chafas)
##
## Coefficients:
##   Estimate Std. Error z value     Pr(z)
## m 11.74005    0.32578  36.036 < 2.2e-16 ***
## s  1.81273    0.10175  17.816 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 5688.677
```

Use AIC to determine the difference between models

```
AICtab(fit.norm.disc, fit.exp.disc, fit.pois.disc, fit.nbinom.disc)
```

```
##                   dAIC   df
## fit.nbinom.disc    0.0 2
## fit.exp.disc      37.1 1
## fit.norm.disc    367.3 2
## fit.pois.disc   3341.0 1
```

Graph the data and estimated distributions of your data based on likelihood estimates of parameters