

## APPENDIX 2

Sequence of commands for spatial distribution of count data (Negative Binomial Distribution).  
Aggregated distribution of individuals of *Scolecopsis squamata* in a sandy beach on the Caribbean, Costa Rica.

#The data file is called beach.fauna in format .csv, and it is available at :

<http://www.kerwa.ucr.ac.cr/handle/10669/29428>

#Commands based in Krebs (1999) and Crawley (2007)

```
attach(beach_fauna)
```

```
#STEP 1) SAMPLE SIZE.
```

```
n<-length(Scolecopsis)
```

```
n
```

```
#STEP 2) POPULATION PARAMETERS.
```

```
Mean<-mean(Scolecopsis)
```

```
Mean
```

```
Variance<-var(Scolecopsis)
```

```
Variance
```

```
#STEP 3) OBSERVED VALUES
```

```
#Table of observed values = first row: number of individuals, and second row: number of sampling units.
```

```
table(Scolecopsis)
```

```
#Observed values including the full range of number of individuals.
```

```
Scolecopsis1 <- factor(Scolecopsis, levels = c(0:max(Scolecopsis+1)))
```

```
table(Scolecopsis1)
```

```
#Frequency of sampling units by counts of individuals.
```

```
TableCounts<-data.frame(table(Scolecopsis1))
```

```
freq<- c(TableCounts$Freq)
```

```
freq
```

```
#STEP 4) INDEX OF DISPERSION:
```

```
#random pattern (ID=1), uniform pattern (ID < 1), and clumped pattern (ID > 1).
```

```
ID<-var(Scolecopsis)/mean(Scolecopsis)
```

```
ID
```

```
#Index of dispersion test, based in a chi-squared.
```

```
CHI<-ID*(n-1)
```

```
CHI
```

```
#if chi-squared value is between tabular values the null hypothesis of random pattern (ID = 1) is not rejected.
```

```
qchisq(0.975,(n-1), lower.tail = FALSE) #tabular value of chi-squared 0.975
```

```
qchisq(0.025,(n-1), lower.tail = FALSE) #tabular value of chi-squared 0.025
```

```
#In this case, if the value of the index of dispersion is higher than tabulated values the pattern is clumped (ID > 1).
```

```
#STEP 5) NEGATIVE BINOMIAL FIT FOR CLUMPED PATTERN (ID > 1).
```

```
#Probability of obtaining X individuals in a sample follows a clumped pattern modeled by the binomial negative distribution.
```

```
#To fit the data to the negative binomial, the approximate negative-binomial exponent need be calculated:
```

```
k=mean(Scolecopsis)^2/ (var(Scolecopsis)- mean(Scolecopsis))
```

```
k
```

```
#Negative-binomial exponent for large samples (n > 20)
```

#The following commands for Crawley (2007) calculate the maximum likelihood estimate of the aggregation parameter, K:

```
x<-0:max(Scolecipis+1)
kfit <-function(x)
{lhs<-numeric()
 rhs<-numeric()
 y <-0:(length(x) - 1)
 j<-0:(length(x)-2)
 m <-sum(x * y)/(sum(x))
 s2 <-((sum(x * y^2) - sum(x * y)^2/sum(x))/(sum(x)- 1)
 k1 <-m^2/(s2 - m)
 a<-numeric(length(x)-1)
 for(i in 1:(length(x) - 1)) a[i] <-sum(x [- c(1:i)])
 i<-0
 for (k in seq(k1/1.2,2*k1,0.001)) {
   i<-i+1
   lhs[i] <-sum(x) * log(1 + m/k)
   rhs[i] <-sum(a/(k + j))}
 k<-seq(k1/1.2,2*k1,0.001)
 plot(k, abs(lhs-rhs),xlab="k",ylab="Difference",type="l")
 d<-min(abs(lhs-rhs))
 sdd<-which(abs(lhs-rhs)==d)
 k[sdd]}
K<-kfit(freq)
```

#approximate k

```
k
#maximum likelihood K
K
```

#STEP 6) AGGREGATION PROBABILITIES BY NEGATIVE BINOMIAL

#density function to obtain the p-value of found  $x$  individuals in a sample, with a mean ( $\mu$ ) and Negative-binomial exponent ( $k$ )

$$p(x) = \left(1 + \frac{\mu}{k}\right)^{-k} \frac{(k+x-1)!}{x! (k-1)!} \left(\frac{\mu}{\mu+k}\right)^x$$

```
binneg<-function(x,Mean,K) (1+ Mean /K)^(-K)*( Mean /( Mean +K))^ x *gamma(K+ x)/(factorial(x)*gamma(K))
binnegprob<-sapply(x, function(i) binneg(i,Mean,K))
```

#STEP 7) EXPECTED FREQUENCIES FOR A THEORETICAL NEGATIVE BINOMIAL

#The negative binomial probabilities can be expressed as frequencies by multiplying each by the sample size (n)

```
binnegfreq<-n*(1+ Mean /K)^(-K)*( Mean /( Mean +K))^ x *gamma(K+ x)/(factorial(x)*gamma(K))
```

#STEP 8) SUMMARY TABLE WITH OBSERVED AND EXPECTED VALUES

```
SummaryTable<-data.frame(x=TableCounts,binnegfreq,binnegprob)
SummaryTable
```

#STEP 9) COMPARISON OF THE OBSERVED AND EXPECTED FREQUENCIES WITH A BARPLOT

```
E<-(max(Scolecipis+1)+1)*2
both<-numeric(E)
both[1:E %% 2 != 0]<-freq
both[1:E %% 2 == 0]<-binnegfreq
```

```

labels<-character(E)
labels[1:E %% 2 == 0]<-as.character(0:max(Scoelepis+1))
par(mfrow=c(1,1), pch=22)
Fig<-barplot(both,col=rep(c("black","white"),max(Scoelepis+1)),names=labels, ylab="Frequency of samples",
  xlab="Counts")
legend(x = "topright",c("observed","expected"),fill=c("black","white"))

#STEP 10) GOODNESS OF FIT (G TEST):
#The null hypothesis is that observed values follow a negative binomial distribution
#To follow the assumption of all expected values should be 1 or more, the last observations need be pooled (Krebs, 1999):

cs<-factor(0:max(Scoelepis+1))
a<-min(which(SummaryTable$binnegfreq<1))
#for a first tail with expected values bellow of one use the command: a<-max(which(SummaryTable$binnegfreq>1))

levels(cs)[a:max(Scoelepis+a)]<-"+";levels(cs)

#expected values for G test:
ef<-as.vector(tapply(SummaryTable$binnegfreq,cs,sum))

#observed values for G test:
of<-as.vector(tapply(SummaryTable$x.Freq,cs,sum))

#Table of contingency for G test:

p<-as.vector(tapply(SummaryTable$binnegprob,cs,sum))
ContingenceTable<-data.frame(of,ef,p)
ContingenceTable

#G test:
G<-2*sum((of)*log(of/ef))
G
# Degrees of freedom= frequency classes (counts) - 3 statistics (total, mean and k) used to fit the negative binomial.
df<-sum(ef>0)-3
df
#Theoretical values of G test (using chi-squared distribution).
qchisq(0.95, df)
#probability of the null hypothesis.
pchisq(G, df, lower.tail = FALSE)
#In this case p>0.05, the null hypothesis can not be rejected.
#Conclusion: The data follows a negative binomial distribution.

#G test adjusted by Williams' correction for continuity for small samples (n < 200)(Krebs, 1999).
q<-1+((a+1)/(6*sum(of)*df))
q
Gadj<-G/q
Gadj
qchisq(0.95, df)
pchisq(Gadj, df, lower.tail = FALSE)
#in this case p > 0.05, and null hypothesis can not be rejected.
#Conclusion: The data follows a negative binomial distribution.

```