

### APPENDIX 3

Sequence of commands for spatial distribution of count data (Binomial Distribution).  
Uniform pattern of number of taxa 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(Taxa)
```

```
n
```

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

```
Mean<-mean(Taxa)
```

```
Mean
```

```
Variance<-var(Taxa)
```

```
Variance
```

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

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

```
table(Taxa)
```

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

```
Taxa1 <- factor(Taxa, levels = c(0:max(Taxa+1)))
```

```
table(Taxa1)
```

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

```
TableCounts<-data.frame(table(Taxa1))
```

```
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(Taxa)/mean(Taxa)
```

```
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 lower than tabulated values the pattern is uniform (ID < 1)
```

```
#STEP 5) BINOMIAL FIT FOR UNIFORM PATTERN (ID < 1)
```

```
#The binomial distribution is determined by the equation:  $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$ 
```

```
# with  $p(x) = \binom{\bar{x}}{k}$ 
```

```
#To fit the data to the binomial distribution, first k is estimated:  $k = \left( \frac{\bar{x}^2}{\bar{x}-s^2} \right)$ ,
```

```
#k is the maximum number of individuals can be found in a sampling unit, and a estimate of n.
```

```

k<-round(Mean^2/(Mean-Variance))
k
#The probability (p) of an event x occurs is calculated
p<-Mean/k
p
#and the probability of not occurrence of an event (q)
q<-1-p
q
#number of individuals
x<-0:max(Taxa+1)
x

# STEP 6) BINOMIAL DISTRIBUTION PROBABILITIES
#Probability of found an X number of individuals in a sampling unit.

binomialp<-dbinom(x, k, p)

#STEP 7) EXPECTED FREQUENCIES FOR A THEORETICAL BINOMIAL DISTRIBUTION
#The binomial probabilities can be expressed as frequencies by multiplying each by the sample size (n)
binomialfreq<-dbinom(x, k, p)*n

# STEP 8) SUMMARY TABLE WITH OBSERVED AND EXPECTED VALUES
SummaryTable<-data.frame(x=TableCounts,binomialfreq,binomialp)
SummaryTable

#STEP 9) COMPARISON OF THE OBSERVED AND EXPECTED FREQUENCIES WITH A BARPLOT
E<-(max(Taxa+1)+1)*2
both<-numeric(E)
both[1:E %% 2 != 0]<-freq
both[1:E %% 2 == 0]<-binomialfreq
labels<-character(E)
labels[1:E %% 2 == 0]<-as.character(0:max(Taxa+1))

Fig<-barplot(both,col=rep(c("black","white"),max(Taxa+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):

#USING G TEST
#The null hypothesis is that observed values follow a 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(Taxa+1))
a<-max(which(SummaryTable$binomialfreq>1))
levels(cs)[a:max(Taxa+a)]<-"+";levels(cs)

#expected values for G test:
ef<-as.vector(tapply(SummaryTable$binomialfreq,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$binomialp,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, p and k) used to fit the binomial distribution.
gl<-sum(ef>0)-3
gl
#Theoretical values of G test (using chi-squared distribution).
qchisq(0.95,gl)
#probability of the null hypothesis.
pchisq(G,gl, lower.tail = FALSE)
#in this case p>0.05, and null hypothesis can not be rejected.
#Conclusion: The data follows a binomial distribution.

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

```