Tutorial: functions for biocultural complexity

M. Humberto Reyes-Valdés

2/6/2019

This tutorial shows the use of the functions coded in the script *biocultural.R*, to analyzie biocultural complexity from a table of cultures and species usage. In this case, I use the file *toyset.csv*, which is an artificial simple data set. The platform and language is R.

1. Open an R console
2. Read the data set

```
dat<-read.csv("data/toyset.csv",head=T)
```

3. Visualize data features

```
dat
```

```
##   culture rep sp1 sp2 sp3 sp4 sp5
## 1      c1   1   0  25  39  27   9
## 2      c2   1 110  10   0   0   0
## 3      c3   1   0  26  24  28  10
## 4      c4   1   0  11 119   0   0
```

```
#rep is the representation frequency of cultural groups.
#In this case the assumption is that all groups are equally represented,
#and one can write 1 for each entry.
#culture is the column of cultural group names
#sp1 to sp5 are species names, with columns
#representing the observed field counts
str(dat)
```

```
## 'data.frame':    4 obs. of  7 variables:
##  $ culture: Factor w/ 4 levels "c1","c2","c3",..: 1 2 3 4
##  $ rep    : int  1 1 1 1
##  $ sp1    : int  0 110 0 0
```

```
##  $ sp2    : int  25 10 26 11
##  $ sp3    : int  39 0 24 119
##  $ sp4    : int  27 0 28 0
##  $ sp5    : int  9 0 10 0
```

4. Source script

```
source("biocultural.R")
```

```
##
## Attaching package: 'tidyr'
```

```
## The following object is masked from 'package:reshape2':
##
##     smiths
```

5. The function bc.info creates an object with basic estimated biocultural parameters

```
ob.sim<-bc.info(dat)
names(ob.sim)
```

```
##  [1] "cu"                "species"           "specificities"
##  [4] "r.specificities"   "s.diversity"       "specializations"
##  [7] "r.specializations" "mutualInfo"        "BC"
## [10] "r.BC"
```

6. A more complete and practical function is *bc.tables*, which performs bootstrap for bias correction, and estimates standard errors and confidence intervals. Definitions for table columns are in the respective paper (in review).

```
#Run bc.tables with 1000 bootstrap resamplings
set.seed(123)
x<-bc.tables(dat,1000)
names(x)
```

```
## [1] "tabSpecies"  "tabCultures" "tabBC"       "taBC_CI_1"   "taBC_CI_2"
```

7. Display table for species

```
library(knitr)
kable(x$tabSpecies)
```

Species	Specificity	BCorSpec	SE.S	R.Specificity	BCorRSpec	SE.RS
sp1	2.0000000	2.0000000	0.0000000	1.0000000	1.0000000	0.0000000
sp2	0.2165054	0.1902675	0.0807451	0.1082527	0.0951338	0.0403725

Species	Specificity	BCorSpec	SE.S	R.Specificity	BCorRSpec	SE.RS
sp3	0.6081559	0.6014965	0.0423920	0.3040780	0.3007482	0.0211960
sp4	1.0048459	0.9935196	0.0227816	0.5024230	0.4967598	0.0113908
sp5	1.0097404	0.9752105	0.0619537	0.5048702	0.4876052	0.0309769

8. Display table for cultural groups

```
kable(x$tabCultures)
```

Culture	Specialization	BCorSpecia	SE.S	R.Specialization	BCorRSpecia	SE.RS	S.Diversity	BCorSDiver	SE.SD
c1	0.6534922	0.6415892	0.0397434	0.3267461	0.3207946	0.0198717	1.8524724	1.8749198	0.0599532
c2	1.8513755	1.8497396	0.0510622	0.9256877	0.9248698	0.0255311	0.4138169	0.4220676	0.0919829
c3	0.6642951	0.6466176	0.0462059	0.3321476	0.3233088	0.0231029	1.9131163	1.9399870	0.0519947
c4	0.5750163	0.5677502	0.0459490	0.2875081	0.2838751	0.0229745	0.4182366	0.4232946	0.0873122

9. Display a table for Biocultural Complexity and Relative Biocultural Complexity , and another one with confidence intervals (see See Bryan FJ Manly, Randomization, Bootstrap and Monte Carlo methods in Biology, 2006, page 44).

```
#Biocultural Complexity and its relative counterpart
kable(x$tabBC)
```

BC	BCorBC	SE.BC	R.BC	BCorRBC	SE.RBC
1.913276	1.899936	0.0455588	0.4783189	0.4749841	0.0113897

```
#Confidence intervals
kable(x$taBC_CI_2)
```

BC_CL	BC_CH	RBC_CL	RBC_CH
1.810772	1.986325	0.452693	0.4965812

```
#Note: the attribute taBC_CI_1 contains a table with confidence intervals,
#based solely on quantiles.
#taBC_CI_2 gives intervals by the second method described by Manly (2006),
#which appears to work better for this application.
```

10. Perform a chi square test for association between cultures and species. A montecarlo simulation is used to get the p value.

```
chisq.test(dat[-c(1,2)],simulate.p.value=T)
```

```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  dat[-c(1, 2)]
## X-squared = 547.36, df = NA, p-value = 0.0004998
```