bor: An R Package for Transforming Behavioral Observation Records into Data Matrices

David N. Sousa[1] **and João R. Daniel**[1]

1 William James Center for Research, ISPA - Instituto Universitário

Summary

`bor` (Sousa & Daniel, 2018) is an R (R Core Team, 2016) package developed to transform raw focal sampling data of social interaction events into asymmetric data matrices (one for each type of social interaction recorded). For example, the number of times (events) a chimpanzee observes other group members. Focal sampling is a widely used sampling method in observational studies (Altmann, 1974,Lehner (1998),Martin, Bateson, & Bateson (1993)). When referring to social interactions, focal samples detail with whom the focal subject interacted with, the type of social interactions that took place, and who was responsible for initiating these interactions.

Converting hundreds of social behavior records, of who does what to whom, to data structures more appropriate for data analysis is a cumbersome process. Here, we present how `bor` can be used for wrangling social interactions' events into data matrices. In these matrices, cells provide counts on the number of times a specific type of social interaction was initiated by the row subject and directed to the column subject. Transforming focal sampling data into matrices allows users to easily use R base functions and other packages' functionalities to further analyze their data.

Example

The following installs and loads `bor` package:

```
# install bor from CRAN
install.packages("bor")

# load bor
library(bor)
```

The package includes three objects:

- `dtable` a function that tidies raw focal observations' data into a data frame more suitable for data processing.

- `countb` a function that computes asymmetric data matrices, separately for each observer and type of social interaction recorded, from a data frame with a structure similar to that of the output of `dtable` function.

- `ex_field_data` a data frame containing an example of raw focal observations' data (data frames passed to `dtable` function should have this structure).

Coding scheme and raw data structure

`ex_field_data` is an hypothetical data set that illustrates the required coding scheme and the structure of raw focal observations' data to be passed to the `dtable` function. This example contains 100 focal samples, recorded in a group of 9 subjects (coded s1 to s9).

```
head(ex_field_data)
```

```
#>   id1        act obs
#> 1  s2          x   1
#> 2  s3          0   1
#> 3  s5          x   1
#> 4  s1       +.s4   1
#> 5  s1          x   1
#> 6  s5  s4.-;s2.+   1
```

Here, **id1** refers to the focal subject's identification code, **act** refers to the recorded social interactions in each focal sample, and **obs** refers to the observer's identification code. In this example two different social interactions were recorded (coded "+" and "-"; social interactions' codes, subjects' id codes, and separation characters should not overlap). The "." character in **act** is used to separate a subject's id code from the social interaction code, and the ";" character is used to separate different social interactions occurring within the same focal sample (other characters can be used; please see the example below detailing the use of `dtable` function). Whenever a subject's id appears before a social interaction code it means that, that subject initiated a social interaction with the focal subject (see row 6) and whenever a subject's id appears after the social interaction code it means that the focal subject (id1 column) initiated a social interaction with that subject (see row 4).

For example, the 4th row indicates that the focal subject **s1** initiated a "+" social interaction with subject **s4**, while the 6th row indicates that the focal subject **s5** was involved in two social interactions- the first initiated by **s4** ("-"), the second ("+") initiated by **s2**. A "0" in this column refers to a focal sample where no social interactions were recorded, and an "x" refers to focal sample where the focal subject was unavailable for observation.

`dtable` function does not require that the input data frame has matching column names to that of `ex_field_data`. Nevertheless, the input data frame should include three columns, in the same order as above, containing the same information, and with the same type of coding scheme for social interactions (i.e., the subject id that interacted with the focal should appear before/after the social interaction code, if he was the initiator/target of the social interaction; focal subject's id is omitted in **act** column).

Tyding raw data

The following code uses `dtable` function to tidy `ex_field_data` data frame (see details above) in a new data frame (here, named data) that can be passed to `countb` function. `dtable` function requires that recorded social interactions' codes are provided (**bset** argument). In `ex_field_data` these codes are "+" and "-".

```
## specify the set of behaviors recorded and that will be used in bset argument
b <- c("+","-")
```

We leave all `dtable` remaining arguments: **bsep**, **asep**, **missing** and **noc** at their default values (".", ";", "x", "0" respectively). **bsep** and **asep** arguments refer to separation characters used in the coding scheme of social interactions. The first refers to the character (default is ".") used to separate a subject's id code from the social interaction code, and the second, to the character (default is ";") used to separate different social interactions

occurring within the same focal sample (see `ex_field_data` above). The **missing** argument (default is "x" character) refers to the character used to indicate that the focal subject was unavailable for observation, while the **noc** argument (default is "0" character) refers to the character used to indicate that no social interaction was recorded in the focal sample.

```
##  convert raw focal observations' data in ex_field_data
data <- dtable(ex_field_data, bset = b)
head(data)

   id1  id2 sender_id1 behavior no_occurrence missing observer
1  s2 <NA>         NA     <NA>            NA       1        1
2  s3 <NA>         NA     <NA>             1      NA        1
3  s5 <NA>         NA     <NA>            NA       1        1
4  s1   s4          1        +            NA      NA        1
5  s1 <NA>         NA     <NA>            NA       1        1
6  s5   s4          0        -            NA      NA        1
```

`data` object has 7 columns: **id1** is the focal subject's identification code; **id2** is the identification code of the social interaction partner; **sender_id1** indicates whether the focal subject was the initiator/sender (coded "1") or the target of the social interaction (coded "0"); **behavior** indicates the code of the social interaction recorded; **no_occurrence** indicates whether no social interactions were recorded (coded "1"; "NA" otherwise); **missing** indicates whether the focal subject was unavailable for observation (coded "1"}; "NA" otherwise); and **observer** is the observer's identification code.

Computing matrices

`countb` function can now be used on `data` to compute asymmetric data matrices (stored inside a list object), containing the number of times a specific type of social interaction was initiated by the row subject and directed to the column subject (target), separately for each social interaction and for each observer.

```
data2 <- countb(data)
data2

$`1`
$`1`$`-`
   s1 s2 s3 s4 s5
s1  0  1  0  0  2
s2  0  0  0  0  1
s3  0  0  0  1  0
s4  0  0  1  0  1
s5  0  1  1  1  0

$`1`$`+`
   s1 s2 s3 s4 s5
s1  0  1  1  1  0
s2  0  0  2  0  2
s3  2  0  0  0  0
s4  0  1  0  0  0
s5  0  0  0  1  0

$`2`
$`2`$`-`
   s1 s2 s3 s4 s5
```

Sousa et al., (2018). bor: An R Package for Transforming Behavioral Observation Records into Data Matrices. *Journal of Open Source Software*, 3(28), 909. https://doi.org/10.21105/joss.00909

```
s1  0  0  1  2  0
s2  0  0  0  3  0
s3  0  1  0  1  0
s4  0  2  1  0  1
s5  1  1  1  0  0

$`2`$`+`
    s1 s2 s3 s4 s5
s1  0  0  1  0  0
s2  0  0  0  2  0
s3  0  1  0  2  1
s4  0  0  1  0  0
s5  1  0  1  0  0
```

data2 is a list of lists: one for each observer (in this example `data2\$`1`` and `data2\$`2``).
Inside each list there is one asymmetric interaction matrix per social interaction recorded
(e.g., `data2\$`1`\$`-``, `data2\$`1`\$`+``).

Acknowledgements

References

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*,
49(3), 227–266.

Lehner, P. N. (1998). *Handbook of ethological methods*. Cambridge University Press.

Martin, P., Bateson, P. P. G., & Bateson, P. (1993). *Measuring behaviour: An introductory
guide*. Cambridge University Press.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna,
Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.
org/

Sousa, D. N., & Daniel, J. R. (2018). *Bor: Transforming behavioral observation records
into data matrices*.