

Analysis of activity kernels

David

6/30/2021

Overview

The point of this document is to look a little more in depth at an appropriate method to determine if the similarity between two activity kernels is statistically significant. I'll look at a couple different options:

- 1) Watson's U2 statistic, which seems to be the method of choice for papers that have compared activity kernels from camera trap data, and
- 2) `compareKern`: the randomization test from the `activity` package.

Both options use an asymptotic distribution of a test statistic to test the null hypothesis that the test statistic could have arose by random chance alone; i.e., that if the test was repeated a large number of times, you'd have a probability (the p value) to get a test statistic as or more extreme than the test statistic.

The `activity::compareKern` function generates a test statistic which is the overlap index (\hat{D}_{4}). This is the area under the curve that is formed by taking the minimum of two density functions. No overlap would be 0 and perfect overlap would be 1. Then a null distribution of overlap indices is estimated by resampling with replacement from both kernels combined. This randomised distribution is then used to define an empirical probability distribution against which the probability that the observed overlap arose by chance is judged.

The Watson's U2 stat, aka Watson's two-sample test of homogeneity, uses the sum of squares of deviations between the sample distribution functions as test statistics. The p-value is derived from the likelihood ratio test that if the null is supported by the observed data, the two likelihoods shouldn't differ by more than the sampling error. To determine statistical significance you can specify what alpha level of significance to test for. There isn't a ton of description in the documentation of the function in the `circStats` package, but it appears to also use an empirical cdf to test whether the test statistic arose by chance alone.

Bottom line, my opinion is that both methods are statistically valid and get us the info we're looking for. They seem to be either very similar or actually identical. It would take more digging to find out and that doesn't seem like a great use of time.

A code example

```
library(activity)

## Warning: package 'activity' was built under R version 4.0.5
library(CircStats)

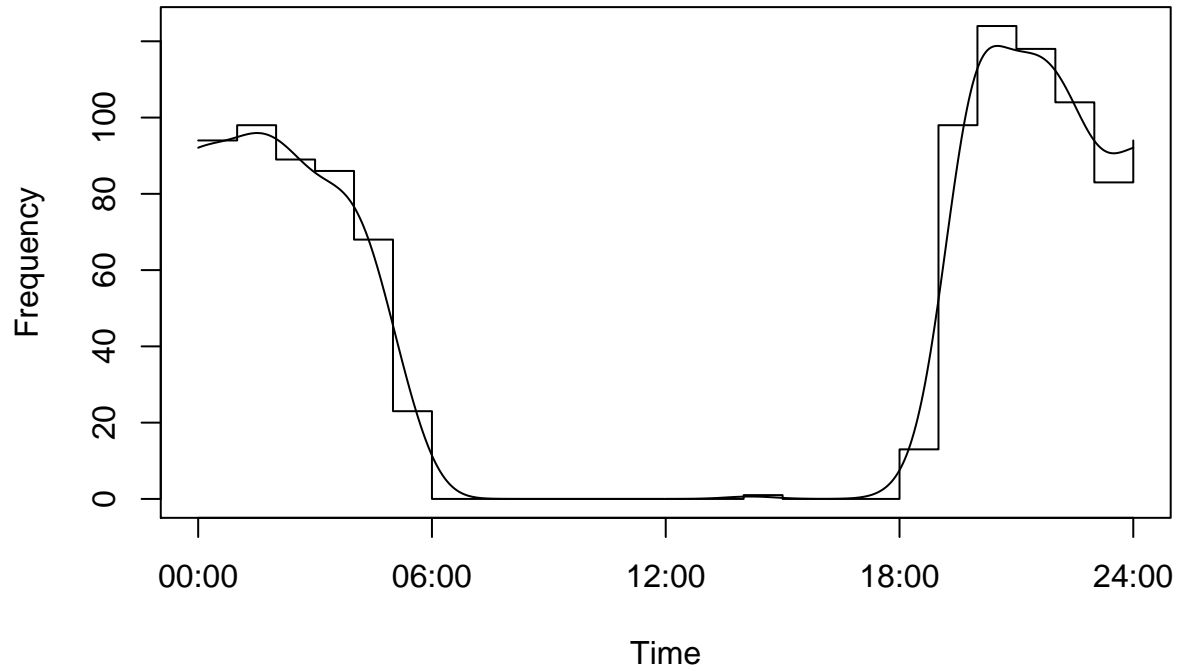
## Warning: package 'CircStats' was built under R version 4.0.5
## Loading required package: MASS
## Loading required package: boot
data(BCItime)
```

```

tPaca <- 2*pi*BCitime$time[BCitime$species=="paca"]
tRat <- 2*pi*BCitime$time[BCitime$species=="rat"]

fPaca <- fitact(tPaca)
fRat <- fitact(tRat)
plot(fPaca)

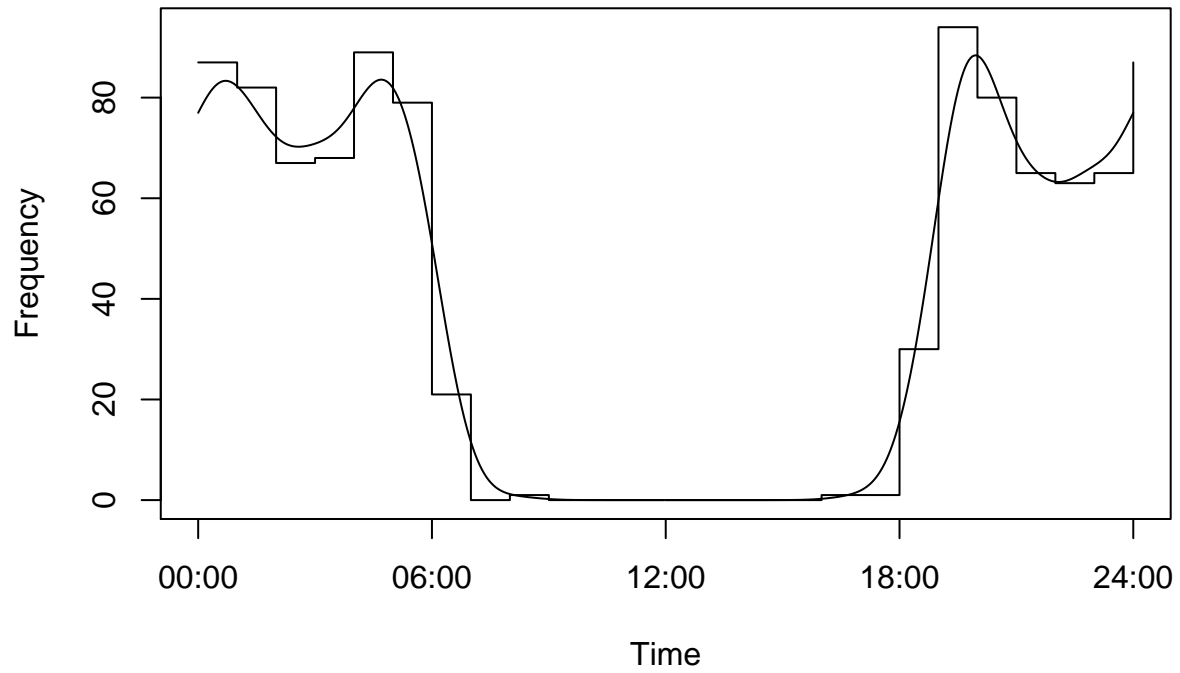
```



```

plot(fRat)

```



Randomization test

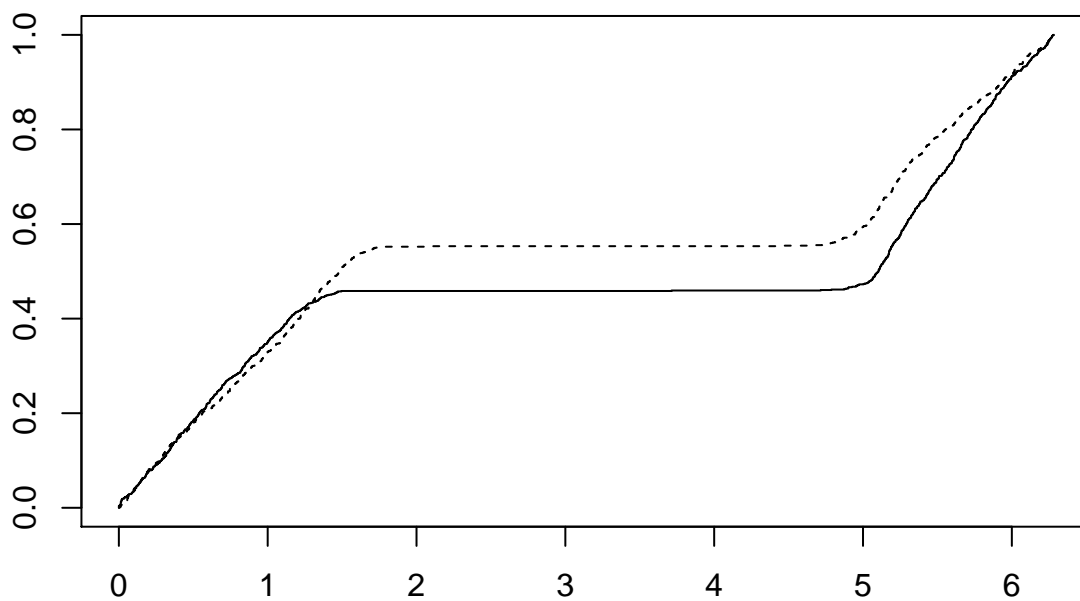
```
compareCkern(fPaca,fRat, reps=10)
```

```
##          obs          null          seNull          pNull
## 0.863063731 0.960661082 0.007081724 0.000000000
```

Watson's two-sample test for homogeneity

```
watson.two(tPaca, tRat, alpha=0.05, plot=TRUE)
```

Comparison of Empirical CDFs



```
##
##      Watson's Two-Sample Test of Homogeneity
##
## Test Statistic: 1.3207
## Level 0.05 Critical Value: 0.187
## Reject Null Hypothesis
##
```

```
watson.two(tPaca, tRat)
```

```
##
##      Watson's Two-Sample Test of Homogeneity
##
## Test Statistic: 1.3207
## P-value < 0.001
##
```