Low-Order Analysis of Discrete Behavioral Sequence Data

Ben de Bivort
Danylo Lavrentovich
JAX Workshop Oct 12 2022

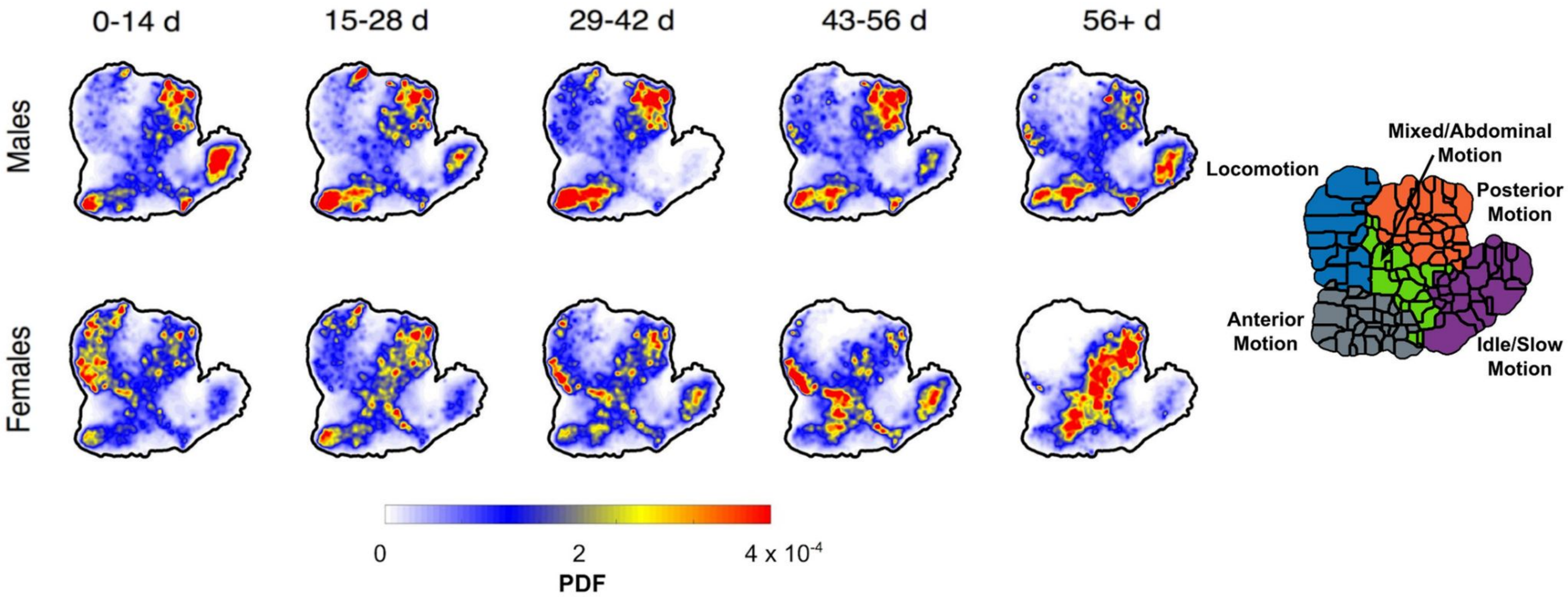
What are we doing in today’s workshop?

Given time-indexed sequences of discrete behavioral states (potentially the output of MoSeq or MotionMapper):

animal 1	[1, 1, 1, 0, 0, 57, 44, 44, 44, 44, 96, 96, 0, …]
animal N	[2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, 25, 14, 105, 105, …]

- Zero-order analysis:
  - What are the abundances of each state?
  - How do two groups of animals differ in state abundances?
- First-order analysis:
  - What do the transitions from one state to another look like?
  - How well does a basic first-order model, a Markov process, explain our data?
- Higher-order analysis:
  - Scott’s workshop right after this one – latent states underlying behaviors

Case example: Overman, …, Berman PLOS Comp Bio 2022



But you can also use your own time-indexed discretized behavior data if you format it appropriately

Notebook logistics

- The emoji code:
    - 🏃: run this code without editing it
    - 🤔: go through this code carefully, editing variables above the ######################
    - ❓: questions at the end of 🤔 sections for you to think about, talk with a partner, and go back into the code for toggling/investigation
- We'll periodically come together as a group for discussion
    - 🛑: stop signs in the notebook
- Be careful not to skip running a cell! Some are linked to each other
    - Run cells by hitting shift + enter, or the play button on the left hand side
- Left side table of contents panel shows you where you are in the notebook

Student’s t test

- What is the hypothesis being tested?
  - The null hypothesis that there isn't a mean difference between two groups
- What is a p value?
  - In frequentist statistics, the frequency of events more extreme than the event of interest, under the null hypothesis
- What are assumptions for the t test?
  - Distributions have to be normal
  - Variance of both groups is equal to that of the null distribution

Bootstrapping

- What is bootstrapping?
  - Method for performing frequentist test of any distribution
- What are the assumptions?
  - That the true distribution that produced our data is literally the observed data distribution
- How do you do it?
  - Sample the observed data *with replacement,* up to the size of the data
  - Compute statistic of interest
  - Repeat the experiment, sampling the observed data with replacement again
  - Examine variation of your statistic across bootstrap experiment replicates
    - standard deviation of resampled statistic is the standard error of the statistic
    - you may be familiar with the parametric standard error of the mean: sigma / sqrt(n)

Markov models

- Basic first-order model of discrete states in time
  - States need to be discrete, could potentially be outputs from MoSeq / MotionMapper
  - Time can be continuous or discrete (for the case example data, it'll be discrete)
- Probabilities of future transitions don't depend on any past data, only the current state of the system
  - $P(X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \ldots, X_0) = P(X_{t+1} \mid X_t)$
- What will we do today?
  - Take our data and infer transition probabilities by counting state-state transitions
  - Fit a Markov model to the transition matrix, sample from it to generate a Markov sequence
  - Make comparisons:
    - How does the observed sequence of states compare to the Markov sequence?
    - Do large step-size transitions in the data follow the Markov prediction (powers of the single-step transition matrix)?
    - Do dwell times in the observed sequence follow the Markov predicted exponential distribution?

Dwell times are exponentially distributed in Markov processes

- $P(X > t + s \mid X > t) = P(X > s)$
  - Memoryless property
  - prob. of observing an event at 15 minutes given you've waited 10 minutes = prob. of observing an event in 5 minutes given you're at time 0
- $P(X > t + s \mid X > t)\, P(X > t) = P(X > s)\, P(X > t)$
  - multiply both sides by $P(X > t)$
- $P(X > t + s) = P(X > s)\, P(X > t)$
  - definition of conditional probability
- Let $G(t) = P(X > t)$. Substituting into the above:
- $G(t+s) = G(s)\, G(t)$
  - what functions have this property? Exponential family functions: $G(a) = e^{\Lambda a} = P(X > t)$