**Analysis of Armadillo Acceleration Data**

Armadillo acceleration (measured as activity counts per recording interval; 5 min) was analyzed in bayesmove using the segmentation and clustering (LDA) procedures.

*Data preparation*

These data were first cleaned up to remove all rows with missing data for “Activity Counts” and duplicate observations. Observations were also filtered to only retain observations at the primary time interval (5 min), which only resulted in the removal of 5 observations. This left 140,161 observations from 7 individuals for analysis by the model.

Since the data (Activity Counts) were highly right-skewed, quantiles were used to bin this variable. Specifically, I discretized the data into 6 bins using the 0, 0.50, 0.75, 0.80, 0.85, 0.95, and 1 quantiles. This resulted in bin widths that increased in size appeared to capture natural breaks in biologically relevant movement. On the scale of Activity Counts, these bin limits were 0, 1, 12, 36, 96, 200, and 300 counts of deviation from the previous values of a triaxial accelerometer.

A picture containing table, computer

Description automatically generated

A picture containing screenshot

Description automatically generated

*Segmentation model*

Based on traceplots of the log marginal likelihood for each individual, the model appeared to converge after 15000 iterations.

Although the segmentation results (as heatmaps) are not shown here, results appeared to match relatively well with the data. However, it was also difficult to inspect these results graphically since time series for each individual armadillo was ranged from 10,000 to 30,000 observations with 200 to 300 breakpoints.

*LDA model*

Following the clustering of the segments of Activity Counts across all individuals, plots were inspected to determine the optimal number of behavioral states as well as the shapes of the Activity Count distributions.

A picture containing clock

Description automatically generated

Based on this boxplot of results from the theta matrix, the first 3 states/clusters accounted for 95.4% of all observations on average. Therefore, these clusters were retained for further analysis.

A picture containing screenshot

Description automatically generated

Plots of the bin distributions for each of the 5 possible states (where the first 3 are shown in color) indicate that these first 3 states are biologically distinct and relevant. Based on these distributions, state 1 appears to represent “High Activity”, state 2 represents “Moderate Activity”, and state 3 represents “Low Activity”.

Time series plots of these behavior proportions for each individual armadillo are shown next and clearly depict regular intervals of higher and lower activity. The long periods in the middle of the time series with minimal change in behavior proportion is likely an artifact of the visualization function used, meaning that there are probably large gaps of missing data during these periods.

A picture containing implement, pencil, bookshelf

Description automatically generated

*Conclusions*

Based on these findings, it appears that our model performs well at characterizing relative activity levels from acceleration Activity Counts. We would need to consult further with Nina to determine if this matches with her intuition and what she has found from her previous analysis of this data.

A potential issue with this dataset is that the geographic coordinates are missing for these observations. While they may be present in a separate dataset from what I have, it makes it impossible for us to do any spatial analyses, connectivity/resistance, and resource selection analyses. However, we could characterize temporal patterns and activity budgets based on these data. Additionally, we could make comparisons among individuals based on certain intrinsic traits (size, sex, maturity, etc).