# markov_after_comments_parsed

February 5, 2021

## 1 Markov after comments

### 1.1 States and transition probabilities

The transition matrices on the previous report were calculated with transitions between "Greater" and "Smaller" states, **but also from a null state**, which is the first track of the album. This is how I calculated the likelyhood of track 2 given track 1.

The final empirical transition matrices are displayed below, and it shows the probability of transition between a row to a column (e.g. prob of going from any state to the null state (i.e. start) is 0, because only the first track is in state "start").

A matrix: $3 \times 3$ of type dbl

| | greater | smaller | start |
|---|---|---|---|
| greater | 0.3205376 | 0.6794624 | 0 |
| smaller | 0.6677269 | 0.3322731 | 0 |
| start | 0.5336611 | 0.4663389 | 0 |

A matrix: $3 \times 3$ of type dbl

| | greater | smaller | start |
|---|---|---|---|
| greater | 0.3118086 | 0.6881914 | 0 |
| smaller | 0.6616254 | 0.3383746 | 0 |
| start | 0.5075643 | 0.4924357 | 0 |

A matrix: $3 \times 3$ of type dbl

| | greater | smaller | start |
|---|---|---|---|
| greater | 0.3223561 | 0.6776439 | 0 |
| smaller | 0.6515744 | 0.3484256 | 0 |
| start | 0.5279879 | 0.4720121 | 0 |

A matrix: 3 × 3 of type dbl

| | greater | smaller | start |
|---|---|---|---|
| greater | 0.3328422 | 0.6671578 | 0 |
| smaller | 0.6619533 | 0.3380467 | 0 |
| start | 0.5075643 | 0.4924357 | 0 |

### 1.2 State distributions

Here I computed the overal frequency of each state, State distribution seems to be stable accross different features, which strengthens the empirically derived transition matrices presented in section 1.1 (e.g. the transition between "smaller" to "smaller" in valence is not due to a generally low ammount of "smaller" states in the dataset - see frequency tables below)

A data.table: 3 × 4

| | valence <dbl> | energy <dbl> | loudnes <dbl> | tempo <dbl> |
|---|---|---|---|---|
| Greater | 0.4551 | 0.4493 | 0.4495 | 0.4495 |
| Smaller | 0.4600 | 0.4658 | 0.4656 | 0.4656 |
| Start | 0.0849 | 0.0849 | 0.0849 | 0.0849 |

## 2 Up-down by section (New report)

### 2.1 Method 1

- Each album was divided into 3 sections;
- Within each section and each album, features received a tag of "greater" or "smaller".
- The tags "greater" or "smaller" related to the overall (mean) value of the feature within that album/track section.
- For instance, given album $A$, if the valence of track_1 was 10, and the mean valence in position_1 was 15, the track received the tag "smaller". A few lines from the dataset are displayed below:

A tibble: 10 × 5

| track_number <int> | position <chr> | valence <dbl> | overall_valence <dbl> | valence_cat <chr> |
|---|---|---|---|---|
| 1 | 1st | 0.0673 | 0.1645750 | smaller |
| 2 | 1st | 0.2010 | 0.1645750 | greater |
| 3 | 1st | 0.1770 | 0.1645750 | greater |
| 4 | 1st | 0.2130 | 0.1645750 | greater |
| 5 | 2nd | 0.5800 | 0.3900000 | greater |
| 6 | 2nd | 0.5310 | 0.3900000 | greater |
| 7 | 2nd | 0.2800 | 0.3900000 | smaller |
| 8 | 2nd | 0.1690 | 0.3900000 | smaller |
| 9 | 3d | 0.0816 | 0.1798667 | smaller |
| 10 | 3d | 0.3130 | 0.1798667 | greater |

### 2.2 Method 2

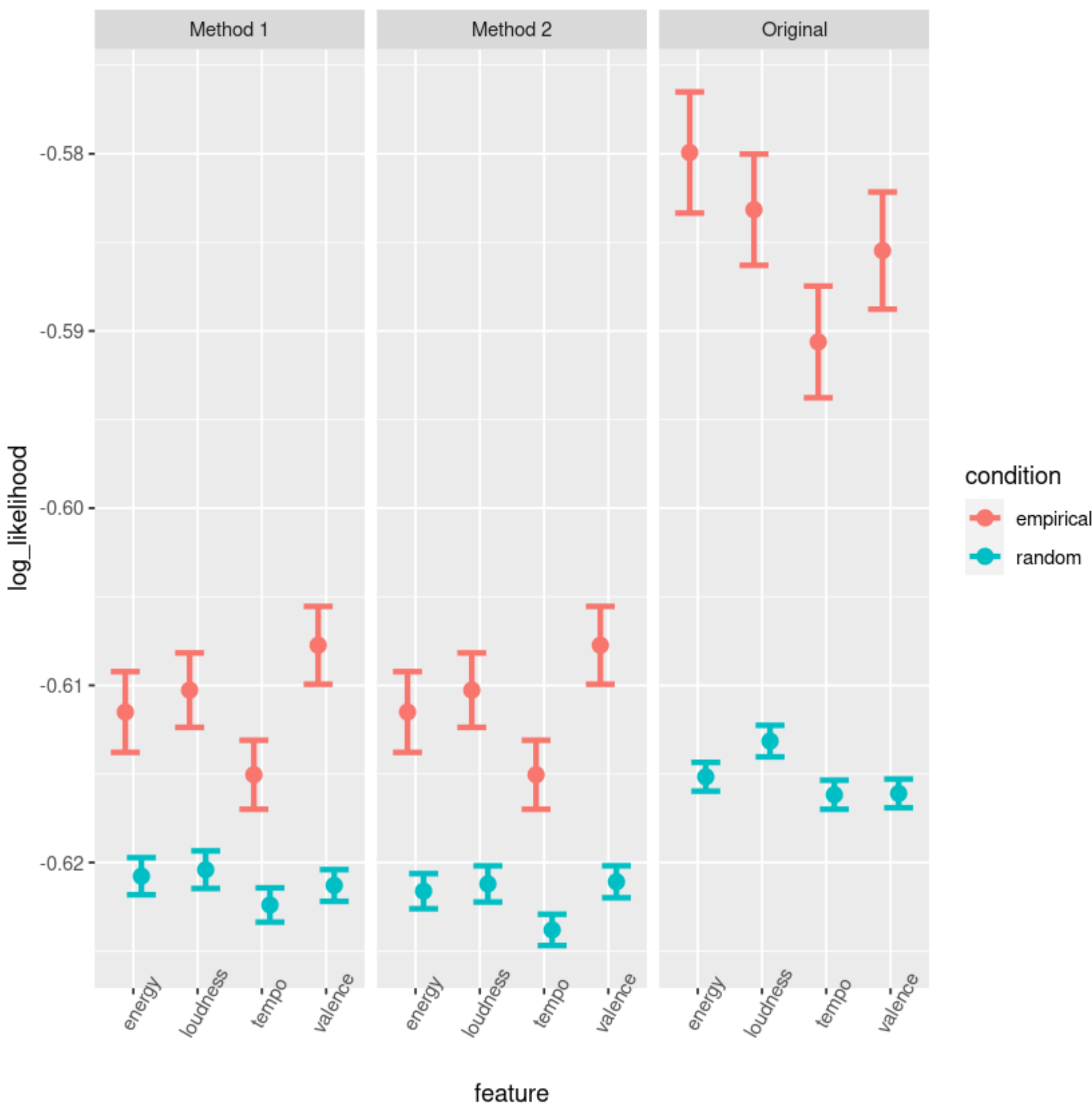
Same thing as method 1, but colapsing each album in 4 sections.

### 2.3 Evaluation

Here I'm comparing the difference in log-likelihood between the different methods of "feature tagging":

- Method 1: separating the album in 3 sections and comparing each track with the next section
- Method 2: separating the album in 4 sections and comparing each track with the next section
- Original: comparing each track with its nearest neighbors

Again, the statistics below show the likelyhood of the data given the transition matrices fitted on 1) the original dataset, and 2) a shuffled version of the same data set. I computed the mean log likelyhood for each album, and standard errors between them.



It seems like the difference in likelyhood between the model fitted on the permutated dataset is largest when we use the 1st method. Perhaps we can try some different grouping methods.