Developing a potential biomarker for mood prediction using typing dynamics from smartphone data collected in-the-wild

Homa Rashidisabet[1], John Zulueta[1], Jonathan Stange[1], Faraz Hussain[1], Scott A. Langenecker[2], Shannon Young[3], Larsson Omberg[3], Peter Nelson[1], Raeanne C. Moore[4], Olusola Ajilore[1], Alex Leow[1], Alexander P. Demos[1]

1. University of Illinois at Chicago
2. University of Utah
3. Sage Bionetworks
4. University of California San Diego

Research questions

1. Does the typing dynamics follow a power law (PL) distribution?
2. Is the PL scaling exponent a reliable feature intra- and inter-participants?

Background

Why typing dynamics is important?

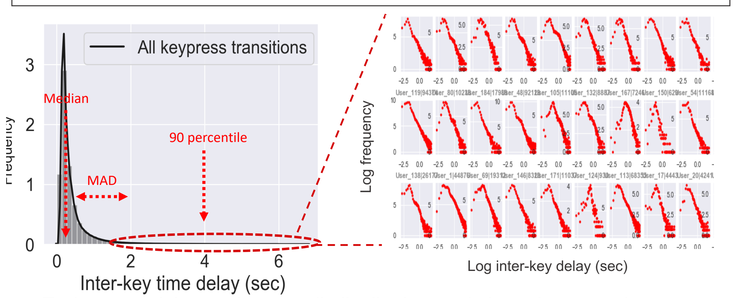
- Prominent typing phenotype that can differentiate mood state changes, predict affective states, discriminate between various emotional states

Why the interest in PL distribution?

- PL behavior is observed in human motor control, locomotor activity, and subserving cognitive processes
- PL exponent has been used to differentiate patients with mental disorders and healthy control. It has been also shown that it degrades with aging and under various pathological conditions.

Why reliability analysis is essential?

- Reliability analysis allows for confidence in the stability of this measure collected, reduce the change to draw false inferences and spurious interpretations



Typing speed has a heavy tail distribution and there is a consistent straight-line form in the log-log plot of all participants, suggesting PL.

Data and Participants

1. Typing dynamics data in this study is collected passively and unobtrusively via a mobile health application, named BiAffect.
2. Data were collected within the first 18 months of our ongoing open-science BiAffect study.

Sample Size	Gender			Age (M, +/-SD)	Min # of data
	Female	Male	Non-binary		
256	74.2%	22.6%	2.4%	34.30 (+/- 10.97)	4,056
86 *	72.1%	25.6%	2.3%	39.19 (+/- 11.40)	4,624
42**	76.2%	19.0%	4.8%	40.0 (+/- 12.30)	11,160

* Subsample of the n = 256 participants, ** subsample of the 86 participants.

Methods

We developed a framework to estimate PL distributions reliably for empirical data by adopting rules and modifications to existing approaches. Our method is as follows:

1) Bootstrap PL-fit per participant

2) Apply clustering methods to bootstrapped PL estimates per participant

3) Finding a consistent threshold across participants to identify clusters of stable and unstable estimates

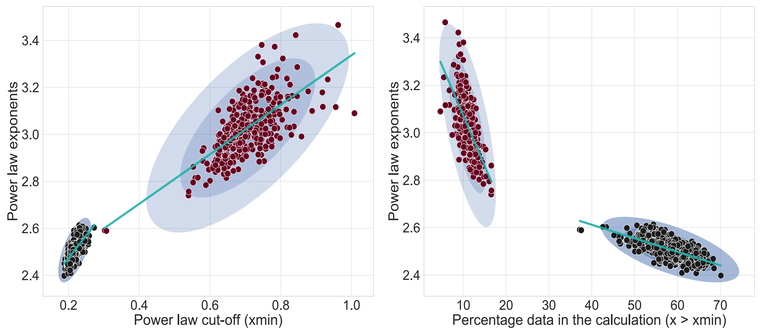
4) Employing our proposed "rejection rule" to filter bootstrapped exponents associated with unstable cluster

5) Finding the minimum required data volume for reliable estimation of PL distribution

6) Measuring inter- and intra-participant reliability using Pearson correlations

7) Comparing the goodness-of-fit of PL against competing distributions (exponential, Weibull distribution)

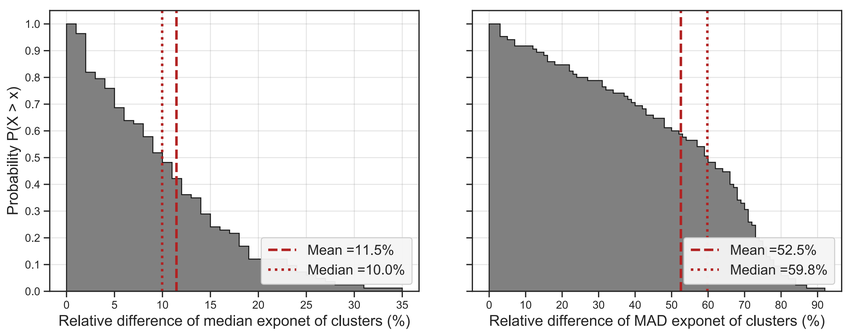
Results

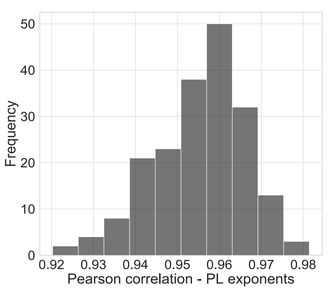


- Consistently across participants, there is a cluster pattern between bootstrapped PL parameters and the volume of data in calculations
- When there is a low number of data on the tail, the PL exponent will be greater and PL behavior starts at a higher cut-off value
- Averaged across participants, we found that unstable cluster (in red) is associated with 41% fewer data in calculations after controlling for the number of keypresses (n = 4,000) in calculations

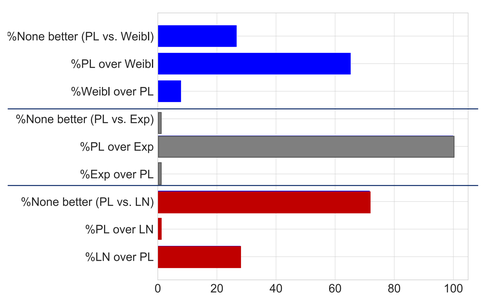
Complementary CDF of the relative difference of median and MAD between unstable and stable clusters supports that:

- For 50% of participants, the MAD and median of unstable clusters are respectively 60% and 10% higher than the MAD and median of the stable clusters
- Estimates in the unstable clusters are variable due to lack of data in calculations and they need to be filtered for a reliable analysis





- Histogram of Pearson correlations as a measure of inter-participant reliability supports that PL scaling exponent is a reliable feature inter-participants with an average correlation 0.95 (SD = 0.01)
- Based on the percentage of bootstrapped likelihood ratios, we found that PL distribution best represents human typing dynamics relative to competing distributions (exponential, Weibull distribution)



Conclusions

1. We showed that PL distribution, among competing heavy-tail distributions, is a plausible model for representing typing dynamics
2. We demonstrated that PL exponent is a reliable feature intra- and inter-participants

Future directions

When a larger sample will be available, we would like to replicate whether PL exponent can differentiate between controls and participants with mental disorders using typing dynamics data