Analysing inter-key intervals: Beyond means, medians and pause frequencies

Jens Roeser   Sven De Maeyer   Mark Torrance

Mariëlle Leijten   Luuk Van Waes

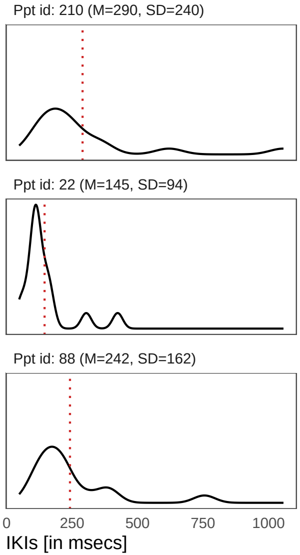
jens.roeser@ntu.ac.uk

SIG 27 Conference
University of Antwerp

Dec 15, 2020

The problem: what's a pause?

- Keystroke data are heavily skewed.
- Skew reflects cognitive processes.
- How can we distinguish between fluent and disfluent key transitions?
- Fixed thresholds: 0.5 or 2 secs?
- A key transition normal for learner might be longer than a pause of an experienced writer.



The problem: what's a pause?

- Keystroke data are heavily skewed.
- Skew reflects cognitive processes.
- How can we distinguish between fluent and disfluent key transitions?
- Fixed thresholds: 0.5 or 2 secs?
- A key transition normal for learner might be longer than a pause of an experienced writer.

- Pause size depends on:
  - writing skills / style
  - position in text, sentence, word
  - experience with target language (in L2)
  - process of interest (lexical, motor, semantic)
  - writing task
  - …

Research focus

- How do we deal with the heavy tail statistically without losing information, trimming data, imposing pause thresholds?
- …and extract pause frequencies in a principled way?
- Mixture of log-normals (Almond et al., 2012; Baaijen et al., 2012)

→ More than one underlying data-generating process; i.e. normal key transitions and disfluencies.

- Compared to standard log-normal treatment.

→ One underlying data-generating process.

Method

- Implementation of the copy-typing process as statistical model in Stan (Carpenter et al., 2016); code based on Sorensen et al. (2016) and Vasishth et al. (2017).
- Bigram-keystroke interval data: Dutch subset of copy-task corpus (Van Waes et al., 2019; Van Waes et al., 2020)
- Lexical and non-lexical copy-typing context

Consonants task

tjxgfl pgkfkq dtdrgt npwdvf

LF-bigrams task

een chaotische cowboy

LF-bigrams task

een chaotische cowboy

$$\Downarrow$$

e^e^n   c^h^a^o^t^i^s^c^h^e   c^o^w^b^o^y

LF-bigrams task

een chaotische cowboy

$$\Downarrow$$

e^e^n   c^h^a^o^t^i^s^c^h^e   c^o^w^b^o^y

$$\Downarrow$$

162 97 107 141 800 148 278 132 199 94 154 177 870 88 274 611

LF-bigrams task

een chaotische cowboy

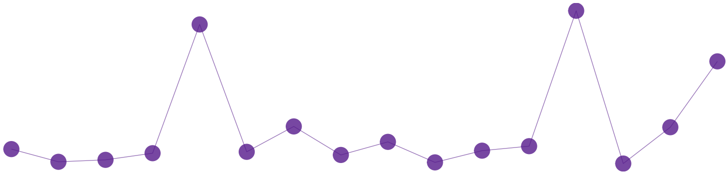
$\Downarrow$

e^e^n c^h^a^o^t^i^s^c^h^e c^o^w^b^o^y

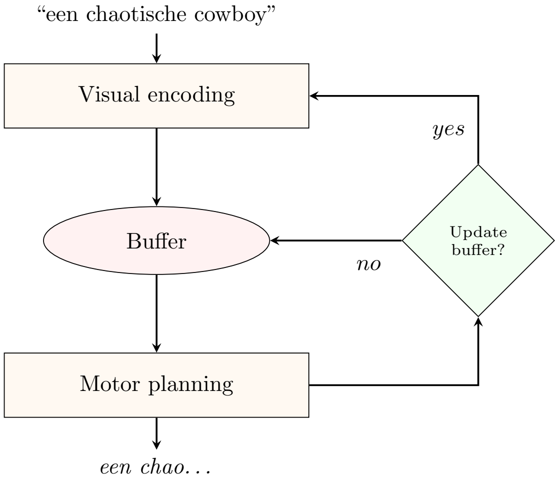
$\Downarrow$

162 97 107 141 800 148 278 132 199 94 154 177 870 88 274 611

$\Downarrow$

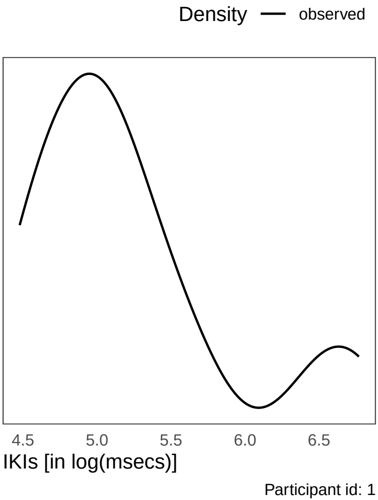


Disfluencies in a basic model of copy-typing



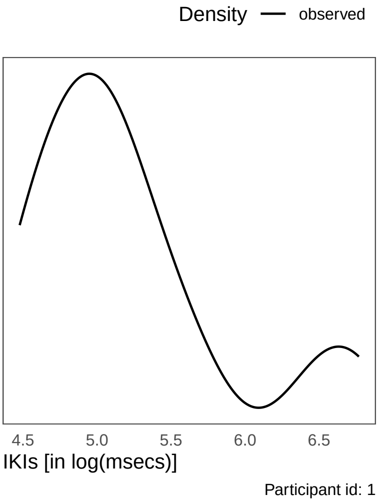
Mixed-Effects Model

$$y_{ij} \sim LogNormal(\mu_{ij}, \sigma_e^2)$$



Mixed-Effects Model

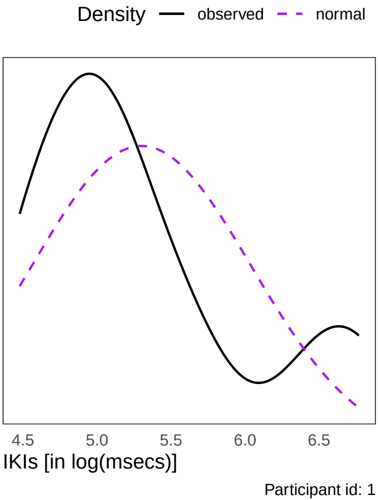
$$y_{ij} \sim LogNormal(\mu_{ij}, \sigma_e^2)$$
$$\mu_{ij} = \alpha + u_i + w_j$$



Mixed-Effects Model

$$y_{ij} \sim LogNormal(\mu_{ij}, \sigma_e^2)$$
$$\mu_{ij} = \alpha + u_i + w_j$$

- α: average IKI
- σ_e^2: error variance
- Participants: u_i
- Bigrams: w_j



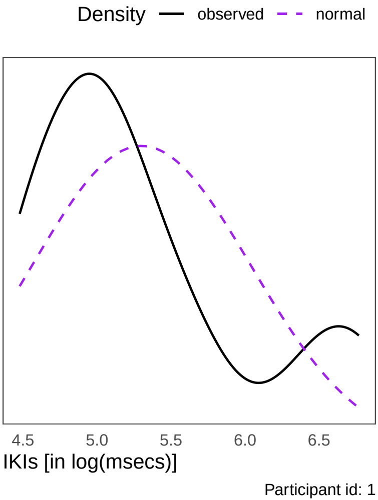
Mixed-Effects Model

$$y_{ij} \sim LogNormal(\mu_{ij}, \sigma_e^2)$$
$$\mu_{ij} = \alpha + u_i + w_j$$

- α: average IKI
- σ_e^2: error variance
- Participants: u_i
- Bigrams: w_j

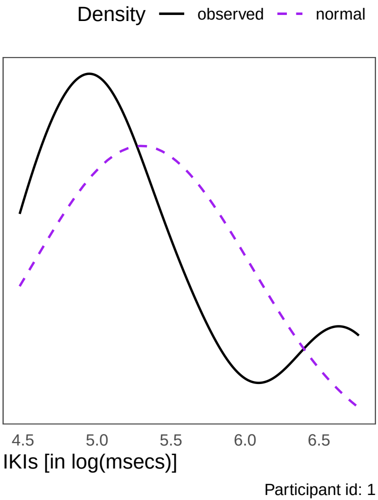
Extending mixed models to finite mixtures

$$y \sim LogNormal(\mu, \sigma_e^2)$$



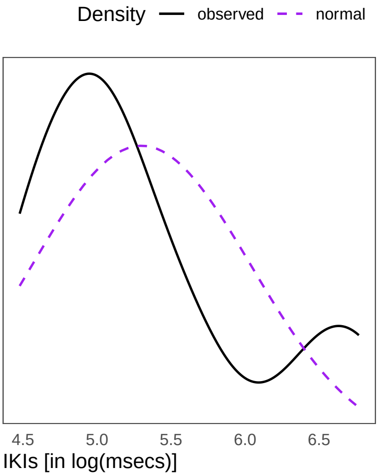
Extending mixed models to finite mixtures

$$y \sim LogNormal(\mu, \sigma_e^2)$$



Extending mixed models to finite mixtures

$$y \sim \boxed{\theta} \cdot LogNormal(\mu, \sigma_e^2)$$
$$\theta = 1$$

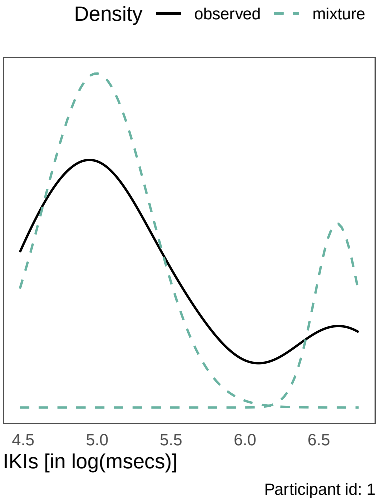


Participant id: 1

Extending mixed models to finite mixtures

$$y \sim \theta \cdot LogNormal(\mu_1, \sigma^2_{e_1}) + (1 - \theta) \cdot LogNormal(\mu_2, \sigma^2_{e_2})$$

$$\theta = 0.81$$

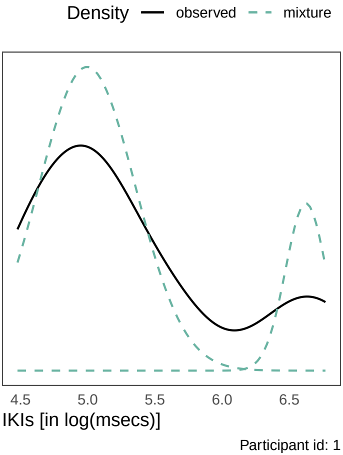


Finite mixture of two log-Gaussians

$$y_{ij} \sim \theta_i \cdot LogNormal(\mu_{ij} + \delta, \sigma^2_{e'}) + (1 - \theta_i) \cdot LogNormal(\mu_{ij}, \sigma^2_e)$$

$$\mu_{ij} = \alpha + u_i + w_j$$

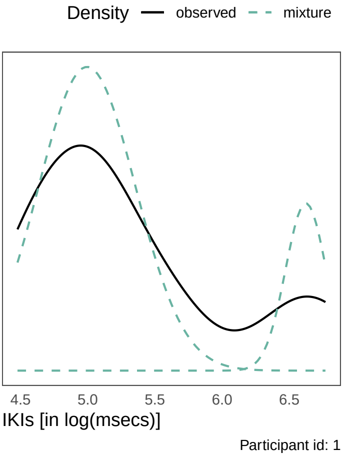
- α: fluent IKI
- δ: slowdown
- θ: disfluency probability
- $\sigma^2_{e'}$: larger variance



Finite mixture of two log-Gaussians

$$y_{ij} \sim \theta_i \cdot LogNormal(\mu_{ij} + \delta, \sigma^2_{e'}) +$$
$$(1 - \theta_i) \cdot \boxed{LogNormal(\mu_{ij}, \sigma^2_e)}$$
$$\mu_{ij} = \alpha + u_i + w_j$$

- α: fluent IKI
- δ: slowdown
- θ: disfluency probability
- $\sigma^2_{e'}$: larger variance

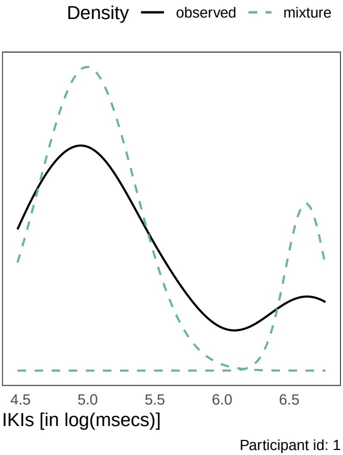


Finite mixture of two log-Gaussians

$$y_{ij} \sim \theta_i \cdot \boxed{LogNormal(\mu_{ij} + \delta, \sigma^2_{e'})} + (1 - \theta_i) \cdot LogNormal(\mu_{ij}, \sigma^2_e)$$

$$\mu_{ij} = \alpha + u_i + w_j$$

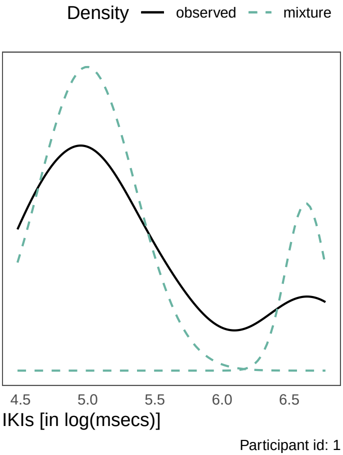
- α: fluent IKI
- δ: slowdown
- θ: disfluency probability
- $\sigma^2_{e'}$: larger variance



Finite mixture of two log-Gaussians

$$
\begin{aligned}
y_{ij} \sim \boxed{\theta_i} \cdot LogNormal(\mu_{ij} + \delta, \sigma^2_{e'}) + \\
(1 - \theta_i) \cdot LogNormal(\mu_{ij}, \sigma^2_e) \\
\mu_{ij} = \alpha + u_i + w_j
\end{aligned}
$$

- α: fluent IKI
- δ: slowdown
- θ: disfluency probability
- $\sigma^2_{e'}$: larger variance



Model comparisons

Predictive performance estimated as the *expected log predictive density* ($\widehat{elpd}$) (Vehtari et al., 2015, 2017). Models are ordered by predictive performance (model with highest predictive performance in top row). Standard error in parentheses.

		Consonants task		LF-bigrams task	
Models	Distribution	$\Delta\widehat{elpd}$	$\widehat{elpd}$	$\Delta\widehat{elpd}$	$\widehat{elpd}$
MoG	$2 \times$ Log-normal				
LMM	Log-normal				

Note. LMM = Linear mixed effects model; MoG = Mixture of Gaussians

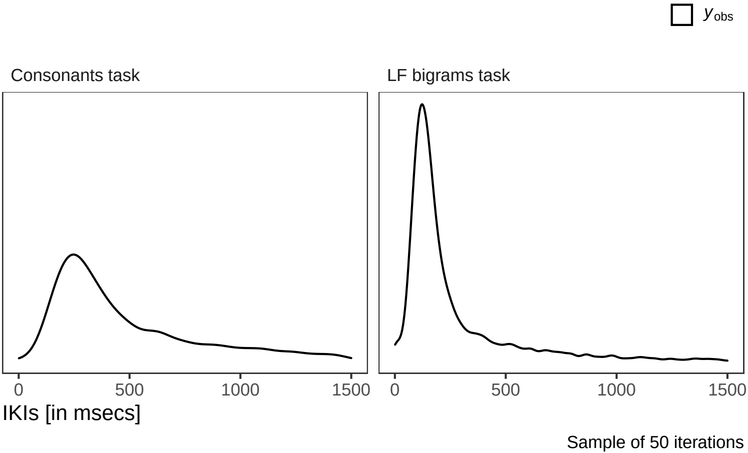
Model comparisons

Predictive performance estimated as the *expected log predictive density* ($\widehat{elpd}$) (Vehtari et al., 2015, 2017). Models are ordered by predictive performance (model with highest predictive performance in top row). Standard error in parentheses.

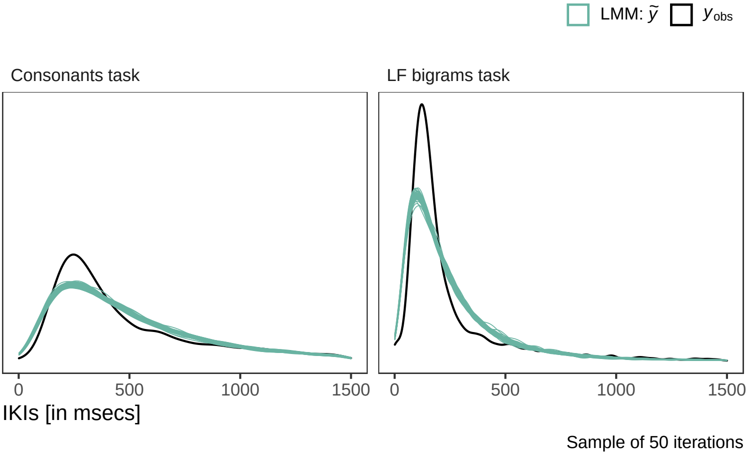
		Consonants task		LF-bigrams task	
Models	Distribution	$\Delta\widehat{elpd}$	$\widehat{elpd}$	$\Delta\widehat{elpd}$	$\widehat{elpd}$
MoG	2 × Log-normal	–	-37,069 (101)	–	-33,178 (113)
LMM	Log-normal	-281 (25)	-37,350 (99)	-994 (63)	-34,173 (121)

Note. LMM = Linear mixed effects model; MoG = Mixture of Gaussians

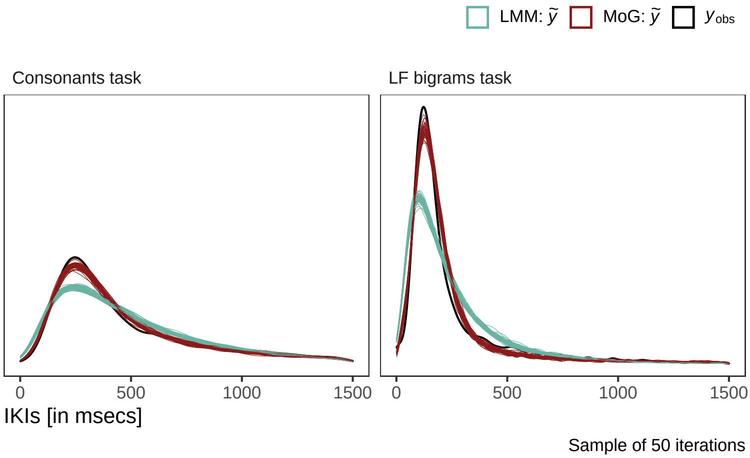
Observed vs. predicted IKIs



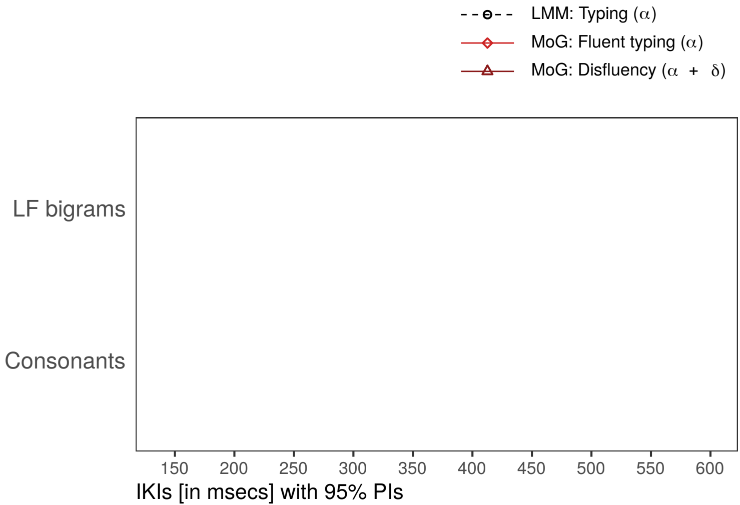
Observed vs. predicted IKIs



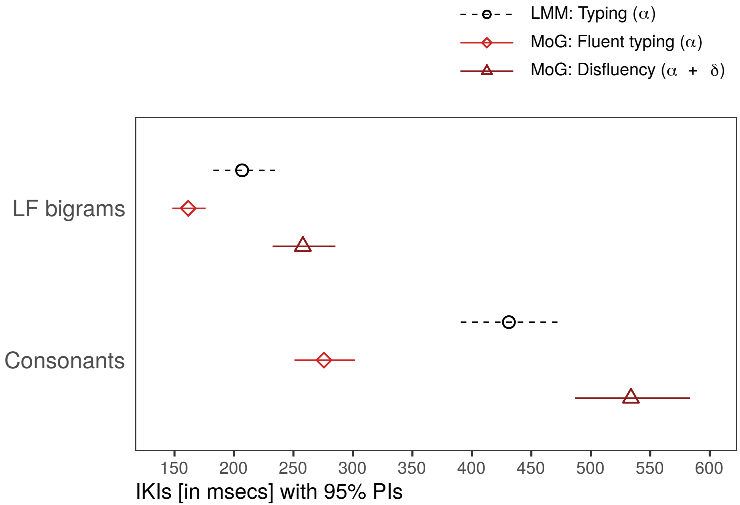
Observed vs. predicted IKIs



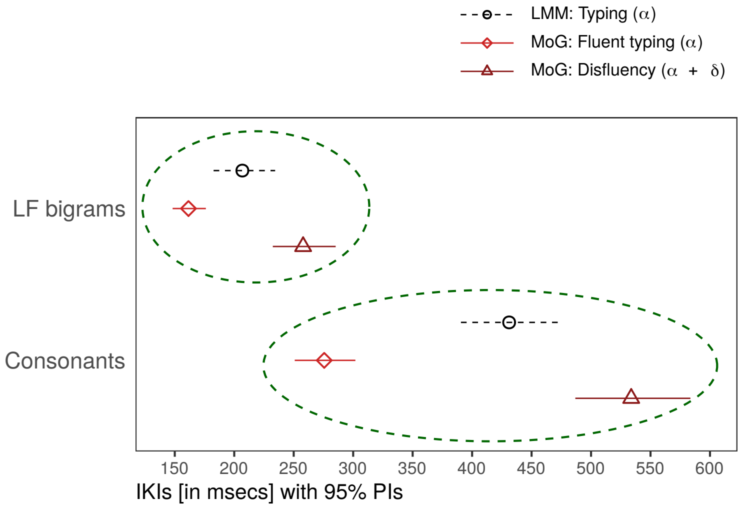
Population estimates



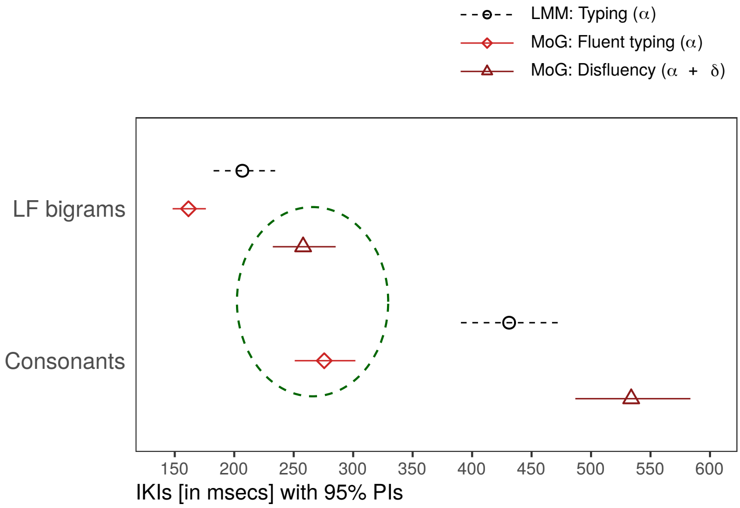
Population estimates



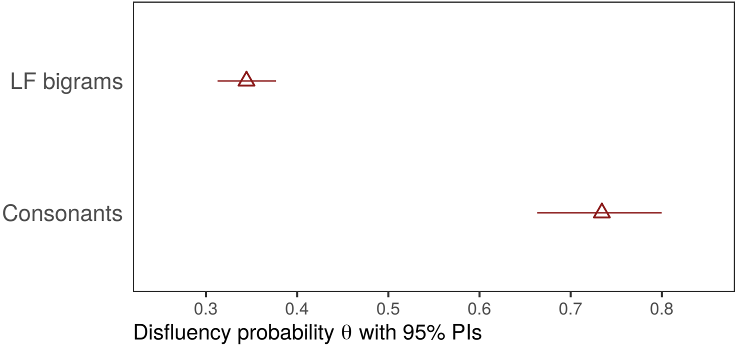
Population estimates



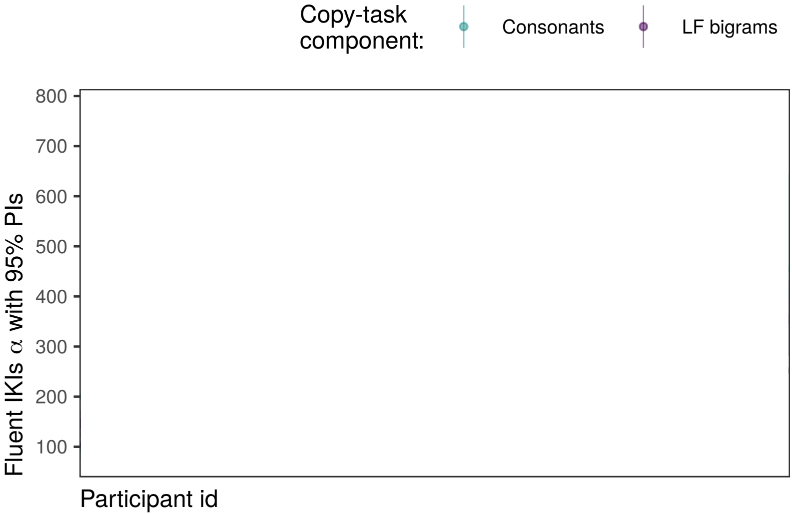
Population estimates



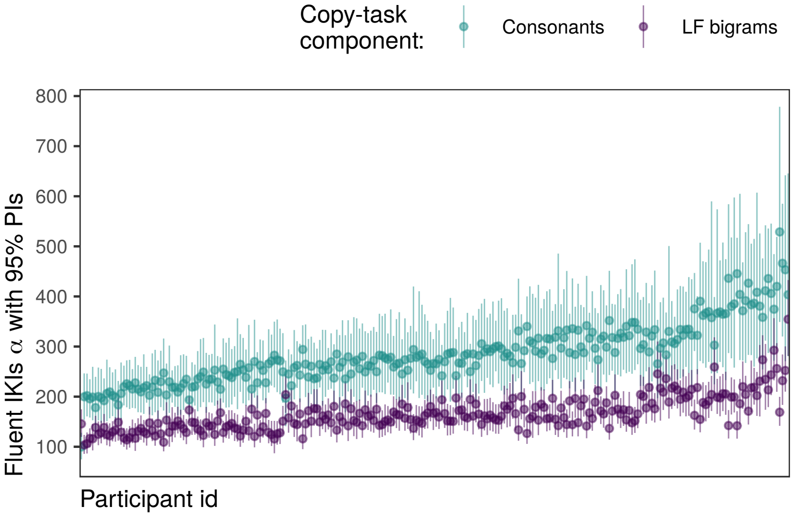
Population estimates



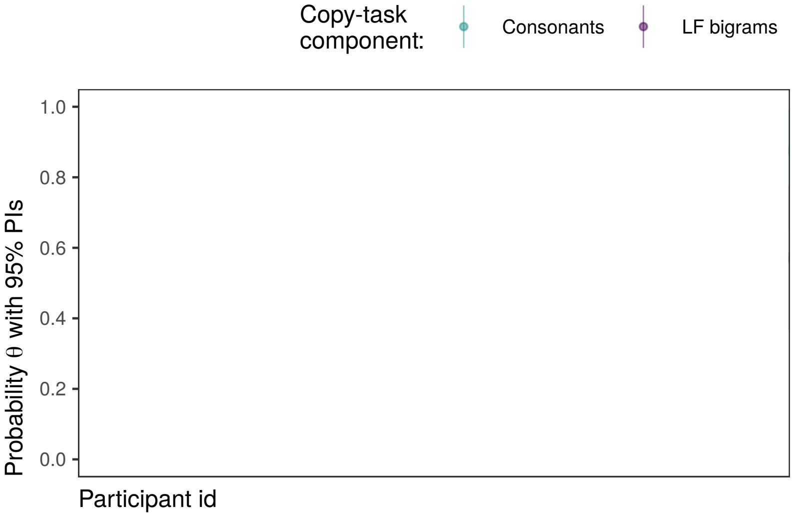
By-participant fluent-typing intervals



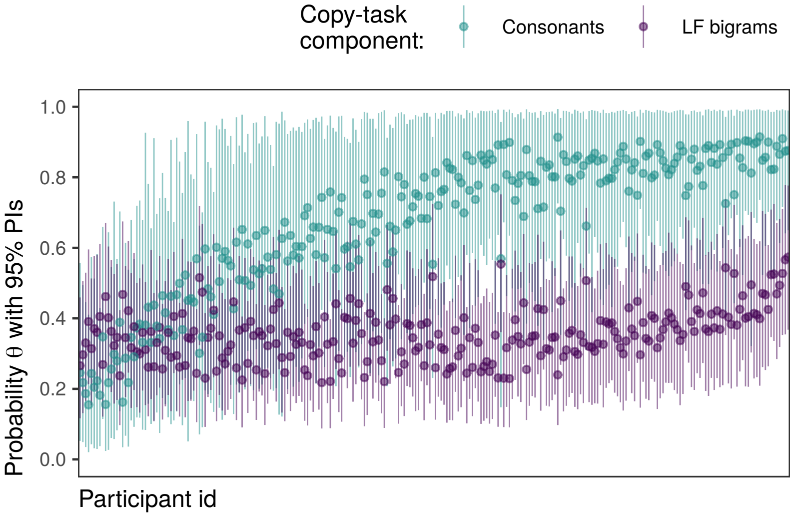
By-participant fluent-typing intervals



By-participant disfluency probability



By-participant disfluency probability



Conclusion

- Better fit for mixture models over standard analysis.
- Data come from a mixture of processes reflected in fluent and disfluent keystroke transitions.
- Advantages for writing research:
  1. map on cascading models of writing (ask me why).
  2. represent the probabilistic nature of disfluencies.
  3. capture disfluencies in a principled way.
  4. provide reliable typing estimates and pause frequencies

Discussion

- Our statistical techniques need to align closely with the cognitive process we're trying to understand.
- …and represent our current understanding of the underlying cognitive process.
- To achieve this we need to model the raw data possible rather than summary statistics.
- Otherwise we risk incorrect conclusions about our data.

Prospects

- Shiny-app: calculate disfluencies from keystroke data.
- Manuscript: extension to autoregression.
- Application of models for data in which pause frequencies play a central role: group comparisons, diagnostic tool?
- Application picture description data with spelling manipulation.

Thanks for listening!

email: jens.roeser@ntu.ac.uk
nottinghamtrent.academia.edu/JensRoeser

R-scripts, *Stan*-code, slides, preprint:
https://github.com/jensroes/Typing-disfluency

References I

Almond, R., Deane, P., Quinlan, T., Wagner, M. & Sydorenko, T. (2012). *A preliminary analysis of keystroke log data from a timed writing task* (tech. rep. Research Report No. RR-12-23). Princeton, NJ, Educational Testing Service.

Baaijen, V. M., Galbraith, D. & de Glopper, K. (2012). Keystroke analysis: Reflections on procedures and measures. *Written Communication*, *29*(3), 246–277.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. & Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *20*.

Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M. & Bennett, R. E. (2018). Modeling basic writing processes from keystroke logs. *Journal of Educational Measurement*, *55*(2), 194–216.

Sorensen, T., Hohenstein, S. & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *Quantitative Methods for Psychology*, *12*(3), 175–200.

Van Waes, L., Leijten, M., Pauwaert, T. & Van Horenbeeck, E. (2019). A multilingual copy task: Measuring typing and motor skills in writing with inputlog. *Journal of open research software*, *7*(30), 1–8.

Van Waes, L., Leijten, M., Roeser, J., Olive, T. & Grabowski, J. (2020). Designing a copy task to measure typing and motor skills in writing research [submitted]. *Journal of Writing Research*.

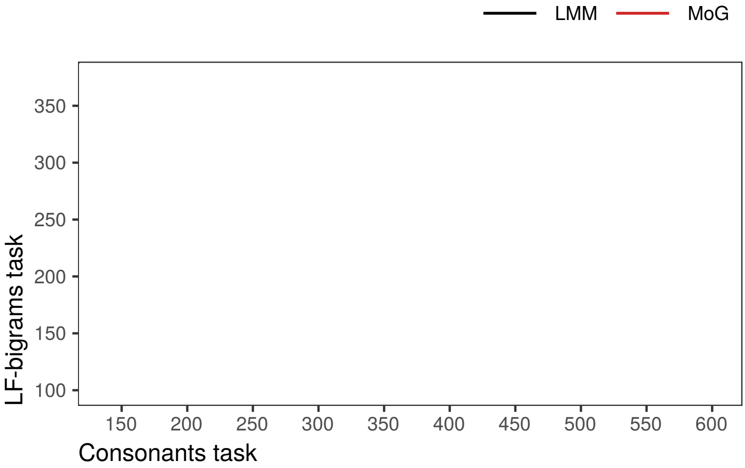
References II

Vasishth, S., Chopin, N., Ryder, R. & Nicenboim, B. (2017). Modelling dependency completion in sentence comprehension as a Bayesian hierarchical mixture process: A case study involving Chinese relative clauses. *ArXiv e-prints*.

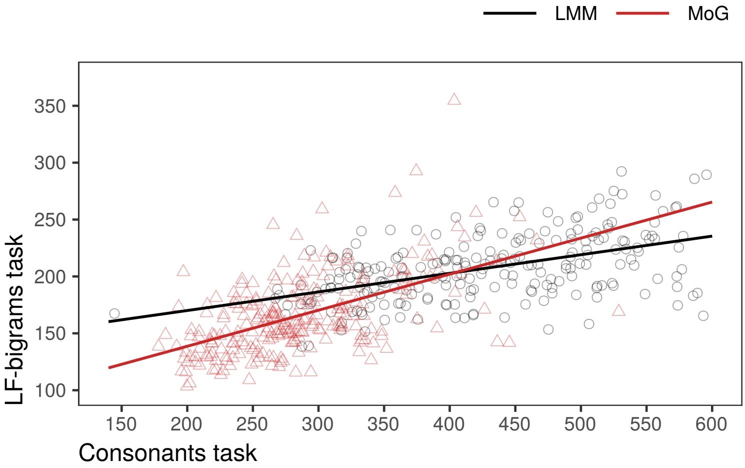
Vehtari, A., Gelman, A. & Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*.

Vehtari, A., Gelman, A. & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

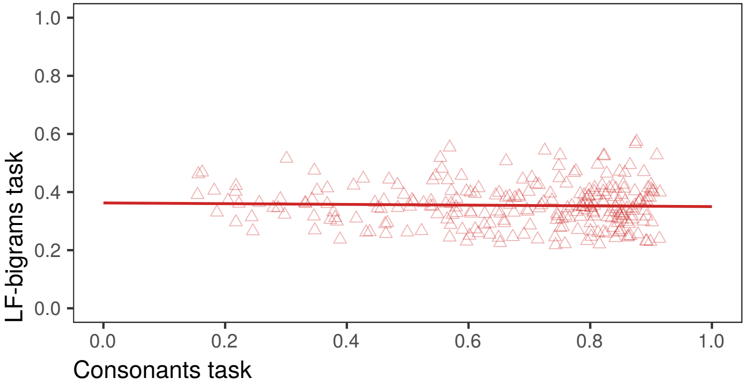
Estimated (fluent) keystroke transitions



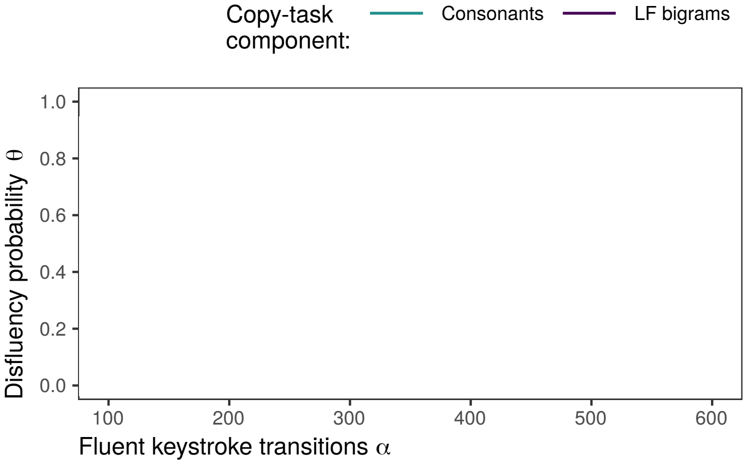
Estimated (fluent) keystroke transitions



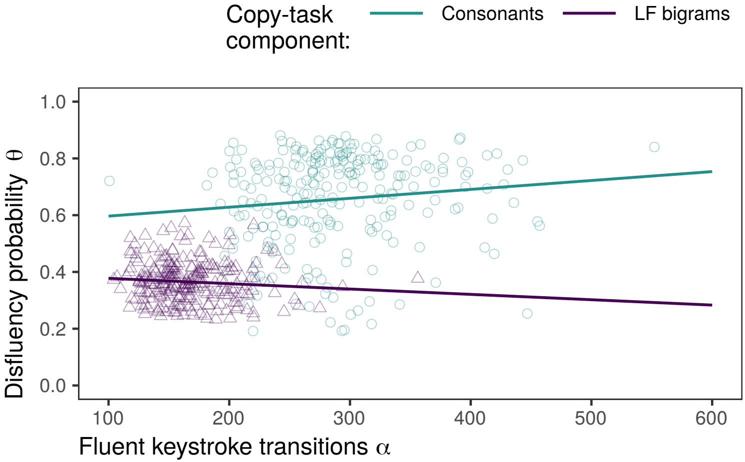
Estimated disfluency probability



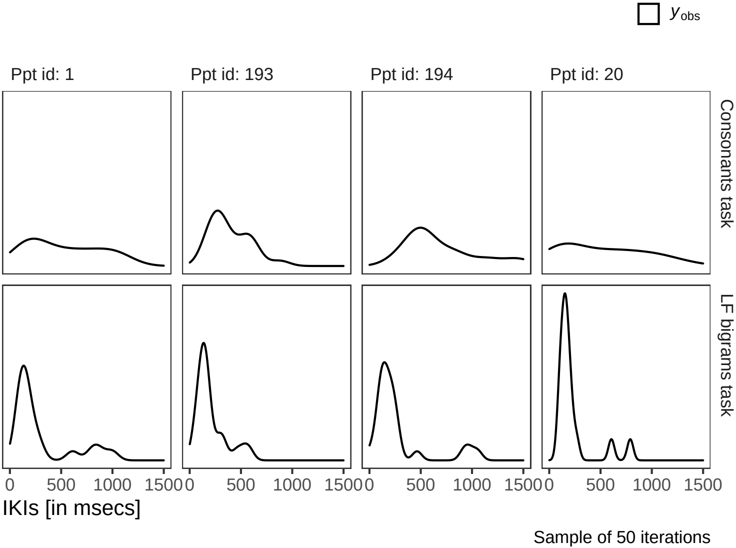
Disfluency typing-speed trade-off



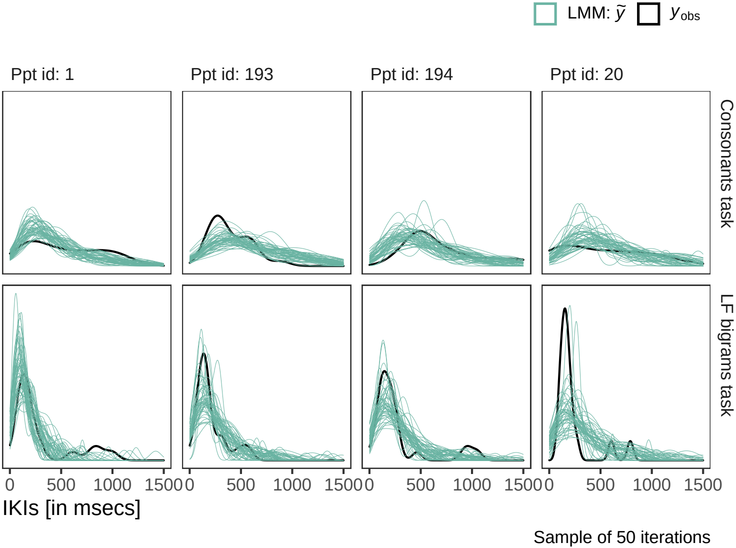
Disfluency typing-speed trade-off



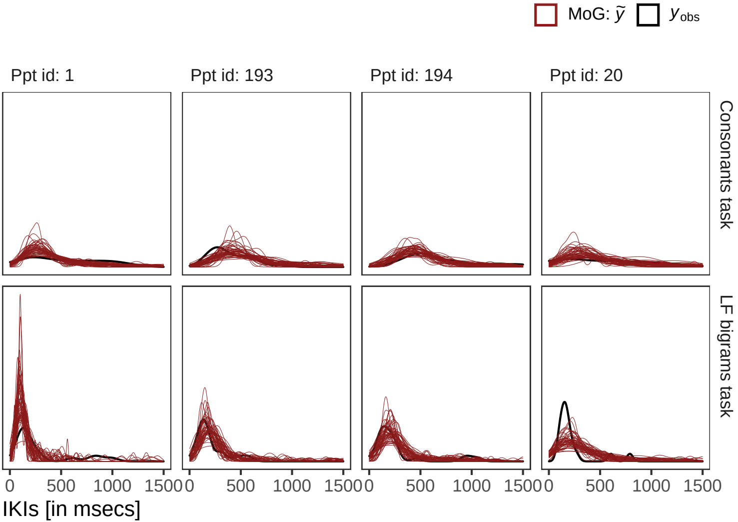
Observed vs. predicted IKIs



Observed vs. predicted IKIs

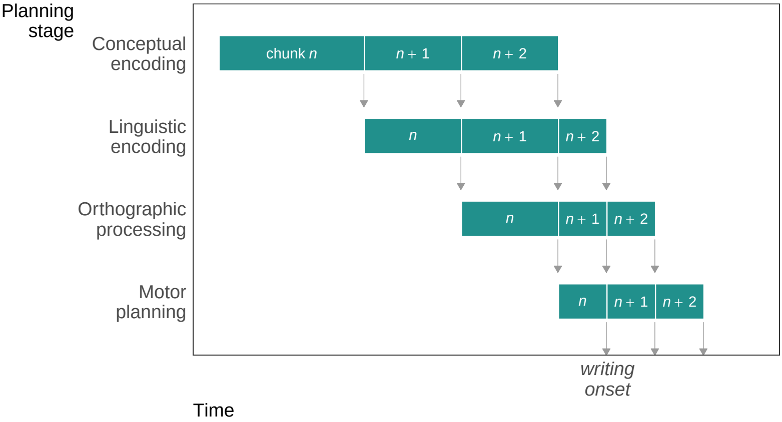


Observed vs. predicted IKIs

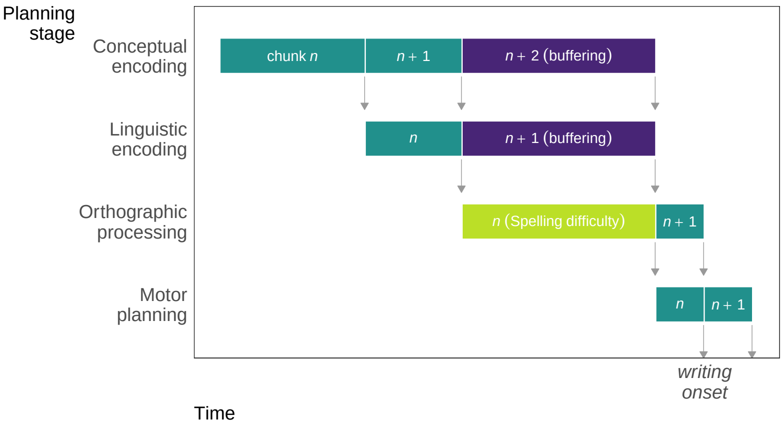


Sample of 50 iterations

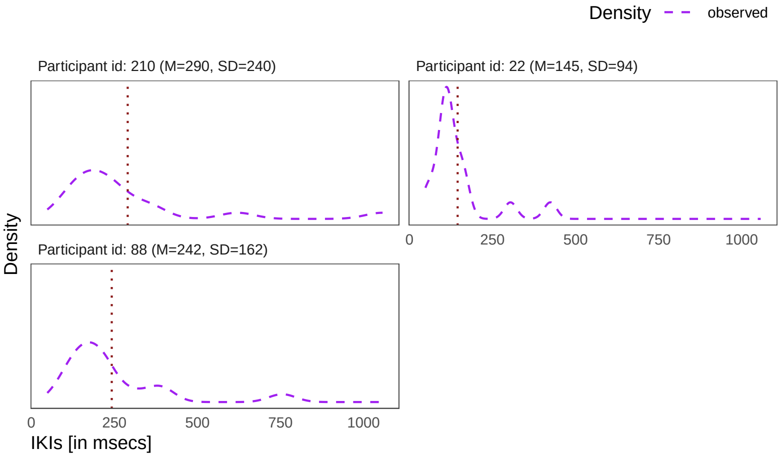
Planning cascade in writing



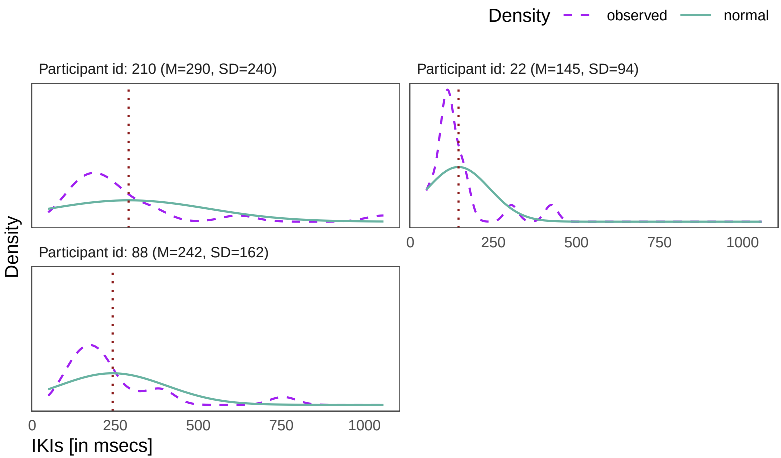
Planning cascade in writing



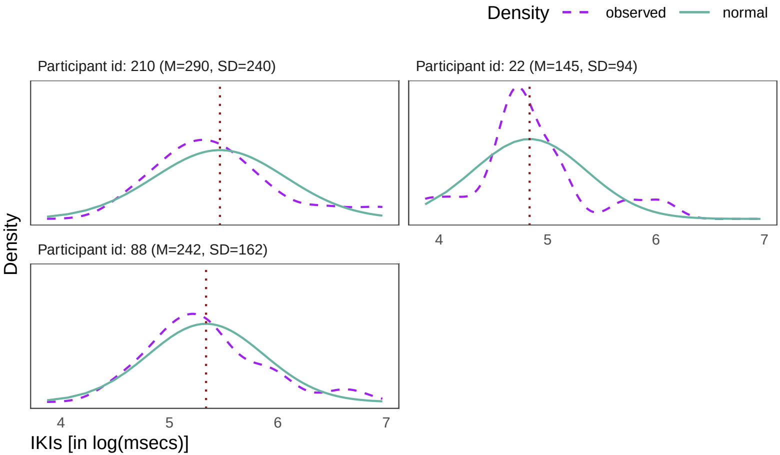
Keystroke transitions are not normal distributed



Keystroke transitions are not normal distributed



Keystroke transitions are not normal distributed



Long intervals are not bigram specific