

Finite mixture modeling: A case study involving retrieval processes in sentence comprehension

Shravan Vasishth

with: Nicolas Chopin, Robin Ryder, Bruno Nicenboim

University of Potsdam, Germany / CEREMADE, Université Dauphine

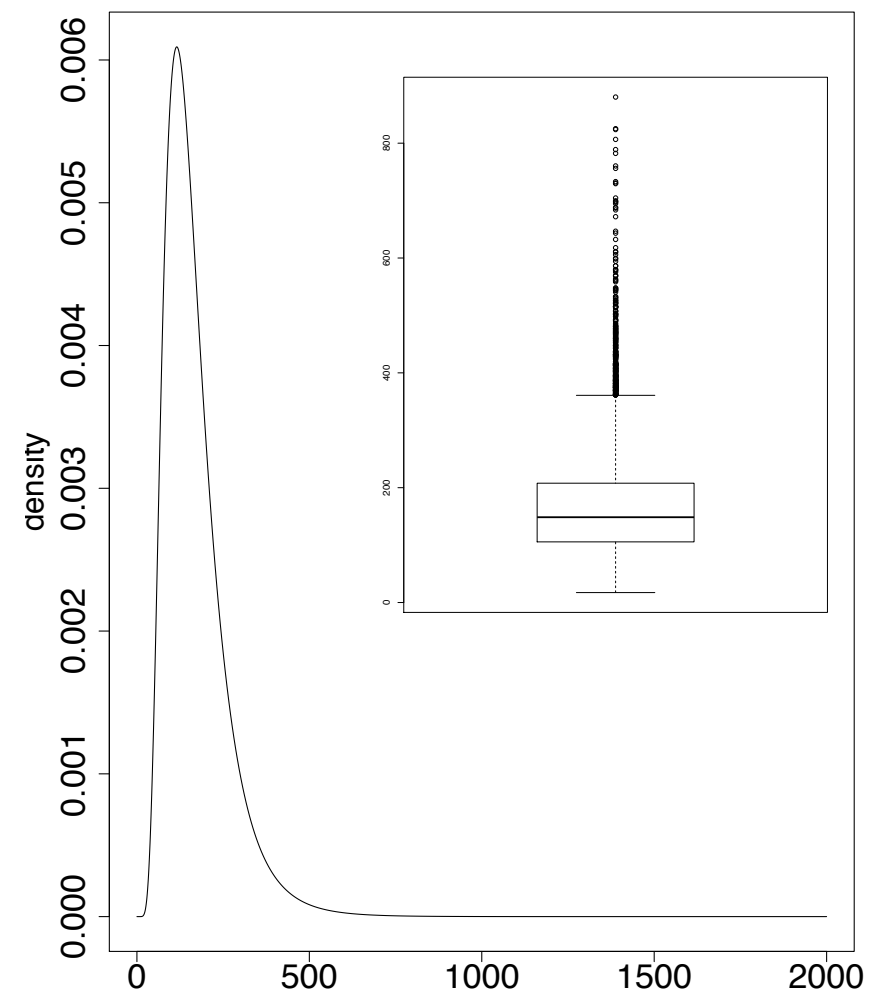
A simple data generation process

Imagine a simple data generation process

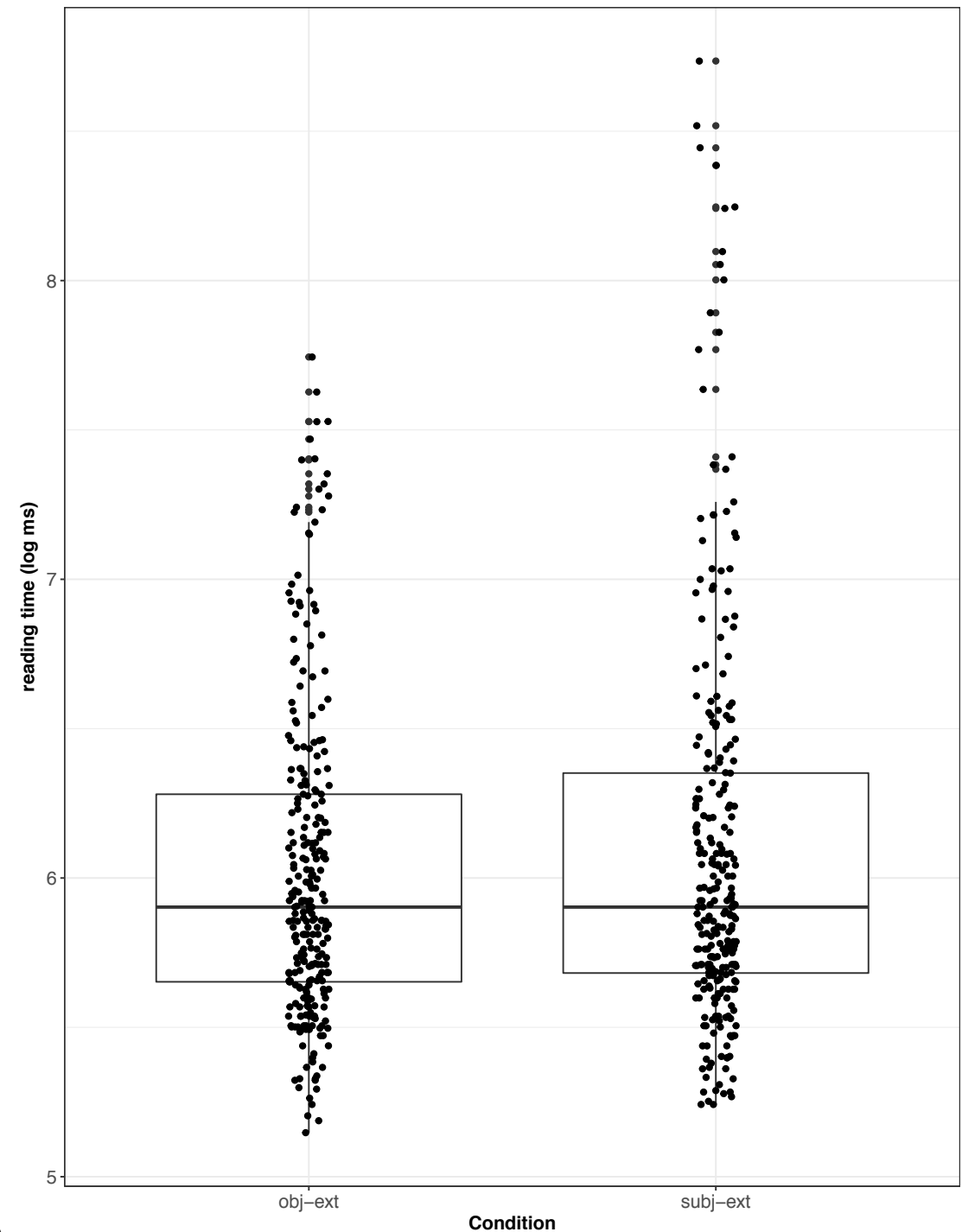
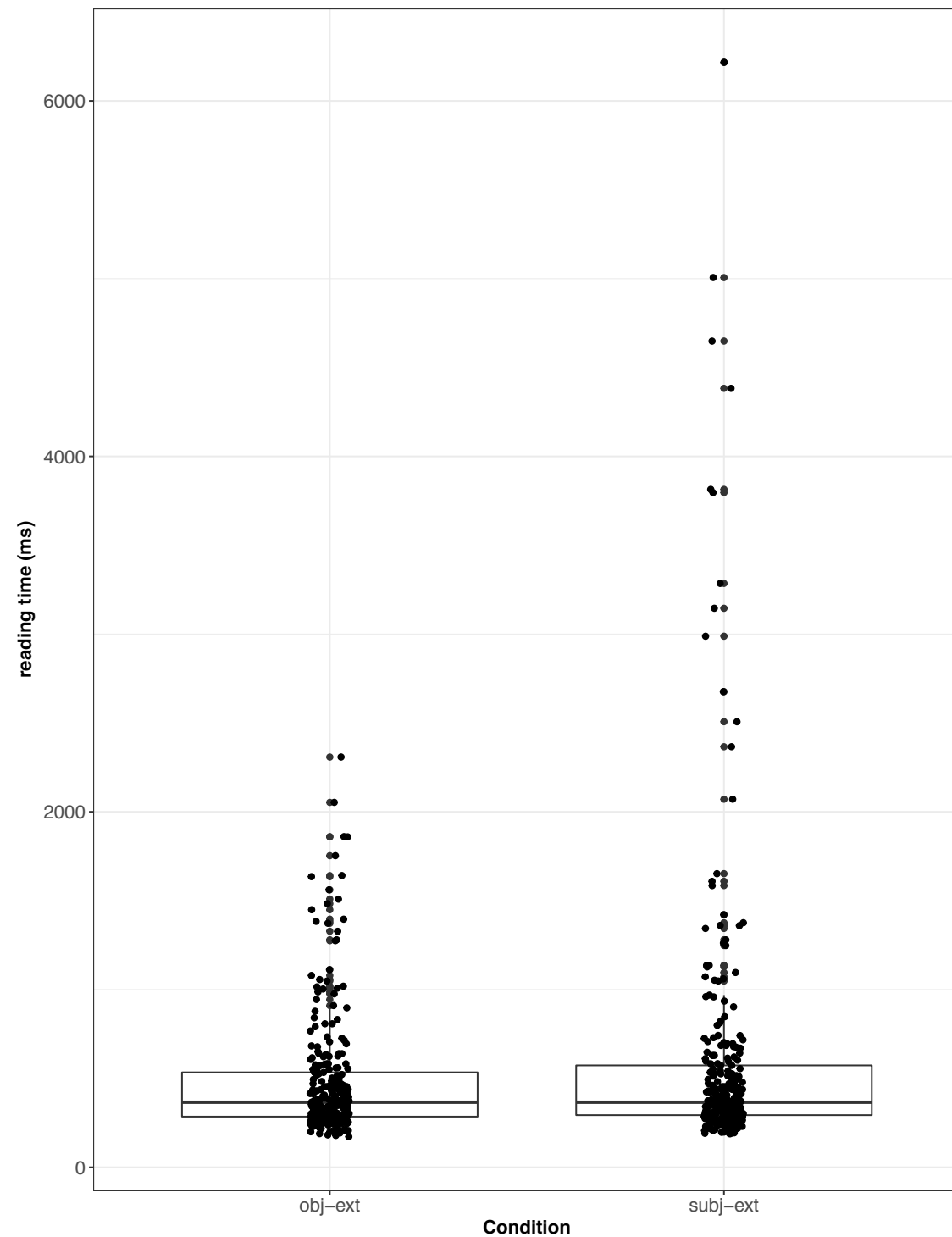
$$y_i \sim \text{LogNormal}(\mu, \sigma^2) \quad i = 1, \dots, N$$

Example:

Reading time in milliseconds



Comparing means from two conditions in reading time data



Comparing means from two conditions in reading time data

Option 1: A simple linear regression:

$$y_i \stackrel{iid}{\sim} \text{LogNormal}(\beta_0 + \beta_1 x_i, \sigma^2)$$

x_i is coded $+1/2$ or $-1/2$

	Estimate	SE
Intercept	6.1	0.03
x	-0.08	0.05

Comparing means from two conditions in reading time data

Option 2: A hierarchical regression:

$$y_{jk} \sim \text{LogNormal}(\beta_0 + \beta_1 x_{jk} + \textcolor{red}{b}_j + \textcolor{red}{c}_k, \sigma^2)$$

x_{jk} is coded ± 0.5

$$b_j \sim \text{Normal}(0, \sigma_b^2) \quad c_k \sim \text{Normal}(0, \sigma_c^2)$$

j: subject, k: items

$\hat{\beta}_1$ is the estimated difference in means

Comparing means from two conditions in reading time data: A concrete example



a. The **nurse** who the doctor scolded **resigned**



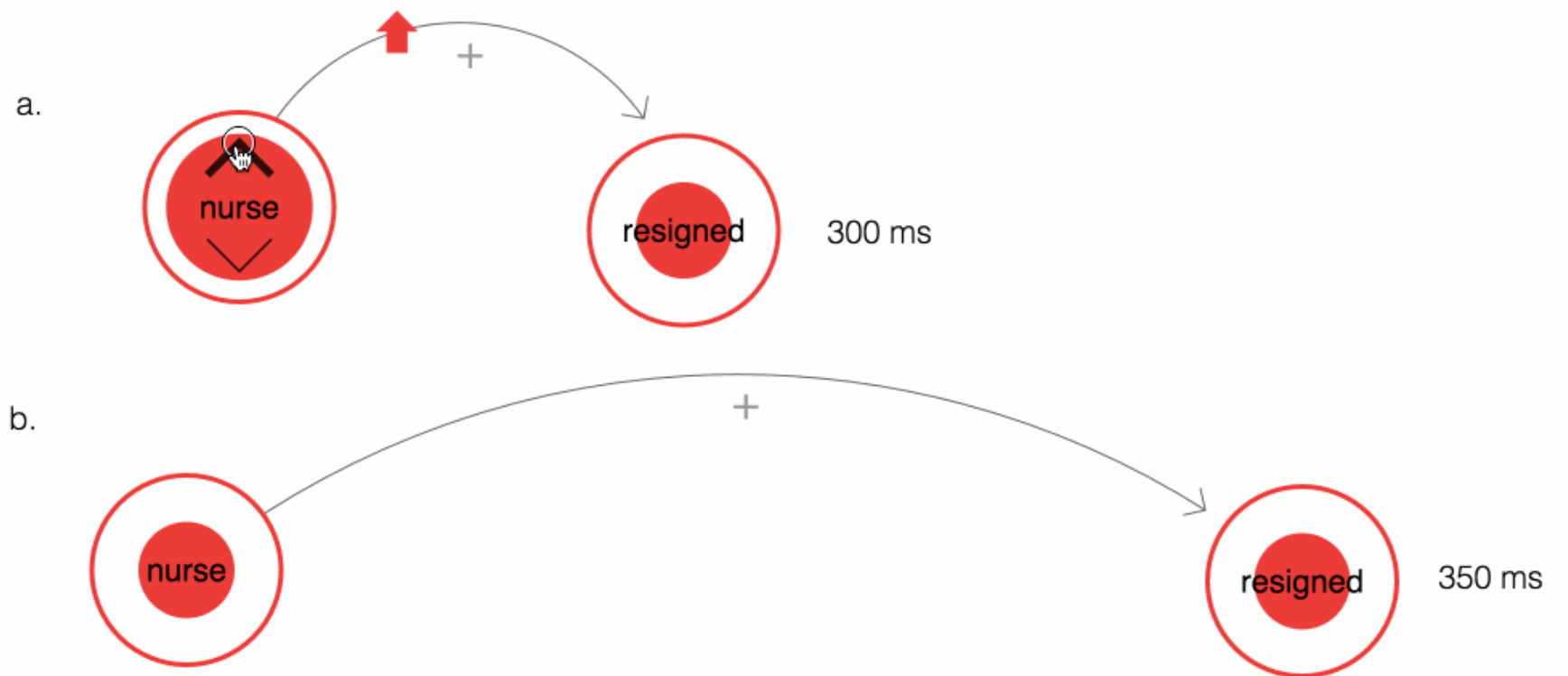
b. The **nurse** who the doctor from the clinic scolded **resigned**

Mean reading times at *resigned* are longer in b vs a. Why?

Two theories about dependency completion

Theory 1: Decay in working memory

A distance-based explanation for reading time differences
(Just & Carpenter 1992; Gibson 1998, 2000; Lewis & Vasishth 2005)

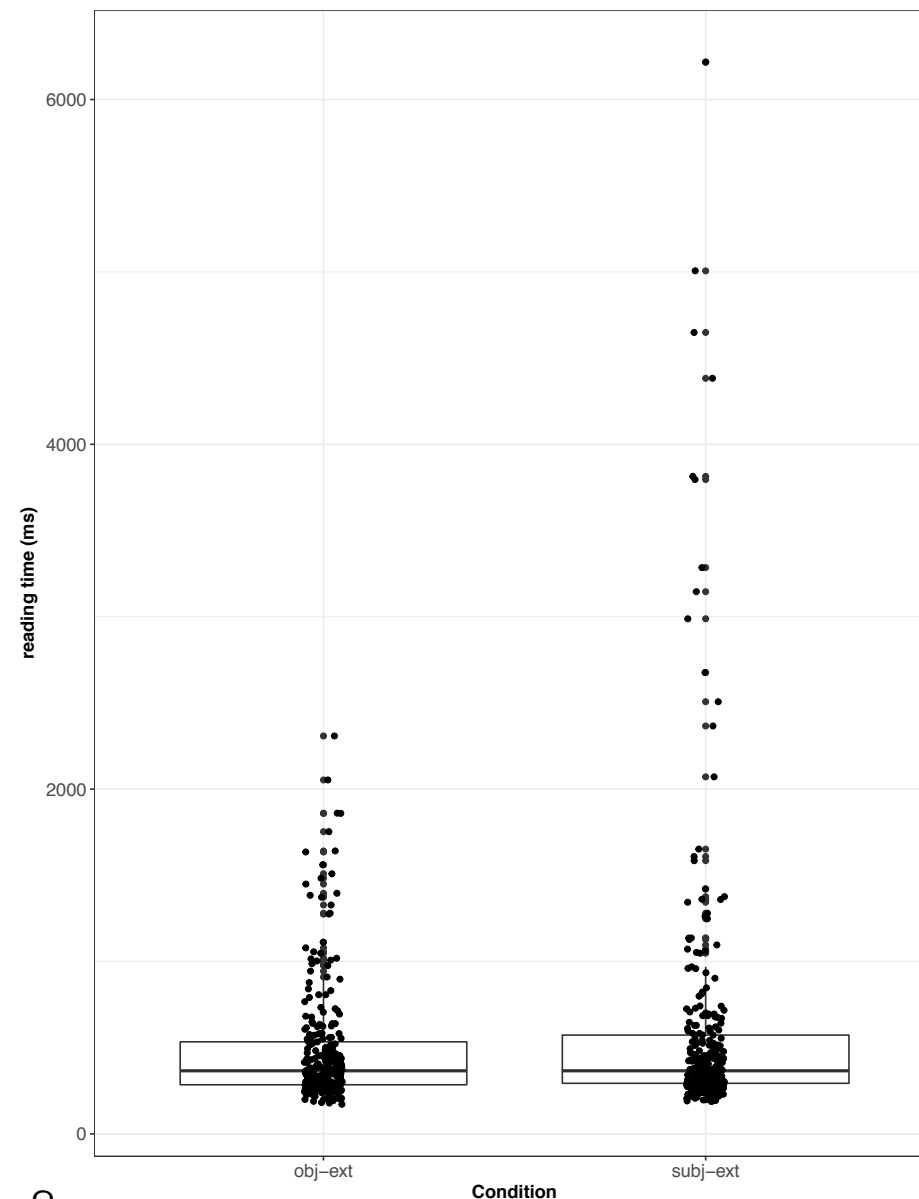


0:00

The decay model as a hierarchical linear model

$$y_{jk} \sim \text{LogNormal}(\beta_0 + \beta_1 + b_j + c_k, \sigma^2)$$

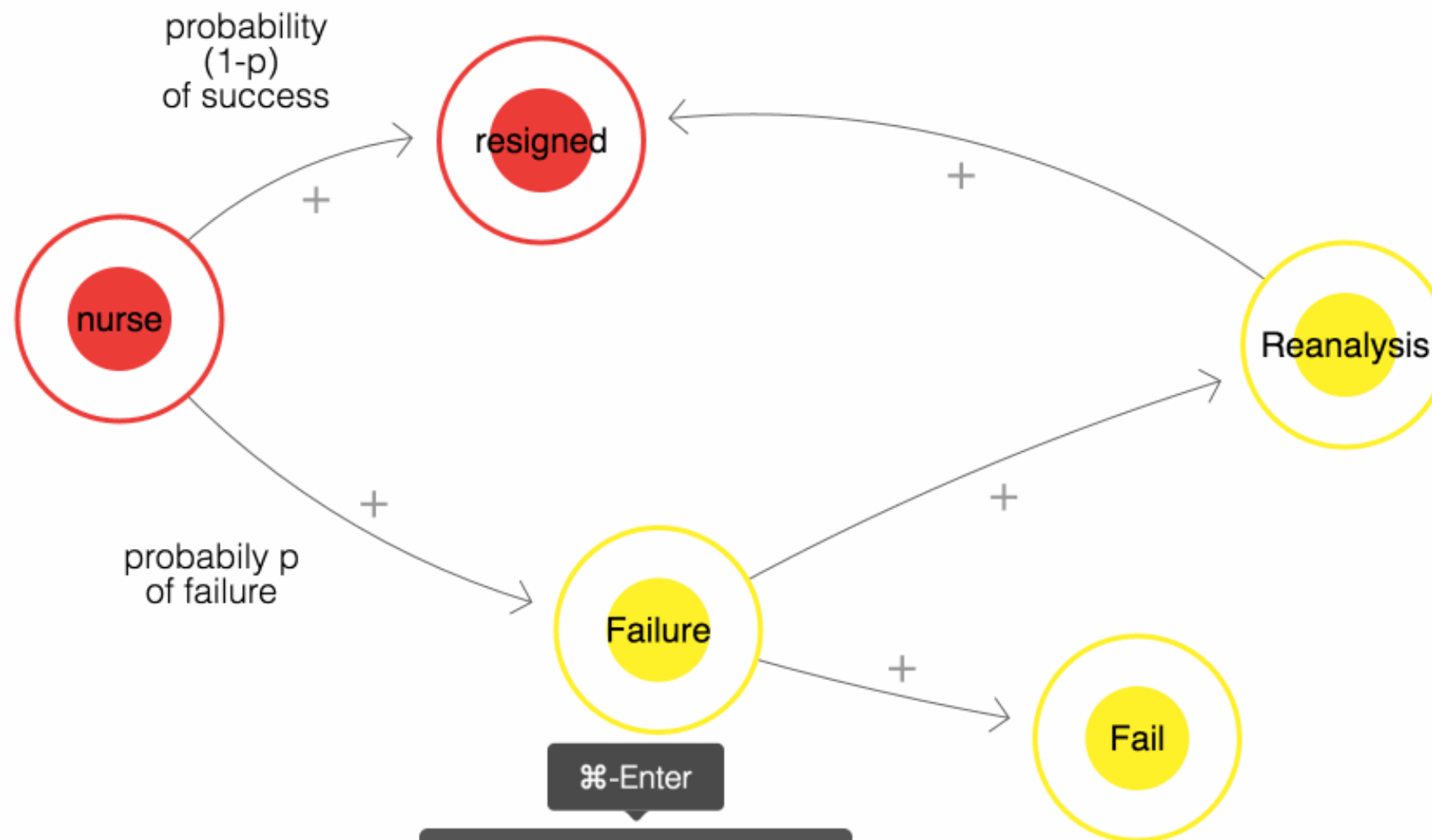
$\hat{\beta}_1$ estimates the
difference in means



Two theories about dependency completion

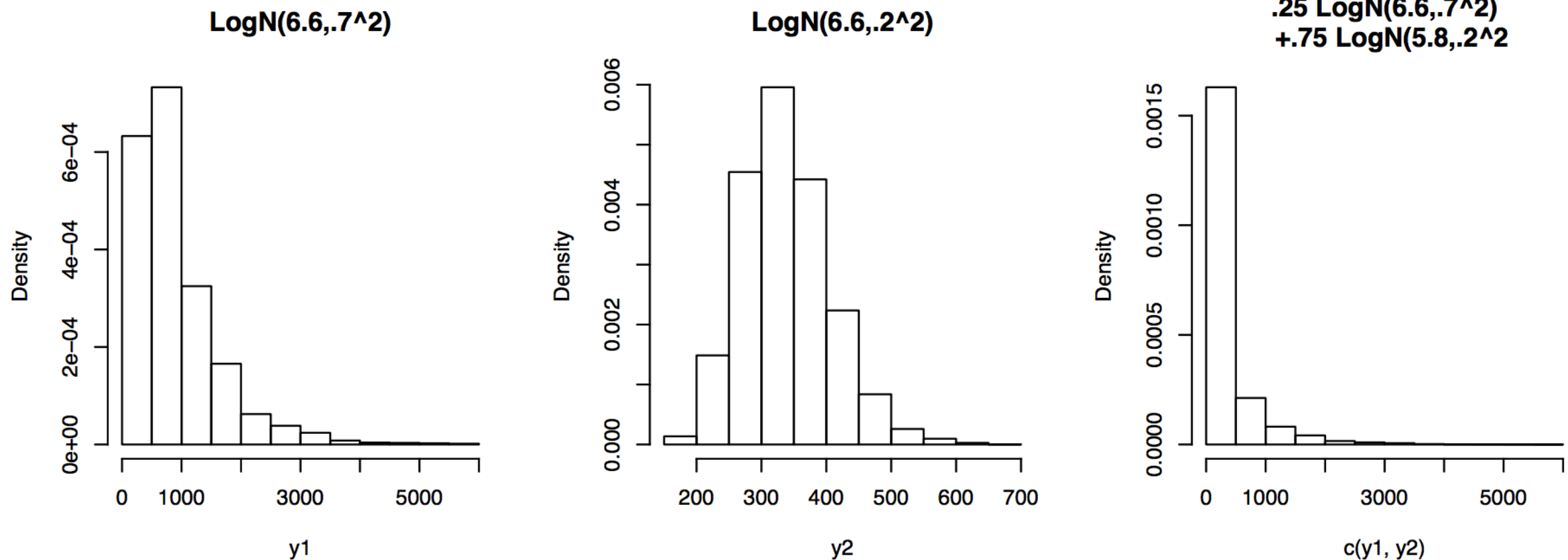
Theory 2: Direct-access model

The direct access model of McElree et al., 2003 (see Nicenboim & Vasishth 2017).



The direct-access model as a mixture process

$$y_{jk} \sim p_1 \cdot \text{LogNormal}(\mu'_{jk}, \sigma'^2) + (1 - p_1) \cdot \text{LogNormal}(\mu_{jk}, \sigma^2)$$



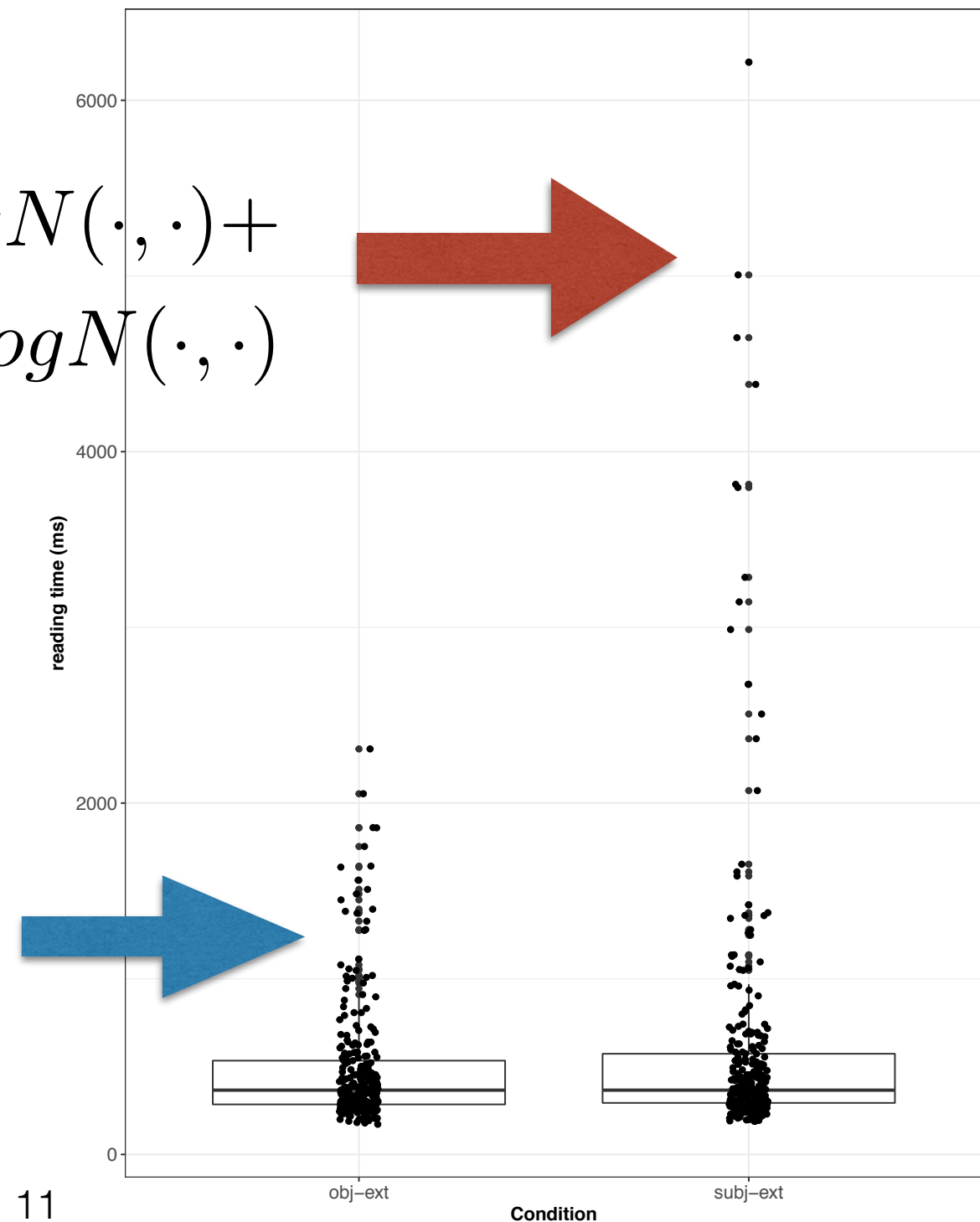
The direct-access model as a mixture process

(also see: Nicenboim & Vasishth, StanCon 2017)

Expected: $p_1 > p_2$

$$p_1 \text{Log}N(\cdot, \cdot) + (1 - p_1) \text{Log}N(\cdot, \cdot)$$

$$p_2 \text{Log}N(\cdot, \cdot) + (1 - p_2) \text{Log}N(\cdot, \cdot)$$



Implementing both models in Stan

Decay model

$$y_{jk} \sim \text{LogNormal}(\beta_0 + \beta_1 x + b_j + c_k, \sigma^2)$$

Direct-access model

$$y_{jk} \sim p_1 \cdot \text{LogNormal}(\mu'_{jk}, \sigma'^2) + \\ (1 - p_1) \cdot \text{LogNormal}(\mu_{jk}, \sigma^2))$$

$$y_{jk} \sim p_2 \cdot \text{LogNormal}(\mu'_{jk}, \sigma'^2) + \\ (1 - p_2) \cdot \text{LogNormal}(\mu_{jk}, \sigma^2))$$



mc-stan.org

Research question

Which of the decay model and the direct-access model characterises the data better?

Four steps needed

- Step 1: Use fake data to validate mixture model
- Step 2: Estimate parameters of mixture model from real data
- Step 3: Compare mixture model to hierarchical model
- Step 4: Repeat steps 2,3 using new real data.

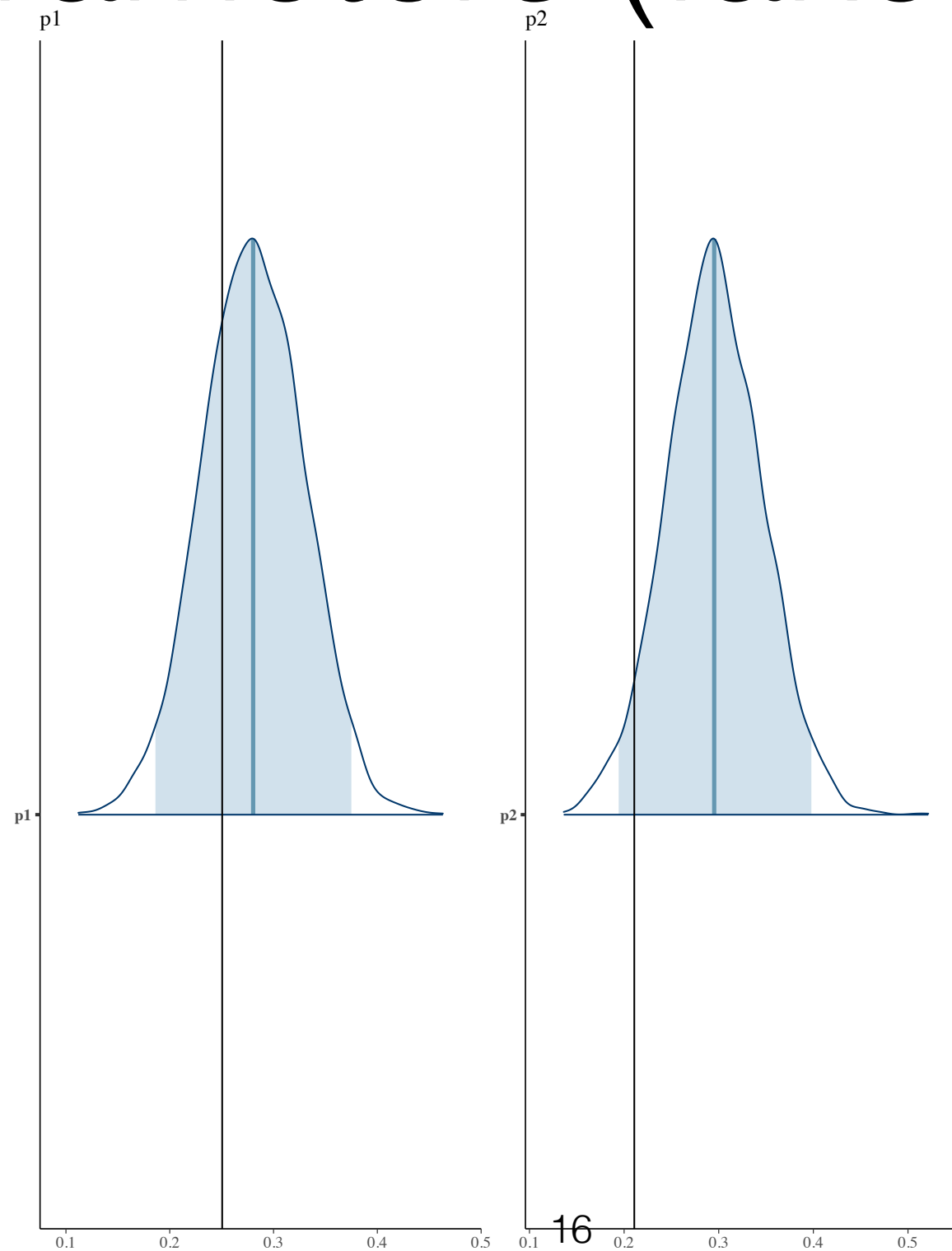
Step 1: Validation of model using fake-data simulation

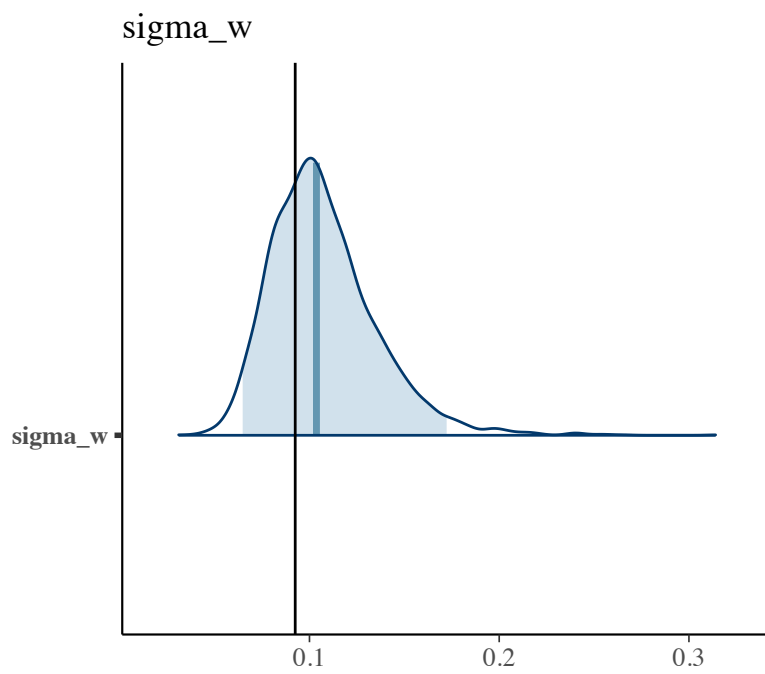
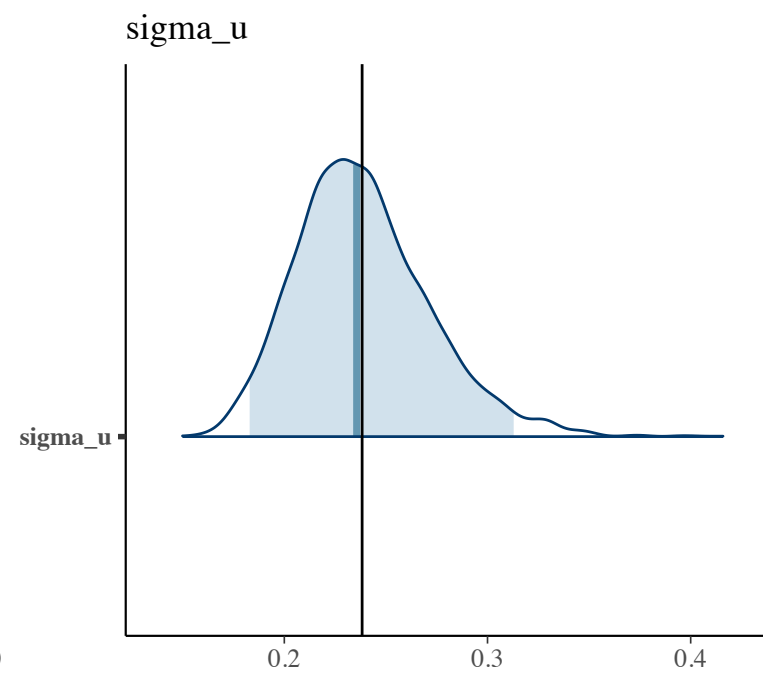
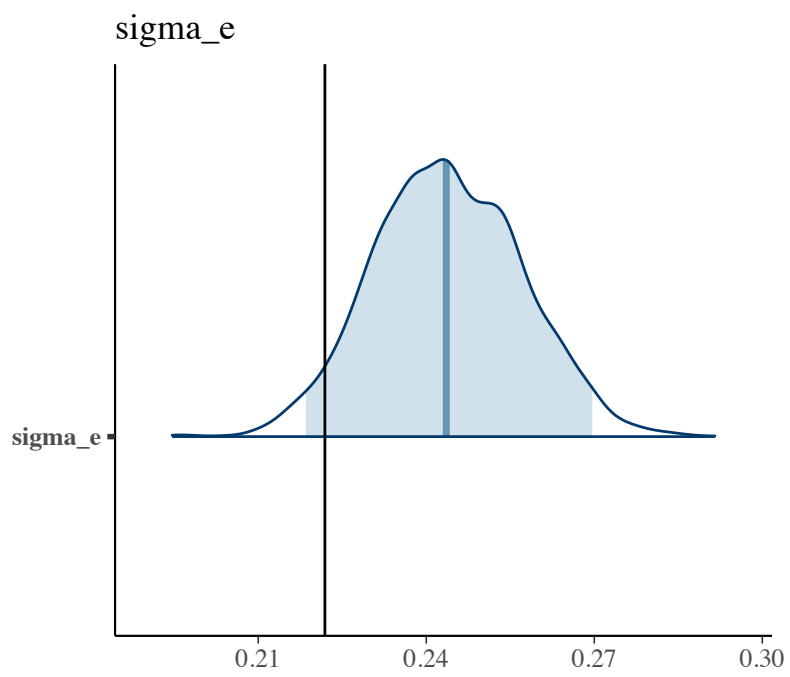
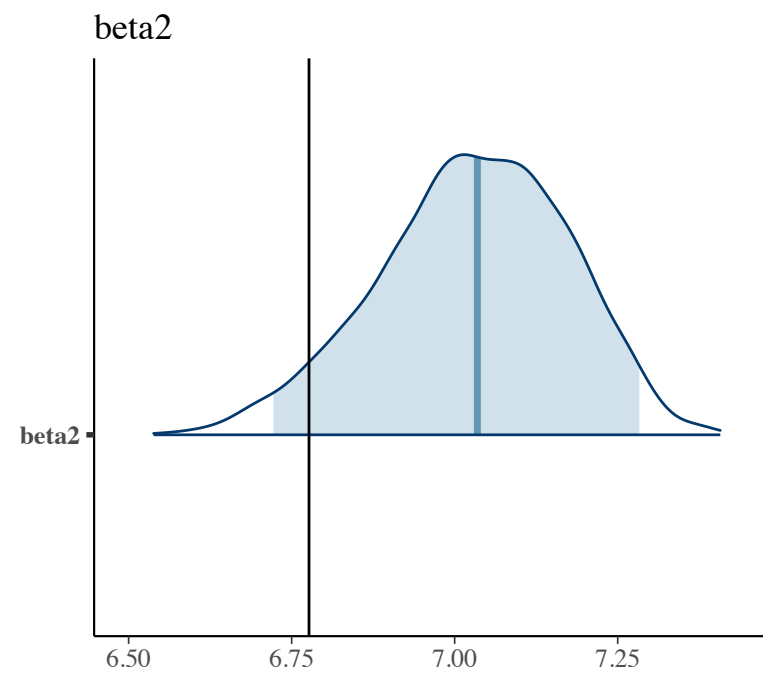
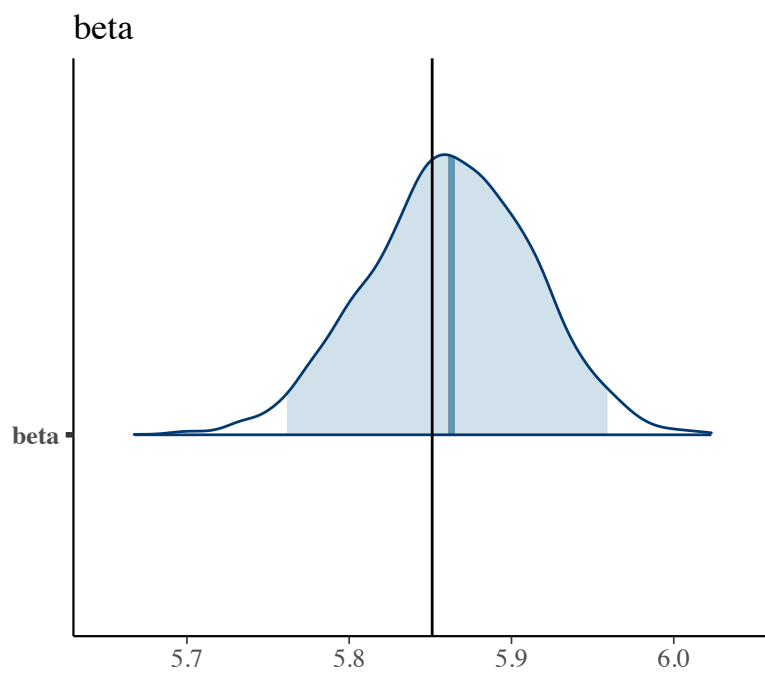
(does mixture model recover parameters?)

Method

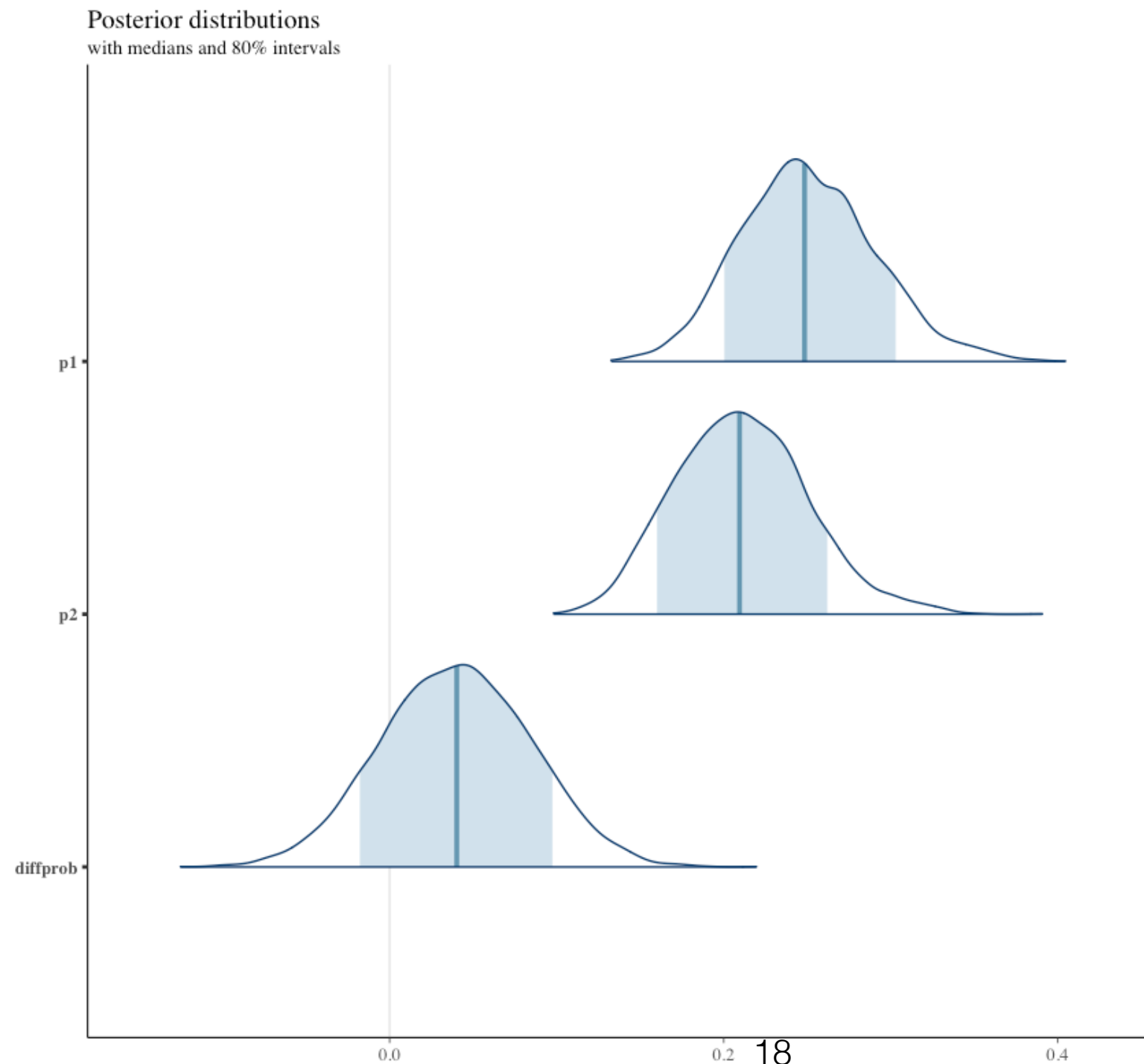
- Generate fake data with fixed parameter values for mixture distribution.
- Plot posterior distributions and determine whether true parameters lie within 95% credible interval.

Mixture probabilities: parameters (fake data)





Step 2: Estimate mixture probabilities: parameters (real data)



Step 3: Model comparison (Vehtari et al 2016)

- To compare models, we compare **predictive performance** of each model.
- We use an approximation of leave-one-out cross-validation (Pareto-smoothed importance sampling).
- **Expected log pointwise density (elpd)** is a measure of predictive accuracy.
- Higher *elpd* implies better predictive performance.

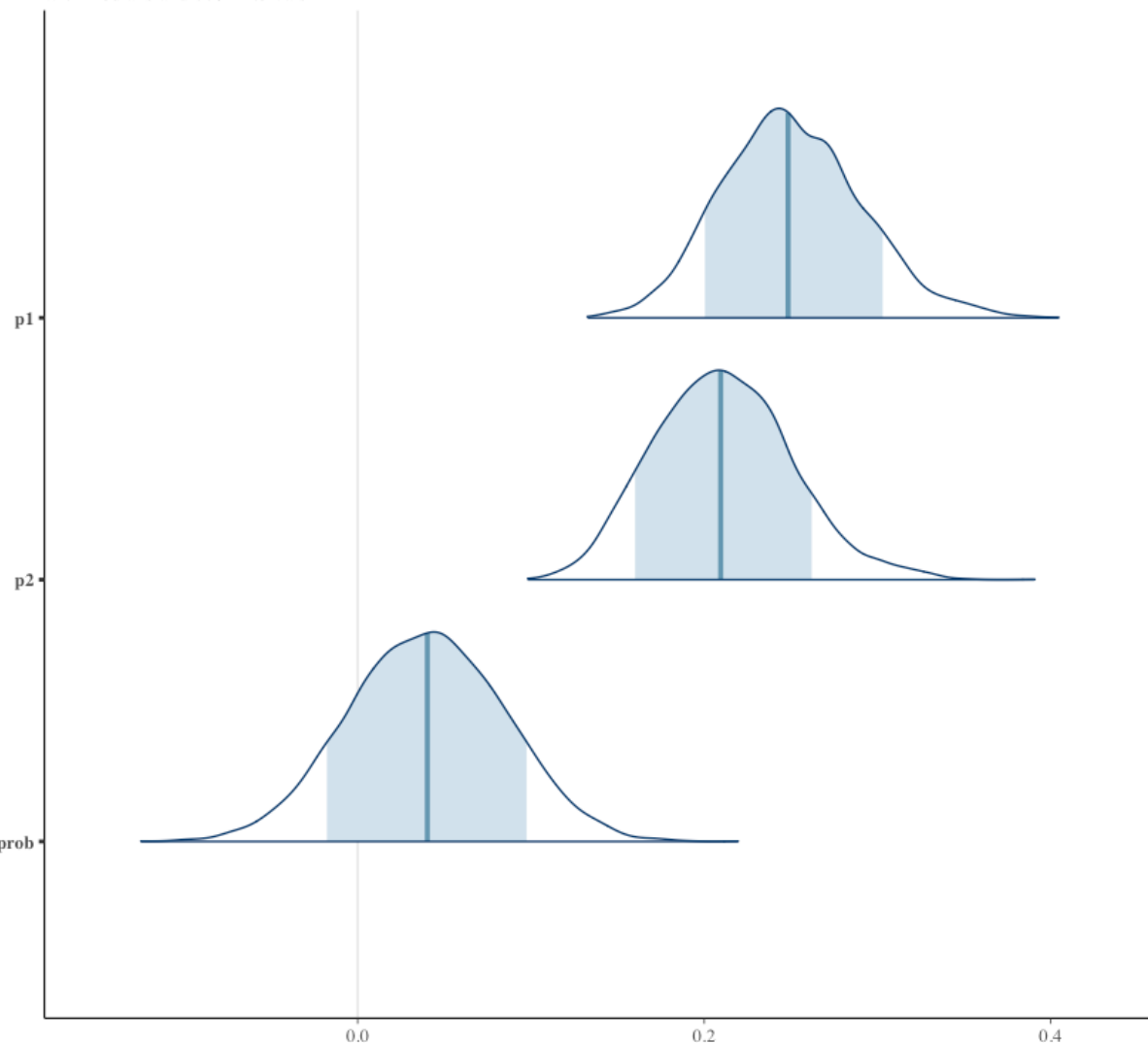
HLM vs Mixture ELPD difference: 147.6 (17.2)

Step 4: Do we obtain the same results with new data?

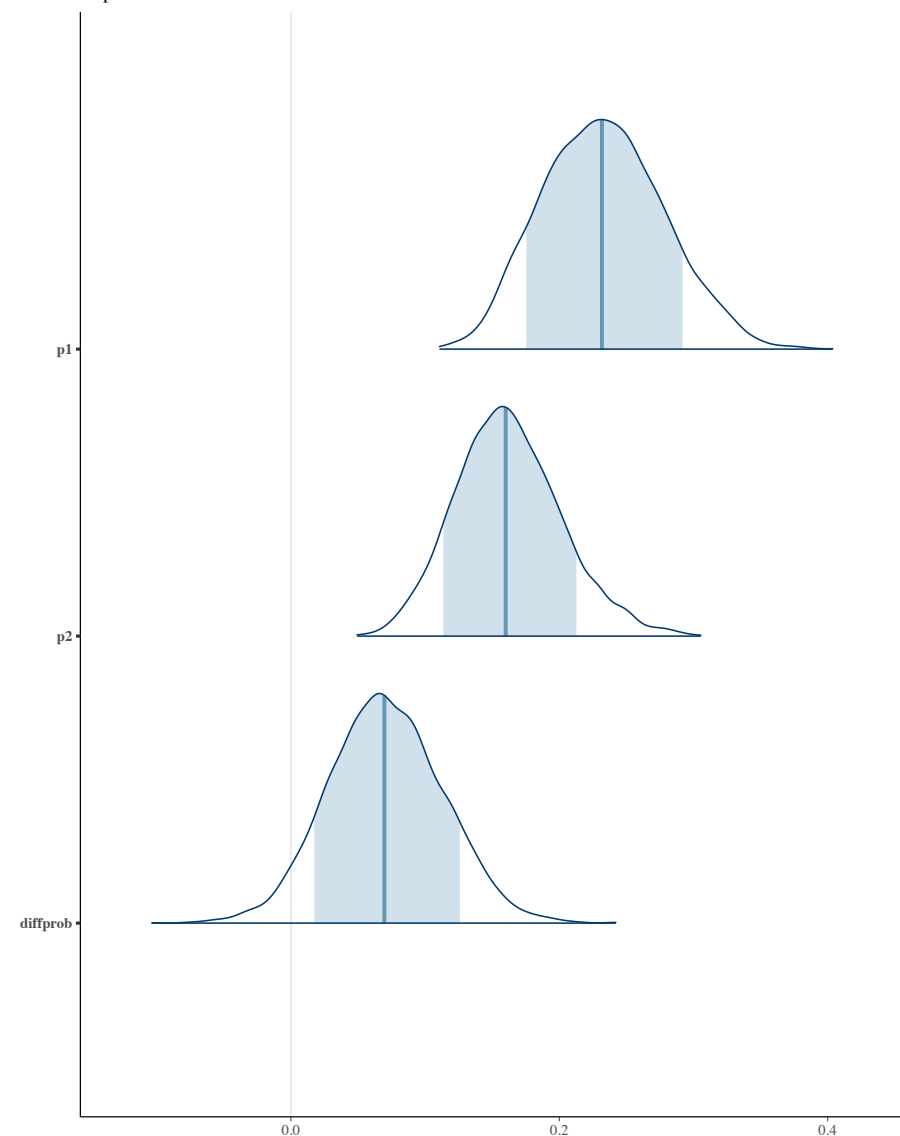
ELPD diff: 147.6 (17.2)

ELPD diff: 156.3 (28.9)

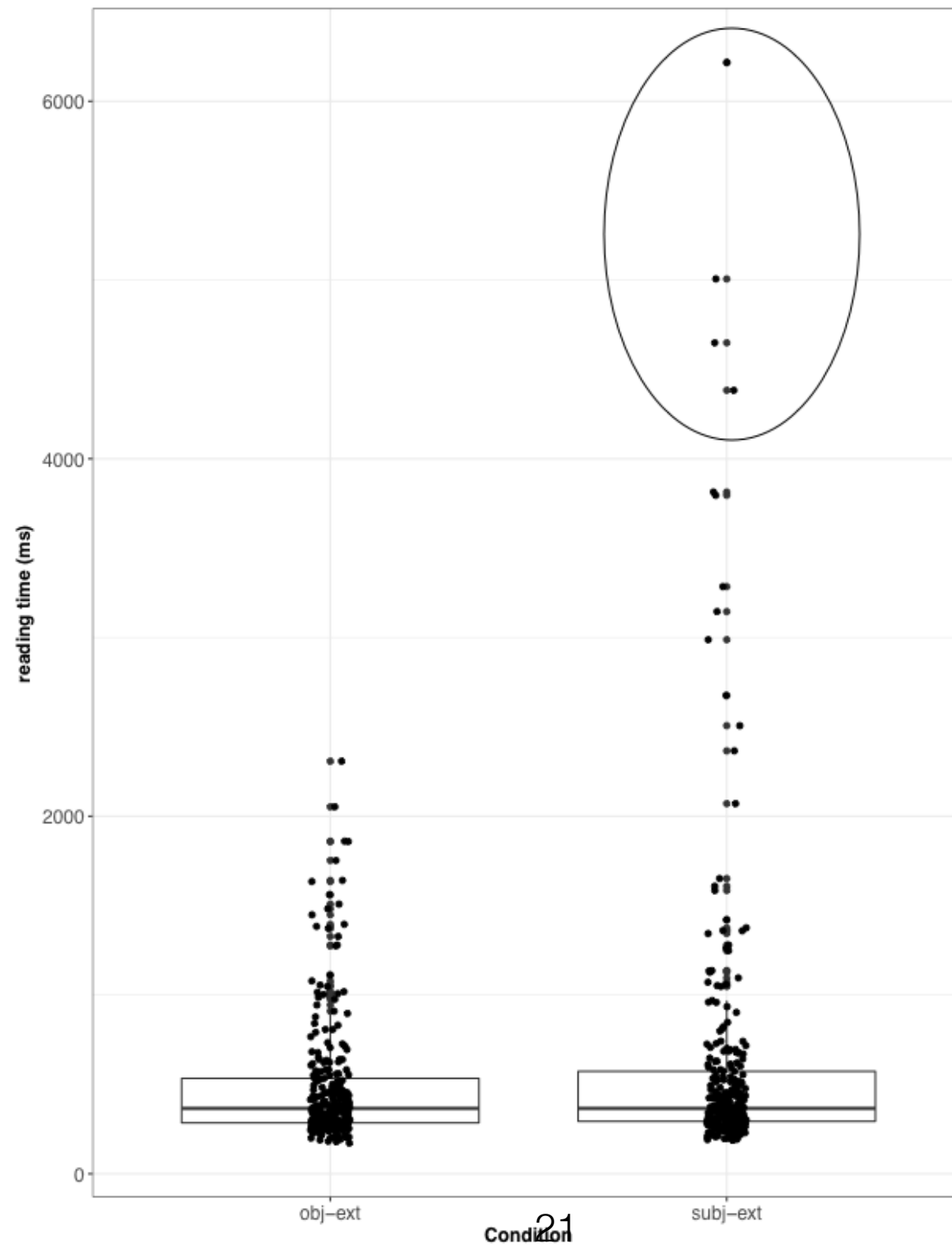
Posterior distributions
with medians and 80% intervals



Replication data



Conclusion



Conclusion

The evidence for the direct-access model is stronger than for the decay model.

Full paper: <http://arxiv.org/abs/1702.00564>

Conclusion

Other evidence consistent with
direct-access model:

Nicenboim & Vasisht, StanCon 2017

Video, code:

<http://bit.ly/NicenboimVasishtStanCon2017>

Conclusion



Michael Betancourt @betanalpha · Apr 10

Friends don't let friends fit mixture models, at least not for problems that matter.



4



1



14



Conclusion

I hope that I at least convinced you to take a look at Stan for Bayesian modeling.

It's a powerful framework for building process models and thinking in terms of the generative process underlying your data.

We will teach a one-day Stan tutorial in Tübingen, Sept 17, 2017: <http://fgme2017.de/>