Bayesian Partition Models for Local Inference in Longitudinal and Survival Data

DISSERTATION DEFENSE

Giorgio Paulon

July 28, 2021

The University of Texas at Austin
Department of Statistics and Data Sciences
College of Natural Sciences

Acknowledgments

Committee members:

- Abhra Sarkar*
- Peter Müller†
- Bharath Chandrasekaran‡
- Mingyuan Zhou§

Funding sources:

- Department of Statistics and Data Sciences

* Department of Statistics and Data Sciences, UT Austin
† Department of Mathematics, UT Austin
‡ Department of Communication Science and Disorders, University of Pittsburgh
§ Department of Information, Risk, and Operations Management, UT Austin

Outline

Thesis organization:

I. Locally varying longitudinal mixed models

II. Drift-diffusion models for tone learning

III. Bivariate survival regression for current status data

Common features: Partition models that share **dependence** across time (longitudinal data) or across outcomes (survival data)

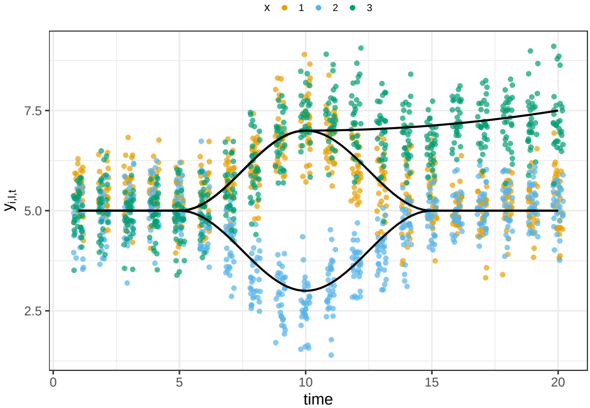
Locally varying longitudinal mixed models

Paulon, G., Llanos, F., Chandrasekaran, B., & Sarkar, A. (2020). Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults. *Journal of the American Statistical Association*.

Paulon, G., Müller, P., & Sarkar, A. (2021). Bayesian semiparametric hidden Markov tensor partition models for local variable selection in longitudinal data. *Submitted*.

The setting

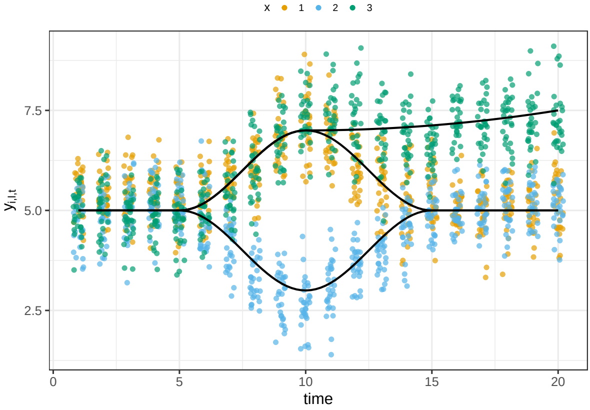
- Continuous response y varying smoothly over time
- Associated categorical predictor x which may vary with time
- The levels of x may affect y differently in the longitudinal stages



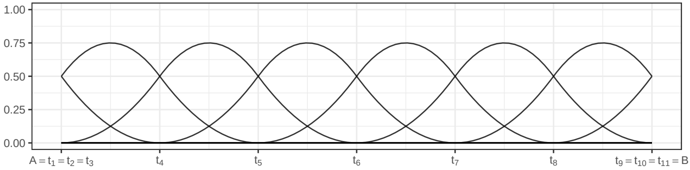
The setting

$$y_{i,\ell,t} \mid (x_{i,\ell,t} = x) = \underbrace{f_x(t)}_{\text{fixed effects}} + \overbrace{u_i(t)}^{\text{random effects}} + \underbrace{\varepsilon_{i,\ell,t}}_{\text{residuals}}, \quad \varepsilon_{i,\ell,t} \sim f_\varepsilon$$

Goal: partition model for the covariate space that evolves dynamically



Penalized B-splines: fixed effects



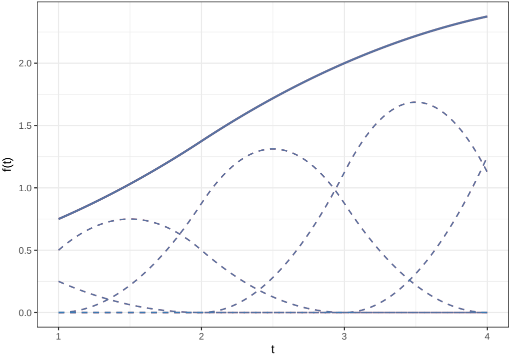
$$y_{i,\ell,t} = f_x(t) + u_i(t) + \varepsilon_{i,\ell,t}$$
$$\{f_x(t) \mid \boldsymbol{\beta}_x\} = \sum_{k=1}^{K} \beta_{k,x} B_k(t)$$

- Smoothness of curves favored by penalty for 'roughness'
- B-splines have local support
- We can cluster the curves by allowing the spline coefficients to have identical values

Local clustering: fixed effects

Curves can cluster ‘globally’[1].

$$\beta_1 = (0.5, 1, 1.75, 2.25, 2.5)^\intercal$$
$$\beta_2 = (0.5, 1, 1.75, 2.25, 2.5)^\intercal$$

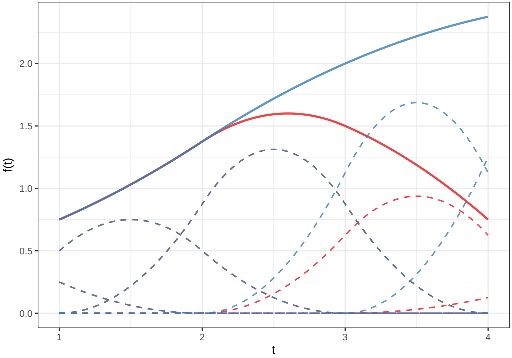


[1] Gelfand, A. E., Kottas, A., & MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, *100*, 1021–1035.

Local clustering: fixed effects

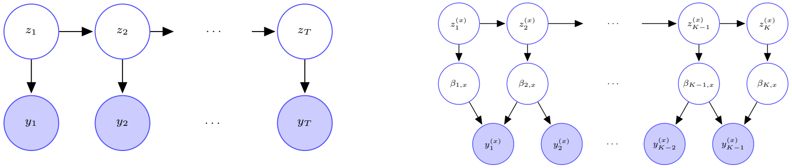
Curves can merge and branch at knot points, i.e. cluster 'locally'[2].

$$\boldsymbol{\beta}_1 = (0.5, 1, 1.75, 2.25, 2.5)^\intercal$$
$$\boldsymbol{\beta}_2 = (0.5, 1, 1.75, 1.25, 0.25)^\intercal$$



[2] Petrone, S., Guindani, M., & Gelfand, A. E. (2009). Hybrid Dirichlet mixture models for functional data. *Journal of the Royal Statistical Society: Series B, 71*, 755–782.

Local clustering: fixed effects

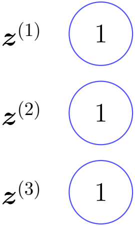


Conventional HMM and proposed HMM

$$
\begin{aligned}
&y_{i,\ell,t} = f_x(t) + u_i(t) + \varepsilon_{i,\ell,t} \\
&\{f_x(t) \mid \boldsymbol{\beta}_x\} = \textstyle\sum_{k=1}^{K} \beta_{k,x} B_k(t) \\
&\beta_{k,x} \mid (z_k^{(x)} = z_k) \sim \mathbb{1}\{\beta_{k,x} = \beta^{\star}_{k,z_k}\}
\end{aligned}
$$

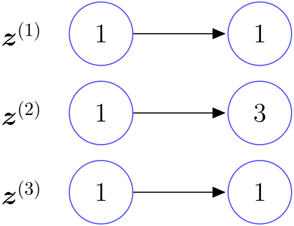
- Dynamic clustering given by time evolving latent variables

Local clustering: example



$$\beta_{1,1} = \beta_{1,2} = \beta_{1,3} = \beta^{\star}_{1,1}$$

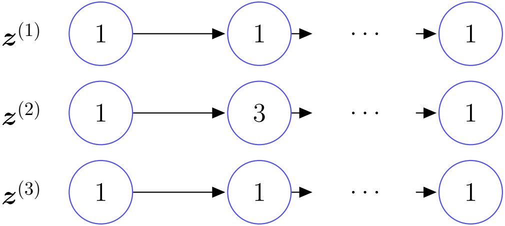
Local clustering: example



$$\beta_{2,1} = \beta_{2,3} = \beta^{\star}_{2,1}$$

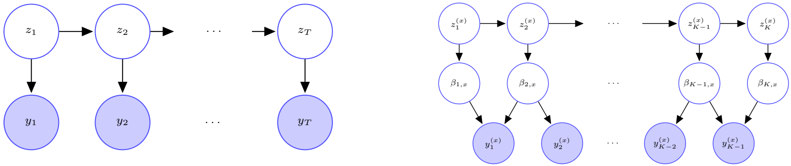
$$\beta_{2,2} = \beta^{\star}_{1,3}$$

Local clustering: example



$$\beta_{K,1} = \beta_{K,2} = \beta_{K,3} = \beta^{\star}_{K,1}$$

Local clustering: fixed effects

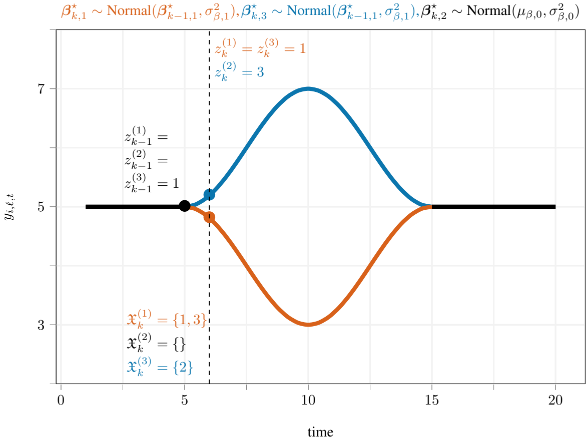


Conventional HMM and proposed HMM

$$y_{i,\ell,t} = f_x(t) + u_i(t) + \varepsilon_{i,\ell,t}$$
$$z_k^{(x)} \mid z_{k-1}^{(x)} = z_{k-1} \sim \text{Mult}(\pi_{z_{k-1},1}, \ldots, \pi_{z_{k-1},z_{max}})$$
$$(\pi_{z,1}, \ldots, \pi_{z,z_{max}}) \sim \text{Dir}(\alpha/z_{max}, \ldots, \alpha/z_{max})$$

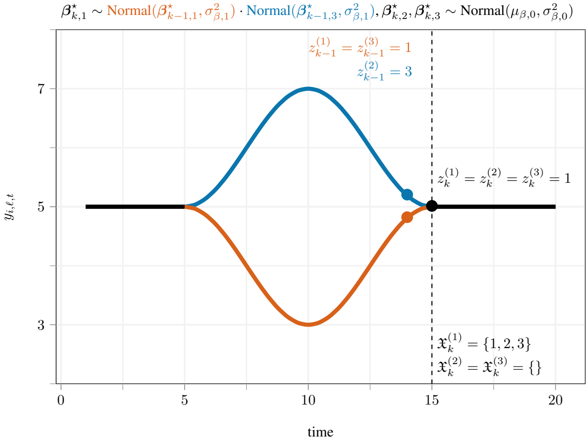
- Each level of the categorical predictor x is associated to a HMM for the group membership variables

Smoothing prior: fixed effects



- Markovian prior on the 'atoms' $\beta^\star_k$ to penalize first-order differences: favor smoothness

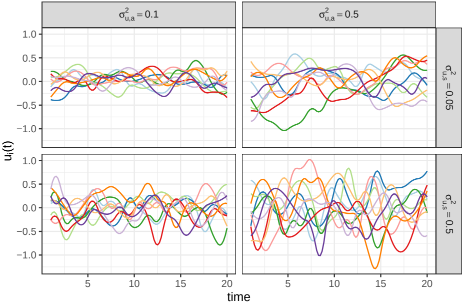
Smoothing prior: fixed effects



- Markovian prior on the 'atoms' $\beta^{\star}_k$ to penalize first-order differences: favor smoothness

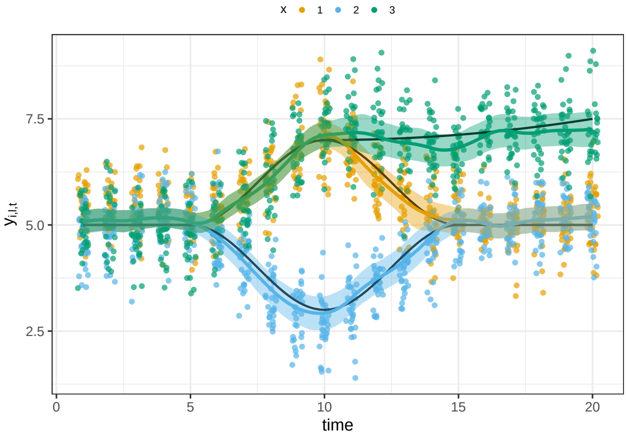
Random effects

$$y_{i,\ell,t} = f_x(t) + u_i(t) + \varepsilon_{i,\ell,t}$$
$$u_i(t) = \sum_{k=1}^{K} \beta_{k,u,i} B_k(t)$$
$$\boldsymbol{\beta}_{u,i} \sim \text{MVN}_K\{\mathbf{0}, (\sigma_{u,a}^{-2}\boldsymbol{I}_K + \sigma_{u,s}^{-2}\boldsymbol{P}_u)^{-1}\}$$
$$\sigma_{u,s}^2 \sim \mathcal{C}^+(0,1), \quad \sigma_{u,a}^2 \sim \mathcal{C}^+(0,1)$$



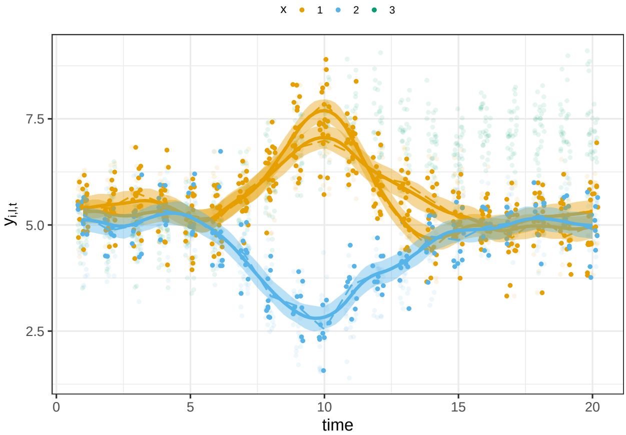
Each panel shows a collection of 10 random draws from the random effects distribution for a combination of $(\sigma_{u,s}^2, \sigma_{u,a}^2)$

Results: fixed effects



$$y_{i,\ell,t} = f_x(t) + u_i(t) + \varepsilon_{i,\ell,t}$$

Results: individual effects



$$y_{i,\ell,t} = f_x(t) + u_i(t) + \varepsilon_{i,\ell,t}$$

Possible extensions

How to generalize the previous ideas to multiple predictors $x_1, \ldots, x_p$?

- Redefine each combination of the levels of $(x_1, \ldots, x_p)$ as a level of a new predictor $x^\star$, and use the model for a single predictor.

  Drawbacks: cumbersome computation; impossibility to characterize local and global variable importance of the individual predictors.

Possible extensions

How to generalize the previous ideas to multiple predictors $x_1, \ldots, x_p$?

- Redefine each combination of the levels of $(x_1, \ldots, x_p)$ as a level of a new predictor $x^\star$, and use the model for a single predictor.

  Drawbacks: cumbersome computation; impossibility to characterize local and global variable importance of the individual predictors.

- Use an additive model: $y_{i,\ell,t} = f_{x_1}(t) + \cdots + f_{x_p}(t) + u_i(t) + \varepsilon_{i,\ell,t}$.

  Drawbacks: no interactions, no borrowing of information across levels of different predictors.

Possible extensions

How to generalize the previous ideas to multiple predictors $x_1, \ldots, x_p$?

- Redefine each combination of the levels of $(x_1, \ldots, x_p)$ as a level of a new predictor $x^\star$, and use the model for a single predictor.

  Drawbacks: cumbersome computation; impossibility to characterize local and global variable importance of the individual predictors.

- Use an additive model: $y_{i,\ell,t} = f_{x_1}(t) + \cdots + f_{x_p}(t) + u_i(t) + \varepsilon_{i,\ell,t}$.

  Drawbacks: no interactions, no borrowing of information across levels of different predictors.

- Use a **joint model**: $y_{i,\ell,t} = f_{x_1,\ldots,x_p}(t) + u_i(t) + \varepsilon_{i,\ell,t}$.

A joint longitudinal mixed model

- Multiple predictors $x_j \in \{1, \ldots, x_{j,\max}\}, j = 1, \ldots, p$

We consider the following generic class of longitudinal mixed models

$$\{y_{i,\ell,t} \mid x_{j,i,\ell,t} = x_j, j = 1, \ldots, p\} = \underbrace{f_{x_1,\ldots,x_p}(t)}_{\text{fixed effects}} + \overbrace{u_i(t)}^{\text{random effects}} + \underbrace{\varepsilon_{i,\ell,t}}_{\text{residuals}}$$

$$f_{x_1,\ldots,x_p}(t) = \sum_{k=1}^{K} \beta_{k,x_1,\ldots,x_p} B_k(t)$$

Proposed model:

- allows for **dynamic variable selection** and **partitioning** of the levels of the categorical variables
- accommodates **all order interactions** among the predictors

Why clustering?

Modeling all $K \prod_{j=1}^{p} x_{j,\max}$ parameters is impractical.

If we allow for identical values in some of the spline coefficients:

- we can reduce the number of parameters to be modeled
- we can establish that the some combinations of levels of $(x_1, \ldots, x_p)$ have no differential effect on the data generating mechanism
- we can borrow information across predictors and use more data to estimate the spline coefficients

Example

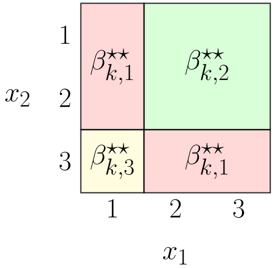
If $\boldsymbol{\beta}_{x_1,\ldots,x_{j,1},\ldots,x_p} = \boldsymbol{\beta}_{x_1,\ldots,x_{j,2},\ldots,x_p}$ for all combinations of $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p)$, then two levels $x_{j,1}$ and $x_{j,2}$ of x_j have no differential effect on the response.

Local clustering: fixed effects

We introduce local random partitions $\boldsymbol{\rho}_k = \{S_{k,1}, \ldots, S_{k,m_k}\}$ of $\mathcal{X}$, where m_k denotes the cardinality of $\boldsymbol{\rho}_k$.

$$\{y_{i,\ell,t} \mid x_{j,i,\ell,t} = x_j, j = 1, \ldots, p\} = f_{x_1,\ldots,x_p}(t) + u_i(t) + \varepsilon_{i,\ell,t}$$

$$\{f_{x_1,\ldots,x_p}(t) \mid (x_1, \ldots, x_p) \in S_{k,h}, k = 1, \ldots, K\} = \sum_{k=1}^{K} \beta^{\star\star}_{k,h} B_k(t)$$



$S_{k,1} = \{(1,1), (1,2), (2,3), (3,3)\}$

$S_{k,2} = \{(2,1), (2,2), (3,1), (3,2)\}$

$S_{k,3} = \{(1,3)\}$

Local clustering: multiple predictors

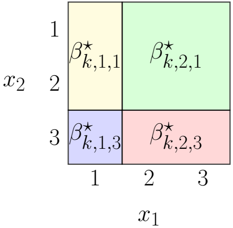
How to induce a partition $\boldsymbol{\rho}_k = \{S_{k,1}, \ldots, S_{k,m_k}\}$ of the predictor space $\mathcal{X}$?

$$z_{1,k}^{(1)} = 1$$

$$z_{1,k}^{(2)} = z_{1,k}^{(3)} = 2$$

$$z_{2,k}^{(1)} = z_{2,k}^{(2)} = 1$$

$$z_{2,k}^{(3)} = 3$$



Local clustering: multiple predictors

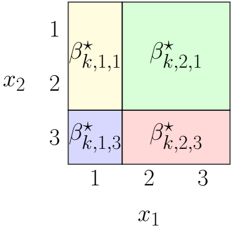
How to induce a partition $\boldsymbol{\rho}_k = \{S_{k,1}, \ldots, S_{k,m_k}\}$ of the predictor space $\mathcal{X}$?

$$z_{1,k}^{(1)} = 1$$

$$z_{1,k}^{(2)} = z_{1,k}^{(3)} = 2$$

$$z_{2,k}^{(1)} = z_{2,k}^{(2)} = 1$$

$$z_{2,k}^{(3)} = 3$$

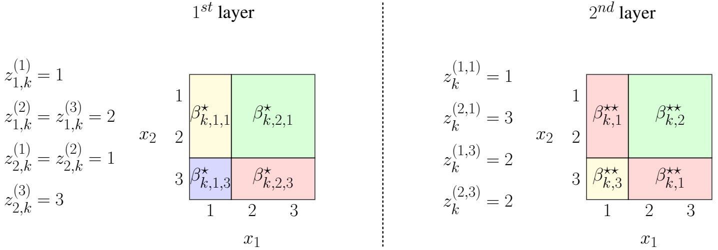


Limitation: joint partition is expressed as the product of p independent marginal partitions.

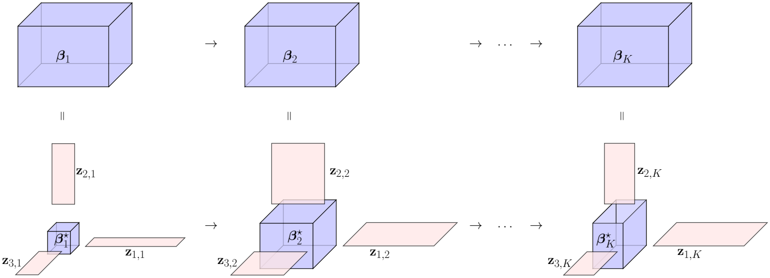
Dynamic partition model

Solution: introduce an additional latent allocation variable $z_k^{(z_{1,k},\ldots,z_{p,k})}$ to make clustering fully flexible. It allows to find coarser partitions by grouping levels across categorical predictors.

$$\{\beta^{\star}_{k,z_{1,k},\ldots,z_{p,k}} \mid (z_{1,k}^{(x_1)},\ldots,z_{p,k}^{(x_p)}) = (z_{1,k},\ldots,z_{p,k}), z_k^{(z_{1,k},\ldots,z_{p,k})} = z_k\} = \beta^{\star\star}_{k,z_k}$$



Tensor view



$$\{\beta_{k,x_1,\ldots,x_p} \mid z_{j,k}^{(x_j)}, j = 1, \ldots, p\} = \sum_{z_{1,k}} \cdots \sum_{z_{p,k}} \beta^{\star}_{k,z_{1,k},\ldots,z_{p,k}} \prod_{j=1}^{p} 1\{z_{j,k}^{(x_j)} = z_{j,k}\}$$

Theoretical properties

Consider $||f||_{2,g,loc} = \sum_{\boldsymbol{x} \in \mathcal{X}} g(x) \sum_{k=1}^{K} f_{\boldsymbol{x}}^2(k)$.

Theoretical properties

Consider $||f||_{2,g,loc} = \sum_{\boldsymbol{x} \in \mathcal{X}} g(x) \sum_{k=1}^{K} f_{\boldsymbol{x}}^2(k)$.

We can correctly estimate the underlying functions.

Function estimation

For any $\varepsilon > 0$, $\Pi(||f - f_0||_{2,g,loc} < \varepsilon \mid \text{data}) \to 1$.

Theoretical properties

Consider $||f||_{2,g,loc} = \sum_{\boldsymbol{x} \in \mathcal{X}} g(x) \sum_{k=1}^{K} f_{\boldsymbol{x}}^2(k)$.

We can correctly estimate the underlying functions.

Function estimation

For any $\varepsilon > 0$, $\Pi(||f - f_0||_{2,g,loc} < \varepsilon \mid \text{data}) \to 1$.

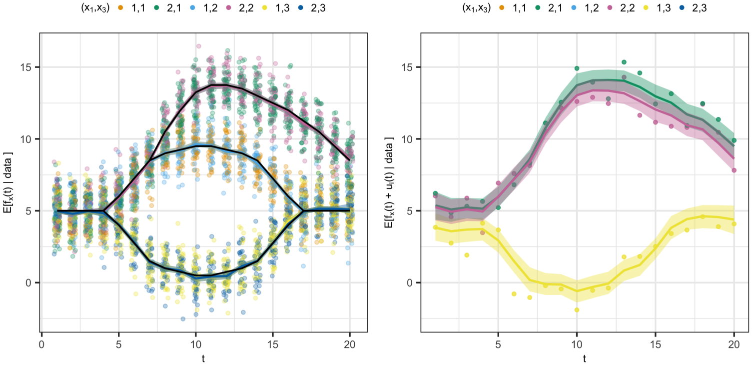
We can also identify the local partitions.

Local partitions

$\Pi(\boldsymbol{\rho}_k = \boldsymbol{\rho}_{k,0} \mid \text{data}) \to 1$.

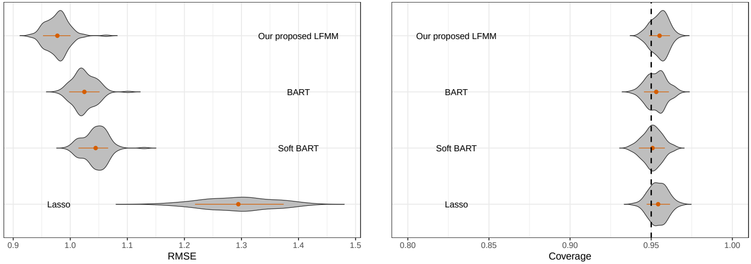
Consistency in recovering the local partitions implies consistency in **local variable selection**.

Results: synthetic data



- The model correctly identifies the only two important predictors and does not include the eight redundant ones

Results: synthetic data



- The out-of-sample predictive performance is uniformly better than the one of popular models for nonparametric regression

Drift-diffusion models for tone learning

Paulon, G., Reetzke, R., Chandrasekaran, B., & Sarkar, A. (2018). Functional logistic mixed-effects models for learning curves from longitudinal binary data. *Journal of Speech, Language, and Hearing Research, 62, 543-553*.

Paulon, G., Llanos, F., Chandrasekaran, B., & Sarkar, A. (2020). Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults. *Journal of the American Statistical Association*.

Roark, C., Paulon, G., Sarkar, A., & Chandrasekaran, B. (2021). Comparing perceptual category learning across modalities in the same individuals. *Psychonomic Bulletin & Review*.

Motivating study

- Mandarin is a **tonal language**: every syllable has four different tones which convey different lexical meanings

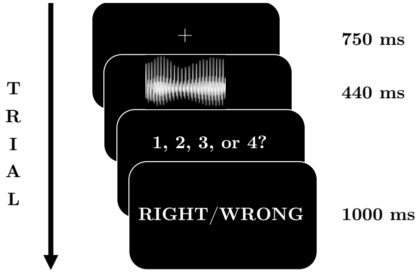
Goal: understand learning of novel speech categories in adulthood

In general: neural commitment to native-language speech sounds may preclude the learning of novel speech categories in adulthood.

Auditory tone learning experiments

Features of **tone learning** experiments:

- exposure to perceptually variable tones
- trial-by-trial corrective feedback

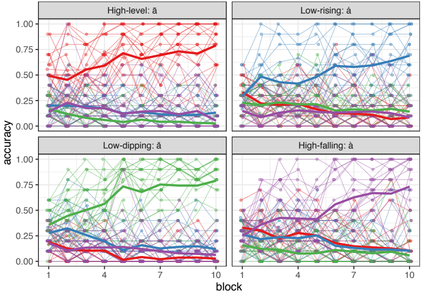


t	i	ℓ	s	d	τ
1	1	1	2	1	1.2
1	1	2	3	3	0.9
			⋮		

Improvement of tone categorization within a few hundred trials

Auditory tone learning experiments

- Learned by 20 non-native speakers over a period of $\approx$ 20 days
- We focus on the first two days, when critical differences emerge
- Data comprise **final responses** and associated **response times**

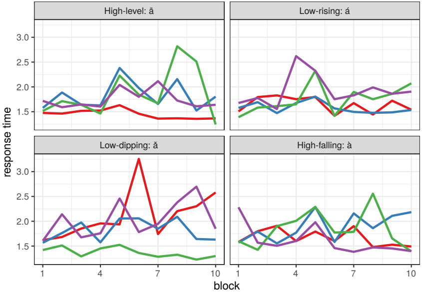


Proportions of times an input tone was classified into different tone categories by different subjects

How do the stimuli affect the perceptual mechanisms at different longitudinal stages?

Auditory tone learning experiments

- Learned by 20 non-native speakers over a period of $\approx$ 20 days
- We focus on the first two days, when critical differences emerge
- Data comprise **final responses** and associated **response times**



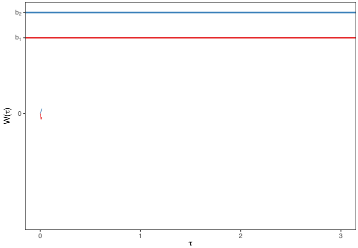
Response times averaged across subjects

How do the stimuli affect the perceptual mechanisms at different longitudinal stages?

Two-choice decision tasks

Neural assumptions of perceptual decision making:

1. evidence is accumulated over time via increased firing of the neurons ⇒ multiple racing[3] stochastic processes $W(\tau)$
2. a decision is made when firing rates reach a threshold ⇒ sufficient evidence has accumulated favoring one alternative over the other



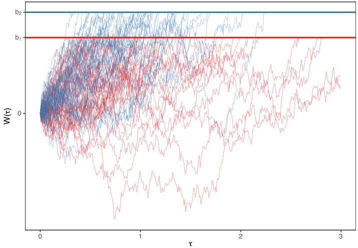
- Response **category**: first threshold to be reached
- Response **time** (RT): first-passage time

[3] Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological review, 108*.

Two-choice decision tasks

Neural assumptions of perceptual decision making:

1. evidence is accumulated over time via increased firing of the neurons $\Rightarrow$ multiple racing[3] stochastic processes $W(\tau)$
2. a decision is made when firing rates reach a threshold $\Rightarrow$ sufficient evidence has accumulated favoring one alternative over the other



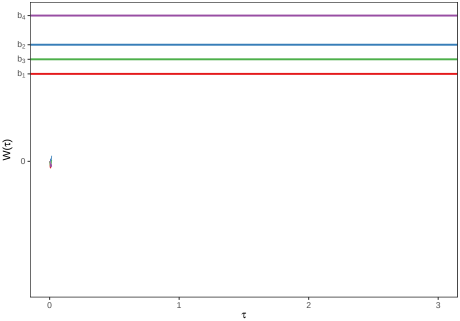
- Response **category**: first threshold to be reached
- Response **time** (RT): first-passage time

[3] Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological review, 108*.

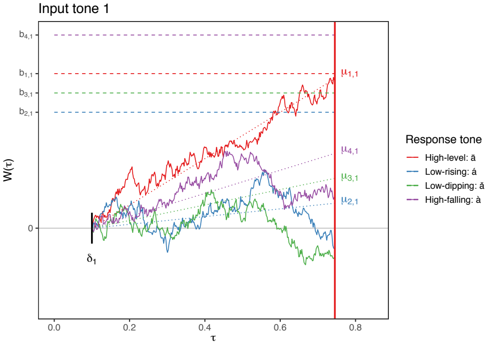
Our proposal

Extend the existing literature to:

1. multiple-choice tasks
2. dynamic settings: $\boldsymbol{\theta}_{d,s} \longrightarrow \boldsymbol{\theta}_{d,s}(t)$
3. subject-specific heterogeneity: $\boldsymbol{\theta}_{d,s}(t) \longrightarrow \boldsymbol{\theta}_{d,s}^{(i)}(t)$
4. local clustering



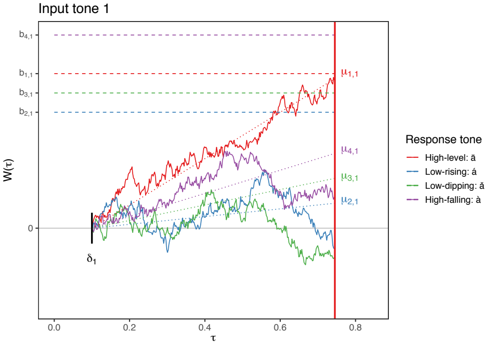
Drift-diffusion models



- When presented with the stimulus s, a decision d is reached at response time τ if the corresponding threshold $b_{d,s}$ is crossed **first**

$$\begin{cases} \tau = \min_{d'} \tau_{d'} \\ d = \arg\min_{d'} \tau_{d'} \end{cases}$$

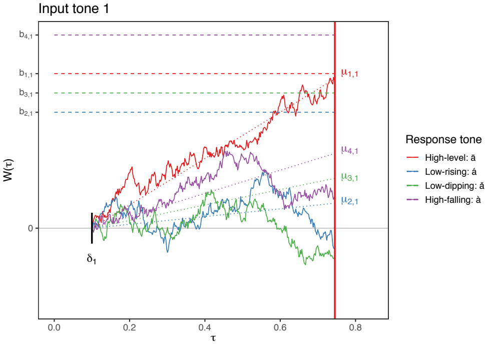
Drift-diffusion models



- δ_s: **offset time** irrelevant to the decision making process (e.g. encode an input signal s before accumulation starts, press a computer key).

  Here the auditory stimulus $s = 1$, i.e. we focus on the input tone T1. In our experiment, $s \in \{1, \ldots, 4\}$.

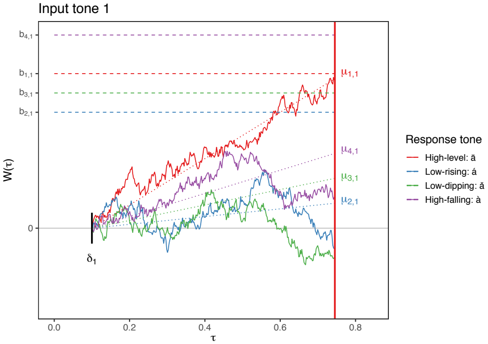
Drift-diffusion models



- $\mu_{d,s}$: **drift rate** at which the brain accumulates evidence in favor of d when the stimulus is s. Related to the firing rate of neurons.

  Here the auditory stimulus $s = 1$, i.e. we focus on the input tone T1. In our experiment, $s \in \{1, \ldots, 4\}; d \in \{1, \ldots, 4\}$.

Drift-diffusion models



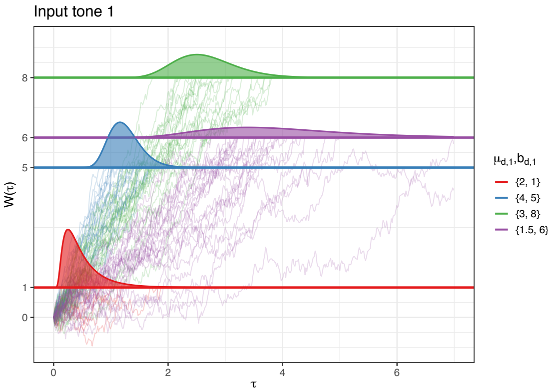
- $b_{d,s}$: **boundary parameter**, i.e. amount of evidence needed to decide in favor of d when the stimulus is s.

  Here the auditory stimulus $s = 1$, i.e. we focus on the input tone T1. In our experiment, $s \in \{1, \ldots, 4\}; d \in \{1, \ldots, 4\}$.

Drift-diffusion models

The time τ_d to reach the d^{th} response boundary under the s^{th} input stimulus is distributed as an inverse-Gaussian with p.d.f.

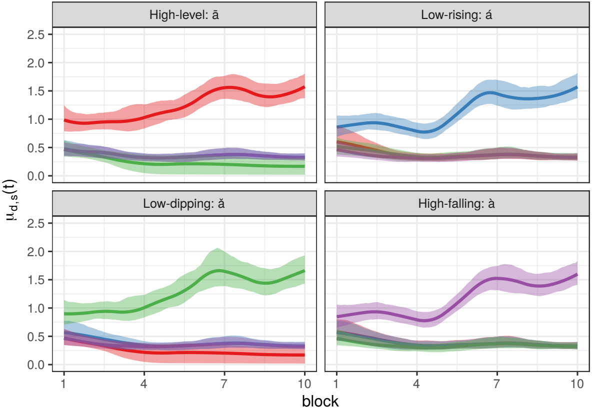
$$g(\tau_d \mid \boldsymbol{\theta}_{d,s}) = \frac{b_{d,s}}{\sqrt{2\pi}}(\tau_d - \delta_s)^{-3/2} \exp\left[-\frac{\{b_{d,s} - \mu_{d,s}(\tau_d - \delta_s)\}^2}{2(\tau_d - \delta_s)}\right].$$



$$\mathbb{E}[\tau_d \mid \boldsymbol{\theta}_{d,s}] = b_{d,s}/\mu_{d,s}$$

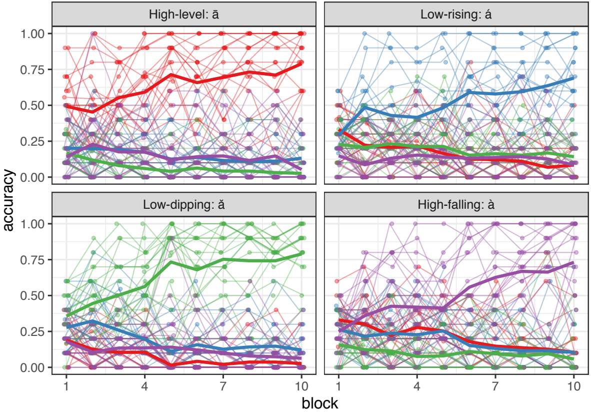
$$\text{var}(\tau_d \mid \boldsymbol{\theta}_{d,s}) = b_{d,s}/\mu_{d,s}^3$$

Results: drift parameters



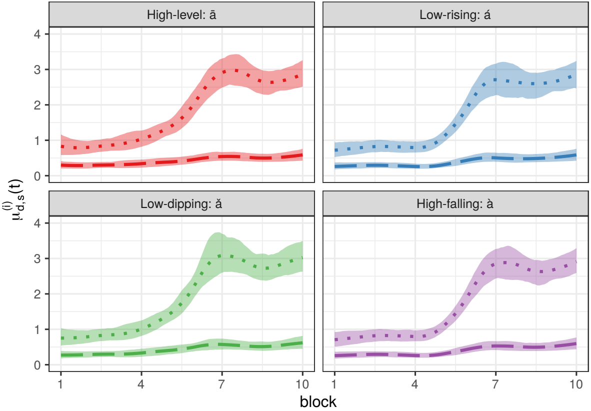
Estimated posterior mean and 95% CI for the population drift parameters $\mu_{d,s}(t)$

Results: drift parameters



Proportions of times an input tone was classified into different tone categories by different subjects

Results: individual level parameters



Estimated posterior mean for the individual drift parameters $\mu_{d,s}^{(i)}(t)$ obtained for two different participants, indicated by different line types

Bivariate survival regression for current status data

Paulon, G., Müller, P., & Sal y Rosas, V. G. (2021). Bayesian nonparametric bivariate survival regression for current status data. *Submitted*.

Motivating study

Randomized controlled trial to assess the effect of partner-notification strategies[4]:

- **expedited-treatment**: offered medication to give to their partners, or staff members contacted partners and provided them with medication without a clinical examination
- **standard partner referral**: advised to refer their partners for treatment

Follow-up visits for $n = 1864$ participants treated for recurrent infections:

- 933 assigned to standard therapy
- 931 assigned to expedited partner therapy

[4] Golden, M. R., Whittington, W. L., Handsfield, H. H., Hughes, J. P., Stamm, W. E., Hogben, M., … Thomas, K. K., et al. (2005). Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infection. *New England Journal of Medicine*, *352*, 676–685.

Current status data

Primary endpoints of the study: time to symptoms S_i and time to re-infection I_i.

When visiting the hospital at time C_i, two outcomes are recorded:

- Does the patient test positive for re-infection? $\Delta_{I_i} = \mathbb{1}(I_i < C_i)$
- Is the patient experiencing symptoms? $\Delta_{S_i} = \mathbb{1}(S_i < C_i)$

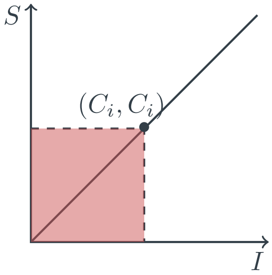
Available information on the outcomes: whether or not they exceed a common monitoring time $C_i \rightarrow$ bivariate current status data.

This is a very common source of data

Current status data

Four possible cases:

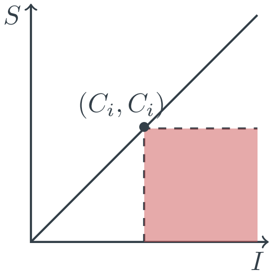
1. $(\Delta_{I_i}, \Delta_{S_i}) = (1, 1)$: positive test and presence of symptoms (symptomatic infections)



Current status data

Four possible cases:

2. $(\Delta_{I_i}, \Delta_{S_i}) = (0, 1)$: negative test but presence of symptoms (due to other causes)

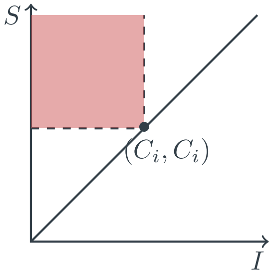


In this case $I_i > S_i$ and the symptoms must be due to other causes.

Current status data

Four possible cases:

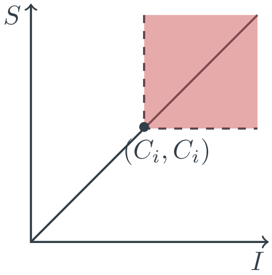
3. $(\Delta_{I_i}, \Delta_{S_i}) = (1, 0)$: positive test but absence of symptoms (asymptomatic infections)



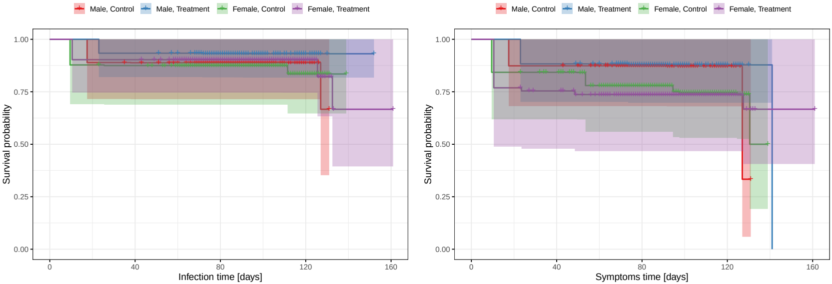
Current status data

Four possible cases:

4. $(\Delta_{I_i}, \Delta_{S_i}) = (0, 0)$: negative test and absence of symptoms



Issue 1: censoring misspecification



Nonparametric MLE

Let us focus on the univariate case:

- S_i: time to symptoms for patient i
- Δ_{S_i}: censoring indicator
- C_i: censoring time
  - $\Delta_{S_i} = 1$ if $S_i \leq C_i$ (left censoring)
  - $\Delta_{S_i} = 0$ if $S_i > C_i$ (right censoring)

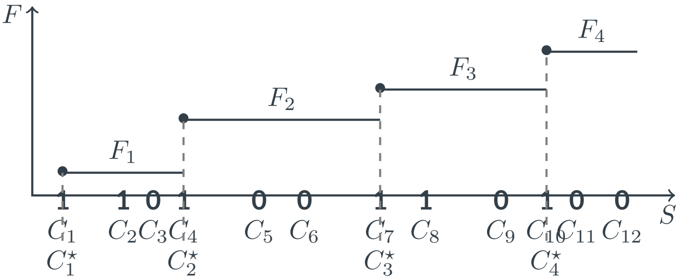
Goal: infer the unknown distribution $f_S(s)$ based on $\{(C_i, \Delta_{S_i})\}_{i=1}^n$, $i = 1, \ldots, n$.

W.l.o.g., we assume that the censoring times are ordered, $C_{i-1} \leq C_i$.

Nonparametric MLE

Define $A = \{i > 1 \text{ s.t. } \Delta_{S_i} = 1, \Delta_{S_{i-1}} = 0\} \cup \{1\}$, the set of indices of left censored observations immediately following a right censored observation.

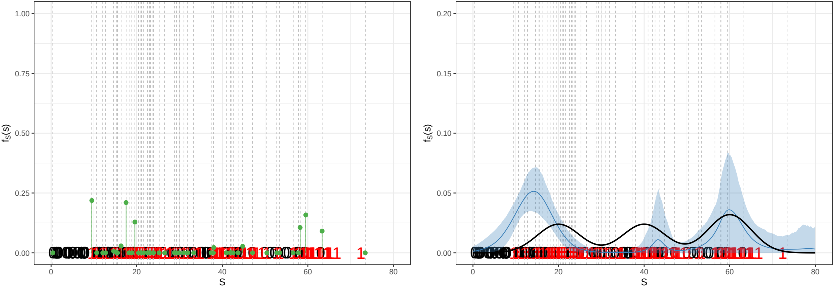
Let $J = |A|$ and $\boldsymbol{C}^\star = (C_1^\star, \ldots, C_J^\star) = (C_i, i \in A)$ denote the corresponding censoring times.



Easy to show that $f_S(S) = \sum_{j=1}^{J+1} p_k \delta_{C_j^\star}$.

Nonparametric MLE

We can implement a simple EM algorithm[5].



[5] Groeneboom, P., & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation.*

Proposed solution

Time of visit to the hospital can either occur:

- **uniformly** in the observation range (visit by protocol)

$$C_i \mid S_i, \lambda = \min\{\textsf{Unif}(A, B);$$

Proposed solution

Time of visit to the hospital can either occur:

- **uniformly** in the observation range (visit by protocol)
- closely **following** the symptoms onset (visit prompted by symptoms)

$$C_i \mid S_i, \lambda = \min\{\mathsf{Unif}(A, B); S_i + \mathsf{Exp}(\lambda)\}$$

Proposed solution

Time of visit to the hospital can either occur:

- **uniformly** in the observation range (visit by protocol)
- closely **following** the symptoms onset (visit prompted by symptoms)

$$C_i \mid S_i, \lambda = \min\{\text{Unif}(A, B); S_i + \text{Exp}(\lambda)\}$$

The p.d.f. of the conditional distribution of censoring times given the event times is

$$f_{C|S}(c \mid s) = \frac{\mathbb{1}\{c \leq s\}}{B - A} + \frac{\mathbb{1}\{c > s\}}{B - A} e^{-\lambda(c-s)}\{1 + \lambda(B - c)\}.$$

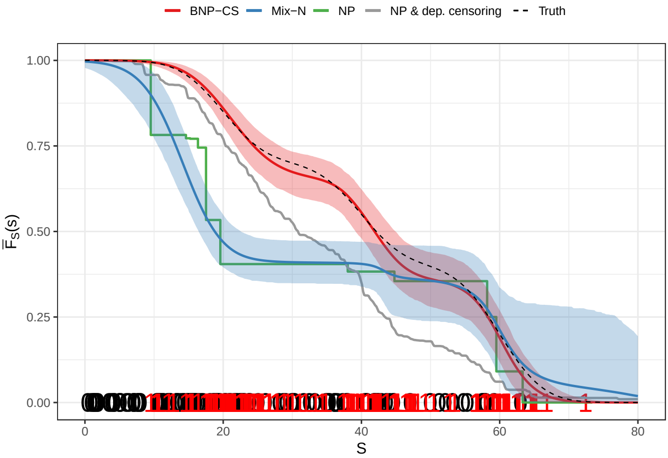
The setting

Prior shrinkage via BNP model

$$
\begin{aligned}
& S_i \mid H \sim \int \mathsf{N}(S_i \mid \mu, \sigma^2) dH(\mu, \sigma^2), \quad H \sim \mathsf{DP}(M, H_0) \\
\Rightarrow \quad & S_i \mid \{\mu_k, \sigma_k^2, \pi_k\}_{k=1}^{+\infty} \sim \sum_{k=1}^{+\infty} \pi_k \mathsf{N}(S_i \mid \mu_k, \sigma_k^2)
\end{aligned}
$$

- Introduce heterogeneity
- Can include regression on covariates

Results: synthetic data



Issue 2: nonparametric unidentifiability

Back to the bivariate case: we only observe the common monitoring time C_i and the two indicators $\Delta_{I_i} = \mathbb{1}\{I_i < C_i\}$, $\Delta_{S_i} = \mathbb{1}\{S_i < C_i\}$.

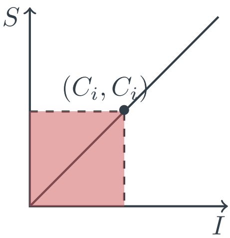
The conditional likelihood only depends on $F_1(c) = P(I_i \leq c)$, $F_2(c) = P(S_i \leq c)$, $F_3(c) = P(I_i \leq c, S_i \leq c)$.

Issue 2: nonparametric unidentifiability

Back to the bivariate case: we only observe the common monitoring time C_i and the two indicators $\Delta_{I_i} = \mathbb{1}\{I_i < C_i\}, \Delta_{S_i} = \mathbb{1}\{S_i < C_i\}$.

The conditional likelihood only depends on $F_1(c) = P(I_i \le c)$, $F_2(c) = P(S_i \le c)$, $F_3(c) = P(I_i \le c, S_i \le c)$.

$$F = \prod_{i=1}^{n} \{F_3(c_i)^{\Delta_{I_i}\Delta_{S_i}}$$

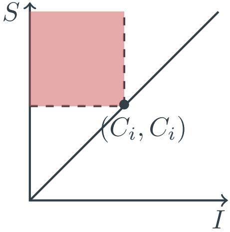


Issue 2: nonparametric unidentifiability

Back to the bivariate case: we only observe the common monitoring time C_i and the two indicators $\Delta_{I_i} = \mathbb{1}\{I_i < C_i\}, \Delta_{S_i} = \mathbb{1}\{S_i < C_i\}$.

The conditional likelihood only depends on $F_1(c) = P(I_i \leq c)$, $F_2(c) = P(S_i \leq c)$, $F_3(c) = P(I_i \leq c, S_i \leq c)$.

$$F = \prod_{i=1}^{n} \{F_3(c_i)^{\Delta_{I_i}\Delta_{S_i}} (F_1 - F_3)(c_i)^{\Delta_{I_i}(1-\Delta_{S_i})}$$



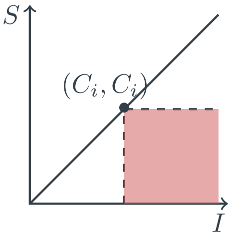
Issue 2: nonparametric unidentifiability

Back to the bivariate case: we only observe the common monitoring time C_i and the two indicators $\Delta_{I_i} = \mathbb{1}\{I_i < C_i\}$, $\Delta_{S_i} = \mathbb{1}\{S_i < C_i\}$.

The conditional likelihood only depends on $F_1(c) = P(I_i \leq c)$, $F_2(c) = P(S_i \leq c)$, $F_3(c) = P(I_i \leq c, S_i \leq c)$.

$$F = \prod_{i=1}^{n} \{F_3(c_i)^{\Delta_{I_i}\Delta_{S_i}} (F_1 - F_3)(c_i)^{\Delta_{I_i}(1-\Delta_{S_i})}$$

$$(F_2 - F_3)(c_i)^{(1-\Delta_{I_i})\Delta_{S_i}}$$

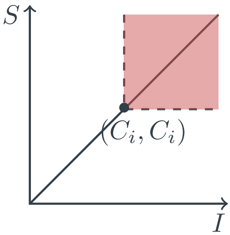


Issue 2: nonparametric unidentifiability

Back to the bivariate case: we only observe the common monitoring time C_i and the two indicators $\Delta_{I_i} = \mathbb{1}\{I_i < C_i\}, \Delta_{S_i} = \mathbb{1}\{S_i < C_i\}$.

The conditional likelihood only depends on $F_1(c) = P(I_i \leq c)$, $F_2(c) = P(S_i \leq c)$, $F_3(c) = P(I_i \leq c, S_i \leq c)$.

$$F = \prod_{i=1}^{n} \{F_3(c_i)^{\Delta_{I_i}\Delta_{S_i}} (F_1 - F_3)(c_i)^{\Delta_{I_i}(1-\Delta_{S_i})}$$
$$(F_2 - F_3)(c_i)^{(1-\Delta_{I_i})\Delta_{S_i}} (1 + F_3 - F_1 - F_2)(c_i)^{(1-\Delta_{I_i})(1-\Delta_{S_i})}\}$$



Possible solutions

Identifiability issues with the nonparametric m.l.e. We can only identify:

- the marginals $F_1(c)$ and $F_2(c)$
- the measure of dependence $(F_3 - F_1 F_2)(c)$.

Possible strategies:

1. estimate the joint distribution under parametric or semiparametric assumptions
2. build the joint model from the two identifiable marginal distributions and a choice for their dependence structure[6] that reflects the underlying biology

[6] Wang, W., & Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika*, *87*, 879–893.

A bivariate event time model

Two causes of symptoms:

- due to other causes

- due to disease

$$f_{IS}(I,S) = w f^{\star}_{IS}(I,S) \; + \; (1-w) \underbrace{f'_{IS}(I,S)}_{\text{constrained to } I<S}$$

A bivariate event time model

Two causes of symptoms:

- due to other causes $\rightarrow$ time to infection and time to symptoms are independent
- due to disease $\rightarrow$ time to infection and latency time are independent

$$f_{IS}(I,S) = w \underbrace{f^{\star}_{IS}(I,S)}_{I \perp S} + (1-w) \underbrace{f'_{IS}(I,S)}_{\substack{I < S \\ I \perp L = S - I}}$$

A bivariate event time model

Two causes of symptoms:

- due to other causes $\rightarrow$ time to infection and time to symptoms are independent
- due to disease $\rightarrow$ time to infection and latency time are independent

$$f_{IS}(I,S) = w \underbrace{f^{\star}_{IS}(I,S)}_{I \perp S} + (1-w) \underbrace{f'_{IS}(I,S)}_{\substack{I < S \\ I \perp L = S - I}}$$

$$= w f_I(I) f^{\star}_S(S) + (1-w) f_I(I) f_L(S-I)$$

A bivariate event time model

Two causes of symptoms:

- due to other causes $\rightarrow$ time to infection and time to symptoms are independent
- due to disease $\rightarrow$ time to infection and latency time are independent

$$f_{IS}(I,S) = w \underbrace{f^{\star}_{IS}(I,S)}_{I \perp S} + (1-w) \underbrace{f'_{IS}(I,S)}_{\substack{I < S \\ I \perp L = S - I}}$$
$$= w f_I(I) f^{\star}_S(S) + (1-w) f_I(I) f_L(S-I)$$

Positivity constraint on the latency time:

$$L \mid \lambda_L \sim \mathsf{Exp}(\lambda_L)$$

Building dependence structure

Nonparametric priors:

$$f_S^\star(S) = \sum_{k=1}^{+\infty} \pi_k^{(S)} \mathrm{N}(S \mid \mu_k^{(S)}, \sigma_k^{(S)2})$$

$$f_I(I) = \sum_{k=1}^{+\infty} \pi_k^{(I)} \mathrm{N}(I \mid \mu_k^{(I)}, \sigma_k^{(I)2})$$

Building dependence structure

Nonparametric priors:

$$f^{\star}_S(S) = \sum_{k=1}^{+\infty} \pi_k^{(S)} \mathrm{N}(S \mid \mu_k^{(S)}, \sigma_k^{(S)2})$$

$$f_I(I) = \sum_{k=1}^{+\infty} \pi_k^{(I)} \mathrm{N}(I \mid \mu_k^{(I)}, \sigma_k^{(I)2})$$

Dependent censoring on S only:

$$C_i \mid S_i, \lambda = \min\{\mathsf{Unif}(A, B); S_i + \mathsf{Exp}(\lambda)\}$$

Building dependence structure

Nonparametric priors:

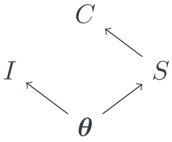
$$f_S^\star(S) = \sum_{k=1}^{+\infty} \pi_k^{(S)} \mathrm{N}(S \mid \mu_k^{(S)}, \sigma_k^{(S)2})$$

$$f_I(I) = \sum_{k=1}^{+\infty} \pi_k^{(I)} \mathrm{N}(I \mid \mu_k^{(I)}, \sigma_k^{(I)2})$$

Dependent censoring on S only:

$$C_i \mid S_i, \lambda = \min\{\mathsf{Unif}(A, B); S_i + \mathsf{Exp}(\lambda)\}$$

Censoring regularizes inference on both f_I and f_S:



Building dependence structure

The implied marginal distributions are

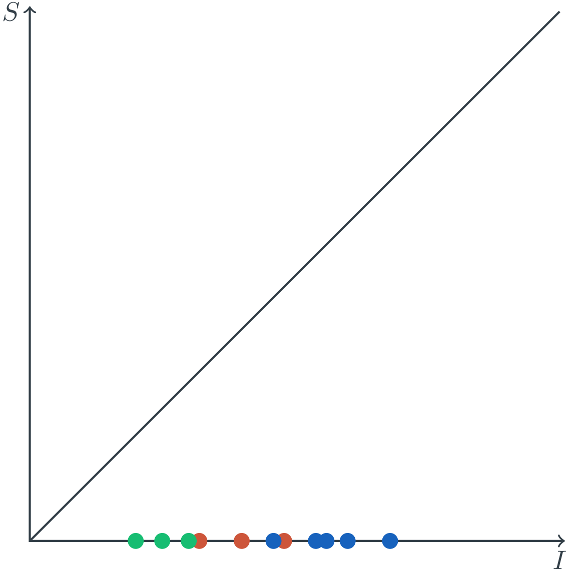
$$f_I(I) = \sum_{k=1}^{+\infty} \pi_k^{(I)} \mathrm{N}(I \mid \mu_k^{(I)}, \sigma_k^{(I)2})$$

$$f_S(S) = w \sum_{k=1}^{+\infty} \pi_k^{(S)} \mathrm{N}(S \mid \mu_k^{(S)}, \sigma_k^{(S)2}) + (1-w) \sum_{k=1}^{+\infty} \pi_k^{(I)} \mathrm{EMG}(S \mid \mu_k^{(I)}, \sigma_k^{(I)2}, \lambda_L),$$

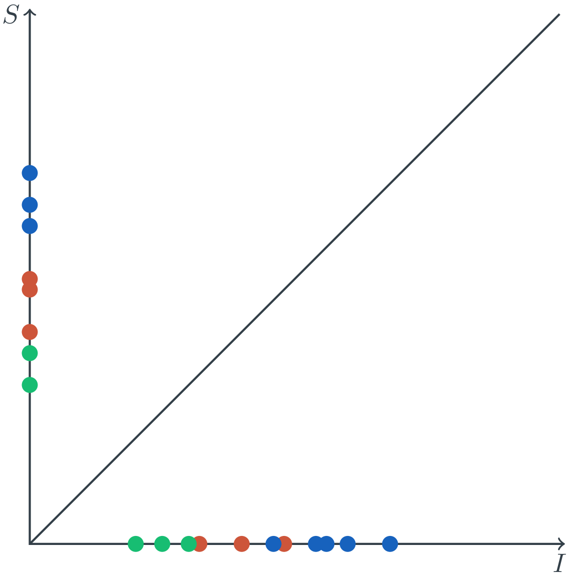
where $\mathrm{EMG}(\mu, \sigma^2, \lambda)$ denotes the exponentially modified Gaussian distribution[7].

[7] Grushka, E. (1972). Characterization of exponentially modified Gaussian peaks in chromatography. *Analytical Chemistry*, *44*, 1733–1738.

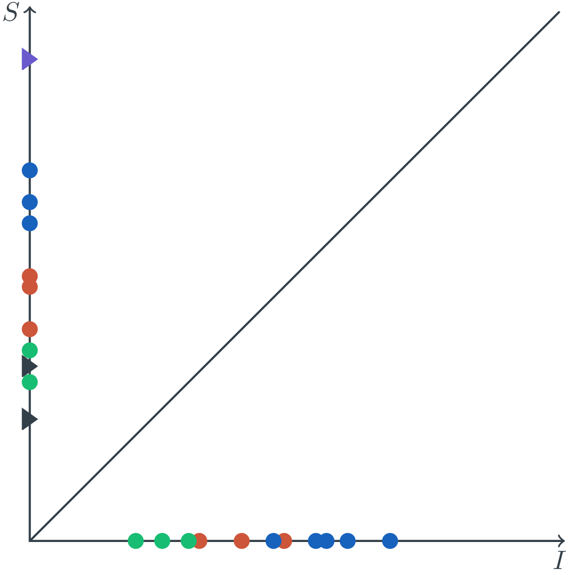
Dependent partitions



Dependent partitions



Dependent partitions



Regression on covariates

- Predictors: $\boldsymbol{x}_i = \{gender, arm, age\}$; $\boldsymbol{d} = (1, x_1, x_2, x_3)^\intercal$.

Extend the BNP priors to the families of r.p.m.s $\{H_{\boldsymbol{x}}^{(I)}, H_{\boldsymbol{x}}^{(S)}, \boldsymbol{x} \in \mathcal{X}\}$.

Regression on covariates

- Predictors: $\boldsymbol{x}_i = \{gender, arm, age\}$; $\boldsymbol{d} = (1, x_1, x_2, x_3)^\intercal$.

Extend the BNP priors to the families of r.p.m.s $\{H_{\boldsymbol{x}}^{(I)}, H_{\boldsymbol{x}}^{(S)}, \boldsymbol{x} \in \mathcal{X}\}$.

We use a dependent DP (DDP) prior with common weights and covariate dependent atoms, i.e.

$$H_x = \sum_k \pi_k \delta_{\boldsymbol{d}^\intercal \boldsymbol{m}_k}$$

where

$$\boldsymbol{d}^\intercal \boldsymbol{m}_k = \delta_k + \alpha_k x_1 + \beta_k x_2 + \gamma_k x_3$$

Regression on covariates

- Predictors: $\boldsymbol{x}_i = \{gender, arm, age\}$; $\boldsymbol{d} = (1, x_1, x_2, x_3)^\intercal$.

Extend the BNP priors to the families of r.p.m.s $\{H_{\boldsymbol{x}}^{(I)}, H_{\boldsymbol{x}}^{(S)}, \boldsymbol{x} \in \mathcal{X}\}$.

We use a dependent DP (DDP) prior with common weights and covariate dependent atoms, i.e.

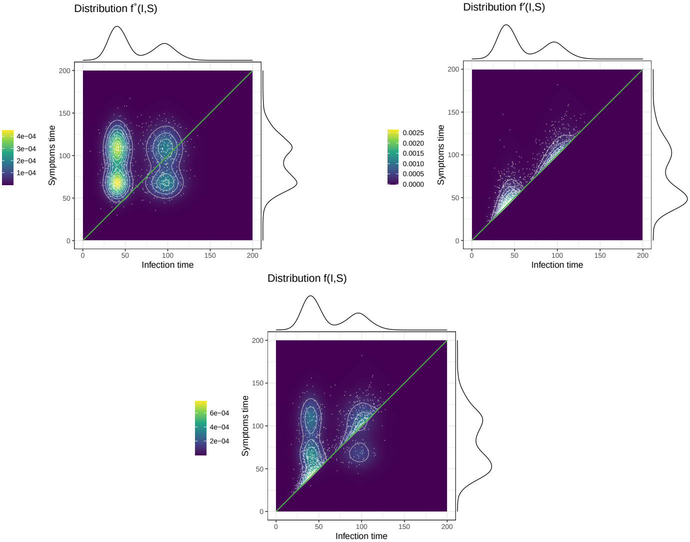
$$H_x = \sum_k \pi_k \delta_{\boldsymbol{d}^\intercal \boldsymbol{m}_k}$$

where

$$\boldsymbol{d}^\intercal \boldsymbol{m}_k = \delta_k + \alpha_k x_1 + \beta_k x_2 + \gamma_k x_3$$

- **Heterogeneous treatment effect**: H_x implies $H_\beta = \sum_k \pi_k \delta_{\beta_k}$

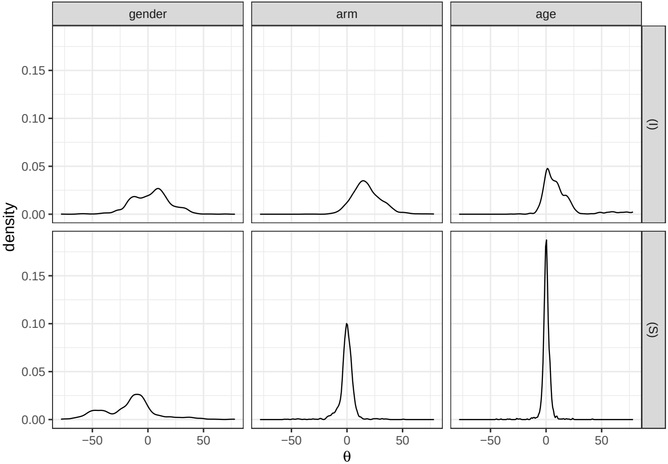
Results: synthetic data



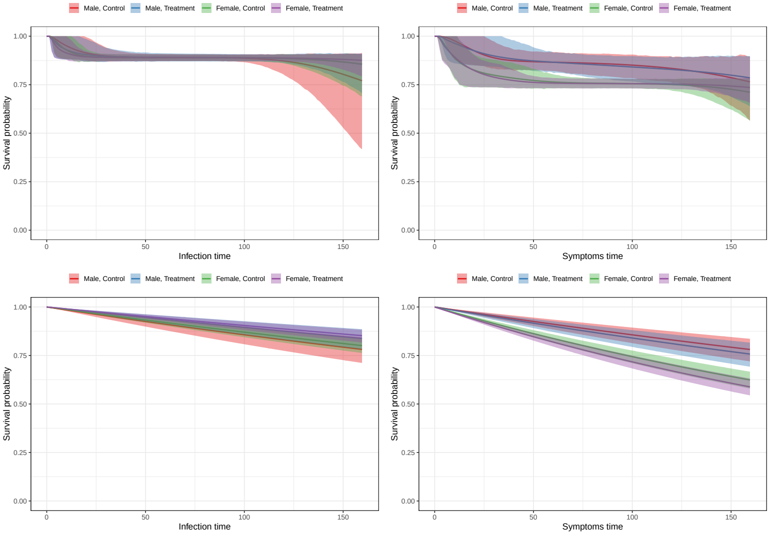
Results: synthetic data

	Sample Size	Distr.	De Iorio et al.	Bivariate Gumbel	Our method
(I)	$n = 250$	Inf.	1.64 (0.92, 3.01)	4.01 (3.14, 5.59)	1.10 (0.09, 2.24)
		Sym.	2.98 (1.11, 5.01)	6.15 (5.31, 8.77)	1.33 (0.18, 3.72)
	$n = 1000$	Inf.	1.32 (0.73, 1.90)	3.76 (3.19, 4.54)	0.50 (0.04, 1.80)
		Sym.	2.32 (1.19, 3.25)	5.99 (5.31, 6.99)	1.30 (0.54, 2.66)
(II)	$n = 250$	Inf.	0.96 (0.74, 1.56)	3.44 (3.08, 4.59)	0.99 (0.13, 2.07)
		Sym.	8.44 (5.21, 12.30)	11.75 (9.18, 18.01)	**0.76** (0.22, 2.16)
	$n = 1000$	Inf.	0.80 (0.50, 1.10)	3.12 (3.03, 3.41)	**0.19** (0.05, 0.50)
		Sym.	8.18 (6.28, 10.32)	10.74 (9.58, 12.49)	**0.12** (0.02, 0.37)
(III)	$n = 250$	Inf.	4.45 (3.00, 6.30)	4.24 (3.09, 5.79)	**0.45** (0.08, 1.14)
		Sym.	9.82 (6.70, 13.20)	8.08 (5.72, 12.15)	**0.24** (0.03, 0.81)
	$n = 1000$	Inf.	4.10 (3.18, 4.96)	3.96 (3.24, 4.81)	**0.13** (0.01, 0.35)
		Sym.	9.94 (8.44, 11.71)	7.98 (6.31, 10.06)	**0.05** (0.01, 0.15)

Results: partner notification study



Results: partner notification study



To sum up

Our approach:

- similar to copula models, with added interpretability, e.g. exponential rates
- combining BNP model with known structure provides sufficient **regularization**

Work in progress:

- generalization to different **correlation** structures

Conclusions

Conclusions

- Longitudinal partition models & Auditory neuroscience

Paulon, G., Reetzke, R., Chandrasekaran, B., & Sarkar, A. (2018). Functional logistic mixed-effects models for learning curves from longitudinal binary data. *Journal of Speech, Language, and Hearing Research, 62, 543-553*.

Paulon, G., Llanos, F., Chandrasekaran, B., & Sarkar, A. (2020). Bayesian semiparametric longitudinal drift-diffusion mixed models for tone learning in adults. *Journal of the American Statistical Association*.

Roark, C., Paulon, G., Sarkar, A., & Chandrasekaran, B. (2021). Comparing perceptual category learning across modalities in the same individuals. *Psychonomic Bulletin & Review*.

Paulon, G., Müller, P., & Sarkar, A. (2021). Bayesian semiparametric hidden Markov tensor partition models for local variable selection in longitudinal data. *Submitted*.

Conclusions

- Longitudinal partition models & Auditory neuroscience
- Bivariate survival regression & Recurrent hospitalization data

---

Paulon, G., Müller, P., & Sal y Rosas, V. G. (2021). Bayesian nonparametric bivariate survival regression for current status data. *Submitted*.

Conclusions

- Longitudinal partition models & Auditory neuroscience
- Bivariate survival regression & Recurrent hospitalization data
- Dependent mixture models

Paulon, G., Trippa, L., & Müller, P. (2018). Invited comment on "Bayesian cluster analysis: Point estimation and credible balls". *Bayesian Analysis*, 13, 590-593.

Pagani Zanini, C. T., Paulon, G., & Müller, P. (2021). Dependent mixtures: Modeling cell lineages. *In preparation*.

Thank you!

References

Gelfand, A. E., Kottas, A., & MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, *100*, 1021–1035.

Golden, M. R., Whittington, W. L., Handsfield, H. H., Hughes, J. P., Stamm, W. E., Hogben, M., … Thomas, K. K., et al. (2005). Effect of expedited treatment of sex partners on recurrent or persistent gonorrhea or chlamydial infection. *New England Journal of Medicine*, *352*, 676–685.

Groeneboom, P., & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*.

Grushka, E. (1972). Characterization of exponentially modified Gaussian peaks in chromatography. *Analytical Chemistry*, *44*, 1733–1738.

Petrone, S., Guindani, M., & Gelfand, A. E. (2009). Hybrid Dirichlet mixture models for functional data. *Journal of the Royal Statistical Society: Series B*, *71*, 755–782.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological review*, *108*.

Wang, W., & Ding, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika*, *87*, 879–893.

Appendix

Smoothing priors: temporal dynamics

- HMM for the temporal evolution of the latent local cluster indicators with p independent dynamics:

$$(z_{j,k}^{(x_j)} \mid z_{j,k-1}^{(x_j)} = z_{k-1}) \sim \mathsf{Mult}(\pi^{(j)}_{z_{k-1},1}, \ldots, \pi^{(j)}_{z_{k-1},z_{max}})$$
$$\boldsymbol{\pi}^{(j)} = (\pi^{(j)}_{z,1}, \ldots, \pi^{(j)}_{z,z_{max}})^{\mathsf{T}} \sim \mathsf{Dir}(\alpha^{(j)}/z_{max}, \ldots, \alpha^{(j)}/z_{max})$$
$$\alpha^{(j)} \sim \mathsf{Ga}(a_\alpha, b_\alpha)$$

Smoothing priors: temporal dynamics

- HMM for the temporal evolution of the latent local cluster indicators with p independent dynamics:

$$(z_{j,k}^{(x_j)} \mid z_{j,k-1}^{(x_j)} = z_{k-1}) \sim \mathsf{Mult}(\pi_{z_{k-1},1}^{(j)}, \ldots, \pi_{z_{k-1},z_{max}}^{(j)})$$

$$\boldsymbol{\pi}^{(j)} = (\pi_{z,1}^{(j)}, \ldots, \pi_{z,z_{max}}^{(j)})^{\mathsf{T}} \sim \mathsf{Dir}(\alpha^{(j)}/z_{max}, \ldots, \alpha^{(j)}/z_{max})$$

$$\alpha^{(j)} \sim \mathsf{Ga}(a_\alpha, b_\alpha)$$

- Explicit shrinkage priors on the covariate importance indicators:

$$\ell_{j,k} \propto \exp(-\varphi_j \ell_{j,k}) \mathbb{1}_{\{1,\ldots,x_{j,\max}\}}(\ell_{j,k}), \quad \varphi_j \sim \mathsf{Ga}(a_\varphi, b_\varphi)$$

Smoothing priors: temporal dynamics

- HMM for the temporal evolution of the latent local cluster indicators with p independent dynamics:

$$
\begin{aligned}
&(z_{j,k}^{(x_j)} \mid z_{j,k-1}^{(x_j)} = z_{k-1}) \sim \mathsf{Mult}(\pi_{z_{k-1},1}^{(j)}, \ldots, \pi_{z_{k-1},z_{max}}^{(j)}) \\
&\boldsymbol{\pi}^{(j)} = (\pi_{z,1}^{(j)}, \ldots, \pi_{z,z_{max}}^{(j)})^\mathsf{T} \sim \mathsf{Dir}(\alpha^{(j)}/z_{max}, \ldots, \alpha^{(j)}/z_{max}) \\
&\alpha^{(j)} \sim \mathsf{Ga}(a_\alpha, b_\alpha)
\end{aligned}
$$

- Explicit shrinkage priors on the covariate importance indicators:

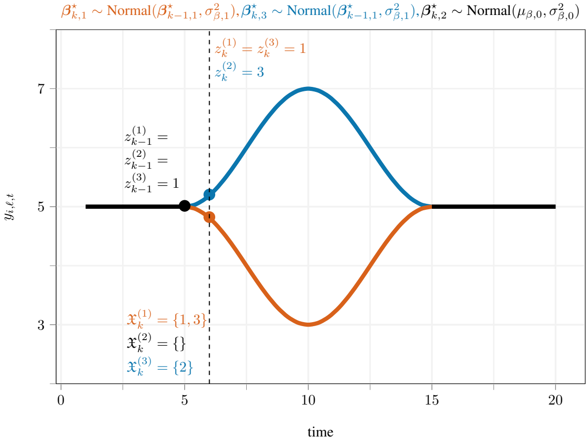
$$
\ell_{j,k} \propto \exp(-\varphi_j \ell_{j,k}) \mathbb{1}_{\{1,\ldots,x_{j,\max}\}}(\ell_{j,k}), \quad \varphi_j \sim \mathsf{Ga}(a_\varphi, b_\varphi)
$$

- Shrinkage priors on the second layer of allocation variables

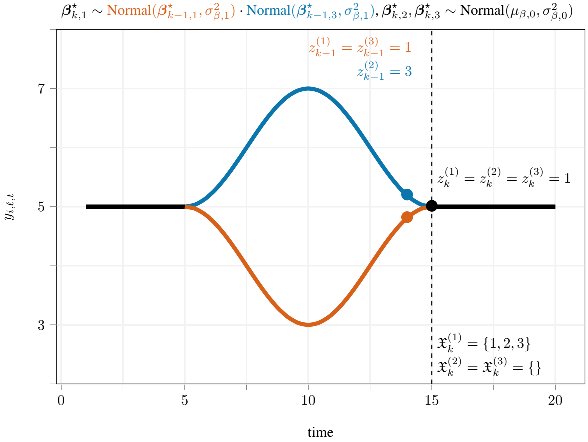
$$
\begin{aligned}
&(z_k^{(z_{1,k},\ldots,z_{p,k})} \mid \boldsymbol{\pi}^\star) \sim \mathsf{Mult}(\pi_1^\star, \ldots, \pi_{\ell_k}^\star) \\
&\boldsymbol{\pi}^\star = (\pi_1^\star, \ldots, \pi_{\ell_k}^\star)^\mathsf{T} \sim \mathsf{Dir}(\alpha^\star/\ell_k, \ldots, \alpha^\star/\ell_k) \\
&\alpha^\star \sim \mathsf{Ga}(a_{\alpha^\star}, b_{\alpha^\star})
\end{aligned}
$$

  where $\ell_k = \prod_{j=1}^p \ell_{j,k}$.

Smoothing prior: fixed effects



Smoothing prior: fixed effects



Posterior inference

Varying values of $\ell_{j,k}$ result in varying model dimensions.

Trans-dimensional step to update $(\boldsymbol{\rho}_k, \boldsymbol{\beta}_k^{\star\star})$:

1. propose a change to the partition structure $\boldsymbol{\rho}_k$
2. conditional on $\boldsymbol{\rho}_k$, sample from the posterior of the spline coefficients $\boldsymbol{\beta}_k^{\star\star} = \{\beta_{k,h}^{\star\star}\}_{h=1}^{M_k}$

Step 1: for each predictor j, perform the following M-H step

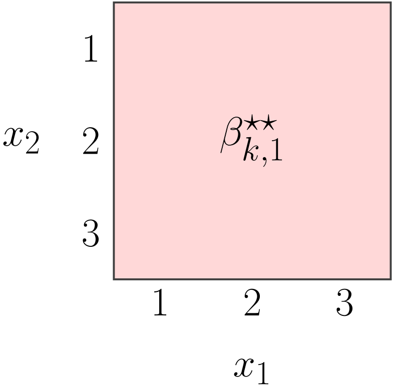
(i) propose a change to the marginal partition for the levels of x_j (split or merge)

(ii) propose a corresponding change to the joint partition $\boldsymbol{\rho}_k$

(iii) evaluate the acceptance rate, integrating out the curve-specific parameters

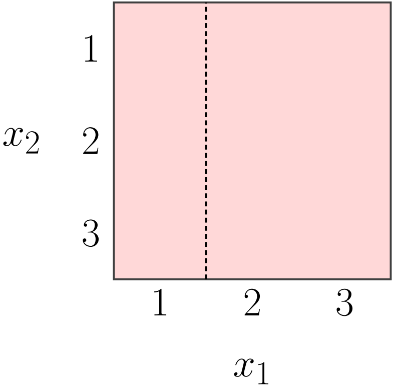
Posterior inference

Initial configuration



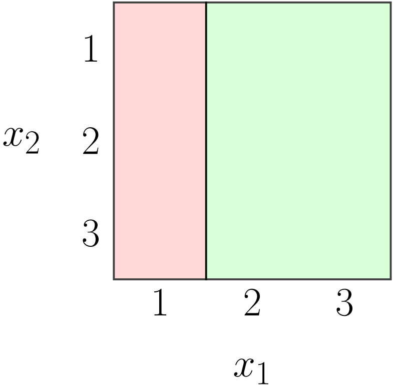
Posterior inference

Propose marginal partition for x_1: split



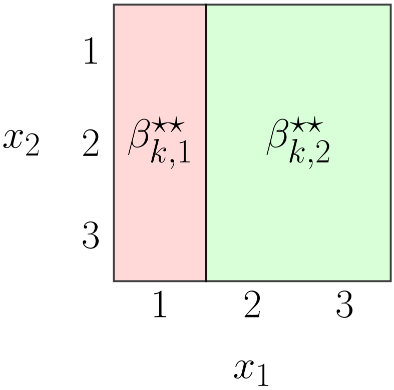
Posterior inference

Conditional on the marginal partition, propose joint partition



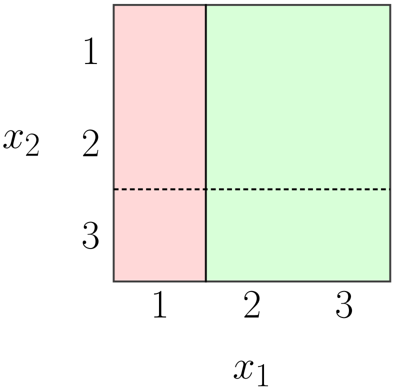
Posterior inference

If accept, update curve-specific parameters



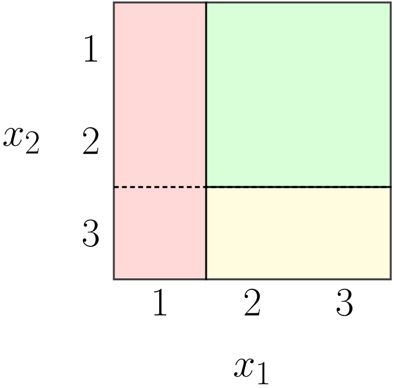
Posterior inference

Propose marginal partition for x_2: split



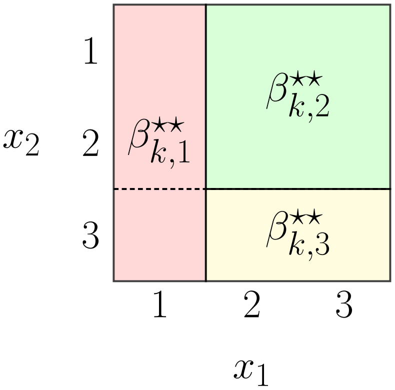
Posterior inference

Conditional on the marginal partition, propose joint partition



Posterior inference

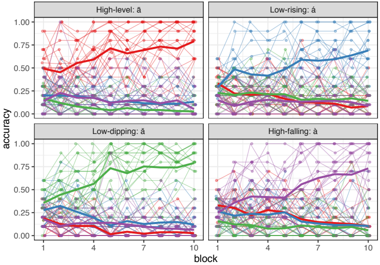
If accept, update curve-specific parameters

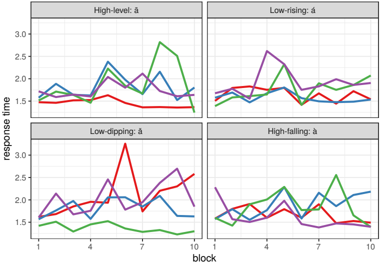


Advantages of our approach

Compared to the existing literature, our approach does:

- introduce a biologically interpretable class of **multi-category** DDM
- accommodate flexible **random effects** for subject heterogeneity (good vs poor learners)
- allow to study the **longitudinal** evolution of the parameters as the subjects get trained
- assess local similarities/**dissimilarities** in the model parameters





Longitudinal drift-diffusion mixed models

Notation:

- Time points (blocks): $t \in \{1, \ldots, T = 10\}$
- Individuals: $i \in \{1, \ldots, n = 20\}$
- Trials: $\ell \in \{1, \ldots, L = 40\}$
- Observed data: $\boldsymbol{y}_{i,\ell,t} = (d_{i,\ell,t}, \tau_{i,\ell,t})$

Longitudinal drift-diffusion mixed models

Notation:

- Time points (blocks): $t \in \{1, \dots, T = 10\}$
- Individuals: $i \in \{1, \dots, n = 20\}$
- Trials: $\ell \in \{1, \dots, L = 40\}$
- Observed data: $\boldsymbol{y}_{i,\ell,t} = (d_{i,\ell,t}, \tau_{i,\ell,t})$

The likelihood function of the **longitudinal** drift-diffusion **mixed** model is

$$L(\boldsymbol{d}, \boldsymbol{\tau} \mid \boldsymbol{s}, \boldsymbol{\theta}) = \prod_{d=1}^{d_0} \prod_{s=1}^{d_0} \prod_{t=1}^{T} \prod_{i=1}^{n} \prod_{\ell=1}^{L} \left[g\{\tau_{i,\ell,t} \mid \boldsymbol{\theta}_{d,s}^{(i)}(t)\} \prod_{d' \neq d} \bar{G}\{\tau_{i,\ell,t} \mid \boldsymbol{\theta}_{d',s}^{(i)}(t)\} \right]^{\mathbb{1}\{d_{i,\ell,t}=d, s_{i,\ell,t}=s\}}$$

where $\boldsymbol{\theta}_{d,s}^{(i)}(t) = (\delta_s, \mu_{d,s}^{(i)}(t), b_{d,s}^{(i)}(t))^{\mathsf{T}}$.

Longitudinal drift-diffusion mixed models

Notation:

- Time points (blocks): $t \in \{1, \dots, T = 10\}$
- Individuals: $i \in \{1, \dots, n = 20\}$
- Trials: $\ell \in \{1, \dots, L = 40\}$
- Observed data: $\boldsymbol{y}_{i,\ell,t} = (d_{i,\ell,t}, \tau_{i,\ell,t})$

The likelihood function of the **longitudinal** drift-diffusion **mixed** model is

$$L(\boldsymbol{d}, \boldsymbol{\tau} \mid \boldsymbol{s}, \boldsymbol{\theta}) = \prod_{d=1}^{d_0} \prod_{s=1}^{d_0} \prod_{t=1}^{T} \prod_{i=1}^{n} \prod_{\ell=1}^{L} \left[g\{\tau_{i,\ell,t} \mid \boldsymbol{\theta}_{d,s}^{(i)}(t)\} \prod_{d' \neq d} \bar{G}\{\tau_{i,\ell,t} \mid \boldsymbol{\theta}_{d',s}^{(i)}(t)\} \right]^{\mathbb{1}\{d_{i,\ell,t}=d, s_{i,\ell,t}=s\}}$$

where $\boldsymbol{\theta}_{d,s}^{(i)}(t) = (\delta_s, \mu_{d,s}^{(i)}(t), b_{d,s}^{(i)}(t))^\intercal$.

We need to enforce a **positivity constraint** on $\{\mu_{d,s}^{(i)}(t), b_{d,s}^{(i)}(t)\}$

Parameter modeling

With $\boldsymbol{x} = (d, s)$, we let

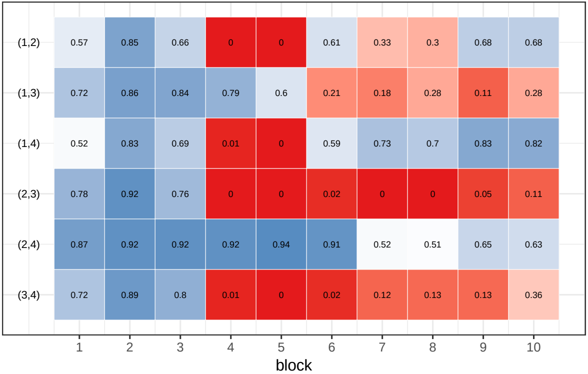
$$\mu_{\boldsymbol{x}}^{(i)}(t) = \exp\{f_{\mu,\boldsymbol{x}}(t) + u_{\mu,\boldsymbol{x}}^{(i)}(t)\}$$

$$\{f_{\mu,\boldsymbol{x}}(t) \mid z_k^{(\boldsymbol{x})} = z_k\} = \sum_{k=1}^{K} \beta_{\mu,k,z_k}^{\star} B_k(t)$$

$$u_{\mu,\boldsymbol{x}}^{(i)}(t) = \begin{cases} u_{\mu,C}^{(i)}(t) & \text{if } s = d \\ u_{\mu,I}^{(i)}(t) & \text{if } s \neq d \end{cases}$$

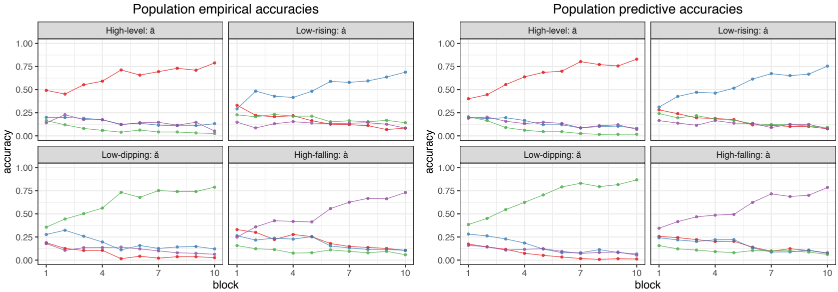
- Same specification for the boundary parameters $b_{\boldsymbol{x}}^{(i)}(t)$
- $\delta_s \sim$ Unif$(0, \delta_{s,max})$, where $\delta_{s,max}$ is the minimum of all response times under stimulus s

Results: co-clustering probabilities



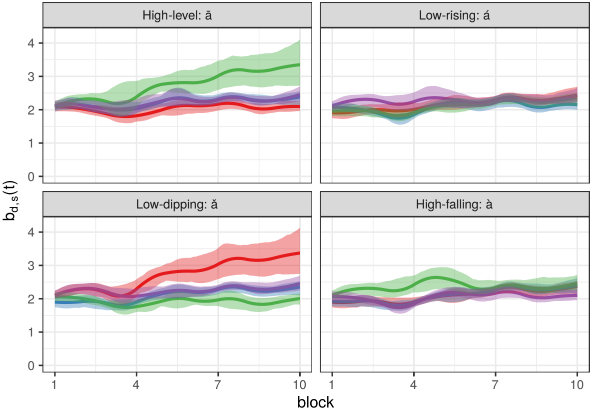
Matrix containing the probabilities of pairwise co-clustering between tones. On the y-axis, each pair of success parameters is considered

Results: predictive check



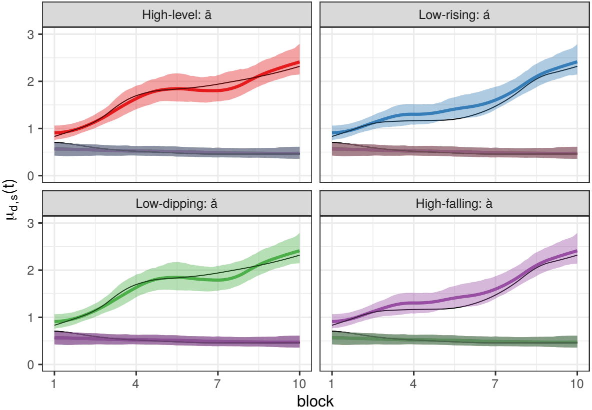
Empirical (left) and predicted (right) population level accuracies

Results: boundary parameters



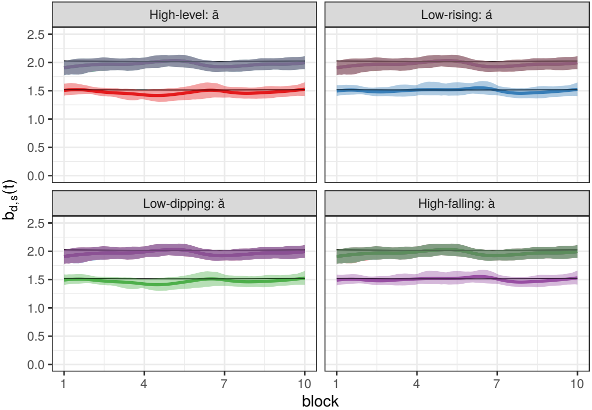
Estimated posterior mean and 95% CI for the population boundaries $b_{d,s}(t)$

Synthetic data: population level drifts



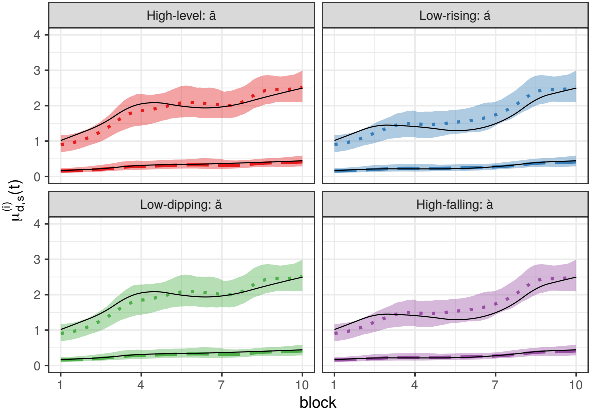
Estimated posterior mean and 95% CI for the population drift parameters $\mu_{d,s}(t)$

Synthetic data: population level boundaries



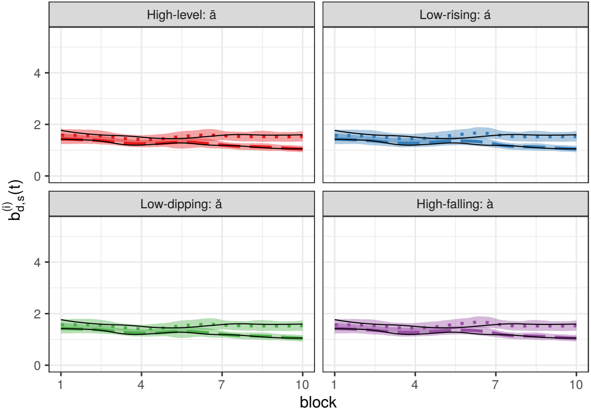
Estimated posterior mean and 95% CI for the population boundary parameters $b_{d,s}(t)$

Synthetic data: individual level drifts



Estimated posterior mean and 95% CI for the individual level drift parameters $\mu_{d,s}^{(i)}(t)$

Synthetic data: individual level boundaries



Estimated posterior mean and 95% CI for the individual level boundary parameters $b_{d,s}^{(i)}(t)$