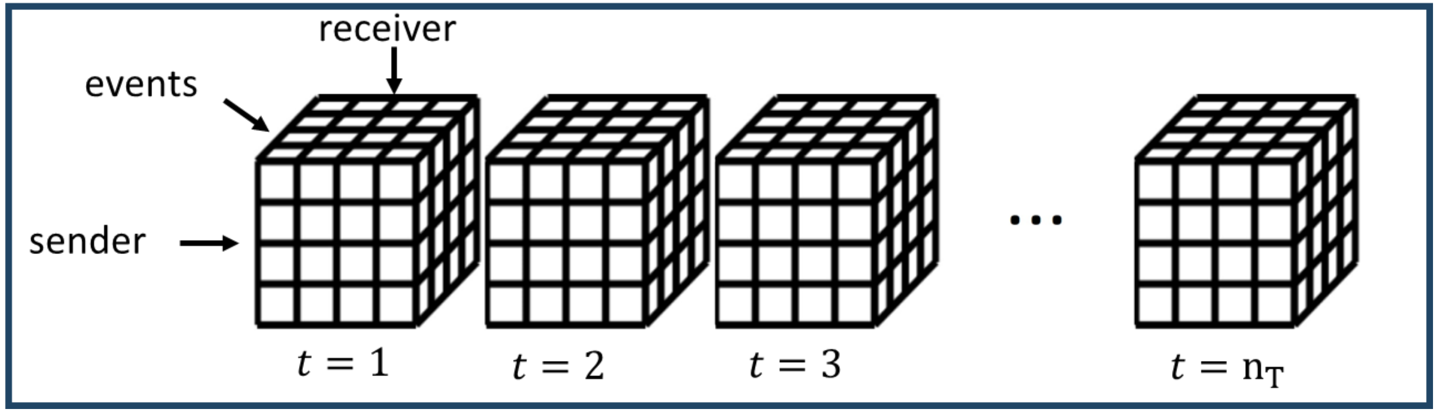
Sandia National Laboratories

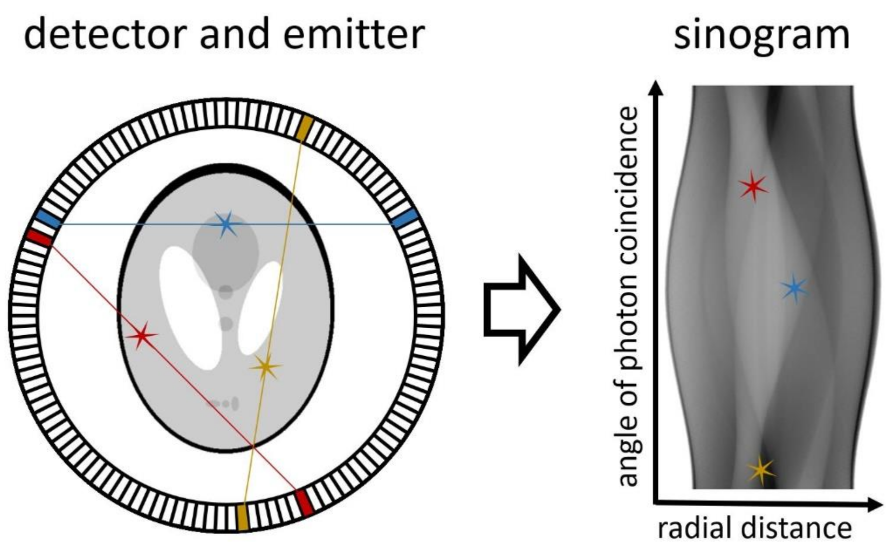
Poisson-response Tensor-on-Tensor Regression and Applications

Carlos Llosa and Daniel Dunlavy

SAND2023-09329C

Motivating problems





Temporal prediction of dyadic relationships

ICEWS database [1]
- Countries as receivers and senders
- Events such as threats or aid
- **Can we predict future relations?**

Change-point detection for multi-way relationships

Enron email database [2]
- Employees as senders and receivers
- Events such as words or topics of conversation
- **Are there changes in communication at different times?**

Positron emission tomography (PET) reconstruction [3]

- Poisson distribution commonly used for photon counts in emission tomography
- Statistical model for PET reconstruction is ill-posed without restrictions
- **Can we reformulate and improve the model using low-rank tensors?**

[1] O'Brien, *Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research*, ISR, 2010.
[2] Enron Email Dataset: https://www.cs.cmu.edu/~enron/
[3] Shepp and Vardi, *Maximum Likelihood Reconstruction for Emission Tomography,* IEEE TMI, 1982.

From Linear to Multilinear Regression

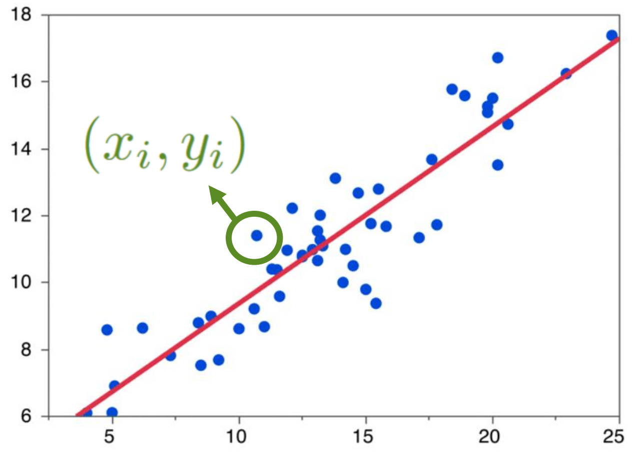
Linear regression: $y_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i | \boldsymbol{\beta} \rangle, \sigma^2)$

Tensor regression: $y_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \sigma^2)$

Multivariate linear regression: $\boldsymbol{y}_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i | B \rangle, \Sigma)$

Tensor-on-Tensor regression [4,5]: $\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$

$m_1 \times \cdots \times m_p$ $h_1 \times \cdots \times h_l$ $h_1 \times \cdots \times h_l \times m_1 \times \cdots \times m_p$



Linear regression example

$$\langle S|T \rangle = \begin{cases} \text{inner product} & dim(S) = dim(T) \\ \text{contraction} & dim(S) \neq dim(T) \end{cases}$$

$$\mathcal{Y}_i \in \mathbb{R}^{10\times10\times10} \quad \mathcal{X}_i \in \mathbb{R}^{20\times20\times20}$$
$$\mathcal{B} \in \mathbb{R}^{10\times10\times10\times20\times20\times20}$$

Tensor models not tractable without using a very large sample size

[4] Llosa and Maitra, *Reduced-Rank Tensor-on-Tensor Regression and Tensor-Variate Analysis of Variance*, IEEE TPAMI 2022.

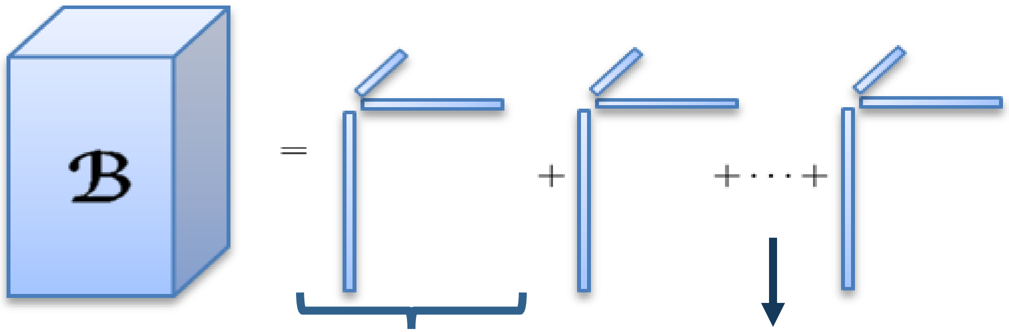
[5] Lock, *Tensor-on-Tensor Regression*, JCGS, 2018

Tensor-on-Tensor Regression (ToTR)

$$\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$$

Canonical polyadic (CP) ToTR [4,5] : $\mathcal{B} = [\![\boldsymbol{\lambda}; U_1, \ldots, U_l, V_1, \ldots, V_p]\!]$

3-dimensional tensor example



Rank-1 tensor

Rank summation

Each rank-1 tensor is the outer product of $p + q$ vector

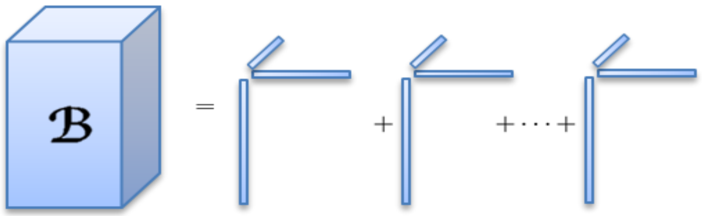
Other low-rank tensor structures studied in [4]

Make the model tractable with low-rank tensors!

[4] Llosa and Maitra, *Reduced-Rank Tensor-on-Tensor Regression and Tensor-Variate Analysis of Variance*, IEEE TPAMI 2022.
[5] Lock, *Tensor-on-Tensor Regression*, JCGS, 2018

Existing: Gaussian ToTR

$$\mathcal{Y}_i \overset{ind.}{\sim} N(\langle \mathcal{X}_i | \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$$



Loglikelihood:

$$\ell(\mathcal{B}) = \sum_{i=1}^{n} \text{vec}(\mathcal{Y}_i - \langle \mathcal{X}_i | \mathcal{B} \rangle)' \left[\bigotimes_{k=1}^{p} \Sigma_k^{-1} \right] \text{vec}(\mathcal{Y}_i - \langle \mathcal{X}_i | \mathcal{B} \rangle)$$

Least squares loss

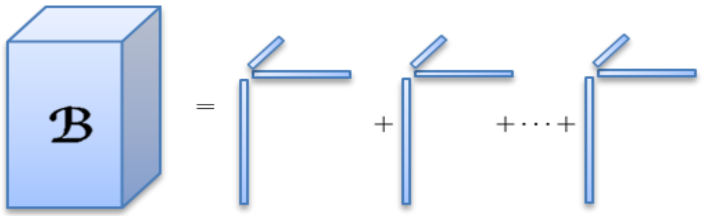
- CP decomposition constraints

New: Poisson ToTR (PToTR)

$$\mathcal{Y}_{i\boldsymbol{j}} \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}})$$

$$\Downarrow$$

$$\mathcal{Y}_i \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i | \mathcal{B} \rangle)$$



Loglikelihood:

$$\ell(\mathcal{B}) = \sum_{i=1}^{n} \sum_{\boldsymbol{j}} [\mathcal{Y}_{i\boldsymbol{j}} \log(\langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}) - \langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}]$$

Kullback-Leibler divergence loss

- CP decomposition constraints
- **Strictly positive $\mathcal{B}$ constraints**

Analogous to the existing model, harder to optimize loss function

Optimization procedure for PToTR

Model

$$\mathcal{Y}_i \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i | \mathcal{B} \rangle)$$

Loglikelihood

$$\ell(\mathcal{B}) = \sum_{i=1}^{n} \sum_{\boldsymbol{j}} [\mathcal{Y}_{i\boldsymbol{j}} \log(\langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}) - \langle \mathcal{X}_i | \mathcal{B} \rangle_{\boldsymbol{j}}]$$

Constraints

$$\mathcal{B} = [\![\boldsymbol{\lambda}; U_1, \ldots, U_l, V_1, \ldots, V_p]\!] > 0$$

Our new multiplicative update rules extend those in the CP-alternating Poisson regression (CP-APR) algorithm [6]

	Estimation of V_k	Estimation of U_k
Alternative expression for Loglikelihood:	$\sum_{i=1}^{n} \mathbf{1}' \left[V_k^* G_{ik} - \mathcal{Y}_{i(k)} * \log(V_k^* G_{ik}) \right] \mathbf{1}$	$\sum_{i=1}^{n} [(\text{ vec } U_k^*)' H_{ik} - (\text{ vec } \mathcal{Y}_i)' * \log((\text{ vec } U_k^*)' H_{ik})] \mathbf{1}$
Multiplicative update: (non-decreasing loglikelihood)	$\widehat{V}_k^* \leftarrow \widehat{V}_k^* * \left\{ \sum_{i=1}^{n} \left[\left(\mathcal{Y}_{i(k)} \oslash (\widehat{V}_k^* G_{ik}) \right) G'_{ik} \right] \right\} \oslash \left\{ \mathbf{1} (\sum_{i=1}^{n} \boldsymbol{w}_i)' \right\}$	$(\text{ vec } \widehat{U}_k^*) \leftarrow \left\{ \sum_{i=1}^{n} \left[H_{ik} \left(\text{vec}(\mathcal{Y}_i) \oslash \left(H'_{ik} (\text{ vec } \widehat{U}_k^*) \right) \right) \right] \right\} * \text{ vec} \left(\widehat{U}_k^* \oslash \sum_{i=1}^{n} W_i \right).$

Now we have the tools for tackling our three PToTR applications!

[6] Chi and Kolda, *On Tensors, Sparsity, and Nonnegative Factorizations*, SIMAX 2012.

Application I: Autoregressive Model for the ICEWS Database

$\mathcal{Y}_t(i_1, i_2, i_3)$ = # times action i_3 was taken by country i_1 on country i_2 at week t.

$$\mathcal{Y}_t \overset{indep.}{\sim} \text{Poisson}(\langle \mathcal{Y}_{t-1} | \mathcal{B} \rangle)^*$$

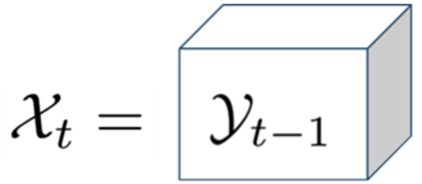
$\mathcal{B}(i_1, i_2, i_3, j_1, j_2, j_3)$ = Effect that the **previous** i_3 event from i_1 towards i_2 has on **current** j_3 event from j_1 towards j_2.

*Not well defined without accounting for trend

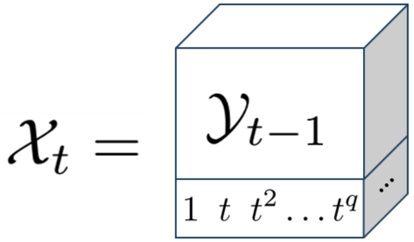
Other longitudinal effects by choosing $\mathcal{X}_t$ in:

$$\mathcal{Y}_t \overset{indep.}{\sim} \text{Poisson}(\langle \mathcal{X}_t | \mathcal{B} \rangle)$$

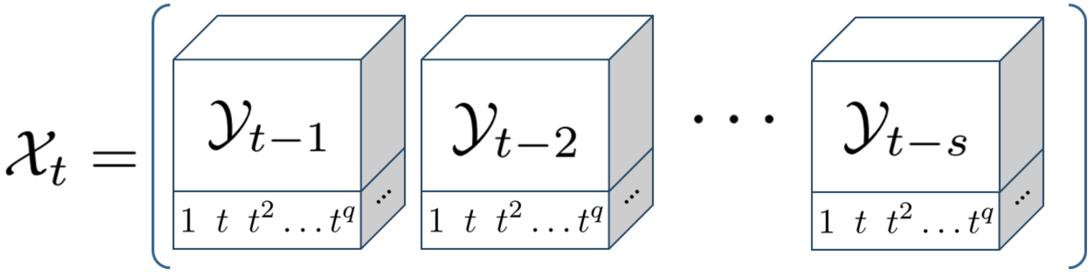
- **AR(1) with no trend (as before):**



- **AR(1) with q-th order polynomial trend:**



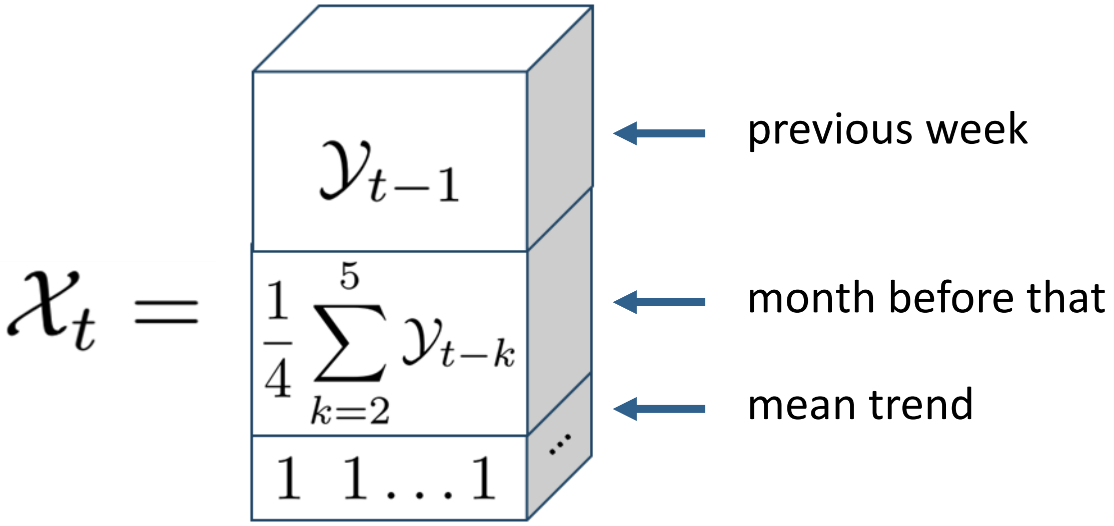
- **AR(s) with q-th order polynomial trend:**



Application I: Autoregressive Model for the ICEWS Database

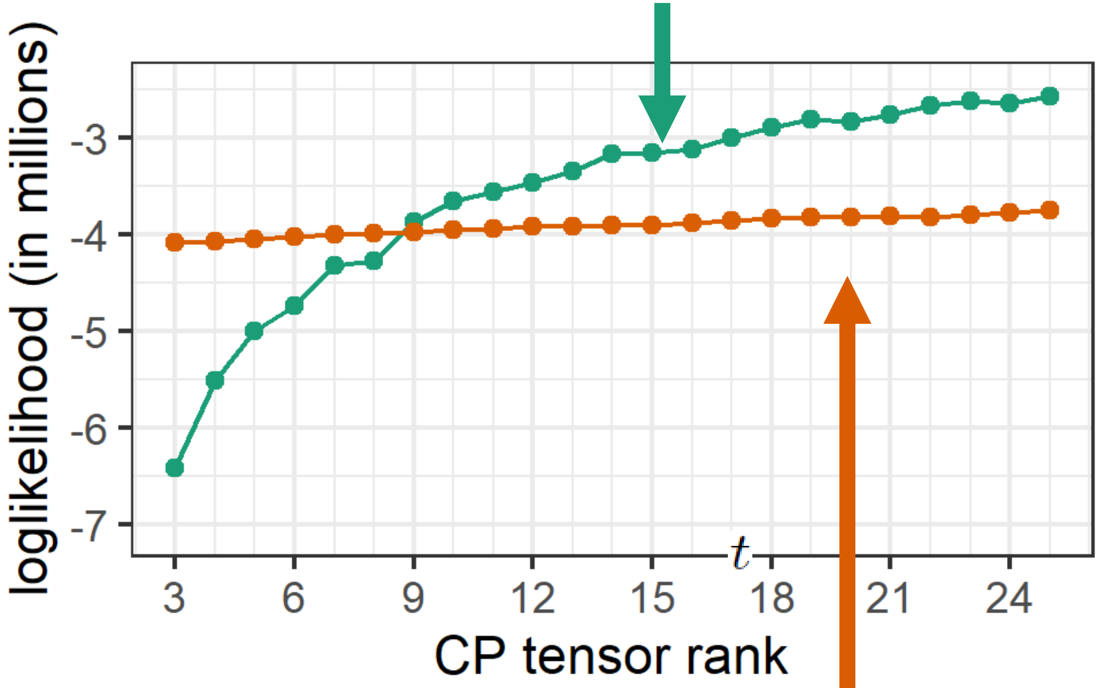
Data selected as in [7]:

- Weekly data from 2004 to mid-2014
- 25 countries
- 4 quad classes (type of events)



Poisson ToTR :

$$\mathcal{Y}_i \overset{indep.}{\sim} \text{Poisson}(\langle \mathcal{X}_i | \mathcal{B} \rangle)$$



Gaussian ToTR:

$$\mathcal{Y}_i \overset{indep.}{\sim} N(\, \langle \mathcal{X}_i | \mathcal{B} \rangle \,, \Sigma_1, \ldots, \Sigma_p)$$

Our Poisson model fits the count data better!

[7] Hoff, *Multilinear tensor regression for longitudinal relational data*, Ann. Appl. Stat., 2015

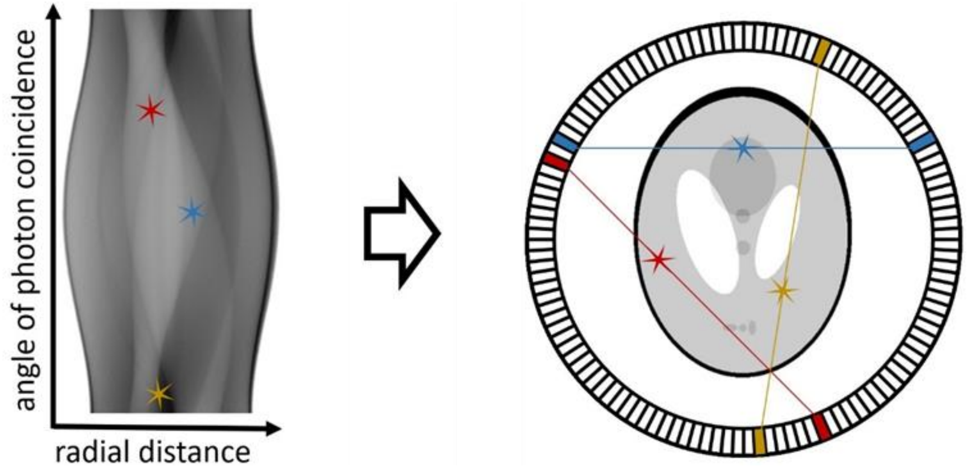
Application II: Poisson model for Positron Emission Tomography [3]

Radon transform of B

$$Y \sim \text{Poisson}(\langle B | \mathcal{K} \rangle)$$

Observed sinogram containing counts

Image to reconstruct



- Poisson distribution for number of photon coincidences in sinogram
- Radon basis $\mathcal{K}$ is a 4D tensor with combined dimensions
- Element-wise formulated as:

$$y_{i_1 i_2} \overset{ind.}{\sim} \text{Poisson}(\langle B, K_{i_1 i_2} \rangle)$$

- Here $K_{i_1 i_2}$ is one matrix slice of $\mathcal{K}$

2D PET is scalar-response, matrix-predictor Poisson regression!

[3] Shepp and Vardi, *Maximum Likelihood Reconstruction for Emission Tomography,* IEEE TMI, 1982.

Application II: From 2D to 4D PET

2D PET model – elementwise

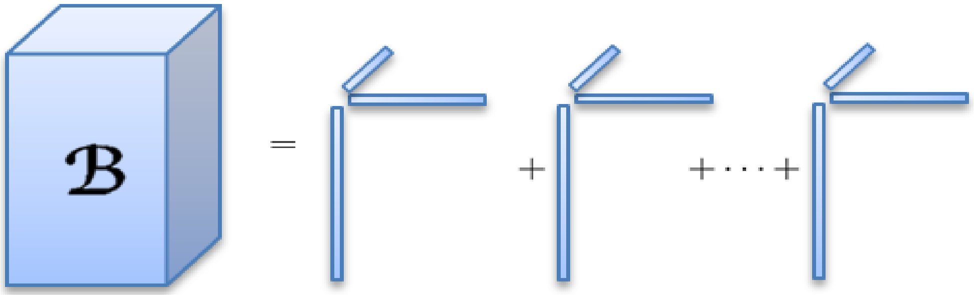
$$y_{i_1 i_2} \overset{ind.}{\sim} \text{Poisson}(\langle K_{i_1,i_2}, B \rangle)$$

- Same model as before
- Reconstructs 2D image B
- Scalar-response, matrix-predictor
- 2D image to reconstruct
- Ill-posed without additional restrictions [8]

4D PET model : volume across time

$$Y_{i_1,i_2} \overset{ind.}{\sim} \text{Poisson}(\langle K_{i_1,i_2} | \mathcal{B} \rangle)$$

- Extension of 2D PET for depth and time
- Reconstruct 4D image $\mathcal{B}$
- Matrix-response, matrix-predictor
- 4D image to reconstruct
- Ill-posed without additional restrictions [8]
- Other types of N-D PET (such as using 3D Radon transforms) can also be framed as PToTR

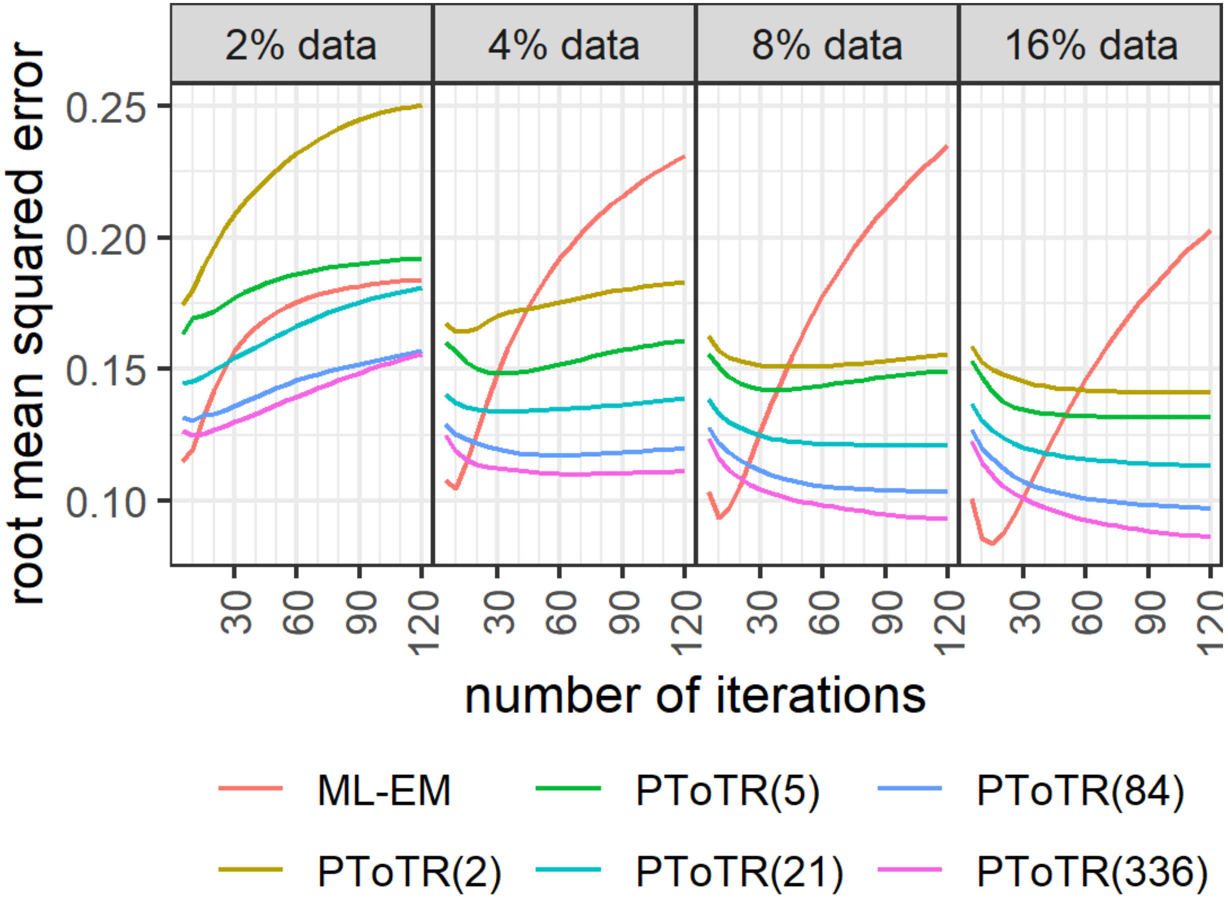


[8] Snyder et. al., Noise and Edge Artifacts in Maximum-Likelihood Reconstructions for Emission Tomography, IEEE TMI, 1987.

Application II: A simulation of low-rank 4D PET

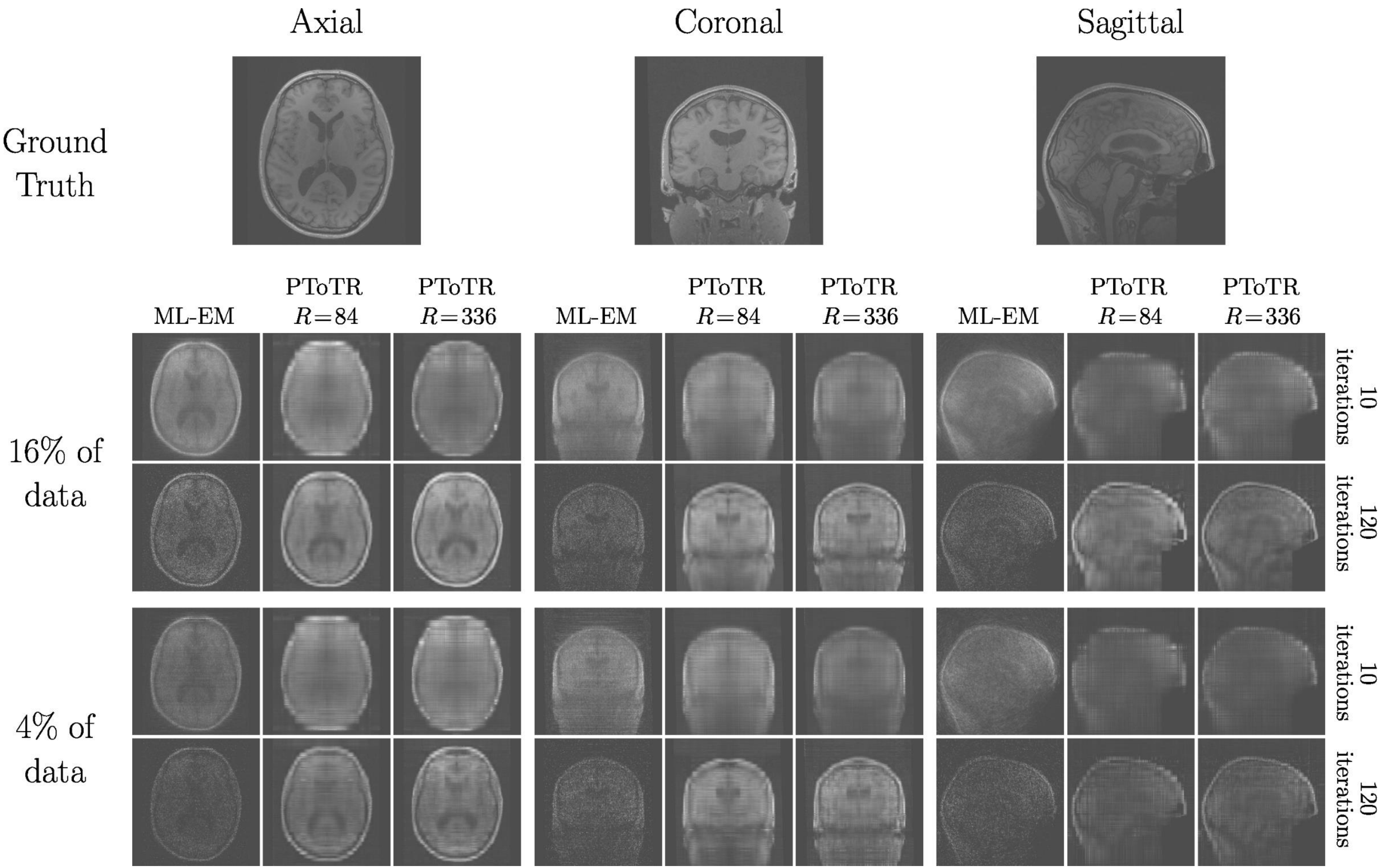
$$Y_{i_1,i_2} \overset{ind.}{\sim} \text{Poisson}(\langle K_{i_1,i_2} | \mathcal{B} \rangle)$$

- True 4D image $\mathcal{B}$: four MRI measurements on the same subject and scanner [9]
- 256 x 256 matrix slices → 256 x 1024 sinograms
- We use 2%, 8% and 16% of the data.
- We implement different methods:
  - Full MLEM: no restrictions on $\mathcal{B}$:
    number of parameters: ~ 63 million
  - PToTR: CP-restricted $\mathcal{B}$ with rank 84:
    Number of parameters: ~ 63 thousand
  - We try ranks 2, 5, 21, 84, 336



Low-rank image reconstruction stabilizes the estimation !

[9] Hawco, et. al., A longitudinal multi-scanner multimodal human neuroimaging dataset, Scientific Data, 2022


Axial
Coronal
Sagittal
Ground Truth
ML-EM
PToTR $R=84$
PToTR $R=336$
16% of data
4% of data
10 iterations
120 iterations

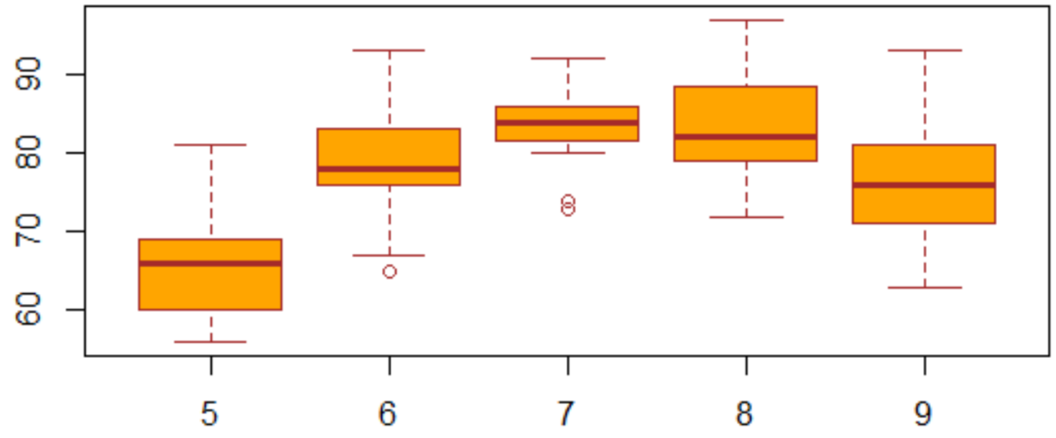
Application III: A path towards change-point detection

Analysis of variance (ANOVA):
Tool for distinguishing between groups in **scalar** data.

Multivariate ANOVA (MANOVA):
Tool for distinguishing between groups in **vector** data.

Tensor-variate ANOVA (TANOVA):
Tool for distinguishing between groups in **tensor** data, and across **multiple** factors.

Poisson-response TANOVA (PTANOVA):
Tool for distinguishing between groups in **tensor count** data, and across **multiple** factors.



ANOVA example

Change-point detection:

- **Group** data before and after some point in time

Let's go through the math in the same way we did for regression

Application III: Tensor-Variate Analysis of Variance (TANOVA)[4]

TANOVA is a special case of ToTR ($\mathcal{X}_i$ are indicator tensors) that generalizes ANOVA

General Model	Definition	Special Case (Indicator $\mathcal{X}_i$)
Linear regression (LR)	$y_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i \vert \boldsymbol{\beta} \rangle, \sigma^2)$	**ANOVA**
Multivariate LR	$\boldsymbol{y}_i \overset{indep.}{\sim} N(\langle \boldsymbol{x}_i \vert B \rangle, \Sigma)$	**MANOVA**
Tensor regression	$y_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i \vert \mathcal{B} \rangle, \sigma^2)$	**Factorial designs**
Tensor-on-tensor regression	$\mathcal{Y}_i \overset{indep.}{\sim} N(\langle \mathcal{X}_i \vert \mathcal{B} \rangle, \Sigma_1, \ldots, \Sigma_p)$	**TANOVA**

Poisson ToTR (PToTR)

$\mathcal{Y}_i \overset{ind.}{\sim} Poisson(\langle \mathcal{X}_i \vert \mathcal{B} \rangle)$ → **Poisson TANOVA**

We can do change-point detection through PTANOVA!

[4] Llosa and Maitra, *Reduced-Rank Tensor-on-Tensor Regression and Tensor-Variate Analysis of Variance*, IEEE TPAMI 2022.

Application III: PTANOVA and Change-Point Detection

Poisson-tensor change-point detection

$\mathcal{Y}_t$ changes in mean at time τ_1

$$\mathcal{Y}_t \sim \left\{ \begin{array}{ll} \text{Poisson}(\mathcal{M}_1) & t = 1, \ldots, \tau_1 \\ \text{Poisson}(\mathcal{M}_2) & t = \tau_1 + 1, \ldots, n_T \end{array} \right\}$$

- Mean before change-point: $\mathcal{M}_1$
- Mean after change-point: $\mathcal{M}_2$
- Change-point location: τ_1

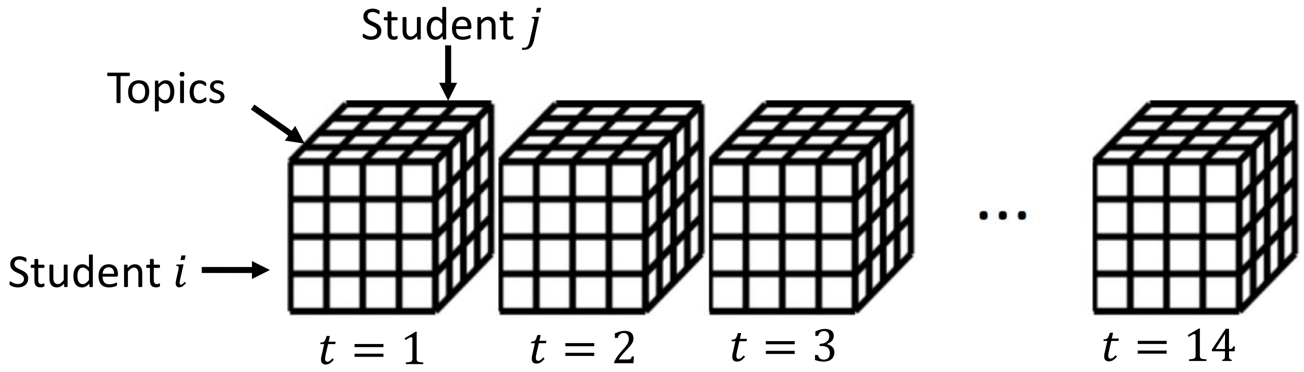
Equivalent PTANOVA formulation

$$\mathcal{Y}_t \sim \text{Poisson}(\langle \boldsymbol{x}_t | \mathcal{B} \rangle)$$

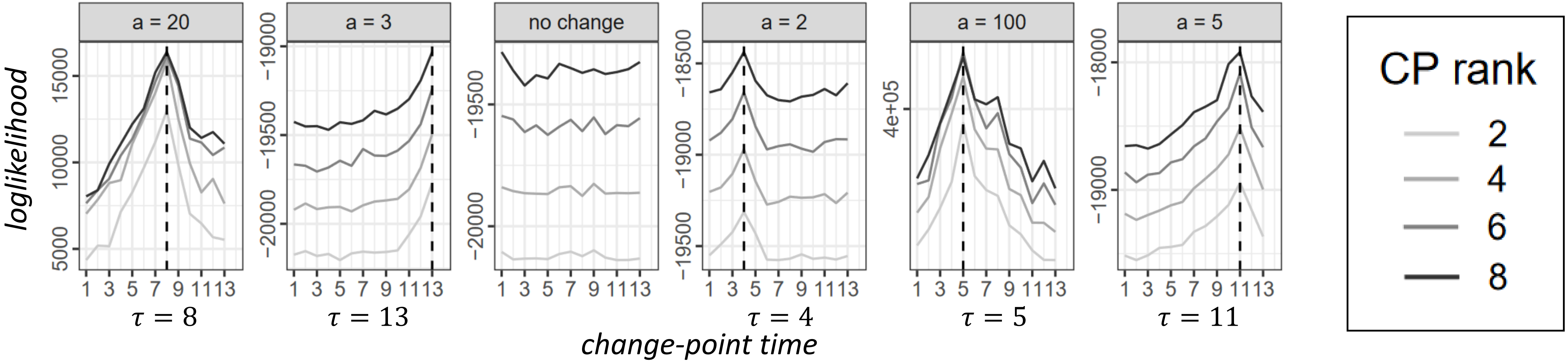
$$\boldsymbol{x}_t = \left\{ \begin{array}{ll} (1,0)' & t = 1, \ldots, \tau_1 \\ (0,1)' & t = \tau_1 + 1, \ldots, n_T \end{array} \right\}$$

- Mean before change-point: $\langle (1,0)' | \mathcal{B} \rangle$
- Mean after change-point: $\langle (0,1)' | \mathcal{B} \rangle$
- Change-point location: τ_1

Application III: Change-point detection experiment



- 10 Students talking about 15 topics over 14 time-steps: 14 count tensors of size 10 x 10 x 15.
- Event occurs at time τ that changes communication pattern:
  - All conversation generated independently from Poisson(λ), $\lambda = 5$.
  - One topic changes to Poisson($\lambda \times a$) after time τ, where $a = 1, 2, 3, 5, 20, 100$.



Conclusions and path forward

- Model: Linear regression → Tensor regression → Poisson tensor regression
  - Low-rank tensors and constrained optimization

- Methods:
  - Tensor autoregressive models: temporal prediction
  - 2D PET → 4D PET: stable estimation
  - ANOVA → TANOVA → PTANOVA: change-point detection for tensor count data

- Where we are going
  - Efficient model-fitting algorithms
  - Multiple change-point detection
  - Inference on estimated coefficients

Thank You!

For follow-up questions reach out to:

- Carlos Llosa cjllosa@sandia.gov,
- Danny Dunlavy dmdunla@sandia.gov