

# II-Probabilities

## Empirical Distributions:

## Representation of Samples

Logik für Erklärbare KI: Technische Einführung in das ENEXA Projekt

Maria Schnödt-Fuchs, Alex Goëßmann

Funded by the  
European Union



15.+16. July, 2024

## Example (Students and their semester)

We ask students about their semester and want to capture the statistics.

### Task: Representation of Observations

Let there be observations of the system by a set of states

$$\{(i_0^j, \dots, i_{d-1}^j) : j \in [m]\}.$$

How to capture them by tensors?

Aim:

- ▶ Represent the observations for querying
- ▶ Find models making sense out of the observations

Given a dataset  $\{(i_0^j, \dots, i_{d-1}^j) : j \in [m]\}$  of samples of the factored system we define the sample selector map

$$\theta^D : [m] \rightarrow \bigtimes_{k \in [d]} [m_k]$$

### Idea

Enrich the factored system with variables  $X_0, \dots, X_{d-1}$  by a data selection variable  $L$  and encode  $\theta^D$  using the one-hot encoding of the larger system.

# More general: Encoding of maps between Factored Systems

## Definition

Let  $f$  be a function

$$f : \underset{k \in [d]}{\times} [m_k] \rightarrow \underset{l \in [p]}{\times} [m_l]$$

which maps the states of a factored system to variables  $X_0, \dots, X_{d-1}$  to the states of another factored system with variables  $Y_0, \dots, Y_{p-1}$ . Then the tensor representation of  $f$  is a tensor

$$\theta^f \in \left( \bigotimes_{l \in [p]} \mathbb{R}^{m_l} \right) \otimes \left( \bigotimes_{k \in [d]} \mathbb{R}^{m_k} \right)$$

defined by

$$\theta^f = \sum_{i_0 \in [m_0]} \cdots \sum_{i_{k-1} \in [m_{k-1}]} e_{f(i_0, \dots, i_{d-1})} \otimes e_{i_0, \dots, i_{d-1}}.$$

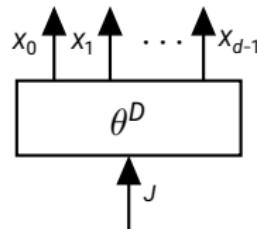
In the case of the sample selection map

$$D : [m] \rightarrow \bigtimes_{k \in [d]} [m_k]$$

we have a representation tensor called the **data tensor**

$$\theta^D = \sum_{j \in [m]} e_j \otimes e_{i_0^j, \dots, i_{d-1}^j}$$

depicted as



## Representation of Data

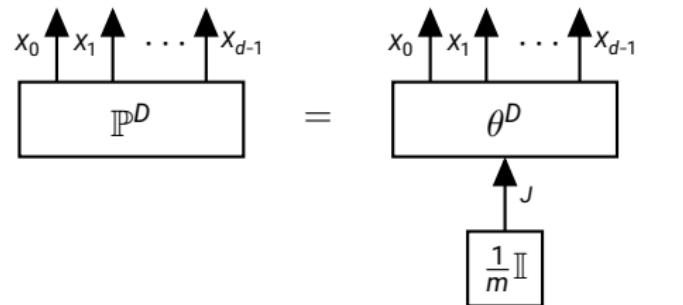
The one-hot encoding of each observation is stored on a slice of the data tensor.

# Average of the Observation

The **empirical distribution** is the average of the one-hot encoded observed states:

$$\mathbb{P}^D := \frac{1}{m} \sum_{j \in [m]} e_{i_0^j, \dots, i_{d-1}^j}$$

In a contraction diagram we denote the average by



# More efficient data representation

The storage demand of  $\theta^D$  increases exponentially

$$\dim [\theta^D] = \left( \prod_{k \in [d]} m_k \right) \cdot m.$$

Store the values  $X_k$  separately for each  $k \in [d]$ !

The collection of encoding tensors

$$\theta^{D_k} \in \mathbb{R}^{m_k \times m}$$

to the maps

$$D_k : [m] \rightarrow [m_k] \quad , \quad D_k(j) = i_k^j$$

has a linearly increasing storage demand of

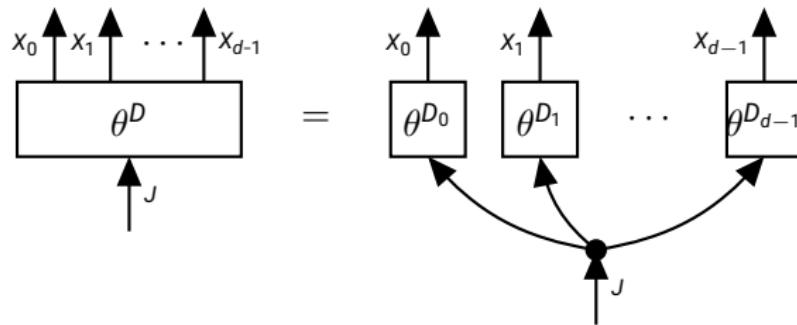
$$\sum_{k \in [d]} \dim [\theta^{D_k}] = \sum_{k \in [d]} m_k \cdot m.$$

# Representation of the Empirical in a CP Format

The data tensor can be stored by the contraction

$$\mathcal{C} \left( \{\theta^D\}, \{X_0, \dots, X_{k-1}, J\} \right) = \mathcal{C} \left( \{\theta^{D_0}, \dots, \theta^{D_{d-1}}\}, \{X_0, \dots, X_{d-1}, J\} \right)$$

depicted as



This is called a CP Tensor Decomposition.

## CP Decomposition of Empirical Distributions

Instead of storing the datacore  $\theta^D$ , store the encodings  $\theta^{D_k}$ .