II-Probabilities Empirical Distributions: Representation of Samples

Foundations of Neuro-Symbolic AI

Alex Goessmann

University of Applied Science Würzburg-Schweinfurt

Summer Term 2026

Observations of a Factored System

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

$$\{(x_0^j, \ldots, x_{d-1}^j) \; : \; j \in [m]\} \, .$$

How to capture them by tensors?

Aim:

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

Formalization with a sample selector map

Given a dataset $\{(x_0^j, \ldots, x_{d-1}^j) \,:\, j \in [m]\}$ of samples of the factored system we define the sample selector map

$$\beta^D : [m] \to \bigtimes_{k \in [d]} [m_k]$$

Idea

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

More general: Encoding of maps between Factored Systems

Definition

Let q be a function

$$q : \bigtimes_{k\in[d]} [m_k] \to \bigtimes_{l\in[r]} [m_l]$$

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 $Y_0, \dots, Y_{p\text{-}1}$. Then the tensor representation of q is a tensor

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

defined by

$$\beta^f = \sum_{x_0\in[m_0]} \cdots \sum_{x_{k-1}\in[m_{k-1}]} \epsilon_{q(x_0,\dots,x_{d-1})} \otimes \epsilon_{x_0,\dots,x_{d-1}} .$$

Data tensor

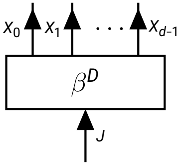
In the case of the sample selection map

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

we have a representation tensor called the data tensor

$$\beta^D = \sum_{j \in [m]} \epsilon_j \otimes \epsilon_{x_0^j, \dots, x_{d-1}^j}$$

depicted as



Representation of Data

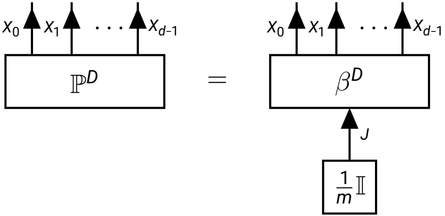
The one-hot encoding of each observation is stored on a slice of the data tensor.

Average of the Observation

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

$$\mathbb{P}^D := \frac{1}{m}\sum_{j\in[m]} \epsilon_{x^j_0,\dots,x^j_{d-1}}$$

In a contraction diagram we denote the average by



.

More efficient data representation

The storage demand of β^D increases exponentially

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

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

The collection of encoding tensors

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

to the maps

$$D_k : [m] \to [m_k] \quad , \quad D_k(j) = x_k^j$$

has a linearly increasing storage demand of

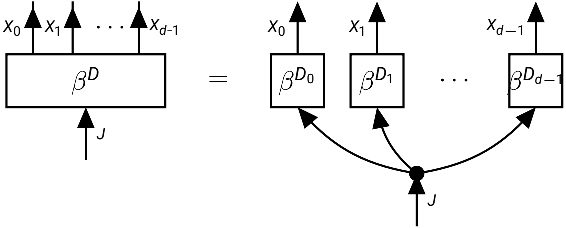
$$\sum_{k\in[d]} \dim\left[\beta^{D_k}\right] = \sum_{k\in[d]} m_k \cdot m\,.$$

Representation of the Empirical in a CP Format

The data tensor can be stored by the contraction

$$\left\langle \beta^{D} \right\rangle_{[X_0,\ldots,X_{k-1},J]} = \left\langle \beta^{D_0},\ldots,\beta^{D_{d-1}} \right\rangle_{[X_0,\ldots,X_{d-1},J]}$$

depicted as



This is called a CP Tensor Decomposition.

CP Decomposition of Empirical Distributions

Instead of storing the datacore β^D, store the encodings β^{D_k}.