# II-Probabilities Graphical Models: Representing Probabilities as Tensor Networks

## Foundations of Neuro-Symbolic AI

Alex Goessmann

University of Applied Science Würzburg-Schweinfurt

Summer Term 2026

# How to beat the curse of dimensions?

Probability distribution of factored systems with states $\bigtimes_{k\in[d]}[m_k]$ has

$$\left(\prod_{k\in[d]} m_k\right) - 1$$

degrees of freedom (coordinates to specify and store).

Mitigation: Tensor Network Decompositions

**Independencies of Random Variables**

Decompositions of Probability Tensors correspond with independencies of (hidden) random variables.

# Example: Being at a dentist

Add a variable Cloud, denoting the weather outside the dentists lab.

- This adds an additional axis to $\mathbb{P}$, thus the number of coordinates increases by a factor of 2.
- But: Intuitively, knowing Cloud should not affect the probability of having a cavity, so why shall we care?

**Independence of Cloud to the other Variables**

After showing that cavity, catch and toothache are independent of cloud, we do not have to consider cloud any more.

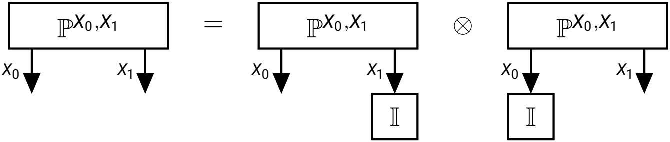
# Formal definition of Independencies

## Definition (Independence)

Given a joint distribution of variables $X_0$ and $X_1$, we say that $X_0$ is independent from $X_1$ if for any values $x_{X_0}, x_{X_1}$ we have

$$\mathbb{P}^{X_0=x_{X_0}, X_1=x_{X_1}} = \mathbb{P}\left[X_0 = x_{X_0}\right] X_0 \cdot \mathbb{P}\left[X_1 = x_{X_1}\right] X_1 \,.$$

In the tensor network decomposition we depict this by



# Contraction criterion for independence

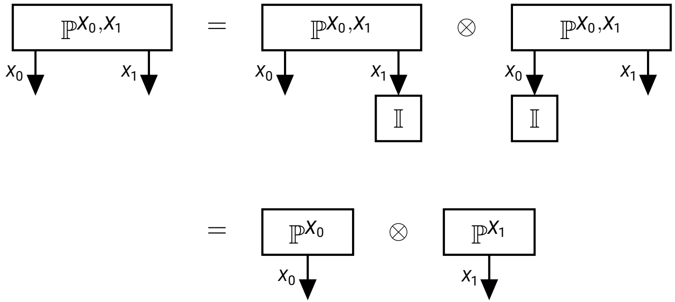
**Theorem**

*Given a probability distribution $\mathbb{P}$, $X_0$ is independent from $X_1$, if and only if*

$$\mathbb{P} = \langle \mathbb{P} \rangle_{[X_0]} \otimes \langle \mathbb{P} \rangle_{[X_1]} \, .$$

# Decomposition into Marginal Probability Tensors

Independence allows the decomposition into




## Exponential to linear storage demand

Instead of storing $m_{X_0} \cdot m_{X_1}$ coordindates, we can store $\mathbb{P}$ with $m_{X_0} + m_{X_1}$ demand.

# Formal definition of Conditional Independencies

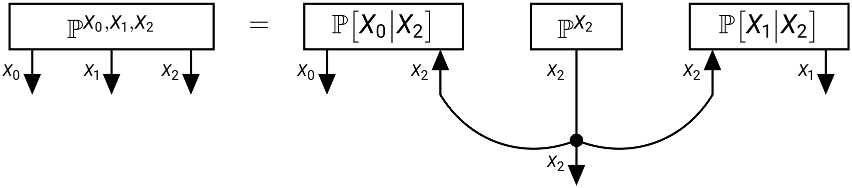
**Definition (Conditional Independence)**

Given a joint distribution of variables $X_0, X_1$ and $X_2$, we say $X_0$ is independent from $X_1$ conditioned on $X_2$ if

$$\mathbb{P}\left[X_0, X_1 \middle| X_2\right] = \mathbb{P}\left[X_0 \middle| X_2\right] \cdot \mathbb{P}\left[X_1 \middle| X_2\right] .$$

# Decomposition given conditional independence

We depict conditional independence by tensor network decompositions:



# Chain Rule: Decomposing Probabilities

**Theorem (Chain Rule)**

*For any joint probability distribution* $\mathbb{P}$ *of the variables* $\mathbb{P}^{X_0,\ldots,X_{d-1}}$ *we have*

$$\mathbb{P} = \left\langle \{\mathbb{P}\left[X_k \middle| X_0,\ldots,X_{k-1}\right] \;:\; k \in [d]\} \right\rangle_{\left[\{f^0,\ldots,f^{d-1}\}\right]}$$

*where for* $k = 1$ *we denote by* $\mathbb{P}\left[X_0 \middle| X_0,\ldots,X_{-1}\right]$ *the marginal distribution* $\mathbb{P}^{X_0}$.

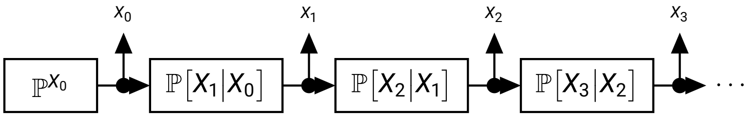
# Markov Chains

## Theorem (Markov Chain)

*Let there be a set of variables $X_t$ where $t \in [T]$. When $X_t$ is independent of $X_{0:t-2}$ conditioned on $X_{t-1}$ (the Markov Property), then*

$$\mathbb{P} = \left\langle \mathbb{P}\left[X_t \middle| X_{t-1}\right] \;:\; t \in [T] \right\rangle_{[X_0,\ldots,X_{T-1}]} .$$

We depict this by



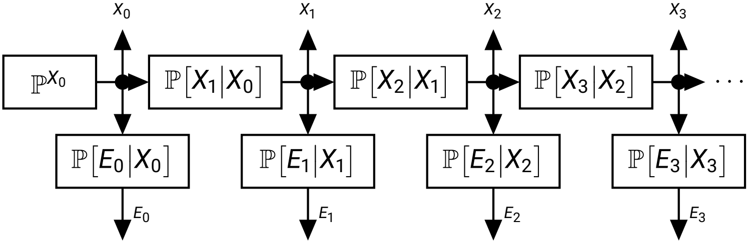
# Hidden Markov Models

Hidden Markov Models extend Markov Chains by limited observation $E_t$ of the variables $X_t$.

The independence assumptions are

- $X_{t+1}$ is independent of $X_{0:t-1}$ and $E_{0:t}$ conditioned on $X_t$
- $E_t$ is independent of all other variables conditioned on $X_t$

The independence assumptions are exploited in the decomposition



# Directed Graphical Models: Bayesian Networks

## Definition (Bayesian Networks)

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed acyclic graph and for each node $v \in \mathcal{V}$ a random variable $X_v$. Further let there be for each node $v \in \mathcal{V}$ with parents $\mathrm{Pa}(v)$ a conditional probability distribution

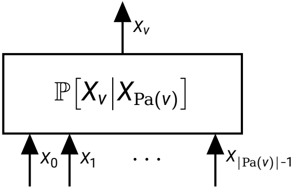
$$\mathbb{P}\left[X_v \middle| X_{\mathrm{Pa}(v)}\right] .$$

Then the Bayesian Network with respect to $\mathcal{G}$ and the conditional probability terms is the distribution

$$\mathbb{P} = \left\langle \mathbb{P}\left[X_v \middle| X_{\mathrm{Pa}(v)}\right] \;:\; v \in \mathcal{V} \right\rangle_{[X_v \,:\, v \in \mathcal{V}]} .$$

# Directed Graphical Models: Bayesian Networks

For each variable we build the conditional probability tensor




The Bayesian Network is then the contraction

$$\mathbb{P} = \left\langle \mathbb{P}\left[X_v \middle| X_{\mathrm{Pa}(v)}\right] \; : \; v \in \mathcal{V} \right\rangle_{[X_v \, : \, v \in \mathcal{V}]} \, .$$

# Undirected Graphical Models: Markov Networks

## Definition (Markov Networks)

Let $\tau^{\mathcal{G}}$ be a Tensor Network on a hypergraph $\mathcal{G}$. The associated Markov Network is the probability distribution of $\{X_v \,:\, v \in \mathcal{V}\}$ defined by

$$\mathbb{P}^{\mathcal{G}} = \frac{\left\langle \{\tau^e : e \in \mathcal{E}\} \right\rangle_{[\mathcal{V}]}}{\left\langle \{\tau^e : e \in \mathcal{E}\} \right\rangle_{[\varnothing]}} = \left\langle \{\tau^e : e \in \mathcal{E}\} \right\rangle_{[\mathcal{V}|\varnothing]} .$$

We call the denominator

$$\mathcal{Z}(\mathcal{G}) = \left\langle \{\tau^e : e \in \mathcal{E}\} \right\rangle_{[\varnothing]}$$

the partition function of the Markov Network.