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

## Logik für Erklärbare KI: Technische Einführung in das ENEXA Projekt

Maria Schnödt-Fuchs, Alex Goeßmann

Funded by the European Union

15.+16. July, 2024

# 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.

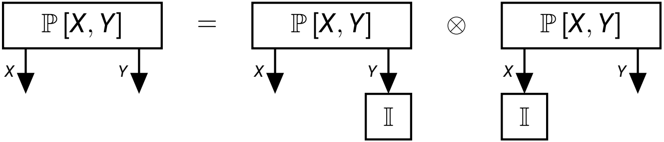
# Formal definition of Independencies

## Definition (Independence)

Given a joint distribution of variables $X$ and $Y$, we say that $X$ is independent from $Y$ if for any values $i_X, i_Y$ we have

$$\mathbb{P}\left[X = i_X, Y = i_Y\right] = \mathbb{P}\left[X = i_X\right] \cdot \mathbb{P}\left[Y = i_Y\right] \, .$$

In the tensor network decomposition we depict this by



# Contraction criterion for independence

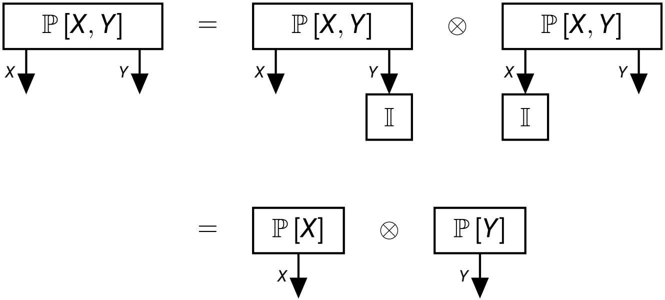
**Theorem**

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

$$\mathbb{P} = \mathcal{C}\left(\{\mathbb{P}\}, \{X\}\right) \otimes \mathcal{C}\left(\{\mathbb{P}\}, \{Y\}\right) .$$

# Decomposition into Marginal Probability Tensors

Independence allows the decomposition into




## Exponential to linear storage demand

Instead of storing $m_X \cdot m_Y$ coordindates, we can store $\mathbb{P}$ with $m_X + m_Y$ demand.

# Formal definition of Conditional Independencies

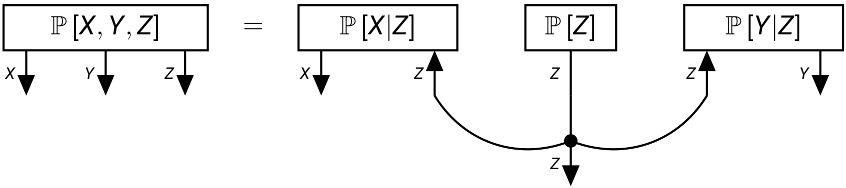
**Definition (Conditional Independence)**

Given a joint distribution of variables $X$, $Y$ and $Z$, we say $X$ is independent from $Y$ conditioned on $Z$ if

$$\mathbb{P}[X, Y|Z] = \mathbb{P}[X|Z] \cdot \mathbb{P}[Y|Z] \ .$$

# 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* $\mathbb{P}\left[X_0, \ldots, X_{d-1}\right]$ *we have*

$$\mathbb{P} = \mathcal{C}\left(\left\{\mathbb{P}\left[X_k | X_0, \ldots, X_{k-1}\right] \,:\, k \in [d]\right\}, \left\{X_0, \ldots, X_{d-1}\right\}\right)$$

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

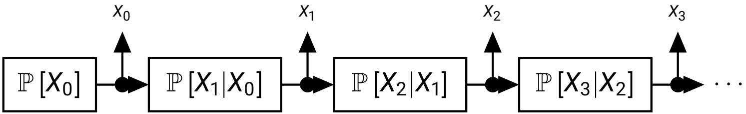
# Markov Chains

## 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*

$$\mathbb{P} = \mathcal{C}\left(\{\mathbb{P}\left[X_t | X_{t-1}\right] \,:\, t \in [T]\}, \{X_0, \ldots, X_{T-1}\}\right) .$$

We depict this by



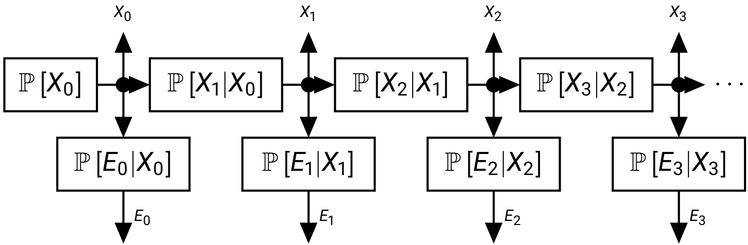
# Hidden Markov Models

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

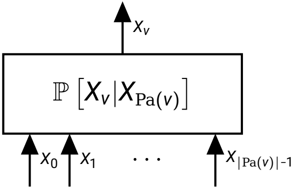
$$\mathbb{P}\left[X_v | X_{\mathrm{Pa}(v)}\right] .$$

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

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

# Directed Graphical Models: Bayesian Networks

For each variable we build the conditional probability tensor




The Bayesian Network is then the contraction

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

# Undirected Graphical Models: Markov Networks

## Definition (Markov Networks)

Let $\mathcal{T}^{\mathcal{G}}$ 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

$$\mathbb{P}^{\mathcal{G}} = \frac{\mathcal{C}\left(\{T^e : e \in \mathcal{E}\}, \mathcal{V}\right)}{\mathcal{C}\left(\{T^e : e \in \mathcal{E}\}, \varnothing\right)} = \mathcal{N}\left(\{T^e : e \in \mathcal{E}\}, \mathcal{V}, \varnothing\right) .$$

We call the denominator

$$\mathcal{Z}(\mathcal{G}) = \mathcal{C}\left(\{T^e : e \in \mathcal{E}\}, \varnothing\right)$$

the partition function of the Markov Network.