# Rank-one decomposition of probability tables

Petr Savický and Jiří Vomlel

Academy of Sciences of the Czech Republic (AV ČR)

Mikulov, September 20, 2006

# Notation

- An index set $N = \{1, 2, \ldots, n\}$
- Variables $X_i, i \in N$
- Each variable $X_i$ takes values from a finite set $\mathcal{X}_i$
- For $A \subseteq N$ we use $X_A$ to denote the multidimensional variable $(X_i)_{i \in A}$
- For $A \subseteq N$ we use $x_A$ to denote the vector of values of variable $X_A$
- For $A \subseteq N$ we use $\mathcal{X}_A$ to denote $\times_{i \in A} \mathcal{X}_i$

# Notation

- An index set $N = \{1, 2, \ldots, n\}$
- Variables $X_i, i \in N$
- Each variable $X_i$ takes values from a finite set $\mathcal{X}_i$
- For $A \subseteq N$ we use $X_A$ to denote the multidimensional variable $(X_i)_{i \in A}$
- For $A \subseteq N$ we use $x_A$ to denote the vector of values of variable $X_A$
- For $A \subseteq N$ we use $\mathcal{X}_A$ to denote $\times_{i \in A} \mathcal{X}_i$

# Notation

- An index set $N = \{1, 2, \ldots, n\}$
- Variables $X_i, i \in N$
- Each variable $X_i$ takes values from a finite set $\mathcal{X}_i$
- For $A \subseteq N$ we use $X_A$ to denote the multidimensional variable $(X_i)_{i \in A}$
- For $A \subseteq N$ we use $x_A$ to denote the vector of values of variable $X_A$
- For $A \subseteq N$ we use $\mathcal{X}_A$ to denote $\times_{i \in A} \mathcal{X}_i$

# Notation

- An index set $N = \{1, 2, \ldots, n\}$
- Variables $X_i, i \in N$
- Each variable $X_i$ takes values from a finite set $\mathcal{X}_i$
- For $A \subseteq N$ we use $X_A$ to denote the multidimensional variable $(X_i)_{i \in A}$
- For $A \subseteq N$ we use $x_A$ to denote the vector of values of variable $X_A$
- For $A \subseteq N$ we use $\mathcal{X}_A$ to denote $\times_{i \in A} \mathcal{X}_i$

# Notation

- An index set $N = \{1, 2, \ldots, n\}$
- Variables $X_i, i \in N$
- Each variable $X_i$ takes values from a finite set $\mathcal{X}_i$
- For $A \subseteq N$ we use $X_A$ to denote the multidimensional variable $(X_i)_{i \in A}$
- For $A \subseteq N$ we use $x_A$ to denote the vector of values of variable $X_A$
- For $A \subseteq N$ we use $\mathcal{X}_A$ to denote $\times_{i \in A} \mathcal{X}_i$

# Notation

- An index set $N = \{1, 2, \ldots, n\}$
- Variables $X_i, i \in N$
- Each variable $X_i$ takes values from a finite set $\mathcal{X}_i$
- For $A \subseteq N$ we use $X_A$ to denote the multidimensional variable $(X_i)_{i \in A}$
- For $A \subseteq N$ we use $x_A$ to denote the vector of values of variable $X_A$
- For $A \subseteq N$ we use $\mathcal{X}_A$ to denote $\times_{i \in A} \mathcal{X}_i$

# Tables

**Definition**

Assume $A \subseteq N$. Table is a function $\psi : \mathcal{X}_A \rightarrow \mathbb{A}$.

- If $\mathbb{A} = [0,1]$ and $\sum_{x \in \mathcal{X}} \psi(x) = 1$ then $\psi$ is a probability table.
- If $\mathbb{A} = [0,1]$ and there exists a set $B \subset A \subseteq N$ such that for every $x_B$ it holds $\sum_{x_C} \psi(x_B, x_C) = 1$, where $C = A \setminus B$, then $\psi(X_C | X_B)$ is a conditional probability table (CPT).

Example: $A = \{1,2,3\}$ and $B = \{1,2\}$. The table $\psi(X_3 \mid X_1, X_2)$

$$\begin{pmatrix} (0.1, 0.9) & (0.3, 0.7) \\ (0.4, 0.6) & (0.9, 0.1) \end{pmatrix}$$

is a conditional probability table.

- $\mathbb{A} = \mathbb{R}$, the set of all real numbers,
- $\mathbb{A} = \mathbb{C}$, the set of all complex numbers.

# Tables

**Definition**

Assume $A \subseteq N$. Table is a function $\psi : \mathcal{X}_A \to \mathbb{A}$.

- If $\mathbb{A} = [0,1]$ and $\sum_{x \in \mathcal{X}} \psi(x) = 1$ then $\psi$ is a probability table.
- If $\mathbb{A} = [0,1]$ and there exists a set $B \subset A \subseteq N$ such that for every $x_B$ it holds $\sum_{x_C} \psi(x_B, x_C) = 1$, where $C = A \setminus B$, then $\psi(X_C | X_B)$ is a conditional probability table (CPT).

Example: $A = \{1,2,3\}$ and $B = \{1,2\}$. The table $\psi(X_3 \mid X_1, X_2)$

$$\begin{pmatrix} (0.1, 0.9) & (0.3, 0.7) \\ (0.4, 0.6) & (0.9, 0.1) \end{pmatrix}$$

is a conditional probability table.

- $\mathbb{A} = \mathbb{R}$, the set of all real numbers,
- $\mathbb{A} = \mathbb{C}$, the set of all complex numbers.

# Tables

**Definition**

Assume $A \subseteq N$. Table is a function $\psi : \mathcal{X}_A \to \mathbb{A}$.

- If $\mathbb{A} = [0,1]$ and $\sum_{x \in \mathcal{X}} \psi(x) = 1$ then $\psi$ is a probability table.
- If $\mathbb{A} = [0,1]$ and there exists a set $B \subset A \subseteq N$ such that for every $x_B$ it holds $\sum_{x_C} \psi(x_B, x_C) = 1$, where $C = A \setminus B$, then $\psi(X_C|X_B)$ is a conditional probability table (CPT).

Example: $A = \{1,2,3\}$ and $B = \{1,2\}$. The table $\psi(X_3 \mid X_1, X_2)$

$$\begin{pmatrix} (0.1, 0.9) & (0.3, 0.7) \\ (0.4, 0.6) & (0.9, 0.1) \end{pmatrix}$$

is a conditional probability table.

- $\mathbb{A} = \mathbb{R}$, the set of all real numbers,
- $\mathbb{A} = \mathbb{C}$, the set of all complex numbers.

# Tables

**Definition**

Assume $A \subseteq N$. Table is a function $\psi : \mathcal{X}_A \to \mathbb{A}$.

- If $\mathbb{A} = [0,1]$ and $\sum_{x \in \mathcal{X}} \psi(x) = 1$ then $\psi$ is a probability table.
- If $\mathbb{A} = [0,1]$ and there exists a set $B \subset A \subseteq N$ such that for every $x_B$ it holds $\sum_{x_C} \psi(x_B, x_C) = 1$, where $C = A \setminus B$, then $\psi(X_C | X_B)$ is a conditional probability table (CPT).

Example: $A = \{1,2,3\}$ and $B = \{1,2\}$. The table $\psi(X_3 \mid X_1, X_2)$

$$\begin{pmatrix} (0.1, 0.9) & (0.3, 0.7) \\ (0.4, 0.6) & (0.9, 0.1) \end{pmatrix}$$

is a conditional probability table.

- $\mathbb{A} = \mathbb{R}$, the set of all real numbers,
- $\mathbb{A} = \mathbb{C}$, the set of all complex numbers.

# Tables

## Definition

Assume $A \subseteq N$. Table is a function $\psi : \mathcal{X}_A \to \mathbb{A}$.

- If $\mathbb{A} = [0,1]$ and $\sum_{x \in \mathcal{X}} \psi(x) = 1$ then $\psi$ is a probability table.
- If $\mathbb{A} = [0,1]$ and there exists a set $B \subset A \subseteq N$ such that for every $x_B$ it holds $\sum_{x_C} \psi(x_B, x_C) = 1$, where $C = A \setminus B$, then $\psi(X_C | X_B)$ is a conditional probability table (CPT).

Example: $A = \{1,2,3\}$ and $B = \{1,2\}$. The table $\psi(X_3 \mid X_1, X_2)$

$$\begin{pmatrix} (0.1, 0.9) & (0.3, 0.7) \\ (0.4, 0.6) & (0.9, 0.1) \end{pmatrix}$$

is a conditional probability table.

- $\mathbb{A} = \mathbb{R}$, the set of all real numbers,
- $\mathbb{A} = \mathbb{C}$, the set of all complex numbers.

# Tables

**Definition**

Assume $A \subseteq N$. Table is a function $\psi : \mathcal{X}_A \to \mathbb{A}$.

- If $\mathbb{A} = [0,1]$ and $\sum_{x \in \mathcal{X}} \psi(x) = 1$ then $\psi$ is a probability table.
- If $\mathbb{A} = [0,1]$ and there exists a set $B \subset A \subseteq N$ such that for every $x_B$ it holds $\sum_{x_C} \psi(x_B, x_C) = 1$, where $C = A \setminus B$, then $\psi(X_C | X_B)$ is a conditional probability table (CPT).

Example: $A = \{1,2,3\}$ and $B = \{1,2\}$. The table $\psi(X_3 \mid X_1, X_2)$

$$\begin{pmatrix} (0.1, 0.9) & (0.3, 0.7) \\ (0.4, 0.6) & (0.9, 0.1) \end{pmatrix}$$

is a conditional probability table.

- $\mathbb{A} = \mathbb{R}$, the set of all real numbers,
- $\mathbb{A} = \mathbb{C}$, the set of all complex numbers.

# Multiplication of tables

## Definition

Assume two tables $\psi_1 : \mathcal{X}_Q \to \mathbb{A}$ and $\psi_2 : \mathcal{X}_R \to \mathbb{A}$, where $Q, R \subseteq N$.
The product $\psi_1 \otimes \psi_2$ is a table $\psi_3 : \mathcal{X}_{Q \cup R} \to \mathbb{A}$ such that for all $x_{Q \cup R}$:

$$\psi_3(x_{Q \cup R}) \quad = \quad \psi_1(x_Q) \cdot \psi_2(x_R) \ .$$

Example: $Q = \{1, 2\}$ and $R = \{1, 3\}$

$$\begin{pmatrix} \mathrm{a} & \mathrm{b} \\ \mathrm{c} & \mathrm{d} \end{pmatrix} \otimes \begin{pmatrix} (e, f) & (g, h) \end{pmatrix} = \begin{pmatrix} (\mathrm{ae},\mathrm{af}) & (\mathrm{bg},\mathrm{bh}) \\ (\mathrm{ce},\mathrm{cf}) & (\mathrm{dg},\mathrm{dh}) \end{pmatrix}$$

# Multiplication of tables

## Definition

Assume two tables $\psi_1 : \mathcal{X}_Q \to \mathbb{A}$ and $\psi_2 : \mathcal{X}_R \to \mathbb{A}$, where $Q, R \subseteq N$. The product $\psi_1 \otimes \psi_2$ is a table $\psi_3 : \mathcal{X}_{Q \cup R} \to \mathbb{A}$ such that for all $x_{Q \cup R}$:

$$\psi_3(x_{Q\cup R}) \quad = \quad \psi_1(x_Q) \cdot \psi_2(x_R) \enspace .$$

Example: $Q = \{1,2\}$ and $R = \{1,3\}$

$$\begin{pmatrix} \mathrm{a} & \mathrm{b} \\ \mathrm{c} & \mathrm{d} \end{pmatrix} \otimes \begin{pmatrix} (e,f) & (g,h) \end{pmatrix} = \begin{pmatrix} (\mathrm{ae},\mathrm{af}) & (\mathrm{bg},\mathrm{bh}) \\ (\mathrm{ce},\mathrm{cf}) & (\mathrm{dg},\mathrm{dh}) \end{pmatrix}$$

# Addition of tables

## Definition

Assume two tables $\psi_1, \psi_2 : \mathcal{X}_Q \to \mathbb{A}$, where $Q \subseteq N$.
The sum $\psi_1 \oplus \psi_2$ is a table $\psi_3 : \mathcal{X}_Q \to \mathbb{A}$ such that for all $x_Q$:

$$\psi_3(x_Q) \quad = \quad \psi_1(x_Q) + \psi_2(x_Q) \ .$$

Example: $Q = \{1, 2\}$.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \oplus \begin{pmatrix} e & f \\ g & h \end{pmatrix} \quad = \quad \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix}$$

# Addition of tables

**Definition**

Assume two tables $\psi_1, \psi_2 : \mathcal{X}_Q \to \mathbb{A}$, where $Q \subseteq N$.
The sum $\psi_1 \oplus \psi_2$ is a table $\psi_3 : \mathcal{X}_Q \to \mathbb{A}$ such that for all $x_Q$:

$$\psi_3(x_Q) = \psi_1(x_Q) + \psi_2(x_Q) \ .$$

Example: $Q = \{1, 2\}$.

$$\begin{pmatrix} \mathsf{a} & \mathsf{b} \\ \mathsf{c} & \mathsf{d} \end{pmatrix} \oplus \begin{pmatrix} \mathsf{e} & \mathsf{f} \\ \mathsf{g} & \mathsf{h} \end{pmatrix} = \begin{pmatrix} \mathsf{a+e} & \mathsf{b+f} \\ \mathsf{c+g} & \mathsf{d+h} \end{pmatrix}$$

# Multiplicative decomposition of a table

In the rest of the presentation we will assume (without loss of generality) that $A = N = \{1, 2, \ldots, n\}$.

**Definition**

A table $\psi : \mathcal{X}_N \to \mathbb{A}$ factorizes with respect to a system $\{C_1, \ldots, C_k\}$, $C_j \subseteq N, j = 1, \ldots, k$ iff there exist tables $\psi_j : \mathcal{X}_{C_j} \to \mathbb{A}, j = 1, \ldots, k$ such that $\psi = \psi_1 \otimes \ldots \otimes \psi_k$.

# Multiplicative decomposition of a table

In the rest of the presentation we will assume (without loss of generality) that $A = N = \{1, 2, \ldots, n\}$.

## Definition

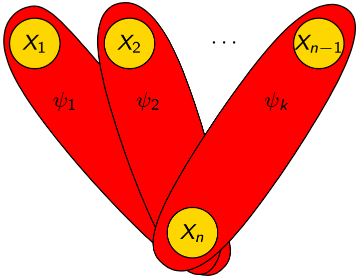
A table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ factorizes with respect to a system $\{C_1, \ldots, C_k\}$, $C_j \subseteq N, j = 1, \ldots, k$ iff there exist tables $\psi_j : \mathcal{X}_{C_j} \rightarrow \mathbb{A}, j = 1, \ldots, k$ such that $\psi \ = \ \psi_1 \otimes \ldots \otimes \psi_k$.

# Multiplicative decomposition of a table

In the rest of the presentation we will assume (without loss of generality) that $A = N = \{1, 2, \ldots, n\}$.

## Definition

A table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ factorizes with respect to a system $\{C_1, \ldots, C_k\}$, $C_j \subseteq N, j = 1, \ldots, k$ iff there exist tables $\psi_j : \mathcal{X}_{C_j} \rightarrow \mathbb{A}, j = 1, \ldots, k$ such that $\psi = \psi_1 \otimes \ldots \otimes \psi_k$.



# Rank-one tables

**Definition**

If a table $\psi : \mathcal{X}_N \to \mathbb{A}$ factorizes with respect to the system of all singletons (i.e., $C_j = \{j\}, j = 1, \ldots, n$) then we say that table $\psi$ has rank one.

Example:

$$\begin{pmatrix} (0,1) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} = (1,1) \otimes (1,1) \otimes (0,1)$$

therefore it has rank one.

# Rank-one tables

## Definition

If a table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ factorizes with respect to the system of all singletons (i.e., $C_j = \{j\}, j = 1, \ldots, n$) then we say that table $\psi$ has rank one.

Example:

$$\left( \begin{array}{cc} (0,1) & (0,1) \\ (0,1) & (0,1) \end{array} \right) = (1,1) \otimes (1,1) \otimes (0,1)$$

therefore it has rank one.

# Rank-one decomposition of a table

**Definition**

We say that table $\psi : \mathcal{X}_N \to \mathbb{A}$ has degree $r$ (in $\mathbb{A}$) if there exist tables $\psi_j : \mathcal{X}_N \to \mathbb{A}, j = 1, \ldots, r$ of rank one such that $\psi = \psi_1 \oplus \ldots \oplus \psi_r$.

**Definition**

The minimal degree of a table $\psi : \mathcal{X}_N \to \mathbb{A}$ is called rank of $\psi$ (in $\mathbb{A}$).

Example:

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} = \begin{pmatrix} (0,1) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} + \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (0,0) \end{pmatrix}$$

$$= ((1,1) \otimes (1,1) \otimes (0,1)) \oplus ((1,0) \otimes (1,0) \otimes (1,-1))$$

therefore it has degree 2 in $\mathbb{R}$.

# Rank-one decomposition of a table

## Definition

We say that table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ has degree $r$ (in $\mathbb{A}$) if there exist tables $\psi_j : \mathcal{X}_N \rightarrow \mathbb{A}, j = 1, \ldots, r$ of rank one such that $\psi = \psi_1 \oplus \ldots \oplus \psi_r$.

## Definition

The minimal degree of a table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ is called rank of $\psi$ (in $\mathbb{A}$).

Example:

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} = \begin{pmatrix} (0,1) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} + \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (0,0) \end{pmatrix}$$

$$= ((1,1) \otimes (1,1) \otimes (0,1)) \oplus ((1,0) \otimes (1,0) \otimes (1,-1))$$

therefore it has degree 2 in $\mathbb{R}$.

## Definition

We say that table $\psi : \mathcal{X}_N \to \mathbb{A}$ has degree $r$ (in $\mathbb{A}$) if there exist tables $\psi_j : \mathcal{X}_N \to \mathbb{A}, j = 1, \ldots, r$ of rank one such that $\psi = \psi_1 \oplus \ldots \oplus \psi_r$.

## Definition

The minimal degree of a table $\psi : \mathcal{X}_N \to \mathbb{A}$ is called rank of $\psi$ (in $\mathbb{A}$).

Example:

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} = \begin{pmatrix} (0,1) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} + \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (0,0) \end{pmatrix}$$

$$= ((1,1) \otimes (1,1) \otimes (0,1)) \oplus ((1,0) \otimes (1,0) \otimes (1,-1))$$

therefore it has degree 2 in $\mathbb{R}$.

# Rank-one decomposition of a table

**Definition**

We say that table $\psi : \mathcal{X}_N \to \mathbb{A}$ has degree $r$ (in $\mathbb{A}$) if there exist tables $\psi_j : \mathcal{X}_N \to \mathbb{A}, j = 1, \ldots, r$ of rank one such that $\psi = \psi_1 \oplus \ldots \oplus \psi_r$.

**Definition**

The minimal degree of a table $\psi : \mathcal{X}_N \to \mathbb{A}$ is called rank of $\psi$ (in $\mathbb{A}$).

Example:

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} = \begin{pmatrix} (0,1) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} + \begin{pmatrix} (1,-1) & (0,0) \\ (0,0) & (0,0) \end{pmatrix}$$

$$= ((1,1) \otimes (1,1) \otimes (0,1)) \oplus ((1,0) \otimes (1,0) \otimes (1,-1))$$

therefore it has degree 2 in $\mathbb{R}$.

# Rank-one decomposition of a table

**Definition**

We say that table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ has degree $r$ (in $\mathbb{A}$) if there exist tables $\psi_j : \mathcal{X}_N \rightarrow \mathbb{A}, j = 1, \ldots, r$ of rank one such that $\psi = \psi_1 \oplus \ldots \oplus \psi_r$.

**Definition**

The minimal degree of a table $\psi : \mathcal{X}_N \rightarrow \mathbb{A}$ is called rank of $\psi$ (in $\mathbb{A}$).

Example:

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} = \begin{pmatrix} (0,1) & (0,1) \\ (0,1) & (0,1) \end{pmatrix} + \begin{pmatrix} (1,\text{-}1) & (0,0) \\ (0,0) & (0,0) \end{pmatrix}$$

$$= ((1,1) \otimes (1,1) \otimes (0,1)) \oplus ((1,0) \otimes (1,0) \otimes (1,-1))$$

therefore it has degree 2 in $\mathbb{R}$.

# Auxiliary variable $X'_n$

Rank-one decomposition of $\psi : \mathcal{X}_N \to \mathbb{A}$ of degree $r$ can be also written as

$$\psi(X_N) \quad = \quad \sum_{X'_n} \psi_i(X_1, X'_n) \otimes \ldots \otimes \psi_i(X_n, X'_n) \ ,$$

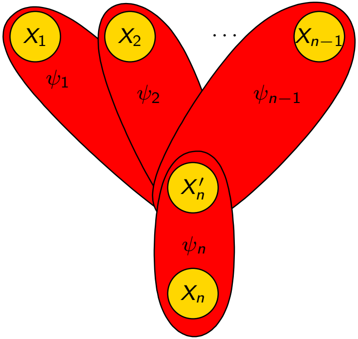
where $X'_n$ is an auxiliary variable having states from $\{1, \ldots, r\}$ and $\sum_X \psi(X, Y)$ is a shorthand notation for the sum over the states of variable $X$.

# Auxiliary variable $X'_n$

Rank-one decomposition of $\psi : \mathcal{X}_N \to \mathbb{A}$ of degree $r$ can be also written as

$$\psi(X_N) \quad = \quad \sum_{X'_n} \psi_i(X_1, X'_n) \otimes \ldots \otimes \psi_i(X_n, X'_n) \ ,$$

where $X'_n$ is an auxiliary variable having states from $\{1, \ldots, r\}$ and $\sum_X \psi(X, Y)$ is a shorthand notation for the sum over the states of variable $X$.



# Motivation for table decompositions

- Decompositions allows more compact representation of tables. This reduces space requirements.
- Computations with decomposed tables are more efficient. This reduces time requirements during inference.
- The estimation of tables from data or from domain experts is easier and more reliable if they are decomposed.

Multiplicative decomposition is known and efficiently used for inference in probabilistic graphical models (Pearl, 1988) and (Lauritzen, Spiegelhalter, 1988).

Rank-one decomposition is relatively new (Díez and Galán, IJIS, 2002), (Vomlel, UAI'02), and (Savicky, Vomlel, IPMU'06). At this moment it allows efficient inference with some tables that cannot be multiplicatively decomposed.

# Motivation for table decompositions

- Decompositions allows more compact representation of tables. This reduces space requirements.
- Computations with decomposed tables are more efficient. This reduces time requirements during inference.
- The estimation of tables from data or from domain experts is easier and more reliable if they are decomposed.

Multiplicative decomposition is known and efficiently used for inference in probabilistic graphical models (Pearl, 1988) and (Lauritzen, Spiegelhalter, 1988).
Rank-one decomposition is relatively new (Díez and Galán, IJIS, 2002), (Vomlel, UAI'02), and (Savicky, Vomlel, IPMU'06). At this moment it allows efficient inference with some tables that cannot be multiplicatively decomposed.

# Motivation for table decompositions

- Decompositions allows more compact representation of tables. This reduces space requirements.
- Computations with decomposed tables are more efficient. This reduces time requirements during inference.
- The estimation of tables from data or from domain experts is easier and more reliable if they are decomposed.

Multiplicative decomposition is known and efficiently used for inference in probabilistic graphical models (Pearl, 1988) and (Lauritzen, Spiegelhalter, 1988).

Rank-one decomposition is relatively new (Díez and Galán, IJIS, 2002), (Vomlel, UAI'02), and (Savicky, Vomlel, IPMU'06). At this moment it allows efficient inference with some tables that cannot be multiplicatively decomposed.

# Motivation for table decompositions

- Decompositions allows more compact representation of tables. This reduces space requirements.
- Computations with decomposed tables are more efficient. This reduces time requirements during inference.
- The estimation of tables from data or from domain experts is easier and more reliable if they are decomposed.

Multiplicative decomposition is known and efficiently used for inference in probabilistic graphical models (Pearl, 1988) and (Lauritzen, Spiegelhalter, 1988).

Rank-one decomposition is relatively new (Díez and Galán, IJIS, 2002), (Vomlel, UAI'02), and (Savicky, Vomlel, IPMU'06). At this moment it allows efficient inference with some tables that cannot be multiplicatively decomposed.

# Motivation for table decompositions

- Decompositions allows more compact representation of tables. This reduces space requirements.
- Computations with decomposed tables are more efficient. This reduces time requirements during inference.
- The estimation of tables from data or from domain experts is easier and more reliable if they are decomposed.

Multiplicative decomposition is known and efficiently used for inference in probabilistic graphical models (Pearl, 1988) and (Lauritzen, Spiegelhalter, 1988).
Rank-one decomposition is relatively new (Díez and Galán, IJIS, 2002), (Vomlel, UAI'02), and (Savicky, Vomlel, IPMU'06). At this moment it allows efficient inference with some tables that cannot be multiplicatively decomposed.

# Deterministic CPTs generated by functions

Now we are going to discuss tables of the form $\psi(X_a \mid X_{N\setminus\{a\}})$ where $a \in N$.

Without loss of generality assume that $a = n$.

From now $M$ will denote $N \setminus \{n\}$.

**Definition**

Assume a function $f : \mathcal{X}_M \to \mathcal{X}_n$. We say that $\psi_f(X_n \mid X_M)$ is generated by function $f$ if

$$\psi_f(x_n \mid x_M) = \begin{cases} 1 & \text{if } x_n = f(x_M), \\ 0 & \text{otherwise.} \end{cases}$$

Example: Assume $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 = \{0,1\}$. Then table

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix}$$

is generated by function $f(x_1, x_2) = x_1 \vee x_2$.

# Deterministic CPTs generated by functions

Now we are going to discuss tables of the form $\psi(X_a \mid X_{N\setminus\{a\}})$ where $a \in N$.

Without loss of generality assume that $a = n$.

From now $M$ will denote $N \setminus \{n\}$.

**Definition**

Assume a function $f : \mathcal{X}_M \to \mathcal{X}_n$. We say that $\psi_f(X_n \mid X_M)$ is generated by function $f$ if

$$\psi_f(x_n \mid x_M) = \begin{cases} 1 & \text{if } x_n = f(x_M), \\ 0 & \text{otherwise.} \end{cases}$$

Example: Assume $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 = \{0,1\}$. Then table

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix}$$

is generated by function $f(x_1, x_2) = x_1 \vee x_2$.

# Deterministic CPTs generated by functions

Now we are going to discuss tables of the form $\psi(X_a \mid X_{N\setminus\{a\}})$ where $a \in N$.
Without loss of generality assume that $a = n$.
From now $M$ will denote $N \setminus \{n\}$.

**Definition**

Assume a function $f : \mathcal{X}_M \to \mathcal{X}_n$. We say that $\psi_f(X_n \mid X_M)$ is generated by function $f$ if

$$\psi_f(x_n \mid x_M) = \begin{cases} 1 & \text{if } x_n = f(x_M), \\ 0 & \text{otherwise.} \end{cases}$$

Example: Assume $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 = \{0,1\}$. Then table

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix}$$

is generated by function $f(x_1, x_2) = x_1 \vee x_2$.

# Deterministic CPTs generated by functions

Now we are going to discuss tables of the form $\psi(X_a \mid X_{N\setminus\{a\}})$ where $a \in N$.
Without loss of generality assume that $a = n$.
From now $M$ will denote $N \setminus \{n\}$.

## Definition

Assume a function $f : \mathcal{X}_M \to \mathcal{X}_n$. We say that $\psi_f(X_n \mid X_M)$ is generated by function $f$ if

$$\psi_f(x_n \mid x_M) = \begin{cases} 1 & \text{if } x_n = f(x_M), \\ 0 & \text{otherwise.} \end{cases}$$

Example: Assume $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 = \{0,1\}$. Then table

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix}$$

is generated by function $f(x_1, x_2) = x_1 \vee x_2$.

# Deterministic CPTs generated by functions

Now we are going to discuss tables of the form $\psi(X_a \mid X_{N \setminus \{a\}})$ where $a \in N$.

Without loss of generality assume that $a = n$.

From now $M$ will denote $N \setminus \{n\}$.

## Definition

Assume a function $f : \mathcal{X}_M \to \mathcal{X}_n$. We say that $\psi_f(X_n \mid X_M)$ is generated by function $f$ if

$$\psi_f(x_n \mid x_M) = \begin{cases} 1 & \text{if } x_n = f(x_M), \\ 0 & \text{otherwise.} \end{cases}$$

Example: Assume $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3 = \{0, 1\}$. Then table

$$\begin{pmatrix} (1,0) & (0,1) \\ (0,1) & (0,1) \end{pmatrix}$$

is generated by function $f(x_1, x_2) = x_1 \vee x_2$.

# Rank of some deterministic CPTs

$\mathcal{X}_i = \{0, 1, \ldots, a_i\}, a_i \in \mathbb{N}$, for $i = 1, \ldots, n-1$,

**Lemma (Rank lower bound)**

*The rank of $\psi_f$ is greater or equal to $|\mathcal{X}_n|$.*

**Theorem (Rank of some deterministic tables)**

*If $f : \mathcal{X}_M \to \mathbb{R}$ is*

- $\min\{x_1, \ldots, x_{n-1}\}$ *or*
- $\max\{x_1, \ldots, x_{n-1}\}$ *or*
- $x_1 + \ldots + x_{n-1}$

*then rank of $\psi_f$ is $|\mathcal{X}_n|$.*

# Rank of some deterministic CPTs

$\mathcal{X}_i = \{0, 1, \ldots, a_i\}, a_i \in \mathbb{N}$, for $i = 1, \ldots, n-1$,

**Lemma (Rank lower bound)**

*The rank of $\psi_f$ is greater or equal to $|\mathcal{X}_n|$.*

**Theorem (Rank of some deterministic tables)**

*If $f : \mathcal{X}_M \to \mathbb{R}$ is*

- $\min\{x_1, \ldots, x_{n-1}\}$ *or*
- $\max\{x_1, \ldots, x_{n-1}\}$ *or*
- $x_1 + \ldots + x_{n-1}$

*then rank of $\psi_f$ is $|\mathcal{X}_n|$.*

# Rank of some deterministic CPTs

$\mathcal{X}_i = \{0, 1, \ldots, a_i\}, a_i \in \mathbb{N}$, for $i = 1, \ldots, n-1$,

## Lemma (Rank lower bound)

*The rank of $\psi_f$ is greater or equal to $|\mathcal{X}_n|$.*

## Theorem (Rank of some deterministic tables)

*If $f : \mathcal{X}_M \to \mathbb{R}$ is*

- $\min\{x_1, \ldots, x_{n-1}\}$ *or*
- $\max\{x_1, \ldots, x_{n-1}\}$ *or*
- $x_1 + \ldots + x_{n-1}$

*then rank of $\psi_f$ is $|\mathcal{X}_n|$.*

# Probabilistic tables of noisy functional dependence

## Definition

Assume a function $f : \mathcal{X}_M \to \mathcal{X}_n$. We say that $\xi_f(X_n \mid X_M)$ represents noisy functional dependence generated by function $f$ if there exists $\psi_f(X_n \mid X_M)$ (deterministic core) generated by function $f$ and tables $\varphi_i : \mathcal{X}_i \times \mathcal{X}_i \to [0, 1]$, $i = 1, \ldots, n$ such that

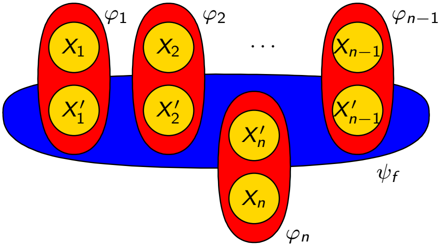
$$\xi_f(X_n \mid X_M) \quad = \quad \sum_{X'_M} \psi_f(X_n \mid X'_M) \otimes_{i=1}^{n} \varphi_i(X'_i, X_i) \ .$$

# Probabilistic tables of noisy functional dependence

**Definition**

Assume a function $f : \mathcal{X}_M \rightarrow \mathcal{X}_n$. We say that $\xi_f(X_n \mid X_M)$ represents noisy functional dependence generated by function $f$ if there exists $\psi_f(X_n \mid X_M)$ (deterministic core) generated by function $f$ and tables $\varphi_i : \mathcal{X}_i \times \mathcal{X}_i \rightarrow [0, 1]$, $i = 1, \ldots, n$ such that

$$\xi_f(X_n \mid X_M) \quad = \quad \sum_{X'_M} \psi_f(X_n \mid X'_M) \otimes_{i=1}^{n} \varphi_i(X'_i, X_i) \ .$$



# Rank of tables of noisy functional dependence

**Theorem**

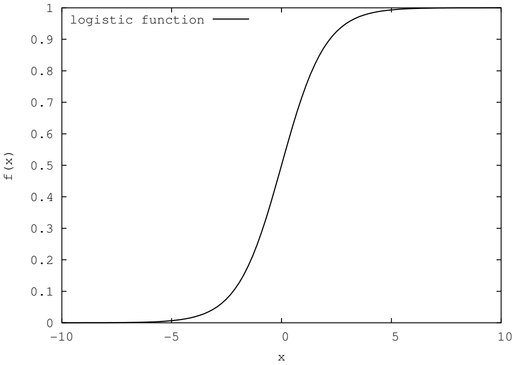
*Let tensor* $\xi$ *represent the noisy functional dependence generated by function* $f$. *Then rank of* $\xi$ *is less or equal to the rank of its deterministic core* $\psi_f$.

# Logistic function

$$\sigma(x) = \frac{1}{1 + e^{-(ax+b)}}$$

# Logistic function

$$\sigma(x) = \frac{1}{1 + e^{-(ax+b)}}$$



# Tables of soft threshold with low rank

Logistic function is well studied and broadly used, but it seems that tables generated by this function have a high rank.

**Definition**

Assume $\mathcal{X}_n = \{0, 1\}$. $\psi_\sigma(X_n \mid X_M)$ is table representing soft threshold if

$$\psi_\sigma(x_n \mid x_M) = \begin{cases} \sigma\left(\sum_{i=1}^{n-1} \alpha_i x_i + \beta\right) & \text{if } x_n = 1 \\ 1 - \sigma\left(\sum_{i=1}^{n-1} \alpha_i x_i + \beta\right) & \text{if } x_n = 0, \end{cases}$$

where $\sigma$ is a function with sigmoidal shape.

# Tables of soft threshold with low rank

Logistic function is well studied and broadly used, but it seems that tables generated by this function have a high rank.

**Definition**

Assume $\mathcal{X}_n = \{0,1\}$. $\psi_\sigma(X_n \mid X_M)$ is table representing soft threshold if

$$\psi_\sigma(x_n \mid x_M) = \begin{cases} \sigma\left(\sum_{i=1}^{n-1} \alpha_i x_i + \beta\right) & \text{if } x_n = 1 \\ 1 - \sigma\left(\sum_{i=1}^{n-1} \alpha_i x_i + \beta\right) & \text{if } x_n = 0, \end{cases}$$

where $\sigma$ is a function with sigmoidal shape.

# Cosine soft threshold function

Consider a family of functions parametrized by $\ell = 1, 2, \ldots$

$$\omega_\ell(t) \;=\; \sum_{j=1}^{\ell} c_j \cdot \cos(2j-1) \;,$$

where the coefficients $c_j, j = 1, \ldots, \ell$ are defined as the solution of the system of linear equations:

$$\omega_\ell(0) \;=\; 1 \qquad \frac{\partial^{2k}\omega_\ell(t)}{\partial t}|_{t=0} \;=\; 0 \text{ for } k = 1, 2, \ldots, \ell - 1,$$

which gives

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 3^2 & 5^2 & \ldots & (2\ell-1)^2 & 0 \\ \ldots & & & & & \\ 1 & 3^{2(\ell-1)} & 5^{2(\ell-1)} & \ldots & (2\ell-1)^{2(\ell-1)} & 0 \end{array}\right)$$

By the properties of Vandermonde determinant the system has a unique solution.

# Cosine soft threshold function

Consider a family of functions parametrized by $\ell = 1, 2, \ldots$

$$\omega_\ell(t) \;=\; \sum_{j=1}^{\ell} c_j \cdot \cos(2j-1) \;,$$

where the coefficients $c_j, j = 1, \ldots, \ell$ are defined as the solution of the system of linear equations:

$$\omega_\ell(0) \;=\; 1 \qquad \frac{\partial^{2k}\omega_\ell(t)}{\partial t}\Big|_{t=0} \;=\; 0 \text{ for } k = 1, 2, \ldots, \ell - 1,$$

which gives

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 3^2 & 5^2 & \ldots & (2\ell-1)^2 & 0 \\ \ldots & & & & & \\ 1 & 3^{2(\ell-1)} & 5^{2(\ell-1)} & \ldots & (2\ell-1)^{2(\ell-1)} & 0 \end{array}\right)$$

By the properties of Vandermonde determinant the system has a unique solution.

# Cosine soft threshold function

Consider a family of functions parametrized by $\ell = 1, 2, \ldots$

$$\omega_\ell(t) \;=\; \sum_{j=1}^{\ell} c_j \cdot \cos(2j-1) \;,$$

where the coefficients $c_j, j = 1, \ldots, \ell$ are defined as the solution of the system of linear equations:

$$\omega_\ell(0) \;=\; 1 \qquad \frac{\partial^{2k}\omega_\ell(t)}{\partial t}\Big|_{t=0} \;=\; 0 \text{ for } k = 1, 2, \ldots, \ell - 1,$$

which gives

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 3^2 & 5^2 & \ldots & (2\ell-1)^2 & 0 \\ \ldots & & & & & \\ 1 & 3^{2(\ell-1)} & 5^{2(\ell-1)} & \ldots & (2\ell-1)^{2(\ell-1)} & 0 \end{array}\right)$$

By the properties of Vandermonde determinant the system has a unique solution.

# Cosine soft threshold function

Consider a family of functions parametrized by $\ell = 1, 2, \ldots$

$$\omega_\ell(t) \;=\; \sum_{j=1}^{\ell} c_j \cdot \cos(2j-1)\;,$$

where the coefficients $c_j, j = 1, \ldots, \ell$ are defined as the solution of the system of linear equations:

$$\omega_\ell(0) \;=\; 1 \qquad \frac{\partial^{2k}\omega_\ell(t)}{\partial t}\Big|_{t=0} \;=\; 0 \text{ for } k = 1, 2, \ldots, \ell-1,$$

which gives

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 3^2 & 5^2 & \ldots & (2\ell-1)^2 & 0 \\ \ldots & & & & & \\ 1 & 3^{2(\ell-1)} & 5^{2(\ell-1)} & \ldots & (2\ell-1)^{2(\ell-1)} & 0 \end{array}\right)$$

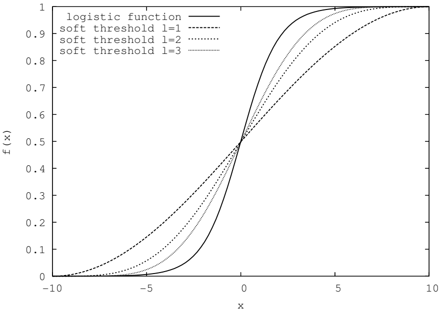
By the properties of Vandermonde determinant the system has a unique solution.

# Cosine soft threshold

Define $\sigma_\ell(t) = \frac{1}{2}(\omega_\ell(\pi\frac{t+30}{40}) + 1)$ for $\ell = 1, 2, 3$.

# Cosine soft threshold

Define $\sigma_\ell(t) = \frac{1}{2}(\omega_\ell(\pi \frac{t+30}{40}) + 1)$ for $\ell = 1, 2, 3$.



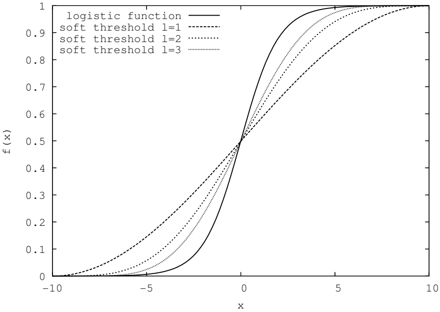

**Theorem**

*The rank of the tables representing the cosine soft threshold function $\sigma_\ell(t)$ is at most $2(\ell + 1)$ in complex numbers.*

Note that $\cos(t) = \frac{1}{2}(e^{it} + e^{-it})$.

# Cosine soft threshold

Define $\sigma_\ell(t) = \frac{1}{2}(\omega_\ell(\pi\frac{t+30}{40}) + 1)$ for $\ell = 1, 2, 3$.



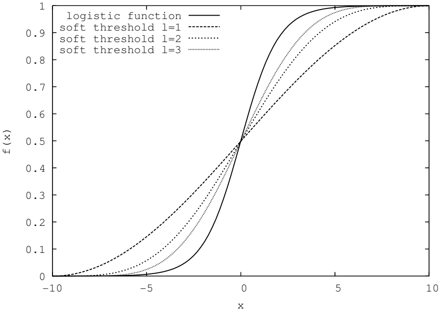

## Theorem

*The rank of the tables representing the cosine soft threshold function $\sigma_\ell(t)$ is at most $2(\ell + 1)$ in complex numbers.*

Note that $\cos(t) = \frac{1}{2}(e^{\mathrm{i}t} + e^{-\mathrm{i}t})$.

# Cosine soft threshold

Define $\sigma_\ell(t) = \frac{1}{2}(\omega_\ell(\pi \frac{t+30}{40}) + 1)$ for $\ell = 1, 2, 3$.




## Theorem

*The rank of the tables representing the cosine soft threshold function $\sigma_\ell(t)$ is at most $2(\ell + 1)$ in complex numbers.*

Note that $\cos(t) = \frac{1}{2}(e^{\mathrm{i}t} + e^{-\mathrm{i}t})$.