Maximum Likelihood Estimation for Toric Varieties

Centro de Matemática Universidade do Porto

CIMPA Summer School 2022

Mathematical Methods in Data Analysis

Eliana Duarte

CENTRO DE MATEMÁTICA UNIVERSIDADE DO PORTO

Lecture #1:

Toric varieties are exponential families + Examples

Lecture #2:

Fundamentals of MLE for toric varieties, an algebraic approach

Lecture #3:

Toric varieties with rational MLE, recent trends and open questions

Toric Varieties

Def 1.1: $c \in \mathbb{R}^r_{>0}$, $A = (a_{ij}) \in \mathbb{Z}^{k \times r}$ = integer matrix

$\mathbf{1} = (1, \dots, 1) \in \mathrm{Rowspan}(A)$

$$\begin{aligned} \phi^{A,c} : (\mathbb{C}^*)^k &\longrightarrow \mathbb{P}^{r-1} \\ \mathbf{t} &\longmapsto (c_1 \mathbf{t}^{a_1}, \dots, c_r \mathbf{t}^{a_r}) \end{aligned}$$

where $\mathbf{t} := (t_1, \dots, t_k)$, $a_j = (a_{1j} \ a_{2j} \ \cdots \ a_{kj})^T$

monomial exponent notation $\longrightarrow$ $c_j \mathbf{t}^{a_j} = c_j \prod_{i=1}^{r} t_i^{a_{ij}}$

$X_{A,c} := \overline{\operatorname{im} \phi^{A,c}}$ is the scaled projective toric variety associated to A, c.

$X_A := X_{A,\mathbf{1}}$

Defining ideals of toric varieties

$R := \mathbb{R}[p_1, \dots, p_r]$ is the coordinate ring of $\mathbb{P}^{r-1}$

Def 1.2: The ideal $I_{A,c} = I(X_{A,c}) \subseteq R$ is called the toric ideal associated to the pair (A, c). $I_A := I_{A,1}$

- Obtain gens of $I_{A,c}$ from gens of I_A, $p_j \mapsto p_j / c_j$

Prop 1.3: The toric ideal I_A is a binomial ideal and

$$I_A = \langle p^u - p^v : u, v \in \mathbb{N}^r \text{ and } Au = Av \rangle$$

If $\mathbf{1} \in \text{rowspan}(A) \Rightarrow I_A$ is homogeneous

Defining ideals of toric varieties

$R := \mathbb{R}[p_1, \ldots, p_r]$ is the coordinate ring of $\mathbb{P}^{r-1}$

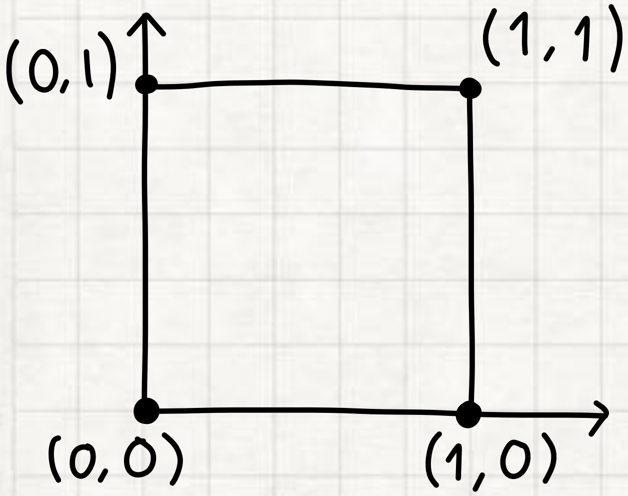
Def 1.2: The ideal $I_{A,c} = I(X_{A,c}) \subseteq R$ is called the toric ideal associated to the pair (A, c). $I_A := I_{A,1}$

- Obtain gens of $I_{A,c}$ from gens of I_A, $p_j \mapsto p_j / c_j$

Prop 1.3: The toric ideal I_A is a binomial ideal and
$$I_A = \langle p^u - p^v : u, v \in \mathbb{N}^r \text{ and } Au = Av \rangle$$
If $\mathbf{1} \in \text{rowspan}(A) \Rightarrow I_A$ is homogeneous

The beauty of toric varieties is the close relation between the discrete geometry of $P = \text{conv}(A)$ and the algebraic geometry of $X_{A,c}$

Example 1.4: $P = \mathrm{conv}(A) =$ unit square



$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Add row 1

$$c = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\phi_A : \mathbb{R}^3 \longrightarrow \mathbb{P}^3, \quad (t_1, t_2, z) \mapsto (\underset{p_1}{z}, \underset{p_2}{t_1 z}, \underset{p_3}{t_2 z}, \underset{p_4}{t_1 t_2 z})$$

$$u = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \qquad v = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$Au = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \qquad Av = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \qquad p^u - p^v = p_4 p_1 - p_2 p_3$$

$$X_A = \overline{\mathrm{im}(\phi_A)} = V(p_1 p_4 - p_2 p_3) \subseteq \mathbb{P}^3$$

Example 1.4:

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \qquad X_A = \overline{\mathrm{im}(\phi_A)} = V(p_1p_4 - p_2p_3) \subseteq \mathbb{P}^3$$

- The variety X_A only depends on the rowspan of A.

$$A' = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$ defines the same variety $P' = \mathrm{con}(A') \subseteq \mathbb{R}^4$ is equivalent to P.

$$(t_1, t_2, t_3, t_4) \mapsto (t_1t_3, t_1t_4, t_2t_3, t_2t_4)$$

AKA: Segre embedding $\mathbb{P}^1 \times \mathbb{P}^1 \longrightarrow \mathbb{P}^3$

Some Probability and Statistics

X = random variable, X discrete $\leadsto$ probability mass function
X continuous $\leadsto$ density function

Example 1.5:

(1) Suppose $\theta \in (0,1)$ $X \sim \text{Bernoulli}(\theta)$

$\mathcal{X}$ = state space $= \{0,1\}$ $\quad p_0 = P(X=0) = \theta$, $\quad p_1 = P(X=1) = 1-\theta$

$(p_0, p_1) \in \mathbb{R}^2$, $\quad p_0, p_1 > 0 \quad p_0 + p_1 = 1$

(2) X = univariate Gaussian mean $= \mu$, variance $= \sigma^2$

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Def 1.6:

- A statistical model $\mathcal{M}$ is a collection of p.d.fs or p.m.fs over a given space
- Θ = fin. dim parameter space $\subseteq \mathbb{R}^d$

$\mathcal{M}_\Theta$ = parametric statistical model $\subseteq$ space of pdfs/pmfs

$= $ image of a map $P_\cdot : \Theta \longrightarrow \mathcal{M}_\Theta$, $\theta \longmapsto P_\theta$

$= \{ P_\theta : \theta \in \Theta \}$

- identifiable: P_θ is uniquely defined by θ

Def 1.6:

- A statistical model $\mathcal{M}$ is a collection of p.d.fs or p.m.fs over a given space
- Θ = fin. dim parameter space $\subseteq \mathbb{R}^d$

$\mathcal{M}_\Theta$ = parametric statistical model $\subseteq$ space of pdfs/pmfs

$= \text{image of a map } p_\cdot : \Theta \to \mathcal{M}_\Theta , \ \theta \mapsto p_\theta$

$= \{ p_\theta : \theta \in \Theta \}$

- identifiable: p_θ is uniquely defined by θ

Example 1.5 (again)

(1) Bernoulli model: $\mathcal{M}_\Theta = \{ p_\theta = (\theta, 1-\theta) : \theta \in (0,1) \} \subseteq \mathbb{R}^2$

(2) Gaussian model:

$\mathcal{M}_\Theta = \{ f(x \mid \mu, \sigma^2) : (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \} \subseteq$ space of densities

The probability simplex is

$$\mathring{\Delta}_{r-1} := \{ (p_1, p_2, \dots, p_r) \in \mathbb{R}^r : p_i > 0 , \ p_1 + \cdots + p_r = 1 \}$$

<u>Def 1.7:</u> A discrete statistical model is a subset $\mathcal{M} \subseteq \mathring{\Delta}_{r-1}$.

Example 1.5 (continued)

(3) The independence model $\mathcal{M}_{X \perp\!\!\!\perp Y} = \{ p_\theta : \theta \in \mathring{\Delta}_{n-1} \times \mathring{\Delta}_{m-1} \}$

X = discrete r.v state space $\{1, \dots, n\}$

Y = " ______________ " $\{1, \dots, m\}$

$$p_\bullet : \mathring{\Delta}_{n-1} \times \mathring{\Delta}_{m-1} \longrightarrow \mathring{\Delta}_{mn-1}$$

$$(a_1, \dots, a_n) \times (b_1, \dots, b_m) \longmapsto (a_1 b_1, \dots, a_n b_m)$$

If $n = m = 2 \Rightarrow (a_1, a_2) \times (b_1, b_2) \to (a_1 b_1, a_1 b_2, a_2 b_1, a_2 b_2)$

$$\mathcal{M}_{X \perp\!\!\!\perp Y} = V(p_{11} p_{00} - p_{01} p_{10}) \cap \mathring{\Delta}_3$$

In algebraic statistics we identify $\Delta_{r-1} \sim \mathbb{P}^{r-1}_{\mathbb{C}}$ (Why? exercise)

Exponential Families

Def 1.8: A family of pdfs/pmfs is called an exponential family if it can be expressed as:

$$f(x) = h(x)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x) \right), \qquad \Theta = \text{parameter space}$$

- $h(x) \geq 0$
- $t_1(x), \dots, t_k(x)$ are real valued functs. of x (do not depend on θ)
- $c(\theta) \geq 0$
- $w_1(\theta), \dots, w_k(\theta)$ are real valued functs. (possibly depend on θ)

Exponential Families

Def 1.8: A family of pdfs/pmfs is called an exponential family if it can be expressed as:

$$f(x) = h(x)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x) \right), \qquad \Theta = \text{parameter space}$$

- $h(x) \geq 0$
- $t_1(x), \ldots, t_k(x)$ are real valued functs. of x (do not depend on θ)
- $c(\theta) \geq 0$
- $w_1(\theta), \ldots, w_k(\theta)$ are real valued functs. (possibly depend on θ)

Example 1.9

$\{\text{Binomial}(n, \theta) : \theta \in (0,1)\} =: B_n$

Perform Bernoulli experiment n times, θ = probability of success

$\mathcal{X} = \{0, 1, \ldots, n\}$, $x \in \mathcal{X}$

$$f(x|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{n}{x} (1-\theta)^n \left(\frac{\theta}{1-\theta}\right)^x$$

(Write on the board)

Θ = parameter space of the family

$$f(x|\theta) = h(x)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x) \right), \quad \theta \in \Theta$$

Consider

$$\tilde{\Theta} := \{ (\eta_1, \dots, \eta_k) \in \mathbb{R}^k : \ \eta_1 = w_1(\theta), \dots, \ \eta_k = w_k(\theta), \quad \theta \in \Theta \}$$
$$\subseteq \{ (\eta_1, \dots, \eta_k) : \ \eta_i \in \mathbb{R} \} \subseteq \mathbb{R}^k$$

Extend the parameter space and consider:

$$f(x|\eta) = h(x) \exp\left(\sum_{i=1}^{k} \eta_i\, t_i(x) \right) \cdot \left(\frac{1}{Z(\eta)} \right)$$

$h(x)$, $t_i(x)$ as before.

$$Z(\eta) = 1 \Big/ \int_{-\infty}^{\infty} h(x) \exp\left(\sum_{i=1}^{k} \eta_i\, t_i(x) \right) dx =: \text{log- partition function}$$

Natural Parameter space $= \mathcal{H} := \{ (\eta_1, \dots, \eta_k) : \ \int_{-\infty}^{\infty} f(x|\eta)\, dx < \infty \}$

k-dimensional family if $\dim \mathcal{H} = k$

Discrete Regular Exponential families

$$f(x|\eta) = h(x) \frac{1}{Z(\eta)} e^{\eta^t T(x)} \qquad \eta^t = (\eta_1, \dots, \eta_K) \; , \; T(x) = \begin{pmatrix} t_1(x) \\ \vdots \\ t_K(x) \end{pmatrix}$$

$\mathcal{X} = [r] := \{1, \dots, r\} \ni x$

$T : \mathcal{X} \longrightarrow \mathbb{R}^K \quad \Rightarrow \quad T(x)$ is a vector

$h : \mathcal{X} \longrightarrow \mathbb{R} \quad \Rightarrow \quad (h(1), h(2), \dots, h(r))$ is a vector

For $\eta \in \mathbb{R}^K$, the normalizing constant is a sum

$$Z(\eta) = \sum_{x \in [r]} h(x) e^{\eta^t T(x)}$$

Let $Z(\eta) = e^{\phi(\eta)}$

For $x \in [r]$ $\quad p_\eta(x) = h(x) e^{\eta^t T(x) - \phi(\eta)}$

For $x \in [r]$ $\quad p_\eta(x) = h(x)\, e^{\eta^t T(x) - \phi(\eta)}$

$$(\eta_1, \dots, \eta_k) \cdot \underbrace{\begin{pmatrix} a_{1x} \\ a_{2x} \\ \vdots \\ a_{kx} \end{pmatrix}}_{a_x := T(x)} \qquad x \in [r]$$

Different θ from before

$$e^{\sum_{i=1}^{n} \eta_i a_{ix}} = e^{\eta_1 a_{1x}} e^{\eta_2 a_{2x}} \cdots e^{\eta_k a_{kx}}, \qquad \theta_i = \exp(\eta_i)$$

$$p_\eta(x) = \frac{h(x)\, e^{\eta_1 a_{1x}} \cdots e^{\eta_k a_{kx}}}{Z(\theta)}$$

$$= \frac{h_x\, \theta_1^{a_{1x}} \cdots \theta_k^{a_{kx}}}{Z(\theta)} = \frac{h_x\, \boldsymbol{\theta}^{a_x}}{Z(\theta)}, \qquad \boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$$

For $x \in [r]$ $\quad P_\eta(x) = h(x)\, e^{\eta^t T(x) - \phi(\eta)}$

$$(\eta_1, \dots, \eta_k) \cdot \underbrace{\begin{pmatrix} a_{1x} \\ a_{2x} \\ \vdots \\ a_{kx} \end{pmatrix}}_{a_x := T(x)} \qquad x \in [r]$$

Different θ from before

$$e^{\sum_{i=1}^{n} \eta_i a_{ix}} = e^{\eta_1 a_{1x}} e^{\eta_2 a_{2x}} \cdots e^{\eta_k a_{kx}} , \qquad \theta_i = \exp(\eta_i)$$

$$P_\eta(x) = \frac{h(x)\, e^{\eta_1 a_{1x}} \cdots e^{\eta_k a_{kx}}}{Z(\theta)}$$

$$= \frac{h_x\, \theta_1^{a_{1x}} \cdots \theta_k^{a_{kx}}}{Z(\theta)} = \frac{h_x\, \boldsymbol{\theta}^{a_x}}{Z(\theta)} , \qquad \boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$$

if a_{jx} are integers $\Rightarrow$

$$c = (h_1, \dots, h_r)$$

THIS IS A TORIC VARIETY

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ | & | & & | \end{pmatrix}$$

CIMPA Summer School 2022

MATHEMATICAL METHODS IN DATA ANALYSIS

dates: July 18–July 29, 2022

BREAK

Examples

(1) Log-linear models

(2) Undirected graphical models

(3) Bayesian Networks / DAG models

(4) Staged tree models

All these models are algebraic varieties and are subsets of a probability simplex

Log-linear models

Def 1.1: $c \in \mathbb{R}^r_{>0}$, $A = (a_{ij}) \in \mathbb{Z}^{k \times r}$ = integer matrix

$\mathbf{1} = (1, \dots, 1) \in \text{Rowspan}(A)$

$$\begin{aligned} \phi^{A,c} : (\mathbb{C}^*)^k &\longrightarrow \mathbb{P}^{r-1} \\ \mathbf{t} &\longmapsto (c_1 \mathbf{t}^{a_1}, \dots, c_r \mathbf{t}^{a_r}) \end{aligned}$$

$X_{A,c} := \overline{\operatorname{im} \phi^{A,c}}$ is the scaled projective toric variety associated to A, c.

Log-linear models

Def 1.1: $c \in \mathbb{R}^r_{>0}$, $A = (a_{ij}) \in \mathbb{Z}^{k\times r}$ = integer matrix

$\mathbf{1} = (1, \dots, 1) \in \mathrm{Rowspan}(A)$

$$\begin{aligned} \phi^{A,c} : (\mathbb{C}^*)^k &\longrightarrow \mathbb{P}^{r-1} \\ \mathbf{t} &\longmapsto (c_1 \mathbf{t}^{a_1}, \dots, c_r \mathbf{t}^{a_r}) \end{aligned}$$

$X_{A,c} := \overline{\mathrm{im}\, \phi^{A,c}}$ is the scaled projective toric variety associated to A, c.

Let $W = V(x_1 \cdots x_r (x_1 + \cdots + x_r)) \subseteq \mathbb{P}^{r-1}$ define the map

$$\begin{aligned} X_{A,c} \setminus W &\to (\mathbb{C}^*)^r \\ (x_1 : \cdots : x_r) &\mapsto \frac{1}{x_1 + \cdots + x_r} (x_1, \dots, x_s) \end{aligned}$$

Its image is closed and denoted by $Y_{A,c}$

$\mathcal{M}_{A,c} = Y_{A,c} \cap \mathbb{R}^r_{>0} \subseteq \Delta^{\circ}_{r-1}$ ← Log-linear model

Undirected Graphical Models

$G=(V,E)$ an undirected graph

$V=[p] \sim$ vector of discrete random variables

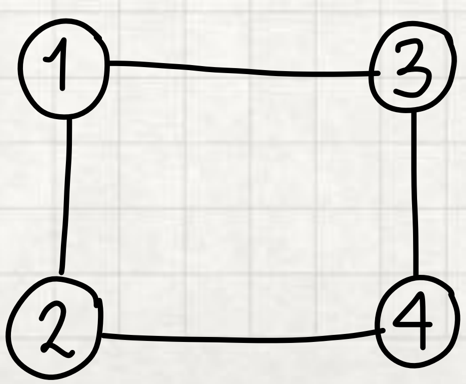
$\sim (X_1, X_2, \dots, X_p) = X_{[p]}$ $\quad [d_i] :=$ state space of X_i, $i \in [p]$

$\mathcal{X} = \{ \mathbf{x} = x_1 \cdots x_p \in \mathcal{R} = \prod_{i=1}^{p} [d_i] \}$

- A clique is a set $C \subseteq V$ such that $(i,j) \in E$ for all $i,j \in C$
- $\mathcal{C}(G) = \{$ all maximal cliques $\}$
- $B \subseteq V \quad X_B = (X_i)_{i \in B}$
- For each $C \in \mathcal{C}(G)$ introduce a potential function $\phi_C(x_C) \geq 0$ on $\mathcal{X}_C$, the state space of X_C

Example 1.10

$X_{[4]} = (X_1, X_2, X_3, X_4)$ vector of binary r.v.



$\mathcal{C}(G) = \{ \{1,2\}, \{1,3\}, \{3,4\}, \{2,4\} \}$, $C = \{1,2\}$

$\phi_C : \{0,1\}^2 \longrightarrow \mathbb{R}_{\geq 0} \qquad \phi_C(00), \phi_C(01), \phi_C(10), \phi_C(11)$

Def 1.11: The parametrized undirected graphical model $\mathcal{M}(G)$ consists of all pdfs on $\mathcal{X}$ of the form

$$f(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)$$

for some potential functions $\phi_C(x_C)$, where $Z = \int_{\mathcal{X}} \prod_{C \in \mathcal{C}(G)} \phi_C(x_C)\, d\mu(x)$

parameter space = tuples of potential functions s.t $Z < \infty$

f factorizes according to G if it can be written as above for some potential functions

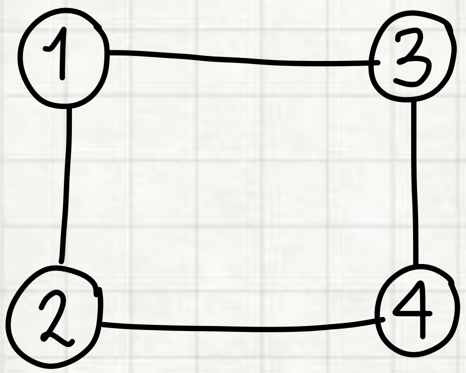
$$\mathcal{M}(G) \subseteq \Delta^{0}_{|\mathcal{R}|-1} = \{ (f_{\mathbf{x}})_{\mathbf{x} \in \mathcal{R}} : f_{\mathbf{x}} \in (0,1) \ \forall \mathbf{x} \in \mathcal{R} \}$$

$$\mathbf{x} = x_1 \cdots x_p \in \mathcal{R} = \prod_{i=1}^{p} [d_i]$$

THIS IS A MONOMIAL PARAMETRIZATION $\Rightarrow$ TORIC VARIETY

Example 1.10

$X_{[4]} = (X_1, X_2, X_3, X_4)$ vector of binary r.v.



$$\mathcal{C}(G) = \{ \underbrace{\{1,2\}}_{C_1}, \underbrace{\{1,3\}}_{C_2}, \underbrace{\{3,4\}}_{C_3}, \underbrace{\{2,4\}}_{C_4} \},$$

$$\mathring{\Delta}_{|\mathcal{R}|-1} = \mathring{\Delta}_{2^4-1} = \mathring{\Delta}_{15} \supseteq \mathcal{M}(G)$$

Parametric Description

Clique factorization $\quad \mathbf{x} = x_1 x_2 x_3 x_4 \in \mathcal{R}$

$$f(\mathbf{x}) = f(x_1 x_2 x_3 x_4) = \phi_{C_1}(\mathbf{x}_{C_1})\, \phi_{C_2}(\mathbf{x}_{C_2})\, \phi_{C_3}(\mathbf{x}_{C_3})\, \phi(\mathbf{x}_{C_4})$$

$$= \phi_{C_1}(x_1 x_2)\, \phi_{C_2}(x_1 x_3)\, \phi_{C_3}(x_3 x_4)\, \phi(x_2 x_4)$$

For each clique C_i, treat $\phi_{C_i}(00), \phi_{C_i}(01), \phi_{C_i}(10), \phi_{C_i}(11)$, as indeterminates

For each $\mathbf{x} \in \mathcal{R}$ $f(\mathbf{x})$ is a monomial in the symbols $\phi_{C_i}(\mathbf{x}_{C_i})$ $\quad i = 1, 2, 3, 4.$

Exercise: Write the exponent matrix A for this model.

Implicit Description of $\mathcal{M}(G)$

$$\mathbb{R}[p_{0000}, \ldots, p_{1111}] \xrightarrow{\psi_G} \mathbb{R}\left[\theta^{C_i}_{x_{C_i}} : \begin{array}{l} C_i \in \mathcal{C}(G) \\ x_{C_i} \in \mathcal{R}_{C_i} \end{array}\right]$$

$$p_{0000} \mapsto \theta^{C_1}_{00}\,\theta^{C_2}_{00}\,\theta^{C_3}_{00}\,\theta^{C_4}_{00}$$

$$\vdots$$

$$p_{0101} \mapsto \theta^{C_1}_{01}\,\theta^{C_2}_{00}\,\theta^{C_3}_{01}\,\theta^{C_4}_{11}$$

$$\vdots$$

$\mathrm{Ker}(\psi_G)$ is a toric ideal that defines $\mathcal{M}(G)$ implicitly

To write A, find the parametrization, then use M2 to find $\mathrm{Ker}(\psi_G)$.

Directed Graphical Models, Bayesian Networks

$G=(V,E)$ directed acyclic graph $E=$ directed edges

$V=[p] \sim$ vector of discrete random variables

$\sim (X_1, X_2, \dots, X_p) = X_{[p]}$

$\mathbf{x} \in \mathcal{R}$, $B \subseteq [p]$ $\quad f(X_K \mid \mathbf{x}_B) :=$ Conditional distribution of X_K given $\mathbf{x}_B$.

$i \in [p]$ $\quad pa_G(i) = \{ j : j \to i \in E \}$

A distribution $f \in \Delta^{o}_{|\mathcal{R}|-1}$ factors according to G if

$$f(\mathbf{x}) = \prod_{i=1}^{p} f(x_i \mid \mathbf{x}_{pa_G(i)}), \text{ for all } \mathbf{x} = x_1 x_2 \cdots x_p \in \mathcal{R}$$

$$f(X_1, \dots, X_p) = \prod_{i=1}^{p} f(X_i \mid X_{pa_G(i)})$$

The DAG model $\mathcal{M}(G) \subseteq \Delta^{\circ}_{|\mathcal{R}|-1}$ consists of all $f \in \Delta^{\circ}_{|\mathcal{R}|-1}$ that factor according to G.

Example 1.11: $X_{[4]} = (X_1, X_2, X_3, X_4)$ vector of binary r.v.

$$\begin{tikzcd} 1 \arrow[r] \arrow[d] & 2 \arrow[d] \\ 3 \arrow[r] & 4 \end{tikzcd}$$

$$f(X_{[4]}) = f(X_1)f(X_2|X_1)f(X_3|X_1)f(X_4|X_2,X_3)$$

$\mathbf{x} \in \mathcal{R}$ $\quad \mathbf{x} = x_1 x_2 x_3 x_4$

$$f(\mathbf{x}) = f(X_1 = x_1)\, f(X_2 = x_2 | X_1 = x_1)\, f(X_3 = x_3 | X_2 = x_2)\, f(X_4 = x_4 | X_{23} = x_2 x_3)$$

↳ Recursive factorization

$f(X_1=0)=s_0 \quad f(X_1=1)=s_1 \quad , \quad s_0+s_1=1$

$f(X_2=0\,|\,X_1=0)=s_2 \quad f(X_2=1\,|\,X_1=0)=s_3 \quad , \quad s_2+s_3=1$

$f(X_2=0\,|\,X_1=1)=s_4 \quad f(X_2=1\,|\,X_1=1)=s_5 \,, \quad s_4+s_5=1$

$f(X_3=0\,|\,X_1=0)=s_6 \quad f(X_3=1\,|\,X_1=0)=s_7 \,, \quad s_6+s_7=1$

$f(X_3=0\,|\,X_1=1)=s_8 \quad f(X_3=1\,|\,X_1=1)=s_9 \,, \quad s_8+s_9=1$

$f(X_4\,|\,X_{2,3}=00) \quad f(X_4\,|\,X_{2,3}=01) \quad f(X_4\,|\,X_{2,3}=10) \quad f(X_4\,|\,X_{2,3}=11)$

$s_9+s_{10}=1 \qquad s_{11}+s_{12}=1 \qquad s_{13}+s_{14}=1 \qquad s_{15}+s_{16}=1$

$f(0101)=s_0 s_3 s_6 s_{14}$

Parametric Description

9 simplices

$$\Delta_1^\circ \times \cdots \times \Delta_1^\circ \longrightarrow \Delta_{16}^\circ$$

$$(s_0,\dots,s_{16}) \longrightarrow \left(f(\mathbf{x})=\prod_{i=1}^{4} f(x_i\,|\,\mathbf{x}_{pa_G(i)})\right)_{\mathbf{x}\in\mathcal{R}}$$

$\mathcal{M}(G)=$ 9-dimensional model in Δ_{16}°

Implicit Description of $\mathcal{M}(G)$

$$\mathbb{R}[p_{0000}, \ldots, p_{1111}] \xrightarrow{\Phi_G} \mathbb{R}[s_0, s_1, \ldots, s_{16}] / \text{sum-to-one ideal}$$

$$p_{0000} \mapsto s_0 s_2 s_6 s_9$$

$$\vdots$$

$$p_{0101} \mapsto s_0 s_3 s_6 s_{14}$$

$\mathrm{Ker}(\Phi_G)$ is a prime ideal, not toric, that difenes $\mathcal{M}(G)$ implicitly

Exercise: Compare $\mathrm{Ker}(\Phi_G)$, $\mathrm{Ker}(\psi_G)$.

WARNING: DAG MODELS ARE NOT ALWAYS TORIC

In general undirected models $\neq$ directed models
$G =$ DAG

$\mathcal{M}(G)$ is a decomposable model if G is a chordal graph (i.e every induced cycle of length 4 has a chord)
$\widetilde{G} =$ undirected skeleton of G, $i \to j \mapsto i - j$

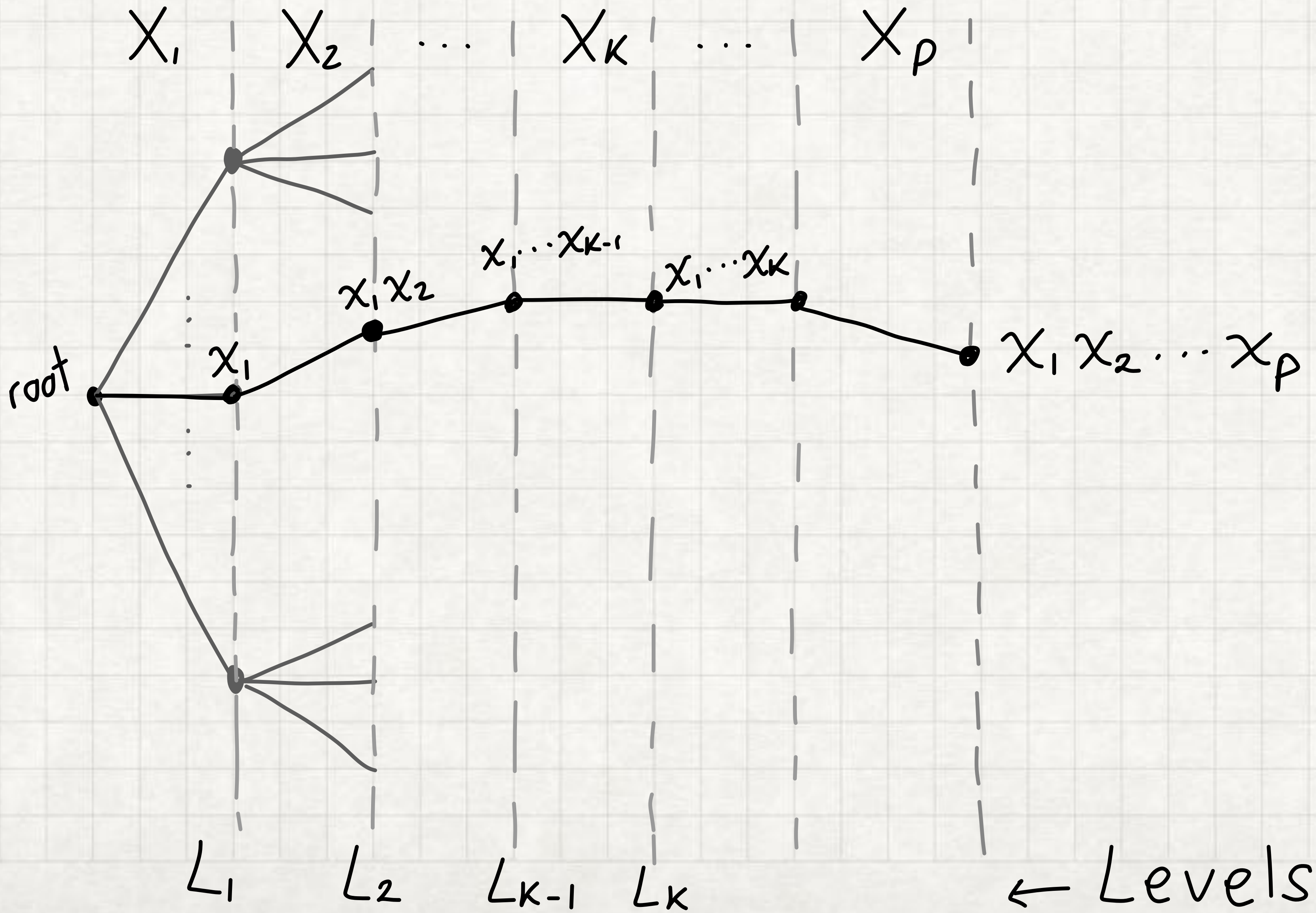
<u>Theorem:</u> (Geiger, Meek, Sturmfels 2006) T.F.A.E
(1) G is chordal
(2) $\mathcal{M}(G) = \mathcal{M}(\widetilde{G})$
in this case $\mathcal{M}(G)$ is a toric variety i.e it is defined by binomials.

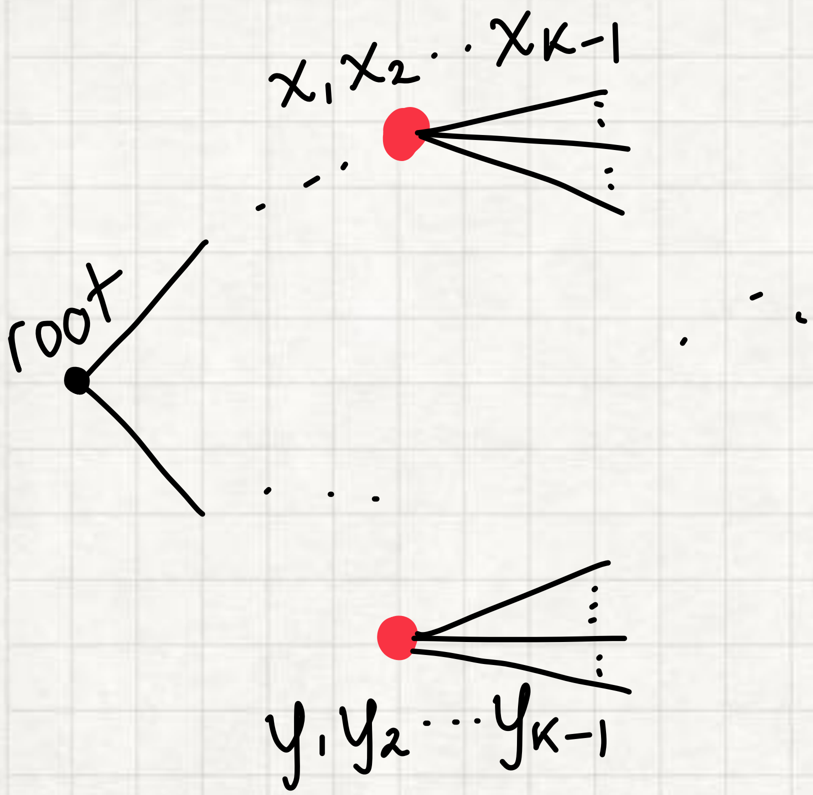
Staged Trees

$X_{[p]} = (X_1, \dots, X_p)$

$\mathbf{x} \in \mathcal{R} \qquad f(\mathbf{x}) = f(x_1, \dots, x_p) = f(x_1) f(x_2 | x_1) \cdots f(x_p | x_1 \cdots x_{p-1})$

$\mathcal{T}$ = tree representation of $\mathcal{R}$





In a staged tree $(\mathcal{T}, \theta)$ you color vertices on the same level to represent equality of conditional distributions

$$x_1 x_2 \cdots x_{k-1} \sim y_1 \cdots y_{k-1}$$

$$\Longleftrightarrow$$

$$f(X_k \mid x_1 \cdots x_{k-1}) = f(X_k \mid y_1 \cdots y_{k-1})$$

Vertices with the same color are in the same stage

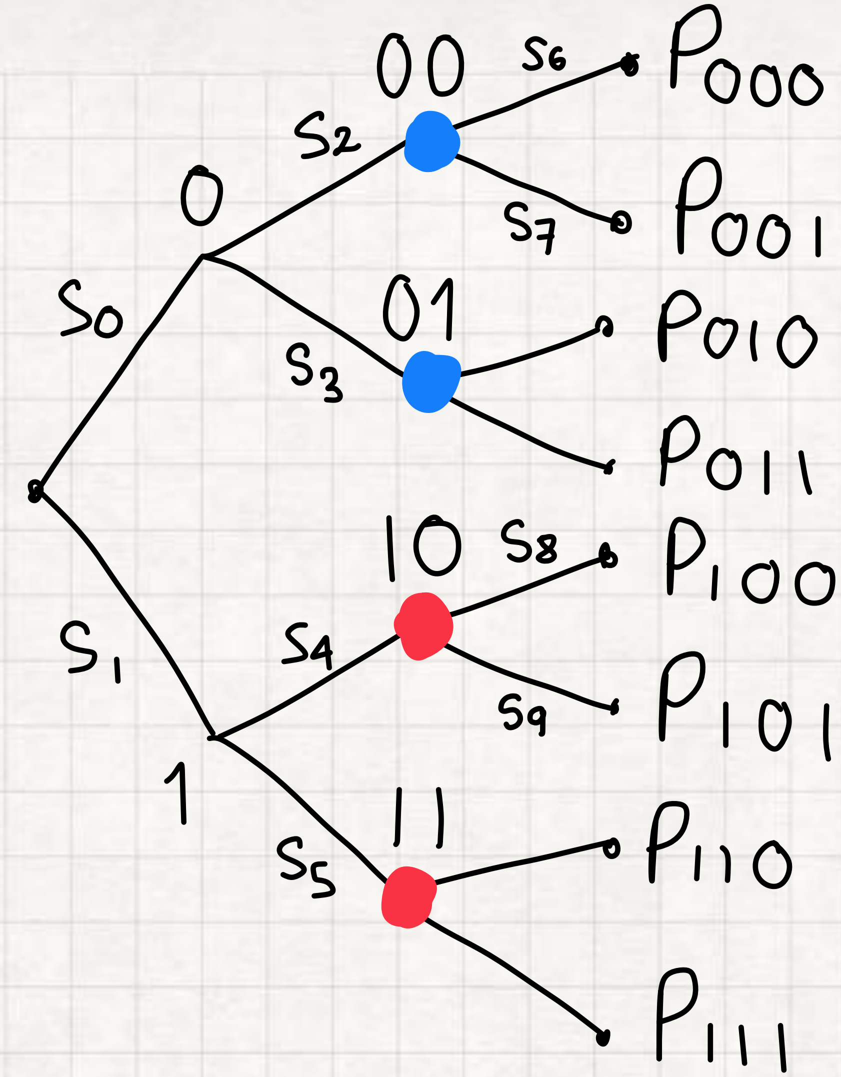
$$\mathcal{M}(\mathcal{T}) = \{ f \in \Delta^{\circ}_{|\mathcal{R}|-1} : f \text{ factors according to } \mathcal{T} \}$$

Every discrete DAG model is a staged tree model.

Example 1.12:

$X_{[3]} = (X_1, X_2, X_3)$

- $f(X_3 | 00) = f(X_3 | 01)$
- $f(X_3 | 10) = f(X_3 | 11)$



$s_0 + s_1 = 1$, $s_2 + s_3 = 1$, $s_4 + s_5 = 1$

$s_6 + s_7 = 1$, $s_8 + s_9 = 1$

$$\mathring{\Delta}_1 \times \cdots \times \mathring{\Delta}_1 \longrightarrow \mathring{\Delta}_7$$

Parametric Description

$$(s_0, s_1, \ldots, s_8, s_9) \longrightarrow (s_0 s_2 s_6, s_0 s_2 s_7, s_0 s_3 s_6, s_0 s_3 s_7, s_1 s_4 s_8, s_1 s_4 s_9, s_1 s_5 s_8, s_1 s_5 s_9)$$

$$\mathcal{M}(\tau) = V(p_{000} p_{011} - p_{010} p_{001}, p_{100} p_{111} - p_{110} p_{100}) \cap \mathring{\Delta}_7$$

Implicit Description

$$\mathbb{R}[p_{000}, \dots, p_{111}] \xrightarrow{\psi_\tau} \mathbb{R}[s_0, s_1, \dots, s_9] \Big/ \text{sum-to-one conditions}$$

$$\begin{aligned}
p_{000} &\mapsto s_0 s_2 s_6 \\
p_{001} &\mapsto s_0 s_2 s_7 \\
p_{010} &\mapsto s_0 s_3 s_6 \\
p_{011} &\mapsto s_0 s_3 s_7 \\
p_{100} &\mapsto s_1 s_4 s_8 \\
p_{101} &\mapsto s_1 s_4 s_9 \\
p_{110} &\mapsto s_1 s_5 s_8 \\
p_{111} &\mapsto s_1 s_5 s_9
\end{aligned}$$

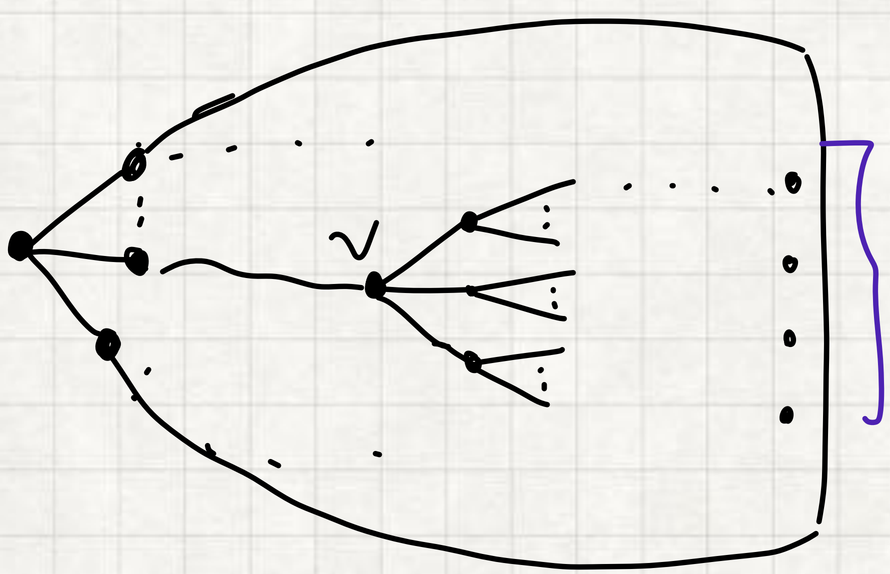
$$\operatorname{Ker}(\psi_\tau) = \langle p_{000} p_{011} - p_{010} p_{001},\ p_{100} p_{111} - p_{110} p_{100} \rangle$$

$(\mathcal{T}, \theta)$ = a staged tree $\qquad \mathcal{T} = (V, E)$

$\theta(e)$ is the label on edge e

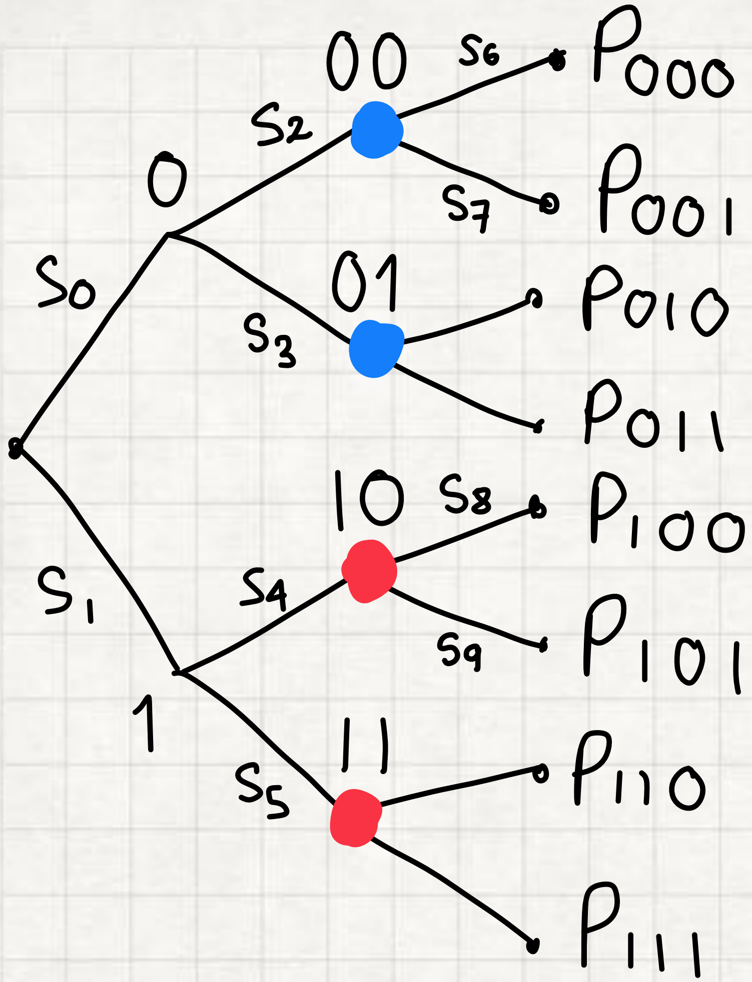
for $v \in V$, v in level $k-1$

$$t(v) := \sum_{\mathbf{y} \in \mathcal{R}_{[p]\setminus[k-1]}} \left(\prod_{e \in E(v \to v\mathbf{z})} \theta(e) \right) \in \mathbb{R}[\theta(e) : e \in E]$$



sum over v-to-leaf paths of the product of the edge labels in each path

$t(v) := 1$ if v is a leaf



$$t(0) = s_2 s_6 + s_2 s_7 + s_3 s_6 + s_3 s_7$$

$$t(00) = s_6 + s_7$$

$$t(000) := 1$$

A pair of vertices $v = x_1 \cdots x_{k-1}$, $w = y_1 \cdots y_{k-1}$ in the same stage is balanced if for all $i, j \in [d_k]$

$$t(vi)t(wj) = t(wi)t(vj) \quad \text{in } \mathbb{R}[\theta(e) : e \in E]$$

$\mathcal{T}$ is balanced if all vertices in the same stage are balanced

Theorem: (D, Görgen 2020)

$(\mathcal{T},\theta)$ is balanced $\Leftrightarrow$ $\mathcal{M}(\mathcal{T})$ is defined by binomials (i.e. toric)

Theorem: (D, Solus 2021) TFAE

(1) G is chordal

(2) $\mathcal{T}_G$ is balanced

(3) $\mathcal{M}(G)$ is defined by binomials

$\mathcal{T}_G$ = staged tree representation of G.