Recall: the 0-1 case.

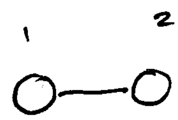
- turned into linear program.
- guaranteed that ~~ver~~ solution would hit integer verts only in special case ("cheaper to agree than disagree")

Q: could we build a polytope that
a) had only integer verts
b) contained all possible sol'n's?

A: (seems like NO ~~but~~ – no free lunch – but likely worth discussion).

Extend: to multilabel case.

eg.



0-1 model.

x_1 , x_2 , $x_1 \times x_2$

equiv to

$$p(x_1, x_2) \propto \exp\left[f_1(x_1) + f_2(x_2) + g(x_1, x_2) \right]$$

where $x_i \in \{0, 1\}$

$$p(x_1, x_2) \propto \exp\left[\begin{array}{l} f_1(1).x_1 + (1-x_1)\, f_1(0) \\ + f_2(1)\, x_2 + (1-x_2)\, f_2(0) \\ + x_1 x_2\, g(1,1) + (1-x_1)\, x_2\, g(0,1) \\ + (1-x_1)(1-x_2)\, g(0,0) \\ + x_1 (1-x_2)\, g(1,0) \end{array} \right]$$

Notice the log prob is a fn of $x_1, x_2, x_1 x_2$

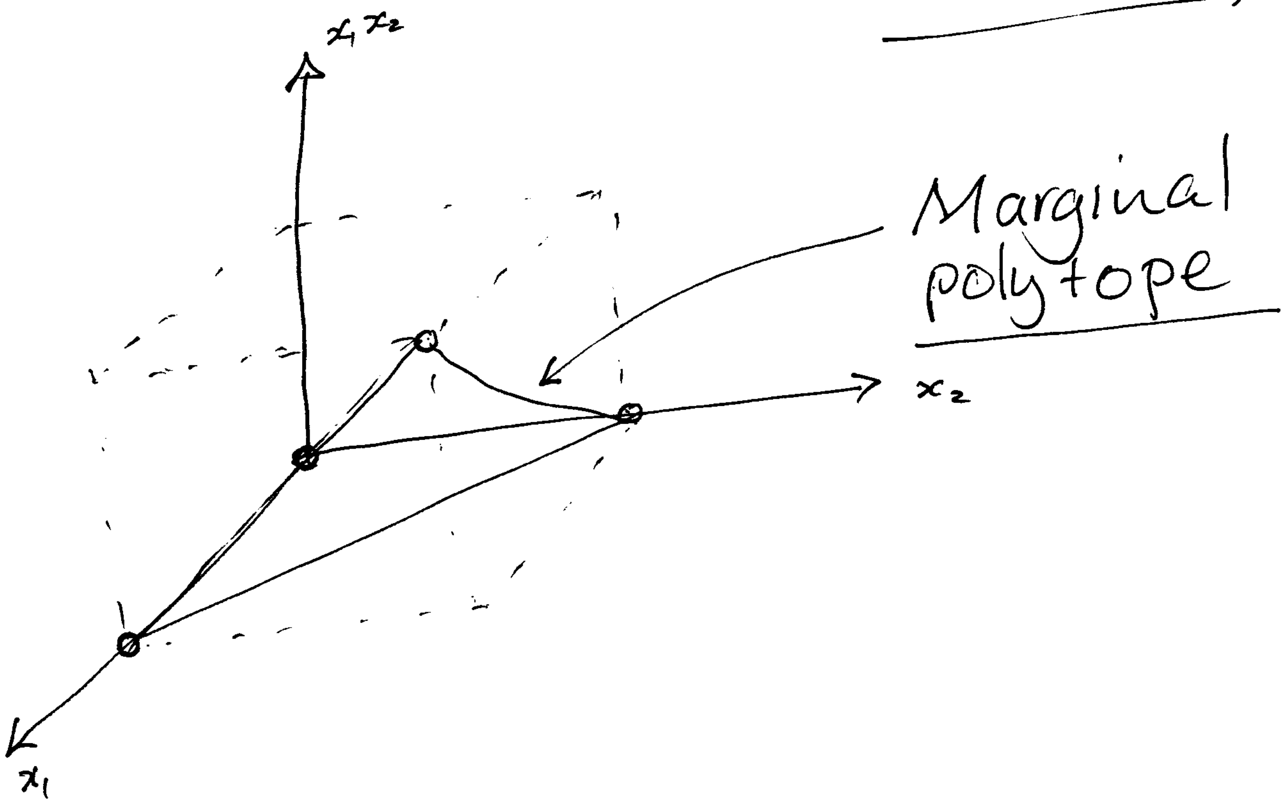
so

$$a_1 x_1 + a_2 x_2 + a_3 \nu_{12}$$

where $\nu_{12} = x_1 x_2$ ← remember our linear constraints

So I can encode with a 3d vector

$$\underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}}_{\text{states of model}}$$

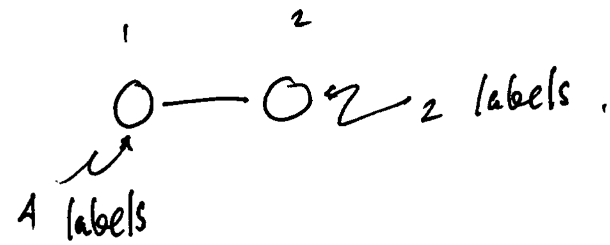


for more vars, this could get exciting

3 vars, 0,1	—	6 D
4	—	10 D
N ..	—	$N + \frac{N(N-1)}{2}$

If there are more labels, it matters how we represent the multi label problem.

eg.



Natural repn for unary: terms

$$\left[\underbrace{\text{a 1 in here}}_{4} \;\middle|\; \underbrace{\text{a 1 in here}}_{2}\right]$$

≡ "one hot vector"

binary term is a table

~~4~~ 2

4 a 1 in here

and this must be consistent with the unary term.

i.e. if x_1 repn = 1 in third place

x_2 repn = 1 in 2 "

then binary repn has 1 at (4, 2)

write $\mu_1 \quad \cdots \quad \mu_N$ for vector repn $x_1 \cdots x_N$

then $\mathbf{1}^T \mu_i = 1$, $\mu_i \in (0, 1)$

write ν_{ij} for table repn $x_i x_j$ term

then $\sum_j \nu_{ij} = \mu_i$; $\sum_i \nu_{ij} = \mu_j$

$\nu_{ij} \in \{0, 1\}$

Problem: how do we ~~mak~~ define marginal polytope?

think of μ_i, ν_{ij} as functions of

$\underline{x}_1 \cdots \underline{x}_N$

underlying labels

eg x_1 has $\{a, b\}$ — then $\mu_1(x_1) = \begin{cases} \begin{pmatrix}1\\0\end{pmatrix} & x_1 = a \\ \begin{pmatrix}0\\1\end{pmatrix} & x_1 = b \end{cases}$

etc.

Marginal polytope

$$= \text{Set of all } E_p\begin{bmatrix}\mu_1\\ \vdots \\ \mu_N \\ \nu_{ij} \\ \vdots \\ \nu\end{bmatrix} \text{ for any } p$$

1) its convex

2) its verts are integer

3) its verts meet consistency constraints

The marginal polytope for a tree structured model

- for each var x_i, construct $\underline{\mu}_i$
  so that $\underline{\mu}_i = \begin{bmatrix}\text{Vector of expected values of 1-hot}\end{bmatrix}$

  so: $\underline{\mu}_i \geq 0$; $1^T \underline{\mu}_i = 1$.

- for each pair x_i, x_j **ON AN EDGE**
  construct $\underline{\mu}_{ij} = \begin{bmatrix}\text{Table of expected values of 1-hot table } \nu_{ij}\end{bmatrix}$

  so $\mu_{ij} \geq 0$; $1^T \mu_{ij} 1 = 0$; $1^T \mu_{ij} = \mu_j^T$

  $\mu_{ij} 1 = \mu_i$

(last two impose consistency).

Now overload notation slightly

$\underline{\mu}_i(\underline{x}_i) \equiv$ element of $\underline{\mu}_i$ identified by 1-hot vector $\underline{x}_i$

then:

$$p(x_1 \cdot \cdot x_N) = \prod_{i \in V} \mu_i(x_i) \cdot \prod_{(i,j) \in E} \frac{\mu_{ij}(x_i, x_j)}{\mu_i(x_i) \cdot \mu_j(x_j)}$$

is a probability distribution that has marginals μ_i, μ_{ij}

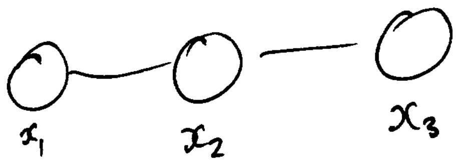
$$\Rightarrow \mathcal{M}(T) \equiv \{\mu_i, \mu_{ij} \text{ constructed in this way}\}$$

marginal polytope

tree

This means we know the ineq's that define $\mathcal{M}$, and there aren't that many proof by 2 examples, induction

eg 1



$$p(x_1, x_2, x_3) = \mu_1(x_1)\,\mu_2(x_2)\,\mu_3(x_3)\,\frac{\mu_{12}(x_1,x_2)}{\mu_1(x_1)\,\mu_2(x_2)}\,\frac{\mu_{23}(x_2,x_3)}{\mu_2(x_2)\,\mu_3(x_3)}$$

$$= \frac{\mu_{12}\cdot\mu_{23}}{\mu_2}$$

$$p(x_1) = \sum_{x_2} \frac{\mu_{12}}{\mu_2}\left[\sum_{x_3} \mu_{23}\right]$$

$= \mu_2$, constraints

$$= \sum_{x_2} \mu_{12}$$

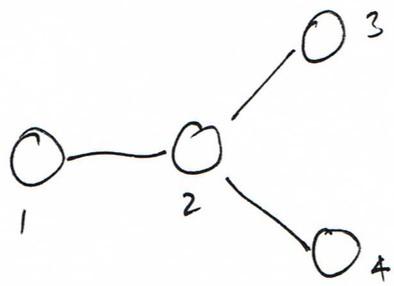
$$= \mu_1$$

$\rightarrow$ constraints

Same reasoning gives $p(x_1, x_2) = \mu_{12}$

[Q: why are these expectations?
A: x is 1-hot!]

eg 2



$$p(x_1\ x_2\ x_3\ x_4) = \cancel{\ } \frac{\mu_{12}\,\mu_{23}\,\mu_{24}}{\mu_2^2}$$

$$p(x_1) = \sum_{x_2} \frac{\mu_{12}}{\mu_2^2} \underbrace{\left[\sum_{x_3} \mu_{23}\right]}\underbrace{\left[\sum_{x_4} \mu_{24}\right]}$$

each $= \mu_2$, constra

$$= \sum_{x_2} \mu_{12}$$

$$= \mu_1$$ constraints

Same reasoning yields

- $p(x_1, x_2) = \cancel{\ }\ \mu_{12}$, etc

+ induction
  - OK for trees

Where we're going:

- all methods recover information from some form of approximating dist
- mean field methods find this dist from an approximation to INTERIOR of MP
- bp + variants find this dist from an approx that is mostly bigger than MP