KTH VETENSKAP OCH KONST

Royal Institute of Technology

DGM MLE AND POSTERIOR

ORDER OF PRODUCTS

$$p(x \mid \theta) = \prod_{v=1}^{V} p(x_v \mid x_{\mathrm{pa}(v)}, \theta_v)$$

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v) = \prod_{v=1}^{V} p(\mathscr{D}_v \mid \mathscr{D}_{\mathrm{pa}(v)}, \theta_v)$$

CATEGORICAL – NOTATION

★ For a v ∈ [V],

values $k \in S_v$

combined values $c \in C_v = \prod_{s \in \text{pa}(v)} S_s$

Cartesian product

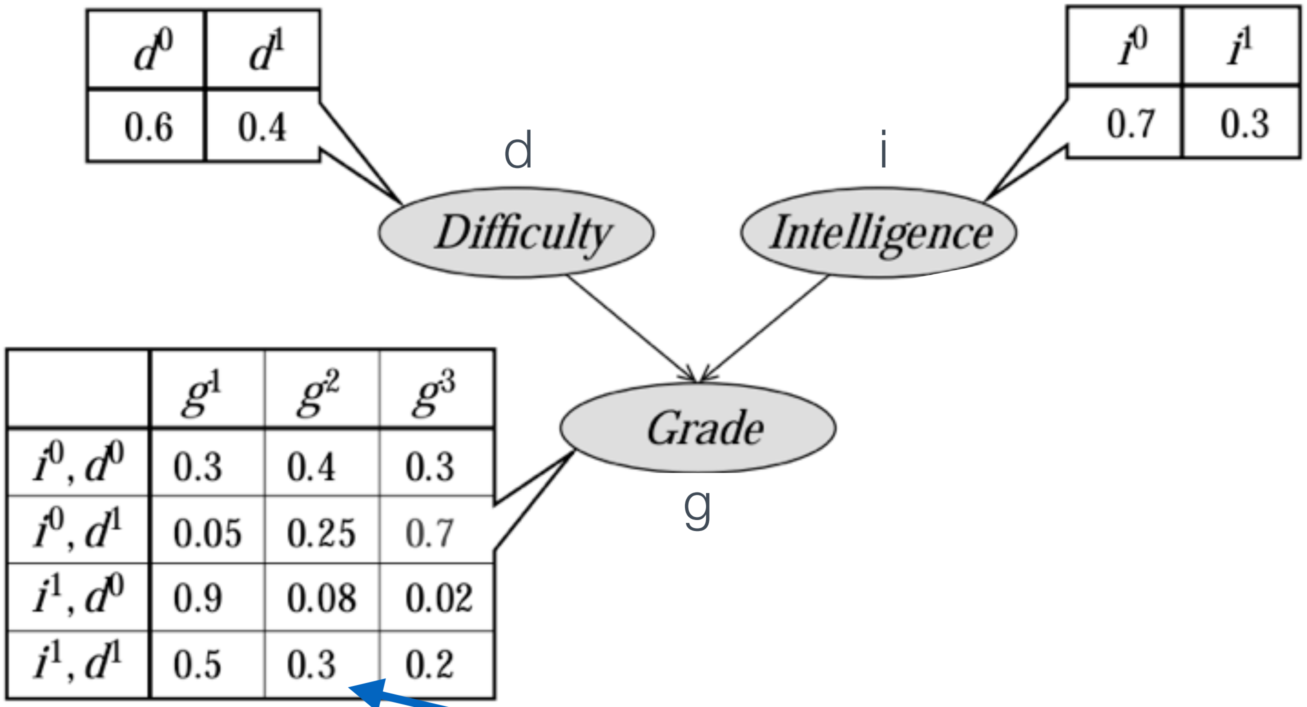
★ Cat CPDs

where $P(x_v | x_{\text{pa}(v)} = c) = \text{Cat}(\boldsymbol{\theta}_{vc})$

and $\theta_{vck} = P(X_v = k \,|\, X_{\text{pa}(v)} = c)$

NOTATION EXAMPLE

$S_d = \{0,1\}$ $S_i = \{0,1\}$



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

		g^1	g^2	g^3
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 0,0\rangle})$	i^0, d^0	0.3	0.4	0.3
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 0,1\rangle})$	i^0, d^1	0.05	0.25	0.7
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 1,0\rangle})$	i^1, d^0	0.9	0.08	0.02
$\mathrm{Cat}(\boldsymbol{\theta}_{g\langle 1,1\rangle})$	i^1, d^1	0.5	0.3	0.2

$\theta_{g\langle 1,1\rangle 2}$

$S_g = \{1,2,3\}$

$$C_g = \prod_{s \in \mathrm{pa}(g)} S_s = S_i \times S_d = \{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,0\rangle, \langle 1,1\rangle\}$$

THE DGM LIKELIHOOD FACTORIZES

- ★ Data $\mathscr{D} = \{x^1, \ldots, x^N\}$ $x^n = x_1^n, \ldots, x_V^n$

- ★ Likelihood

$$p(\mathscr{D} \,|\, \theta) = \prod_{n=1}^{N} p(x^n \,|\, \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \,|\, x_{\mathrm{pa}(v)}^n, \theta_v)$$

- ★ Notation

$$p(X_v = k \,|\, X_{\mathrm{pa}} = c) = \theta_{vck}$$

- ★ Counts

$$N_{vck} = \sum_{n=1}^{N} I(x_v^n = k, x_{\mathrm{pa(v)}}^n = c) \qquad N_{vc} = \sum_{n=1}^{N} I(x_{\mathrm{pa}(v)}^n = c)$$

THE DGM LIKELIHOOD FACTORIZES

- Complete data

$$\mathscr{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- Likelihood

v’s CPD

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

THE DGM LIKELIHOOD FACTORIZES

- ★ Complete data

$$\mathcal{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- ★ Likelihood

$$p(\mathcal{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

THE DGM LIKELIHOOD FACTORIZES

- ★ Complete data

$$\mathcal{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x_1^n, \ldots, x_V^n$$

- ★ Likelihood

$$p(\mathcal{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x_v^n \mid x_{\mathrm{pa}(v)}^n, \theta_v) = \prod_{v=1}^{V} p(\underset{\substack{\uparrow \\ \text{values of v}}}{\mathcal{D}_v} \mid \underset{\substack{\uparrow \\ \text{values of v's parents}}}{\mathcal{D}_{\mathrm{pa}(v)}}, \theta_v)$$

Called: decomposable likelihood (factorizes into family-factors)

PARAMETER AND COUNTS

$$N_{vc} = \sum_{n=1}^{N} I(x^n_{\mathrm{pa}(v)} = c)$$

$$N_{vck} = \sum_{n=1}^{N} I(x^n_v = k, x^n_{\mathrm{pa(v)}} = c)$$

- Complete data

$$\mathscr{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x^n_1, \ldots, x^n_V$$

- Likelihood

$$p(\mathscr{D} \,|\, \theta) = \prod_{n=1}^{N} p(x^n \,|\, \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x^n_v \,|\, x^n_{\mathrm{pa}(v)}, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x^n_v \,|\, x^n_{\mathrm{pa}(v)}, \theta_v) = \prod_{v=1}^{V} p(\mathscr{D}_v \,|\, \mathscr{D}_{\mathrm{pa}(v)}, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{c \in C_v} \prod_{k \in S_v} \theta_{vck}^{N_{vck}}$$

THE LIKELIHOOD FACTORIZES

$$N_{vc} = \sum_{n=1}^{N} I(x^n_{\mathrm{pa}(v)} = c)$$

$$N_{vck} = \sum_{n=1}^{N} I(x^n_v = k, x^n_{\mathrm{pa(v)}} = c)$$

★ Complete data

$$\mathscr{D} = \{x^1, \ldots, x^N\}$$

$$x^n = x^n_1, \ldots, x^n_V$$

★ Likelihood

$$p(\mathscr{D} \mid \theta) = \prod_{n=1}^{N} p(x^n \mid \theta) = \prod_{n=1}^{N} \prod_{v=1}^{V} p(x^n_v \mid x^n_{\mathrm{pa}(v)}, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{n=1}^{N} p(x^n_v \mid x^n_{\mathrm{pa}(v)}, \theta_v) = \prod_{v=1}^{V} p(\mathscr{D}_v \mid \mathscr{D}_{\mathrm{pa}(v)}, \theta_v)$$

$$= \prod_{v=1}^{V} \prod_{c \in C_v} \prod_{k \in S_v} \theta_{vck}^{N_{vck}}$$

So ML estimate

$$\theta_{vck} = N_{vck} / N_{vc}$$

BAYESIAN PARAMETER LEARNING

- ★ Decomposable prior

$$p(\boldsymbol{\theta}) = \prod_{v=1}^{V} p(\boldsymbol{\theta}_v) \quad = \prod_{\substack{v \in [V] \\ c \in S_{\mathbf{pa}(v)}}} \mathrm{Dir}(\boldsymbol{\theta}_{v\boldsymbol{c}} | \boldsymbol{\alpha}_{v\boldsymbol{c}})$$

- ★ Gives decomposable posterior

$$p(\theta \,|\, \mathscr{D}) \propto p(\mathscr{D} \,|\, \theta) p(\theta)$$

$$= \prod_{v=1}^{V} p(\mathscr{D}_v \,|\, \mathscr{D}_{\mathrm{pa}(v)}, \theta_v) p(\theta_v)$$

BAYESIAN PARAMETER LEARNING

- ★ Decomposable prior

$$p(\boldsymbol{\theta}) = \prod_{v=1}^{V} p(\boldsymbol{\theta}_v) \quad = \quad \prod_{\substack{v \in [V] \\ c \in S_{\mathrm{pa}(v)}}} \mathrm{Dir}(\boldsymbol{\theta}_{v\boldsymbol{c}} | \boldsymbol{\alpha}_{v\boldsymbol{c}})$$

- ★ Gives decomposable posterior

$$\begin{aligned}
p(\theta \,|\, \mathscr{D}) &\propto p(\mathscr{D} \,|\, \theta) p(\theta) \\
&= \prod_{v=1}^{V} p(\mathscr{D}_v \,|\, \mathscr{D}_{\mathrm{pa}(v)}, \theta_v) p(\theta_v) \\
&= \prod_{\substack{v \in [V] \\ c \in S_{\mathrm{pa}(v)}}} \mathrm{Cat}(N_{v\boldsymbol{c}1}, \ldots, N_{v\boldsymbol{c}|S_v|} | \boldsymbol{\theta}_{vc}) \mathrm{Dir}(\boldsymbol{\theta}_{v\boldsymbol{c}} | \boldsymbol{\alpha}_{v\boldsymbol{c}}) \\
&\propto \prod_{\substack{v \in [V] \\ c \in S_{\mathrm{pa}(v)}}} \mathrm{Dir}(\boldsymbol{\theta}_{v\boldsymbol{c}} | N_{v\boldsymbol{c}1} + \alpha_{v\boldsymbol{c}1}, \ldots, N_{v\boldsymbol{c}|S_v|} + \alpha_{v\boldsymbol{c}|S_v|})
\end{aligned}$$

BAYESIAN PARAMETER LEARNING

- ★ Decomposable prior

$$p(\boldsymbol{\theta}) = \prod_{v=1}^{V} p(\boldsymbol{\theta}_v) \quad = \prod_{\substack{v \in [V] \\ c \in S_{\mathrm{pa}(v)}}} \mathrm{Dir}(\boldsymbol{\theta}_{v\boldsymbol{c}} | \boldsymbol{\alpha}_{v\boldsymbol{c}})$$

- ★ Gives decomposable posterior

$$p(\theta \,|\, \mathscr{D}) \propto p(\mathscr{D} \,|\, \theta) p(\theta)$$

$$= \prod_{v=1}^{V} p(\mathscr{D}_v \,|\, \mathscr{D}_{\mathrm{pa}(v)}, \theta_v) p(\theta_v)$$

$$= \prod_{\substack{v \in [V] \\ c \in S_{\mathrm{pa}(v)}}} \mathrm{Cat}(N_{v\boldsymbol{c}1}, \ldots, N_{v\boldsymbol{c}|S_v|} | \boldsymbol{\theta}_{vc}) \mathrm{Dir}(\boldsymbol{\theta}_{v\boldsymbol{c}} | \boldsymbol{\alpha}_{v\boldsymbol{c}})$$