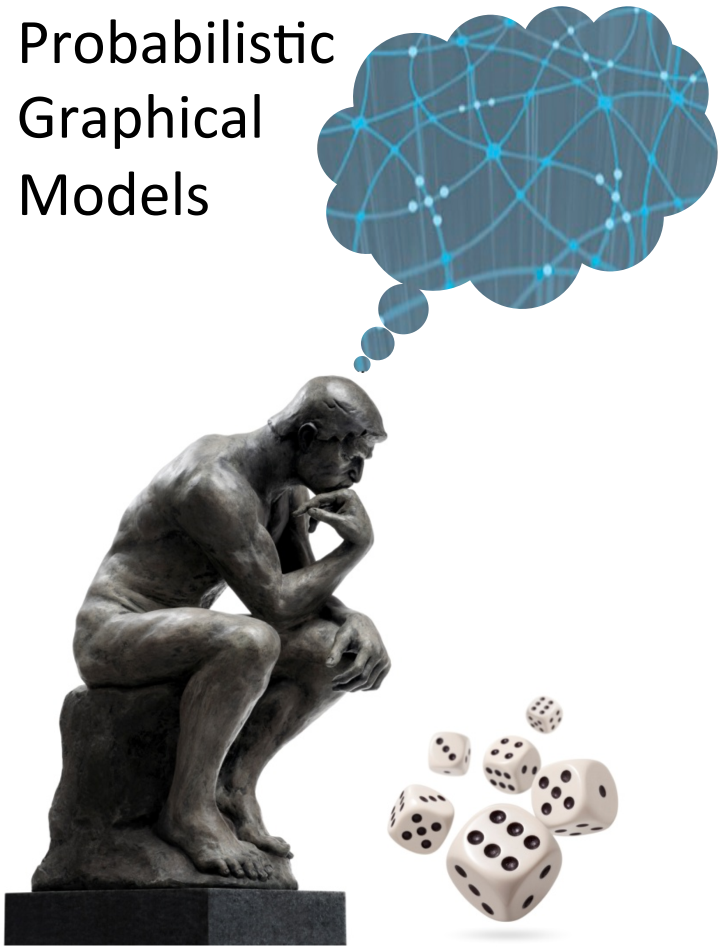
Probabilistic
Graphical
Models



Representation

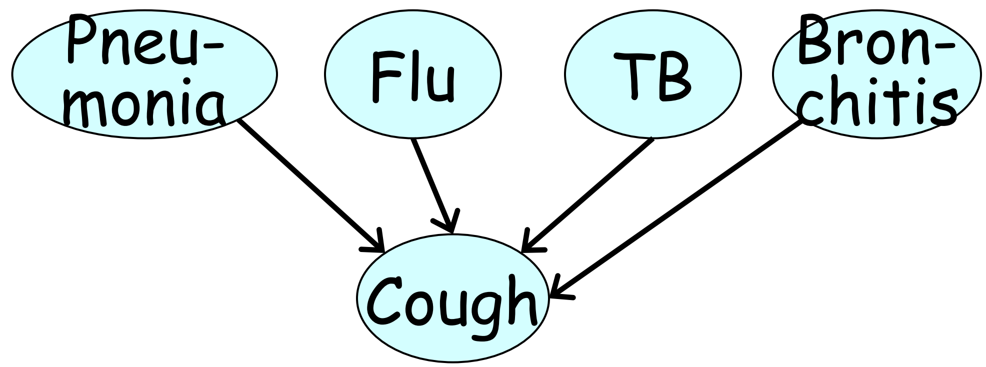
Local Structure

Overview

Tabular Representations

I D
G

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2



k parents
$O(2^k)$ entries

General CPD

- CPD $P(X \mid Y_1, \ldots, Y_k)$ specifies distribution over X for each assignment $y_1, \ldots, y_k$
- Can use any function to specify a factor $\phi(X, Y_1, \ldots, Y_k)$ such that

$$\sum_x \phi(x, y_1, \ldots, y_k) = 1 \text{ for all } y_1, \ldots, y_k$$

Many Models

- Deterministic CPDs
- Tree-structured CPDs
- Logistic CPDs & generalizations
- Noisy OR / AND
- Linear Gaussians & generalizations

Context-Specific Independence

$$P \models (X \perp_c Y \mid \boldsymbol{Z}, \boldsymbol{c})$$

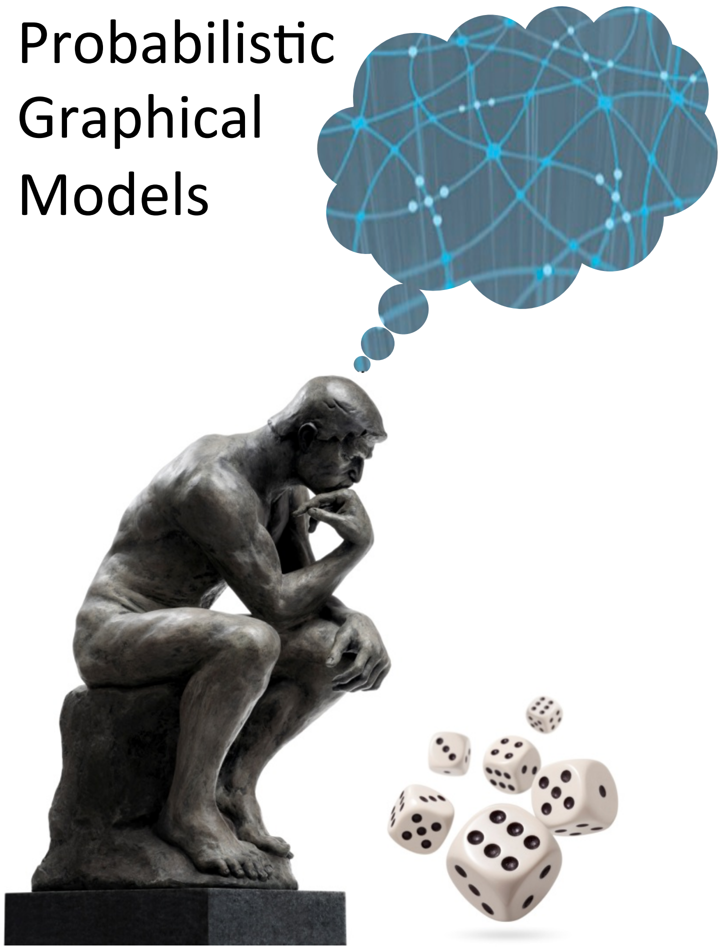
assignment to C

$$P(\boldsymbol{X}, \boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{c}) = P(\boldsymbol{X} \mid \boldsymbol{Z}, \boldsymbol{c}) P(\boldsymbol{y} \mid \boldsymbol{Z}, \boldsymbol{c})$$

$$P(\boldsymbol{X} \mid \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{c}) = P(\boldsymbol{X} \mid \boldsymbol{Z}, \boldsymbol{c})$$

$$P(\boldsymbol{Y} \mid \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{c}) = P(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{c})$$

Probabilistic Graphical Models

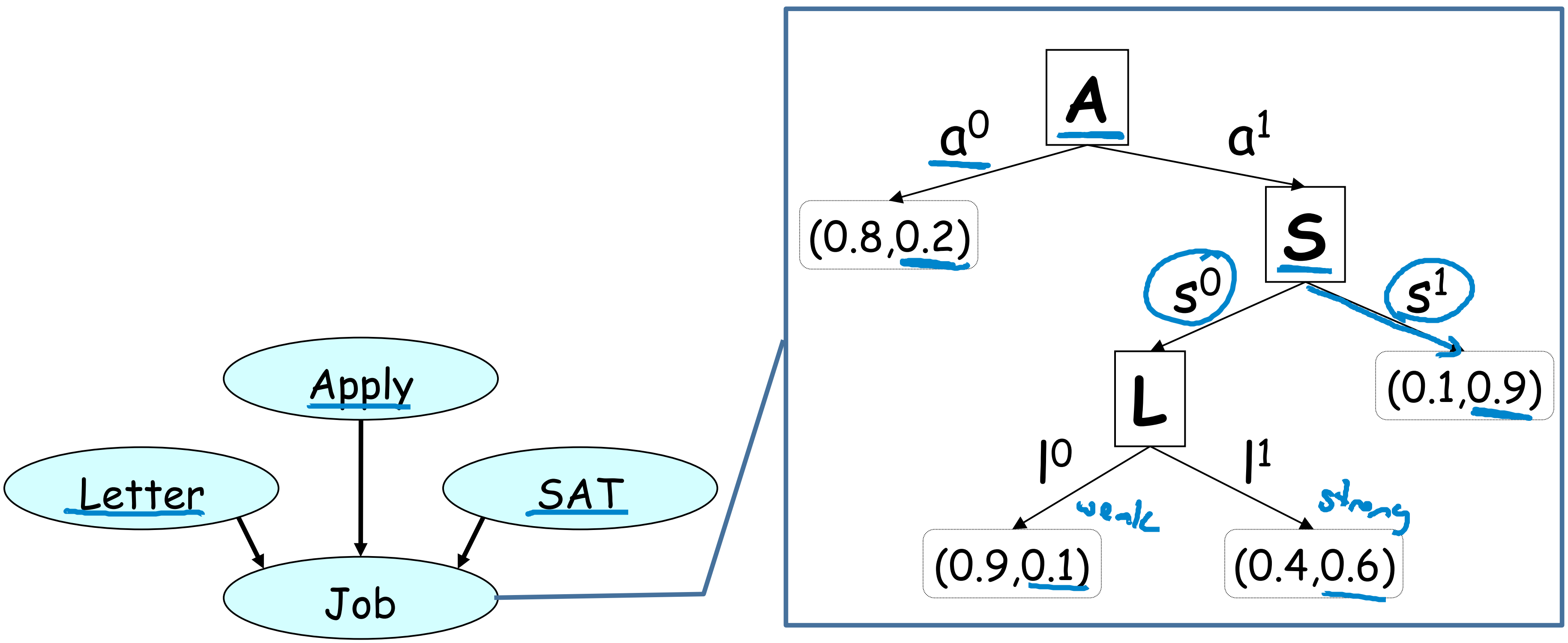


Representation

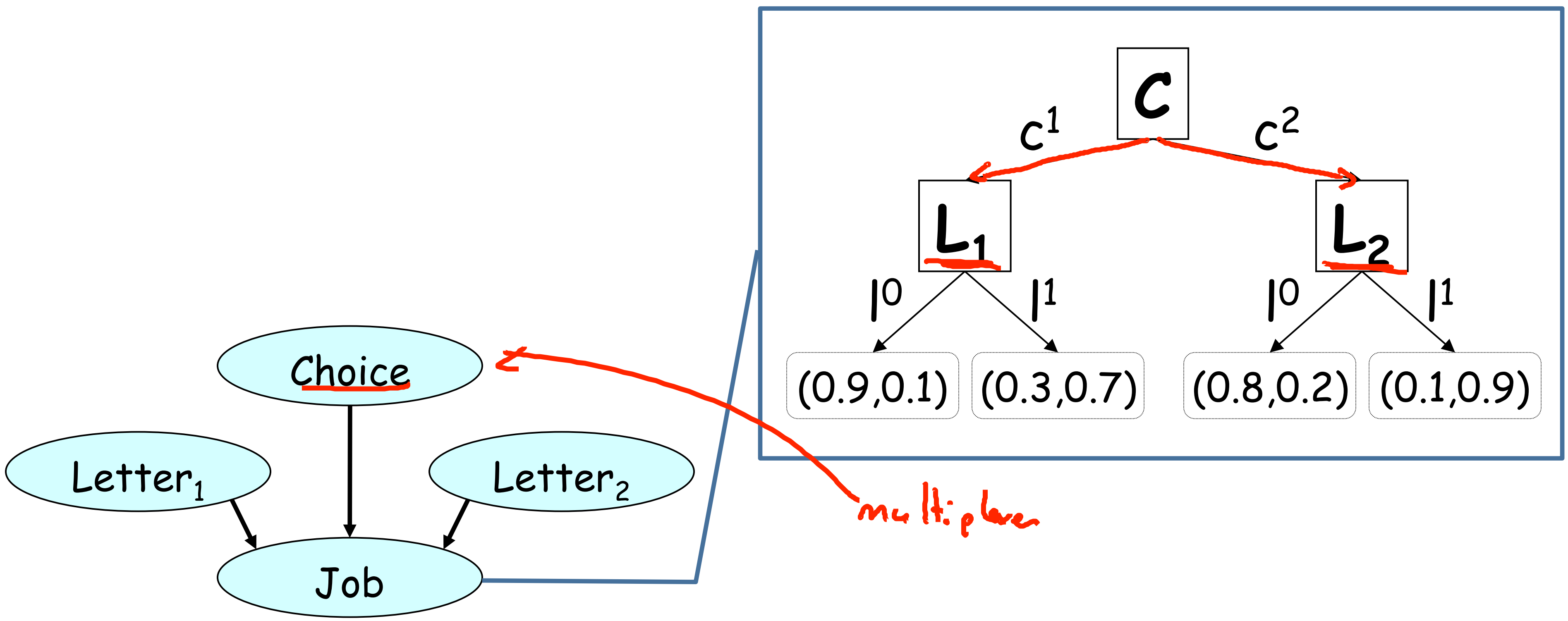
Local Structure

Tree Structured CPDs

Tree CPD


Apply
Letter
SAT
Job
A
a^0
a^1
(0.8,0.2)
S
s^0
s^1
(0.1,0.9)
L
l^0
l^1
weak
strong
(0.9,0.1)
(0.4,0.6)

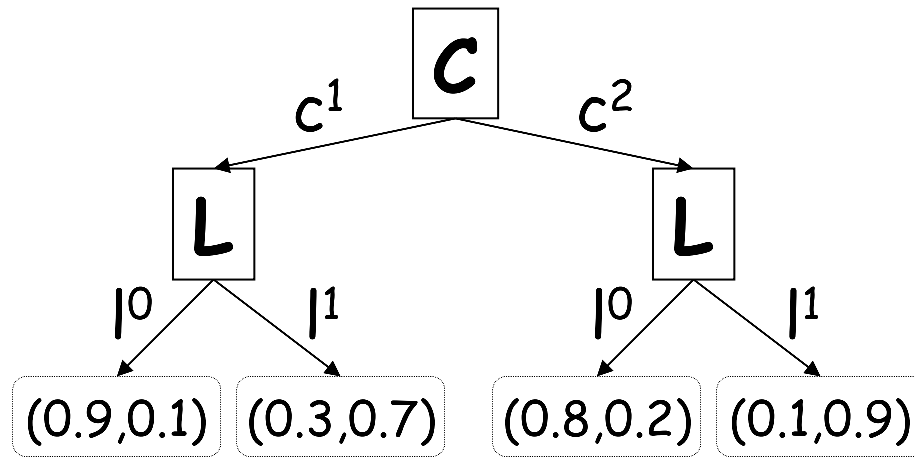
Tree CPD

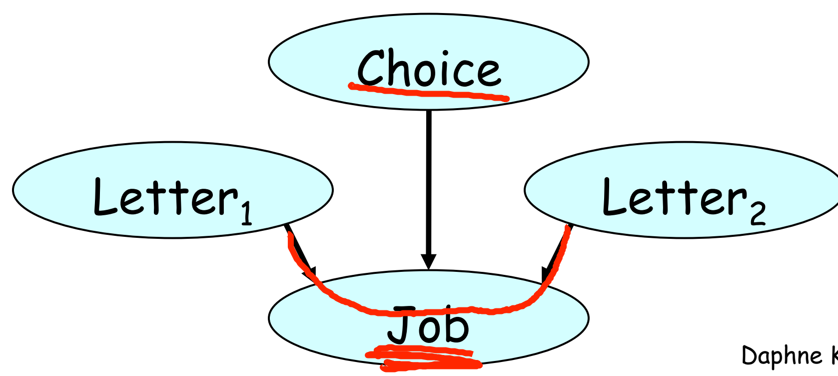


Daphne Koller

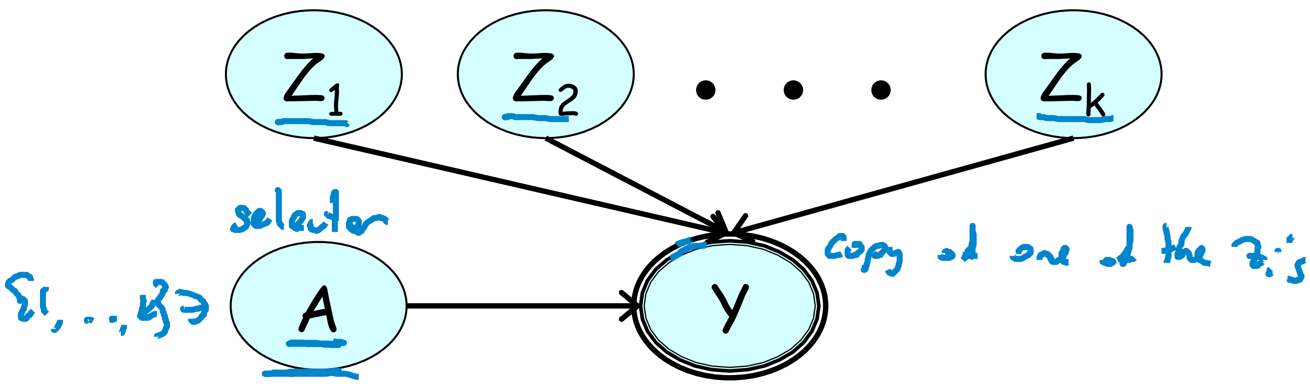
$$(L_1 \perp L_2 \mid J, C)$$

$$(L_1 \perp_c L_2 \mid J, c_1)$$
$$(L_1 \perp_c L_2 \mid J, c_2)$$





Multiplexer CPD

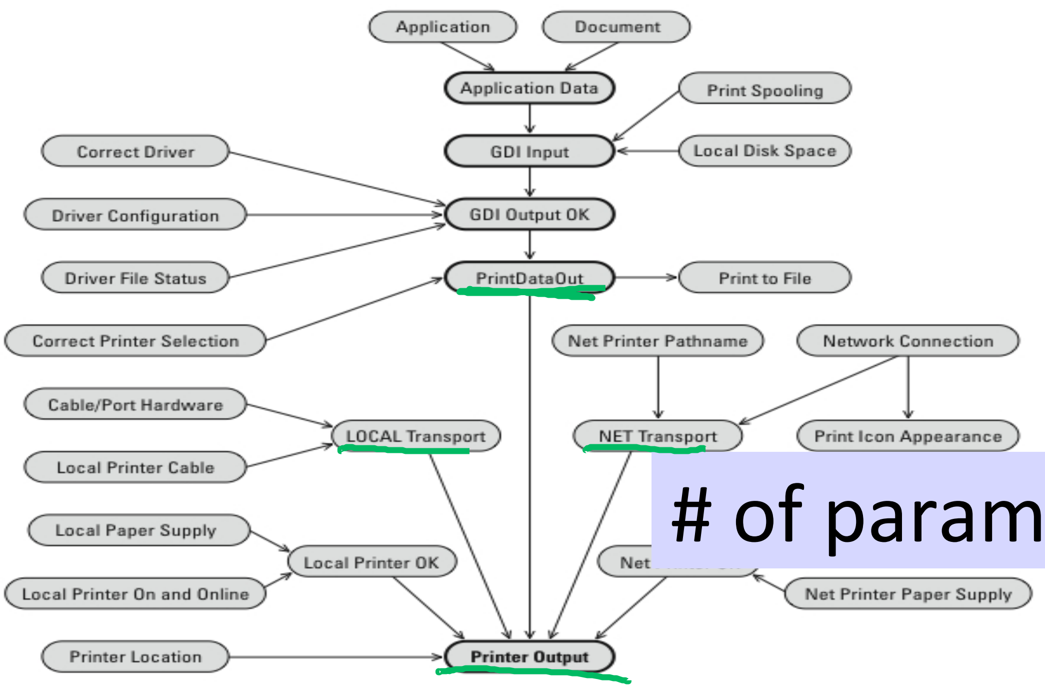


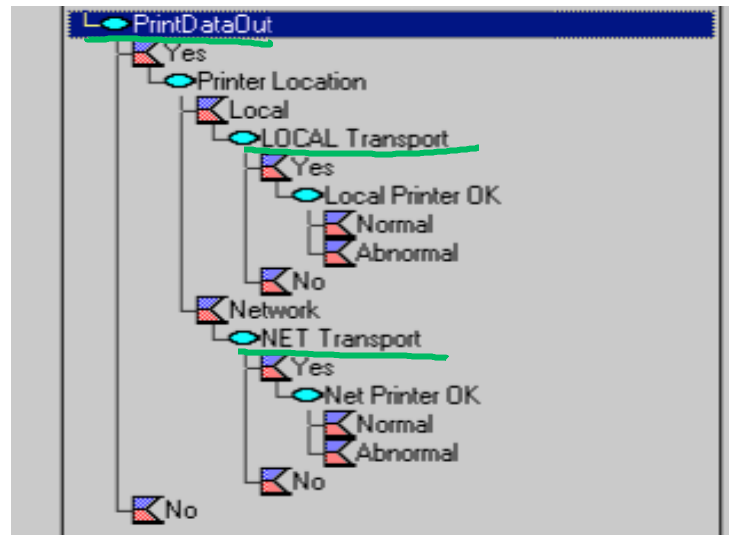
$A = a \Rightarrow Y = Z_a$

$$P(Y \mid A, Z_1, \ldots, Z_k) = \begin{cases} 1 & Y = Z_A \\ 0 & \text{otherwise} \end{cases}$$

Thanks to: Eric Horvitz, Microsoft Research

Microsoft Troubleshooters



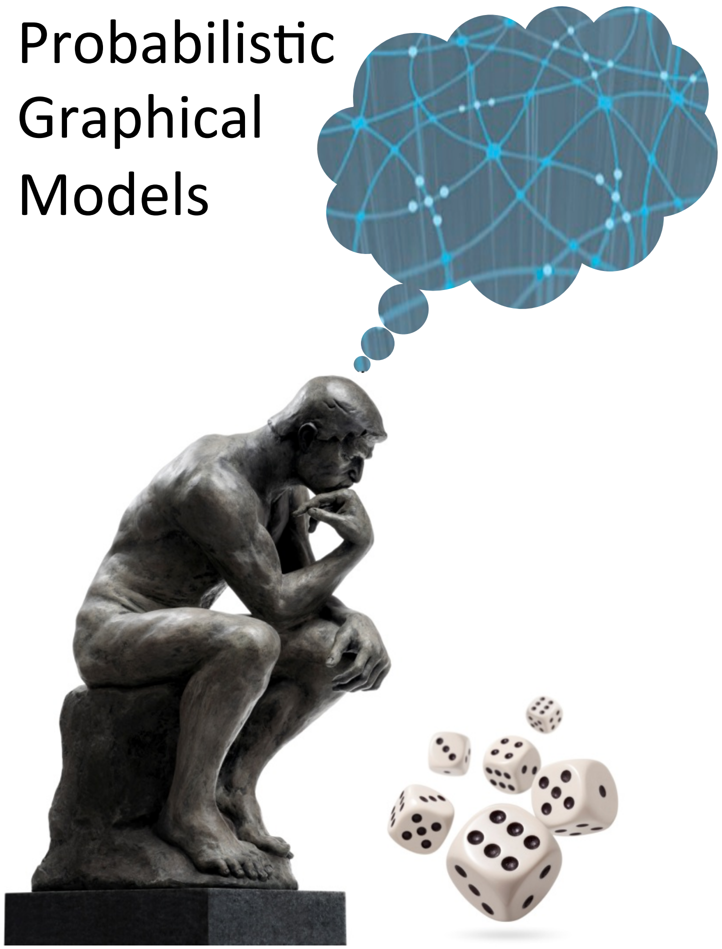


# of parameters: 145 to 55

Summary

- Compact CPD representation that captures context-specific dependencies
- Relevant in multiple applications:
  - Hardware configuration variables
  - Medical settings
  - Dependence on agent's action
  - Perceptual ambiguity

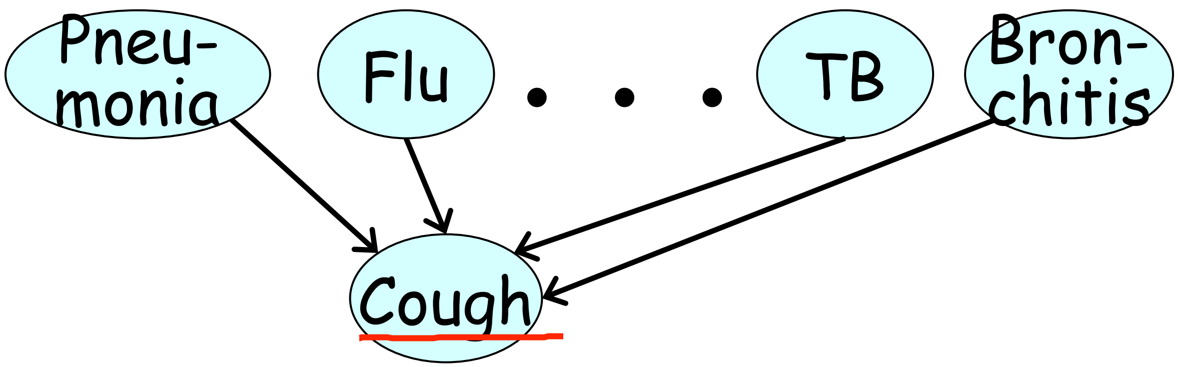
Probabilistic Graphical Models



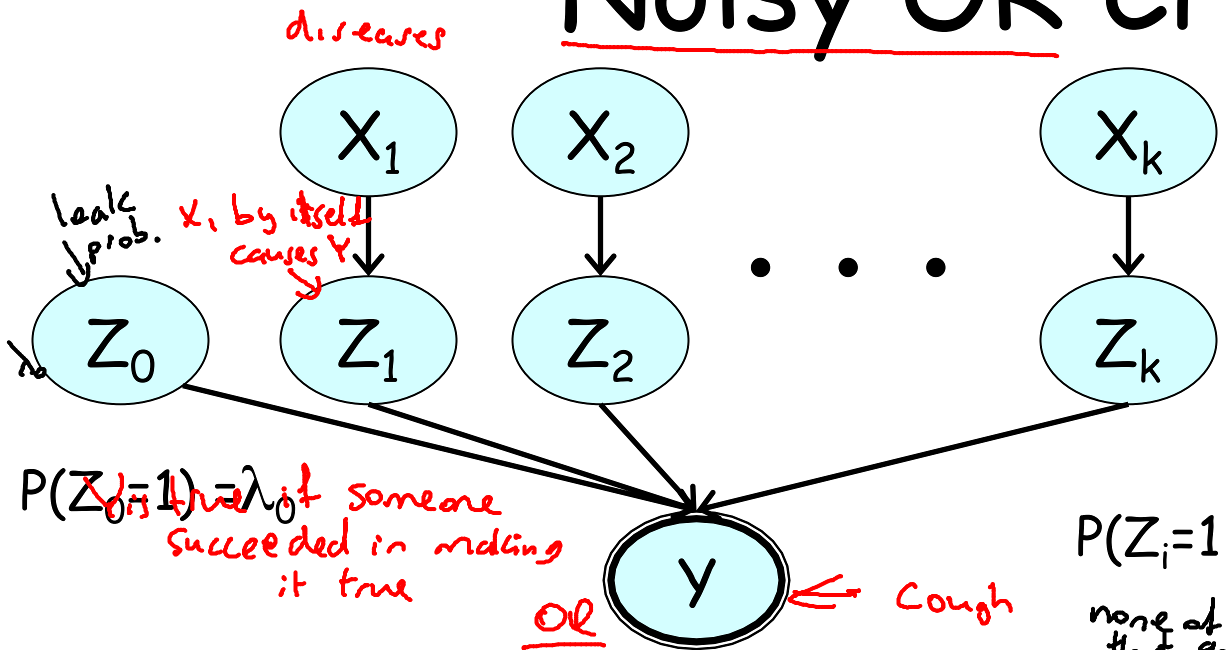
Representation

Local Structure

Independence of Causal Influence


Pneu-
monia
Flu
TB
Bron-
chitis
Cough

Noisy OR CPD



$P(Z_0=1) = \lambda_0$

$$P(Z_i=1 \mid X_i) = \begin{cases} 0 & X_i=0 \\ \lambda_i \in [0,1] & X_i=1 \end{cases}$$

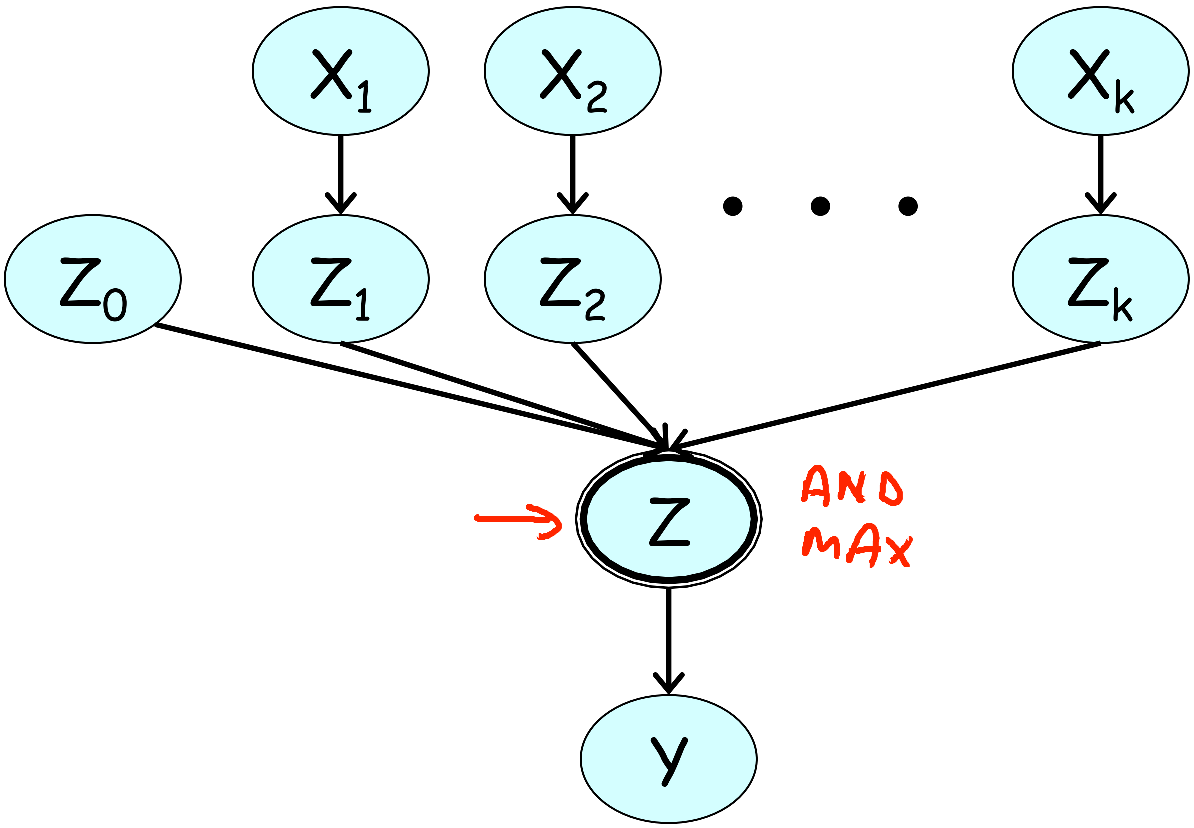
none of causes that are on fail

$$P(Y=0 \mid X_1,\ldots,X_k) = (1-\lambda_0) \prod_{i:X_i=1} (1-\lambda_i)$$

$$P(Y=1 \mid X_1,\ldots,X_k) = 1 - P(Y=0 \mid X_1,\ldots,X_k)$$

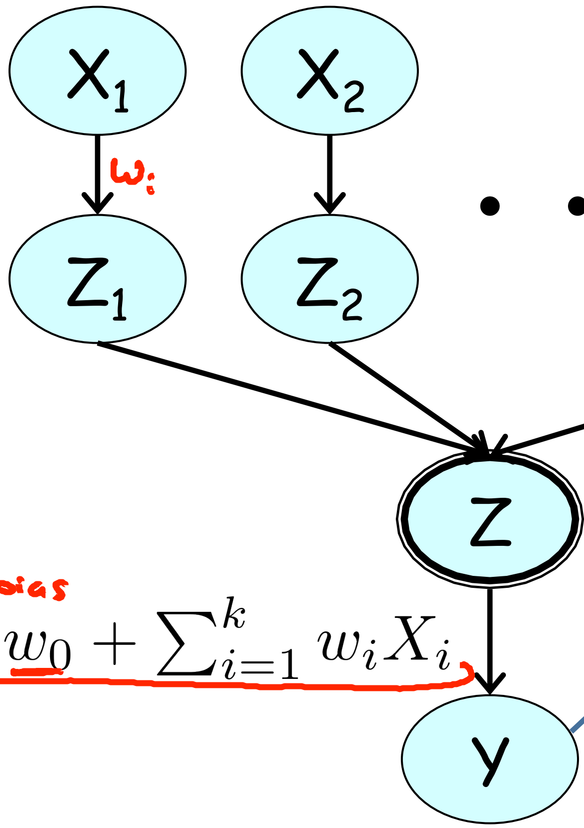
Daphne Koller

Independence of Causal Influence



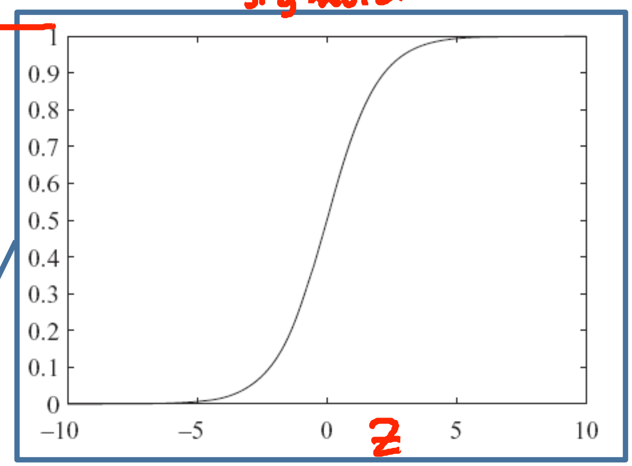
Daphne Koller

Sigmoid CPD



bias

$$Z = w_0 + \sum_{i=1}^{k} w_i X_i$$

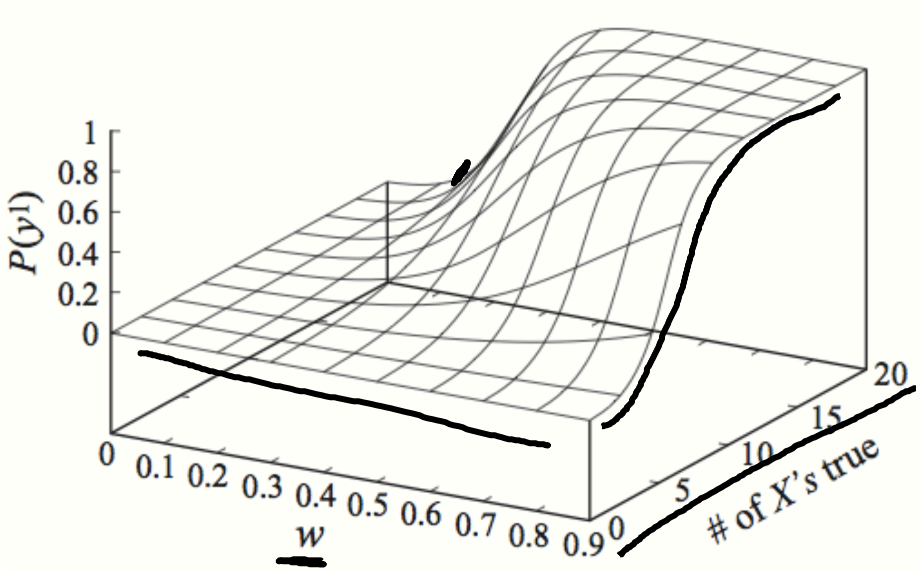


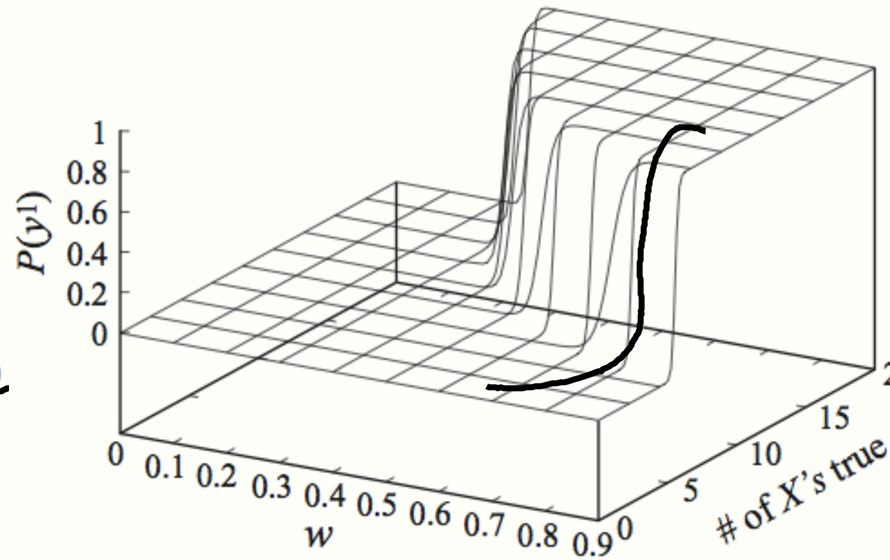
$$\text{sigmoid}(z) = \frac{e^z}{1 + e^z}$$ (0,1)

$$P(y^1 \mid X_1, \ldots, X_k) = \text{sigmoid}(Z)$$

Daphne Koller

Behavior of Sigmoid CPD





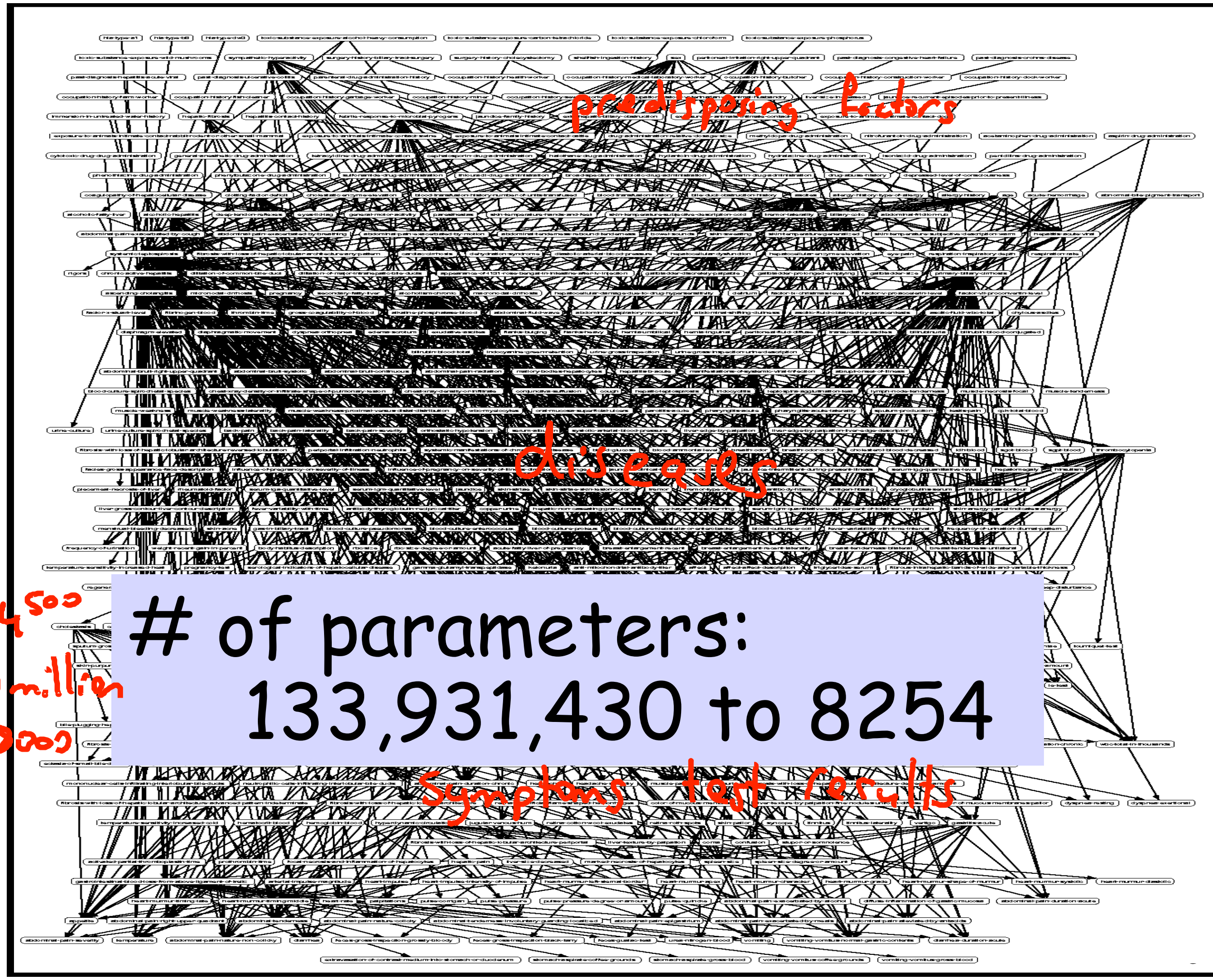
$w_0 = -5$

multiply w and w_0 by 10

CPCS

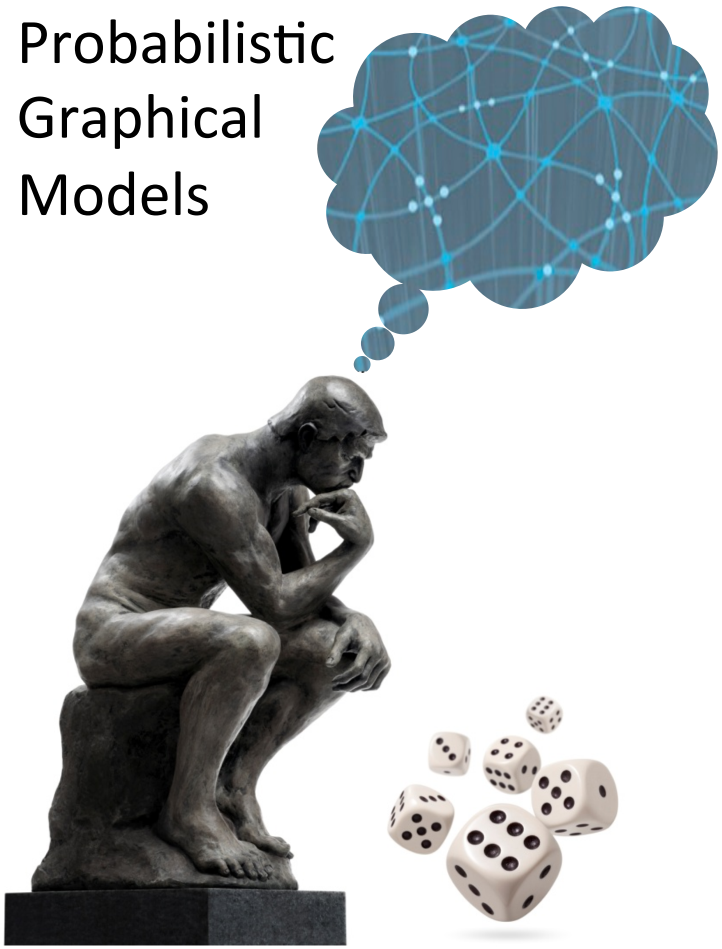
M. Pradhan
G. Provan
B. Middleton
M. Henrion
UAI 1994

joint dist $\approx 4^{500}$
factorized $\approx$ 134 million
noisy max CPD $\approx$ 8000



# of parameters:
133,931,430 to 8254

Probabilistic Graphical Models

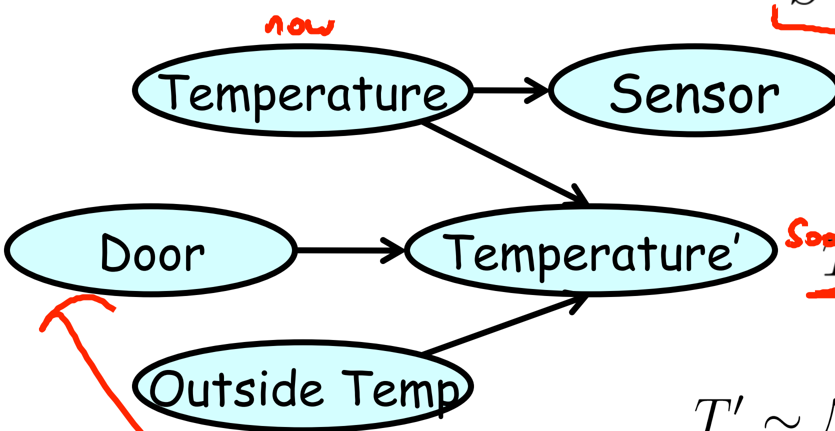


Representation

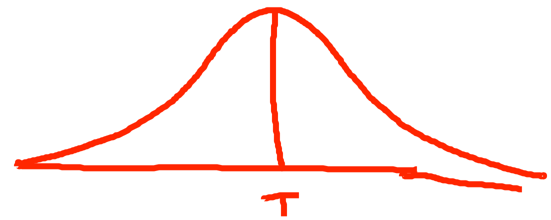
Local Structure

Continuous Variables

Continuous Variables



$$S \sim \mathcal{N}\left(T; \sigma_S^2\right)$$



linear Gaussian

soon $T' \sim \mathcal{N}\left(\alpha T + (1-\alpha)O; \sigma_T^2\right)$

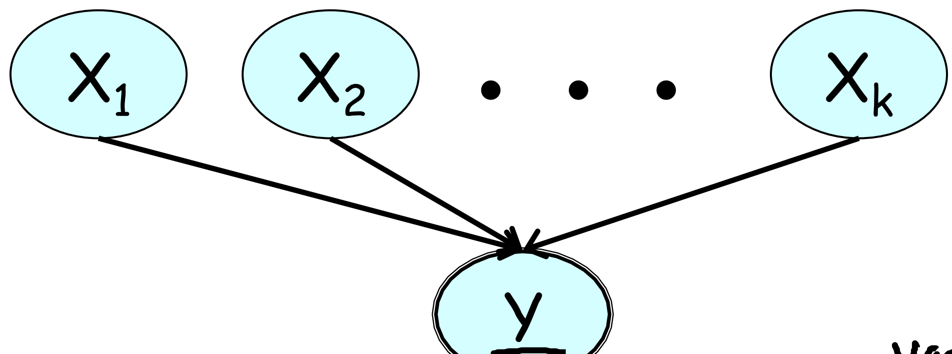
Gaussian around

$$T' \sim \mathcal{N}\left(\alpha_0 T + (1-\alpha_0)O; \sigma_{0T}^2\right) \quad \text{if } D=0$$

$$T' \sim \mathcal{N}\left(\alpha_1 T + (1-\alpha_1)O; \sigma_{1T}^2\right) \quad \text{if } D=1$$

conditional linear Gaussian

Linear Gaussian

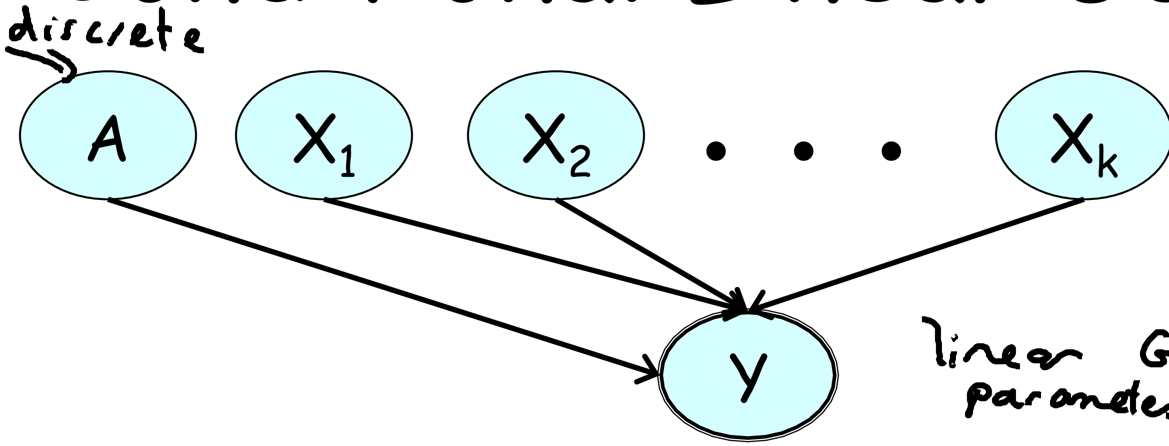


$$Y \sim \mathcal{N}\left(\underbrace{w_0 + \sum w_i X_i}_{\text{mean}}; \sigma^2\right)$$

variance doesn't depend on parents

linear function of parents

Conditional Linear Gaussian

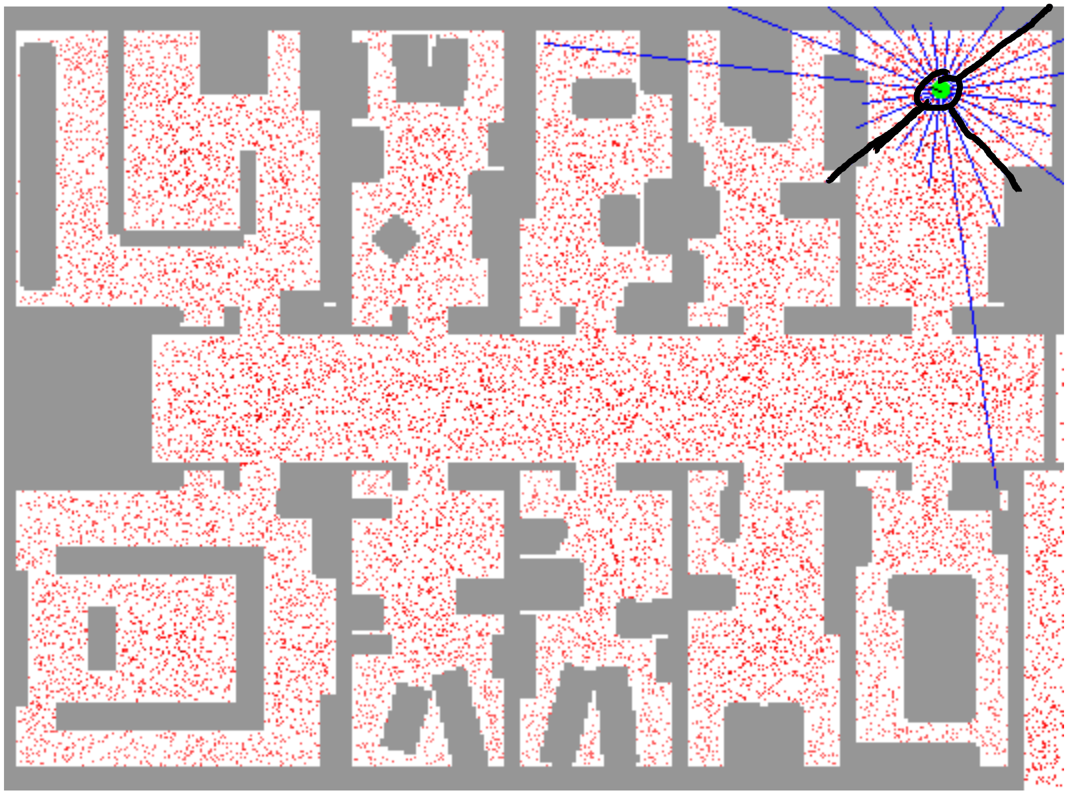


linear Gaussian whose parameters depend on A

$$Y \sim \mathcal{N}\left(w_{\underline{a}0} + \sum w_{\underline{a}i} X_i ; \sigma^2_{\underline{a}}\right)$$

variance can depend on A

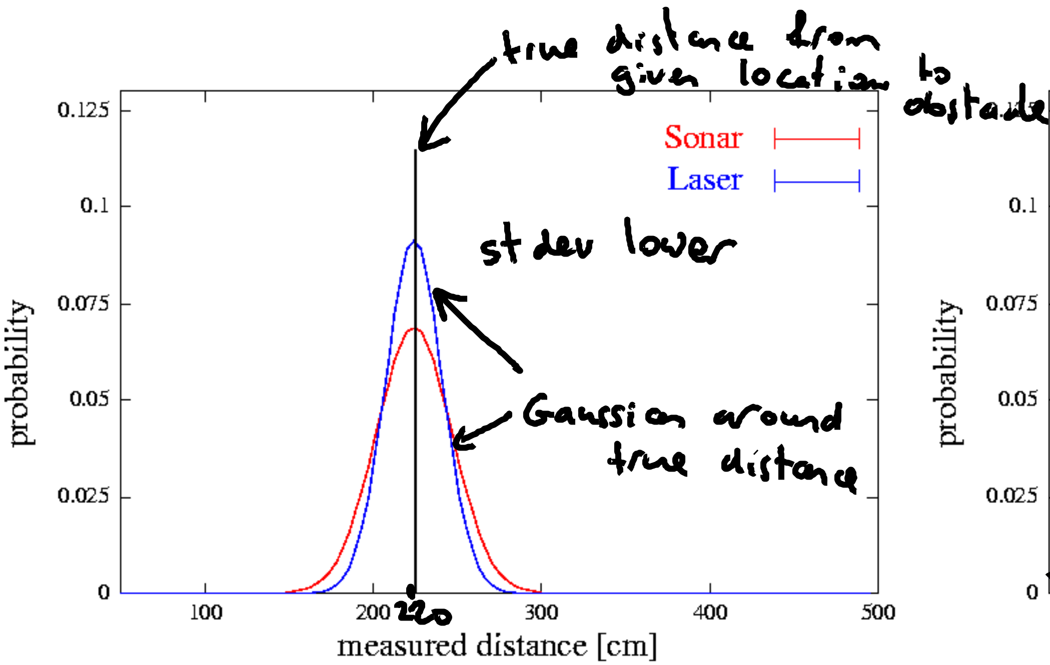
Robot Localization

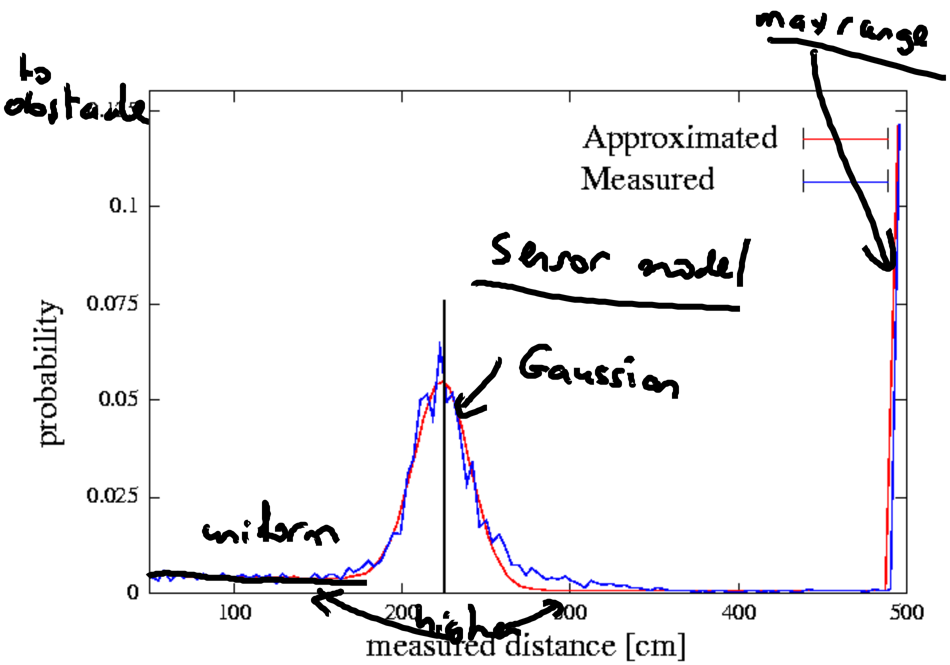


Fox, Burgard, Thrun

Daphne Koller

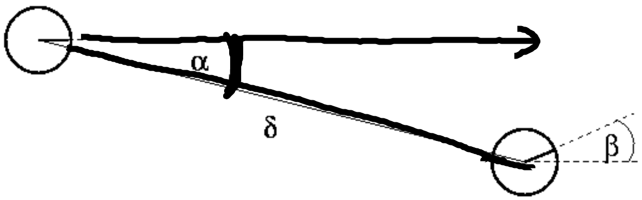
Nonlinear Gaussians

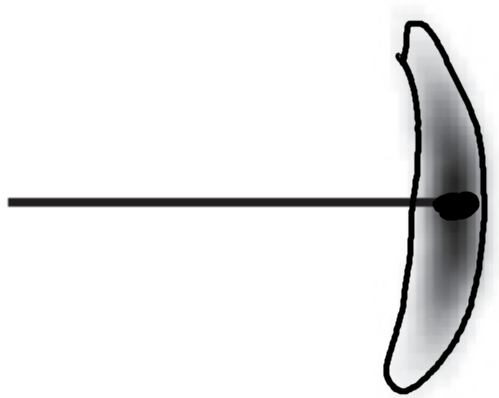


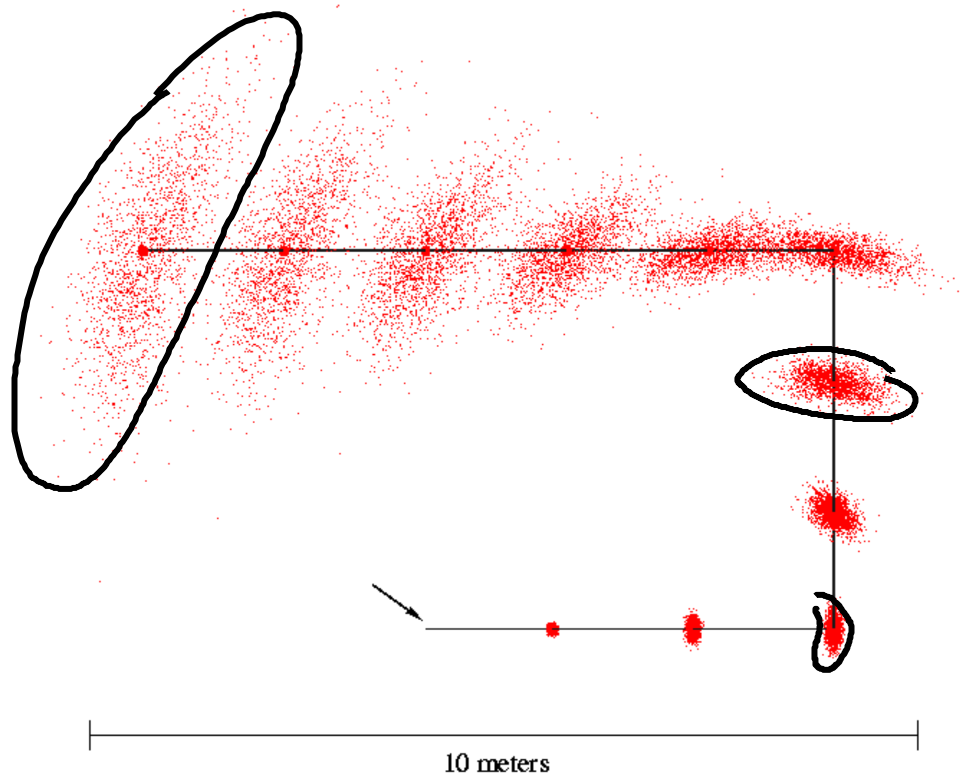


Fox, Burgard, Thrun

Robot Motion Model







Fox, Burgard, Thrun