Boston Bayesians

Microsoft NERD – August 25, 2016

Intro to Bayesian Networks: The Monty Hall Problem

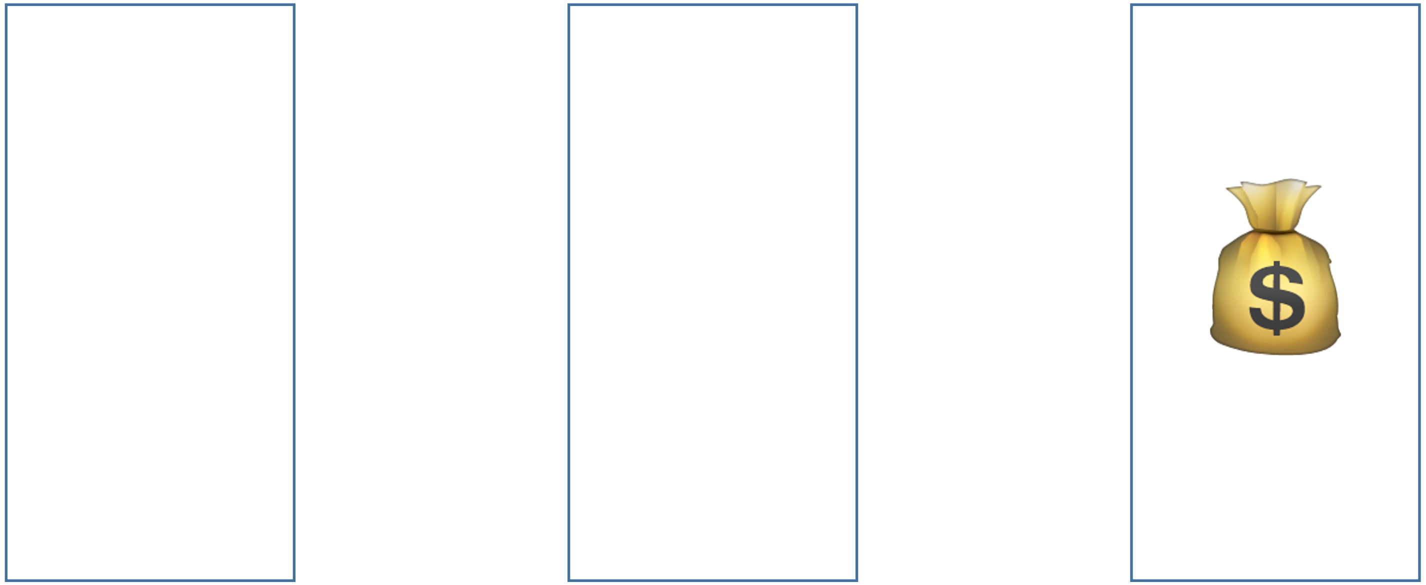
Byron Galbraith, PhD

Chief Data Scientist, Talla

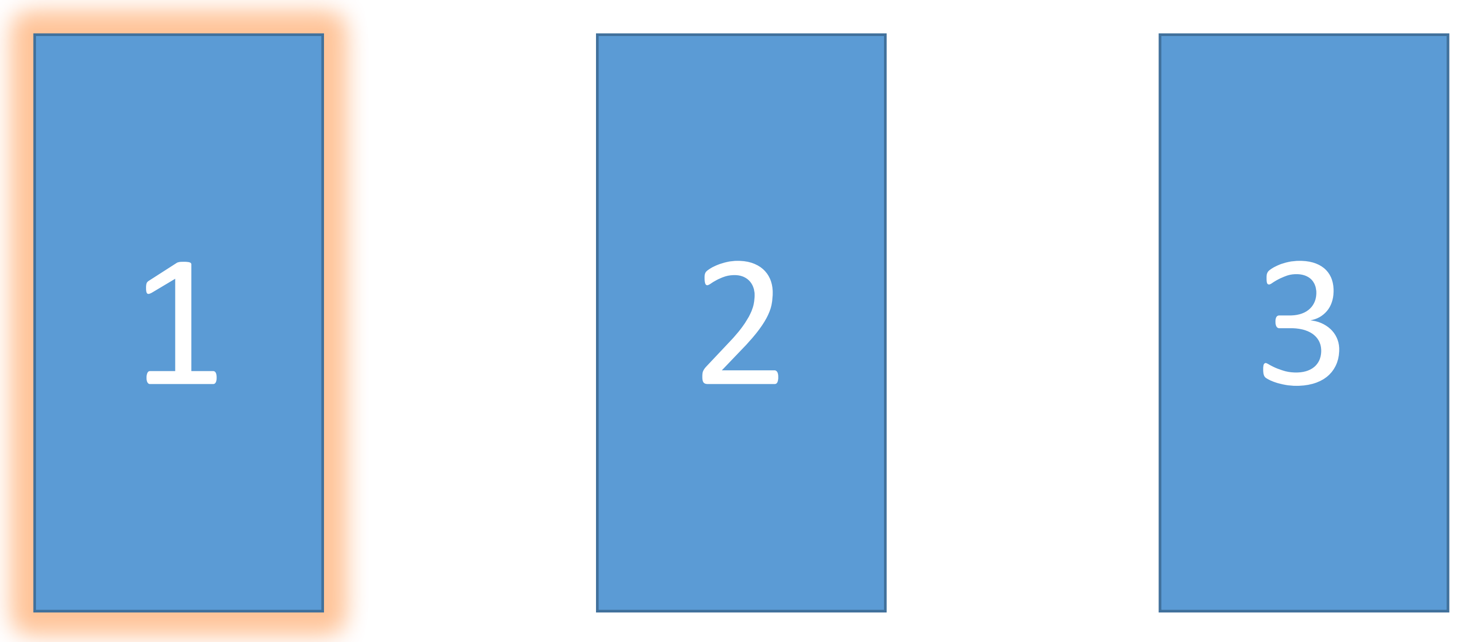
@bgalbraith

The Monty Hall Problem

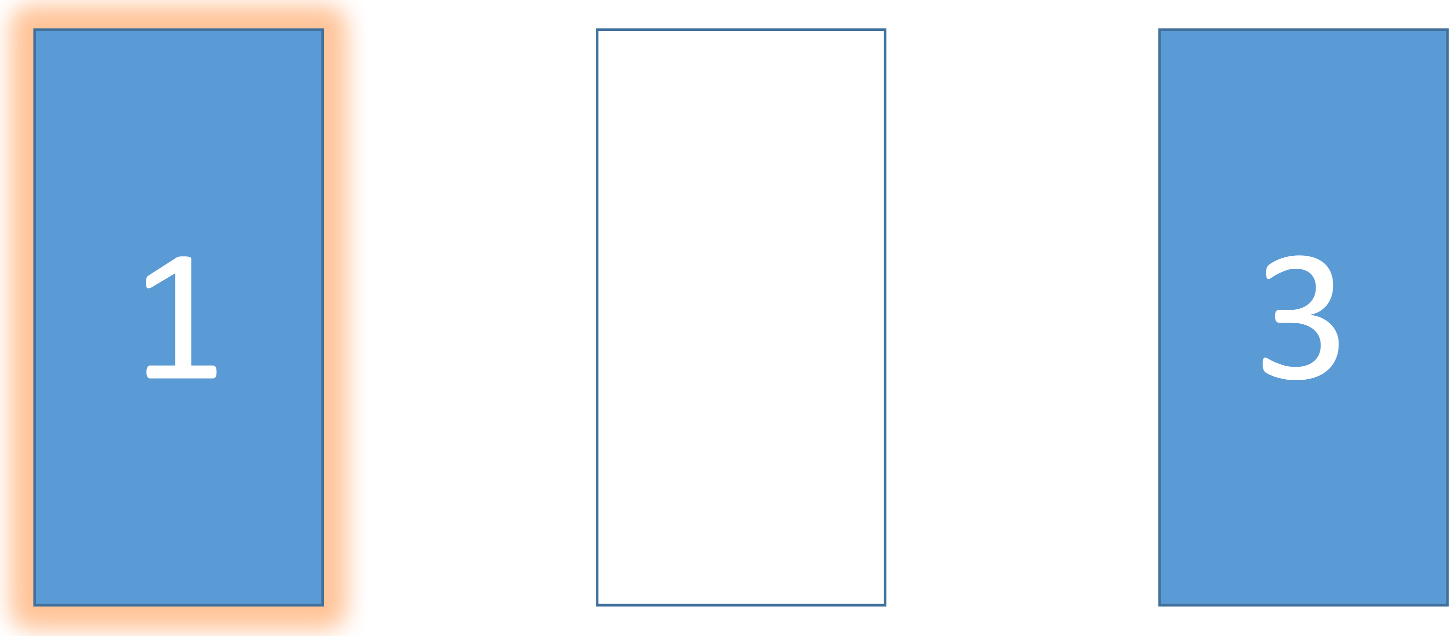
The **prize** is hidden behind a door



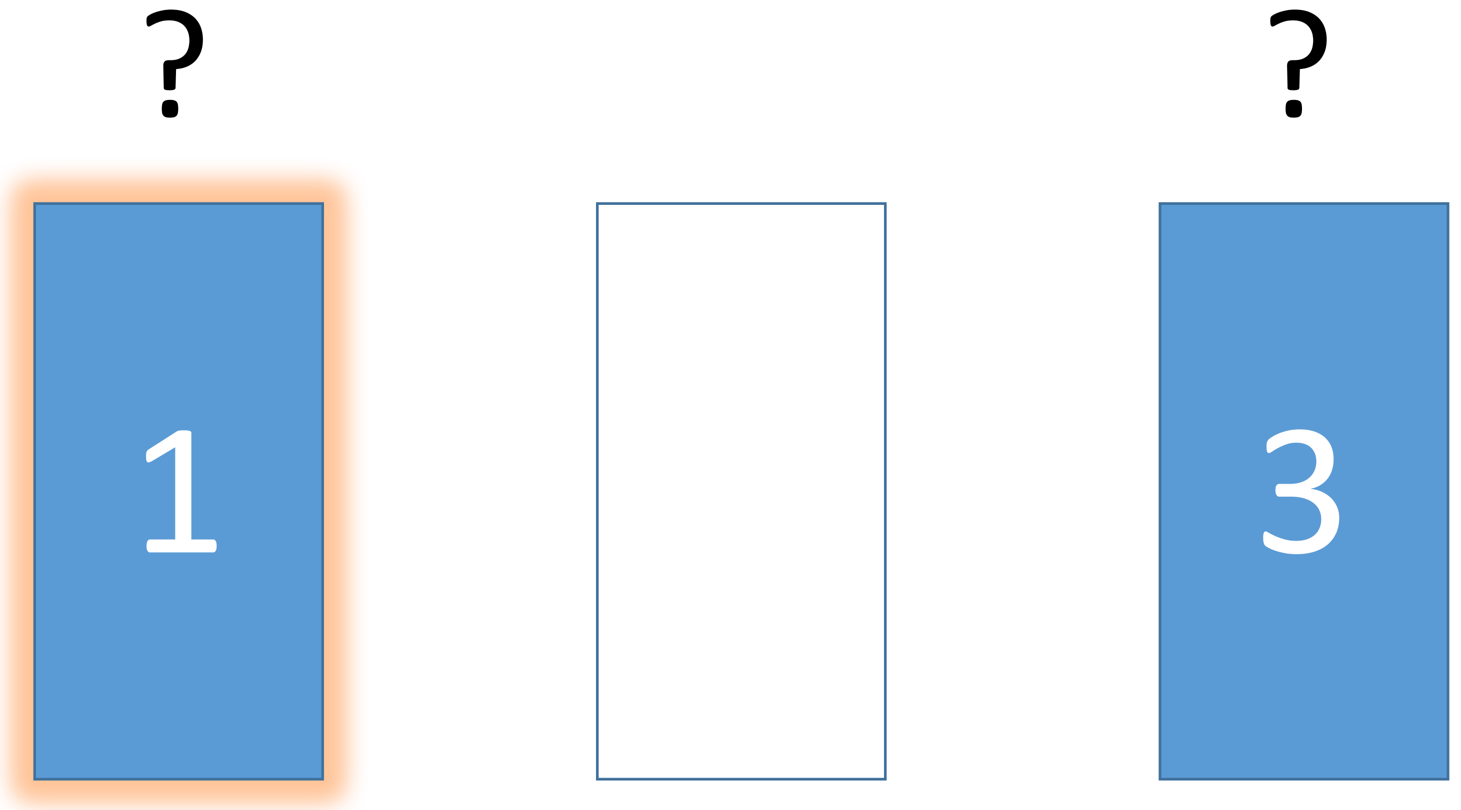
The **contestant** chooses a door



The **host** reveals a non-winning door



Should the contestant **switch** or **stay**?



Random Variables

$P(X)$ $P(Y)$

X Y

Independent Random Variables

$$P(X, Y) = P(X)P(Y)$$

X Y

Conditional Random Variables

$$P(X, Y) = P(X)P(Y|X)$$

X Y

Bayesian Network

$$P(X,Y) = P(X)P(Y|X)$$

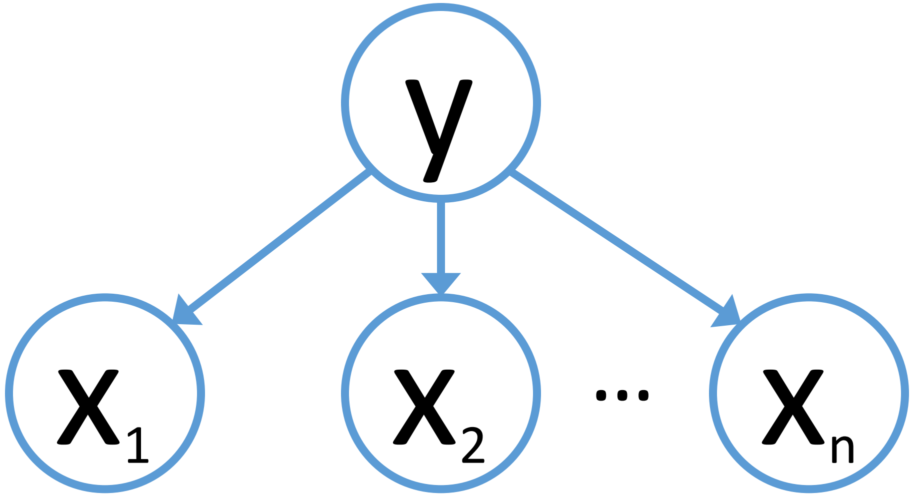
Bayesian Networks are Graphical Models

$$P(X,Y) = P(X)P(Y|X)$$

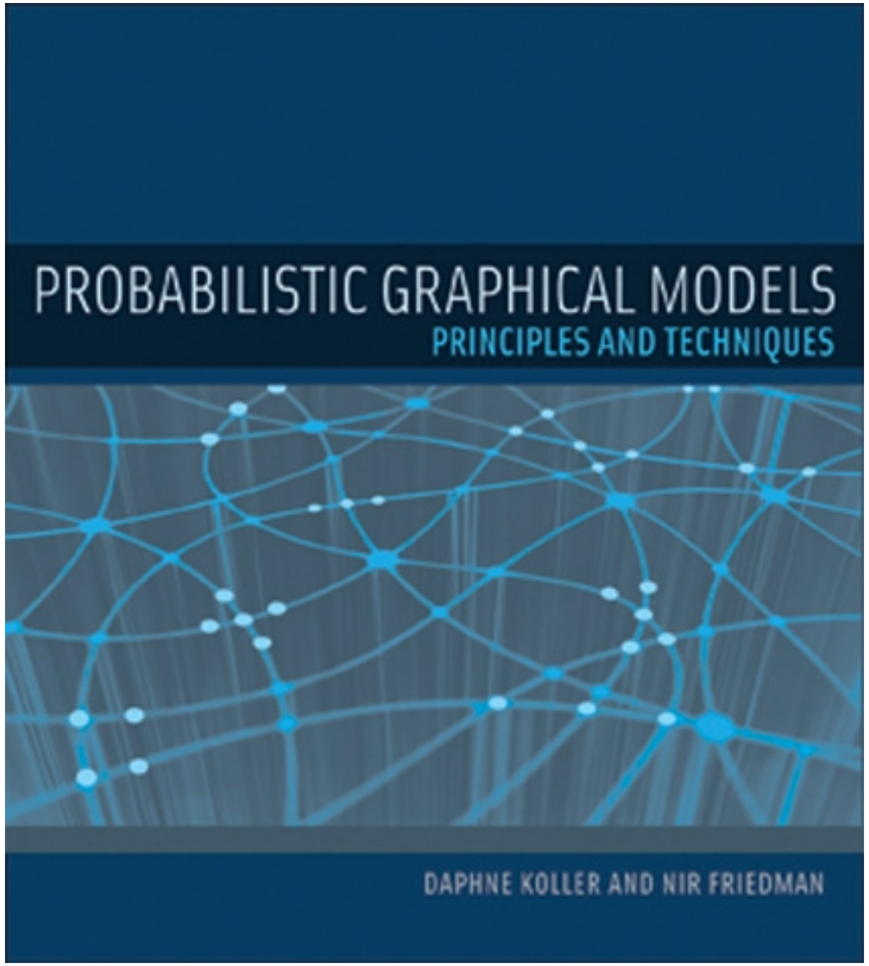
Node = Random variable
Edge = Conditional dependency

Bayesian Network: Naïve Bayes

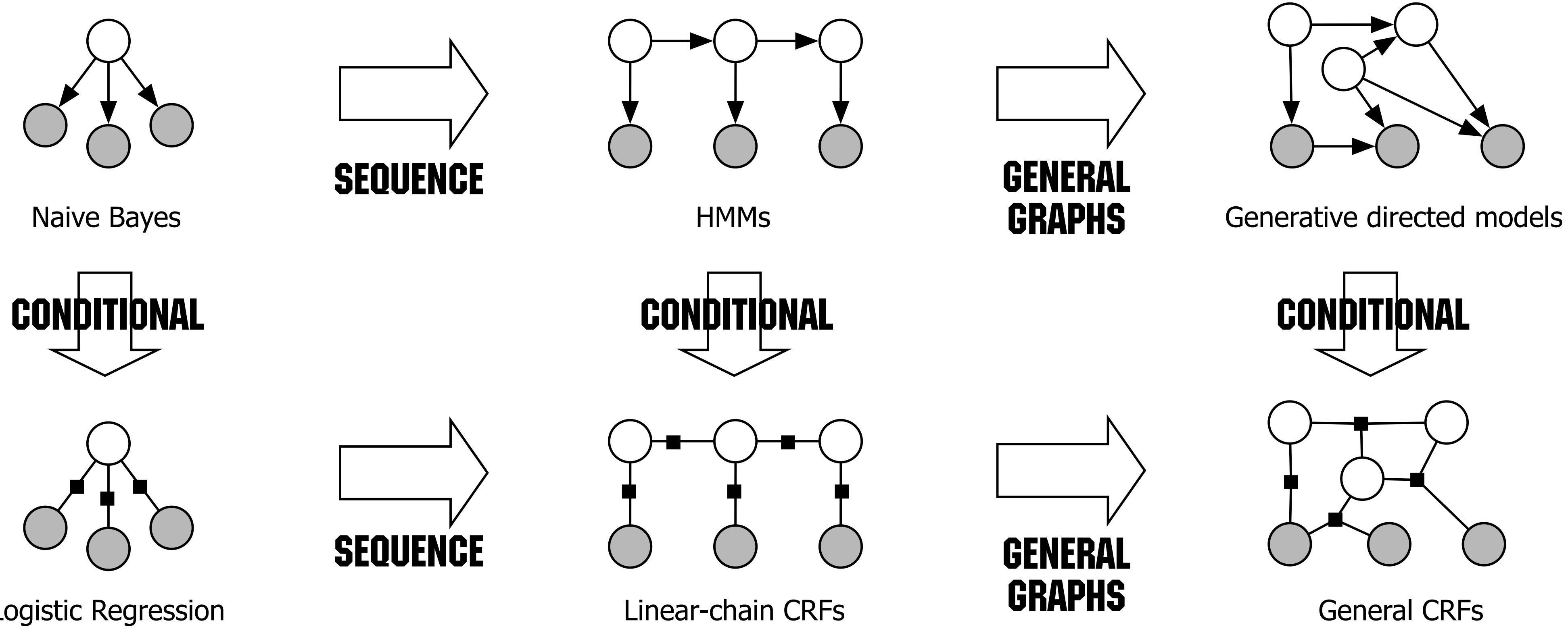
$$P(y, x_1, x_2, \ldots, x_n) = P(y) \prod_{n} P(x_n | y)$$



Probabilistic Graphical Models



Some PGM Applications

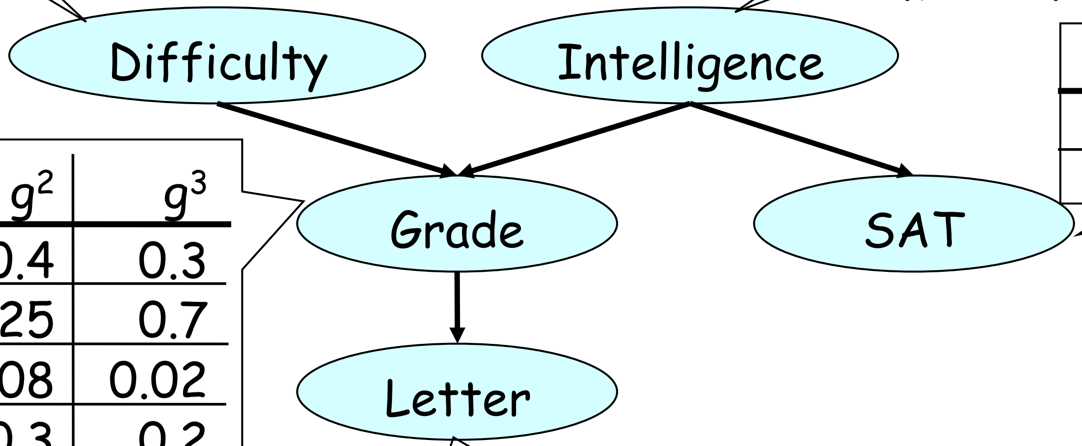


Sutton and McCallum (2012)

Conditional Probability Tables

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	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

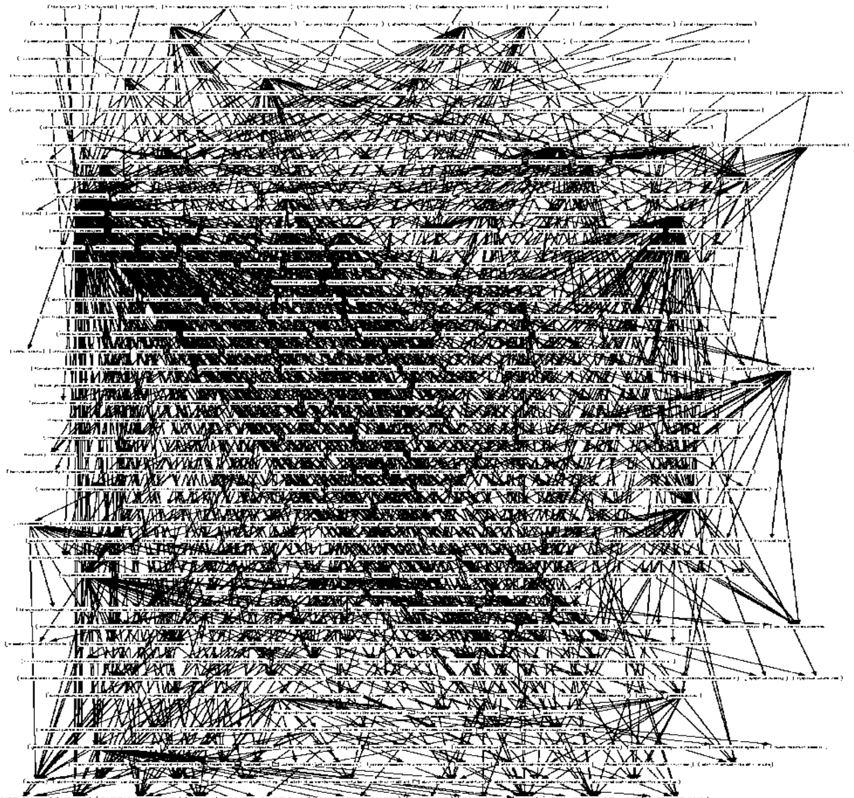
	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$P(d^0, i^1, g^3, s^1, l^1) =$

$0.6 * 0.3 * 0.02 * 0.01 * 0.8$

Daphne Koller

BNs suffer from Exponential Complexity

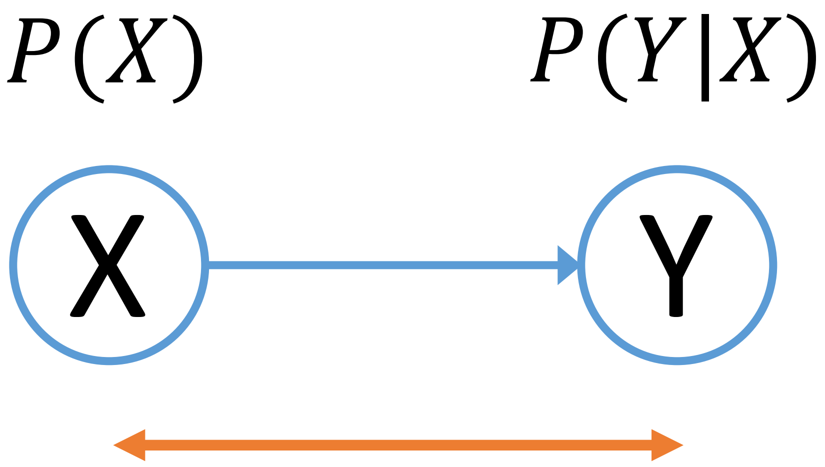


M. Pradhan, G. Provan, B. Middleton, M. Henrion, UAI 94

Observing Random Variables

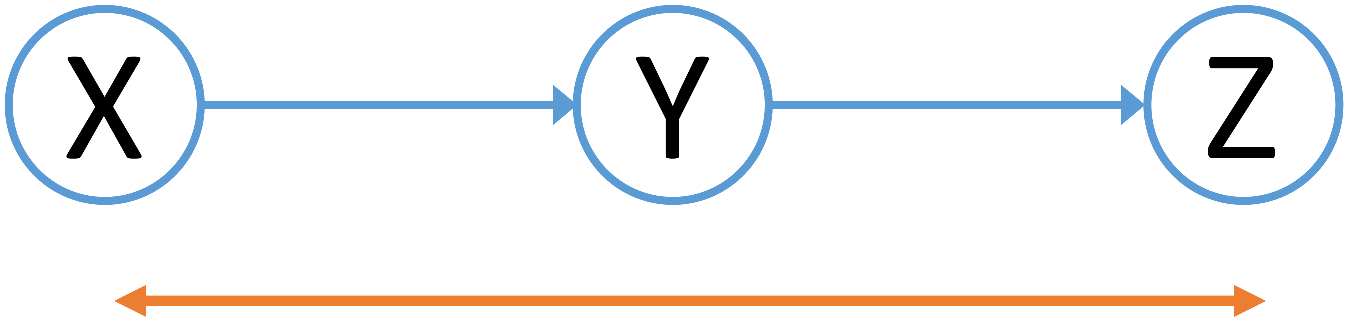
$$P(X = x, Y) = P(X = x)P(Y|X = x)$$

Observations influence belief

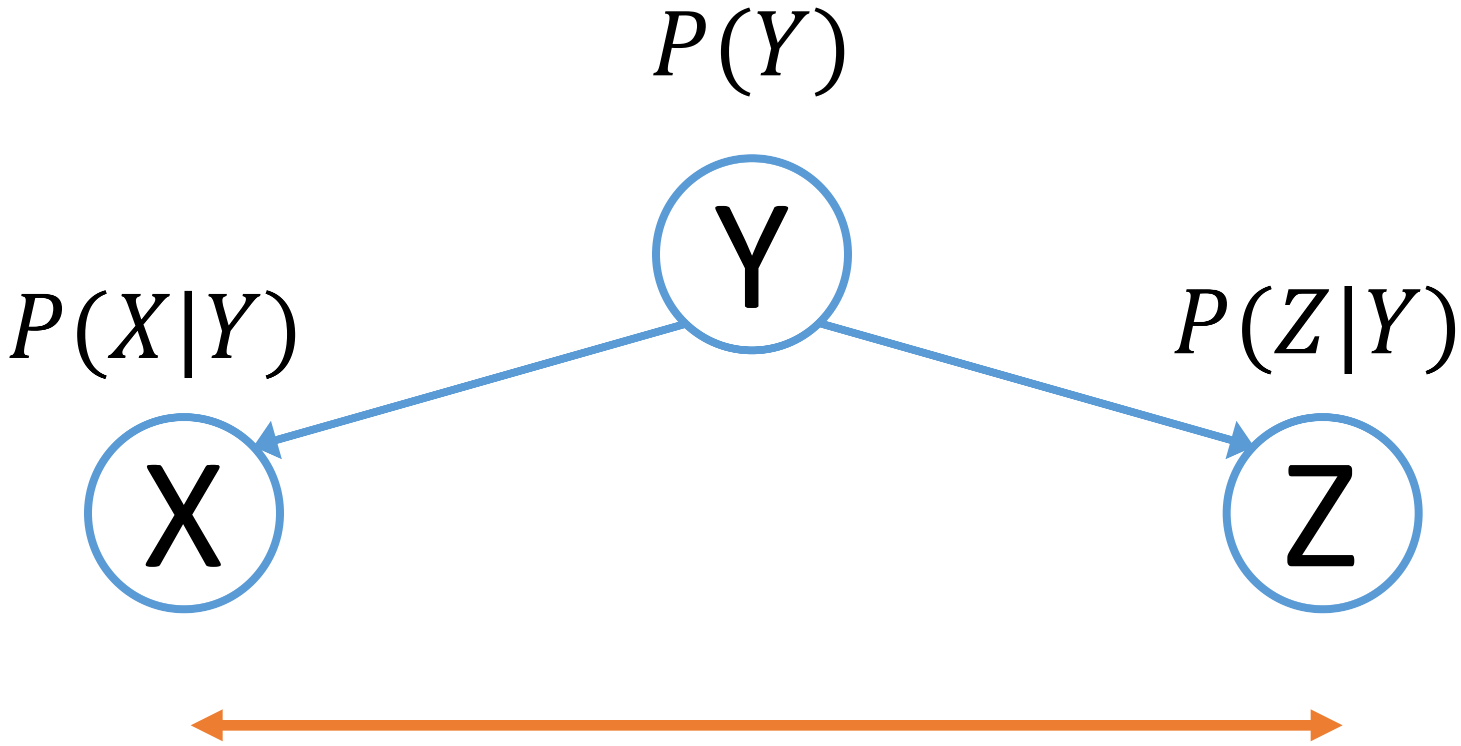


Flow of Influence follows **active trails**

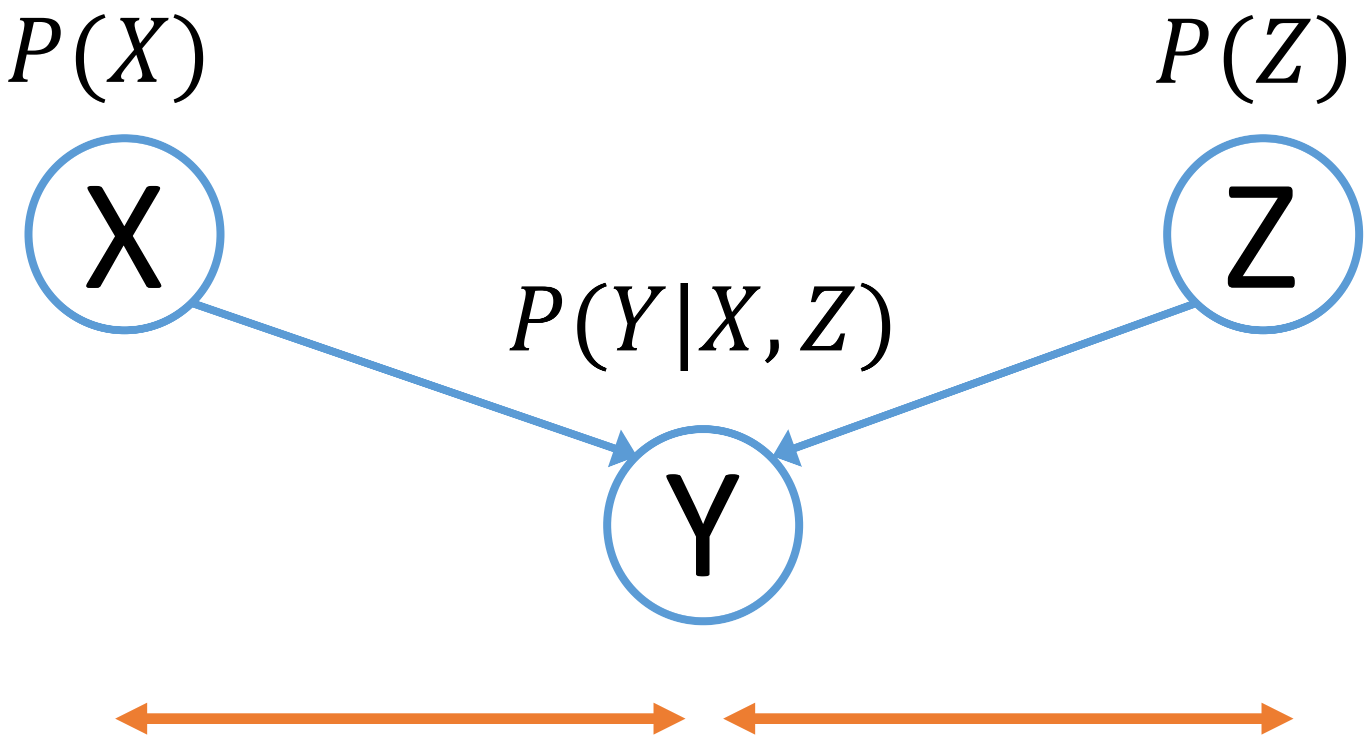
$P(X)$ $P(Y|X)$ $P(Z|Y)$



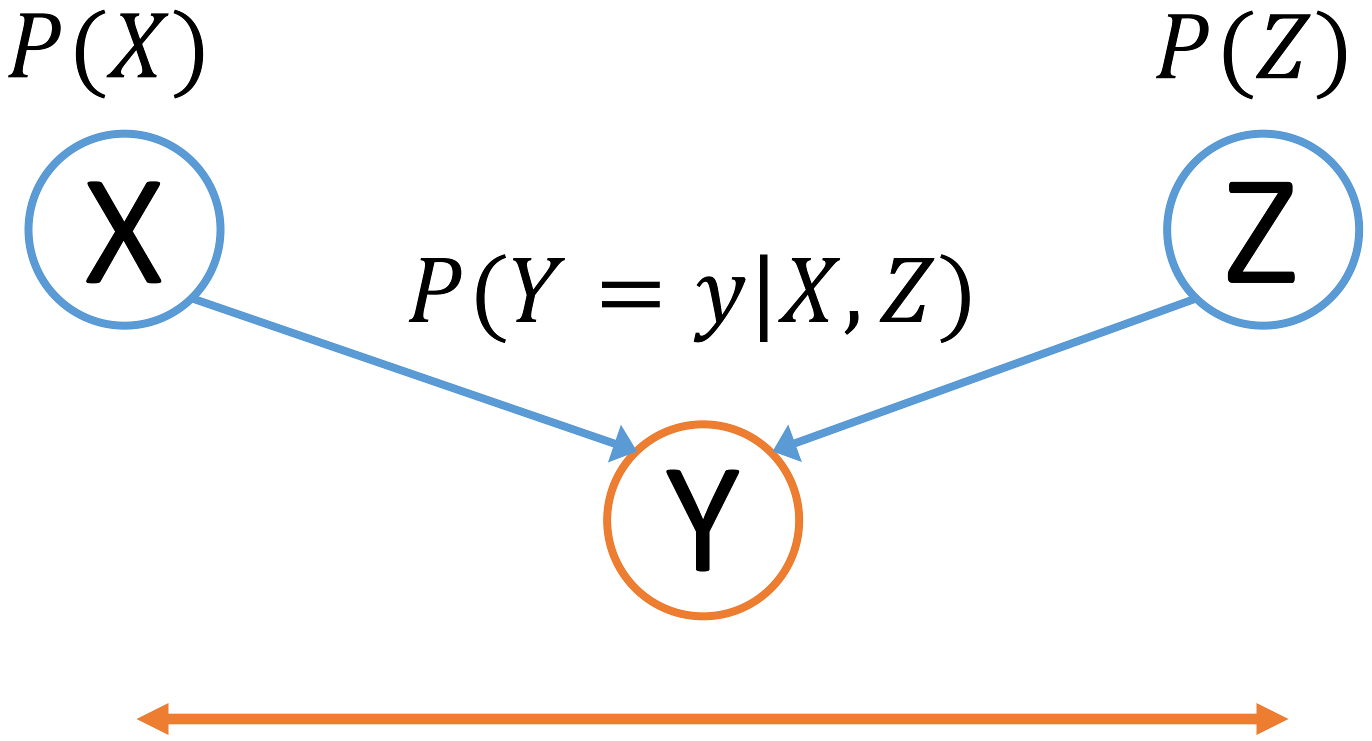
Siblings can influence each other



But independent parents cannot



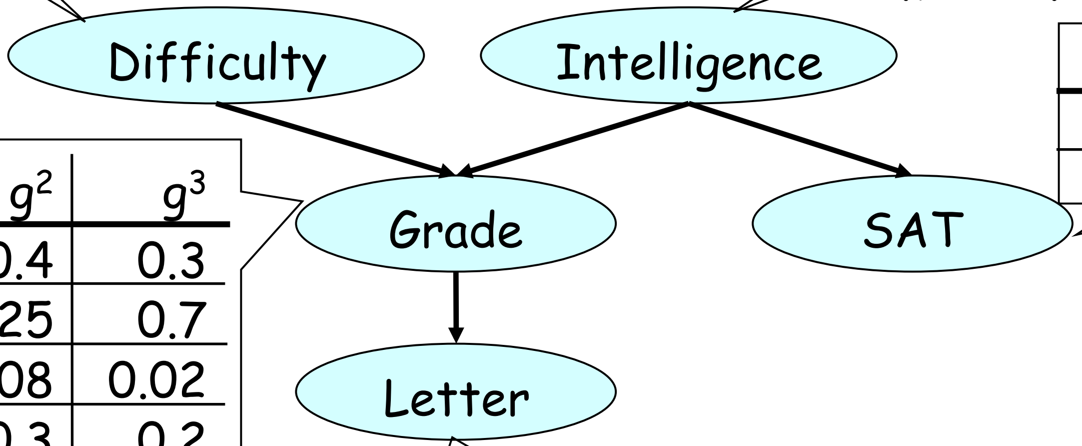
Unless we observe the child



Recomputing probabilities gets difficult

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	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

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$P(d^0, i^1, g^3, s^1, l^1) =$

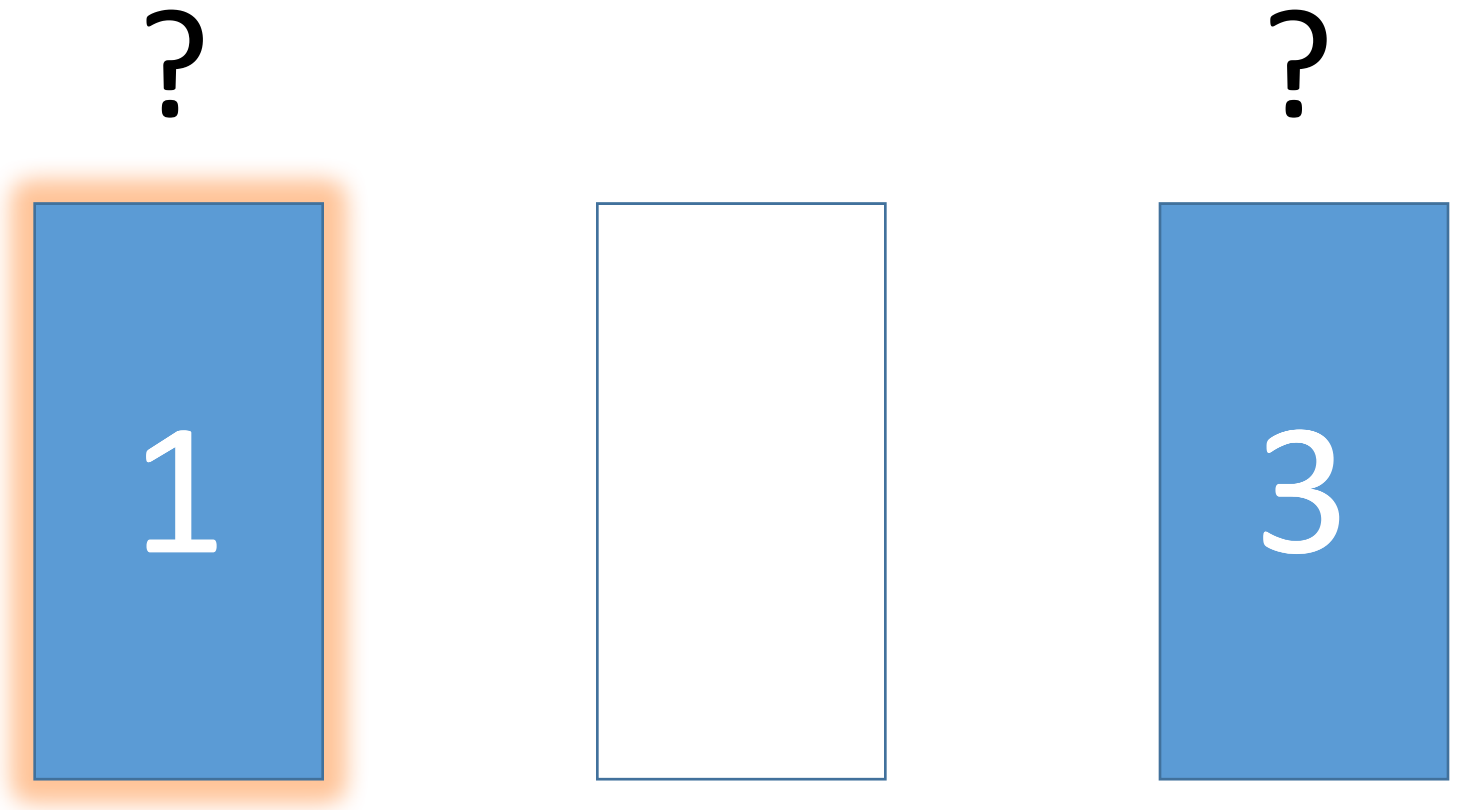
$0.6 * 0.3 * 0.02 * 0.01 * 0.8$

Daphne Koller

Factors are real-valued functions of RVs

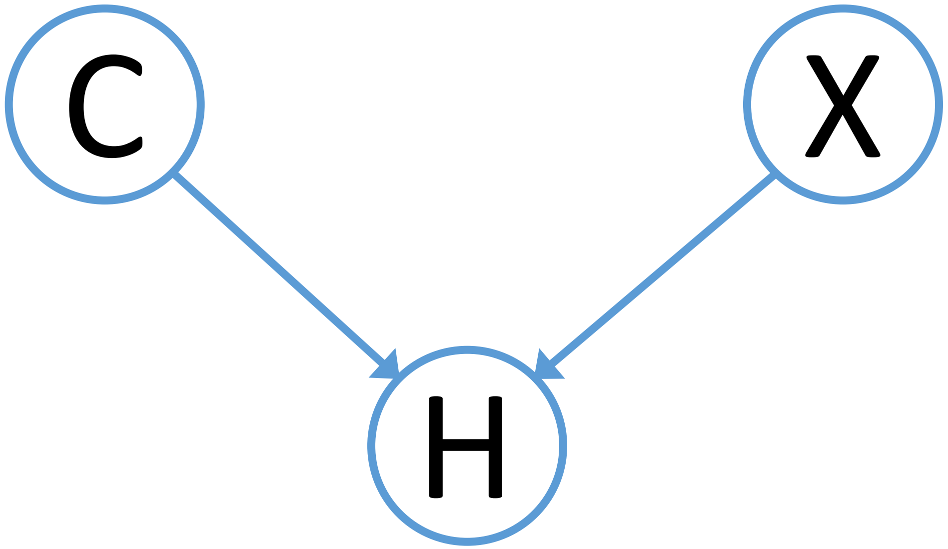
X	Y	ϕ
1	1	**0.2**
1	2	**100**
1	3	**42**
2	1	**0.7**
2	2	**0**
2	3	**9001**
3	1	**0.1**
3	2	**1**
3	3	**2**

Should the contestant switch or stay?



Monty Hall Problem: Bayesian Network

X – Prize
C – Contestant
H – Host



$$P(C, X, H) = P(C)P(X)P(H|C, X)$$

Challenge: Find $P(X|C = c, H = h)$

We completely describe the game

$\phi(C)$

C	ϕ
1	**1/3**
2	**1/3**
3	**1/3**

$\phi(X)$

X	ϕ
1	**1/3**
2	**1/3**
3	**1/3**

$\phi(C, X, H)$

C	X	H	ϕ
1	1	1	**0**
1	1	2	**1/2**
1	1	3	**1/2**
1	2	1	**0**
1	2	2	**0**
1	2	3	**1**
1	3	1	**0**
1	3	2	**1**
1	3	3	**0**

C	X	H	ϕ
2	1	1	**0**
2	1	2	**0**
2	1	3	**1**
2	2	1	**1/2**
2	2	2	**0**
2	2	3	**1/2**
2	3	1	**1**
2	3	2	**0**
2	3	3	**0**

C	X	H	ϕ
3	1	1	**0**
3	1	2	**1**
3	1	3	**0**
3	2	1	**1**
3	2	2	**0**
3	2	3	**0**
3	3	1	**1/2**
3	3	2	**1/2**
3	3	3	**0**

Switching has 67% chance of winning

$\phi(X, H, C = 1)$

C	X	H	ϕ
1	1	1	**0**
1	1	2	**1/2**
1	1	3	**1/2**
1	2	1	**0**
1	2	2	**0**
1	2	3	**1**
1	3	1	**0**
1	3	2	**1**
1	3	3	**0**

reduce

$\phi(X, C = 1, H = 2)$

C	X	H	ϕ
1	1	2	**1/2**
1	2	2	**0**
1	3	2	**1**

normalize

$P(X | C = 1, H = 2)$

X	ϕ
1	**1/3**
2	**0**
3	**2/3**

But only if we observe the contestant

$\phi(C, X, H = 2)$

C	X	H	ϕ
1	1	2	**1/2**
1	2	2	**0**
1	3	2	**1**
2	1	2	**0**
2	2	2	**0**
2	3	2	**0**
3	1	2	**1**
3	2	2	**0**
3	3	2	**1/2**

→

X	H	ϕ
1	2	**1/2 + 0 + 1**
2	2	**0 + 0 + 0**
3	2	**1 + 0 + 1/2**

marginalize

→

$\phi(X, H = 2)$

X	H	ϕ
1	2	**3/2**
2	2	**0**
3	2	**3/2**

→

$P(X|H = 2)$

X	ϕ
1	**1/2**
2	**0**
3	**1/2**

normalize

Thanks!

Slides & Notebooks

- https://github.com/bgalbraith/intro-to-bayesian-networks

Daphne Koller’s Coursera Course

- https://www.coursera.org/learn/probabilistic-graphical-models

Python Libraries

- https://github.com/jmschrei/pomegranate
- https://github.com/pgmpy/pgmpy