



Department of Biomedical Informatics

[Home](#)[About](#)[People](#)[Research](#)[Centers & Labs](#)[Training Program](#)[Publications](#)[Grants](#)[News](#)[Events](#)[Jobs](#)

Xia Jiang, PhD

Research Interests:

Application of AI, Machine Learning, and Bayesian Networks in biomedical domain, clinical informatics, causal learning, prediction and decision support, biomarker/risk factors discovery via learning from data, design and development of computational methods/algorithms, and cancer and translational informatics.

Associate Professor, Department of Biomedical Informatics

University of Pittsburgh School of Medicine

Faculty, Pitt Intelligent Systems Program

Faculty, CMU-Pitt Computational and Systems Biology PhD Program

Publications:



About MBIL

Xia Jiang, PhD

Definition of Bayesian Network

Suppose we have the following:

1. a directed acyclic graph (DAG) \mathbb{G} whose nodes are random variables.
2. A joint probability distribution P of the variables in \mathbb{G} .
3. Each node is conditionally independent of its nondescendants given its parents.

Then we say that \mathbb{G} and P satisfy the *Markov condition* and we call (\mathbb{G}, P) a *Bayesian network*.

BN is a probability based graphic model that represents conditional independent (dependent) relationships among variables.

Markov Condition

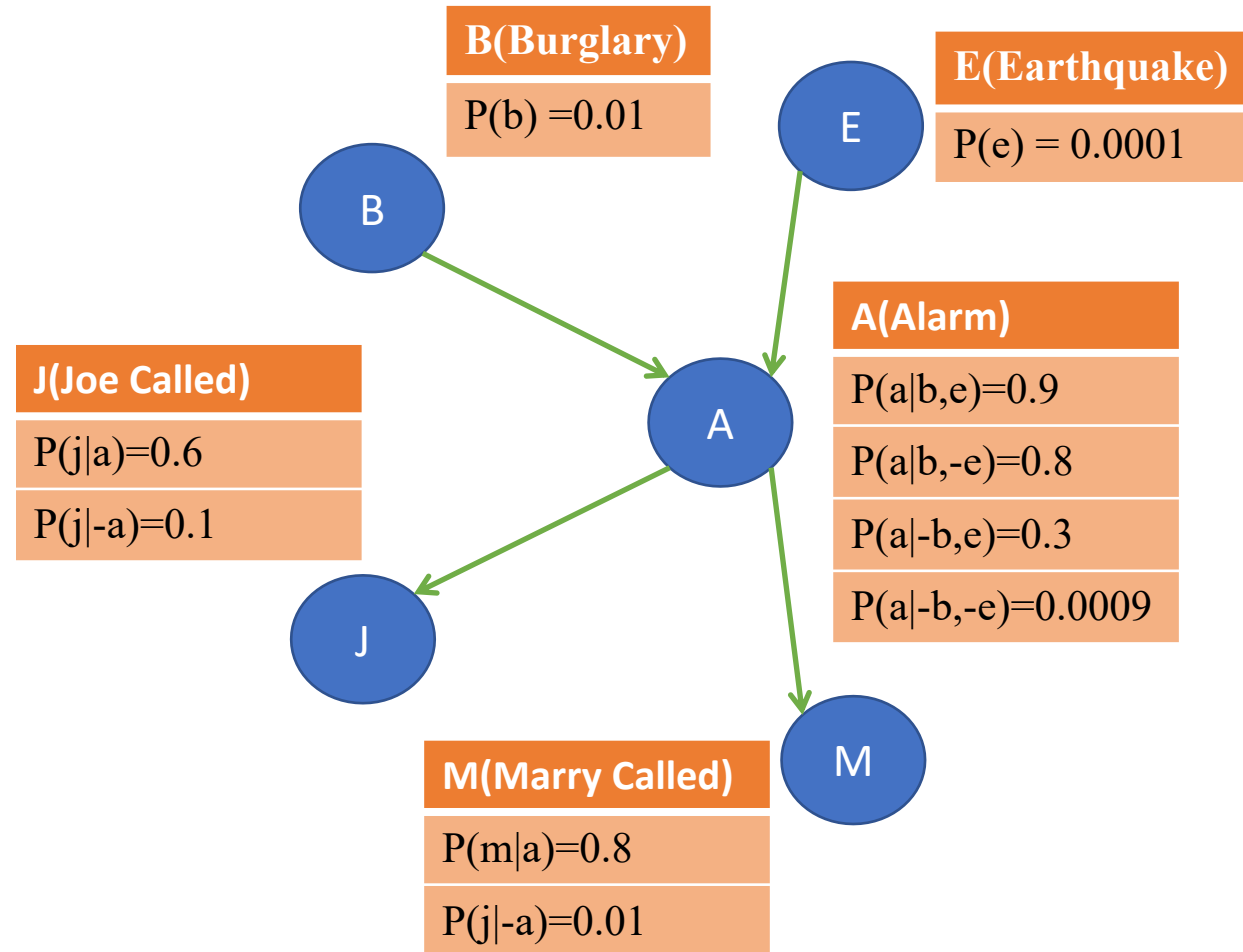
Suppose we have a joint probability distribution P of the random variables in some set V and a DAG $G = (V, E)$. We say that (G, P) satisfies the Markov condition if for each variable $X \in V$, $\{X\}$ is independent of the set of all its other nondescendants given the set of all its parents

$$I_G(\{X\}, \text{ND}_X \mid \text{PA}_X)$$

$$I_G(\{X\}, \{Y\} \mid Z) \Rightarrow I_P(\{X\}, \{Y\} \mid Z)$$



We spent quite some time to discuss $I_P(\{X\}, \{Y\} \mid Z)$ in the first lecture !



Learning BN Structure from Data : The Score-based Approach

- Model selection is to select a "good" causal model from among all possible models, and use it as if it were the correct model.
- When we apply model selection to causal Bayesian network structure learning, how do we decide whether or not a model is "good"?
- The answer is that we can use a Bayesian scoring criterion to measure the relative strength of a model.

A Bayesian Score Criterion

- A Bayesian score criterion for a DAG is a function that assigns a value to the DAG under consideration based on data.
- Assuming G is a variable that represents a possible DAG, $P(data|G)$ is a Bayesian score criterion that we can use in our model selection approach of learning BN structures.
- In this class, we will introduce two formulas for obtaining $P(data|G)$, that is, the K2 Score and the BDeu Score.

The Probability of a DAG given Data

Assuming *Data* is a variable that represents our observed data, assuming G_1 is a variable that represents a possible DAG of the “true” model, G_2 is a variable that represents another possible DAG of the “true” model, then

if $P(G_1|Data) > P(G_2|Data)$

We can say that between G_1 and G_2 , G_1 is more likely to be the DAG of the “true” model that is responsible for the data we observed.



Why a
Bayesian score
criterion can be
a measure of
the strength of a
Model?

The Probability of a DAG given Data

Bayes' Rule

$$P(G_1|Data) = \frac{P(Data|G_1) \times P(G_1)}{P(Data)}$$

$$P(G_2|Data) = \frac{P(Data|G_2) \times P(G_2)}{P(Data)}$$

The Probability of a DAG given Data

$$P(G_1|Data) > P(G_2|Data)$$



Likelihood
Function

$$\frac{P(Data|G_1) \times P(G_1)}{P(Data)} > \frac{P(Data|G_2) \times P(G_2)}{P(Data)}$$

The Probability of a DAG given Data

$$P(G_1|Data) > P(G_2|Data)$$

$$\frac{P(Data|G_1) \times P(G_1)}{P(Data)} > \frac{P(Data|G_2) \times P(G_2)}{P(Data)}$$

Normalization
Constant

***P(Data)* can be
treated as a
constant in our
comparison.**



The Probability of a DAG given Data

$$P(G_1|Data) > P(G_2|Data)$$

$$\frac{P(Data|G_1) \times P(G_1)}{P(Data)} > \frac{P(Data|G_2) \times P(G_2)}{P(Data)}$$

Prior Probability
of the DAG
structure

So $P(Data|G)$ is
used as the score
criterion

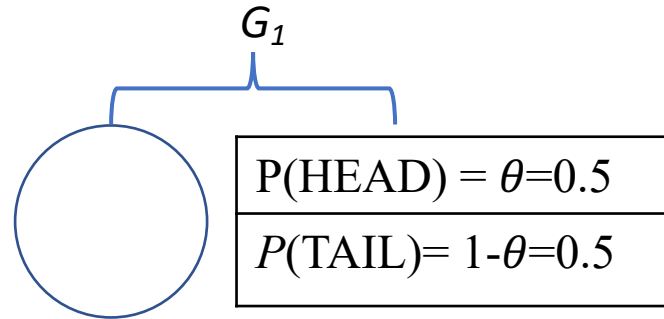
When we have no prior knowledge about the DAG structures, we can assume that $P(G_1)$ and $P(G_2)$ are equal. In that case, can you think of a principle that we previously covered that allows us to make this assumption?



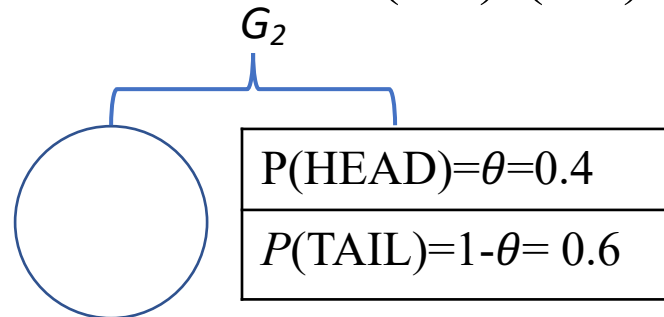
How do We Compute $P(Data|G)$?

Example

HEAD
HEAD
TAIL
TAIL
TAIL
HEAD
TAIL



$$P(Data | \theta, G_1) = \theta^3 (1 - \theta)^4 \\ = (0.5)^3 (0.5)^4$$



$$P(Data | \theta, G_1) = (0.4)^3 (0.6)^4$$



IID stands for
Independently obtained
and following **I**dentical
Distribution.




Data should be IID!

How do We Compute $P(Data|G)$? A More general case

Assuming $Data$ is a variable that represents our data concerning one **single multinomial variable X** that has r different values, assuming G is a variable that represents a possible DAG of the “true” model, assuming IID, how do we compute $P(Data|G)$?

$$\theta = \{\theta_1, \dots, \theta_r\}$$


$$P(Data|G) = \int P(Data, \theta | G) d\theta = \int P(Data | \theta, G) P(\theta | G) d\theta = \int P(Data | \theta) P(\theta) d\theta$$


What is this based on?



How do We Compute $P(Data|G)$? A More general case

Assuming $Data$ is a variable that represents our data concerning one **single multinomial variable X** that has r different values, assuming G is a variable that represents a possible DAG of the “true” model, assuming IID, how do we compute $P(Data|G)$?

$$\theta = \{\theta_1, \dots, \theta_r\}$$

$$P(Data|G) = \int P(Data, \theta | G) d\theta = \int P(Data | \theta, G) P(\theta | G) d\theta = \int P(Data | \theta) P(\theta) d\theta$$


What is this based on?



How do we compute $P(Data|G)$? A More general case

Assuming $Data$ is a variable that represents our data concerning one **single multinomial variable X** that has r different values, assuming G is a variable that represents a possible DAG of the “true” model, assuming IID, how do we compute $P(Data|G)$?

$$P(Data|\boldsymbol{\theta}) = \prod_{k=1}^r \theta_k^{N_k}$$

N_k is the count of the k^{th} value in Data

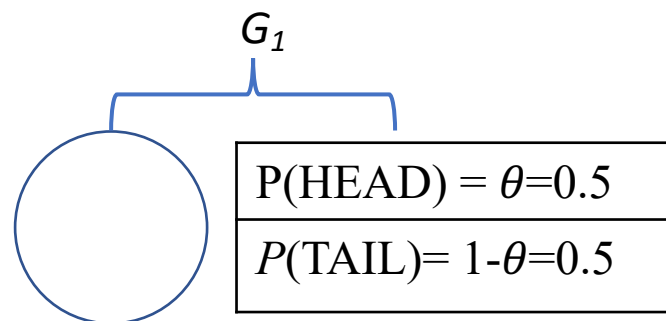
***True for IID Data
and single
multinomial variable
!***



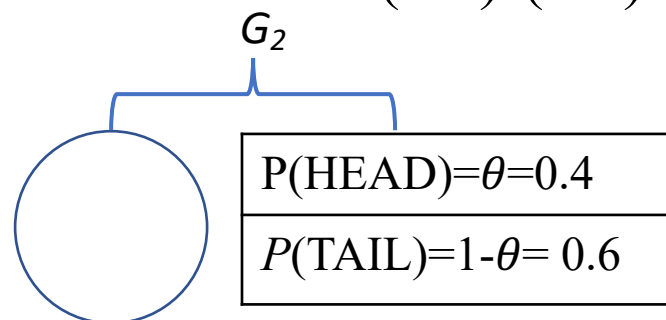
How do we compute $P(Data|G)$?

Example

HEAD
HEAD
TAIL
TAIL
TAIL
HEAD
TAIL



$$P(Data|\theta, G_1) = \theta^3 (1-\theta)^4$$
$$= (0.5)^3 (0.5)^4$$



$$P(Data|\theta, G_2) = (0.4)^3 (0.6)^4$$



IID stands for
Independently obtained
and following **I**dentical
Distribution.

Data should be IID!



How do we compute $P(Data|G)$? A More general case

Assuming $Data$ is a variable that represents our data concerning **one single multinomial variable X** that has r different values, assuming G is a variable that represents a possible DAG of the “true” model, assuming IID, how do we compute $P(Data|G)$?

$$\boldsymbol{\theta} = \{\theta_1, \dots, \theta_r\}$$

$$\alpha = \sum_{k=1}^r \alpha_k$$

$$P(\boldsymbol{\theta}) = Dir(\boldsymbol{\theta} | \alpha_1, \dots, \alpha_r) = \frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r \theta_k^{\alpha_k - 1}$$

$$P(Data | \boldsymbol{\theta}) = \prod_{k=1}^r \theta_k^{N_k}$$

N_k is the count of the k^{th} value in Data

True when G contains a single multinomial variable and IID data!

$$P(Data | G) = \int P(Data | \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \frac{\prod_{k=1}^r \Gamma(\alpha_k + N_k)}{\Gamma(\sum_{k=1}^r (\alpha_k + N_k))}$$

$$= \frac{\Gamma(\alpha)}{\Gamma(\sum_{k=1}^r (\alpha_k + N_k))} \prod_{k=1}^r \frac{\Gamma(\alpha_k + N_k)}{\Gamma(\alpha_k)}$$

How do we compute $P(Data|G)$? A More general case

Assuming $Data$ is a variable that represents our data concerning **one single multinomial variable X** that has r different values, assuming G is a variable that represents a possible DAG of the “true” model, assuming IID, how do we compute $P(Data|G)$?

$$\boldsymbol{\theta} = \{\theta_1, \dots, \theta_r\}$$

$$\alpha = \sum_{k=1}^r \alpha_k$$

$$P(\boldsymbol{\theta}) = Dir(\boldsymbol{\theta} | \alpha_1, \dots, \alpha_r) = \frac{\Gamma(\alpha)}{\prod_{k=1}^r \Gamma(\alpha_k)} \prod_{k=1}^r \theta_k^{\alpha_k - 1}$$

$$P(Data | \boldsymbol{\theta}) = \prod_{k=1}^r \theta_k^{N_k}$$

N_k is the count of the k^{th} value in Data

$$P(Data | G) = \int P(Data | \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta} = \frac{\Gamma(\alpha)}{\Gamma(\sum_{k=1}^r (\alpha_k + N_k))} \prod_{k=1}^r \frac{\Gamma(\alpha_k + N_k)}{\Gamma(\alpha_k)}$$

$$B(s, t) = \int x^{s-1} (1-x)^{t-1} dx, \quad s > 0, t > 0$$

$$B(s, t) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$$

$$\Gamma(n) = (n-1)!, \quad \text{when } n \geq 1 \text{ and } n \text{ is an integer}$$

True when G contains a single multinomial variable and IID data!

The Bayesian score

$$score_{Bayes}(G : Data) = P(Data | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} a_{ijk})}{\Gamma(\sum_{k=1}^{r_i} a_{ijk} + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})},$$

where r_i is the number of states of X_i , q_i is the number of different instantiations of the parents of X_i , a_{ijk} is the ascertained prior belief concerning the number of times X_i took its k th value when the parents of X_i had their j th instantiation, and s_{ijk} is the number of times in the data that X_i took its k th value when the parents of X_i had their j th instantiation.



Recall we have previously derived this for the single variable case.

K2 Score

$$score_{Bayes}(G : Data) = P(Data | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} a_{ijk})}{\Gamma(\sum_{k=1}^{r_i} a_{ijk} + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})}$$

$$score_{K2}(G : Data) = P(Data | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(r_i)}{\Gamma(r_i + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \Gamma(1 + s_{ijk})$$

In the K2 score

$$a_{ijk} = 1 \quad \text{for all } i, j, k.$$

BDeu Score

$$score_{Bayes}(G : Data) = P(Data | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\sum_{k=1}^{r_i} a_{ijk})}{\Gamma(\sum_{k=1}^{r_i} a_{ijk} + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + s_{ijk})}{\Gamma(a_{ijk})}$$

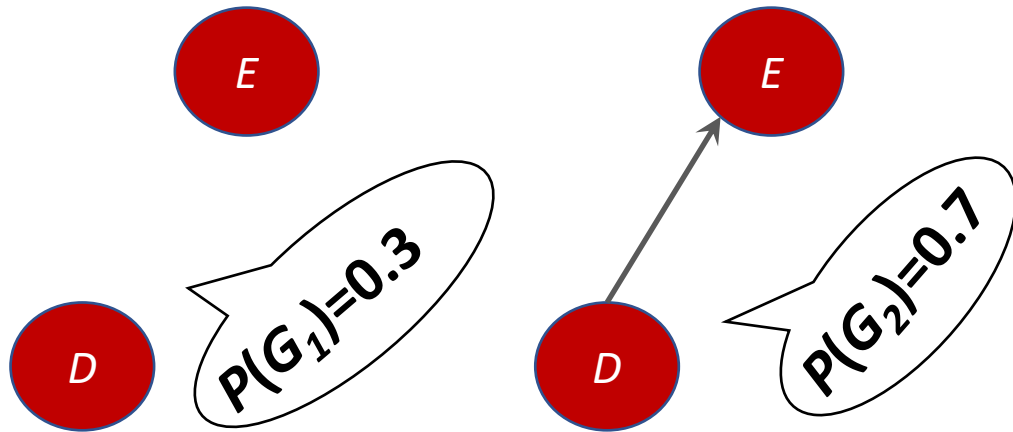
For the BDeu score, the values of the hyperparameters a_{ijk} are calculated from a single parameter α called the *prior equivalent sample size*. If we want to use a prior equivalent sample size α and represent a prior uniform distribution for each variable in the network, for all i , j , and k we set $a_{ijk} = \alpha / r_i q_i$.

$$score_{\alpha}(G : Data) = P(Data | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha / q_i)}{\Gamma(\alpha / q_i + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha / r_i q_i + s_{ijk})}{\Gamma(\alpha / r_i q_i)}.$$

Example

D	E
T	F
T	T
F	T
T	T
F	F

Consider a binary disease variable D and a binary symptom E . Both variables have the values T (present) and F (absent). Suppose a physician, Dr. Roberts, believes that the following two causal Bayesian network structures G_1 and G_2 are possible, with the prior probabilities shown. Use the BDeu score with $\alpha=4$ to learn from the given data the most probable DAG structure, that is to obtain and compare the $P(G_1 \mid \text{data})$ and $P(G_2 \mid \text{data})$.



Example

<i>D</i>	<i>E</i>
T	F
T	T
F	T
T	T
F	F

$$\text{score}_\alpha(G : \text{Data}) = P(\text{Data} | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha / q_i)}{\Gamma(\alpha / q_i + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha / r_i q_i + s_{ijk})}{\Gamma(\alpha / r_i q_i)}.$$

What is the value of $P(\text{data} | G_1)$?

$\alpha = 4$

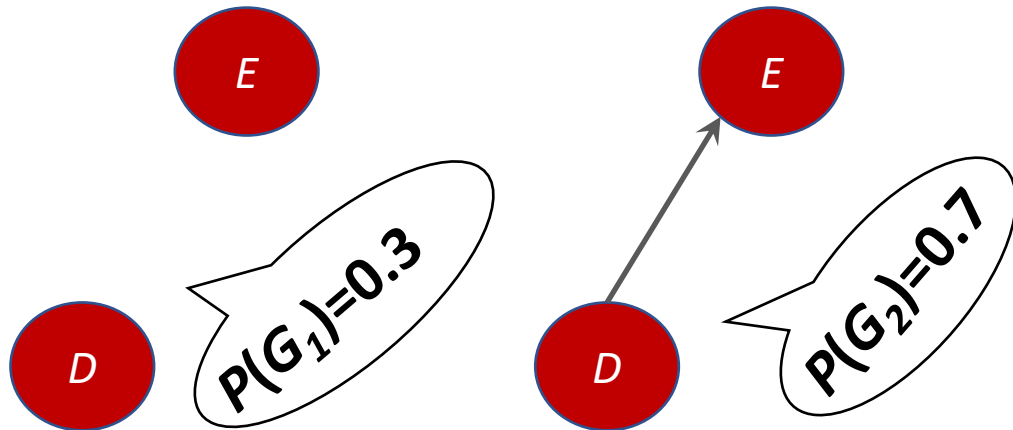
Term for Node D

$$\frac{\Gamma(\frac{4}{1})}{\Gamma(\frac{4}{1}+5)} \frac{\Gamma(\frac{4}{2 \times 1}+3)\Gamma(\frac{4}{2 \times 1}+2)}{\Gamma(\frac{4}{2 \times 1})\Gamma(\frac{4}{2 \times 1})} = \frac{3!}{8!} \frac{4!3!}{1!1!} = \frac{3 \times 2 \times 3 \times 2}{8 \times 7 \times 6 \times 5} = \frac{3}{140}$$

$$\frac{\Gamma(\frac{4}{1})}{\Gamma(\frac{4}{1}+5)} \frac{\Gamma(\frac{4}{2 \times 1}+3)\Gamma(\frac{4}{2 \times 1}+2)}{\Gamma(\frac{4}{2 \times 1})\Gamma(\frac{4}{2 \times 1})} = \frac{3!}{8!} \frac{4!3!}{1!1!} = \frac{3 \times 2 \times 3 \times 2}{8 \times 7 \times 6 \times 5} = \frac{3}{140}$$

Term for Node E

$$P(\text{Data} | G_1) = \frac{9}{19600} = 4.5918 \times 10^{-4}$$

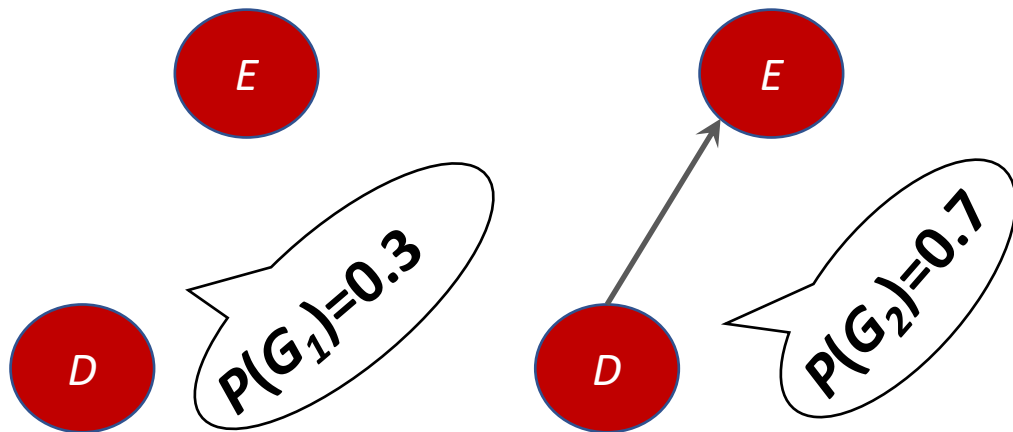


<i>D</i>	<i>E</i>
T	F
T	T
F	T
T	T
F	F

$$score_{\alpha}(G : Data) = P(Data | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha / q_i)}{\Gamma(\alpha / q_i + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha / r_i q_i + s_{ijk})}{\Gamma(\alpha / r_i q_i)}.$$

What is the value of $P(data | G_2)$?

$\alpha = 4$



<i>D</i>	<i>E</i>
T	F
T	T
F	T
T	T
F	F

$$\text{score}_\alpha(G : \text{Data}) = P(\text{Data} | G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\alpha / q_i)}{\Gamma(\alpha / q_i + \sum_{k=1}^{r_i} s_{ijk})} \prod_{k=1}^{r_i} \frac{\Gamma(\alpha / r_i q_i + s_{ijk})}{\Gamma(\alpha / r_i q_i)}.$$

What is the value of $P(\text{data} | G_2)$

$$\alpha = 4$$

Term for Node D

$$\frac{\Gamma(\frac{4}{1})}{\Gamma(\frac{4}{1}+5)} \frac{\Gamma(\frac{4}{2 \times 1}+3)\Gamma(\frac{4}{2 \times 1}+2)}{\Gamma(\frac{4}{2 \times 1})\Gamma(\frac{4}{2 \times 1})} = \frac{3! \ 4!3!}{8! \ 1!1!} = \frac{3 \times 2 \times 3 \times 2}{8 \times 7 \times 6 \times 5} = \frac{3}{140}$$

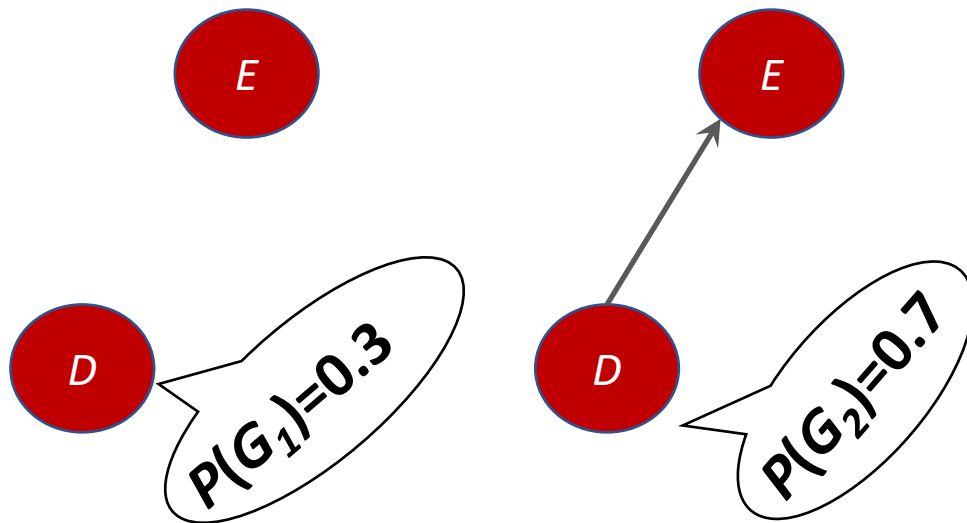
$$\frac{\Gamma(\frac{4}{2})}{\Gamma(\frac{4}{2}+3)} \frac{\Gamma(\frac{4}{2 \times 2}+2)\Gamma(\frac{4}{2 \times 2}+1)}{\Gamma(\frac{4}{2 \times 2})\Gamma(\frac{4}{2 \times 2})} = \frac{1! \ 2!1!}{4! \ 0!0!} = \frac{1}{12}$$

$$\frac{\Gamma(\frac{4}{2})}{\Gamma(\frac{4}{2}+2)} \frac{\Gamma(\frac{4}{2 \times 2}+1)\Gamma(\frac{4}{2 \times 2}+1)}{\Gamma(\frac{4}{2 \times 2})\Gamma(\frac{4}{2 \times 2})} = \frac{1! \ 1!1!}{3! \ 0!0!} = \frac{1}{6}$$

Term for Node E when $j=1$, ie, $D=T$

Term for Node E when $j=2$, ie, $D=F$

$$P(\text{Data} | G_2) = \frac{3}{12 \times 6 \times 140} = 2.976 \times 10^{-4}$$



Example

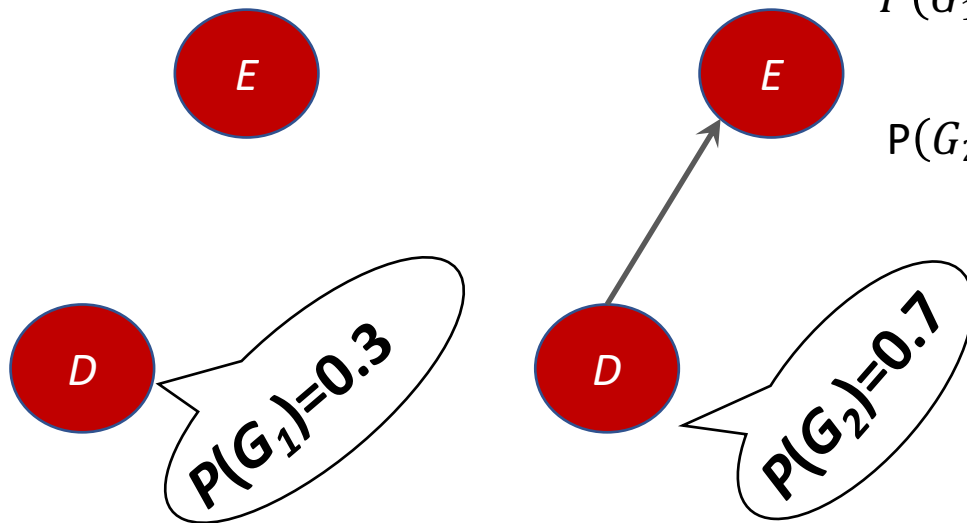
<i>D</i>	<i>E</i>
T	F
T	T
F	T
T	T
F	F

What is the $P(G_1 | \text{data})$ and what is $P(G_2 | \text{data})$, using the likelihoods that we just derived?

$$P(\text{Data}) = \sum_{i=1}^2 P(\text{Data}, G_i) = \sum_{i=1}^2 P(\text{Data}|G_i)P(G_i) = 4.5918 \times 10^{-4} \times 0.3 + 2.976 \times 10^{-4} \times 0.7$$

$$P(G_1 | \text{Data}) = \frac{P(\text{Data}|G_1)P(G_1)}{P(\text{Data})} = \frac{4.5918 \times 10^{-4} \times 0.3}{4.5918 \times 10^{-4} \times 0.3 + 2.976 \times 10^{-4} \times 0.7} = 0.398$$

$$P(G_2 | \text{Data}) = 1 - 0.398 = 0.602$$



$$P(\text{Data}|G_1) = \frac{9}{19600} = 4.5918 \times 10^{-4}$$

$$P(\text{Data}|G_2) = \frac{3}{12 \times 6 \times 140} = 2.976 \times 10^{-4}$$



Thought about Bayesian priors?

How Big is the DAG Search Space?

The number of DAGs containing n nodes can be computed using the following recurrence:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i)$$

$n > 2$
 $f(0)=1$
 $f(1)=1$

n	$f(n)$
1	1
2	3
3	

$$f(3) = \sum_{i=1}^3 (-1)^{i+1} \binom{3}{i} 2^{i(3-i)} f(3-i) = 25$$

RW, Robinson “Continuing Unlabeled Acyclic Digraphs,” in Little, C.H.C. (Ed): *Lecture Notes in Mathematics* Springer-Verlag, 1977

How Big is the DAG Search Space?

The number of DAGs containing n nodes can be computed using the following recurrence:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i)$$

$n > 2$
 $f(0)=1$
 $f(1)=1$

n	$f(n)$
1	1
2	3
3	25
4	

$$f(3) = \sum_{i=1}^3 (-1)^{i+1} \binom{3}{i} 2^{i(3-i)} f(3-i) = 25$$

$$f(4) = \sum_{i=1}^4 (-1)^{i+1} \binom{4}{i} 2^{i(4-i)} f(4-i) = 544$$

RW, Robinson “Continuing Unlabeled Acyclic Digraphs,” in Little, C.H.C.
 (Ed): *Lecture Notes in Mathematics* Springer-Verlag, 1977

How Big is the DAG Search Space?

The number of DAGs containing n nodes can be computed using the following recurrence:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i)$$

$n > 2$
 $f(0)=1$
 $f(1)=1$

n	$f(n)$
1	1
2	3
3	25
4	544
5	

$$f(3) = \sum_{i=1}^3 (-1)^{i+1} \binom{3}{i} 2^{i(3-i)} f(3-i) = 25$$

$$f(4) = \sum_{i=1}^4 (-1)^{i+1} \binom{4}{i} 2^{i(4-i)} f(4-i) = 544$$

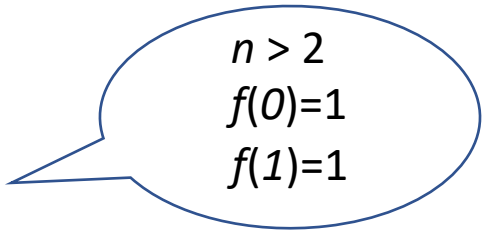
$$f(5) = \sum_{i=1}^5 (-1)^{i+1} \binom{5}{i} 2^{i(5-i)} f(5-i) = 29000$$

RW, Robinson “Continuing Unlabeled Acyclic Digraphs,” in Little, C.H.C.
 (Ed): *Lecture Notes in Mathematics* Springer-Verlag, 1977

How Big is the DAG Search Space?

The number of DAGs containing n nodes can be computed using the following recurrence:

$$f(n) = \sum_{i=1}^n (-1)^{i+1} \binom{n}{i} 2^{i(n-i)} f(n-i)$$


$$\begin{aligned} n &> 2 \\ f(0) &= 1 \\ f(1) &= 1 \end{aligned}$$

n	$f(n)$
1	1
2	3
3	25
4	544
5	29000

$$f(3) = \sum_{i=1}^3 (-1)^{i+1} \binom{3}{i} 2^{i(3-i)} f(3-i) = 25$$

$$f(4) = \sum_{i=1}^4 (-1)^{i+1} \binom{4}{i} 2^{i(4-i)} f(4-i) = 544$$

$$f(5) = \sum_{i=1}^5 (-1)^{i+1} \binom{5}{i} 2^{i(5-i)} f(5-i) = 29000$$

...

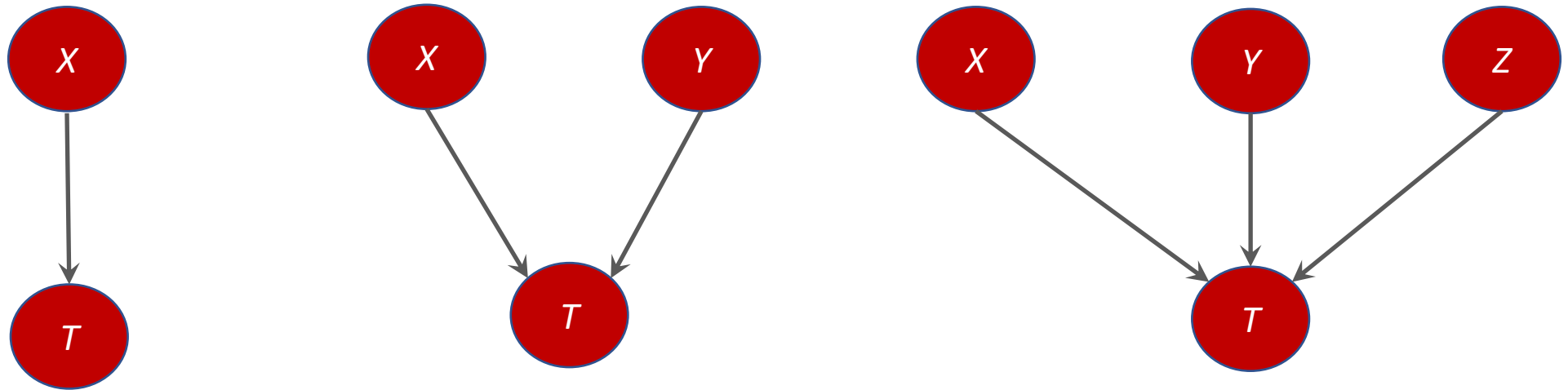
$$f(10) > 10^{18}$$

MBIL : A Score-based Approach to Learn A Causal DAG that Contains One Outcome/Effect Node and its Parents

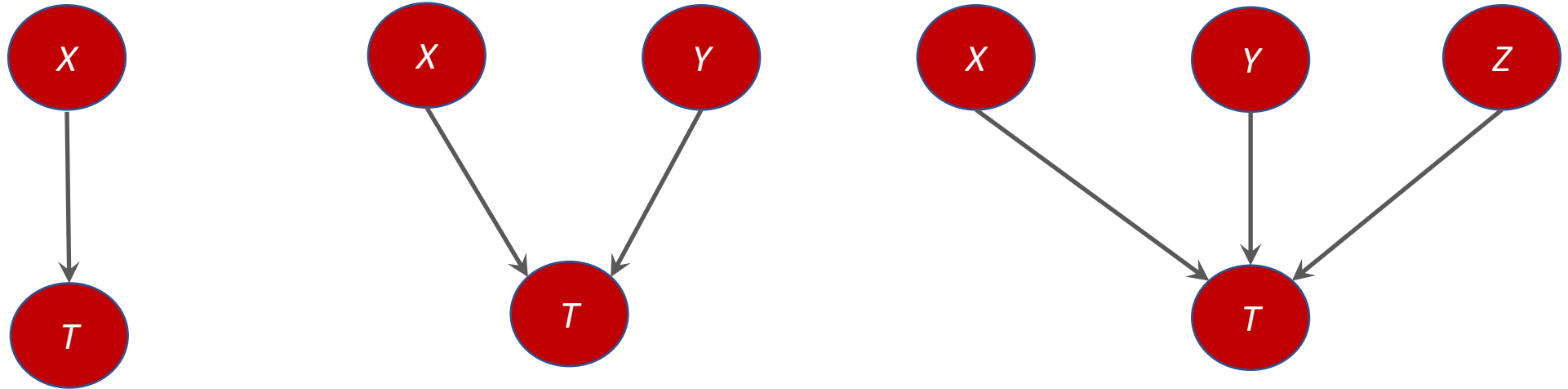
- MBIL stands for the *Markov Blanket and Interactive risk factor Learner*.
- MBIL learns single and interactive risk factors having a direct influence on a patient's outcome.
- The algorithm does not merely find correlation, but rather identifies direct interactive risk factors.

Examples of Models Learned by MBIL

Examples are as follows.



Examples of Interactive Models Learned by MBIL



Markov Boundary

Let V be a set of random variables, P be their joint probability distribution, and X is in V , then a **Markov boundary** of X is any Markov blanket of X such that none of its proper subsets is a Markov blanket of X .

In a Bayesian Network, for each variable X in the set of all variables V , the set of all **parents** of X , **children** of X , and **parents of children** of X is a Markov blanket and a Markov boundary of X

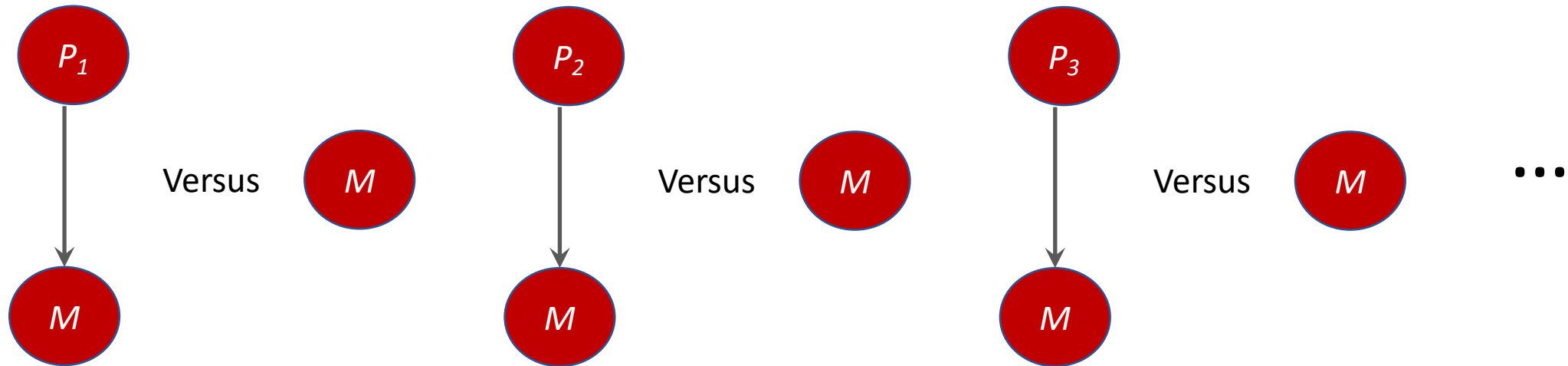
Markov Blanket

Let (\mathbb{G}, P) represent a Bayesian network, then a **Markov blanket** M_X of a node X is any set of variables such that X is conditionally independent of all the other nodes given M_X .

$$I_{\mathbb{G}}(\{X\}, V - (\{X\} \cup M_X) \mid M_X)$$

MBIL Procedures

1. *determine_single_predictors* learns the set $PA1$ of single-variable risk factors of a target T . The procedure does this by determining whether the BDeu score of the model in which the variable has an edge to T is greater than the BDeu score of the model in which T has no parents.



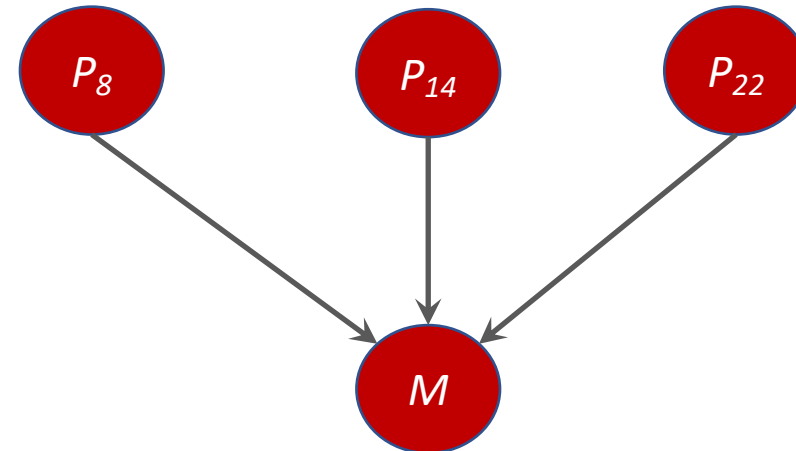
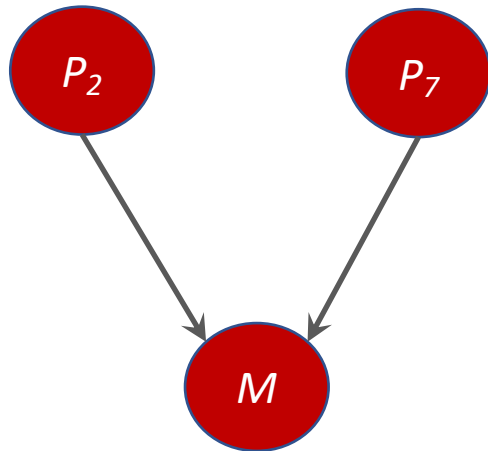
MBIL Procedures

1. *determine_single_predictors* learns the set $PA1$ of single-variable risk factors of a target T . The procedure does this by determining whether the BDeu score of the model in which the variable has an edge to T is greater than the BDeu score of the model in which T has no parents. For example, we learned that,



MBIL Procedures

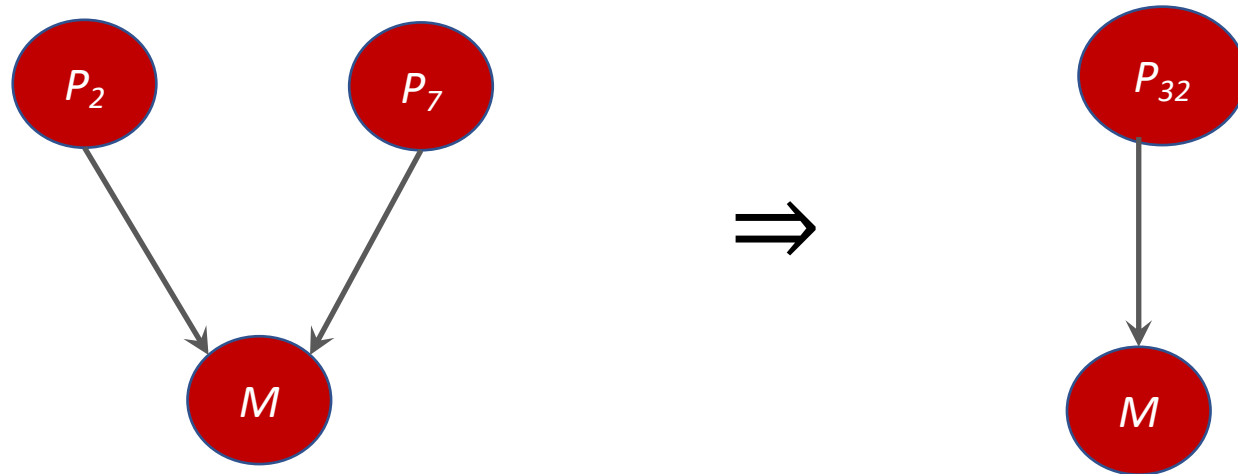
2. *determine_interactions* uses Exhaustive_IGain to learn interactive risk factors of target T. Examples are as follows. These modes are included in one of my output files named as “...learned_interactions...”.



MBIL Procedures

3. *transform_from_interactions* collapses each interaction to a single variable. The set of these new single variables is $PA2$, and the set $PA = PA1 \cup PA2$ is the set of all candidate direct risk factors of T .

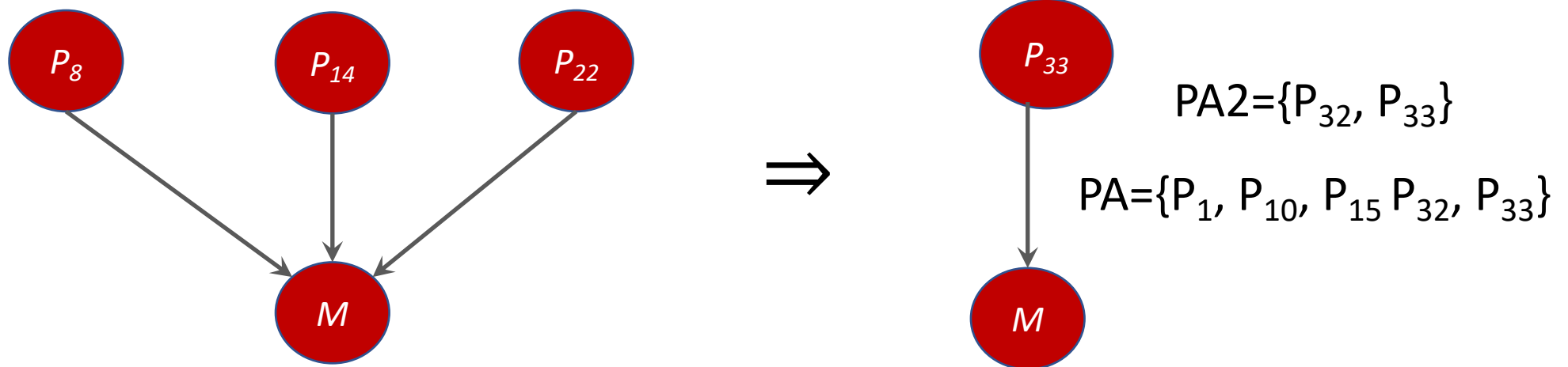
Assuming we have n candidate predictors, we will create new predictors ($n+1, n+2, \dots$) by transformation. For example,



MBIL Procedures

3. *transform_from_interactions* collapses each interaction to a single variable. The set of these new single variables is $PA2$, and the set $PA = PA1 \cup PA2$ is the set of all candidate direct risk factors of T .

Assuming we have n candidate predictors, we will create new predictors ($n+1, n+2, \dots$) by transformation. For example,



MBIL Procedures

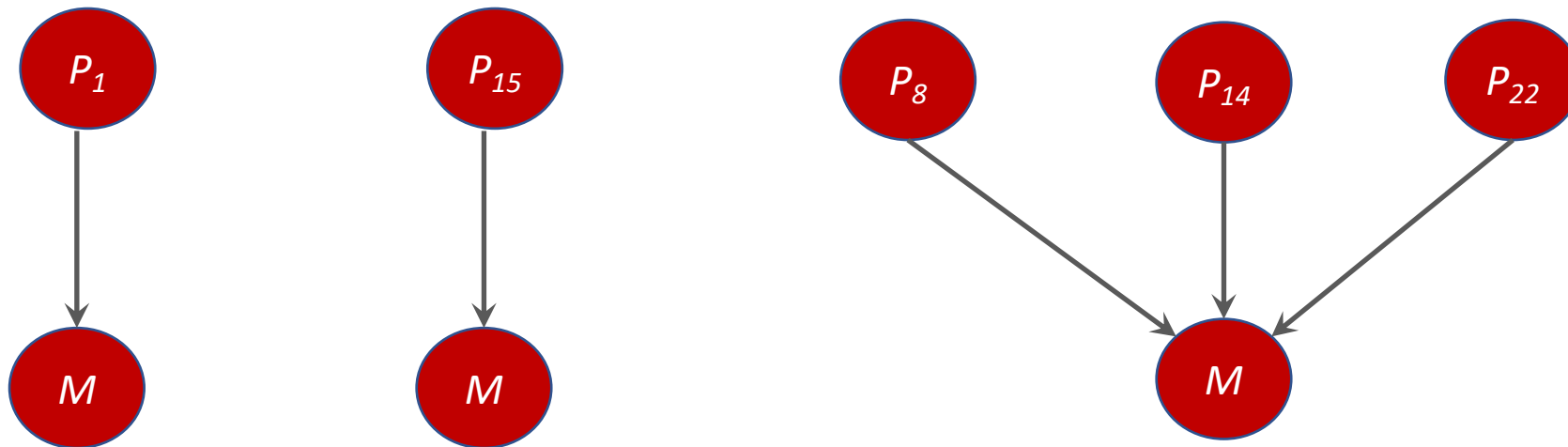
4. *find_parents* then removes indirect risk factors of T from PA . It does this as follows. When $i=1$, the procedure checks whether each one parent model has a lower score than the model with no parents. If it does, the variable in the model is not a direct risk factor and is removed from PA . When $i=2$ it checks, for each two parent models, whether removing one of the variables in the model increases the score of the model. If so, T is independent of the removed variable given the other variable, and so the removed variable is not a direct risk factor and is removed from PA . Next when $i=3$ the algorithm checks, for each three parent model, whether removing one of the variables in the model increases the score of the model. If so, T is independent of the removed variable given the other two variables, and so the removed variable is not a direct risk factor and is removed from PA . The procedure continues in this fashion until i equals the size of PA or some maximum number R of Markov blanket variables is reached.

MBIL Procedures

5. *transform_to_interactions* transforms the variables representing interactions back to interactions. For example, assume P_1 and P_{32} get removed during procedure 4, then

$$PA = \{ P_{10}, P_{15}, P_{33} \}$$

The mt odels we will learn are as follows, these are included in one of my output files called “...causal_sets ...”



- Some questions

1. Input

