COMP424 Tutorial

A3

Rudolf Lam

Today’s Menu

- Bayesian Networks
  - Revision
  - Performing inferences
  - Learning with Bayes Nets
  - More Examples

Revision

What is a Bayes Net?

Revision

What is a Bayes Net?

Bayes Net : G x Θ

Where:

- G is a DAG
- Θ is the set parameters for all conditional probability distributions in G

Revision

What does Bayes Net do?

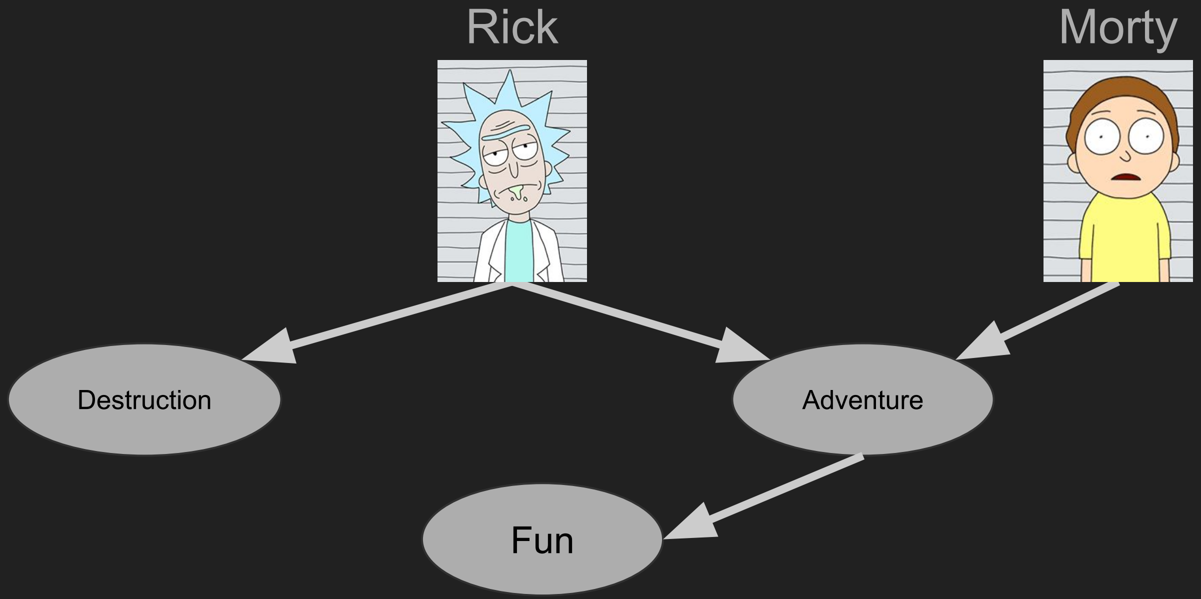
Revision

What does Bayes Net do?

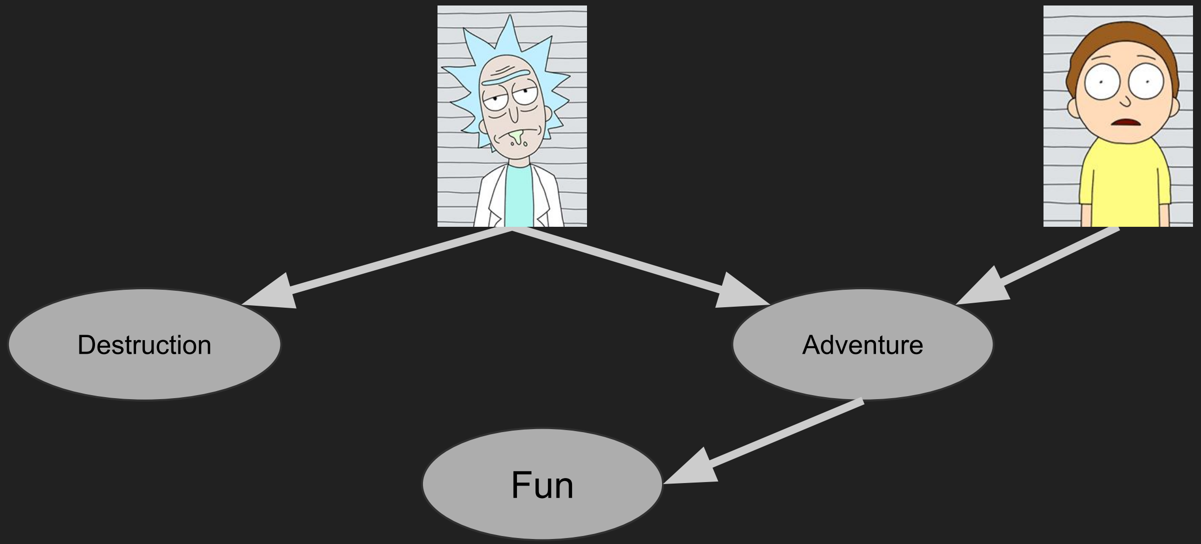
Recall:

Bayes Net : $G \times \Theta$

Example

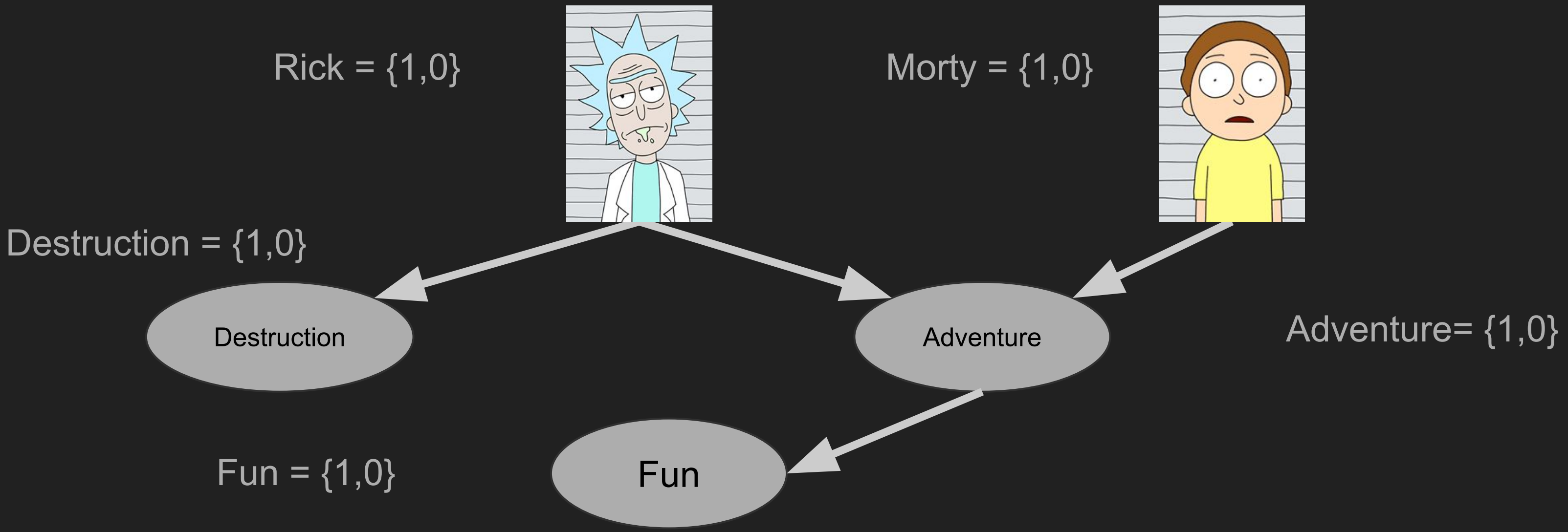


Example



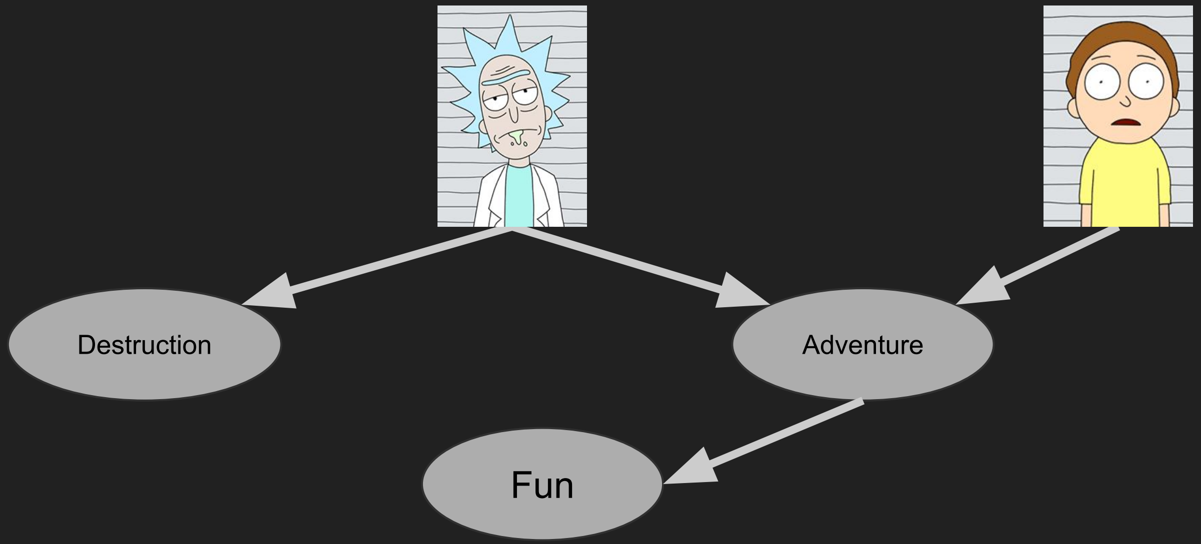
What do the nodes represent?

Example



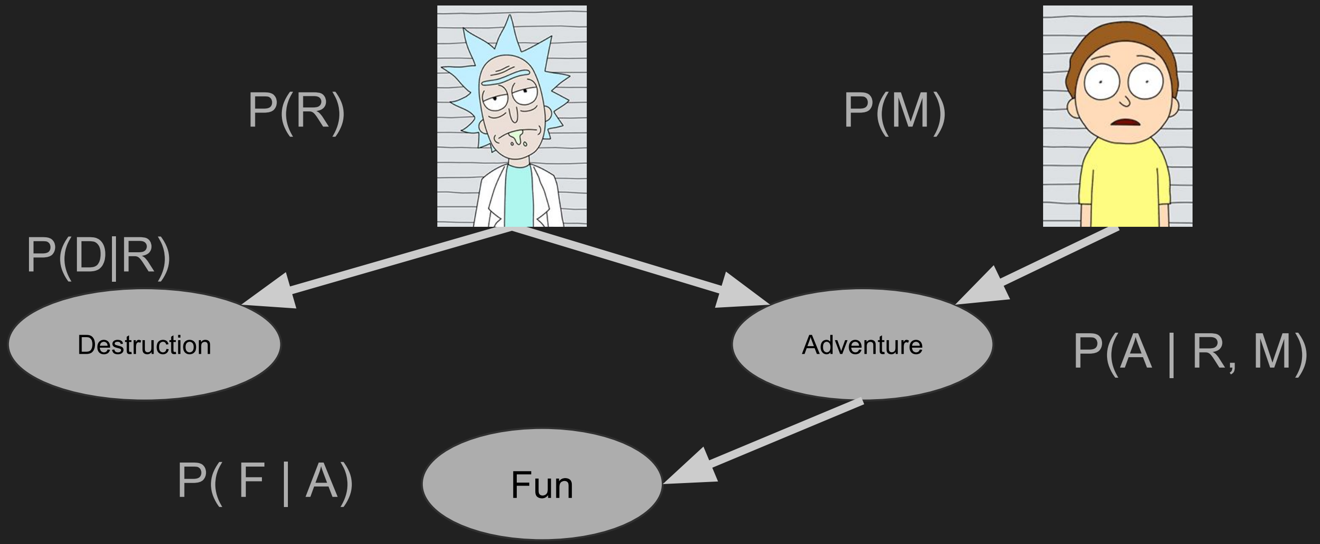
The random variables

Example



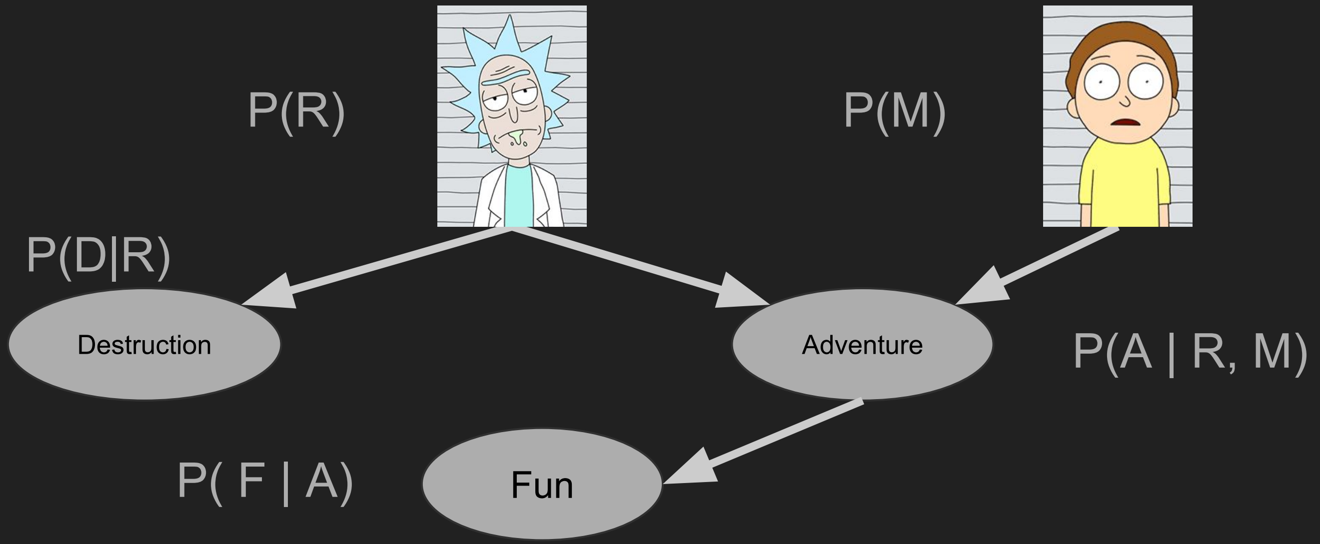
What do the edges represent?

Example



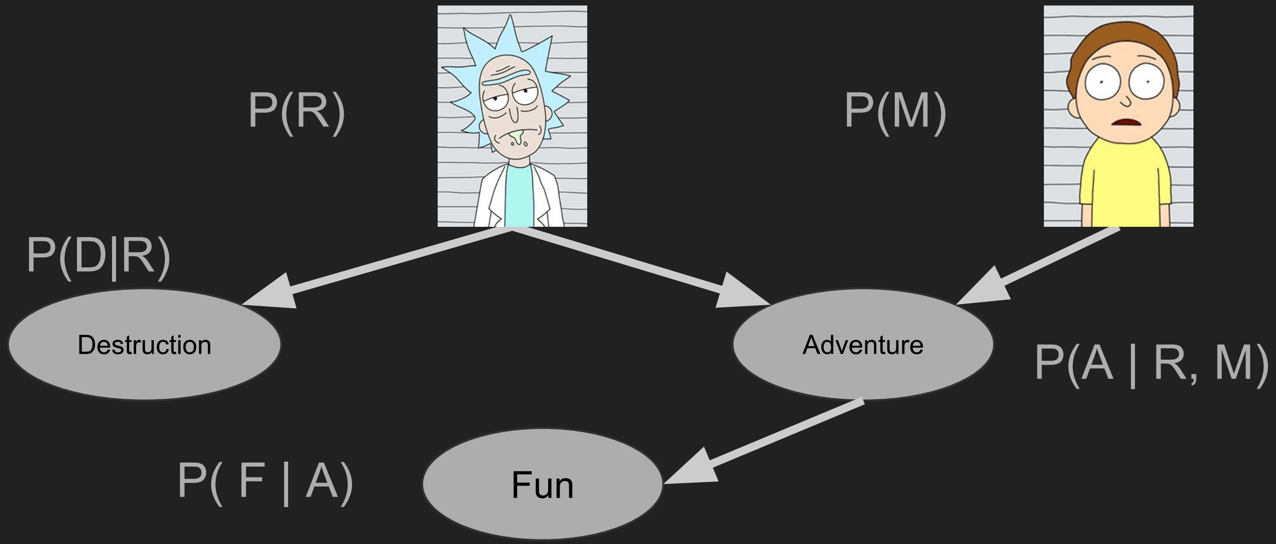
Conditional probability of a node’s random variable conditioned on its parents

Example



What are we missing for a Bayes Net?

Example

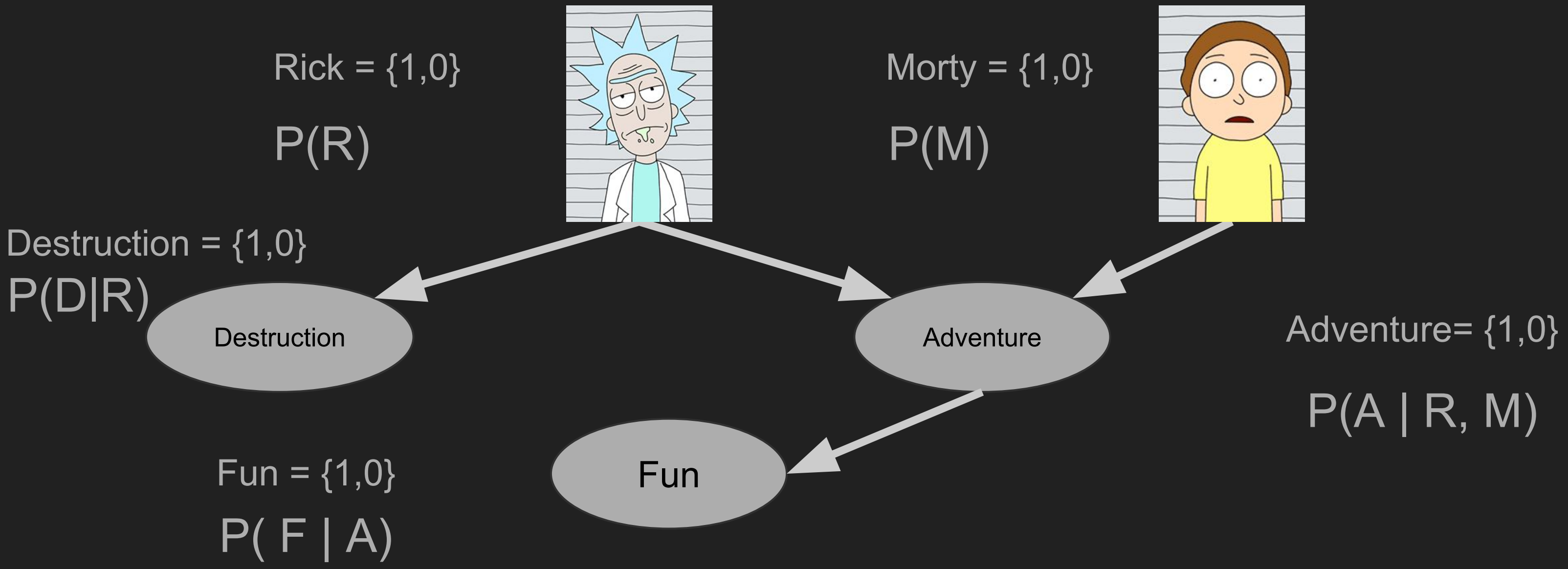


Θ_A

R	M	A
0	0	0.01
1	0	0.80
0	1	0.05
1	1	0.90

What are we missing for a Bayes Net?

Question for you



How many parameters do we have?

Parameters

1 Rick = {1,0}

P(R)

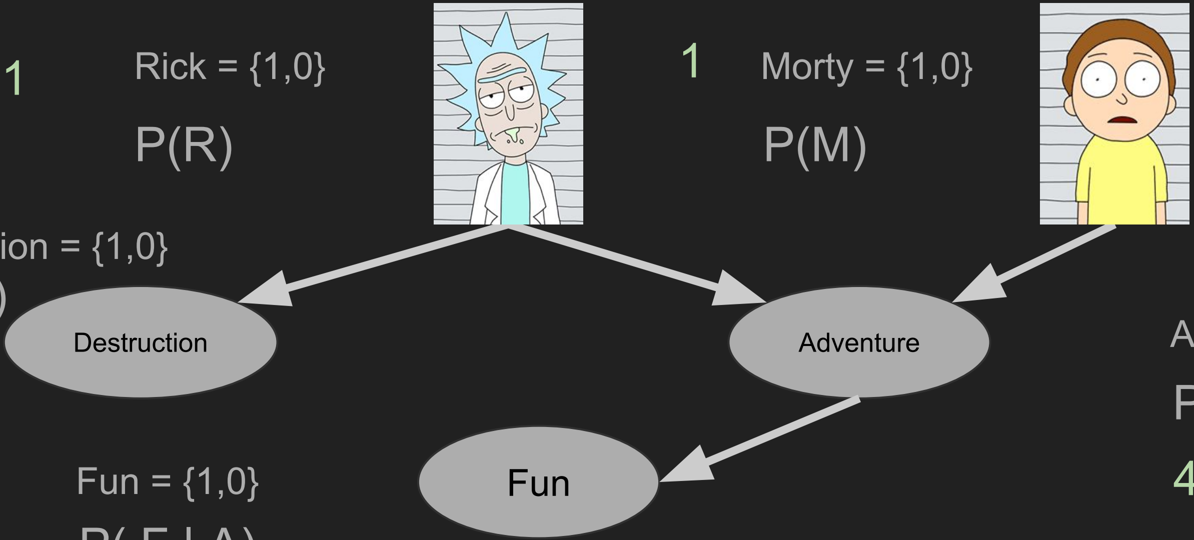
1 Morty = {1,0}

P(M)

Destruction = {1,0}

P(D|R)

2



Adventure= {1,0}

P(A | R, M)

4

Fun = {1,0}

P(F | A)

2

The number of entries in all of the Θs

So what,

What can we do with this Bayes Net?

What to do

What can we do with this Bayes Net?

Compute the probability of a specific state

What to do

What can we do with this Bayes Net?

We can query :

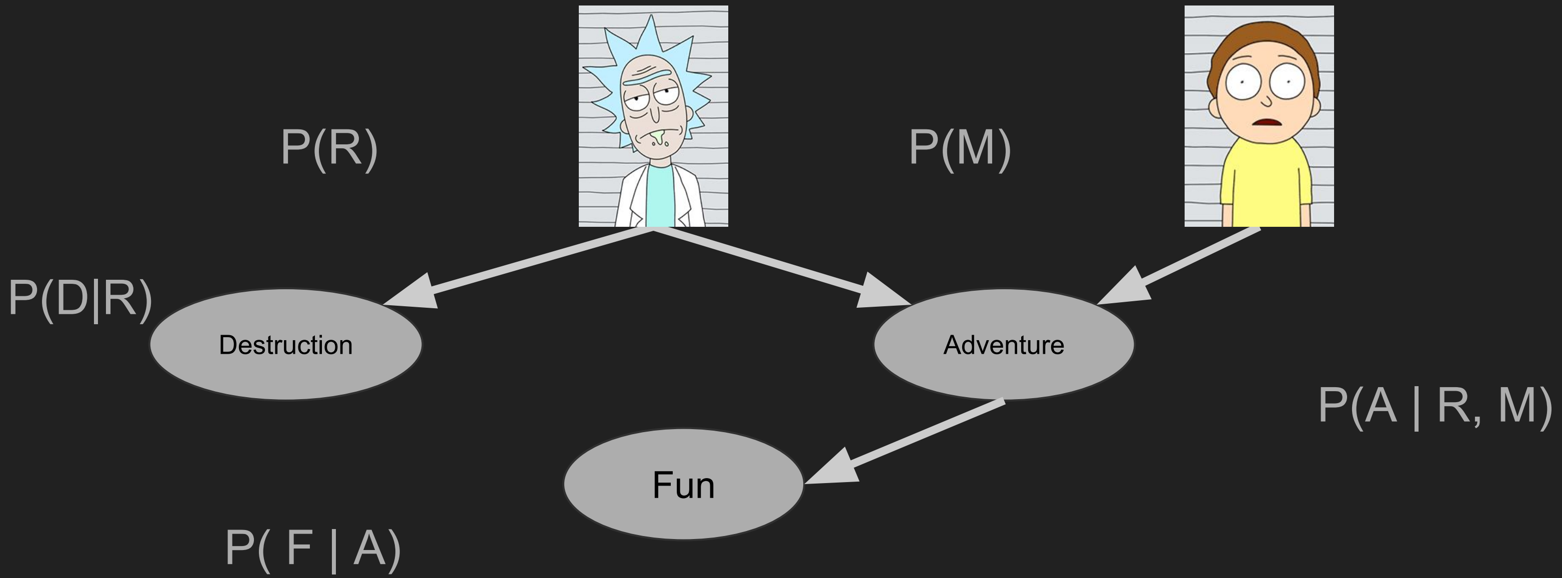
P(Rick=1, Morty=1, Destruction=1, Adventure=1, Fun=0)

What to do

Joint probability

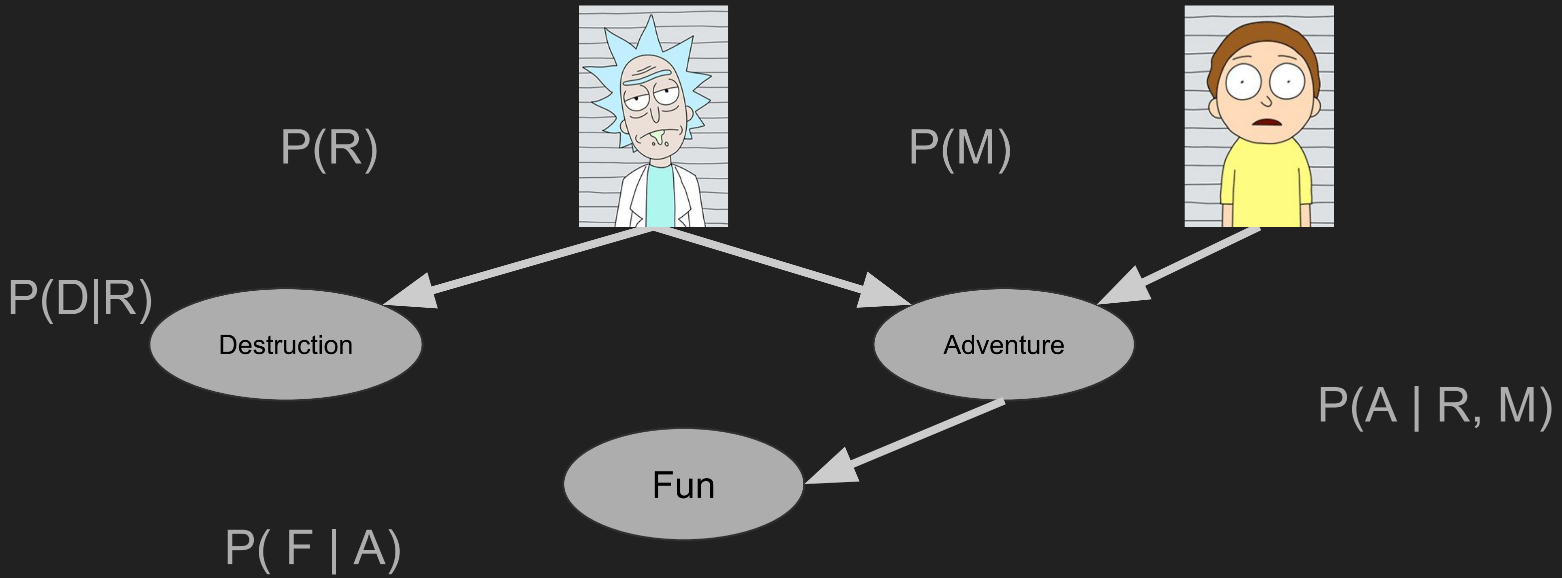
$$P(v_1, \ldots, v_n) = \prod_k P(v_k \mid \text{Parents}(v_k))$$

Joint Probability



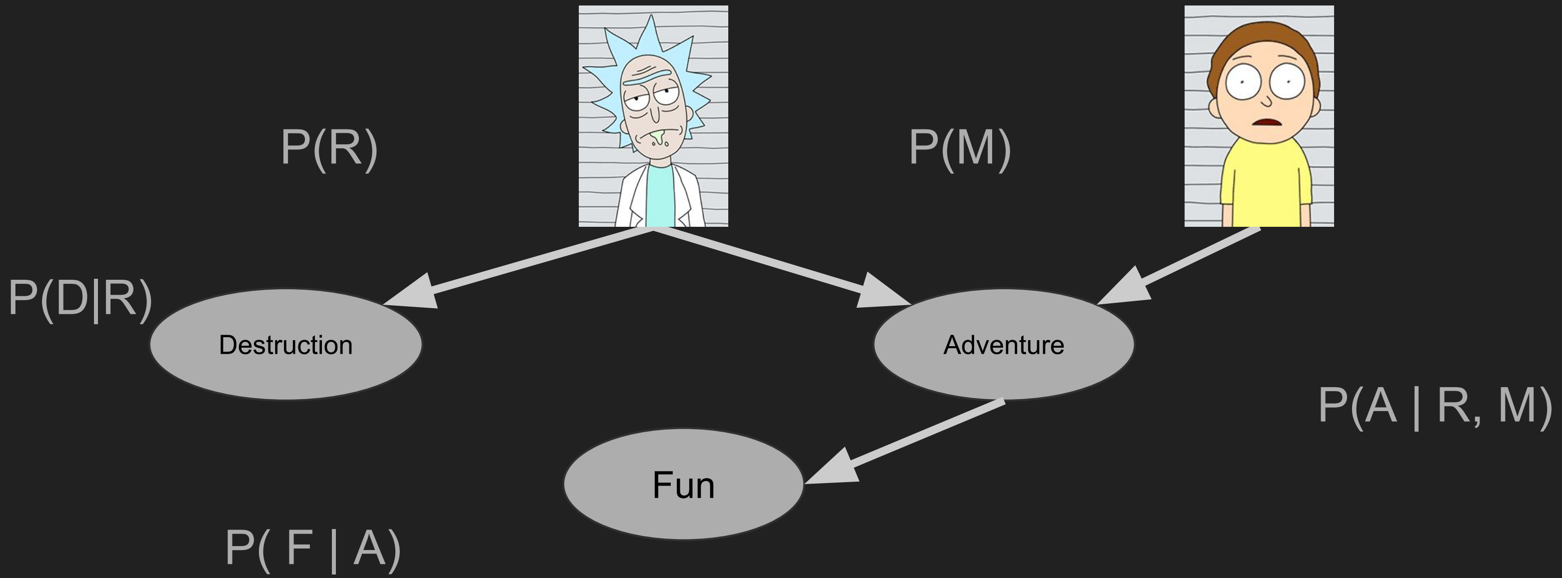
$P(\text{Rick}=1, \text{Morty}=1, \text{Destruction}=1, \text{Adventure}=1, \text{Fun}=0) = ?$

Joint Probability



$$P(R=1)\ P(M=1)\ P(D=1|R=1)\ P(A=1|\ R=1, M=1)\ P(F=0|A=1)$$

Question for you



Is this a polytree?

Polytree?

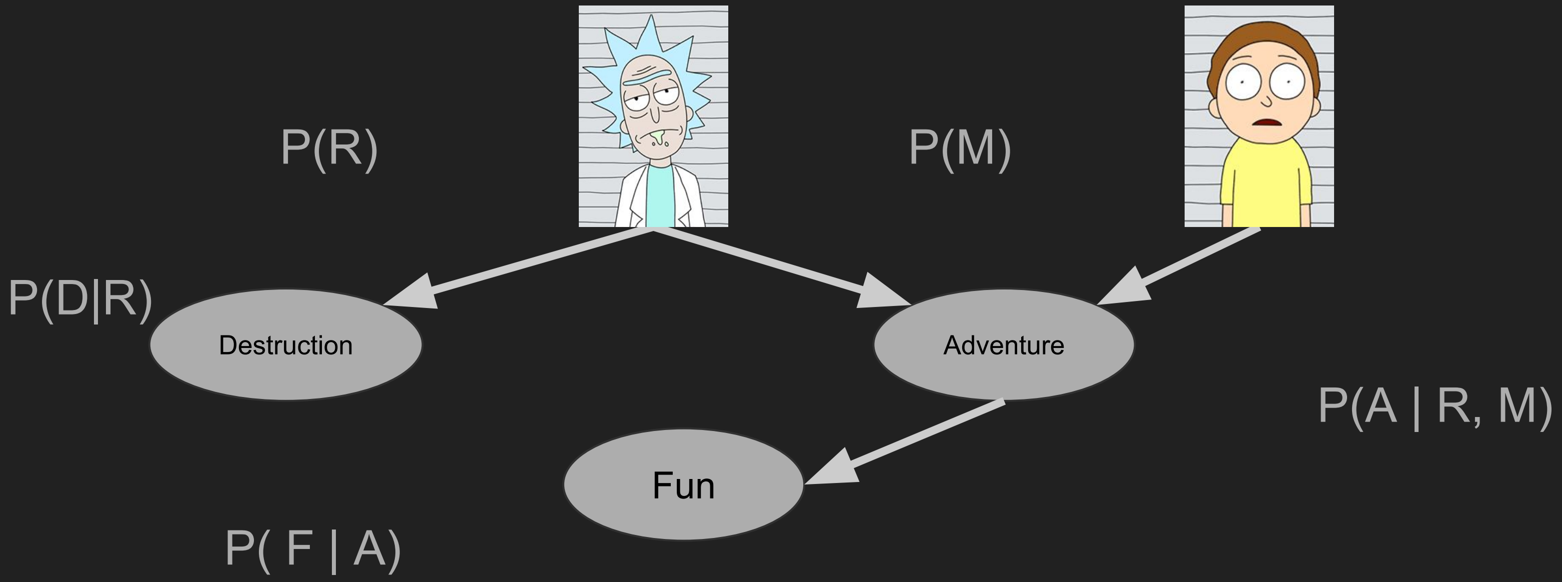
Directed Acyclic Singly-connected Tree

is a PolyTree

Polytree?

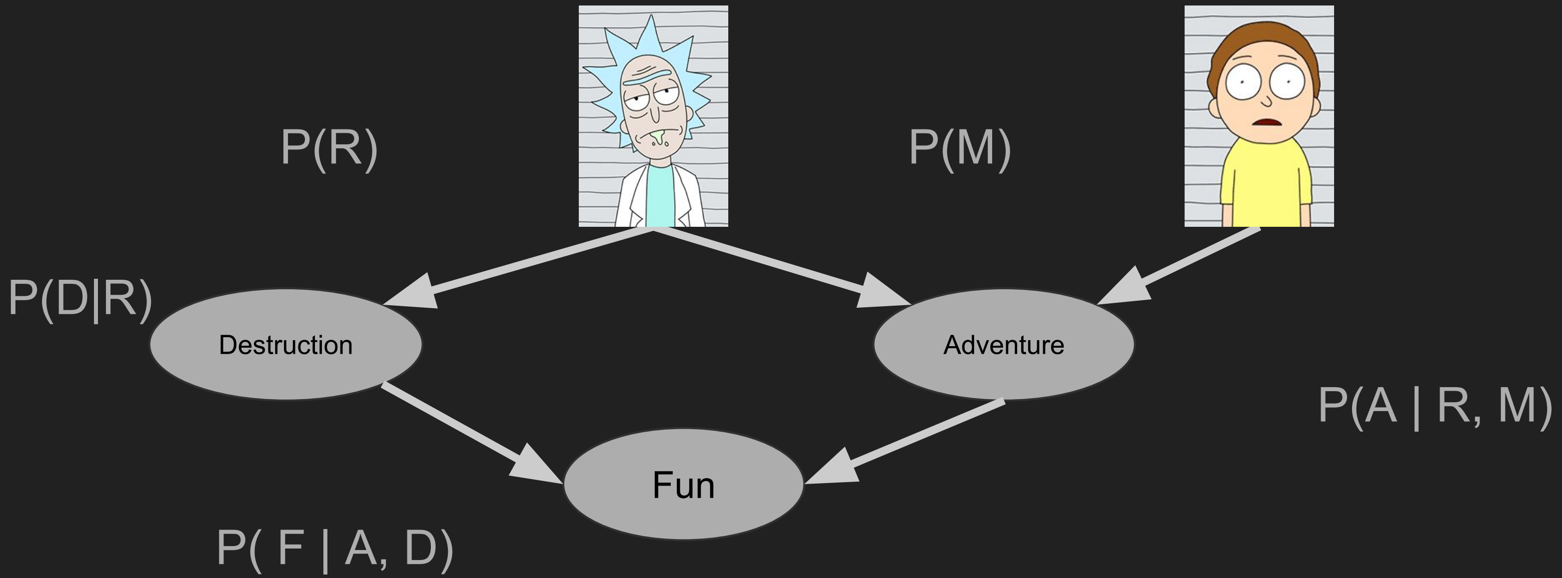
You can reach at least one node from more than one path

Question for you



Is this a polytree?

Consider a sadistic observer model



Is this still a polytree? Why?

Inferences

Let’s says we want the probability of a particular random variable.

How do we do it?

Inferences

Marginalization over the joint distribution.

Recall:

Joint probability

$$P(\, v_1, \ldots , v_n) = \prod_k P(v_k \mid \text{Parents}(v_k)\,)$$

Example

Given our model, can we find out the probability of Destruction being there?

Probability of which random variable are we looking for?

P(D)

Example

How do we do it?

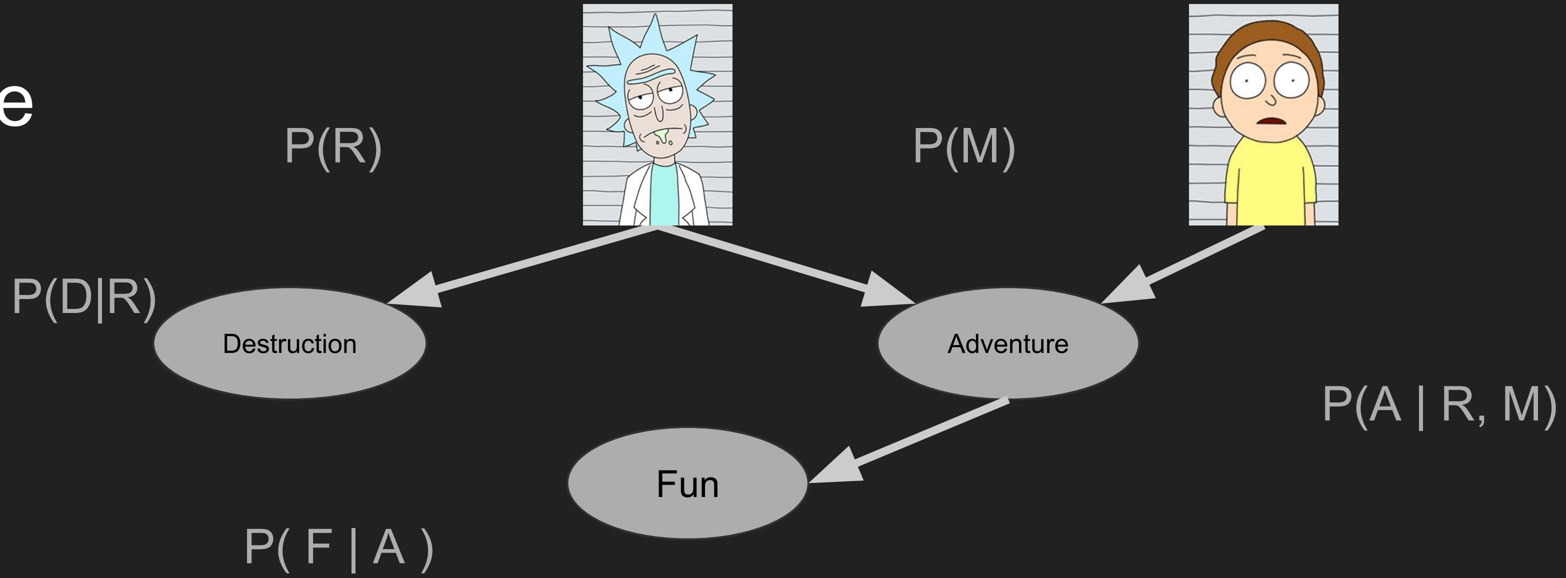
Example

Marginalization

Filtering out variables that we do not want

Like Morty, Fun, Adventure, etc.

Example



$$P(d) = \sum_{R,M,A,F} P(R, M, d, A, F)$$

$$P(d) = \sum_{R,M,A,F} P(R)\, P(M)\, P(d|R)\, P(A|R, M)\, P(F|A)$$

$$P(d) = \sum_{M} P(M) \sum_{R} P(R)\, P(d|R) \sum_{A} P(A|R, M) \sum_{F} P(F|A)$$

Abuse of notations alert

Example

Θ_R

R
0.90

Θ_D

R	D
0	0.01
1	0.95

$$P(d) = \sum_R P(R)\, P(d|R) \sum_M P(M) \sum_A P(A|R, M) \sum_F P(F|A)$$

$$P(d) = \sum_R P(R)\, P(d|R)$$

$$P(d) = P(r)\, P(d|r) + P(\neg r)\, P(d|\neg r)$$

$$P(d) = \Theta_R(\ )\, \Theta_D(R{=}1) + (\ 1 - \Theta_R(\ )\ )\, \Theta_D(R{=}0)$$

Abuse of notations alert

Another example

Now, let’s say we see that there is Fun can we tell what is the probability of Destruction?

Which conditional probability are we looking for?

Another example

Now, let's say we see that there is Fun can we tell what is the probability of Destruction?

Which conditional probability are we looking for?

$P(d|f)$

Abuse of notations alert

Going back to the definition

$$P(d|f) = P(d,f) / P(f)$$

Abuse of notations alert

Going back to the definition

$$P(d|f) = P(d,f) / P(f)$$

We know how to find P(f). How do we find P(d,f)

Abuse of notations alert

Going back to marginalization

Again, marginalization

$$P(d,f) = \sum_{R,M,A} P(R)\,P(M)\,P(d|R)\,P(A|R, M)\,P(f|A)$$

Abuse of notations alert

Ooh wee

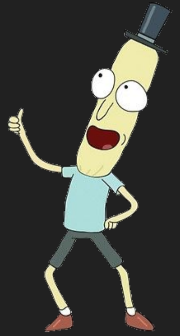
We get:

$$\left(\sum_{R,M,A} P(R)\, P(M)\, P(d|R)\, P(A|R, M)\, P(f|A) \right) / P(f)$$

something really complicated;

we didn’t even expand P(f)

Abuse of notations alert



Maximum a posteriori

Now, let’s say, instead, given Fun can we tell what is it more likely to have Destruction than not?

In this case, we are no longer looking for $P(D=1|F=1)$. What are we looking for?

Maximum a posteriori

Now, let's say, instead, given Fun can we tell what is the most likely assignment of Destruction?

In this case, we are no longer looking for $P(D=1|F=1)$. What are we looking for?

$$\text{argmax}_d \; P(D=d \mid F=1)$$

Variable Elimination

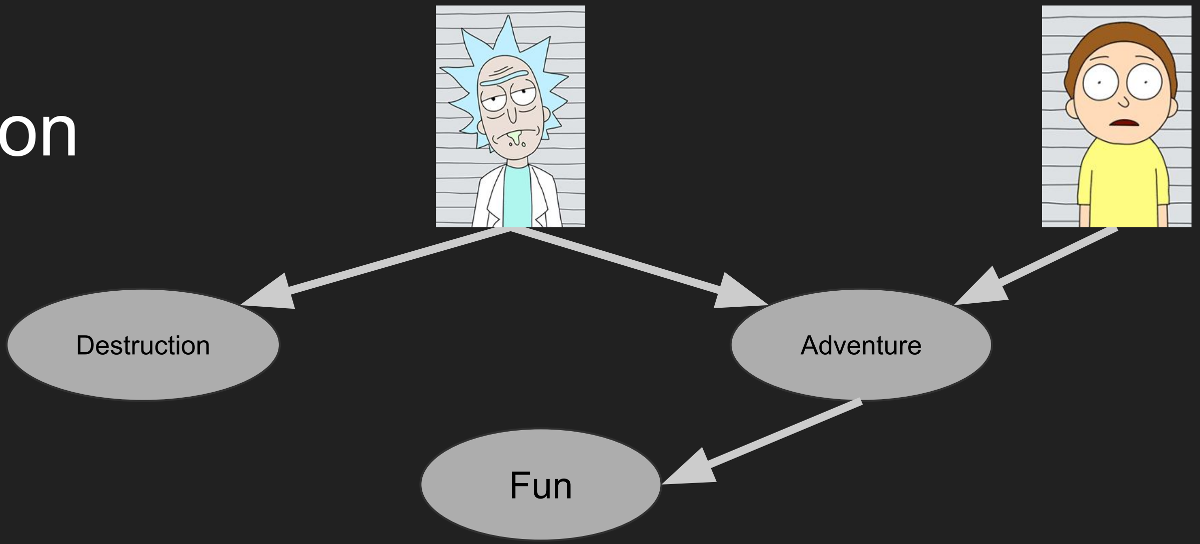
Variable elimination is marginalization with caching

Variable Elimination

Variable Elimination:

1. Define an ordering, with the query being the last variable
2. Repeat until no more factors
   a. Set up a list of factors
   b. Marginalize base on the ordering
   c. Cache each marginalized sum

Variable Elimination

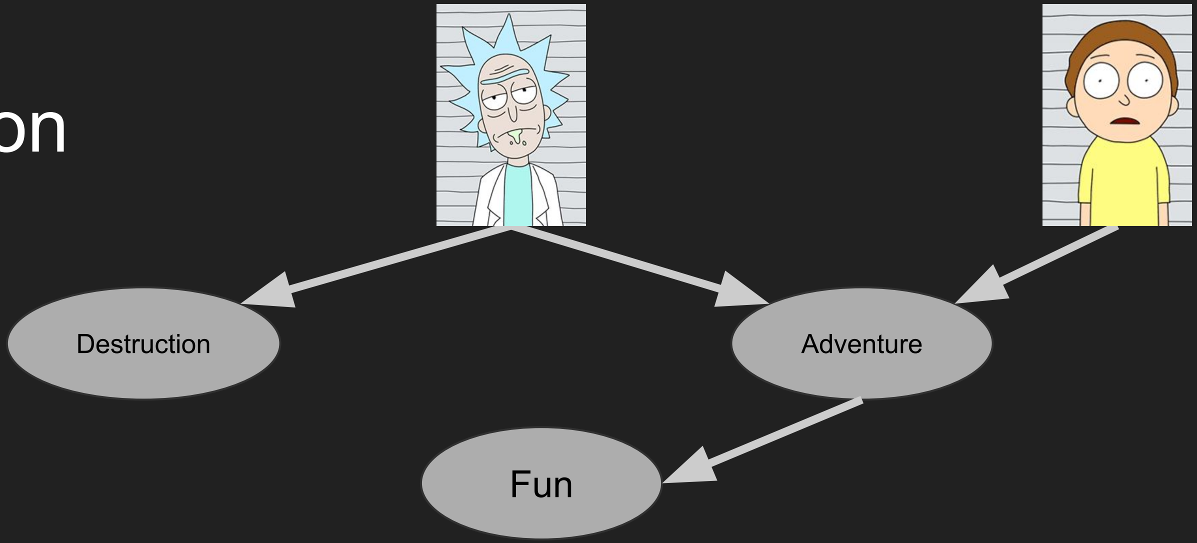


Variable Elimination:

1. Define an ordering, with the query being the last variable
2. Repeat until no more factors
    a. Set up a list of factors
    b. Marginalize base on the ordering
    c. Cache each marginalized sum

Let’s choose an ordering. But first, what should we marginalize out for P(D|f)?

Variable Elimination



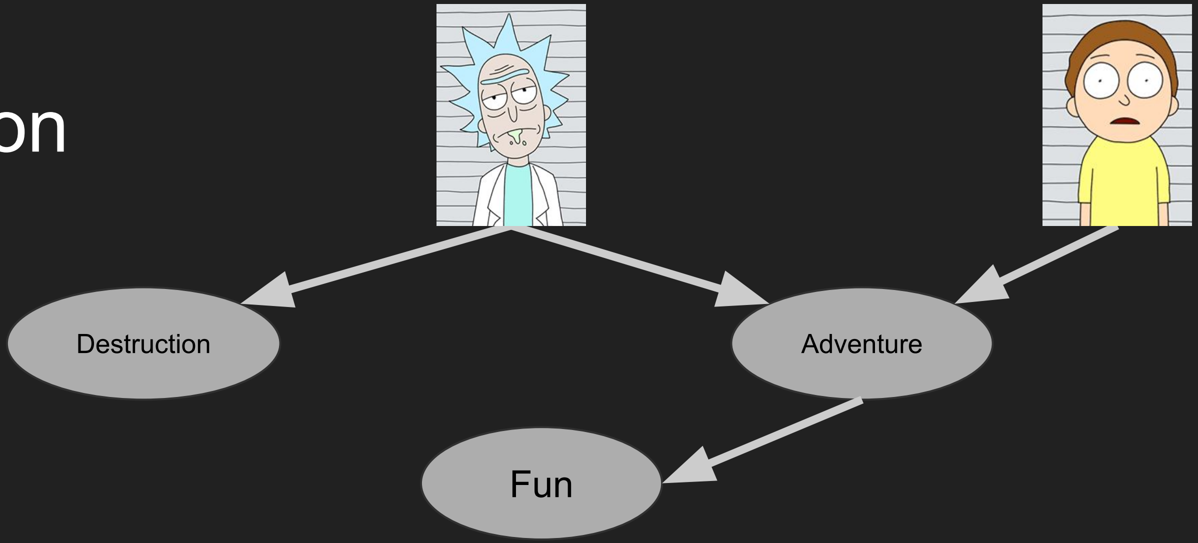
Variable Elimination:

1. Define an ordering, with the query being the last variable
2. Repeat until no more factors
   a. Set up a list of factors
   b. Marginalize base on the ordering
   c. Cache each marginalized sum

Let’s choose an ordering.

<A, R, M, F, D>

Variable Elimination



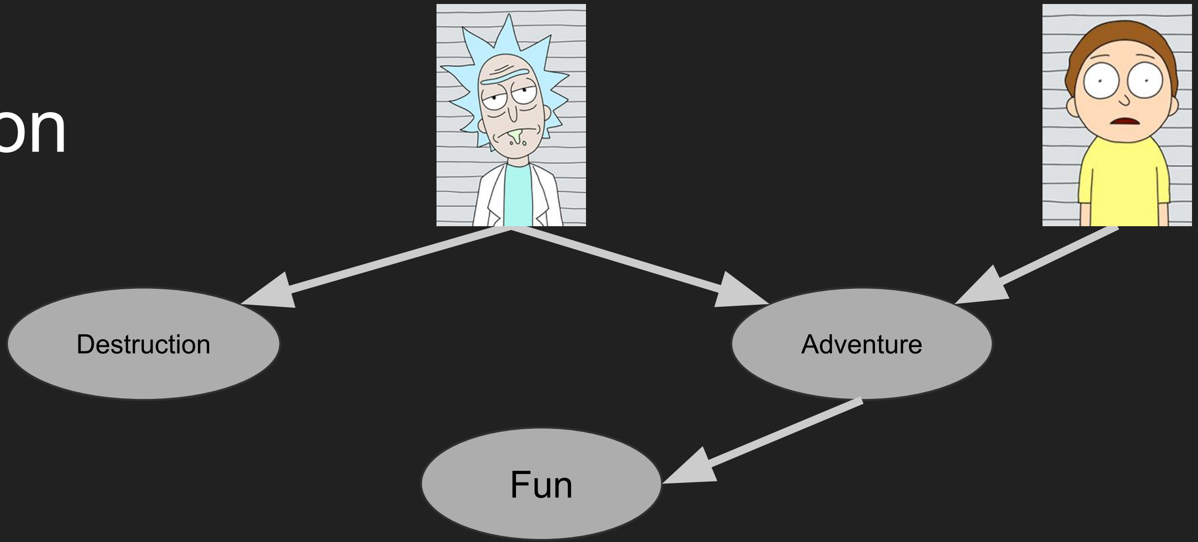
What factors are available to us now?

P(R) , P(D|R), P(M), P(A|R,M), P(F|A)

we add additional δ(F, 1) for observation; we have

P(R) , P(D|R), P(M), P(A|R,M), P(F|A) , δ(F, 1)

Variable Elimination



Variable Elimination:

1. Define an ordering, with the query being the last variable
2. Repeat until no more factors
   a. Set up a list of factors
   b. Marginalize base on the ordering
   c. Cache each marginalized sum

$$P(R)\ ,\ P(D|R),\ P(M),\ P(A|R,M),\ P(F|A)\ ,\ \delta(F,\ 1)$$

Marginalize base on the ordering

Θ_A

R	M	A
0	0	0.01
1	0	0.80
0	1	0.05
1	1	0.90

Θ_F

A	F
0	0.30
1	0.90

Ordering : <A, R, M, F, D>

Factors: $P(R)$, $P(D|R)$, $P(M)$, $P(A|R,M)$, $P(F|A)$, $\delta(F, 1)$

Marginalize A: $P(R)$, $P(D|R)$, $P(M)$, $P(A|R,M)$, $P(F|A)$, $\delta(F, 1)$

$$m_A(F,R,M) = \sum_a P(a|R,M)\, P(F|a)$$

Trick

Θ_A

R	M	A
0	0	0.01
1	0	0.80
0	1	0.05
1	1	0.90

Θ_F

A	F
0	0.30
1	0.90

Use matrices.

$m_A(F,R,M)$ is a 2x2x2 matrix

Trick

Θ_A

R	M	A
0	0	0.01
1	0	0.80
0	1	0.05
1	1	0.90

Θ_F

A	F
0	0.30
1	0.90

$m_A(F,R,M) =$

F	R	M	P(A=1|R,M) P(F|A=1)+P(A=0|R,M) P(F|A=0)
0	0	0	?
0	1	0	?
0	0	1	?
...	...	...	...

Marginalize base on the ordering

Θ_A

R	M	A
0	0	0.01
1	0	0.80
0	1	0.05
1	1	0.90

Θ_F

A	F
0	0.30
1	0.90

Ordering : <A, R, M, F, D>

Factors: P(R) , P(D|R), P(M), P(A|R,M), P(F|A) , δ(F, 1)

Marginalize A: P(R) , P(D|R), P(M), P(A|R,M), P(F|A) , δ(F, 1)

Update Factors: ?

Continued

Ordering : <~~A~~, R, M, F, D>

Factors: $P(R)$, $P(D|R)$, $P(M)$, $m_A(F,R,M)$, $\delta(F, 1)$

Marginalize R: $P(R)$, $P(D|R)$, $P(M)$, $m_A(F,R,M)$, $\delta(F, 1)$

$$m_R(F,D,M) = \sum_R P(R)\, P(D|R)\, m_A(F,R,M)$$

We do the same thing...

Continued...

Ordering : $\langle \cancel{A, R}, M, F, D \rangle$

Factors: $P(M)$, $m_R\ (F,D,M)$, $\delta(F, 1)$

Marginalize M: $P(M)$, $m_R\ (F,D,M)$, $\delta(F, 1)$

$$m_M\ (F,D) = \sum_M P(M)\ m_R\ (F,D,M)$$

How many entries do we have in this table?

Continued...

Ordering : <~~A, R, M~~, F, D>

Factors: m_M (F,D), δ(F, 1)

Marginalize F: m_M (F,D), δ(F, 1)

$$m_F(?) = \sum_F m_M(F,D)\,\delta(F, 1)$$

How many entries do we have in this table?

Continued...

Ordering : <~~A, R, M, F,~~ D>

Factors: $m_F(D)$

We have no more factors.

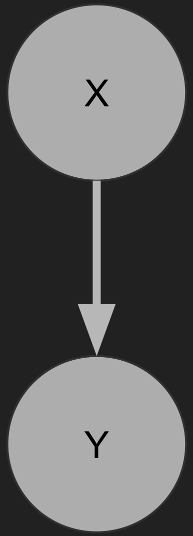
What remains is the 2 element vector $m_F(D) = \begin{bmatrix} d \\ \neg d \end{bmatrix}$

Continued...

Now we just pick D=1 if d is bigger than ¬d,

otherwise we pick D=0

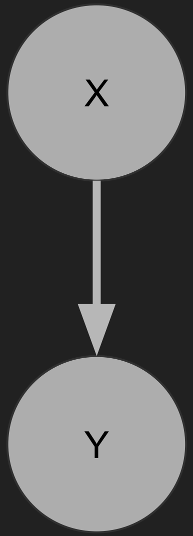
Setting up Bayes Net with data



Let's say we don't start with the complete model.

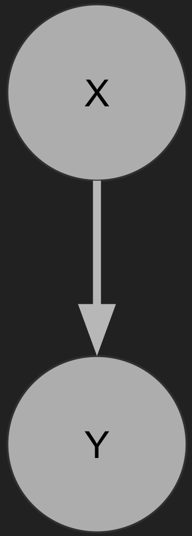
We are only given G, but not Θ

Setting up Bayes Net with data



Which parameters must we learn?

Setting up Bayes Net with data

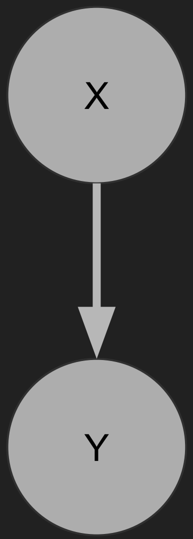


Which parameters must we learn?

$\theta_X = P(X)$

$\theta_Y = P(Y|X)$

Setting up Bayes Net with data



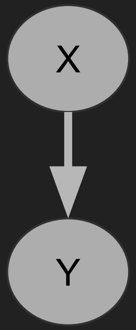
X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4

How do we learn these?

$\theta_X = P(X)$

$\theta_Y = P(Y|X)$

MLE

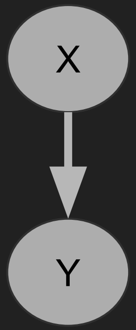


X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4

$\theta_X = P(X)$

$P(X=1) = 7/10$

MLE

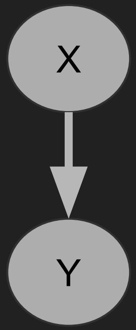


X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4

$\theta_Y = P(Y|X)$

We have two cases here, one for each X=0 and X=1

MLE



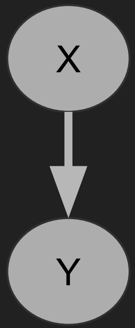
X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4

$\theta_Y = P(Y|X)$

$P(Y=1 \mid X=0) = 2/3$

$P(Y=1 \mid X=1) = 4/7$

Laplace smoothing



X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4

$\theta_X = P(X)$

$P(X=1) = (7 + 1) / (10 + |dom(X)|) = 7 + 1 / (10 + |\{0,1\}|) = 8/12$

$\theta_Y = P(Y|X)$

$P(Y=1 \mid X=0) = (2 + 1) / (3 + |dom(X)|) = 3/5$

$P(Y=1 \mid X=1) = (4 + 1) / (7 + |dom(X)|) = 5/9$

Extra information

What if we have extra incomplete data that we want to use?

X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4
0	?	1

EM algorithm

We can impute the missing piece and then integrate the new data using

EM algorithm, soft version.

EM algorithm

We had:

$$P(Y=1 \mid X=0) = \frac{2}{3}$$

We have weights:

- $Y=1$: $\frac{2}{3}$
- $Y=0$: $\frac{1}{3}$

EM algorithm

Without this data:

$$P(Y=1 \mid X=0) = 2/3$$

We had:

$$P(Y=1 \mid X=0) = (2 + \tfrac{2}{3} * 1) / (3+1) = \tfrac{2}{3}$$

X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4
0	?	1

EM algorithm

Without this data:

$P(Y=1 \mid X=0) = 2/3$

We had:

$P(Y=1 \mid X=0) = (2 + \frac{2}{3} * 1) / (3+1) = \frac{2}{3}$

X	Y	# of instances
0	0	1
0	1	2
1	0	3
1	1	4
0	?	1

For this example, the imputed data does not provide us with more information and has converged on the first step. So, no further E-steps are necessary.

And we are done for today