Music Machine Learning

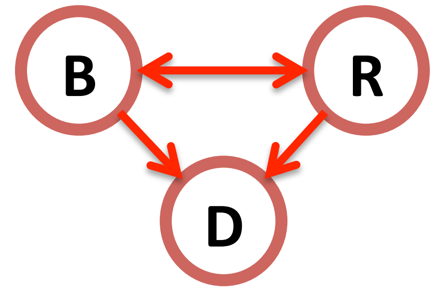
VI – Probabilistic models

Master ATIAM - Informatique

Philippe Esling (esling@ircam.fr)
Maître de conférences – UPMC
Equipe représentations musicales (IRCAM, Paris)

Probabilistic inference

- Let say that you have a dog
- If a raccoon shows up, your dog will bark
- If a burglar shows up, your dog will bark
- If I want to know the probabilities …
- I need to build a table and keep the counts
- So **how big of a table do I need to keep?**

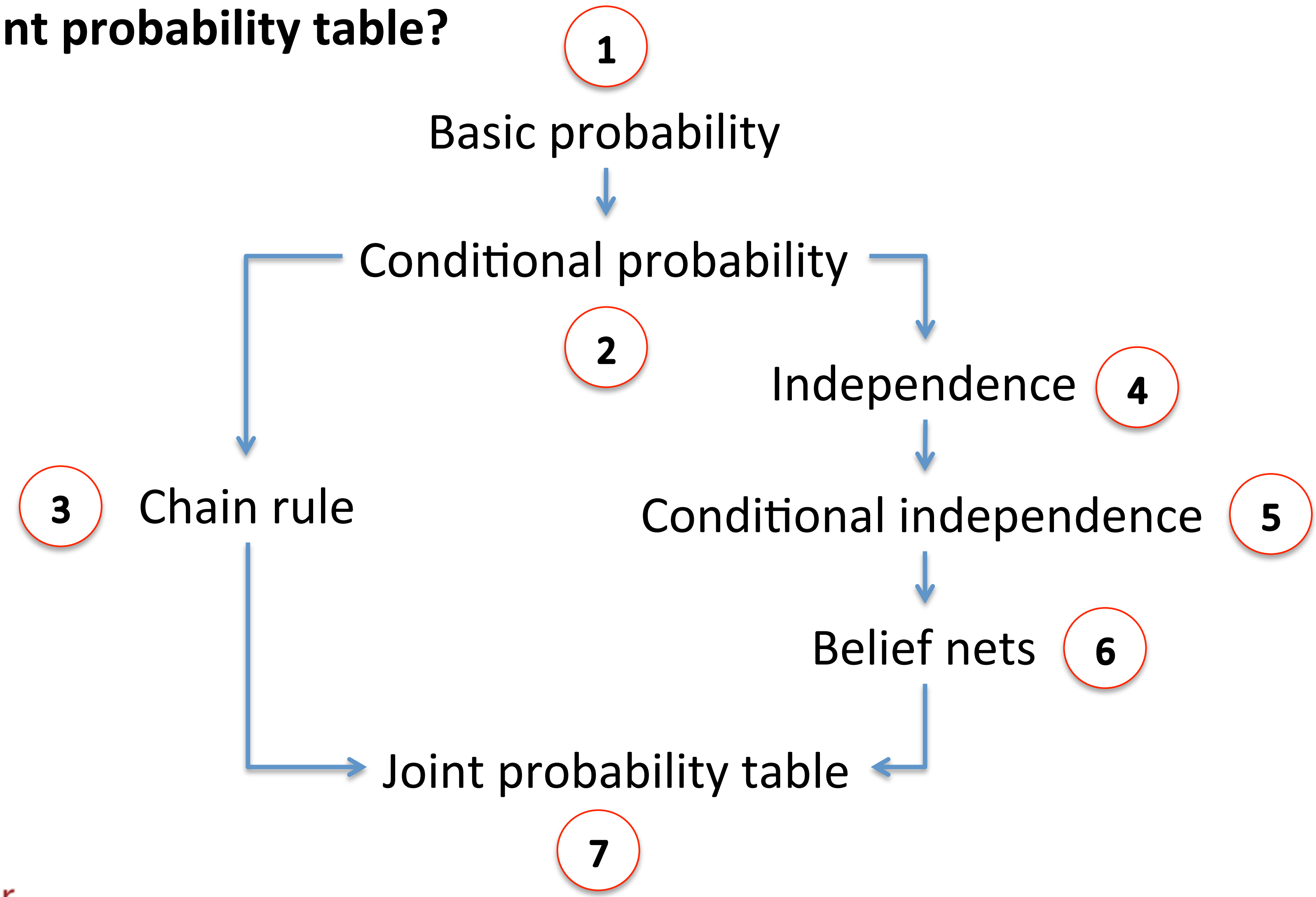


B	R	D	Seen
T	T	T	|
T	F	T	||
F	T	T	
F	F	T	|
T	T	F	
T	F	F	|
F	T	F	
F	F	F	||||

$\mathcal{P}$
.1
.2
.05
.1
.01
.09
.05
.4

Probabilistic Inference

Joint probability table?



1. Basic probabilities

$$0 \leq \mathcal{P}(a) \leq 1 \qquad \mathcal{P}(True) = 1$$
$$\mathcal{P}(False) = 0$$
$$\mathcal{P}(a) + \mathcal{P}(b) - \mathcal{P}(a \wedge b) = \mathcal{P}(a \vee b)$$

Intuition: notion of space

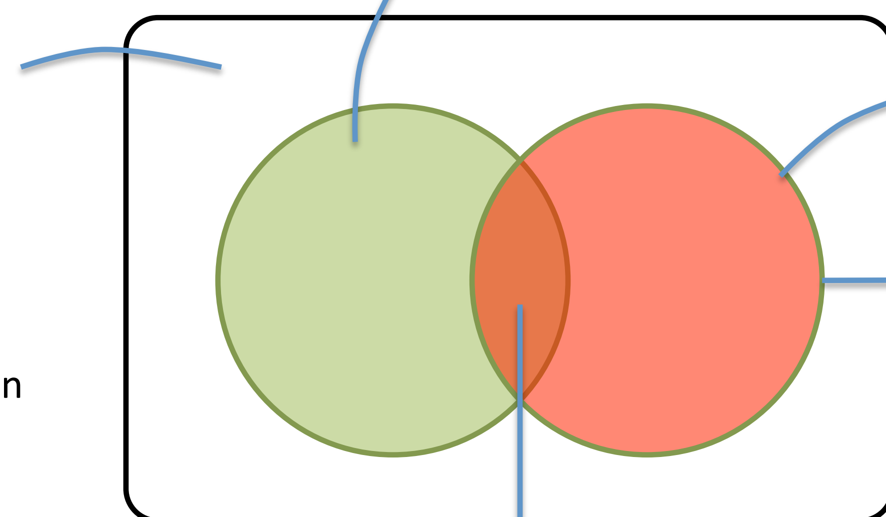
$\mathcal{P}(a)$ « Probability of landing in the circle »
area of the circle divided by area of space

Space of possibilities
Probability of landing anywhere in the space is going to be 1

$$\mathcal{P}(True) = 1$$

Probability of landing on an infinitesimal point

$$\mathcal{P}(False) = 0$$



If we add the two areas, we see that the intersection is counted twice

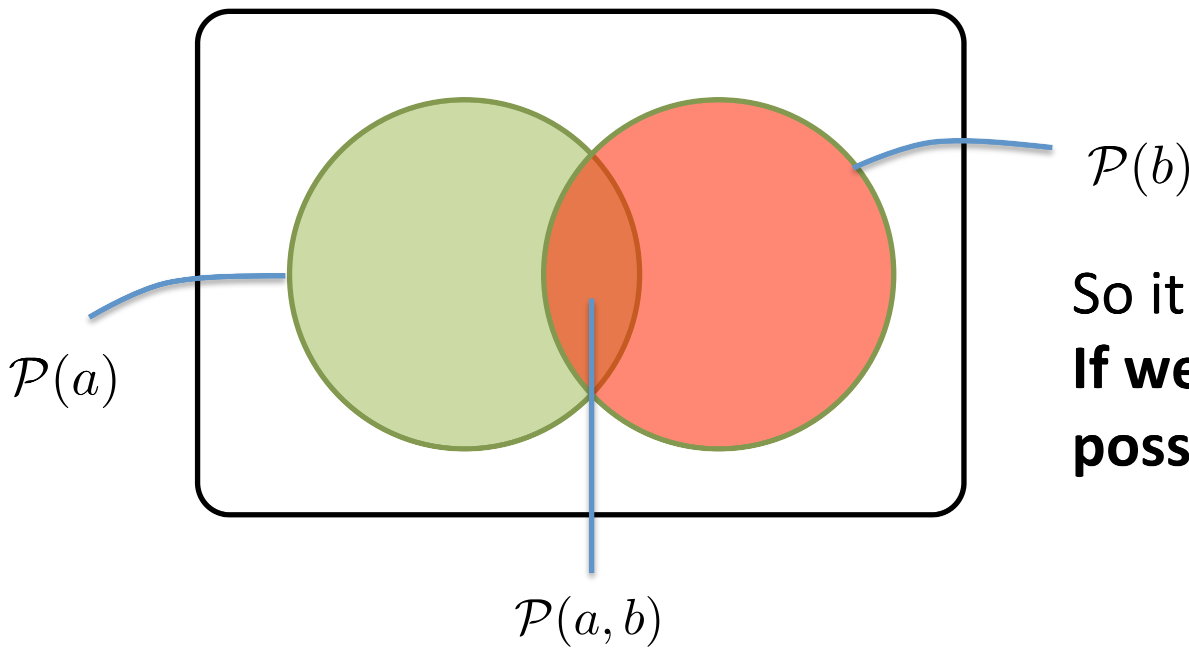
2. Conditional probability

$$\mathcal{P}(a \wedge b)$$

Definition: $\mathcal{P}(a|b) = \dfrac{\mathcal{P}(a,b)}{\mathcal{P}(b)}$

2. Conditional probability

Definition: $\mathcal{P}(a|b) = \dfrac{\mathcal{P}(a,b)}{\mathcal{P}(b)}$



So it is the probability of a **If we restrict the world of possibilities to b**

2. Conditional probability

Definition: $\mathcal{P}(a|b) = \dfrac{\mathcal{P}(a,b)}{\mathcal{P}(b)}$ $\quad \mathcal{P}(a,b) = \mathcal{P}(a|b).\mathcal{P}(b)$

$$\mathcal{P}(a,b,c) =? \quad y = b,c$$

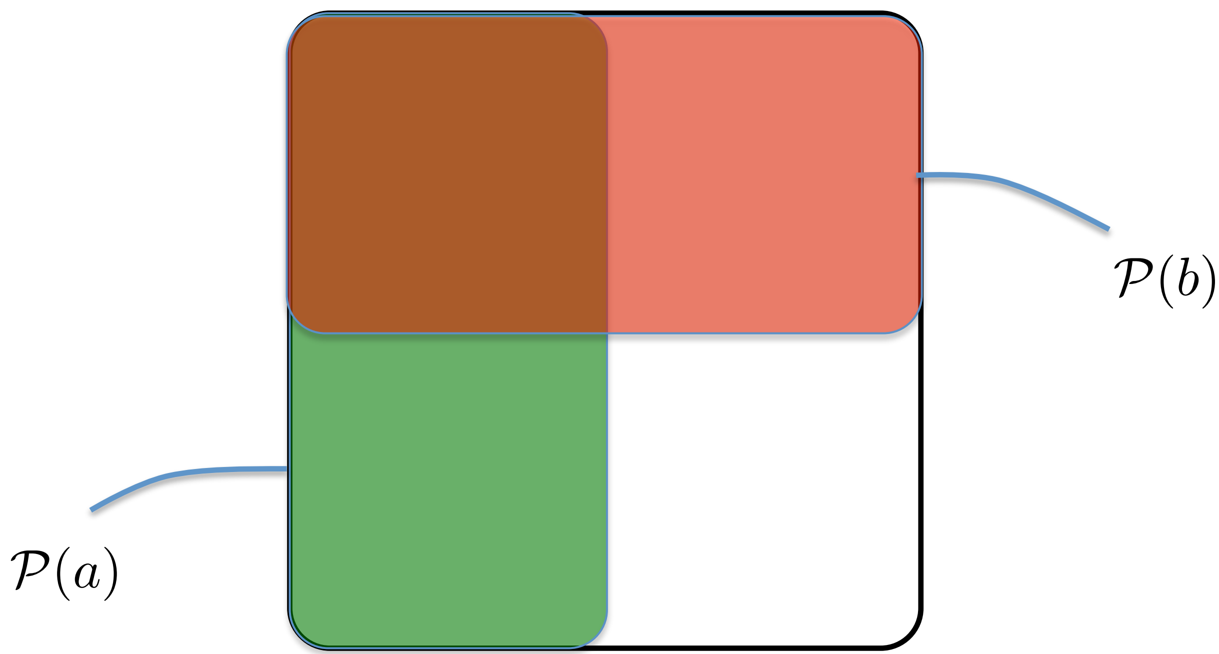
$$\begin{aligned}\mathcal{P}(a,b,c) &= \mathcal{P}(a,y) \\ &= \mathcal{P}(a|y)\mathcal{P}(y) \\ &= \mathcal{P}(a|b,c)\mathcal{P}(b,c) \\ &= \mathcal{P}(a|b,c)\mathcal{P}(b|c)\mathcal{P}(c)\end{aligned}$$

$$\mathcal{P}(x_1,\ldots,x_n) = \prod_{i=n}^{1} \mathcal{P}(x_i|x_{i-1},\ldots,x_1)$$

3. Chain rule

4. Independence

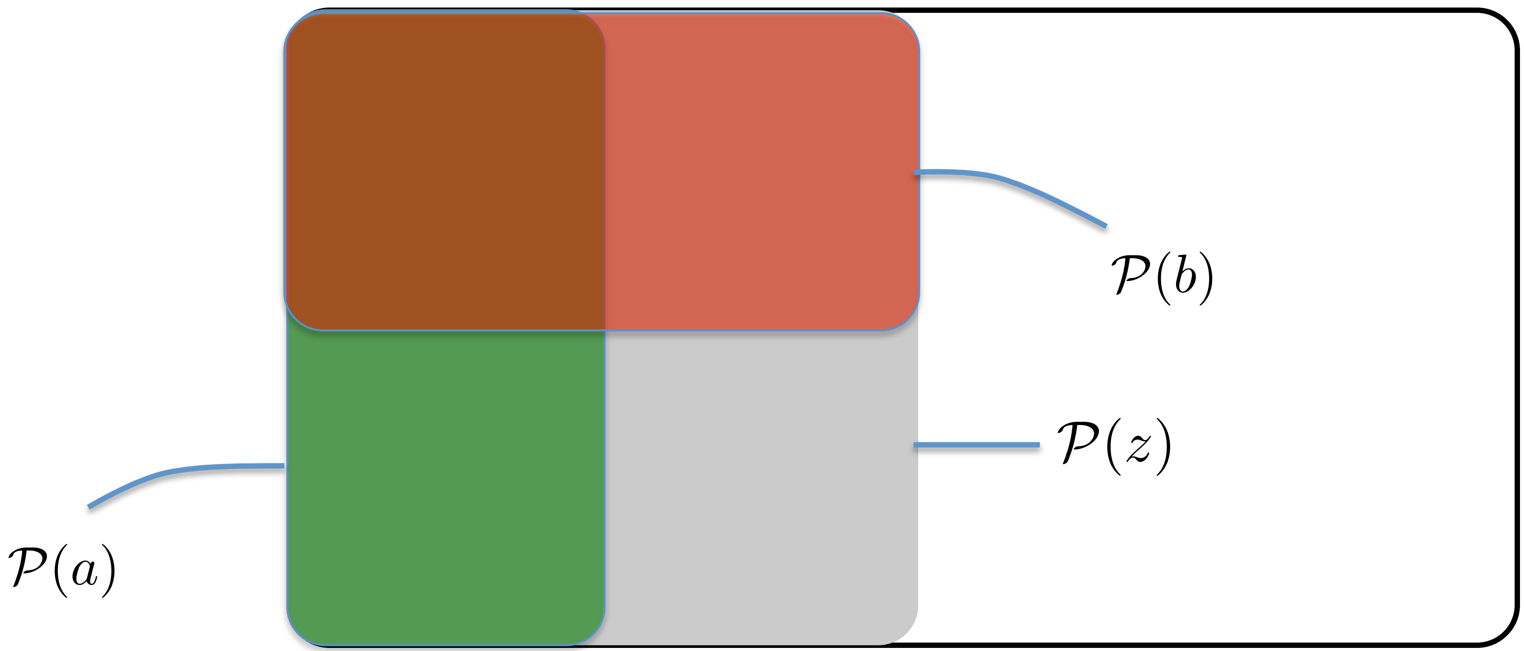
Definition: $\mathcal{P}(a|b) = \mathcal{P}(a)$ a independent of b



5. Conditional independence

Definition: $\mathcal{P}(a|bz) = \mathcal{P}(a|z)$

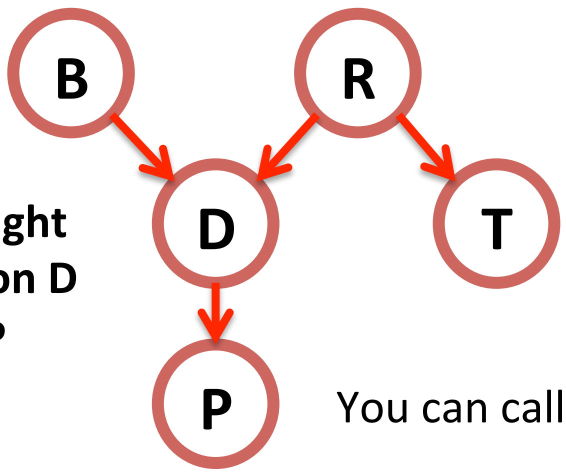
$$\mathcal{P}(ab|z) = \mathcal{P}(a|z)\mathcal{P}(b|z)$$



6. Belief network

- Let say that you have a dog
- If a raccoon shows up, your dog will bark
- If a burglar shows up, your dog will bark
- The raccoon will not show up as the dog is barking
- The burglar will not show up as the dog is barking
- So the causality flows from the raccoon and the burglar to the dog
- So we can make a diagram of this

But B might depend on D and P



Given its parents every node is independent of all other non-descendents

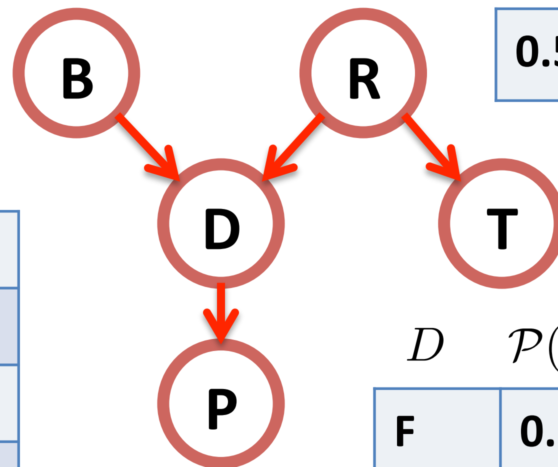
Racoons can make the trashcans fall

You can call the police

Eg. Calling the police depends **solely on the probability of the dog** (causality flows in)

How big is the probability table I need if I want to keep the tallies ?

7. Joint probability table



$\mathcal{P}(B)$

0.1

$\mathcal{P}(R)$

0.5

B	R	$\mathcal{P}(D)$	
F	F	0.1	
T	F	1.0	
F	T	0.5	
T	T	1.0	

R	$\mathcal{P}(T)$
F	0.01
T	0.8

D	$\mathcal{P}(P)$
F	0.01
T	0.5

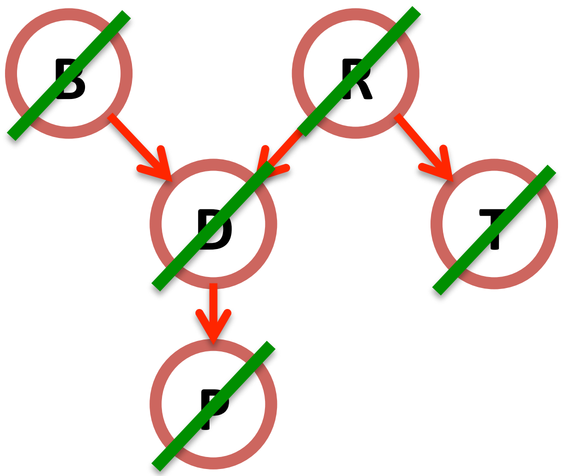
- **We only had to specify 10 numbers instead of 32 !**

$$\mathcal{P}(p,d,b,t,r) = \mathcal{P}(p|d\cancel{b}\cancel{t}\cancel{r})\mathcal{P}(d|b\cancel{t}r)\mathcal{P}(b|\cancel{t}\cancel{r})\mathcal{P}(t|r)\mathcal{P}(r)$$

- The use of belief nets (Bayes nets) alleviates the full table
- As the table grows exponentially …
- But we still need this full table to perform calculations !

7. Joint probability table

- We devise the belief net to represent causality
- Very important property of belief nets … There is no loops
- So we can work from the bottom and go up



P D B T R

- Now we are sure that no descendent are on the left …

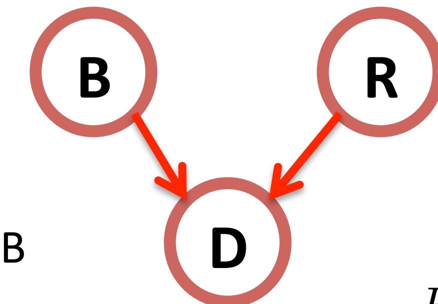
$$\mathcal{P}(p, d, b, t, r) = \mathcal{P}(p|d\cancel{b}\cancel{t}\cancel{r})\mathcal{P}(d|b\cancel{t}r)\mathcal{P}(b|\cancel{t}\cancel{r})\mathcal{P}(t|r)\mathcal{P}(r)$$

7. Joint probability table

- Let's look closer to a sub-problem of this

$\mathcal{P}(B)$	B is true	All
	II	III

A priori probability of B



Experimental observations

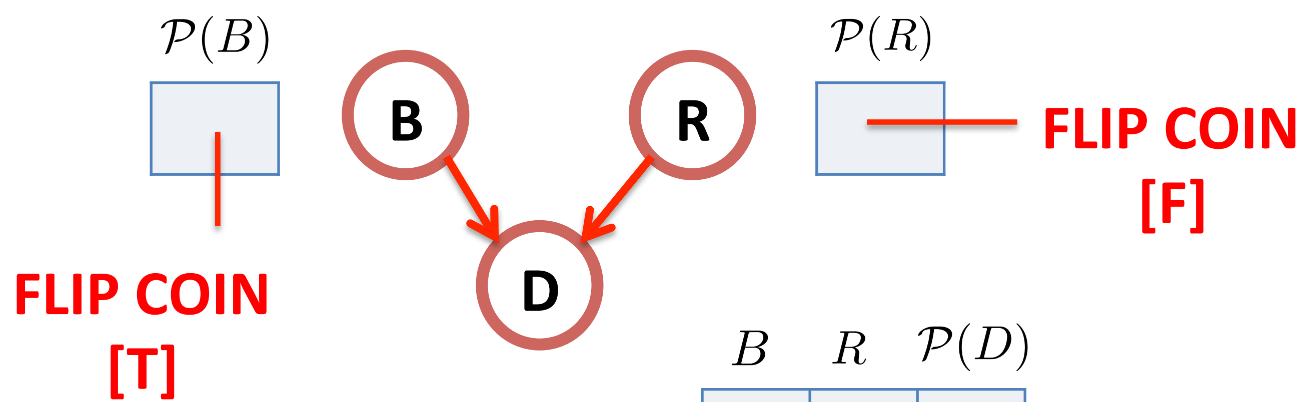
1. **T T T**
2. **F F F**
3. **T T F**

B	R	$\mathcal{P}(D)$	D is true	All
T	T		I	II
T	F			
F	T			
F	F			I

- By making experimental observations, we update probabilities
- By going to infinite, we approximate the real probabilities

7. Joint probability table

- But what we want is to create a **model**
- So imagine that we filled in all probabilities tables



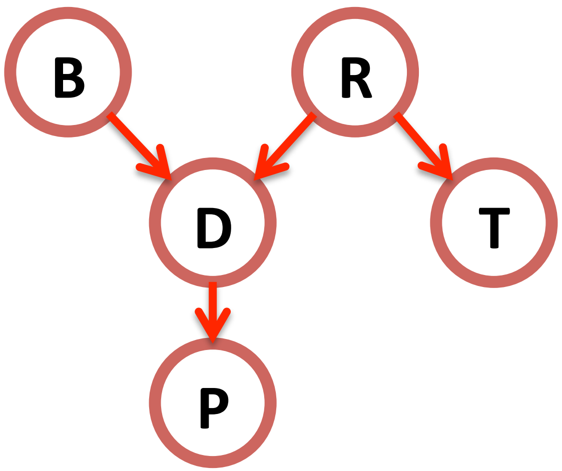
B	R	$\mathcal{P}(D)$
T	T	
T	F	
F	T	
F	F	

FLIP COIN

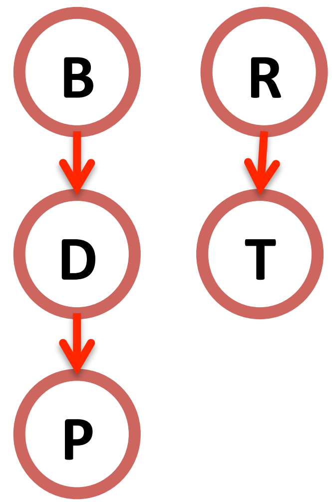
- So here we have an **experimental trial**
- But of course this depends **if the model is right**

7. Joint probability table

You start with a model



But someone says
you got it all wrong



- So how could we decide which one is right ?
- World of **naïve Bayes inference**