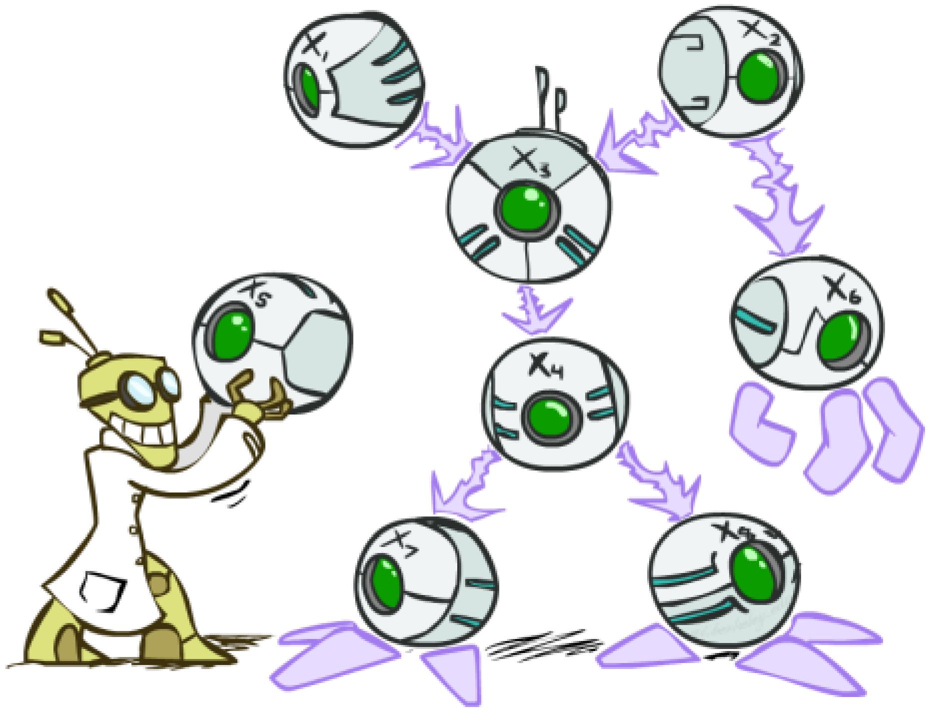
AI: Representation and Problem Solving

Bayes Nets



Instructors: Tuomas Sandholm and Vincent Conitzer

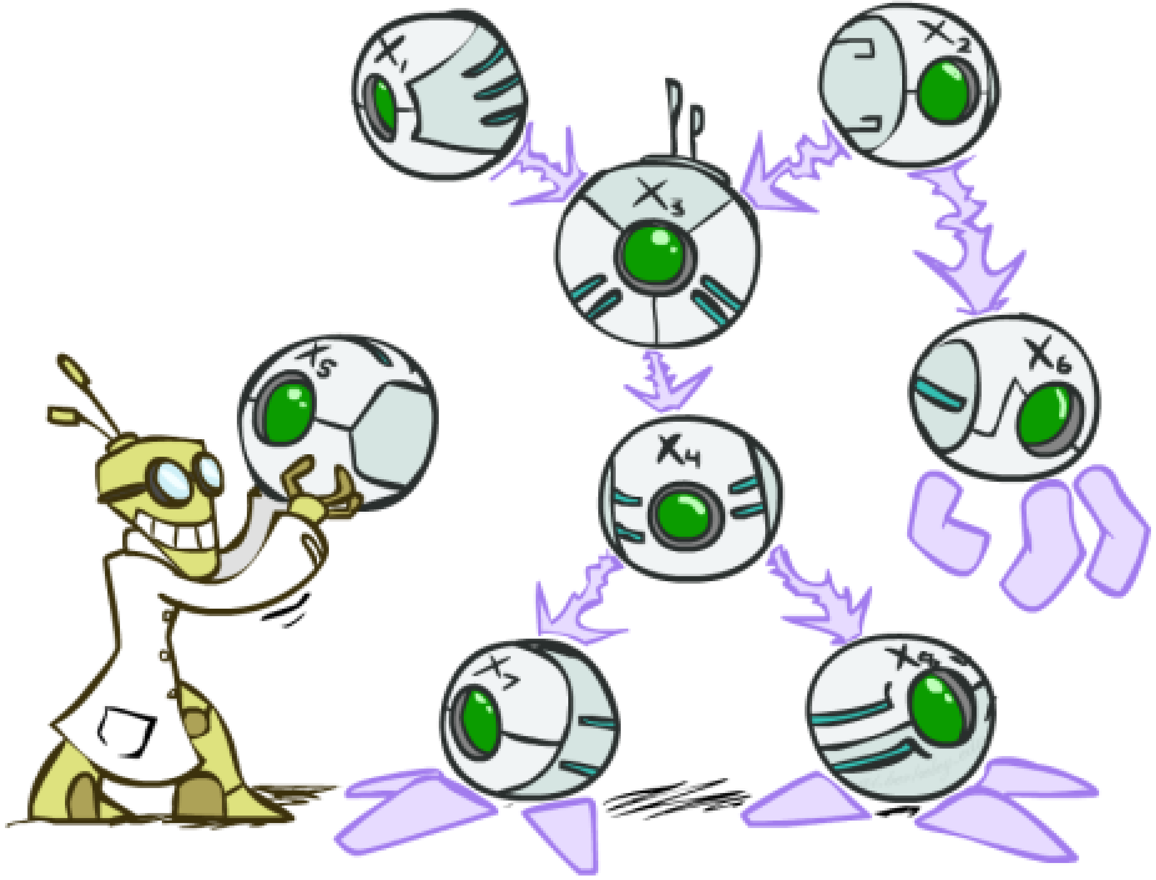
Slide credits: CMU AI and http://ai.berkeley.edu

Announcements

- Midterm 2 Wednesday March 26
  - covers Classical Planning, MDPs, RL, this lecture
  - 80 minutes in class, same general rules
  - cheatsheet should be reasonable size writing

- HW7 (online) due Thursday (but you may want to do it for practice before the midterm)
- P4 is due Monday (March 31)

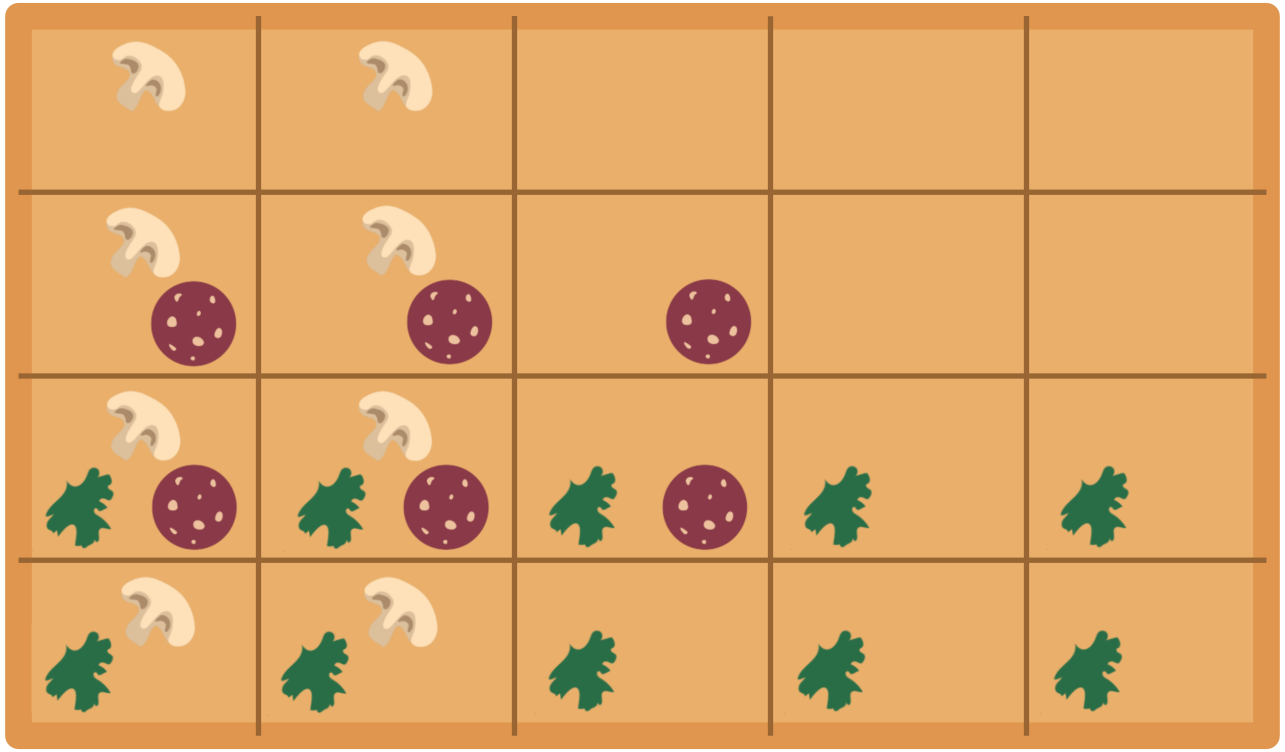
Bayes nets



Omega Pizzeria!

What is the probability of getting a slice with:

1) No mushrooms
2) Spinach and no mushrooms
3) Spinach, when asking for slice with no mushrooms

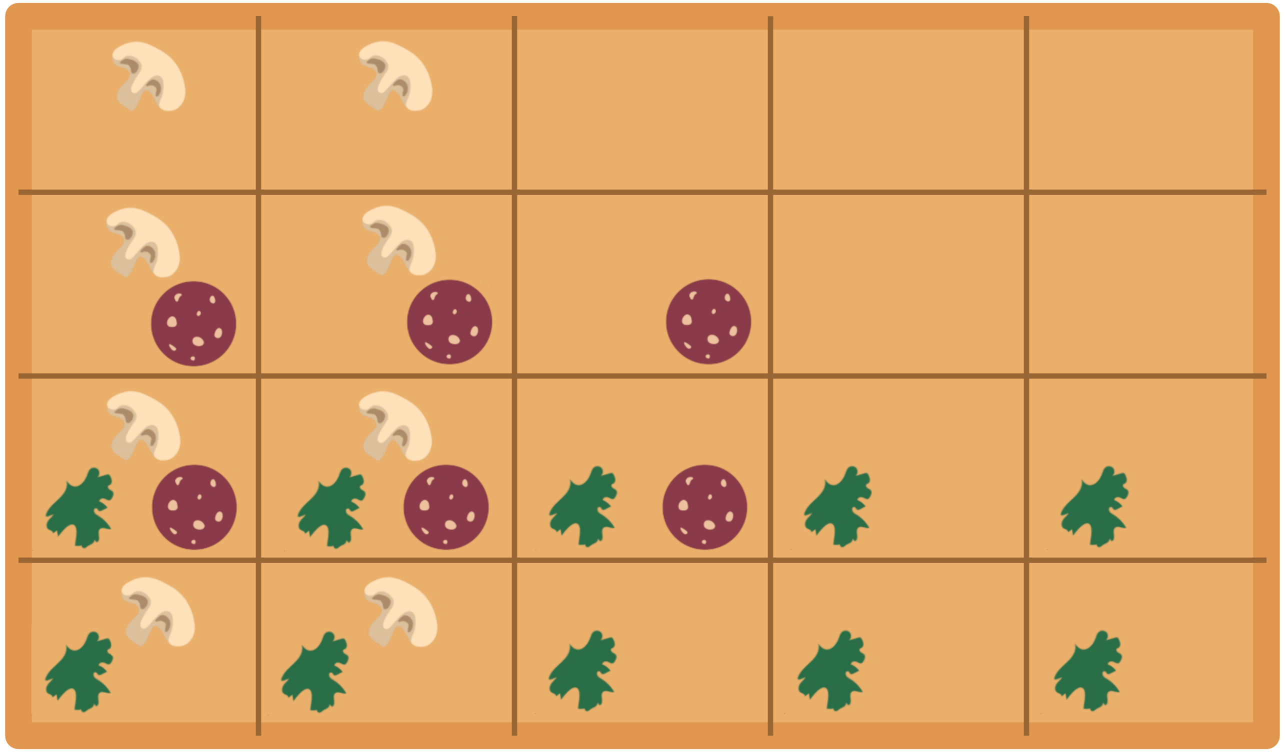


Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Omega Pizzeria!

What is the probability of getting a slice with:

- Mushrooms
- Spinach
- No spinach
- No spinach and mushrooms
- No spinach when asking for no mushrooms
- No spinach when asking for mushrooms
- Spinach when asking for mushrooms
- No mushrooms and no spinach



Notations and conventions in the course

- Random variables:
  - Capitalized
  - Represents all potential outcomes
    - e.g., C
- Outcomes (values):
  - lowercase
  - e.g., $+b$, a_1, a_2, a_3
- Variables for values:
  - lowercase
  - e.g., a

Discrete Probability Distributions

For each random variable

- Discrete outcomes
- Disjoint outcomes
- Accounts for entire event space
- Not always binary

Discrete Random Variables (and their domains)

$A \in \{a_1, a_2, a_3\}$

$B \in \{+b, -b\}$

$C \in \{+c, -c\}$

Probability notation

- In this lecture, we use uppercase letters such as A to denote random variables
- For a random variable A taking values $\{a_1, a_2, a_3\}$:
- $P(A) = \begin{pmatrix} 0.1 \\ 0.5 \\ 0.4 \end{pmatrix}$
- represents the set of probabilities for each value that A can take on (this is a *function* mapping values of A to numbers that sum to one)
- Conversely, we will use lowercase to denote a specific *value* of A (or a *variable* taking such a value, i.e., for above example $a \in \{a_1, a_2, a_3\}$)
- $P(A = a_1)$ or just $P(a_1)$ refers to a *number* (the corresponding entry of $P(A)$)

Probability notation

- Given two random variables: B with values in $\{+b, -b\}$ and C with values in $\{+c, -c\}$:

  $P(B, C)$ refers to the *joint distribution*, i.e., a set of 4 possible values, i.e., a function mapping $(+b, +c), (+b, -c), (-b, +c), \dots$ to corresponding probabilities

  $P(+b, -c)$ is a *number:* probability that $B = +b$ and $C = -c$

  $P(B, -c)$ is a set of 2 values, the probabilities for all values of B together with the given value $C = -c$, i.e., it is a function mapping $+b, -b$ to numbers (note: *not* probability distribution, it will not sum to one – **why?**)

Probability notation

- Three random variables: $A \in \{a_1, a_2, a_3\}, B \in \{+b, -b\}, C \in \{+c, -c\}$
- $P(+b, +c) = \sum_{a \in \{a_1, a_2, a_3\}} P(a, +b, +c)$
- $P(+b, C) = \sum_{a \in \{a_1, a_2, a_3\}} P(a, +b, C)$

Joint probability distribution

- Table representing all values

$A \in \{+a, -a\}$

$B \in \{+b, -b\}$

$C \in \{+c, -c\}$

A	B	C	P(A=a, B=b, C=c)
$+a$	$+b$	$+c$	.1
$+a$	$+b$	$-c$	.1
$+a$	$-b$	$+c$	.1
$+a$	$-b$	$-c$	.1
$-a$	$+b$	$+c$	.1
$-a$	$+b$	$-c$	.1
$-a$	$-b$	$+c$	.2
$-a$	$-b$	$-c$	.2

Marginalization

- For random variables B, C with joint distribution $P(B, C)$, the marginal probabilities $P(B), P(C)$ are as follows
  - $P(B) = \sum_{c \in \{+c, -c\}} P(B, c)$
  - $P(C) = \sum_{b \in \{+b, -b\}} P(b, C)$

- $P(+b) = \sum_{c \in \{+c, -c\}} P(+b, c) = P(+b, +c) + P(+b, -c)$

Marginalization from table

A	B	C	P(A=a, B=b, C=c)
+a	$+b$	$+c$	.1
+a	$+b$	$-c$	.1
+a	$-b$	$+c$	.1
+a	$-b$	$-c$	.1
−a	$+b$	$+c$	.1
−a	$+b$	$-c$	.1
−a	$-b$	$+c$	.2
−a	$-b$	$-c$	.2

$A \in \{+a, -a\}$

$B \in \{+b, -b\}$

$C \in \{+c, -c\}$

What is $P(B)$?

Marginalization from table

A	B	C	P(A=a, B=b, C=c)
+a	$+b$	$+c$	.1
+a	$+b$	$-c$	.1
+a	$-b$	$+c$	.1
+a	$-b$	$-c$	.1
−a	$+b$	$+c$	.1
−a	$+b$	$-c$	.1
−a	$-b$	$+c$	.2
−a	$-b$	$-c$	.2

$A \in \{+a, -a\}$

$B \in \{+b, -b\}$

$C \in \{+c, -c\}$

Marginalization to get $P(B)$: aggregating rows that share the same value of B

Conditional probability

- The conditional probability $P(B \mid C)$ (the conditional probability of B given C) is defined as

$$P(B|C) = \frac{P(B,C)}{P(C)}$$

- $P(+b \mid +c) = \frac{P(+b,+c)}{P(+c)}$
- $P(-b \mid +c) = \frac{P(-b,+c)}{P(+c)}$

Conditional probability from table

A	B	C	P(A=a, B=b, C=c)
+a	$+b$	$+c$	.1
+a	$+b$	$-c$	.1
+a	$-b$	$+c$	.1
+a	$-b$	$-c$	.1
−a	$+b$	$+c$	.1
−a	$+b$	$-c$	.1
−a	$-b$	$+c$	.2
−a	$-b$	$-c$	.2

$A \in \{+a, -a\}$

$B \in \{+b, -b\}$

$C \in \{+c, -c\}$

What is $P(B| + c)$?

Conditional probability from table

A	B	C	P(A=a, B=b, C=c)
+a	$+b$	$+c$	
+a	$+b$	$-c$	
+a	$-b$	$+c$	
+a	$-b$	$-c$	
−a	$+b$	$+c$	
−a	$+b$	$-c$	
−a	$-b$	$+c$	
−a	$-b$	$-c$	

$A \in \{+a, -a\}$

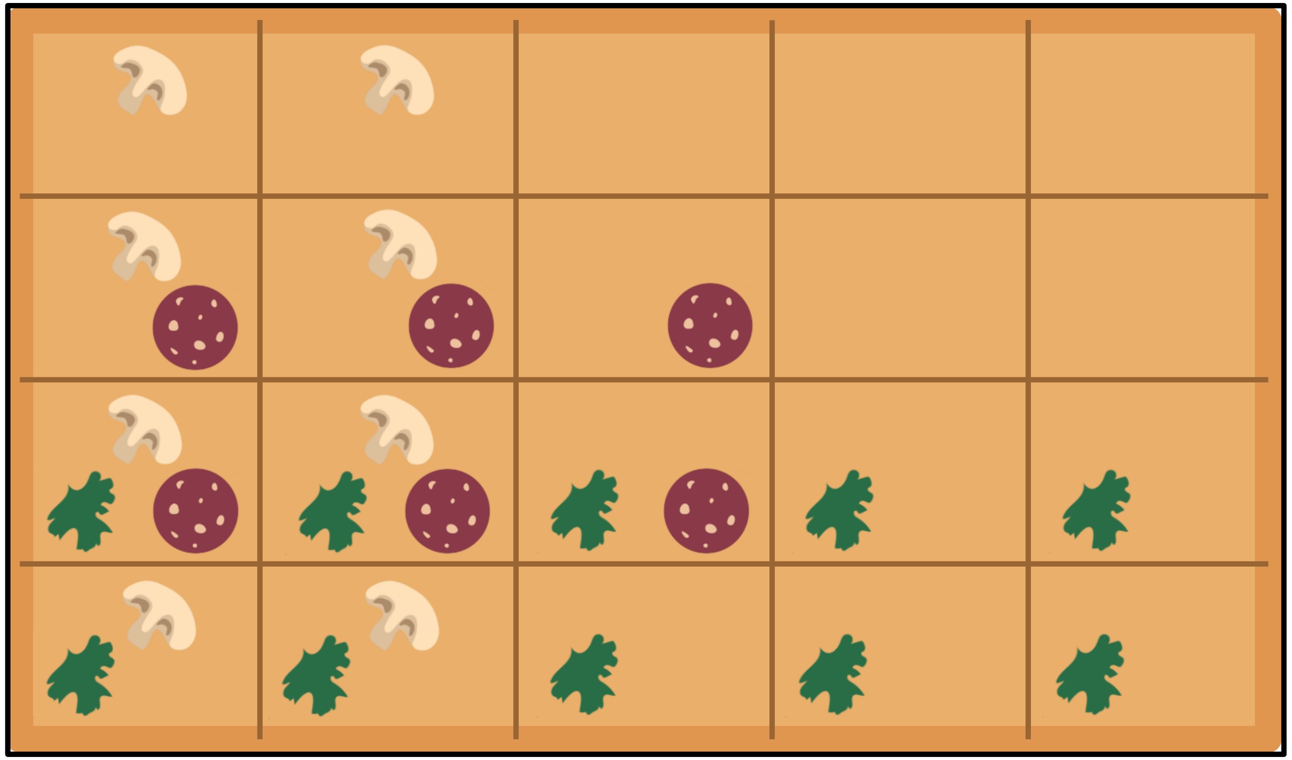
$B \in \{+b, -b\}$

$C \in \{+c, -c\}$

We restrict our attention to rows that satisfy the "given" condition, and then normalize the values so that they sum to 1 (form a distribution)

Discrete probability distributions

- Joint distribution $P(M, S, R)$



Discrete Random Variables

(and their domains)

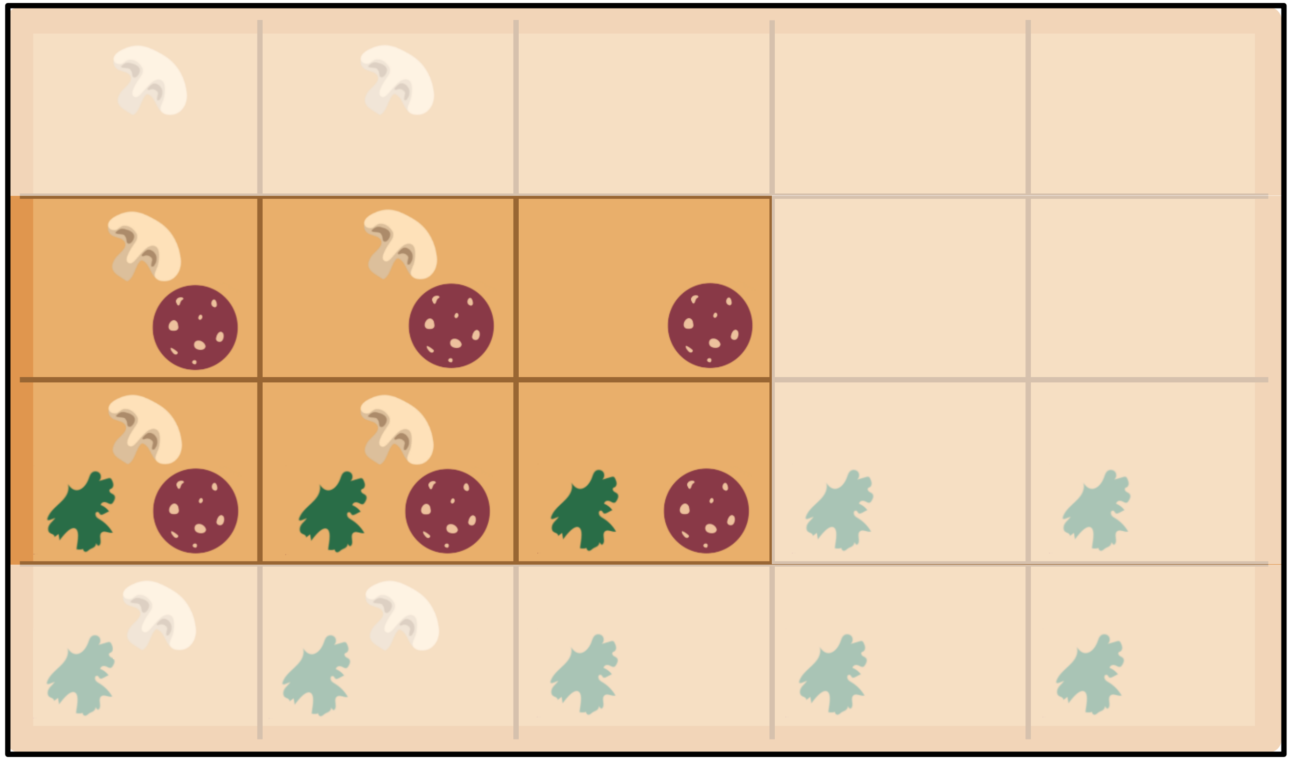
$M \in \{+m, -m\}$

$S \in \{+s, -s\}$

$R \in \{+r, -r\}$

Discrete probability distributions

- Marginal distribution $P(+r)$



Discrete Random Variables

(and their domains)

$M \in \{+m, -m\}$

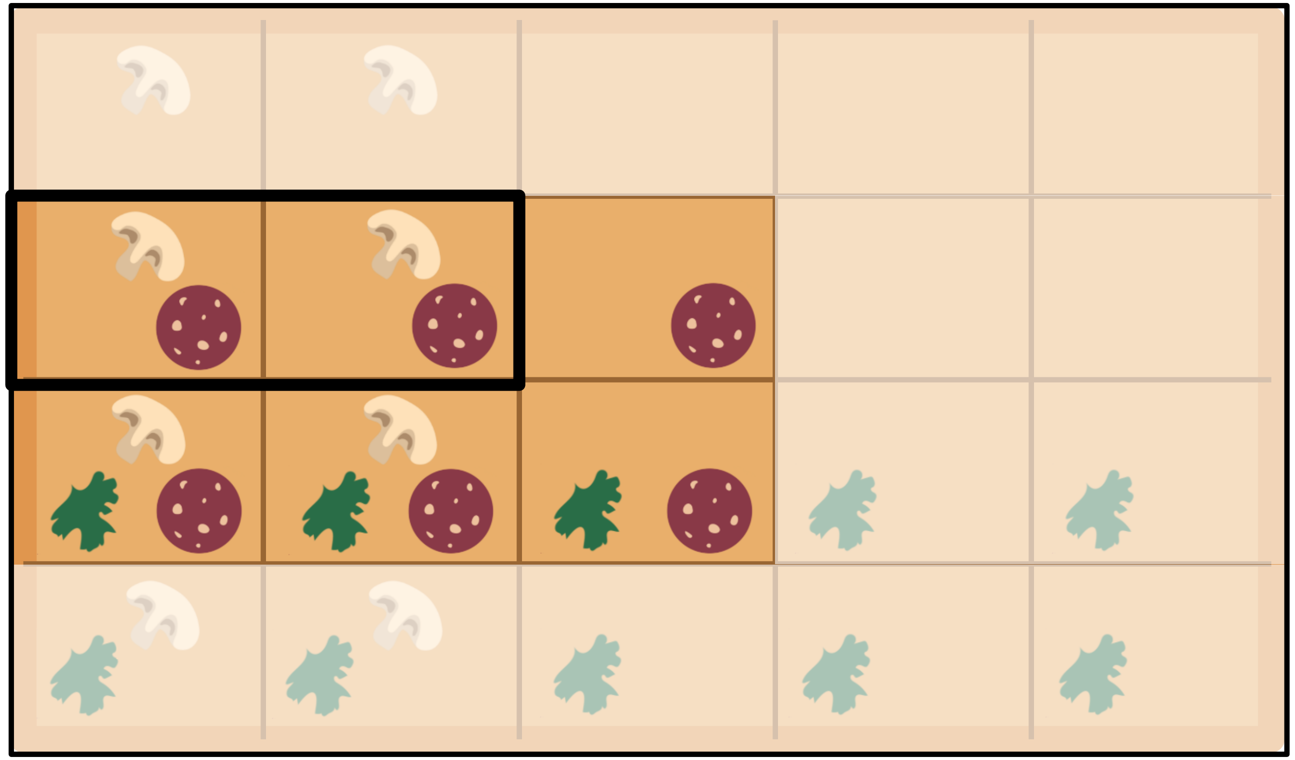
$S \in \{+s, -s\}$

$R \in \{+r, -r\}$

Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Discrete probability distributions

- Conditional distribution $P(+m, -s \mid +r)$



Discrete Random Variables

(and their domains)

$M \in \{+m, -m\}$

$S \in \{+s, -s\}$

$R \in \{+r, -r\}$

Poll 1

- Which of the following probability tables sum to one?
- Select all that apply

i. $P(A \mid b)$

ii. $P(A, b, C)$

iii. $P(A, C \mid b)$

iv. $P(a, c \mid b)$

Bayes rule

$$P(B|C) = \frac{P(C|B)P(B)}{P(C)}$$

Intuition: follow the expression for conditional probability

$$P(B|C) = \frac{P(B,C)}{P(C)} = \frac{P(C|B)P(B)}{P(C)}$$

When is it used? Go from $P(B|C)$ to $P(C|B)$

Poll 2

- I want to know if I have come down with a rare strain of flu (occurring in only 1/10,000 people). There is an "accurate" test for the flu: if I have the flu, it will tell me I have 99% of the time, and if I do not have it, it will tell me I do not have it 99% of the time. I go to the doctor and test positive. What is the probability I have this flu?
  - (A) ≈ 99%
  - (B) ≈ 10%
  - (C) ≈ 1%
  - (D) ≈ 0.1%

Probability Tools Summary

1. Definition of conditional probability $$P(A|B) = \frac{P(A,B)}{P(B)}$$
2. Product Rule $$P(A,B) = P(A|B)P(B)$$
3. Bayes' theorem $$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$
4. Chain Rule $$P(X_1, \dots, X_N) = \prod_{n=1}^{N} P(X_n \mid X_1, \dots, X_{n-1})$$

Additional probability tools

- Marginalization (law of total probability) (summing out)

$$P(A) = \sum_b \sum_c P(A, b, c)$$

- Normalization

$$P(B \mid a) = \frac{P(a, B)}{P(a)}$$

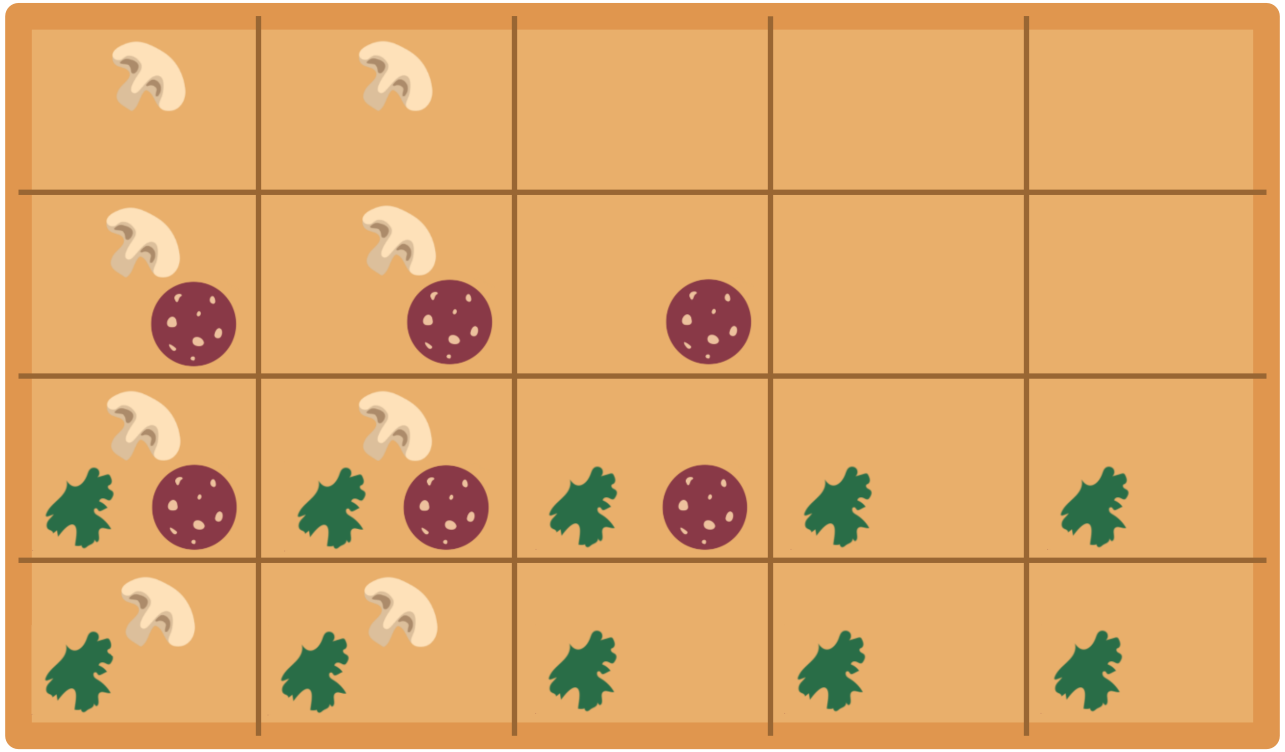
$$P(B \mid a) \propto P(a, B)$$

$$P(B \mid a) = \frac{1}{z} P(a, B)$$

$$z = P(a) = \sum_b P(a, b)$$

More practice

What is the probability of getting a slice with:



Icons: CC, https://openclipart.org/detail/296791/pizza-slice

Answer queries from joint distribution

- You can answer all of these questions:

$P(M)$

$+m$	12/20
$-m$	

$P(S)$

$+s$	
$-s$	

$P(M,S)$

$+m+s$	
$+m-s$	
$-m+s$	
$-m-s$	

$P(M|+s)$

$+m$	
$-m$	

$P(M|-s)$

$+m$	
$-m$	

$P(S|+m)$

$+s$	
$-s$	6/12

$P(S|-m)$

$+s$	
$-s$	

More practice

- P(Weather)?
- P(Weather | winter)?
- P(Weather | winter, hot)?

Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

- P(Weather)?

Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

- P(Weather | winter)?

Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

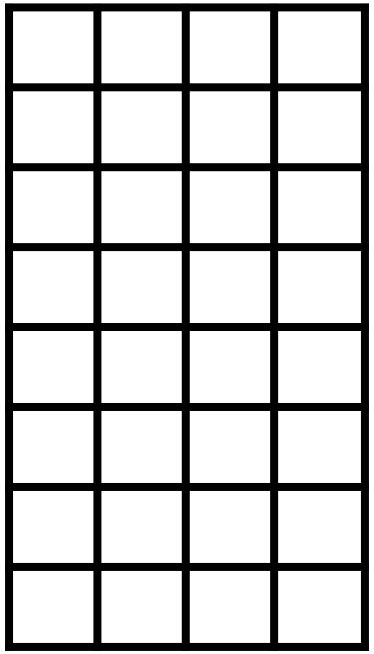
- P(Weather | winter, hot)?

Season	Temp	Weather	P(S, T, W)
summer	hot	sun	0.30
summer	hot	rain	0.05
summer	cold	sun	0.10
summer	cold	rain	0.05
winter	hot	sun	0.10
winter	hot	rain	0.05
winter	cold	sun	0.15
winter	cold	rain	0.20

Answer Any Query from Joint Distribution

- Joint distributions are the best!

Joint



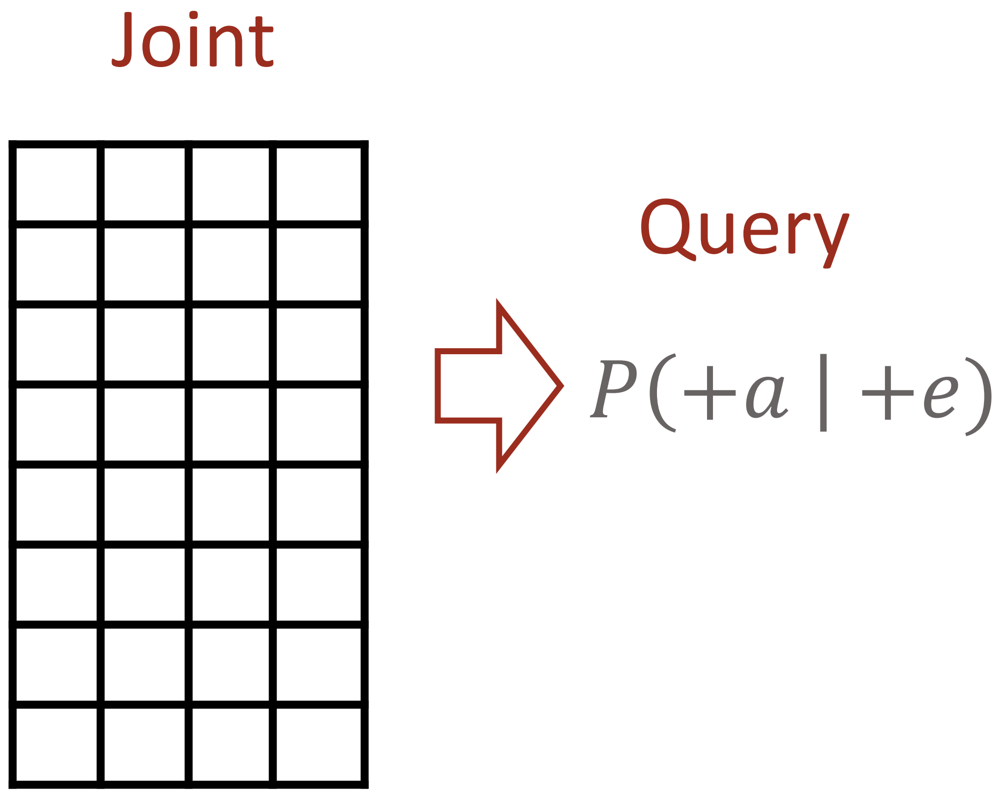
Query

$$P(+q_1, +q_2 \mid +e_1, +e_2, +e_3)$$

Joint distributions

- Joint distributions are the best!

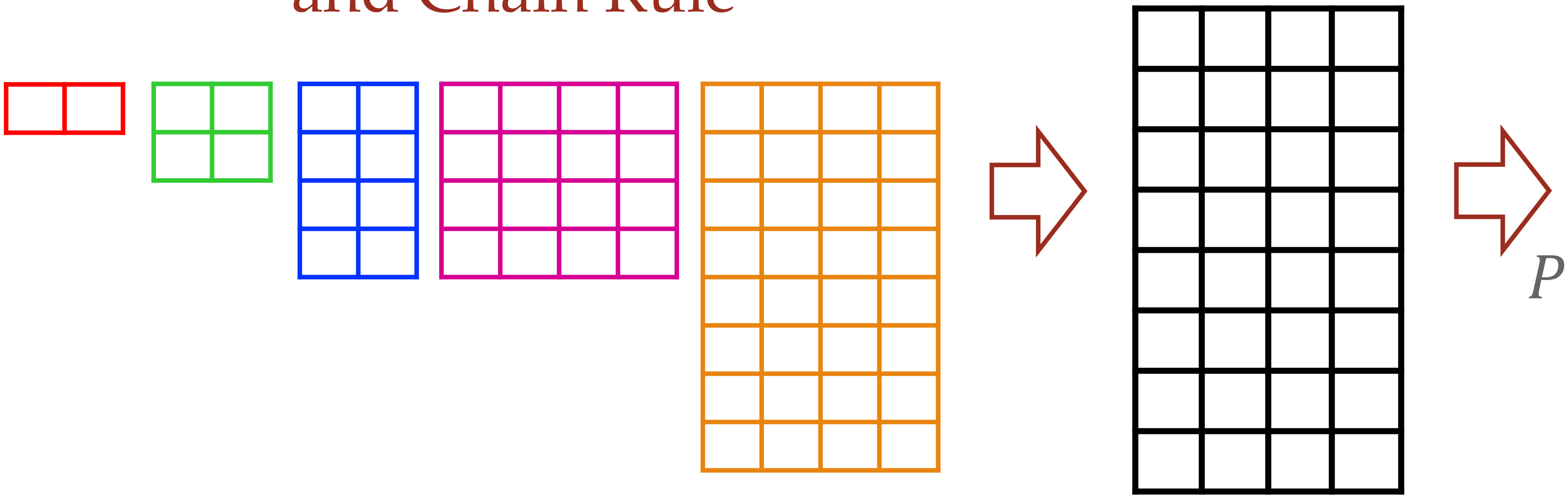
- Problems with joints
  - We aren't given the joint table
    - Usually some set of conditional probability tables



Build joint distribution using CPTs

Conditional Probability Tables and Chain Rule

Joint



Query

$P(+a \mid +e)$

$P(A)$ $P(B|A)$ $P(C|A,B)$ $P(D|A,B,C)$ $P(E|A,B,C,D)$

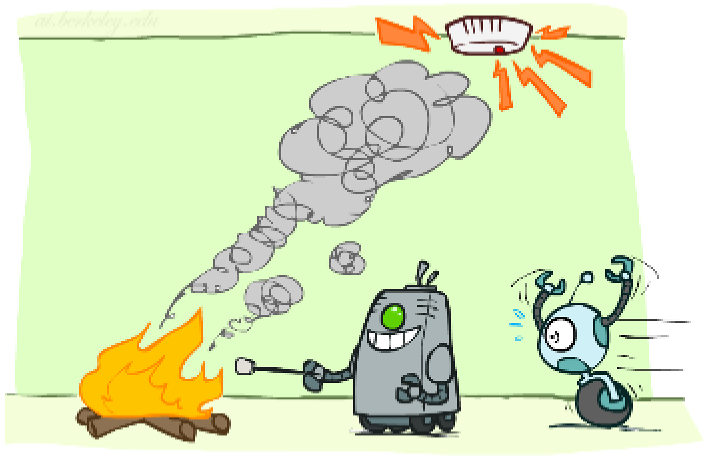
Build joint distribution using chain rule

Two tools to construct joint distribution

1. Product rule
   - $P(A, B) = P(A \mid B)P(B)$
   - $P(A, B) = P(B \mid A)P(A)$
2. Chain rule
   - $P(X_1, X_2, \dots, X_n) = \prod_i P(X_i \mid X_1, \dots, X_{i-1})$
   - $P(A, B, C) = P(A)P(B \mid A)P(C \mid A, B)$ for ordering *A, B, C*
   - $P(A, B, C) = P(A)P(C \mid A)P(B \mid A, C)$ for ordering *A, C, B*
   - $P(A, B, C) = P(C)P(B \mid C)P(A \mid C, B)$ for ordering *C, B, A*
   - …

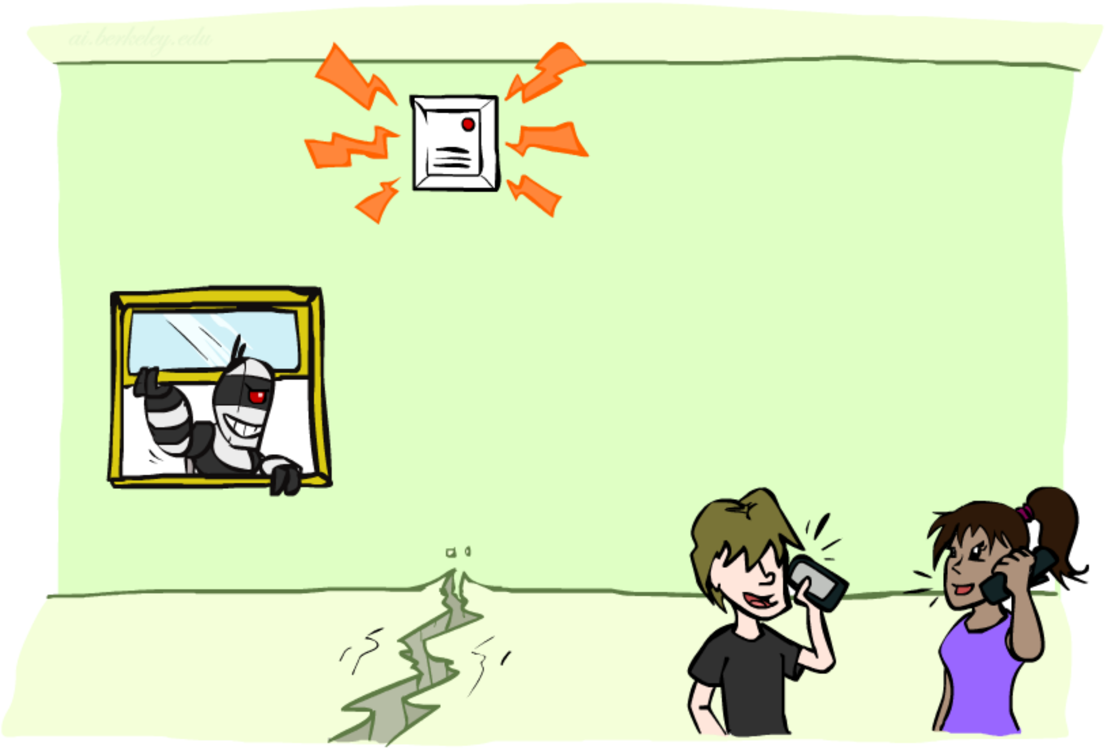
Build joint distribution using chain rule

- Binary random variables
  - Fire
  - Smoke
  - Alarm



Poll 3

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!



How many different ways can we write the chain rule?

A. 1

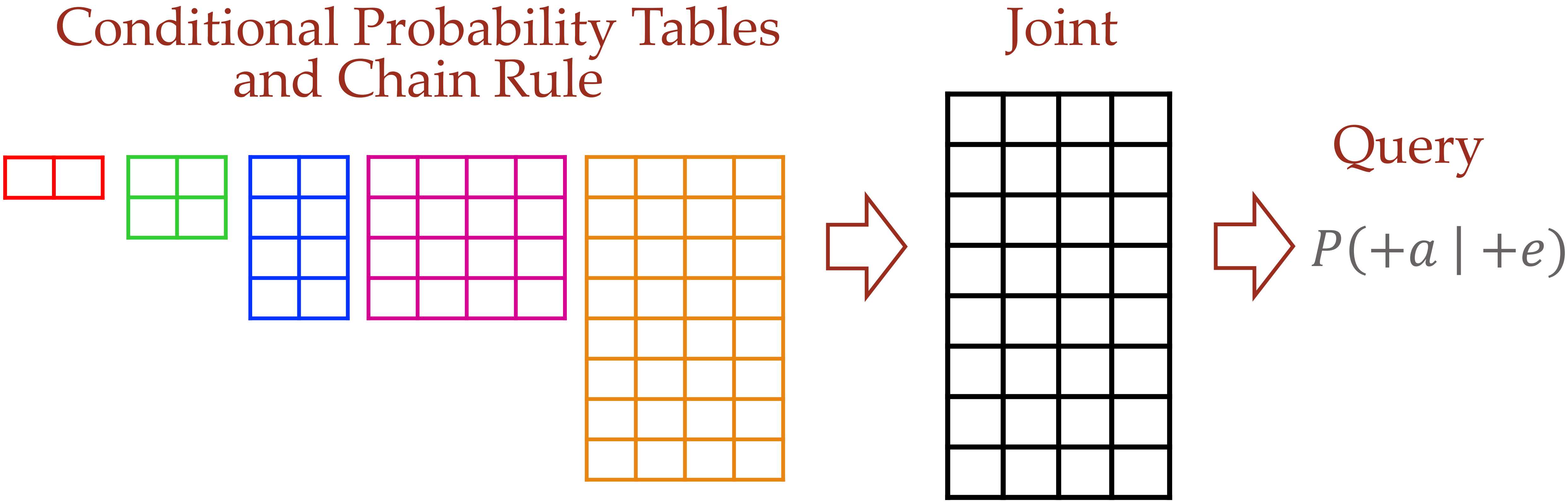
B. 5

C. 5 *choose* 5

D. 5!

E. 5^5

Build joint distribution using chain rule



$P(A)\ P(B|A)\ P(C|A,B)\ P(D|A,B,C)\ P(E|A,B,C,D)$

Queries from CPTs

- Process to go from (specific) conditional probability tables to query

1. Construct the joint distribution
   1. Product Rule or Chain Rule
2. Answer query from joint
   1. Definition of conditional probability
   2. Law of total probability (marginalization, summing out)

Queries from CPTs

- Bayes' rule as an example
- Given: $P(E|Q)$, $P(Q)$ Query: $P(Q \mid +e)$

1. Construct the joint distribution
   1. Product Rule or Chain Rule

$$P(E,Q) = P(E|Q)P(Q)$$

2. Answer query from joint
   1. Definition of conditional probability

$$P(Q \mid +e) = \frac{P(+e,Q)}{P(+e)}$$

   2. Law of total probability (marginalization, summing out)

- $P(Q \mid +e) = \frac{P(+e,Q)}{\sum_q P(+e,q)}$

Bayes nets

- Graphical representation of conditional probability tables
- One node per random variable
- Directed acyclic graph
- Each node corresponds to a conditional probability distribution

Bayes nets

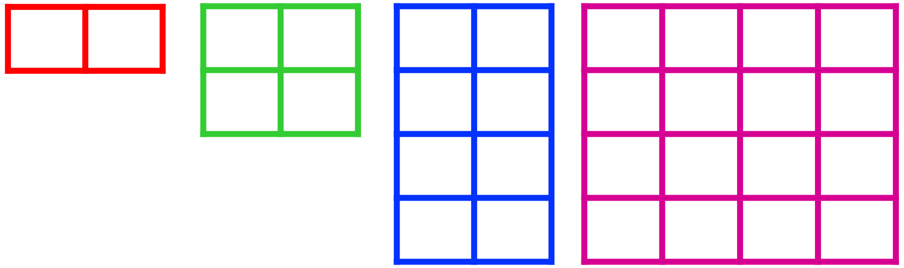
- Each node corresponds to a conditional probability distribution
- Bayes nets encode joint distributions as product of conditional distributions on each variable

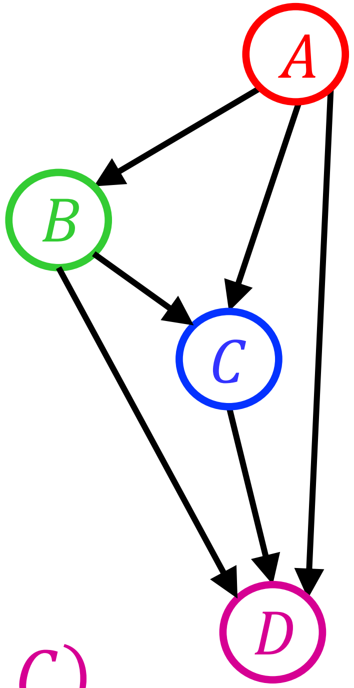
$$P(\text{node} \mid \text{parents}(\text{node}))$$

- Encode joint distributions as product of conditional distributions on each variable

$$P(X_1, \ldots, X_N) = \prod_i P(X_i \mid Parents(X_i))$$

Bayes nets

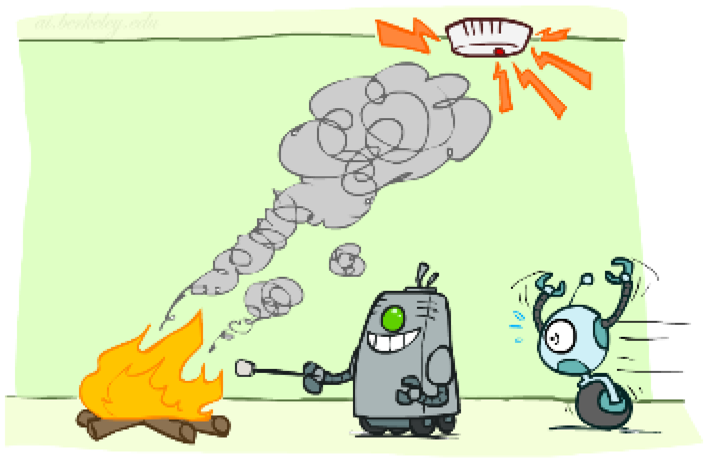




$$P(A,B,C,D) = P(A)\, P(B|A)\, P(C|A,B)\, P(D|A,B,C)$$

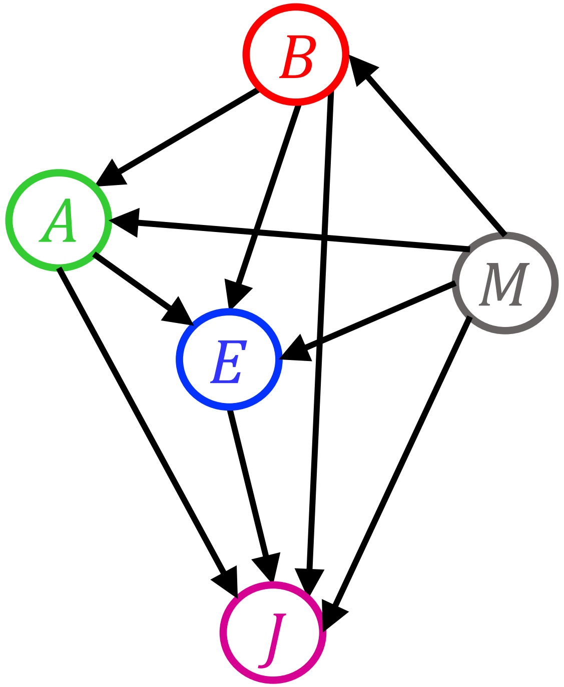
Build Bayes Net Using Chain Rule

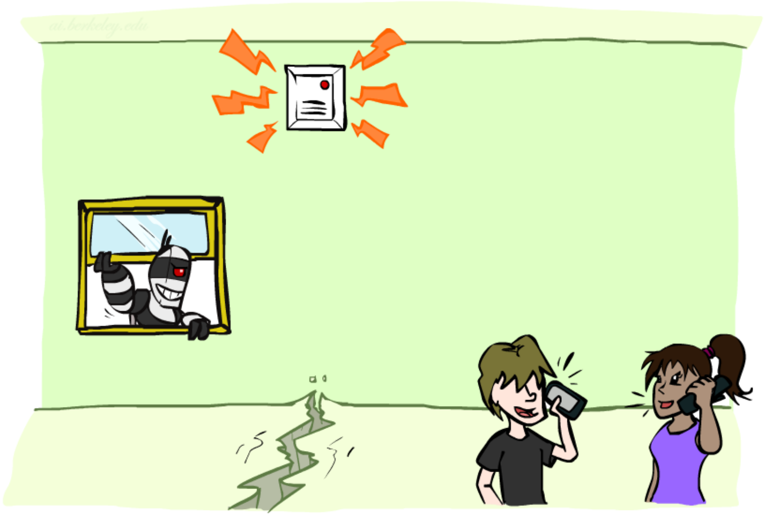
- Binary random variables
  - Fire
  - Smoke
  - Alarm



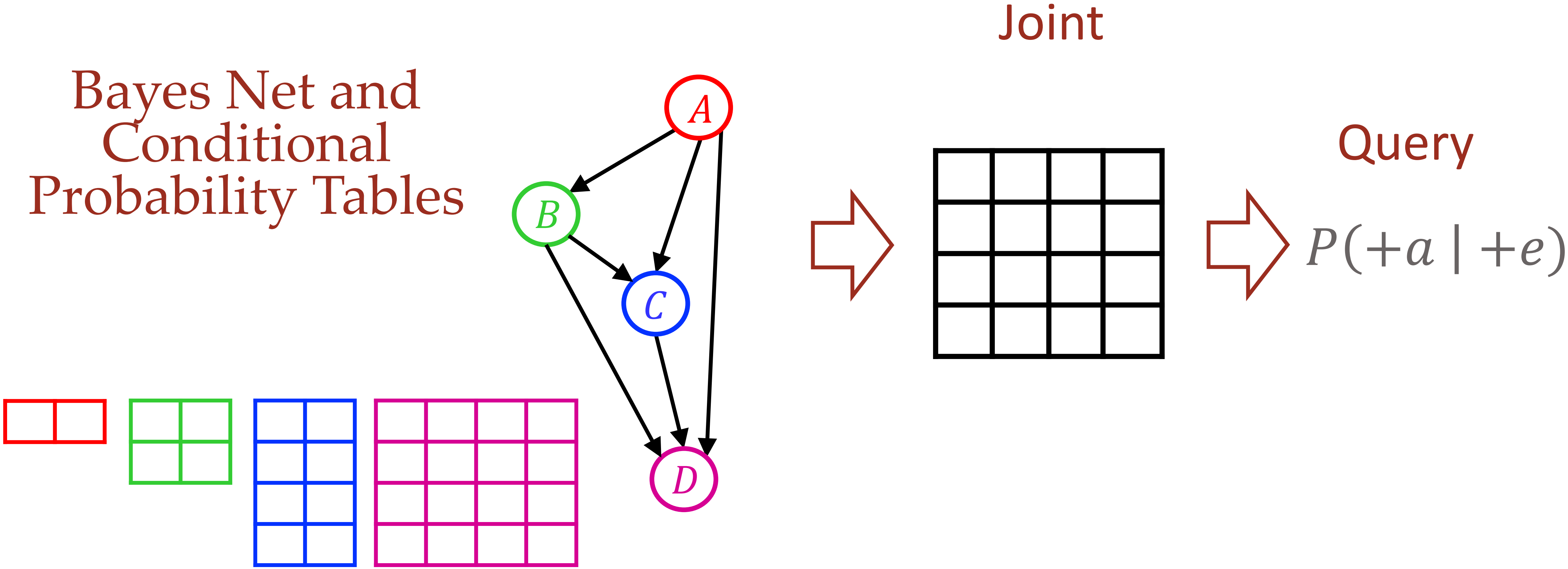
Question

- Variables
  - B: Burglary
  - A: Alarm goes off
  - M: Mary calls
  - J: John calls
  - E: Earthquake!





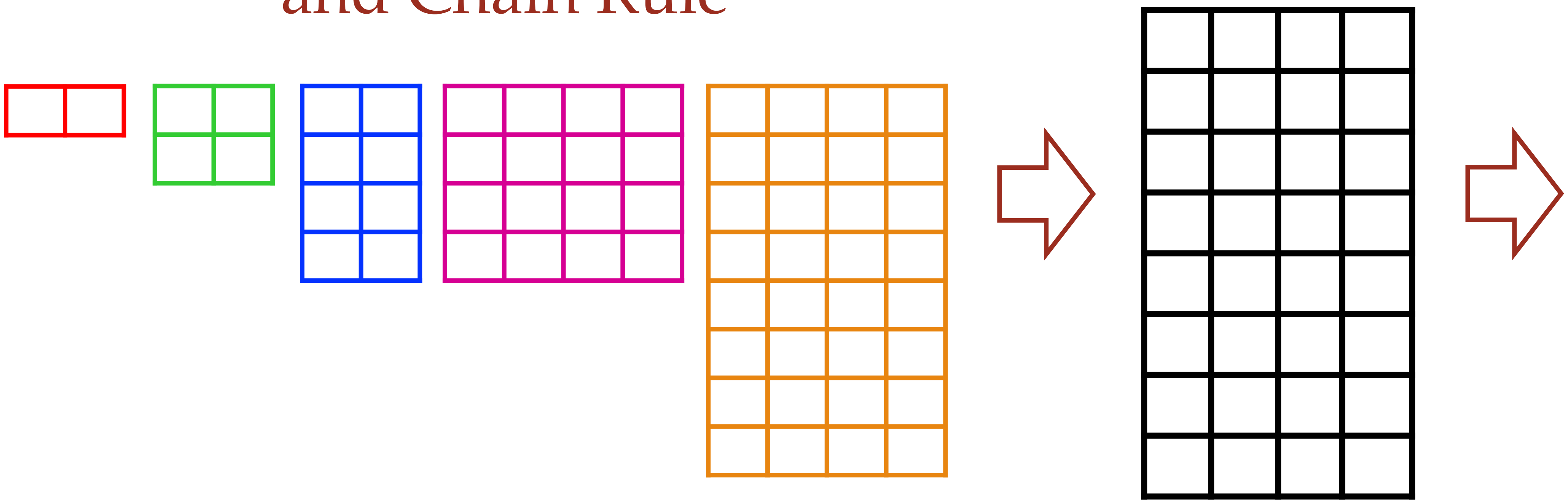
Answer Any Query from Bayes Net



Answer any query from CPTs

Conditional Probability Tables and Chain Rule

Joint



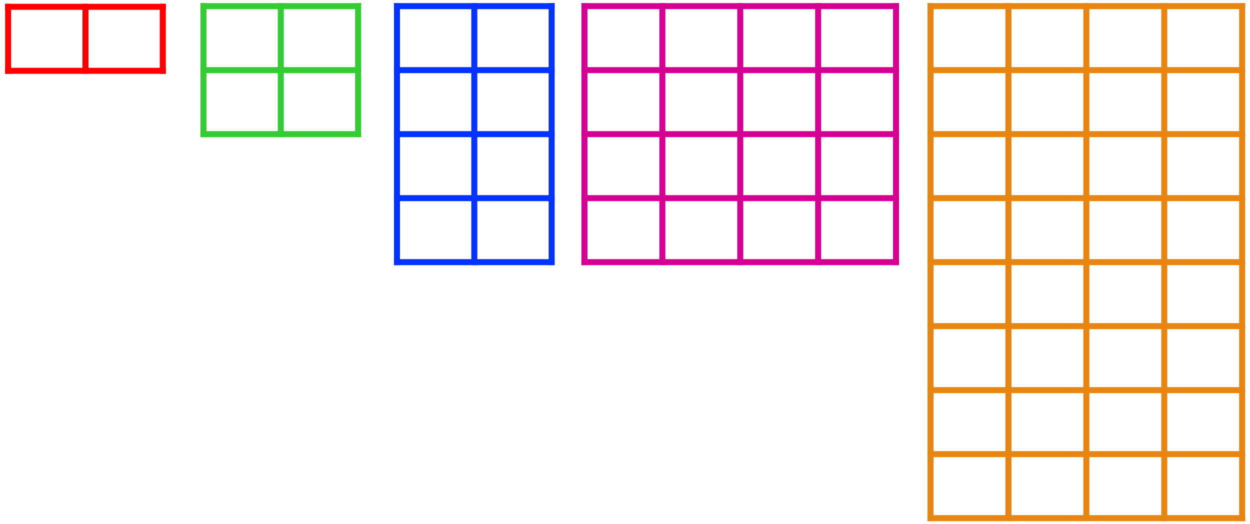
Query

$P(+a \mid +e)$

$P(A)$ $P(B|A)$ $P(C|A,B)$ $P(D|A,B,C)$ $P(E|A,B,C,D)$

Answer Any Query from Conditional Probability Tables

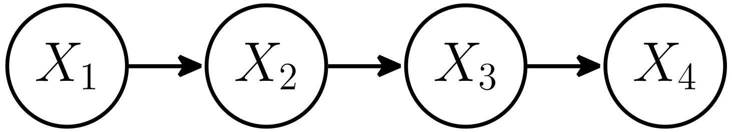
Conditional Probability Tables and Chain Rule



$P(A)$ $P(B|A)$ $P(C|A,B)$ $P(D|A,B,C)$ $P(E|A,B,C,D)$

- Problems
  - Huge
    - n variables with d values
    - d^n entries
  - We aren't given the right tables

Do we need the full chain rule?



- Some Bayes Nets represent simpler distributions

- $p(X_1, X_2, X_3, X_4) = p(X_1)p(X_2|X_1)p(X_3|X_1, X_2)p(X_4|X_1, X_2, X_3)$

$$= p(X_1)p(X_2|X_1)p(X_3|X_2)p(X_4|X_3)$$

The power of conditional independencies

- For a Bayes net $P(X_1, \ldots, X_N) = \prod_i P(X_i | \, Parents(X_i))$

- When $Parents(X_i)$ is small, the conditional probability tables are much smaller
  - Makes them easier to estimate from data

- This and much more in the next class