Announcements

- **Midterm is next week, 3/14 from 7-9PM**
  - Scope is everything from the first lecture up until and including today’s (3/7) lecture
  - Some important policies:
    - One double sided handwritten crib sheet is allowed
    - NO calculators are allowed
  - There will be an exam review session this Saturday from 2-5PM at Hearst Annex A1
  - There will be a practice midterm
  - More information (rooms, slides, etc.) will be posted tonight in a large piazza post
CS 188: Artificial Intelligence

Bayes’ Nets: Independence

Instructors: Sergey Levine and Anca Dragan --- University of California, Berkeley

[These slides were created by Dan Klein and Pieter Abbeel, and edited by Sergey Levine. All CS188 materials are available at http://ai.berkeley.edu.]
Probability Recap

- **Conditional probability**
  \[ P(x|y) = \frac{P(x, y)}{P(y)} \]

- **Product rule**
  \[ P(x, y) = P(x|y)P(y) \]

- **Chain rule**
  \[ P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \ldots = \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \]

- **X, Y independent if and only if:**
  \[ \forall x, y : P(x, y) = P(x)P(y) \]

- **X and Y are conditionally independent given Z if and only if:**
  \[ X \independent Y|Z \quad \forall x, y, z : P(x, y|z) = P(x|z)P(y|z) \]
A Bayes’ net is an efficient encoding of a probabilistic model of a domain

Questions we can ask:

- Inference: given a fixed BN, what is $P(X \mid e)$?
- Representation: given a BN graph, what kinds of distributions can it encode?
- Modeling: what BN is most appropriate for a given domain?
Bayes’ Net Semantics

- A directed, acyclic graph, one node per random variable
- A conditional probability table (CPT) for each node
  - A collection of distributions over $X$, one for each combination of parents’ values
    \[
    P(X|a_1 \ldots a_n)
    \]
- Bayes’ nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:
    \[
    P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\text{parents}(X_i))
    \]
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) =
\]

<table>
<thead>
<tr>
<th>B</th>
<th>P(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+b</td>
<td>0.001</td>
</tr>
<tr>
<td>-b</td>
<td>0.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E</th>
<th>P(E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+e</td>
<td>0.002</td>
</tr>
<tr>
<td>-e</td>
<td>0.998</td>
</tr>
</tbody>
</table>

| A  | J   | P(J|A) |
|----|-----|------|
| +a | +j  | 0.9  |
| +a | -j  | 0.1  |
| -a | +j  | 0.05 |
| -a | -j  | 0.95 |

| A  | M   | P(M|A) |
|----|-----|-------|
| +a | +m  | 0.7   |
| +a | -m  | 0.3   |
| -a | +m  | 0.01  |
| -a | -m  | 0.99  |

| B  | E  | A  | P(A|B,E) |
|----|----|----|----------|
| +b | +e | +a | 0.95     |
| +b | +e | -a | 0.05     |
| +b | -e | +a | 0.94     |
| +b | -e | -a | 0.06     |
| -b | +e | +a | 0.29     |
| -b | +e | -a | 0.71     |
| -b | -e | +a | 0.001    |
| -b | -e | -a | 0.999    |
Example: Alarm Network

\[
P(+b, -e, +a, -j, +m) = P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) = 0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7
\]
Size of a Bayes’ Net

- How big is a joint distribution over N Boolean variables?
  \[ 2^N \]

- How big is an N-node net if nodes have up to k parents?
  \[ O(N * 2^{k+1}) \]

- Both give you the power to calculate
  \[ P(X_1, X_2, \ldots X_n) \]

- BNs: Huge space savings!

- Also easier to elicit local CPTs

- Also faster to answer queries (coming)
Bayes’ Nets

- Representation
  - Conditional Independences
  - Probabilistic Inference
  - Learning Bayes’ Nets from Data
Conditional Independence

- X and Y are independent if

\[ \forall x, y \quad P(x, y) = P(x)P(y) \quad \implies \quad X \perp Y \]

- X and Y are conditionally independent given Z

\[ \forall x, y, z \quad P(x, y|z) = P(x|z)P(y|z) \quad \implies \quad X \perp Y|Z \]

- (Conditional) independence is a property of a distribution

- Example: \( Alarm \perp Fire|Smoke \)
Bayes Nets: Assumptions

- Assumptions we are required to make to define the Bayes net when given the graph:
  \[ P(x_i|x_1 \cdots x_{i-1}) = P(x_i|\text{parents}(X_i)) \]

- Beyond above “chain rule → Bayes net” conditional independence assumptions
  - Often additional conditional independences
  - They can be read off the graph

- Important for modeling: understand assumptions made when choosing a Bayes net graph
Example

- Conditional independence assumptions directly from simplifications in chain rule:

\[ P(X, Y, Z, W) = P(X)P(Y|X)P(Z|Y)P(W|Z) \]
\[ = P(X)P(Y|X)P(Z|X, Y)P(W|X, Y, Z) \]
\[ Z \perp X|Y \quad W \perp X, Y|Z \]

- Additional implied conditional independence assumptions?

\[ P(W|X, Y) = \frac{P(W, X, Y)}{P(X, Y)} = \frac{\sum_Z P(W, X, Y, Z)}{P(X, Y)} = \sum_Z \frac{P(X)P(Y|X)P(Z|Y)P(W|Z)}{P(X)P(Y|X)} \]
\[ = \sum_Z P(Z|Y)P(W|Z) = \sum_Z P(Z|Y)P(W|Z, Y) = \sum_Z P(Z, W|Y) = P(W|Y) \]
Independence in a BN

- **Important question about a BN:**
  - Are two nodes independent given certain evidence?
  - If yes, can prove using algebra (tedious in general)
  - If no, can prove with a counter example
  - Example:

```
X -> Y -> Z
```

- Question: are X and Z necessarily independent?
  - Answer: no. Example: low pressure causes rain, which causes traffic.
  - X can influence Z, Z can influence X (via Y)
  - Addendum: they *could* be independent: how?
D-separation: Outline
D-separation: Outline

- Study independence properties for triples
- Analyze complex cases in terms of member triples
- D-separation: a condition / algorithm for answering such queries
Causal Chains

- This configuration is a “causal chain”
  
\[
P(x, y, z) = P(x)P(y|x)P(z|y)
\]

- Guaranteed X independent of Z?  No!
  
- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

- Example:
  
  Low pressure causes rain causes traffic, high pressure causes no rain causes no traffic

- In numbers:
  
  \[
P(+y \mid +x) = 1, P(-y \mid -x) = 1, P(+z \mid +y) = 1, P(-z \mid -y) = 1
\]
Causal Chains

- This configuration is a “causal chain”

\[ P(x, y, z) = P(x)P(y|x)P(z|y) \]

- Guaranteed X independent of Z given Y?

\[
\begin{align*}
P(z|x, y) &= \frac{P(x, y, z)}{P(x, y)} \\
&= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)} \\
&= P(z|y)
\end{align*}
\]

Yes!

- Evidence along the chain “blocks” the influence
Common Cause

- This configuration is a “common cause”

\[ P(x, y, z) = P(y)P(x|y)P(z|y) \]

- Guaranteed X independent of Z? No!

- One example set of CPTs for which X is not independent of Z is sufficient to show this independence is not guaranteed.

- Example:
  - Project due causes both forums busy and lab full

- In numbers:
  \[ P(+x | +y) = 1, \ P(-x | -y) = 1, \ P(+z | +y) = 1, \ P(-z | -y) = 1 \]
Common Cause

- This configuration is a “common cause”

\[ P(x, y, z) = P(y)P(x|y)P(z|y) \]

- Guaranteed X and Z independent given Y?

\[
P(z|x, y) = \frac{P(x, y, z)}{P(x, y)}
\]

\[= \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)}
\]

\[= P(z|y)\]

Yes!

- Observing the cause blocks influence between effects.
Last configuration: two causes of one effect (v-structures)

Are X and Y independent?

\[ P(X, Y) = \sum_Z P(X)P(Y)P(Z|X, Y) = P(X)P(Y) \sum_Z P(Z|X, Y) \]

- **Yes**: the ballgame and the rain cause traffic, but they are not correlated
- Still need to prove they must be (try it!)

Are X and Y independent given Z?

- **No**: seeing traffic puts the rain and the ballgame in competition as explanation.

This is backwards from the other cases

- Observing an effect activates influence between possible causes.
The General Case
The General Case

- General question: in a given BN, are two variables independent (given evidence)?
- Solution: analyze the graph
- Any complex example can be broken into repetitions of the three canonical cases
Recipe: shade evidence nodes, look for paths in the resulting graph

Attempt 1: if two nodes are connected by an undirected path not blocked by a shaded node, they are conditionally independent

Almost works, but not quite
- Where does it break?
- Answer: the v-structure at T doesn’t count as a link in a path unless “active”
**Question:** Are X and Y conditionally independent given evidence variables \{Z\}?
- Yes, if X and Y “d-separated” by Z
- Consider all (undirected) paths from X to Y
- No active paths = independence!

**A path is active if each triple is active:**
- Causal chain \(A \rightarrow B \rightarrow C\) where B is unobserved (either direction)
- Common cause \(A \leftarrow B \rightarrow C\) where B is unobserved
- Common effect (aka v-structure)
  - \(A \rightarrow B \leftarrow C\) where B or one of its descendents is observed

**All it takes to block a path is a single inactive segment**
D-Separation

- Query: \( X_i \perp\!
\!
\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \) ?

- Check all (undirected!) paths between \( X_i \) and \( X_j \):
  - If one or more active, then independence not guaranteed
    \[ X_i \not\perp\!
\!
\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \]
  - Otherwise (i.e. if all paths are inactive),
    then independence is guaranteed
    \[ X_i \perp\!
\!
\perp X_j \mid \{X_{k_1}, \ldots, X_{k_n}\} \]
Example

\[ R \perp B \]
\[ R \perp B | T \]
\[ R \perp B | T' \]
Example

\[ L \perp T' \mid T \quad \text{Yes} \]
\[ L \perp B \quad \text{Yes} \]
\[ L \perp B \mid T \]
\[ L \perp B \mid T' \]
\[ L \perp B \mid T, R \quad \text{Yes} \]
Example

- **Variables:**
  - R: Raining
  - T: Traffic
  - D: Roof drips
  - S: I’m sad

- **Questions:**

  \[
  T \perp D \\
  T \perp D | R \quad \text{Yes} \\
  T \perp D | R, S
  \]
- Given a Bayes net structure, can run d-separation algorithm to build a complete list of conditional independences that are necessarily true of the form

\[ X_i \perp \!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
\!
Computing All Independences

Compute **ALL THE INDEPENDENCES**!
Given some graph topology $G$, only certain joint distributions can be encoded.

The graph structure guarantees certain (conditional) independences.

(There might be more independence)

Adding arcs increases the set of distributions, but has several costs.

Full conditioning can encode any distribution.
Bayes Nets Representation Summary

- Bayes nets compactly encode joint distributions

- Guaranteed independencies of distributions can be deduced from BN graph structure

- D-separation gives precise conditional independence guarantees from graph alone

- A Bayes’ net’s joint distribution may have further (conditional) independence that is not detectable until you inspect its specific distribution
Bayes’ Nets

- Representation
- Conditional Independences
  - Probabilistic Inference
    - Enumeration (exact, exponential complexity)
    - Variable elimination (exact, worst-case exponential complexity, often better)
    - Probabilistic inference is NP-complete
    - Sampling (approximate)
  - Learning Bayes’ Nets from Data
Inference

- Inference: calculating some useful quantity from a joint probability distribution

Examples:

- Posterior probability
  \[ P(Q|E_1 = e_1, \ldots E_k = e_k) \]

- Most likely explanation:
  \[ \text{argmax}_q P(Q = q|E_1 = e_1 \ldots) \]
Inference by Enumeration

- **General case:**
  - Evidence variables: $E_1 \ldots E_k = e_1 \ldots e_k$
  - Query* variable: $Q$
  - Hidden variables: $H_1 \ldots H_r$

  \[
  \begin{align*}
  &X_1, X_2, \ldots X_n \\
  &\text{All variables}
  \end{align*}
  \]

- **Step 1:** Select the entries consistent with the evidence

- **Step 2:** Sum out H to get joint of Query and evidence

- **Step 3:** Normalize

\[
\begin{align*}
P(Q|e_1 \ldots e_k) &= \frac{1}{Z} P(Q, e_1 \ldots e_k) \\
Z &= \sum_{q} P(Q, e_1 \ldots e_k)
\end{align*}
\]

* Works fine with multiple query variables, too
Inference by Enumeration in Bayes’ Net

- Given unlimited time, inference in BNs is easy
- Reminder of inference by enumeration by example:

\[
P(B \mid +j, +m) \propto_B P(B, +j, +m) \\
= \sum_{e,a} P(B, e, a, +j, +m) \\
= \sum_{e,a} P(B) P(e) P(a \mid B, e) P(+j \mid a) P(+m \mid a)
\]

\[
= P(B) P(+e) P(+a \mid B, +e) P(+j \mid +a) P(+m \mid +a) \\
+ P(B) P(+e) P(-a \mid B, +e) P(+j \mid -a) P(+m \mid -a) \\
+ P(B) P(-e) P(+a \mid B, -e) P(+j \mid +a) P(+m \mid +a) \\
+ P(B) P(-e) P(-a \mid B, -e) P(+j \mid -a) P(+m \mid -a)
\]
Inference by Enumeration?

\[ P(\text{Antilock}|\text{observed variables}) = ? \]
Inference by Enumeration vs. Variable Elimination

- **Why is inference by enumeration so slow?**
  - You join up the whole joint distribution before you sum out the hidden variables

- **Idea:** interleave joining and marginalizing!
  - Called “Variable Elimination”
  - Still NP-hard, but usually much faster than inference by enumeration

- First we’ll need some new notation: factors
Factor Zoo
Factor Zoo I

- Joint distribution: $P(X,Y)$
  - Entries $P(x,y)$ for all $x, y$
  - Sums to 1

- Selected joint: $P(x,Y)$
  - A slice of the joint distribution
  - Entries $P(x,y)$ for fixed $x$, all $y$
  - Sums to $P(x)$

- Number of capitals = dimensionality of the table

\[
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{hot} & \text{sun} & 0.4 \\
\hline
\text{hot} & \text{rain} & 0.1 \\
\hline
\text{cold} & \text{sun} & 0.2 \\
\hline
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
T & W & P \\
\hline
\text{cold} & \text{sun} & 0.2 \\
\hline
\text{cold} & \text{rain} & 0.3 \\
\hline
\end{array}
\]
**Factor Zoo II**

- **Single conditional:** $P(Y \mid x)$
  - Entries $P(y \mid x)$ for fixed $x$, all
  - Sums to 1

- **Family of conditionals:** $P(X \mid Y)$
  - Multiple conditionals
  - Entries $P(x \mid y)$ for all $x, y$
  - Sums to $|Y|$
Specified family: \( P(y | X) \)
- Entries \( P(y | x) \) for fixed \( y \), but for all \( x \)
- Sums to ... who knows!

\[
P(rain | T')
\]

\[
\begin{array}{ccc}
T & W & P \\
\hline
\text{hot} & \text{rain} & 0.2 \\
\text{cold} & \text{rain} & 0.6 \\
\end{array}
\]

\[
\begin{align*}
P(rain | \text{hot}) & = 0.2 \\
P(rain | \text{cold}) & = 0.6
\end{align*}
\]
In general, when we write \( P(Y_1 \ldots Y_N \mid X_1 \ldots X_M) \):

- It is a “factor,” a multi-dimensional array
- Its values are \( P(y_1 \ldots y_N \mid x_1 \ldots x_M) \)
- Any assigned (=lower-case) \( X \) or \( Y \) is a dimension missing (selected) from the array
Example: Traffic Domain

- **Random Variables**
  - R: Raining
  - T: Traffic
  - L: Late for class!

\[
P(L) = ?
= \sum_{r,t} P(r, t, L)
= \sum_{r,t} P(r)P(t|r)P(L|t)
\]
Inference by Enumeration: Procedural Outline

- Track objects called factors
- Initial factors are local CPTs (one per node)

\[
\begin{array}{c|c}
\text{P}(R) & \\
\hline
+r & 0.1 \\
-r & 0.9 \\
\end{array}
\quad
\begin{array}{c|c|c}
\text{P}(T|R) & & \\
\hline
+r & +t & 0.8 \\
+r & -t & 0.2 \\
-r & +t & 0.1 \\
-r & -t & 0.9 \\
\end{array}
\quad
\begin{array}{c|c|c}
\text{P}(L|T) & & \\
\hline
+t & +l & 0.3 \\
+t & -l & 0.7 \\
-t & +l & 0.1 \\
-t & -l & 0.9 \\
\end{array}
\]

- Any known values are selected
  - E.g. if we know \( L = +\ell \), the initial factors are

\[
\begin{array}{c|c}
\text{P}(R) & \\
\hline
+r & 0.1 \\
-r & 0.9 \\
\end{array}
\quad
\begin{array}{c|c|c}
\text{P}(T|R) & & \\
\hline
+r & +t & 0.8 \\
+r & -t & 0.2 \\
-r & +t & 0.1 \\
-r & -t & 0.9 \\
\end{array}
\quad
\begin{array}{c|c|c}
\text{P}(+\ell|T) & & \\
\hline
+t & +l & 0.3 \\
-t & +l & 0.1 \\
\end{array}
\]

- Procedure: Join all factors, then eliminate all hidden variables
Operation 1: Join Factors

- First basic operation: joining factors
- Combining factors:
  - Just like a database join
  - Get all factors over the joining variable
  - Build a new factor over the union of the variables involved

- Example: Join on R

\[
P(R) \times P(T|R) \rightarrow P(R, T)
\]

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>-r</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>+t</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>0.8</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>+t</td>
<td>0.1</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>-t</td>
<td>0.1</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

\[P(R, T) = P(r) \cdot P(t|r)\]
Example: Multiple Joins
Example: Multiple Joins

$P(R)$

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

$P(T|R)$

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>+t</td>
<td></td>
</tr>
<tr>
<td>+r</td>
<td>-t</td>
<td>0.2</td>
</tr>
<tr>
<td>-r</td>
<td>+t</td>
<td>0.1</td>
</tr>
<tr>
<td>-r</td>
<td>-t</td>
<td>0.9</td>
</tr>
</tbody>
</table>

$P(L|T)$

<table>
<thead>
<tr>
<th></th>
<th>+t</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>+t</td>
<td>+l</td>
<td>0.3</td>
</tr>
<tr>
<td>+t</td>
<td>-l</td>
<td>0.7</td>
</tr>
<tr>
<td>-t</td>
<td>+l</td>
<td>0.1</td>
</tr>
<tr>
<td>-t</td>
<td>-l</td>
<td>0.9</td>
</tr>
</tbody>
</table>

$P(R,T)$

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>+t</th>
<th>-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+r</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P(R,T,L)$

<table>
<thead>
<tr>
<th></th>
<th>+r</th>
<th>+t</th>
<th>+l</th>
<th>-l</th>
</tr>
</thead>
<tbody>
<tr>
<td>+r</td>
<td>0.024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+r</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+r</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+r</td>
<td>0.018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.027</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.081</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-r</td>
<td>0.729</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Operation 2: Eliminate

- Second basic operation: **marginalization**
- Take a factor and sum out a variable
  - Shrinks a factor to a smaller one
  - A *projection* operation
- Example:

  \[
  P(R, T) = \begin{array}{ccc}
  +r & +t & 0.08 \\
  +r & -t & 0.02 \\
  -r & +t & 0.09 \\
  -r & -t & 0.81 \\
  \end{array}
  \]

  sum \( R \)  \quad \Rightarrow \quad \begin{array}{c}
  P(T) \\
  +t & 0.17 \\
  -t & 0.83 \\
  \end{array}
Multiple Elimination

\[ P(R, T, L) \]

\[
\begin{array}{ccc}
+r & +t & +l & 0.024 \\
+r & +t & -l & 0.056 \\
+r & -t & +l & 0.002 \\
+r & -t & -l & 0.018 \\
-r & +t & +l & 0.027 \\
-r & +t & -l & 0.063 \\
-r & -t & +l & 0.081 \\
-r & -t & -l & 0.729 \\
\end{array}
\]

Sum out R

\[ P(T, L) \]

\[
\begin{array}{cc}
+t & +l & 0.051 \\
+t & -l & 0.119 \\
-t & +l & 0.083 \\
-t & -l & 0.747 \\
\end{array}
\]

Sum out T

\[ P(L) \]

\[
\begin{array}{c}
+l & 0.134 \\
-l & 0.886 \\
\end{array}
\]
Thus Far: Multiple Join, Multiple Eliminate (= Inference by Enumeration)
Marginalizing Early (= Variable Elimination)