First-order probabilistic inference

David Poole

University of British Columbia

Overview

- Simple representation: parametrized belief networks.
  Means grounding.

- Inference: combine variable elimination and unification
  - One step of first-order variable elimination corresponds to an unbounded number of VE steps.
  - Allows for new queries, for example queries that depend on population size.

Bayesians

- Probability is a measure of belief.
- All of the individuals about which we have the same information have the same probability.
  - Idea: share probability tables both initially and during inference.

Background: Belief (Bayesian) networks

- Totally order the variables of interest: $X_1, \ldots, X_n$
- Theorem of probability theory (chain rule):

$$\begin{aligned} P(X_1, \ldots, X_n) &= P(X_1)P(X_2|X_1)\cdots P(X_n|X_1, \ldots, X_{n-1}) \\ &= \prod_{i=1}^{n} P(X_i|X_1, \ldots, X_{i-1}) \end{aligned}$$

- The parents of X_i $\pi_i \subseteq X_1, \ldots, X_{i-1}$ such that

$$P(X_i|\pi_i) = P(X_i|X_1, \ldots, X_{i-1})$$

- So $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\pi_i)$

⇒ Belief network nodes are variables, arcs from parents

Background: variable elimination

To compute the probability of a variable X given evidence $\overline{Z} = \overline{e}$:

$$P(X|\overline{Z} = \overline{e}) = \frac{P(X \wedge \overline{Z} = \overline{e})}{P(\overline{Z} = \overline{e})}$$

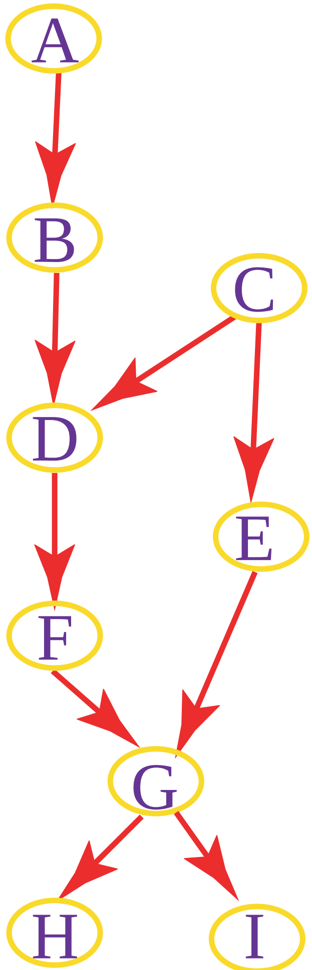
Suppose the other variables are $Y_1, \ldots, Y_m$:

$$\begin{aligned} & P(X \wedge \overline{Z}) \\ & \quad = \sum_{Y_m} \cdots \sum_{Y_1} P(X_1, \ldots, X_n) \\ & \quad = \sum_{Y_m} \cdots \sum_{Y_1} \prod_{i=1}^{n} P(X_i|\pi_i) \end{aligned}$$

Eliminating a variable

- to compute $AB + AC$ efficiently, distribute out A: $A(B + C)$.
- to compute
$$\sum_{Y_j} \prod_{i=1}^{n} P(X_i | \pi_i)$$
distribute out those factors that don't involve Y_j.
- Closely related to nonserial dynamic programming [Bertelè & Brioschi, 1972]

Variable Elimination Example



$$\left.\begin{array}{l} P(A) \\ P(B|A) \end{array}\right\} \xrightarrow{\text{elim } A} f_1(B)$$

$$\left.\begin{array}{l} P(C) \\ P(D|BC) \\ P(E|C) \end{array}\right\} \xrightarrow{\text{elim } C} f_2(BDE)$$

$$P(F|D)$$

$$P(G|FE)$$

$$P(H|G) \left.\right\} \xrightarrow{\text{obs } H} f_3(G)$$

$$P(I|G) \left.\right\} \xrightarrow{\text{elim } I} f_4(G)$$

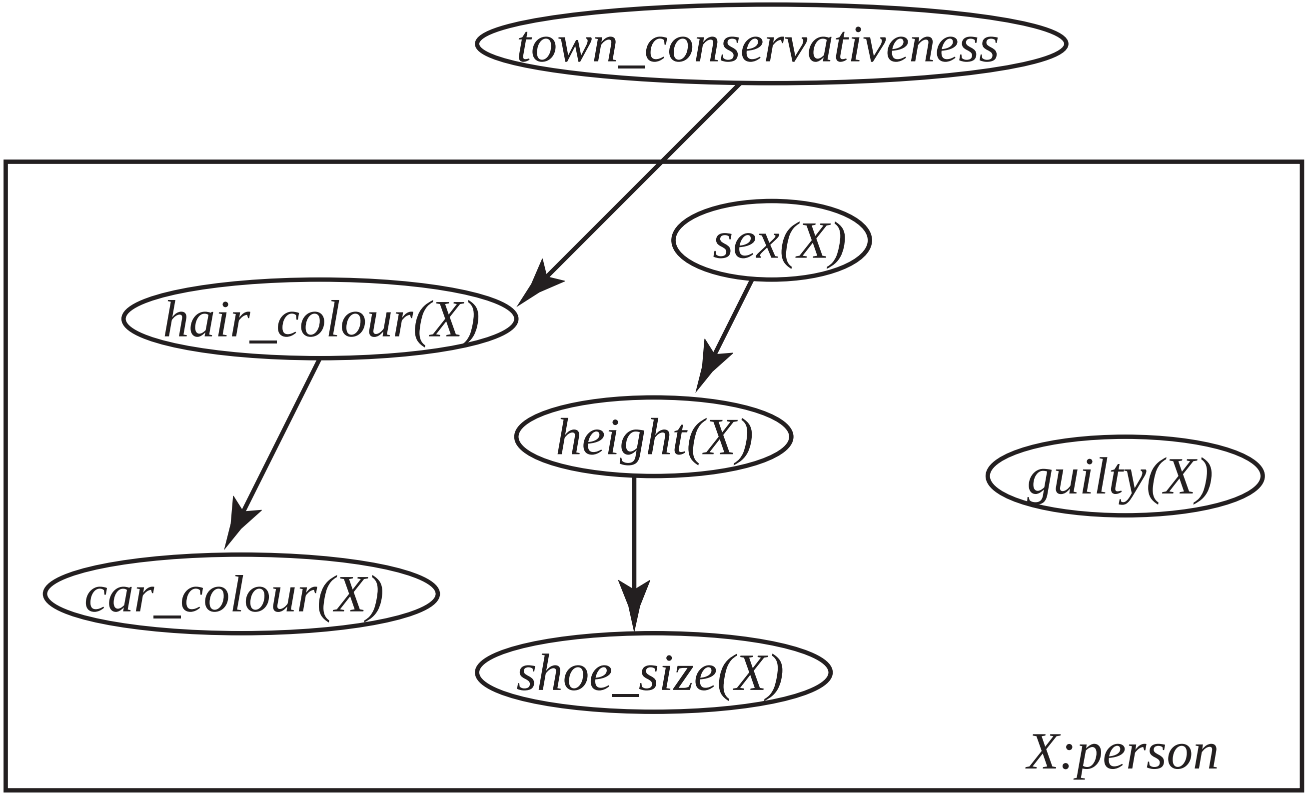
Variable Elimination: basic operations

- conditioning on observations (local to each factor)
- multiplying factors
- summing a variable from a factor

Parametrized belief networks

- Allow random variables to be parametrized. $height(X)$
- Parameters correspond to logical variables. X
- Each parameter is typed with a population. $X : person$
- Each population has a size. $|person| = 1000000$
- Parametrized belief network means its grounding: for each combination of parameters, an instance of each random variable for each member of parameters' population. $height(p_1) \ldots height(p_{1000000})$
- Instances are independent (but can have common ancestors and descendents).

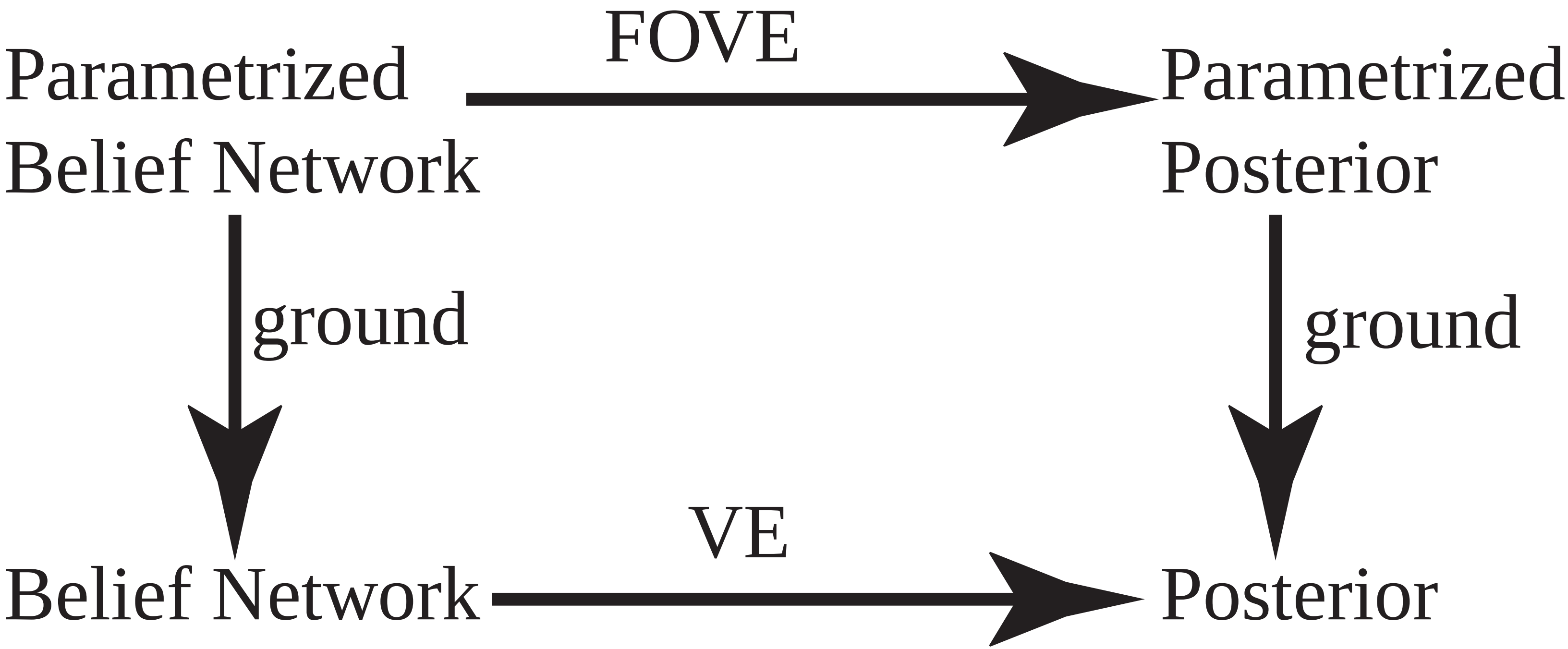
Example parametrized belief network



$$\forall X\ P(car\_colour(X){=}pink \mid hair\_colour(X){=}pink) = 0.1$$

$$\forall X\ P(hair\_colour(X){=}pink \mid town\_conservative) = 0.001.$$

First-order probabilistic inference



Theorem Proving and Unification

In 1965, Robinson showed how unification allows many ground steps with one step:

$$\underbrace{f(X, Z) \vee p(X, a) \qquad \neg p(b, Y) \vee g(Y, W)}$$

$$f(b, Z) \vee g(a, W)$$

Substitution $\{X/b, Y/a\}$ is the most general unifier of $p(X, a)$ and $p(b, Y)$.

Variable Elimination and Unification

- Multiplying parametrized factors:

$$\underbrace{[f(X,Z), p(X,a)] \quad \times \quad [p(b,Y), g(Y,W)]}_{[f(b,Z),\, p(b,a),\, g(a,W)]}$$

  Doesn't quite work because the first parametrized factor can't be used for $X = b$ but can be used for other instances of X.

- Intuitively, we want to add the constraint $X \neq b$ to $[f(X,Z), p(X,a)]$ after the above multiplication.

Parametric Factors

A **parametric factor** is a triple $\langle C, V, t\rangle$ where

- C is a set of inequality constraints on parameters,
- V is a set of parametrized random variables
- t is a table representing a factor from the random variables to the non-negative reals.

$$\left\langle \{X \neq sue\}, \{hair\_col(X), cons\}, \begin{array}{|cc|c|} \hline hair\_col & cons & Val \\ \hline purple & yes & 0.001 \\ purple & no & 0.01 \\ & \ldots & \\ \hline \end{array} \right\rangle$$

Splitting

Instead of applying substitutions to parametric factors, we split the parametric factors on the substitution.

A **split** of $\langle C, V, t\rangle$ on $X = \gamma$, results in parametric factors:

$$\langle C[X/\gamma], V[X/\gamma], t\rangle$$

$$\langle \{X \neq \gamma\} \cup C, V, t\rangle \qquad \longleftarrow \text{residual}$$

where $V[X/\gamma]$ is V with γ substituted for X.

Splitting on a substitution

- Splitting on a substitution, means splitting on each equality in the substitution.

- Different orders of splitting give the same final result, but may give different residuals.

- Example: Split

$$\langle \{\}, \{foo(X, Y, Z)\}, t_1 \rangle$$

on $\{X = Z, Y = b\}$ results in

$$\langle \{\}, \{foo(X, b, X)\}, t_1 \rangle$$

$$\langle \{X \neq Z\}, \{foo(X, Y, Z)\}, t_1 \rangle$$

$$\langle \{Y \neq b\}, \{foo(X, Y, X)\}, t_1 \rangle$$

Multiplying Parametric Factors

Suppose we were to eliminate p and multiply the two parametric factors:

$$\langle \{\}, \{p(X, a), q(Y, c), s(X, Y)\}, t_1 \rangle$$

$$\langle \{W \neq d\}, \{p(b, Z), q(W, T), r(W, T)\}, t_2 \rangle$$

- If we grounded these, then did VE, some instances of these would be multiplied and some wouldn't.
- We unify $p(X, a)$ and $p(b, Z)$ resulting in the substitution $\theta = \{X/b, Z/a\}$.
- Unification finds the most general instances that need to be multiplied.

Splitting when Multiplying I

We are multiplying the two parametric factors:

$$\langle \{\}, \{p(X, a), q(Y, c), s(X, Y)\}, t_1 \rangle \tag{1}$$

$$\langle \{W \neq d\}, \{p(b, Z), q(W, T), r(W, T)\}, t_2 \rangle \tag{2}$$

We split parametric factor (1) on $\theta = \{X/b, Z/a\}$:

$$\langle \{\}, \{p(b, a), q(Y, c), s(b, Y)\}, t_1 \rangle \tag{3}$$

$$\langle \{X \neq b\}, \{p(X, a), q(Y, c), s(X, Y)\}, t_1 \rangle \tag{4}$$

We can split (2) on θ resulting in:

$$\langle \{W \neq d\}, \{p(b, a), q(W, T), r(W, T)\}, t_2 \rangle \tag{5}$$

$$\langle \{Z \neq a, W \neq d\}, \{p(b, Z), q(W, T), r(W, T)\}, t_2 \rangle \tag{6}$$

Splitting when Multiplying II

When we are multiplying:

$$\langle \{\}, \{p(b,a), q(Y,c), s(b,Y)\}, t_1 \rangle$$

$$\langle \{W \neq d\}, \{p(b,a), q(W,T), r(W,T)\}, t_2 \rangle$$

- All ground instances would need to be multiplied.
- Not all instances have the same number of variables: some will have two different q instances, and some have one.
- We need to split again on the most general unifier of $q(Y,c)$ and $q(W,T)$.

Summing out variables

- If we are not removing a parameter, we sum out as normal. E.g., summing out p:

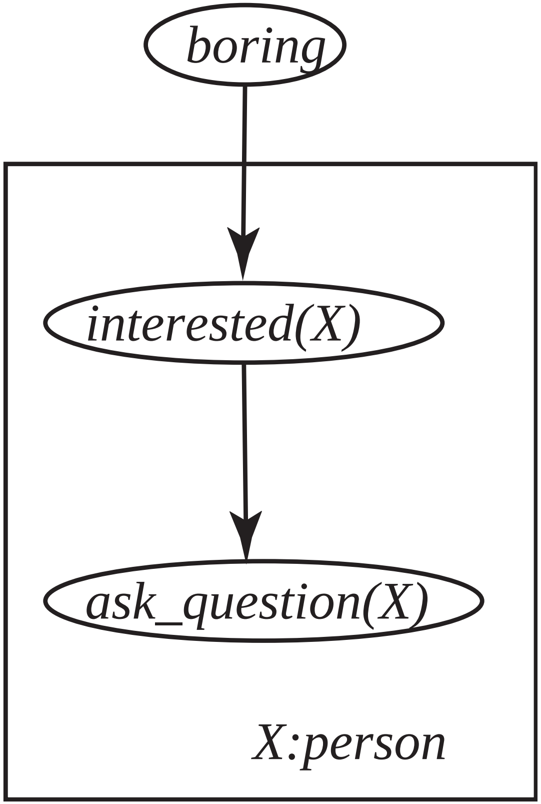
$$\langle \{\}, \{p(X), q(X)\}, t[p, q] \rangle$$

- If we are removing a parameter, we must take to the power of the effective population size. E.g., summing out p:

$$\langle \{Y \neq a\}, \{p(X, Y), q(X)\}, t[p, q] \rangle$$

- Other functions such as noisy-or, you need to take into account the population size.

Removing a parameter when summing



Eliminate *interested*:

$$\langle \{\}, \{boring, interested(X)\}, t_1 \rangle$$

$$\langle \{\}, \{interested(X)\}, t_2 \rangle$$

$$\downarrow$$

$$\left\langle \{\}, \{boring\}, (t_1 \times t_2)^{100} \right\rangle$$

$|people| = 100$

observe no questions

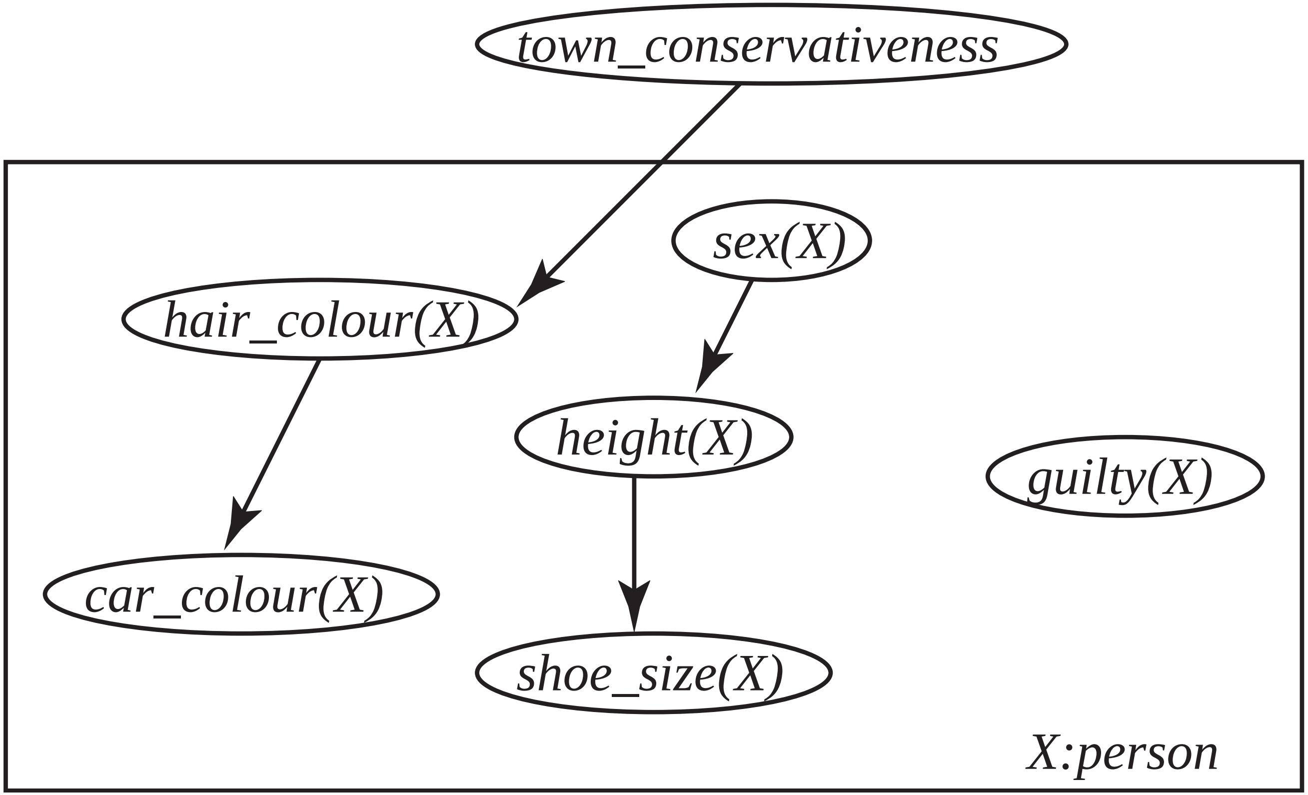
Existential Observations

Suppose we observe:

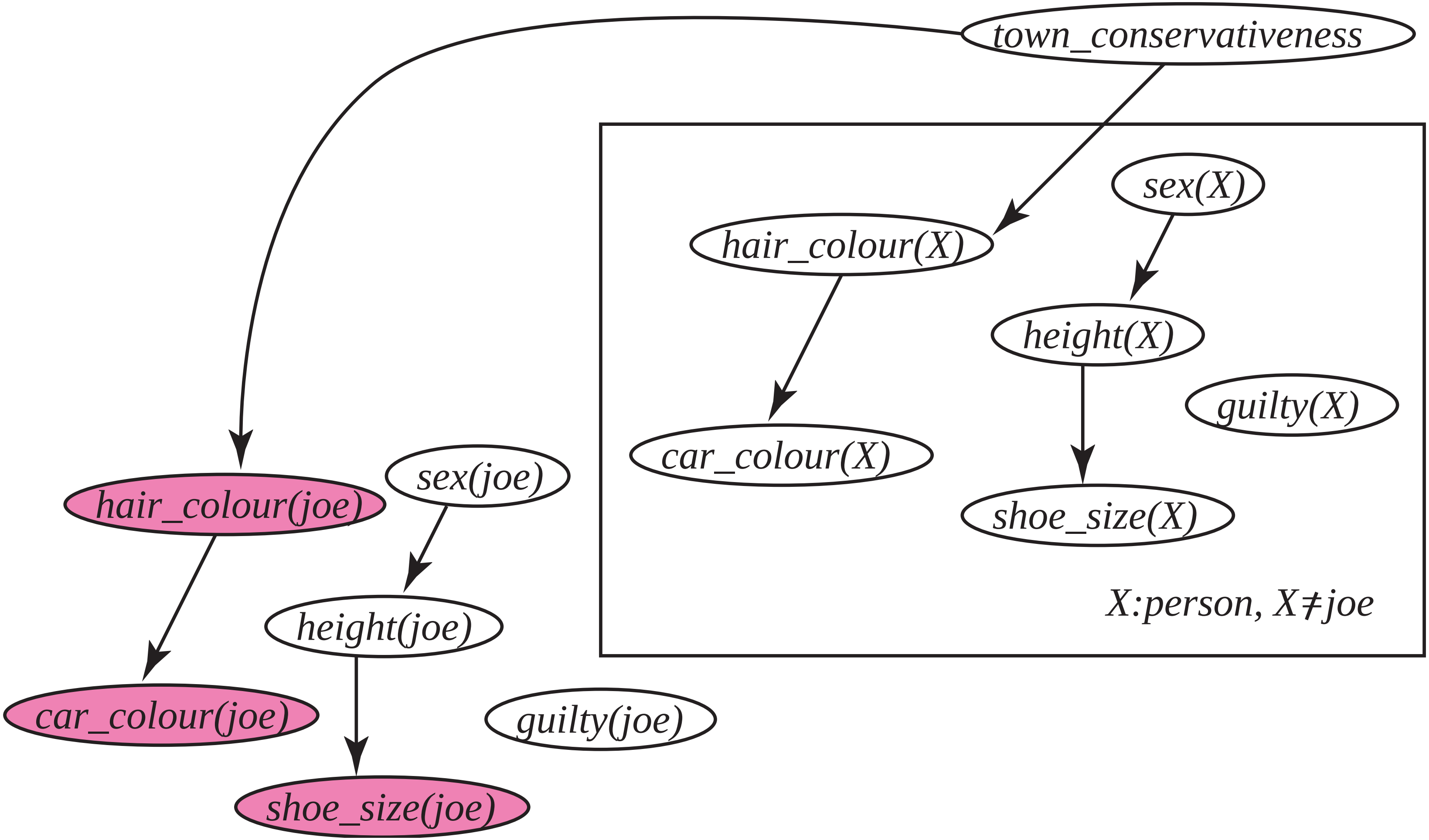
- Joes has purple hair, a purple car, and has big feet.
- A person with purple hair, a purple car, and who is very tall was seen committing a crime.

What is the probability that Joe is guilty?

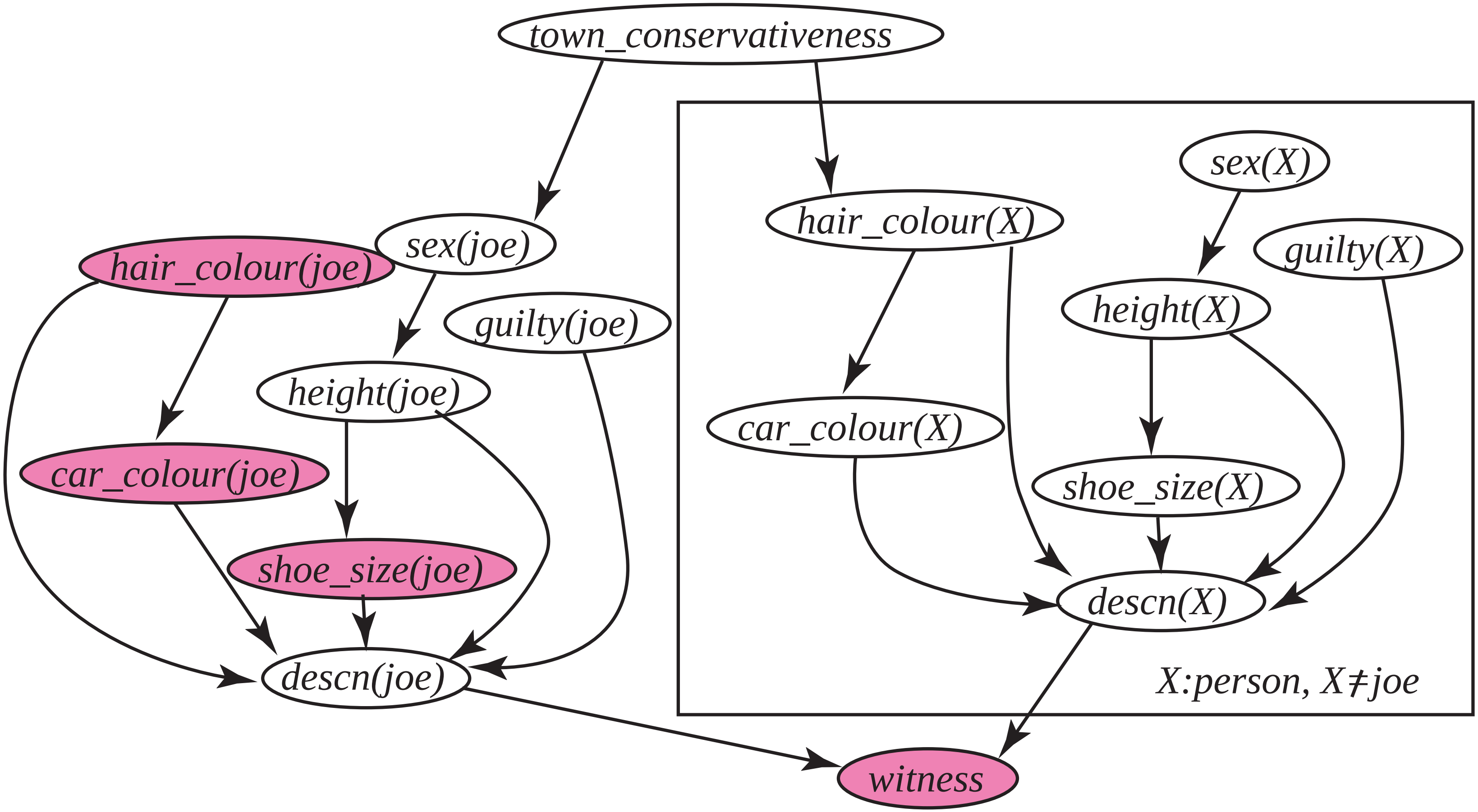
Background parametrized belief network



Observing information about Joe



Observing Joe and the crime



Last Steps

We end up with parametric Factors:

$$\langle \{\}, \{guilty(joe), descn(joe), conservativeness\}, t_1 \rangle$$

$$\langle \{X \neq joe\}, \{descn(X), conservativeness\}, t_2 \rangle$$

$$\langle \{\}, \{descn(X), witness\}, t_3 \rangle$$

$$\langle \{\}, \{conservativeness\}, t_4 \rangle$$

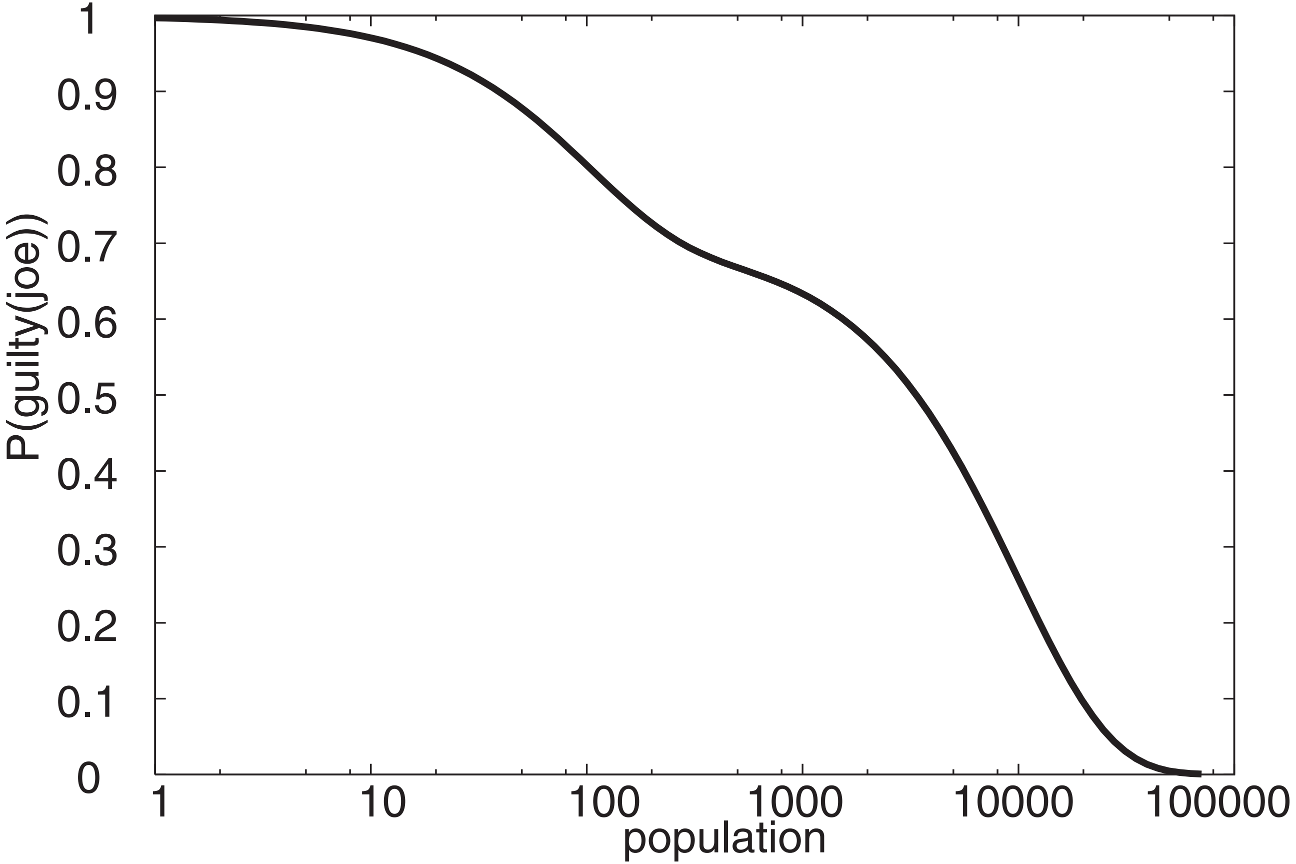
We eliminate $descn(X)$:

$$\langle \{\}, \{guilty(joe), witness, conservativeness\}, t_5 \rangle$$

We sum out *conservativeness* and condition on *witness*:

$$\langle \{\}, \{guilty(joe)\}, t_6 \rangle$$

Guilty as a function of population



Conclusions

- We combine variable elimination + unification.
  - One step of first-order variable elimination corresponds to many steps in ground representation.
  - We can condition on existential and universal observations.