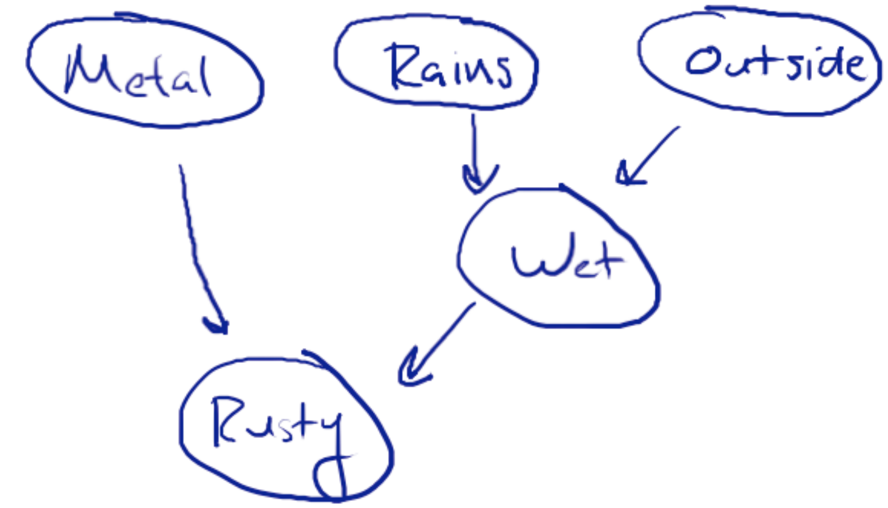
Review: graphical models

- Represent a distribution over some RVs
  - using both diagrams and numbers
- Chief problem: given a GM (the ***prior***) and some evidence (***data***), compute properties of the conditional distribution P(RVs | data) (the ***posterior***)
  - called ***inference***

Review: Bayes nets



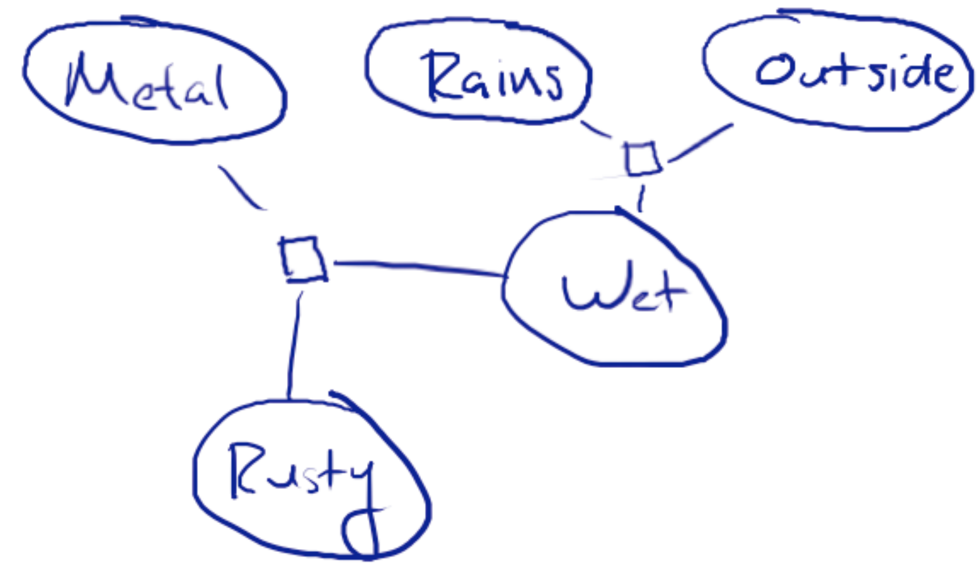
- Bayes net = DAG + CPT
- Independence
  - from DAG alone v. accidental
  - d-separation
    - blocking, explaining away
  - Markov blanket

Review: CPTs

- $P(W \mid Ra, O) =$

- Represents probability distribution(s)
  - for:
  - sums to 1:

Review: factor graphs



- Undirected, bipartite graph
  - factor & variable nodes
- Both Bayes nets and factor graphs can represent ***any*** distribution
  - either may be more efficient
  - conversion is easier bnet → factor graph
  - accidental v. graphical indep's may differ

Review: factors

- sum constraints:
- often results from:
- note: many ways to display same table!

Review: parameter learning

- Bayes net, when fully observed
  - counting, Laplace smoothing
- Missing data: harder
- Factor graph: harder (even if fully observed)

Admin

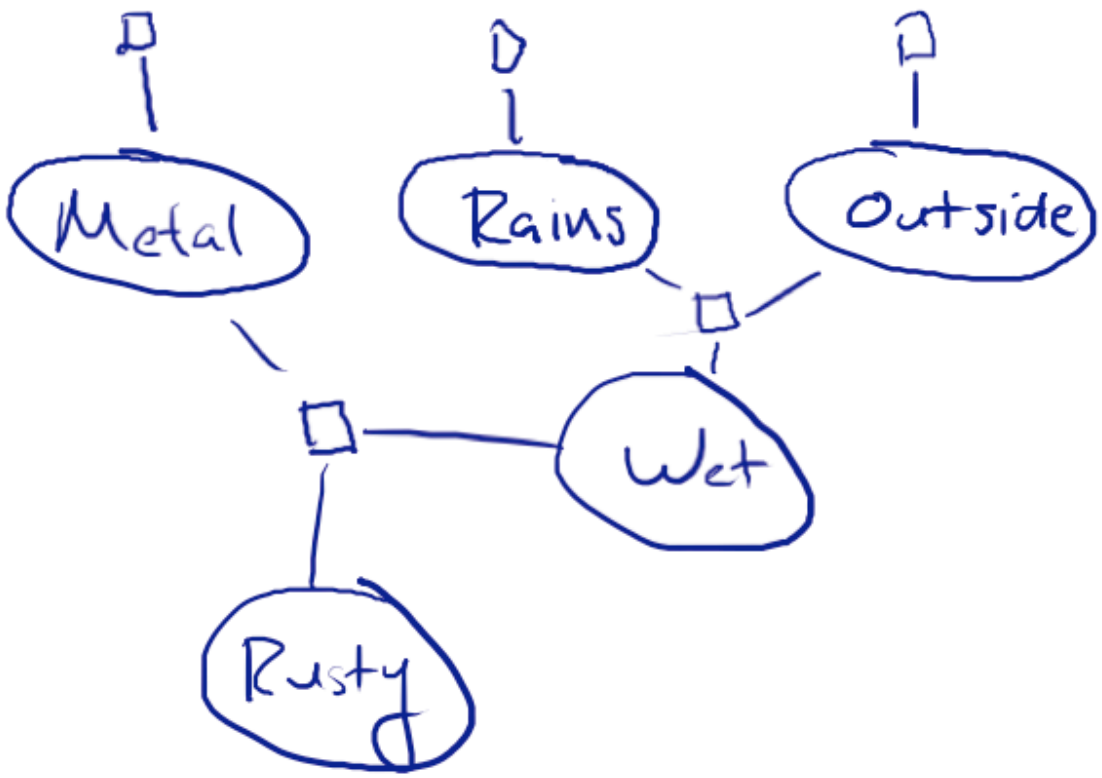
- HWs are due at 10:30
  - don’t skip class to work on it and turn it in at noon
- Late HWs are due at 10:30 (+ n days)
  - must use a whole number of late days
- HWs should be complete at 10:30

Inference

- Inference: prior + evidence → posterior
- We gave examples of inference in a Bayes net, but not a general algorithm
- Reason: general algorithm uses factor-graph representation
- Steps: instantiate evidence, eliminate nuisance nodes, normalize, answer query

Inference

$$P(M, Ra, O, W, Ru) = \phi_1(M)\,\phi_2(Ra)\,\phi_3(O)\,\phi_4(Ra, O, W)\,\phi_5(M, W, Ru)\,/\,Z$$



$\phi_1(M) =$

T	0.9
F	0.1

$\phi_2(Ra) =$

T	0.7
F	0.3

$\phi_3(O) =$

T	0.2
F	0.8

$\phi_4(Ra, O, W) =$

TTT	0.9
TTF	0.1
TFT	0.1
TFF	0.9
FTT	0.1
FTF	0.9
FFT	0.1
FFF	0.9

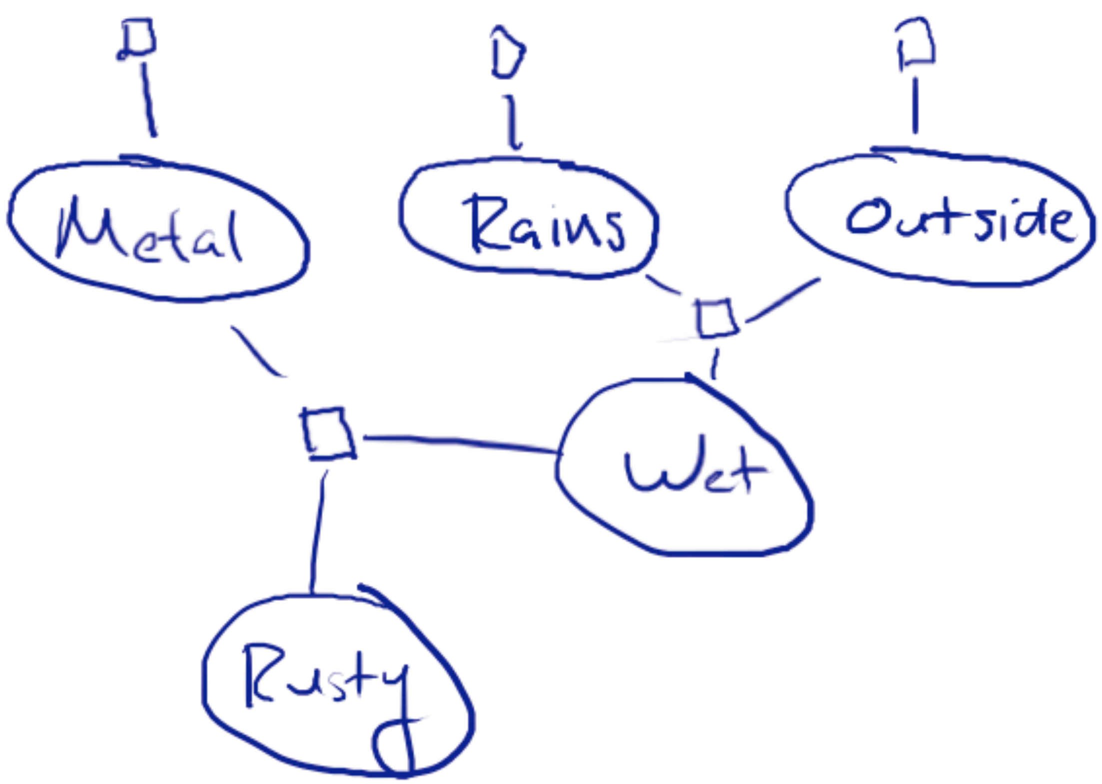
$\phi_5(M, W, Ru) =$

TTT	0.8
TTF	0.2
TFT	0.1
TFF	0.9
FTT	0
FTF	1
FFT	0
FFF	1

- Typical Q: given Ra=F, Ru=T, what is P(W)?

Incorporate evidence

$$P(M, Ra, O, W, Ru) = \phi_1(M)\,\phi_2(Ra)\,\phi_3(O)\,\phi_4(Ra, O, W)\,\phi_5(M, W, Ru)/Z$$



$\phi_1(M) =$ T 0.9, F 0.1

$\phi_2(Ra) =$ T 0.7, F 0.3

$\phi_3(O) =$ T 0.2, F 0.8

$\phi_4(Ra, O, W) =$

TTT	0.9
TTF	0.1
TFT	0.1
TFF	0.9
FTT	0.1
FTF	0.9
FFT	0.1
FFF	0.9

$\phi_5(M, W, Ru) =$

TTT	0.8
TTF	0.2
TFT	0.1
TFF	0.9
FTT	0
FTF	1
FFT	0
FFF	1

Condition on Ra=F, Ru=T

Eliminate nuisance nodes

$$P(M, Ra, O, W, Ru) = \phi_1(M)\,\phi_2(Ra)\,\phi_3(O)\,\phi_4(Ra, O, W)\,\phi_5(M, W, Ru) / Z$$

- Remaining nodes: M, O, W
- Query: P(W)
- So, O&M are nuisance—marginalize away
- Marginal =

Elimination order

$$\sum_{M} \sum_{O} \phi_1(M)\,\phi_3(O)\,\phi_4(O,W)\,\phi_5(M,W) \,/\, Z$$

- Sum out the nuisance variables in turn
- Can do it in any order, but some orders may be easier than others
- Let's do O, then M

$\phi_3(O) =$

T	0.2
F	0.8

$\phi_4(O,W) =$

T T	0.1
T F	0.9
F T	0.1
F F	0.9

One last elimination

$\phi_1(M) =$

T	0.9
F	0.1

$\phi_6(W) =$

T	0.1
F	0.9

$\phi_5(M, W, \cancel{A}) =$

T	T	~~T~~	0.8
T	F	~~T~~	0.1
F	T	~~T~~	0
F	F	~~T~~	0

Checking our work

- http://www.aispace.org/bayes/version5.1.6/bayes.jnlp

Discussion

- FLOP count
- Steps: instantiate evidence, eliminate nuisance nodes, normalize, answer query
  - each elimination introduces:
- Normalization
- Each elimination order:
  - some tables:

Example: elim order

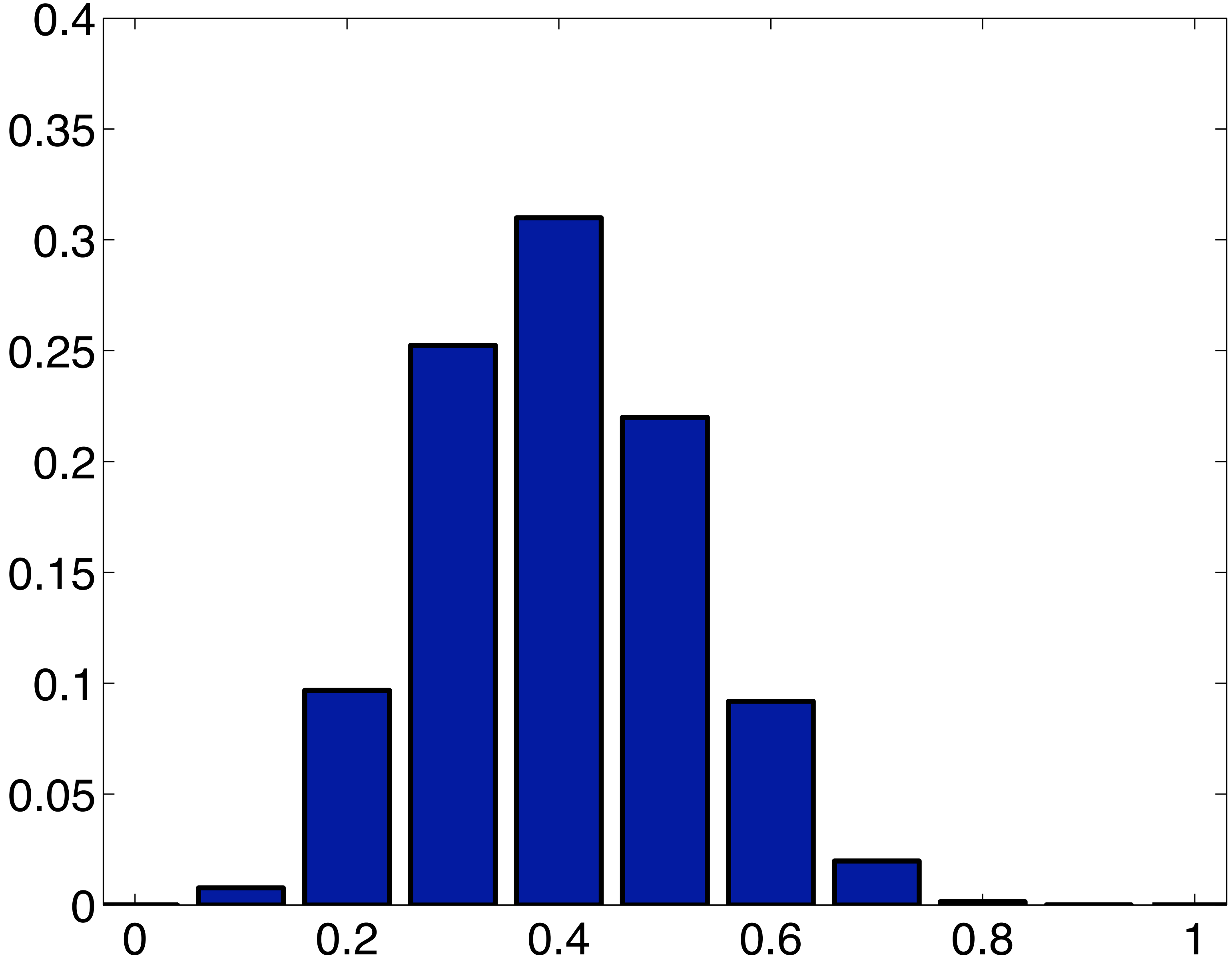
Example: elim order

- Compare: B,C,D vs. C,D, B

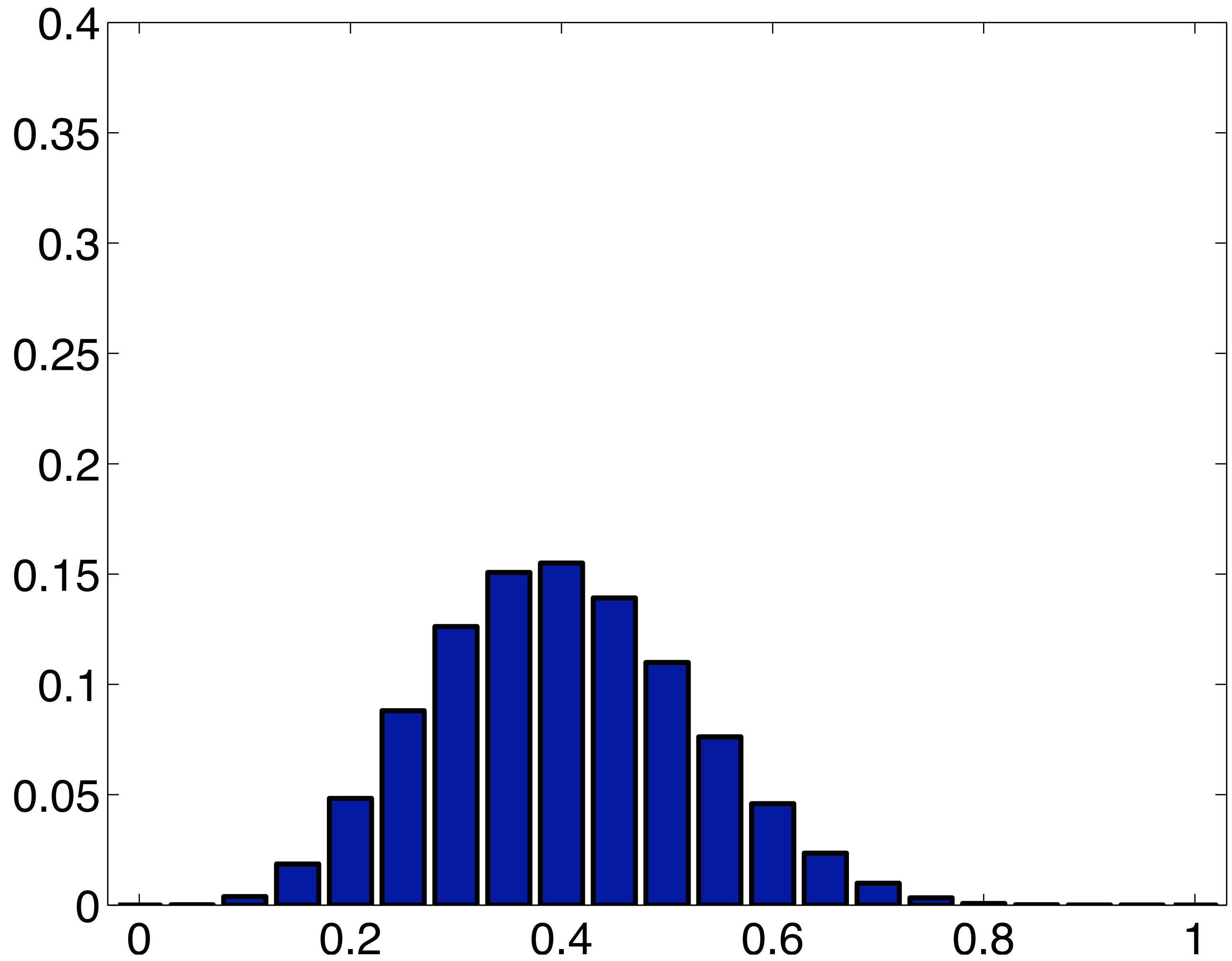
Continuous RVs

- All RVs we've used so far have been discrete
- Occasionally, we used a continuous one by ***discretization***
- We'll want to use truly continuous ones below

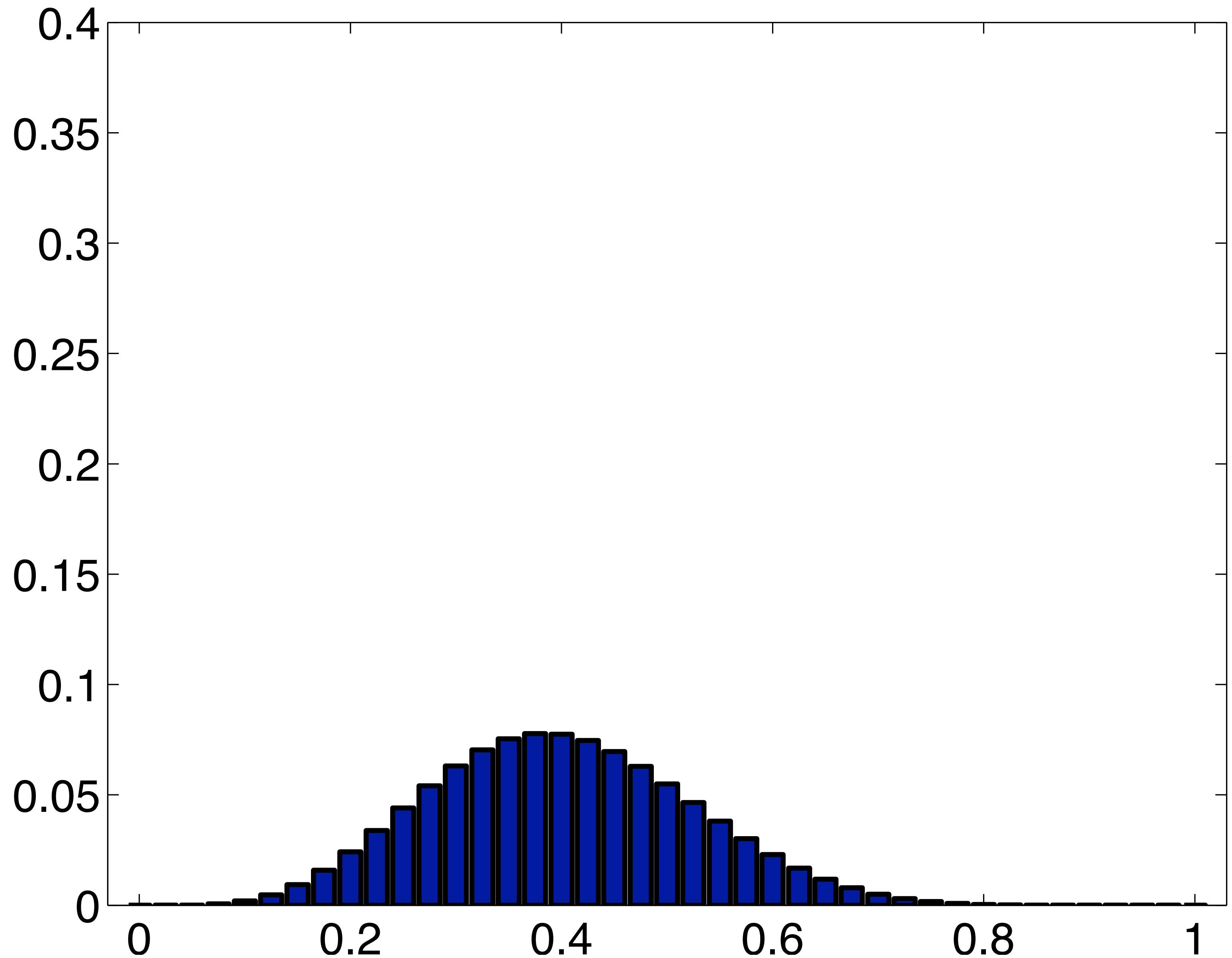
Finer & finer discretization



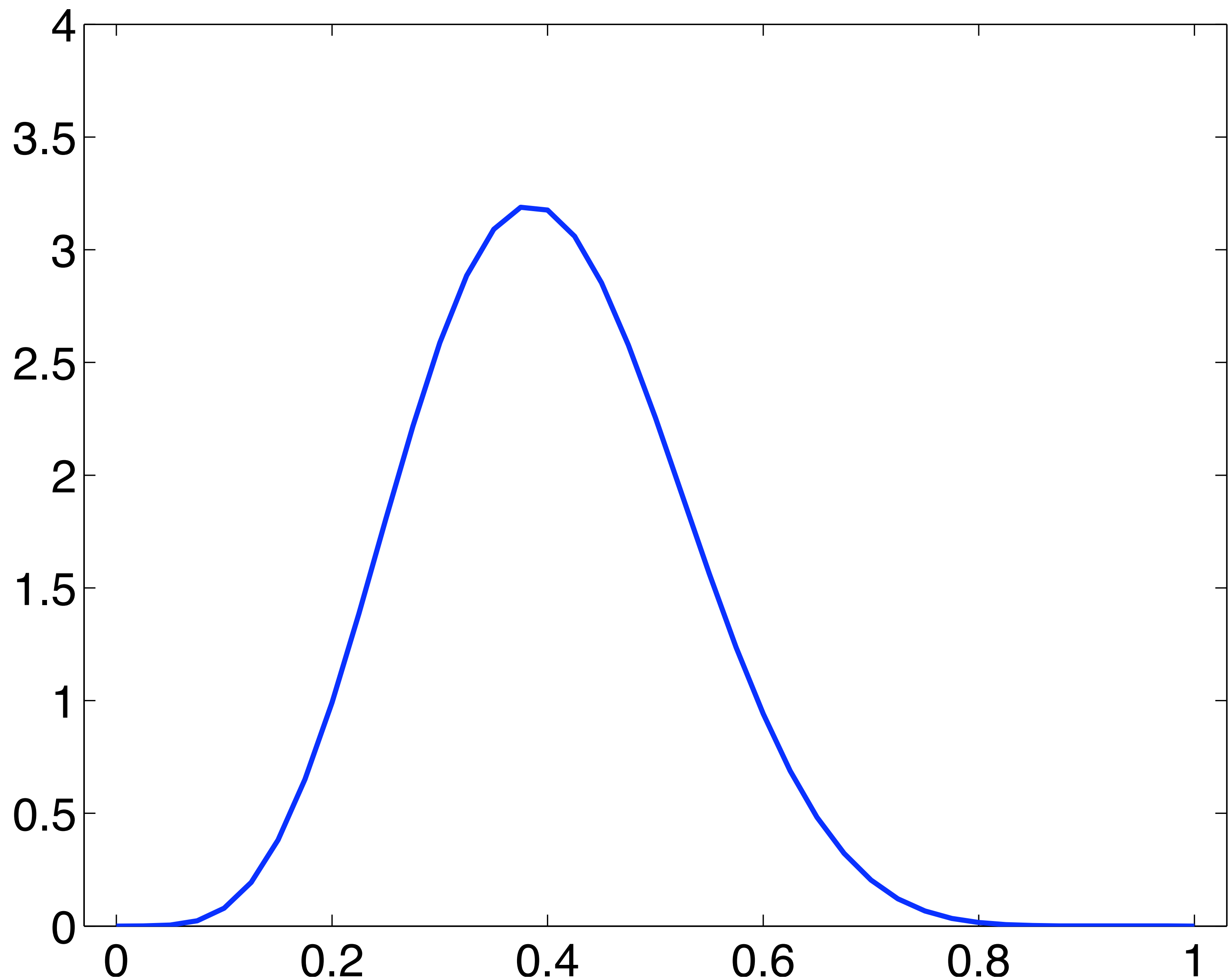
Finer & finer discretization



Finer & finer discretization



In the limit: density



- $\lim P(x \le X \le x+h) / h = P(x)$

Properties of densities

- instead of sum to 1,
- density may be
- PDF =
- Confusingly, we use $P(\cdot)$ for both, and sometimes people say distribution to mean either discrete or continuous

Events

- For continuous RVs X, Y:
- Sample space Ω = { }
- Event = subset of Ω
- Density: events → $\mathbb{R}_+$
  - disjoint union: additive
  - P(Ω) = 1

Continuous RVs in graphical models

- Very useful to have continuous RVs in GMs
- CPTs or potentials are now ***functions*** (tables where some dimensions are infinite)
- E.g.: $(X, Y) \in [0, 1]^2$
- $\phi(X, Y) =$
- $P(X, Y) =$

Continuous GM example