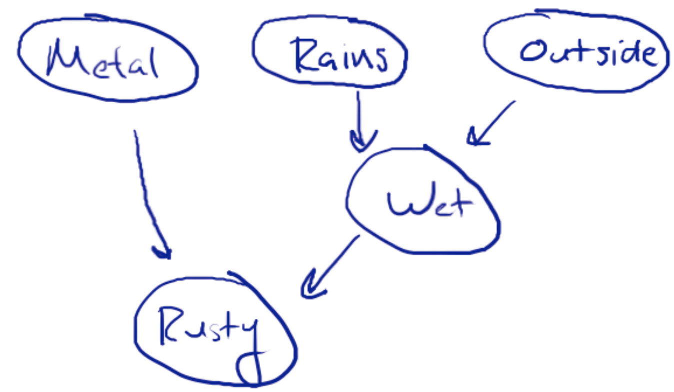
Review: graphical models

- 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

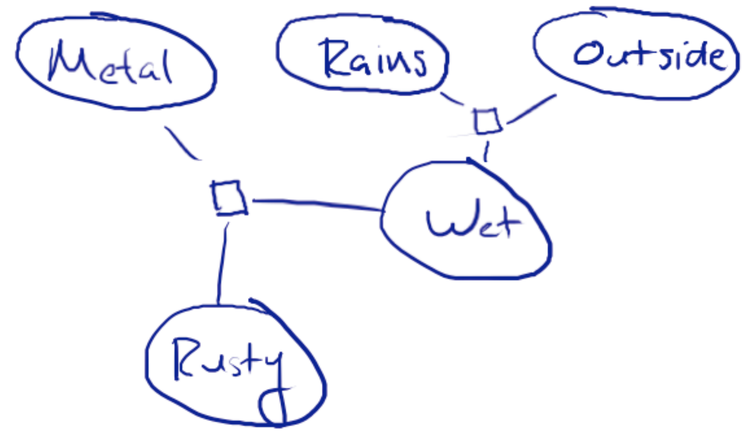
Review: CPTs

- P(W | Ra, O) =

Ra O	W: T	F
T T	.5	.5
T F	.5	.5
F T	.5	.5
F F	.33	.67

- 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

$$\phi(Ra, O) = \begin{array}{cc} Ra, O & \\ TT & 1/2 \\ TF & 1/2 \\ FT & 1/2 \\ FF & 2/3 \end{array}$$

$$\phi(Ra, O) \quad \begin{array}{c|cc} & & O \\ Ra & T & F \\ \hline T & 1/2 & 1/2 \\ F & 1/2 & 2/3 \end{array}$$

- sum constraints: none!!
- often results from: conditioning
- 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

poster session
Dec 4th 1-4 PM

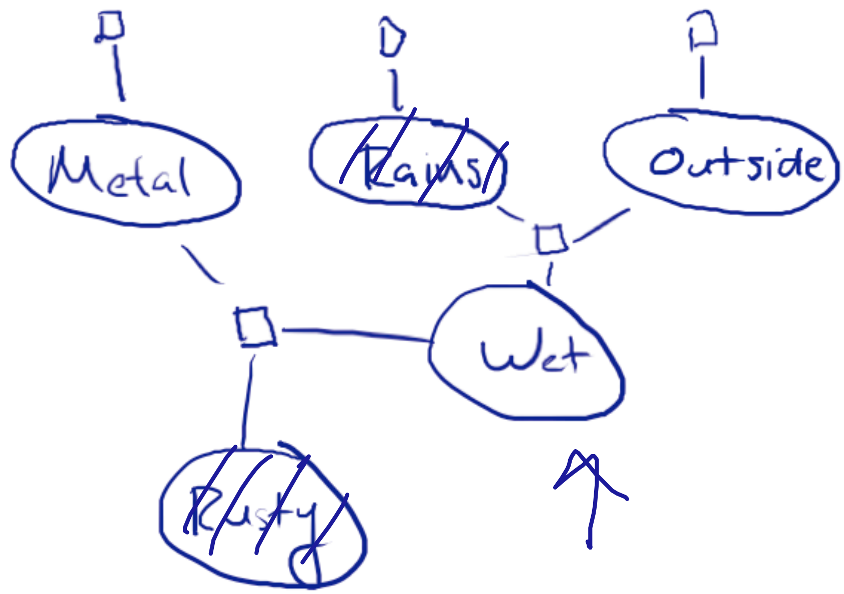
- HWs are due at 10:30
  - don't skip class to work on it and turn it in at noon
- Late HWs are due at 10:30 (+ n days)
  - must use a whole number of late days
- HWs should be complete at 10:30

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

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



$\phi_1(M) =$

T	0.9
F	0.1

$\cancel{\phi_2(Ra)} =$

~~T~~	~~0.7~~
~~F~~	~~0.3~~

$\phi_3(O) =$

T	0.2
F	0.8

$\phi_4(\cancel{Ra}, O, W) =$

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

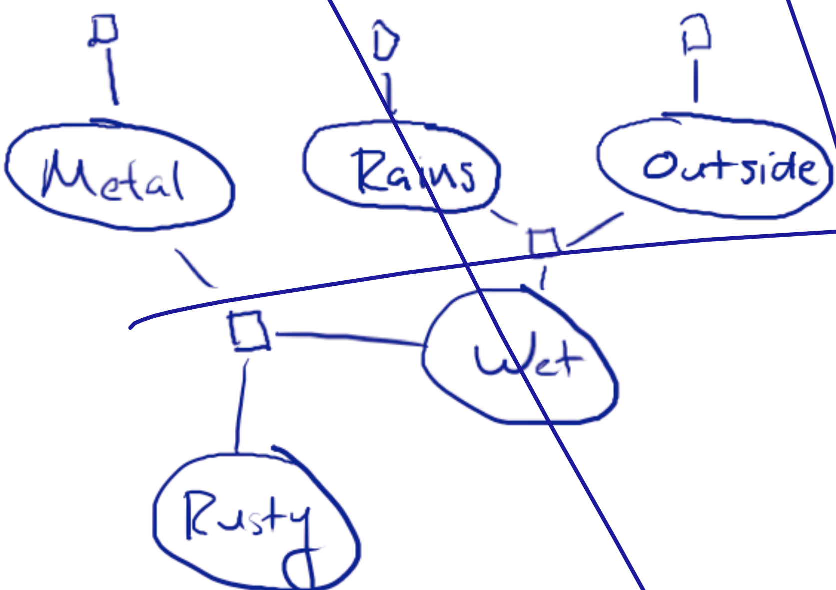
$\phi_5(M, W, \cancel{Ru}) =$

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

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

- Remaining nodes: M, O, W
- Query: P(W)
- So, O&M are nuisance—marginalize away
- Marginal = $\sum_{O \in T,F} \sum_{M \in T,F} \phi_1(M)\, \phi_3(O)\, \phi_4(O, W)\, \phi_5(M, W) / Z$

6 FLOPs

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

$$\phi_3(O) = \begin{array}{cc} T & 0.2 \\ F & 0.8 \end{array}$$

$$\sum_M \phi_1(M)\phi_5(M,W)\underbrace{\sum_O \phi_3(O)\phi_4(O,W)}_{\phi_6(W)}/Z$$

$\phi_4(\cancel{M}, O, W) =$

~~T~~	TT	0.1
~~T~~	TF	0.9
~~T~~	FT	0.1
~~F~~	FF	0.9

$\phi_6(W)$

W=T: $.1\times.2+.1\times.8=.1$

F: $.9\times.2+.9\times.8=.9$

10 Flops

One last elimination

+ 3 Flops

HH = 19 Flops

$$\sum_M \phi_1(M)\,\phi_5(M,W)\,\phi_6(W)/Z$$

$W=T$: $.9 \times .1 \times .8 + 0 = .072$

$W=F$: $.9 \times .9 \times 1 + 0 = .081$

$\Rightarrow Z = .072 + .081$

$P(W=T) = .072/Z = 8/17$

$P(W=F) = .081/Z = 9/17$

$\phi_1(M) =$

T	0.9
F	0.1

$\phi_6(W) =$

T	0.1
F	0.9

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

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

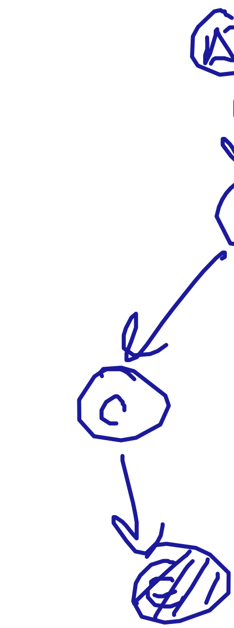
Checking our work

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

Discussion

- FLOP count 19 FLOPs vs. 221 for naive method
- Steps: instantiate evidence, eliminate nuisance nodes, normalize, answer query
  - each elimination introduces: new ϕ_t (some RV)
- Normalization
- Each elimination order: different set of ϕs
  - some tables: bigger!!

Example: elim order



$P(A) = .8 \qquad \phi_1(A) = .8 \;\; .2$

$P(B|A) = \begin{array}{ll} T & .3 \\ F & .5 \end{array}$

$\phi_2(A,B)$

		B: T	B: F
A	T	.3	.7
A	F	.5	.5

. . . .

$$\frac{1}{Z} \cdot \phi_1(A)\,\phi_2(A,B)\,\phi_3(B,C)\,\phi_4(B,D)\,\phi_5(C,\cancel{E})$$

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

Q: $P(A)$

BCD v. CDB

$$\frac{1}{Z} \sum_B \sum_C \sum_D \phi_1(A)\,\phi_2(A,B)\,\phi_3(B,C)\,\phi_4(B,D)\,\phi_5(C)$$

$$= \frac{1}{Z} \sum_C \sum_D \phi_1(A)\,\phi_5(C) \underbrace{\sum_B \phi_2(A,B)\,\phi_3(B,C)\,\phi_4(B,D)}_{\phi_6(A,C,D)}$$

Example: elim order

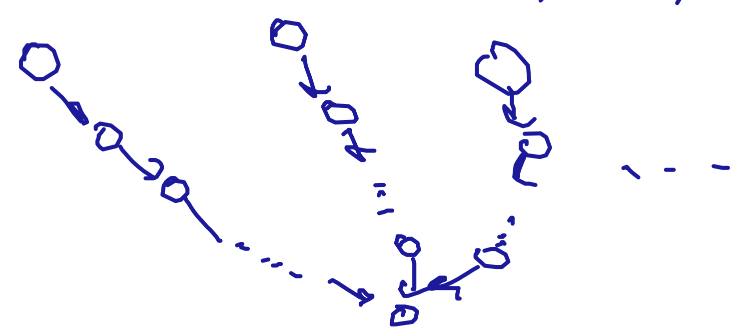
- Compare: B,C,D vs. C,D, B

$$\frac{1}{Z}\sum_B\sum_C\sum_D \phi_1(A)\,\phi_5(C)\,\phi_2(A,B)\,\phi_3(B,C)\,\phi_4(B,D)$$

$$= \frac{1}{Z}\sum_B\sum_D \phi_1(A)\,\phi_2(A,B)\,\phi_4(B,D)\underbrace{\sum_C \phi_5(C)\,\phi_3(B,C)}_{\phi_6'(B)}$$

$$= \frac{1}{Z}\sum_B \phi_1(A)\,\phi_2(A,B)\,\phi_6'(B)\underbrace{\sum_D \phi_4(B,D)}_{\phi_7'(B)}$$

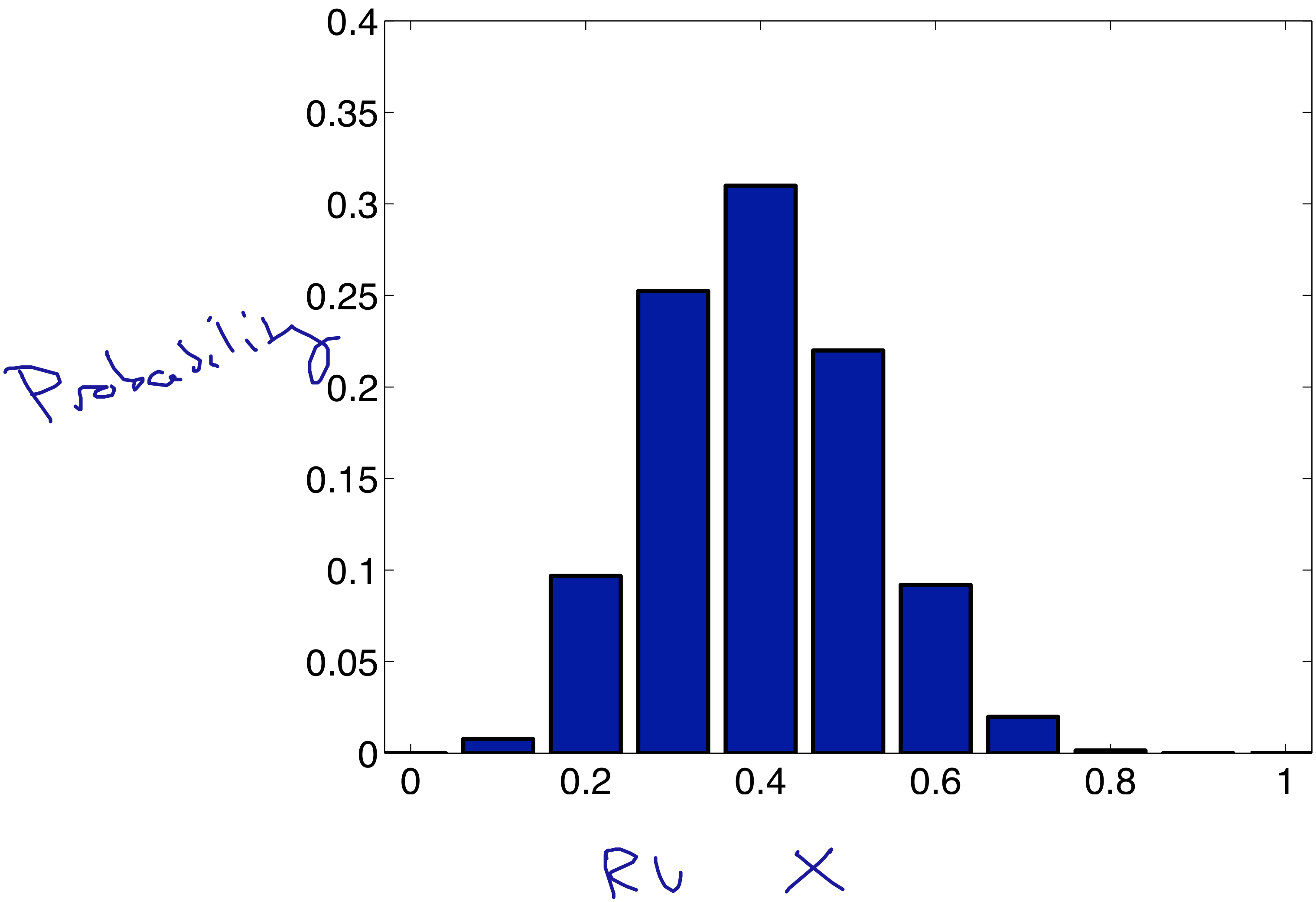
$$= \frac{1}{Z}\phi_8'(A)$$



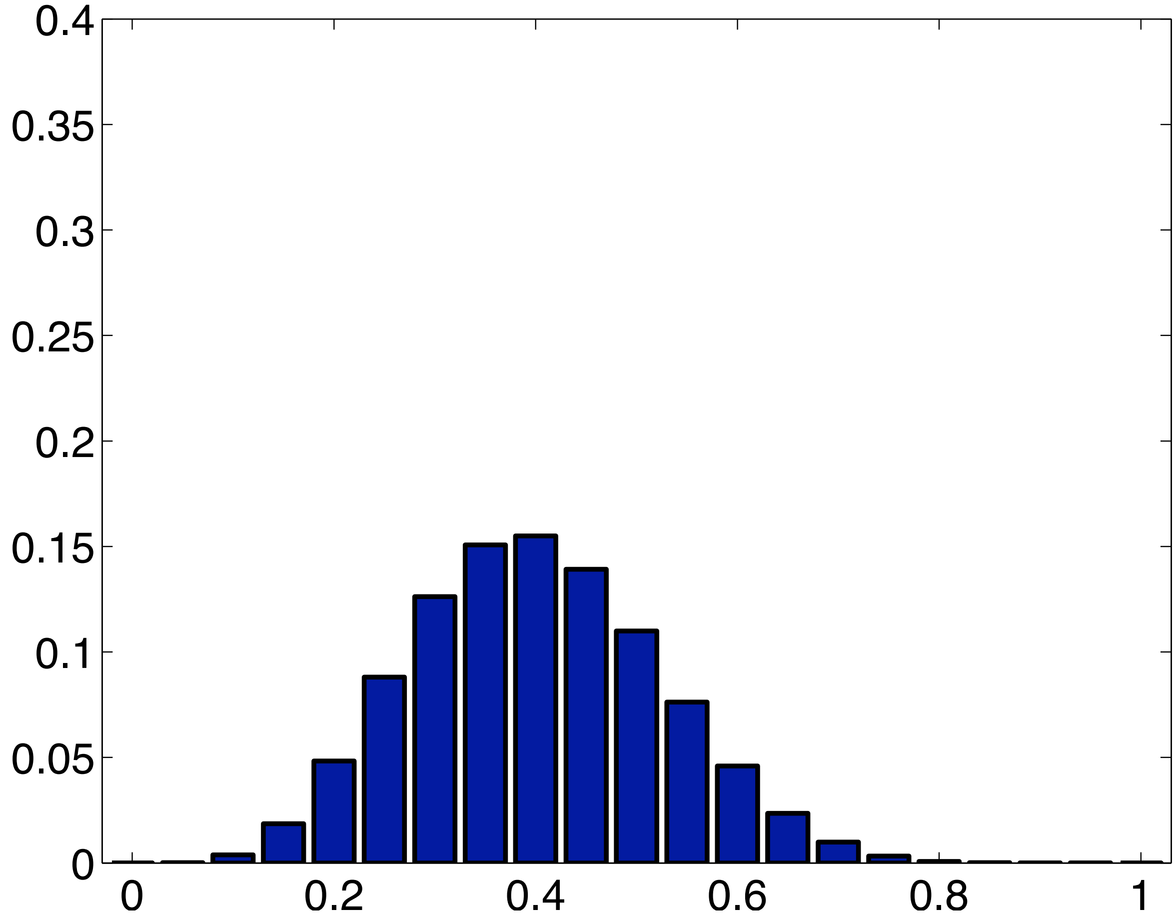
Continuous RVs

- 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

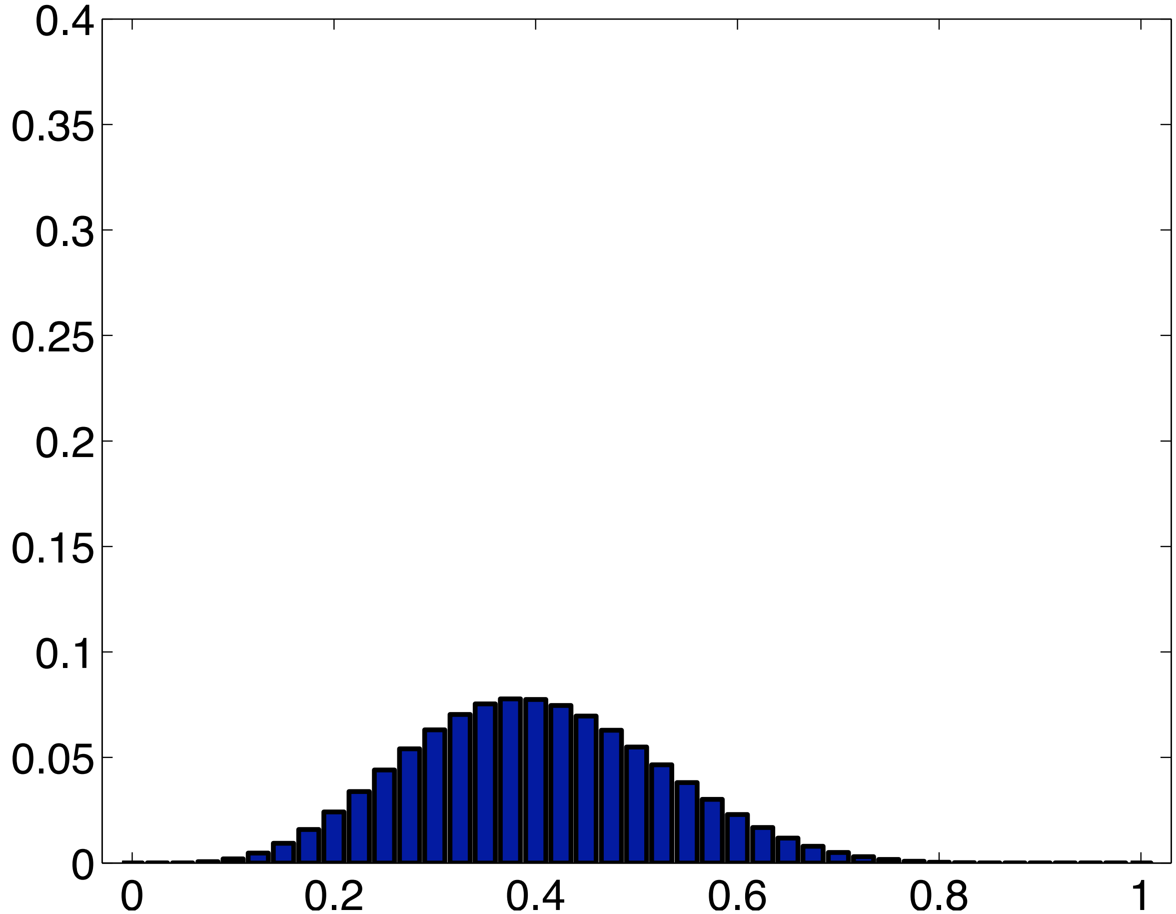
Finer & finer discretization



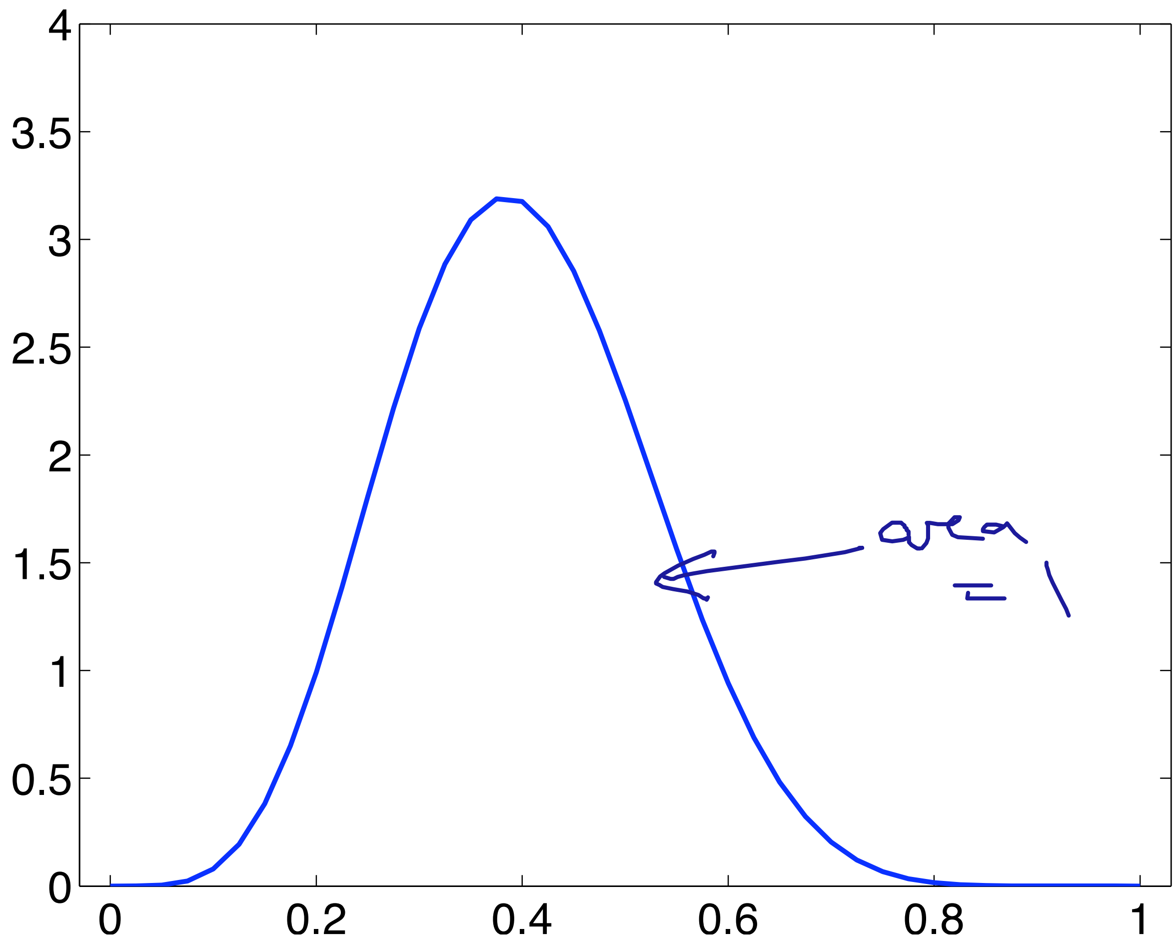
Finer & finer discretization



Finer & finer discretization



In the limit: density



$\int_0^1 P(x)\,dx = 1$

- $\lim_{h \to 0} P(x \le X \le x+h) / h = P(x)$ ← density

Properties of densities

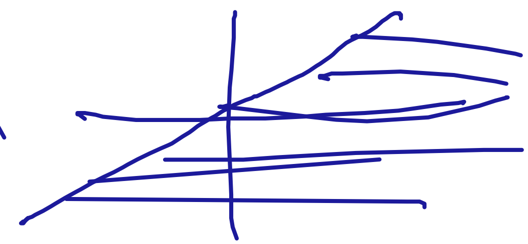
- instead of sum to 1, $\int dx$ to 1
- density may be > 1
- PDF = probability $\begin{cases} \text{density} \\ \text{distribution} \end{cases}$ function
- Confusingly, we use P(·) for both, and sometimes people say distribution to mean either discrete or continuous

mass ⇒ discrete

Events

$\mathbb{R}^2$

- For continuous RVs X, Y:
- Sample space Ω = { all settings of (X,Y) }
- Event = subset of Ω $\quad X \geq Y$

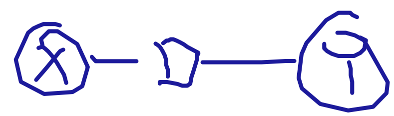


- Density: events $\rightarrow \mathbb{R}_+$
  - disjoint union: additive $\quad P(A) + P(B) = P(A \cup B)$ if $A \cap B = \emptyset$
  - $P(\Omega) = 1$

integrate to 1

Continuous RVs in graphical models

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



- $\phi(X,Y) = e^{-\pi(x-y)^2} \geq 0$
- $P(X,Y) = \frac{1}{Z} \phi(x,y)$

$$Z = \int_0^1 \int_0^1 \phi(x,y)\,dx\,dy$$

Continuous GM example