ECE521 W17 Tutorial 9

Shenlong Wang and Yuhuai (Tony) Wu

UNIVERSITY OF TORONTO

Outline

- Graphical models
- HMM example
- Family tree example

Big picture

- Probabilistic modelling is useful (K-means v.s. MoG)

Big picture

- Probabilistic modelling is useful (K-means v.s. MoG)

- Probabilistic models are represented by their joint probability distributions

Why we need it?

Big picture

- Probabilistic modelling is useful (K-means v.s. MoG)

- Probabilistic models are represented by their joint probability distributions

- Learning probabilistic models involves:
  - E-step: compute posterior distribution
  - M-step: adjust model parameters using posterior distribution

Big picture

- Probabilistic models are represented by their joint probability distributions

- Learning probabilistic models involves:
    - E-step: compute posterior distribution
    - M-step: adjust model parameters using posterior distribution

Big picture

- Probabilistic models are represented by their joint probability distributions

    Simple models only have a few conditional probabilities

- Learning probabilistic models involves:
    - E-step: compute posterior distribution
    - M-step: adjust model parameters using posterior distribution

Derive the expressions by hand using Bayes rule

Derivatives are computed using calculus chain rules (Efficient! Mechanical!)

Big picture

- Probabilistic models are represented by their joint probability distributions

  Simple models only have a few conditional probabilities

  How to represent large complicated models?

- Learning probabilistic models involves:
  - E-step: compute posterior distribution

    Derive the expressions by hand using Bayes rule

    Need efficient and mechanical algorithm for large models

  - M-step: adjust model parameters using posterior distribution

    Derivatives are computed using calculus chain rules (Efficient! Mechanical!)

Big picture

- Probabilistic models are represented by their joint probability distributions

  Simple models only have a few conditional probabilities

  How to represent large complicated models? (some circuit diagrams?)

- Learning probabilistic models involves:
  - E-step: compute posterior distribution
  - M-step: adjust model parameters using posterior distribution

Derive the expressions by hand using Bayes rule

Need efficient and mechanical algorithm for large models

Derivatives are computed using calculus chain rules (Efficient! Mechanical!)

(something like calculus chain rule?)

Big picture

- Probabilistic models are represented by their joint probability distributions
    - Represent joint distributions in terms of **graphical models,** Bayes Nets, MRFs, factor graphs
        - **Easy to spot conditional independence**
        - **Easy to spot factorization**

Make joint probability compact.

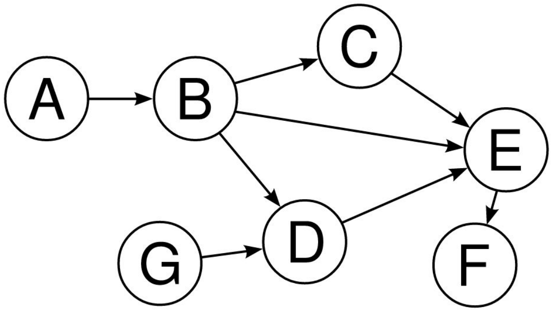
- Learning probabilistic models involves:
    - E-step: compute posterior distribution
    - M-step: adjust model parameters using posterior distribution

Big picture

- Probabilistic models are represented by their joint probability distributions
  - Represent joint distributions in terms of **graphical models,** Bayes Nets, MRFs, factor graphs
    - **Easy to spot conditional independence**
    - **Easy to spot factorization**
- Learning probabilistic models involves:
  - E-step: compute posterior distribution
    - Compute marginal and conditional probabilities with **message-passing algorithms**
      1. **Mechanical**
      2. **Computationally efficient (just like calculus chain rules / back-propagation)**
  - M-step: adjust model parameters using posterior distribution

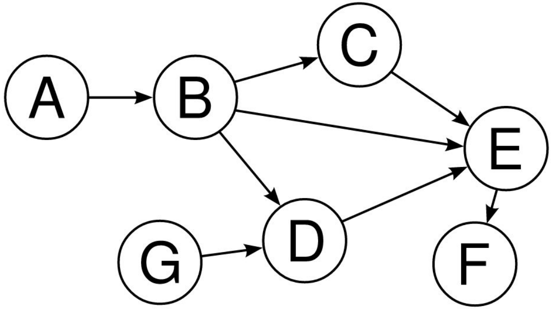
Graphical models

- Bayesian networks (i.e. BN, BayesNet), directed-acyclic-graph (DAG)

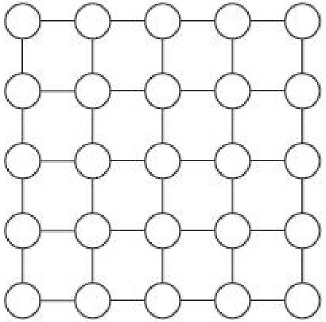


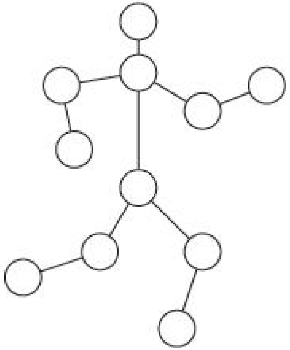
Graphical models

- Bayesian networks (i.e. BN, BayesNet), directed-acyclic-graph (DAG)



- Markov random fields, undirected graph



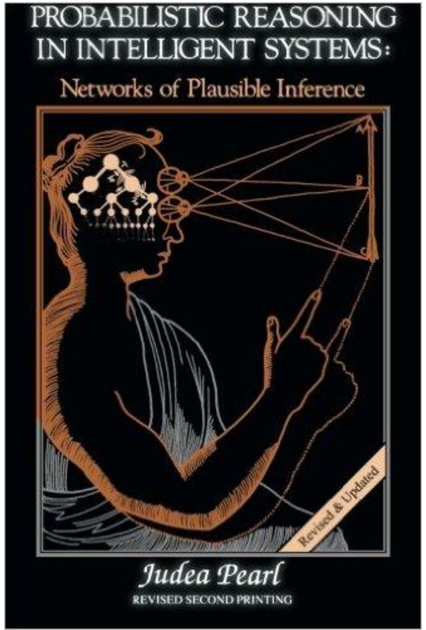


Graphical models

- Bayesian networks (i.e. BN, BayesNet)
  - Invented by computer scientists, e.g. Judea Pearl, who worked on classical AI in 1980s
- Markov random fields
  - Invented in 1970s by statisticians and statistical physicists
  - Re-invented by Geoffrey Hinton in 1980s as Boltzmann Machine

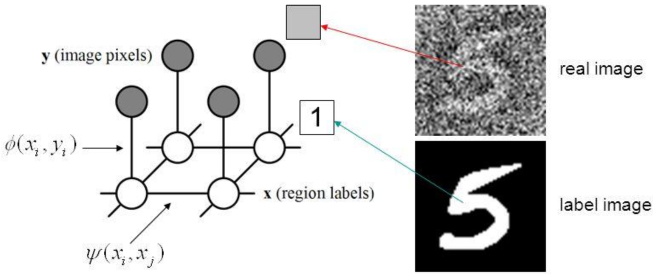
Graphical models

- Bayesian networks (i.e. BN, BayesNet)
  - Invented by computer scientists, e.g. Judea Pearl, who worked on classical AI in 1980s
- Markov random fields
  - Invented in 1970s by statisticians and statistical physicists
  - Re-invented by Geoffrey Hinton in 1980s as Boltzmann Machine



Graphical models

- Bayesian networks (i.e. BN, BayesNet)
  - Invented by computer scientists, e.g. Judea Pearl, who worked on classical AI in 1980s
- Markov random fields
  - Invented in 1970s by statisticians and statistical physicists
  - Re-invented by Geoffrey Hinton in 1980s as Boltzmann Machine

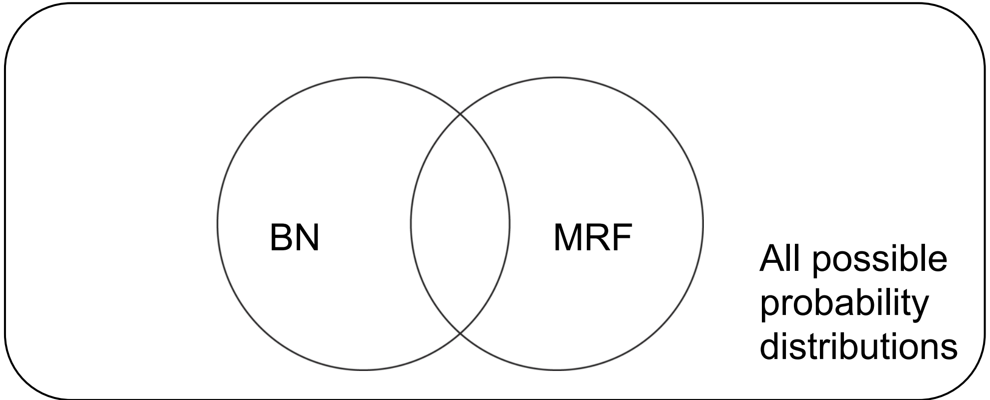


Graphical models

- A graphical model expresses two properties about a joint distribution:
  - Conditional independence
  - Factorization
- Neither BNs nor MRFs can represent all the possible conditional independence and factorization properties

Graphical models

- A graphical model expresses two properties about a joint distribution:
  - Conditional independence
  - Factorization
- Neither BNs nor MRFs can represent all the possible conditional independence and factorization properties



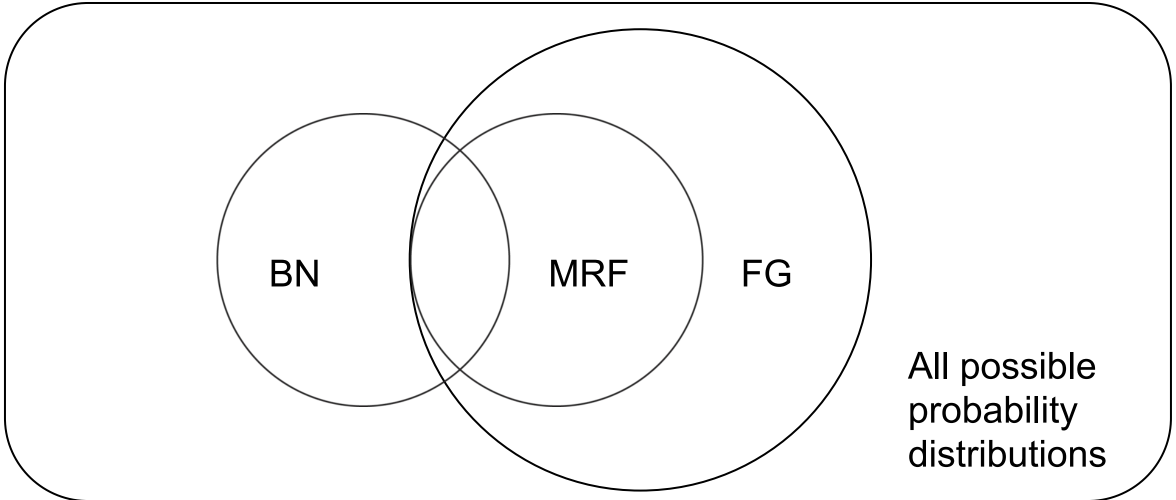
Graphical models

- BN and MRFs are not one-to-one mapping for both conditional independence and factorization properties

- i.e. if we convert the graph representations back and forth, we would obtain a different graph from what we started

Graphical models

- Factor graph is meant to unify both BN and MRF
- But...

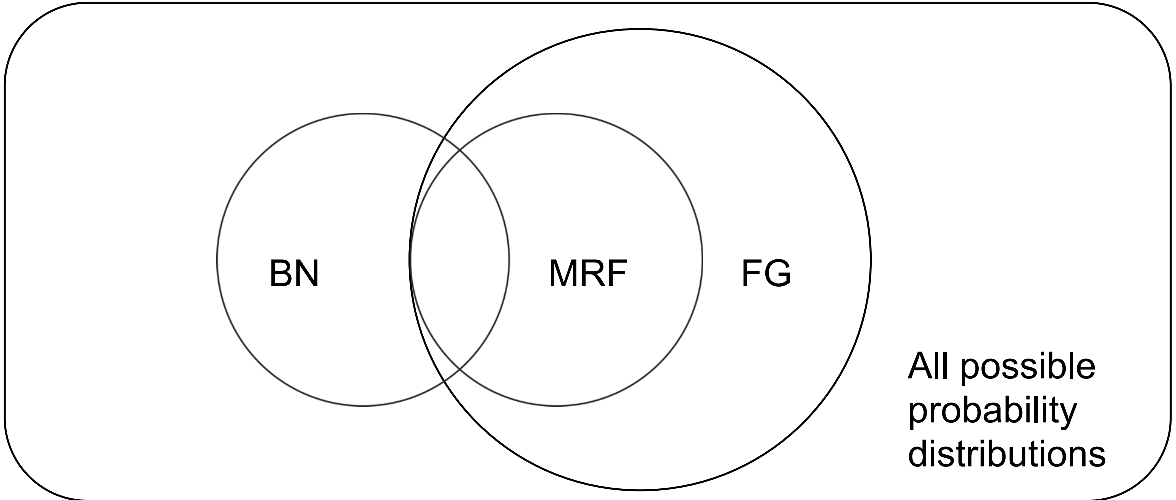


Graphical models

- Directed factor graph eventually unifies both BN and MRF (Frey, B., 2003)

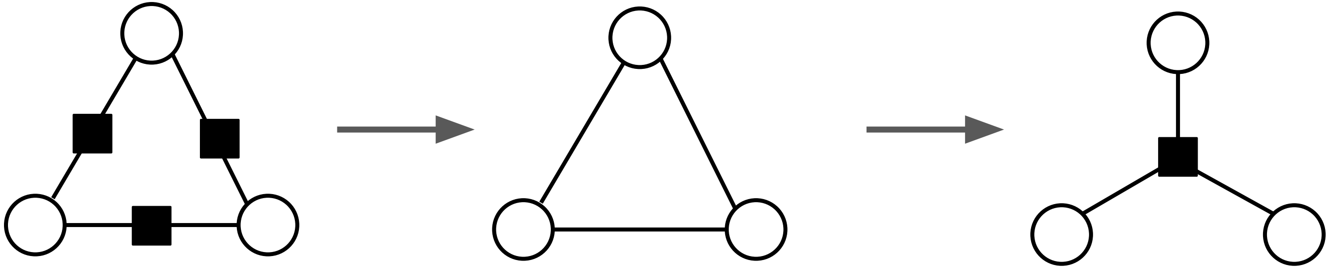
Graphical models

- Factor graph has one-to-one mapping to MRF **ONLY** in terms of conditional independence properties
- The factorization properties does not carry over during conversion



Graphical models

- Factor graph has one-to-one mapping to MRF **ONLY** in terms of conditional independence properties
- The factorization properties does not carry over during conversion



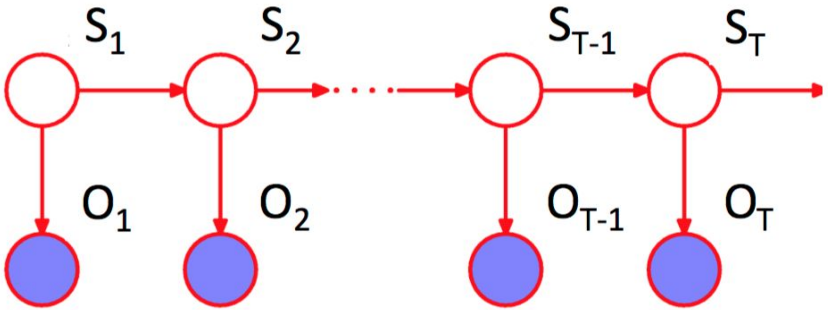
a FG expresses pair-wise factorization

Convert the FG on the left to a MRF

Convert the MRF back to a FG, we lost the factorization property

Hidden Markov Models

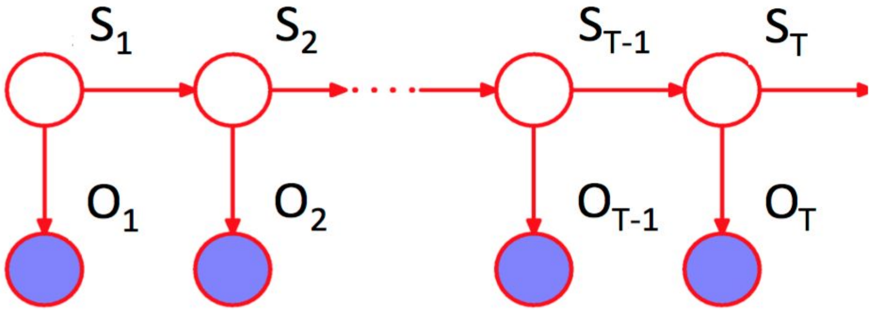
- Distributions that characterize sequential data with few parameters but are not limited by strong Markov assumptions.



Observation space $O_t \in \{y_1, y_2, \ldots, y_K\}$

Hidden states $S_t \in \{1, \ldots, I\}$

Hidden Markov Models



$$p(S_1,\ldots,S_T,O_1,\ldots,O_T) \quad = \quad \prod_{t=1}^{T} p(O_t|S_t) \prod_{t=1}^{T} p(S_t|S_{t-1})$$

1st order Markov assumption on hidden states $\{S_t\}$ t = 1, …, T (can be extended to higher order).

Note: O_t depends on all previous observations $\{O_{t-1},\ldots O_1\}$

Hidden Markov Models

- Parameters – stationary/homogeneous markov model (independent of time t)

Initial probabilities

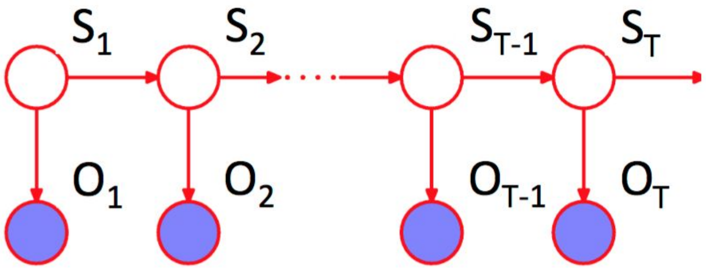
$p(S_1 = i) = \pi_i$

Transition probabilities

$p(S_t = j | S_{t-1} = i) = p_{ij}$

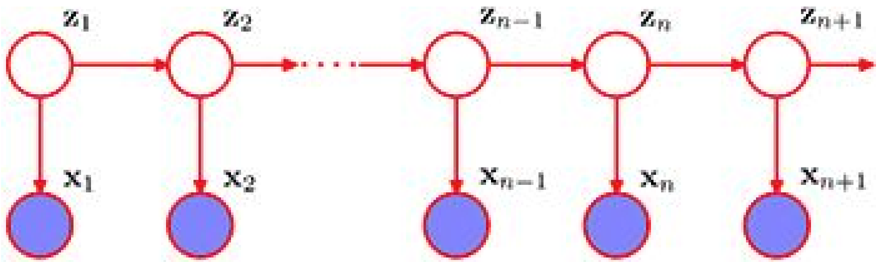
Emission probabilities

$p(O_t = y | S_t = i) = q_i^y$



$$p(\{S_t\}_{t=1}^T, \{O_t\}_{t=1}^T) = p(S_1) \prod_{t=2}^T p(S_t|S_{t-1}) \prod_{t=1}^T p(O_t|S_t)$$

Example of Hidden Markov Model



- States Z: L/H (atmospheric pressure).
- Observations X: R/D .
- Transition probabilities:

t|t-1	L	H
L	0.3	0.2
H	0.7	0.8

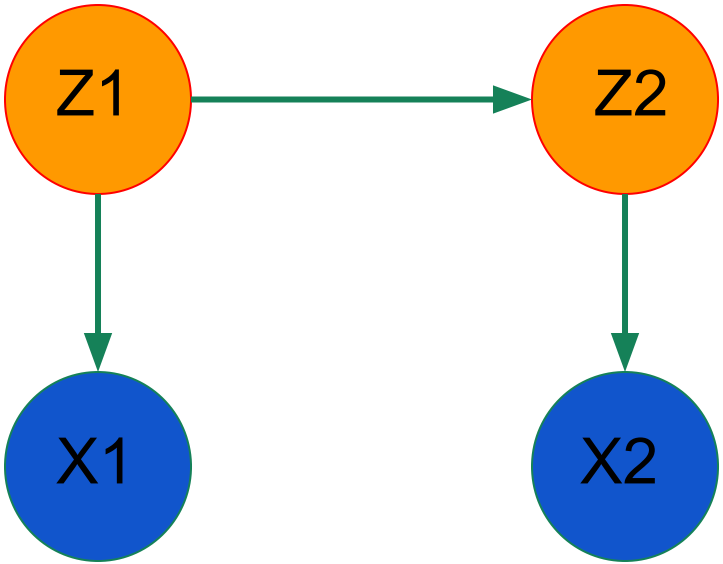
Observation probabilities:

X|Z	L	H
R	0.6	0.4
D	0.4	0.6

- Initial probabilities: say P(Z_1=L)=0.4 , P(Z_1=H)=0.6 .

Example of Hidden Markov Model

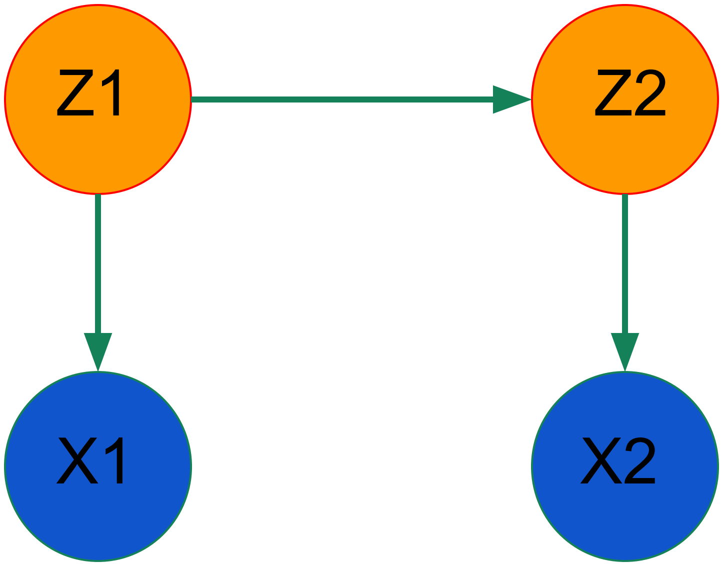
- Ex1:
- What is P(X1=D,X2=R)?



Example of Hidden Markov Model

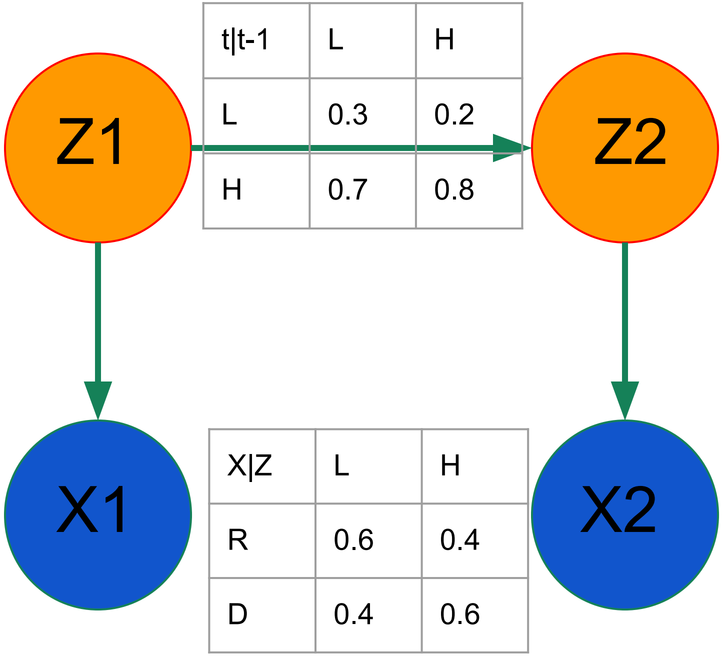
- Ex1:
- What is $P(X1=D,X2=R)$?

- $P(X1=D,X2=R) =$
  $P(X1=D,X2=R,Z1=L,Z2=L)+$
  $P(X1=D,X2=R,Z1=L,Z2=H)+$
  $P(X1=D,X2=R,Z1=H,Z2=L)+$
  $P(X1=D,X2=R,Z1=H,Z2=H)$



Example of Hidden Markov Model

- First term:
  P(X1=D,X2=R,Z1=L,Z2=L)
  = P(X1=D,X2=R| Z1=L,Z2=L)*P(Z1=L,Z2=L)
  = P(X1=D| Z1=L)*
  P(X2=R| Z2=L)*P(Z1=L)*P(Z2=L|Z1=L)
  = 0.4*0.6*0.4*0.3 = 0.0288



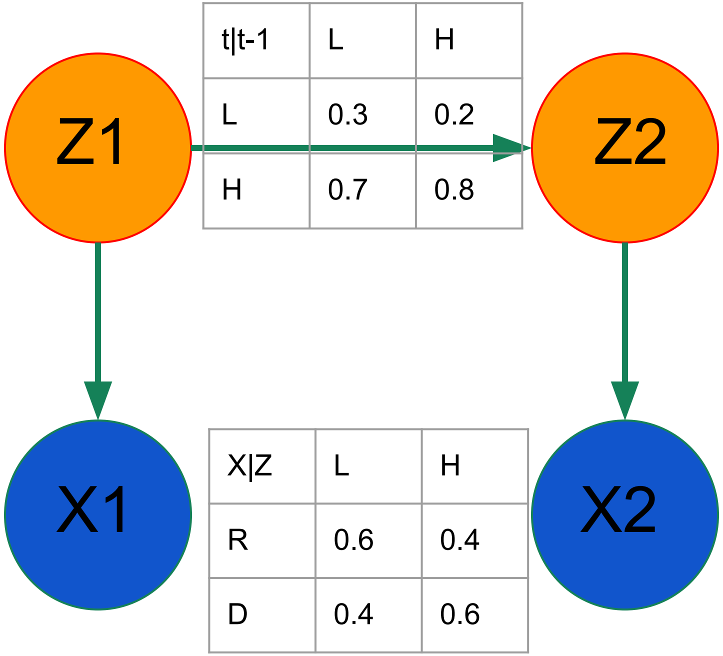
t|t-1	L	H
L	0.3	0.2
H	0.7	0.8

X|Z	L	H
R	0.6	0.4
D	0.4	0.6

Example of Hidden Markov Model

- Ex1:
- What is P(X1=D,X2=R)?

- P(X1=D,X2=R) =
  P(X1=D,X2=R,Z1=L,Z2=L)+
  P(X1=D,X2=R,Z1=L,Z2=H)+
  P(X1=D,X2=R,Z1=H,Z2=L)+
  P(X1=D,X2=R,Z1=H,Z2=H)
  = 0.4*0.6*0.4*0.3+
  0.4*0.4*0.4*0.7+
  0.6*0.6*0.6*0.2+
  0.6*0.4*0.6*0.8 = 0.232



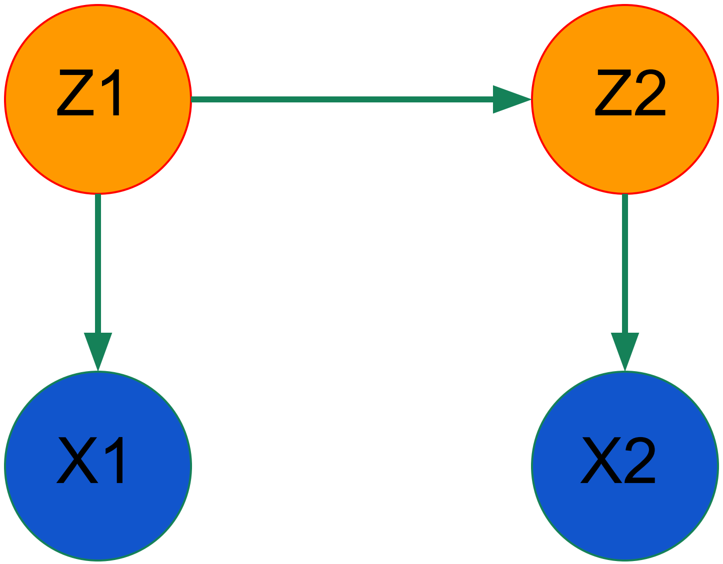
t|t-1	L	H
L	0.3	0.2
H	0.7	0.8

X|Z	L	H
R	0.6	0.4
D	0.4	0.6

Example of Hidden Markov Model

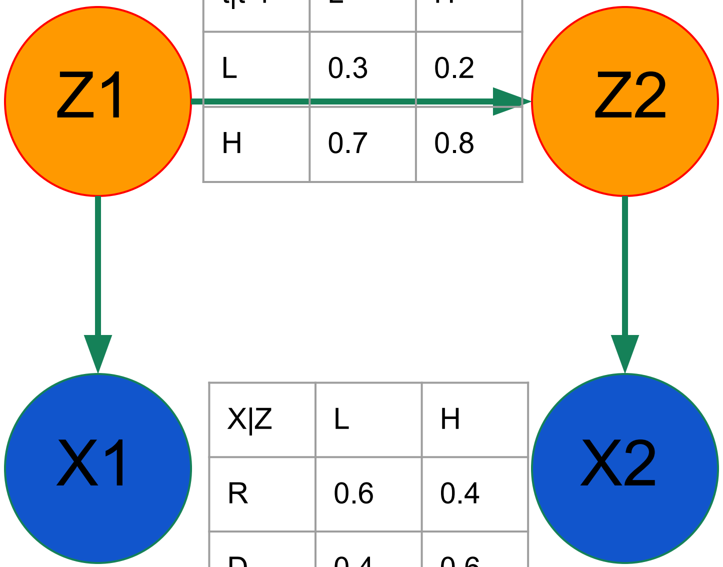
- Ex2:
- What is P(Z1=H|X1=D,X2=R)?

- P(Z1=H|X1=D,X2=R)=
  P(Z1=H,X1=D,X2=R)/P(X1=D,X2=R)



Example of Hidden Markov Model

- P(Z1=H,X1=D,X2=R)=
  P(Z1=H,Z2=L,X1=D,X2=R)+
  P(Z1=H,Z2=H,X1=D,X2=R)
  =0.6*0.6*0.6*0.2+
  0.6*0.4*0.6*0.8
  =0.1584



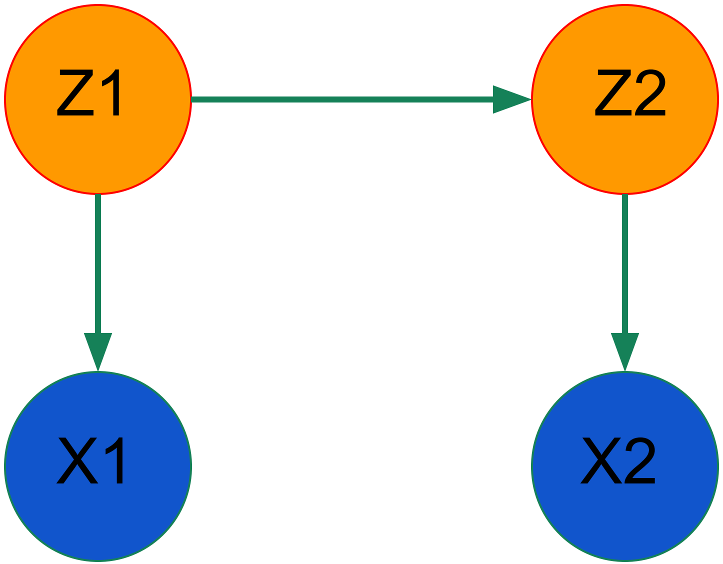
t|t-1	L	H
L	0.3	0.2
H	0.7	0.8

X|Z	L	H
R	0.6	0.4
D	0.4	0.6

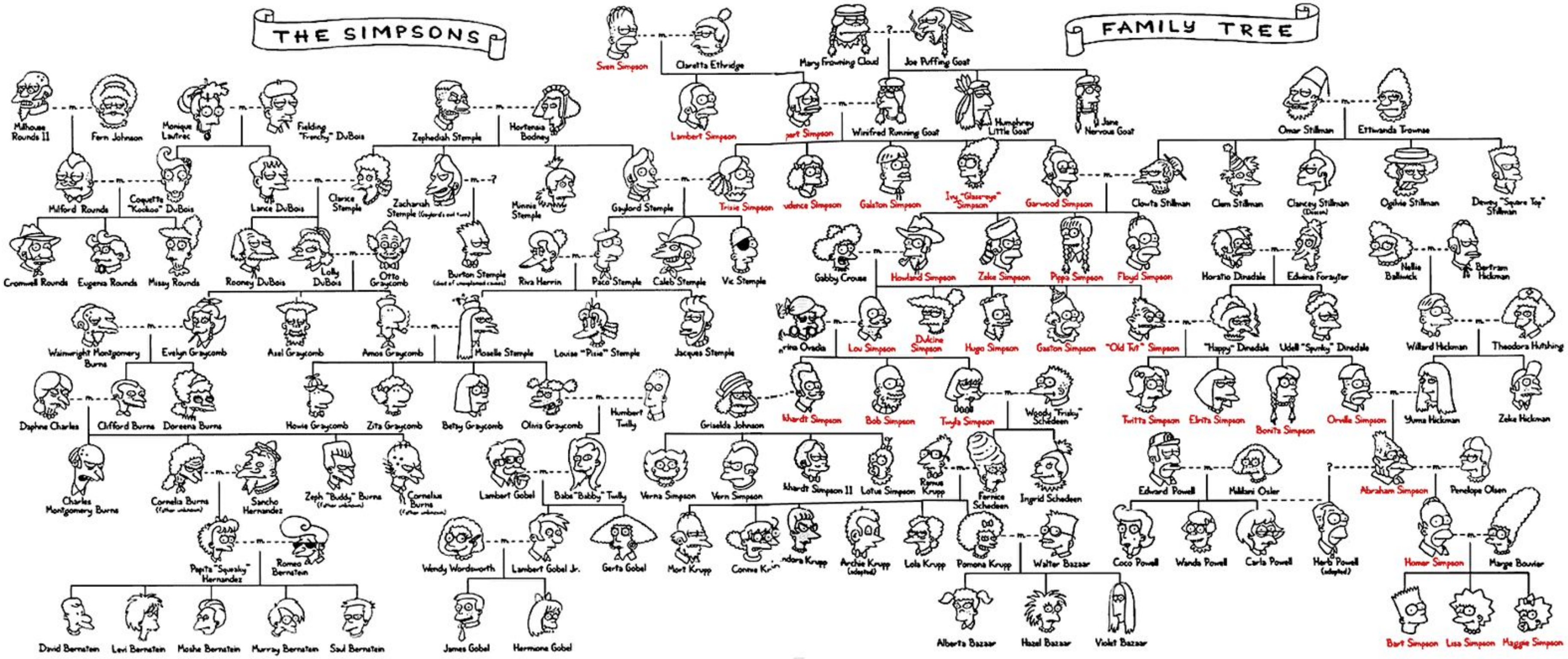
Example of Hidden Markov Model

- Ex2:
- What is $P(Z1=H|X1=D,X2=R)$?

- $P(Z1=H|X1=D,X2=R)=$
  $P(Z1=H,X1=D,X2=R)/P(X1=D,X2=R)$
  $=0.1584/0.232=0.683$

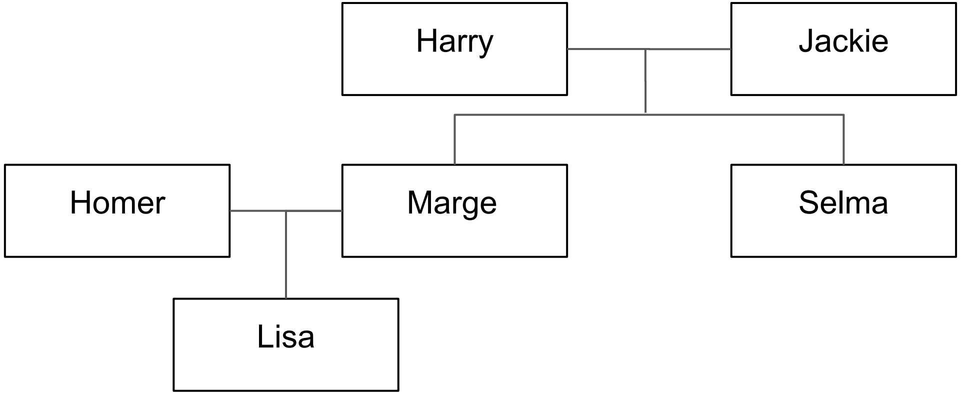


Family Tree


THE SIMPSONS
FAMILY TREE
Sven Simpson
Claretta Ethridge
Mary Frowning Cloud
Joe Puffing Goat
Lambert Simpson
Winifred Running Goat
Humphrey Little Goat
Jane Nervous Goat
Omar Stillman
Ettwanda Townse
Zephediah Stemple
Hortensia Bodney
Fern Johnson
Milford Rounds
Lance DuBois
Gaylord Stemple
Trixie Simpson
Galston Simpson
Garwood Simpson
Clowta Stillman
Clem Stillman
Clancey Stillman
Ogilvie Stillman
Cromwell Rounds
Eugena Rounds
Missy Rounds
Rooney DuBois
Otto Graycomb
Burton Stemple
Riva Herrin
Paco Stemple
Caleb Stemple
Vic Stemple
Gabby Crouse
Howland Simpson
Zeke Simpson
Pippa Simpson
Floyd Simpson
Horatio Dinadale
Edwina Forayter
Evelyn Graycomb
Axel Graycomb
Amos Graycomb
Moselle Stemple
Jacques Stemple
Lou Simpson
Hugo Simpson
Gaston Simpson
Willard Hickman
Theodora Hutshing
Daphne Charles
Clifford Burns
Doreena Burns
Howie Graycomb
Zita Graycomb
Betsy Graycomb
Olivia Graycomb
Griselda Johnson
Bob Simpson
Orville Simpson
Yuma Hickman
Zeke Hickman
Charles Montgomery Burns
Cornelia Burns
Sancho Hernandez
Lambert Gobel
Verna Simpson
Vern Simpson
Lotus Simpson
Ingrid Schedeen
Edward Powell
Abraham Simpson
Penelope Olsen
Romeo Bernstein
Wendy Wordsworth
Lambert Gobel Jr.
Gerta Gobel
Mort Krupp
Archie Krupp
Lola Krupp
Pomona Krupp
Walter Bazaar
Coco Powell
Wanda Powell
Carla Powell
Herb Powell
Homer Simpson
Marge Bouvier
David Bernstein
Levi Bernstein
Moshe Bernstein
Murray Bernstein
Saul Bernstein
James Gobel
Hermione Gobel
Alberta Bazaar
Hazel Bazaar
Violet Bazaar
Bart Simpson
Lisa Simpson
Maggie Simpson

Family Tree

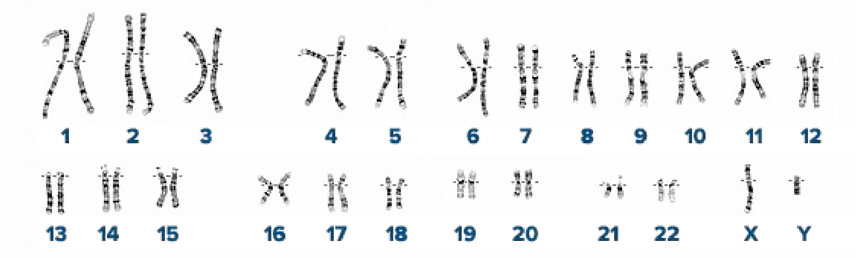
Source: Wikia

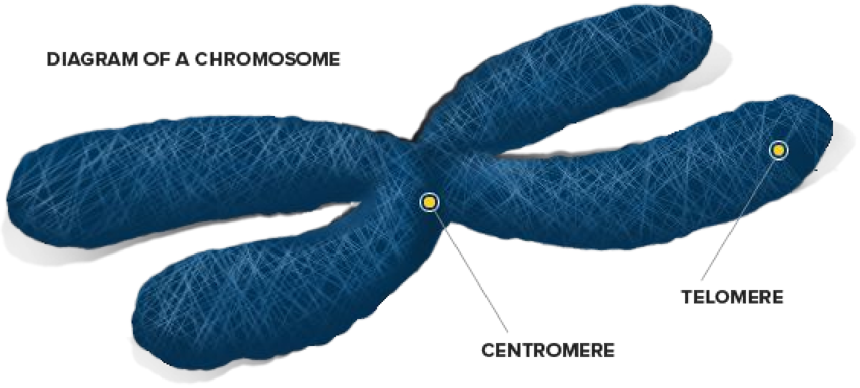


Background Knowledge

Human genetic material

- 23 pairs of chromosomes
- consisting genes that determine a person’s property
- A region of interest is called a locus, which may have several variants
- One of alleles is from father and one is from mother





Background Knowledge

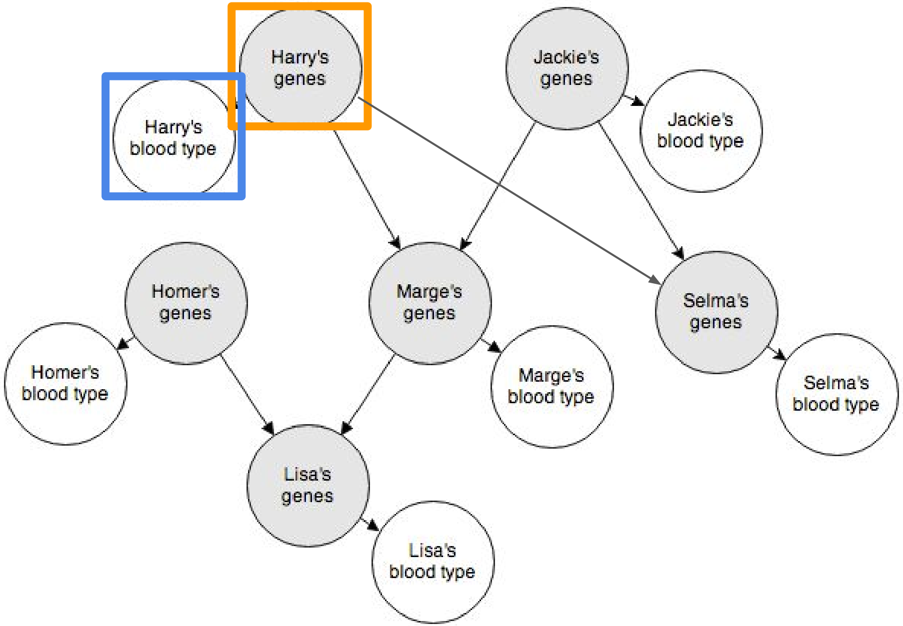
Gene for ABO Blood type

- Three alleles: A, B, O
- Everyone has an ordered pair, one from mother and one from father
- Genotype: 6 types (order-independent): (A, A), (A, B), … (O, O)
- Genotype → blood type is deterministic:
    - (A, A) → A type blood
    - (A, O) → A type blood
    - (A, B) → AB type blood
    - (O, O) → O type blood
    - (B, B) → B type blood
    - (B, O) → B type blood

Background Knowledge

How to model the problem?

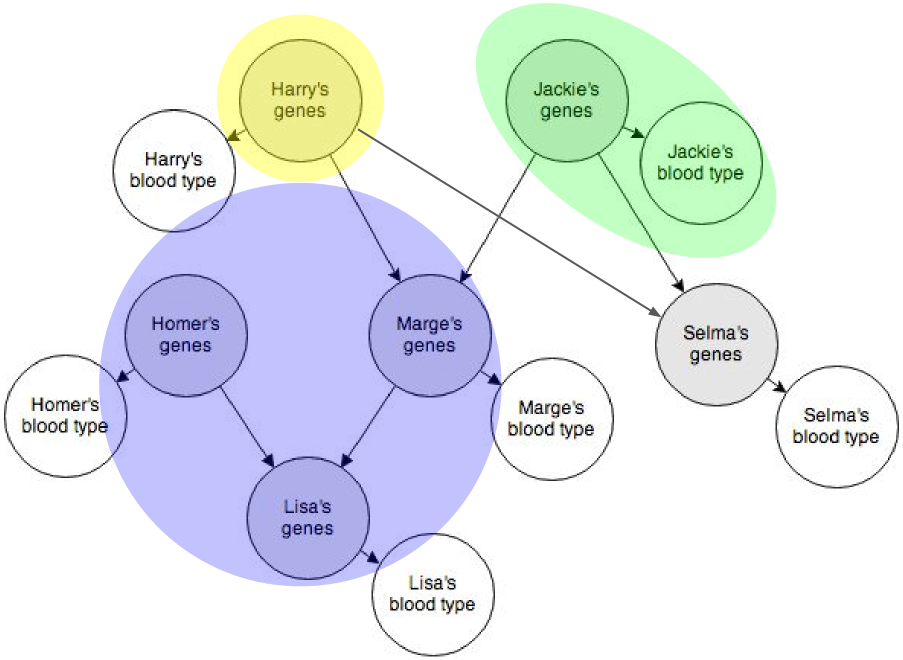
- Two types of variables: **genotype** G and **blood type** B



Background Knowledge

How to model the problem?

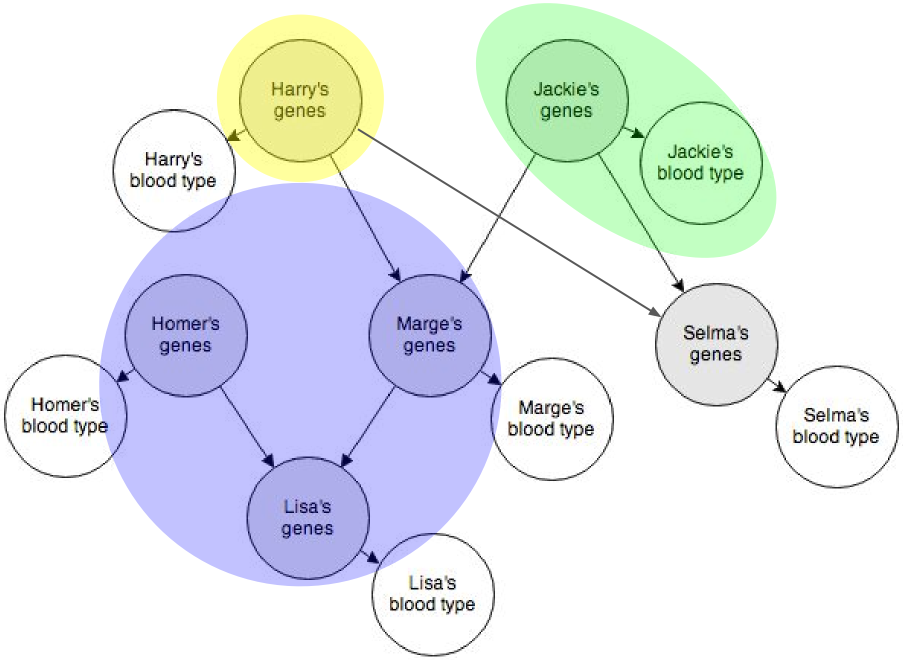
- Two types of variables: **genotype** G and **blood type** B
- Three types of conditional probability distribution:
  - **Transmission** model: P(G(c) | G(p), G(m))
  - **Penetrance** model: P(B(c) | G(c))
  - **Prior** model: P(G(c))



Background Knowledge

How to model the problem?

- Two types of variables: **genotype** G and **blood type** B
- Three types of conditional probability distribution:
  - **Transmission** model: P(G(c) | G(p), G(m))
  - **Penetrance** model: P(B(c) | G(c))
  - **Prior** model: P(G(c))



Quiz: what is the size of the conditional probability table for each model?

Background Knowledge

Conditional probability table for transmission model

- We simply use six states, instead of nine, since the order here does not matter

Parent Genes	AA	BB	AB	OO	AO	BO
AA	AA	AB	AA, AB	AO	AA, AO	AB, AO
BB	AB	BB	AB, BB	BO 1	AB, BO	BB, BO
AB	AA, AB	AB, BB	AA, BB, AB	AO, BO	AA, AO, BO, AB	BB, BO, AB, AO
OO	AO	BO	AO 0.5, BO 0.5	OO	AO, OO	BO, OO
AO	AA, AO	AB, BO	AA, AB, AO, BO	AO, OO	AA 0.25, AO 0.5, OO 0.25	AO, BO, AB, OO
BO	AO, AB	BB, BO	AB, AO, BB, BO	BO, OO	AB, AO, BO, OO	BB, BO, OO

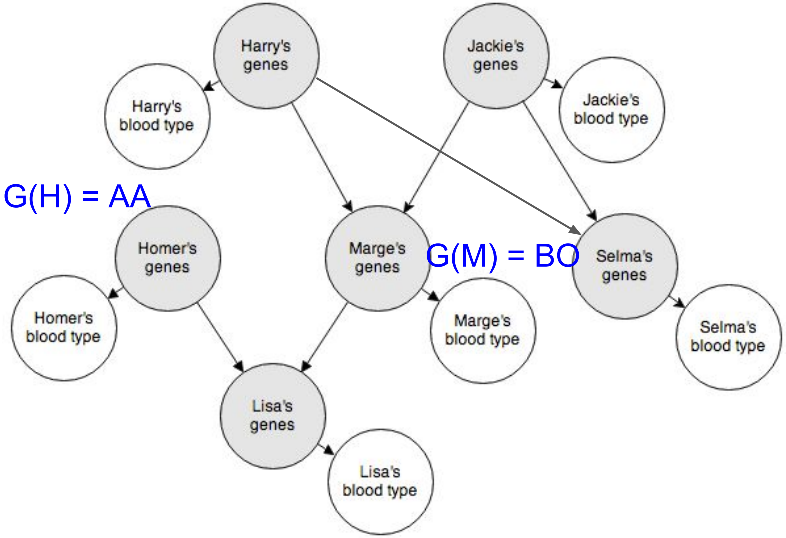
Background Knowledge

Prior probability for root nodes

- We simply assume equal prior probability over the six states

Bayesian Network for Genetics Inheritance

- Warm-up Example-a
  - Observations:
    - G(H) = G(Homer) = AA
    - G(M) = G(Marge) = BO
  - Question:
    - p(B(L)|obs) = p(B(Lisa)|obs) = ?



Bayesian Network for Genetics Inheritance

- Warm-up Example-a
  - Observations:
    - G(H) = G(Homer) = AA
    - G(M) = G(Marge) = BO
  - Question:
    - p(B(L)|obs) = p(B(Lisa)|obs) = ?

p(B(L)=j | G(H)=AA, G(M)=BO)

Marginalize out G(L)

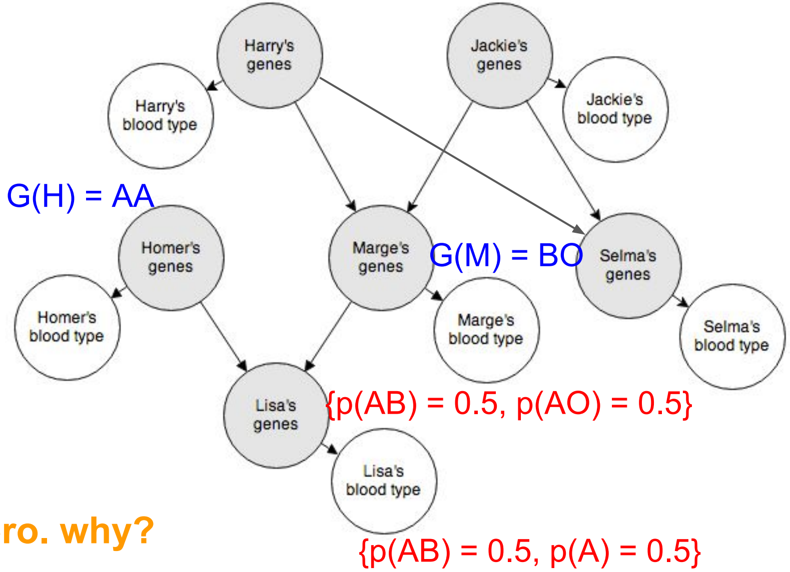
= sum_i P(B(L) | G(L) = i) p(G(L) = i | G(H) = AA, G(M) = BO)

= P(B(L)= j | G(L) = AB) p (G(L) = AB | G(H) = AA, G(M) = BO) +
P(B(L) = j | G(L) = AO) p (G(L) = BO | G(H) = AA, G(M) = BO)

Rest terms = zero. why?

→ P(B(L) = AB | G(H)=AA, G(M)=BO) = 1 * ½ + 0 = ½

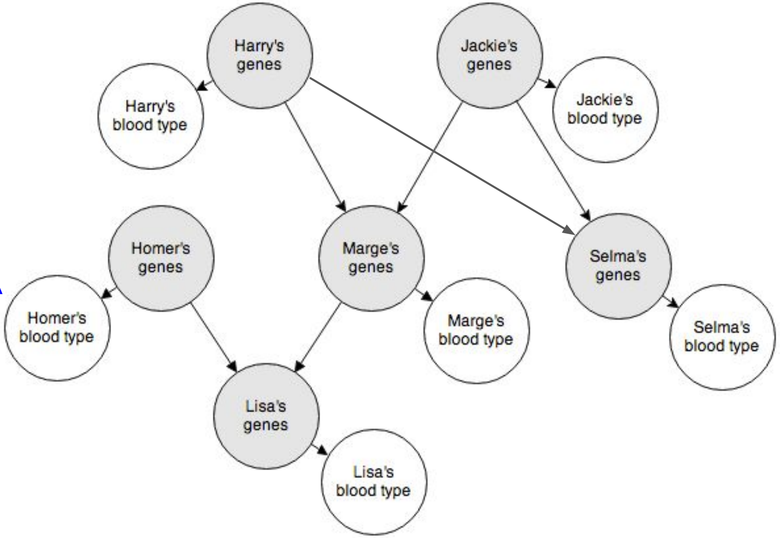
P(B(L) = A | G(H)=AA, G(M)=BO) = 0 + 1 * ½ = ½



Bayesian Network for Genetics Inheritance

- Warm-up Example-b
  - Observations:
    - B(H) = B(Homer) = A
  - Question:
    - p(G(H)|obs) = p(G(Homer)|obs) = ?

B(H) = A

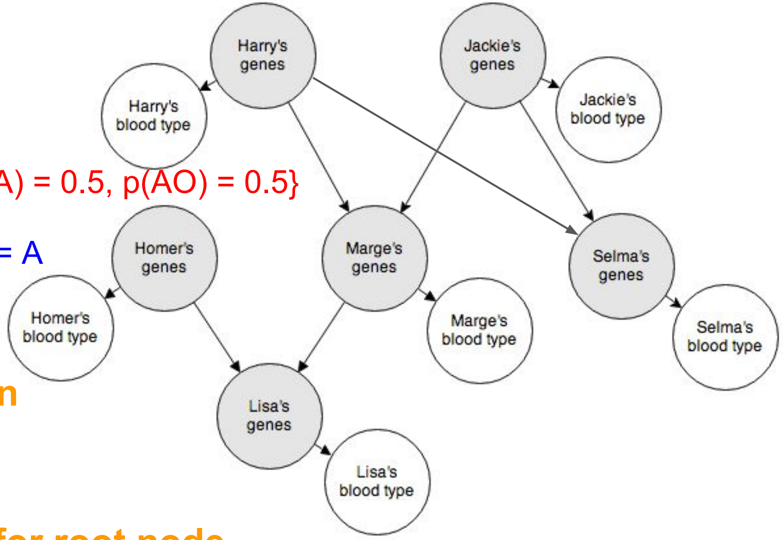


Bayesian Network for Genetics Inheritance

- Warm-up Example-b
  - Observations:
    - B(H) = B(Homer) = A
  - Question:
    - p(G(H)|obs) = p(G(Homer)|obs) = ?

{p(AA) = 0.5, p(AO) = 0.5}

B(H) = A



p(G(H)=j | B(H)=A)

Bayes Rule

= P(B(H) = A | G(H) = j) p(G(H) = j) / P(B(H) = A)

Marginalization

= P(B(H) = A | G(H) = j) p(G(H) = j) / (sum_i P(B(H) = A | G(H) = i)

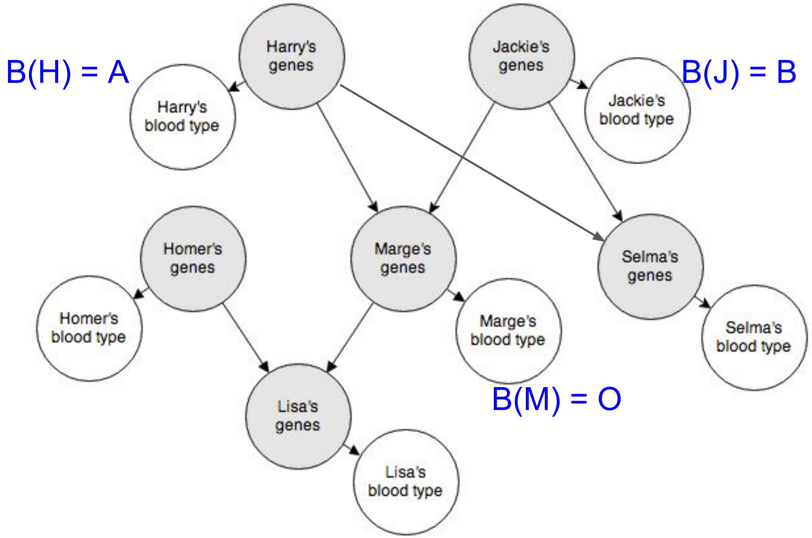
$\rightarrow$ $P(G(H) = AA \mid B(H) = A) = (1 * \frac{1}{6}) / (1 * \frac{1}{6} + 1 * \frac{1}{6}) = \frac{1}{2}$

Prior for root node

$P(G(H) = AO \mid B(H) = A) = (1 * \frac{1}{6}) / (1 * \frac{1}{6} + 1 * \frac{1}{6}) = \frac{1}{2}$

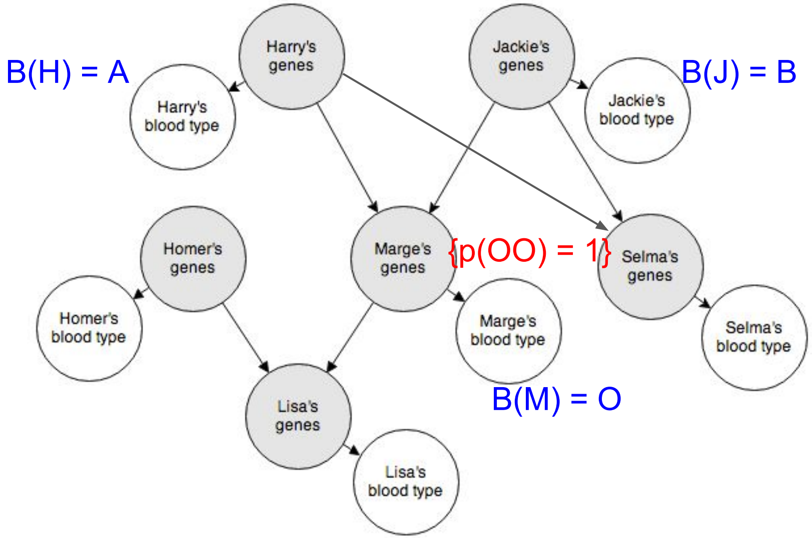
Bayesian Network for Genetics Inheritance

- Example 1
  - Observations:
    - B(Harry) = A
    - B(Jackie) = B
    - B(Marge) = O
  - Question:
    - p(B(Selma)|obs) = ?



Bayesian Network for Genetics Inheritance

- Example 1
  - Observations:
    - B(Harry) = A
    - B(Jackie) = B
    - B(Marge) = O
  - Question:
    - p(B(Selma)|obs) = ?



p(G(M)=OO | B(M)=O) = 1

Try to derive this by yourself using the warm-up example-b

Bayesian Network for Genetics Inheritance

$p(G(H)=i, G(J)=j \mid G(M)=OO, B(H) = A, B(J) = B)$

$= P(G(M)=OO, B(H) = A, B(J) = B \mid G(H)=i, G(J)=j)\, p(G(H)=i, G(J)=j) \,/\, P(G(M)=OO, B(H) = A, B(J) = B)$

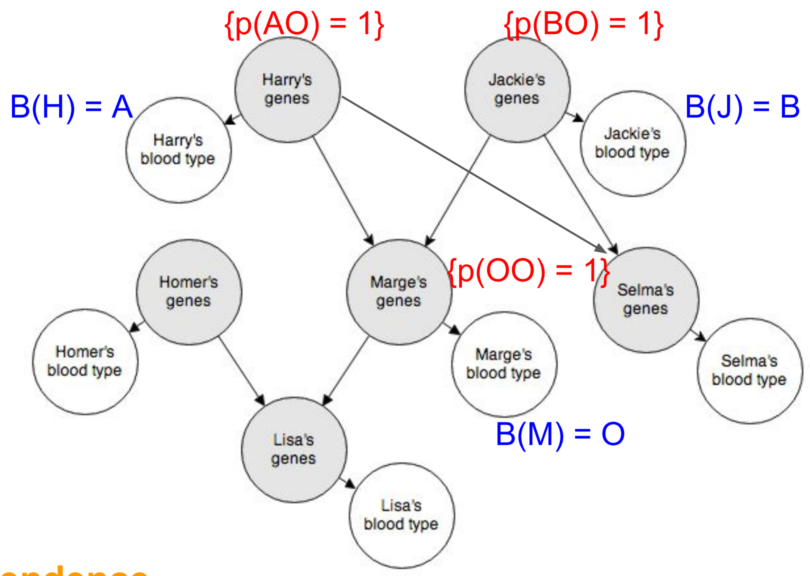
Nothing but a huge bayes rule

$= P(G(M)=OO, B(H) = A, B(J) = B \mid G(H)=i, G(J)=j)\, p(G(H)=i)\, p(G(J)=j) \,/\, \text{sum}_{\{m, n\}} (P(G(M)=OO, B(H) = A, B(J) = B \mid G(H)=m, G(J)=n)\, p(G(H)=m)\, p(G(J)=n))$

Independence **Marginalization**

$= P(G(M)=OO \mid G(H)=i, G(J)=j)\, P(B(H)=A \mid G(H)=i)\, P(B(J)=B \mid G(J)=j)\, p(G(H)=i)\, p(G(J)=j) \,/\, \text{sum}_{\{m,n\}} (P(G(M)=OO \mid G(H)=m, G(J)=n)\, P(B(H)=A \mid G(H)=i)\, P(B(J)=B \mid G(J)=j))\, p(G(H)=m)\, p(G(J)=n))$

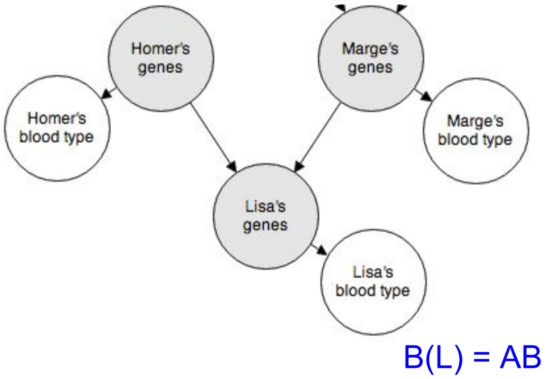
Conditional independence



$p(G(H)=AO, G(J)=BO \mid G(M)=OO, B(H) = A, B(J) = B) = \frac{1}{4} * 1 * 1 * \frac{1}{6} * \frac{1}{6} \,/\, (\frac{1}{4} * 1 * 1 * \frac{1}{6} * \frac{1}{6}) = 1$

Bayesian Network for Genetics Inheritance

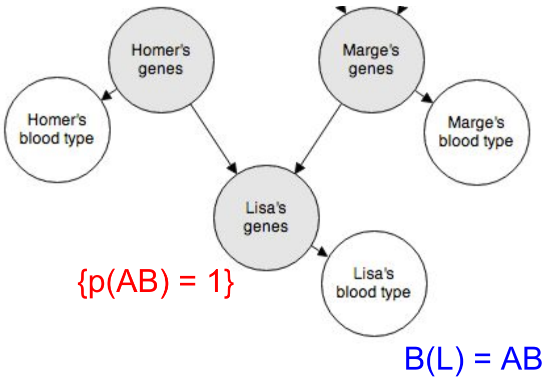
- Example 2
  - Observations:
    - **B(Lisa) = AB**
  - Question:
    - Who is more likely to have a B-type blood given the observation, Homer or Marge?



Bayesian Network for Genetics Inheritance

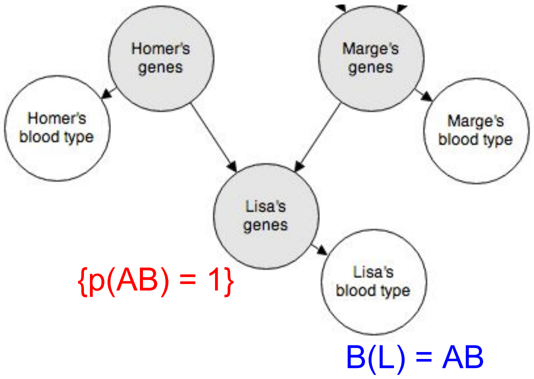
- Example 2
  - Observations:
    - **B(Lisa) = AB**
  - Question:
    - Who is more likely to have a B-type blood given Lisa has an AB type, Homer or Marge?

Bayes rule, similar to warm-up example-b



Bayesian Network for Genetics Inheritance

- Example 2
  - Observations:
    - **B(Lisa) = AB**
  - Question:
    - Who is more likely to have a B-type blood given Lisa has a AB type, Homer or Marge?



p(G(H) = i, G(M) = j | G(L) = AB)

= p(G(L) = AB | G(H) = i, G(M) = j) p(G(H) = i, G(M) = j) / P(G(L) = AB)

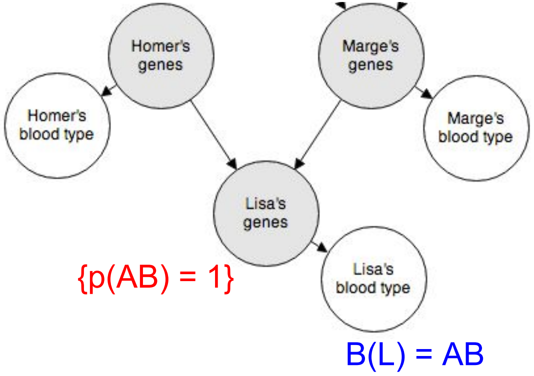
= p(G(L) = AB | G(H) = i, G(M) = j) p(G(H) = i) p(G(M) = j) / sum_{m, n} p(G(L) = AB | G(H) = m, G(M) = n) p(G(H) = m) p(G(M) = n)

Quiz: how many possible combinations of G(H) = i, G(M) = j that could make an AB baby?

Bayesian Network for Genetics Inheritance

G(H)	G(M)	P(G(H)=i)	P(G(M)=j)	P(G(L) = AB| G(M), G(H))
AB	AA	1/6	1/6	1/2
AB	AB	1/6	1/6	1/2
AB	BB	1/6	1/6	1/2

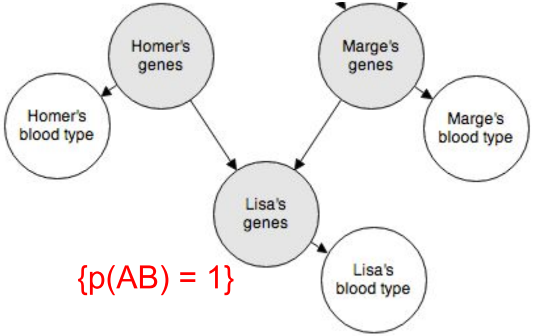
We build a large table for all combinations (cont'd).



Bayesian Network for Genetics Inheritance

G(H)	G(M)	P(G(H)=i)	P(G(M)=j)	P(G(L) = AB| G(M), G(H))
AA	AB	1/6	1/6	1/2
AA	BB	1/6	1/6	1
AA	BO	1/6	1/6	1/2
AO	BB	1/6	1/6	1/2
AO	AB	1/6	1/6	1/4
AO	BO	1/6	1/6	1/4

We build a large table for all combinations (cont'd).



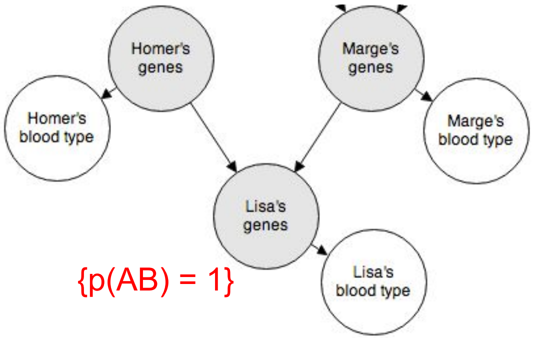
{p(AB) = 1}

B(L) = AB

Bayesian Network for Genetics Inheritance

G(H)	G(M)	P(G(H)=i)	P(G(M)=j)	P(G(L) = AB| G(M), G(H))
BO	AA	1/6	1/6	1/2
BO	AB	1/6	1/6	1/4
BO	AO	1/6	1/6	1/4
BB	AB	1/6	1/6	1/2
BB	AA	1/6	1/6	1
BB	AO	1/6	1/6	1/2

We build a large table for all combinations (cont'd).

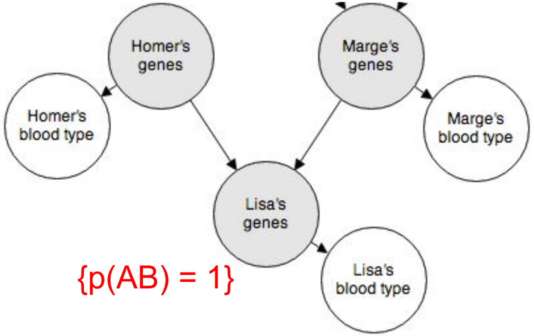


{p(AB) = 1}

B(L) = AB

Bayesian Network for Genetics Inheritance

G(H)	G(M)	P(G(H)=i)	P(G(M)=j)	P(G(L) = AB| G(M), G(H))
BO	**AA**	**1/6**	**1/6**	**1/2**
BO	AB	1/6	1/6	1/4
BO	AO	1/6	1/6	1/4
BB	AB	1/6	1/6	1/2
BB	AA	1/6	1/6	1
BB	AO	1/6	1/6	1/2



{p(AB) = 1}

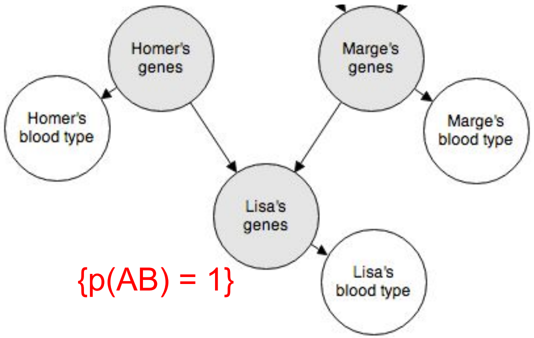
B(L) = AB

p(G(H) = i, G(M) = j | G(L) = AB) = p(G(L) = AB | G(H) = i, G(M) = j) p(G(H) = i) p(G(M) = j) / big_sum

E.g. P(G(H) = BO, G(M) = AA | G(L) = AB) = (½ * ⅙ * ⅙) / big_sum

Bayesian Network for Genetics Inheritance

G(H)	G(M)	P(G(H)=i)	P(G(M)=j)	P(G(L) = AB| G(M), G(H))
BO	**AA**	**1/6**	**1/6**	**1/2**
BO	**AB**	**1/6**	**1/6**	**1/4**
BO	**AO**	**1/6**	**1/6**	**1/4**
BB	AB	1/6	1/6	1/2
BB	AA	1/6	1/6	1
BB	AO	1/6	1/6	1/2



{p(AB) = 1}

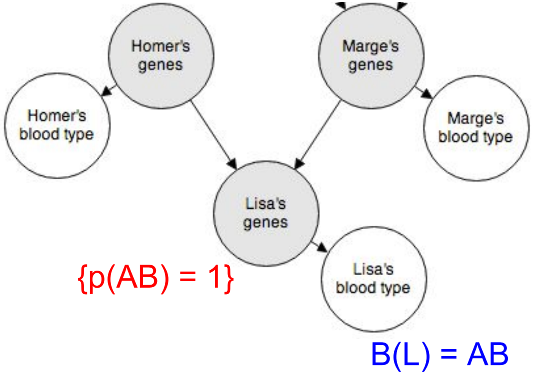
p(G(H) = i | G(L) = AB) = sum_j p(G(H) = i, G(M) = j | G(L) = AB) **Marginalization**

B(L) = AB

E.g. P(G(H) = BO | G(L) = AB) = (⅙ * ⅙ * ½) / big_sum + (⅙ * ⅙ * ¼) / big_sum + (⅙ * ⅙ * ¼) / big_sum

Bayesian Network for Genetics Inheritance

G(H)	G(M)	P(G(H)=i)	P(G(M)=j)	P(G(L) = AB| G(M), G(H))
BO	**AA**	**1/6**	**1/6**	**1/2**
BO	**AB**	**1/6**	**1/6**	**1/4**
BO	**AO**	**1/6**	**1/6**	**1/4**
BB	**AB**	**1/6**	**1/6**	**1/2**
BB	**AA**	**1/6**	**1/6**	**1**
BB	**AO**	**1/6**	**1/6**	**1/2**



Finally, we can get p(B(H) = i | B(L) = AB) by marginalizing out G(H)

p(B(H) = i | G(L) = AB) = sum_j p(B(H) = i| G(H) = j, G(L) = AB) p(G(H) = j | G(L) = AB)

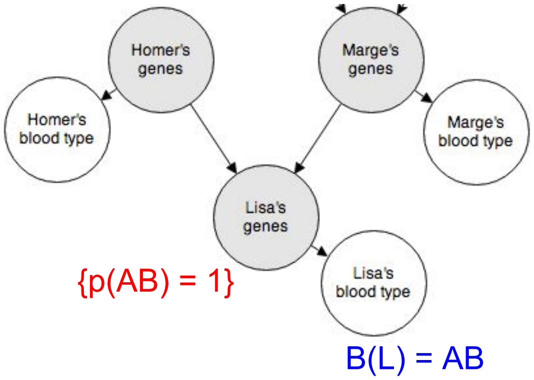
Bayesian Network for Genetics Inheritance

- Example 2
  - Observations:
    - **B(Lisa) = AB**
  - Question:
    - Who is more likely to have a B-type blood given Lisa has a AB type, Homer or Marge?

p(B(H) = A | B(L) = AB) = (3/36) / (3/72+3/36+3/36) = 0.4 = p(B(M) = A| B(L) = AB)

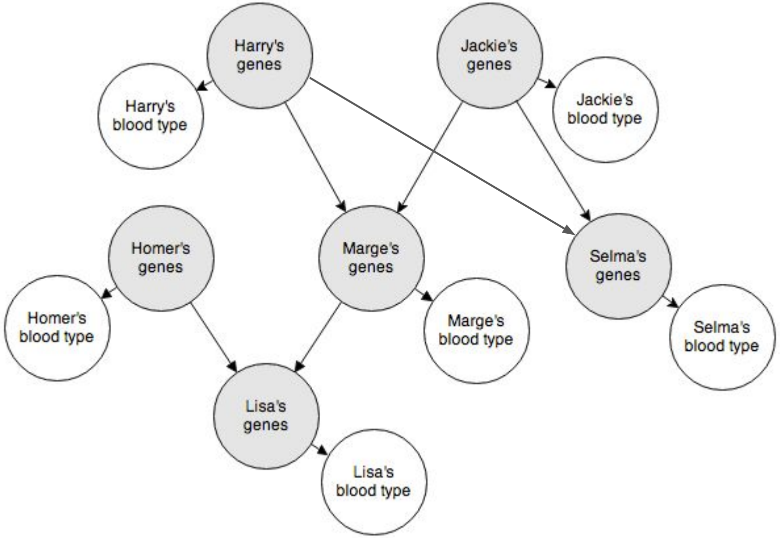
p(B(H) = B | B(L) = AB) = (3/36) / (3/72+3/36+3/36) = 0.4 = p(B(M) = B | B(L) = AB)

p(B(H) = AB | B(L) = AB) = (3/72) / (3/72+3/36+3/36) = 0.2 = p(B(M) = AB | B(L) = AB)



Bayesian Network for Genetics Inheritance

- Example 3
  - Observations:
    - B(Lisa) = AB
    - **B(Selma) = B**
  - Question:
    - Who is more likely to have a B-type blood given the observations, Homer or Marge?



What if you have one more observation?
Try to solve this problem by yourself.

Quiz: do you need to marginalize the grandma out?