CS 480

Introduction to Artificial Intelligence

October 26th, 2021

Announcements / Reminders

- **Programming Assignment #01:**
  - **due: ~~October 17th~~ ~~October 22th~~ ~~October 24th~~ October 24th, 11:00 PM CST**
- **Programming Assignment #02:**
  - **This week. Topic: CSPs**
- **Written Assignment #03:**
  - **This week**
- **Grading TA assignment:**

  https://docs.google.com/spreadsheets/d/1Cav_GBTGC7fLGzxuBCAUmEuJYPeF-HMLCYvwPbq8Fus/edit?usp=sharing

Programming Assignment 01 Grading

- File naming correct and code comments: [5/100]
- Report comparison / summary: [5/100]
- Greedy Best First Search algorithm: [45/100]
  - has to find an existing path correctly
  - has to report failure for non-existing path correctly
- A* algorithm: [45/100]
  - has to find an existing path correctly
  - has to report failure for non-existing path correctly
- Execution time will not be graded

Plan for Today

- **Quantifying and dealing with uncertainty**

Bayes' Rule: Another Interpretation

Another way to think about Baye's rule: it allows us to update the hypothesis H in light of some new data/evidence e.

$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

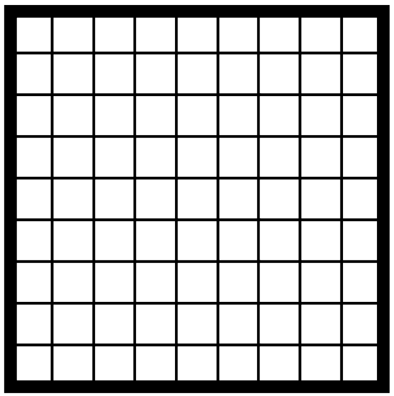
$$P(Hypothesis \mid evidence) = \frac{P(evidence \mid Hypothesis) * P(Hypothesis)}{P(evidence)}$$

where:

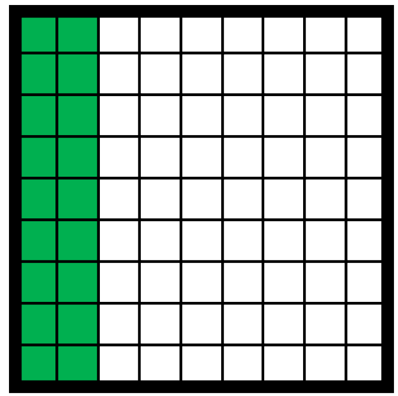
- P(H) - **probability of the** Hypothesis H **being true BEFORE we see new data/**evidence e **(<u>prior</u> probability)**
- P(H | e) - **probability of the** Hypothesis H **being true AFTER we see new data/**evidence e **(<u>posterior</u> probability)**
- P(e | H) - **probability of new data/**evidence e **being true under the** Hypothesis H **(<u>likelihood</u>)**
- P(e) - **probability of new data/**evidence e **being true under ANY hypothesis (<u>normalizing constant</u>)**

Bayes' Rule: Visual Interpretation

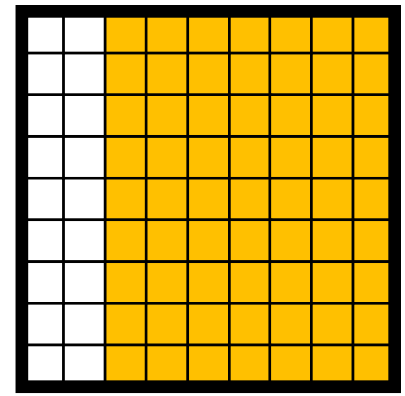
All possible cases



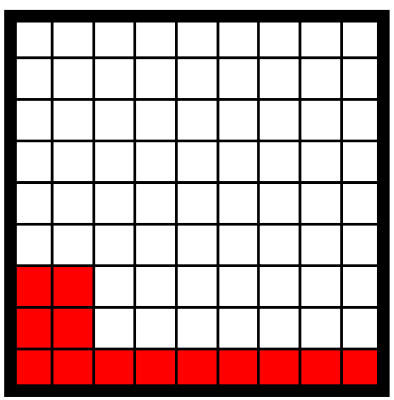
Cases where Hypothesis H **is true**
P(H)



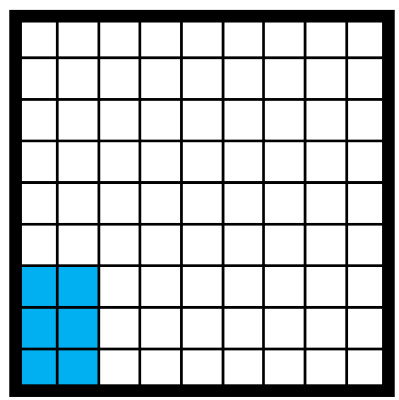
Cases where Hypothesis H **is false**
P(¬H)



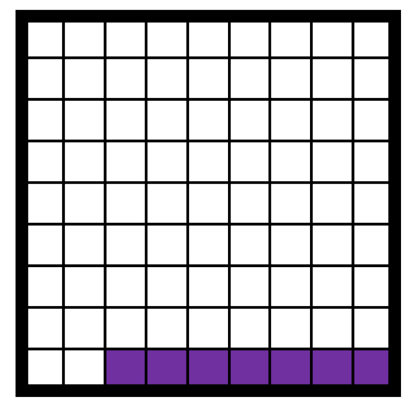
Cases where evidence e **is true**
P(e)



Cases where evidence e **is true**
given Hypothesis H **true** P(e | H)



Cases where evidence e **is true**
given Hypothesis H **false** P(e | ¬H)



Bayes' Rule: Visual Interpretation

Bayes' Rule:

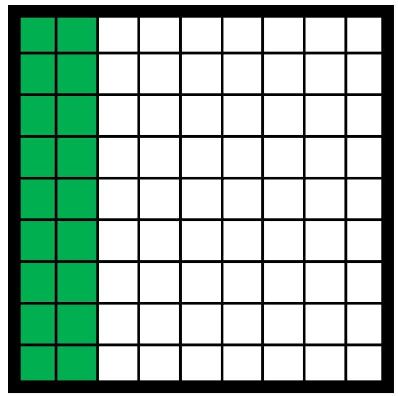
$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(H) * P(e \mid H) + P(\neg H) * P(e \mid \neg H)}$$

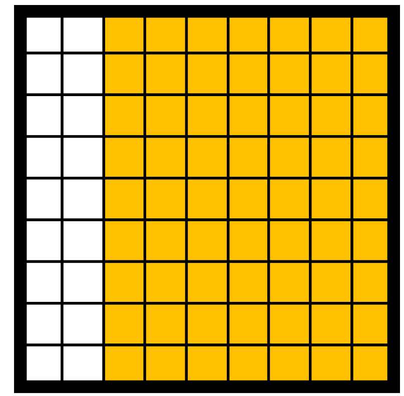
Cases where Hypothesis H is true

$P(H)$



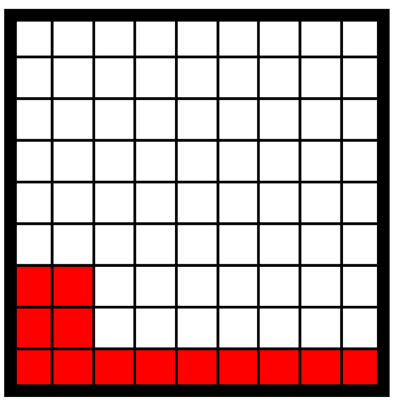
Cases where Hypothesis H is false

$P(\neg H)$

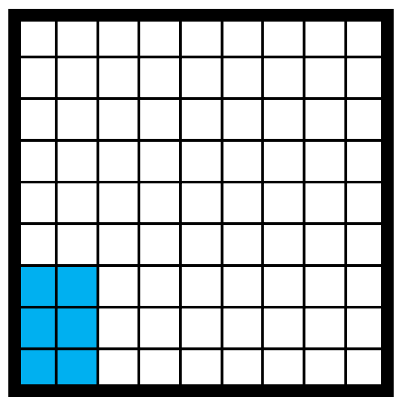


Cases where evidence e is true

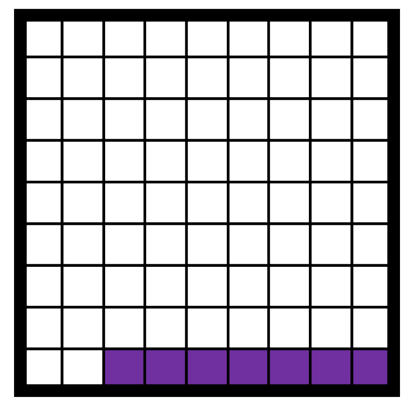
$P(e)$



Cases where evidence e is true given Hypothesis H true $P(e \mid H)$

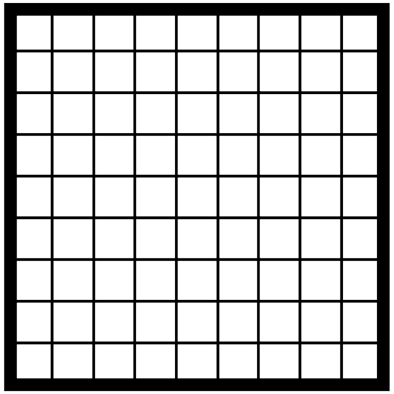


Cases where evidence e is true given Hypothesis H false $P(e \mid \neg H)$

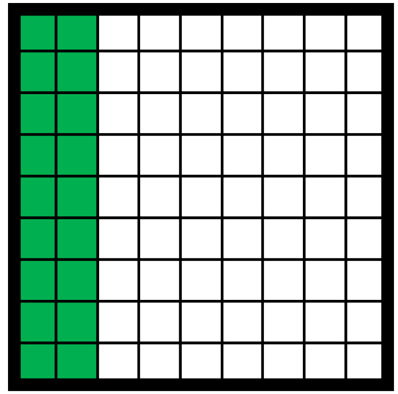


Bayes’ Rule: Visual Interpretation

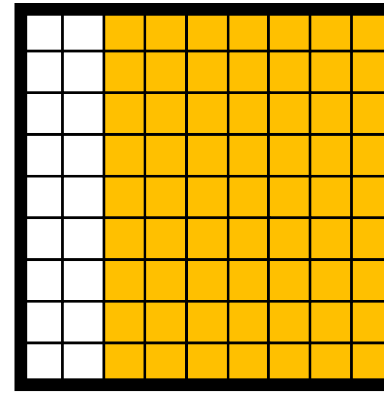
All CS 480 Students
Hypothesis H: **graduate student**



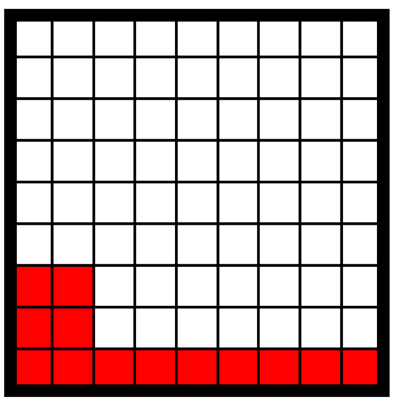
Cases where Hypothesis H **is true**
P(H) = P(grad = true)



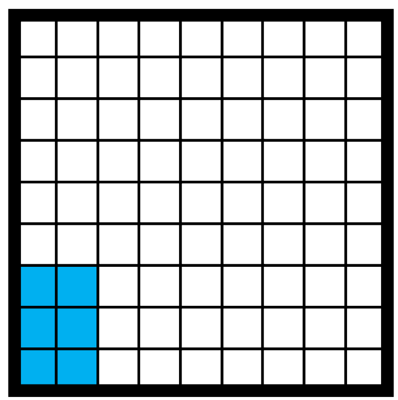
Cases where Hypothesis H **is false**
P(¬H) = P(grad = false)



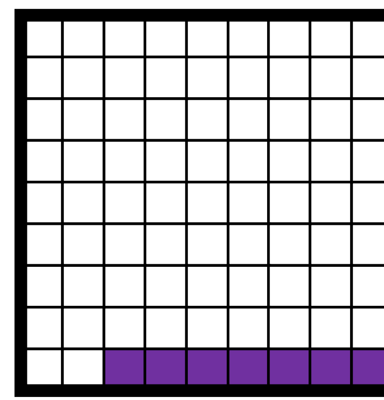
Cases where evidence e **is true**
P(e) = P(female = true)



Cases where e **true given** H **true**
P(e | H)=P(female = true | grad = true)



Cases where e **true given** H **false**
P(e |¬H)=P(female = true | grad = false)



Bayes' Rule: Visual Interpretation

Given (made up roster data):

%of G students: $P(H)$

%of UG students: $P(\neg H)$

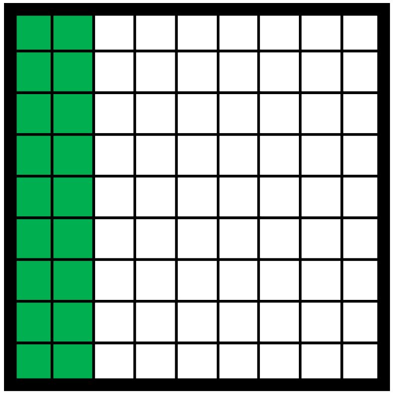
%of female students: $P(e)$

%of female G students: $P(e \mid H)$

%of female UG students: $P(e \mid \neg H)$

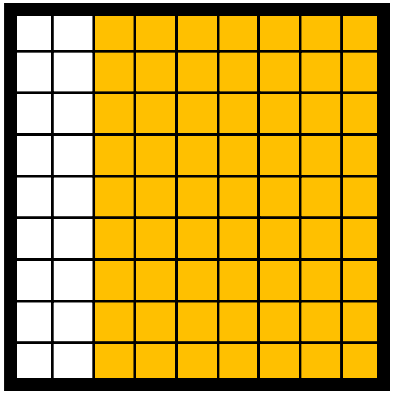
Cases where Hypothesis H is true

$P(H) = 18 / 81$



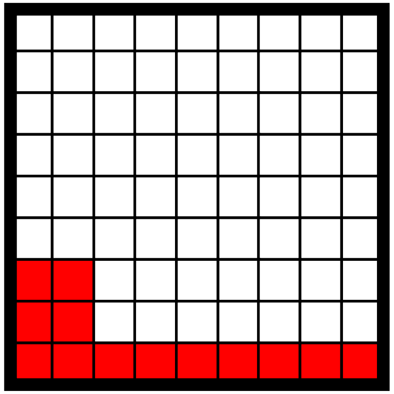
Cases where Hypothesis H is false

$P(\neg H) = 63 / 81$



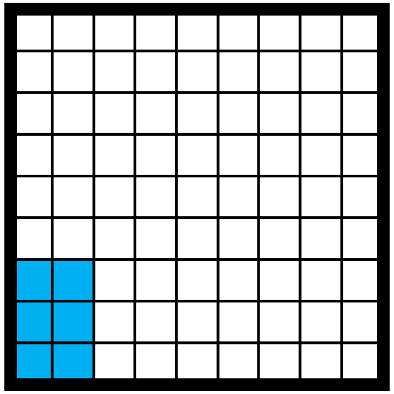
Cases where evidence e is true

$P(e) = 13 / 81$



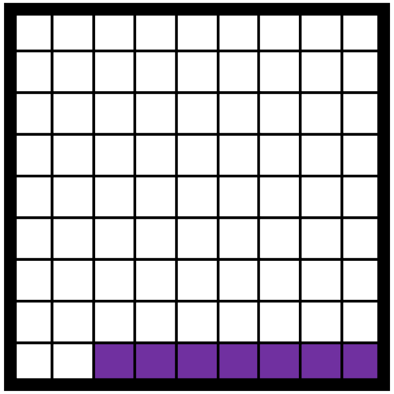
Cases where e true given H true

$P(e \mid H) = 6 / 18$



Cases where e true given H false

$P(e \mid \neg H) = 7 / 18$



Bayes' Rule: Visual Interpretation

Bayes' Rule:

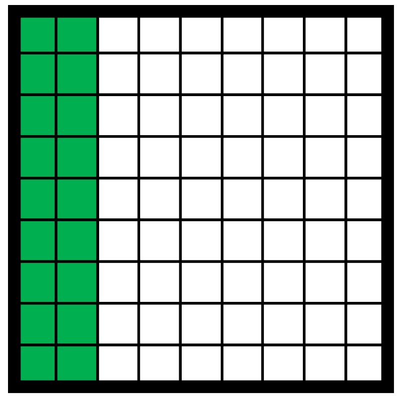
$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(H) * P(e \mid H) + P(\neg H) * P(e \mid \neg H)}$$

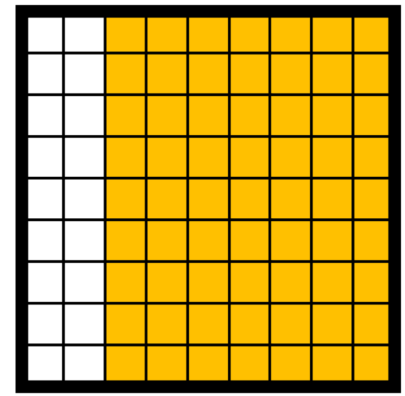
Cases where Hypothesis H is true

P(H) = 18 / 81



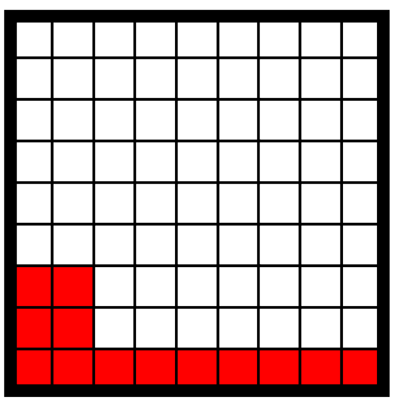
Cases where Hypothesis H is false

P(¬H) = 63 / 81



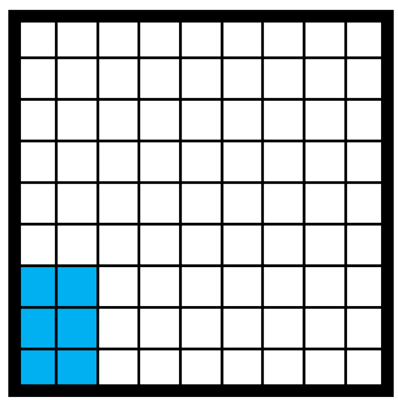
Cases where evidence e is true

P(e) = 13 / 81



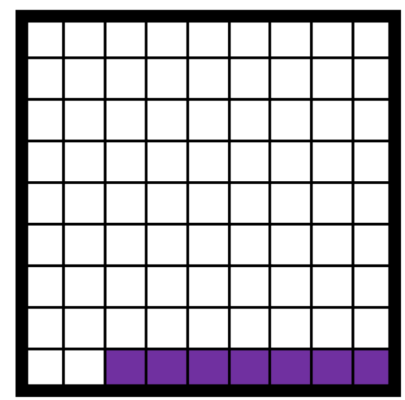
Cases where e true given H true

P(e | H) = 6 / 18



Cases where e true given H false

P(e | ¬H) = 7 / 63



Bayes' Rule: Visual Interpretation

Bayes' Rule:

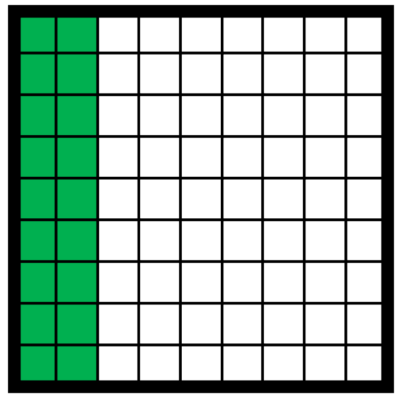
$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

$$P(H \mid e) = \frac{6 / 18 * 18 / 81}{13 / 81}$$

$$P(H \mid e) = \frac{6 / 18 * 18 / 81}{18 / 81 * 6 / 18 + 63 / 81 * 7 / 63}$$

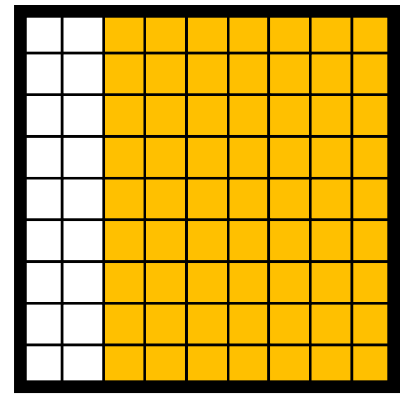
Cases where Hypothesis H **is true**

P(H) = 18 / 81



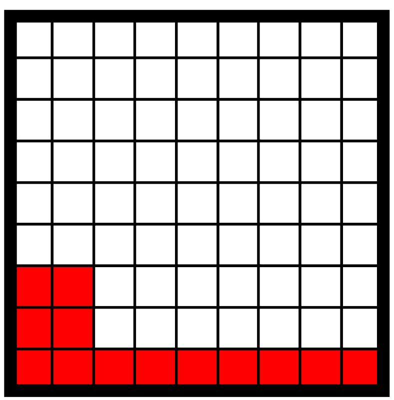
Cases where Hypothesis H **is false**

P(¬H) = 63 / 81



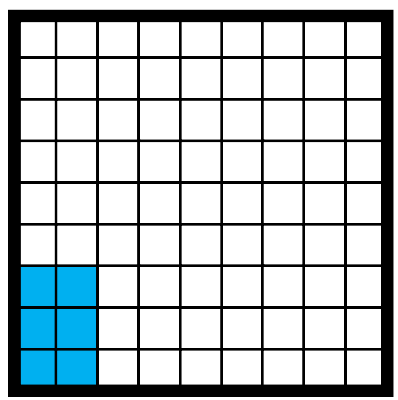
Cases where evidence e **is true**

P(e) = 13 / 81



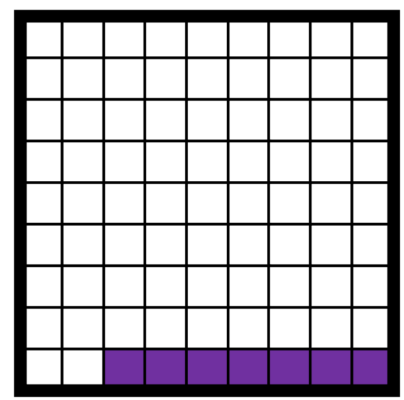
Cases where e **true given** H **true**

P(e | H) = 6 / 18



Cases where e **true given** H **false**

P(e | ¬H) = 7 / 63



Bayes' Rule: Visual Interpretation

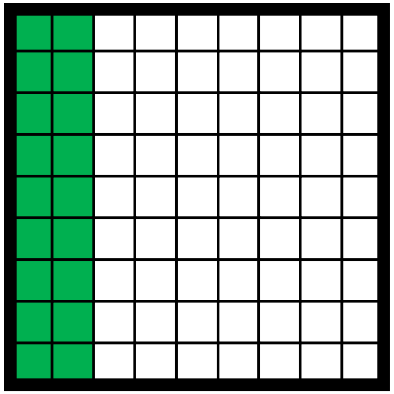
Bayes' Rule:

$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

$$P(H \mid e) \approx 0.462$$

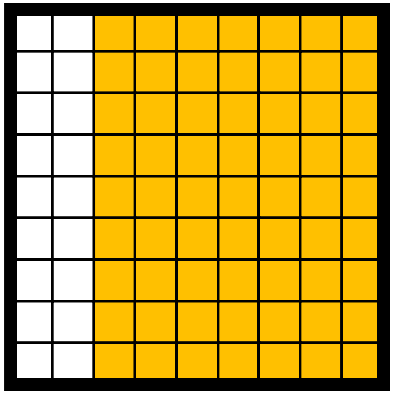
Cases where Hypothesis H is true

P(H) = 18 / 81



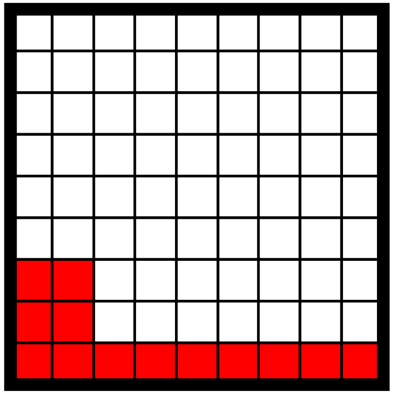
Cases where Hypothesis H is false

P(¬H) = 63 / 81



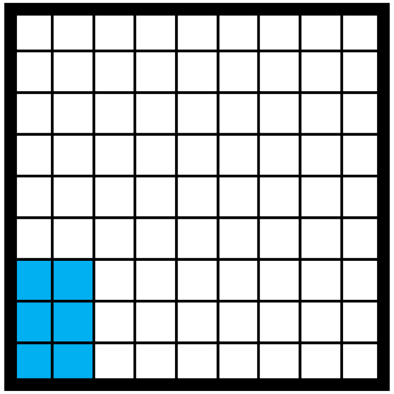
Cases where evidence e is true

P(e) = 13 / 81



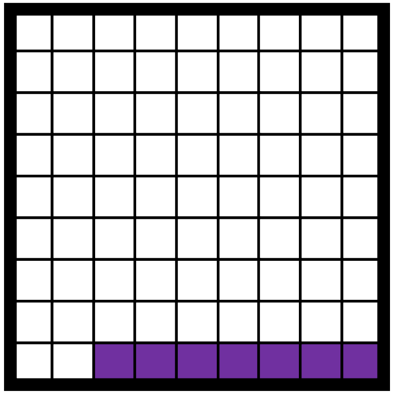
Cases where e true given H true

P(e | H) = 6 / 18



Cases where e true given H false

P(e | ¬H) = 7 / 63



Bayes' Rule: Visual Interpretation

Prior probability:

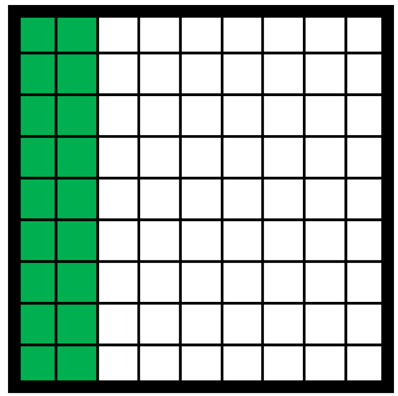
$$P(H) = 18\ /\ 81 \approx 0.222$$

Posterior probability:

$$P(H \mid e) \approx 0.462$$

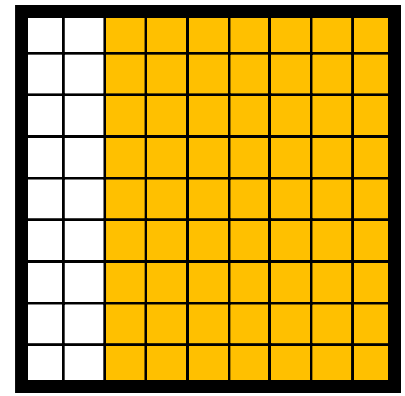
Cases where Hypothesis H **is true**

P(H) = 18 / 81



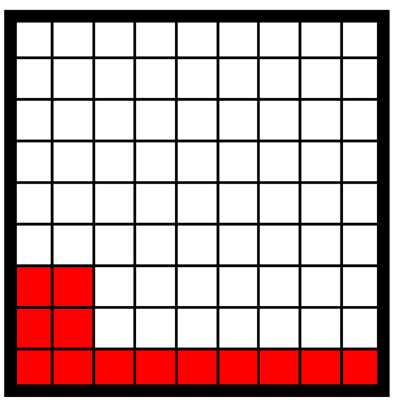
Cases where Hypothesis H **is false**

P(¬H) = 63 / 81



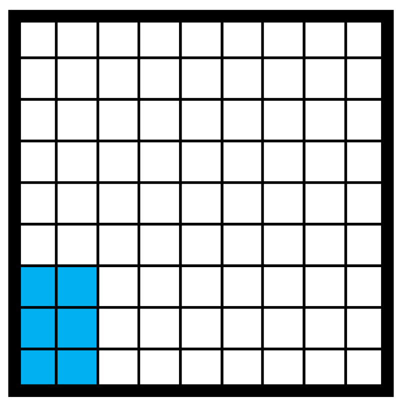
Cases where evidence e **is true**

P(e) = 13 / 81



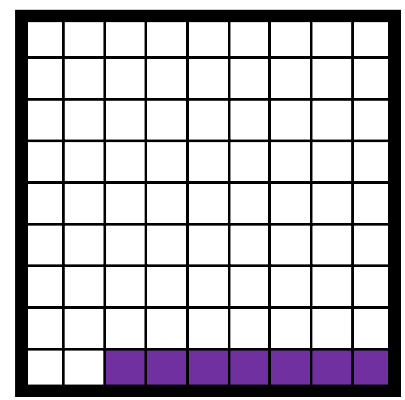
Cases where e **true given** H **true**

P(e | H) = 6 / 18



Cases where e **true given** H **false**

P(e | ¬H) = 7 / 63



Bayes' Rule: Belief/Probability Update

A student approaches the podium. Without looking I create a hypothesis H:

this is a grad student (grad = true)

My belief in H being true is based on prior probability:

$$P(H) = 18\,/\,81 \approx 0.222$$

I look up and see a female student, which is new data / evidence e (female = true). Bayes' Rule helps me update my belief in H being true with posterior probability:

$$P(H \mid e) = \frac{6\,/\,18 * 18\,/\,81}{18\,/\,81 * 6\,/\,18 + 63\,/\,81 * 7\,/\,63} \approx 0.462$$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)	Conditional probabilities
true	true	$P(H \mid e)*P(e) \approx 0.074$	$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)} = \frac{6/18 * 18/81}{13/81} \approx 0.462$
true	false	$P(H \mid \neg e)*P(\neg e) \approx 0.148$	$P(H \mid \neg e) = \frac{P(\neg e \mid H) * P(H)}{P(\neg e)} = \frac{12/18 * 18/81}{68/81} \approx 0.176$
false	true	$P(\neg H \mid e)*P(e) \approx 0.086$	$P(\neg H \mid e) = \frac{P(e \mid \neg H) * P(\neg H)}{P(e)} = \frac{7/63 * 63/81}{13/81} \approx 0.538$
false	false	$P(\neg H \mid \neg e)*P(\neg e) \approx 0.691$	$P(\neg H \mid \neg e) = \frac{P(\neg e \mid \neg H) * P(\neg H)}{P(\neg e)} = \frac{56/63 * 63/81}{68/81} \approx 0.824$
		SUM = 1	

Joint probabilities calculated using the Product Rule:

$$P(A \wedge B) = P(A \mid B) * P(B)$$

Conditional probabilities calculated using Bayes' Rule:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(grad = true \wedge female = true) = P(H, e) = P(H \wedge e) = P(H \mid e) * P(e) \approx 0.074$
true	false	$P(grad = true \wedge female = false) = P(H, \neg e) = P(H \mid \neg e) * P(\neg e) \approx 0.148$
false	true	$P(grad = false \wedge female = true) = P(\neg H, e) = P(\neg H \mid e) * P(e) \approx 0.086$
false	false	$P(grad = false \wedge female = false) = P(\neg H, \neg e) = P(\neg H \mid \neg e) * P(\neg e) \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Product Rule:

$$P(A \wedge B) = P(A \mid B) * P(B)$$

Conditional probabilities calculated using Bayes' Rule:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Joint Probability Distribution

H: grad	e: female	$P(H, e) = P(H \wedge e)$: $P(grad \wedge female)$
true	true	**0.074**
true	false	**0.148**
false	true	**0.086**
false	false	**0.691**
		SUM = 1

If we know the joint probability distribution, we can infer:

- **marginal probabilities** $P(H)$, $P(\neg H)$, $P(e)$, **and** $P(\neg e)$
- **conditional probabilities** $P(H \mid e)$, $P(H \mid \neg e)$, $P(\neg H \mid e)$, **and** $P(\neg H \mid \neg e)$

Joint Probability: Marginalization

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	0.074
true	false	0.148
false	true	0.086
false	false	0.691
		SUM = 1

Probability P(H)**:**

$$P(H) = P(grad = true) = 0.074 + 0.148 \approx 18 / 81$$

Probability P(H)**: "sum of all probabilities where** H **true"**

Joint Probability: Marginalization

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	**0.074**
true	false	**0.148**
false	true	**0.086**
false	false	**0.691**
		SUM = 1

Probability P(e)**:**

$$P(e) = P(female = true) = 0.074 + 0.086 \approx 13 / 81$$

Probability P(e)**: "sum of all probabilities where** e **true"**

Marginal Probability

Marginal probability: the probability of an event occurring $P(A)$.

It may be thought of as an unconditional probability.

It is not conditioned on another event.

Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	**0.074**
true	false	**0.148**
false	true	**0.086**
false	false	**0.691**
		SUM = 1

Joint probabilities calculated using the Product Rule:

$$P(A \wedge B) \ = \ P(A \mid B) * \ P(B)$$

Conditional probabilities calculated using Bayes' Rule:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Joint Probability: Conditionals

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	**0.074**
true	false	**0.148**
false	true	**0.086**
false	false	**0.691**
		SUM = 1

From product rule:

$$P(H \wedge e) = P(H \mid e) * P(e)$$

we can derive:

$$P(H \mid e) = \frac{P(H \wedge e)}{P(e)}$$

Joint Probability: Conditionals

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	**0.074**
true	false	**0.148**
false	true	**0.086**
false	false	**0.691**
		SUM = 1

From product rule:

$$P(H \wedge e) = P(H \mid e) * P(e)$$

we can derive:

$$P(H \mid e) = \frac{P(H \wedge e)}{P(e)} = \frac{0.074}{0.074 + 0.086} \approx 0.462$$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	**0.074**
true	false	**0.148**
false	true	**0.086**
false	false	**0.691**
		SUM = 1

Joint probabilities calculated using the Product Rule:

$$P(A \wedge B) \; = \; P(A \mid B) * \; P(B)$$

Conditional probabilities calculated using Bayes' Rule:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)}$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Random variables:

Toothache **- Boolean**

Cavity **- Boolean**

Catch **(dentist's probe catches tooth) - Boolean**

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Probability P(Cavity ∨ Toothache)**:**

$$P(Cavity = true \vee Toothache = true) =$$
$$= 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064$$
$$= 0.28$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Marginal probability P(Cavity)**:**

$$P(Cavity = \ true) = 0.108 + 0.012 + 0.072 + 0.008 = 0.2$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Conditional probability P(Cavity | Toothache)**:**

$$P(Cavity = true \mid Toothache = true) =$$

$$= \frac{P(Cavity = true \wedge Toothache = true)}{P(Toothache = true)} =$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Conditional probability P(¬Cavity | Toothache):

$$P(\neg Cavity = true \mid Toothache = true) =$$

$$= \frac{P(\neg Cavity = true \wedge Toothache = true)}{P(Toothache = true)} =$$

$$= \frac{0.016 + 0.164}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Note that:

$$P(Cavity \mid Toothache) = \frac{P(Cavity \wedge Toothache)}{P(Toothache)} = 0.6$$

$$P(\neg Cavity \mid Toothache) = \frac{P(\neg Cavity \wedge Toothache)}{P(Toothache)} = 0.4$$

add up to 1 and the same denominator is involved.

Full Joint Probability Distribution

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Note that $\mathrm{P}()$ is the distribution, NOT individual probability:

$$\boldsymbol{P}(Cavity \mid Toothache) = \alpha * \boldsymbol{P}(Cavity, Toothache) =$$

$$= \alpha * [\boldsymbol{P}(Cavity, Toothache, Catch) + \boldsymbol{P}(Cavity, Toothache, \neg Catch)] =$$

$$= \alpha * [\langle 0.108, 0.016 \rangle + \langle 0.012, 0.064 \rangle] =$$

$$= \alpha * \langle 0.12, 0.08 \rangle =$$

$$= \langle 0.6, 0.4 \rangle$$

General Inference Procedure

Given:

- **a query involving a single variable X (in our example: Cavity),**
- **a <u>list</u> of evidence variables $\mathbf{E}$ (in our example: just Toothache),**
- **a <u>list</u> of observed values e for $\mathbf{E}$,**
- **a list of remaining unobserved variables $\mathbf{Y}$ (in our example: just Catch),**

where X, $\mathbf{E}$, and $\mathbf{Y}$ together are a COMPLETE set of variables for the domain, the probability $P(X \mid \mathbf{E})$ can be evaluated as:

$$\boldsymbol{P}(X \mid \boldsymbol{e}) = \alpha * \boldsymbol{P}(X, \boldsymbol{e}) = \alpha * \sum_{\boldsymbol{y}} \boldsymbol{P}(X, \boldsymbol{e}, \boldsymbol{y})$$

where ys are all possible values for $\mathbf{Y}$s, α - normalization constant.

$\mathbf{P}(X, \mathbf{e}, \mathbf{y})$ is a subset of probabilities from the joint distribution

Complex Joint Distributions

Consider a complex joint probability distribution involving N random variables P_1, P_2, P_3, ..., P_{N-1}, Pp_N .

2^N Possible Worlds (Models)	N Random Variables						Joint Probability	
	P_1	P_2	P_3	...	P_{N-1}	P_N		2^N values
	true	true	true	...	true	true	false	
	true	true	true	...	true	false	true	
	true	true	false	...	false	true	false	
	...	...	...	...	...	...	...	
	false	false	true	...	true	false	true	
	false	false	true	...	false	true	true	
	false	false	false	...	false	false	false	

Non-binary / Non-Boolean RVs

Some Random Variables are going to have more than two possible, discrete, values:

- height **-> short, average, tall**
- size **-> S, M, L, XL**
- streetLight **-> green, orange, red**
- vision **-> 20/20, 15/15, etc.**
- continent **-> Africa, Antarctica, Asia, Australia, Europe, North America, South America**

Non-binary RVs increase the complexity.

This May Be Impossible to Manage!

	N Random Variables						Joint Probability	
	P_1	P_2	P_3	...	P_{N-1}	P_N		
2^N Possible Worlds (Models)	true	true	true	...	true	true	false	2^N values
	true	true	true	...	true	false	true	
	true	true	false	...	false	true	false	
	...	...	...	...	...	...	...	
	false	false	true	...	true	false	true	
	false	false	true	...	false	true	true	
	false	false	false	...	false	false	false	

Independent Variable

		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

Let's introduce another random variable Cloudy **representing some weather conditions. It is difficult to imagine the other random variables here (**Toothache**,** Cavity**,** Catch**) being dependent on** Cloudy **and vice versa.**

Independent Variable

		Toothache		¬Toothache	
¬Cloudy		Catch	¬Catch	Catch	¬Catch
	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

		Toothache		¬Toothache	
Cloudy		Catch	¬Catch	Catch	¬Catch
	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

Let's try to calculate the following probability:

$$P(Toothache, Catch, Cavity, Cloudy)$$

using the Product Rule:

$$P(Toothache, Catch, Cavity, Cloudy) =$$
$$= P(Cloudy \mid Toothache, Catch, Cavity) * P(Toothache, Catch, Cavity)$$

Independent Variable

		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

It's hard to imagine Cloudy influencing other variables, so:

$$P(Cloudy \mid Toothache, Catch, Cavity) = P(Cloudy)$$

and then:

$$P(Toothache, Catch, Cavity, Cloudy) =$$
$$= P(Cloudy) * P(Toothache, Catch, Cavity)$$

Independent Variable / Factoring

¬Cloudy		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

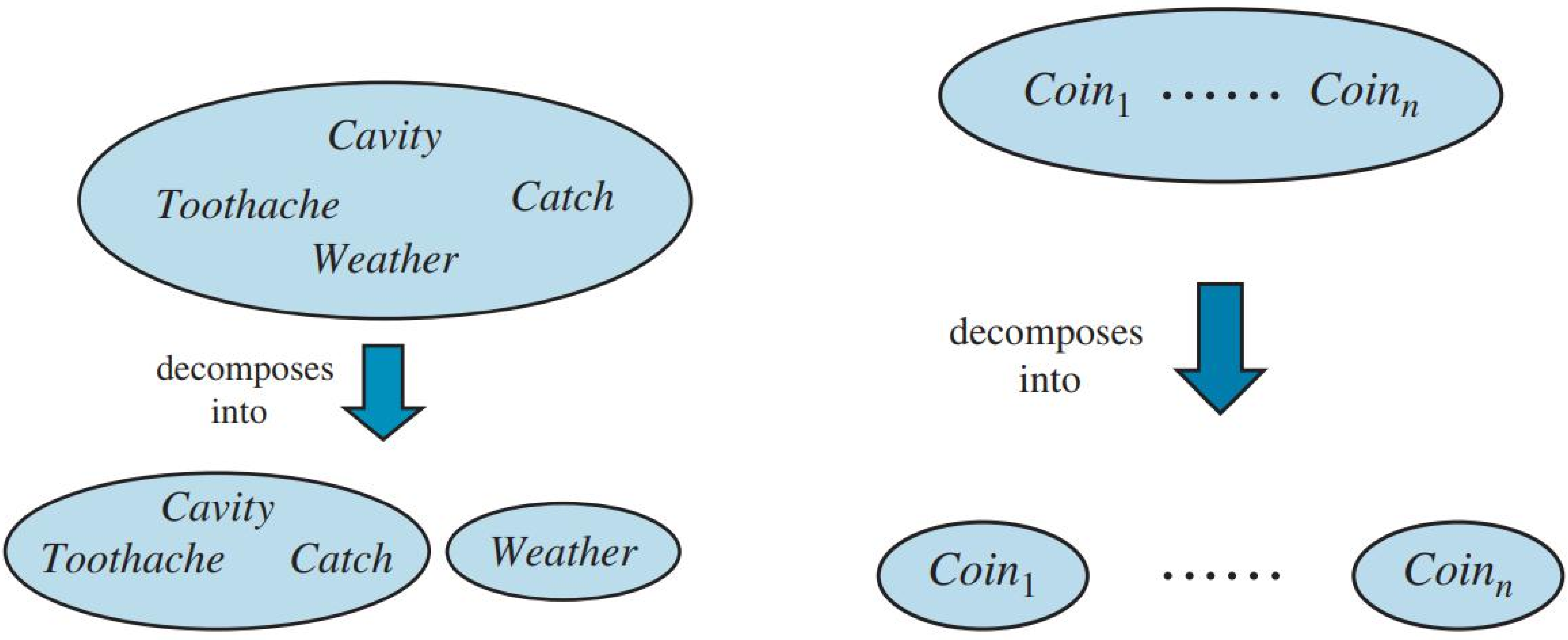
Cloudy		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

It's hard to imagine Cloudy influencing other variables, so:

$$P(Toothache, Catch, Cavity, Cloudy) =$$
$$= P(Cloudy) * P(Toothache, Catch, Cavity)$$

This shows that Cloudy **is INDEPENDENT of other variables and factoring can be applied.**

Factoring / Decomposition



Use Chain Rule To Decompose

N Random Variables						Joint Probability
P_1	P_2	P_3	…	P_{N-1}	P_N	
true	true	true	…	true	true	false
true	true	true	…	true	false	true
true	true	false	…	false	true	false
…	…	…	…	…	…	…
false	false	true	…	true	false	true
false	false	true	…	false	true	true
false	false	false	…	false	false	false

Chain Rule

Conditional probabilities can be used to decompose conjunctions using the chain rule. For any propositions $f_1, f_2, \ldots, f_n$:

$$P(f_1 \wedge f_2 \wedge \ldots \wedge f_n) = \prod_{i=1}^{n} P(f_i \mid f_1 \wedge \ldots \wedge f_{i-1})$$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(grad = true \wedge female = true) = P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) \approx 0.074$
true	false	$P(grad = true \wedge female = false) = P(H, \neg e) = P(H) * P(\neg e \mid H) \approx 0.148$
false	true	$P(grad = false \wedge female = true) = P(\neg H, e) = P(\neg H) * P(e \mid \neg H) \approx 0.086$
false	false	$P(grad = false \wedge female = false) = P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Chain Rule:

$$P(f_1 \wedge f_2) = \prod_{i=1}^{2} P(f_i \mid f_1 \wedge \ldots \wedge f_{i-1})$$

$$P(f_1 \wedge f_2) = P(f_1) * P(f_2 \mid f_1)$$

so: $P(grad \wedge female) = P(H \wedge e) = P(H) * P(e \mid H)$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Chain Rule:

$$P(f_1 \wedge f_2) = \prod_{i=1}^{2} P(f_i \mid f_1 \wedge \ldots \wedge f_{i-1})$$

$$P(f_1 \wedge f_2) = P(f_1) * P(f_2 \mid f_1)$$

so: $P(grad \wedge female) = P(H \wedge e) = P(H) * P(e \mid H)$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18/81 * 6/18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18/81 * 12/18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63/81 * 7/63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63/81 * 56/63 \approx 0.691$
		SUM = 1

Joint probabilities calculated using the Chain Rule:

$$P(f_1 \wedge f_2) = \prod_{i=1}^{2} P(f_i \mid parents(f_i))$$

$$P(f_1 \wedge f_2) = P(f_1) * P(f_2 \mid parents(f_i))$$

so: $P(H \wedge e) = P(H) * P(e \mid parents(e)) = P(H) * P(e \mid H)$

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Conditional Probability Table (CPT)

Bayesian (Belief) Network

A Bayesian belief network describes the joint probability distribution for a set of variables.

A Bayesian network is an acyclic, directed graph (DAG), where the nodes are random variables (propositions). There is an edge (arc) from each elements of $\mathrm{parents}(X_i)$ into X_i. Associated with the Bayesian network is a set of conditional probability distributions - the conditional probability of each variable given its parents (which includes the prior probabilities of those variables with no parents).

Consists of:

- **a graph (DAG) with nodes corresponding to random variables**
- **a domain for each random variable**
- **a set of conditional probability distributions $P(X_i \mid \mathrm{parents}(X_i))$**

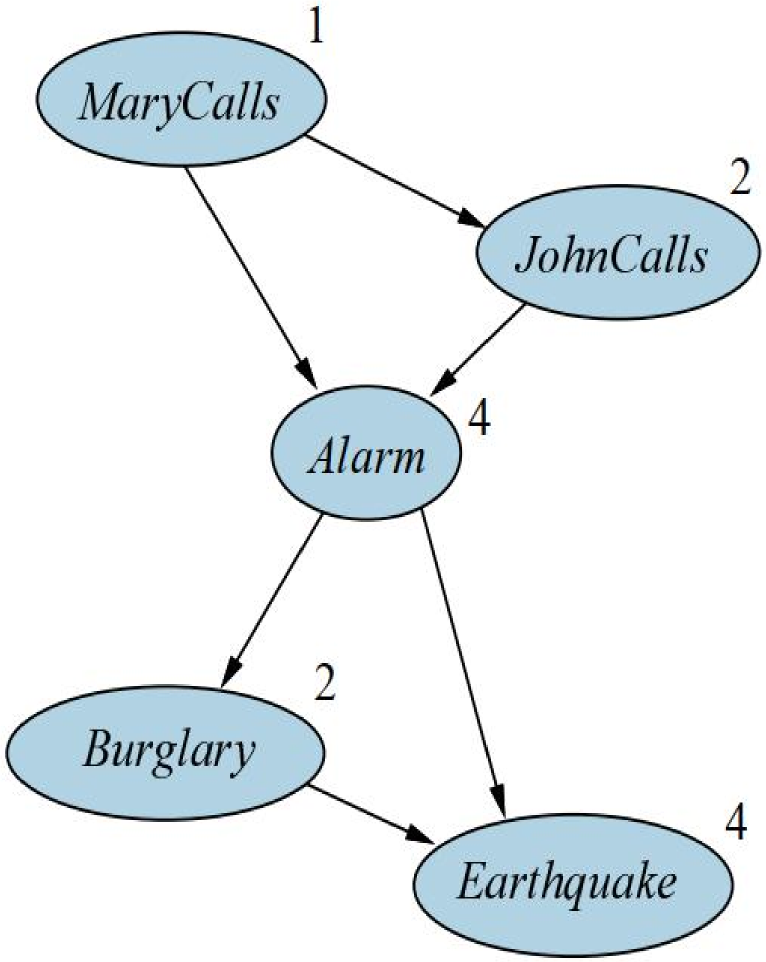
Building Bayesian (Belief) Network

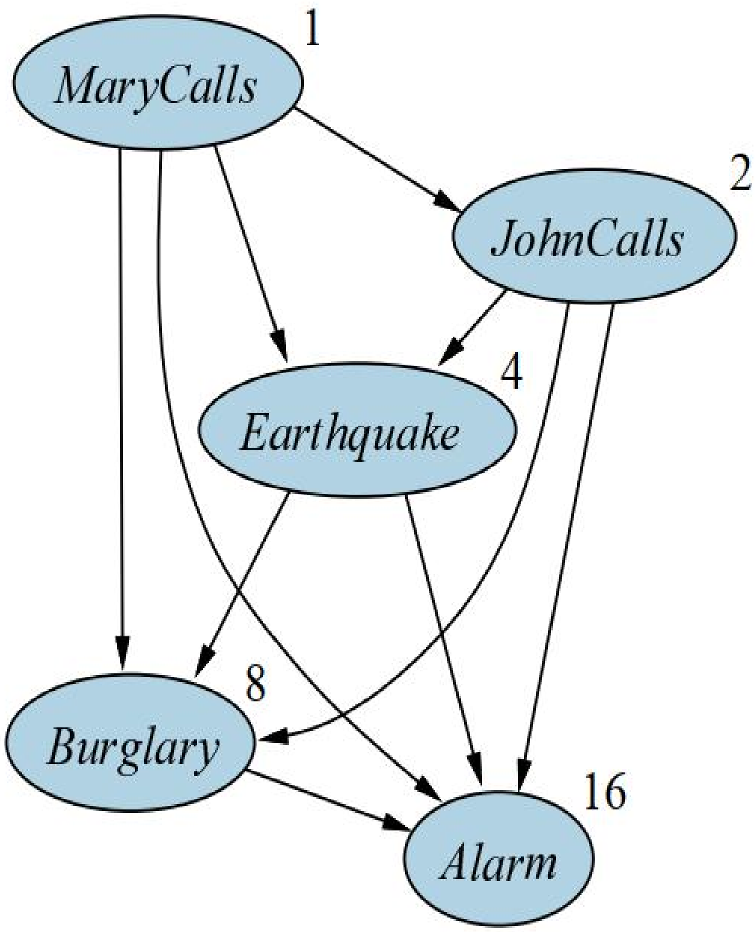
1. **Order Random Variables (ordering matters!)**
2. **Create network nodes for each Random Variable**
3. **Add edges between parent nodes and children nodes**
   - **For every node node X_i:**
     - **choose a minimal set S of parents for X_i**
     - **for each parent node Y in S add an edge from Y to X_i**

4. **Add Conditional Probability Tables**

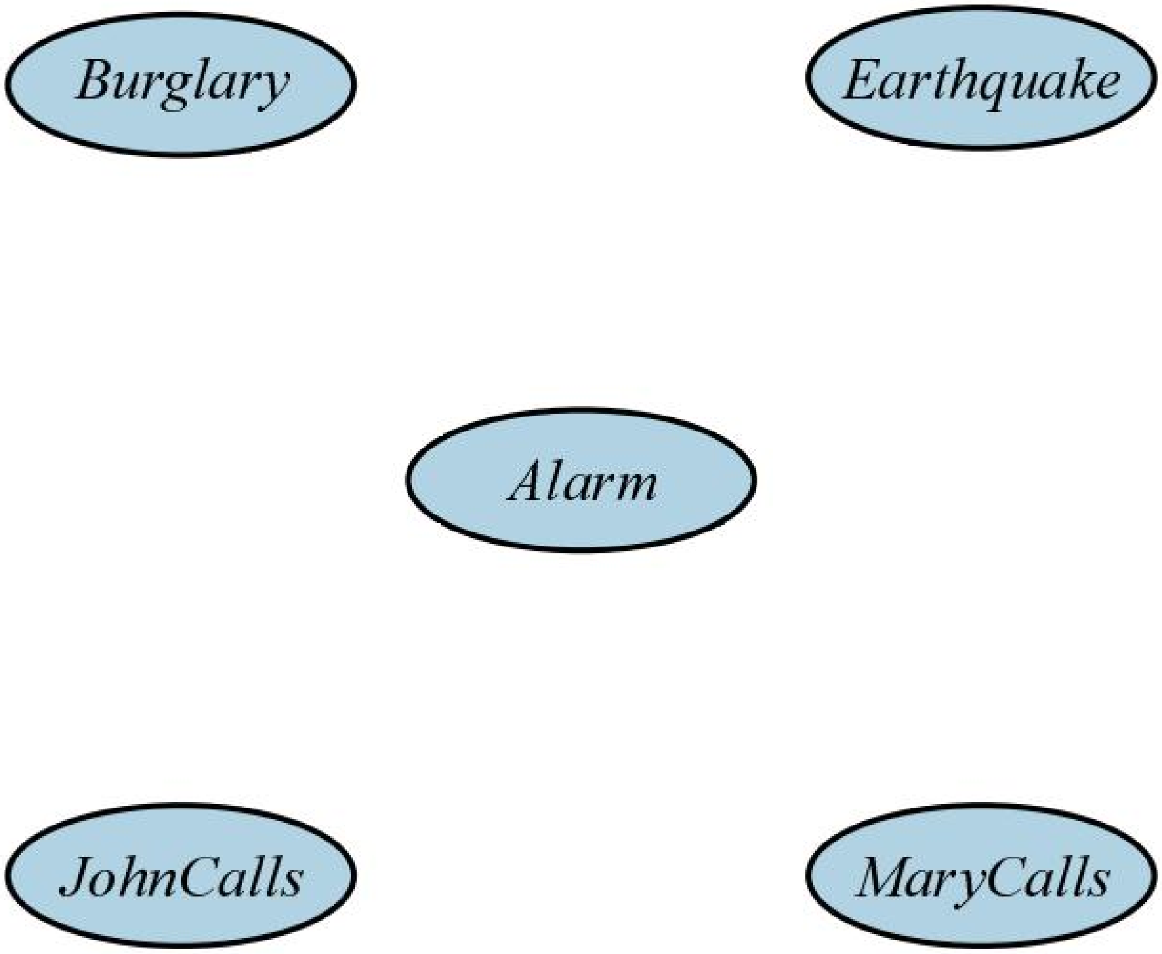
Make it compact / sparse: choose your Random Variable ordering wisely.

Ordering Matters!

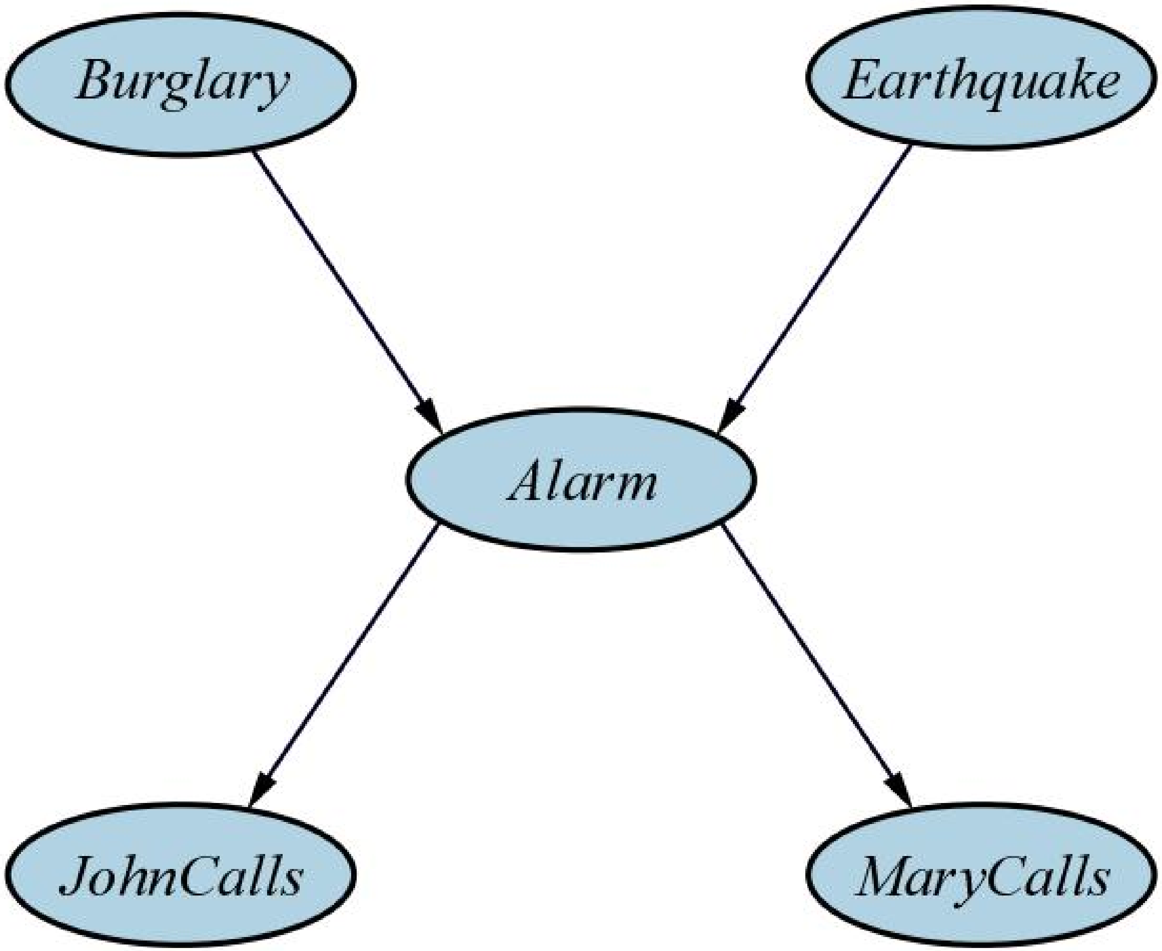




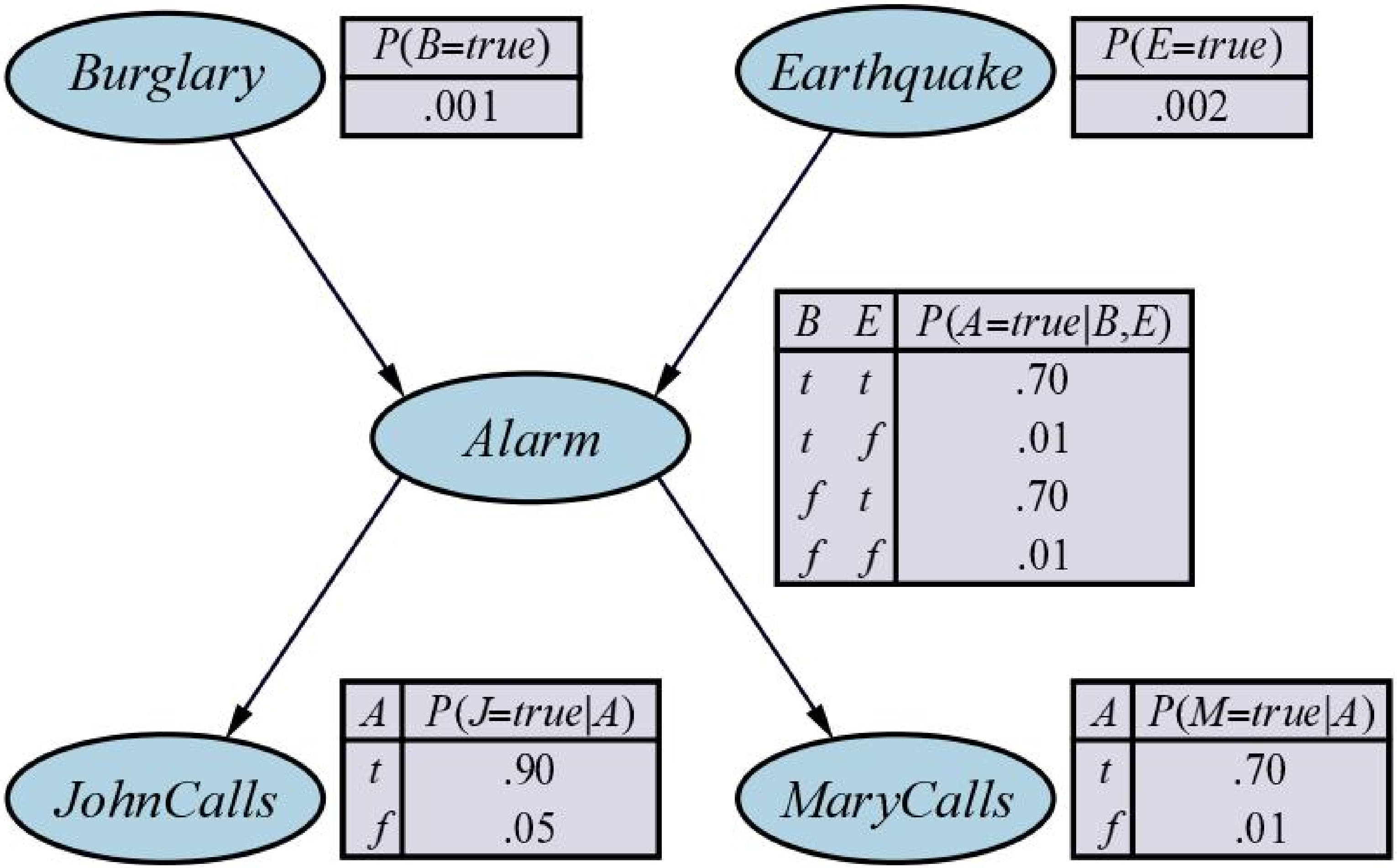
Create Vertices / Node / Random Vars



Add Edges



Add Conditional Probability Tables



P(B=true)
.001

P(E=true)
.002

B	E	P(A=true|B,E)
t	t	.70
t	f	.01
f	t	.70
f	f	.01

A	P(J=true|A)
t	.90
f	.05

A	P(M=true|A)
t	.70
f	.01

Full Joint Probability Distribution

H: grad	e: female	P(H, e) = P(H ∧ e): P(grad ∧ female)
true	true	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \approx 0.074$
true	false	$P(H, \neg e) = P(H) * P(\neg e \mid H) = 18 / 81 * 12 / 18 \approx 0.148$
false	true	$P(\neg H, e) = P(\neg H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \approx 0.086$
false	false	$P(\neg H, \neg e) = P(\neg H) * P(\neg e \mid \neg H) = 63 / 81 * 56 / 63 \approx 0.691$
		SUM = 1

$$P(H \wedge e) = P(H) * P(e \mid parents(e)) = \boxed{P(H)} * \boxed{P(e \mid H)}$$

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Conditional Probability Table (CPT)

Create Vertices / Node / Random Vars

Create Vertices / Node / Random Vars

Add Edges

Add Conditional Probability Tables

H: grad	¬H: ¬grad
18 / 81 ≈ 0.22	63 / 81 ≈ 0.78

H: grad	e: female	P(e | H)
true	true	6 / 18 ≈ 0.333
true	false	12 / 18 ≈ 0.667
false	true	7 / 63 ≈ 0.111
false	false	56 / 63 ≈ 0.889

Bayesian Network: Car Insurance