CS 480

Introduction to Artificial Intelligence

February 22, 2024

Announcements / Reminders

- Please follow the Week 06 To Do List instructions (if you haven't already):
- Quiz #05: due on Sunday (02/25/24) at 11:59 PM CST
  - NO QUIZ next week
- Written Assignment #03: due on Sunday (02/25/24) at 11:59 PM CST
- Midterm Exam: 02/27/2024
  - Section 02 – Make arrangements with Mr. Charles Scott

Midterm Exam: Rules

- Exam will be pen and paper
- No electronic devices allowed
  - including AirPods, earbuds, etc.
  - exception: REGULAR calculator (not a phone app)
  - all your electronic devices need to be hidden from view
- No communication allowed
- Closed book / closed notes
  - you can bring ONE letter-sized double-sided cheat sheet
- NO programming will be involved, however you are expected to understand algorithms to work out solutions by hand
- Material: everything I covered in class without saying “this is not going to be on the exam” is fair game

Plan for Today

- **Probability Refresher**

Joint Probability

The probability of event A and event B occurring (or more than two events). It is the probability of the intersection of two or more events.

$$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}, \mathrm{B}) = P(A \text{ and } B) = P(A \wedge B)$$

Random Variable

A Random Variable is a mathematical formalization of a quantity or object which depends on random events

A Random Variable X is a function mapping events/outcomes from the sample space S to a measurable space (such as $\mathbb{R}$) :

$$\mathrm{X}: \mathrm{S} \rightarrow \mathbb{R}$$

Random Variable

Sample Space

Random Variable

1

-1

Probability

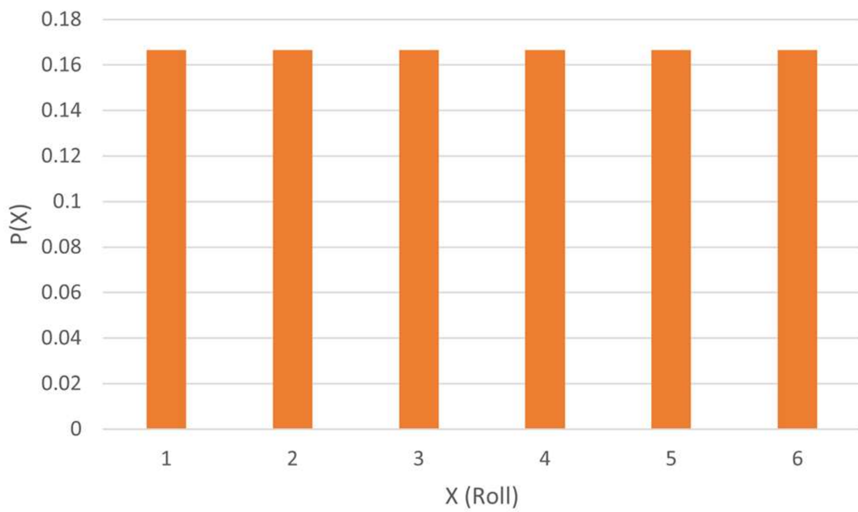
0.5

Random Variable Distribution

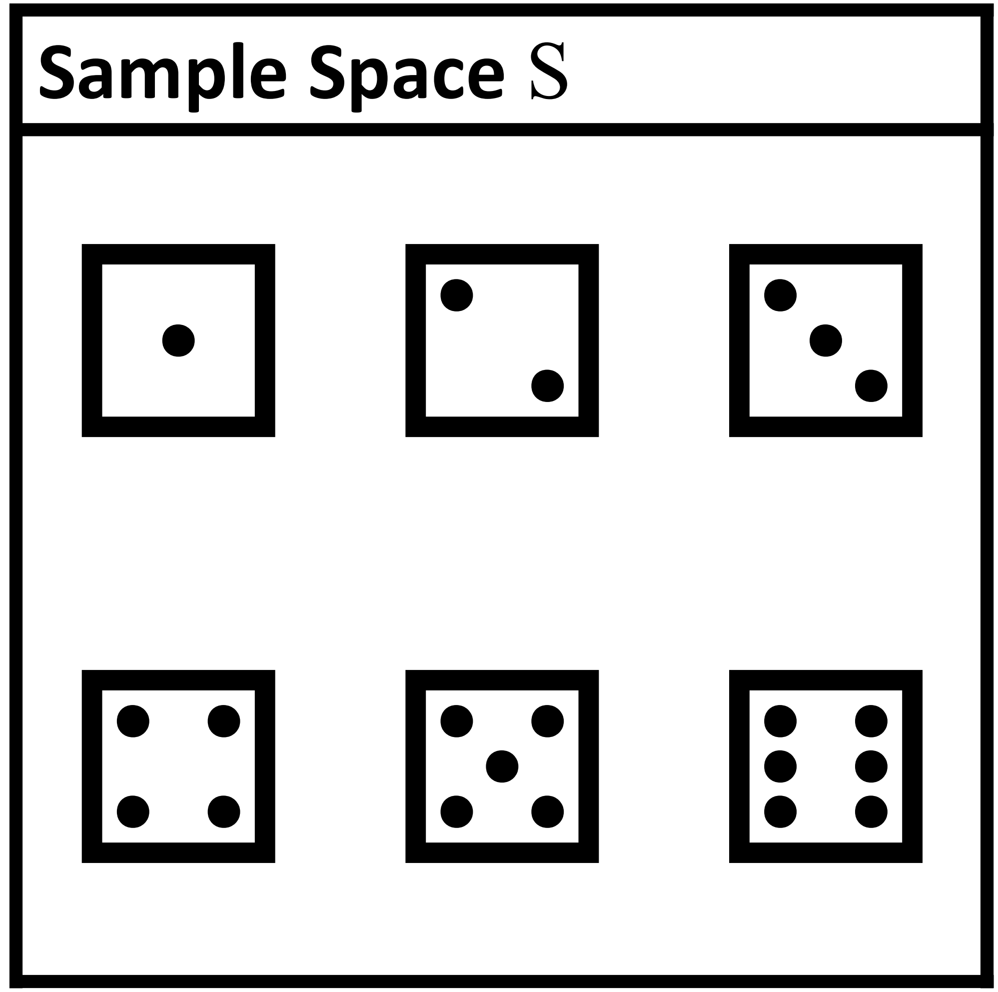
The probability distribution for a discrete random variable X can be perceived as a frequency distributions.

It is a graph, table or formula that gives the possible values of X and the probability $P(X)$ associated with each value of X.

Single Die Roll: Distribution



X	P (X)
⚀	1/6
⚁	1/6
⚂	1/6
⚃	1/6
⚄	1/6
⚅	1/6



Random Variable: Typical Notation

- Capital: X: a variable
- Lowercase: x: a particular value of X
- $Val(X)$: the set of values X can take
- Bold Capital: $\mathbf{X}$: a set of variables
- Bold lowercase: $\mathbf{x}$: an assignment to all variables in $\mathbf{X}$
- $P(X = x)$ will be shortened as $P(x)$
- $P(X = x \cap Y = y)$ will be shortened as $P(x, y)$
- $\mathbf{P}(X)$: probability distribution for X

Random Variable: Typical Notation

- Pick variables of interest/relevance
  - Medical diagnosis
    - Age, gender, weight, temperature, …
  - Loan application
    - Income, savings, payment history, …
  - other
- Every variable has a domain
  - Binary (e.g., True/False)
  - Categorical (e.g., Red/Green/Blue)
  - Real-valued (e.g., 97.8)
- Possible world
  - An assignment to all variables of interest

Joint Probability

The probability of event A and event B occurring (or more than two events). It is the probability of the intersection of two or more events.

$$\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\mathrm{A}, \mathrm{B}) = \mathbf{P}(\mathrm{A} \text{ and } \mathrm{B}) = \mathbf{P}(\mathrm{A} \wedge \mathrm{B})$$

For example (specific probability shown):

$$\mathrm{P}(\text{pressure} = 90, \text{ temperature} = 100, \text{ volume} = 6) = 0.1$$

For any random variables: $f_1, f_2, \ldots, f_n$ **:**

$$\boldsymbol{P}(f_1, f_2, \ldots, f_n)$$

Probability Model

A fully specified probability model associates a numerical probability $\mathrm{P}(\omega)$ with each possible world (assume there is a finite number of such worlds):

$$0 \leq \mathrm{P}(\omega) \leq 1, \text{for every} \sum_{\omega \in S} \mathrm{P}(\omega) = 1$$

Joint Probability

The probability of event A and event B occurring (or more than two events). It is the probability of the intersection of two or more events.

$$\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\mathrm{A}, \mathrm{B}) = \mathbf{P}(\mathrm{A} \text{ and } \mathrm{B}) = \mathbf{P}(\mathrm{A} \wedge \mathrm{B})$$

For example (specific probability shown):

ONE POSSIBLE "WORLD":

$$\mathrm{P}(\text{pressure} = 90, \text{temperature} = 100, \text{volume} = 6) = 0.1$$

For any random variables: $f_1, f_2, \ldots, f_n$ **:**

$$\boldsymbol{P}(f_1, f_2, \ldots, f_n)$$

Random Variables, Events, Logic

An **event** is the set of possible worlds where a given predicate is true

- Roll two dice
  - The possible worlds are (1,1), (1,2), …, (6,6); 36 possible worlds
  - Predicate = two dice sum to 10
  - Event = {(4,6), (5,5), (6,4)}
- Toothache and cavity
  - Four possible worlds: $(t, c), (t, \sim c), (\sim t, c), (\sim t, \sim c)$
  - Some worlds are more likely than others
  - Predicate can be anything about these variables: $t \wedge c, t, t \vee \sim c,$

Complex Joint Probability Distribution

Consider a complex joint probability distribution involving N random variables $f_1, f_2, f_3, \ldots,$ f_{N-1}, f_N . [values can be OTHER than true/false and non-binary]

2^N **Possible Worlds (Models)**

N Random Variables						Joint Probability
f_1	f_2	f_3	…	f_{N-1}	f_N	
true	true	true	…	true	true	0.0011
true	true	true	…	true	false	0.0451
true	true	false	…	false	true	0.1011
…	…	…	…	…	…	…
false	false	true	…	true	false	0.0909
false	false	true	…	false	true	0.0651
false	false	false	…	false	false	0.2021

2^N **values**

Frequentist versus Causal Perspective

- **Frequentist view:**

  Probability represents long-run frequencies of repeatable events.

- **Causal perspective:**

  Probability is a measure of belief.

Prior (Unconditional) Probabilities

Degree of belief that some event A is occurred *in the absence of any other related information* is called unconditional or prior probability (or "prior" for short) $P(A)$.

Examples:

$$P(\text{isRaining})$$

$$P(\text{dieRoll} = 5)$$

$$P(\text{CourseFinalGrade} = \text{'A'})$$

$$P(\text{toothache})$$

Conditioning

Conditioning is a process of revising beliefs based on new evidence e:

- **start by taking all background information (prior probabilities) into account**
- **if new evidence e is acquired, a conditional probability of some proposition A given evidence e can be calculated (posterior probability):** $P(A \mid e)$

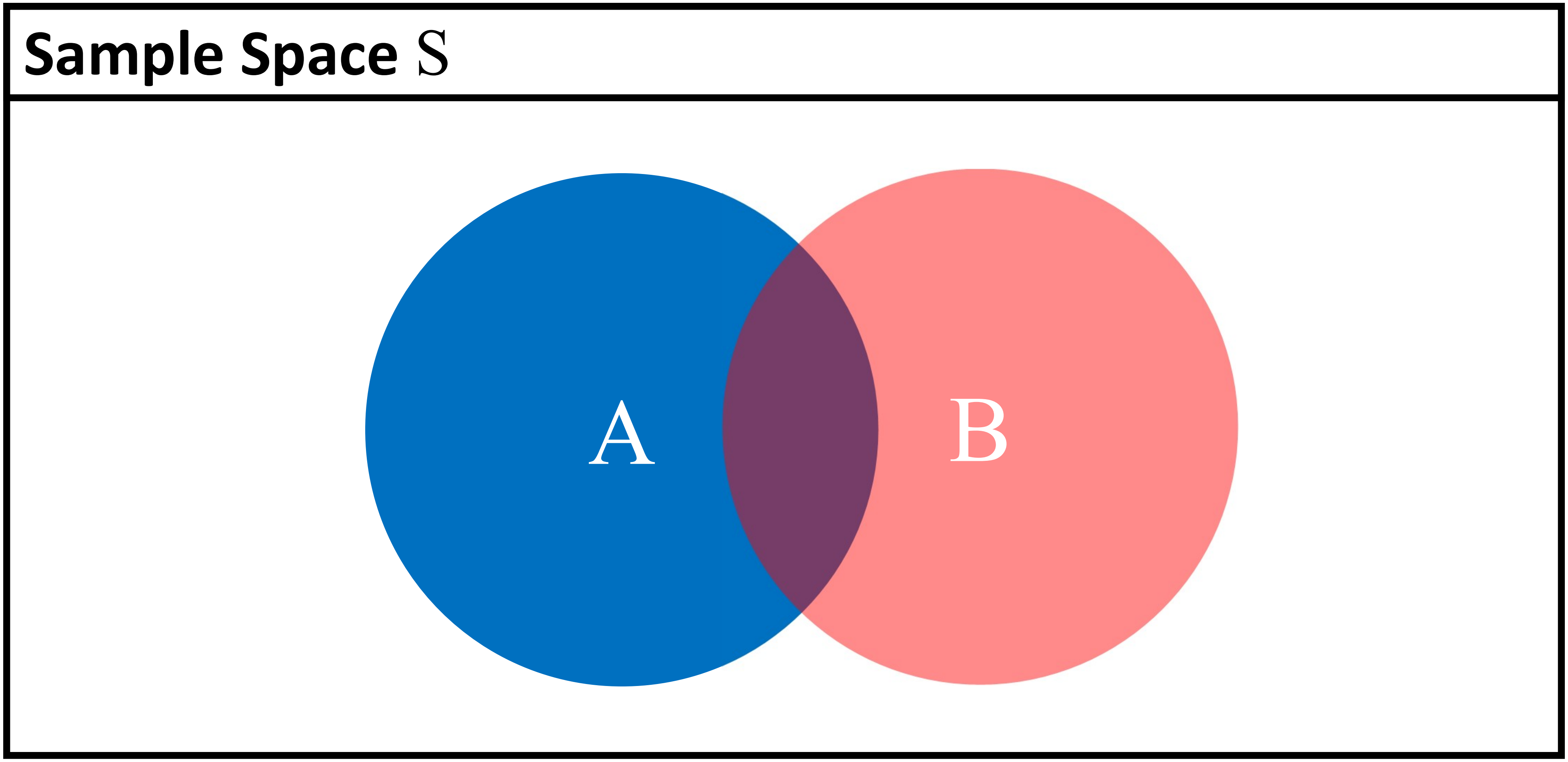
Conditional Probability

If A and B are two events in sample space S, then conditional probability of A given B is defined as:

$$P(A \text{ given } B) = P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

where: $P(B) > 0$

Conditional Probability: Venn Diagram



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A, B)}{P(B)} = \frac{P(A \wedge B)}{P(B)}$$

Conditional Probability

If A and B are two events in sample space S, then conditional probability of A given B is defined as:

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

where: $P(B) > 0$ ← [Otherwise B is impossible]

Conditional Probability: Notation

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

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

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

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

Conditional Probability

If A **and** $\mathrm{evidence}$ **are two events in sample space** S**, then conditional probability of** A **given** $\mathrm{evidence}$ **is defined as:**

$$\mathrm{P(A \mid evidence)} = \frac{\mathrm{P(A \cap evidence)}}{\mathrm{P(evidence)}}$$

where: $\mathrm{P(evidence)} > 0$

Posterior (Conditional) Probabilities

Typically, there is going to be some information, called evidence e, that affects our degree of belief about some event A being occurring. This allows us to also consider conditional or posterior probability (or "posterior" for short) $P(A \mid e)$.

Examples ($P(A \text{ given } e)$):

$$P(\text{isRaining} \mid \text{cloudy})$$

$$P(\text{CourseFinalGrade} = \text{'A'} \mid \text{CoursePA1Score} > 80)$$

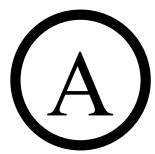
$$P(\text{cavity} \mid \text{toothache})$$

Evidence e

Evidence e rules out possible worlds incompatible with e.

Prior vs. Posterior Probabilities

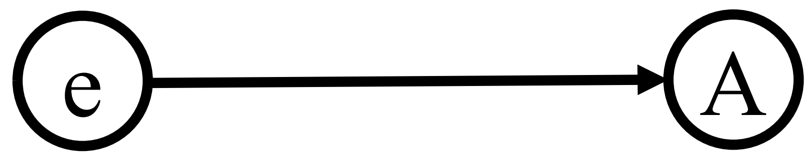
Prior Probability



$$P(A)$$

BTW: it is also $P(A \mid T)$

Posterior Probability



$$P(A \mid e)$$

Conditional Probability: Notation

$$P(A \mid \text{evidence}) = \frac{P(A \cap \text{evidence})}{P(\text{evidence})}$$

$$P(A \mid \text{evidence}) = \frac{P(A \text{ and evidence})}{P(\text{evidence})}$$

$$P(A \mid \text{evidence}) = \frac{P(A, \text{evidence})}{P(\text{evidence})}$$

$$P(A \mid \text{evidence}) = \frac{P(A \wedge \text{evidence})}{P(\text{evidence})}$$

Conditional Probability: Notation

$$P(A, B, C, D \mid E, F, G) = \frac{P(A, B, C, D, E, F, G)}{P(E, F, G)}$$

Axioms of Conditional Probability

Axiom 1:

For any event A**,** $P(A \mid B) \geq 0$

Axiom 2:

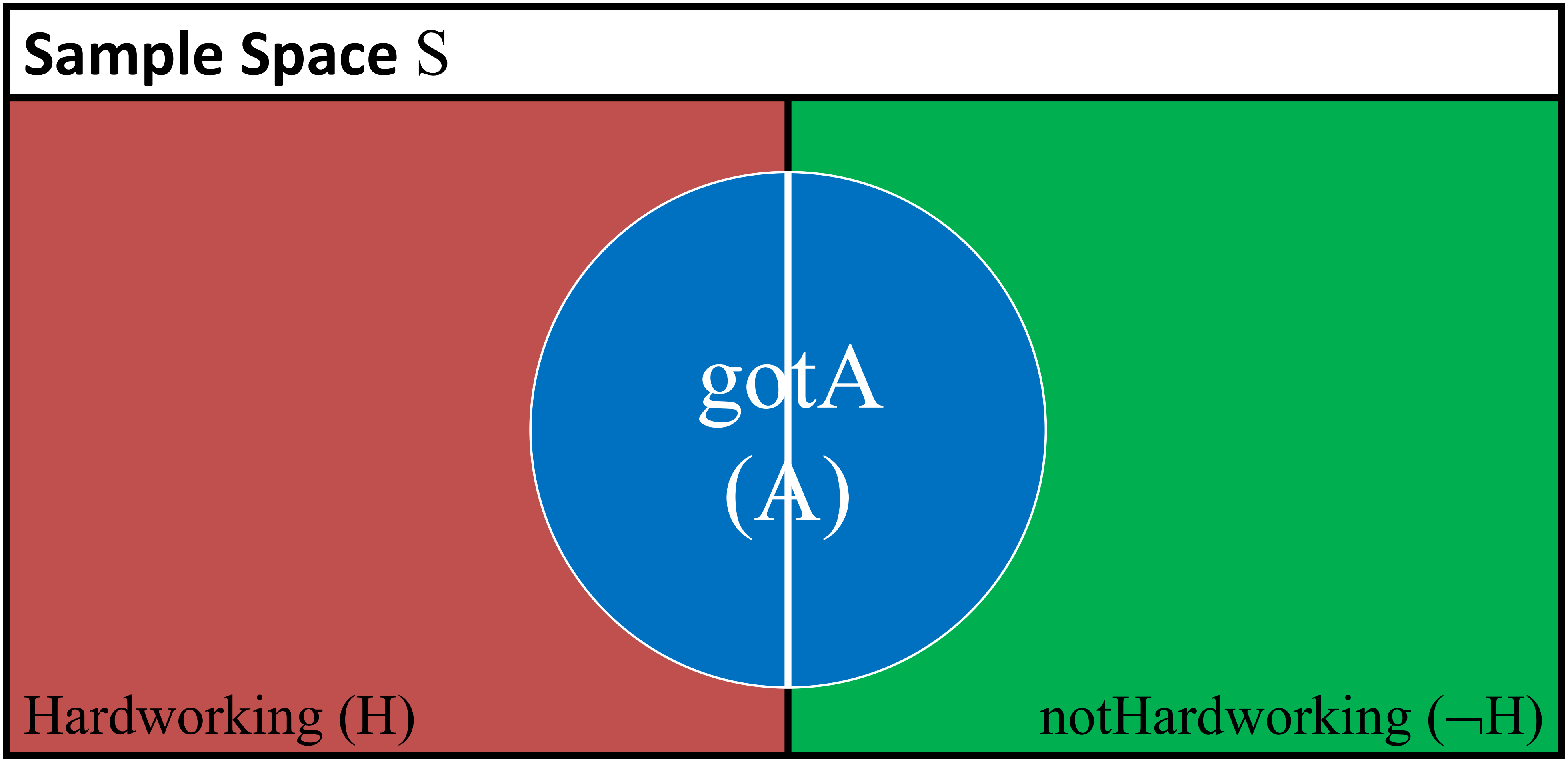
Conditional probability of B **given** B **is** $P(B \mid B) = 1$

Axiom 3:

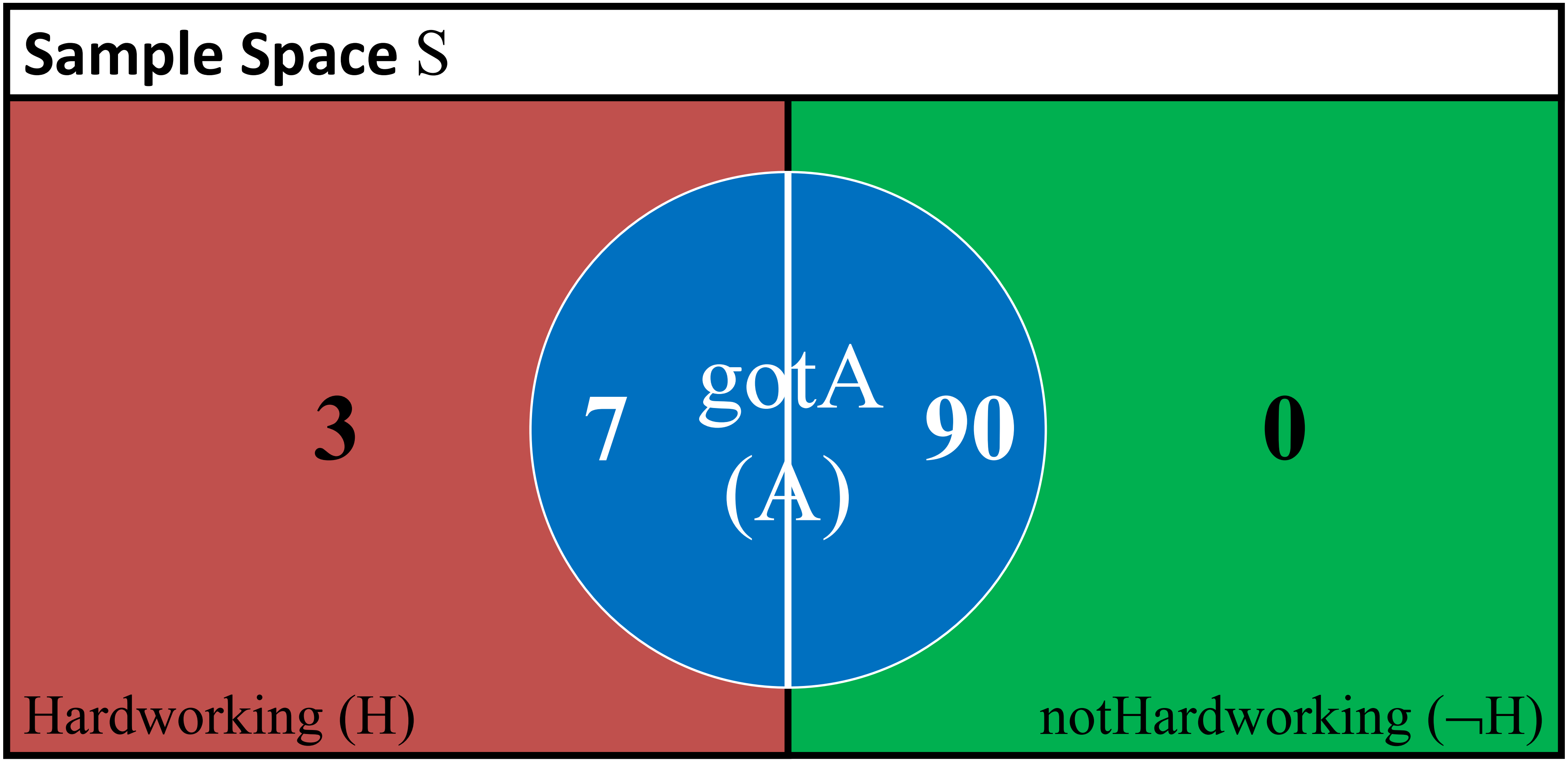
If A_1**,** A_2**, … are disjoint events, then**

$$P(A_1 \cup A_2 \cup \ldots \mid B) = P(A_1 \mid B) + P(A_2 \mid B) + \ldots$$

Conditional Probability: Visualization

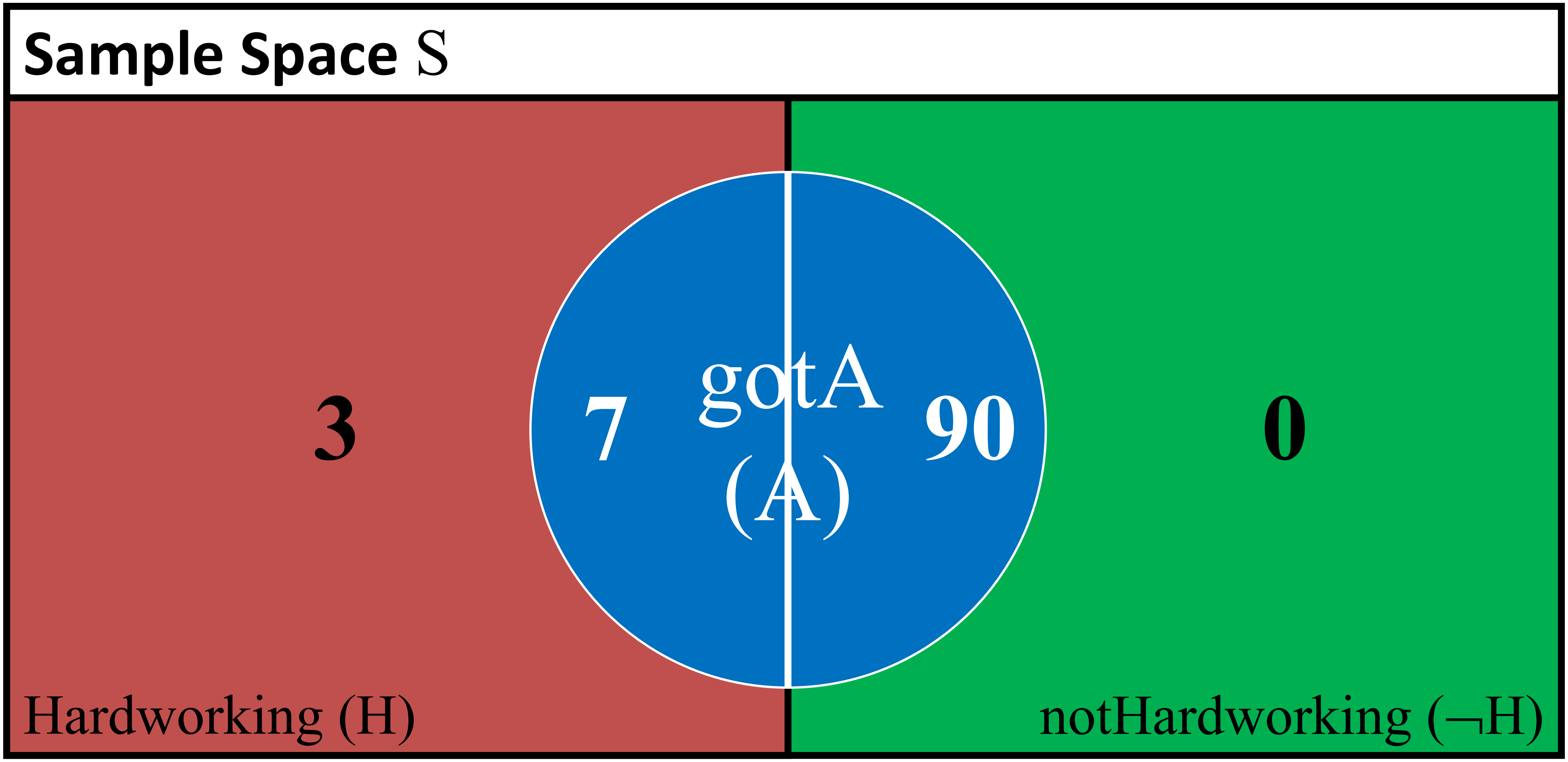


Conditional Probability: Visualization



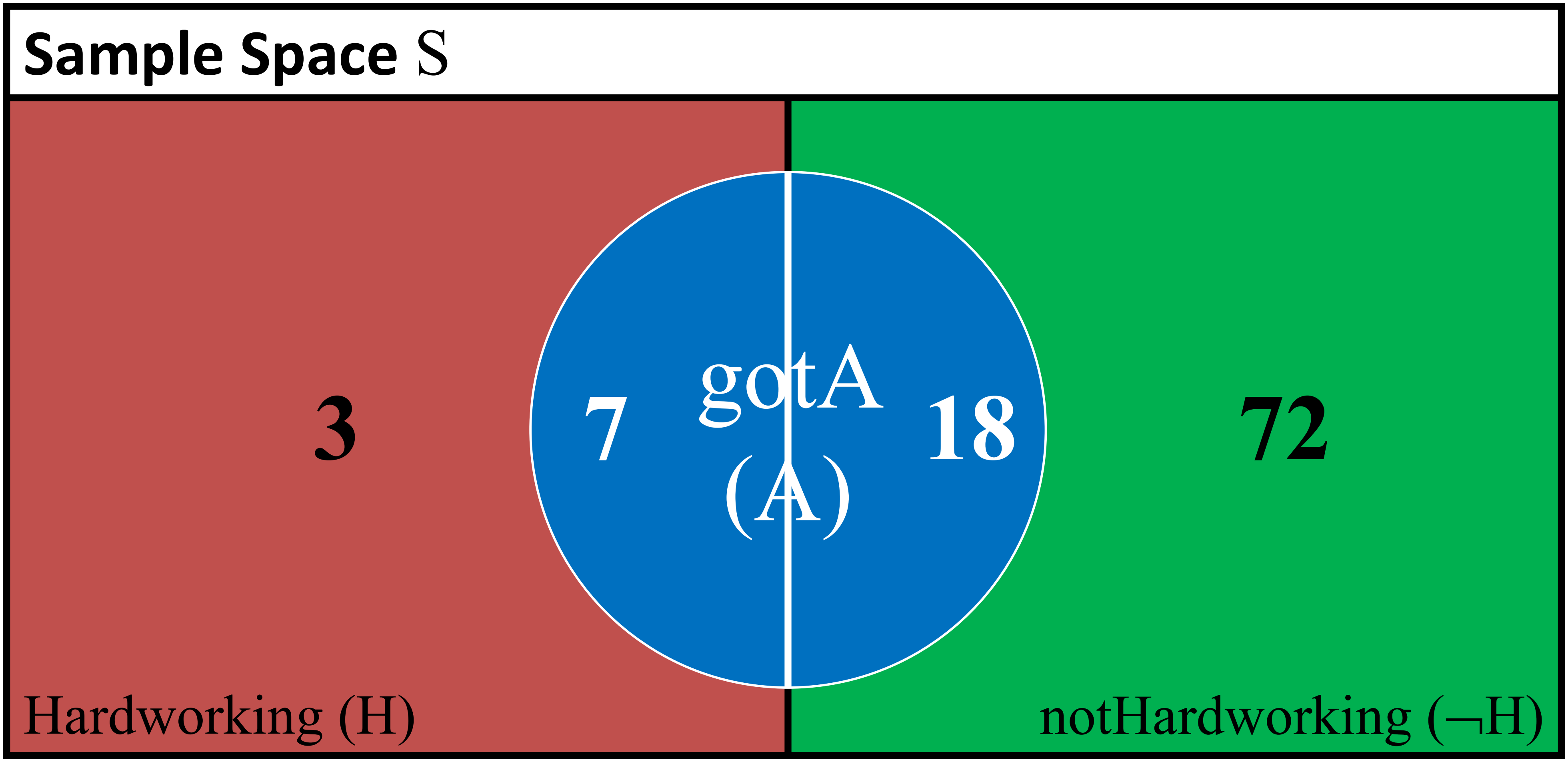
$$P(H \mid A) = ?$$

Conditional Probability: Visualization



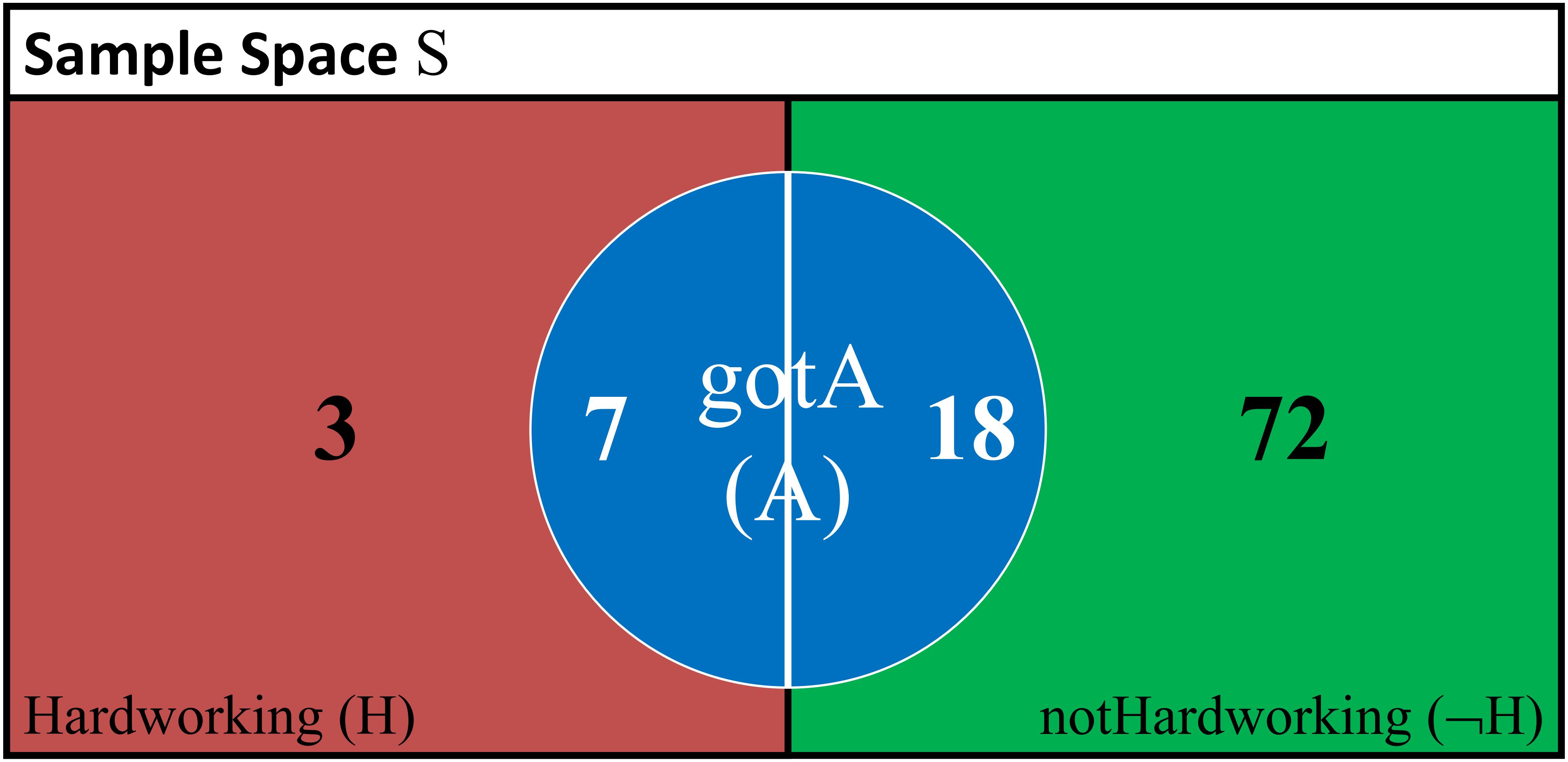
$$P(H \mid A) = \frac{P(H \cap A)}{P(A)} = \frac{7/100}{97/100} = \frac{7}{97}$$

Conditional Probability: Visualization



$$P(H \mid A) = ?$$

Conditional Probability: Visualization



$$P(H \mid A) = \frac{P(H \cap A)}{P(A)} = \frac{7/100}{25/100} = \frac{7}{25}$$

Chain Rule

Conditional probabilities can be used to decompose joint probabilities using the chain rule. For any random variables $f_1, f_2, \ldots, f_n$ and values $x_1, x_2, \ldots, x_n$:

$$P(f_1 = x_1, f_2 = x_2, \ldots, f_n = xn) =$$

$$P(f_1 = x_1) *$$

$$P(f_2 \mid f_1 = x_1) *$$

$$P(f_3 \mid f_1 = x_1, f_2 = x_2) *$$

$$\ldots$$

$$P(f_n = xn \mid f_1 = x_1, \ldots, f_{n-1} = x_{n-1}) =$$

$$= \prod_{i=1}^{n} P(f_i = xi \mid f_1 = x_1, \ldots, f_{i-1} = x_{i-1})$$

Independence

Two events are independent if one does not convey any information about the other.

Two events A and B are independent if:

$$P(A \cap B) = P(A) * P(B)$$

Independence

Two events A and B are independent if:

$$P(A \cap B) = P(A) * P(B)$$

So (from conditional probability formula):

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

Disjointment vs. Independence

Concept	Meaning	Formulas
Disjoint	Events A and B cannot occur at the same time	$A \cap B = \varnothing$ $P(A \cup B) = P(A) + P(B)$
Independent	Event A does not give any information about event B	$P(A \mid B) = P(A)$ $P(B \mid A) = P(B)$ $P(A \cap B) = P(A) * P(B)$

Independence

If two events A and B are independent:

- **events A and B' are independent**
- **events A' and B are independent**
- **events A' and B' are independent**

Independence

If A_1, A_2, … , A_N are independent events:

$$P(A_1 \cup A_2 \cup \ldots \cup A_N) =$$

$$= 1 - (1\text{-}P(A_1)) * (1\text{-}P(A_1)) * \ldots * (1\text{-}P(A_N))$$

Conditional Independence

Random variable X is **conditionally independent** **of random variable Y given Z if for all $x \in Dx$, for all $y \in Dy$, and for all $z \in Dz$, such that**

$$P(Y = y \wedge Z = z) > 0 \text{ and } P(Y = y' \wedge Z = z) > 0$$

$$P(X = x \mid Y = y \wedge Z = z) = P(X = x \mid Y = y' \wedge Z = z)$$

In other words, given a value of Z, knowing Y's value DOES NOT affect your belief in value of X.

Conditional Independence

The following four statements are equivalent as long as conditional probabilities:

1. X **is conditionally independent of** Y **given** Z
2. Y **is conditionally independent of** X **given** Z
3. $P(X \mid Y, Z) = P(X \mid Z)$
4. $P(X, Y \mid Z) = P(X \mid Z) * P(Y \mid Z)$

Conditional Independence

Consider three random variables: P(owerful), H(appy), R(ich)

with domains:

$$D_P = \{\text{powerful}, \text{powerless}\}, D_H = \{\text{happy}, \text{unhappy}\}, D_R = \{\text{rich}, \text{poor}\}$$

Now, when:

$$P(H = \text{happy}, R = \text{rich}) > 0 \textbf{ and } P(H = \text{unhappy}, R = \text{rich}) > 0$$

and:

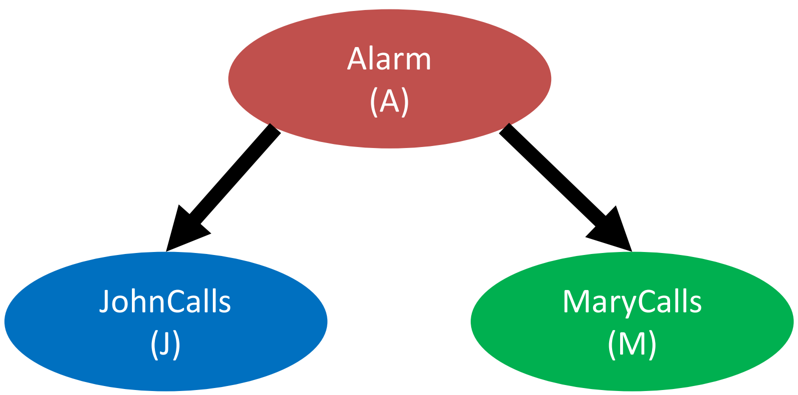
$$P(P = \text{powerful} \mid H = \text{happy}, R = \text{rich}) = P(P = \text{powerful} \mid H = \text{unhappy}, R = \text{rich})$$

In other words, given a value of R, knowing Y's value DOES NOT affect your belief in the value of X.

"Being un/happy does not make you less powerful, if you are rich."

More On Conditional Independence

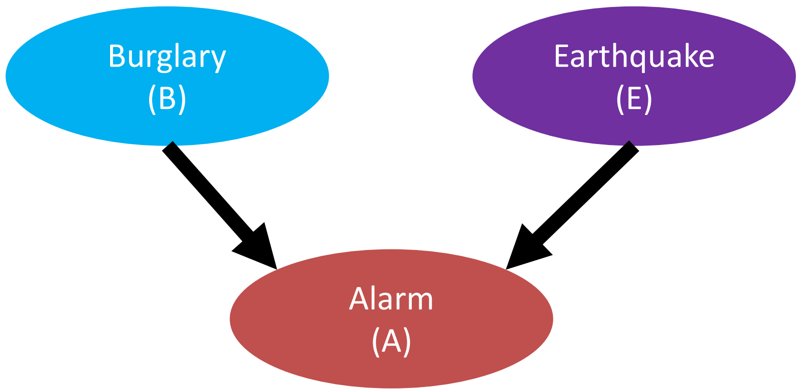
Common Cause:



JohnCalls and MaryCalls are NOT independent

JohnCalls and MaryCalls are CONDITIONALLY independent given Alarm

Common Effect:



Burglary and Earthquake are independent

Burglary and Earthquake are NOT CONDITIONALLY independent given Alarm

Bayes' Rule

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

Bayes' Rule

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

Bayes' Rule

$$P(cause \mid effect) = \frac{P(effect \mid cause) * P(cause)}{P(effect)}$$

Bayes' Rule

$P(cause \mid effect)$ **diagnostic direction relation**

$$P(cause \mid effect) = \frac{P(effect \mid cause) * P(cause)}{P(effect)}$$

$P(effect \mid cause)$ **causal direction relation**

Bayes' Rule

$P(disease \mid symptoms)$ **diagnostic direction relation**

$$P(disease \mid symptoms) = \frac{P(symptoms \mid disease) * P(disease)}{P(symptoms)}$$

$P(symptoms \mid disease)$ **causal direction relation**

Bayes' Rule

Why is this useful?

- **Because in practice it is easier to get probabilities for** P(effect|cause) **and** P(cause) **than for** P(cause|effect)

$$P(disease \mid symptoms) = \frac{P(symptoms \mid disease) * P(disease)}{P(symptoms)}$$

  - **It is easier to know what symptoms diseases cause. It is harder to diagnose a disease given symptoms**

Bayes' Rule

$$P(cause \mid effect) = \frac{P(effect \mid cause) * P(cause)}{P(effect)}$$

Problem: a single card is drawn from a standard deck of cards. What is the probability that we drew a queen if we know that a face card (J, Q, K) was drawn?

$$P(queen \mid face) = \frac{P(face \mid queen) * P(queen)}{P(face)}$$

$$P(queen \mid face) = \frac{1 * 4 / 52}{12 / 52} = \frac{1}{3}$$

Bayes' Rule

$$P(cause \mid effect) = \frac{P(effect \mid cause) * P(cause)}{P(effect)}$$

Problem: Calculate probability that a patient has meningitis if a patient has stiff neck. Meningitis is a cause of neck stiffness in 70% of cases, probability of having meningitis is 1/50000. Stiff neck happens to 1% of patients.

$$P(m \mid s) = \frac{P(s \mid m) * P(m)}{P(s)}$$

$$P(m \mid s) = \frac{0.7 * 1/50000}{0.01} = 0.0014$$

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.

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

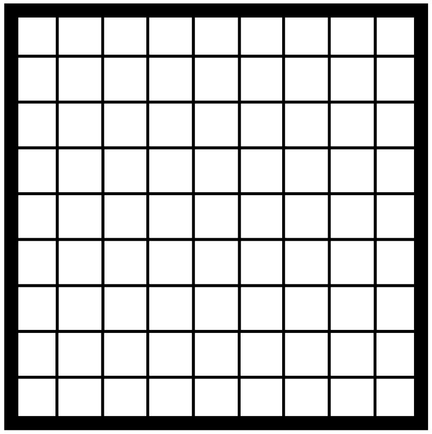
$$P(Hypothesis \mid evidence) = \frac{P(evidence \mid Hypothesis) * P(Hypothesis)}{P(evidence)}$$

where:

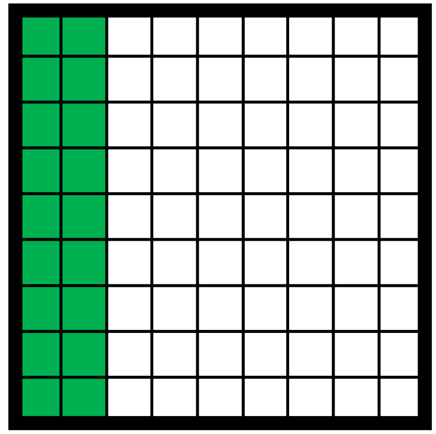
- P(H) - **probability of the** Hypothesis H **being true BEFORE we see new data/**evidence e **(prior probability)**
- P(H | e) - **probability of the** Hypothesis H **being true AFTER we see new data/**evidence e **(posterior probability)**
- P(e | H) - **probability of new data/**evidence e **being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

Bayes' Rule: Visual Interpretation

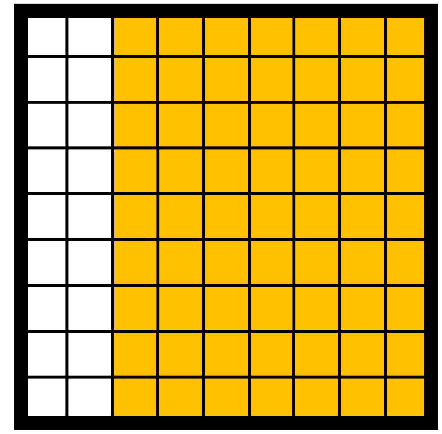
All possible cases



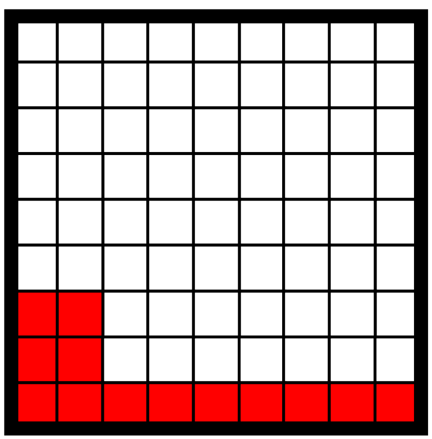
Cases where Hypothesis H **is true**
$P(H)$



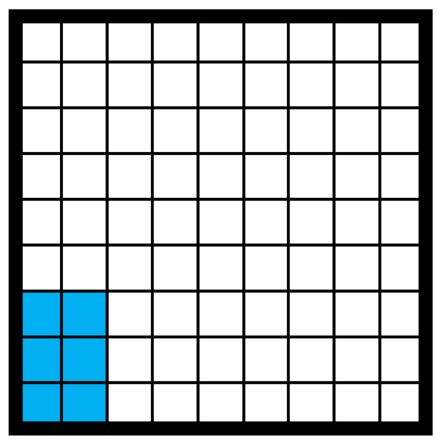
Cases where Hypothesis H **is false**
$P(\neg H)$



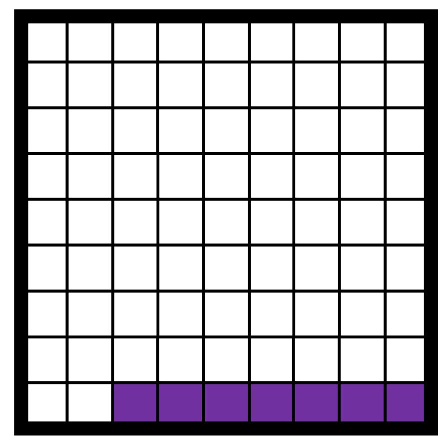
Cases where evidence e **is true**
$P(e)$



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

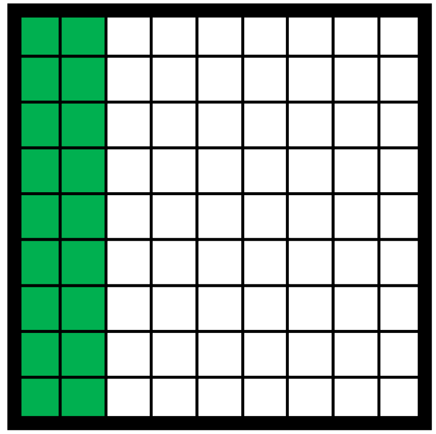
Bayes' Rule:

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

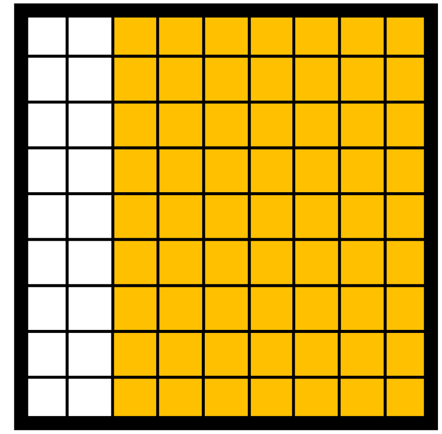
$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

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

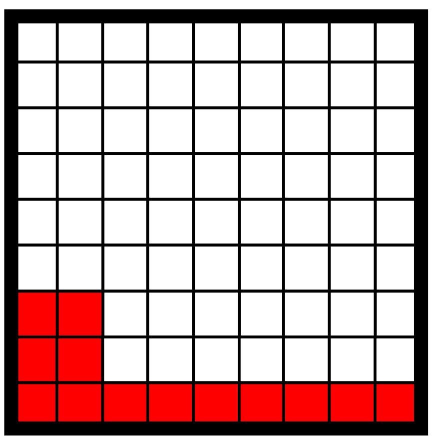
Cases where Hypothesis H **is true**
$P(H)$



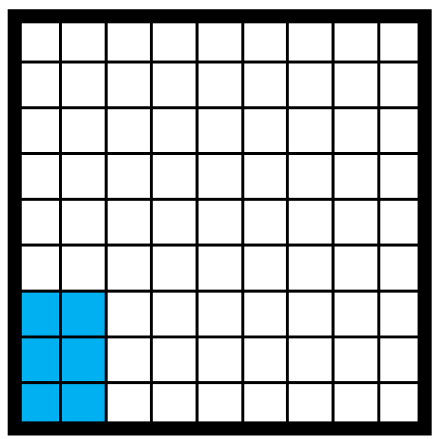
Cases where Hypothesis H **is false**
$P(\neg H)$



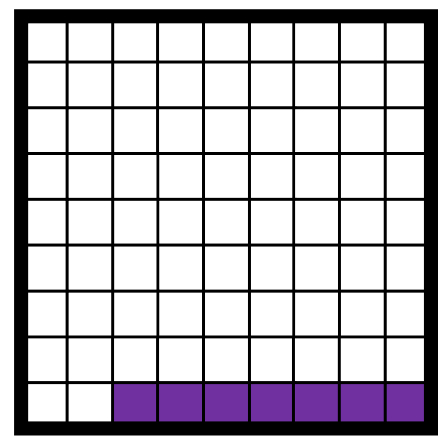
Cases where evidence e **is true**
$P(e)$



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



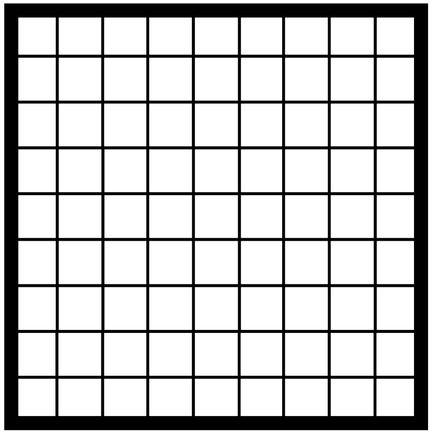
Cases where evidence e **is true given** Hypothesis H **false** $P(e \mid \neg H)$



Bayes' Rule: Visual Interpretation

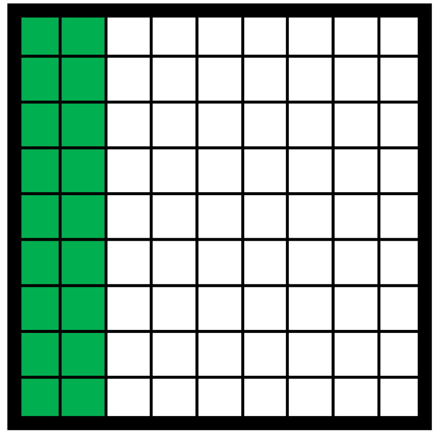
All Students

Hypothesis H: **graduate student**



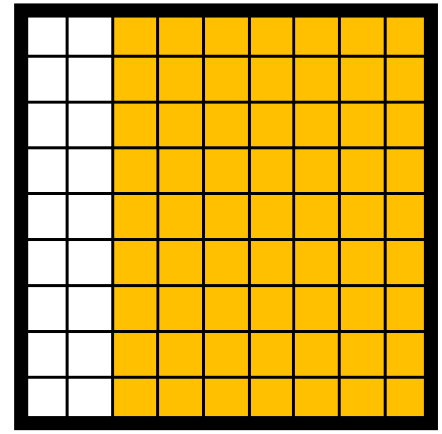
Cases where Hypothesis H **is true**

P(H) = P(grad = true)



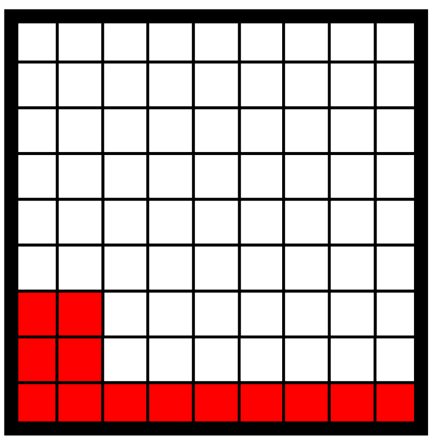
Cases where Hypothesis H **is false**

P(¬H) = P(grad = false)



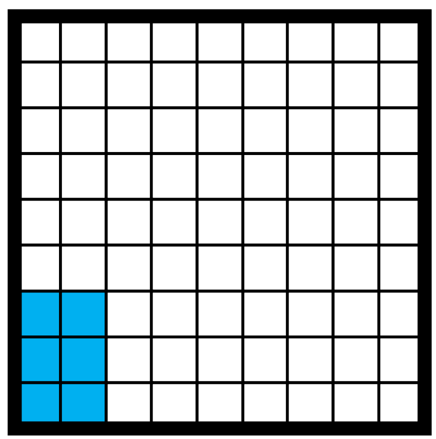
Cases where evidence e **is true**

P(e) = P(female = true)



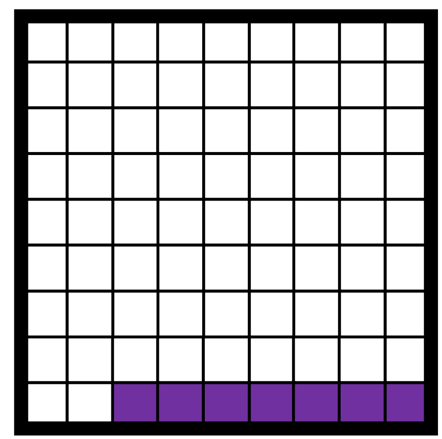
Cases where e **true given** H **true**

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 (<u>made up</u> roster data):

%of G students: P(H)

%of UG students: P(¬H)

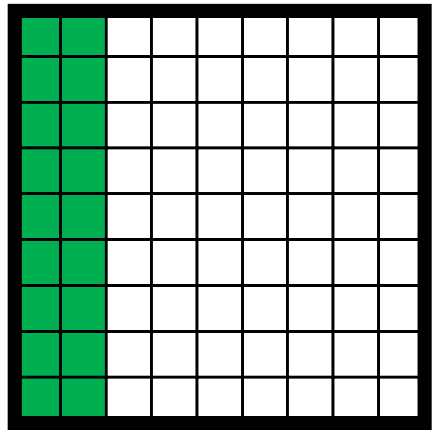
%of female students: P(e)

%of female G students: P(e | H)

%of female UG students: P(e | ¬H)

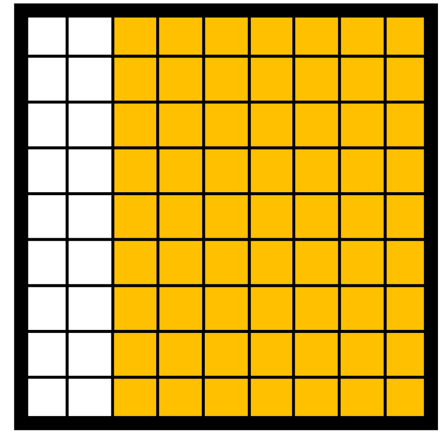
Cases where Hypothesis H **is true**

P(H) = 18 / 81



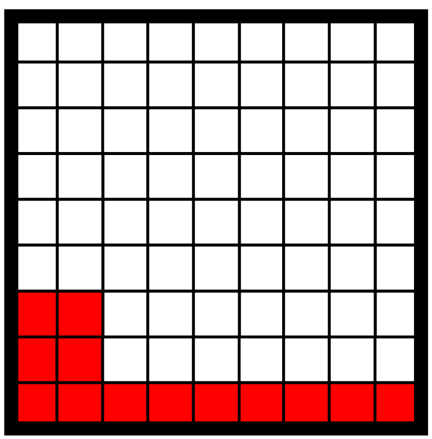
Cases where Hypothesis H **is false**

P(¬H) = 63 / 81



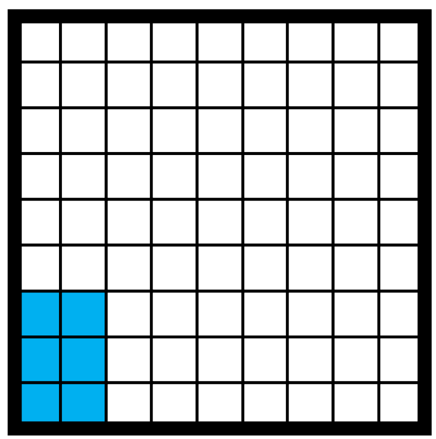
Cases where evidence e **is true**

P(e) = 13 / 81



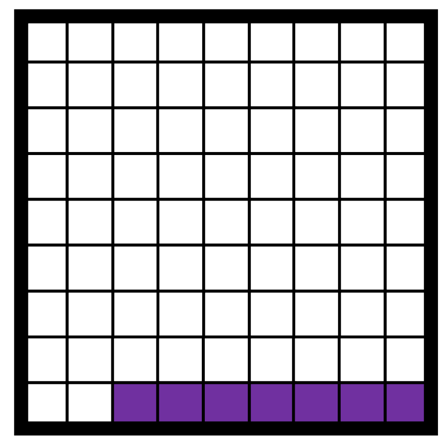
Cases where e **true given** H **true**

P(e | H) = 6 / 18



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

P(e | ¬H) = 7 / 63



Bayes' Rule: Visual Interpretation

Bayes' Rule:

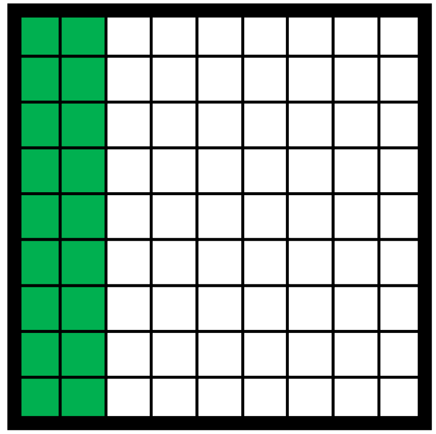
$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

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

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

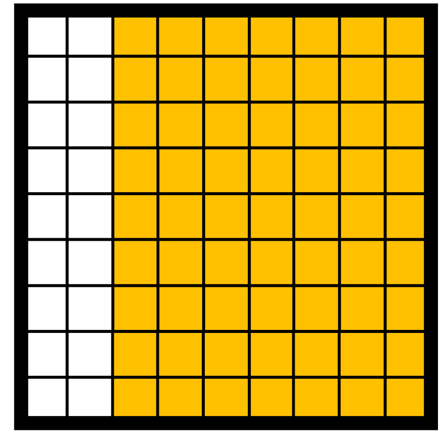
Cases where Hypothesis H **is true**

P(H) = 18 / 81



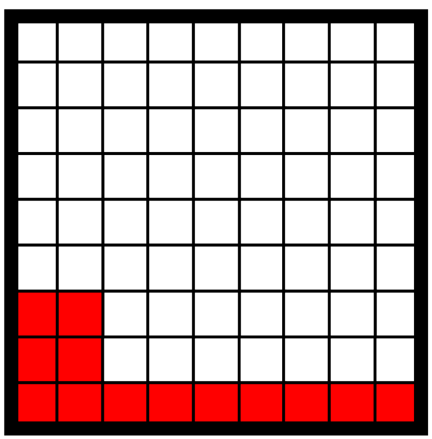
Cases where Hypothesis H **is false**

P(¬H) = 63 / 81



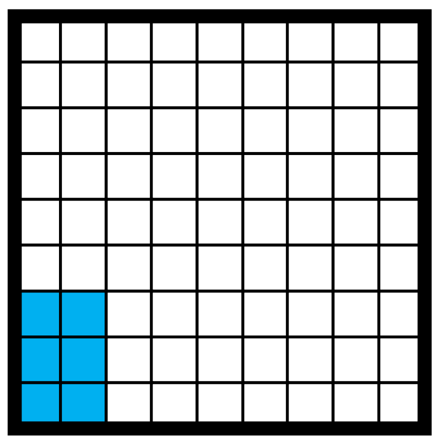
Cases where evidence e **is true**

P(e) = 13 / 81



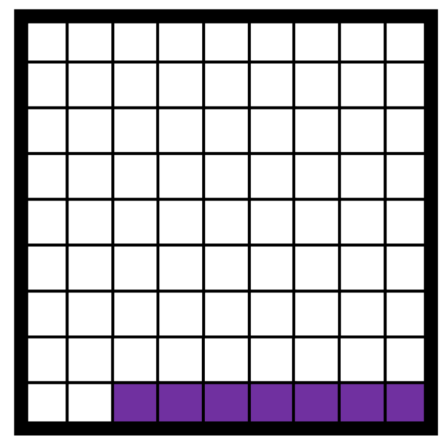
Cases where e **true given** H **true**

P(e | H) = 6 / 18



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

P(e | ¬H) = 7 / 63



Bayes’ Rule: Visual Interpretation

Bayes’ Rule:

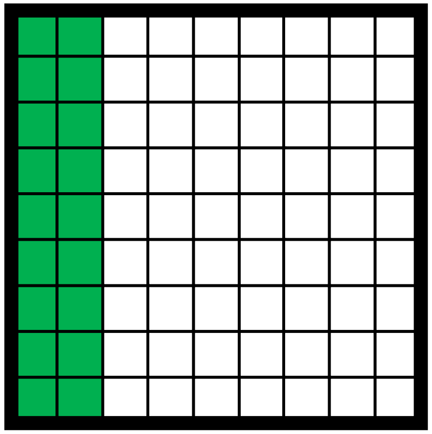
$$P(H \mid e) = \frac{P(e \mid H) * P(H)}{P(e)}$$

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

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

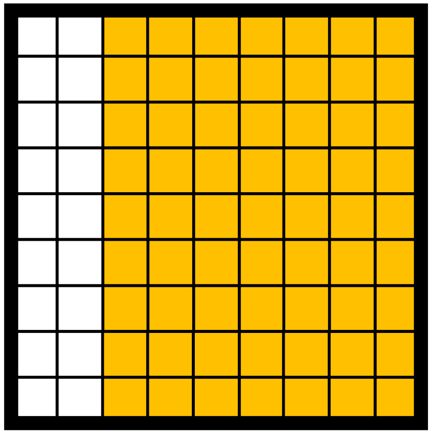
Cases where Hypothesis H is true

P(H) = 18 / 81



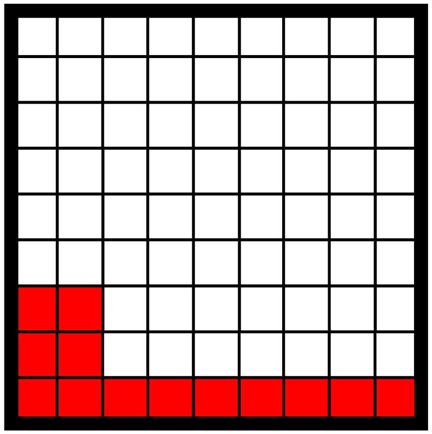
Cases where Hypothesis H is false

P(¬H) = 63 / 81



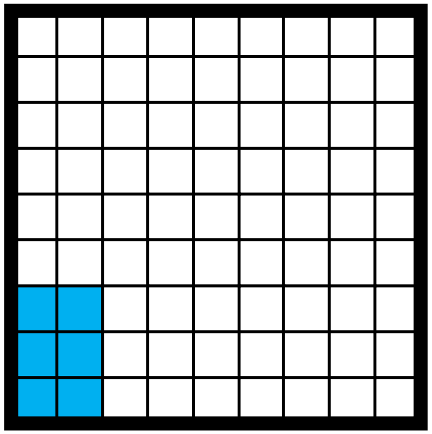
Cases where evidence e is true

P(e) = 13 / 81



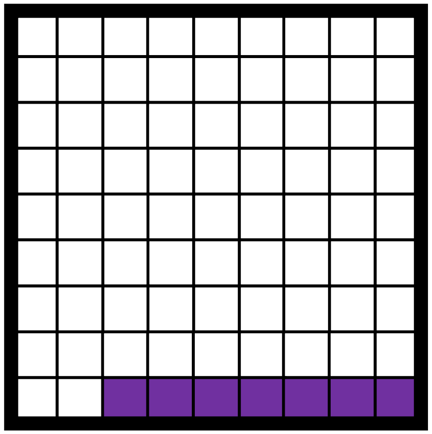
Cases where e true given H true

P(e | H) = 6 / 18



Cases where e true given H false

P(e | ¬H) = 7 / 63



Bayes' Rule: Visual Interpretation

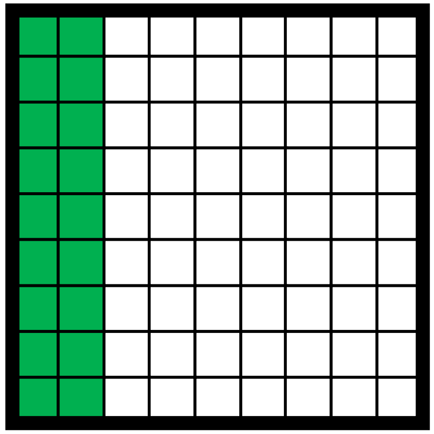
Bayes' Rule:

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

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

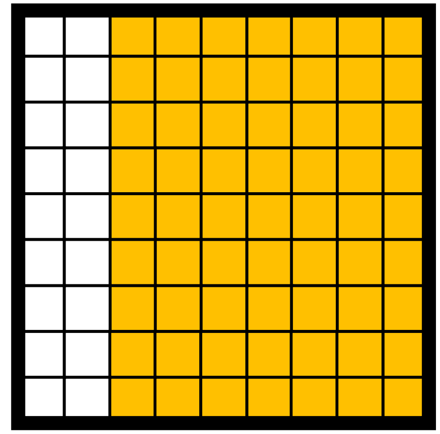
Cases where Hypothesis H **is true**

P(H) = 18 / 81



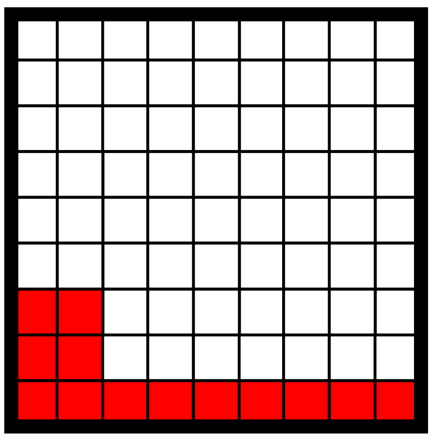
Cases where Hypothesis H **is false**

P(¬H) = 63 / 81



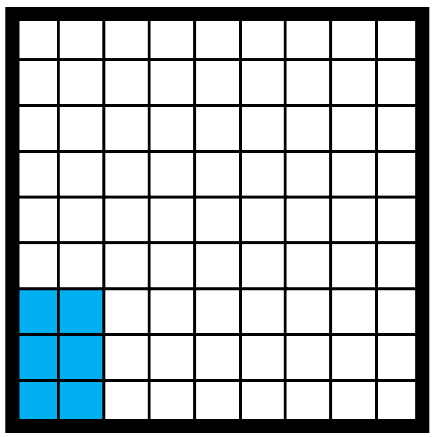
Cases where evidence e **is true**

P(e) = 13 / 81



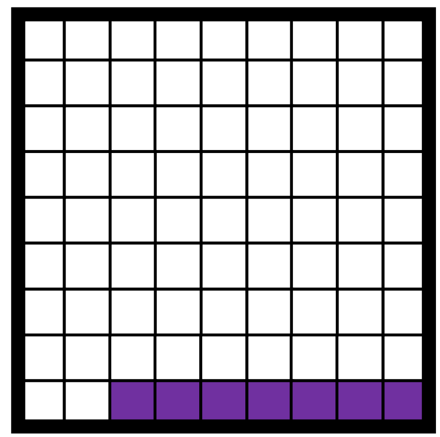
Cases where e **true given** H **true**

P(e | H) = 6 / 18



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

P(e | ¬H) = 7 / 63



Bayes' Rule: Visual Interpretation

Prior probability:

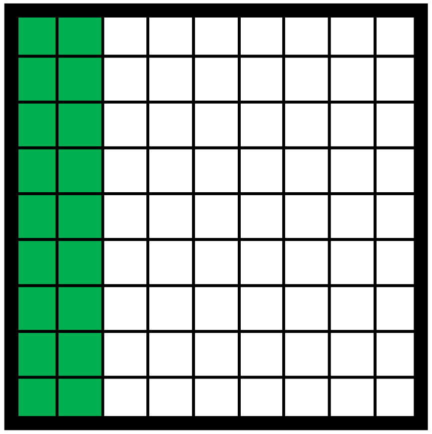
$$P(H) = 18 / 81 \approx 0.222$$

Posterior probability:

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

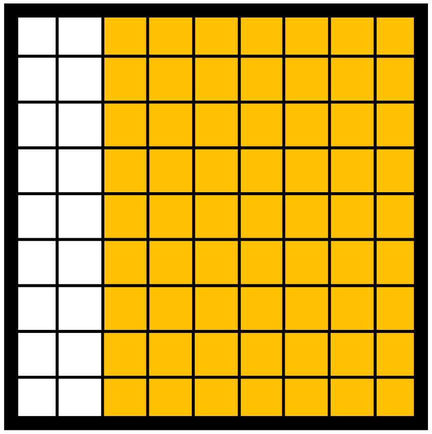
Cases where Hypothesis H is true

$P(H) = 18 / 81$



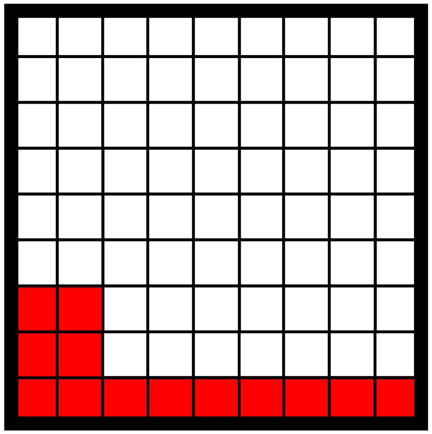
Cases where Hypothesis H is false

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



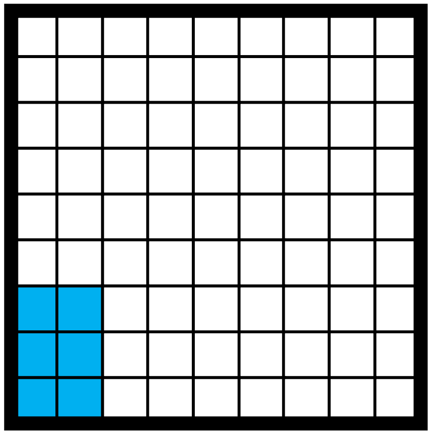
Cases where evidence e is true

$P(e) = 13 / 81$



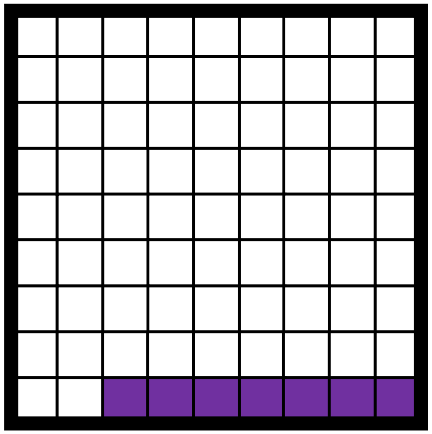
Cases where e true given H true

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



Cases where e true given H false

$P(e \mid \neg H) = 7 / 63$



Bayes' Rule: Belief/Probability Update

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

this is a grad student ($\mathrm{grad} = \mathrm{true}$)

My <u>belief</u> in H **being true is based on prior probability:**

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

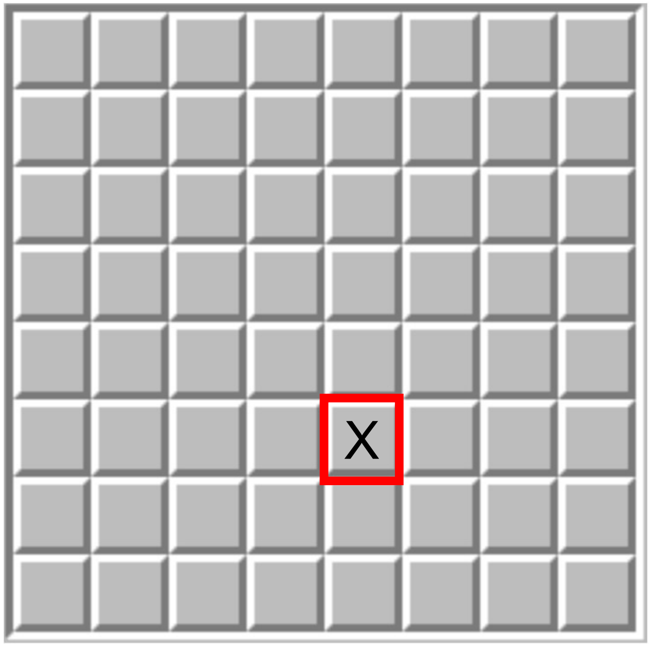
I look up and see a female student, which is <u>new data / evidence</u> e ($\mathrm{female} = \mathrm{true}$)**. Bayes' Rule helps me update my <u>belief</u> in** H **being true with <u>posterior</u> probability:**

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

Playing Minesweeper with Bayes' Rule

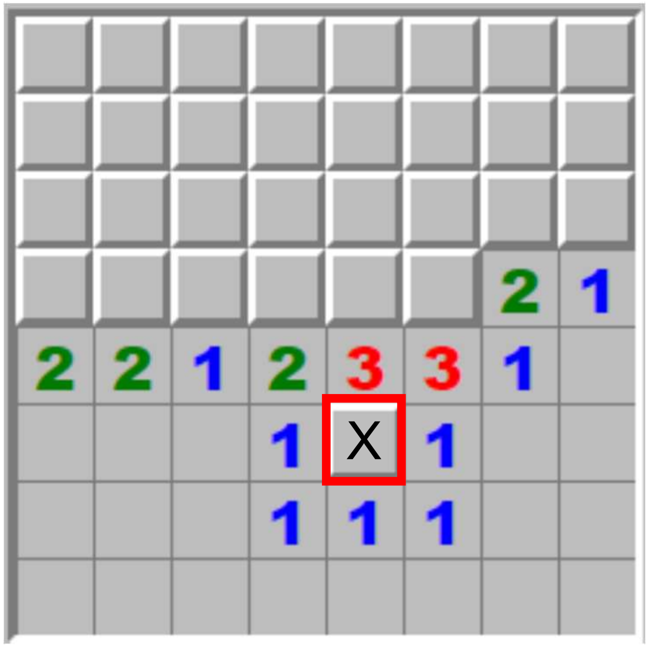
Prior probability / belief:

$$P(X = mine) = 0.5$$



Posterior probability / belief:

$$P(X = mine \mid \boldsymbol{evidence}) = 1.0$$



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.

Full Joint Probability Distribution

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

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:

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

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 P() is the distribution, NOT individual probability:

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

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

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

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

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

Complex Joint Distributions

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

	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	

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!

2^N Possible Worlds (Models)

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

2^N values

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	
		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 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

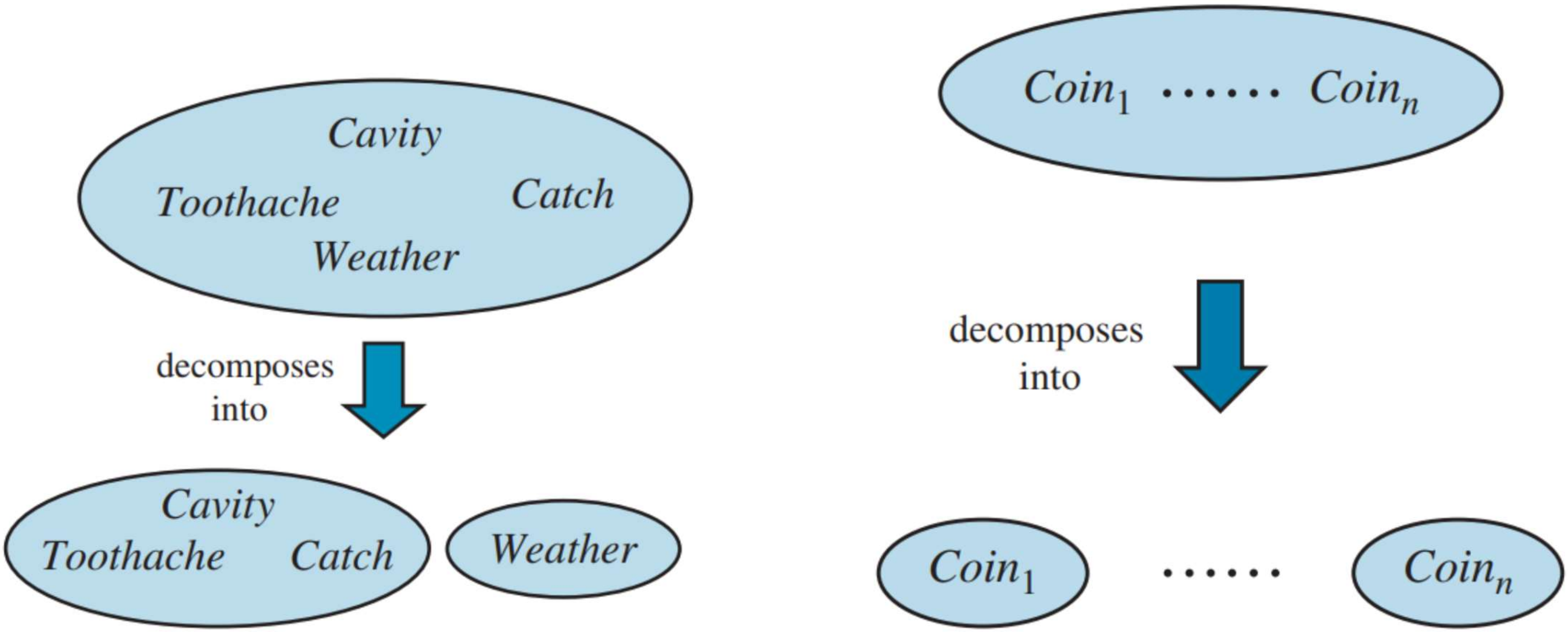
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
¬Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576
		Toothache		¬Toothache	
		Catch	¬Catch	Catch	¬Catch
Cloudy	Cavity	0.108	0.012	0.072	0.008
	¬Cavity	0.016	0.064	0.144	0.576

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

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

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

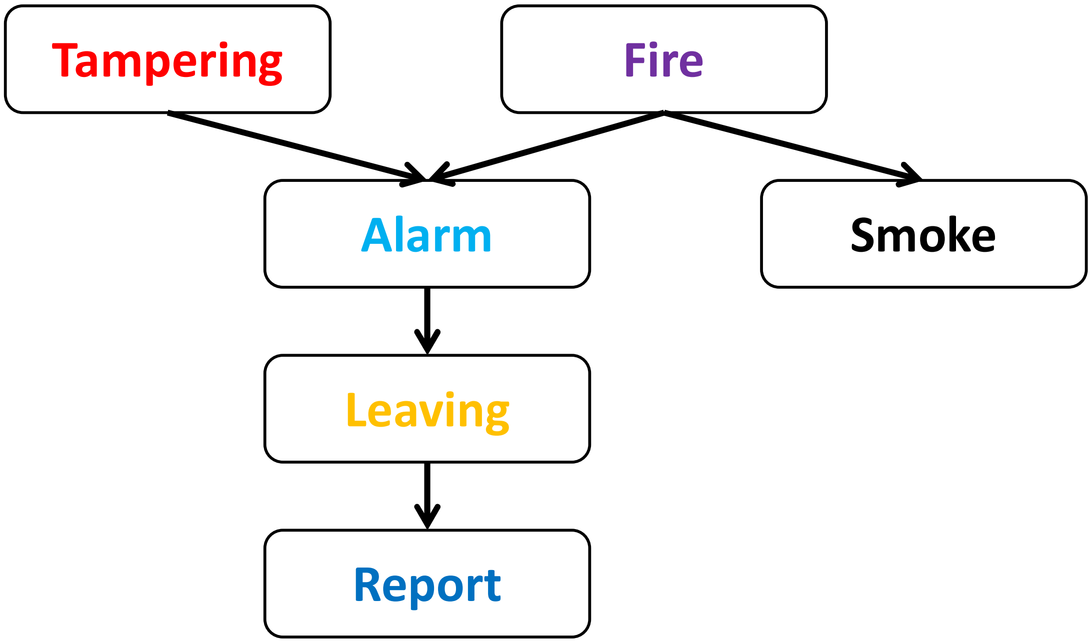
A Bayesian belief network describes the joint probability distribution for a set of variables.

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 $P(X_i \mid \mathrm{parents}(X_i))$

Bayesian (Belief) Network: Example



Random Variables (Propositions):

- **Tampering: true if the alarm is tampered with**
- **Fire: true if there is a fire**
- **Alarm: true if the alarm sounds**
- **Smoke: true if there is smoke**
- **Leaving: true if people leaving the building at once**
- **Report: true if someone who left the building reports fire**

Domain for all variables: $\{\text{true}, \text{false}\}$

NOTE: RVs don't have to be Boolean

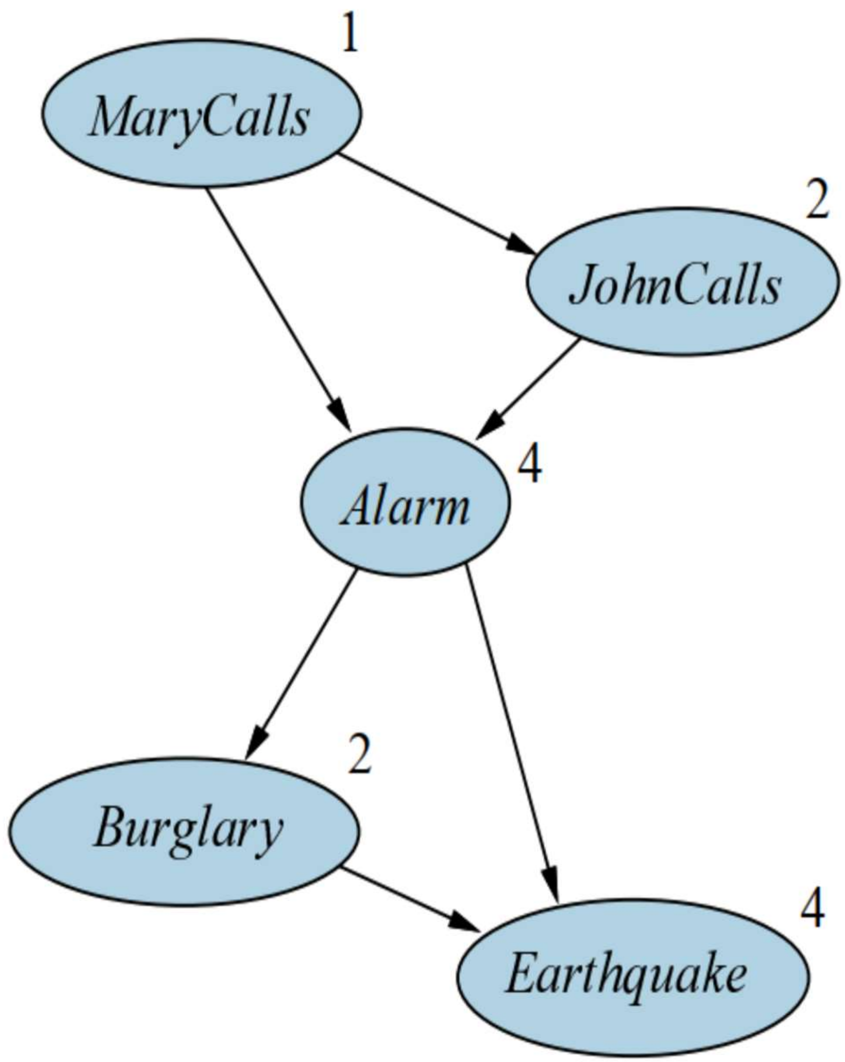
Building Bayesian (Belief) Network

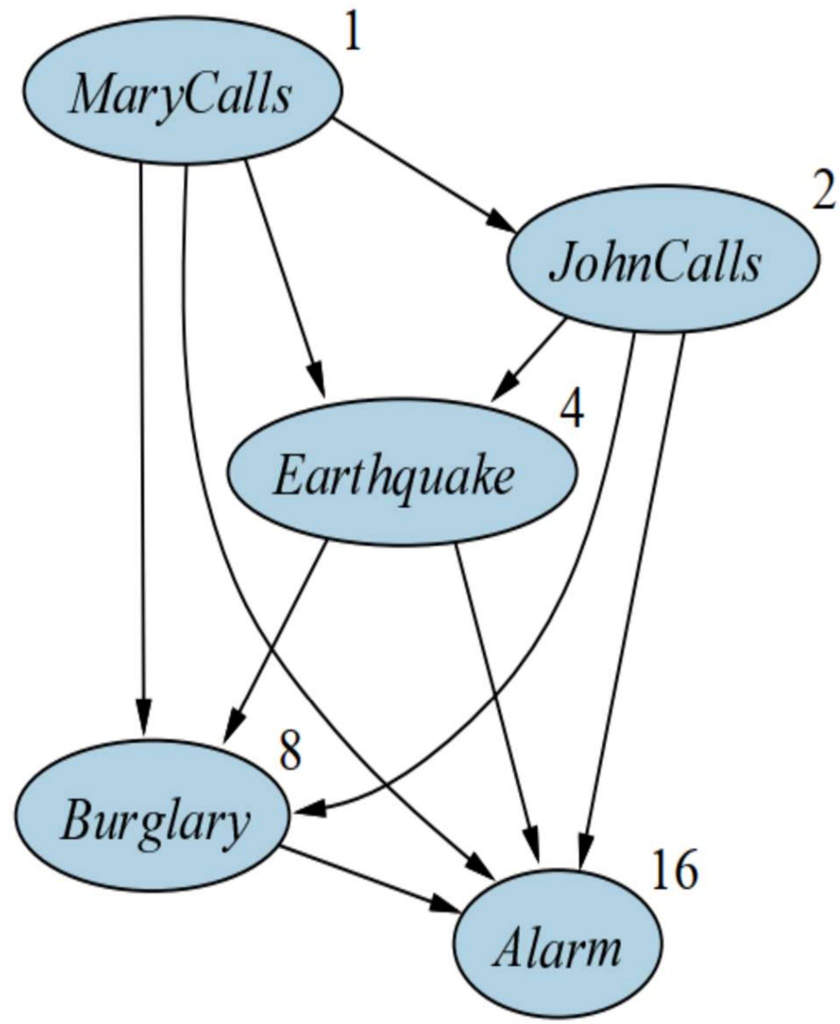
1. Order Random Variables (ordering matters!)
2. Create network nodes for each Random Variable
3. Add edges between parent nodes and children nodes
   - For every node node X_i:
     - choose a minimal set S of parents for X_i
     - for each parent node Y in S add an edge from Y to X_i

4. Add Conditional Probability Tables

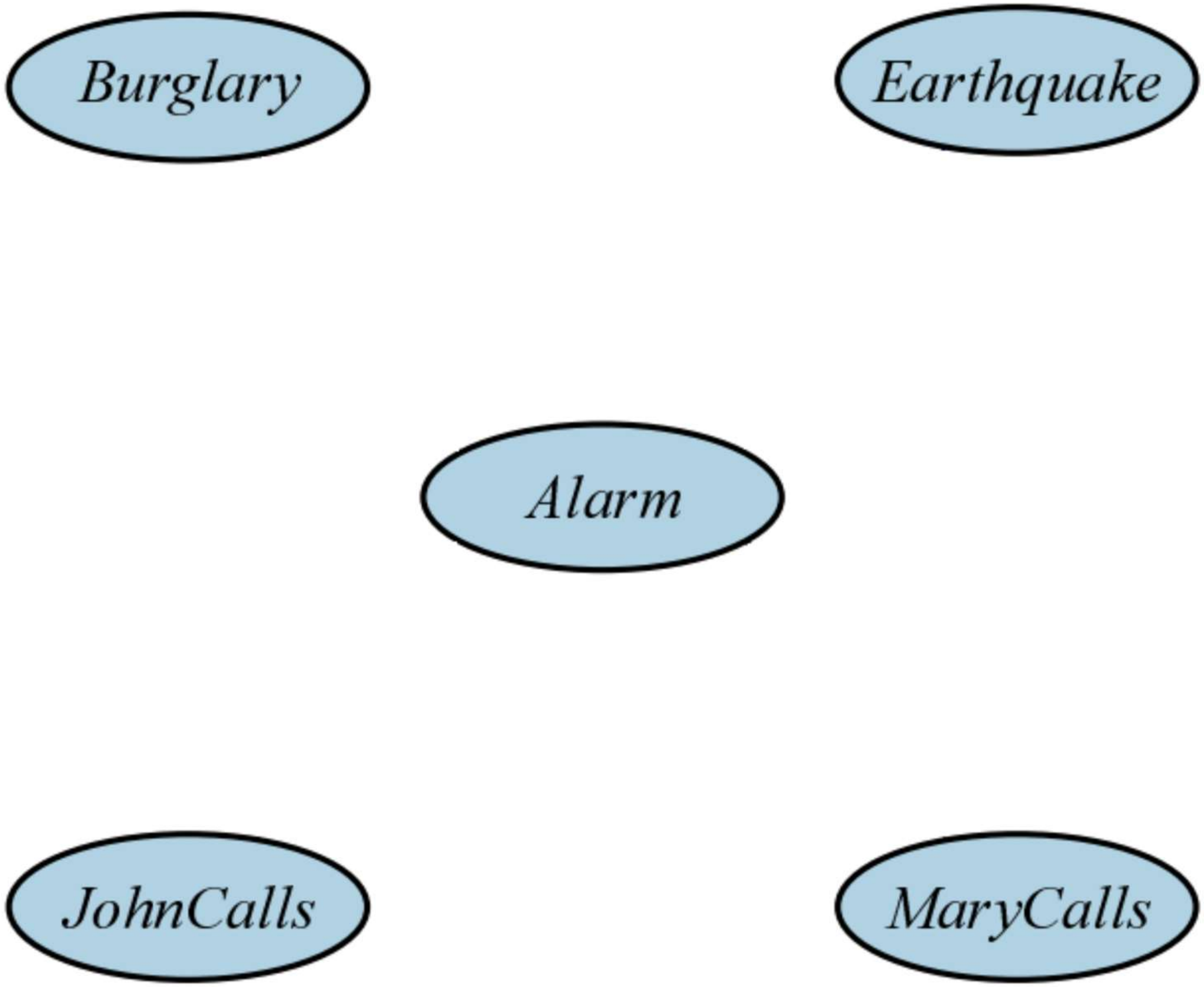
Make it compact / sparse: choose your Random Variable ordering wisely.

Ordering Matters!

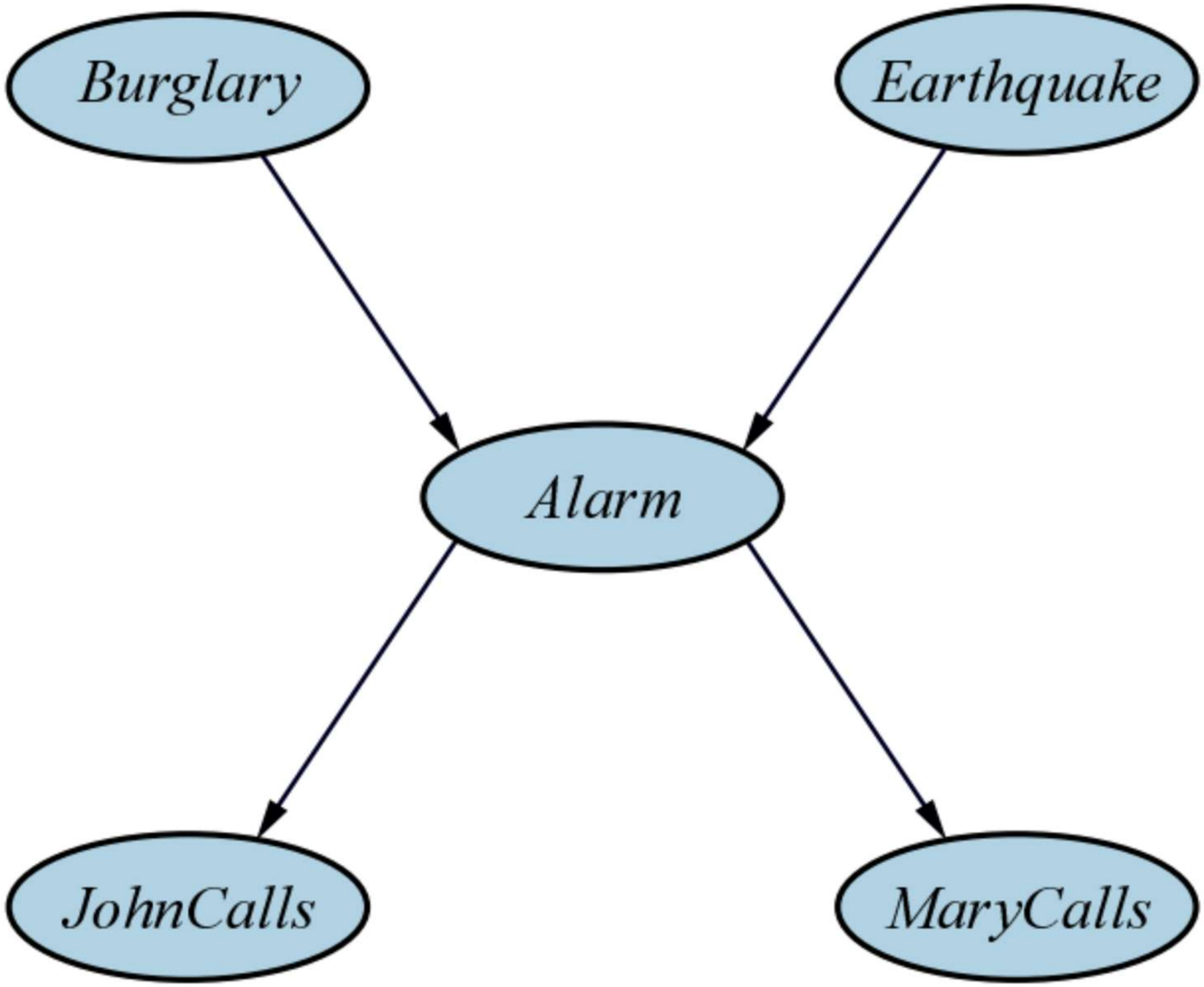




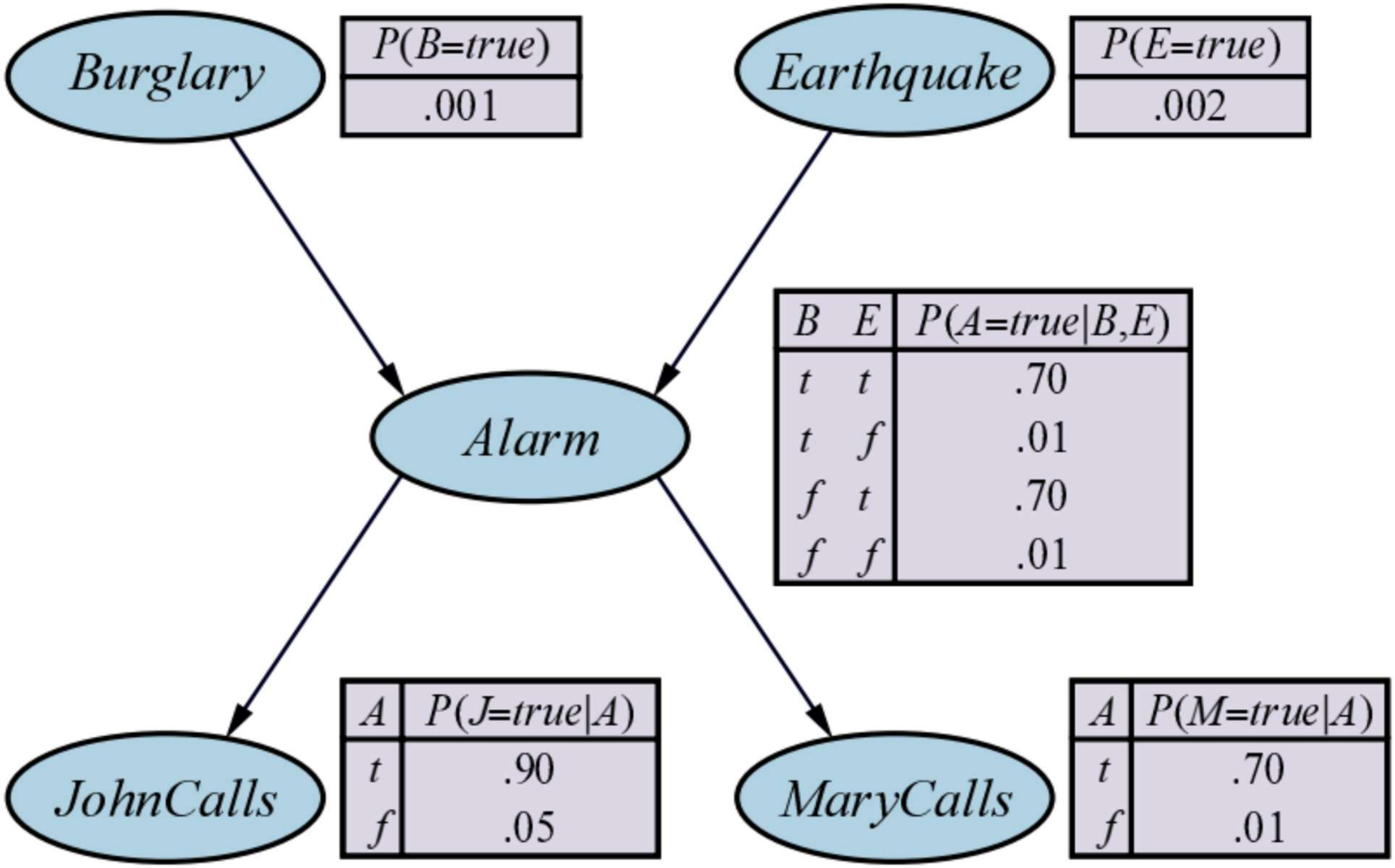
Create Vertices / Node / Random Vars



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	$P(H, e) = P(H \wedge e) = P(H) * P(e \mid H) = 18 / 81 * 6 / 18 \quad 0.074$
true	false	$P(H, \neg e) = P(H) * P(\ \ e \mid H) = 18 / 81 * 12 / 18 \quad 0.148$
false	true	$P(\ \ H, e) = P(\ \ H) * P(e \mid \neg H) = 63 / 81 * 7 / 63 \quad 0.086$
false	false	$P(\ \ H, \neg e) = P(\ \ H) * P(\ \ e \mid \neg H) = 63 / 81 * 56 / 63 \quad 0.691$
		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