Lecture notes for *Principles of Artificial Intelligence* (COMS 4720/5720)
Yan-Bin Jia, Iowa State University

# Bayes' Rule

## Outline

I. Bayes' rule

II. Conditional independence

* Figures are either from the textbook site or by the instructor.

# I. Bayes' Rule

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$ (product rule)

# I. Bayes' Rule

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$ (product rule)

$$\Downarrow$$

$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

# I. Bayes' Rule

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$ (product rule)

$$\Downarrow$$

$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

$$\Downarrow$$

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$ (Bayes' rule)

Probability of
$b$ given $a$

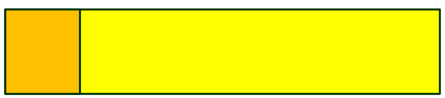
# I. Bayes' Rule

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$ (product rule)

$$\Downarrow$$

$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

$$\Downarrow$$



$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$ (Bayes' rule)

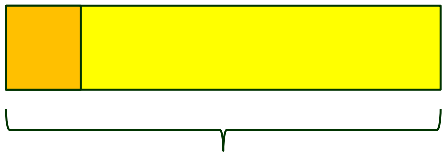
Probability of
$b$ given $a$

# I. Bayes' Rule

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$ (product rule)

$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$ (Bayes' rule)



Probability of $a$: $P(a)$

Probability of $b$ given $a$

# I. Bayes' Rule

$$P(a \wedge b) = P\,(a \mid b)P(b) = P\,(b \mid a)P(a)$$ (product rule)

⇓

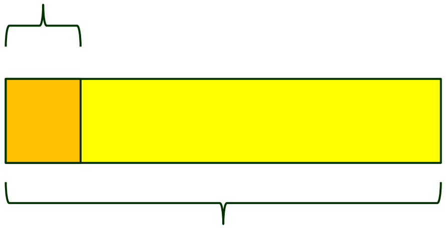
$$P\,(a \mid b)P(b) = P\,(b \mid a)P(a)$$

⇓

$$P\,(b \mid a) = \frac{P\,(a \mid b)P(b)}{P(a)}$$ (Bayes' rule)

Probability of $b$ given $a$

Portion caused by $b$: $P(a \mid b)P(b)$



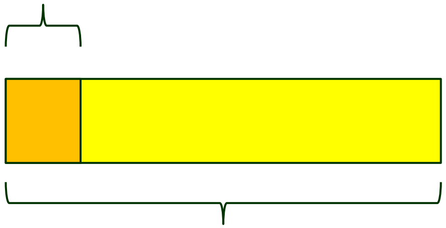
Probability of $a$: $P(a)$

# I. Bayes' Rule

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$ (product rule)

$$P(a \mid b)P(b) = P(b \mid a)P(a)$$

Portion caused by $b$: $P(a \mid b)P(b)$



Probability of $a$: $P(a)$

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$ (Bayes' rule)

Probability of $b$ given $a$

It tells us how often $b$ happens given that $a$ happens, when we know:

- how often $a$ happens given that $b$ happens, and
- how likely $a$ is on its own, and
- how likely $b$ is on its own.

# Bayes' Rule for Multivalued Variables

$$\mathbf{P}(Y \mid X) = \frac{\mathbf{P}(X \mid Y)\mathbf{P}(Y)}{\mathbf{P}(X)}$$

The above is a set of equations, each for a pair of possible values of $X$ and $Y$.

# Bayes' Rule for Multivalued Variables

$$\boldsymbol{P}(Y \mid X) = \frac{\boldsymbol{P}(X \mid Y)\boldsymbol{P}(Y)}{\boldsymbol{P}(X)}$$

The above is a set of equations, each for a pair of possible values of $X$ and $Y$.

A more generalized version conditionalized on some evidence $\boldsymbol{e}$:

$$\boldsymbol{P}(Y \mid X, \boldsymbol{e}) = \frac{\boldsymbol{P}(X \mid Y, \boldsymbol{e})\boldsymbol{P}(Y \mid \boldsymbol{e})}{\boldsymbol{P}(X \mid \boldsymbol{e})} = \frac{\boldsymbol{P}(X, Y \mid \boldsymbol{e})}{\boldsymbol{P}(X \mid \boldsymbol{e})}$$

# Applying Bayes' Rule

- Perceive as the evidence the *effect* of some unknown *cause*.
- Determine the cause.

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

# Applying Bayes' Rule

- Perceive as the evidence the *effect* of some unknown *cause*.
- Determine the cause.

causal relationship

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

# Applying Bayes' Rule

- Perceive as the evidence the *effect* of some unknown *cause*.
- Determine the cause.

diagnostic relationship causal relationship

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

# Applying Bayes' Rule

- Perceive as the evidence the *effect* of some unknown *cause*.
- Determine the cause.

diagnostic relationship    causal relationship

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

Probabilistic information is often available in the form $P(effect \mid cause)$ not $P(cause \mid effect)$.

# Applying Bayes' Rule

- Perceive as the evidence the *effect* of some unknown *cause*.
- Determine the cause.

diagnostic relationship causal relationship

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

Probabilistic information is often available in the form $P(effect \mid cause)$ not $P(cause \mid effect)$.

The doctor knows $P(symptoms \mid disease)$ and wants to derive a diagnosis $P(disease \mid symptoms)$.

# Example 1

The doctor knows:

- The disease meningitis causes a patient to have a stiff neck 70% of time.
- The prior probability that any patient has meningitis is 1/50,000.
- The prior probability that any patient has a stiff neck is 1/100.

# Example 1

The doctor knows:

- The disease meningitis causes a patient to have a stiff neck 70% of time.
- The prior probability that any patient has meningitis is 1/50,000.
- The prior probability that any patient has a stiff neck is 1/100.

$s$: The patient has a stiff neck.
$m$: The patient has meningitis.

# Example 1

The doctor knows:

- The disease meningitis causes a patient to have a stiff neck 70% of time.
- The prior probability that any patient has meningitis is 1/50,000.
- The prior probability that any patient has a stiff neck is 1/100.

$s$: The patient has a stiff neck.
$m$: The patient has meningitis.

$$P(s \mid m) = 0.7$$
$$P(m) = 1/50000$$
$$P(s) = 0.01$$

# Example 1

The doctor knows:

- The disease meningitis causes a patient to have a stiff neck 70% of time.
- The prior probability that any patient has meningitis is 1/50,000.
- The prior probability that any patient has a stiff neck is 1/100.

$s$: The patient has a stiff neck.
$m$: The patient has meningitis.

$$P(s \mid m) = 0.7$$

$$P(m) = 1/50000$$

$$P(s) = 0.01$$

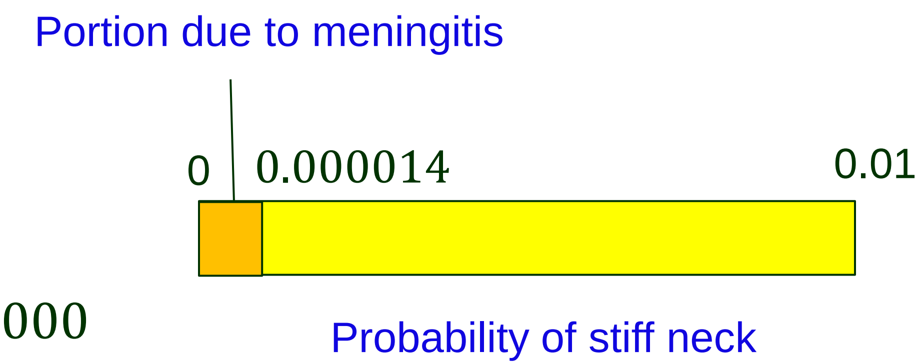
$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \cdot 1/50000}{0.01} = 0.0014$$

# Example 1

The doctor knows:

- The disease meningitis causes a patient to have a stiff neck 70% of time.
- The prior probability that any patient has meningitis is 1/50,000.
- The prior probability that any patient has a stiff neck is 1/100.

$s$: The patient has a stiff neck.
$m$: The patient has meningitis.




$$P(s \mid m) = 0.7$$

$$P(m) = 1/50000$$

$$P(s) = 0.01$$

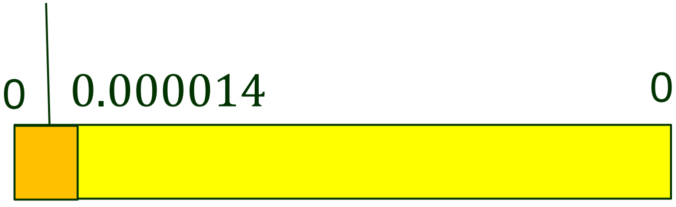
$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \cdot 1/50000}{0.01} = 0.0014$$

# Example 1

The doctor knows:

- The disease meningitis causes a patient to have a stiff neck 70% of time.
- The prior probability that any patient has meningitis is 1/50,000.
- The prior probability that any patient has a stiff neck is 1/100.

$s$: The patient has a stiff neck.
$m$: The patient has meningitis.




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

Identify the probability of meningitis given a stiff neck with the disease' portion of contribution to the symptom.

# Prior Probability of the Evidence

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

# Prior Probability of the Evidence

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

# Prior Probability of the Evidence

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

We obtain $P(s)$ from computing $P(s \mid m)$ and $P(s \mid \neg m)$:

# Prior Probability of the Evidence

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

We obtain $P(s)$ from computing $P(s \mid m)$ and $P(s \mid \neg m)$:

$$P(m, s) = P(m \mid s)\, P(s) = P(s \mid m)\, P(m) \qquad P(\neg m \mid s)\, P(s) = P(s \mid \neg m)\, P(\neg m)$$

# Prior Probability of the Evidence

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

We obtain $P(s)$ from computing $P(s \mid m)$ and $P(s \mid \neg m)$:

$$P(m, s) = P(m \mid s)\, P(s) = P(s \mid m)\, P(m) \qquad P(\neg m \mid s)\, P(s) = P(s \mid \neg m)\, P(\neg m)$$

$$\Downarrow$$

$$P(m \mid s)\, P(s) + P(\neg m \mid s)\, P(s) = P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)$$

# Prior Probability of the Evidence

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

We obtain $P(s)$ from computing $P(s \mid m)$ and $P(s \mid \neg m)$:

$$P(m, s) = P(m \mid s)\, P(s) = P(s \mid m)\, P(m) \qquad P(\neg m \mid s)\, P(s) = P(s \mid \neg m)\, P(\neg m)$$

$$P(m \mid s)\, P(s) + P(\neg m \mid s)\, P(s) = P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)$$

$$P(m \mid s) + P(\neg m \mid s) = 1$$

$$P(s) = P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)$$

# Example 2

**Beliefs**

- If a patient has lung cancer, there is a 60% chance that an X-ray test will come back positive; and a 40% chance negative.
- If a patient does not have lung cancer, there is 2% percent chance that an X-ray test will come back positive ; and a 98% percent chance negative.
- Population cancer rate is 1/1000.

**Observation** X-ray test came back positive.

**Inference task** What is the chance that the patient has lung cancer?

# Example 2

**Beliefs**

- If a patient has lung cancer, there is a 60% chance that an X-ray test will come back positive; and a 40% chance negative.
- If a patient does not have lung cancer, there is 2% percent chance that an X-ray test will come back positive ; and a 98% percent chance negative.
- Population cancer rate is 1/1000.

**Observation** X-ray test came back positive.

**Inference task** What is the chance that the patient has lung cancer?

$c$: The patient has cancer.

$x$: The patient's X-ray test comes back positive.

# Example 2

**Beliefs**

- If a patient has lung cancer, there is a 60% chance that an X-ray test will come back positive; and a 40% chance negative.
- If a patient does not have lung cancer, there is 2% percent chance that an X-ray test will come back positive ; and a 98% percent chance negative.
- Population cancer rate is 1/1000.

**Observation** X-ray test came back positive.

**Inference task** What is the chance that the patient has lung cancer?

$c$: The patient has cancer.

$x$: The patient's X-ray test comes back positive.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} = \frac{P(x \mid c)P(c)}{P(x \mid c)P(c) + P(x \mid \neg c)P(\neg c)}$$

# Example 2

**Beliefs**

- If a patient has lung cancer, there is a 60% chance that an X-ray test will come back positive; and a 40% chance negative.
- If a patient does not have lung cancer, there is 2% percent chance that an X-ray test will come back positive ; and a 98% percent chance negative.
- Population cancer rate is 1/1000.

**Observation** X-ray test came back positive.

**Inference task** What is the chance that the patient has lung cancer?

$c$: The patient has cancer.

$x$: The patient's X-ray test comes back positive.

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)} = \frac{P(x \mid c)P(c)}{P(x \mid c)P(c) + P(x \mid \neg c)P(\neg c)}$$

$$= \frac{0.6 \cdot 0.001}{0.6 \cdot 0.001 + 0.02 \cdot 0.999} \approx 0.02915$$

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\boldsymbol{P}(M \mid s) = \alpha \langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m) \rangle$$

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\mathbf{P}(M \mid s) = \alpha\langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m)\rangle$$

random variable with two values $m$ and $\neg m$.

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\boldsymbol{P}(M \mid s) = \alpha \langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m) \rangle$$

random variable with two values $m$ and $\neg m$.

$$\alpha = \frac{1}{P(s)} = \frac{1}{P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)}$$

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\boldsymbol{P}(M \mid s) = \alpha \langle P(s \mid m) P(m), P(s \mid \neg m) P(\neg m) \rangle$$

random variable with two values $m$ and $\neg m$.

$$\alpha = \frac{1}{P(s)} = \frac{1}{P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)}$$

- $P(\neg m) = 1 - P(m)$

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\mathbf{P}(M \mid s) = \alpha \langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m)\rangle$$

random variable with two values $m$ and $\neg m$.

$$\alpha = \frac{1}{P(s)} = \frac{1}{P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)}$$

- $P(\neg m) = 1 - P(m)$
- $P(s \mid \neg m)$ needs to be estimated instead of $P(s)$.

  Sometimes this is easier, sometimes harder.

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\mathbf{P}(M \mid s) = \alpha \langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m) \rangle$$

random variable with two values $m$ and $\neg m$.

$$\alpha = \frac{1}{P(s)} = \frac{1}{P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)}$$

- $P(\neg m) = 1 - P(m)$
- $P(s \mid \neg m)$ needs to be estimated instead of $P(s)$.

  Sometimes this is easier, sometimes harder.

$$\mathbf{P}(Y \mid X) = \alpha\, \mathbf{P}(X \mid Y)\mathbf{P}(Y)$$

# Bayes' Rule (General Form)

Apply normalization to Bayes' rule when $P(s)$ is unknown:

$$\mathbf{P}(M \mid s) = \alpha \langle P(s \mid m)P(m), P(s \mid \neg m)P(\neg m) \rangle$$

random variable with two values $m$ and $\neg m$.

$$\alpha = \frac{1}{P(s)} = \frac{1}{P(s \mid m)\, P(m) + P(s \mid \neg m)\, P(\neg m)}$$

- $P(\neg m) = 1 - P(m)$
- $P(s \mid \neg m)$ needs to be estimated instead of $P(s)$.

  Sometimes this is easier, sometimes harder.

$$\mathbf{P}(Y \mid X) = \alpha\, \mathbf{P}(X \mid Y)\mathbf{P}(Y)$$

normalization constant to make the entries in $\mathbf{P}(Y \mid X)$ sum to 1.

# II. Combining Evidence

What happens when we have two or more pieces of evidences?

- Suppose we know the full joint distribution.

| | $toothache$ | | $\neg toothache$ | |
|---|---|---|---|---|
| | $catch$ | $\neg catch$ | $catch$ | $\neg catch$ |
| $cavity$ | 0.108 | 0.012 | 0.072 | 0.008 |
| $\neg cavity$ | 0.016 | 0.064 | 0.144 | 0.576 |

# II. Combining Evidence

What happens when we have two or more pieces of evidences?

- Suppose we know the full joint distribution.

| | *toothache* | | $\neg$*toothache* | |
|---|---|---|---|---|
| | *catch* | $\neg$*catch* | *catch* | $\neg$*catch* |
| *cavity* | 0.108 | 0.012 | 0.072 | 0.008 |
| $\neg$*cavity* | 0.016 | 0.064 | 0.144 | 0.576 |

$$\boldsymbol{P}(Cavity \mid toothache \wedge catch) = \alpha\langle 0.108, 0.016\rangle \approx \langle 0.871, 0.129\rangle$$

$$\frac{1}{\alpha} = P(toothache \wedge catch) = 0.108 + 0.016 = 0.124$$

# II. Combining Evidence

What happens when we have two or more pieces of evidences?

- Suppose we know the full joint 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 |

$$\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\langle 0.108, 0.016\rangle \approx \langle 0.871, 0.129\rangle$$

$$\frac{1}{\alpha} = P(toothache \wedge catch) = 0.108 + 0.016 = 0.124$$

♠ Does not scale up to a large number of evidence variables.

# Exponential Growth

- Apply Bayes' rule:

$$\boldsymbol{P}(Cavity \mid toothache \wedge catch) = \alpha \boldsymbol{P}(toothache \wedge catch \mid Cavity)\, \boldsymbol{P}(Cavity)$$

namely,

$$\langle P(cavity \mid toothache \wedge catch), P(\neg cavity \mid toothache \wedge catch) \rangle$$

$$= \alpha \langle P(toothache \wedge catch \mid cavity), P(toothache \wedge catch \mid \neg cavity) \rangle$$

$$\cdot \langle P(cavity), P(\neg cavity) \rangle$$ // elementwise product

# Exponential Growth

- Apply Bayes' rule:

$$\boldsymbol{P}(\mathit{Cavity} \mid \mathit{toothache} \wedge \mathit{catch}) = \alpha \boldsymbol{P}(\mathit{toothache} \wedge \mathit{catch} \mid \mathit{Cavity})\, \boldsymbol{P}(\mathit{Cavity})$$

namely,

$$\langle P(\mathit{cavity} \mid \mathit{toothache} \wedge \mathit{catch}), P(\neg \mathit{cavity} \mid \mathit{toothache} \wedge \mathit{catch}) \rangle$$

$$= \alpha \langle P(\mathit{toothache} \wedge \mathit{catch} \mid \mathit{cavity}), P(\mathit{toothache} \wedge \mathit{catch} \mid \neg \mathit{cavity}) \rangle$$

$$\cdot \langle P(\mathit{cavity}), P(\neg \mathit{cavity}) \rangle$$ // elementwise product

- $n$ (Boolean) evidence variables: *Toothache*, *Catch*, *X-rays*, *Diet*, …

# Exponential Growth

- Apply Bayes' rule:

$$\boldsymbol{P}(Cavity \mid toothache \wedge catch) = \alpha \boldsymbol{P}(toothache \wedge catch \mid Cavity)\, \boldsymbol{P}(Cavity)$$

namely,

$$\langle P(cavity \mid toothache \wedge catch), P(\neg cavity \mid toothache \wedge catch)\rangle$$

$$= \alpha \langle P(toothache \wedge catch \mid cavity), P(toothache \wedge catch \mid \neg cavity)\rangle$$

$$\cdot \langle P(cavity), P(\neg cavity)\rangle$$ // elementwise product

- $n$ (Boolean) evidence variables: *Toothache*, *Catch*, *X-rays*, *Diet*, …

⇒ $2^n$ possible combinations of observed values $\boldsymbol{P}(toothache \wedge catch \mid Cavity \wedge Toothache \wedge \cdots)$ are needed!

# Conditional Independence

- *Toothache* and *Catch* are independent *given the presence or absence of a cavity*.

# Conditional Independence

- *Toothache* and *Catch* are independent *given the presence or absence of a cavity*.
  - Each is directly caused by the cavity.
  - Neither has a direct effect on the other.

    Toothache depends on the state of the nerves in the tooth.

    The probe's accuracy depends on the dentist's skill irrelevant to the toothache.

# Conditional Independence

- *Toothache* and *Catch* are independent *given the presence or absence of a cavity*.
  - Each is directly caused by the cavity.
  - Neither has a direct effect on the other.

    Toothache depends on the state of the nerves in the tooth.

    The probe's accuracy depends on the dentist's skill irrelevant to the toothache.

- Conditional independence of *toothache* and *catch* given *Cavity*:

$$\mathbf{P}(toothache \wedge catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)$$

# Conditional Independence

- *Toothache* and *Catch* are independent *given the presence or absence of a cavity*.
  - Each is directly caused by the cavity.
  - Neither has a direct effect on the other.

    Toothache depends on the state of the nerves in the tooth.

    The probe's accuracy depends on the dentist's skill irrelevant to the toothache.

- Conditional independence of *toothache* and *catch* given *Cavity*:

$$\mathbf{P}(toothache \wedge catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)$$

Bayes' rule: $\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$

# Conditional Independence

- *Toothache* and *Catch* are independent *given the presence or absence of a cavity*.
  - Each is directly caused by the cavity.
  - Neither has a direct effect on the other.

    Toothache depends on the state of the nerves in the tooth.

    The probe's accuracy depends on the dentist's skill irrelevant to the toothache.

- Conditional independence of *toothache* and *catch* given *Cavity*:

$$\mathbf{P}(toothache \wedge catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)$$

Bayes' rule: $\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\, \boxed{\mathbf{P}(toothache \wedge catch \mid Cavity)}\, \mathbf{P}(Cavity)$

# Conditional Independence

- *Toothache* and *Catch* are independent *given the presence or absence of a cavity*.
  - Each is directly caused by the cavity.
  - Neither has a direct effect on the other.

    Toothache depends on the state of the nerves in the tooth.

    The probe's accuracy depends on the dentist's skill irrelevant to the toothache.

- Conditional independence of *toothache* and *catch* given *Cavity*:

$$\mathbf{P}(toothache \wedge catch \mid Cavity) = \mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(catch \mid Cavity)$$

Bayes' rule: $\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$

$$\Downarrow$$

$$\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha \mathbf{P}(toothache \mid Cavity)\ \mathbf{P}(catch \mid Cavity)\mathbf{P}(Cavity)$$

# Definition of Conditional Independence

Conditional independence of two variables $X$ and $Y$ given a third variable $Z$:

$$\mathbf{P}(\underbrace{X, Y}_{X \wedge Y} \mid Z) = \mathbf{P}(X \mid Z)\, \mathbf{P}(Y \mid Z)$$

# Definition of Conditional Independence

Conditional independence of two variables $X$ and $Y$ given a third variable $Z$:

$$\boldsymbol{P}(\underbrace{X, Y}_{X \wedge Y} \mid Z) = \boldsymbol{P}(X \mid Z)\, \boldsymbol{P}(Y \mid Z)$$

$$\boldsymbol{P}(\mathit{Toothache}, \mathit{Catch} \mid \mathit{Cavity}) = \boldsymbol{P}(\mathit{Toothache} \mid \mathit{Cavity})\, \boldsymbol{P}(\mathit{Catch} \mid \mathit{Cavity})$$

// asserts independence for all possible combinations of values of
// *Toothache* and *Catch*.

# Definition of Conditional Independence

Conditional independence of two variables $X$ and $Y$ given a third variable $Z$:

$$\boldsymbol{P}(\underbrace{X, Y}_{X \wedge Y} \mid Z) = \boldsymbol{P}(X \mid Z)\, \boldsymbol{P}(Y \mid Z)$$

$$\boldsymbol{P}(\mathit{Toothache}, \mathit{Catch} \mid \mathit{Cavity}) = \boldsymbol{P}(\mathit{Toothache} \mid \mathit{Cavity})\, \boldsymbol{P}(\mathit{Catch} \mid \mathit{Cavity})$$

// asserts independence for all possible combinations of values of
// *Toothache* and *Catch*.

Equivalent forms:

# Definition of Conditional Independence

Conditional independence of two variables $X$ and $Y$ given a third variable $Z$:

$$\boldsymbol{P}(\underbrace{X, Y}_{X \wedge Y} \mid Z) = \boldsymbol{P}(X \mid Z)\, \boldsymbol{P}(Y \mid Z)$$

$$\boldsymbol{P}(\mathit{Toothache}, \mathit{Catch} \mid \mathit{Cavity}) = \boldsymbol{P}(\mathit{Toothache} \mid \mathit{Cavity})\, \boldsymbol{P}(\mathit{Catch} \mid \mathit{Cavity})$$

// asserts independence for all possible combinations of values of
// *Toothache* and *Catch*.

Equivalent forms:

$$\boldsymbol{P}(X \mid Y, Z) = \boldsymbol{P}(X \mid Z)$$ and $$\boldsymbol{P}(Y \mid X, Z) = \boldsymbol{P}(Y \mid Z)$$

# Definition of Conditional Independence

Conditional independence of two variables $X$ and $Y$ given a third variable $Z$:

$$\mathbf{P}(\underbrace{X, Y}_{X \wedge Y} \mid Z) = \mathbf{P}(X \mid Z)\, \mathbf{P}(Y \mid Z)$$

$\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)$

// asserts independence for all possible combinations of values of
// *Toothache* and *Catch*.

Equivalent forms:

$\mathbf{P}(X \mid Y, Z) = \mathbf{P}(X \mid Z)$ and $\mathbf{P}(Y \mid X, Z) = \mathbf{P}(Y \mid Z)$

Decomposition into smaller conditional assertions.

# Decomposition of Joint Distribution

$\boldsymbol{P}(Toothache, Catch, Cavity)$

# Decomposition of Joint Distribution

$$\mathbf{P}(Toothache, Catch, Cavity)$$

$$= \mathbf{P}(Toothache, Catch \mid Cavity)\ \mathbf{P}(Cavity)$$

# Decomposition of Joint Distribution

$$\mathbf{P}(Toothache, Catch, Cavity)$$

$$= \mathbf{P}(Toothache, Catch \mid Cavity)\ \mathbf{P}(Cavity)$$

$$= \mathbf{P}(Toothache \mid Cavity)\ \mathbf{P}(Catch \mid Cavity)\ \mathbf{P}(Cavity)$$

# Decomposition of Joint Distribution

$$\boldsymbol{P}(Toothache, Catch, Cavity)$$

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

$$= \underbrace{\boldsymbol{P}(Toothache \mid Cavity)\ \boldsymbol{P}(Catch \mid Cavity)\ \boldsymbol{P}(Cavity)}$$

3 tables of dimensions $2 \times 2, 2 \times 2$, and $2 \times 1$
with a total of $2 + 2 + 1 = 5$ independent numbers
(which appear in the first row of every table).

# Decomposition of Joint Distribution

$$\boldsymbol{P}(Toothache, Catch, Cavity)$$

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

$$= \underbrace{\boldsymbol{P}(Toothache \mid Cavity)\ \boldsymbol{P}(Catch \mid Cavity)\ \boldsymbol{P}(Cavity)}$$

3 tables of dimensions $2 \times 2$, $2 \times 2$, and $2 \times 1$
with a total of $2 + 2 + 1 = 5$ independent numbers
(which appear in the first row of every table).

With $n$ symptoms, the size grows as $O(n)$ instead of $O(2^n)$.

# Decomposition of Joint Distribution

$$\boldsymbol{P}(Toothache, Catch, Cavity)$$

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

$$= \underbrace{\boldsymbol{P}(Toothache \mid Cavity)\ \boldsymbol{P}(Catch \mid Cavity)\ \boldsymbol{P}(Cavity)}$$

3 tables of dimensions $2 \times 2$, $2 \times 2$, and $2 \times 1$
with a total of $2 + 2 + 1 = 5$ independent numbers
(which appear in the first row of every table).

With $n$ symptoms, the size grows as $O(n)$ instead of $O(2^n)$.

- Conditional independence assertions allow probabilistic systems to scale up.
- They are more commonly available than absolute independence assertions.
- Decomposition of large probabilistic domains through conditional independence is one of the most important recent developments in AI.