Artificial Intelligence

Probability: Introduction

Jendrik Seipp

Linköping University

based on slides by Thomas Keller

Discuss Feedback

post feedback and ask questions anonymously at

`https://padlet.com/jendrikseipp/tddc17`

Unconditional Probability

Uncertainty Example I

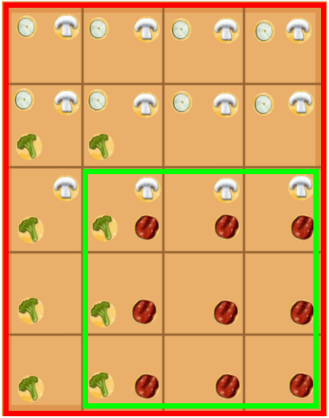
Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami?
- without salami?
- with salami and broccoli?
- with salami or brocoli?

Uncertainty Example I



Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami? $\rightsquigarrow \frac{9}{20} = 0.45$
- without salami?
- with salami and broccoli?
- with salami or brocoli?

Uncertainty Example I

Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami? $\rightsquigarrow$ $\frac{9}{20} = 0.45$
- without salami? $\rightsquigarrow$ $\frac{11}{20} = 0.55$
- with salami and broccoli?
- with salami or brocoli?

Uncertainty Example I

Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami? $\rightsquigarrow \frac{9}{20} = 0.45$
- without salami? $\rightsquigarrow \frac{11}{20} = 0.55$
- with salami and broccoli? $\rightsquigarrow \frac{3}{20} = 0.15$
- with salami or brocoli?

Uncertainty Example I

Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami? $\rightsquigarrow \frac{9}{20} = 0.45$
- without salami? $\rightsquigarrow \frac{11}{20} = 0.55$
- with salami and broccoli? $\rightsquigarrow \frac{3}{20} = 0.15$
- with salami or brocoli? $\rightsquigarrow \frac{14}{20} = 0.7$

Worlds and Sample Space

- a world ω is one possible state of the agent's environment
- often described as assignments to (propositional or finite-domain) variables

Worlds and Sample Space

- a world ω is one possible state of the agent's environment
- often described as assignments to (propositional or finite-domain) variables

  e.g., $\omega_{1,1}$ or $\omega_{3,4}$ are worlds

Worlds and Sample Space

- a world ω is one possible state of the agent's environment
- often described as assignments to (propositional or finite-domain) variables

  e.g., $\omega_{1,1}$ or $\omega_{3,4}$ are worlds

- the sample space Ω is the (finite) set of all possible worlds

  $\rightsquigarrow \Omega = \{\omega_{x,y} \mid 1 \leq x \leq 4 \text{ and } 1 \leq y \leq 5\}$

Probability Model

- a probability model associates probability $P(\omega)$ to each $\omega \in \Omega$ s.t.
  - $0 \leq P(\omega) \leq 1$ for all $\omega \in \Omega$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$

each slice is picked with equal probability:

$$P(\omega) = \frac{1}{20} \text{ for all } \omega \in \Omega$$

Events

- an event φ is a subset of Ω, e.g.:
  - $\varphi = \{\omega_{x,y} \mid 1 \le x \le 4 \text{ and } 1 \le y \le 2\}$
  - we also describe φ verbally, e.g., as the event that the slice "has onions on it"
- probability of an event is the sum of probabilities over its worlds

$$P(\varphi) = \sum_{1 \le x \le 4} \sum_{1 \le y \le 2} P(\omega_{x,y}) = \frac{8}{20}$$

Random Variables

- a random variable X is a property of the world about which we may be uncertain
- formally: $X : \Omega \mapsto dom(X)$, where $dom(X)$ is the domain (also range) of X
- by convention:
  - random variables use upper-case names
  - values use lower-case names (if possible)
  - $dom(X) = \{+x, -x\}$ if X can be true ($+x$) or false ($-x$)

Random Variables

- a random variable X is a property of the world about which we may be uncertain
- formally: $X : \Omega \mapsto dom(X)$, where $dom(X)$ is the domain (also range) of X
- by convention:
  - random variables use upper-case names
  - values use lower-case names (if possible)
  - $dom(X) = \{+x, -x\}$ if X can be true ($+x$) or false ($-x$)
- examples:
  - O (for Onion) with $dom(O) = \{+o, -o\}$
    $O(\omega_{1,1}) = +o$, $O(\omega_{2,3}) = -o$
  - $NumTops$ with $dom(NumTops) = \{1, 2, 3\}$
    $NumTops(\omega_{1,1}) = 2$, $NumTops(\omega_{2,3}) = 3$

(Probability) Distributions

How likely are the values of a random variable?

O	$P(O)$
+o	0.4
-o	0.6

M	$P(M)$
+m	0.6
-m	0.4

B	$P(B)$
+b	0.4
-b	0.6

S	$P(S)$
+s	0.45
-s	0.55

(Probability) Distributions

How likely are the values of a random variable?

O	$P(O)$
+o	0.4
-o	0.6

M	$P(M)$
+m	0.6
-m	0.4

B	$P(B)$
+b	0.4
-b	0.6

S	$P(S)$
+s	0.45
-s	0.55

$$= \sum_{\omega \in \Omega : S(\omega) = +s} P(\omega)$$

(Probability) Distributions

How likely are the values of a random variable?

O	$P(O)$
+o	0.4
-o	0.6

M	$P(M)$
+m	0.6
-m	0.4

B	$P(B)$
+b	0.4
-b	0.6

S	$P(S)$
+s	0.45
-s	0.55

$$= \sum_{\omega \in \Omega : S(\omega) = +s} P(\omega)$$

equivalent alternatives:

- $P(B = +b) = 0.4, P(B = -b) = 0.6$
- $P(+b) = 0.4, P(-b) = 0.6$
- $P(B) = \langle 0.4, 0.6 \rangle$ (requires ordered domain)

(Probability) Distributions

How likely are the values of a random variable?

O	$P(O)$
+o	0.4
-o	0.6

M	$P(M)$
+m	0.6
-m	0.4

$\sum_{m \in dom(M)} P(m) = 1$

B	$P(B)$
+b	0.4
-b	0.6

S	$P(S)$
+s	0.45
-s	0.55

$= \sum_{\omega \in \Omega : S(\omega) = +s} P(\omega)$

equivalent alternatives:

- $P(B = +b) = 0.4, P(B = -b) = 0.6$
- $P(+b) = 0.4, P(-b) = 0.6$
- $P(B) = \langle 0.4, 0.6 \rangle$ (requires ordered domain)

(Probability) Distributions

How likely are the values of a random variable?

O	$P(O)$
+o	0.4
-o	0.6

M	$P(M)$
+m	0.6
-m	0.4

$\sum_{m \in dom(M)} P(m) = 1$

short: $\sum_m P(m) = 1$

B	$P(B)$
+b	0.4
-b	0.6

S	$P(S)$
+s	0.45
-s	0.55

$= \sum_{\omega \in \Omega : S(\omega) = +s} P(\omega)$

equivalent alternatives:

- $P(B = +b) = 0.4, P(B = -b) = 0.6$
- $P(+b) = 0.4, P(-b) = 0.6$
- $P(B) = \langle 0.4, 0.6 \rangle$ (requires ordered domain)

Joint Distributions

How likely are all value combinations of a set of random variables?

S	B	$P(S,B)$
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

S	O	M	$P(S,O,M)$
+s	+o	+m	0.0
+s	+o	-m	0.0
+s	-o	+m	0.15
+s	-o	-m	0.3
-s	+o	+m	0.4
-s	+o	-m	0.0
-s	-o	+m	0.05
-s	-o	-m	0.1

Joint Distributions

How likely are all value combinations of a set of random variables?

S	B	$P(S,B)$
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

$= \sum_{\omega\in\Omega:S(\omega)=+s \text{ and } B(\omega)=-b} P(\omega)$

S	O	M	$P(S,O,M)$
+s	+o	+m	0.0
+s	+o	-m	0.0
+s	-o	+m	0.15
+s	-o	-m	0.3
-s	+o	+m	0.4
-s	+o	-m	0.0
-s	-o	+m	0.05
-s	-o	-m	0.1

Joint Distributions

How likely are all value combinations of a set of random variables?

S	B	$P(S,B)$
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

$$= \sum_{\omega \in \Omega : S(\omega)=+s \text{ and } B(\omega)=-b} P(\omega)$$

S	O	M	$P(S,O,M)$
+s	+o	+m	0.0
+s	+o	-m	0.0
+s	-o	+m	0.15
+s	-o	-m	0.3
-s	+o	+m	0.4
-s	+o	-m	0.0
-s	-o	+m	0.05
-s	-o	-m	0.1

notation: $P(-s,+o,+m) = 0.4$

Full Joint Distribution

O	M	S	B	$P(O, M, S, B)$
+o	+m	+s	+b	0.0
+o	+m	+s	-b	0.0
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
+o	-m	+s	+b	0.0
+o	-m	+s	-b	0.0
+o	-m	-s	+b	0.0
+o	-m	-s	-b	0.0
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	+m	-s	-b	0.0
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1
-o	-m	-s	-b	0.0

Full Joint Distribution

size with n variables and maximal domain size k?

O	M	S	B	$P(O, M, S, B)$
+o	+m	+s	+b	0.0
+o	+m	+s	-b	0.0
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
+o	-m	+s	+b	0.0
+o	-m	+s	-b	0.0
+o	-m	-s	+b	0.0
+o	-m	-s	-b	0.0
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	+m	-s	-b	0.0
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1
-o	-m	-s	-b	0.0

Full Joint Distribution

size with n variables and maximal domain size k? $\rightsquigarrow O(k^n)$

O	M	S	B	$P(O, M, S, B)$
+o	+m	+s	+b	0.0
+o	+m	+s	-b	0.0
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
+o	-m	+s	+b	0.0
+o	-m	+s	-b	0.0
+o	-m	-s	+b	0.0
+o	-m	-s	-b	0.0
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	+m	-s	-b	0.0
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1
-o	-m	-s	-b	0.0

Full Joint Distribution

too big even for this small example

O	M	S	B	$P(O, M, S, B)$
+o	+m	+s	+b	0.0
+o	+m	+s	-b	0.0
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
+o	-m	+s	+b	0.0
+o	-m	+s	-b	0.0
+o	-m	-s	+b	0.0
+o	-m	-s	-b	0.0
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	+m	-s	-b	0.0
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1
-o	-m	-s	-b	0.0

Full Joint Distribution

too big even for this small example

$\rightsquigarrow$ omit 0.0 entries in the following (no solution for size problem!)

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

Marginalization

- full joint distribution "contains" all (joint) distributions
- What if we don't care about some variables?
- retrieve by summing out irrelevant variables $Y_1, \ldots, Y_m$:

$$P(X_1, \ldots, X_n) = \sum_{y_1, \ldots, y_m} P(X_1, \ldots, X_n, y_1, \ldots, y_m)$$

- this is called marginalization and yields marginal distribution (a subtable where some variables have been eliminated)

S	P
+s	?
-s	?

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

Marginalization

- full joint distribution "contains" all (joint) distributions
- What if we don't care about some variables?
- retrieve by summing out irrelevant variables $Y_1, \ldots, Y_m$:

$$P(X_1, \ldots, X_n) = \sum_{y_1,\ldots,y_m} P(X_1, \ldots, X_n, y_1, \ldots, y_m)$$

- this is called marginalization and yields marginal distribution (a subtable where some variables have been eliminated)

S	P
+s	0.45
-s	0.55

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

Marginalization

- full joint distribution "contains" all (joint) distributions
- What if we don't care about some variables?
- retrieve by summing out irrelevant variables $Y_1, \ldots, Y_m$:

$$P(X_1, \ldots, X_n) = \sum_{y_1,\ldots,y_m} P(X_1, \ldots, X_n, y_1, \ldots, y_m)$$

- this is called marginalization and yields marginal distribution (a subtable where some variables have been eliminated)

S	P
+s	0.45
-s	0.55

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

S	B	P
+s	+b	?
+s	-b	?
-s	+b	?
-s	-b	?

Marginalization

- full joint distribution "contains" all (joint) distributions
- What if we don't care about some variables?
- retrieve by summing out irrelevant variables $Y_1, \ldots, Y_m$:

$$P(X_1, \ldots, X_n) = \sum_{y_1,\ldots,y_m} P(X_1, \ldots, X_n, y_1, \ldots, y_m)$$

- this is called marginalization and yields marginal distribution (a subtable where some variables have been eliminated)

S	*P*
+s	0.45
-s	0.55

O	*M*	*S*	*B*	*P*
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

S	*B*	*P*
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

Uncertainty Example I Revisited

Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami? $\rightsquigarrow$ 0.45
- without salami? $\rightsquigarrow$ 0.55
- with salami and broccoli? $\rightsquigarrow$ 0.15
- with salami or broccoli? $\rightsquigarrow$ 0.7

Uncertainty Example I Revisited

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

Adam is in the pizzeria.

Today's special offer I is to get a random piece of the depicted pizza.

What are his chances to get a slice

- with salami? $\rightsquigarrow$ 0.45
- without salami? $\rightsquigarrow$ 0.55
- with salami and broccoli? $\rightsquigarrow$ 0.15
- with salami or broccoli? $\rightsquigarrow$ 0.7

$\rightsquigarrow$ full joint distribution sufficient for all (unconditional) queries

Conditional Probability

Uncertainty Example II

Adam is in the pizzeria and still undecided.

Special offer II is to name one topping and get a random piece among those with the topping.

Which topping should he name to maximize the probability to get salami and broccoli?

- onions
- mushrooms
- salami
- broccoli

Uncertainty Example II

Adam is in the pizzeria and still undecided.

Special offer II is to name one topping and get a random piece among those with the topping.

Which topping should he name to maximize the probability to get salami and broccoli?

- onions $\rightsquigarrow \frac{0}{8}$
- mushrooms
- salami
- broccoli

Uncertainty Example II

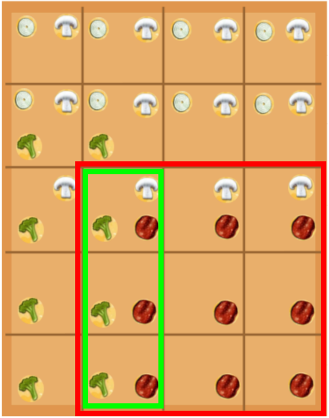
Adam is in the pizzeria and still undecided.

Special offer II is to name one topping and get a random piece among those with the topping.

Which topping should he name to maximize the probability to get salami and broccoli?

- onions $\rightsquigarrow \frac{0}{8}$
- mushrooms $\rightsquigarrow \frac{1}{12}$
- salami
- broccoli

Uncertainty Example II



Adam is in the pizzeria and still undecided.

Special offer II is to name one topping and get a random piece among those with the topping.

Which topping should he name to maximize the probability to get salami and broccoli?

- onions $\rightsquigarrow \frac{0}{8}$
- mushrooms $\rightsquigarrow \frac{1}{12}$
- salami $\rightsquigarrow \frac{1}{3}$
- broccoli

Uncertainty Example II

Adam is in the pizzeria and still undecided.

Special offer II is to name one topping and get a random piece among those with the topping.

Which topping should he name to maximize the probability to get salami and broccoli?

- onions $\rightsquigarrow \frac{0}{8}$
- mushrooms $\rightsquigarrow \frac{1}{12}$
- salami $\rightsquigarrow \frac{1}{3}$
- broccoli $\rightsquigarrow \frac{3}{8}$

Conditional Probabilities

- $P(x)$ is called prior probability of x because it assumes no additional information
- if we learn that variable Y has value y (i.e., we obtain evidence), our belief on x changes

$\rightsquigarrow$ we can update to conditional probability $P(x \mid y)$,
which is defined in terms of joint probabilities:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

Conditional Probabilities

- $P(x)$ is called prior probability of x because it assumes no additional information
- if we learn that variable Y has value y (i.e., we obtain evidence), our belief on x changes

$\rightsquigarrow$ we can update to conditional probability $P(x \mid y)$, which is defined in terms of joint probabilities:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

does this match our intuition?

- name mushrooms to get salami and broccoli: $P(+s, +b \mid +m) = \frac{1}{12}$

Conditional Probabilities

- $P(x)$ is called prior probability of x because it assumes no additional information
- if we learn that variable Y has value y (i.e., we obtain evidence), our belief on x changes

$\rightsquigarrow$ we can update to conditional probability $P(x \mid y)$, which is defined in terms of joint probabilities:

$$P(x \mid y) = \frac{P(x,y)}{P(y)}$$

does this match our intuition?

- name mushrooms to get salami and broccoli: $P(+s,+b \mid +m) = \frac{1}{12} = \frac{1/20}{12/20} = \frac{P(+s,+b,+m)}{P(+m)}$ ✓

Conditional Probabilities

- $P(x)$ is called prior probability of x because it assumes no additional information
- if we learn that variable Y has value y (i.e., we obtain evidence), our belief on x changes

⇝ we can update to conditional probability $P(x \mid y)$, which is defined in terms of joint probabilities:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

does this match our intuition?

- name mushrooms to get salami and broccoli:
  $P(+s, +b \mid +m) = \frac{1}{12} = \frac{1/20}{12/20} = \frac{P(+s,+b,+m)}{P(+m)}$ ✓
- name broccoli to get salami and broccoli:
  $P(+s, +b \mid +b) = \frac{3}{8}$

Conditional Probabilities

- $P(x)$ is called prior probability of x because it assumes no additional information
- if we learn that variable Y has value y (i.e., we obtain evidence), our belief on x changes

$\leadsto$ we can update to conditional probability $P(x \mid y)$, which is defined in terms of joint probabilities:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

does this match our intuition?

- name mushrooms to get salami and broccoli:
  $P(+s, +b \mid +m) = \frac{1}{12} = \frac{1/20}{12/20} = \frac{P(+s,+b,+m)}{P(+m)}$ ✓
- name broccoli to get salami and broccoli:
  $P(+s, +b \mid +b) = \frac{3}{8} = \frac{3/20}{8/20} = \frac{P(+s,+b,+b)}{P(+b)}$ ✓

Conditional Probabilities

- $P(x)$ is called prior probability of x because it assumes no additional information
- if we learn that variable Y has value y (i.e., we obtain evidence), our belief on x changes

⇝ we can update to conditional probability $P(x \mid y)$, which is defined in terms of joint probabilities:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

does this match our intuition?

- name mushrooms to get salami and broccoli: $P(+s, +b \mid +m) = \frac{1}{12} = \frac{1/20}{12/20} = \frac{P(+s,+b,+m)}{P(+m)}$ ✓
- name broccoli to get salami and broccoli: $P(+s, +b \mid +b) = \frac{3}{8} = \frac{3/20}{8/20} = \frac{P(+s,+b,+b)}{P(+b)}$ ✓

⇝ full joint distribution sufficient for all ~~(unconditional)~~ queries

Conditional Distributions

conditional distributions are a distribution over random variables given evidence

$P(S, B)$

S	B	P
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

$$\left.\begin{array}{c} P(S \mid +b) \\ \begin{array}{|c|c|} \hline S & P \\ \hline +s & \frac{3}{8} \\ -s & \frac{5}{8} \\ \hline \end{array} \\ \\ P(S \mid -b) \\ \begin{array}{|c|c|} \hline S & P \\ \hline +s & 0.5 \\ -s & 0.5 \\ \hline \end{array} \end{array}\right\} P(S \mid B)$$

Conditional Distributions: Normalization Trick

conditional distributions are a distribution over random variables given evidence

$P(S, B)$

S	B	P
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

$P(S, +b)$

S	B	P
+s	+b	0.15
-s	+b	0.25

1. select entries matching evidence

$P(S, -b)$

S	B	P
+s	-b	0.3
-s	-b	0.3

$P(S \mid +b)$

S	P
+s	$\frac{3}{8}$
-s	$\frac{5}{8}$

$P(S \mid -b)$

S	P
+s	0.5
-s	0.5

$P(S \mid B)$

Conditional Distributions: Normalization Trick

conditional distributions are a distribution over random variables given evidence

$P(S, B)$

S	B	P
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

$P(S, +b)$

S	B	P
+s	+b	0.15
-s	+b	0.25

$P(S, -b)$

S	B	P
+s	-b	0.3
-s	-b	0.3

1. select entries matching evidence

2. normalize selected entries

(divide each entry by sum)

$P(S \mid +b)$

S	P
+s	$\frac{3}{8}$
-s	$\frac{5}{8}$

$P(S \mid -b)$

S	P
+s	0.5
-s	0.5

$P(S \mid B)$

Conditional Distributions: Normalization Trick

conditional distributions are a distribution over random variables given evidence

$P(S, B)$

S	B	P
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

1.

$P(S, +b)$

S	B	P
+s	+b	0.15
-s	+b	0.25

2.

$P(S \mid +b)$

S	P
+s	$\frac{3}{8}$
-s	$\frac{5}{8}$

1.

$P(S, -b)$

S	B	P
+s	-b	0.3
-s	-b	0.3

2.

$P(S \mid -b)$

S	P
+s	0.5
-s	0.5

$P(S \mid B)$

1. select entries matching evidence

2. normalize selected entries

(divide each entry by sum)

Why does this work? e.g., $P(+s \mid +b) = \frac{P(+s,+b)}{P(+b)} = \frac{P(+s,+b)}{\sum_s P(s,+b)} = \frac{0.15}{0.15+0.25} = \frac{3}{8}$

Conditional Distributions: Normalization Trick

conditional distributions are a distribution over random variables given evidence

$P(S, B)$

S	B	P
+s	+b	0.15
+s	-b	0.3
-s	+b	0.25
-s	-b	0.3

$P(S, +b)$

S	B	P
+s	+b	0.15
-s	+b	0.25

$P(S, -b)$

S	B	P
+s	-b	0.3
-s	-b	0.3

1. select entries matching evidence

2. normalize selected entries
(divide each entry by sum)

$\rightsquigarrow$ write $P(S \mid B) \propto P(S, B)$
or $P(S \mid B) = \alpha P(S, B)$ if
we plan to normalize later

$P(S \mid B)$:

$P(S \mid +b)$

S	P
+s	$\frac{3}{8}$
-s	$\frac{5}{8}$

$P(S \mid -b)$

S	P
+s	0.5
-s	0.5

Why does this work? e.g., $P(+s \mid +b) = \frac{P(+s,+b)}{P(+b)} = \frac{P(+s,+b)}{\sum_s P(s,+b)} = \frac{0.15}{0.15+0.25} = \frac{3}{8}$

Product and Chain Rule

two important rules that are often used can be derived:

- definition of conditional probability:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

- reorder to obtain product rule:

$$P(x, y) = P(x \mid y) \cdot P(y)$$

- recursive application yields chain rule:

$$\begin{aligned} P(x_1, \ldots, x_n) &= P(x_n \mid x_1, \ldots, x_{n-1}) \cdot P(x_1, \ldots, x_{n-1}) \\ &= P(x_n \mid x_1, \ldots, x_{n-1}) \cdot P(x_{n-1} \mid x_1, \ldots, x_{n-2}) \cdot P(x_1, \ldots, x_{n-2}) \\ &= P(x_n \mid x_1, \ldots, x_{n-1}) \cdot P(x_{n-1} \mid x_1, \ldots, x_{n-2}) \cdot \ldots \cdot P(x_2 \mid x_1) \cdot P(x_1) \\ &= \prod_{i=1}^{n} P(x_i \mid x_1, \ldots, x_{i-1}) \end{aligned}$$

Probabilistic Inference

Probabilistic Inference

- probabilistic inference: compute a desired probability from other known probabilities (e.g., conditional from joint)
- we usually compute conditional probabilities that represent the agent's belief given the evidence
- probabilities change with new evidence, i.e., the agent's belief is updated

Probabilistic Inference

- probabilistic inference: compute a desired probability from other known probabilities (e.g., conditional from joint)
- we usually compute conditional probabilities that represent the agent's belief given the evidence
- probabilities change with new evidence, i.e., the agent's belief is updated

probabilistic inference example:

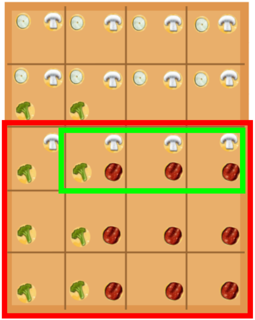
- Adam's initial belief to get mushroom and salami with special offer I is $P(+m, +s) = 0.15$

Probabilistic Inference

- probabilistic inference: compute a desired probability from other known probabilities (e.g., conditional from joint)
- we usually compute conditional probabilities that represent the agent's belief given the evidence
- probabilities change with new evidence, i.e., the agent's belief is updated

probabilistic inference example:

- Adam's initial belief to get mushroom and salami with special offer I is $P(+m, +s) = 0.15$
- Adam learns that the used onions were bad and cannot be sold $\rightsquigarrow$ $P(+m, +s \mid -o) = 0.25$

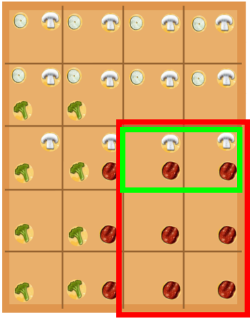


Probabilistic Inference

- probabilistic inference: compute a desired probability from other known probabilities (e.g., conditional from joint)
- we usually compute conditional probabilities that represent the agent's belief given the evidence
- probabilities change with new evidence, i.e., the agent's belief is updated

probabilistic inference example:

- Adam's initial belief to get mushroom and salami with special offer I is $P(+m, +s) = 0.15$
- Adam learns that the used onions were bad and cannot be sold $\rightsquigarrow$ $P(+m, +s \mid -o) = 0.25$
- the neighbor table orders all remaining pieces with broccoli $\rightsquigarrow$ $P(+m, +s \mid -o, -b) = \frac{1}{3}$
- ...



Probabilistic Inference

- probabilistic inference: compute a desired probability from other known probabilities (e.g., conditional from joint)
- we usually compute conditional probabilities that represent the agent's belief given the evidence
- probabilities change with new evidence, i.e., the agent's belief is updated

we consider three probabilistic inference methods:

- using the full joint distribution
- using Bayes' rule
- using Bayesian networks

Probabilistic Inference Using Full Joint Distribution

How can we answer a query from the full joint distribution?

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

- given: full joint distribution $P(O, M, S, B)$ and query $P(S, B \mid m)$

Probabilistic Inference Using Full Joint Distribution

How can we answer a query from the full joint distribution?

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

- given: full joint distribution $P(O, M, S, B)$ and query $P(S, B \mid m)$
- step 1: partition variables in query, evidence and hidden variables

Probabilistic Inference Using Full Joint Distribution

How can we answer a query from the full joint distribution?

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05

- given: full joint distribution $P(O, M, S, B)$ and query $P(S, B \mid m)$
- step 1: partition variables in query, evidence and hidden variables
- step 2: select entries consistent with evidence

Probabilistic Inference Using Full Joint Distribution

How can we answer a query from the full joint distribution?

O	*M*	*S*	*B*	*P*
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

O	*M*	*S*	*B*	*P*
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05

M	*S*	*B*	*P*
+m	-s	+b	0.15
+m	-s	-b	0.3
+m	+s	+b	0.05
+m	+s	-b	0.1

- given: full joint distribution $P(O, M, S, B)$ and query $P(S, B \mid m)$
- step 1: partition variables in query, evidence and hidden variables
- step 2: select entries consistent with evidence
- step 3: sum out hidden variables

Probabilistic Inference Using Full Joint Distribution

How can we answer a query from the full joint distribution?

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05
-o	-m	+s	+b	0.1
-o	-m	+s	-b	0.2
-o	-m	-s	+b	0.1

O	M	S	B	P
+o	+m	-s	+b	0.1
+o	+m	-s	-b	0.3
-o	+m	+s	+b	0.05
-o	+m	+s	-b	0.1
-o	+m	-s	+b	0.05

M	S	B	P
+m	-s	+b	0.15
+m	-s	-b	0.3
+m	+s	+b	0.05
+m	+s	-b	0.1

S	B	P
-s	+b	0.25
-s	-b	0.5
+s	+b	$\frac{1}{12}$
+s	-b	$\frac{1}{6}$

- given: full joint distribution $P(O, M, S, B)$ and query $P(S, B \mid m)$
- step 1: partition variables in query, evidence and hidden variables
- step 2: select entries consistent with evidence
- step 3: sum out hidden variables
- step 4: normalize (e.g., $P(+s, +b \mid +m) = \frac{P(+s,+b,+m)}{P(+m)} = \frac{0.15}{0.15+0.3+0.05+0.1} = \frac{1}{4}$)

Bayes' Rule

- recall product rule:

$$P(x, y) = P(x \mid y) \cdot P(y)$$

- x and y are symmetrical in $P(x, y)$:

$$P(x \mid y) \cdot P(y) = P(x, y) = P(y \mid x) \cdot P(x)$$

- from this, we can derive Bayes' rule (or Bayes' theorem):

$$P(x \mid y) = \frac{P(y \mid x) \cdot P(x)}{P(y)}$$

- allows to compute conditional probability from its reverse

Bayes' Rule

- recall product rule:

$$P(x, y) = P(x \mid y) \cdot P(y)$$

- x and y are symmetrical in $P(x, y)$:

$$P(x \mid y) \cdot P(y) = P(x, y) = P(y \mid x) \cdot P(x)$$

- from this, we can derive Bayes' rule (or Bayes' theorem):

$$P(x \mid y) = \frac{P(y \mid x) \cdot P(x)}{P(y)}$$

- allows to compute conditional probability from its reverse

Why is this useful?

$\rightsquigarrow$ often hard to obtain one conditional while the reverse is simple

Probabilistic Inference With Bayes' Rule

a typical use case for Bayes' rule is medical diagnosis, where

$$P(\text{illness} \mid \text{symptom}) = \frac{P(\text{symptom} \mid \text{illness}) \cdot P(\text{illness})}{P(\text{symptom})}$$

Probabilistic Inference With Bayes' Rule

a typical use case for Bayes' rule is medical diagnosis, where

$$P(\text{illness} \mid \text{symptom}) = \frac{P(\text{symptom} \mid \text{illness}) \cdot P(\text{illness})}{P(\text{symptom})}$$

example

- variables M (meningitis) and S (stiff neck)
- (statistical) evidence:
  - 1 in 10000 patients suffers from meningitis $\rightsquigarrow P(+m) = 0.0001$
  - 80% of meningitis patients suffer from a stiff neck $\rightsquigarrow P(+s \mid +m) = 0.8$
  - 1 in 100 patients have a stiff neck but no meningitis $\rightsquigarrow P(+s \mid -m) = 0.01$

$$P(+m \mid +s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m)+P(+s|-m)P(-m)} \approx 0.008$$

$\rightsquigarrow$ only 0.8% of patients with stiff neck have meningitis