Quantifying Uncertainty
Advanced Artificial Intelligence: Lecture

Heriberto Cuayáhuitl
http://staff.lincoln.ac.uk/hcuayahuitl

School of Computer Science, University of Lincoln

2 October 2019

About Me

- Dr. Heriberto Cuayáhuitl
- PhD in informatics, University of Edinburgh, UK
- Research Interests:
  - Machine Intelligence
  - Dialogue Systems
  - Interactive Robots
- Surgery Hours: Wednesday 13:00-14:00, or arrange meeting
- Office Location: Isaac Newton Building, room INB3130
- Email: HCuayahuitl@lincoln.ac.uk

Overview

Topics:

- Introduction to Probability
- Probabilistic Reasoning

Readings:

1. Russel, S.& Norvig, P. (2013). "Artificial Intelligence: A Modern Approach", **chapter13** (main reading)
2. Bishop, C. "Model-Based Machine Learning", **chapter1** (optional reading)

Probability Basics

- Let Ω be the **sample space**, e.g., 6 possible rolls of a die, and $\omega \in \Omega$ a **sample point** or **possible world** or **atomic event**

Probability Basics

- Let Ω be the **sample space**, e.g., 6 possible rolls of a die, and $\omega \in \Omega$ a **sample point** or **possible world** or **atomic event**

- A **probability space** or **probability model** is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t. $0 \leq P(\omega) \leq 1$ $\sum_\omega P(\omega) = 1$, e.g. $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

Probability Basics

- Let Ω be the **sample space**, e.g., 6 possible rolls of a die, and $\omega \in \Omega$ a **sample point** or **possible world** or **atomic event**

- A **probability space** or **probability model** is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t. $0 \leq P(\omega) \leq 1$ $\sum_\omega P(\omega){=}1$, e.g. $P(1){=}P(2){=}P(3){=}P(4){=}P(5){=}P(6){=}1/6$

- An **event** A is any subset of Ω, where $P(A){=}\sum_{\omega \in A} P(\omega)$ e.g. $P(\text{die roll} < 4) = P(1){+}P(2){+}P(3) = 1/6{+}1/6{+}1/6 = 1/2$

Probability Basics

- Let Ω be the **sample space**, e.g., 6 possible rolls of a die, and $\omega \in \Omega$ a **sample point** or **possible world** or **atomic event**

- A **probability space** or **probability model** is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t. $0 \leq P(\omega) \leq 1$ $\sum_{\omega} P(\omega){=}1$, e.g. $P(1){=}P(2){=}P(3){=}P(4){=}P(5){=}P(6){=}1/6$

- An **event** A is any subset of Ω, where $P(A){=}\sum_{\omega \in A} P(\omega)$ e.g. $P(\text{die roll} < 4) = P(1)+P(2)+P(3) = 1/6+1/6+1/6 = 1/2$

- Variables in probability theory are called **random variables**, their names begin with an uppercase letter (e.g. $Die_1{=}\{1, ..., 6\}$, and have domain values (e.g. $\{true, false\}$)

Propositions

- Think of a proposition as an event (set of sample points) where the proposition is true

Propositions

- Think of a proposition as an event (set of sample points) where the proposition is true
- Given Boolean random variables A and B:
  - event $a =$ set of sample points where $A(\omega) = \mathit{true}$
  - event $\neg a =$ set of sample points where $A(\omega) = \mathit{false}$
  - event $a \wedge b =$ points where $A(\omega) = \mathit{true}$ and $B(\omega) = \mathit{true}$

Propositions

- Think of a proposition as an event (set of sample points) where the proposition is true
- Given Boolean random variables A and B:
  - event a = set of sample points where $A(\omega) = true$
  - event $\neg a$ = set of sample points where $A(\omega) = false$
  - event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$
- Sample points are *defined* by the values of random variables, i.e. the sample space is the Cartesian product of their ranges

Propositions

- Think of a proposition as an event (set of sample points) where the proposition is true
- Given Boolean random variables A and B:
  - event a = set of sample points where $A(\omega) = \mathit{true}$
  - event $\neg a$ = set of sample points where $A(\omega) = \mathit{false}$
  - event $a \wedge b$ = points where $A(\omega) = \mathit{true}$ and $B(\omega) = \mathit{true}$
- Sample points are *defined* by the values of random variables, i.e. the sample space is the Cartesian product of their ranges
- Proposition = disjunction of atomic events in which it is true, e.g. $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
  $\implies P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Syntax for Propositions

- **Boolean** random variables
  e.g., *Cavity* (do I have a cavity?)
  $Cavity = true$ is a proposition, also written *cavity*

Syntax for Propositions

- **Boolean** random variables
  e.g., *Cavity* (do I have a cavity?)
  $Cavity = true$ is a proposition, also written *cavity*

- **Discrete** random variables (*finite* or *infinite*)
  e.g., *Weather* is one of $< sunny, rain, cloudy, snow >$
  $Weather = rain$ is a proposition
  Values must be exhaustive and mutually exclusive

Syntax for Propositions

- **Boolean** random variables
  e.g., *Cavity* (do I have a cavity?)
  $Cavity = true$ is a proposition, also written *cavity*

- **Discrete** random variables (*finite* or *infinite*)
  e.g., *Weather* is one of $< sunny, rain, cloudy, snow >$
  $Weather = rain$ is a proposition
  Values must be exhaustive and mutually exclusive

- **Continuous** random variables (*bounded* or *unbounded*)
  e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Prior Probability and (Joint) Probability Distribution

- **Prior** or **unconditional probabilities** of propositions, e.g. $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$ correspond to belief prior to arrival of any (new) evidence

[1] Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Prior Probability and (Joint) Probability Distribution

- **Prior** or **unconditional probabilities** of propositions, e.g. $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$ correspond to belief prior to arrival of any (new) evidence

- **Probability distribution** gives values for all possible assignments: $\mathbf{P}(Weather) =< 0.72, 0.1, 0.08, 0.1 >$ (*normalised*, i.e. sums to 1)

[1] Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Prior Probability and (Joint) Probability Distribution

- **Prior** or **unconditional probabilities** of propositions, e.g. $P(Cavity = true) = 0.1$ and $P(Weather = sunny) = 0.72$ correspond to belief prior to arrival of any (new) evidence

- **Probability distribution** gives values for all possible assignments: $\mathbf{P}(Weather) =< 0.72, 0.1, 0.08, 0.1 >$ (*normalised*, i.e. sums to 1)

- **Joint probability distribution**[1] gives the probability of every atomic event on a set of random variables (i.e. every sample point), e.g. $\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

Weather =	sunny	rainy	cloudy	snowy
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

[1]Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Conditional Probability (1/3)

- **Conditional** or **posterior probabilities** e.g., $P(cavity|toothache) = 0.8$, i.e. *given that toothache is all I know* ... *NOT* "if *toothache* then 80% chance of *cavity*"

Conditional Probability (1/3)

- **Conditional** or **posterior probabilities** e.g., $P(cavity|toothache) = 0.8$, i.e. *given that toothache is all I know* ... *NOT* "if *toothache* then 80% chance of *cavity*"

- Notation for conditional distributions: $\mathbf{P}(Cavity|toothache) =$ $< P(cavity|toothache), P(\neg cavity|toothache) >$

Conditional Probability (1/3)

- **Conditional** or **posterior probabilities** e.g., $P(cavity|toothache) = 0.8$, i.e. *given that toothache is all I know* ... *NOT* "if *toothache* then 80% chance of *cavity*"

- Notation for conditional distributions:
$\mathbf{P}(Cavity|toothache) =$
$< P(cavity|toothache), P(\neg cavity|toothache) >$

- New evidence may be irrelevant, allowing simplification, e.g. $P(cavity|toothache, Weather) = P(cavity|toothache) = 0.8$

Conditional Probability (2/3)

- Conditional probability: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$

Conditional Probability (2/3)

- Conditional probability: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$
- The **product rule** gives an alternative formulation: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

Conditional Probability (2/3)

- Conditional probability: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$ if $P(b) \neq 0$
- The **product rule** gives an alternative formulation:
  $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$
- A general version holds for whole distributions, e.g.,
  $\mathbf{P}(\mathit{Weather}, \mathit{Cavity}) = \mathbf{P}(\mathit{Weather}|\mathit{Cavity})\mathbf{P}(\mathit{Cavity})$
  (View as a 4×2 set of equations, *not* matrix multiplication)

$$
\begin{aligned}
&P(W = \mathit{sunny} \wedge C = \mathit{true}) = P(W = \mathit{sunny}|C = \mathit{true})\, P(C = \mathit{true})\\
&P(W = \mathit{rain} \wedge C = \mathit{true}) = P(W = \mathit{rain}|C = \mathit{true})\, P(C = \mathit{true})\\
&P(W = \mathit{cloudy} \wedge C = \mathit{true}) = P(W = \mathit{cloudy}|C = \mathit{true})\, P(C = \mathit{true})\\
&P(W = \mathit{snow} \wedge C = \mathit{true}) = P(W = \mathit{snow}|C = \mathit{true})\, P(C = \mathit{true})\\
&P(W = \mathit{sunny} \wedge C = \mathit{false}) = P(W = \mathit{sunny}|C = \mathit{false})\, P(C = \mathit{false})\\
&P(W = \mathit{rain} \wedge C = \mathit{false}) = P(W = \mathit{rain}|C = \mathit{false})\, P(C = \mathit{false})\\
&P(W = \mathit{cloudy} \wedge C = \mathit{false}) = P(W = \mathit{cloudy}|C = \mathit{false})\, P(C = \mathit{false})\\
&P(W = \mathit{snow} \wedge C = \mathit{false}) = P(W = \mathit{snow}|C = \mathit{false})\, P(C = \mathit{false})\,.
\end{aligned}
$$

Conditional Probability (3/3)

- The **chain rule** is derived by the successive application of the product rule:
  - $P(X_1, X_2) = P(X_2|X_1)P(X_1)$
  - $P(X_1, X_2, X_3) = P(X_3|X_2, X_1)P(X_2, X_1)$
  - $\qquad = P(X_3|X_2, X_1)P(X_2|X_1)P(X_1)$
  - $P(X_1, X_2, X_3, X_4) = ?$
  - ...
  - $P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1)\ldots$
  - $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_{i-1}, \ldots, X_1)$

Conditional Probability (3/3)

- The **chain rule** is derived by the successive application of the product rule:
  - $P(X_1, X_2) = P(X_2|X_1)P(X_1)$
  - $P(X_1, X_2, X_3) = P(X_3|X_2, X_1)P(X_2, X_1)$
  - $\qquad\qquad\qquad = P(X_3|X_2, X_1)P(X_2|X_1)P(X_1)$
  - $P(X_1, X_2, X_3, X_4) = ?$
  - ...
  - $P(X_1, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_3|X_2, X_1)\ldots$
  - $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_{i-1}, \ldots, X_1)$

- **Bayes Rule**: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

$$= \frac{P(B|A)P(A)}{\sum_i^n P(B|A_i)P(A_i)}$$

Marginal Probability

- Marginalisation is the process of summing up the probabilities of the other variables—i.e. taking them out of the equation.

Marginal Probability

- Marginalisation is the process of summing up the probabilities of the other variables—i.e. taking them out of the equation.
- General rule for any set of variables Y and Z:
  $\mathbf{P}(Y) = \sum_{z in Z} \mathbf{P}(Y, z)$

Marginal Probability

- Marginalisation is the process of summing up the probabilities of the other variables—i.e. taking them out of the equation.
- General rule for any set of variables Y and Z:
  $\mathbf{P}(\mathrm{Y}) = \sum_{\mathrm{z} in \mathrm{Z}} \mathbf{P}(\mathrm{Y}, \mathrm{z})$
- For example: $\mathbf{P}(Cavity) = \sum_{\mathrm{z} \in \{Catch, Toothache\}} \mathbf{P}(Cavity, \mathrm{z})$

Marginal Probability

- Marginalisation is the process of summing up the probabilities of the other variables—i.e. taking them out of the equation.
- General rule for any set of variables Y and Z:
  $\mathbf{P}(Y) = \sum_{zinZ} \mathbf{P}(Y, z)$
- For example: $\mathbf{P}(Cavity) = \sum_{z \in \{Catch, Toothache\}} \mathbf{P}(Cavity, z)$
- Consider the following joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Inference by Enumeration (1/3)

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by Enumeration (2/3)

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

$P(cavity \vee toothache) =$
$0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$

Inference by Enumeration (3/3)

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

We can also compute conditional probabilities:

$$
\begin{aligned}
P(\neg cavity | toothache) &= \frac{P(\neg cavity \wedge toothache)}{P(toothache)} \\
&= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4
\end{aligned}
$$

Normalisation

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

The denominator can be viewed as a *normalization constant* α

$$\mathbf{P}(Cavity|toothache) = \alpha\, \mathbf{P}(Cavity, toothache)$$
$$= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg catch)]$$
$$= \alpha\, [< 0.108, 0.016 > + < 0.012, 0.064 >]$$
$$= \alpha\ < 0.12, 0.08 >=< 0.6, 0.4 >$$

General idea: compute distribution on query variable by fixing **evidence variables** and summing over **hidden variables**

Keywords

- **Probability Distribution**
- **Joint Probability**
- **Conditional Probability**
- **Prior Probability**
- **Posterior Probability**
- **Product Rule**
- **Sum Rule**
- **Chain Rule**
- **Marginalisation**
- **Normalisation**
- **Inference**
- **Bayes Rule**
- **Probabilistic Model**

Next

Workshop:

- Excercises about quantifying uncertainty, Q&A
- Wednesday from 10:30-12hrs, room=INB1301

Readings:

1. Russel, S.& Norvig, P. (2013). "Artificial Intelligence: A Modern Approach", **chapter13** (main reading)
2. Bishop, C. "Model-Based Machine Learning", **chapter1** (optional reading)