CS 1571 Introduction to AI
Lecture 19

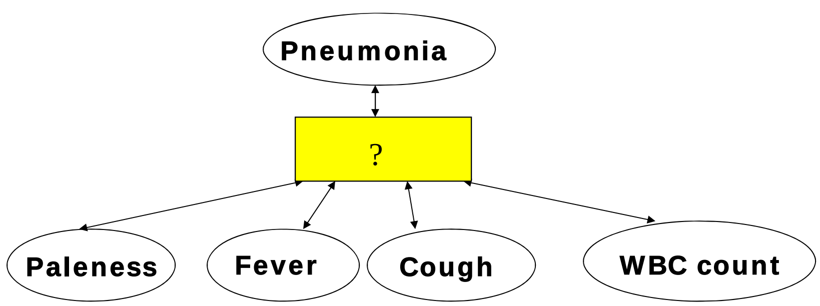
Modeling uncertainty with probabilities

Milos Hauskrecht
milos@cs.pitt.edu
5329 Sennott Square

Modeling the uncertainty.

Key challenges:

- How to represent the relations in the presence of uncertainty?
- How to manipulate such knowledge to make inferences?
  - **Humans can reason with uncertainty.**



Methods for representing uncertainty

Probability theory

- A well defined theory for modeling and reasoning in the presence of uncertainty
- A natural choice to replace certainty factors

Facts (propositional statements)

- Are represented via **random variables** with two or more values

  Example: *Pneumonia* is a random variable

  values: ***True*** **and** ***False***
- Each value can be achieved **with some probability:**

$$P(Pneumonia = True) = 0.001$$

$$P(WBCcount = high) = 0.005$$

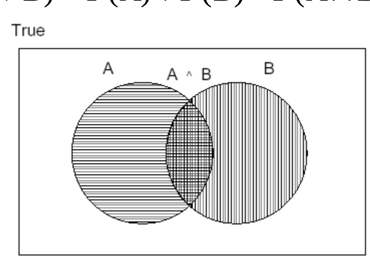
Probability theory

- Well-defined theory for representing and manipulating statements with uncertainty

Axioms of probability:

For any two propositions A, B.

1. $0 \leq P(A) \leq 1$
2. $P(True) = 1$ and $P(False) = 0$
3. $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Methods for representing uncertainty

Probabilistic extension of propositional logic

- **Propositions:**
  - statements about the world
  - Statements are represented by the assignment of values to **random variables**
- **Random variables:**

! – **Boolean** $Pneumonia$ is either $True, False$

Random variable **Values**

! – **Multi-valued** $Pain$ is one of $\{Nopain, Mild, Moderate, Severe\}$

Random variable **Values**

– **Continuous** $HeartRate$ is a value in $<0;180>$

Random variable **Values**

Probabilities

Unconditional probabilities (prior probabilities)

$P(Pneumonia) = 0.001$ or $P(Pneumonia = True) = 0.001$

$P(Pneumonia = False) = 0.999$

$P(WBCcount = high) = 0.005$

Probability distribution

- Defines probabilities **for all possible value assignments to a random variable**
- Values are mutually exclusive

$P(Pneumonia = True) = 0.001$
$P(Pneumonia = False) = 0.999$

$Pneumonia$	$\mathbf{P}(Pneumonia)$
$True$	0.001
$False$	0.999

Probability distribution

Defines probability for **all possible value assignments**

Example 1:

$P(Pneumonia = True) = 0.001$
$P(Pneumonia = False) = 0.999$

$Pneumonia$	$\mathbf{P}(Pneumonia)$
$True$	0.001
$False$	0.999

$$P(Pneumonia = True) + P(Pneumonia = False) = 1$$

Probabilities sum to 1 !!!

Example 2:

$P(WBCcount = high) = 0.005$
$P(WBCcount = normal) = 0.993$
$P(WBCcount = high) = 0.002$

$WBCcount$	$\mathbf{P}(WBCcount)$
$high$	0.005
$normal$	0.993
low	0.002

Joint probability distribution

Joint probability distribution (for a set variables)

- Defines probabilities for **all possible assignments of values to variables in the set**

Example: variables *Pneumonia* and *WBCcount*

$\mathbf{P}(pneumonia, WBCcount)$

Is represented by 2×3 array(matrix)

$Pneumonia$ \ $WBCcount$	$high$	$normal$	low
$True$	0.0008	0.0001	0.0001
$False$	0.0042	0.9929	0.0019

Joint probability distribution

Joint probability distribution (for a set variables)

- Defines probabilities for **all possible assignments of values to variables in the set**

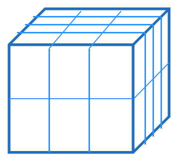
Example 2: Assume variables:

Pneumonia (2 values)

WBCcount (3 values)

Pain (4 values)

$\mathbf{P}(pneumonia, WBCcount, Pain)$ is represented by $2 \times 3 \times 4$ array



Example of an entry in the array

$P(pneumonia = T, WBCcount = high, Pain = severe)$

Joint probabilities: marginalization

Marginalization

- reduces the dimension of the joint distribution
- Sums variables out

$\mathbf{P}(pneumonia, WBCcount)$ 2×3 matrix

WBCcount

		high	*normal*	*low*	$\mathbf{P}(Pneumonia)$
Pneumonia	*True*	0.0008	0.0001	0.0001	0.001
	False	0.0042	0.9929	0.0019	0.999
$\mathbf{P}(WBCcount)$		0.005	0.993	0.002	

Marginalization (here summing of columns or rows)

Marginalization

Marginalization

- reduces the dimension of the joint distribution

$$P(X_1, X_2, \ldots X_{n-1}) = \sum_{\{X_n\}} P(X_1, X_2, \ldots X_{n-1}, X_n)$$

- We can continue doing this

$$P(X_1, \ldots X_{n-2}) = \sum_{\{X_{n-1}, X_n\}} P(X_1, X_2, \ldots X_{n-1}, X_n)$$

What is the maximal joint probability distribution?

- **Full joint probability**

Full joint distribution

- **the joint distribution for all variables in the problem**
  - It defines the complete probability model for the problem

Example: pneumonia diagnosis

- **Variables:** *Pneumonia, Fever, Paleness, WBCcount, Cough*
- Full joint probability: P(*Pneumonia, Fever, Paleness, WBCcount, Cough)*
  - defines the probability for all possible assignments of values to these variables

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = T)$$
$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = F)$$
$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = F, Paleness = T)$$

$\ldots$ etc

- **How many probabilities are there?**

Full joint distribution

- **the joint distribution for all variables in the problem**
  - It defines the complete probability model for the problem

Example: pneumonia diagnosis

Variables: *Pneumonia, Fever, Paleness, WBCcount, Cough*

Full joint probability: P(*Pneumonia, Fever, Paleness, WBCcount, Cough)*
  - defines the probability for all possible assignments of values to these variables

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = T)$$

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = T, Paleness = F)$$

$$P(Pneumonia = T, WBCcount = High, Fever = T, Cough = F, Paleness = T)$$

$\ldots$ etc

- **How many probabilities are there?**
- Exponential in the number of variables

Full joint distribution

- **Any joint probability over a subset of variables can be obtained via marginalization**

$$P(Pneumonia, WBCcount, Fever) = \sum_{c,p=\{T,F\}} P(Pneumonia, WBCcount, Fever, Cough = c, Paleness = p)$$

- **Is it possible to recover the full joint from the joint probabilities over a subset of variables?**

Joint probabilities

- **Is it possible to recover the full joint from the joint probabilities over a subset of variables?**

$\mathbf{P}(pneumonia, WBCcount)$ 2×3 matrix

		WBCcount			$\mathbf{P}(Pneumonia)$
		high	*normal*	*low*	
Pneumonia	*True*	?	?	?	0.001
	False	?	?	?	0.999
$\mathbf{P}(WBCcount)$		0.005	0.993	0.002	

Joint probabilities and independence

- **Is it possible to recover the full joint from the joint probabilities over a subset of variables?**
- Only if the variables are independent !!!

$\mathbf{P}(pneumonia, WBCcount)$ 2×3 matrix

		WBCcount			$\mathbf{P}(Pneumonia)$
		high	*normal*	*low*	
Pneumonia	*True*	?	?	?	0.001
	False	?	?	?	0.999
$\mathbf{P}(WBCcount)$		0.005	0.993	0.002	

Variable independence

- **The two events A, B are said to be independent if:**

  $P(A, B) = P(A)P(B)$
- **The variables X, Y are said to be independent if their joint can be expressed as a product of marginals:**

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

Conditional probabilities

- **Conditional probability distribution**

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

Conditional probabilities

- **Conditional probability distribution**

$$P(A \mid B) = \frac{P(A,B)}{P(B)} \quad \text{s.t. } P(B) \neq 0$$

- **Product rule.** Join probability can be expressed in terms of conditional probabilities

$$P(A,B) = P(A \mid B)P(B)$$

Conditional probabilities

- **Conditional probability distribution**

$$P(A \mid B) = \frac{P(A,B)}{P(B)} \quad \text{s.t. } P(B) \neq 0$$

- **Product rule.** Join probability can be expressed in terms of conditional probabilities

$$P(A,B) = P(A \mid B)P(B)$$

- **Chain rule**. Any joint probability can be expressed as a product of conditionals

$$P(X_1, X_2, \ldots X_n) = P(X_n \mid X_{1,} \ldots X_{n-1})P(X_{1,} \ldots X_{n-1})$$

$$= P(X_n \mid X_{1,} \ldots X_{n-1})P(X_{n-1} \mid X_{1,} \ldots X_{n-2})P(X_{1,} \ldots X_{n-2})$$

$$= \prod_{i=1}^{n} P(X_i \mid X_{1,} \ldots X_{i-1})$$

Conditional probabilities

Conditional probability

- Is defined in terms of the joint probability:

$$P(A\mid B)=\frac{P(A,B)}{P(B)} \quad \text{s.t. } P(B)\neq 0$$

- **Example:**

$$P(pneumonia=true\mid WBCcount=high)=\frac{P(pneumonia=true,WBCcount=high)}{P(WBCcount=high)}$$

$$P(pneumonia=false\mid WBCcount=high)=\frac{P(pneumonia=false,WBCcount=high)}{P(WBCcount=high)}$$

Conditional probabilities

Conditional probability distribution

- Defines probabilities for all possible assignments, given a fixed assignment to some other variable values

$P(Pneumonia=true\mid WBCcount=high)$

$\mathbf{P}(Pneumonia\mid WBCcount)$ 3 element vector of 2 elements

Pneumonia

	True	*False*	
high	0.08	0.92	1.0
normal	0.0001	0.9999	1.0
low	0.0001	0.9999	1.0

WBCcount ← Variable we condition on

$$P(Pneumonia=true\mid WBCcount=high) + P(Pneumonia=false\mid WBCcount=high)$$

Bayes rule

Conditional probability.

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

Bayes rule:

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

When is it useful?

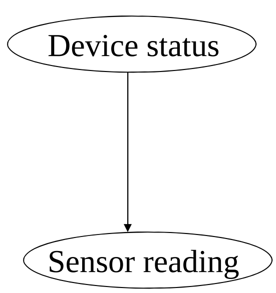
- When we are interested in computing the diagnostic query from the causal probability

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

- **Reason:** It is often easier to assess causal probability
  - E.g. Probability of pneumonia causing fever
    vs. probability of pneumonia given fever

Bayes Rule in a simple diagnostic inference

- **Device** (equipment) operating *normally* or *malfunctioning*.
  - Operation of the device sensed indirectly via a sensor
- **Sensor reading** is either *High* or *Low*



P(Device status)

normal	malfunctioning
0.9	0.1

P(Sensor reading| Device status)

Status\Sensor	High	Low
normal	0.1	0.9
malfunc	0.6	0.4

Bayes Rule in a simple diagnostic inference.

- **Diagnostic inference:** compute the probability of device operating normally or malfunctioning given a sensor reading

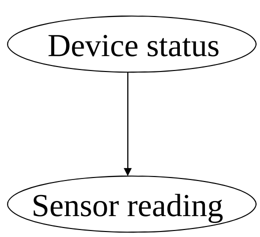
$$\mathbf{P}(\text{Device status} \mid \text{Sensor reading} = \textit{high}) = ?$$

$$= \begin{pmatrix} P(\text{Device status} = \textit{normal} \mid \text{Sensor reading} = \textit{high}) \\ P(\text{Device status} = \textit{malfunctioning} \mid \text{Sensor reading} = \textit{high}) \end{pmatrix}$$

- Note that typically the opposite conditional probabilities are given to us: they are much easier to estimate
- **Solution:** apply **Bayes rule** to reverse the conditioning variables

Bayes Rule in a simple diagnostic inference

- **Device** (equipment) operating *normally* or *malfunctioning*.
  - Operation of the device sensed indirectly via a sensor
- **Sensor reading** is either *High* or *Low*



P(Device status)

normal	malfunctioning
0.9	0.1

P(Sensor reading| Device status)

Status\Sensor	High	Low
normal	0.1	0.9
malfunc	0.6	0.4

$$\mathbf{P}(\text{Device status} \mid \text{Sensor reading} = \textit{high}) = ?$$

Bayes rule

Assume a variable A with multiple values $a_1, a_2, \dots a_k$

Bayes rule can be rewritten as:

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

$$= \frac{P(B = b \mid A = a_j) P(A = a_j)}{\sum_{i=1}^{k} P(B = b \mid A = a_j) P(A = a_j)}$$

Used in practice when we want to compute:

$\mathbf{P}(A \mid B = b)$ for all values of $a_1, a_2, \dots a_k$

Probabilistic inference

Various inference tasks:

- **Diagnostic task. (from effect to cause)**

$$\mathbf{P}(Pneumonia \mid Fever = T)$$

- **Prediction task. (from cause to effect)**

$$\mathbf{P}(Fever \mid Pneumonia = T)$$

- **Other probabilistic queries (**queries on joint distributions).

$$\mathbf{P}(Fever)$$

$$\mathbf{P}(Fever, ChestPain)$$

Inference

Any query can be computed from the full joint distribution !!!

- **Joint over a subset of variables** is obtained through marginalization

$$P(A=a, C=c) = \sum_i \sum_j P(A=a, B=b_i, C=c, D=d_j)$$

- **Conditional probability over set of variables**, given other variables' values is obtained through marginalization and definition of conditionals

$$P(D=d \mid A=a, C=c) = \frac{P(A=a, C=c, D=d)}{P(A=a, C=c)}$$

$$= \frac{\sum_i P(A=a, B=b_i, C=c, D=d)}{\sum_i \sum_j P(A=a, B=b_i, C=c, D=d_j)}$$

Inference

Any query can be computed from the full joint distribution !!!

- Any joint probability can be expressed as a product of conditionals via the **chain rule**.

$$P(X_1, X_2, \ldots X_n) = P(X_n \mid X_{1,} \ldots X_{n-1}) P(X_{1,} \ldots X_{n-1})$$

$$= P(X_n \mid X_{1,} \ldots X_{n-1}) P(X_{n-1} \mid X_{1,} \ldots X_{n-2}) P(X_{1,} \ldots X_{n-2})$$

$$= \prod_{i=1}^{n} P(X_i \mid X_{1,} \ldots X_{i-1})$$

- Sometimes it is easier to define the distribution in terms of conditional probabilities:
  - E.g. $\mathbf{P}(Fever \mid Pneumonia = T)$
    $\mathbf{P}(Fever \mid Pneumonia = F)$

Modeling uncertainty with probabilities

- Defining the **full joint distribution** makes it possible to represent and reason with uncertainty in a uniform way
- We are able to handle an arbitrary inference problem

Problems:

- **Space complexity.** To store a full joint distribution we need to remember $O(d^n)$ numbers.

  n – number of random variables, d – number of values
- **Inference (time) complexity.** To compute some queries requires $O(d^n)$ steps.
- **Acquisition problem.** Who is going to define all of the probability entries?

Medical diagnosis example

- **Space complexity**.
  - Pneumonia (2 values: T,F), Fever (2: T,F), Cough (2: T,F), WBCcount (3: high, normal, low), paleness (2: T,F)
  - Number of assignments: 2*2*2*3*2=48
  - We need to define at least 47 probabilities.
- **Time complexity**.
  - Assume we need to compute the marginal of Pneumonia=T from the full joint

$$P(Pneumonia = T) =$$
$$= \sum_{i \in T,F} \sum_{j \in T,F} \sum_{k=h,n,l} \sum_{u \in T,F} P(Fever = i, Cough = j, WBCcount = k, Pale = u)$$

  - Sum over: 2*2*3*2=24 combinations

Modeling uncertainty with probabilities

- **Knowledge based system era (70s – early 80's)**
  - **Extensional non-probabilistic models**
  - Solve the space, time and acquisition bottlenecks in probability-based models
  - froze the development and advancement of KB systems and contributed to the slow-down of AI in 80s in general

- Breakthrough (late 80s, beginning of 90s)
  - **Bayesian belief networks**
    - Give solutions to the space, acquisition bottlenecks
    - Partial solutions for time complexities
- Bayesian belief network

Bayesian belief networks (BBNs)

Bayesian belief networks.

- Represent the full joint distribution over the variables more compactly with a **smaller number of parameters**.
- Take advantage of **conditional and marginal independences** among random variables

- **A and B are independent**

$$P(A, B) = P(A)P(B)$$

- **A and B are conditionally independent given C**

$$P(A, B \mid C) = P(A \mid C)P(B \mid C)$$
$$P(A \mid C, B) = P(A \mid C)$$