# CS 1571 Introduction to AI
# Lecture 20

# Uncertainty.
# Bayesian belief networks.

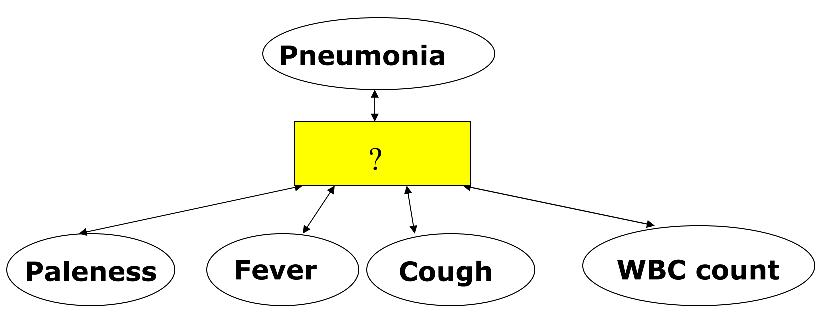
**Milos Hauskrecht**
milos@cs.pitt.edu
5329 Sennott Square



## Modeling the uncertainty.

**Key challenges:**

- How to represent nondeterministic (stochastic) relations?
- How to manipulate such knowledge to make inferences?
  - **Humans can reason with uncertainty.**



# Modeling uncertainty with probabilities

**Probabilistic extension of propositional logic.**

- **Propositions:**
  - statements about the world
  - Represented by the assignment of values to **random variables**
- **Random variables:**
  - ! – **Boolean** $\underbrace{Pneumonia}_{\textbf{Random variable}}$ is either $\underbrace{True, False}_{\textbf{Values}}$
  - ! – **Multi-valued** $\underbrace{Pain}_{\textbf{Random variable}}$ is one of $\underbrace{\{Nopain, Mild, Moderate, Severe\}}_{\textbf{Values}}$
  - – **Continuous** $\underbrace{HeartRate}_{\textbf{Random variable}}$ is a value in $\underbrace{<0;250>}_{\textbf{Values}}$



# Probabilities

**Unconditional 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 $2 \times 3$ matrix

*WBCcount*

| *Pneumonia* | | $high$ | $normal$ | $low$ |
|---|---|---|---|---|
| | $True$ | 0.0008 | 0.0001 | 0.0001 |
| | $False$ | 0.0042 | 0.9929 | 0.0019 |

# Joint probabilities

**Marginalization**

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

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

$\mathbf{P}(Pneumonia)$

| | | $WBCcount$ | | | |
|---|---|---|---|---|---|
| | | $high$ | $normal$ | $low$ | |
| $Pneumonia$ | $True$ | 0.0008 | 0.0001 | 0.0001 | 0.001 |
| | $False$ | 0.0042 | 0.9929 | 0.0019 | 0.999 |
| | | 0.005 | 0.993 | 0.002 | |

$\mathbf{P}(WBCcount)$

**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

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

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 defines the probability for all possible assignments of values to *Pneumonia, Fever, Paleness, WBCcount, Cough*

5 variables: full joint is a captured by a 5-dimensional table



# Full joint distribution

- **Any joint probability for a subset of variables can be obtained from the full joint via marginalization**

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

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

# Full joint distribution

- **Any joint probability for a subset of variables can be obtained from the full joint 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 joint distribution for a set of variables from joint probabilities defined for its subsets?**



# Relations among joint distributions

**Assume:**

| | **True** | **False** |
|---|---|---|
| P(Pneumonia) | 0.001 | 0.999 |

| | **True** | **False** |
|---|---|---|
| P(Fever) | 0.05 | 0.95 |

**Can we unambiguously compute the joint over the two variables?**

| P(Fever, Pneumonia) | **True** | **False** |
|---|---|---|
| **True** | ? | ? |
| **False** | ? | ? |

## Relations among joint distributions

**Assume:**

| | **True** | **False** |
|---|---|---|
| P(Pneumonia) | 0.001 | 0.999 |

| | **True** | **False** |
|---|---|---|
| P(Fever) | 0.05 | 0.95 |

**Can we unambiguously compute the joint over the two variables?**

**No! More than one probability value is possible for joint table entries**

P(Fever, Pneumonia)

| | **True** | **False** |
|---|---|---|
| **True** | ? | ? |
| **False** | ? | ? |



## Relations among joint distribution

**Assume:**

| | **True** | **False** | |
|---|---|---|---|
| P(Pneumonia) | 0.001 | 0.999 | ← 1 free parameter |

| | **True** | **False** | |
|---|---|---|---|
| P(Fever) | 0.05 | 0.95 | ← 1 free parameter |

**Can we unambiguously compute the joint over the two variables?**

**The joint has more free parameters, the two individual distributions together**

P(Fever, Pneumonia)

| | **True** | **False** |
|---|---|---|
| **True** | ? | ? |
| **False** | ? | ? |

← 3 free parameters

## Relations among joint distribution

**Assume:**

| | **True** | **False** | |
|---|---|---|---|
| P(Pneumonia) | 0.001 | 0.999 | ← 1 free parameter |

| | **True** | **False** | |
|---|---|---|---|
| P(Fever) | 0.05 | 0.95 | ← 1 free parameter |

**Is there a condition that would let us unambiguously compute the joint over two variables?**

P(Fever, Pneumonia)

| | **True** | **False** |
|---|---|---|
| **True** | ? | ? |
| **False** | ? | ? |

← 3 free parameters



## Relations among joint distribution

**Assume:**

| | **True** | **False** | |
|---|---|---|---|
| P(Pneumonia) | 0.001 | 0.999 | ← 1 free parameter |

| | **True** | **False** | |
|---|---|---|---|
| P(Fever) | 0.05 | 0.95 | ← 1 free parameter |

**Is there a condition that would let us unambiguously compute the joint over two variables?**

**Yes. When the two variables are independent!**

P(Fever, Pneumonia)

| | **True** | **False** |
|---|---|---|
| **True** | ? | ? |
| **False** | ? | ? |

← 3 free parameters

CS 1571 Intro to AI M. Hauskrecht

## Relations among joint distribution

**Assume:**

| | **True** | **False** | |
|---|---|---|---|
| P(Pneumonia) | 0.001 | 0.999 | ← 1 free parameter |

| | **True** | **False** | |
|---|---|---|---|
| P(Fever) | 0.05 | 0.95 | ← 1 free parameter |

P(Pneumonia=True)*P(Fever=True)

P(Fever, Pneumonia)

| | **True** | **False** |
|---|---|---|
| **True** | 0.00005 | 0.0495 |
| **False** | 0.0095 | ? |

← 3 free parameters



## 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

| | | *WBCcount* | | |
|---|---|---|---|---|
| | | *high* | *normal* | *low* |
| *Pneumonia* | *True* | 0.08 | 0.0001 | 0.0001 |
| | *False* | 0.92 | 0.9999 | 0.9999 |
| | | 1.0 | 1.0 | 1.0 |

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

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

## Conditional probabilities

### Conditional probability

- Is defined in terms of the joint probability:

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

- **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.**

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

- **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

$$\begin{aligned} P(X_1, X_2, \ldots X_n) &= P(X_n \mid X_{1,} \ldots X_{n-1}) P(X_{1,} \ldots X_{n-1}) \\ &= P(X_n \mid X_{1,} \ldots X_{n-1}) P(X_{n-1} \mid X_{1,} \ldots X_{n-2}) P(X_{1,} \ldots X_{n-2}) \\ &= \prod_{i=1}^{n} P(X_i \mid X_{1,} \ldots X_{i-1}) \end{aligned}$$

# 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

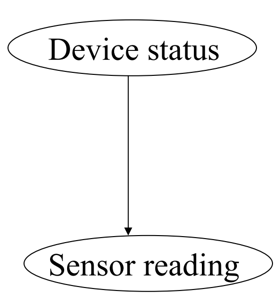
$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

- **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)

| Device\Sensor | high | low |
|---|---|---|
| normal | 0.1 | 0.9 |
| malfunctioning | 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

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

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

- 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

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

**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 $a_1, a_2, \ldots a_k$

## 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**.

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

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

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

- 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 $O(\mathrm{d}^n)$ numbers.
  $n$ – number of random variables, $d$ – number of values
- **Inference (time) complexity.** To compute some queries requires $O(\mathrm{d}^n)$ 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 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**

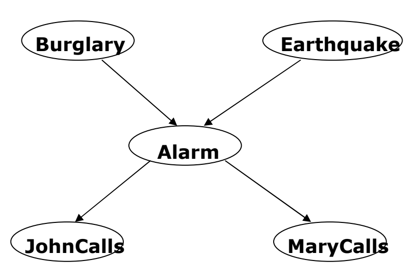
$$P(A,B \mid C) = P(A \mid C)P(B \mid C)$$
$$P(A \mid C,B) = P(A \mid C)$$



# Alarm system example.

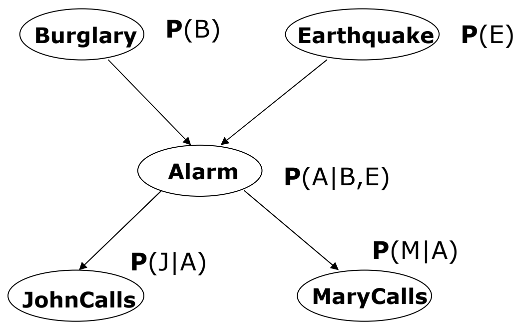
- Assume your house has an **alarm system** against **burglary**. You live in the seismically active area and the alarm system can get occasionally set off by an **earthquake**. You have two neighbors, **Mary** and **John**, who do not know each other. If they hear the alarm they call you, but this is not guaranteed.
- We want to represent the probability distribution of events:
  - Burglary, Earthquake, Alarm, Mary calls and John calls

**Causal relations**



# Bayesian belief network.

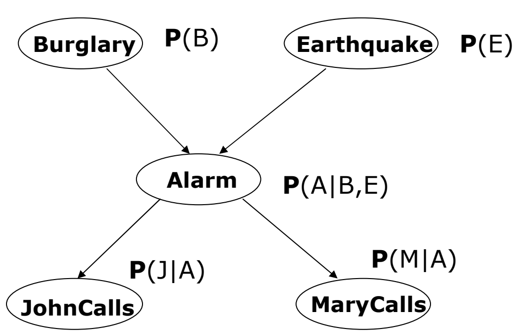
1. **Directed acyclic graph**
   - **Nodes** = random variables
     Burglary, Earthquake, Alarm, Mary calls and John calls
   - **Links** = direct (causal) dependencies between variables.
     The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm



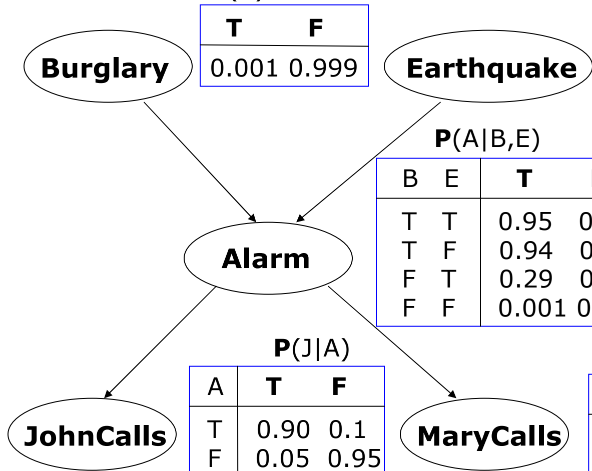



# Bayesian belief network.

2. **Local conditional distributions**
   - relate variables and their parents



# Bayesian belief network.




**P**(B)

| **T** | **F** |
|---|---|
| 0.001 | 0.999 |

**P**(E)

| **T** | **F** |
|---|---|
| 0.002 | 0.998 |

**P**(A|B,E)

| B | E | **T** | **F** |
|---|---|---|---|
| T | T | 0.95 | 0.05 |
| T | F | 0.94 | 0.06 |
| F | T | 0.29 | 0.71 |
| F | F | 0.001 | 0.999 |

**P**(J|A)

| A | **T** | **F** |
|---|---|---|
| T | 0.90 | 0.1 |
| F | 0.05 | 0.95 |

**P**(M|A)

| A | T | F |
|---|---|---|
| T | 0.7 | 0.3 |
| F | 0.01 | 0.99 |

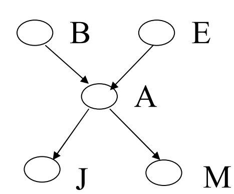


# Full joint distribution in BBNs

**Full joint distribution** is defined in terms of local conditional distributions (obtained via the chain rule):

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

**Example:**




Assume the following assignment of values to random variables

$B=T, E=T, A=T, J=T, M=F$

Then its probability is:

$$P(B=T, E=T, A=T, J=T, M=F) =$$
$$P(B=T)P(E=T)P(A=T \mid B=T, E=T)P(J=T \mid A=T)P(M=F \mid A=T)$$