Universität Stuttgart
IPVS – Institute for Parallel and Distributed Systems
Analytic Computing

Advanced Topics in Machine Learning
2 Representation: Bayesian Networks

Prof. Dr. Steffen Staab
Dr. Rafika Boutalbi
Zihao Wang
https://www.ipvs.uni-stuttgart.de/departments/ac/

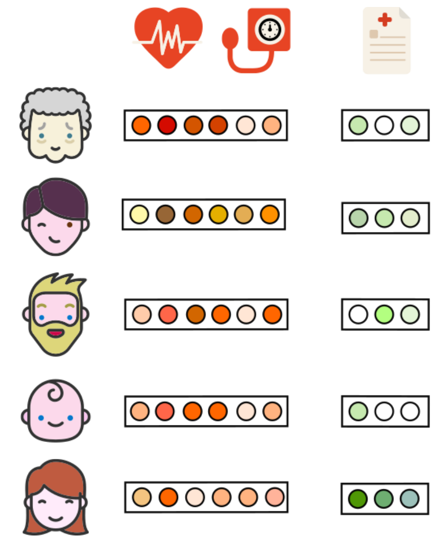
Disclaimer

In this slide deck I reused slides by Dr. Matthias Niepert

1 Random Variables and Probabilistic Models

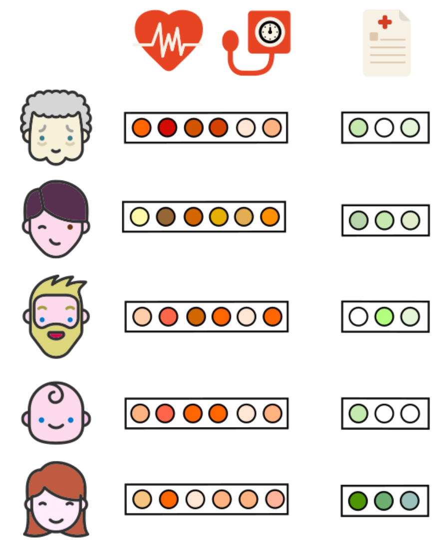
Example:

- Setting: Patient medical records from hospital
- Observations: heart rate, pH, temp, …
- Goal: Make **diagnosis using ML** but also:
  - What is the degree of uncertainty?
  - What tests could reduce uncertainty?
  - What are the reasons?
  - We need to handle missing data



Motivation

- We need a language to represent **observations** and their **uncertainty**
- We need a language to model the **relationships** (dependence) between **observations**
- We need to answer **questions (queries)** using said language such as:
  - “What is the probability of *Sepsis*, given that *pH*>7.4 and *Temp=41.0* ?”
  - “What is the most likely *Diagnosis*, given observations?”
- **Probabilistic models** provide such a (family of) language(s)



Random Variables (quick rehash)

- Atomic building blocks of probabilistic models
  - **Example:** Die_1, *Sum*, *Sepsis*
- Each random variable has a **domain (dom)**
  – the set of possible values it can take on
  - **Example:** {1,…,6}, {2,…,12}, {true, false}
  - Similar to variables in programming languages *but*
- Random variables are equipped with a **distribution P** over their **domain** (probability P for each element)
  - We write P for probability of assignment, and **P** for a distribution
  - **Example:** $\mathbf{P}(Die_1) = <0.1,0.1,0.2,0.2,0.2,0.2>$ or, equivalently, $P(Die_1{=}1)=0.1$, …, $P(Die_1{=}6)=0.2$
- **P** also determines probability of subsets of domain
  - **Example:** $P(Die_1{<=}2)=0.1+0.1=0.2$

Die_1: 

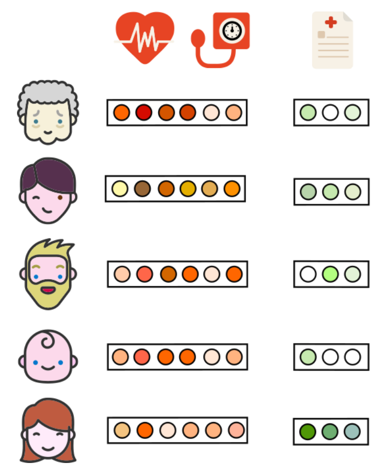
Sum:  + 

Initially, we assume random variables have a **finite** domain.

Random Variables (quick rehash)

- Probability distributions on *multiple variables* are called **joint probability distributions**
  - **Example: P**(*Sepsis*, *PH*, *Temp*), **P**(*Weather*, *Sepsis*)
- A **possible world (state)** is an **assignment** of values to all random variables under consideration
- We write P(**X**=**x**) for the probability of one joint variable assignment **X**=**x**
- **P**($X|Y$) is a *conditional* probability distribution (CPD), and determines all conditional probabilities P(X=x_i|Y=y_j)
- We often use the term **joint probability distribution** and **probabilistic model** interchangeably

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Weather*) = {sunny, rainy}
dom(*Sepsis*) = {true, false}



Question 1:
Random Variables

- Given a random variable that models the sum of two fair dies. What is the conditional probability distribution if the first die is less or equal 3?

- Given a random variable that models the sum of two fair dies. What is the conditional probability distribution if the first die is less or equal 3?
- Sum: <2,3,4,5,6,7,8,9,10,11,12>
- <1/18, 2/18, 3/18, 3/18, 3/18, 3/18, 2/18,1/18,0,0,0>

2 Probabilistic Inference

Queries to Probabilistic Models

- Two basic types of queries (questions) for a given probabilistic model:
  1. "What is a most probable state of the random variables?"

     $$\operatorname{argmax}_{\mathbf{x} \in \mathrm{dom}(X_1) \times \cdots \times \mathrm{dom}(X_n)} P(\mathbf{X} = \mathbf{x})$$

  2. "What is the marginal probability of a (set of) variable(s)?"

$$P(\mathbf{Y} = \mathbf{y}) = \sum_{\mathbf{z} \in \mathrm{dom}(Z_1) \times \cdots \times \mathrm{dom}(Z_k)} P(\mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z})$$

Answering such queries is called **probabilistic inference**

$$\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$$
$$\mathbf{Y} \cap \mathbf{Z} = \emptyset$$

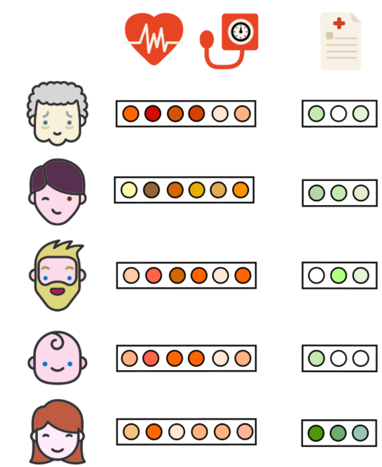
- All queries (including conditional probability queries) can be answered when we can answer the two queries above

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{E} = \mathbf{e}) = \alpha \sum_{\mathbf{z} \in \mathrm{dom}(Z_1) \times \cdots \times \mathrm{dom}(Z_k)} P(\mathbf{Y} = \mathbf{y}, \mathbf{E} = \mathbf{e}, \mathbf{Z} = \mathbf{z})$$

Inference Using Joint Distribution

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

	Temp=high		*Temp*=normal	
	PH=high	*PH*=normal	*PH*=high	*PH*=normal
Sepsis=true	0.108	0.012	0.072	0.008
Sepsis=false	0.016	0.064	0.144	0.576



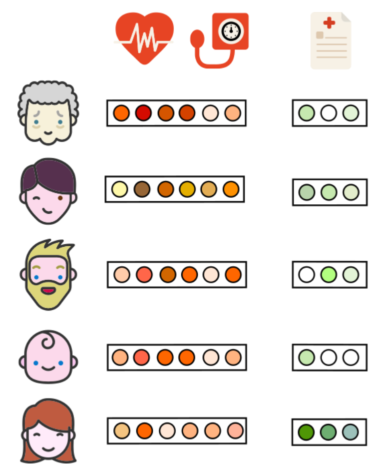
"What is the probability of Sepsis?"

$$P(\text{Sepsis} = true)\ ?$$

Inference Using Joint Distribution

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

	Temp=high		*Temp*=normal	
	PH=high	*PH*=normal	*PH*=high	*PH*=normal
Sepsis=true	0.108	0.012	0.072	0.008
Sepsis=false	0.016	0.064	0.144	0.576



"What is the probability of Sepsis, given PH is high?"

$$P(\text{Sepsis} = true | PH = high)\ ?$$

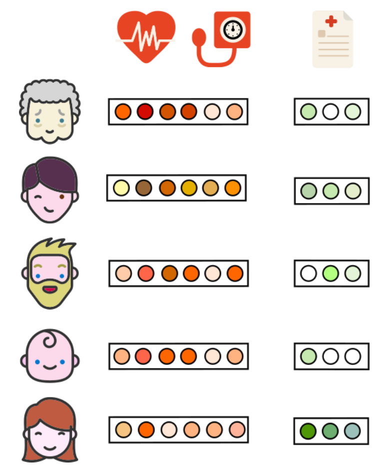
Inference Using Joint Distribution

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

	Temp=high		*Temp*=normal	
	PH=high	*PH*=normal	*PH*=high	*PH*=normal
Sepsis=true	0.108	0.012	0.072	0.008
Sepsis=false	0.016	0.064	0.144	0.576

“What is the most probable state given that the temperature is high?”

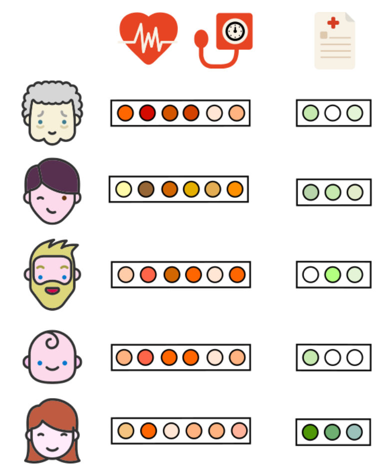
$$\text{argmax}_{?}\, P(?)$$



MAP: Maximum A Posteriori - Query

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

	Temp=high		*Temp*=normal	
	PH=high	*PH*=normal	*PH*=high	*PH*=normal
Sepsis=true	0.108	0.012	0.072	0.008
Sepsis=false	0.016	0.064	0.144	0.576



"What is the most probable state given that the temperature is high?"

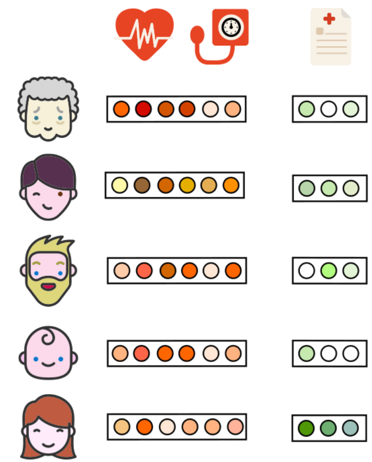
$$MAP(W|e) = \text{argmax}_{w \in W} P(w, e)$$

$$MAP(PH, Sepsis|Temp = high) = \text{argmax}_{w \in PH \times Sepsis} P(W = w, Temp = high)$$

Marginal MAP-Query

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

	Temp=high		*Temp*=normal	
	PH=high	*PH*=normal	*PH*=high	*PH*=normal
Sepsis=true	0.108	0.012	0.072	0.008
Sepsis=false	0.016	0.064	0.144	0.576



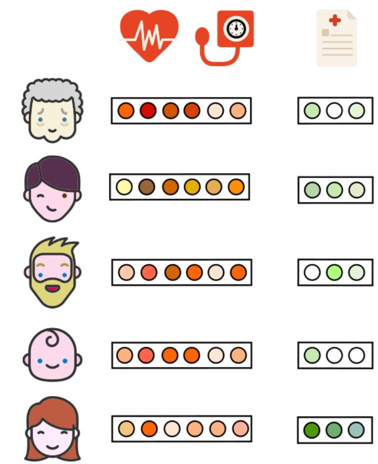
"What is the most probable state wrt Sepsis given that the temperature is high?"

$$MAP(Y|e) = \text{argmax}_{y \in Y} \sum_{z \in Z} P(y, z|e)$$

Semantics of Probabilistic Models

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

	Temp=high		*Temp*=normal	
	PH=high	*PH*=normal	*PH*=high	*PH*=normal
Sepsis=true	0.108	0.012	0.072	0.008
Sepsis=false	0.016	0.064	0.144	0.576



Whatever syntax we choose for:

- language for representing probabilities
- language for querying

The joint probabilities define the underlying semantics

Side note

Computer scientists define

- syntax
- semantics

of languages

all the time

Thus: be aware what is the syntax and what is the semantics of a given language!

Complexity of Probabilistic Inference

- Two basic types of queries (questions) for a probabilistic model:
  1. "What is a most probable state of the random variables?"

     $$argmax_{\mathbf{x} \in \mathrm{dom}(\mathrm{X}_1)\times\cdots\times\mathrm{dom}(\mathrm{X}_n)} P(\mathbf{X} = \mathbf{x})$$

  2. "What is the marginal probability of a (set of) variable(s)?"

     $$P(\mathbf{Y} = \mathbf{y}) = \sum_{\mathbf{z} \in \mathrm{dom}(\mathrm{Z}_1)\times\cdots\times\mathrm{dom}(\mathrm{Z}_k)} P(\mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}) \qquad \begin{array}{c} \mathbf{X} = \mathbf{Y} \cup \mathbf{Z} \\ \mathbf{Y} \cap \mathbf{Z} = \emptyset \end{array}$$

- Have you noticed a potential problem?
  - answering these queries is **exponential** in the number of random variables!
- No algorithm known that computes answers to queries in general and runs in time polynomial in number of random variables

3 Modeling Independence

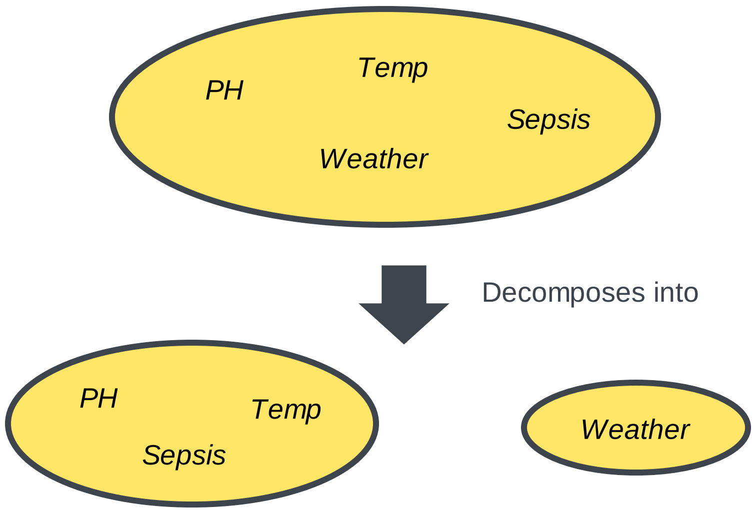
Independence to the Rescue

- Two random variables are independent if $\mathbf{P}(X,Y) = \mathbf{P}(X)\,\mathbf{P}(Y)$
  - We also write: $X \perp Y$
- Can be generalized to sets of random variables:
  - $\mathbf{P}(\mathbf{X},\mathbf{Y}) = \mathbf{P}(\mathbf{X})\,\mathbf{P}(\mathbf{Y}) = \mathbf{P}(X_1,\ldots,X_m,Y_1,\ldots,Y_n) = \mathbf{P}(X_1,\ldots,X_m)\,\mathbf{P}(Y_1,\ldots,Y_n)$
  - We also write: $\mathbf{X} \perp \mathbf{Y}$

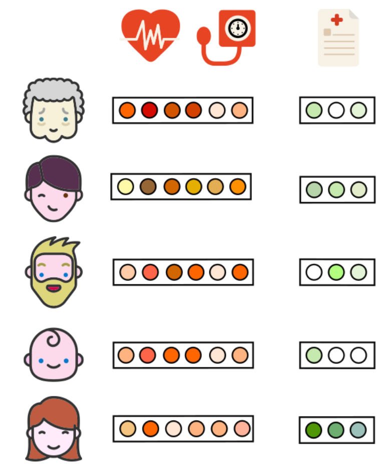
Independence to the Rescue

- Two random variables are independent if $\mathbf{P}(X,Y)=\mathbf{P}(X)\mathbf{P}(Y)$
- Can be generalized to sets of random variables: $\mathbf{P}(\mathbf{X},\mathbf{Y})=\mathbf{P}(\mathbf{X})\mathbf{P}(\mathbf{Y})$
- Independence is often based on knowledge of the domain under consideration but can also be inferred from data
- Independence between random variables can *dramatically* reduce:
  - The amount of information necessary to specify the probability distribution (number of **parameters**!)
  - The **computational complexity** of probabilistic inference

Independence of Random Variables



dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Weather*) = {sunny, rainy}
dom(*Sepsis*) = {true, false}



- Due to the **independence** of {*Weather*} and {*PH*, *Sepsis*, *Temp*} we can write

$$\mathbf{P}(PH, Temp, Sepsis, Weather) = \mathbf{P}(Weather)\,\mathbf{P}(PH, Temp, Sepsis)$$

- The probability distribution **factorizes** (decomposes) into **independent** parts
- Reduction of number of parameters from $2^4-1 = 15$ to $(2-1) + (2^3-1) = 8$

Conditional Independence

- **Problem:** Not many variables in the real world are actually independent
- Two random variables X, Y are conditionally independent given Z if

  $\mathbf{P}(X|Y,Z) = \mathbf{P}(X|Z)$ *or* $\mathbf{P}(Y|X,Z)=\mathbf{P}(Y|Z)$
- Can be generalized to sets of random variables:

  $\mathbf{P}(\mathbf{X}|\mathbf{Y},\mathbf{Z}) = \mathbf{P}(\mathbf{X}|\mathbf{Z})$ *or* $\mathbf{P}(\mathbf{Y}|\mathbf{X},\mathbf{Z})=\mathbf{P}(\mathbf{Y}|\mathbf{Z})$ or $(\mathbf{Y} \perp \mathbf{X} \mid \mathbf{Z})$
- Conditional independencies are **more common** than absolute independencies, yet still *reduce complexity* of model
- Conditional independencies allow probabilistic systems to *scale up* and are represented by **probabilistic graphical models**

4 Bayesian Network

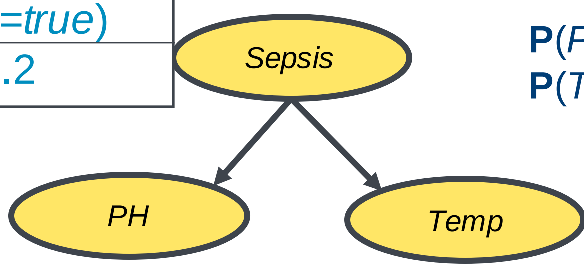
Bayesian Networks

DAG = directed acyclic graph: a directed graph without directed cycles

A Bayesian network is a DAG representing

- A *finite* number of random variables (the nodes)
- Conditional independence assumptions on these random variables:
  "Each variable is conditionally independent of all
  its non-descendants given the value of its parents"
- The (conditional) probability values (parameters)

P(Sepsis=*true*)
0.2



$\mathbf{P}(PH|Sepsis) = \mathbf{P}(PH|Sepsis, Temp)$
$\mathbf{P}(Temp|Sepsis) = \mathbf{P}(Temp|Sepsis, PH)$

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

$\mathbf{P}(Sepsis, Temp, PH) = \mathbf{P}(Sepsis)\mathbf{P}(Temp|Sepsis)\mathbf{P}(PH|Sepsis)$

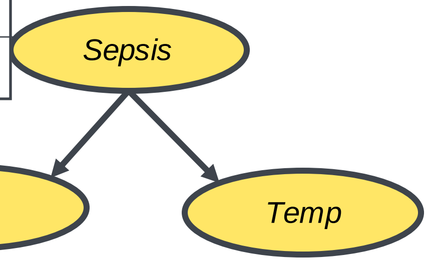
S	P(*PH=high*)
true	0.9
false	0.2

S	P(*Temp=high*)
true	0.6
false	0.1

$$\mathbf{P}(\mathbf{X}) = \prod_{X \in \mathbf{X}} \mathbf{P}(X|\text{parents}(X))$$

Inference Using Bayesian Networks

P(Sepsis=*true*)
0.2



S	P(*PH=high*)
true	0.9
false	0.2

S	P(*Temp=high*)
true	0.6
false	0.1

dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

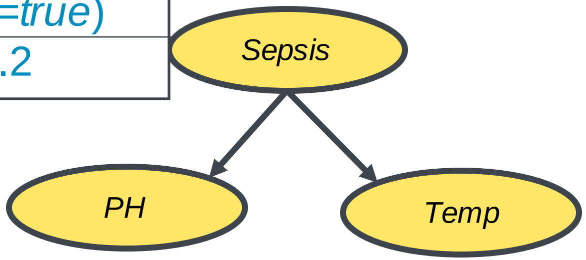
$$\mathbf{P}(\mathbf{X}) = \prod_{X \epsilon \mathbf{X}} \mathbf{P}(X|\text{parents}(X))$$

“What is the probability of Sepsis?”

$$P(\text{Sepsis} = true)\ ?$$

Inference Using Bayesian Networks

P(Sepsis=*true*)
0.2



dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

S	P(*PH=high*)
true	0.9
false	0.2

S	P(*Temp=high*)
true	0.6
false	0.1

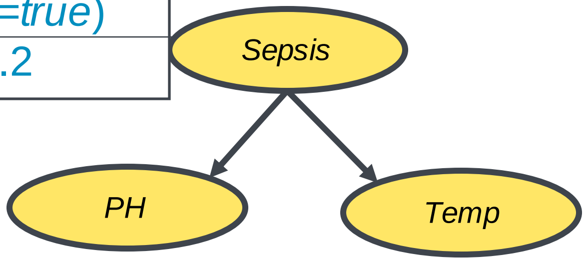
$$\mathbf{P}(\mathbf{X}) = \prod_{X \epsilon \mathbf{X}} \mathbf{P}(X|\text{parents}(X))$$

"What is the probability of Sepsis, given PH is high?"

$$P(\text{Sepsis} = true | PH = high)\ ?$$

Inference Using Bayesian Networks

P(Sepsis=*true*)
0.2



dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

S	P(*PH=high*)
true	0.9
false	0.2

S	P(*Temp=high*)
true	0.6
false	0.1

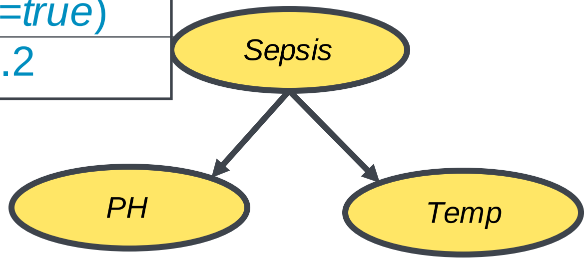
$$\mathbf{P}(\mathbf{X}) = \prod_{X \epsilon \mathbf{X}} \mathbf{P}(X|\text{parents}(X))$$

"What is the probability of Sepsis, given PH is high?"

$$P(\text{Sepsis} = true|PH = high) = P(\text{Sepsis} = true, PH = high) \,/\, P(PH = high)$$

Inference Using Bayesian Networks

P(Sepsis=*true*)
0.2



dom(*PH*) = {normal, high}
dom(*Temp*) = {normal, high}
dom(*Sepsis*) = {true, false}

S	P(*PH=high*)
true	0.9
false	0.2

S	P(*Temp=high*)
true	0.6
false	0.1

$$\mathbf{P(X)} = \prod_{X \epsilon \mathbf{X}} \mathbf{P}(X|\text{parents}(X))$$

Bayes networks reduce the parameter space, but we still need to improve inferencing algorithms!!!

"What is the most probable state, given that the temperature is high?"

$$\text{argmax}_{?}\, P(?)$$

Question 4:

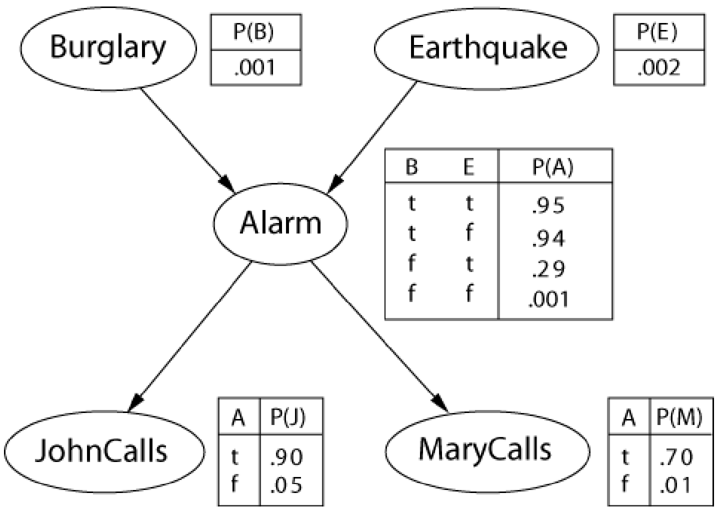
Bayesian Network

Bayesian Networks – Exercise

- Explain how one could sample from the probability distribution
- List a set of conditional independency statements represented by the BN
- Provide the answer to the following probability queries for the BN:

$$\mathrm{argmax}_{\mathbf{x}\in B\times E\times J} P(\mathbf{X}=\mathbf{x}|A=true, M=false)$$

P(E=true)
P(B=true|A=true)
P(M=false|B=true)



The domain of all random variables is {true, false}

Universität Stuttgart
IPVS

Thank you!

Steffen Staab

E-Mail Steffen.staab@ipvs.uni-stuttgart.de
Telefon +49 (0) 711 685-To be defined
www. ipvs.uni-stuttgart.de/departments/ac/

Universität Stuttgart
Analytic Computing, IPVS
Universitätsstraße 32, 50569 Stuttgart