PIP/CA - *Programa Interdisciplinar de Pós-Graduação*
Mestrado em Computação Aplicada da UNISINOS

2000/1 - 2o. Trimestre - AULA 05 / FSO

INTELIGÊNCIA ARTIFICIAL & SISTEMAS INTELIGENTES

- **Professores Responsáveis:**

Parte I - Profa. Dr. Renata Vieira
Web: ***http://www.inf.unisinos.br/~renata/iam.html***
Parte II - Prof. Dr. Fernando Osório
E-Mail: ***osorio@exatas.unisinos.br***
Web: ***http://www.inf.unisinos.br/~osorio/ia.html***

TEMAS DE ESTUDO: Uso da *INCERTEZA nos SISTEMAS INTELIGENTES*

Probabilistic Approaches
Bayesian Approach
Bayesian Belief Networks

- **Conceitos Básicos:**
  Probabilidade, Evidência e Hipótese
- **Teorema de Bayes**
- **Problemas:** **Complexidade, Tabelas JPD, Relações Hipótese/Evidência**
- **Soluções:** **CF, Fuzzy inference, Dempster-Shafer, Bayesian Belief Networks**
- **Bayesian Belief Networks - BBN**
- **Representação Gráfica:** **CPTs - Conditional Probability Tables, Belief Network**
- **Exemplos e Implementações de BBNs**
  Hugin, Netica, MSBN, ...

- **Temas de Pesquisa relacionados aos sistemas tipo Bayesian Belief Networks**

RACIOCÍNIO PROBABILISTA: Conceitos Básicos

Probabilidade: [Russell]

- Resume a nossa incerteza oriunda da falta de conhecimento preciso e completo sobre o problema
- Inabilidade para obter uma decisão definitiva em relação a veracidade da sentença
- Resume a minha crença, de forma numérica (0 à 1, onde 0 e 1 são as “crenças inequívocas”)

Prior Probability:

- Probabilidade Incondicional => P(A) $P(A) = 1 - P(\neg A)$
- A probabilidade de A na ausência de toda outra informação.

Conditional Probability:

- Probabilidade Conditional ou Probabilidade Posterior => P(A|B) $P(A|B) = \frac{P(A \wedge B)}{P(B)}$
- A probabilidade de A dado que tudo que conhecemos é B.

$P(X,Y) = P(X \wedge Y) = P(X|Y)P(Y)$ “Product Rule” => $P(A \wedge B) = P(A|B)\,P(B)$

$P(A \wedge B) = P(B|A)\,P(A)$

• ***Valores numéricos das probabilidades:***

* Interpretação <u>Freqüencista</u>: estimativa baseada em estatísticas (experiências práticas)
* Interpretação <u>Objetivista</u>: regras - aspectos reais do universo (propriedades físicas)
* Interpretação <u>Subjetivista</u>: crenças - baseado na experiência pessoal (grau de crença do agente)
  Exemplo: na minha opinião, a probabilidade de acidentes é de 10%

RACIOCÍNIO PROBABILISTA: Conceitos Básicos

Probabilidade: [Russell]

- Resume a nossa incerteza oriunda da falta de conhecimento preciso e completo sobre o problema
- Inabilidade para obter uma decisão definitiva em relação a veracidade da sentença

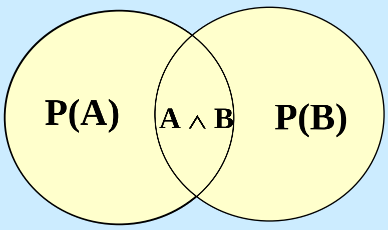


Diagrama de Venn:

P(A), P(B) = Probabilidade de A e de B

Operações: $A \vee B$, $A \wedge B$

Axiomas da Probabilidade:

$P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

$P(X,Y) = P(X \wedge Y) = P(X|Y)P(Y)$ “Product Rule” => $P(A \wedge B) = P(A|B)\,P(B)$

$P(A \wedge B) = P(B|A)\,P(A)$

• ***Valores numéricos das probabilidades:***

* CF - Certainty Factor (±)
* Fuzzy Logic / Fuzzy Sets - Lógica Nebulosa (±)
* Logica Possibilista (Zadeh) (±)
* Dempster-Shafer - Teoria da Evidência (±)
* Teoria Probabilista Bayesiana => Belief Networks

RACIOCÍNIO PROBABILISTA: Conceitos Básicos

Probabilidade: JPD - Joint Probability Distribution

- Um modelo probabilista de um domínio consiste de um conjunto de variáveis aleatórias que podem assumir valores particulares com certas probabilidades.

- A distribuição de probabilidade conjunta P(x1, x2, ... xn) atribui probabilidades para todos os eventos atômicos. Exemplo:

	Dor_de_Dente	¬ Dor_de_Dente
Cárie	0.04	0.06
¬ Cárie	0.01	0.89

- Os eventos atômicos são mutualmente exclusivos e coletivamente exaustivos (axiomas da probab.)

- Para o exemplo acima teremos:

Problema real:
- Enorme quantidade de variáveis randômicas
- Variáveis discretas, e também contínuas!

A soma de todas as probabilidades = 1.0

Probabilidade incondicional P(cárie) = 0.06 + 0.04 = 0.10

Probabilidades incondicional P(cárie ∨ dor_de_dente) = 0.04 + 0.01 + 0.06 = 0.11

Probabilidade condicional $P(\text{cárie}|\text{dor\_de\_dente}) = \dfrac{P(\text{cárie} \wedge \text{dor\_de\_dente})}{P(\text{dor\_de\_dente})} = \dfrac{0.04}{0.04 + 0.01} = 0.80$

Discrete Random Variables

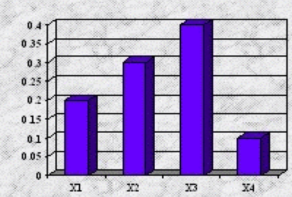
- Finite set of possible outcomes

$$X \in \{x_1, x_2, x_3, \ldots, x_n\}$$

$$P(x_i) \geq 0$$

$$\sum_{i=1}^{n} P(x_i) = 1$$

X binary: $P(x) + P(\bar{x}) = 1$



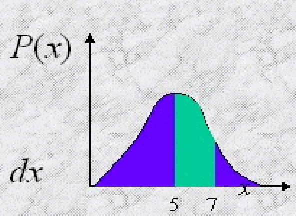
Continuous Random Variable

- Probability distribution (density function) over continuous values

$X \in [0,10]$ $\quad P(x) \geq 0$

$$\int_0^{10} P(x)\,dx = 1$$

$$P(5 \leq x \leq 7) = \int_5^7 P(x)\,dx$$



Variáveis Discretas x Variáveis Contínuas

Hepatitis/Jaundice/BloodTest

J	B	H	P(J, B, H)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

Joint Probability Distribution (JPD)

- Tabela extremamente grande
- Relaciona todas as variáveis
- Completa para todos os valores destas variáveis
- Enorme número de probabilidades condicionais

RACIOCÍNIO PROBABILISTA: Bayesian Reasoning

Regra de Bayes:

- A probabilidade condicional, P(H/E), dos eventos E e H pode ser vista como uma quantificação da relação de causa e efeito entre E e H => E é a *evidência* que suporta a *hipótese* H

$$P(H|E) = \frac{P(H \wedge E)}{P(E)} = \frac{P(H)\,P(E|H)}{P(E)}$$

> A *regra de Bayes* expressa as probabilidades *a priori* em termos das probabilidades posteriores (mais fáceis de obter/estimar)
> >> Permite obter probabilidades desconhecidas <<
> >> a partir de probabilidades conhecidas <<

Para o caso de termos 2 evidências...

$$P(H|E_1 \wedge E_2) = \frac{P(H)\,P(E_1 \wedge E_2 \mid H)}{P(E_1 \wedge E_2)}$$

E de uma maneira geral...

$$P(H|\, E_1 \wedge E_2 \wedge \ldots \wedge E_n) = \frac{P(H)\,P(E_1 \wedge E_2 \wedge \ldots \wedge E_n|H)}{P(E_1 \wedge E_2 \wedge \ldots \wedge E_n)}$$ *Regra de Bayes*

Teorema de Bayes:

$$P(Hi|E) = \frac{P(Hi)\,P(E|Hi)}{\sum_{k=1}^{m} P(Hk)\,P(E|Hk)}$$ *como* $P(E) = \sum_{k=1}^{m} P(Hk)\,P(E|Hk)$ *então* $P(Hi|E) = \frac{P(Hi)\,P(E|Hi)}{P(E)}$

P(E|Hi) = "verossimilhança"

RACIOCÍNIO PROBABILISTA: Bayesian Reasoning

Regra de Bayes: exemplo prático

- Supondo que:
  e1= solteiro, e2= salário_alto, e3= jovem ⇒ apoiam as hipóteses h1= investidor_de_alto_risco
  h2= investidor_de_baixo_risco

  sendo mutualmente exclusivas e exaustivas, isto é: $P(h1 \wedge h2) = 0$ e $P(h1) = 1 - P(h2)$

- Assumindo os conhecimentos do especialista, que estima as probabilidades posteriores:

P(H=h1) = 0.3	P(E=e1|H=h1) = 0.6	P(E=e2|H=h1) = 0.2	P(E=e3|H=h1) = 0.5
P(H=h2) = 0.7	P(E=e1|H=h2) = 0.3	P(E=e2|H=h2) = 0.8	P(E=e3|H=h2) = 0.2

- Probabiliadades a priori (incondicionais):

$$P(h1|e1) = \frac{P(h1)\,P(e1|h1)}{\sum_{k=1}^{m} P(Hk)\,P(e1|Hk)} = \frac{P(h1)\,P(e1|h1)}{P(h1)P(e1|h1) + P(h2)P(e2|h2)} = \frac{0.3 * 0.6}{0.3 * 0.6 + 0.7 * 0.3} = 0.4615$$

Para as demais...

P(h1|e2) = 0.097, P(h1|e3) = 0.517, P(h2|e1) = 0.538, P(h2|e2) = 0.903, P(h2|e3)=0.483

$P(h1|e1 \wedge e3) = 0.681$, $P(h2|e1 \wedge e3) = 0.318$, $P(h1|e1 \wedge e2 \wedge e3) = 0.349$, $P(h1|e1 \wedge e2 \wedge e3)=0.651$

$$P(h1|e1 \wedge e2 \wedge e3) = \frac{P(h1)P(e1|h1)P(e2|h1)P(e3|h1)}{P(h1)P(e1|h1)P(e2|h1)P(e3|h1) + P(h2)P(e1|h2)P(e2|h2)P(e3|h2)} = 0.349$$

RACIOCÍNIO PROBABILISTA: Bayesian Reasoning

Regra de Bayes: exemplo prático

- Supondo que:
e1= solteiro, e2= salário_alto, e3= jovem ⇒ apoiam as hipóteses h1= investidor_de_alto_risco
h2= investidor_de_baixo_risco

sendo mutualmente exclusivas e exaustivas, isto é: $P(h1 \wedge h2) = 0$ e $P(h1) = 1 - P(h2)$

- Assumindo os conhecimentos do especialista, que estima as probabilidades posteriores:
P(H=h1) = 0.3 P(E=e1|H=h1) = 0.6 P(E=e2|H=h1) = 0.2 P(E=e3|H=h1) = 0.5
P(H=h2) = 0.7 P(E=e1|H=h2) = 0.3 P(E=e2|H=h2) = 0.8 P(E=e3|H=h2) = 0.2

- Probabiliadades a priori (incondicionais): ***PROBLEMA EXTREMAMENTE SIMPLES***

	Probabilidade Inicial	Depois da Evidência				
		e1	e2	e3	$e1 \wedge e3$	$e1 \wedge e2 \wedge e3$
h1 = Alto Risco	0.30	0.461	0.097	0.517	0.681	0.349
h2 = Baixo Risco	0.70	0.538	0.90	0.488	0.318	0.651

$$P(h1|e1 \wedge e2 \wedge e3) = \frac{P(h1)P(e1|h1)P(e2|h1)P(e3|h1)}{P(h1)P(e1|h1)P(e2|h1)P(e3|h1) + P(h2)P(e1|h2)P(e2|h2)P(e3|h2)} = 0.349$$

RACIOCÍNIO PROBABILISTA: Bayesian Reasoning Weaknesses

Regra de Bayes: problemas

- Este tipo de método precisa trabalhar com um número MUITO grande de probabilidades - P(Hi) e P(Ej/Hi) - para cada evidência Ej e hipóteses Hi.

- Dificuldade em se estimar estas probabilidades a priori de Ei e Hi...

- A regra de Bayes assume que os antecedentes Ei são independentes. Nem sempre é o caso!

- O uso de probabilidades assume que a presença de uma evidência também afeta a negação da conclusão. Isso pode ser bastante problemático quando usamos uma abordagem baseada em uma "interpretação subjetivista"!

- Se as probabilidades a priori e as probabilidades condicionais são baseadas em contagens de freqüências e estatísticas, temos que assegurar que o número de amostras é representativo o suficiente para obter probabilidades precisas.
(algumas vezes as bases de dados não são corretas e precisas o suficiente para que sua soma seja igual a 1.0)

=> Solução? CF - Certainty Factors, Fuzzy Sets, Dempster-Schafer <==> BBN

Redes de Crenças - *Belief Networks* / Bayesian Belief Networks (BBN)
- Introduz meios de eliminar a independência entre evidências
- A evidência de uma hipótese não é considerada simultaneamente como uma evidência de negação desta hipótese.

RACIOCÍNIO PROBABILISTA: Bayesian Belief Networks - BBN

• *BBN: Conceitos Básicos*

- Permite trabalhar com relações causais de dependência entre variáveis.

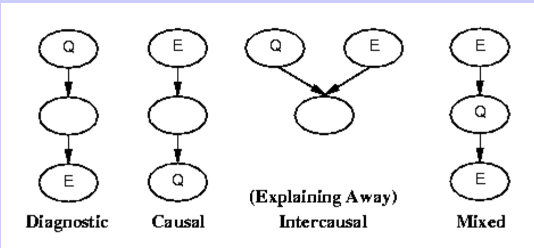
- Representa as tabelas de JPD de forma compacta - Diagramas de influências ("relações locais").

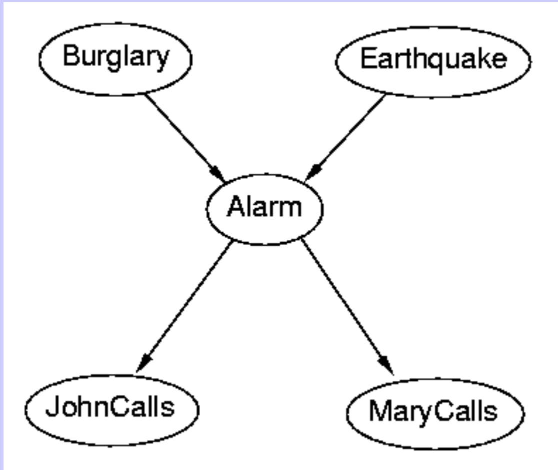
- BBN é um grafo acíclico direcionado => DAG contento:

*** Nodos: representam as variáveis estatísticas**
*** Arcos: representam as relações de dependência causal entre as variáveis**
*** Tabelas: CTP - *Conditional Probability Tables* (representação de JPD decompostas) indicando as probabilidades condicionais de influência entre os nodos conectados.**

• *Padrões de Inferência Probabilista nas Belief Networks:*

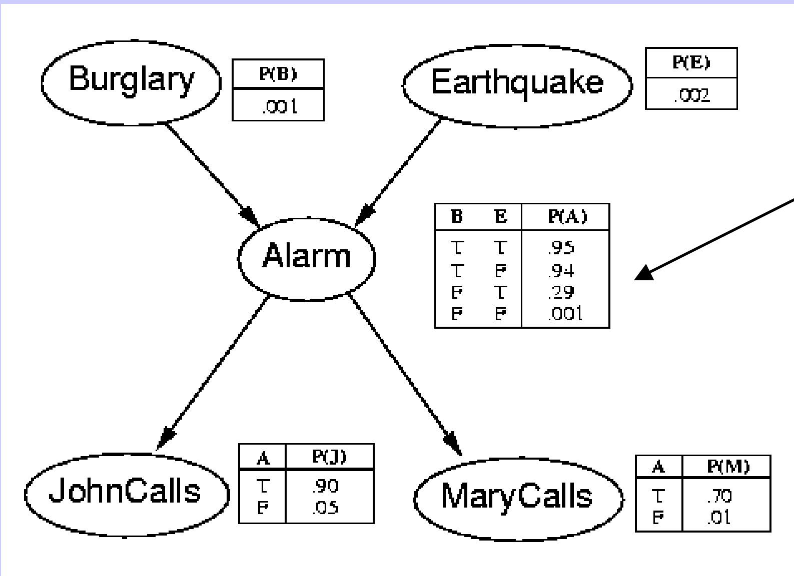
Diagnóstico: evidência => causas (query variable)
Causal : causas => evidência
Intercausal: entre causas de uma mesma evidência
Misto : combinando ambos





RACIOCÍNIO PROBABILISTA: Bayesian Belief Networks [Pearl, Jensen]

• *Exemplo de Rede Bayesiana - BBN*



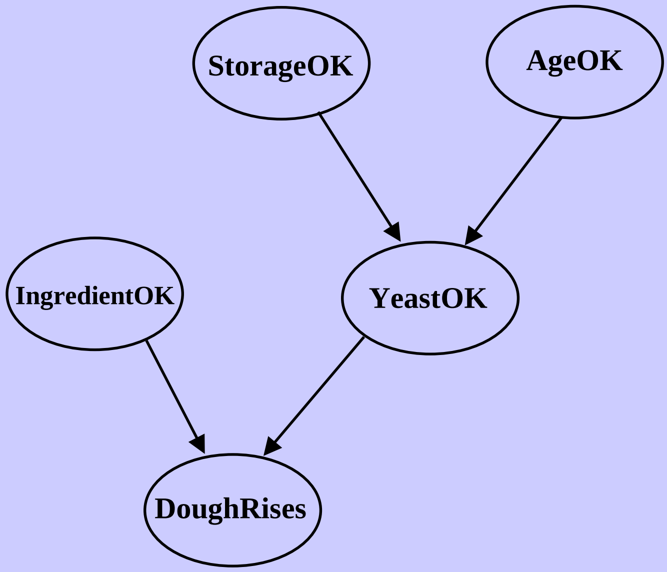
CPT:
Conditional
Probability
Table

BBN is a graphical data structure that compactly represent the joint probability distribution. [Harisson]

- Captura o conhecimento sobre o problema de forma natural e eficiente
- Expressa as relações causais, explorando as dependências condicionais.

RACIOCÍNIO PROBABILISTA: Bayesian Belief Networks - Inferência

P(Storage = True) = 0.8
P(Ingred OK = True) = 0.8
P(Age = True) = 0.8



StorageOK	AgeOK	YeastOK
False	False	0.1
False	True	0.2
True	False	0.3
True	True	0.9

IngredientOK	YeastOK	DoughRises
False	False	0.1
False	True	0.2
True	False	0.3
True	True	0.9

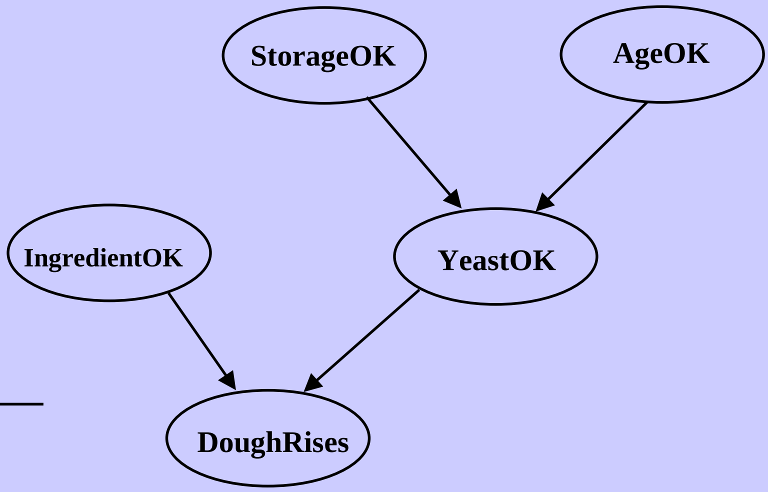
Calculando a JPD através da tabela CPT...

P(StorageOK=True, AgeOK=True, YeastOK=True, IngredientOK=True, DoughRises=True) = ...
P(S=True, A=True, Y=True ,I=True, D=True) = ...

$$P(S, A, Y, I, D) = P(D \mid I, Y, A, S) * P(I \mid Y, A, S) * P(Y \mid A, S) * P(S \mid A) * P(A)$$
$$= P(D \mid I, Y) * P(I) * P(Y \mid A, S) * P(S) * P(A)$$
$$= 0.9 * 0.8 * 0.9 * 08 * 0.9 = 0.46656$$

RACIOCÍNIO PROBABILISTA: Bayesian Belief Networks - Inferência

P(Storage = True) = 0.8
P(Ingred OK = True) = 0.8
P(Age = True) = 0.8



StorageOK	AgeOK	YeastOK
False	False	0.1
False	True	0.2
True	False	0.3
True	True	0.9

IngredientOK	YeastOK	DoughRises
False	False	0.1
False	True	0.2
True	False	0.3
True	True	0.9

Exemplo de uma *query* onde são dadas algumas evidências...

$$P(Y \mid S, I, \neg D) = P(Y, S, I, \neg D) / P(S, I, \neg D)$$

$$P(Y, S, I, \neg D) = P(Y, S, I, \neg D, A) + P(Y, S, I, \neg D, \neg A)$$
$$= P(Y \mid S, A) * P(\neg D \mid I, Y) * P(I) * P(S) * P(\neg A) + P(Y \mid S, \neg A) * P(\neg D \mid I, L) * P(I) * P(S) * P(\neg A)$$
$$= 0.9 * 0.1 * 0.8 * 0.8 * 0.8 + 0.3 * 0.1 * 0.8 * 0.8 * 0.2 = 0.04992$$

$$P(\neg D, I, S) = ... = 0.14848$$

$$P(Y \mid S, I, \neg D) = 0.04882 / 0.14848 = 0.03362$$ [cf. Harrison & Kovalchik]

TEMAS DE PESQUISA SOBRE SISTEMAS ESPECIALISTAS:

*** PAPERS / DOCUMENTAÇÃO:**

- **Artificial Intelligence / Russel & Norvig (Cap. 14 e 15)**
- **Raciocínio Probabilista em Sistemas Inteligentes / Ladeira, Viccari, Coelho (SBC / JAI'99)**
- **Expert Systems / Chris Nikolopoulos (Cap. 5)**
- **The HandBook of Applied Expert Systems / Jay Liebowitz (Cap. 8)**
- **Web: Hugin / Microsoft - MSBN - Heckerman**

*** SOFTWARES:**

- **Hugin - http://www.hugin.dk/**
- **MSBN - http://www.research.microsoft.com/research/dtg/msbn/**
- **Netica - http://www.norsys.com/netica.html**
- **Lista de softwares:** **http://bayes.stat.washington.edu/almond/belief.html**
  http://www.cs.berkeley.edu/~murphyk/Bayes/bnsoft.html

*** TEMAS IMPORTANTES:**

- **Aprendizado em redes bayesianas...** **http://www.research.microsoft.com/research/nips95bn/**
  - **Estimativa das probabilidades (conjuntos de dados estatísticos)**
  - **Grafo (estrutura / dependências entre variáveis)**