Probabilistic Graphical Models (NAIL104)

Pravděpodobnostní grafické modely

Marta Vomlelová
KTIML, S303
marta@ktiml.mff.cuni.cz
http://ktiml.mff.cuni.cz/~marta

Literature

Key:

1. Finn V. Jensen, Thomas D. Nielsen: Bayesian Networks and Decision Graphs, Springer 2007
2. Højsgaard, Søren, Edwards, David, Lauritzen, Steffen : Graphical Models with R, Springer 2012

Additional:

1. http://people.math.aau.dk/~sorenh/misc/2012-Oslo-GMwR/GMwR-notes.pdf
2. John Lafferty, Andrew McCallum, Rernando Pereira: Conditional random fields: Probabilistic models for segmenting and labeling sequence data, Morgan Kaufmann 2001, pp. 282—289
3. Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. Artificial Intelligence, Volume 101, pp. 99-134, 1998
4. S. Russell, P. Norvig: Artificial Intelligence; A Modern Approach, 1995.

Software

- Genie
- R

```
library(gRain)
library(gRbase)
library(RBGL)
library(igraph)
library(ggm)
library(Rgraphviz)
library(depmixS4)
```

- Hugin
- MSBN
- OpenMarkov

Course Outline

- Bayesian Networks (BN)
- Precise Optimized Evaluation (Junction Tree Evaluation)
- Approximate Evaluation
- Probabilistic Classifiers (Naive Bayes, Tree Augmented Naive Bayes)
- Continuous Variables, Conflict Measure, OOBNs, DBNs
- Parameter Learning in BNs
- Structure Learning of BNs
- Decision Graphs (=Influence Diagrams)
- POMDPs

Might be included

- Log-linear Models
- Markov Decision Processes and Reinforcement Learning
- Conditional Markov Fields.

Probability Theory Basics

- **Random variables** Gender, Hair with binary or discrete **domain** *sp*(),
- Cartesian product of random variables $\mathcal{X}$ is the **sample space** (**obor hodnot**) of all **elementary events**.
- Potential set $\mathcal{P}(\mathcal{X})$: set of all subsets is a set of **events** (**jevy**).
- **Probability distribution** P is a mapping $P : \mathcal{P}(\mathcal{X}) \to \langle 0, 1 \rangle$, such that:
  - $P(\mathcal{X}) = 1$ and
  - $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}, \mathcal{A} \cap \mathcal{B} = 0 \Rightarrow P(\mathcal{A} \cup \mathcal{B}) = \mathcal{P}(\mathcal{A}) + \mathcal{P}(\mathcal{B})$.
- Probability distribution is fully specified by the assignment to elementary events.

```
p.HG<-parray(c("Hair","Gender"),
 levels=list(c("none", "short","long"),c("male","female")),
 values=c(.1,.3,.1,.01,.29,.2))
```

p.HG	male	female
none	0.10	0.01
short	0.30	0.29
long	0.10	0.20

```
> p.HG[,1]
none short long
0.1 0.3 0.1
```

```
> p.HG[,2]
none short long
0.01 0.29 0.20
```

Marginalization

- We are interested in a subset of random variables $\{Hair\}$.
- We *sum over* other variables $\{Gender\}$, precisely

$$P(hair) = \Sigma_{gender} P(hair, gender) = P(\{(h,g); h = hair\})$$

- It is often written in multi-dimensional form

$$P(Hair) = \Sigma_{Gender} P(Hair, Gender).$$

We say **we marginalize to *Hair*** or the variable *Gender* **is marginalized out**.

p.HG	male	female
none	0.10	0.01
short	0.30	0.29
long	0.10	0.20

```
> tableMargin(p.HG, "Hair")
Hair
none short long
0.11 0.59 0.30
```

```
> tableMargin(p.HG, "Gender")
Gender
male female
0.5 0.5
```

Marginalization Example

```
p.HGS<-parray(c("Hair","Gender","MFF"),
levels=list(c("none", "short","long"),c("male","female"),c("no",'yes')),
values=c(.1,.3,.1,.01,.19,.2,.01,.06,.02,.001,.005,.004))
```

MFF=no	male	female
none	0.10	0.01
short	0.30	0.19
long	0.10	0.20

MFF=yes	male	female
none	0.010	0.001
short	0.060	0.005
long	0.020	0.004

```
> tableMargin(p.HGS, c("Hair","MFF"))
MFF
Hair no yes
none 0.11 0.011
short 0.49 0.065
long 0.30 0.024
```

```
> tableMargin(p.HGS,
c("Gender"))
Gender
male female
0.59 0.41
```

Evidence (Observations)

- We **enter the evidence** Hair=short by multiplying all inconsistent events by 0, i.e. we cut of the only layer that belongs to the observed evidence.

MFF=no	male	female
none	0.10	0.01
short	0.30	0.19
long	0.10	0.20

MFF=yes	male	female
none	0.010	0.001
short	0.060	0.005
long	0.020	0.004

```
> p.HGS[2,,]
MFF
Gender no yes
male 0.30 0.060
female 0.19 0.005
```

An Algebra of Potentials

- Our tables need not be (conditional) probabilities and they are generally called **potentials**(**tabulky**).

Definition (potential)

A **potential** ϕ is a real-valued function over a domain of finite variables $\mathcal{X}$:

$$\phi : sp(\mathcal{X}) \to \Re.$$

The domain of the potential is denoted by $dom(\phi)$.

- For example, $dom(P(A, B|C)) = \{A, B, C\}$.

Definition

- **Extending** a potential $\phi(A, B)$ to the domain $\{A, B, C\}$ means copying the potential for every possible value of $c \in sp(C)$.
- **Multiplication, division** of potentials can be defined as extending both potentials to the union of their domains and then multiplying pointwise values corresponding to the same element of the domain.

Conditional Probability

- Given the information about Hair, what is the probability of resulting random events?
- **Conditional probability distribution** $P(A|B)$ is a probability distribution that fulfills the *fundamental rule*: $P(A|B)P(B) = P(A,B)$.
- If $P(B = b) > 0$ for every $b \in sp(B)$ then $P(A|B) = P(A,B)/P(B)$.

p.HG	male	female
none	0.10	0.01
short	0.30	0.29
long	0.10	0.20

```
> tableMargin(p.HG, "Hair")
Hair
none short long
0.11 0.59 0.30
```

```
> tableDiv(p.HG, tableMargin(p.HG,"Hair"))
Gender
Hair male female
none 0.9090909 0.09090909
short 0.5084746 0.49152542
long 0.3333333 0.66666667
```

```
tableMult(tableDiv(p.HG, tableMargin(p.HG,"Hair")),tableMargin(p.HG,"Hair"))
```

Bayes Formula

- **Bayes formula:** For any non-zero $P(B|\mathcal{C})$ following holds:

$$P(A|B,\mathcal{C}) = \frac{P(B|A,\mathcal{C}) \cdot P(A|\mathcal{C})}{P(B|\mathcal{C})} = \frac{P(A,B|\mathcal{C})}{P(B|\mathcal{C})} = $$
$$= \frac{P(B|A,\mathcal{C}) \cdot P(A|\mathcal{C})}{\sum_{a \in sp(A)} P(B|A=a,\mathcal{C}) \cdot P(A=a|\mathcal{C})}.$$

Example:

- We have two random variables, Rain, Wet grass, both either yes or no.
- We know:
  - If it rains (Rain=yes), then the grass is wet with probability $\frac{3}{4}$,
  - if it does not rain, the grass is wet with probability $\frac{1}{8}$,
  - The probability of the rain is $\frac{1}{3}$ (**prior**), (**apriorní pr.**).
  - We see the grass is wet in the morning.
- What is the probability it has been raining during the night? (**posterior**) ($\frac{3}{4}$)

Example: Algebra of Potentials

- If it rains (Rain=yes), then the grass is wet with probability $\frac{3}{4}$,
- if it does not rain, the grass is wet with probability $\frac{1}{8}$,
- The probability of the rain is $\frac{1}{3}$.
- We see the grass is wet in the morning.

```
> p.Rain
Rain
yes no
0.3333333 0.6666667
```

p.Wet	yes	no
yes	0.75	0.12
no	0.25	0.88

```
yn=c("yes","no")
p.Rain<-parray(~Rain, values=c(1,2), levels=list(yn), normalize='all')
p.Wet<-parray(~Wet:Rain, values=c(3,1,1,7), levels=list(yn,yn),
normalize='first')
e.Wet<-parray(~Wet,values=c(1,0),levels=list(yn),normalize='none')
un.normalized<- tableMult(tableMult(p.Wet,p.Rain),e.Wet)
un.normalized/sum(un.normalized)
```

Exercise: Monty Hall

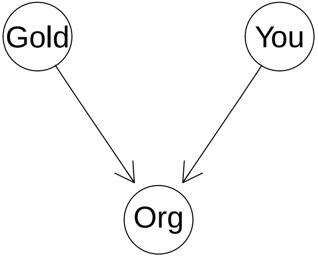
- There are three doors A, B, C. Exactly behind one of them is hidden gold.
- You select one door (C).
- The organizer opens some other door (A, B), there is no gold behind them.
- You may change your mind: select any of closed doors.
- What is the probability of the gold being behind each door? Calculate it.

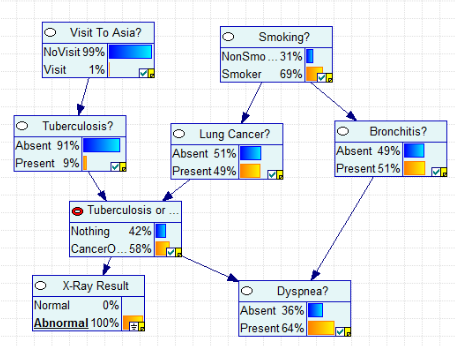
```
v.ABC=c('A',"B",'C')
m.Gold<-parray(~Gold, values=c(1,1,1), levels=list(v.ABC), normalize='all')
m.You<-parray(~You, values=c(0,1,0), levels=list(v.ABC), normalize='none')
m.Org<-parray(~Org:Gold:You, values=c(0,1,1, 0,0,1, 0,1,0,
0,0,1, 1,0,1, 1,0,0,
0,1,0, 1,0,0, 1,1,0
),
levels=list(v.ABC,v.ABC,v.ABC), normalize='first')
tableMargin(tableMult(tableMult(m.Gold,m.Org),m.You),c("Gold",'Org'))
bn=grain(compileCPT(list(m.Gold,m.You, m.Org)));plot(bn)
```

Monty 2

```
> tableMargin(tableMult(tableMult(m.Gold,m.Org),m.You),c("Gold",'Org'))
Org
Gold A B C
A 0.0000000 0 0.3333333
B 0.1666667 0 0.1666667
C 0.3333333 0 0.0000000
```

Graph structure





- The graph structure represents conditional independencies of variables.
- It enables simplification of the model, calculations, estimates.

Independence

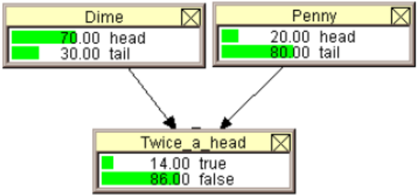
Definition (Independence)

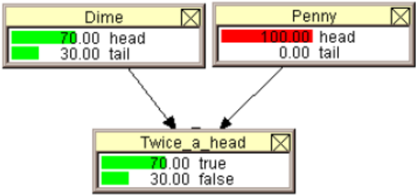
Two random variables A, B are independent iff

$$\forall a \in sp(A), b \in sp(B) : P(A = a, B = b) = P(A = a) \cdot P(B = b).$$

Lemma (independence)

Variables A, B are independent iff
$\forall a \in sp(A), b \in sp(B) : P(A = a|B = b) = P(A = a).$





Example: Idependence, Depencence, Conditional Independence

Fig.: Biased coins: Second flop with the same coin depends on the first flop result.

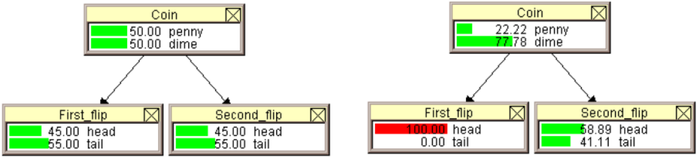
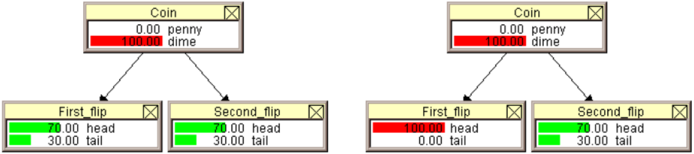


Fig.: Knowing the coin bias, first and second flip are conditionally independent.



Conditional Independence $A \perp\!\!\!\perp B|\mathcal{C}$

Definition (Conditional Independence)

Random variables A, B are **conditionally independent given** $\mathcal{C}$, $A \perp\!\!\!\perp B|\mathcal{C}$, $\mathcal{C} \subseteq V \setminus \{A, B\}$ iff $\forall a \in sp(A), b \in sp(B), c = \langle c_1, \ldots, c_k \rangle \in \mathcal{C}$

$$P(A = a, B = b|\mathcal{C} = c) = P(A = a|\mathcal{C} = c) \cdot P(B = b|\mathcal{C} = c).$$

- Some conditional independencies can be read from the graph structure.
- There may be more independencies hidden in the probability distribution.

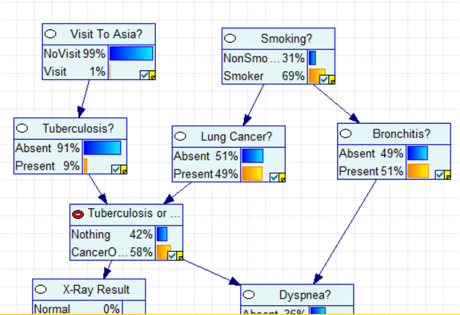
Bayesian Networks

Definition (Bayesian Network)

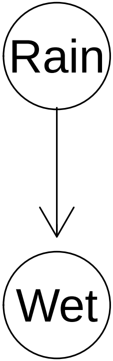
Bayesian network is a pair (G, P), where

- $G = (V, E)$ is a DAG (directed acyclic graph with set of vertexes V and set of edges E) and
- P is a list of conditional probability distributions such that for every vertex X there is a probability table $P(X|pa(X))$ where $pa(X)$ denotes the set of parents of a vertex in G.

The graph represents conditional independencies of the join probability distribution $\Pi_{X \in V} P(X|pa(X))$.



BN Example: Wet grass




```
require(gRain)
yn=c("yes","no")
p.Rain<-cptable(~Rain, values=c(1,2), levels=yn,
normalize=TRUE)
p.Wet<-cptable(~Wet+Rain, values=c(3,1,1,7),
levels=yn, normalize=TRUE)
plist<-compileCPT(list(p.Rain, p.Wet))
bnet<-grain(plist)
plot(bnet)
bnet.f<-setEvidence(bnet,
nodes=c('Wet'),state=c('yes'), propagate=TRUE)
querygrain(bnet.f,nodes=c('Rain'))

> querygrain(bnet.f,nodes=c('Rain'))
$Rain
Rain
yes no
0.75 0.25
```

Chain Rule (**řetězcové pravidlo**)

Definition (Topological Ordering)

Topological ordering of a directed graph is a linear ordering of vertices such that for every edge the index of the head is smaller then the tail index.

- For *any* probability distribution on topologically ordered variables $A_1, \ldots, A_k$ we can write:

$$P(A_1, \ldots, A_k) = P(A_1) \cdot \Pi_{j=2,\ldots,k} P(A_j | A_1, \ldots, A_{j-1})$$ **chain rule**

- For topologically ordered Bayesian network we can write:

$$P(A_1, \ldots, A_k) = P(A_1) \cdot \Pi_{j=2,\ldots,k} P(A_j | pa(A_j))$$

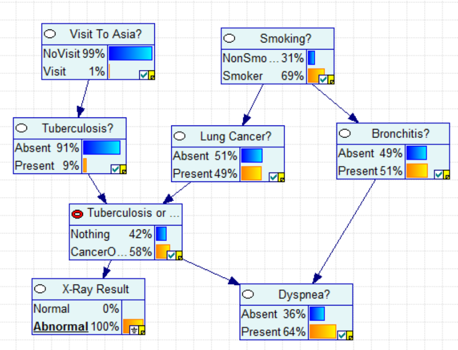
- that is the structure encodes the set of conditional independencies $i < j$, $A_i \notin pa(A_j)$:

$$A_i \perp\!\!\!\perp A_j | pa(A_j).$$

Exercise

- Find a topological ordering.
- Find as many conditional independencies as you can.
- Does the set of independencies depend on selected ordering?

Asia, Smoking, Bronchitis $A \perp\!\!\!\perp_d B|S$ is not discovered in the ordering Smoking, Bronchitis, Asia.



d-separation

Definition (d-separation)

Two nodes $A, B \in V$ of a BN with $G = (V, E)$ a **d-separated** $A \perp\!\!\!\perp_d B|\mathcal{C}$ by $\mathcal{C} \subseteq V \setminus \{A, B\}$ iff each (non-oriented) path from A to B contains at least one node *Blocking* of one of following types:

- $Blocking \in \mathcal{C}$ and *Blocking* is a tail of at least one edge of the path,
- $\{Blocking, succ(Blocking)\} \cap \mathcal{C} = \emptyset$ and *Blocking* is the head of both edges of the path.

Theorem (d-separation)

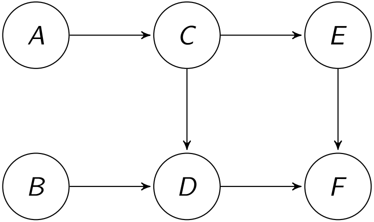
If A, B are d-separated given $\mathcal{C}$ $(A \perp\!\!\!\perp_d B|\mathcal{C})$ in BN B, then they are conditionally independent $(A \perp\!\!\!\perp B|\mathcal{C})$.

d-separation

Definition (d-separation)

Two nodes $A, B \in V$ of a BN with $G = (V, E)$ a **d-separated** $A \perp\!\!\!\perp_d B|\mathcal{C}$ by $\mathcal{C} \subseteq V \setminus \{A, B\}$ iff each (non-oriented) path from A to B contains at least one node *Blocking* of one of following types:

- $Blocking \in \mathcal{C}$ and *Blocking* is a tail of at least one edge of the path,
- $\{Blocking, succ(Blocking)\} \cap \mathcal{C} = \emptyset$ and *Blocking* is the head of both edges of the path.



Does hold following?

- $E \perp\!\!\!\perp_d B$
- $E \perp\!\!\!\perp_d D$
- $E \perp\!\!\!\perp_d D|A$
- $E \perp\!\!\!\perp_d D|C$
- $E \perp\!\!\!\perp_d D|\{C, F\}$
- $E \perp\!\!\!\perp_d B|F$

yes, no, no, yes, no, no

Summary

Summary

- Conditional probability
- Algebra on potentials: multiplication, division, marginalization
- Conditional independence
- Bayesian network
- d-separation.

Thank you for attention.