

Bayesian Networks and Expert Systems

Steffen Lauritzen, University of Oxford

Graphical Models, Lecture 11, Michaelmas Term 2010

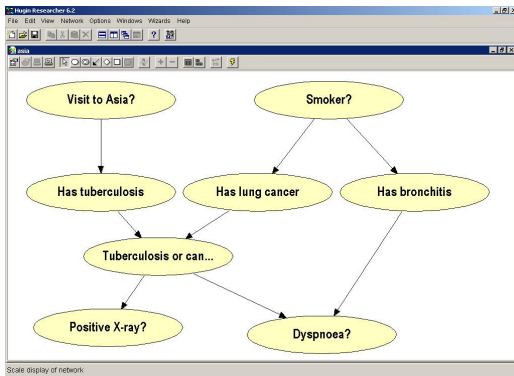
November 15, 2010

A *directed acyclic graph* \mathcal{D} over a finite set V is a simple graph with all edges directed and *no directed cycles*. We use DAG for brevity.

Absence of directed cycles means that, *following arrows in the graph, it is impossible to return to any point*.

Graphical models based on DAGs have proved fundamental and useful in a wealth of interesting applications, including expert systems, genetics, complex biomedical statistics, causal analysis, and machine learning.

Example of a directed graphical model



Local directed Markov property

A semigraphoid relation \perp_{σ} satisfies *the local Markov property* (L) w.r.t. a directed acyclic graph \mathcal{D} if

$$\forall \alpha \in V : \alpha \perp_{\sigma} \{ \text{nd}(\alpha) \setminus \text{pa}(\alpha) \} \mid \text{pa}(\alpha).$$

Here $\text{nd}(\alpha)$ are the *non-descendants* of α .

Ordered Markov property

Suppose the vertices V of a DAG \mathcal{D} are *well-ordered* in the sense that they are linearly ordered in a way which is compatible with \mathcal{D} , i.e. so that

$$\alpha \in \text{pa}(\beta) \Rightarrow \alpha < \beta.$$

We then say semigraphoid relation \perp_{σ} satisfies the *ordered Markov property* (O) w.r.t. a well-ordered DAG \mathcal{D} if

$$\forall \alpha \in V : \alpha \perp_{\sigma} \{\text{pr}(\alpha) \setminus \text{pa}(\alpha)\} \mid \text{pa}(\alpha).$$

Here $\text{pr}(\alpha)$ are the *predecessors* of α , i.e. those which are before α in the well-ordering..

The global Markov property

A semigraphoid relation \perp_{σ} satisfies the *global Markov property* (G) w.r.t. \mathcal{D} if

$$A \perp_{\mathcal{D}} B \mid S \Rightarrow A \perp_{\sigma} B \mid S.$$

It holds for any DAG \mathcal{D} and any semigraphoid relation \perp_{σ} that all directed Markov properties are equivalent:

$$(G) \iff (L) \iff (O).$$

Factorisation

A probability distribution P over $\mathcal{X} = \mathcal{X}_V$ *factorizes* over a DAG \mathcal{D} if its density or probability mass function f has the form

$$(F^*) : \quad f(x) = \prod_{v \in V} f(x_v | x_{\text{pa}(v)}),$$

Markov properties and factorization

In the directed case it is essentially *always true that (F) holds if and only if $\perp\!\!\!\perp_P$ satisfies (G)*,
so all directed Markov properties are equivalent to the factorization property!

$$(F) \iff (G) \iff (L) \iff (O).$$

An *expert system* attempts to crystallise and codify knowledge of experts into a tool, usable by non-specialist.

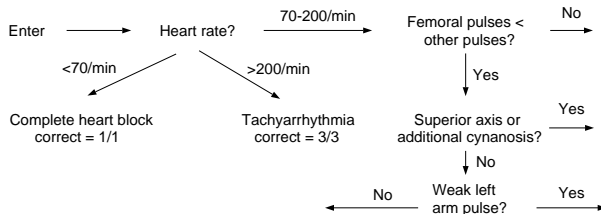
The *knowledge base* encodes the knowledge of the domain.

The *inference engine* consists of algorithms for processing knowledge base and specific information to obtain conclusions.

Classical expert systems *make model of expert*.

Probabilistic expert systems *model the domain* and use Bayesian reasoning.

Classification trees



Not necessarily computerized. Can be constructed using e.g. CART.

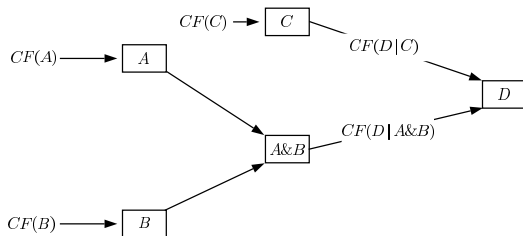
Production systems

Uses *rules*: IF (A_1 & A_2 & ... & A_k) THEN B ; for example

- ▶ IF the animal has hair THEN it is a mammal.
- ▶ IF the animal gives milk THEN it is a mammal.
- ▶ IF the animal has feathers THEN it is a bird.
- ▶ IF the animal flies AND it lays eggs THEN it is a bird.

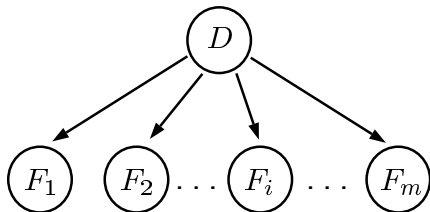
Inference “chaining” (forwards and backwards)

Certainty factors



Production rules with “certainty factor”. Need calculus to combine certainty factors.

Naive Bayes

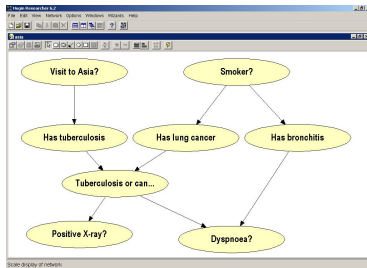


Disease probabilities D used. F_i are findings and $P(F_i | D)$ are specified.

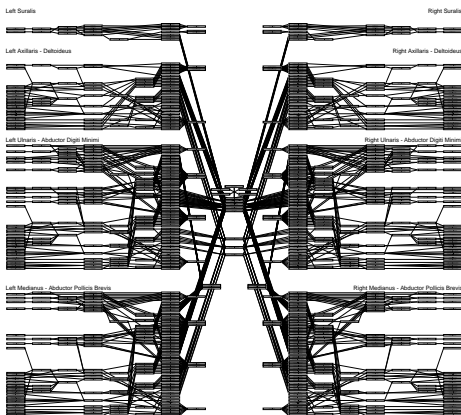
$P(D | F_1, \dots, F_m)$ is calculated by Bayes' formula.

Directed graphical model, to be used for reasoning.

“Bayesian” because it reasons “reversely”, from symptoms to causes, in contrast to feedforward neural networks which were common when BNs were introduced.



MUNIN



Formal definition

A *Bayesian network* represents the *knowledge base* as a directed graphical model:

- ▶ A Directed Acyclic Graph $\mathcal{D} = (V, E)$;
- ▶ Nodes V represent (random) variables $X_v, v \in V$;
- ▶ Specify for all $v \in V$: $p(x_v | x_{\text{pa}(v)})$;
- ▶ Joint distribution is then $p(x) = \prod_{v \in V} p(x_v | x_{\text{pa}(v)})$.

Inference engine exploits junction tree algorithm to calculate $p(x_v | x_E^*)$ for $E \subseteq V$ since $p(x_E^*) = \sum_{y: y_E = x_E^*} p(y)$ has *too many terms*.

Forensic identification

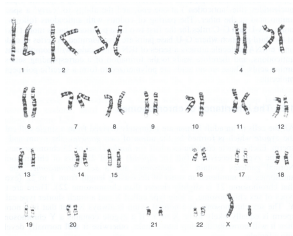
Disputed paternity: Is individual A the father of individual B ?

Immigration cases: Is A the mother of B ? Are A and B related at all? If so, how?

Criminal cases: Did person A contribute to a given stain, found at the scene of the crime? Who contributed to the stain?

Disasters: Was A among the individuals found in a grave? How many of a named subset of individuals were in the grave? Who were found in a grave?

Human chromosomes



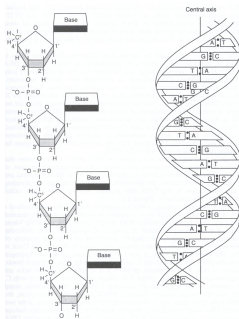
23 pairs of chromosomes in nucleus of human cell.

One pair determines gender: male XY, female XX. Other 22 are *homologous* pairs.

All are DNA molecules.

DNA molecules

A double helix composed by 4 different nucleotides:
 C, A, G, and T, binding in pairs C–G and A–T.



STR markers

An area on a chromosome is a *locus* and the DNA composition on that area is an *allele*.

A locus thus corresponds to a (random) variable and an allele to its realised state.

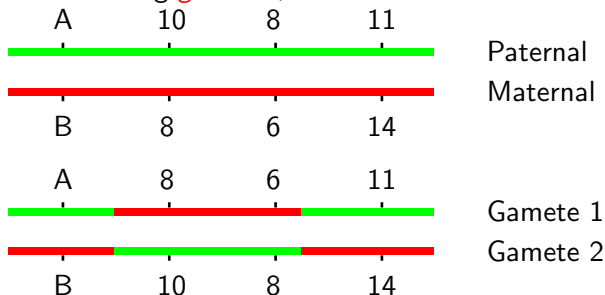
A DNA *marker* is a known locus where the allele can be identified in the laboratory.

Short Tandem Repeats (STR) are markers with alleles given by integers. If an STR allele is 5, a certain word (e.g. CAGGTG) is repeated exactly 5 times at that locus:

... CAGGTGCAGGTGCAGGTGCAGGTGCAGGTG...

Inheritance of DNA

The homologous chromosome pairs are inherited through the process of forming *gametes*, known as *meiosis*:



A child receives one randomly chosen gamete from each parent to form a new homologous pair.

DNA profile and genotypes

The *genotype* of an individual at a given locus is the unordered pair of alleles at that locus. One cannot measure which allele originated from the mother and which from the father.

The genotype is typically reported as $(12, 14)$ or (A, B) , so that the smallest is mentioned first.

A *DNA profile* consists of measurements of the genotype at a number of marker loci. Standard kits use 9 or 10 markers, but occasionally more markers are measured.

Markers are generally chosen on different chromosomes, to avoid problems of *linkage*, i.e. dependence created in the process of meiosis.

Classical paternity case

- ▶ DNA profiles of *mother*, a *child*, and a *male* individual, known as the *putative father*. Denote this *evidence* by E .
- ▶ Query Q to be investigated :
Is the putative father equal to the true father?
- ▶ Weight of evidence reported as a *likelihood ratio*:

$$L = \frac{P(E \mid Q = \text{true})}{P(E \mid Q = \text{false})}.$$

Using Bayesian networks

- ▶ Make BN with $P(E | Q = \text{true})$ determined by laws of inheritance and $P(E | Q = \text{false})$ assuming random genes of putative father.
- ▶ Let $P(Q = \text{true}) = P(Q = \text{false})$ so we have

$$L = \frac{P(E | Q = \text{true})}{P(E | Q = \text{false})} = \frac{P(Q = \text{true} | E)}{P(Q = \text{false} | E)}$$

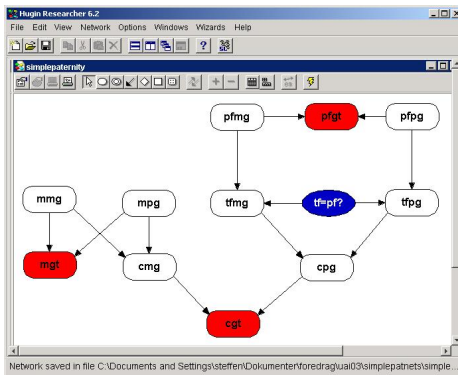
and compute the latter by probability propagation.

We can make a network for each independent marker and multiply likelihood ratios, or we can make a network incorporating all markers at once.

Object-oriented specification of BN

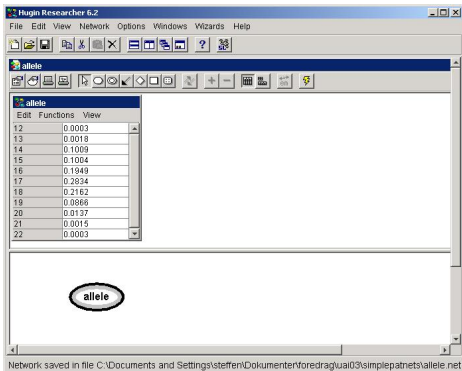
- ▶ Objects are *instances* of BNs of certain class
- ▶ Objects have *input nodes* and *output nodes*, and also ordinary BN nodes
- ▶ Instances of a given class have *identical conditional probability tables* for non-input nodes
- ▶ Objects are connected by directed links from output nodes to input nodes. The links represent *identification* of nodes, so nodes must be of same type and have the same states.

OoBN for paternity case: single marker



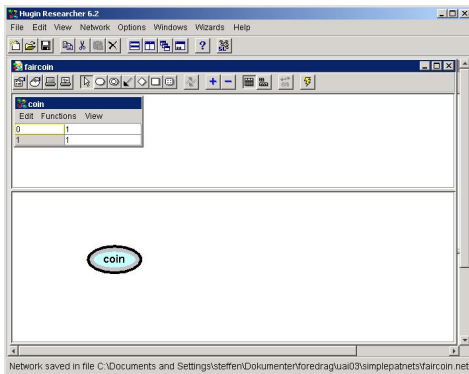
Each node represents itself a Bayesian network.

Allele



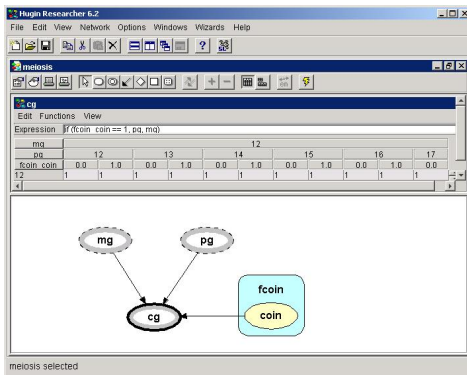
This class represents a randomly chosen allele

Faircoin



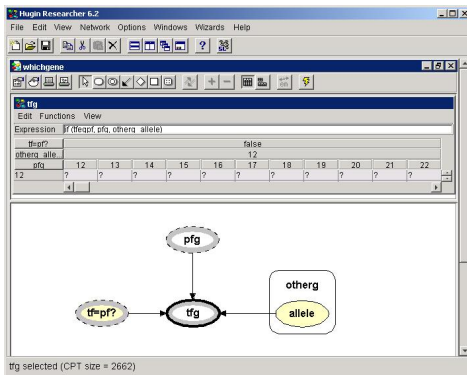
Represents a coin, used to choose allele under meiosis

Meiosis



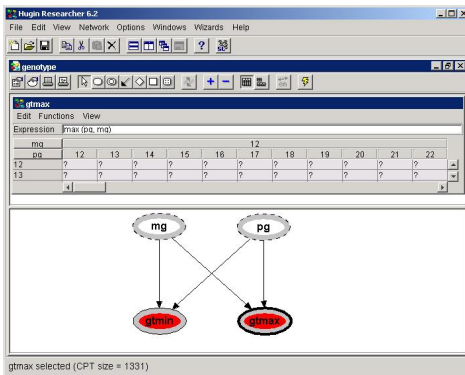
Represents the transmission of allele through meiosis

Who is the father?



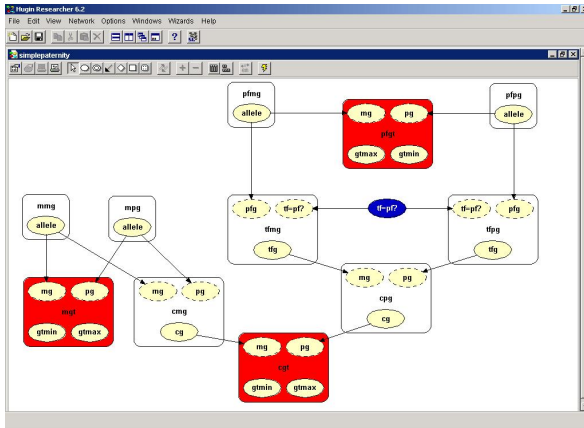
Is the allele from the putative father or random?

Genotype

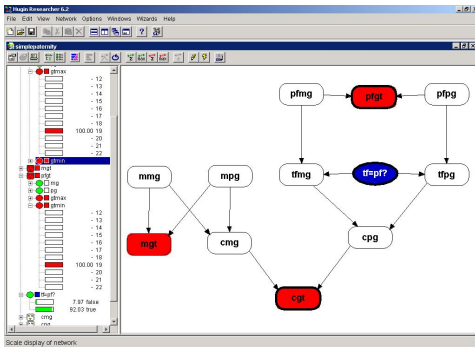


Observation of the smallest and largest allele

Expanded OOBN

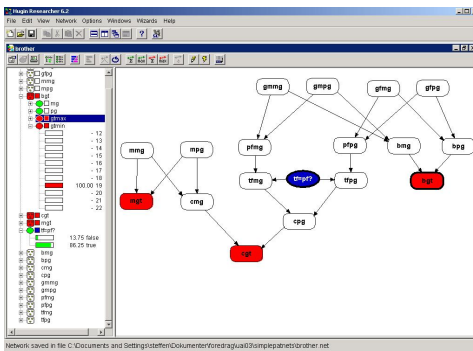


Results



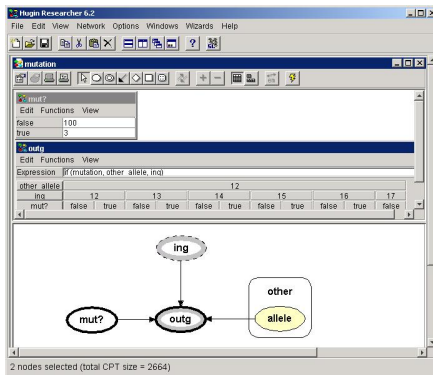
Mother: (15, 16), child: (15, 19), male: (19, 19);
 $L = 92.03/7.97 = 11.55$.

Indirect evidence: only brother available



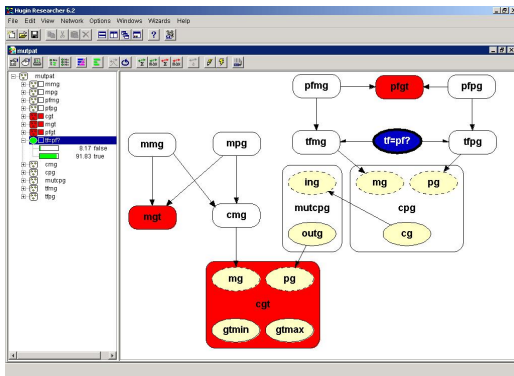
Brother of pf: (19, 19); $L = 86.25/13.75 = 6.27$.

Mutation



Possible mutation in transmission of alleles

Mutation in male germline



$$L = 91.83/8.17 = 11.24.$$

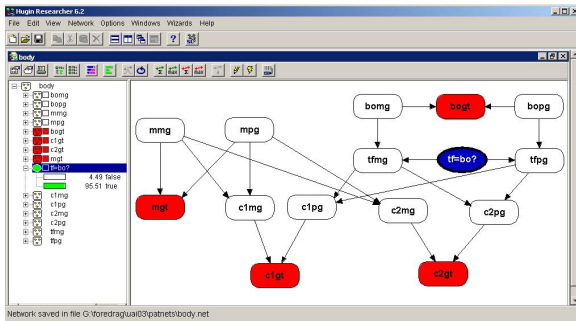
Body identification

Identification of a *single* dead body is not very different for paternity cases.

For example, if a missing person is known to be a specific member of a family (e.g. the father of two children) and DNA profiles can be found for the body, the mother, and the two children, a minor modification of the paternity network yields the solution.

Problems of identification involving *more than one* body, such as in mass graves and in disasters are more difficult because of their complexity.

Unidentified body



Is the body father of the two children? Same data as for paternity.
Second child (16, 19); $L = 95.51/4.49 = 21.27$.