Harvard School of Public Health
Boston, June 16, 2005

Influence diagrams vs. decision trees for medical decision analysis

Francisco Javier Díez
Dept. Artificial Intelligence. UNED
Madrid, Spain
www.ia.uned.es/~fjdiez

OVERVIEW

- Introduction: probabilistic models for medical diagnosis
  - naive-Bayes method
  - Bayesian networks
  - canonical models
- Influence diagrams for decision analysis
  - the project
  - probabilistic reasoning with Elvira: conditional independence, *d*-separation, the Markov property
  - real-world examples: Prostanet, Nasonet, Hepar II
- Causal Bayesian networks for epidemiological research
  - use of Elvira as a pedagogical tool
- Conclusions

A medical problem

- Disease X
  - Prevalence: $P(+x)$ = 0'14
- Therapy D
  - Utility:

$u(x, d)$	$+x$	$\neg x$
$+d$	8	9
$\neg d$	3	10

- Test Y
  - Sensitivity: $P(+y|+x)$ = 0'91
  - Specificity: $P(\neg y|\neg x)$ = 0'97
  - Cost: $u_{\text{test}}(x, d) = u_{\text{not-test}}(x, d) - 0'2$
- Decisions:
  - Is it worthy to do the test?
  - In what cases should we apply the therapy?

The same problem, with a symptom

- Disease X prevalence $P(+x)$
- Therapy D utility $u(x, d)$
- Test Y sensit. $P(+y|+x)$, spec. $P(\neg y|\neg x)$, cost c
- Symtom S sensit. $P(+s|+x)$, spec. $P(\neg s|\neg x)$
- Decisions:
  - Is it worthy to apply the test to a symptomatic patient?
  - Is it worthy to apply the test to an asymptomatic patient?
  - In what cases should we apply the therapy?

Naive-Bayes method for probabilistic diagnosis

- *n* diagnoses, *m* variables representing possible findings
- 1st hypothesis: diagnoses are mutually exclusive
  (i.e., the patient has at most one disease)
- 2nd hypothesis: findings are conditionally independent

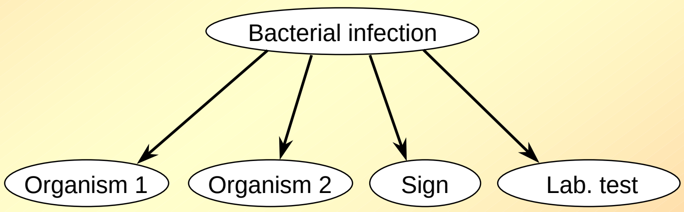
$$P(f_1, \ldots, f_m \mid d_i) = P(f_1|d_i) \cdot \ldots \cdot P(f_m|d_i)$$

- Bayes' theorem (naive method)

$$P(d_i|f_1, \ldots, f_m) = \frac{P(f_1|d_i) \cdot \ldots \cdot P(f_m|d_i) \cdot P(d_i)}{\sum_j P(f_1|d_j) \cdot \ldots \cdot P(f_m|d_j) \cdot P(d_j)}$$

Limitations of the naive-Bayes method

- Some times the diagnoses are not mutually exclusive
- In general, findings are not conditionally independent



- Solution: Bayesian networks

Bayesian network

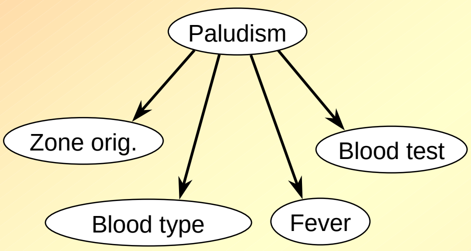
- Elements
  - Set of variables $\{X_i\}$
  - Acyclic directed graph
    - Each node in the graph represents a variable X_i
  - Conditional probability distribution (table) for each variable: $P(x_i \mid pa(x_i))$
    - For a node without parents: $P(x_i \mid pa(x_i)) = P(x_i)$
- Result: join probability distribution

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i | pa(x_i))$$

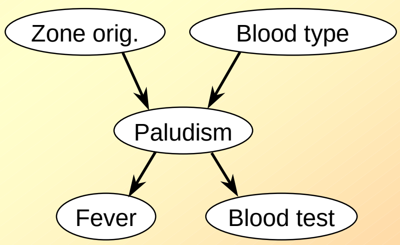
- Markov property
  - Given a set of variables $\{Y_j\}$ such that no Y_j is a descendant of X_i in the graph, it hods that

$$P(x_i \mid pa(x_i), y_1, \ldots, y_n) = P(x_i \mid pa(x_i))$$

Naive-Bayes

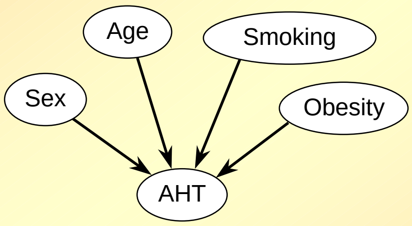


Bayesian network



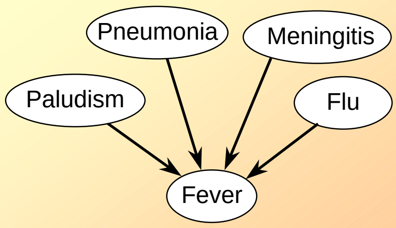
General model

- Probability table:
  $P(y \mid x_1, \ldots, x_n)$
- Factors that influence the prob. of X



Noisy OR

- Efficiency of each link:
  c_i
- Causes that may produce X



Advantages of Bayesian networks

- BNs are usually causal models
  - closer to doctors' reasoning
  - probabilities are in general easier to obtain
- BNs do not assume conditional independence
- BNs can diagnose several diseases simultaneously
- BNs can be learnt from databases
- BNs can combine objective probabilities (frequencies) with subjective estimates
- Specific methods for sensitivity analysis in BNs

(cont.)

Advantages of Bayesian networks (cont.)

- Canonical models facilitate the construction of BNs
  - when the BN is built from human knowledge (subjective estimates)
  - and also when a BN is learnt from a database
  - Díez, Druzdzel. Canonical probabilistic models for knowledge engineering. 2005
- Canonical lead to more efficient inference
  - Díez, Druzdzel. Computational properties of canonical probabilistic models. 2005
- Explanation of reasoning for BNs
  - Lacave, Díez. A review of explanation methods for Bayesian networks. 2002.
- All these advantages are shared with influence diagrams.

Influence diagrams for medical decision making

Use of variables in decision analysis

- The most difficult issue in building decision trees is the use of variables: “trees usually have bugs”
- Three reasons for using variables in decision trees
  1. When the probabilities of several branches depend on a certain parameter (e.g., prevalence, sensitivity, specificity)
  2. Utility functions that depend of several parameters
  3. Sensitivity analysis
- Are variables necessary for IDs?
  1. Each parameter appears only once in the tree
  2. Supervalue nodes allow for utility combination without variables (at least, without extra definitions)
  3. Named variables might be useful only as labels for 2- and 3-way sensitivity analysis

Advantages of influence diagrams (1)

- IDs are more compact
- Explicit representation of causality
- IDs are much easier to <u>build</u> than DTs
  - IDs use direct probabilities (prevalence, sensitivity, specificity...) and costs (mortatility, morbidity, economic cost...)
  - ID can use canonical models (noisy OR, noisy AND, etc)
  - Each parameter only once in the ID (are variables necessary?)
  - IDs can use super-value nodes: explicit combination of utilities
  - “All trees have bugs” (*Primer on MDA*); IDs in general do not.
- No external computation of probabilities is required
  - Algorithms of evaluation compute them when they are needed
- IDs are much easier to <u>modify</u> than DTs

Advantages of influence diagrams (2)

- Having an ID, we can immediately obtain a DT
  - but is the reverse not true in general
- Two possibilities of evaluation:
  1. expand an equivalent decision tree
     - exponential complexity (time and space)
     - many problems cannot be solved by this method
  2. direct algorithms
     - direct manipulation of the graph and/or potentials of the ID
     - similar to the best algoritms for Bayesian networks
     - canonical models can lead to more efficient evaluations
- More possibilities of explanation of reasoning
  - computation of posterior probabilities on the ID (as if it were a BN)
  - value of information (EVPI and other measures) can be computed easily
  - other methods from Bayesian networks and qualitative prob. networks

Limitations of IDs

- Limitations of standard IDs
  - Standard IDs are symmetric
    - They must have artificial “non-observed” or “impossible” states
  - Some software tools (e.g., TreeAge) allow asymmetry
    - but sometimes “arcs of assymmetry” are not intuitive
- Limitations of current software packages
  - Very few packages allow sensitivity analysis directly from IDs
  - No software package allows C.E.A. directly from IDs
- Solutions
  - More flexible representation models
    - Jensen, Nielsen, Shenoy. Sequential influence diagrams. Proc. of PGM-04.
  - More powerful software tools (e.g., future versions of Elvira)

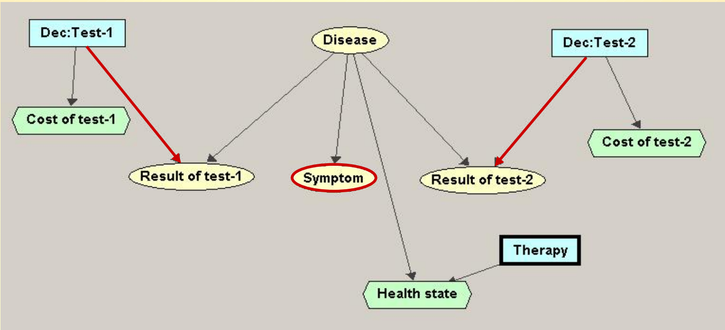
The two-test problem

- Disease X prevalence $P(+x)$
- Therapy D utility $u(x, d)$
- Test Y_1 sensit. $P(+y_1|+x)$, spec. $P(\neg y_1|\neg x)$, cost c_1
- Test Y_2 sensit. $P(+y_2|+x)$, spec. $P(\neg y_2|\neg x)$, cost c_2
- Decisions:
  - Should we do any test?
  - Which one should be done first?
  - If the first test is positive, should we do a second test?
  - If it is negative, should we do the second test?
  - In what cases should we apply the therapy?

This problem cannot be solved in a natural way with standard IDs!

Decision-analysis networks (DANs)

- Very similar to IDs, but:
  - DANs do not require a total ordering of decisions
  - Some nodes are marked as "*known since the beginning*" (for instance, symptoms)
  - DANs may have *revelation arcs*: "Dec:Test"→"Result of test"



Conclusion: utility of Elvira

- Tool for building Bayesian networks and influence diagrams
  - Examples: Prostanet, Nasonet, Hepar II, etc.
- Pedagogical tool
  - illustrate *d*-separation and the Markov property
  - illustrate the problems of analyzing causality given observational data
- Future: tool for epidemiological studies
  - input: causal graph + observational database
  - first step: eliminate unmeasured variables from the graph result: statistical graph
  - second step: obtain the conditional probabilities for the statistical graph from the database
  - output: causal risk ratios

IDs in the literature on MDM

- Journal: *Medical Decision Making*
  - very few papers using IDs in their analyses
- Books that mention decision trees and do not mention IDs
  - Weinstein, Fineberg. *Clinical Decision Making*. 1980.
    [Influence diagrams were first published in (Howard and Matheson, 1984)]
  - Sox et al. *Medical Decision Making*. 1988.
  - Sloan (ed.). *Valuing Health Care*. 1995.
  - Gold et al. *Cost-Effectiveness in Health and Medicine*. 1996.
  - Drummond et al. *Methods for the Economic Evaluation of Health Care Programs*. 2nd ed., 1997.
  - Sacket et al. *Evidence-Based Medicine*. 1997 (and two other books on EBM).
  - Petitti. *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis*. 2nd ed., 2000.
  - Drummond, McGuire (eds.). *Economic Evaluation in Health Care*. 2001.

IDs in the literature on MDM (cont.)

Books that mention decision trees and do not mention IDs (cont.)
- Huning, Glasziou. *Decision Making in Health and Medicine*. 2001.
- Haddix et al. *Prevention Effectiveness*. 2nd ed., 2003.

- One book that mentions IDs
  - Chapman, Sonnenberg (eds.). *Decision Making in Health Care*. 2000 (5 pages out of 421).
- Another books that mentions IDs
  - Muenning. *Designing and Conducting Cost-Effectiveness Analyses in Medicine and Health Care*. 2002.

    "An influence diagram (also known as a tornado diagram) ..." [p. 242]
- Conclusion: informal survey of books on MDM
  - 12 books on MDM published after 1984 speak of decision trees
  - 11 books do not mention IDs
  - only one mentions them, quite briefly.

Influence diagrams at Harvard (HSPH)

Influence diagrams at Harvard (HCRA)