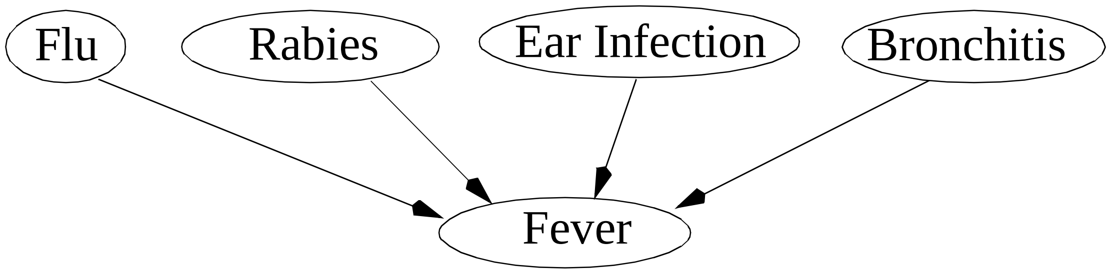
Probability Assessment

Topics:

Causal independence

Accuracy of probability assessment for diagnosis.

Causal Independence



Discrete dist'n: $(|S| - 1) \times |D_1| \times \ldots \times |D_n| \geq 2^n$ independent parameters.

Assessments:

What is the probability that a patient has a fever given that they have flu, rabies, and an ear infection?
"Beats me: I have never seen that combination..."

Statistics:

Require a ferocious number of samples to estimate 2^n parameters.

Causal Independence

Temporal model:

Let $c_1, \ldots, c_n$ be a set of *causes* for *effect* e.

Let $c_{i,t}$ be the variable associated with cause c_i at time t.

Causal independence assumption:

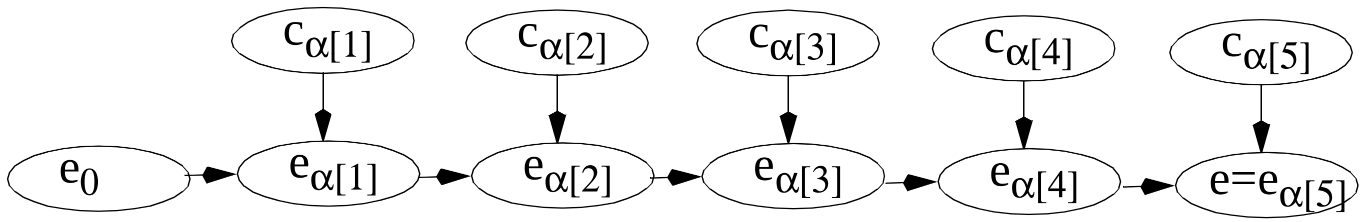
$c_1, \ldots, c_n$ are *causally independent with respect to* e if the set of CISs
$\forall(t < t'), \quad I(e_{t'}, \{c_{1,t'}, \ldots, c_{j-1,t'}, c_{(j+1),t'}, \ldots, c_{n,t'}\} \mid e_t, c_{j,t}, \forall(k \neq j)(c_{k,t} = c_{k,t'}))$ hold.

What this says:

If the only cause that changes state between t and t' is $c_{n,t}$, then the probability distribution over e_t' depends only on e_t and the new value for $c_{n,t}$.

Equivalent Bayes Net

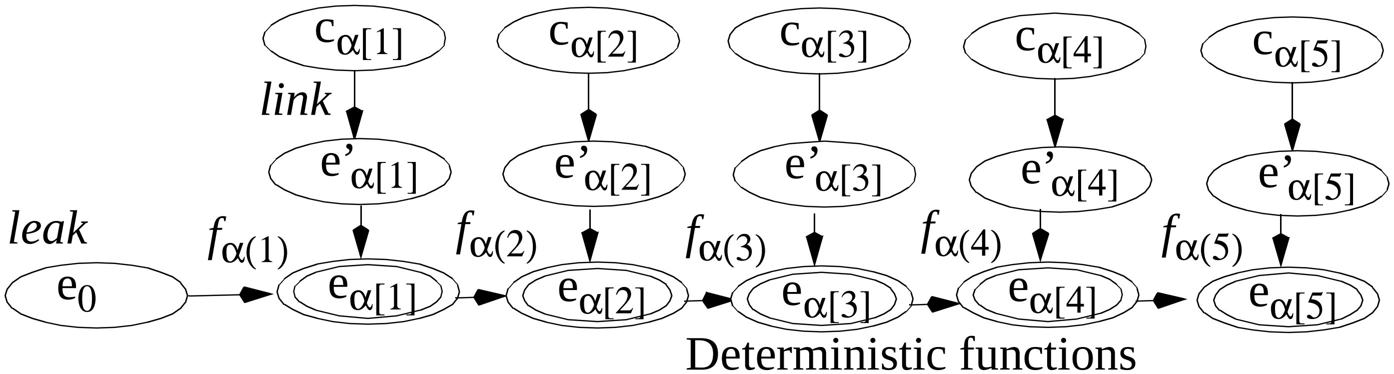
For <u>all</u> orderings α,



Distinguished Value

For each cause c_i designate some state (say $c_i = *$) be a "distinguished" state with the property that if all c_i are in their distinguished state, then, $e = e_0$. This distinguished state often (but not always) corresponds to when the c_i are all "absent".

Atemporal Representation



$$e = f_{\alpha(n)}(e'_{\alpha(n)}, f_{\alpha(n-1)}(e'_{\alpha(n-1)}, \ldots, f_{\alpha(1)}(e'_{\alpha(1)}, e_0))) \quad \text{(Nested decomposition)}$$

More implications

There is a distinguished value $*$ for all $e_{\alpha(n)}$ such that $p\{e'_{\alpha(k)} = * | c_{\alpha(k)} = *\} = 1.0$ and

$e = f_{\alpha(k)}(e, *)$ ($*$ is an identity element for each function $f_{\alpha(k)}$)

Interesting Cases

1. Singly decomposable causal independence:

   There exists some ordering α such that there is a nested decomposition.

2. Fully decomposable causal independence:

   There exists a nested decomposition for all orderings α.

3. Fully decomposable causal independence with equal functions.

   Implies that all outputs are combined using a function that is commutative and associative.
   Examples: boolean or, boolean and, max, min, sum.
   Described in literature as “noisy-or”, “noisy-and”, “noisy-max”, etc.

4. Normal gaussian distribution.

$$e = a + \sum_{i=1}^{n} b_i c_i + \varepsilon .$$

Noisy-Or, Noisy-Max

Combination function is “or” or “max”

Identity element for “or” is FALSE.

Identity element for “max” is $\min_v(e = v)$.

Reasonable model for many models.

Probability assessment:

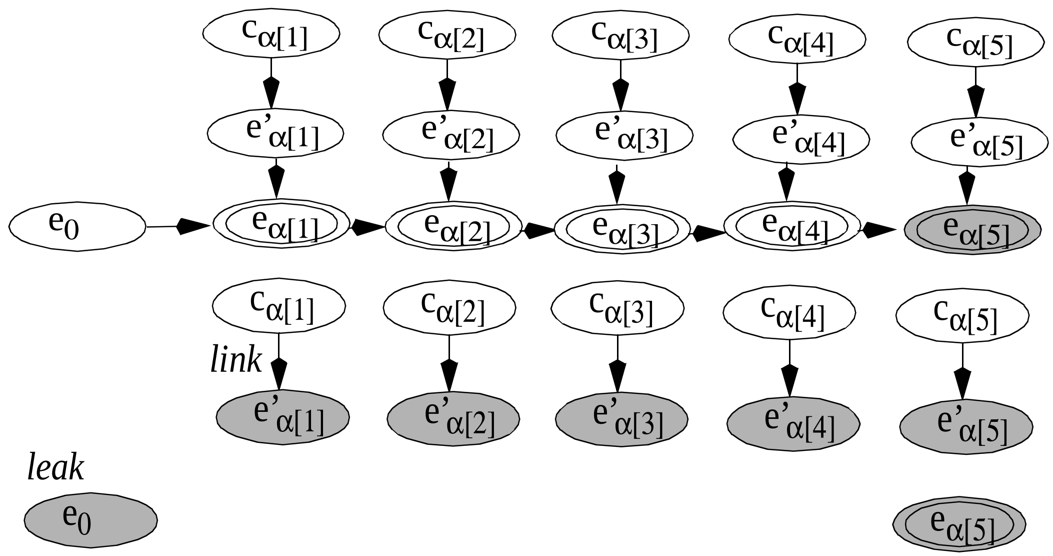
(Leak): What is the probability that the patient has an idiopathic fever (no cause)?

(Link): Say that a patient does not have a fever. They catch disease c_i. What is the probability distribution over their fever once they have caught c_i?

Negative product synergy: $\frac{P\{e|c_i, c_j, E\}}{P\{e|c_i, \neg c_j, E\}} \leq \frac{P\{e|\neg c_i, c_j, E\}}{P\{e|\neg c_i, \neg c_j, E\}}$ implies “explaining away”: $P(c_j|c_i, e, E) \leq P(c_j|\neg c_i, e, E)$ [Henrion&Druzdzel,91].

Additional Conditional Independence

If $P\{c_1, \ldots, c_n | (e = *)\}$, then $e'_1 = *, \ldots, e'_n = *$



Noisy-or: Observing that an effect is “false”, renders the causes conditionally independent.

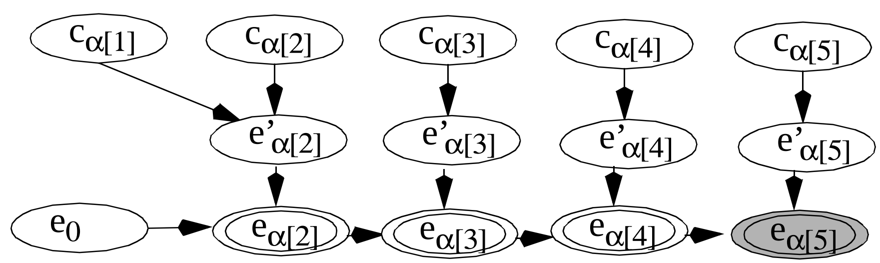
Summary

CI often provides reasonable default behavior when multiple causes.

Valuable since assessment time is $O(n)$ instead of $O(2^n)$.

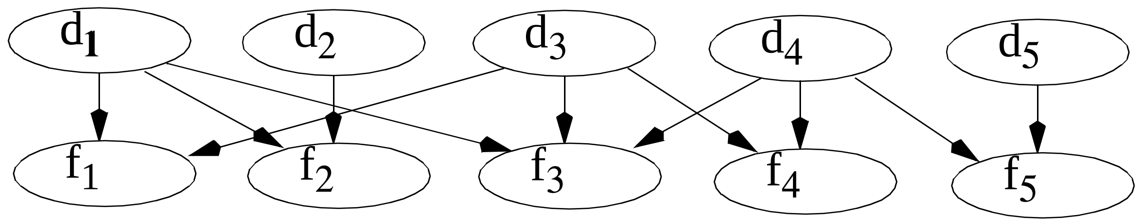
Used for essentially all diagnostic expert systems.

If causal independence doesn't work due to an important cause-cause synergism, try:



Important Large Expert Systems

QMR-DT: (Quick Medical Reference-Decision Theoretic [Shwe, Middleton, Heckerman, Henrion, Horvitz, Lehmann & Cooper, 90])



BN20: Bipartite graph of diseases and findings using noisy-ors.

534 diseases; 4040 findings; 40740 arcs (best known triangulation: ~250 nodes).
Causal independence exploited for inference (Quickscore, [Heckerman, ~90]; Variational approximation [Jordan, ~96], Factoring algorithms [D'Ambrosio], Clustering [Heckerman & Breese, 94]).

CPCS (Computer-based Patient Case Simulation),

Diagnosis is insensitive to imprecision in probability assessment [Henrion et al, 1996]

Critique: “It is too difficult to assess probability distributions”

Are probabilities in a model sensitive to noise?

Motorcycle diagnosis and kappa-calculus: $\kappa = 0.1, 0.01, 0.001$

Noise Experiments

Additive noise inappropriate: strong effect on diagnosis in Pathfinder [study ~1992].

Log-odds noise:

$$\mathrm{Lo}(p) = \log\left(\frac{p}{1-p}\right).$$

Assign $p' = \mathrm{Lo}^{-1}(\mathrm{Lo}(p) + \varepsilon)$ where $\varepsilon = \mathrm{Normal}(0, \sigma)$.

Evidence Weight

$$\mathrm{EW}(f, D) = \log\left[\frac{P\{f|d\}}{P\{f|\neg d\}}\right]$$

$$\mathrm{EW}(\neg f, D) = \log\left[\frac{P\{\neg f|d\}}{P\{\neg f|\neg d\}}\right]$$

Conclusions

Effect of noise on diagnosis is low.

Link probabilities most important, on average.

“Reason to believe” that the result does not hold for predictive models.