Building Bayesian Networks

Thanks to: Rina Dechter
Univ. of Calif. At Irvine

Modeling with Bayesian Networks

Bayesian networks will be constructed in three consecutive steps.

Step 1

Define the network variables and their values.

- A query variable is one which we need to ask questions about, such as compute its posterior marginal.
- An evidence variable is one which we may need to assert evidence about.
- An intermediary variable is neither query nor evidence and is meant to aid the modeling process by detailing the relationship between evidence and query variables.

The distinction between query, evidence and intermediary variables is not a property of the Bayesian network, but of the task at hand.

Modeling with Bayesian Networks

Bayesian networks will be constructed in three consecutive steps.

Step 2

Define the network structure (edges).

We will be guided by a causal interpretation of network structure.

The determination of network structure will be reduced to answering the following question about each network variable X: what set of variables we regard as the direct causes of X?

What about the boundary strata?

Modeling with Bayesian Networks

Bayesian networks will be constructed in three consecutive steps.

Step 3

Define the network CPTs.

The difficulty and objectivity of this step varies considerably from one problem to another:

- CPTs can sometimes be determined completely from the problem statement by objective considerations.
- CPTs can be a reflection of subjective beliefs.
- CPTs can be estimated from data.

Diagnosis I: Model from Expert

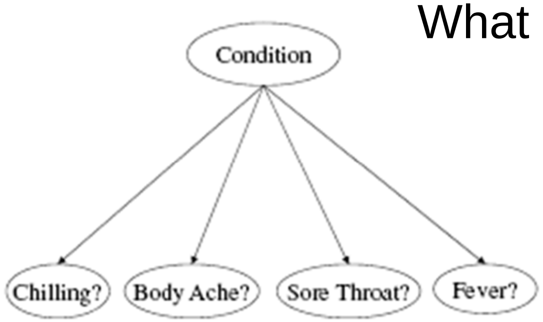
Example

The flu is an acute disease characterized by fever, body aches and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a sore throat. Tonsillitis is inflammation of the tonsils which leads to a sore throat and can be associated with fever.

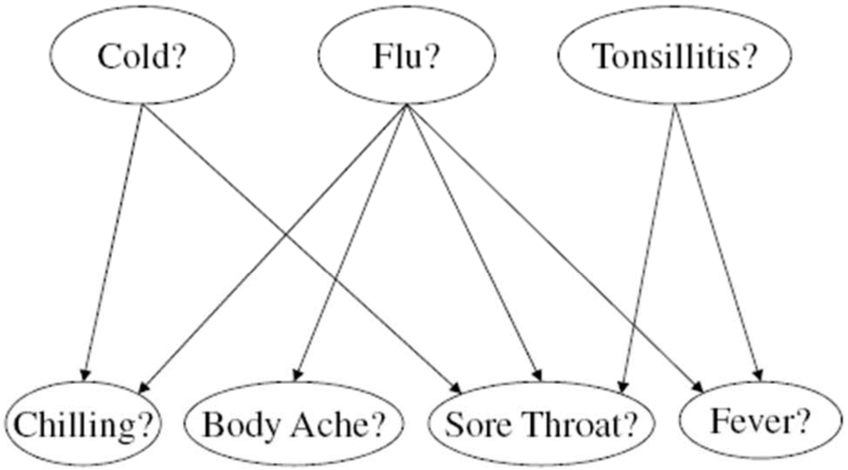
Our goal here is to develop a Bayesian network to capture this knowledge and then use it to diagnose the condition of a patient suffering from some of the symptoms mentioned above.

Variables? Arcs? Try it.

Diagnosis I: Model from Expert

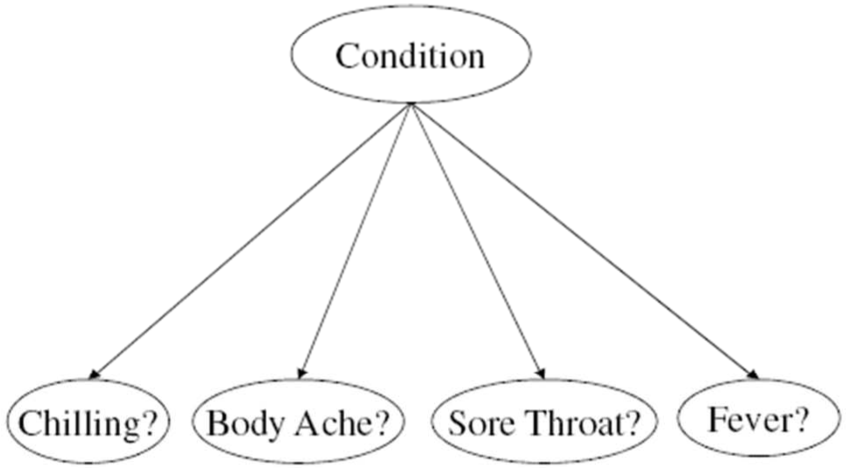


What about?



Variables are binary: values are either true or false. More refined information may suggest different degrees of body ache.

Diagnosis I: Model from Expert



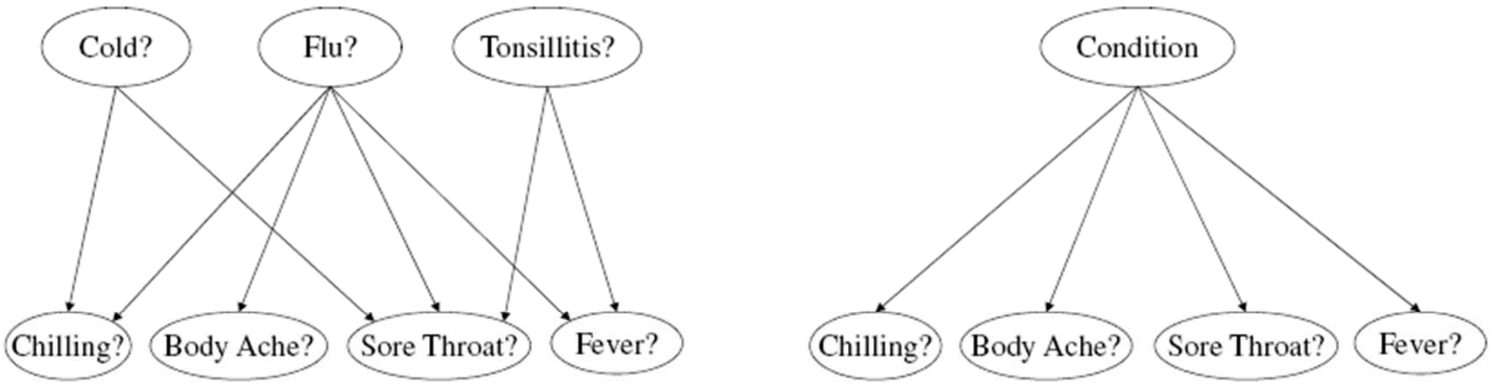
"Condition" has multiple values: normal, cold, flu, and tonsillitis.

A naive Bayes structure

has the following edges $C \rightarrow A_1, \ldots, C \rightarrow A_m$, where C is called the class variable and $A_1, \ldots, A_m$ are called the attributes.

Diagnosis I: Model from Expert

The naive Bayes structure commits to the single-fault assumption.



Suppose the patient is known to have a cold.

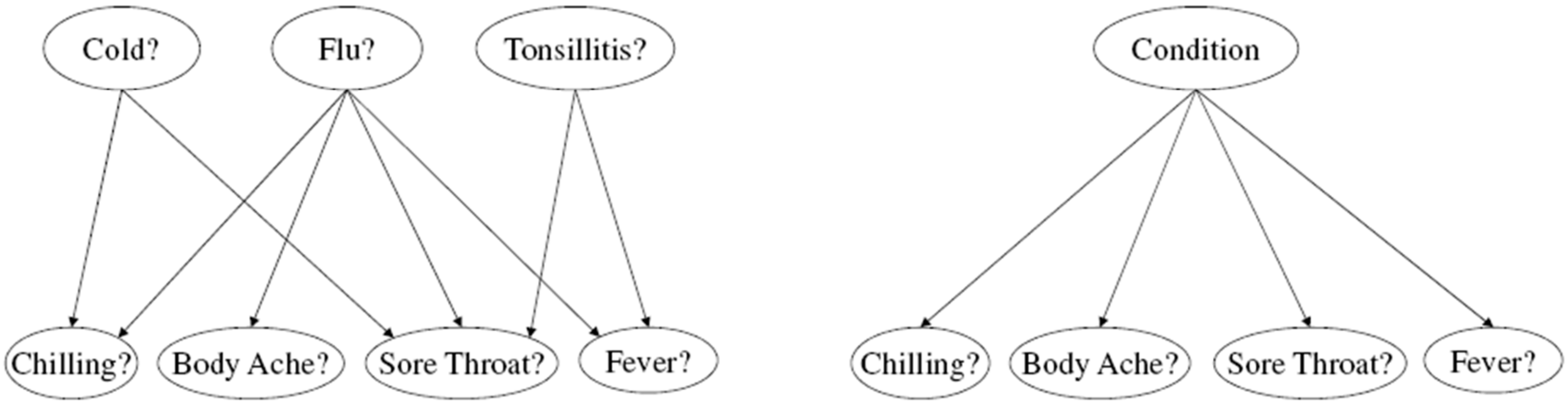
Naive Bayes structure

Fever and sore throat become independent as they are d-separated by "Condition".

Original structure

Fever may increase our belief in tonsillitis, which could then increase our belief in a sore throat.

Diagnosis I: Model from Expert



If the only evidence we have is $\boxed{\text{body ache,}}$ we expect the probability of flu to go up in both networks.

Naive Bayes structure

This leads to dropping the probability of cold or tonsillitis.

Original structure

These probabilities remain the same since both cold and tonsillitis are d-separated from body ache.

Diagnosis I: Model from Expert

CPTs can be obtained from medical experts, who supply this information based on known medical statistics or subjective beliefs gained through practical experience.

CPTs can also be estimated from medical records of previous patients

Case	*Cold?*	*Flu?*	*Tonsillitis?*	*Chilling?*	*Bodyache?*	*Sorethroat?*	*Fever?*
1	true	false	?	true	false	false	false
2	false	true	false	true	true	false	true
3	?	?	true	false	?	true	false
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

? indicates the unavailability of corresponding data for that patient.

Diagnosis II: Model from Expert

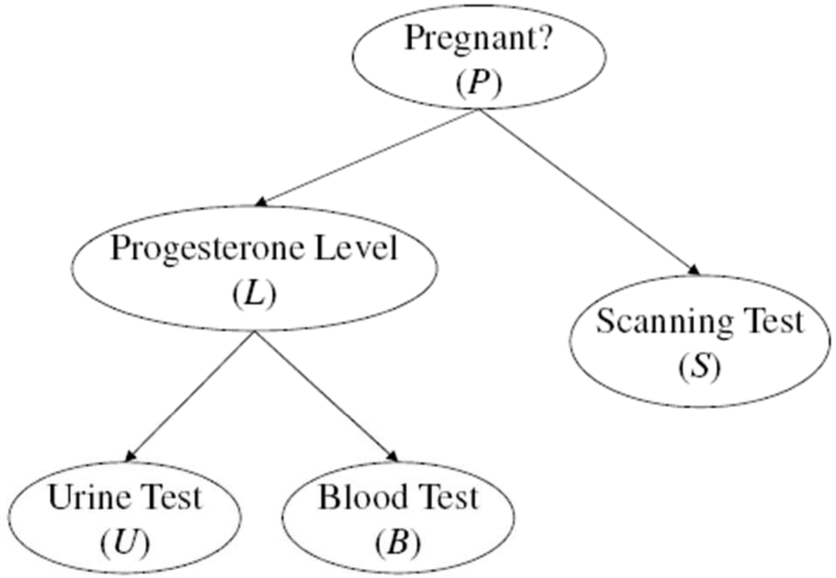
Example

A few weeks after inseminating a cow, we have three possible tests to confirm pregnancy. The first is a scanning test which has a false positive of 1% and a false negative of 10%. The second is a blood test, which detects progesterone with a false positive of 10% and a false negative of 30%. The third test is a urine test, which also detects progesterone with a false positive of 10% and a false negative of 20%. The probability of a detectable progesterone level is 90% given pregnancy, and 1% given no pregnancy. The probability that insemination will impregnate a cow is 87%.

Our task here is to build a Bayesian network and use it to compute the probability of pregnancy given the results of some of these pregnancy tests.

Try it: Variables and values? Structure? CPTs?

Diagnosis II: Model from Expert



P	θ_p
yes	.87

P	S	$\theta_{s\|p}$
yes	−ve	.10
no	+ve	.01

P	L	$\theta_{l\|p}$
yes	undetectable	.10
no	detectable	.01

L	B	$\theta_{b\|l}$
detectable	−ve	.30
undetectable	+ve	.10

L	U	$\theta_{u\|l}$
detectable	−ve	.20
undetectable	+ve	.10

Diagnosis II: Model from Expert

Example

We inseminate a cow, wait for a few weeks, and then perform the three tests which all come out negative:

$$\mathbf{e}: \; S = \text{–ve}, \; B = \text{–ve}, \; U = \text{–ve}.$$

Posterior marginal for pregnancy given this evidence:

P	$\Pr(P \mid \mathbf{e})$
yes	10.21%
no	89.79%

Probability of pregnancy is reduced from 87% to 10.21%, but still relatively high given that all three tests came out negative.

Sensitivity Analysis

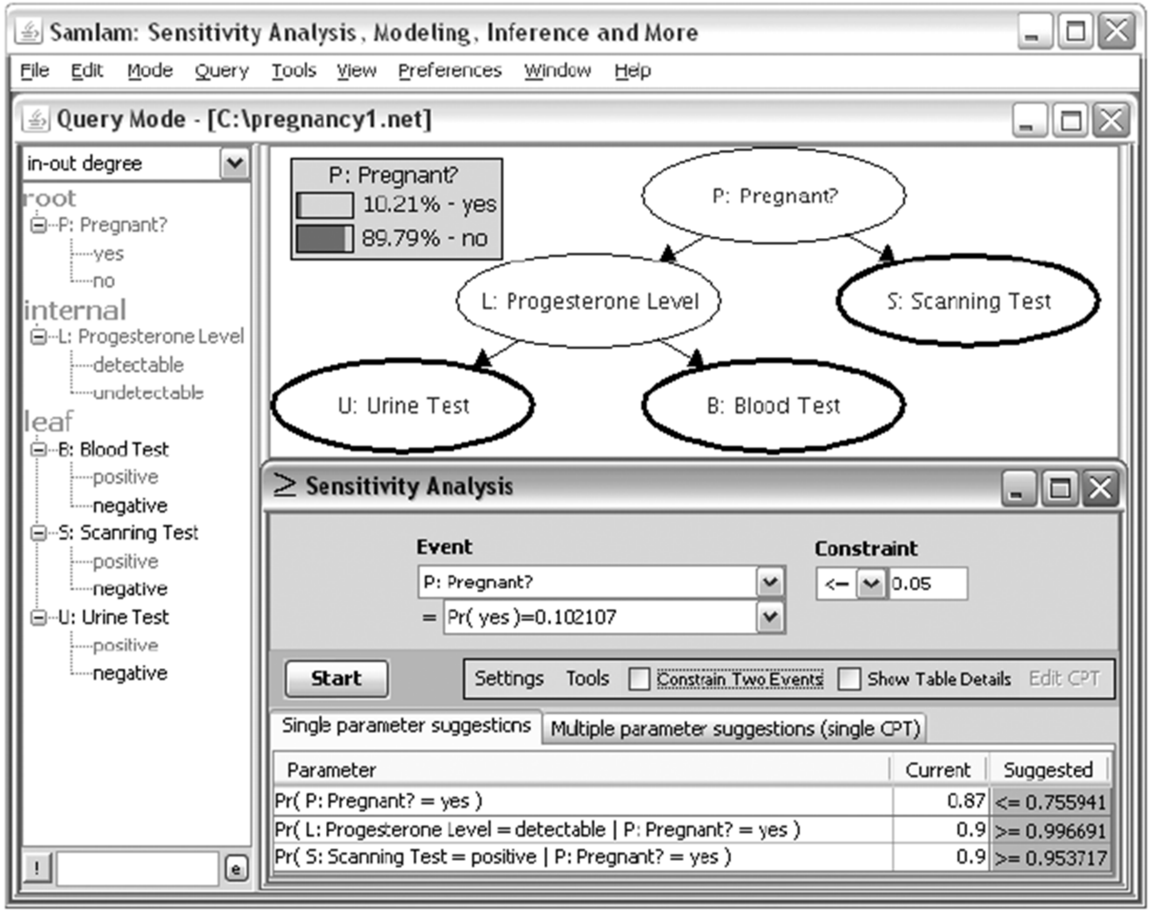
Example

A farmer is not too happy with this and would like three negative tests to drop the probability of pregnancy to no more than 5%. The farmer is willing to replace the test kits for this purpose, but needs to know the false positive and negative rates of the new tests, which would ensure the above constraint.

This is a problem of sensitivity analysis in which we try to understand the relationship between the parameters of a Bayesian network and the conclusions drawn based on the network.

Read in the book.
We will not cover this.

Sensitivity Analysis



Example

Which network parameter do we have to change, and by how much, so as to ensure that the probability of pregnancy would be no more than 5% given three negative tests?

Sensitivity Analysis

Possible (single) parameter changes:

1. If the false negative rate for the scanning test were about 4.63% instead of 10%.
2. If the probability of pregnancy given insemination were about 75.59% instead of 87%.
3. If the probability of a detectable progesterone level given pregnancy were about 99.67% instead of 90%.

The last two changes are not feasible since the farmer does not intend to change the insemination procedure, nor does he control the progesterone level.

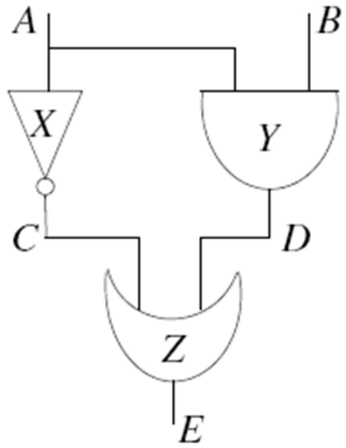
Diagnosis II: Model from Expert

Improving either the blood test or the urine test cannot help.

If our goal is to drop the probability of pregnancy to no more than 8% (instead of 5%), then Samlam identifies the following additional possibilities:

- The false negative for the blood test should be no more than about 12.32% instead of 30%.
- The false negative for the urine test should be no more than about 8.22% instead of 20%.

Diagnosis III: Model from Design

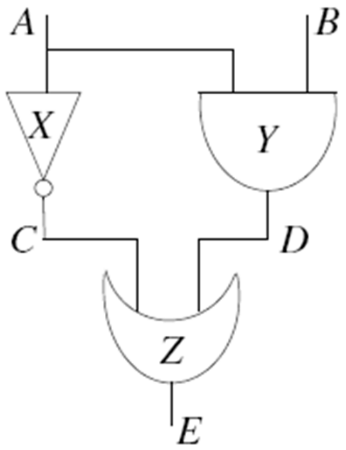


Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

Try it: Variables? Values? Structure?

Diagnosis III: Model from Design



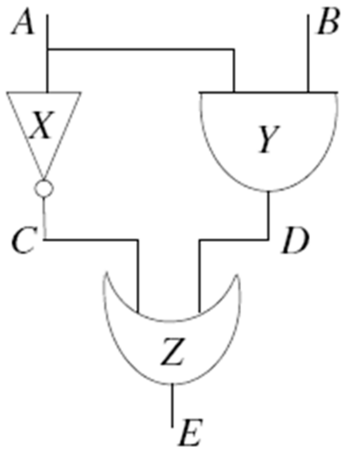
Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

Evidence variables

Primary inputs and output of the circuit, A, B and E.

Diagnosis III: Model from Design



Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

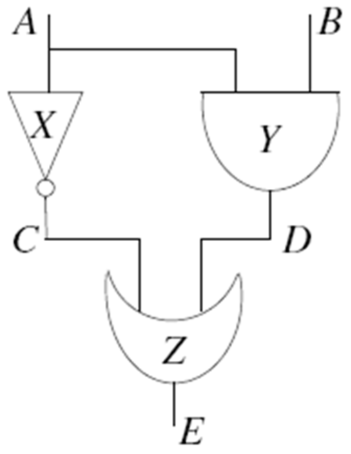
Evidence variables

Primary inputs and output of the circuit, A, B and E.

Query variables

Health of components X, Y and Z.

Diagnosis III: Model from Design



Problem statement

Given some values for the circuit primary inputs and output (test vector), decide if the circuit is behaving normally. If not, find the most likely health states of its components.

Evidence variables

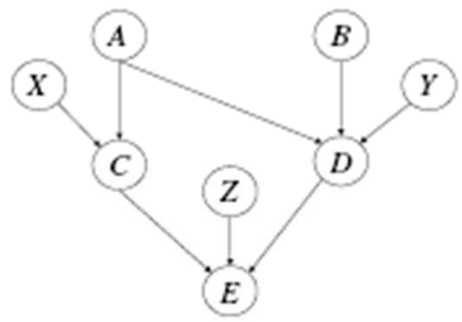
Primary inputs and output of the circuit, A, B and E.

Query variables

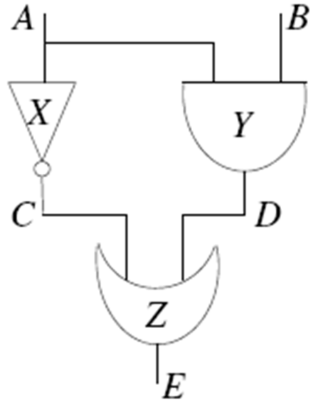
Health of components X, Y and Z.

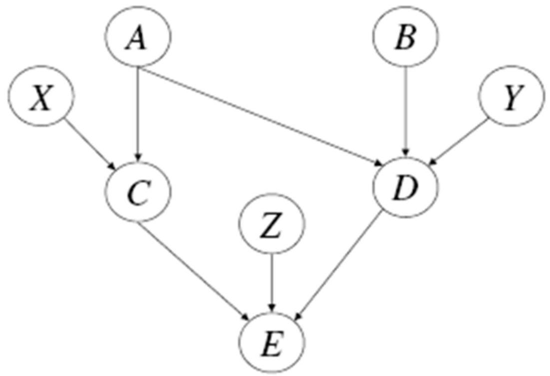
Intermediary variables

Internal wires, C and D.



Diagnosis III: Model from Design

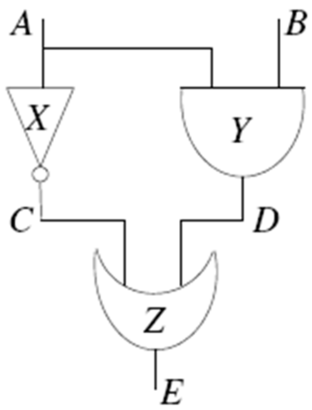


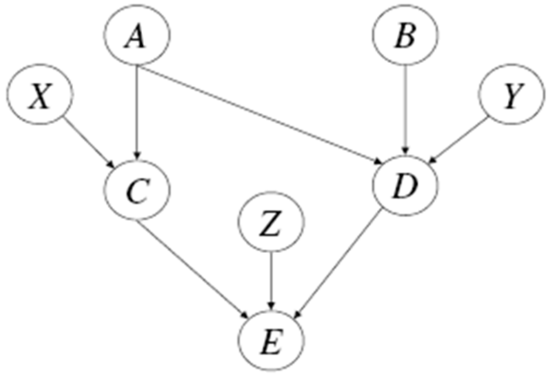


Function blocks

The outputs of each block are determined by its inputs and its state of health.

Diagnosis III: Model from Design





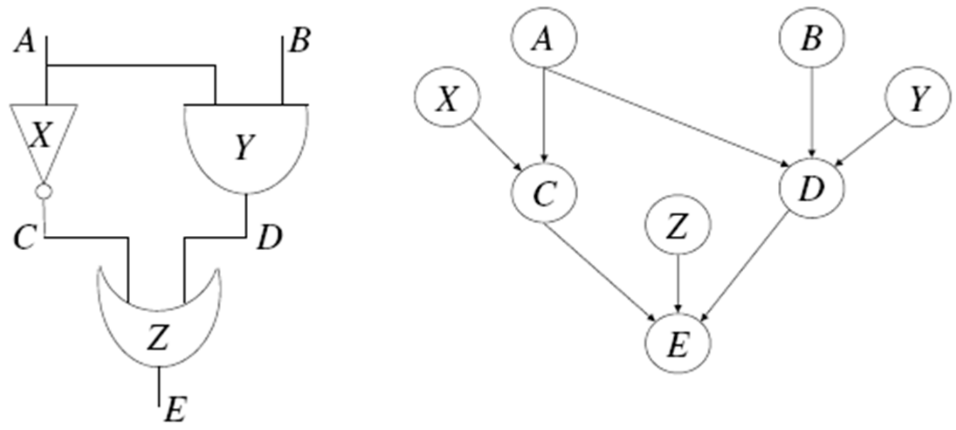
Function blocks

The outputs of each block are determined by its inputs and its state of health.

Primary inputs

No direct causes for primary inputs, A and B: no parents.

Diagnosis III: Model from Design



Function blocks

The outputs of each block are determined by its inputs and its state of health.

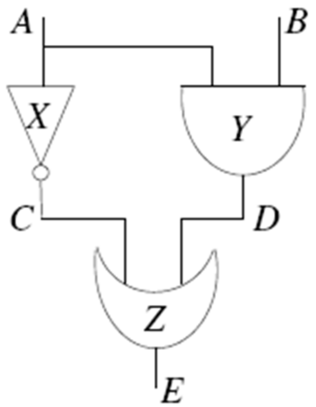
Primary inputs

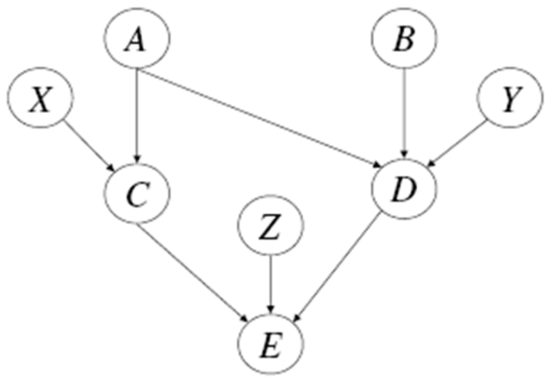
No direct causes for primary inputs, A and B: no parents.

Health states

No direct causes for health of X, Y and Z: no parents.

Diagnosis III: Model from Design





Function blocks

The outputs of each block are determined by its inputs and its state of health.

Primary inputs

No direct causes for primary inputs, A and B: no parents.

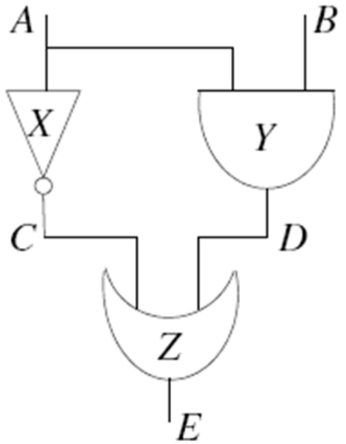
Health states

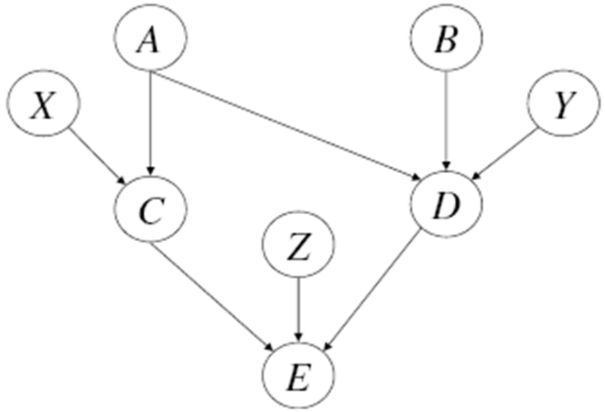
No direct causes for health of X, Y and Z: no parents.

Gate output D

Direct causes of D are gate inputs, A and B, and health of Y.

Diagnosis III: Model from Design





Values of circuit wires:

low or high

Health states: ok or faulty

faulty is too vague as a component may fail in a number of modes.

- stuck-at-zero fault: low output regardless of gate inputs.
- stuck-at-one fault: high output regardless of gate inputs.
- input-output-short fault: inverter shorts input to its output.

Fault modes demand more when specifying the CPTs.

Diagnosis III: Model from Design

Three classes of CPTs

- primary inputs (A, B)
- gate outputs (C, D, E)
- component health (X, Y, Z)

CPTs for health variables depend on their values

X	θ_x
ok	.99
faulty	.01

X	θ_x
ok	.99
stuckat0	.005
stuckat1	.005

Need to know the probabilities of various fault modes.

Diagnosis III: Model from Design

CPTs for component outputs determined from functionality.

Example

CPT for inverter X.

A	X	C	$\theta_{c\|a,x}$
high	ok	high	0
low	ok	high	1
high	stuckat0	high	0
low	stuckat0	high	0
high	stuckat1	high	1
low	stuckat1	high	1

Diagnosis III: Model from Design

CPTs for component outputs determined from functionality.

Example

CPT for inverter X.

A	X	C	$\theta_{c\|a,x}$
high	ok	high	0
low	ok	high	1
high	stuckat0	high	0
low	stuckat0	high	0
high	stuckat1	high	1
low	stuckat1	high	1

If we do not represent health states:

A	X	C	$\theta_{c\|a,x}$
high	ok	high	0
low	ok	high	1
high	faulty	high	?
low	faulty	high	?

Common to use a probability of .50 in this case.

A Diagnosis Example

Example

Given test vector $\mathbf{e}$: $A=$high, $B=$high, $E=$low, compute MAP over health variables X, Y and Z.

A Diagnosis Example

Example

Given test vector **e**: A=high, B=high, E=low, compute MAP over health variables X, Y and Z.

Network with fault modes gives two MAP instantiations:

MAP given **e**	X	Y	Z	
	ok	stuckat0	ok	each probability $\approx$ 49.4%
	ok	ok	stuckat0	

A Diagnosis Example

Example

Given test vector $\mathbf{e}$: A=high, B=high, E=low, compute MAP over health variables X, Y and Z.

Network with fault modes gives two MAP instantiations:

MAP given $\mathbf{e}$	X	Y	Z
	ok	stuckat0	ok
	ok	ok	stuckat0

each probability $\approx$ 49.4%

Network with no fault modes gives two MAP instantiations:

MAP given $\mathbf{e}$	X	Y	Z
	ok	faulty	ok
	ok	ok	faulty

each probability $\approx$ 49.4%

Posterior Marginals

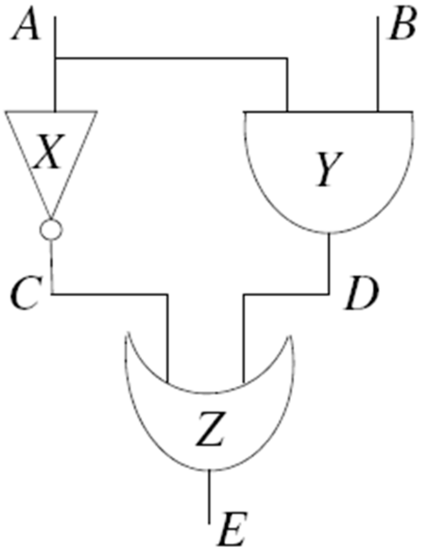
Consider the posterior marginals over the health variables X, Y, Z:

State	X	Y	Z	$\Pr(X, Y, Z \mid \mathbf{e})$
1	ok	ok	ok	0
2	faulty	ok	ok	0
3	ok	faulty	ok	.49374
4	ok	ok	faulty	.49374
5	ok	faulty	faulty	.00499
6	faulty	ok	faulty	.00499
7	faulty	faulty	ok	.00249
8	faulty	faulty	faulty	.00005

- State 2 is impossible.
- Y and Z more likely to be faulty together than Y and X.
- States with faulty Z more likely than states with faulty Y:

$$\Pr(Z = \text{faulty} \mid \mathbf{e}) \approx 50.38\% > \Pr(Y = \text{faulty} \mid \mathbf{e}) \approx 50.13\%.$$

Lack of Symmetry for Double Faults



Test vector

$A = \text{high},\ B = \text{high},\ E = \text{low}$

- If Y and Z are faulty, we have two possible states for C and D: $C = \text{low}$, D either low or high.
- If Y and X are faulty, we have only one possible state for C and D: $C = \text{low}$ and $D = \text{low}$.

Integrating Time

Suppose we have two test vectors instead of only one.

Integrating Time

Suppose we have two test vectors instead of only one.

Additional evidence variables

A', B' and E'

Integrating Time

Suppose we have two test vectors instead of only one.

Additional evidence variables

A', B' and E'

Additional intermediary variables

C' and D'

Integrating Time

Suppose we have two test vectors instead of only one.

Additional evidence variables

A', B' and E'

Additional intermediary variables

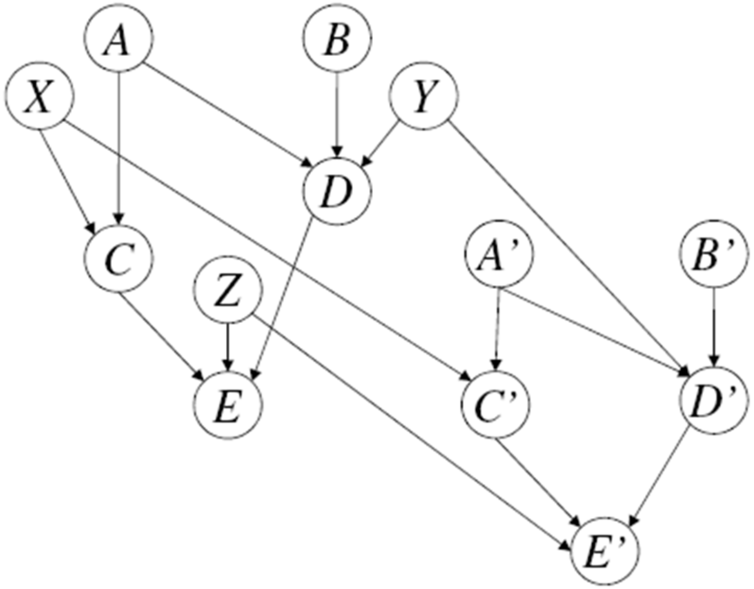
C' and D'

Additional health variables on whether we allow intermittent faults

If health of a component can change from one test to another, we need additional health variables X', Y', and Z'. Otherwise, the original health variables are sufficient.

Variables? Values? Structure?

Integrating Time: No Intermittent Faults

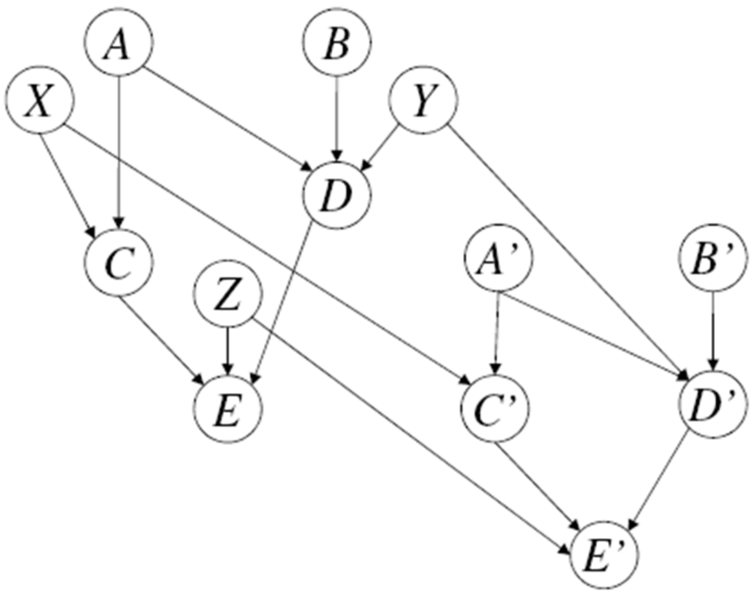


Two test vectors

$\mathbf{e}$: A = high, B = high, E = low

$\mathbf{e}'$: A = low, B = low, E = low.

Integrating Time: No Intermittent Faults



Two test vectors

$\mathbf{e}$: $A\!=\!\text{high},\ B\!=\!\text{high},\ E\!=\!\text{low}$

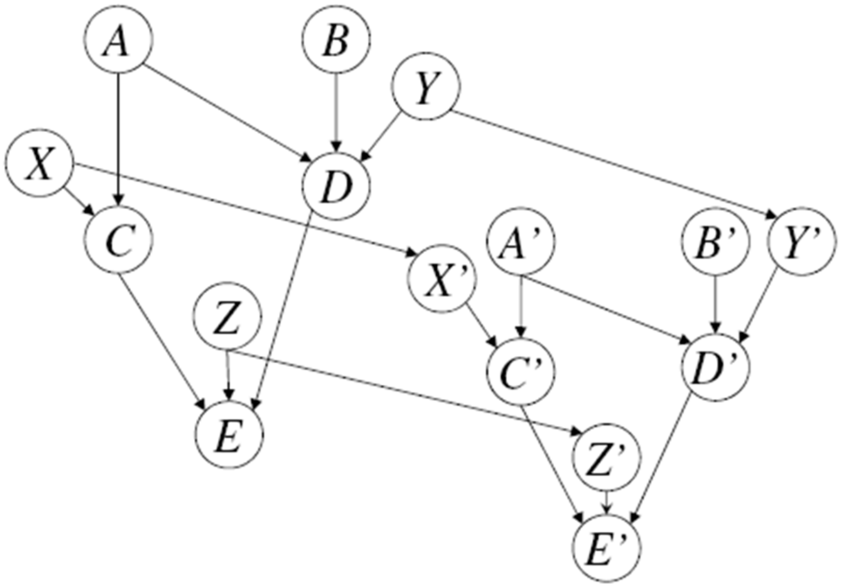
$\mathbf{e}'$: $A\!=\!\text{low},\ B\!=\!\text{low},\ E\!=\!\text{low}.$

MAP using second structure

MAP given $\mathbf{e}, \mathbf{e}'$	X	Y	Z
	ok	ok	faulty

with probability $\approx 97.53\%$

Integrating Time: Intermittent Faults



Dynamic Bayesian network (DBN)

Two test vectors

$\mathbf{e}$: A=high, B=high, E=low
$\mathbf{e}'$: A=low, B=low, E=low.

Persistence model for the health of component X

X	X'	$\theta_{x'\|x}$	
ok	ok	.99	
ok	faulty	.01	healthy component becomes faulty
faulty	ok	.001	faulty component becomes healthy
faulty	faulty	.999	

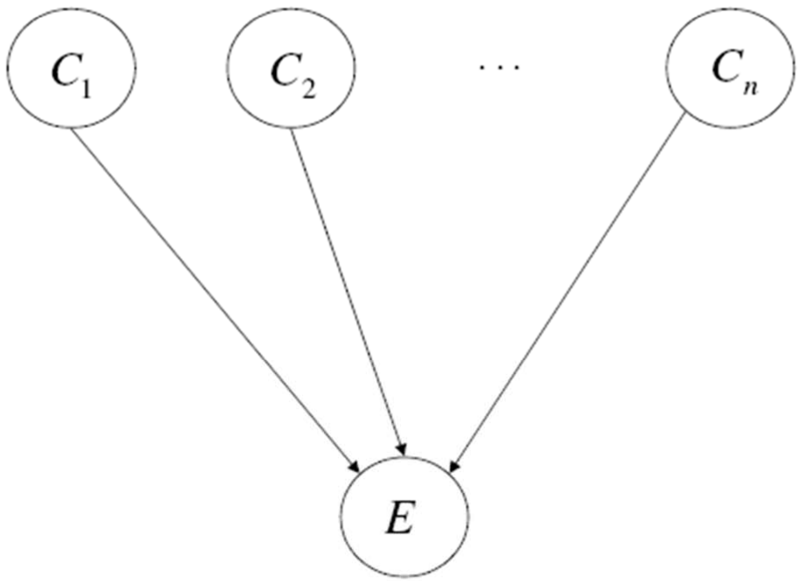
Dealing with Large CPTs

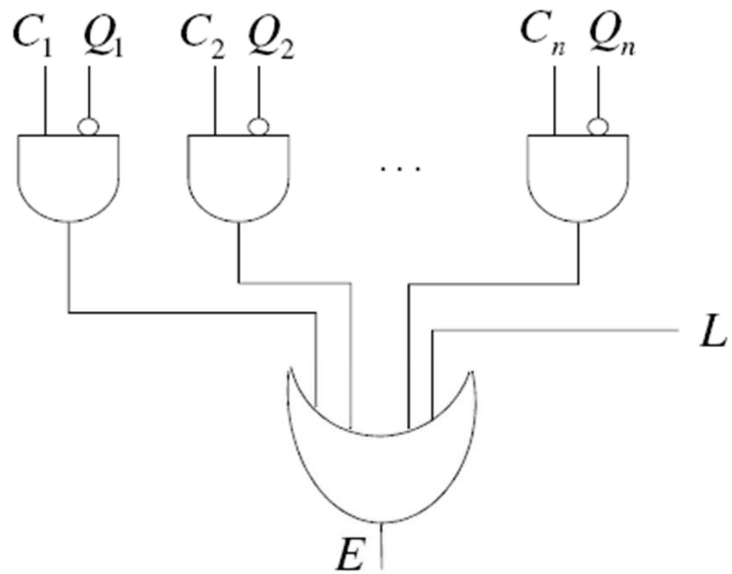
The size of a CPT

for binary variable E with binary parents $C_1, \ldots, C_n$

Number of Parents: n	Parameter Count: 2^n
2	4
3	8
6	64
10	1024
20	$1,048,576$
30	$1,073,741,824$

Micro Model





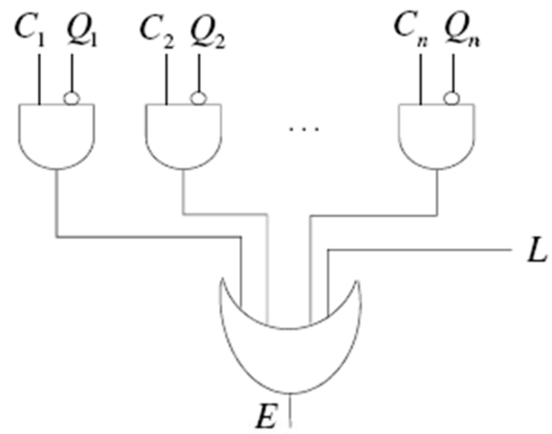
A noisy-or circuit

A micro model

details the relationship between a variable E and its parents $C_1, \ldots, C_n$.

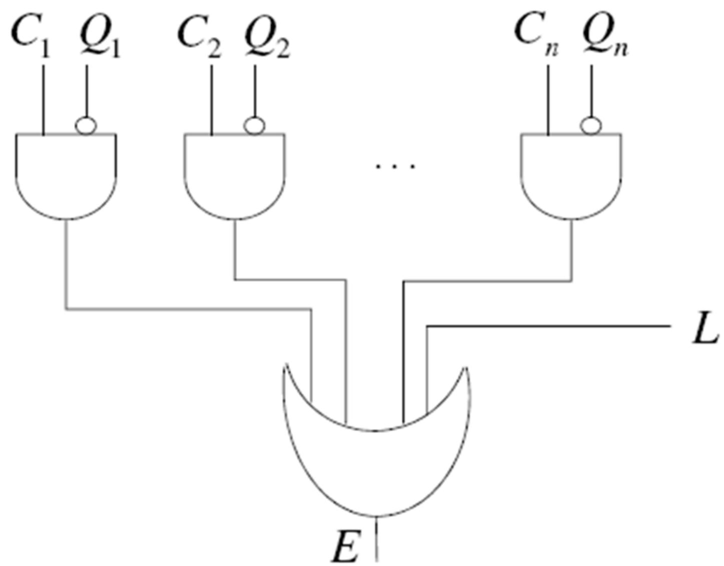
We wish to specify cpt with less parameters

Noisy-or Model



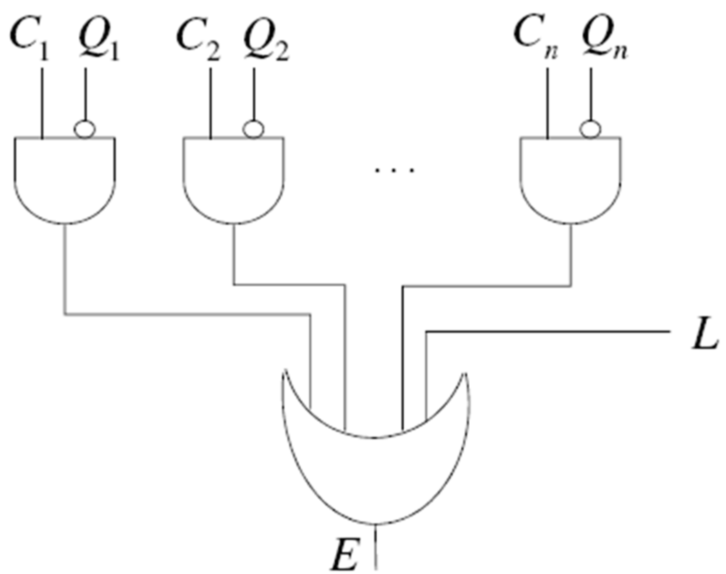
- Cause C_i is capable of establishing effect E, except under some unusual circumstances summarized by suppressor Q_i.
- When suppressor Q_i is active, C_i is no longer able to establish E.
- The leak variable L represents all other causes of E which were not modeled explicitly.
- When none of the causes C_i are active, the effect E may still be established by the leak variable L.

Noisy–or Model



The noisy-or model requires $n + 1$ parameters.

Noisy–or Model



The noisy-or model requires $n + 1$ parameters.

To model the relationship between headache and ten different conditions

- $\theta_{q_i} = \Pr(Q_i = \text{active})$: probability that suppressor of C_i is active.
- $\theta_l = \Pr(L = \text{active})$: probability that leak is active.

Noisy-or Model

- Let I_α be the indices of causes that are active in α.

Noisy-or Model

- Let I_α be the indices of causes that are active in α.
- If

  α: C_1 = active, C_2 = active, C_3 = passive, C_4 = passive, C_5 = active,

  then $I_\alpha = \{1, 2, 5\}$.

Noisy-or Model

- Let I_α be the indices of causes that are active in α.
- If

  α: $C_1 =$ active, $C_2 =$ active, $C_3 =$ passive, $C_4 =$ passive, $C_5 =$ active,

  then $I_\alpha = \{1, 2, 5\}$.
- We then have

$$\begin{aligned} \Pr(E = \text{passive} \mid \alpha) &= (1 - \theta_l) \prod_{i \in I_\alpha} \theta_{q_i} \\ \Pr(E = \text{active} \mid \alpha) &= 1 - \Pr(E = \text{passive} \mid \alpha). \end{aligned}$$

Noisy-or Model

- Let I_α be the indices of causes that are active in α.
- If

  $\alpha: C_1 = \text{active}, C_2 = \text{active}, C_3 = \text{passive}, C_4 = \text{passive}, C_5 = \text{active},$

  then $I_\alpha = \{1, 2, 5\}$.
- We then have

$$\begin{aligned} \Pr(E = \text{passive} | \alpha) &= (1 - \theta_l) \prod_{i \in I_\alpha} \theta_{q_i} \\ \Pr(E = \text{active} | \alpha) &= 1 - \Pr(E = \text{passive} | \alpha). \end{aligned}$$

The full CPT for variable E, with its 2^n parameters, can be induced from the $n + 1$ parameters of the noisy-or model.

Noisy-or Model

Example

Sore throat (S) has three causes: cold (C), flu (F), tonsillitis (T).

Noisy-or Model

Example

Sore throat (S) has three causes: cold (C), flu (F), tonsillitis (T).

If we assume that S is related to its causes by a noisy-or model

we can then specify the CPT for S by the following four probabilities:

- The suppressor probability for cold, say .15
- The suppressor probability for flu, say, .01
- The suppressor probability for tonsillitis, say .05
- The leak probability, say .02

Noisy-or Model

Example

Sore throat (S) has three causes: cold (C), flu (F), tonsillitis (T).

Noisy-or Model

Example

Sore throat (S) has three causes: cold (C), flu (F), tonsillitis (T).

The CPT for sore throat is then determined completely as follows:

C	F	T	S	$\theta_{s\|c,f,t}$	
true	true	true	true	0.9999265	$1-(1-.02)(.15)(.01)(.05)$
true	true	false	true	0.99853	$1-(1-.02)(.15)(.01)$
true	false	true	true	0.99265	$1-(1-.02)(.15)(.05)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
false	false	false	true	.02	$1-(1-.02)$