Chapter 5: Building Bayesian Networks

Adnan Darwiche[1]

October 6, 2009

[1] Lecture slides for *Modeling and Reasoning with Bayesian Networks*, Adnan Darwiche, Cambridge University Press, 2009.

Outline

- Real-world applications, drawn from the domains of diagnosis, reliability, genetics, channel coding, and commonsense reasoning.
- Specific reasoning problems which can be addressed by posing a formal query with respect to a Bayesian network.
- Constructing the required network.
- Identifying the specific queries that need to be applied.

Outline

The construction of a Bayesian network involves three major steps:

- Identify relevant variables and their possible values.
- Build the network structure by connecting variables into DAG.
- Define the CPT for each network variable.

Two issues:

- The potentially large size of CPTs.
- The significance of the specific numbers used to populate them.

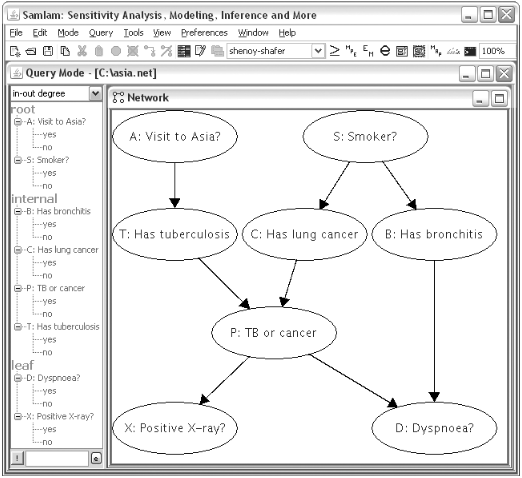
We present techniques for dealing with these issues.

Outline

There are at least four general types of queries which can be posed with respect to a Bayesian network.

Which type of query to use in a specific situation is not always trivial, and some of the queries are guaranteed to be equivalent under certain conditions.

Reasoning with Bayesian Networks

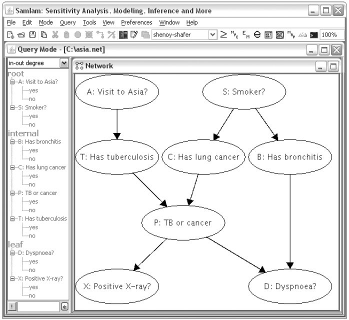


The network Asia will be used as a running example. Screenshot from SamIam.

SamIam available at http://reasoning.cs.ucla.edu/samiam/.

Query: Probability of Evidence

Probability of some variable instantiation $\mathbf{e}$, $\Pr(\mathbf{e})$.



Probability that the patient has a positive X-ray, but no dyspnoea, $\Pr(X = \text{yes}, D = \text{no})$, about 3.96%. Computed by Samlam.

The variables $\mathbf{E} = \{X, D\}$ are called evidence variables. The query $\Pr(\mathbf{e})$ is known as a probability-of-evidence.

Query: Probability of Evidence

There are other types of evidence, beyond variable instantiations.

Any propositional sentence can be used to specify evidence.

Example

We may want to know the probability that the patient has either a positive X-ray or dyspnoea, $X = \text{yes} \vee D = \text{yes}$.

The term evidence is almost always used to mean an instantiation of some variables.

Query: Probability of Evidence

Auxiliary-node method

Bayesian network tools do not usually provide direct support for computing the probability of arbitrary pieces of evidence, but such probabilities can be computed indirectly.

We can add an auxiliary node E, declare nodes X and D as the parents of E, and use the following CPT for E:

X	D	E	$\Pr(e \mid x, d)$
yes	yes	yes	1
yes	no	yes	1
no	yes	yes	1
no	no	yes	0

Event $E = \text{yes}$ is then equivalent to $X = \text{yes} \vee D = \text{yes}$.

Query: Prior and Posterior Marginals

Prior Marginals

Given a joint probability distribution $\Pr(x_1, \ldots, x_n)$, the marginal distribution $\Pr(x_1, \ldots, x_m)$, $m \leq n$, is defined as follows:

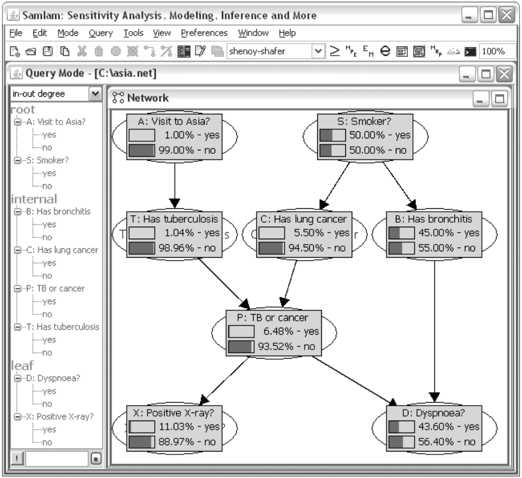
$$\Pr(x_1, \ldots, x_m) = \sum_{x_{m+1}, \ldots, x_n} \Pr(x_1, \ldots, x_n).$$

The marginal distribution can be viewed as a projection of the joint distribution on the smaller set of variables $X_1, \ldots, X_m$.

Posterior marginal given evidence $\mathbf{e}$

$\Pr(x_1, \ldots, x_m | \mathbf{e}) = \sum_{x_{m+1}, \ldots, x_n} \Pr(x_1, \ldots, x_n | \mathbf{e})$.

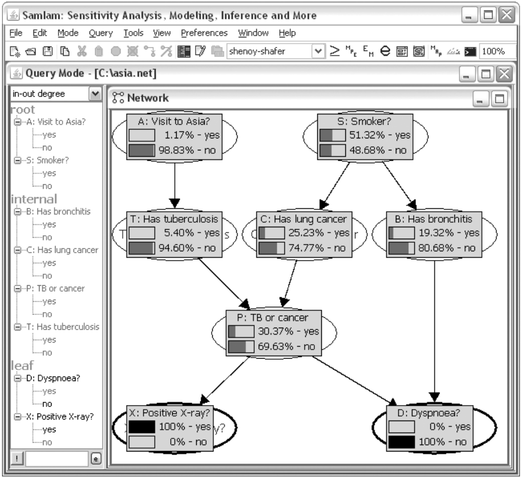
Prior Marginals in the Asia Network



Prior marginal

C	$\Pr(C)$
yes	5.50%
no	94.50%

Query: Posterior Marginals in the Asia Network



Posterior marginal

C	$\Pr(C \mid \mathbf{e})$
yes	25.23%
no	74.77%

$\mathbf{e}: X = \text{yes}, D = \text{no}$

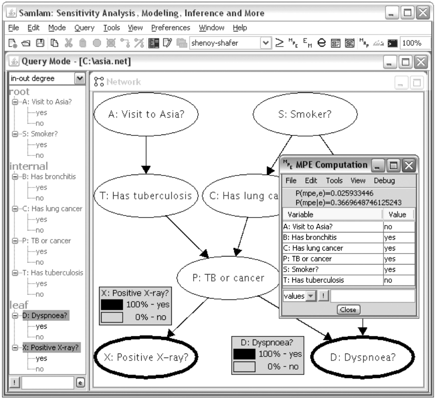
Query: Most Probable Explanation (MPE)

Let $X_1, \ldots, X_n$ be all network variables, and $\mathbf{e}$ be evidence. Identify an instantiation $x_1, \ldots, x_n$ that maximizes the probability $\Pr(x_1, \ldots, x_n|\mathbf{e})$. Instantiation $x_1, \ldots, x_n$ is called a most probable explanation given evidence $\mathbf{e}$.

MPE cannot be obtained directly from posterior marginals.

If $x_1, \ldots, x_n$ is an instantiation obtained by choosing each value x_i so as to maximize the probability $\Pr(x_i|\mathbf{e})$, then $x_1, \ldots, x_n$ is not necessarily an MPE.

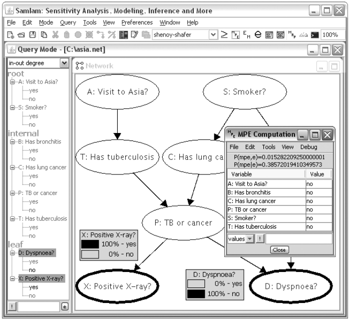
Query: Most Probable Explanation (MPE)



MPE given a positive X-ray and dyspnoea

A patient that made no visit to Asia; is a smoker; has lung cancer and bronchitis; but no tuberculosis.

Query: Most Probable Explanation (MPE)



MPE given a positive X-ray and no dyspnoea ($\approx 38.57\%$)

A patient that made no visit to Asia; is not a smoker; has no lung cancer, no bronchitis and no tuberculosis.

Choosing values with maximal probability, we get:

α: $A=$no, $S=$yes, $T=$no, $C=$no, $B=$no, $P=$no, $X=$yes, $D=$no.

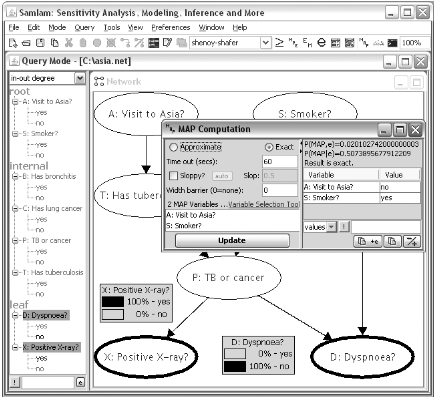
Probability $\approx 20.03\%$ given evidence $\mathbf{e}$: $X=$yes, $D=$no.

Query: Maximum a Posteriori Hypothesis (MAP)

Let $\mathbf{X}$ be all network variables and let $\mathbf{M} \subseteq \mathbf{X}$. Given evidence $\mathbf{e}$, identify an instantiation $\mathbf{m}$ of variables $\mathbf{M}$ for which the probability $\Pr(\mathbf{m}|\mathbf{e})$ is maximal. Instantiation $\mathbf{m}$ is called a maximum a posteriori hypothesis (MAP). Variables $\mathbf{M}$ called MAP variables.

- MPE is a special case of MAP, when $\mathbf{M} = \mathbf{X}$.
- MPE is much easier to compute algorithmically.
- Sometimes, MPE is called MAP.

Query: Maximum a Posteriori Hypothesis (MAP)



MAP variables
$\mathbf{M} = \{A, S\}$ and
evidence
$\mathbf{e}: X = \text{yes}, \ D = \text{no}$

MAP is A=no, S=yes.

MAP has probability of $\approx 50.74\%$ given the evidence.

Query: Maximum a Posteriori Hypothesis (MAP)

A common method for approximating MAP is to compute an MPE and then return the values it assigns to MAP variables. We say in this case that we are projecting the MPE on MAP variables.

Example

MPE given evidence $X = \text{yes}$, $D = \text{no}$:

$$A = \text{no},\ S = \text{no},\ T = \text{no},\ C = \text{no},\ B = \text{no},\ P = \text{no},\ X = \text{yes},\ D = \text{no}$$

Projecting this MPE on MAP variables $\mathbf{M} = \{A, S\}$, we get:

$$A = \text{no},\ S = \text{no},$$

with probability $\approx 48.09\%$ given the evidence.

MAP is $A = \text{no}$, $S = \text{yes}$ with a probability of about 50.74%.

Query: Maximum a Posteriori Hypothesis (MAP)

Let **E** be the evidence variables, **M** be the MAP variables, and let **Y** be all other network variables. If there is at most one instantiation **y** which is compatible with any particular instantiations **m** and **e**, then projecting MPE on MAP variables leads to an exact MAP.

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?

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.

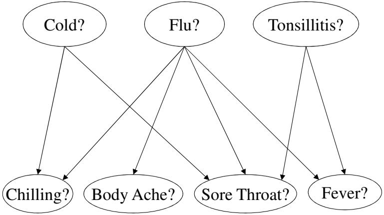
Diagnosis I: Model from Expert

Example

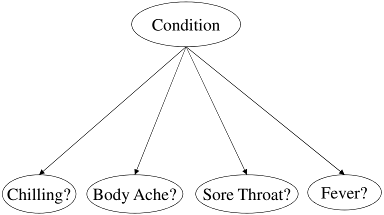
The flu is an acute disease characterized by fever, body aches and pains, and can be associated with chilling and a sore throat. The cold is a bodily disorder popularly associated with chilling and can cause a sore throat. Tonsillitis is inflammation of the tonsils which leads to a sore throat and can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

Diagnosis I: Model from Expert



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



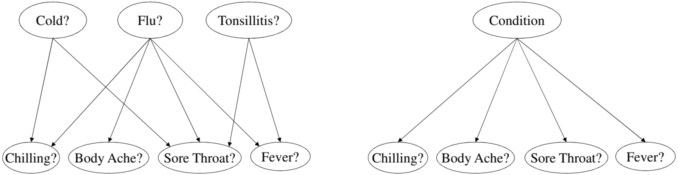
"Condition" has multiple values: normal, cold, flu, and tonsillitis.

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.

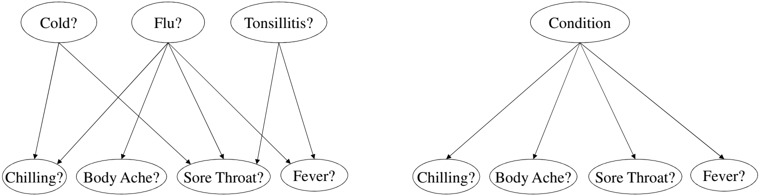
Naive Bayes structure

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 body ache, we expect the probability of flu to go up in both networks.

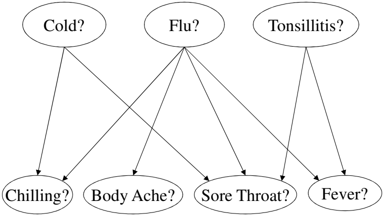
Naive Bayes structure

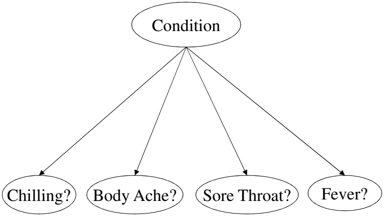
This leads to dropping the probability of cold or tonsillitis.

Original structure

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

Diagnosis I: Model from Expert





Suppose all we know is that the patient has no fever.

Naive Bayes structure

The belief in cold may increase.

Original structure

It remains the same since cold is d-separated from fever.

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 I: Model from Expert

- Tools for Bayesian network inference can generate a network parameterization Θ, which tries to maximize the probability of seeing the given cases.
- If each case is represented by event $\mathbf{d}_i$, such tools will generate a parametrization Θ which leads to a probability distribution $\Pr$ that attempts to maximize:

$$\prod_{i=1}^{N} \Pr(\mathbf{d}_i).$$

- Term $\Pr(\mathbf{d}_i)$ represents the probability of seeing the case i.
- The product represents the probability of seeing all N cases (assuming the cases are independent).
- Parameter estimation techniques will be discussed in Chapters 17 and 18.

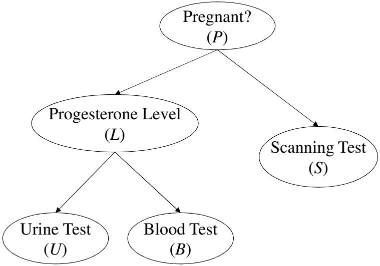
Diagnosis II: Model from Expert

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.

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	$\Pr(P\|\mathbf{e})$
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.

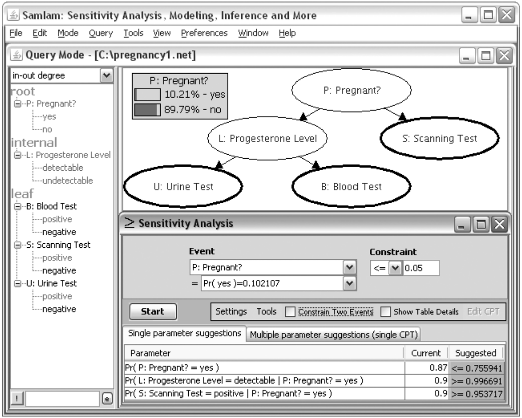
Sensitivity Analysis

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.

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 SamIam 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%.

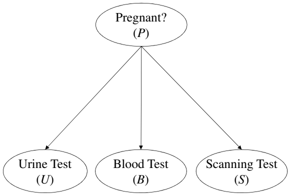
Network Granularity

Progesterone level (L) is neither a query variable nor an evidence variable. We cannot observe the value of this variable, nor are we interested in making inferences about it. The question then is: Why do we need to include it in the network?

We are able to compute the following quantities:

$$\begin{aligned}
\Pr(B = -\text{ve} \mid P = \text{yes}) &= 36\% \\
\Pr(B = +\text{ve} \mid P = \text{no}) &= 10.6\% \\
\Pr(U = -\text{ve} \mid P = \text{yes}) &= 27\% \\
\Pr(U = +\text{ve} \mid P = \text{no}) &= 10.7\%
\end{aligned}$$

We can now build the following network in which the progesterone level is no longer represented explicitly.



P	θ_p
yes	.87

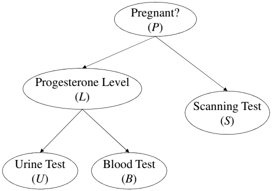
P	S	$\theta_{s\|p}$
yes	−ve	.10
no	+ve	.01

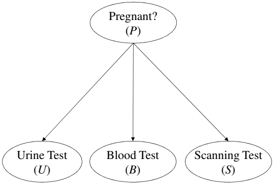
P	B	$\theta_{b\|p}$
yes	−ve	.36
no	+ve	.106

P	U	$\theta_{u\|p}$
yes	−ve	.27
no	+ve	.107

The question now is whether this simpler network is equivalent to the original one from the viewpoint of answering queries.

Network Granularity





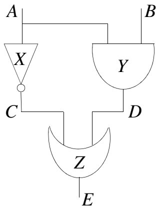
Naive Bayes: blood and urine tests independent given pregnancy

Probability of pregnancy given two negative tests is about 45.09%, given two positive tests is about 99.61%.

Original structure

Probability of pregnancy given these two negative tests is 52.96%, given two positive tests is about 99.54%

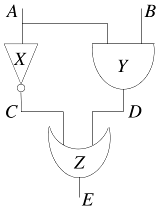
Diagnosis III: Model from Design



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.

Diagnosis III: Model from Design



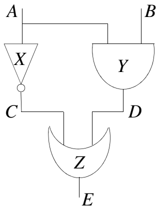
Problem statement

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.

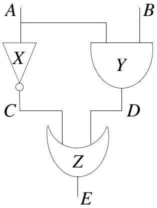
Evidence variables

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.

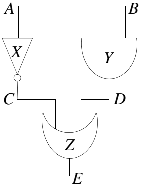
Query variables

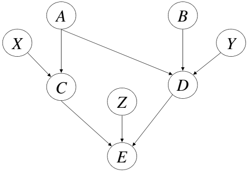
Health of components X, Y and Z.

Intermediary variables

Internal wires, C and D.

Diagnosis III: Model from Design

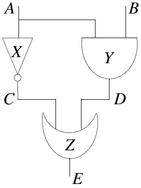


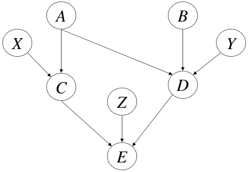


Function blocks

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

Diagnosis III: Model from Design





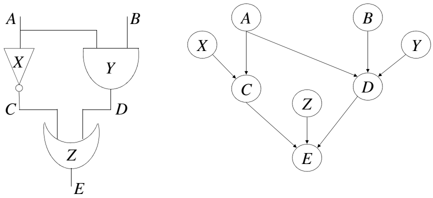
Function blocks

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.

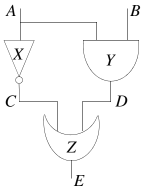
Primary inputs

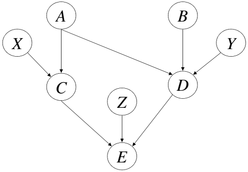
No direct causes for primary inputs, A and B: no parents.

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.

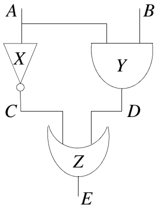
Health states

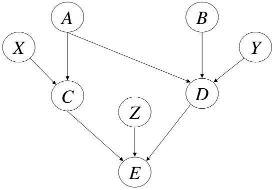
No direct causes for health of X, Y and Z: no parents.

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.

Diagnosis III: Model from Design

Example

CPTs for primary inputs, such as A:

A	θ_a
high	.5
low	.5

Choice for these CPTs does not matter for queries of interest.

Probability of health state x, y, z given test vector a, b, e

is independent of $\Pr(a)$ and $\Pr(b)$ as long as they are not extreme.

Fault Modes Revisited

Choice 1: ok and faulty

X	θ_x
ok	.99
faulty	.01

A	X	C	$\theta_{c\|a,x}$
high	ok	high	0
low	ok	high	1
high	faulty	high	.5
low	faulty	high	.5

Choice 2: ok, stuckat0 and stuckat1

X	θ_x
ok	.99
stuckat0	.005
stuckat1	.005

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

The two choices are equivalent in a particular sense.

A Diagnosis Example

Example

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

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	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:

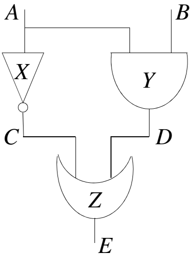
State	X	Y	Z	$\Pr(X, Y, Z \mid \mathbf{e})$
1	ok	ok	ok	0
2	faulty	ok	ok	0
3	ok	faulty	ok	.49374
4	ok	ok	faulty	.49374
5	ok	faulty	faulty	.00499
6	faulty	ok	faulty	.00499
7	faulty	faulty	ok	.00249
8	faulty	faulty	faulty	.00005

- 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:

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

Symmetry when using MAP, but not for posterior marginals.

Lack of Symmetry for Double Faults



Test vector

A = high, B = high, E = low

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