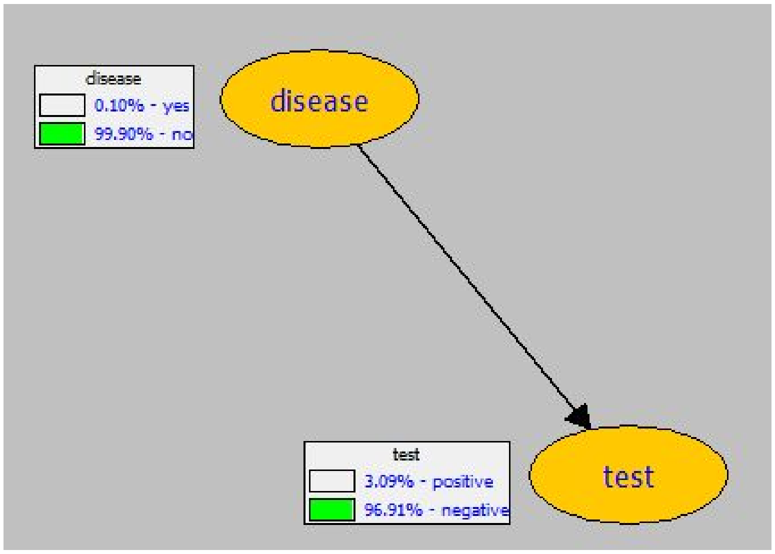
CS 161 Kitu Komya
Homework 8 404-491-375

1. Here is a screenshot of the complete Bayesian structure.



The CPTs are based off of these values:

P(disease)	0.001 (given)
P(~disease) = 1 - P(disease)	0.999
P(test | disease) = 1 - P(~test | disease)	0.94
P(~test | disease) = false negative	0.06 (given)
P(test | ~disease) = false positive	0.03 (given)
P(~test | ~ disease) = 1 - P(test | ~disease)	0.97

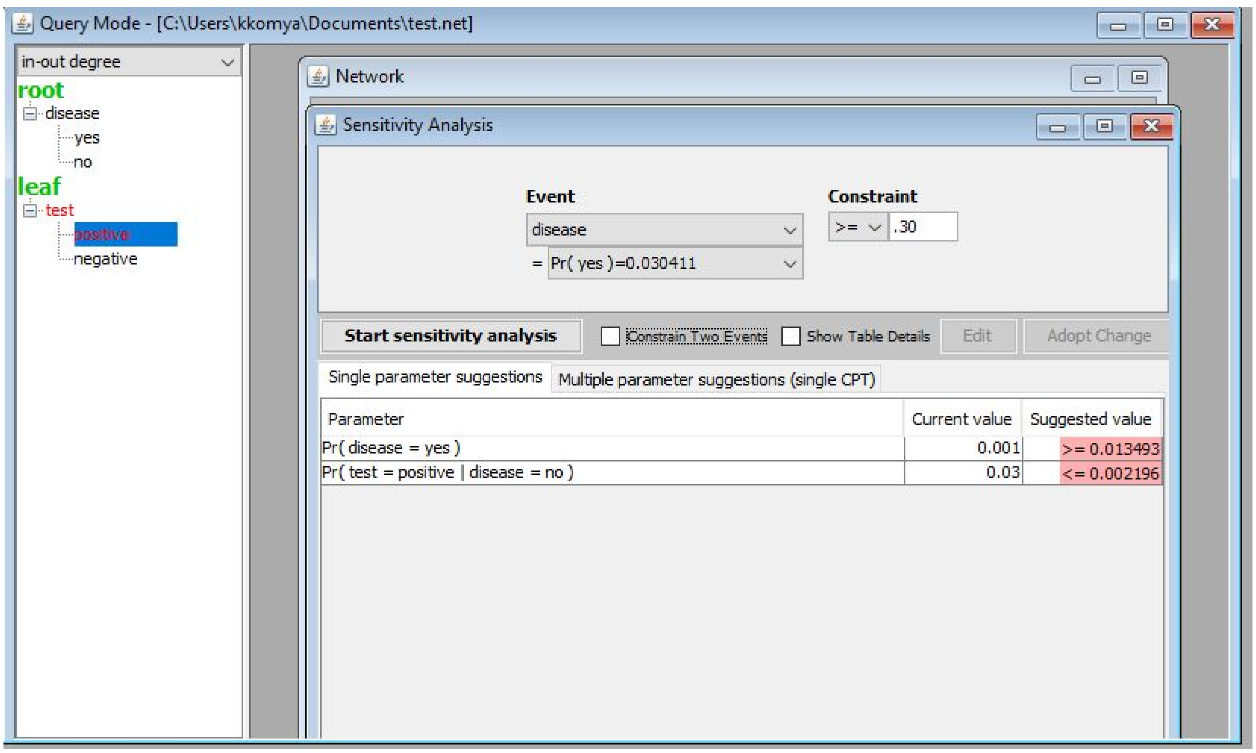
The CPT for disease:

Disease	
Yes	No
0.999	0.001

The CPT for disease and test:

	Has Disease	No Disease
Positive Test	0.94	0.03
Negative Test	0.06	0.97

Sensitivity Analysis:

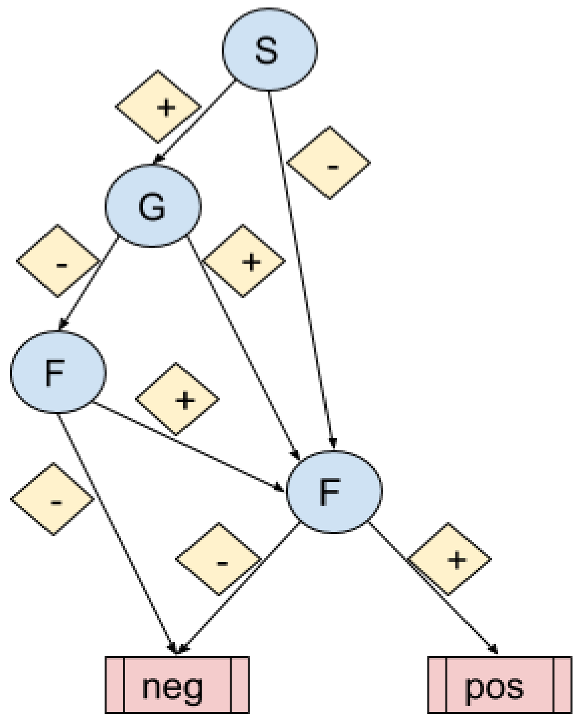


As seen from the above screenshot, in order to attain a probability of at least 30% that a patient has disease given a positive test, two constraints have been suggested:

1. Increase the prevalence of disease (prior probability of the disease) in the population from 0.001 to at least 0.013493. However, infecting a population with disease in order to attain a more accurate reason is neither feasible nor ethical.
2. Decrease the false positive of a test (test reads positive despite patient not having disease) from 0.03 to at most 0.002196. Such a suggestion is ethical but will require a more sensitive instrument, which should be invested in to improve the accuracy.

3. No suggestions on constraining the false negative of the test (test reads negative despite patient having disease) have been suggested. By intuition and ethical responsibility however, a Type II error is much more disastrous, even if reducing this error doesn't improve accuracy to 30%. Patients who are tested negative for a test are much less likely to re-test and confirm that they are not diseased, so minimizing false negatives is crucial for saving humanity.

2. The final, reduced decision graph based on the decision function is illustrated below:



3. <u>Part A</u>

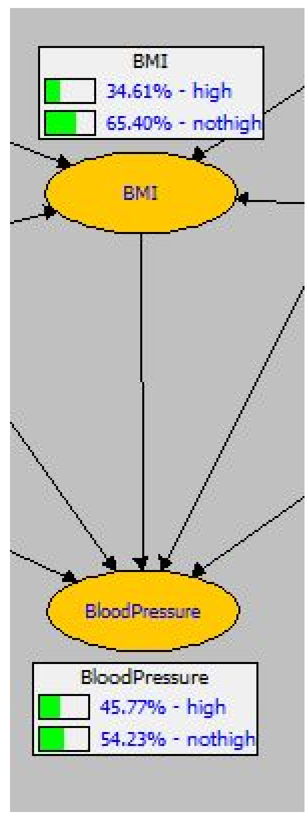
   There are multiple sets of features in which the maximum number of features can be turned off without changing the decision (1) of this instance. The maximum number of features that can be turned off are 3 because if all 4 features are turned off, the dashed-lines of the decision graph lead to the opposite outcome (0).

   One such set is: {S = 0, G = 1, F = 0, M = 0)

since the tree first evaluates S, and given that it's negative, evaluates to the same outcome (1) as long as G is positive. Thus, the values of F and M remain irrelevant.
<u>Part B</u>
Based a bit off the answer from Part A, because the decision outcome is (1), the smallest path to the same outcome is the subset from Part A: {S = 0, G = 1} since once S is defined negative and G is defined positive, the outcome is finalized as (1) regardless of the values of F and M, rendering their values irrelevant.

α is {S = 0, G = 1}
β is {F, M}

4. <u>Part A</u>
The prior marginal distributions are demonstrated from the following screenshot



BMI

High	Not High
0.3461	0.6540

Blood Pressure

High	Not High
0.4577	0.5423

<u>Part B</u>
In order to take into account either high blood pressure or high BMI, a node named bmi_xor_bp was added if one or the other is "high."

bmi_xor_bp

High	Not High

1	0

Posterior marginal distribution of Over60

Older	Younger
.6576	.3424

Part C

Using the SamIam tool, the following results were obtained:

P(MAP, e) = 0.367316415

P(MAP | e) = 0.40812935

where

BloodPressure = Not High

BMI = Not High

Part D

Using the SamIam tool, the following results were obtained:

P(mpe, e) = 0.00781208064

P(mpe | e) = 0.14466815999

where

AgeCausal = Causal

BloodPressure = High

BMI = High

bmi_xor_bp = High

Over60 = Older

Sex = Male

SexCausal = Causal