Exercises for the SMDM 2021 short course:

Cost-effectiveness analysis with probabilistic graphical models

Presenter: Francisco Javier Díez
In collaboration with: Manuel Arias, Manuel Luque, Jorge Pérez-Martín
Dept. Artificial Intelligence, UNED, Spain

This document describes the hands-on exercises that the teachers will solve during the presentation. We recommend that you install OpenMarkov on your computer and solve them yourself right after the course. The file exercises.zip contains some auxiliary material and the solutions, but you should *not* look at them before trying to solve the exercises yourself. If you have any difficulty, ask the teachers by email.

1. Probabilistic diagnosis

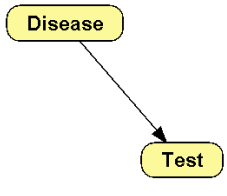
Exercise

The prevalence of a disease in a certain subpopulation is 14%. There is a test for this disease with a sensitivity of 70% and a specificity of 91%.

- What is the positive predictive value of the test?
- What is the negative predictive value?

Hints

- We will create a Bayesian network with two nodes, *Disease* and *Test*, and introduce evidence on the second in order to compute the posterior probabilities of the disease given the test result.
- Create a new network with button 🗋. Check that Network type is *Bayesian network*.
- Create a chance node by clicking on button ○ and then clicking where you wish to place it. Double-click on this node to open the Node properties dialog and change its name to *Disease*. In the Domain tab, check that the states are *present* and *absent*. Click OK.
- In the same way, create the node *Test*. In the Domain tab, assign the states *positive* and *negative*, either by writing the names in the corresponding cells or by selecting them with the Standard domains button.
- Select the button ↗ and draw a link by dragging the pointer from node *Disease* to node *Test*, as shown in the next figure.



- If you wish to move a node, select it with button ↖ and drag it.

- Edit the probability table for *Disease*, either by alt-clicking on this node or by right-clicking on it and selecting the Edit probability option. In the upper cell, write *0.14*, the prevalence given in the statement of the exercise, which is the probability of the disease being present.
- Edit the probability table for *Test*. In the upper right cell, write the sensitivity, 0.70, which is the probability of a positive result when the disease is present. In the lower left cell write the specificity, 0.91, which is the probability of a negative result when the disease is absent.
- Save the network as *BN-one-test.pgmx*, checking that the file type is *ProbModel 0.2 (*.pgmx)*.
- Compile the network with the button  .
- Optional: click  to create a new evidence case. (Each evidence case consists of a set of findings.) This will allow you to compare the posterior probability, i.e., the probability computed after introducing the evidence, with the prior probability.
- Introduce the finding *Test = positive* by double-clicking on the state *positive* of node *Test*. The positive predictive value is the probability of *Disease = present*. This will answer the first question in the statement of the exercise.
- Optional: click  again to create the third evidence case.
- Introduce the finding *Test = negative*. The negative predictive value is the probability of *Disease = absent*. This will answer the second question.

2. Learning Bayesian networks

Exercise

- Learn different Bayesian networks from the dataset *asia10K.csv*, contained in file exercises.zip, using two algorithms: *hill-climbing* and *PC*.
- This dataset contains 10,000 fictitious patients. They were randomly generated by probabilistic sampling from the network *BN-asia.pgmx*, also included in exercises.zip. We will learn a Bayesian network as if we didn't know the true model that generated the data.

Hints

- We recommend that you first open the network *BN-asia.pgmx* (the "true" model) with OpenMarkov and then the dataset *asia10K.csv* with a text editor, to gain some insight about it. The first line of the dataset contains the names of the variables and each of the following lines contains the values for a case (a patient).

Learn a Bayesian network with the hill-climbing algorithm

- In OpenMarkov, go to Tools > Learning and select the dataset *asia10K.csv*. Check that, by default, the algorithm is *hill-climbing* and the metric is *K2*. The alpha parameter (Laplace-like correction) will be used when estimating the parameters of the network, not when learning the structure.
- Click Learn. By default, OpenMarkov will place the nodes in a circle.
- You can accept the edits proposed by OpenMarkov or choose your owns, either by selecting one from the list or by drawing/removing links from the graph. Please note that every time the network is modified, the list of edits is updated.
- Save the network learned, if you wish.

Learn a Bayesian network with the PC algorithm

- Go again to Tools > Learning and replace Hill climbing with PC. Check that initially the significance level is 0.05.
- Click Learn. Accept one by one all the edits proposed by OpenMarkov. Observe that the algorithm has three phases: removing links, orienting pairs of links that converge on a node, and orienting the rest of the links.
- Apply again the PC algorithm for different values of the significance level—for example, 0, 0.001, 0.01, 0.1, 0.5, 0.7, 0.9, 0.99, and 1. You can set it by clicking the Options button after selecting the PC algorithm. Count the number of links in the resulting network for each significance level.

3. Unicriterion decision analysis

Exercise

- For a certain disease, whose prevalence is 2%, there are two tests, with the following properties:

Test	sensitivity	specificity	discomfort	cost
A	0.60	0.92	0.0003 QALY	$100
B	0.80	0.91	0.0001 QALY	$200

- There is also a therapy, costing $7,000, whose effectiveness is:

Disease →	absent	present
therapy	38 QALY	30 QALY
no therapy	40 QALY	20 QALY

- What is the most effective strategy (disregarding costs)?

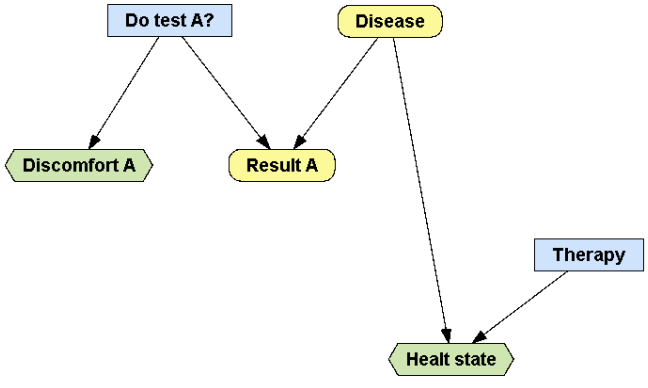
Hints

- Given that the goal is to maximize the effectiveness, we disregard costs at the moment. (We will consider them in the next exercise, which is a cost-effectiveness analysis for the same problem.)

Create the structure of the network (including only test A)

- Click the button 🗋 to create a new model. Network type must be *Decision analysis network*.
- Right-click anywhere on the empty panel to open the Network properties dialog. Go to tab Advanced, select Decision criteria, Standard criteria, and *Cost-effectiveness ($/QALY)*.
- Create the chance node *Disease*; its domain will have the default values, *present* and *absent*.
- Create the chance node *Result A*, with domain {*positive*, *negative*}. You may use the Standard domains button.
- Create the decision node *Do test A?*. Check that its domain is {*yes*, *no*}, the default for decision nodes.
- Create the utility node *Discomfort A*. In the Network properties dialog, the Decision criterion must be "*Effectiveness (QALY)*".
- Create the decision *Therapy*. Its domain will be {*yes*, *no*}, by default.
- Create the utility node *Health state*, whose Decision criterion is also *effectiveness*.

- Draw the causal links, like in the next figure. Remember that you can use the button to select and move nodes, and the button to draw links. If you draw a link by mistake, right-click on it and Remove it.



- Save the network so that you don't lose your work by accident. The name can be, for example, *DAN-two-tests.pgmx*.
- Right-click on the link *Do test A?* → *Result A* and select Add restrictions. Mark the two cells in the left column of the table to indicate that when the test is not done, the result is neither *positive* nor *negative*. The table will look as follows:

Link Restriction: Link between Do test A? and Result A

Do test A?	no	yes
positive	0	1
negative	0	1

- Right-click again on the link *Do test A?* → *Result A*, select Edit revelation conditions, and mark the checkbox for *yes* to indicate that when the test is done, the result (either positive or negative) is known.

Introduce the numerical parameters

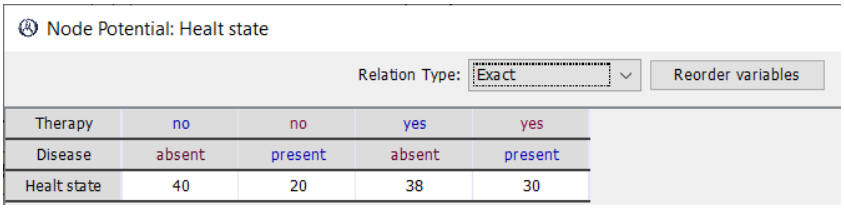
- Open the probability table for *Disease*. Remember that you can do it by alt-clicking or right-clicking on the node. Introduce the prevalence, 0.02.
- Open the probability table for *Result A*. Introduce the sensitivity and specificity of the test, as shown below. You might need to reorder the parents with the Reorder variables button so that the row for *Do test A* appears above the row for *Disease*.

Node Potential: Result A

Relation Type: Table | Reorder variables

Do test A?	no	no	yes	yes
Disease	absent	present	absent	present
positive	0	0	0.08	0.6
negative	0	0	0.92	0.4

- Open the utility table for *Health state*. Set Relation type to *Exact* and introduce the numeric values, as follows:



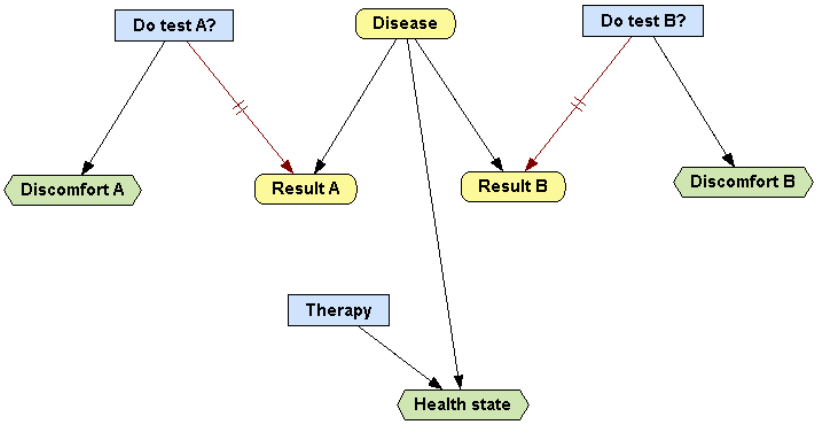
- Open the utility table for *Discomfort A*. Set Relation type to *Exact* and introduce the numeric values: when the test is done, the discomfort is –0.0003 QALY (a negative effectiveness); when the test is not done, there is no discomfort.
- Save the network again.

Evaluate the model

- Click the button  to expand the decision tree. Observe that the option Unicriterion is selected. It means that OpenMarkov will convert cost and effectiveness into a single criterion. In this example there is no cost, so the only criterion will be the effectiveness. Therefore, you must select *QALY* as the unit. Click OK.
- The optimal strategy can be obtained be inspecting the tree and selecting for each decision node the branch that maximizes the effectiveness, which is marked with a red rectangle.
- Another way to get the optimal policy (i.e., an alternative to expanding the decision tree) is to click the S button: OpenMarkov will show the optimal strategy as an intervention tree, which is much more compact than the decision tree.

Add test B and evaluate the model again

- Add now the three nodes (chance, decision, and utility) for test B and the corresponding links, as you did for test A. The network should look like this:



- Save the network as *DAN-two-tests-without-costs.pgmx*.
- If you are unable to complete the exercise, get the network *DAN-two-tests-without-costs.pgmx*, available in exercises.zip in the folder *solutions*.
- Find the optimal strategy for this case, either by expanding the decision tree or by clicking S.

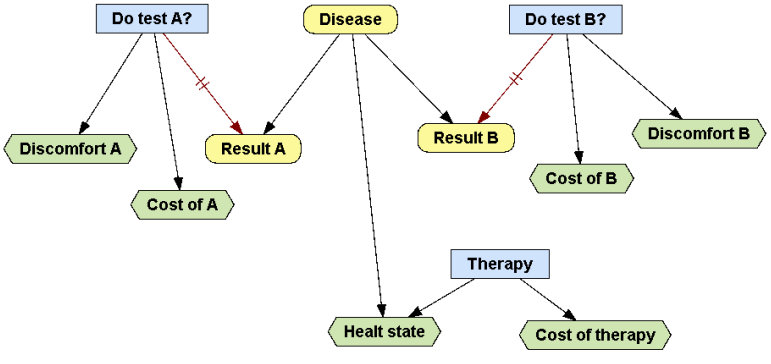
4. Cost-effectiveness analysis (atemporal)

Exercise

- Given the disease and the tests described in the previous exercise, what is the optimal strategy for each value of the willingness to pay (WTP)?

Hints

- Depart from the network *DAN-two-tests-without-costs.pgmx* built in the previous exercise.
- Add the nodes *Cost A*, *Cost B*, and *Cost therapy*, as shown below. Check that the criterion is *Cost ($)* for each of them. Draw the links required and edit the corresponding utility tables. (Note that, even though discomfort was introduced as a negative effectiveness, all costs are positive.)



- Save the network as *DAN-two-tests-with-costs.pgmx*.
- Again, there is a copy of this network available in exercises.zip in the folder *solutions*.
- Click the button  to do a cost-effectiveness analysis and answer the question posed in the statement of the exercise.

5. Cost-effectiveness analysis with a Markov model

Exercise

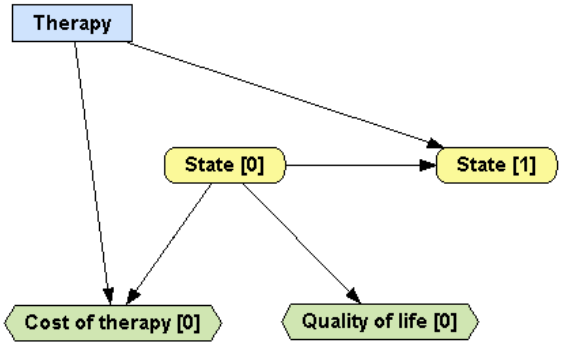
- A disease D may be latent or active. When it is latent, the probability of becoming active in the next month is 11% and that of dying is 2%. When the disease is active, the monthly probability of dying is 15%. Quality of life is 0.9 for latent disease and 0.7 for active. The cost of the standard therapy is $150/month when the disease is latent and $2,500/month when it is active.
- There is a new therapy, with a cost of $950/month, which slows down the progression of the disease, so that the monthly probability of becoming active reduces to 8%, without affecting the probability of death. Unfortunately, this therapy has no effect when the disease is already active.
- With an annual discount rate of 3.5% for both cost and effectiveness, what is the ICER of the new therapy?

Hints

Create the model

- Create a new network of type *Markov influence diagram (MID)*.

- Right-click on an empty zone of the network panel to open the Network properties dialog. Go to the Advanced tab, select Criteria, Standard Criteria, and "Cost-effectiveness ($/QALY)". In the same tab, select Temporal options and set the Cycle length to 1 month.
- Create the chance node *State [0]*; i.e., the name must be "State" and the time slice "0". The domain must have three states: *latent*, *active*, and *dead*. (By default this node has two states. You must add a third one and change their names.)
- Right-click on this node and select Create node in next slice. This will add the node *State [1]*, with the same domain as *State [0]*, as shown in the next figure.



- Create the decision node *Therapy*, with two states in its domain: *new* and *standard*. It represents the decision of whether applying the new therapy when the disease is latent. (When the disease is already active, it is useless to apply it.)
- Create the node *Cost of therapy [0]*; i.e., the name must be "State" and the time slice "0". Check that its criterion is "Cost ($)".
- Create the node *Quality of life [0]*. Change its criterion to *Effectiveness (QALY)*.
- Draw the links from *State [0]* and then those from *Therapy*, as shown in this figure.
- Edit the probability table for *State [0]* to indicate that initially all patients are in state *latent* (this is the subpopulation to which it may be worth applying the therapy).
- Edit the probability table for *State [1]* to indicate the transition probabilities. It must look like this:

Node Potential: State [1]

Relation Type: Table — Reorder variables

State [0]	dead	dead	active	active	latent	latent
Therapy	standard	new	standard	new	standard	new
latent	0	0	0	0	0.87	0.9
active	0	0	0.85	0.85	0.11	0.08
dead	1	1	0.15	0.15	0.02	0.02

- Edit the utility for *Cost of therapy [0]* and set the relation type to *Exact*. The monthly cost, $950, only applies when the decision about therapy is *new*—instead of *standard*—and the disease is *latent*. When the disease is *active*, the standard therapy is applied because, as the statement of the exercise says, it would be useless to apply the new therapy when the disease is already active.

Node Potential: Cost of therapy [0]

Relation Type: Exact — Reorder variables

Therapy	standard	standard	standard	new	new	new
State [0]	dead	active	latent	dead	active	latent
Cost of therapy [0]	0	2500	150	0	2500	950

- Edit the utility for *Quality of life [0]* and set the relation type to *Exact*. The quality of life is 0.0 for *dead*, 0.7 for *active*, and 0.9 for *latent*.
- Save the network as *MID-latent-active.pgmx*.

Check that the model works properly

- Examine how the probability of *State* evolves by right-clicking on *State[0]* or *State [1]* and selecting Temporal evolution. Set the horizon to *60 cycles*, i.e., 5 years, and accept the default value One decision. You will obtain 6 curves, showing the probability of each state along time depending on the therapy applied.

  In order to know the probability of being alive in each cycle (i.e., the proportion of patients that are alive), uncheck the box for *dead* and click the radio button Sum; this way OpenMarkov will show the sum of the probabilities of *latent* and *active*. Observe that the new therapy clearly increases life expectancy.

  Also observe that almost all patients are dead after 60 cycles; the exact values are shown in the Table tab. Therefore, this horizon is expected to give a good approximation to cost and effectiveness. However, we might extend this horizon—for example to *84 cycles*, i.e., 7 years—to obtain more accurate values.

  If you wish to know the probability of the disease being *active* in each cycle, also uncheck the box for *latent*. Observe that the new therapy makes patients transition more slowly to state *active*.
- Examine the temporal evolution of *Cost of therapy*. Observe that it the monthly cost decreases as patients transition from *latent* (the only state in which the new therapy is applied) to the states *active* or *dead*. Click the Cumulative radio button to obtain the accrued cost and observe that, as expected, increases for the new therapy and is null for the standard therapy.
- Examine the temporal evolution of *Quality of life* and observe that, in every cycle, it is higher for the new therapy. Click the Cumulative radio button and observe that the accrued quality of life, i.e., the effectiveness, is higher for the new therapy.

Cost-effectiveness analysis

- Click the button  . Each discount can be set per year or per cycle. You must set both of them to *3.5% per year*, in accordance with the statement of the exercise. During the calculations, OpenMarkov will transform them into per-cycle (i.e., monthly) discounts. In the Scope selector, accept the default value, *One decision*, and click OK.

  The Analysis tab shows the cost and effectiveness of each intervention.

  The CE plane tab shows them on a plot. On the upper right corner, click Display Relative to and select *standard* therapy, to see the incremental cost and effectiveness of the new therapy (instead of the absolute values).

  The Frontier interventions tab shows the non-dominated interventions, as well as the ICER we were looking for, i.e., $3,733.50 per QALY, which means that the new therapy is clearly cost-effective.

6. Sensitivity analysis

Exercise

- The uncertainty about the cost of the new therapy analyzed in the previous exercise can be modeled by a Gamma distribution with a mean of $950 and a standard deviation of 20% of the mean.
- The transition probabilities were obtained from a one-month randomized control trial involving 400 volunteers with latent disease. Half of them received the standard therapy; 4 of them died; the disease

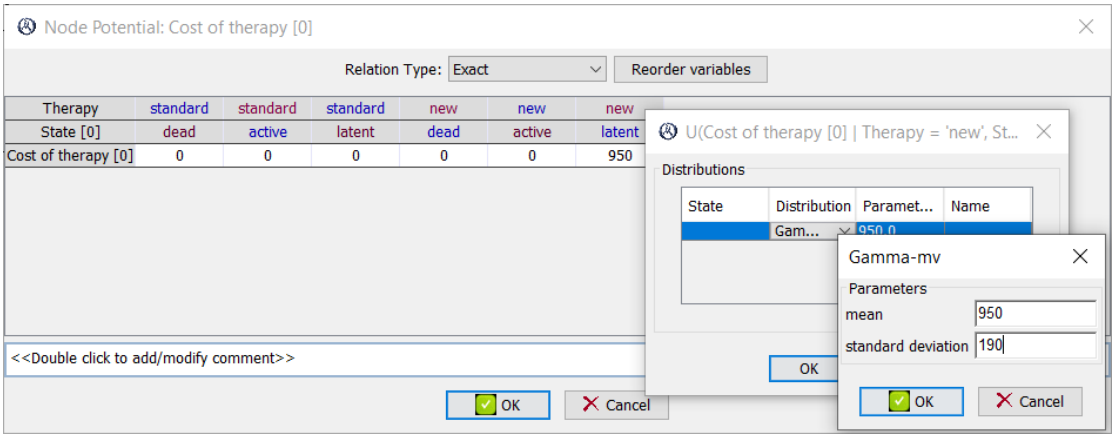
became active in 22 subjects and remained latent in the rest. The other half received the new therapy: 4 of them died (as many as in the control group) but only 16 transitioned to active disease.

- Draw the scatter plot and the acceptability curve.
- Is the new therapy cost-effective for a willingness to pay (WTP) of $50,000/QALY?
- What is the probability of the new therapy being more effective (in QALYs) than the standard one?

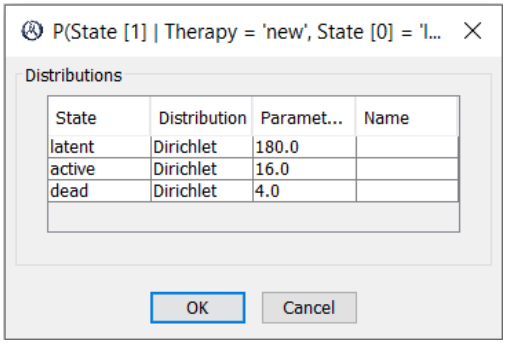
Hints

Encode the second-order uncertainty into the model

- Open the edit probability dialog for *Cost of therapy [0]*. Right-click on the *latent-new* cell, whose value was 950, and select Assign uncertainty. Change the distribution from *Exact* to *Gamma-mv* (where "mv" stands for "mean and variance") and introduce the two parameters—remembering the standard deviation is 20% of the mean—as shown in the next figure. Click OK. The cell in the utility table now has a red triangle, meaning that there is uncertainty about this value.



- Open the utility table for *State [1]*. Right-click on any numeric cell in the *latent-new* column and select Assign uncertainty. Change the distribution of each parameter from *Exact* to *Dirichlet* and introduce the corresponding parameters as shown in this figure. Click OK and observe the red triangles for this column.



- In the same way, assign a Dirichlet distribution to *latent-standard* column, with parameters 174.0, 22.0, and 4.0. Check that the probabilities are the same as in the previous exercise, but now there are red triangles denoting uncertainty about some transition probabilities.
- Save the network as *MID-latent-active-uncert.pgmx*.

Perform a probabilistic sensitivity analysis

- Click the button . Check that Multi criteria selection is *Cost-effectiveness*. (If it were *Unicriterion*, OpenMarkov would offer other sensitivity analysis options, such as tornado diagrams, plots, etc.) Click OK.
- In the next dialog, check that Analysis type is *Cost-effectiveness plot* and that OpenMarkov will run 1,000 simulations. Click OK and wait.
- OpenMarkov will show a scatter plot on the cost-effectiveness plane. All the points corresponding to the standard therapy lie in a horizontal line because there is no uncertainty about its cost.

  On the upper right corner, set the Display as relative to *standard* therapy. This way the plot will show the incremental cost and effectiveness of the new therapy (in blue) with respect to the standard one (in red) for each simulation.
- Select the tab Acceptability curve. On the upper left corner, set the WTP of reference to 50,000/QALY, the value given in the statement of the exercise, and observe that for this threshold the probability of the new therapy being cost-effective is around 76%.
- Setting the WTP of reference to a very high value (for example, $1,000,000/QALY) is equivalent to aiming at the most effective therapy, disregarding its cost. In this case the probability of the new therapy being cost-effective is just the same as the probability of being effective. The plot shows that this probability is around 78%. This percentage corresponds to the proportion of simulations that lie at the right of the coordinate axis in the scatter plot we obtained when setting the Display relative to *standard* therapy.