agena.ai modeller

User Manual

agena.ai

Copyright Notice

Contents

1. Introduction

Agena.ai modeller uses the latest developments from the field of Bayesian artificial intelligence to model complex, uncertain problems and improve how decisions are made. Data scientists, analysts, and AI engineers can easily create and deploy models as cloud-based applications for up to thousands of users.

The agena.ai modeller is a design and execution environment for creating Bayesian networks which runs on Windows, Linux and Macintosh operating systems. Agena.ai modeller delivers state of the art algorithms for Bayesian model computation coupled with an easy-to-use interface. Models developed in agena.ai modeller can then be deployed to the agena.ai cloud environment as web applications or as computational APIs.

Agena.ai modeller:

- Offers numerous advantages over 'big data’ approaches: It copes with incomplete data and represents real world causal interactions. Bayesian models can carry out prediction and abduction (diagnosis) simultaneously and combine both causal and statistical information.

- Combines the benefits of Bayesian Networks (referred to throughout this manual as Bayesian Networks or BNs, but also known as probabilistic graphical models), statistical simulation and decision analysis.

- Provides an extensive library of models covering a huge number of application areas including Project risk, operational risk, stress testing, legal reasoning, medical diagnosis, financial decision making, value of information analysis and more.

- Is visual, easy to use, intuitive and powerful.

1.1 Before you start

Before you start using agena.ai modeller you need to be familiar with the basics of BNs, probability and uncertainty. We recommend that you order the book:

Risk Assessment and Decision Analysis with Bayesian Networks
Norman Fenton and Martin Neil

ISBN 10: 1138035114, ISBN 13: 978-1138035119

You can obtain the book from Amazon:

https://www.amazon.com/Assessment-Decision-Analysis-Bayesian-Networks/dp/1138035114

There is also an agena.ai User Group: https://www.linkedin.com/groups/7473890

1.2 Key features

Key features of agena.ai modeller include:

- **Bayesian network graph drawing,** you can build Bayesian Networks for modelling causal and other relationships
- **Charts** and statistics with zoomable and scalable charts, histogram, and area plots, summary statistics and graphs overlays, as well as percentile and cumulative plots.

- **Learning from data** for Gaussian discrete nodes using the EM (Expectation-Maximization) algorithm and tailored using constructed models.
- **Node probability tables**, expression, and partitioned expressions.
- **Statistical distribution functions**, including Normal, Beta, Binomial etc. using dynamic discretization for greater accuracy when performing inference using continuous variables.
- **Formulae expression parser** including Noisy-OR, Ranked nodes, comparative and other mathematical functions.
- **Sensitivity and multivariate analysis** to assess impact of number of variables on one target variable.
- **Hybrid influence diagrams** for decision making, producing **Decision trees**.
- **Value of Information Analysis**, to help determine how much to pay for information.
- **Bayesian network modelling** for modular model construction.
- **Compound sum analysis** for risk aggregation of event frequency and severity variables.
- **Model library** of comprehensive examples linked to tutorial material in book.
- **Data import and export** via CSV, HTML, XML and JPEG files.
- **Model export to agena.ai cloud services**, to upload your model to be executed over the internet as an API or via a web application.

1.3 Types of modelling you do in agena.ai modeller

The state-of-the-art algorithms implemented in agena.ai modeller allow you to do the following types of problem solving and modelling:

- Simulation of statistical distributions for predictive inference using dynamic discretization as a method of simulation in (Hybrid) Bayesian Networks
- Diagnostic inference for machine learning applications
- Hierarchical modelling as an alternative to Monte Carlo Markov Chains (MCMC)
- Construction of hybrid models containing discrete and continuous uncertain variables
- Representation of expert judgement using subjective probability
- Combing data with expert judgment
- Mixture modelling of discrete and continuous distributions
- Object oriented modelling of complex systems involving multiple objects and interfaces
- Dynamic modelling of time-based or evolving systems (such as Markov analysis)

1.4 Agena.ai versions

- **Agena.ai modeller** is a design and execution environment for Bayesian Networks which runs on Windows, Linux and Macintosh operating systems. Models developed in agena.ai modeller can be deployed within agena.ai cloud services.
- **Agena.ai cloud service**. The agena.ai cloud service includes a computational cloud API that executes models developed using agena.ai modeller, and a design environment to create and deploy Web Apps to your end users. Agena.ai Cloud Service is a hosted service that uses the latest technologies including Kubernetes and Kafka. The Agena.ai cloud service comprises a Web App Designer to design and deploy web applications that use your model and API server for deploying models as part of your own cloud-based applications.
- **Agena.ai APIs** are available under subscription license in Java, Python, and the 'R' statistical programming language to build an execute larger models.

Academic licenses are available for academic institutions for research and teaching purposes.

1.5 Getting started with agena.ai modeller

Part A of this manual provides a set of tutorials to help you get started using agena.ai modeller. These tutorials are also available from within the application by clicking on Tutorials in the Help menu.

In addition, agena.ai modeller comes with a wide range of documented example models, contained in the example model library. These can be found in the Model Library subdirectory of your installation and can be accessed from the File open dialog each time you run agena.ai modeller.

1.6 Buying Agena.ai modeller

To buy a subscription licence for your copy of agena.ai modeller, please visit our website:

agena.ai

1.7 Contacting Agena

For help with installing or running agena.ai modeller, or for general enquiries, please send an email to:

support@agena.ai

We would very much appreciate any feedback you might have about your experiences when using agena.ai modeller. Unfortunately, we cannot offer advice on modelling or algorithms. Agena personal offer training and consulting services if you require help in this area, if interested contact sales@agena.ai.

1.8 About this document

This document is in two parts. Part A contains the tutorials, which take you through, step-by-step, a variety of tasks that cover most of the tool's functionality. Part B is an extended tool reference that covers every aspect of the tool's functionality. In addition, there are appendices covering: Introduction to Modelling using agena.ai, Hints and tips on building Bayesian networks.

Throughout this manual, the formatting conventions in Table 1 are used.

Table 1 Formatting conventions

File	A term appearing in bold will sometimes refer to a menu on the toolbar that you are required to select.
OK	Bold is also used to indicate any button that you are required to click.
23	Bold can be used to indicate any value that you are asked to enter or select.
Menu → Menu Item	Two or more bold terms separated by an arrow indicate a path from a toolbar menu (or from the Start menu) down to a menu item that you are required to click.
Ctrl + W	The use of Ctrl, Alt or Shift in conjunction with a + symbol and a letter indicates that the key should be pressed down simultaneously with the letter to invoke a particular function.
Filename.ast	Names of files and directories (used during opening, saving and importing) are indicated by italics.
Directory / Filename.ast	Two or more italicised terms separated by a forward slash indicate a path from a directory (or folder) down to a particular file.
ℹ	This icon indicates a tip or other piece of useful information.
⚠	This icon indicates a warning; you should pay particular attention to these.

2. The user interface

2.1 User interface overview

The main user interface components of agena.ai modeller are shown in the application window in Figure 1.

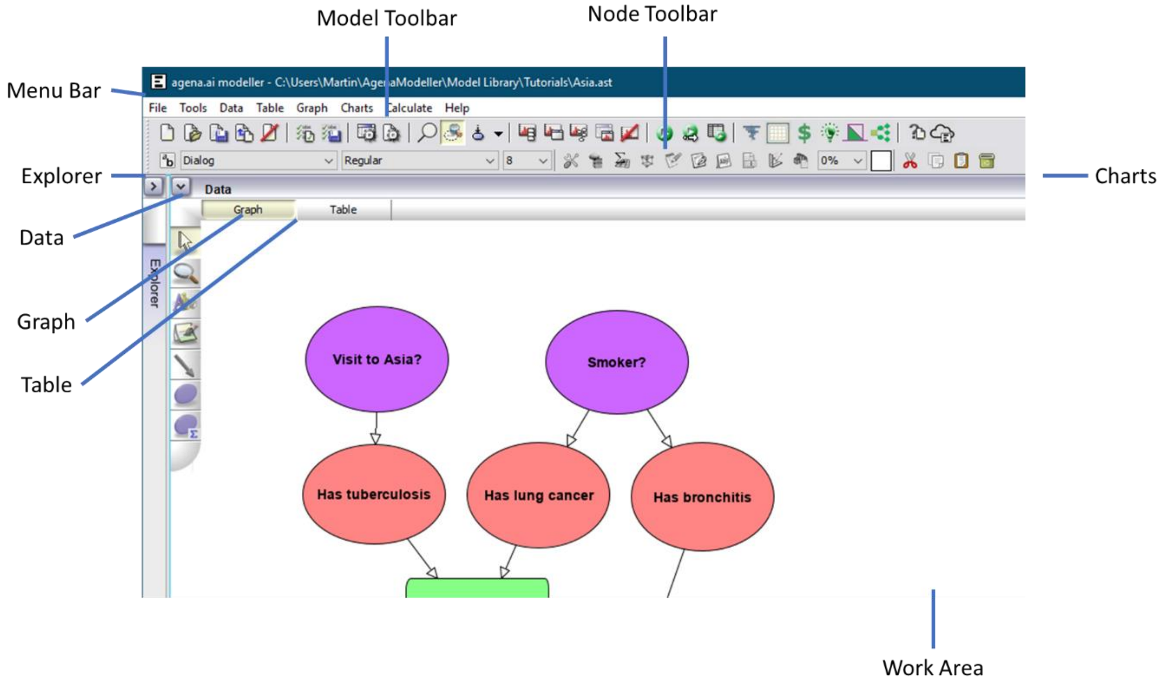


Figure 1 The agena.ai modeller user interface

It has the following main components:

- The **work area.** This has two views (there are tabs for switching between the two), namely a **graph view** and a **table view**. In either case the work area contains the associated details of the item currently selected in the explorer panel.

- The **explorer panel.** This is displayed by clicking on the icon >. It contains a list of Bayesian networks that are part of the currently open model. Each Bayesian network will have its own corresponding table and graph.

- Within the graph view there is the **graph view toolbar** which contains controls to select objects, add nodes, edges, pictures, and text as well as zoom.

- The **Data panel** allows modification and activation of the cases connected to any Bayesian network in the explorer view.

- The **charts panel.** This is displayed by clicking on the icon <. It contains a set of graphs, where a graph is an item that displays the probabilities for a given node in the underlying model after a calculation.

- The **‘model’ toolbar** contains a set of buttons, which allow quick access to functions available via the menu bar or elsewhere in agena.ai modeller.

- The **node toolbar** contains a set of buttons to edit node properties, such as NPTs and expressions, enter constants as well as copy, cut and paste nodes and collections of nodes.

- The **menu bar** includes functions for loading, saving and importing models, propagating models (calculation), report generation, chart manipulation, and table and data management.

- The menu and tool bar (Figure 2) contains functions to load, save and import models and cases; to modify tables; to configure cases; to configure the graph view; and to calculate. This section describes the functions available.

Figure 2 Menu Bar and Tool Bar

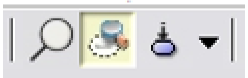
- Note 1: When using agena.ai modeller if you need a quick reminder of the purpose of any toolbar button, simply move the mouse over it and you will see some ‘tool tips’ text that summarises the purpose of the button.

- Note 2: In addition to menu items and tool bar buttons there are keyboard shortcuts for most common functions.

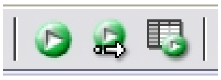
- Note 3: Each toolbar button is associated with a menu option (but not vice versa) as described in this section. It is also worth noting that the toolbar is divided into the following ‘functionally related’ groups of buttons:

   This group deals with file functionality

   This group deals with loading and saving cases and with the configuration of the application and the currently loaded model:

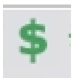
   This group deals with showing status information on the graph view

   This group deals with the graph view

   This group deals with calculation

-  This button launches the Sensitivity analyser.

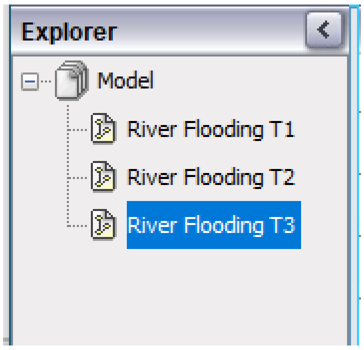
-  This button launches the Multivariate analyser.

-  this button launches the Value of Information analyser

- This button launches the Learning functionality
- This button launches the Compound Sum Analyser
- This button launches this agena.ai user manual in Acrobat Reader or your default pdf file reader.
- This button launches the agena.ai cloud service portal in your web browser
- The graph view toolbar is shown as:

2.2 The Explorer view

The explorer view is on the left-hand side of the application window. It contains a complete list of all the Bayesian networks that are contained within the model, as shown in Figure 3.



Minimize/maximize
Explorer Panel

Figure 3 The explorer panel

Clicking on a Bayesian network in the tree view will select it. Full details of that Bayesian network will subsequently be displayed in the work area (on the right-hand side of the application window). Either the graph view or the table view will be shown depending on which of the corresponding tabs is selected in the work area.

Any item in the explorer can be renamed; this is crucial if you want to create your own tailored version of an existing model. To rename an item, right-click on it and select ‘Rename’ from the pop-up menu. Type the new name into the dialog and click OK.

If the top-level item is selected, then a view of the top-level model will be displayed in the work area.

Any Bayesian networks in the current model can be deleted by right-clicking on the object you wish to remove and selecting the delete function from the pop-up menu.

The explorer panel takes up a reasonable amount of screen real estate. For this reason, it can be minimised and maximised via the arrow button at the top of the panel. The width of the panel can also be changed by dragging the bar that separates the panel from the work area.

2.3 The Graph view

The graph view, as shown in Figure 4, is available when a Bayesian network is selected in the explorer and the graph view tab is selected in the work area. The only exception is if the model in the explorer view is selected, in which case a graphical view encompassing all the objects in the model will be displayed.

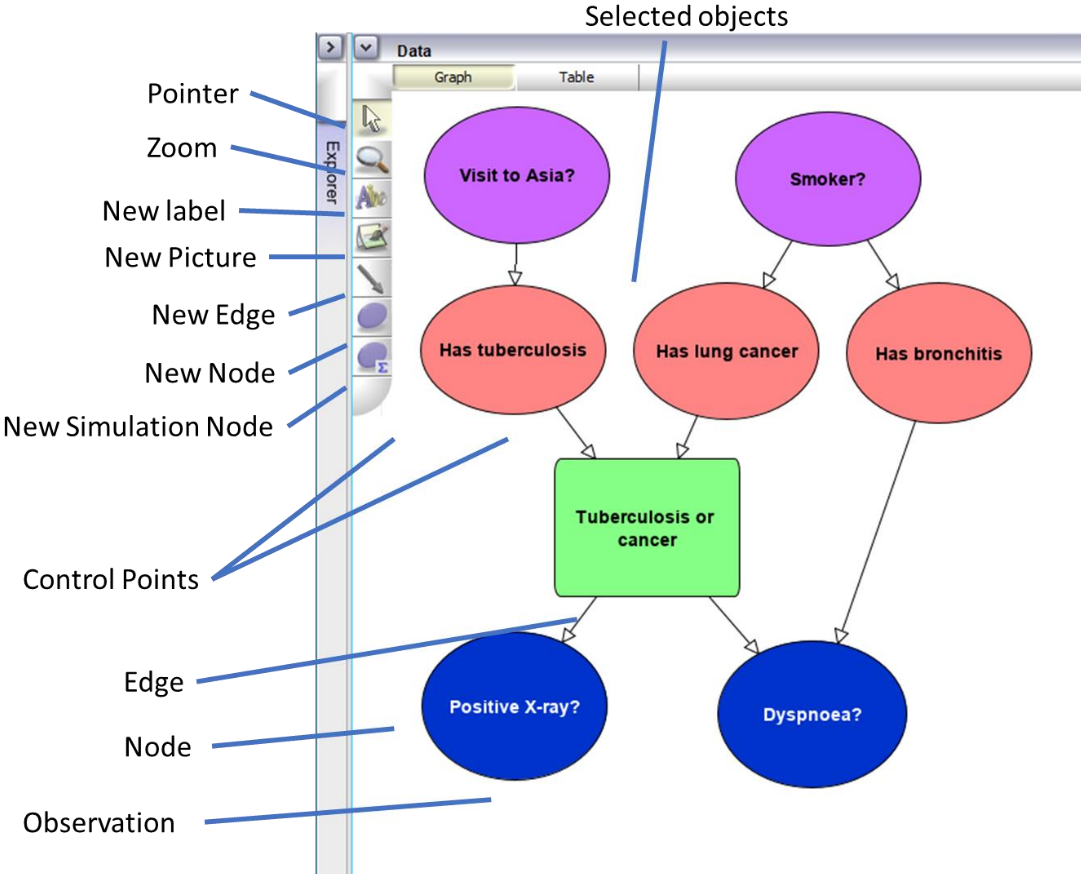
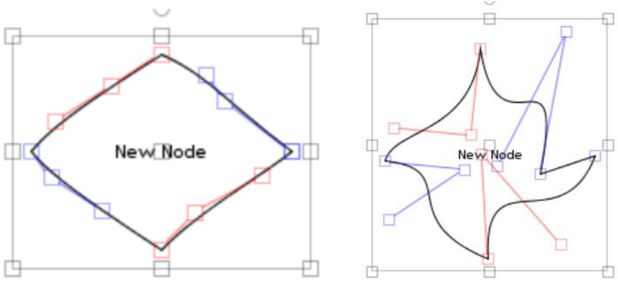


Figure 4 The graph view

The graph view displays a fully interactive graphical view of the object selected in the explorer. The nodes are shown as shapes connected by lines (or edges).

Nodes and edges can be selected in the graph view toolbar. When a node or an edge is selected, several control points will appear around it (they appear as small light grey boxes). Clicking, holding then dragging the mouse over these control points allows you to resize selected nodes. Clicking, holding and dragging the mouse within their outline will move nodes and edges.

If you hold down Shift and Alt and click on a node, several other control points are made available. These allow you to rotate the shape and move its individual parts. This enables you to create nodes of many different shapes.



Edges are generally anchored to nodes. When nodes are moved the edges move with them. They cannot be moved independently of the nodes.

Groups of nodes can be selected by clicking and dragging out over the selected components. Alternatively, multiple objects can also be selected by holding down the Control key while selecting objects. By default, during an area selection the edges are not selected. If you want to select edges as well during the area selection hold down the control key. Right-clicking on a node or an edge will bring up a menu.

The following items are available on the menu:

- **Display Chart** This will bring up the charts for all selected nodes. The choice of item on the sub-menu will determine where the charts are displayed.

- **Enter Data** This allows you to enter data into the selected node. Data are entered into what are called *cases*. When this menu is selected you will be prompted to select which case you want to enter data into and then what the data should be. There are two mechanisms for entering data: selecting from a discrete list or entering a number in a text field. The answering mechanism is inherited from the table entry connected to the node. Both the non-numeric and numeric sub-menus contain a ‘Clear Data’ menu item that will remove any data from the node for the specified case.

- **Properties** This will display a properties dialog for the selected node or edge. If multiple objects are selected, then it will display the properties dialog relevant to all selected objects.

2.4 The node toolbar and node properties

In agena.ai modeller most of the features are accessible via a node’s properties. This node toolbar enables you to edit node properties in the graph view directly without having to right click on the node and bring up the Node properties dialog.

The following example explains the different modes of accessing and editing a node's properties.

> **Example**: Suppose we want to edit the Node Probability Table of the node named "Dyspnoea?" in the example model *asia.ast*. The 'long route' require you to follow the steps shown in Figure 5.

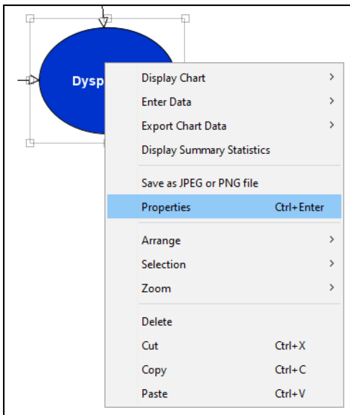
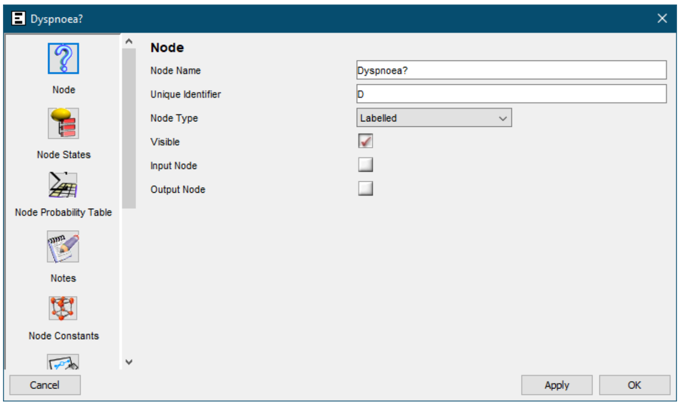
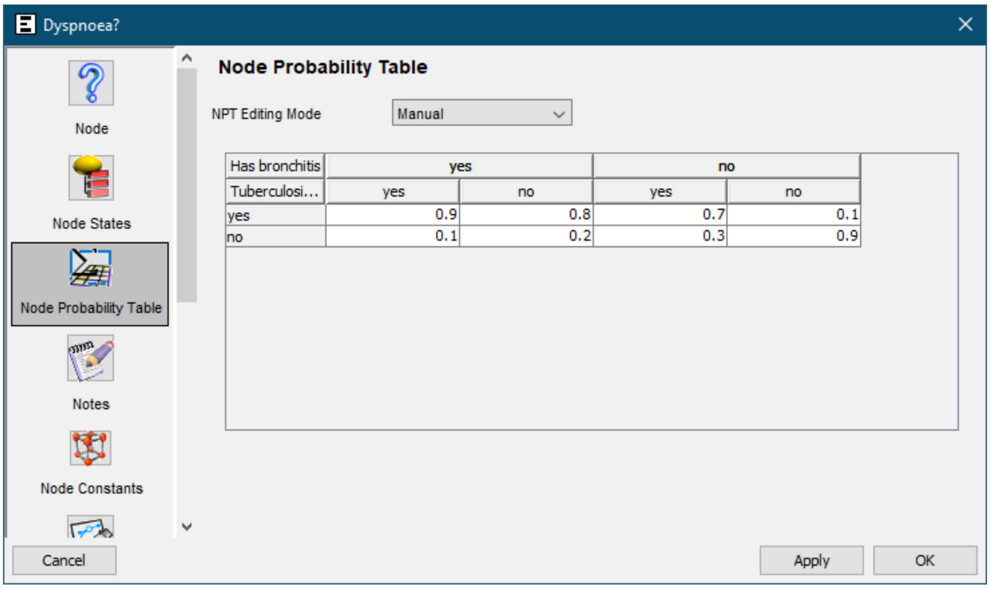
	 **Node** Node Name: Dyspnoea? Unique Identifier: D Node Type: Labelled Visible ☑ Input Node ☐ Output Node ☐
(i) Select node then right click and select "Properties"	Next select the option "Node Probability Table"

Figure 5 'Long route' way of editing NPT

This brings up the Node Probability Table (ready for editing) shown in Figure 6.



Node Probability Table

NPT Editing Mode: Manual

Has bronchitis	yes		no	
Tuberculosi...	yes	no	yes	no
yes	0.9	0.8	0.7	0.1
no	0.1	0.2	0.3	0.9

Figure 6 Node Probability Table

Using the model toolbar, you simply select the node and then click on the Node Probability icon as shown in Figure 7.

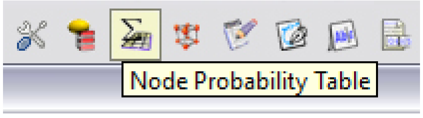


Figure 7 Selecting the Node Probability Menu Icon

This will immediately bring up the Node Probability Table window shown in Figure 6.

All of the related toolbar icons work in exactly the same way. The details are provided in Table 2.

Table 2 Icons in new toolbar

Agency FB	This is the node label's **text font property**. Simply select the font you want using the drop-down menu and this will immediately change the font of the node label text.
Regular	This is the node label's **text style property** (the options you can select are regular, italic, bold). Simply select the style you want using the drop-down menu and this will immediately change the style of the node label text.
8	This is the node label's **text size property**. Simply select the size you want using the drop-down menu and this will immediately change the size of the node label text.
	This brings up the **Node Properties dialog** – selecting this is equivalent to right clicking on the node and then selecting "properties".
	This is the node's **Node States** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Node States".
Σ	This is the node's **Node Probability Table** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Node Probability Table".
	This is the node's **Node Constants** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Node Constants".
	This is the node's **Notes** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Notes".
	This is the node's **Node Appearance** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Node Appearance".
	This is the node's **Text Formal** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Text Format".

	This is the node's **Table Entry** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Table Entry" .
	This is the node's **Chart Defaults** property. Selecting this is equivalent to right clicking on the node, selecting "properties", and then selecting "Graph Defaults".
	This is the node's **Temporary Chart Settings** property, which is set for each case. Selecting this is equivalent to right clicking on the node, selecting "properties", then selecting "Chart Defaults" and finally selecting "Temp Chart Settings".
0% ▾	This is the node's **colour transparency property**. Simply select the percentage you want using the drop-down menu and this will immediately change the transparency. The default is zero; selecting higher values makes the colour more transparent. Selecting this is equivalent to right clicking on the node, selecting "properties", then selecting "Appearance" and then selecting "Transparency".
☐	This is the node's **colour property**. Selecting this will bring up the colour palette. Selecting this is equivalent to right clicking on the node, selecting "properties", then selecting "Appearance" and then selecting "Colour".
✂	These icons are respectively cut, copy, past, delete. The selecting the cut icon, for example is equivalent to right-clicking the node and selecting "Cut".

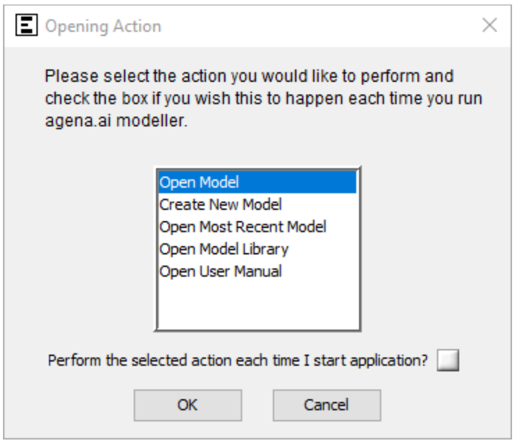
PART A: Tutorials

3. Using the agena.ai modeller user interface

This tutorial teaches you the basics of using the agena.ai modeller user interface. In the first part of the tutorial, you will learn how to open and use a simple model. In the second part, you will learn how to build this model from scratch.

3.1 Opening and using a simple model

1. Start agena.ai modeller by clicking on the agena.ai modeller desktop icon or by selecting **Start → agena.ai modeller → agena.ai modeller**. If this is the first time you have run agena.ai modeller, you will be presented with the dialog shown in Figure 8 below.


Opening Action

Please select the action you would like to perform and check the box if you wish this to happen each time you run agena.ai modeller.

Perform the selected action each time I start application? ☐

Figure 8 Opening dialog

2. Select **Open Example Model** and click **OK**.

3. You will now be presented with a dialog inviting you to select an agena.ai model file to open.

4. Double-click down into the *Tutorials* folder, select the *Asia.ast* file and click **Open** and you will see the open dialog show in Figure 9.

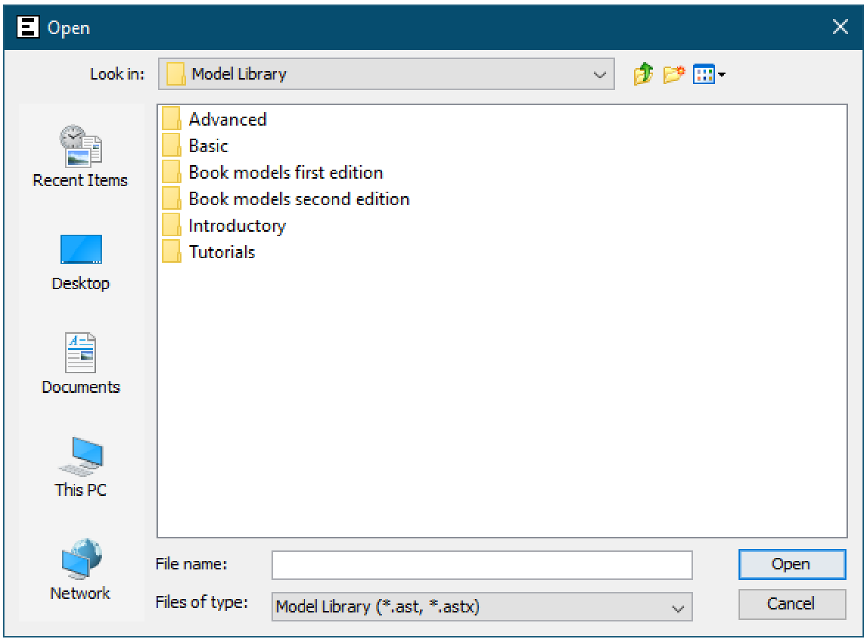


Figure 9 Dialog for choosing file

You will see the dialog for opening files whenever you click on **File** → **Open Model…**, **File** → **Open Example Model…** or **File** → **Import Model….** If you choose **Open Most Recent Model** or choose one of the last four opened files from the **File** menu, no dialog will be shown; the appropriate file will be loaded immediately.

agena.ai model files either have the extension*.ast* or *.cmpx*. The only difference is that *.ast* files are read-only; if you have modified an *.ast* file and want to save it, you will be prompted to save it as a new *.cmpx* file.

5. When the model has loaded, your screen will look like the one shown in Figure 10.

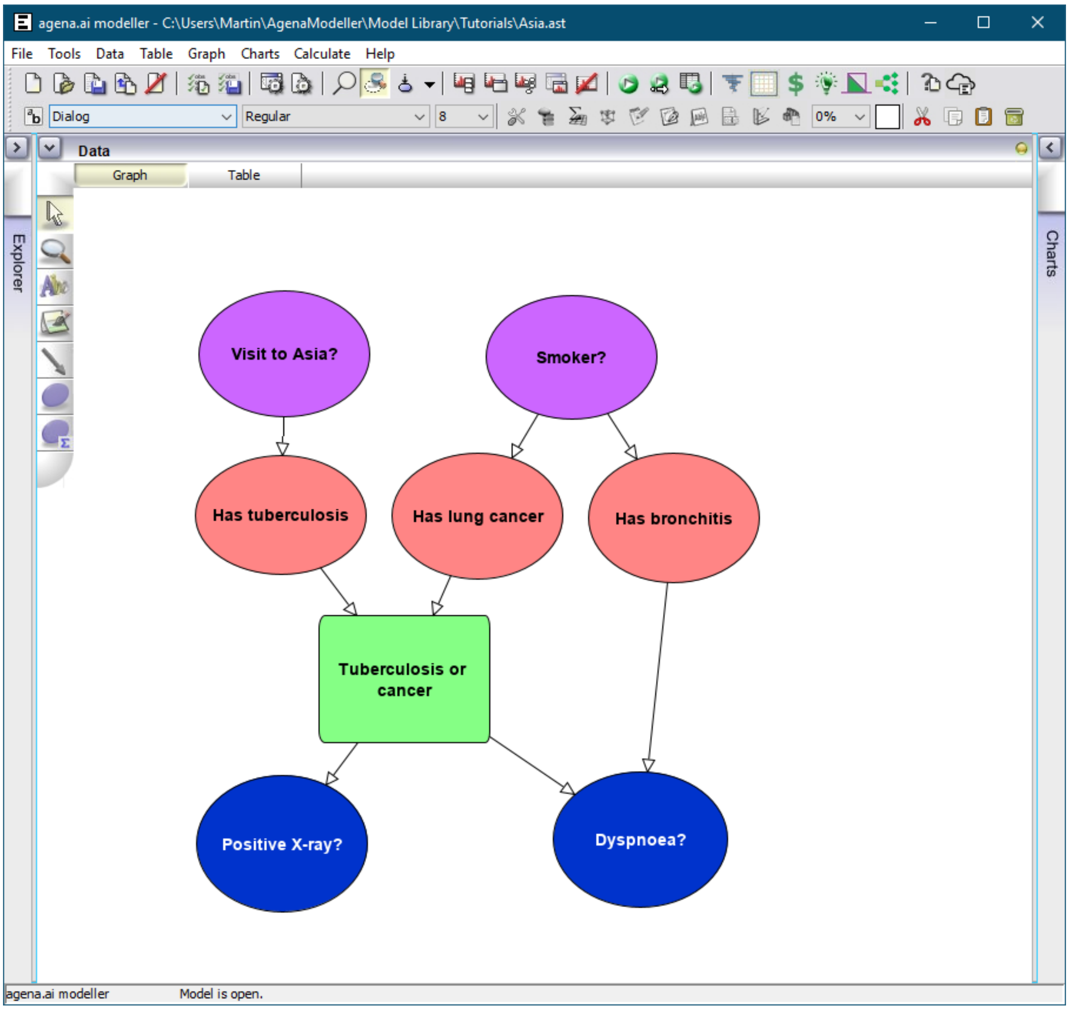


Figure 10 agena.ai modeller after opening *Asia.ast*

When a model is opened, only the graph view is shown by default. The graph represents a single Bayesian network (BN) and it consists of nodes and edges that correspond to the nodes and edges in a BN.

3.2 Using charts

3.2.1 Selecting and moving nodes

1. You can move nodes around easily with your mouse. Click on the *Dyspnoea* node in the bottom right corner of the graph view and, while holding down your mouse button, drag the node so that it is aligned horizontally with the *Tuberculosis or cancer* node and vertically with the *Has bronchitis* node.

2. Nodes can also be moved around in groups. Click on the *Tuberculosis or cancer* node and then, holding down the Ctrl key, click on the *Positive X-ray* node. Both are now selected. Click on either of the selected nodes and move them so that the *Tuberculosis or cancer* node is vertically in line with *Has lung cancer*.

3. Your Bayesian network graph should now look like the one shown in Figure 11.

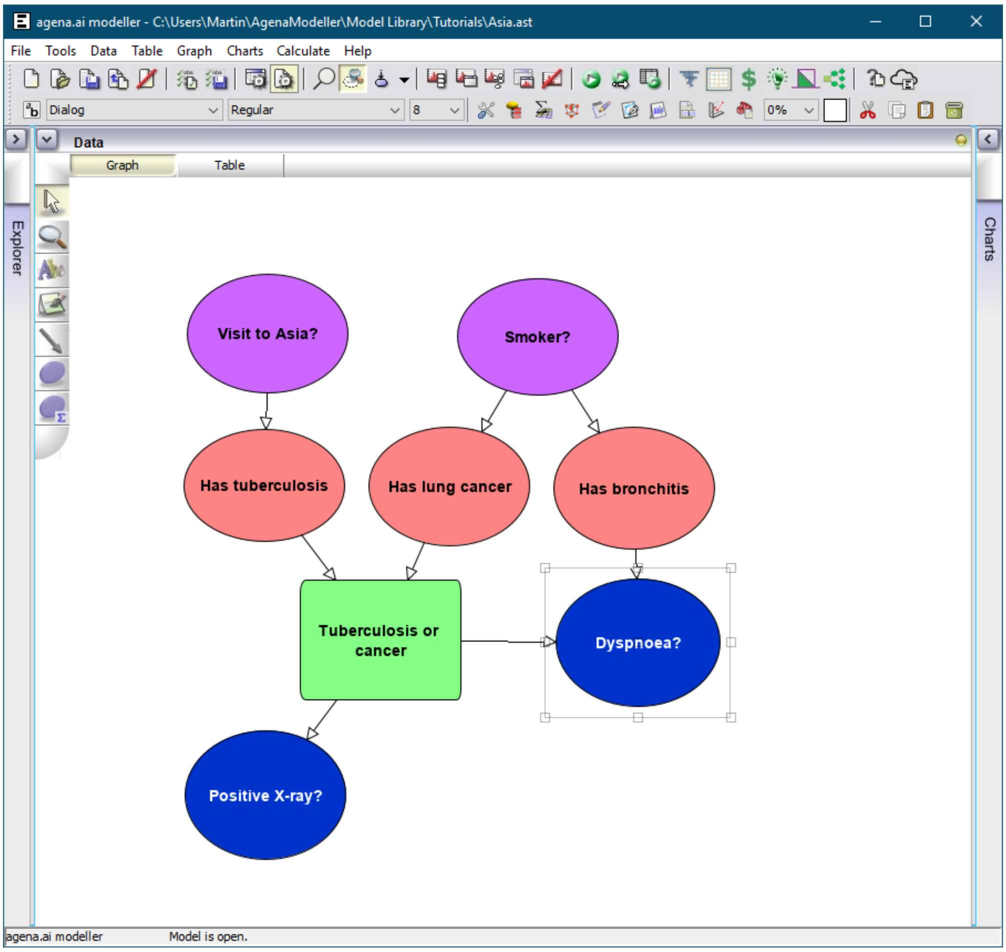


Figure 11 agena.ai modeller after moving nodes

> You can also select multiple nodes by dragging a rectangle around the nodes you require using your mouse.

4. Continue to experiment with multiple selection and moving nodes around until you are comfortable with manipulating the graph.

3.2.2 Zooming in and out

1. Click on the magnifying glass icon on the vertical toolbar to the left of the graph view:

2. Click on the *Tuberculosis or cancer* node. Notice how you have zoomed in on the graph view.

3. Click on the same node again. You zoom in another level.

4. Click twice on the same node with the **right** button of your mouse. You will zoom out to your original view.

5. Click on the mouse pointer icon on the vertical toolbar to return to the selection mode:

> ⚠ It is important that you return to selection mode after zooming; if you don't, the next time you click on the Graph view you will end up zooming in unexpectedly.

3.2.3 Displaying Bayesian networks on the graph view

1. Close the current model without saving it. To do this, click on **File** → **Close Model** and then click **OK** in the confirmation dialog that appears.

> ℹ You can also close the current model by pressing Ctrl + W.

2. Reopen *Asia.ast* by clicking on the **File** menu and selecting it in the list at the bottom.

3. Before proceeding, it is worth briefly describing the *Asia.ast* model. The aim of the model is to predict whether a patient has one of three medical conditions: tuberculosis, lung cancer or bronchitis.

4. The model captures the relationships between these diseases, their causes and their symptoms. It encodes these relationships structurally; an arrow (or edge) between one node and another suggests that one node has a causal influence on the other. It also captures these relationships mathematically; each node has associated with it a node probability table (or NPT). NPTs will be dealt with in detail later on in the tutorial.

5. Each node in the model is Boolean; that is, each node has only two states or possible values: **yes** or **no**.

6. Right-click on the *Has lung cancer* node and select **Display Chart** → **on Graph**. The Bayesian network graph for the node will be superimposed over it as shown in Figure 12.

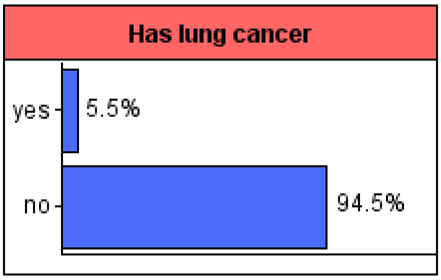


Figure 12 Chart for *Has lung cancer* node

> ℹ Double-clicking on a node is a very quick way of displaying its Chart.

7. The chart shows that, in the absence of any other information, there is a very low probability that the patient has lung cancer (i.e. the **yes** state has a very small bar and the **no** state has a very large bar).

8. In this case, we can see that there is a 0.055 (i.e., approximately 5.5%) chance that the patient has lung cancer and a 0.945 (i.e., approximately 94.5%) chance that the patient does not have lung cancer.

9. Use your mouse to select both the *Has tuberculosis* and the *Has bronchitis* nodes.

10. Right-click on either of the selected nodes and choose **Display Chart → on Graph**. Charts for both of these nodes will be displayed. Verify the probabilities of the patient having each medical condition by moving your mouse over the bars.

3.2.4 Displaying Charts on the Charts Panel

1. As well as displaying Charts directly on the graph, they can be displayed in a dedicated area of the screen called the Charts Panel.

2. Select **Chart → Close Chart** from the toolbar.

3. At the far right of the screen, click on the left-facing arrow that appears just above the text **Charts**:

4. The empty Charts Panel should appear as shown in Figure 13.

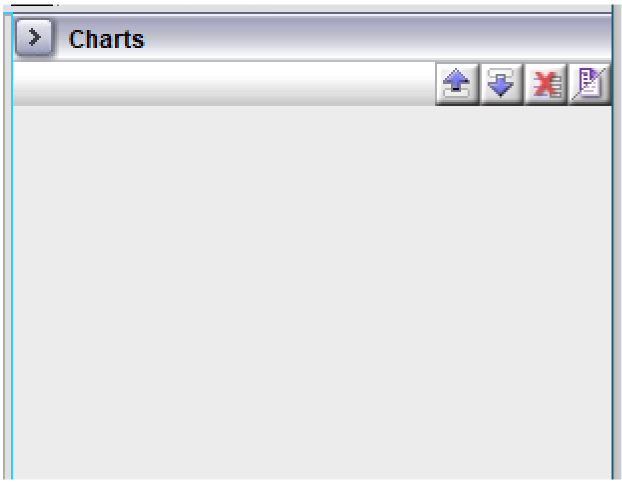


Figure 13 Empty Charts Panel

5. Select all three medical condition nodes (*Has tuberculosis*, *Has lung cancer* and *Has bronchitis*), right-click on any of the selected nodes and choose **Display Chart → on Chart Panel**. All three charts appear in the charts panel as shown in Figure 14.

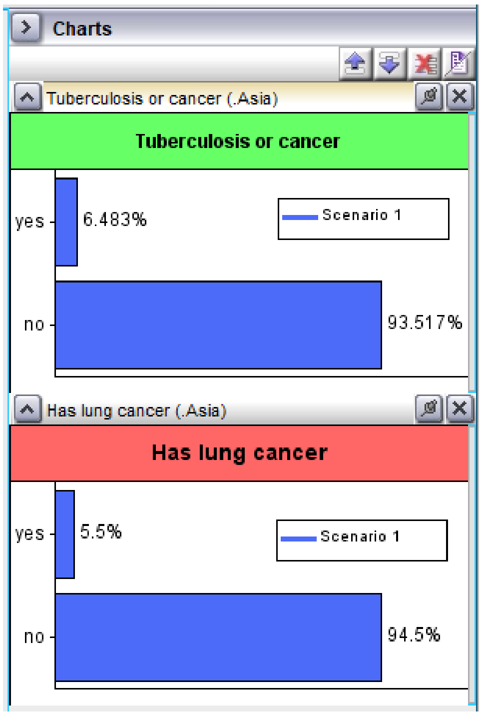


Figure 14 Charts Panel showing two charts

6. Charts can be moved around on the charts panel using the buttons at the top right. Click on the title bar of the *Has lung cancer* chart. You will see that it becomes highlighted.

7. Click twice on the downward arrow:

8. Notice that the chart has now moved to the bottom of the list.

9. Charts can also be detached from the chart Panel and displayed in their own windows. Click on the drawing pin icon next to the X in the top right corner of the *Has tuberculosis* graph:

10. The chart is now displayed in its own window as shown in Figure 15 below.

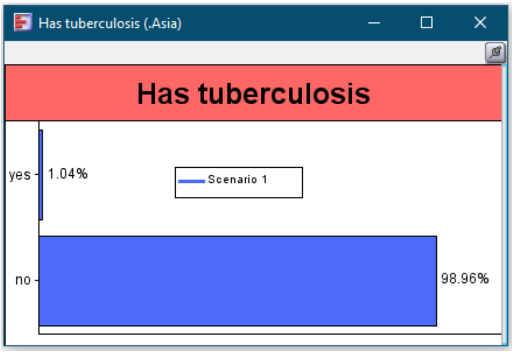


Figure 15 Chart for *Has tuberculosis* shown in own window

11. Click on the drawing pin icon again to dock the chart back on to the charts panel.

12. Now click the red X button on the Charts Panel toolbar:

13. The Charts Panel is empty once more.

3.2.5 Entering data via the graph

1. So far, you have looked at the probability outputs of the *Asia* model as they are when no other information has been entered. To exploit the power of the model, you need to begin entering data.

2. Minimise the Charts Panel by clicking on the right-facing arrow at the top of the panel:

> As well as simply expanding and retracting the Charts Panel using the arrow icons, you can change the width of it by clicking on the bar that separates it from the Graph and dragging it left or right as required.

3. Right-click on an empty area of the Graph and choose **Selection** → **All** from the pop-up menu.

4. Right-click on any of the selected nodes and select **Display Chart** → **on Graph** from the pop-up menu.

5. You should now see chart displayed on top of every node.

6. Ensure that the auto-calculate button on the main application toolbar is pressed down:

   This ensures that the model automatically recalculates each time you enter data.

> If you are using a model that takes a long time to calculate and want to enter a batch of data, you should consider turning off the auto-calculate feature. Otherwise, the model will re-calculate each time a piece of data is entered and this may take a long time.

7. Note the current probabilities of the patient having bronchitis, lung cancer or tuberculosis by inspecting the graphs of the three corresponding nodes.

8. Imagine that the patient we are currently examining has presented with dyspnoea (shortness of breath). We can enter that information into the model by right-clicking on the *Dyspnoea?* node and selecting **Enter Data** → **Case 1** → **yes**.

9. Notice that this data is recorded by a label on the *Dyspnoea?* node and that the probability of the state **yes** is now 1 (or 100%).

10. Look at the charts for the bronchitis, lung cancer and tuberculosis nodes and note how the probabilities have changed. Our belief that the patient has bronchitis has increased significantly.

11. Imagine now that we find out that the patient is not a smoker. Record this information by right-clicking on the *Smoker?* node and selecting **Enter Data** → **Case 1** → **no**. The probability of the patient having lung cancer or bronchitis has fallen slightly and the probability of the patient having tuberculosis has increased slightly but we still believe that bronchitis is the most likely of the three.

12. Next, we find out that the patient has been to Asia. Enter this data by right-clicking on the *Visit to Asia?* node and choosing **Enter Data** → **Case 1** → **yes**. The probability of the patient having tuberculosis has now increased.

13. Finally, imagine that we have carried out an x-ray and that the result is positive. Enter this information by right-clicking on *Positive X-ray* and choosing **Enter Data** → **Case 1** → **yes**. Notice how the belief in the patient having bronchitis has fallen while the probability of tuberculosis has risen.

3.2.6 Entering data via the table

1. Up until now, we have only interacted with the model via the graph. There is another view called the Table. This is useful for entering many data at once.

2. Reopen *Asia.ast* by clicking **File** and then selecting the model from the list at the bottom of the menu.

3. Click **OK** when the confirmation dialog appears.

4. When the model has opened, click on the **Table** tab just next to the **Graph** tab, as show in Figure 16.

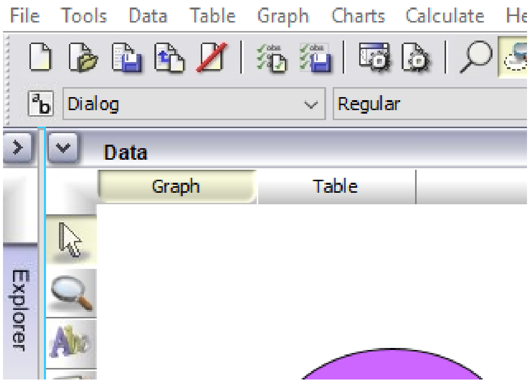
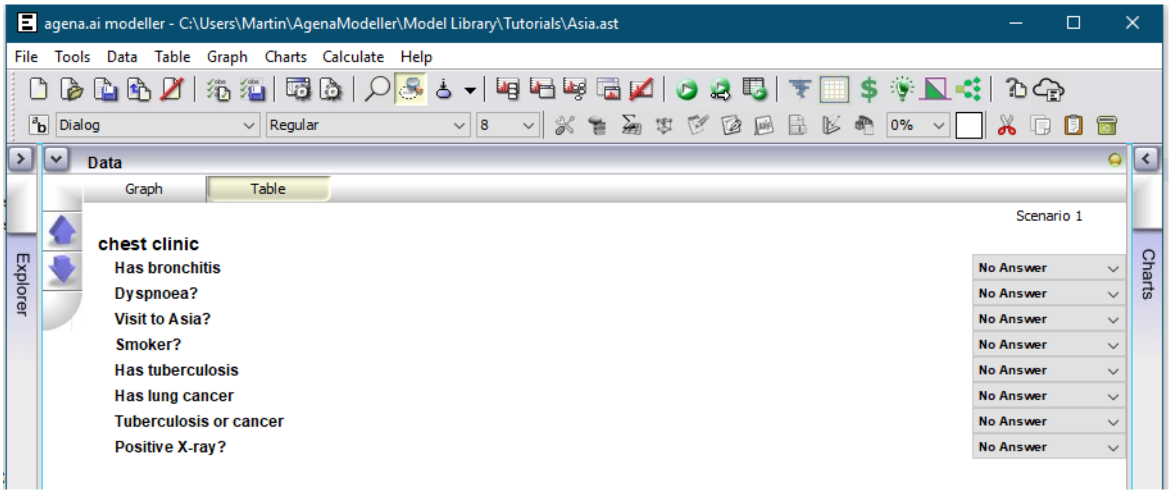


Figure 16 Table tab

5. The Table view now appears (see Figure 17). Note that it contains an entry corresponding to each node in the Graph.



Scenario 1

chest clinic

Has bronchitis: No Answer
Dyspnoea?: No Answer
Visit to Asia?: No Answer
Smoker?: No Answer
Has tuberculosis: No Answer
Has lung cancer: No Answer
Tuberculosis or cancer: No Answer
Positive X-ray?: No Answer

Figure 17 Table view

6. Expand the Charts Panel by clicking on the left-facing arrow:

7. Right-click on the *Has bronchitis* entry and choose **Display Chart**. You will see the chart appear on the Charts Panel.

8. In the drop-down box alongside the *Smoker?* entry, click on **yes**. Notice from the *Has bronchitis* chart that the probability of bronchitis has changed.

3.2.7 Entering batches of data

1. Switch off auto-calculation by toggling the auto-calculate button on the toolbar so that it looks like this:

2. Using the appropriate drop-down boxes, set *Visit to Asia?* to **no** and *Dyspnoea?* and *Positive X-ray* to **yes**. You will see that the *Has bronchitis* chart hasn't been updated.

3. Click on the calculation button on the toolbar to perform a calculation that includes these new data:

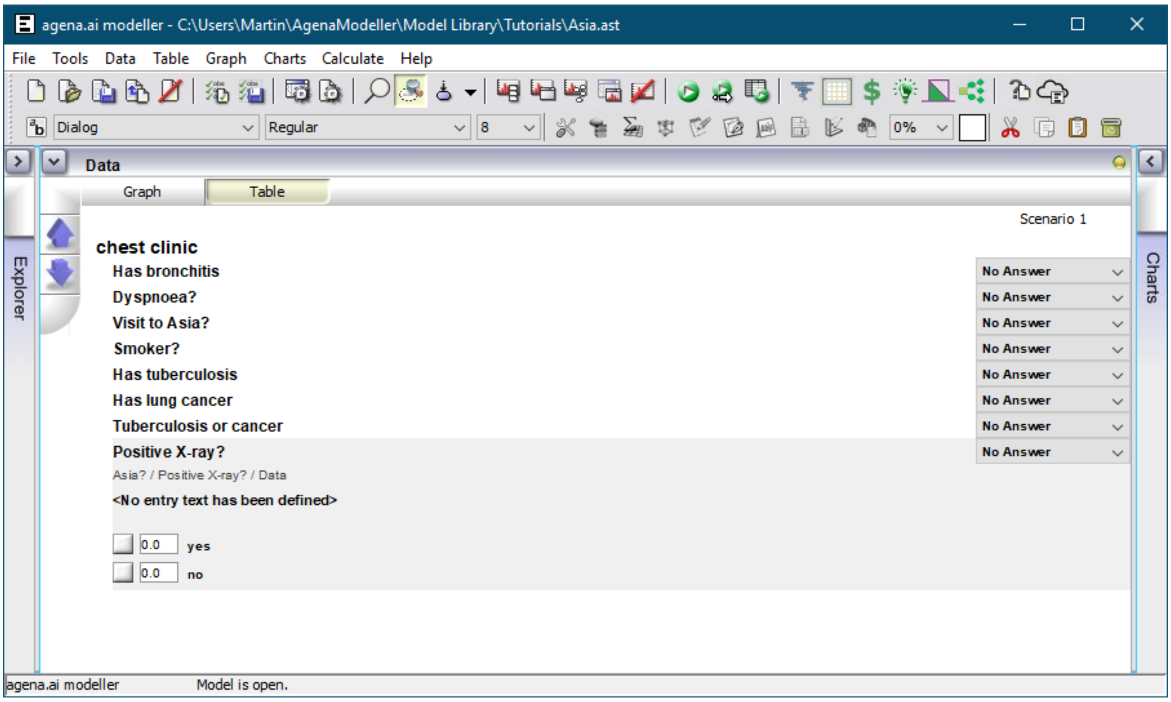
4. You will see that the probability of the patient having bronchitis now increases.

3.2.8 Entering soft evidence

1. So far, you have only entered what is called "hard" evidence. Hard evidence is where you specify that a node takes on an exact value (in this case, either **yes** or **no**). It is also possible, however, to enter "soft" evidence. Soft evidence is where you assign a percentage to two or more of a node's states. For example, an x-ray result may not, in your opinion, be 100% positive; there may be some ambiguity. In which case, you might want to record the fact that the x-ray is only 80% positive.

2. Click on **Tools** → **Clear Entered Data** → **All** from the menu bar and click **Yes** in the confirmation dialog that appears.

3. Double-click on the *Positive X-Ray* entry to expand it as shown in Figure 18 below.



Scenario 1

chest clinic

Has bronchitis	No Answer
Dyspnoea?	No Answer
Visit to Asia?	No Answer
Smoker?	No Answer
Has tuberculosis	No Answer
Has lung cancer	No Answer
Tuberculosis or cancer	No Answer
Positive X-ray?	No Answer

Asia? / Positive X-ray? / Data

<No entry text has been defined>

☐ 0.0 yes
☐ 0.0 no

Figure 18 Expanded Table entry

4. In the box next to **yes** enter the figure 0.8 and in the box next to **no** enter 0.2.

When entering soft evidence, you do not need to ensure that the numbers add up to 1. Agena.ai modeller will automatically normalise the numbers for you.

5. Click on the calculation button on the toolbar to run a calculation with this new data. Note the probability of the patient having bronchitis by inspecting the chart.

6. Now enter 0.9 in the **yes** box and 0.1 in the **no** box and run another calculation. See how the probability of bronchitis has increased slightly.

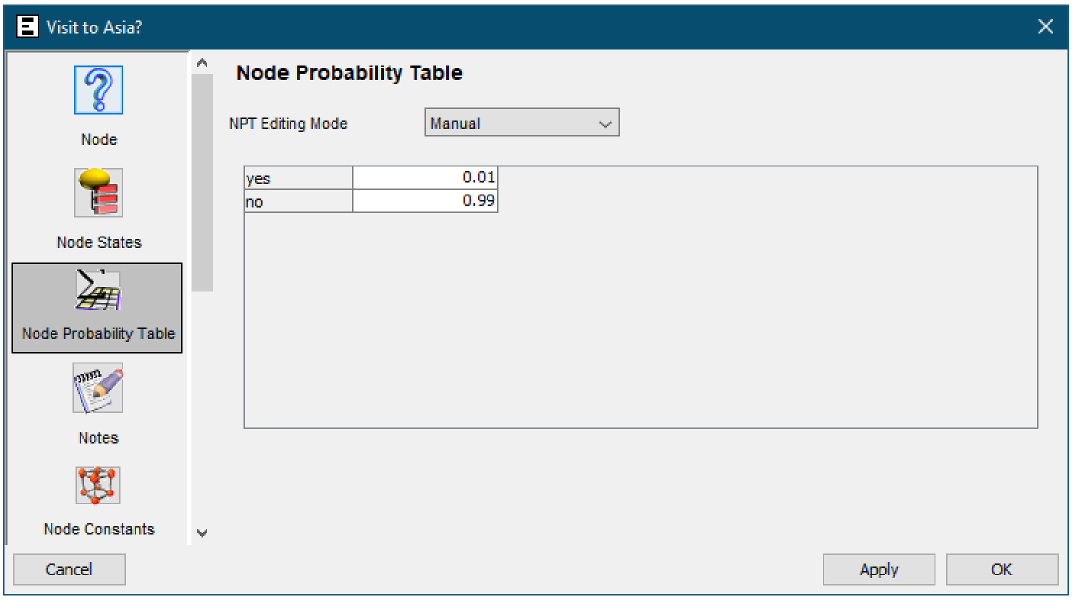
3.2.9 Viewing NPTs

1. Node Probability Tables (NPTs) contain probability information that underpins the structural relationships in a model.

2. For a node that has no parents (i.e., that has no edges coming into it), the NPT is simply a list of probabilities in which each probability corresponds to a state of the node.

3. Switch to the Graph view, right-click on the *Visit to Asia?* node and then select **Properties**.

4. Click on the **Node Probability Table** tab in the left panel. The NPT is shown in Figure 19 below.



Node Probability Table

NPT Editing Mode: Manual

yes	0.01
no	0.99

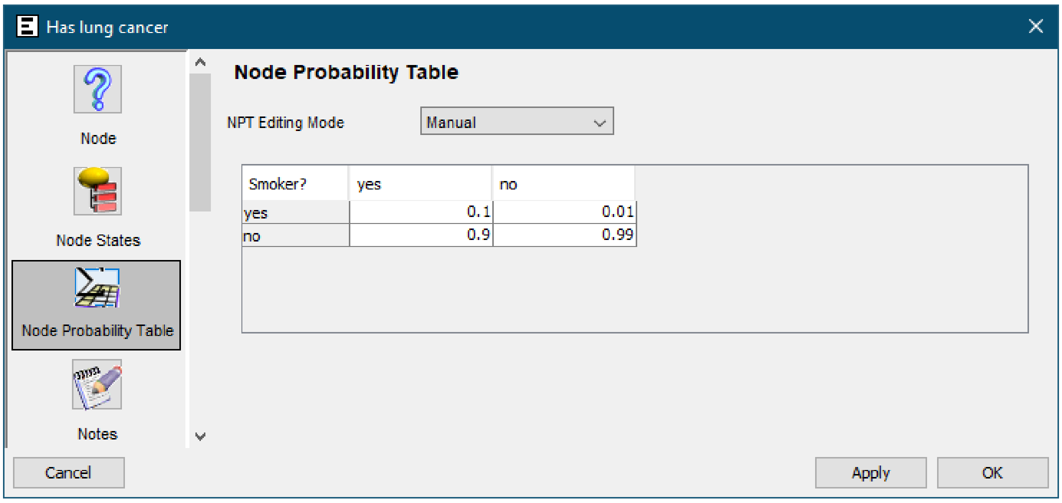
Figure 19 NPT for node with no parents

Alternatively, as explained in Chapter 2, you can use the toolbar to quickly access a node's NPT by selecting the node and clicking the NPT icon

5. The meaning of this NPT is as follows: our initial understanding (based on research, perhaps) is that there is a 0.01 (i.e., 1%) chance that a patient we see has been to Asia. Conversely, there is a 0.99 (i.e., 99%) chance that a patient we see has not been to Asia. These probabilities are called priors because they represent the situation before any other information is known.

6. Click on **OK** to close the dialog.

7. Right-click on the *Has lung cancer* node, select **Properties** and click on the **Node Probability** tab on the left of the panel. The NPT you will see is shown in Figure 20.



Node Probability Table

NPT Editing Mode: Manual

Smoker?	yes	no
yes	0.1	0.01
no	0.9	0.99

Figure 20 NPT for node with one parent

8. This NPT is more complicated and contains so-called “conditional” probabilities; that is, probabilities that are different depending on the state that the parent node is in. The meaning of this NPT is as follows: if we know that the patient is a smoker, then there is a 0.1 (i.e.10%) chance that they will have lung cancer and a 0.9 (i.e. 90%) chance that they will not; if we know that the patient is not a smoker, then there is a 0.01 (i.e. 1%) chance that they will have lung cancer and a 0.99 (i.e. 99%) chance that they will not. These probabilities may be based on historical admissions data at a clinic. Or, alternatively, they could represent the subjective opinion of one or more experts in the field.

9. Agena.ai modeller provides very powerful support for generating NPTs automatically using mathematical expressions, Boolean expressions, statistical distributions and table partitioning. These techniques are beyond the scope of this basic tutorial, however, and are dealt with in the more advanced tutorials and in the agena.ai modeller reference manual.

10. Click on **OK** to close the dialog and then close the model.

3.2.10 Accessing help

1. At any stage, you can view the full Agena.ai modeller user manual by clicking on **Help** → **Open User Manual**.

2. You can also access the user manual by pressing the **F1** key.

> ⚠ To view the user manual, you need to have Adobe Acrobat Reader, or similar, installed on your computer.

3.2.11 Using the agena.ai example models

1. Agena.ai modeller comes with many example models. By exploring these, you can get a better understanding of the different ways in which agena.ai modeller can be used.

2. To open a library model in agena.ai modeller, click on **File** → **Open Model Library…** and click down through the directories until you reach an *.ast* file that you are interested in viewing.

3.3 Creating a new model

1. Click on **File** → **Create New Model** and click on **OK** in the confirmation dialog that appears.

2. A new Bayesian network model is created and all three agena.ai modeller panels are displayed: the Explorer, the Graph and the Charts panel, as shown in Figure 21.

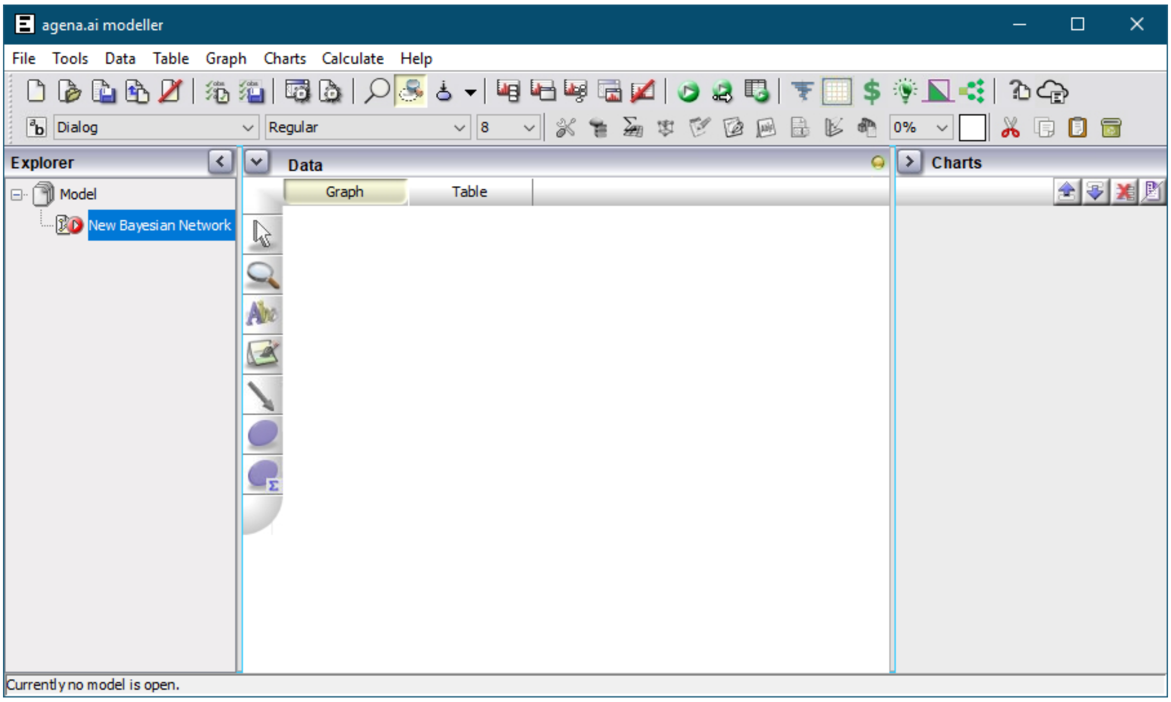
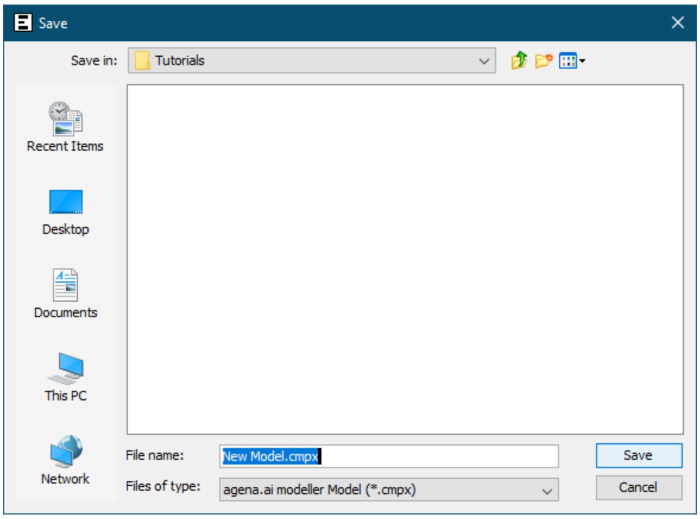


Figure 21 Agena.ai modeller panels shown after new model created

3.3.1 Saving models

1. To save your new model, choose **File** → **Save Model** from the menu bar.

2. A file dialog like the one shown in Figure 22 will appear.


Save in: Tutorials

File name: New Model.cmpx

Files of type: agena.ai modeller Model (*.cmpx)

Figure 22 Dialog for specifying where to save a model

3. Navigate to a writeable directory where you want to save the model and then type the following name in the **File name** field: **Asia From Scratch**.

4. Click **Save**.

5. Click **File** → **Close Model** and then click **OK** in the confirmation dialog that appears.

6. Click **File** → **Open Model…** and verify that your model (*Asia From Scratch.cmp*) is available for selection.

7. Click on the model name and then click **Open**. The model is now displayed in agena.ai modeller.

8. Retract the Explorer view by clicking on the left-facing arrow:

   You can do this whenever you need more space for viewing and using the Graph view.

   The Explorer is only useful when you are dealing with more complex models that have more than one Bayesian network in them. This is covered in the *Advanced Modelling with* Agena.ai modeller tutorial.

9. Expand the Explorer again by clicking on the right-facing arrow

> As with the Charts Panel, you can change the width of the Explorer when it is expanded by clicking on the bar that separates it from the Graph and dragging it.

3.3.2 Adding nodes to the graph

1. Click on the oval icon on the vertical toolbar next to the Graph:

2. Click in the top-left area of the Graph view. You will see that a new node has been added as shown in Figure 23.

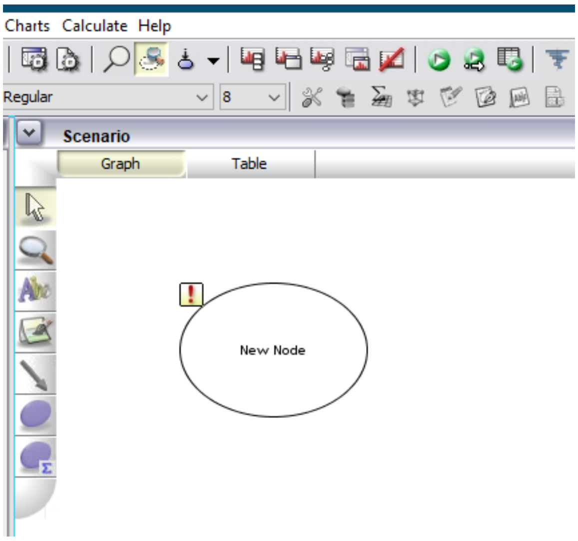
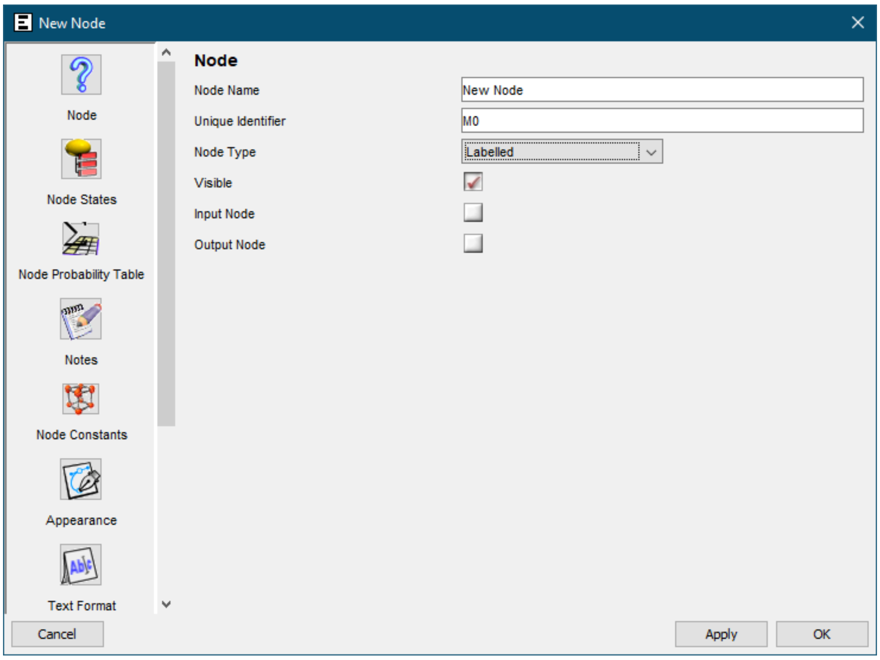


Figure 23 A new node added to the Graph

3.3.3 Changing the name, type and states of a node

1. Right-click on the node you just added and select **Properties** from the menu.
2. In the **Node Name** field, type **Visit to Asia?**
3. In the drop-down box labelled **Node Type** select **Labelled**. The properties dialog should now look like the one shown in Figure 24.



Node

Node Name: New Node

Unique Identifier: M0

Node Type: Labelled

Visible: ☑

Input Node: ☐

Output Node: ☐

Figure 24 Node properties after changing type to Labelled

4. Click on **Apply**.

5. Click on **Node States** in the left panel of the properties dialog. Notice that, by default, the two states of a Labelled node are named: **False** and **True**.

6. In the **States** text box, select all the states by dragging your mouse across them and press the **Backspace** or **Delete** key.

7. Type **yes** on the top line of the empty text box and **no** on the next line.

8. Click on **OK**.

9. The Graph should now look like the one in Figure 25.

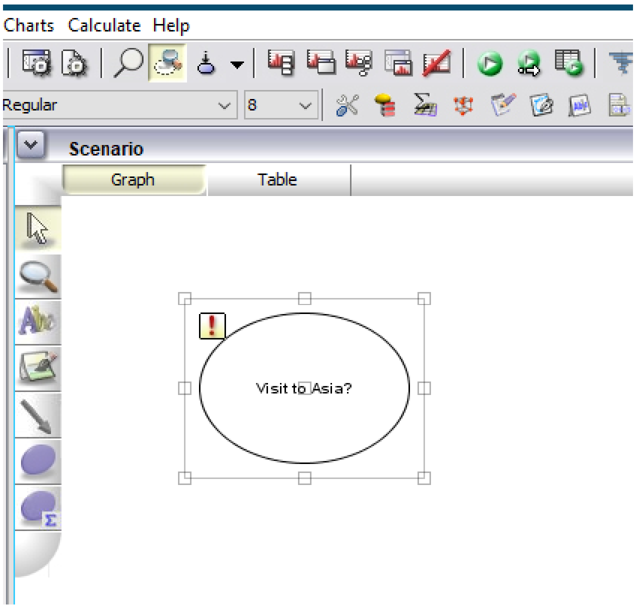


Figure 25 Graph after renaming node

When a new node is created the red exclamation mark icon, !, highlights that the node has not been edited nor been involved in a successful calculation. If it has been edited or an expression declared and then calculated the icon should disappear. Adding and deleting links between nodes, changing expressions and NPTs will result in the return of the icon, again to highlight to you that things have changed.

3.3.4 Using copy and paste

1. Right-click on the *Visit to Asia?* node and select **Copy** from the pop-up menu.

2. Right-click on an empty area of the Graph to the right of the *Visit to Asia?* node and select **Paste**.

3. A copy of the *Visit to Asia?* node (called *Visit to Asia?_1*) is now added to the Graph.

The fastest way of copying and pasting is to use the keyboard shortcuts Ctrl + C and Ctrl + V.

4. Right-click on this new node and select **Properties**. Observe that the node is of Labelled type (rather than the default type for new nodes, Continuous Interval).

5. Click on the **States** tab in the left panel of the dialog and observe that the states are **yes** and **no**.

6. Click on the **Node Details** tab and enter the following name into the **Node Name** field: **Smoker?**

7. Click **OK** and notice that the name of the node has changed.

8. Referring back to Figure 10, use the copy and paste facility to create and lay out the remaining nodes in the Asia model: *Has tuberculosis*, *Has lung cancer*, *Has bronchitis*, *Tuberculosis or cancer*, *Positive X-ray?* and *Dyspnoea?*. Ignore edges for now; you will add these later.

You can copy and paste multiple nodes at one time. Simply select all the nodes you require (by holding down Ctrl while clicking on them) and then copy and paste as for a single node.

9. If you have time, modify the appearance of each node to match Figure 10.

10. Click **File** → **Save Model** or click on the save icon on the toolbar:

    Because you have already saved the model, you do not need to specify a new name; the existing file will be overwritten.

If you want to save a model in a new file at any stage, select **File** → **Save Model As…** from the menu bar.

3.3.5 Adding edges

1. Click on the edge icon on the vertical Graph toolbar:

2. Click on the *Visit to Asia?* node.

3. Move your mouse downwards to the *Has tuberculosis* node. Notice that a dotted line appears to indicate where the edge is going to be created.

4. Click on the *Has Tuberculosis* node to create the edge.

5. The two nodes should now look like Figure 26.

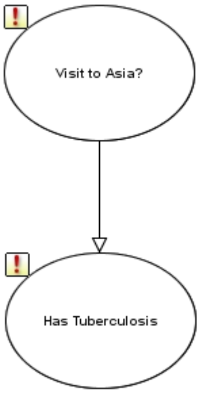


Figure 26 An edge added between two nodes

6. Using Figure 10 in for reference, add the remaining edges to the Graph.

> Edge direction is very important. To create an edge that starts at node A and finishes at node B, you need to click on node A first and then on node B.

3.3.6 Deleting nodes and edges

1. Click on the node icon on the toolbar:

2. Click In an empty area of the Graph view.

3. Right-click on the node that you have just created and select **Delete** from the pop-up menu. Observe that the node disappears.

> The easiest way of deleting objects from the Graph is to press the Delete key on your keyboard.

4. Click on the node icon again and click in an empty area of the Graph view.

5. Repeat this process to create a second node.

6. Click on the edge icon:

7. Draw an edge between the two new nodes.

8. Right-click on the edge and select **Delete**. Notice that the edge disappears.

9. Redraw the edge between the two nodes.

10. Right-click on one of the nodes and select **Delete**. Notice that the edge disappears as well as the node. The edge is deleted because edges can only exist on the Graph if they connect two nodes.

11. Delete the other node so that the Graph shows only the Asia model.

3.3.7 Hiding nodes

1. Sometimes when building models, it is desirable to hide certain nodes to reduce the visual complexity of the Graph. In large models there are often intermediate calculation nodes that don't need to be shown to the user of the model. These nodes can be hidden.

2. Right-click on the *Tuberculosis or cancer* node and select **Properties** from the pop-up menu.

3. Uncheck the box marked **Visible**.

4. Click **OK**.

5. You can see that the *Tuberculosis or cancer* node is now shown with a dotted border. The reason for this is that Agena.ai modeller, by default, will actually show nodes that are marked as hidden. This is typically what you want to do when building models; only when using models do you want truly to hide hidden nodes.

6. To truly hide nodes that are marked as hidden, toggle **Graph → Show Hidden Nodes** on the menu bar. You can also do this by clicking on the following toolbar button:

7. The *Tuberculosis or cancer* node now disappears. Notice that indirect links between nodes are now shown as dotted lines to replace the direct links to and from the hidden node (see Figure 27). Be aware that the underlying structure of the model is the same; only the view has changed.

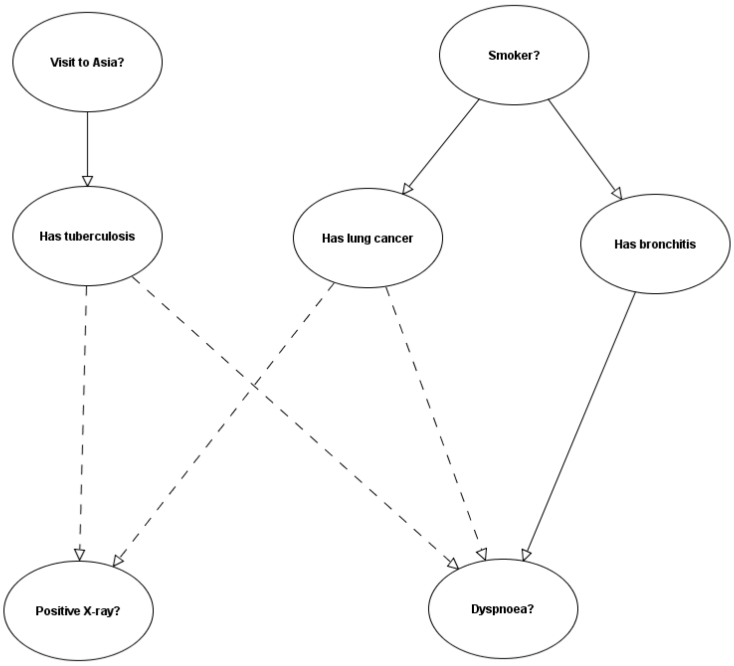


Figure 27 Indirect links shown after a node is hidden

8. Select **Graph → Show Hidden Nodes** from the menu bar (or use the toolbar button) to display the *Tuberculosis or cancer* node again.

9. Right-click on *Tuberculosis or cancer*, select **Properties**, tick the **Visible** box and click **OK**. The node is now restored.

3.3.8 Adding labels

1. Click the text note icon on the Graph vertical toolbar:

2. Click in the space below the *Has bronchitis* node on the Graph. A label containing default text will appear.

3. Right-click on this label and select **Properties** from the pop-up menu.

4. In the box marked **Text**, type the following: **This model is used for diagnosing lung cancer, tuberculosis and bronchitis**.

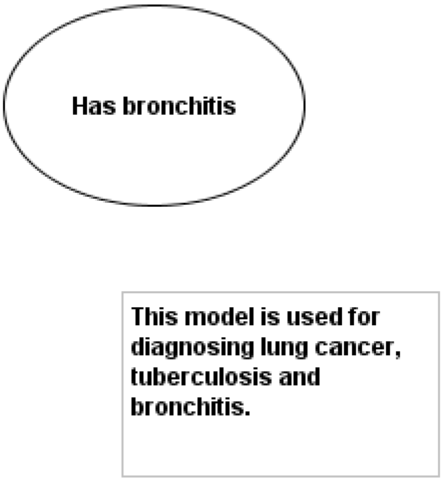
5. In the **Font Size** box, enter the value **12.**

6. In the **Horizontal Text Alignment** drop-down box, select **Centre**.

7. In the **Vertical Text Alignment** drop-down box, select **Centre**.

8. Click **OK**.

9. You will see that the label has changed. However, you can't see all the text now, so drag one of the small grey boxes outwards to enlarge the label. The label should now look like this:



This model is used for diagnosing lung cancer, tuberculosis and bronchitis.

Figure 28 A label added to the Graph

10. Move the label around the Graph and observe that labels behave, in many respects, just like nodes.

11. If you have time, experiment with the other visual properties of the label to see how much control you have over its appearance.

Labels ignore any line breaks that you might enter in the text. To achieve the effect of multiple paragraphs, add a label for each paragraph and use the movement, grouping and alignment functions to space the text as you require.

12. Right-click on the label and, from the pop-up menu, select **Delete**.

3.3.9 Editing NPTs

1. Right-click on the *Visit to Asia?* node and select **Properties**.
2. Click on the **Node Probability Table** tab in the left panel.
3. In the text box to the right of **yes**, enter the value **0.01**.
4. Either click in the box to the right of **no** or press the **Tab** key.
5. Enter the value **0.99** and then press **Return**.
6. Click **OK**.
7. Right-click on the *Tuberculosis or cancer* node and select **Properties**.
8. Click on the **Node Probability Table** tab in the left panel.
9. In the top text box in the first column enter the value **1** and in the bottom text box in the first column, enter the value **0**.
10. Drag your mouse over the two cells in the first column (i.e. the cells that now contain **1.0** and **0.0**).

11. When they are selected, right-click and choose **Copy** from the pop-up menu.
12. Select the two cells in the second column, right-click and choose **Paste**. Observe that the two values from the first column have been entered.
13. Select the two cells in the third column, right-click and choose **Paste** again.
14. In the fourth column, enter **0** in the top text box and enter **1** in the bottom text box.
15. The NPT is now complete. Click **OK**.
16. Complete the NPTs for the remaining nodes using the data in Table 3 below.

> When you are dealing with a large NPT, you can save time by entering the table in Microsoft Excel first, copying the data, selecting the corresponding rows and columns in the NPT and then pasting.

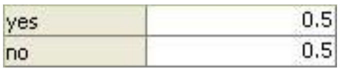
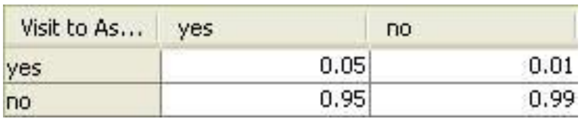
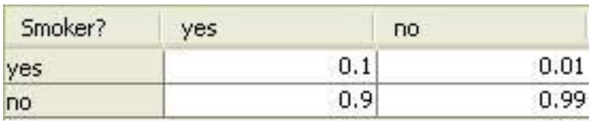
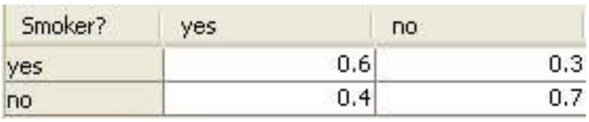
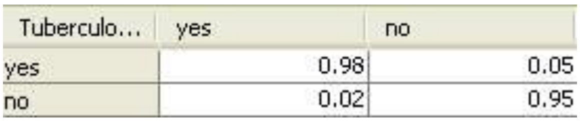
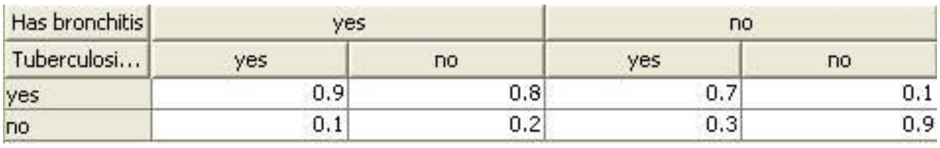
<table>
<tr><td><i>Smoker?</i></td><td>
<table><tr><td>yes</td><td>0.5</td></tr><tr><td>no</td><td>0.5</td></tr></table></td></tr>
<tr><td><i>Has tuberculosis</i></td><td>
<table><tr><th>Visit to As...</th><th>yes</th><th>no</th></tr><tr><td>yes</td><td>0.05</td><td>0.01</td></tr><tr><td>no</td><td>0.95</td><td>0.99</td></tr></table></td></tr>
<tr><td><i>Has lung cancer</i></td><td>
<table><tr><th>Smoker?</th><th>yes</th><th>no</th></tr><tr><td>yes</td><td>0.1</td><td>0.01</td></tr><tr><td>no</td><td>0.9</td><td>0.99</td></tr></table></td></tr>
<tr><td><i>Has bronchitis?</i></td><td>
<table><tr><th>Smoker?</th><th>yes</th><th>no</th></tr><tr><td>yes</td><td>0.6</td><td>0.3</td></tr><tr><td>no</td><td>0.4</td><td>0.7</td></tr></table></td></tr>
<tr><td><i>Positive X-ray?</i></td><td>
<table><tr><th>Tuberculo...</th><th>yes</th><th>no</th></tr><tr><td>yes</td><td>0.98</td><td>0.05</td></tr><tr><td>no</td><td>0.02</td><td>0.95</td></tr></table></td></tr>
<tr><td><i>Dyspnoea?</i></td><td>
<table><tr><th>Has bronchitis</th><th colspan="2">yes</th><th colspan="2">no</th></tr><tr><th>Tuberculosi...</th><th>yes</th><th>no</th><th>yes</th><th>no</th></tr><tr><td>yes</td><td>0.9</td><td>0.8</td><td>0.7</td><td>0.1</td></tr><tr><td>no</td><td>0.1</td><td>0.2</td><td>0.3</td><td>0.9</td></tr></table></td></tr>
</table>

Table 3 Asia model NPT probabilities

17. When you have finished check that the model behaves the same as the original.

3.3.10 Editing table entries

1. Whenever you create a node on the Graph, a corresponding entry is created in the Table. The default names and descriptions given to these entries are not very informative; you can change them so that they present a more informative interface for users who want to enter data via the Table.

2. Click on the Table tab at the top of the screen.

3. Click at the end of the **Visit to Asia?** entry so that the cursor appears after the question mark.

4. Delete the text using the Backspace key and type the following: **Has the patient visited Asia?**

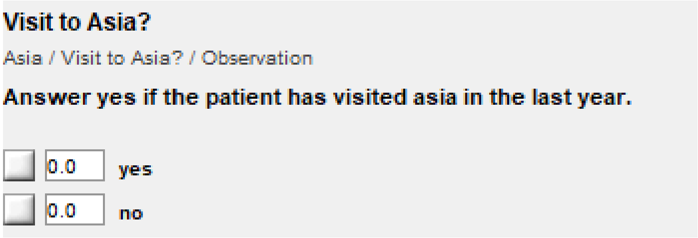
5. Now double-click on the entry to expand it.

6. Click at the end of the text that says **New Node**. This is the long description of the entry where you can put extra explanatory information to help users when they enter data.

7. Delete the text using the Backspace key and type the following: **Answer yes if the patient has visited Asia in the last year. For the purposes of this model, Asia does not include Turkey.**

> You can also edit a Table entry by right-clicking on it and selecting **Properties**.

8. The entry should now look like this:



Visit to Asia?

Asia / Visit to Asia? / Observation

Answer yes if the patient has visited asia in the last year.

☐ 0.0 **yes**
☐ 0.0 **no**

Figure 29 A Table entry after editing

3.3.11 Creating, editing and deleting table headings

1. Headings can be used to organise the entries in the Table. By default, all entries are created under a single heading that has the same name as the Bayesian network (*New Bayesian Network* in this case).

2. Double-click on the **Has the patient visited Asia?** entry to retract it.

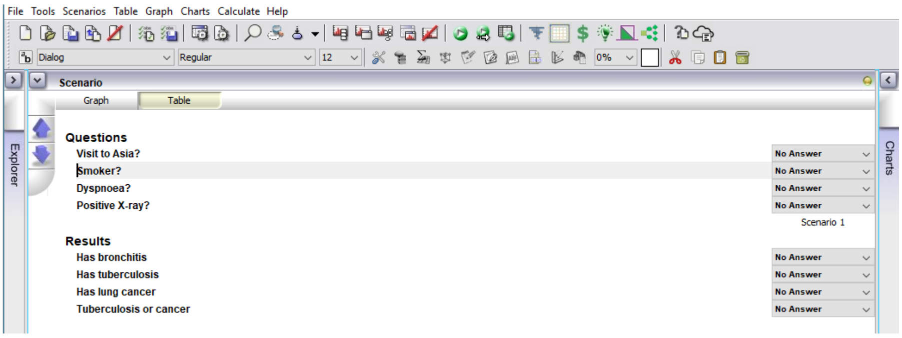
3. Right-click anywhere in the Table and select **Add Heading** from the pop-up menu. Notice that a new heading with the default name **New Heading** has been added at the bottom of the Table.

4. Click at the end of the **New Heading** text so that the cursor appears and delete the text using the **Backspace** key.

5. Type in the following: **Questions**.
6. Create another heading, right-click on it and select **Delete** from the pop-up menu.
7. Rename the **Bayesian Network** heading **Results**.

3.3.12 Moving table entries

1. You can now start to reorganise the Table under the two headings.
2. Click on the **Visit to Asia?** entry and then click on the downward arrow on the Table vertical toolbar:
3. Keep on clicking on the down arrow until the entry appears under the **Questions** heading. Do the same for: **Positive X-ray?**, **Dyspnoea?** and **Smoker?**
4. The Table should now look like this:



Questions
- Visit to Asia? — No Answer
- Smoker? — No Answer
- Dyspnoea? — No Answer
- Positive X-ray? — No Answer

Scenario 1

Results
- Has bronchitis — No Answer
- Has tuberculosis — No Answer
- Has lung cancer — No Answer
- Tuberculosis or cancer — No Answer

Figure 30 A Table entry after editing

5. The entries in your Table might be in a different order depending on which entries you moved first. You can change their order by using the upward arrow on the vertical toolbar

in conjunction with the downward arrow.

3.3.13 Hiding, deleting and adding table entries

1. Often, there are many nodes in a model that don’t really need corresponding Table entries, since users are never going to enter data for these nodes. These entries can either be temporarily hidden or permanently deleted (if you are sure, you are not going to need them).
2. Right-click on the **Tuberculosis or cancer** entry and select **Properties** in the pop-up menu.
3. Uncheck the **Visible** box and click **OK**.
4. You will see that this entry has now been greyed out. This is because, by default, nodes that are marked as hidden are still shown.

5. In order to hide the hidden node, right-click anywhere on the Table. Note that the **Show Hidden Entries** menu item has a tick next to it. Click it, thereby removing the tick.

6. The **Tuberculosis or cancer** entry has now disappeared.

7. Right-click anywhere on the Table and select **Show Hidden Entries** to show the entry again.

8. Right-click on **Tuberculosis or cancer** and select **Delete**. The entry has now been permanently removed from the Table.

9. You can add a new entry to replace it, however. Right-click on the last entry under the **Results** heading and select **Add Entry**. A new entry is added to the Table.

10. Click on the link icon just before the new entry's name:

11. In the properties dialog that appears, fill out the fields so that the dialog looks like that shown in Figure 31 below.

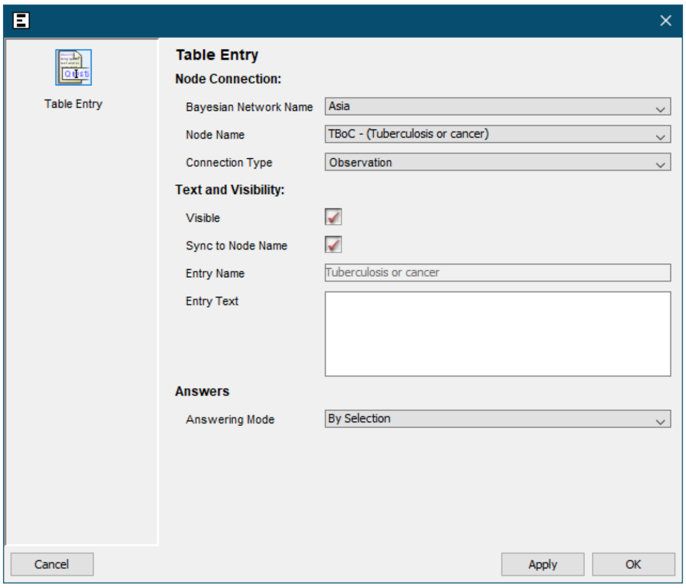


Table Entry

Node Connection:

Bayesian Network Name: Asia

Node Name: TBoC - (Tuberculosis or cancer)

Connection Type: Observation

Text and Visibility:

Visible ☑

Sync to Node Name ☑

Entry Name: Tuberculosis or cancer

Entry Text:

Answers

Answering Mode: By Selection

Figure 31 Properties for a new Table entry

12. Click **OK**. Notice that an entry identical to the previously deleted **Tuberculosis or cancer** entry has now been added.

3.4 Working with horizontal and vertical charts

Agena.ai modeller has some very powerful graph layout capabilities. Area histogram, bar chart layouts are available as is the ability to set the coordinate value ranges, including the use of percentile limits on the ranges plotted. A major feature is the fact that charts can be set so that their probability distributions displayed vertically (as shown in Figure 32) irrespective of the number of states or whether the node was discrete or numeric.

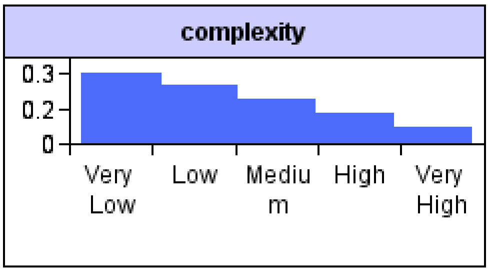
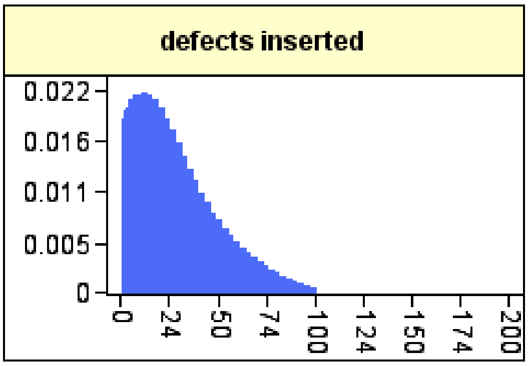
	
Discrete node chart	Numeric Node chart

Figure 32 Vertical charts

Figure 33 shows the new option in the Chart Defaults dialog, with the option to display horizontally circled. By default, Boolean and labelled nodes are displayed horizontally.

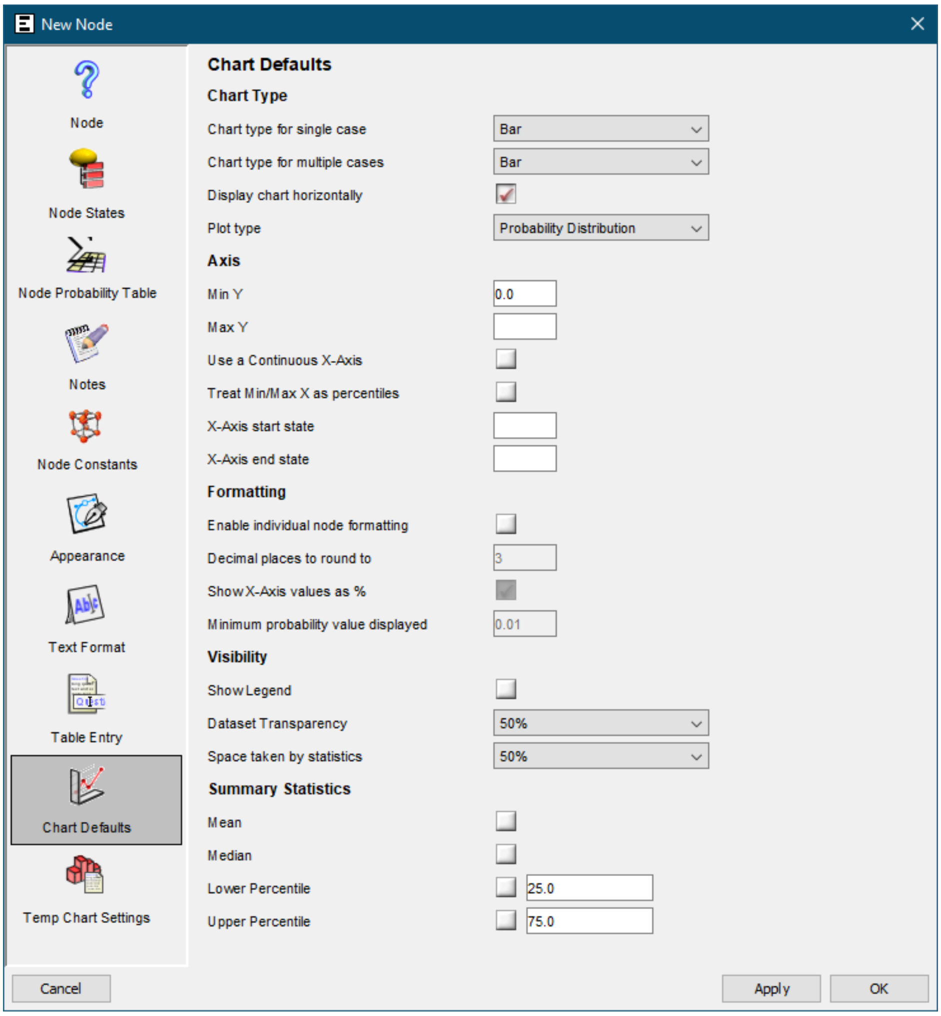


Chart Defaults

Chart Type

Chart type for single case: Bar

Chart type for multiple cases: Bar

Display chart horizontally ☑

Plot type: Probability Distribution

Axis

Min Y: 0.0

Max Y:

Use a Continuous X-Axis ☐

Treat Min/Max X as percentiles ☐

X-Axis start state:

X-Axis end state:

Formatting

Enable individual node formatting ☐

Decimal places to round to: 3

Show X-Axis values as % ☑

Minimum probability value displayed: 0.01

Visibility

Show Legend ☐

Dataset Transparency: 50%

Space taken by statistics: 50%

Summary Statistics

Mean ☐

Median ☐

Lower Percentile ☐ 25.0

Upper Percentile ☐ 75.0

Figure 33 Option for horizontal charts

By default, all new discrete (as opposed to numeric) variable nodes now have probabilities displayed horizontally as shown in Figure 34

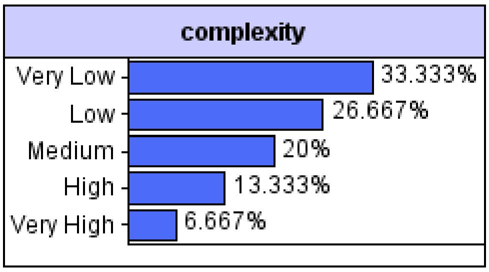
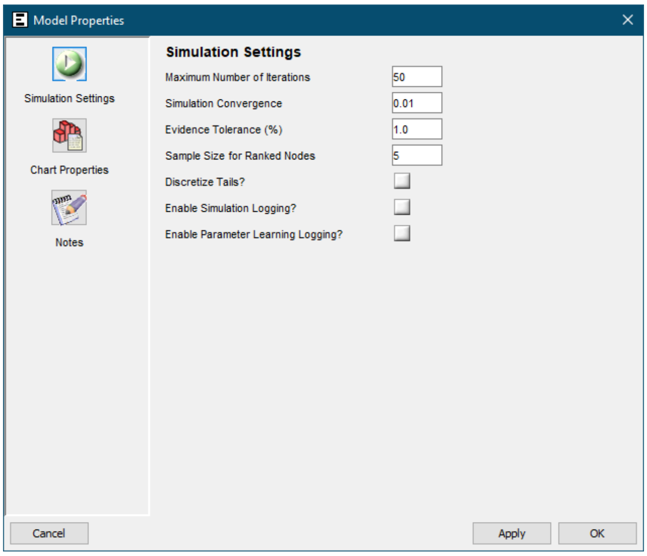


Figure 34 Horizontal chart

For horizontal charts the probability value for each state is written (as a percentage by default) at the right hand end of the bar.

There is a dialog that gives you full control over the way the probability text is displayed for horizontal graphs. To access this dialog simply select the Model Properties button in the main tool bar:

This will display the dialog shown in Figure 35. With this dialog you can set the number of decimal places to be shown as well as the minimum probability value displayed (states whose value is below this minimum will not have any bar displayed).



Simulation Settings

Maximum Number of Iterations: 50

Simulation Convergence: 0.01

Evidence Tolerance (%): 1.0

Sample Size for Ranked Nodes: 5

Discretize Tails? ☐

Enable Simulation Logging? ☐

Enable Parameter Learning Logging? ☐

Figure 35 Dialog for setting properties of horizontal charts

3.5 Exporting charts to CSV

You can export chart values from your models to CSV (Comma Separated Value) files.

You export by right clicking on any node inside the graph, selecting **Export Chart Data as CSV file** and then the case from which the data should be exported, as shown below.

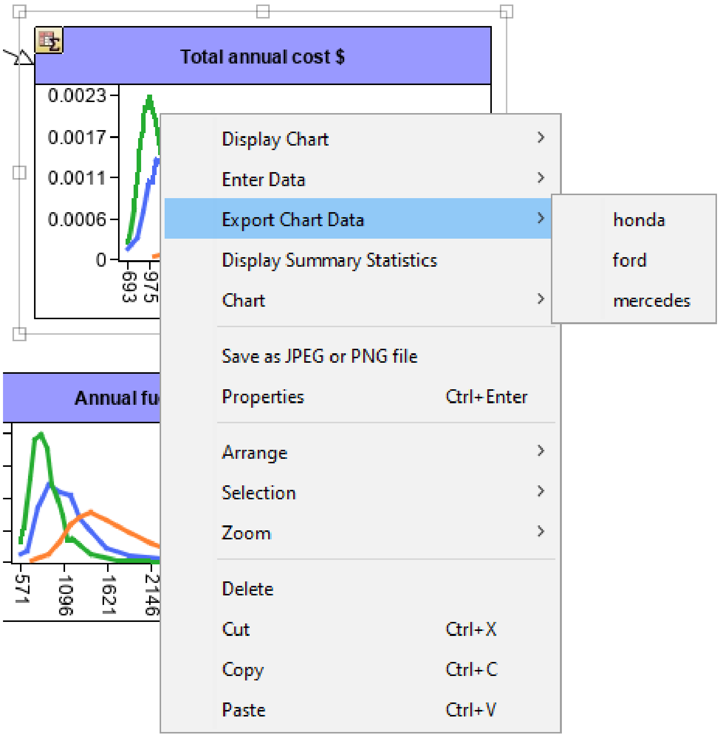


Figure 36 Exporting chart values to CSV

3.6 Next steps

This tutorial has given you a basic introduction to the range of features that agena.ai modeller has to offer. However, it has only skimmed the surface. To benefit fully from agena.ai modelling power and flexibility you should work through the more advanced tutorial, *Modelling with Agena.ai modeller*, and explore the range of well-documented example models that are included with the software.

4. Basic modelling with agena.ai modeller

In the first half of the tutorial, you will explore NPT expressions and simulation in depth by building a powerful hypothesis-testing model. In the second half of this tutorial, you will learn how to build and use models that consist of more than one connected Bayesian network.

Before working through this tutorial, you should ensure that you have a sound working knowledge of agena.ai modeller by working through the previous two chapters.

4.1 Use of NPT expressions and simulation nodes

4.1.1 Introducing the Hypothesis Testing Model

1. Imagine we have two materials, A and B. From sampled data we wish to find which of the two has the best quality; that is, which material has the lowest probability of containing faults.

2. Let’s say that we take 200 samples of material A and 100 samples of material B. Of these we find that 10 of material A and 9 of material B are faulty.

3. Based on this data, can we truly believe that material A is better than material B? We need to test the so-called ‘null hypothesis’:

   H0: pA = pB (meaning there is no difference in the quality of the materials)

   against the so-called ‘alternative hypothesis’:

   H1: pA > pB (meaning material A is better than material B)

4.1.2 Using Conditional Statements

1. Create a new model.

2. In this model, we are going to use the following colour scheme for nodes:

Light Blue	Sample Nodes
Medium Blue	Probability Nodes
Dark Blue	Result Node
Green	Assumption Node

Table 4 Colour scheme for *Hypothesis Testing* model

3. Add a new simulation node and change the following properties:

Node Name	Probability Material A is Faulty
Unique Identifier	p_material_a_faulty
Background Colour	Medium Blue

Table 5 Properties for new node

4. Click **Apply** and then click **OK** on the information message that appears.

5. Copy the node you just created, paste the copy and move it so that it appears to the right of the original.

6. Change the **Node Name** of the new node to **Probability Material B is Faulty** and change the **Unique Identifier** to **p_material_b_faulty**.

7. Copy either of the two nodes and paste the copy beneath them.

8. Make the new node a child of the two existing nodes.

9. Change the **Node Name** of the new node to **p(A is faulty) – p(B is faulty)** and change the **Unique Identifier** to **a_b_difference**.

10. Change the type of expression on the node from **Normal** to **Arithmetic** and, in the **Arithmetic Expression** text box, supply the following expression:

    p_material_a_faulty – p_material_b_faulty

> While you are typing the expression, you will notice that a red border surrounds the box. This indicates that the expression is invalid. If you hold your mouse over the box, a tooltip will appear that explains the cause of the problem. When the expression is complete and valid, the red border will disappear.

11. Click **OK** to complete node properties editing.

12. Create a new node below *p(A is faulty) – p(B is faulty)*. This is going to be a conditional node that represents our two hypotheses and shows the probability of each of them being true.

13. Make the new node a child of *p(A is faulty) – p(B is faulty)*.

14. Open the properties dialog for the new node.

15. Change the **Node Name** to **Hypothesis**, change the **Unique Identifier** to **hypothesis**.

16. Click **Apply**.

17. In the **Node States** tab, change the **State Options** drop down box to **Customised**. Type **Material A better than Material B** in the **Positive Outcome** field and type **Material A not better than Material B** in the **Negative Outcome** field. These two states correspond to the alternative and null hypotheses we want to test.

18. Click **Apply**.

19. Switch to the **Node Probability Table** tab and change the **NPT Editing Type** to **Expression**. Because this is a Boolean node, only one type of expression is available: Comparative.

20. In the **Comparative Expression** field, supply the following expression:

    if(a_b_difference<0,"Material A better than Material B",

    "Material A not better than Material B")

    This expression means that, during calculation, the **Material A better than Material B** state will be chosen when the difference between the two probabilities is less than zero; otherwise, the **Material A not better than Material B** state will be chosen.

> Make sure that you type the exact case-sensitive values of the states. Otherwise, when you generate the NPT, you will get an error message telling that you the NPT contains zero probabilities (because the NPT generator has been unable to match the states in the expression with the states of the node).

21. Click **Apply**.

22. In the **Appearance** tab change the background colour of the node to dark blue and in the **Text Format** tab, change the text colour to white.

23. Click **Apply**.

24. Click **OK**.

25. Run the model, display the charts for all nodes and verify that, in the absence of any information, the probability of material A being better than material b is **0.5**. Your model should now look like the one in Figure 37 below.

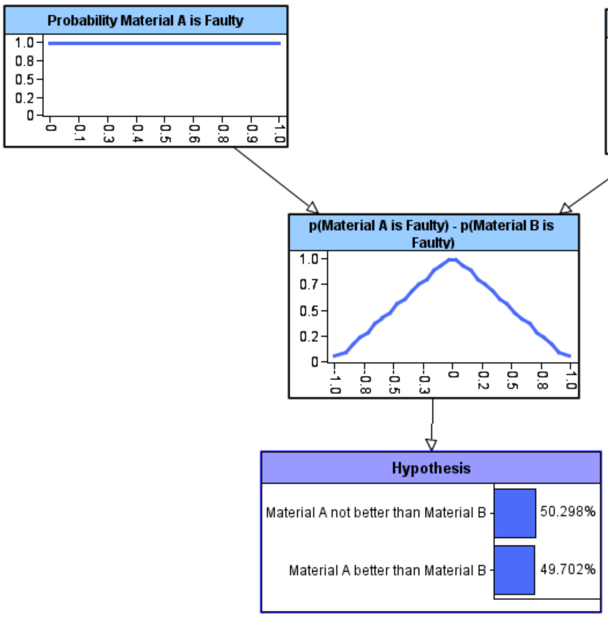


Figure 37 Probabilities of *Hypothesis* node with no data

4.1.3 Using discrete probability distributions

1. You are now going to add nodes that represent the samples of the two materials taken during the test.
2. Create a new simulation node and make it a child of *Probability Material A is Faulty*.
3. Open the properties dialog for this node.
4. Change the **Node Name** to **Material A: faults in 200 trials** and change the **Unique Identifier** to **a_faults**.
5. Change the node type to **Integer Interval**. This type is appropriate because the number of faults observed during sampling is always going to be a whole number.
6. Click **Apply**.
7. Click on the **Node Probability Table** tab and select **Binomial** in the **Expression Type** drop-down box.
8. Enter the value **200** in the **Number of Trials** text box and enter the unique identifier of this node's parent (**p_material_a_faulty**) in the **Probability of Success** text box. This means that the NPT generated for this node will be based on a Binomial distribution of faults observed. When you come to use the model, you will enter hard evidence on this node (in the form of a number of faulty samples observed). Then, via backward calculation, the model will update the probability of material A being faulty based on this evidence.
9. Click **Apply.**
10. In the **Appearance** tab change the background colour of the node to light blue.
11. Click **OK**.
12. Copy *Material A: faults in 200 Trials* and paste it. When you copy and paste a node that has one or more parents, the new copy is connected to the same set of parents by default.
13. Select just the edge between *Probability Material A is Faulty* and the new node and delete it.
14. Move the new node over to the right of the Graph and make it a child of *Probability Material B is Faulty*.
15. Bring up the properties dialog for the new node.
16. Change the value of **Node Name** to **Material B: faults in 100 trials** and change the value of **Unique Identifier** to **b_faults**.
17. Click on the **Node Probability Table** tab. Notice that there is a red warning saying that the NPT needs to be regenerated and that there is a red border around the **Probability of Success** text box. This is because the Binomial expression refers to the parent of the copied node. It needs to be rectified. Clear the **Probability of Success** box and enter **p_material_b_faulty**.
18. In the **Number of Trials** box, enter the value **100**.
19. Click **OK**.

20. The model is now complete and should look similar to the one in Figure 38.

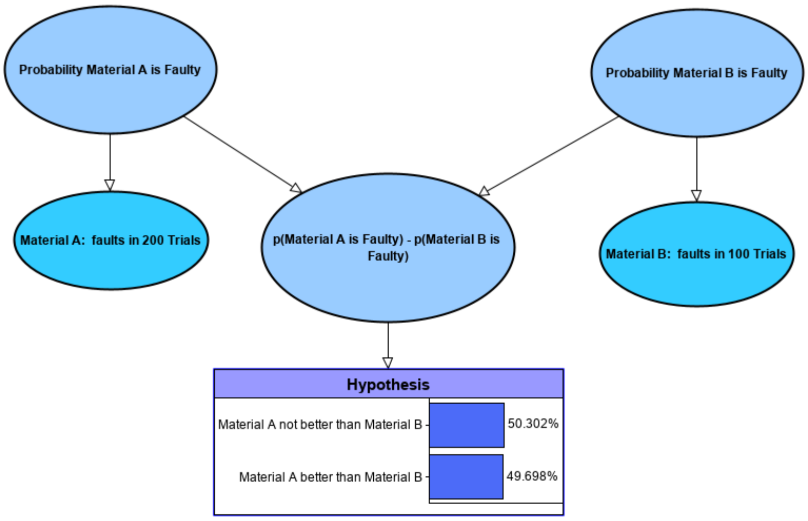


Figure 38 Complete *Hypothesis Testing* model

4.1.4 Calculating the model

1. You can now use the model to test the two hypotheses.

2. If automatic calculation is not already activated, turn it on by toggling the toolbar button:

3. Enter the data **10** for the node *Material A: faults in 200 trials* and enter the data **9** for *Material B: faults in 100 trials*.

4. Look at the chart for the *Hypothesis* node. It should resemble the one in Figure 39 below.

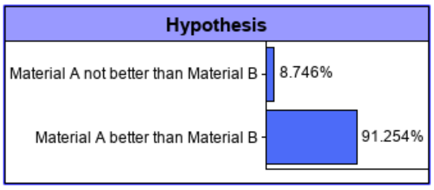


Figure 39 Chart for *Hypothesis* node after evidence has been entered

5. Verify that the probability of material A being better than material B is **0.91254** (i.e., roughly 91%).

6. Experiment with different numbers of samples to see how the result of the hypothesis test changes.

7. Save your model to a writable directory.

8. At this point you might want to compare your model against the solution which can be found in *Model Library / Tutorials / Hypothesis Testing.ast*

4.1.5 Capturing Different Prior Assumptions

1. Close down any open charts.

2. Create a new node at the top of the Graph. (Select all of the existing nodes and move them down if necessary.)

3. Open the properties dialog and change the following properties:

Node Name	Prior Type
Unique Identifier	prior_type
Node Type	Labelled
States	Uniform, Beta
Background Colour	Green

Table 6 Properties for new node

4. Click **OK**.

5. Make *Prior Type* a parent of both *Probability Material A is Faulty* and *Probability Material B is Faulty*.

6. Open the properties dialog for *Probability Material A is Faulty* and switch to the **Node Probability Table** tab.

7. Change the **NPT Editing Mode** to **Partitioned Expression** and move the **Prior Type** entry in the left-hand list over to the right-hand list. You are going to supply one expression to be used when the state of *Prior Type* is *Uniform* and a different expression to be used when the state of *Prior Type* is *Beta*.

8. Double-click on the first cell in the expression table and, in the dialog that appears, specify a **Uniform** distribution with a **Lower Bound** of **0** and an **Upper Bound** of **1**. This is equivalent to a prior assumption of complete ignorance about the likelihood of material A being faulty.

9. Click **OK**.

10. Double-click on the second cell in the expression table and, in the dialog that appears, specify a **Beta** distribution. Enter **1** for **Alpha** and **9** for **Beta**. Specify a **Lower Bound** of **0** and an **Upper Bound** of **1**. This is equivalent to a prior assumption that there is a 1 in 10 chance of material A being faulty.

11. Click **OK**.

12. Define a partitioned expression on *Probability Material B is Faulty* just as you did for *Probability Material A is Faulty*. Specify it identically but make the **Alpha** parameter of the **Beta** distribution **2** and the **Beta** parameter **8**. This encodes the assumption that there is a 1 in 5 chance of material B being faulty.

13. Clear the two existing data from the model (by choosing **Tools** ⟶ **Clear Entered Data** ⟶ **All** from the menu bar).

14. Double-click on the two nodes whose NPTs you just modified and resize the nodes so that you can see the shapes of the graphs. Notice how the two prior assumptions differ as shown in Figure 40.

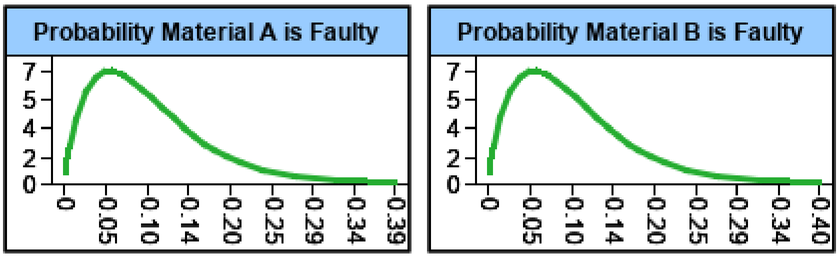


Figure 40 *Hypothesis Testing* model showing different prior assumptions

15. Close the two charts.

16. You are now going to compare two cases side-by-side: one in which the prior assumption is that the distribution of faults is uniform and the other in which a beta distribution of faults is assumed. The latter case might represent the opinion of a materials expert, for example.

17. Expand the Data panel and rename **Case 1** to **Uniform.**

18. Right click the existing case and Add a New Case, rename it **Beta** and tick the **Active** checkbox.

19. Right-click on the *Prior Type* node and choose **Enter Data** ⟶ **Uniform** ⟶ **Uniform**. By doing this, you are entering the data that the *Prior Type* is *Uniform* in the case that you have named **Uniform**.

20. Right-click on the *Prior Type* node again and choose **Enter Data** ⟶ **Beta** ⟶ **Beta**. By doing this, you are entering the data that the *Prior Type* is *Beta* in the case that you have named **Beta**.

21. Double-click on the *Hypothesis* node to show its chart. It should look like the one shown in Figure 41.

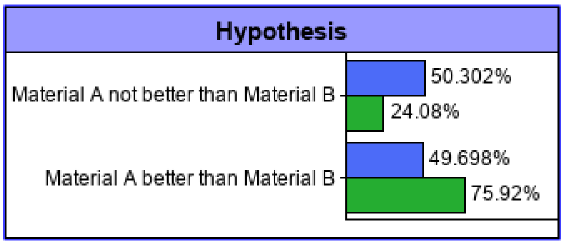


Figure 41 Chart showing results of hypothesis test with two different prior assumptions

Note that the two cases are plotted on the chart. Drag your mouse the points on the two lines and note the values of the two states for each case.

22. The data tells us that, in the absence of any sampling information, when we assume that the probabilities of faultiness are uniformly distributed, there is roughly a **50%** chance that material A is better than material B.

23. However, if our materials expert believes that the prior faultiness probabilities are characterised by beta distributions – Beta(1,9) and Beta(2,8) for material A and material B respectively – then this results in a **76%** chance that material A is better than material B.

4.1.6 Combining data and prior assumptions

1. Open the properties dialog for the node *Material A: faults in 200 trials*.

2. Change the name of the node to *Material A: faults in 10 trials* and change the **Number of Trials** parameter of the Binomial distribution from **200** to **10**.

3. Click **OK**.

4. Open the properties dialog for the node *Material B: faults in 100 trials*.

5. Change the name of the node to *Material B: faults in 10 trials* and change the **Number of Trials** parameter of the Binomial distribution from **100** to **10**.

6. Click **OK**.

7. Turn off the auto-calculate feature (because you are about to enter four data as a “batch” and you don’t need to calculate the model after each one).

8. For both the **Uniform** and the **Beta** case, enter data of **1** fault on *Material A: faults in 10 trials*.

9. For both the **Uniform** and the **Beta** case, enter data of **0** faults on *Material B: faults in 10 trials*.

10. Perform a calculation. The graph for the *Hypothesis* node should now look like this:

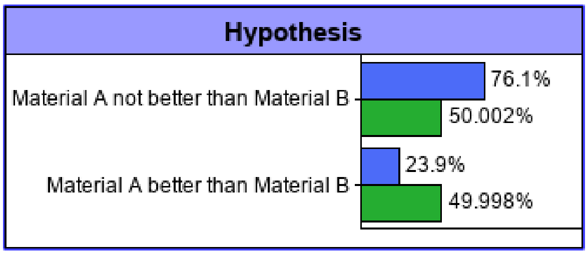


Figure 42 Results of hypothesis test after entering sparse sample data

11. Hold your mouse over the points of the graph to see the state probabilities.

12. Notice that, when we use uniform priors for the two faultiness probabilities, the sparse sample data tells us, with a confidence of roughly **76%**, that material A is not better than material B. However, when we encode the opinion of an expert in the two Beta priors, these priors serve to supplement the sample data and tell us that the quality of the two materials is roughly **equal**.

13. Save your model to a writeable directory.

14. At this point you might want to compare your model against the solution which can be found in *Model Library / Tutorials / Hypothesis Testing with Expert Judgement.ast*

4.1.7 Changing the simulation settings

1. Close all the charts.

2. Open the chart for node *Probability Material B is Faulty*. Drag your mouse over the chart, selecting the **Beta** case and observe that the mean value is **0.10008** and that the variance is **0.004356.** Also notice that the entropy error statistic is **0.0011503**.

3. Open the model properties by clicking on **File ⟶ Model Properties** from the menu bar and switch to the **Simulation Settings** tab.

4. Verify that the current value in the **Simulation Convergence** field is **0.001**. This setting determines how accurate simulation will be (and conversely how much or how little time you wish to wait for the calculation to take). The setting is a threshold for the entropy error for each node. Notice that the entropy error for *Probability Material B is Faulty* is 0.0011503 which is more than the 0.001.

5. Change the value of **Simulation Convergence** to **0.1** and then click **OK**.

6. Perform a calculation.

7. Observe that the calculation takes less time but the results are slightly less accurate: the mean value is now **0.099786** and the variance is **0.0045691**. The entropy error is also now **0.02306**, which is less than the new target of 0.1 but much higher than that achieved before, which was **0.0011503**.

8. Modelling often requires you to make a trade-off between speed and accuracy; in some situations, accuracy will be more important and in others speed will be more important.

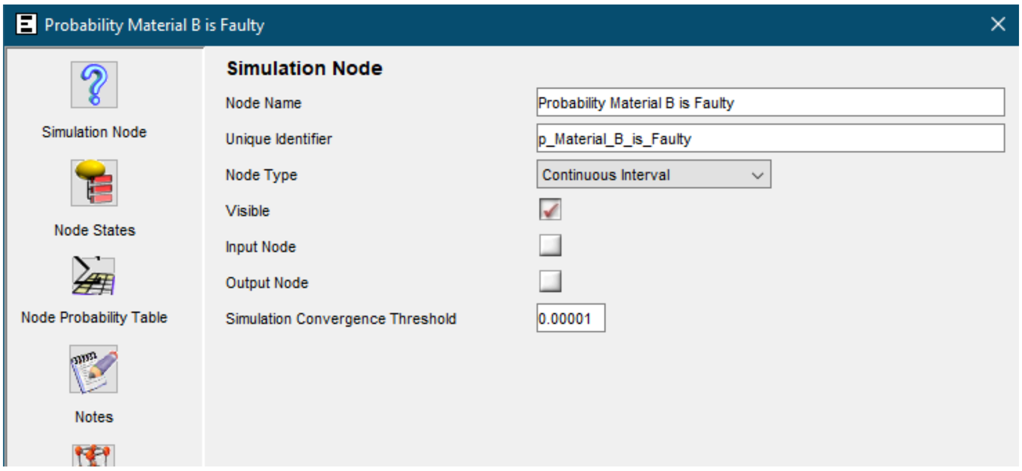
4.1.8 Changing the model simulation settings

1. By default, the model simulation convergence settings are applied to every node. So, in the default case, every simulation node in the model will have the default simulation convergence value of 0.001

2. In large complex models with many simulation nodes, calculating the model at more accurate (i.e., lower) simulation convergence than the default value of 0.001 can result in long calculation times.

3. If you need the higher accuracy associated with lower simulation convergence then, rather than change the model setting, you can define the simulation convergence for individual nodes.

4. For example, Figure 43 shows the simulation convergence threshold for a node “*Probability Material B is Faulty*” being set at 0.00001.



Simulation Node

Node Name: Probability Material B is Faulty
Unique Identifier: p_Material_B_is_Faulty
Node Type: Continuous Interval
Visible ☑
Input Node ☐
Output Node ☐
Simulation Convergence Threshold: 0.00001

Figure 43 Setting simulation convergence for an individual node

4.2 Models with multiple Bayesian networks

4.2.1 Introducing the river flooding model

1. The River Flooding model predicts whether a river will flood using information about the water level of a river, the amount of rainfall and the flood defences in place to prevent flooding.

2. The model is structured in such a way that is possible to link together different instances of it to monitor the flood risk over time.

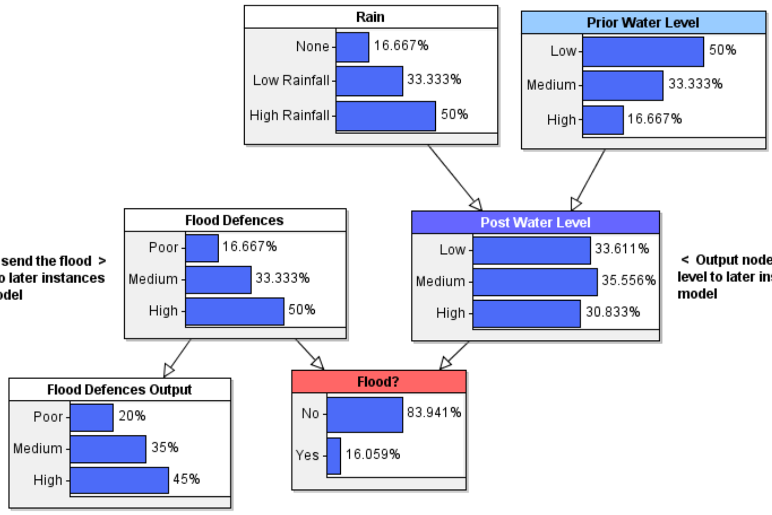
3. The amount of rainfall, together with the “prior” water level (i.e., before rainfall), influences the “post” water level (i.e. after rainfall). This “post” water level can then be used as the “prior” water level in a new instance of the model that represents the next time period.

4. The quality of flood defences and the water level after rainfall determine whether the river will flood. As with the “post” water level, the quality of the flood defences can be linked to the same node in the next time period.

4.2.2 Importing models

1. Start agena.ai modeller and open the *Model Library / Tutorials / River Flooding Basic Object.ast* model.

2. Ensure that the *River Flooding* Bayesian network is selected in the Explorer view. The model should look like the one in Figure 44 below.



< Input node, can take an input from a previous instance of the flooding network

Output node, can send the flood > defence quality to later instances of the flooding model

< Output node, can send the water level to later instances of the flooding model

Figure 44 *River Flooding* model

3. Experiment with different values of *Rain*, *Prior Water Level* and *Flood Defences* to see what impact they have on the probability of a flood occurring.

4. When you are comfortable with how the basic model works, open the properties dialog for *Prior Water Level* and observe the *Input Node* box is checked and then click **OK**.

5. Repeat step 4 for *Flood Defences*.

6. Open the properties dialog for *Post Water Level*, verify that the *Output Node* box is ticked and click **OK**.

7. Repeat step 6 for *Flood Defences Output*.

8. Now save the model to a writeable directory and call it *Dynamic River Flooding.cmpx*.

9. Click on **File** ⟶ **Import Model…**, select the Dynamic *River Flooding.cmpx* file that you just saved and click **OK**.

10. If you click on the root item in the Explorer, you can see both models in the Graph view:

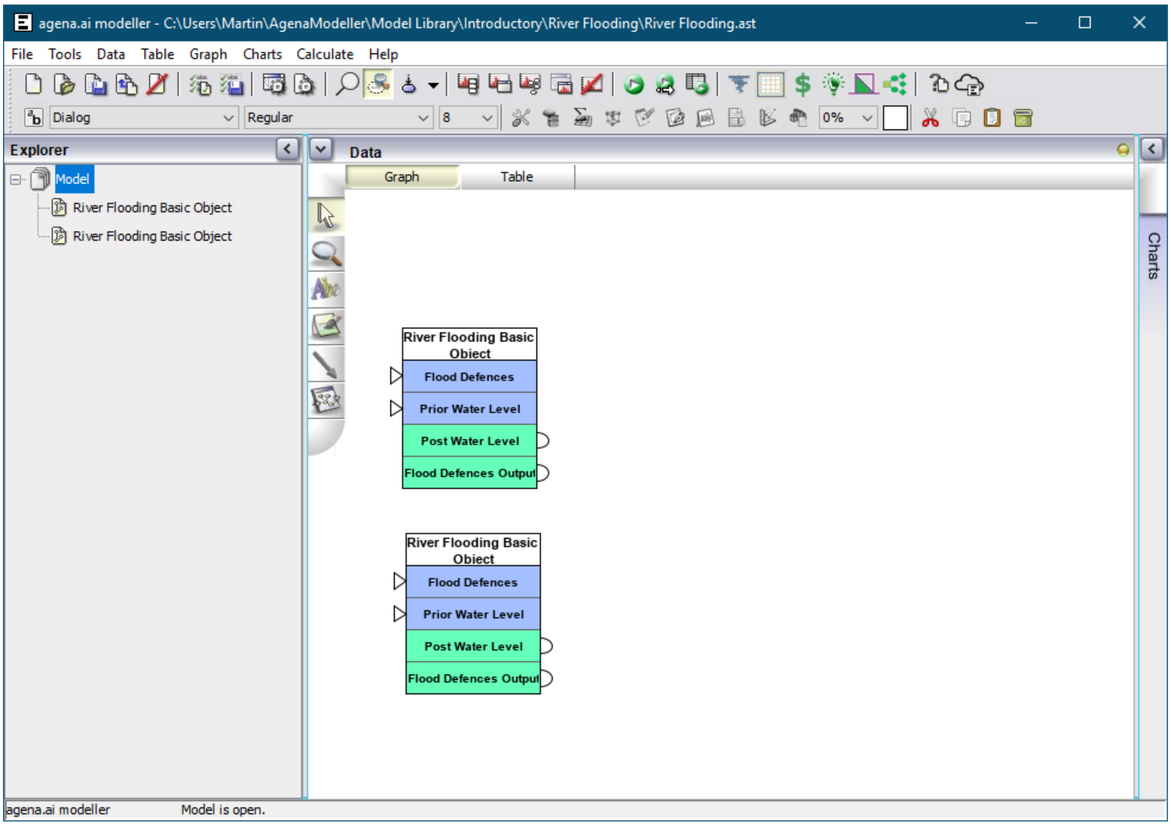


Figure 45 *River Flooding* model showing two Bayesian networks after import

11. Notice that the Graph view shows both Bayesian networks and that each Bayesian network lists its input nodes (shown in blue) and output nodes (shown in green).

12. Import *River Flooding.ast* once more so that your model contains three identical Bayesian networks.

13. In the Explorer, rename the three objects *T1*, *T2* and *T3* (or even *Day 1, Day 2 and Day 3*). Then, in the Graph, line up the three objects horizontally so that they appear in chronological order. Your model should now look like the one shown in Figure 46.

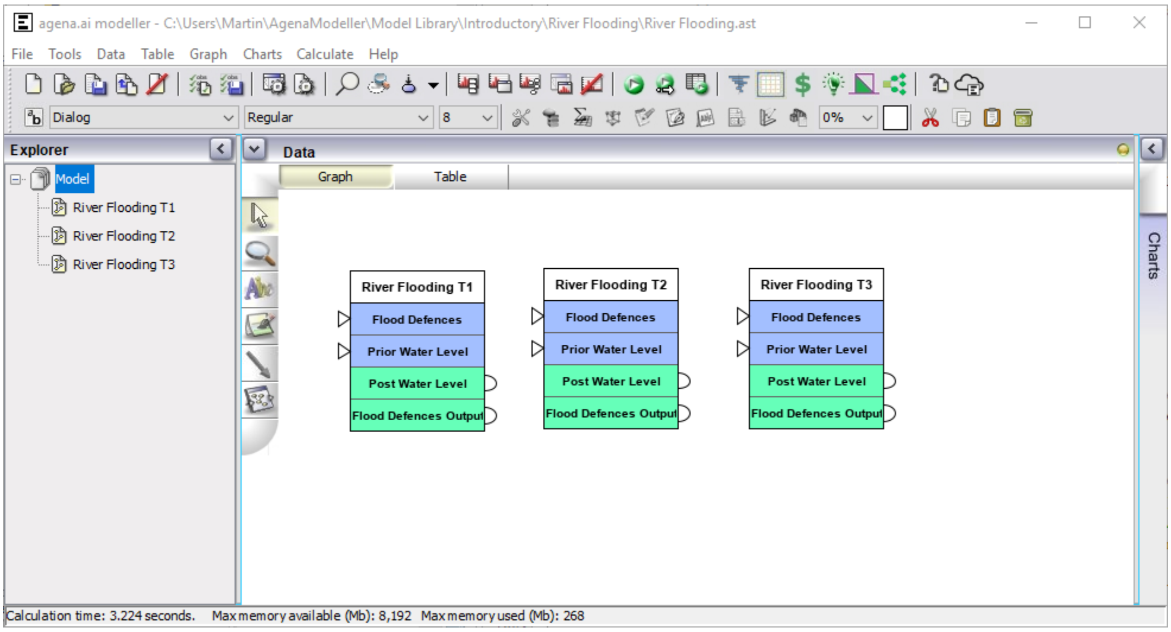


Figure 46 ***River Flooding*** **model showing three Bayesian networks**

14. The River Flooding model now contains three Bayesian networks that are identical in all but name. Select each Bayesian network in the Explorer and verify that the Graph and Table for each object look the same.

4.2.3 Connecting Bayesian networks

1. To capture the time dimension in our model so that you can reason about how changes in the water level each day affect the risk of flooding, you need to connect the three Bayesian networks together.

2. Click on the edge tool on the Graph toolbar:

3. Click on the *Post Water Level* segment of the *Day 1* Bayesian network, move your mouse to the *Prior Water Level* segment of the *Day 2* Bayesian network and click again. This creates a connection between the two Bayesian networks:

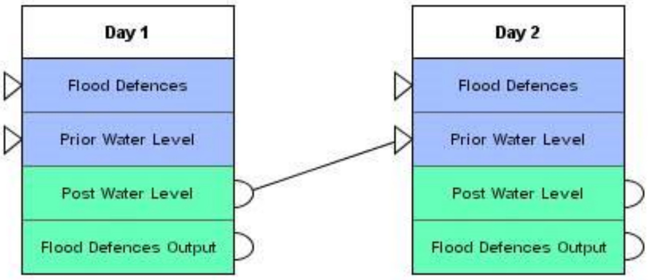


Figure 47 A link between two Bayesian networks

4. This link means that, whenever the model is calculated, the probabilities associated with each state of the *Post Water Level* node in *Day 1* are entered as probability distributions in

the *Prior Water Level* node in *Day 2*. In this way, the water level at the end of the first day becomes the water level at the beginning of the second day.

5. Draw another link between *Flood Defences Output* in *Day 1* and *Flood Defences* in *Day 2* so that the two Bayesian networks look like those in Figure 48.

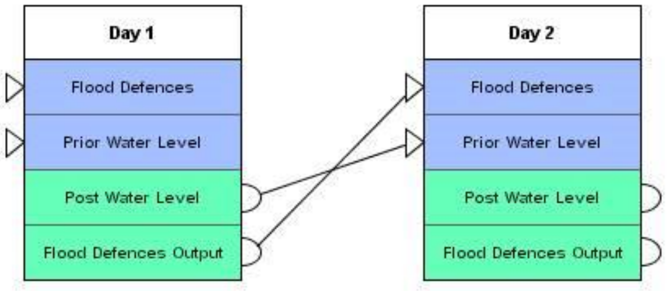


Figure 48 Two connections between two Bayesian networks

6. Connect *Day 2* with *Day 3* in the same way:

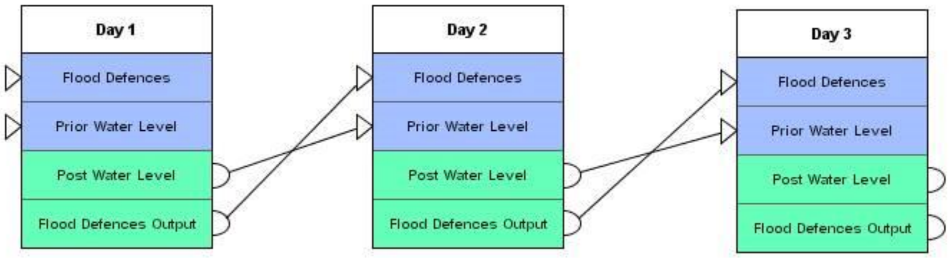


Figure 49 Three connected Bayesian Networks

7. The model is now fully connected and can be used to assess how the risk of flooding changes over a three-day period.

4.2.4 Calculating the model

1. Turn on the auto-calculate option by toggling the toolbar button:

2. Close all charts by selecting **Charts** → **Close All Charts** from the menu bar or by clicking:

3. You are now going to use the model to explore how three days of high rainfall affects the risk of flooding and the quality of the flood defences.

4. Using the Explorer, click on **Day 1** and enter the data **High Rainfall** on the *Rain* node.

5. Enter the same data for **Day 2** and **Day 3**.

6. Return to **Day 1** in the Explorer and double-click on the *Flood?* and *Flood Defences* nodes to show their charts. Hold your mouse over the bars on the graphs to view the exact probability values of the states.

7. Click on **Day 2** in the Explorer and show the same two charts. Notice that the risk of flooding has increased, and that the quality of the flood defences has degraded.

8. Click on **Day 3** in the Explorer. Open the two charts and observe that the risk of flooding has increased again, and that the quality of the flood defences has degraded further still.

9. Suppose, now, that you want to see what effects remedial action on day 3 will have on the risk of flooding.

10. Ensuring that **Day 3** is still selected in the Explorer, enter the data **High** on the *Flood Defences* node.

11. Verify that the risk of flooding on day 3 has now been reduced by over **half**.

12. You might like to check your model against the solution which can be found in *Model Library / Tutorials / River Flooding.ast model*

4.2.5 More powerful methods of passing parameters between Bayesian networks

Before we connected two Bayesian networks by linking an output node of one to an input node of another. For this to work the two nodes had to be the same type with the same set of state values. The result of the linking was to pass the entire set of probability values from the input node to the output node. While this type of linking is the default type of linking between Bayesian networks in Agena.ai modeller, you are not restricted to this. There are the following additional options for linking nodes in different Bayesian networks:

a) From a continuous node to a continuous node, you can either pass the full set of marginals (as was the previous default), or the value of a summary statistic as a constant. So, for example, the output node might represent a variable "height" and the input node might represent a variable "mean height". In this case the link type you would select would be the summary statistic "Mean".

b) From non-continuous node to a continuous node, you can pass the value of a single state as a constant. For example, the node "Flood" in the above tutorial is a Boolean node. We could link this to a continuous node (with a range 0 to 1) called "Flood probability" in another Bayesian network and specify that the value passed is the value of "True". If the value of the state "True" is 0.6 in the node "Flood" then the node "Flood probability" will have the value 0.6.

These options require you to specify what you wish to pass. There is a link type option in node properties which can be accessed by right-clicking on the **input** node and selecting properties. You then select the link type option on the left-hand side as shown in Figure 50. Note that this link type icon only appears if the input node is already joined to an output node, otherwise it will be hidden.

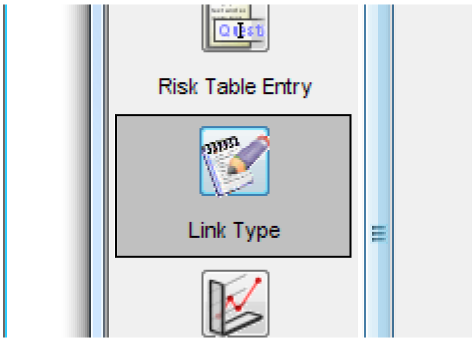
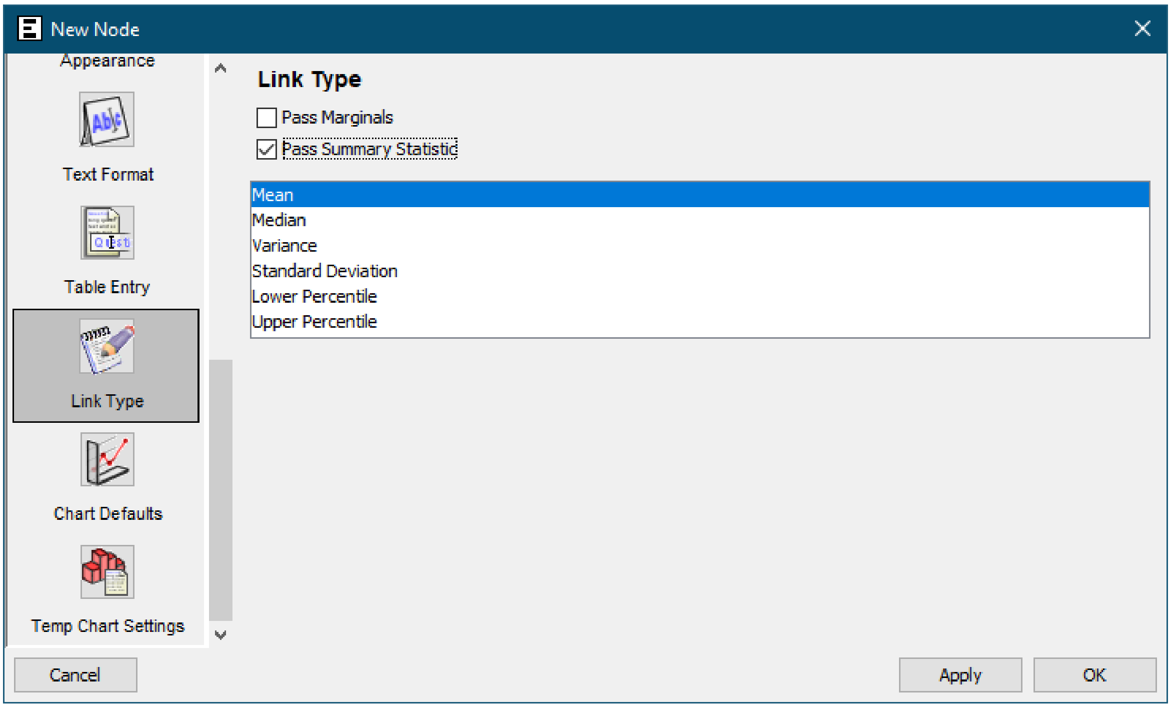


Figure 50 Selecting link type in node properties

The options in link type vary based upon whether it is option **a** or **b**. For option **a** you will see the dialogue in Figure 51.



Link Type

☐ Pass Marginals

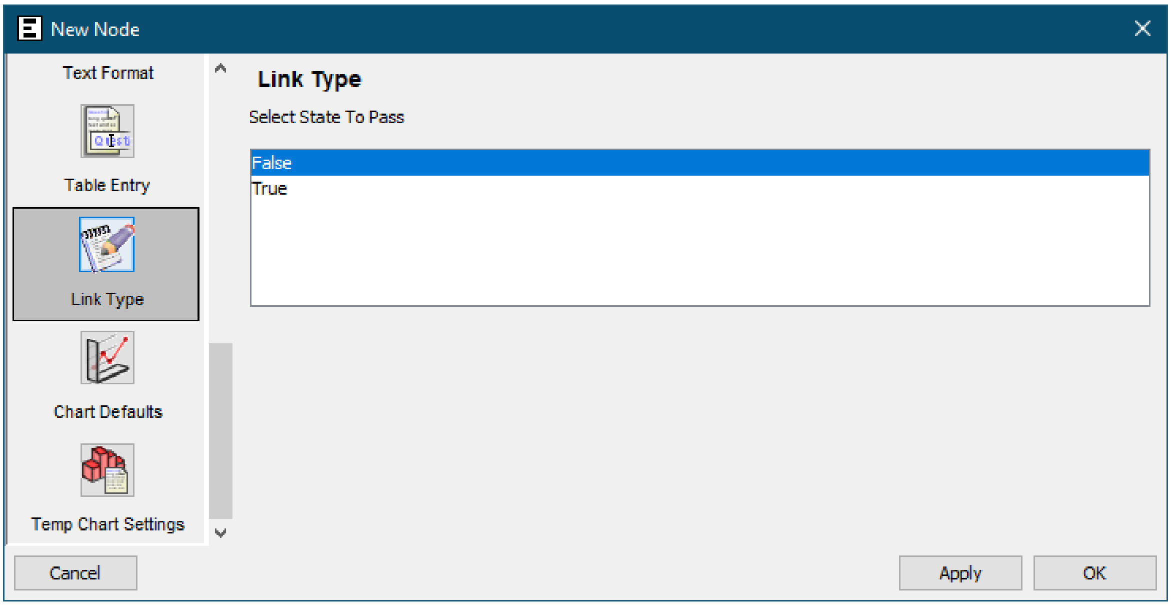
☑ Pass Summary Statistic

Mean
Median
Variance
Standard Deviation
Lower Percentile
Upper Percentile

Figure 51 Link types option for connecting continuous nodes

Here, in addition to passing full marginals you have the option to select which summary statistic to pass.

For option **b**, you can select which state value to pass as shown in Figure 52.



Link Type

Select State To Pass

False

True

Figure 52 Link type options for passing from non-continuous to continuous

Please note that once you select and press apply, the selection of simulation checkbox will be set automatically.

Using a link type, passing a state or summary statistic, disables the node probability table and expression editing.

4.3 Next Steps

This tutorial has shown you some of the advanced aspects of building models using Agena.ai modeller. To learn more about the many types of problem Agena.ai modeller can be used to model, you should explore and interact with the range of well-documented example models that come with the software. To learn more about working with numeric nodes and simulation nodes you should read Chapter 5.

5. Working with numeric nodes and simulation

One of the most powerful features of Agena.ai modeller, compared with other BN and simulation packages, is the way it enables users to include numeric nodes and numeric functions in their models without loss of accuracy. This tutorial explains what you need to know to get the most of this very powerful functionality. ***It is important that you are familiar with this before working with numeric nodes in Agena.ai modeller.***

The tutorial starts by highlighting the fundamental problem with traditional approaches to incorporating numeric nodes in BNs. Hence, by working through this you will find out not just how to implement numeric nodes in the 'traditional way' but also what the limitations of this approach are. Next, you will find out how things are radically improved by using simulation nodes. Finally, we explain how to avoid common errors in using simulation nodes.

5.1 Static discretization

The problem with the previous generation of BN tools is that working with numeric nodes is both complicated and inaccurate. This is because they use a static discretization of numeric nodes into intervals that do not change regardless of the evidence entered into the model. You must guess the state ranges before performing the calculation, thus pre-supposing that you know the resulting probability distribution of the results beforehand. In simple cases this may be quite easy, but in others it will be difficult or even impossible.

As a simple example let's suppose we are trying to build a model to predict the number of customers we will have at the end of the next year. Then in our model we might have two (uncertain) nodes:

A: the number of new customers gained during the year and

B: the number of customers lost during the year

Although A and B may be uncertain, we want to define a node

C: net customers lost during the year

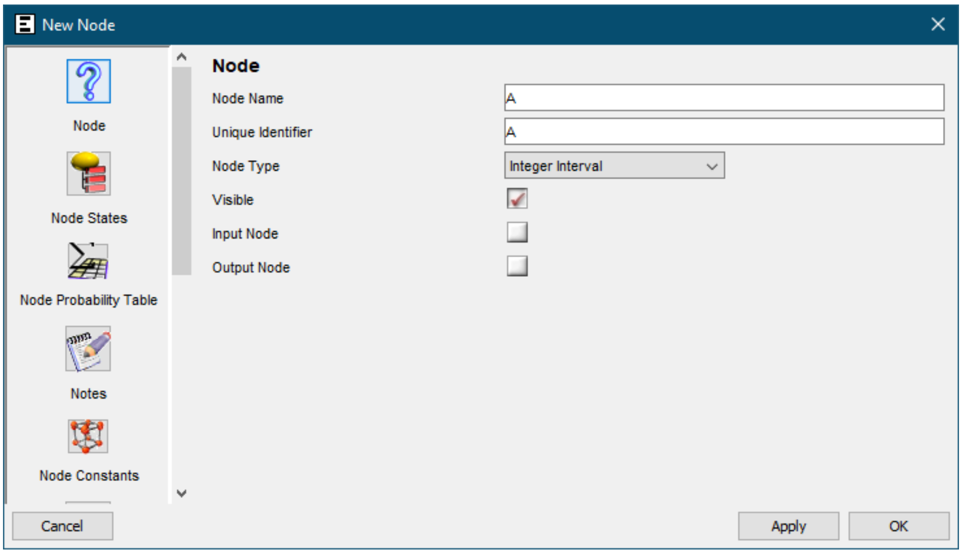
which is deterministic, specifically C = A-B

In the first part of the tutorial, you are going to build this simple model using the 'traditional approach' that is used in other BN tools.

1. Open a new model and create a new node by clicking the icon:

   ⬬

2. Right click on the node to bring up the properties dialog and make the selections shown in Figure 53 then click **Apply** (you will need to close a warning message that tells you it is more efficient to use a simulation node).

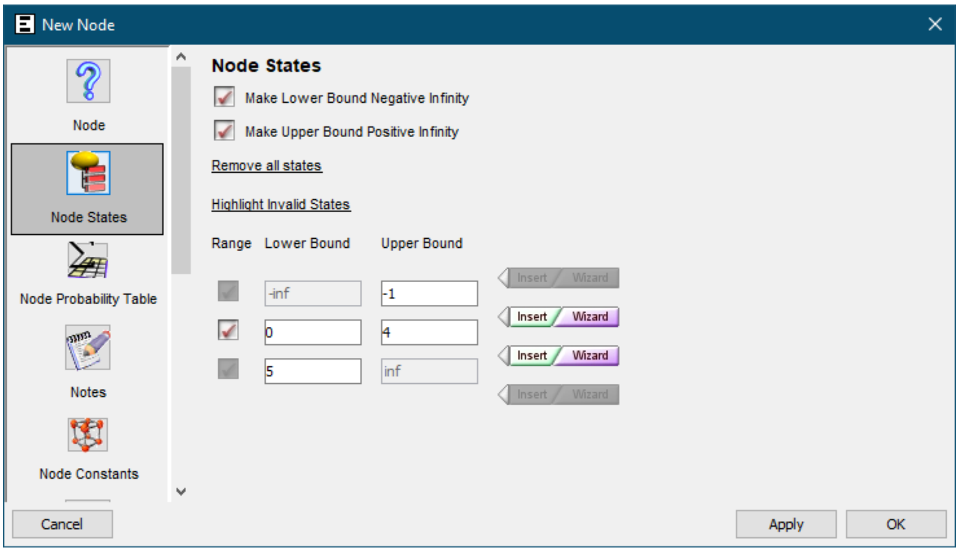


Node

Node Name: A
Unique Identifier: A
Node Type: Integer Interval
Visible ☑
Input Node ☐
Output Node ☐

Figure 53 Create an integer interval node

3. Click on the “Node States” icon in the dialog. Because the node type is now specified as “integer interval” you will see the default node states as shown in Figure 54



Node States

☑ Make Lower Bound Negative Infinity
☑ Make Upper Bound Positive Infinity

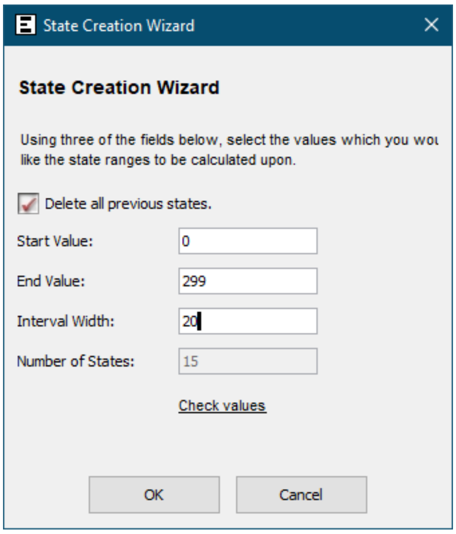
Remove all states

Highlight Invalid States

Range	Lower Bound	Upper Bound
☑	-inf	-1
☑	0	4
☑	5	inf

Figure 54 Default nodes states for integer interval node

4. Because we are using the static discretization approach you have to make a decision now about the range for node A and the relevant set of intervals within this range. Suppose that the typical number of new customers is something like 100, then it might seem reasonable to specify a range of 0 to 299 with intervals of size 20. Fortunately, in Agena.ai modeller you do not have to do this manually. First click on the “Remove all states” option and then simply click on the top ‘wizard’ button to bring up the dialog shown in Figure 55.


State Creation Wizard

State Creation Wizard

Using three of the fields below, select the values which you wou
like the state ranges to be calculated upon.

☑ Delete all previous states.

Start Value: 0

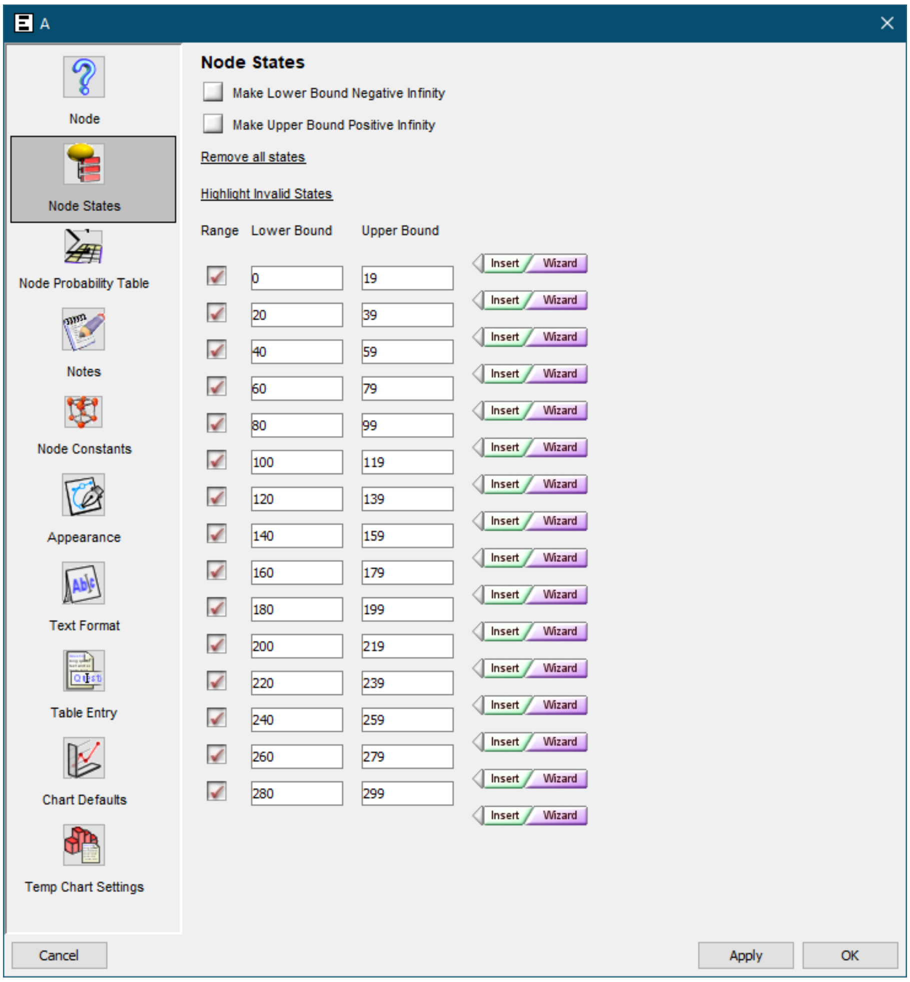
End Value: 299

Interval Width: 20

Number of States: 15

Figure 55 State creation wizard

5. Change the End Value to **299** and the Interval Width to **20**. Also check the box “Delete all previous states”. Press **OK**. You should then see the revised node states shown in Figure 56. Press **Apply**.


A

Node States

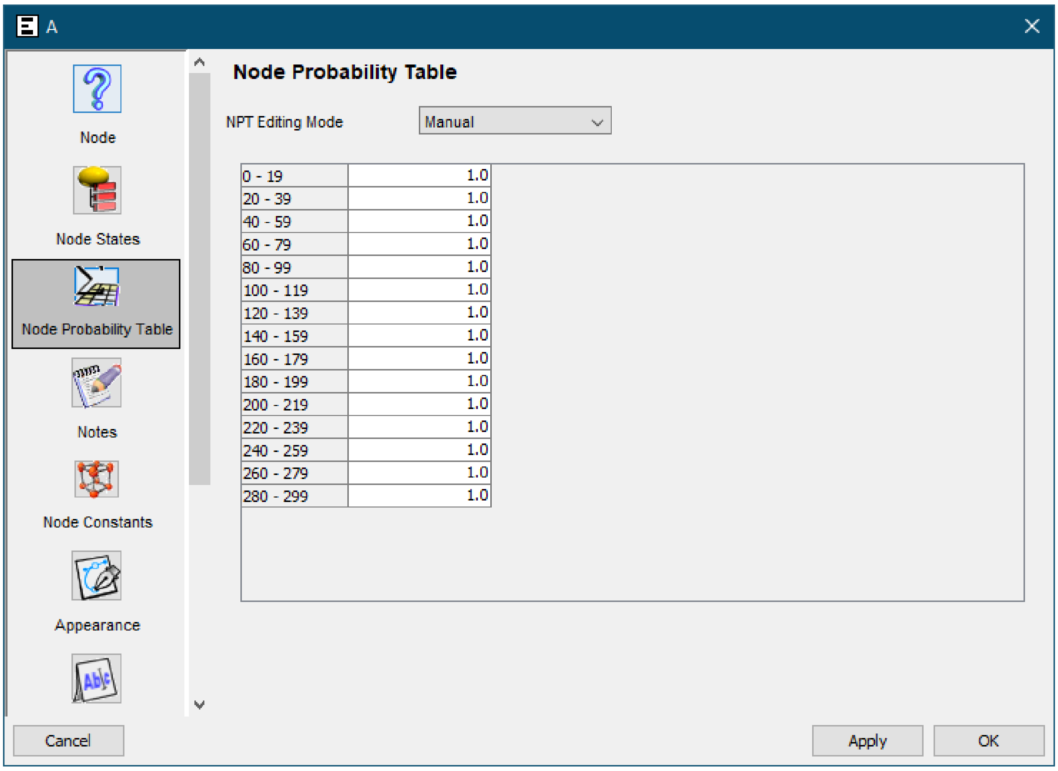
☐ Make Lower Bound Negative Infinity

☐ Make Upper Bound Positive Infinity

Range	Lower Bound	Upper Bound
☑	0	19
☑	20	39
☑	40	59
☑	60	79
☑	80	99
☑	100	119
☑	120	139
☑	140	159
☑	160	179
☑	180	199
☑	200	219
☑	220	239
☑	240	259
☑	260	279
☑	280	299

Figure 56 New states created automatically

6. Now click on the “Node Probability Table” icon. You should see the dialog shown in Figure 57. Note that by default the tool has assigned each of the state intervals equal probability. Note that “Placeholder” text should be overwritten by the expression you enter.



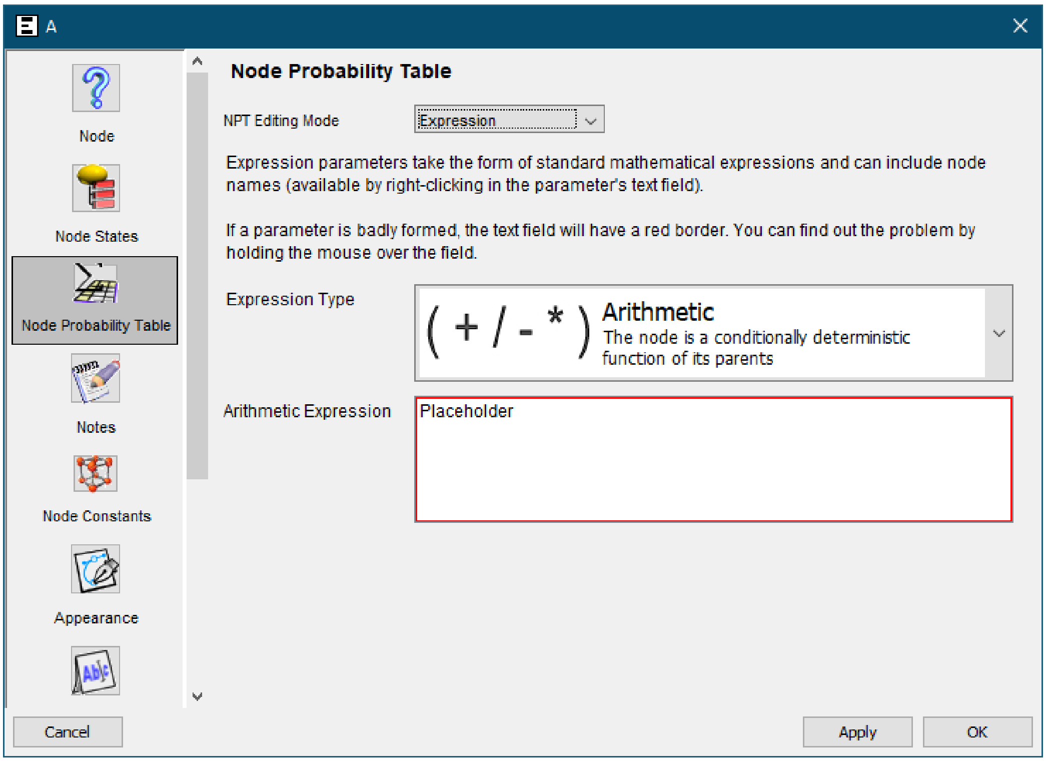
Node Probability Table

NPT Editing Mode: Manual

0 - 19	1.0
20 - 39	1.0
40 - 59	1.0
60 - 79	1.0
80 - 99	1.0
100 - 119	1.0
120 - 139	1.0
140 - 159	1.0
160 - 179	1.0
180 - 199	1.0
200 - 219	1.0
220 - 239	1.0
240 - 259	1.0
260 - 279	1.0
280 - 299	1.0

Figure 57 Node probability table dialog

7. Instead of this manual, uniform distribution, suppose that data for previous years suggests the prior distribution is something like a normal distribution peaking at 100. To capture this, simply select the NPT Editing Mode “Expression” as shown in Figure 58.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

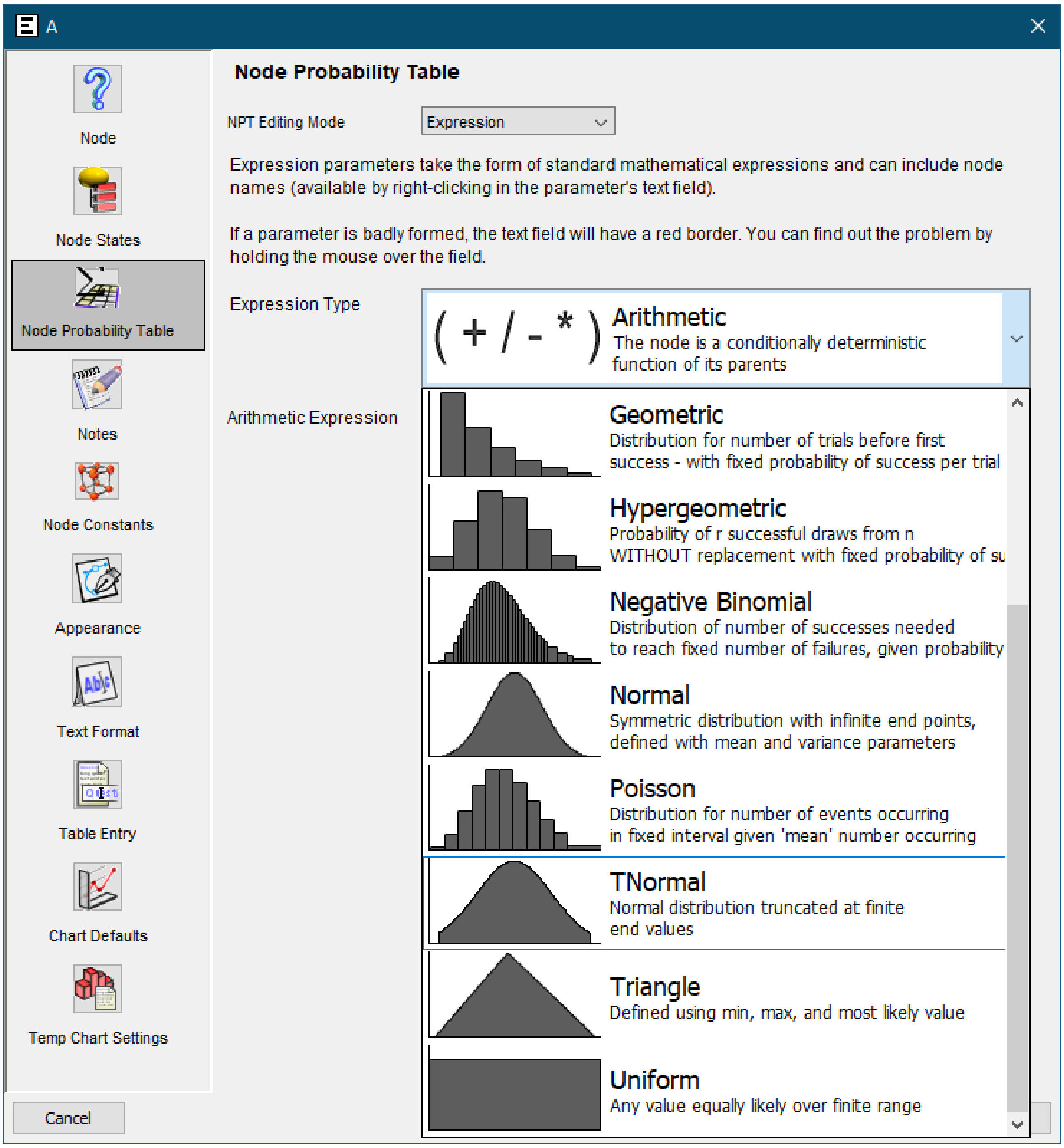
If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: (+ / - *) Arithmetic
The node is a conditionally deterministic function of its parents

Arithmetic Expression: Placeholder

Figure 58 Entering an expression

8. For the expression type, select the TNormal as shown in Figure 59 (the TNormal is the truncated form of the Normal distribution. It makes sense to use this, rather than the Normal distribution since the range of the node is finite, from 0 to 299).



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: Arithmetic — The node is a conditionally deterministic function of its parents

Arithmetic Expression

Geometric — Distribution for number of trials before first success - with fixed probability of success per trial

Hypergeometric — Probability of r successful draws from n WITHOUT replacement with fixed probability of su

Negative Binomial — Distribution of number of successes needed to reach fixed number of failures, given probability

Normal — Symmetric distribution with infinite end points, defined with mean and variance parameters

Poisson — Distribution for number of events occurring in fixed interval given 'mean' number occurring

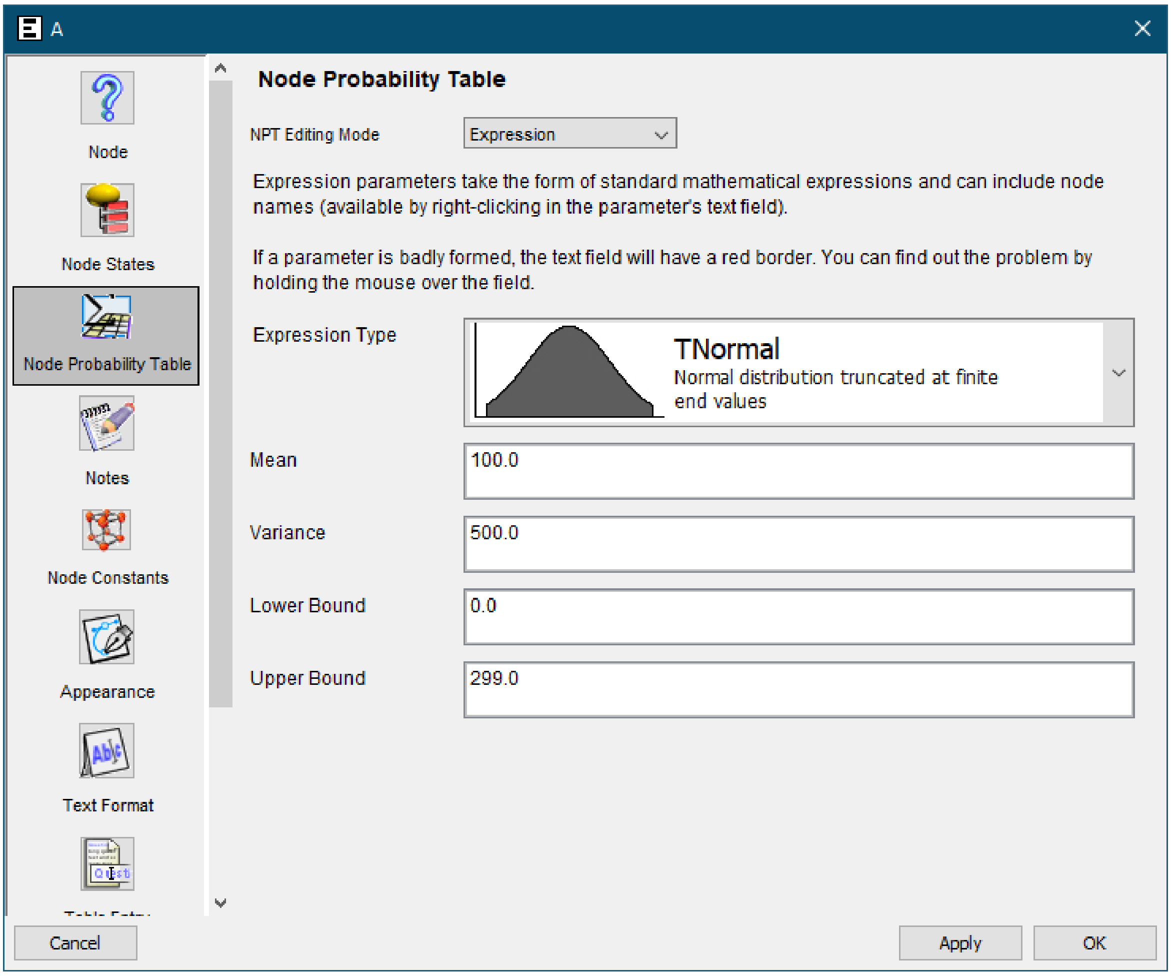
TNormal — Normal distribution truncated at finite end values

Triangle — Defined using min, max, and most likely value

Uniform — Any value equally likely over finite range

Figure 59 Selecting the TNormal distribution

9. Enter the values for the parameters Mean and Variance as shown in Figure 60 and press **OK**.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: TNormal — Normal distribution truncated at finite end values

Mean: 100.0

Variance: 500.0

Lower Bound: 0.0

Upper Bound: 299.0

Figure 60 Entering the parameters for the statistical distribution for node probability table on node A

10. Now calculate the model by clicking the Run Calculation icon

11. Double click on the node A to bring up its chart and enlarge it so that you see something like the chart shown in Figure 61.

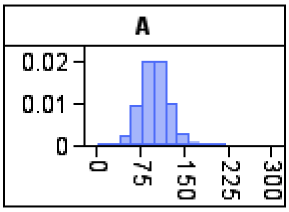
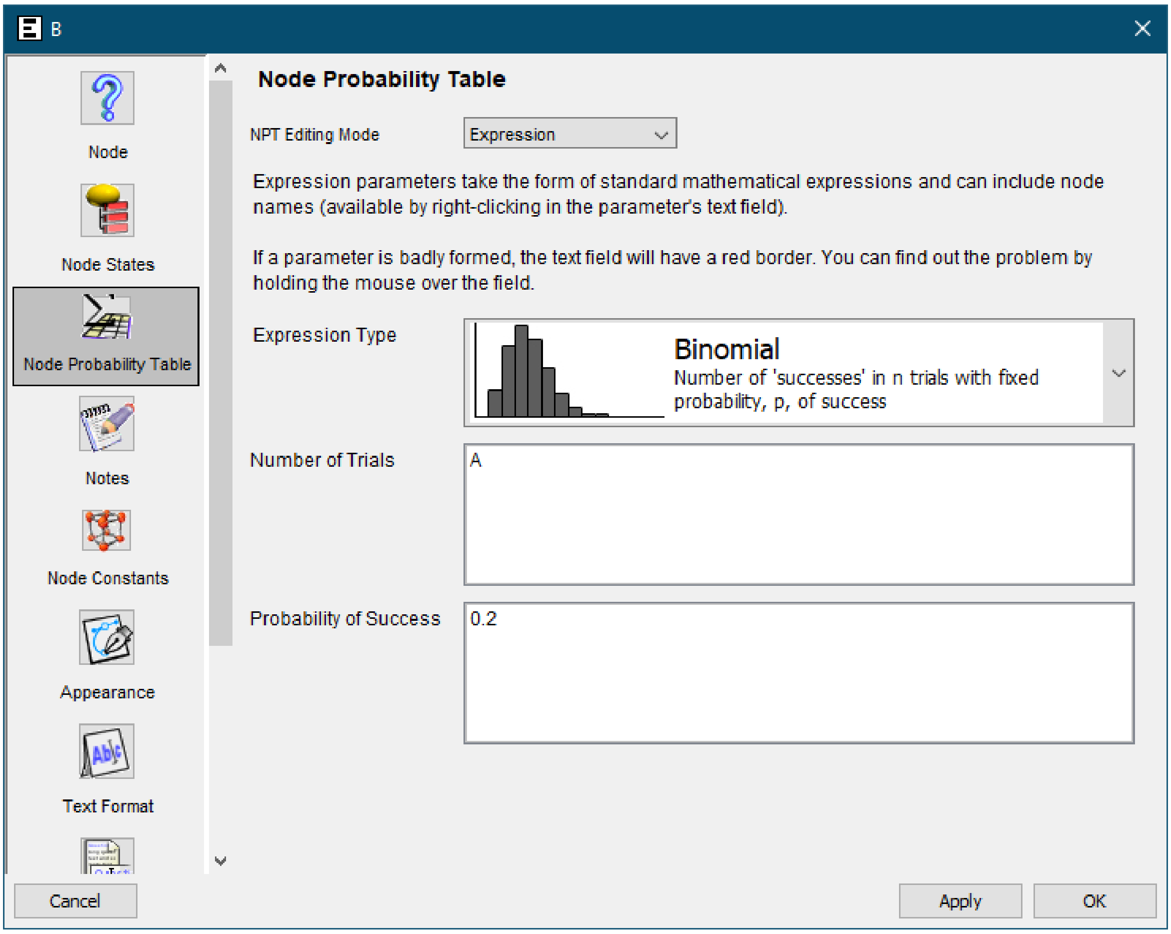


Figure 61 Chart for node A

12. Copy node A and after pasting it rename it node, B, representing ‘Customers lost’. Add a link from node A to node B.

13. Edit the node probability table of node B to be a Binomial distribution with parameters Number of Trials = A and Probability of Success = 0.2 as shown in Figure 62.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: Binomial — Number of 'successes' in n trials with fixed probability, p, of success

Number of Trials: A

Probability of Success: 0.2

Figure 62 Defining Binomial distribution for node B

14. Next copy node A and make node C. Make nodes A and B parents of C as shown in Figure 63.

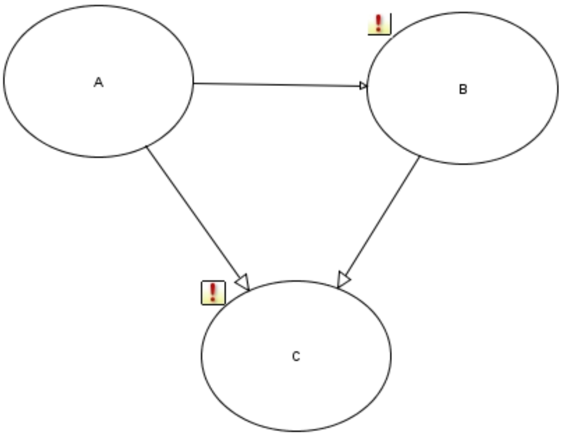
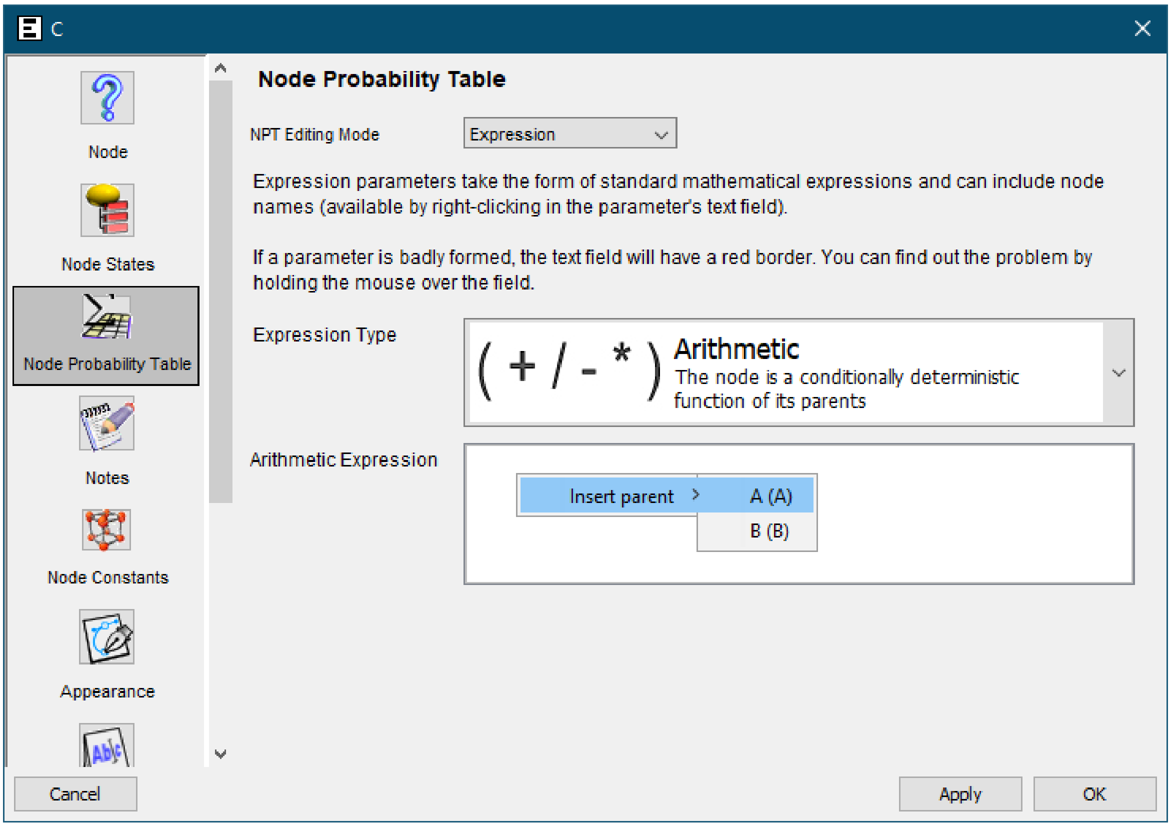


Figure 63 Graph with nodes A, B, and C

15. We now wish to ‘define’ the node probability table for C as the expression A-B. Select the Node Probability Table for node C and change the Expression Type to Arithmetic as shown in Figure 64. Then, if you right click in the text “Arithmetic Expression” you will be provided with a list of the parent nodes that you can select to be included in the expression as shown in Figure 64.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: (+ / - *) Arithmetic The node is a conditionally deterministic function of its parents

Arithmetic Expression

Figure 64 Setting the expression for C

16. Select parent A then input the “-“ symbol and then select parent B so that the arithmetic expression is “A-B”. You can, of course, enter this expression directly without selecting the

parent nodes, but when things get more complex, it is always advisable to select the parents to avoid mistakes.

17. You will probably see an error message warning you that the table is large.

ℹ	When entering an expression that involves parent nodes it always makes sense to use the ‘insert parent’ functionality since this will ensure the correct name of the parent is inserted. It also ensures that only valid variables are used in your expression.

18. Now calculate the model and open the graphs. You should see something like Figure 65.

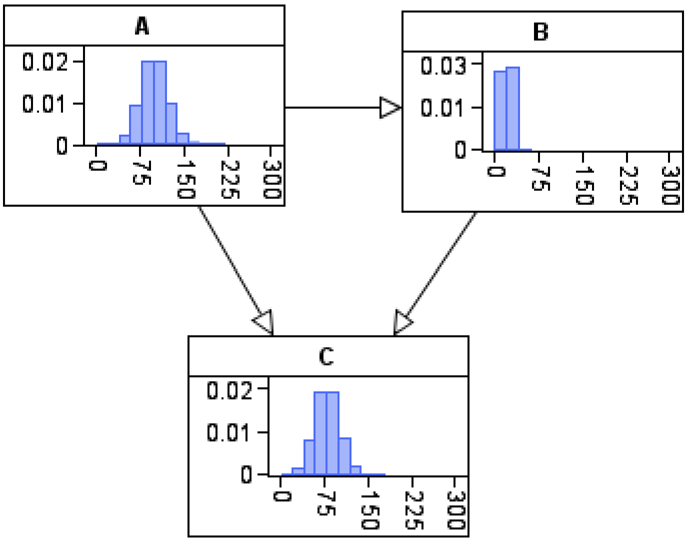
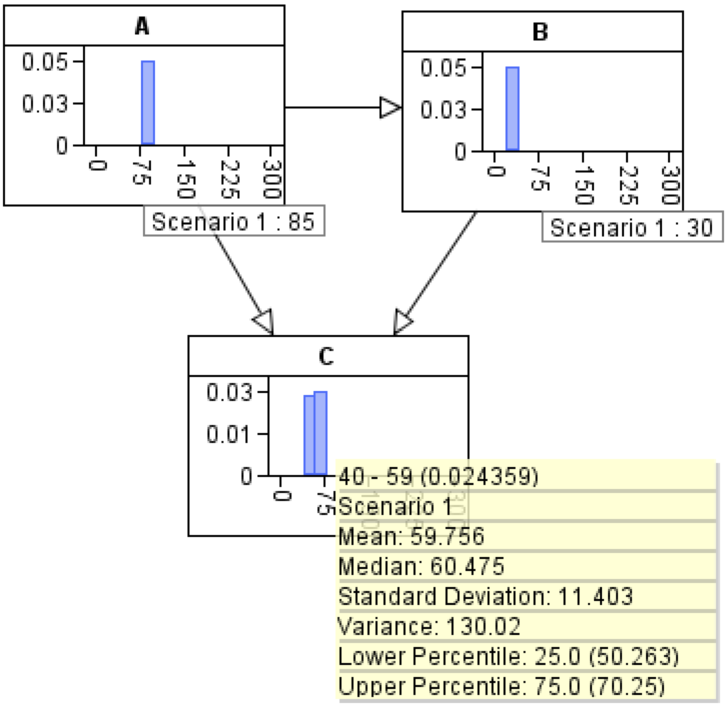


Figure 65 The calculated model

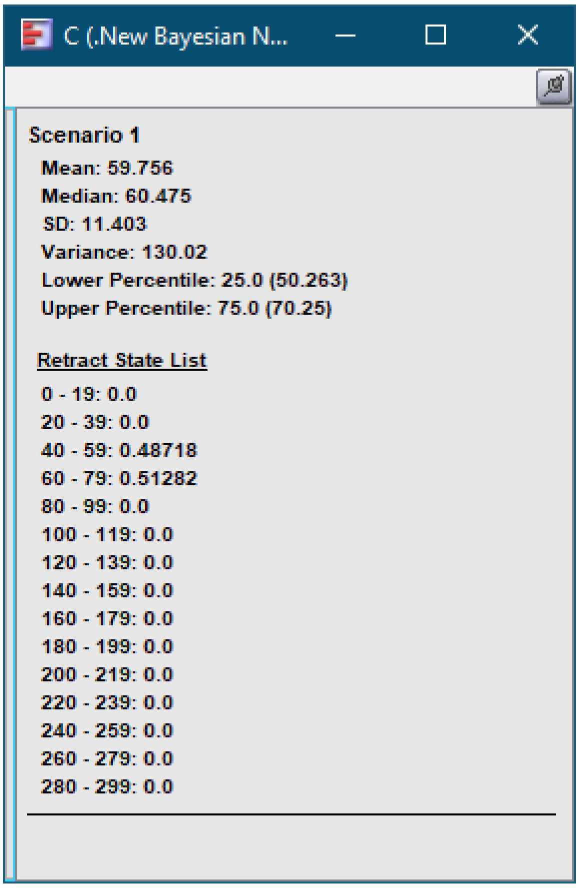
19. Now suppose in a particular year we know that A (the number of new customers) is **85** and B (the number of lost customers) is **30**. Enter these values as data into the nodes A and B respectively and run the calculation. You should see something like Figure 66. If you move the mouse over the graph for C you will see that the mean and median value for the node C is about 60. Hence, the model is calculating the result of 85-30 to be 60.



40 - 59 (0.024359)
Scenario 1
Mean: 59.756
Median: 60.475
Standard Deviation: 11.403
Variance: 130.02
Lower Percentile: 25.0 (50.263)
Upper Percentile: 75.0 (70.25)

Figure 66 Calculation with data entered

20. This approximation may or may not be satisfactory for your application. The reason this result is obtained is because when you enter the value 85 for node A all the model knows is that the value of A lies in the interval containing 85. In other words, the value for A is the interval 80-99. Similarly, the data 35 for node B simply means that the value of node B is the interval 20-39. When Agena.ai modeller runs the calculation, it uses the interval 80 to 99 for node A and the interval 20-39 for node B. For the expression A-B half of the results end up in the interval 40-59 and half end up in the interval 60-79. You can check this by right clicking on the node C and selecting to display summary statistics and then “expand state list” as shown in Figure 67.



Scenario 1

Mean: 59.756
Median: 60.475
SD: 11.403
Variance: 130.02
Lower Percentile: 25.0 (50.263)
Upper Percentile: 75.0 (70.25)

Retract State List

0 - 19: 0.0
20 - 39: 0.0
40 - 59: 0.48718
60 - 79: 0.51282
80 - 99: 0.0
100 - 119: 0.0
120 - 139: 0.0
140 - 159: 0.0
160 - 179: 0.0
180 - 199: 0.0
200 - 219: 0.0
220 - 239: 0.0
240 - 259: 0.0
260 - 279: 0.0
280 - 299: 0.0

Figure 67 Statistics and state list for node C

21. If you think that calculation is inaccurate, then you will be dismayed to discover that things can get much worse. Change the data for A to be **99** and the data of B to be **21**. Instead of the expected result 78 for node C, when you run the calculation, you will get the same result as before (mean 60). This is because 99 and 21 are in the same intervals as 85 and 30 respectively.

22. To increase accuracy, you can edit the set of states of each node to include smaller intervals. As an extreme you could make the interval size just 1 (or equivalently change the node type to be “Integer interval” and make the interval 1) for each of the nodes. When you do this the first thing you will notice is that the model starts to run very slowly. This is not surprising. Since the underlying NPT for node C must store values for every combination of the states this means 300*300*300 different values; that is table of 27 million values. This is simply impractical. Moreover, even if you could make such a model run you may discover that the range 0 to 299 was insufficient since in some years the values might get as high as 2000. You may even find that you cannot set a maximum number and wish to include the range 0 to infinity. Static discretization can suffer from having too many states that have low probability regions and too few states for high probability regions in the results. You will always be fighting a losing battle. The more states in a model the slower its execution and the more memory it demands.

23. At this point you might want to compare your model against the solution which can be found in Model Library / Tutorials / Dynamic discretization (look at the Bayesian network called “Static discretization).

24. Fortunately, although that is the approach you must adopt with other BN tools, agena.ai modeller has a very powerful, yet extremely simple, solution, covered in the next section.

5.2 Dynamic discretization using simulation nodes

1. You are going to create a new model with nodes A, B, and C just as before, but this time the nodes are going to use simulation nodes. To create a simulation node simply click on the simulation node icon:

2. Define the Node probability Table to be TNormal with Mean 100 and Variance 500 with upper and lower bounds 0 and 299 respectively.

3. As before, make two copies of node A calling them B and C. Add a link from A to B and declare the node probability table of B to be an expression: a Binomial distribution with parameters Number of Trials = A and Probability of Success = 0.2.

4. Note that this time around you do NOT need to add states to each of the nodes because there is no option to do so.

5. Add links from A to C and B to C. Define the NPT of node C to be the expression A-B.

6. Now run calculation and display the charts for all nodes. You should see something like Figure 68.

7. At this point you might want to compare your model against the solution which can be found in Model Library / Tutorials / Dynamic discretization (look at the Bayesian network called “Dynamic discretization).

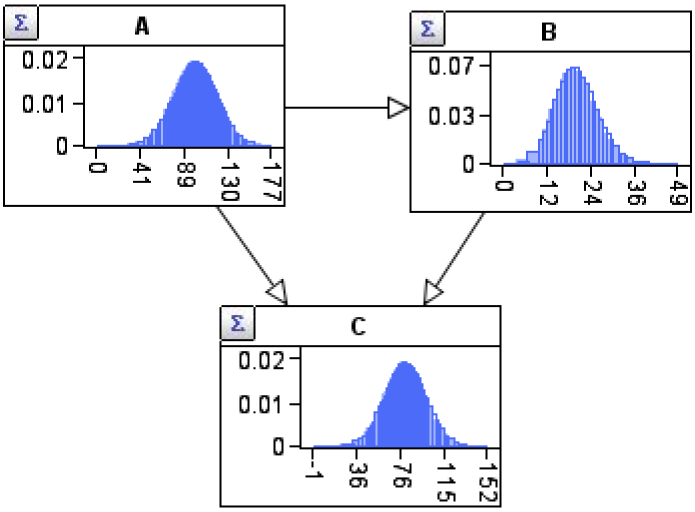
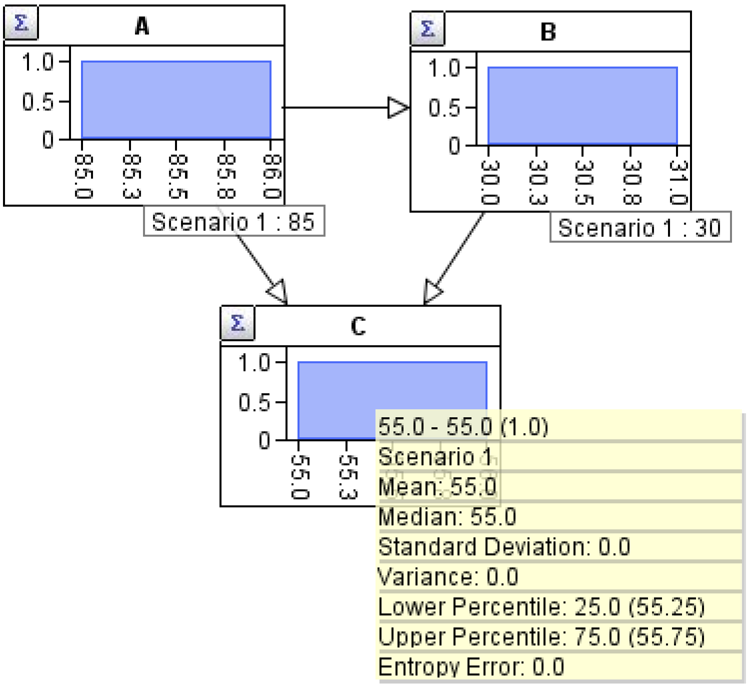


Figure 68 Model with simulation nodes and graphs displayed

8. Now enter the same data for nodes A and B (namely 85 and 30 respectively) and run the model. With the default graph settings, you should see something like Figure 69 (later we will show how to use more appropriate graph settings).



55.0 - 55.0 (1.0)
Scenario 1
Mean: 55.0
Median: 55.0
Standard Deviation: 0.0
Variance: 0.0
Lower Percentile: 25.0 (55.25)
Upper Percentile: 75.0 (55.75)
Entropy Error: 0.0

Figure 69 Result (with default chart settings)

9. The first thing to note is that when you move the mouse over Node C the mean and median have the same value 55 – which is exactly the result expected. Now enter the data 99 and 21 respectively for A and B and you will see that the mean and median for C are both 78, which again is exactly the expected result. Hence, not only is it simpler to build the model using simulation nodes (we do not have to worry about defining the discretization intervals) but we have solved the problem of inaccuracy.

10. There are several other features relating to simulation nodes that need to be explored. The first is that you may be surprised at the look of the charts when the data are entered and the calculated. In fact, they look like this because of the chosen defaults, but you have great flexibility over the way they are displayed.

11. Select the chart defaults property for node A. You will see a dialog as in Figure 70, but you will see that the box “Treat Min/Max X as percentiles” is checked and the Min X value is set as 0 and the Max X value is set as 100. What this means is that the whole chart will be displayed. However, if an data has been entered then all the probability mass is on a single point, at 55, and so the effect of zooming in on a percentile of this will be what you see in Figure 69. To see all the distribution (which makes sense in this case since it has a finite range) simply uncheck the box “Treat Min/Max X as percentiles” (as shown in Figure 70). This now charts the resulting distribution in a sensible x-axis range with Min X equal to zero and Max x equal to 100.

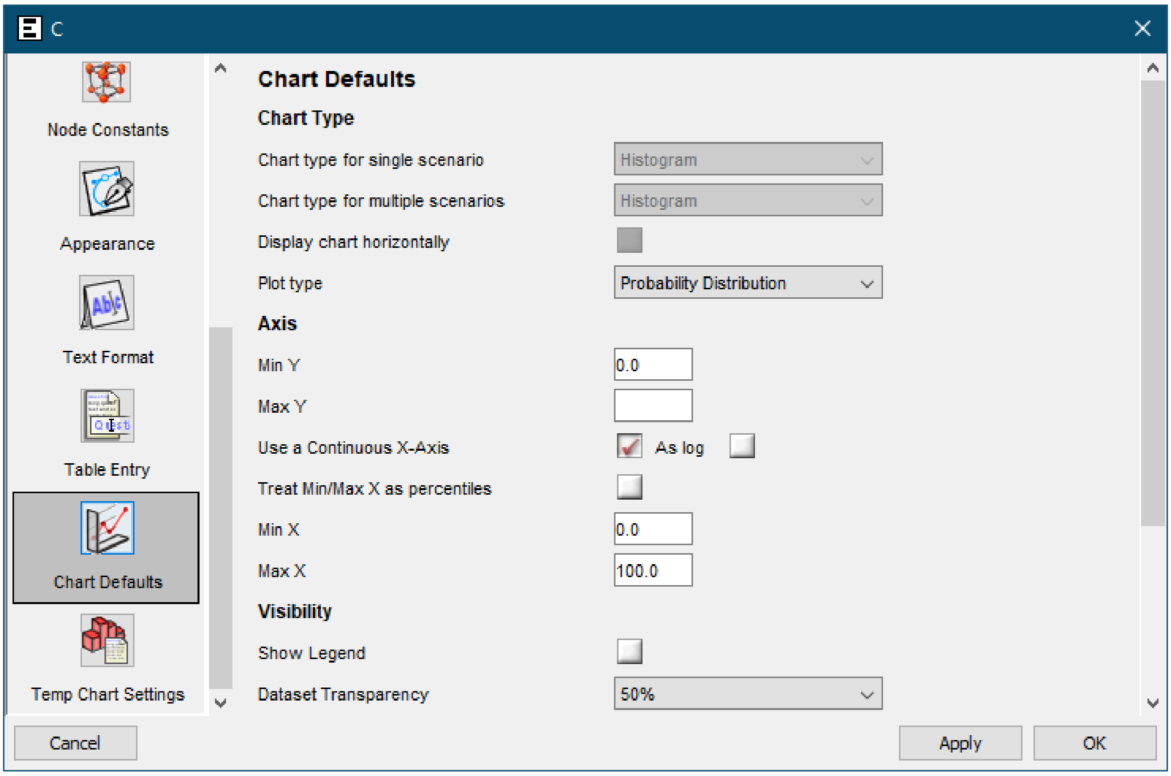


Chart Defaults

Chart Type

Chart type for single scenario: Histogram

Chart type for multiple scenarios: Histogram

Display chart horizontally ☐

Plot type: Probability Distribution

Axis

Min Y: 0.0

Max Y:

Use a Continuous X-Axis ☑ As log ☐

Treat Min/Max X as percentiles ☐

Min X: 0.0

Max X: 100.0

Visibility

Show Legend ☐

Dataset Transparency: 50%

Figure 70 Changing the chart defaults

12. Make the same changes to the graph defaults for nodes B and C. You will now see the revised graphs as shown in Figure 71.

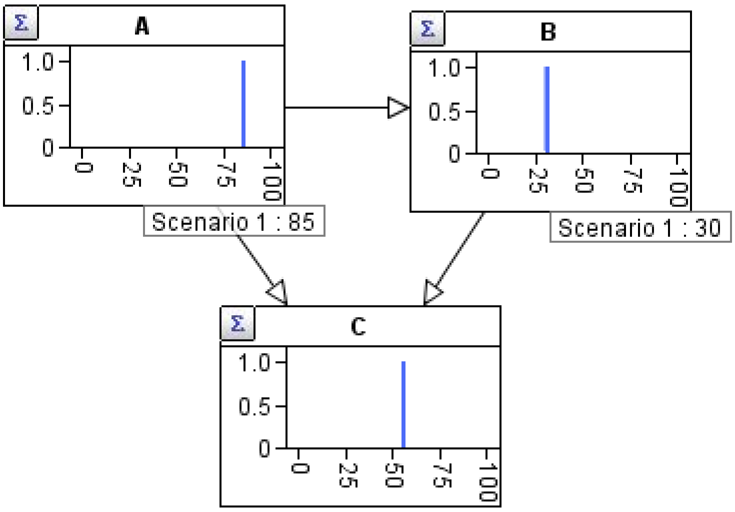


Figure 71 Results with revised graph defaults

13. Remove data (remember you can do these all at the same time using the Menu option **Tools→ Clear Entered Data → All**) and recalculate. You should see the revised charts shown in Figure 72.

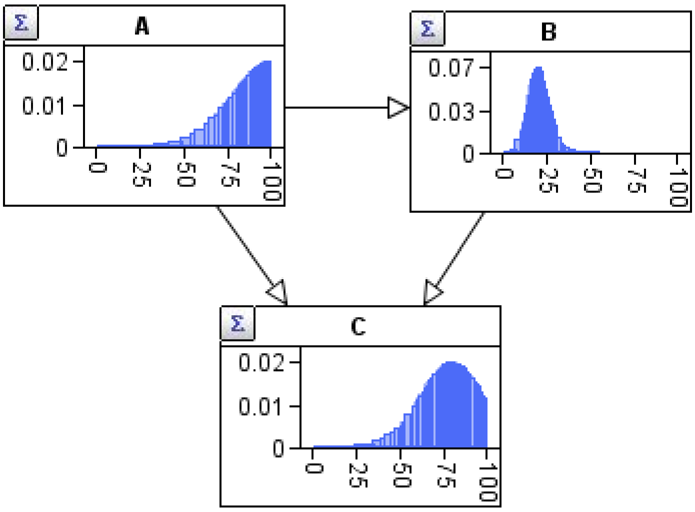


Figure 72 Data removed – partial distributions shown

14. Notice that the distributions on the charts appear truncated for nodes A and C. This is because we truncated the x-axis ranges to between zero and 100.If we revisit the chart defaults for each node and check the box "Treat Min/Max X as percentiles" the full range of the distributions will be shown as in Figure 73, the same as originally shown in Figure 68.

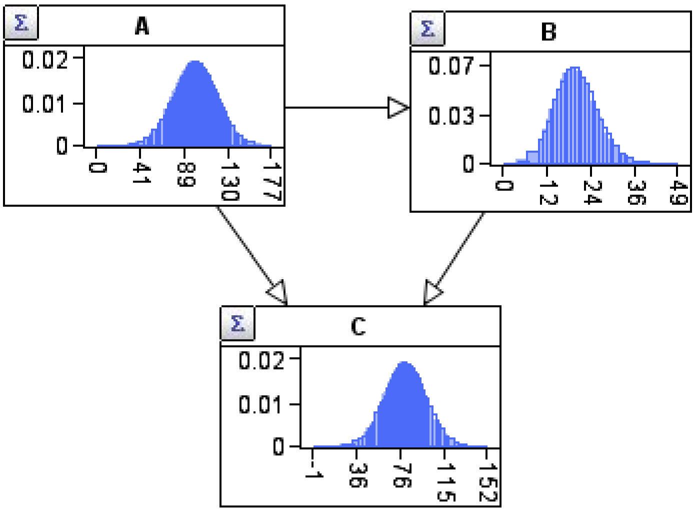
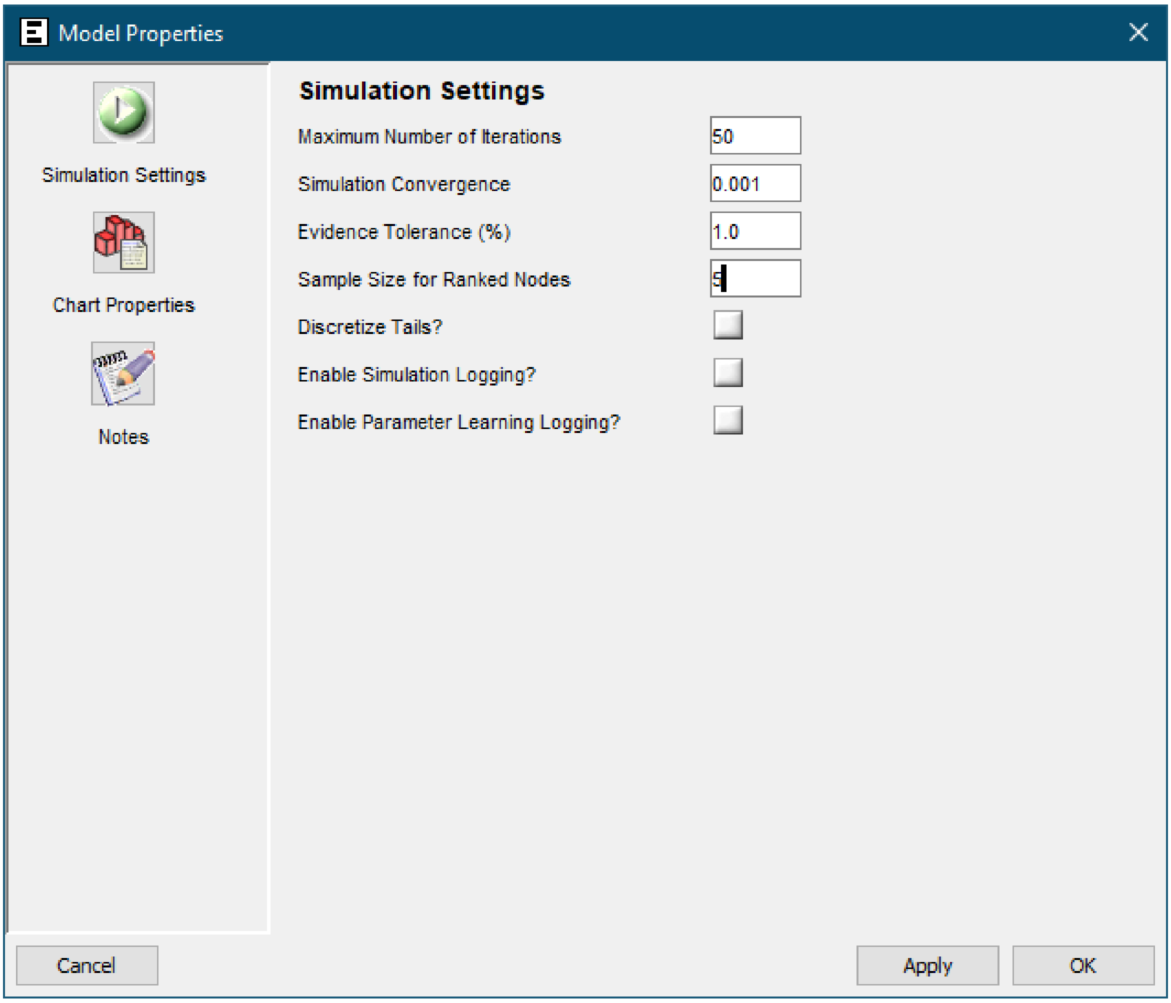


Figure 73 Full distributions restored

15. Now select the Model Properties icon

    from the toolbar menu. You should see the Simulation Settings dialog. Change the "Simulation convergence" to 0.1 (the default value is 0.001). Press **OK** and recalculate.



Simulation Settings

Maximum Number of Iterations 50

Simulation Convergence 0.001

Evidence Tolerance (%) 1.0

Sample Size for Ranked Nodes 5

Discretize Tails? ☐

Enable Simulation Logging? ☐

Enable Parameter Learning Logging? ☐

Figure 74 Simulation setting

16. The effect of increasing the simulation value is to decrease the accuracy of the model, which visibly results in ‘coarser’ distributions as shown in Figure 75.

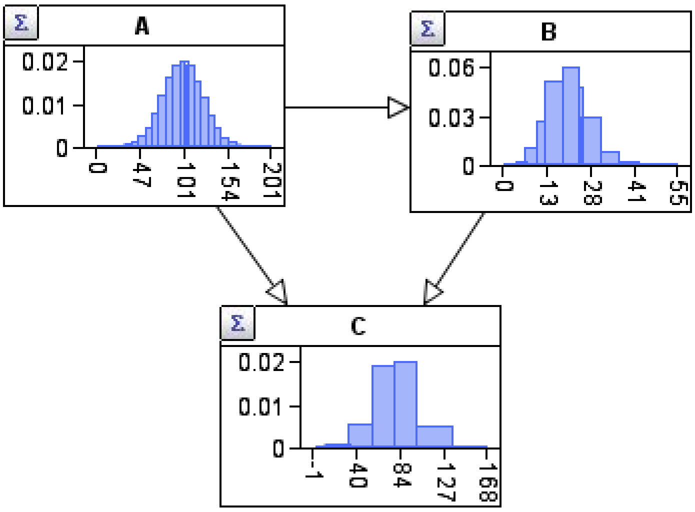


Figure 75 Coarser simulation resulting from simulation convergence change

17. In general, when you use simulation nodes you will need to consider the trade-off between accuracy and calculation speed. In static discretization we saw that there was a similar trade-off where accuracy was determined by the number of discretized states you used. It is much easier with simulation nodes to perform the trade-off since you only need to change one setting: simulation convergence.

18. As explained in Section 4.1.8 the default simulation convergence setting is 0.001. By default, this applies to every simulation node in the model. However, as explained there you can change the setting for each individual node

⚠ For large models with many simulation nodes, we strongly recommend that, while developing and testing the model, you set the simulation convergence to a high value. We recommend 0.1 or 0.01. Only when the model is complete, and you need the highest accuracy in the results should you make the value lower. This makes for fast model development.

6. Importing/Exporting data

You can export data (data) from your models to CSV (Comma Separated Value) files. These can then be opened in any spreadsheet or CSV supported application.

You export by right clicking on any object inside the explorer, selecting **Export Cases** and then the case from which the data should be exported, as shown below. Right clicking on Model or the top-most object in the explorer will export the whole model. Right clicking on individual Bayesian networks will only export that Bayesian network.

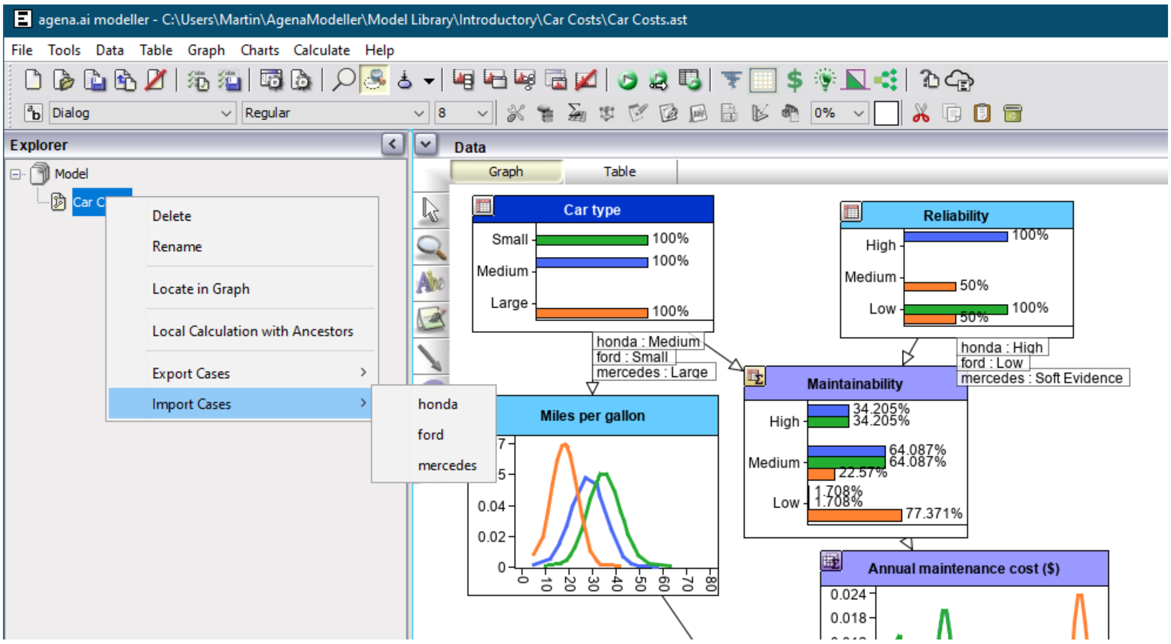
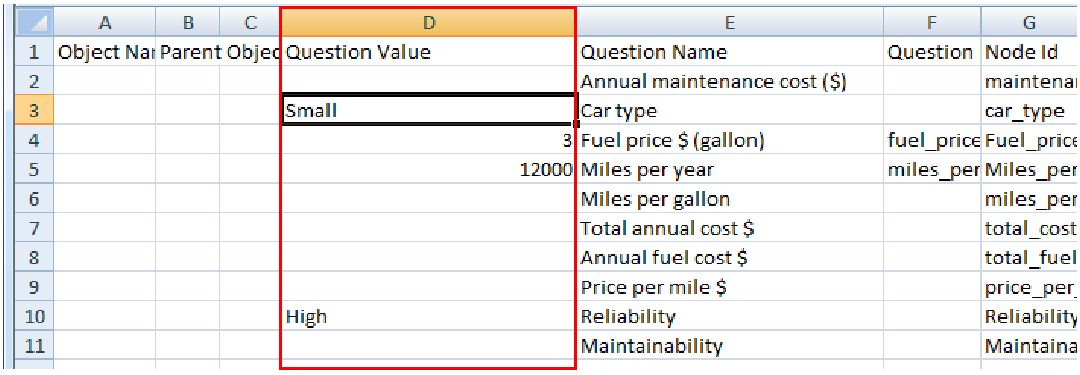


Figure 76 Exporting data for Bayesian network to .json .xml .csv formats

A dialog will open which allows you to then specify the filename that you wish to save the data file as. The choice is .json, .sml and .csv formats. If you exported as .csv file You can then open the created CSV file in Excel (or any other CSV supporting application)



	A	B	C	D	E	F	G
1	Object Nai	Parent Objec	Question Value		Question Name	Question	Node Id
2					Annual maintenance cost ($)		maintenai
3				Small	Car type		car_type
4				3	Fuel price $ (gallon)	fuel_price	Fuel_price
5				12000	Miles per year	miles_per	Miles_per
6					Miles per gallon		miles_per
7					Total annual cost $		total_cost
8					Annual fuel cost $		total_fuel
9					Price per mile $		price_per_
10				High	Reliability		Reliability
11					Maintainability		Maintaina

Figure 77 Editing exported data file in a spreadsheet

Each row represents a question/node and its associated value. The fourth column is the value and is the column that you should edit. The fifth column is the question name and the sixth, if not empty, is any constant associated with the question.

Once you have edited your CSV data file you can then import it into your model once again. You can do this by right clicking on any object in the Explorer, selecting **Import Data file (.CSV)** again as shown below.

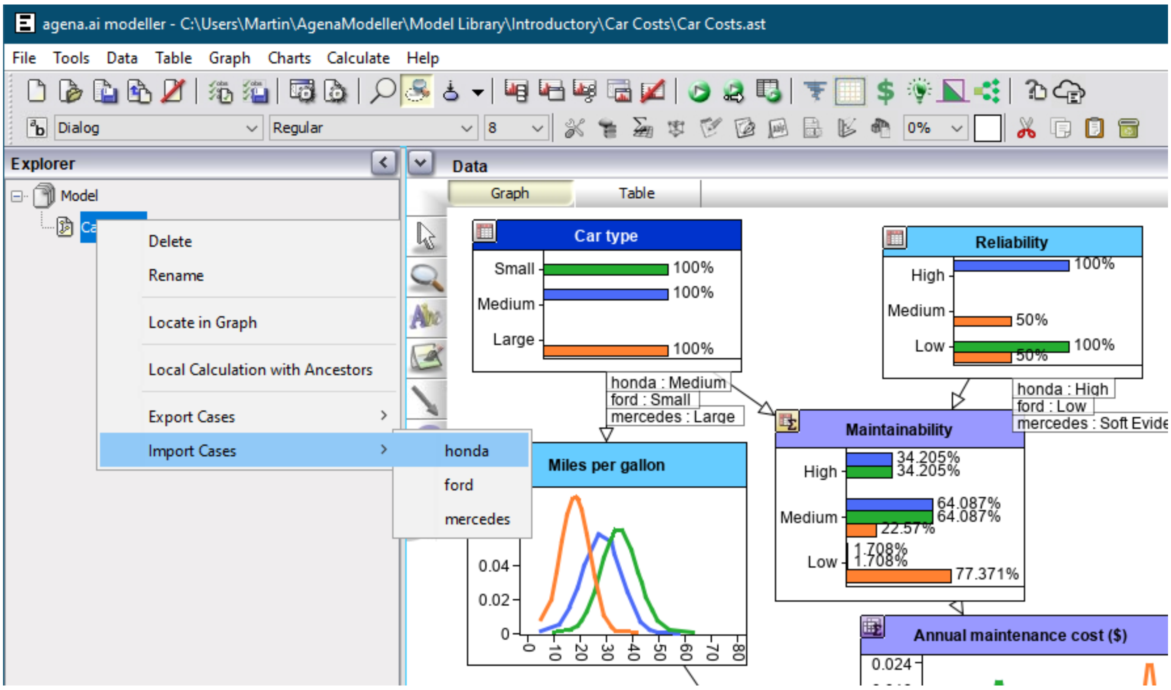
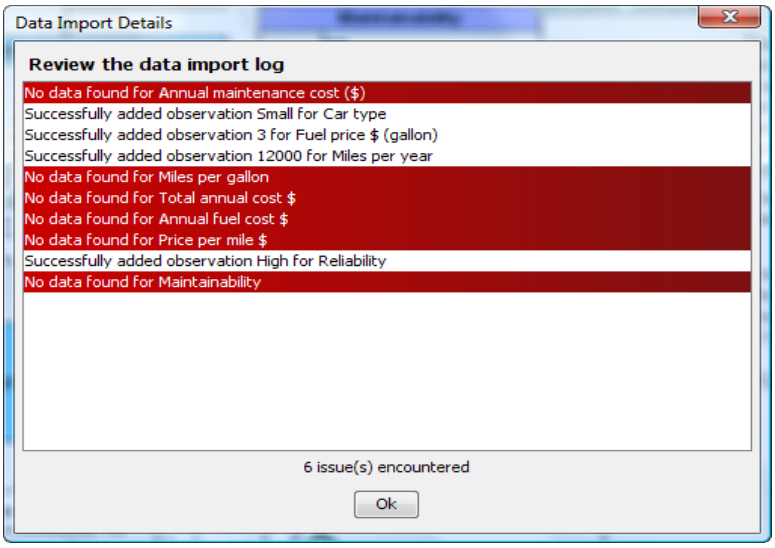


Figure 78 Importing data file into your model

Upon clicking the case into which you wish to import you will be presented with a dialog for you to browse and select your data file.

The data file will then be imported into your model, for the specific case selected, and you will be presented with a log of the data import as shown below.



```
No data found for Annual maintenance cost ($)
Successfully added observation Small for Car type
Successfully added observation 3 for Fuel price $ (gallon)
Successfully added observation 12000 for Miles per year
No data found for Miles per gallon
No data found for Total annual cost $
No data found for Annual fuel cost $
No data found for Price per mile $
Successfully added observation High for Reliability
No data found for Maintainability
```

Figure 79 Data Import Log highlight issues

This log will highlight all issues found with the data file in red. Most of these issues are completely harmless – notably nodes for which there are no data. Please note that the level that you exported at is the same level you must import at. If you exported by right clicking on a Bayesian network you must then import by right clicking on that Bayesian network. Similarly, if you right clicked on the model object to export, you must right click on the model object to import.

ℹ	Note that data can also be imported/exported using the agena.ai modeller proprietary data format, via the toolbar buttons: 

7. Using the Sensitivity Analyser

This tutorial demonstrates how to use the sensitivity analyser, using two of the example models. In the first we will demonstrate the sensitivity analyser with discrete nodes and in the second we will demonstrate the more powerful features of the sensitivity analyser using continuous nodes.

Note that to ensure a fast computational response sensitivity analysis uses a fixed discretisation for continuous nodes, regardless of the evidence, and this provides less accurate results than those obtainable when running the model for each point of evidence you may be interested in.

7.1 Sensitivity analysis for discrete nodes

1. Open the example model *Asia.ast* by following the same instructions as before.

2. There are two ways to launch the sensitivity analyser, as shown in Figure 80.

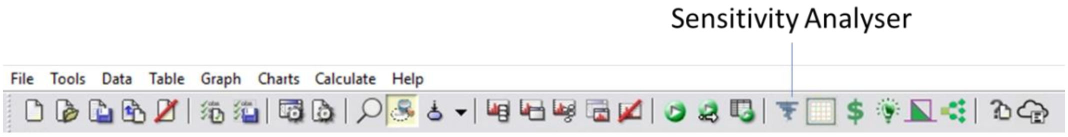
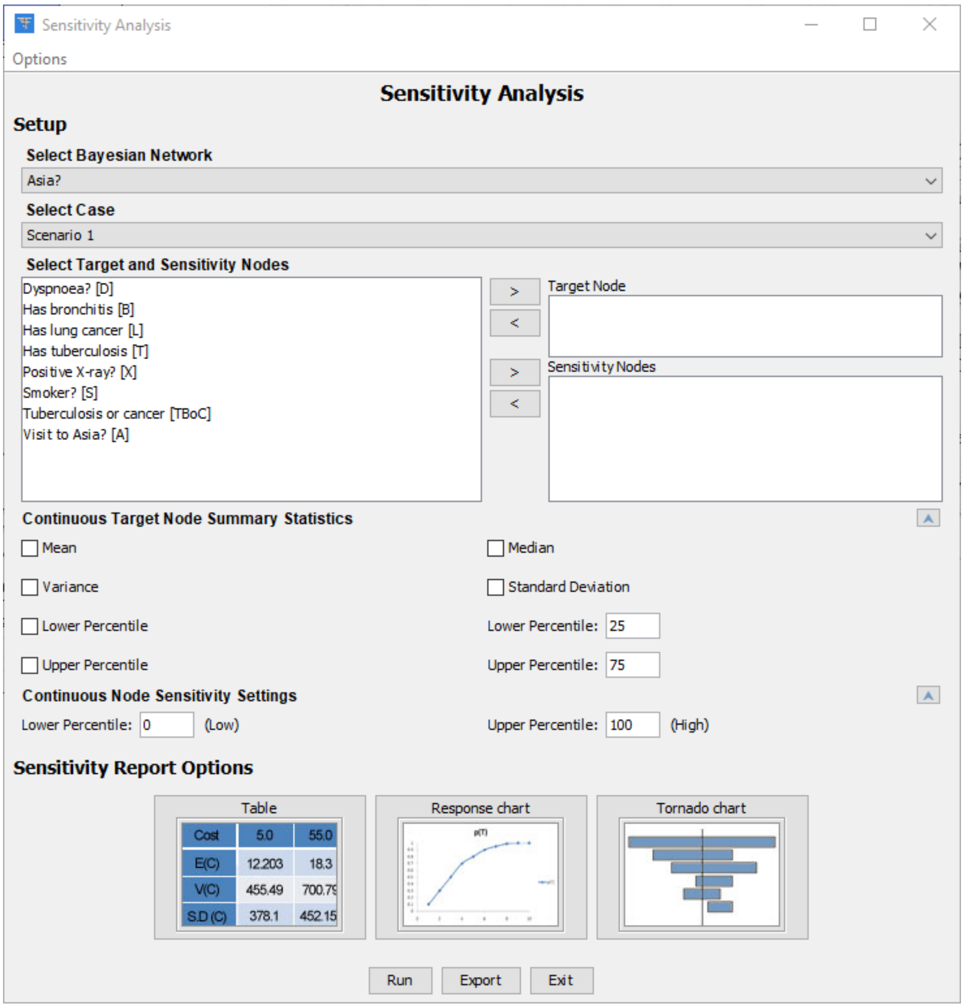


Figure 80 Accessing the sensitivity analyser

Specifically, you can do so either via the Tools menu or, more quickly, click the toolbar menu icon

3. You will then be presented with the sensitivity analysis main screen as shown in Figure 81.



Sensitivity Analysis

Options

Sensitivity Analysis

Setup

Select Bayesian Network: Asia?

Select Case: Scenario 1

Select Target and Sensitivity Nodes

Dyspnoea? [D]
Has bronchitis [B]
Has lung cancer [L]
Has tuberculosis [T]
Positive X-ray? [X]
Smoker? [S]
Tuberculosis or cancer [TBoC]
Visit to Asia? [A]

Target Node

Sensitivity Nodes

Continuous Target Node Summary Statistics

☐ Mean ☐ Median
☐ Variance ☐ Standard Deviation
☐ Lower Percentile Lower Percentile: 25
☐ Upper Percentile Upper Percentile: 75

Continuous Node Sensitivity Settings

Lower Percentile: 0 (Low) Upper Percentile: 100 (High)

Sensitivity Report Options

Figure 81 Sensitivity analysis main screen

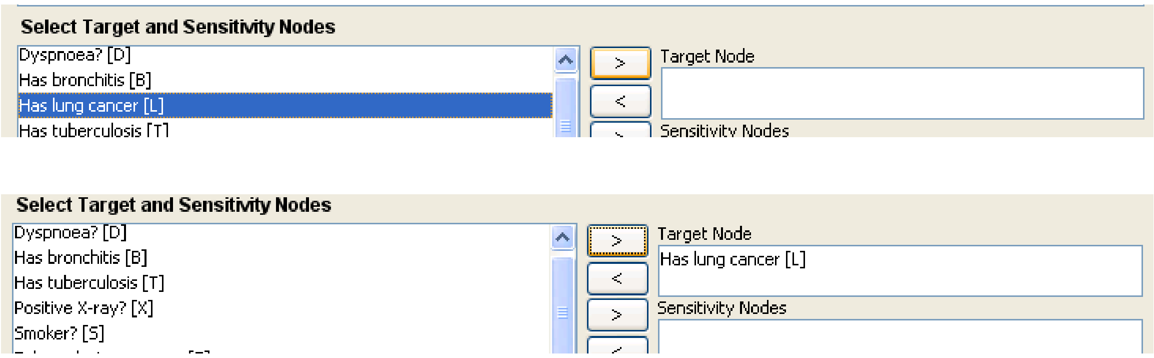
4. In any sensitivity analysis you must select **a single target node** and one or more sensitivity nodes. The objective of the analysis is to get a visual representation of the impact of different sensitivity nodes on the selected target nodes.

	In this example the model only has a single case. When there are multiple cases you can select any combination of them to include in the sensitivity analysis. By default, only the first case is selected. To select or deselect additional cases simply press Ctrl + Click

5. Select "Has lung cancer (L)" as the target node and click on the top arrow

   \>

   as shown in Figure 82



Select Target and Sensitivity Nodes

Dyspnoea? [D]
Has bronchitis [B]
Has lung cancer [L]
Has tuberculosis [T]

Target Node

Sensitivity Nodes

Select Target and Sensitivity Nodes

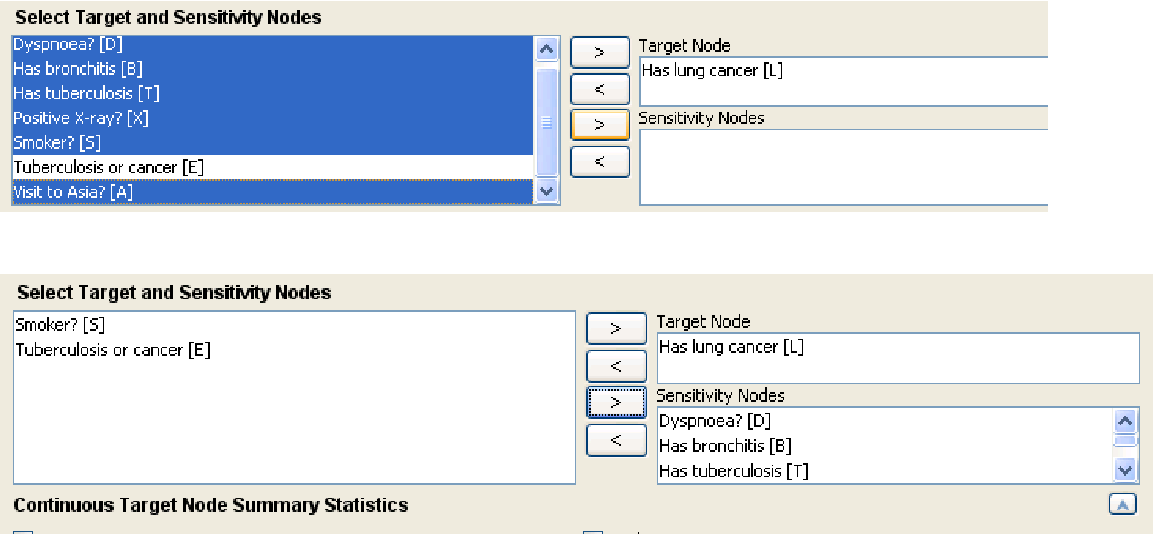
Dyspnoea? [D]
Has bronchitis [B]
Has tuberculosis [T]
Positive X-ray? [X]
Smoker? [S]

Target Node: Has lung cancer [L]

Sensitivity Nodes

Figure 82 Selecting the target node

6. Now select the nodes shown in Figure 83 as sensitivity nodes (you select multiple nodes by clicking on a node while holding the Ctrl key) and click the ***third*** arrow down so that the chosen sensitivity nodes are moved as shown in Figure 83.



Select Target and Sensitivity Nodes

Dyspnoea? [D]
Has bronchitis [B]
Has tuberculosis [T]
Positive X-ray? [X]
Smoker? [S]
Tuberculosis or cancer [E]
Visit to Asia? [A]

Target Node: Has lung cancer [L]

Sensitivity Nodes

Select Target and Sensitivity Nodes

Smoker? [S]
Tuberculosis or cancer [E]

Target Node: Has lung cancer [L]

Sensitivity Nodes:
Dyspnoea? [D]
Has bronchitis [B]
Has tuberculosis [T]

Continuous Target Node Summary Statistics

Figure 83 Selecting sensitivity nodes

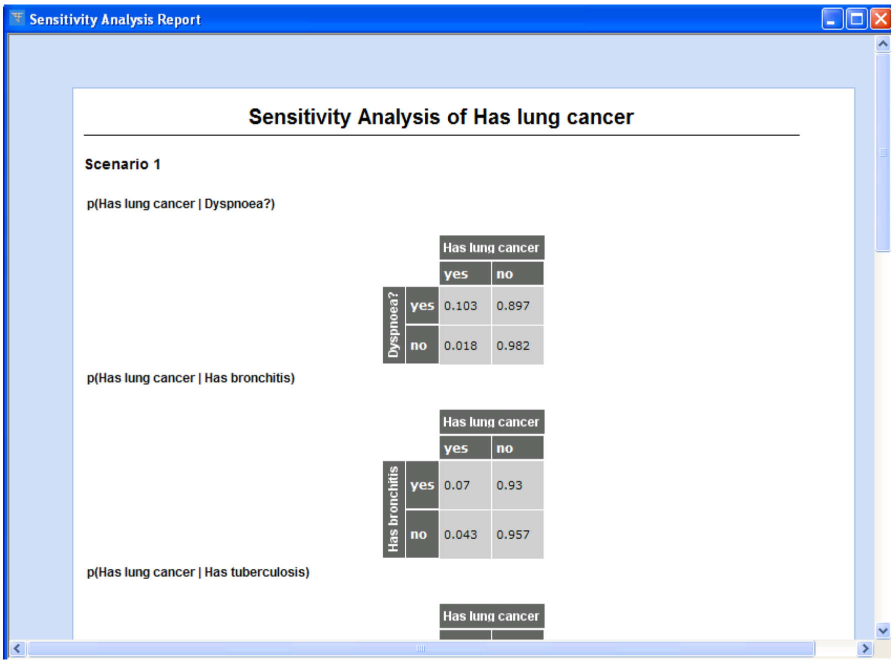
7. Before generating the results you have to specify what kind of results you would like. Since all the nodes in this model are discrete, you can ignore the set of options that deal with continuous nodes. Hence you only need to select the type of report to generate. For this example select both the "Table" and "Tornado Graph" options by clicking on them. This will result in large ticks as shown in Figure 84.

Sensitivity Report Options

Figure 84 Selecting report options

8. You are now ready to generate the sensitivity analysis results by simply clicking the “Run” button.

9. This will produce an html report (in your default web browser or in your agena.ai modeller directory called SensitivityAnalysis.html) as shown in Figure 85.



Sensitivity Analysis of Has lung cancer

Scenario 1

p(Has lung cancer | Dyspnoea?)

		Has lung cancer	
		yes	no
Dyspnoea?	yes	0.103	0.897
	no	0.018	0.982

p(Has lung cancer | Has bronchitis)

		Has lung cancer	
		yes	no
Has bronchitis	yes	0.07	0.93
	no	0.043	0.957

p(Has lung cancer | Has tuberculosis)

Has lung cancer

Figure 85 Sensitivity analysis report (tables)

10. You can set the report to display in your systems default browser by clicking on the menu Options→Report Settings and ensuring that the box labelled ‘View report in internal browser’ is not checked.

11. The first part of the report shows the tables. Specifically, for each of the selected sensitivity nodes a table is produced which shows the probability of the target node given the sensitivity node.

12. Scroll down and you will come to the tornado graphs as shown in Figure 86.

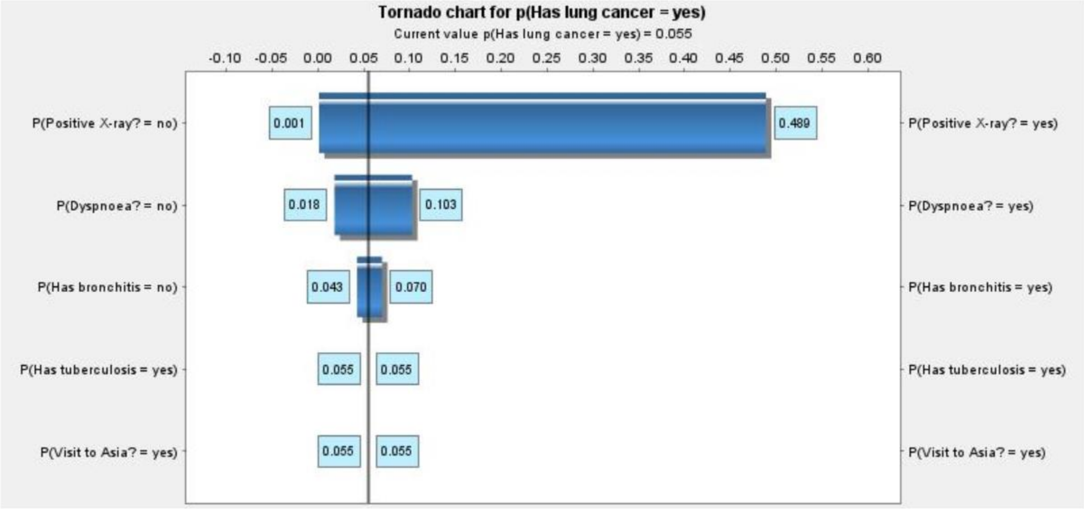
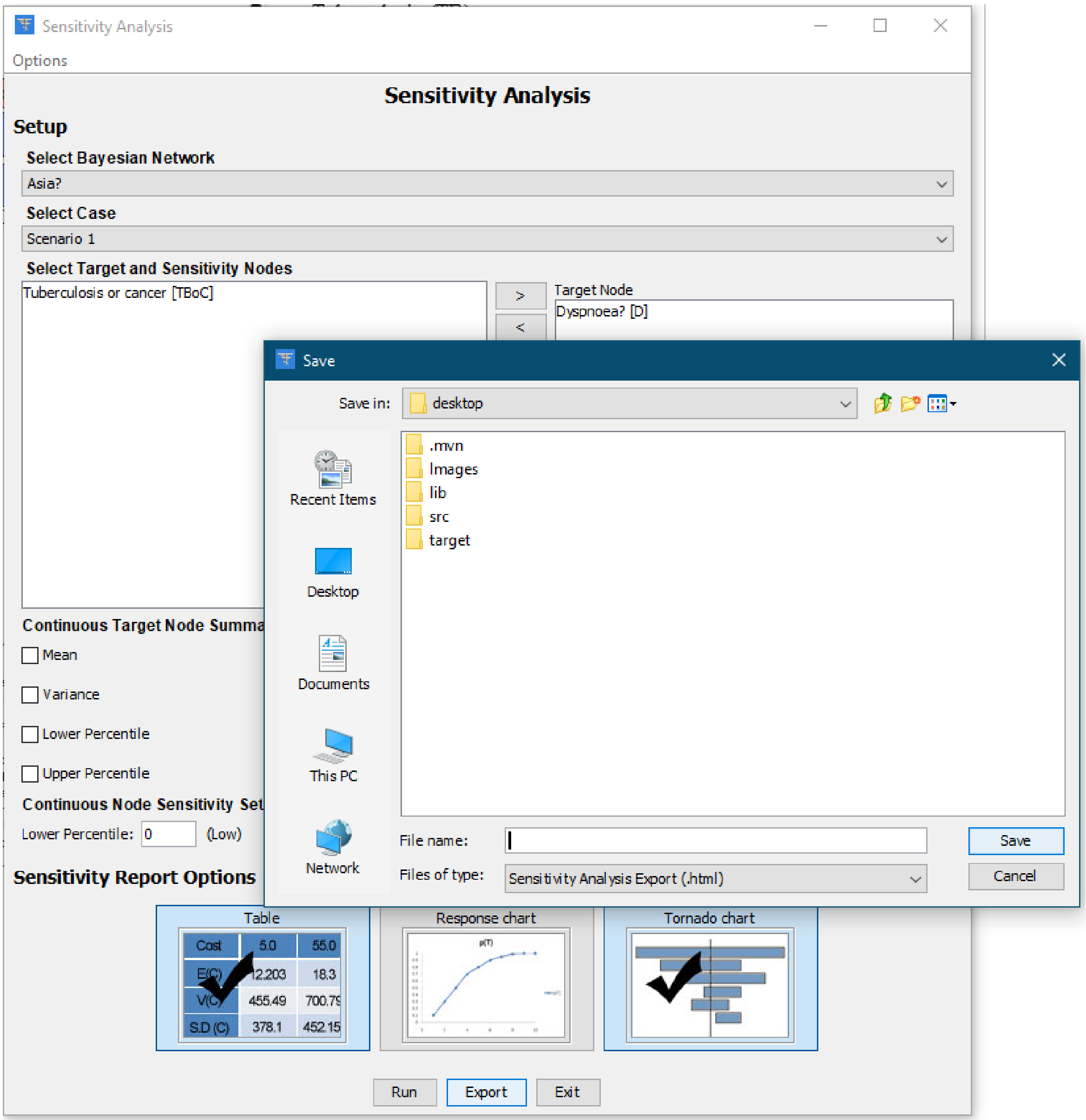


Figure 86 Sensitivity analysis report (tornado graphs)

13. From a purely visual perspective you can think of the length of the bars corresponding to each sensitivity node in the tornado graph as being a measure of the impact of that node on the target node. Thus, the node “Positive X-ray“ has by far the most impact on lung cancer with “Dyspnoea“ and “Has bronchitis“ as poor second and third respectively, and the others having negligible impact.

14. In fact, the bars have a formal interpretation, which comes from the tables. You will see, for example, that the probability of lung cancer given the result of positive X-ray goes from 0.001 (when positive X-ray is *no*) to 0.489 (when positive X-ray is *yes*). This range (from 0.001 to 0.489) is exactly the bar that is plotted for the tornado graph.

15. To export the results of the sensitivity analysis simply click the ‘export’ button at the bottom of the main sensitivity analysis screen. This brings up a file dialog as shown in Figure 87.



Select Bayesian Network: Asia?

Select Case: Scenario 1

Tuberculosis or cancer [TBoC]

Target Node: Dyspnoea? [D]

Lower Percentile: 0 (Low)

Cost	5.0	55.0
E(C)	12.203	18.3
V(C)	455.49	700.79
S.D (C)	378.1	452.15

Save in: desktop

File name:

Files of type: Sensitivity Analysis Export (.html)

Figure 87 Exporting sensitivity analysis results

16. Type in the name “sens_report”. Then the following will be saved in the selected directory:

    - A file called *sens_report.html*
    - A folder called *Images*

17. If you subsequently double-click the file *sens_report.html* file in Windows, the full sensitivity analysis report will appear in your default web browser. The jpg images are in the folder *saimages*.

7.2 Sensitivity Analysis for continuous nodes

1. Open the example model called *Software Project Risk.ast* in the folder:

   Model Library \Advanced\Software Project Risk

2. Note that this is a large model with many nodes of different types.

3. Select the sensitivity analysis tool, accessible using this icon on the toolbar .

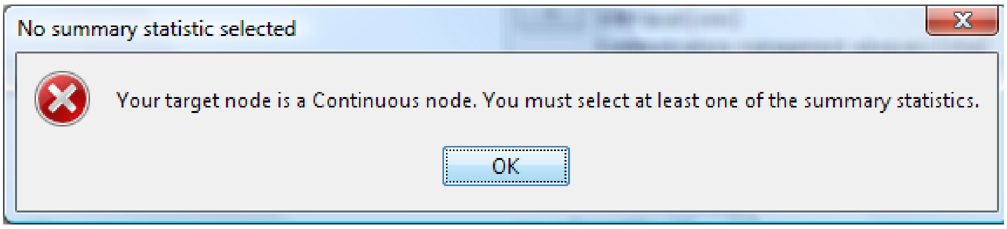
4. Select the node “defects per KLOC post release’ as the target node (note that the nodes are listed alphabetically so it is easy to scroll down to find the node you need). To understand the example better it is useful to note that the number of defects per KLOC post release is a standard measure of software quality (the lower the value the higher the delivered quality).

5. Select all the remaining nodes as Sensitivity nodes.

6. Select the Response curve and Tornado graph as sensitivity report options.

7. Click **Run**

8. Although this was how you set up the sensitivity analysis report in the previous example, you now see an error message as shown in Figure 88. This is because the target node is continuous.



No summary statistic selected

Your target node is a Continuous node. You must select at least one of the summary statistics.

Figure 88 Error message for continuous nodes

9. Because the target node is continuous you need to choose at least one summary statistic on which to perform the sensitivity analysis. Select the **mean** and click run,

10. After a few seconds a long report will be generated. Near the top of the report you should see the graph shown in Figure 89.

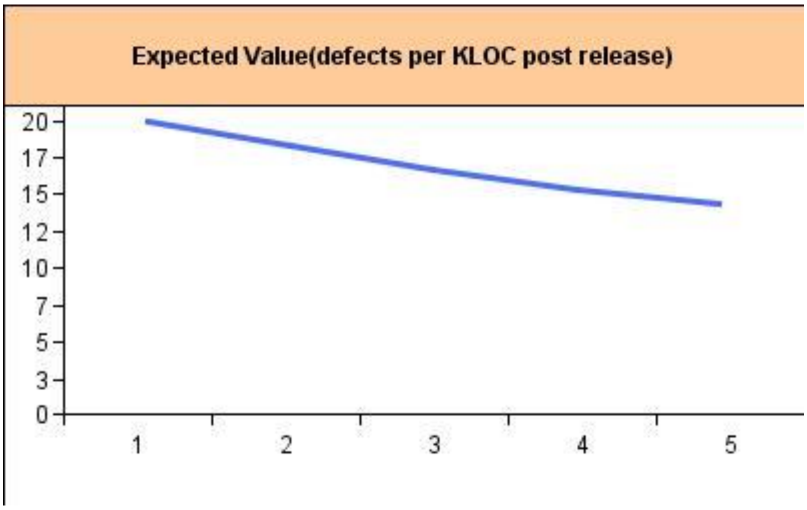


Figure 89 Response graph for target node given CMM level node

11. The node “CMM level“ is actually a ranked node with states 1,2,3,4,5 (CMM is an internationally recognised ranked score measure of process maturity for software organisations; level 5 organisations are considered to be the most mature). What the graph actually means is that, ignoring all other factors, the mean number of defects per KLOC drops fairly significantly as the

CMM level increases. Compare this response graph with the next one in the list (requirements complexity). Here the change in the mean value of defects per KLOC is much less ‘steep’ as the requirements complexity goes from ‘very high’ to ‘very low’.

12. Now scroll down to the bottom of the report. You will see a very compressed Tornado graph as shown in Figure 90. Fortunately, we change the default graph settings to improve this output.

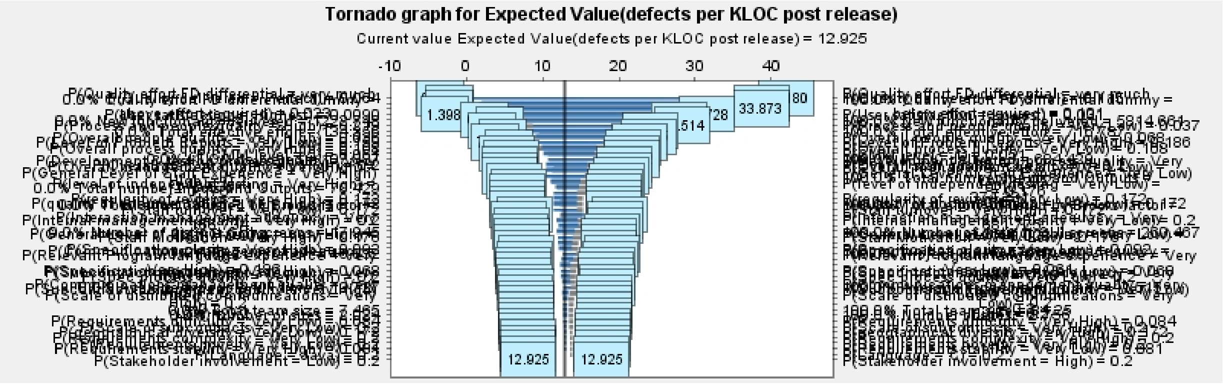


Figure 90 Overly compressed Tornado graph

13. Close the generated report and in the sensitivity analysis dialog window select the menu Options→Chart Settings as shown in Figure 91.

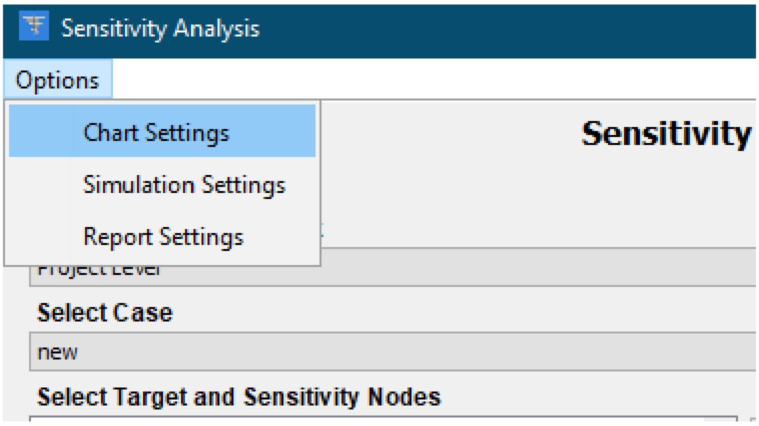


Figure 91 Choosing the chart settings option

14. The dialog in Figure 92 will appear. Change the values for Tornado Width and Height to the ones shown in Figure 92.

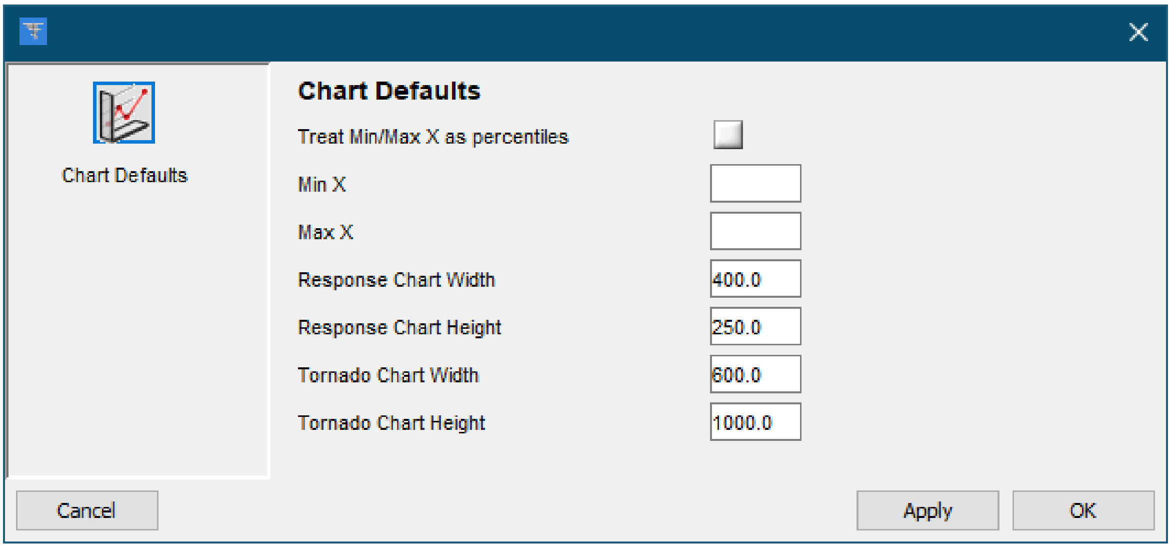


Chart Defaults

Treat Min/Max X as percentiles ☐

Min X

Max X

Response Chart Width 400.0

Response Chart Height 250.0

Tornado Chart Width 600.0

Tornado Chart Height 1000.0

Figure 92 Chart settings dialog

15. Run the report again and scroll down to the bottom. The Tornado graph should now look like the one shown in Figure 93.

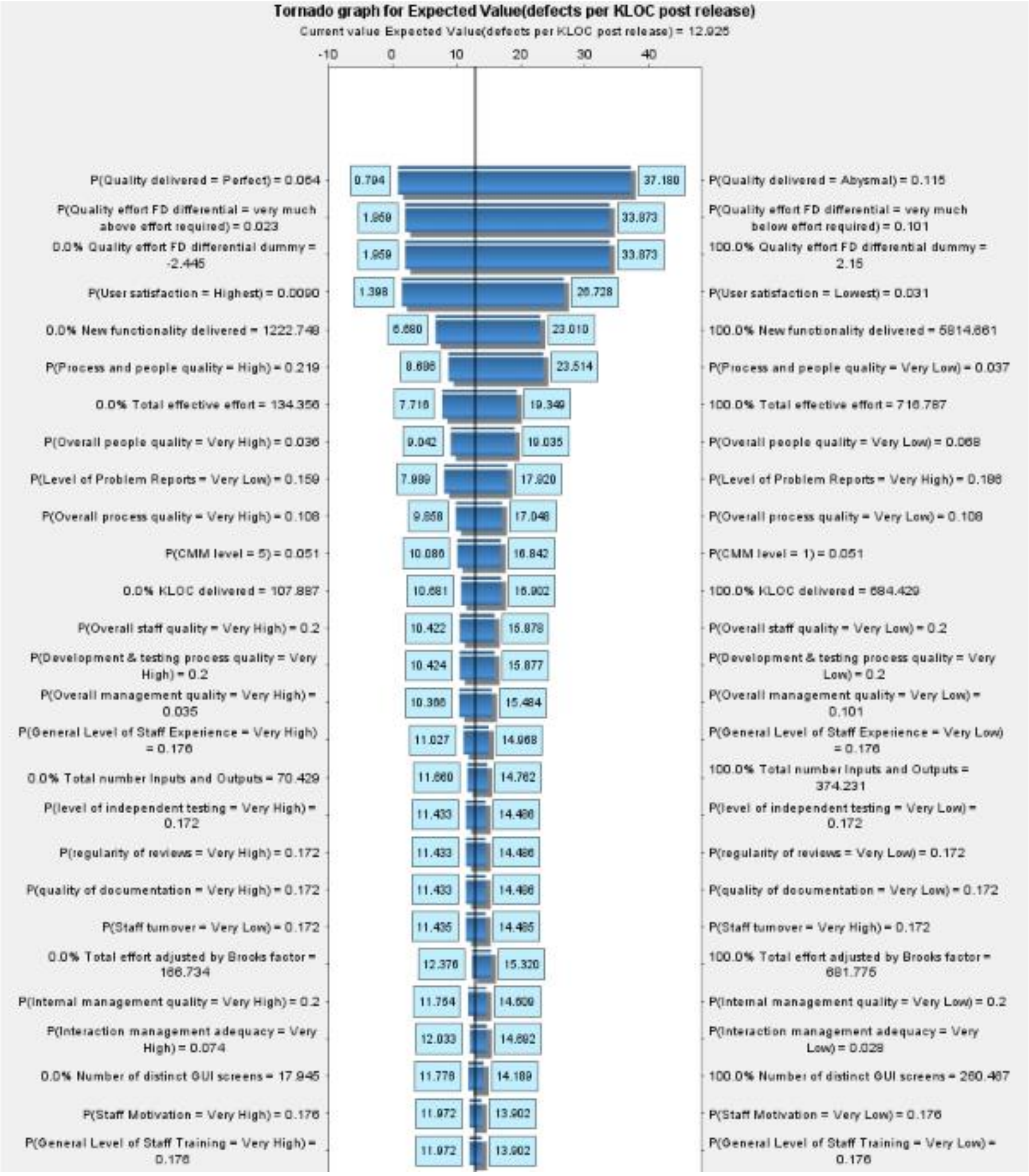


Figure 93 Improved Tornado graph after graph settings changed

16. Without closing the sensitivity analysis window, go back into the model and enter the data **100** for the node “New functionality delivered“ as shown in Figure 94.

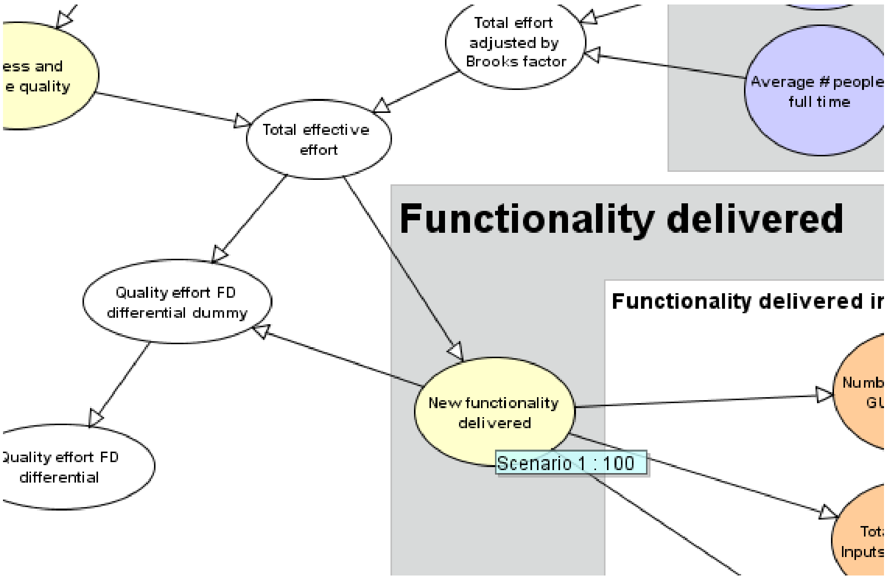


Figure 94 Entering an data

17. When you do this you will see that the target and sensitivity nodes will be reset in the sensitivity analysis window. Make the same selections as before and rerun the report. By comparison with the previous report look at the response curve for CMM level, Figure 95.

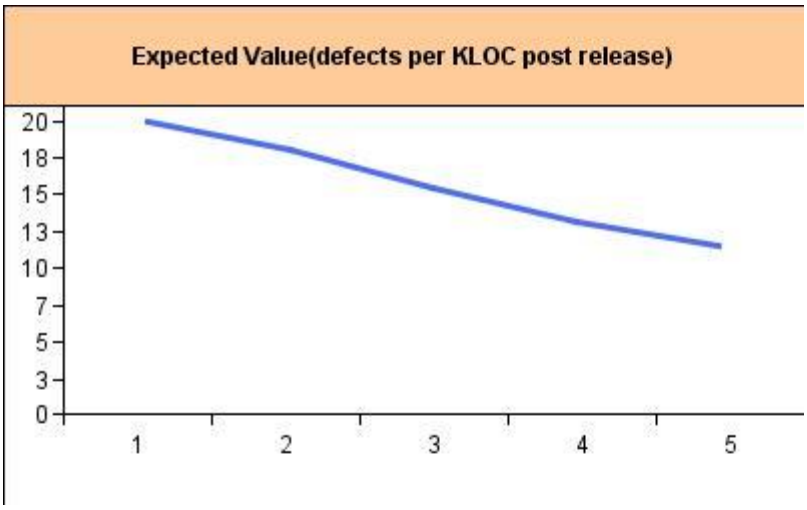


Figure 95 Revised response curve for CMM level

18. Note that the mean values for defects per KLOC has changed significantly (they are lower because of the data entered for functionality delivered) and the curve is now even steeper than before. This suggests that CMM level has an even greater impact on the target node for this level of functionality.

19. You should also compare the resulting Tornado graph with the previous one.

8. Multivariate analysis

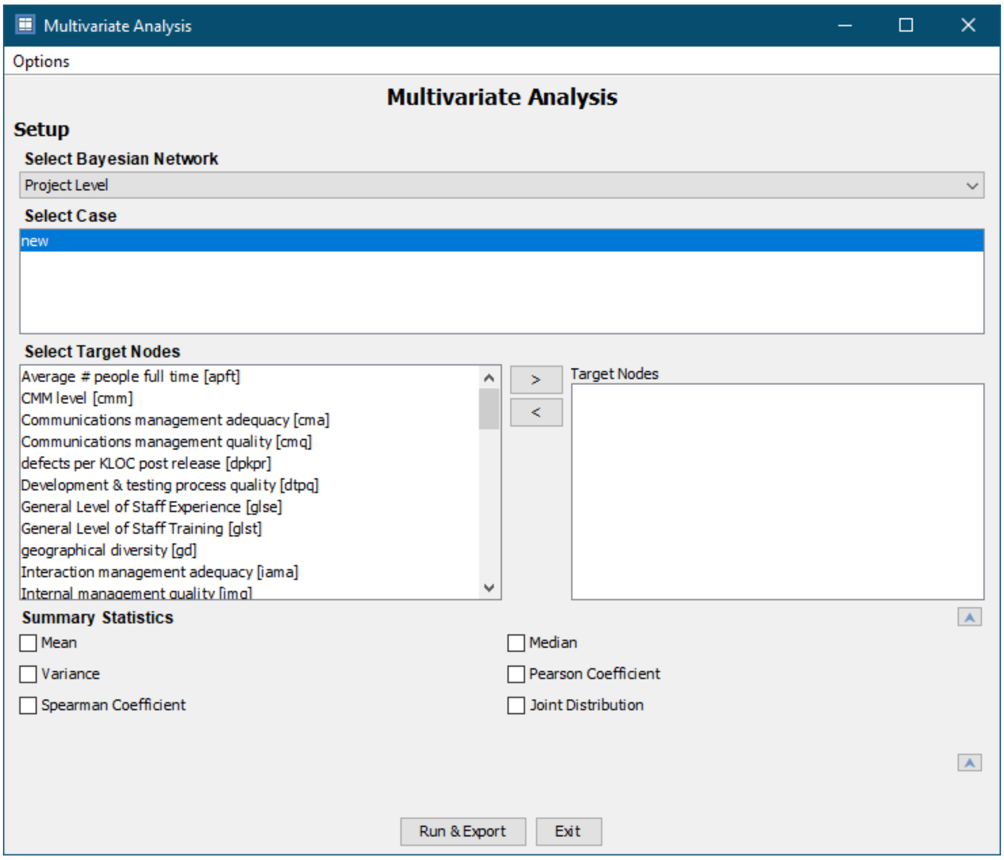
Multivariate analysis works in a similar fashion to sensitivity analysis in that you simply select a set of nodes to use and then run and export the CSV file.

1. Open the example model called *Car Costs.ast* in

   Model Library \Introductory\Car Costs.ast

2. Select the multivariate analysis tool, accessible using this icon on the toolbar .

3. Select the first case, “Honda” and the nodes “annual maintenance cost” and “Total annual cost” as the target nodes, as shown in Figure 96.



Select Bayesian Network: Project Level

Select Case: new

Select Target Nodes:
Average # people full time [apft]
CMM level [cmm]
Communications management adequacy [cma]
Communications management quality [cmq]
defects per KLOC post release [dpkpr]
Development & testing process quality [dtpq]
General Level of Staff Experience [glse]
General Level of Staff Training [glst]
geographical diversity [gd]
Interaction management adequacy [iama]
Internal management quality [imq]

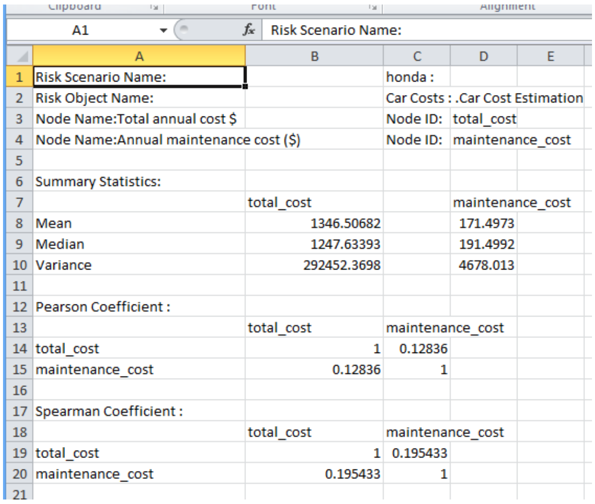
Summary Statistics
☐ Mean
☐ Median
☐ Variance
☐ Pearson Coefficient
☐ Spearman Coefficient
☐ Joint Distribution

Figure 96 Multivariate analysis

4. Select all of the summary statistics and the joint distribution options.

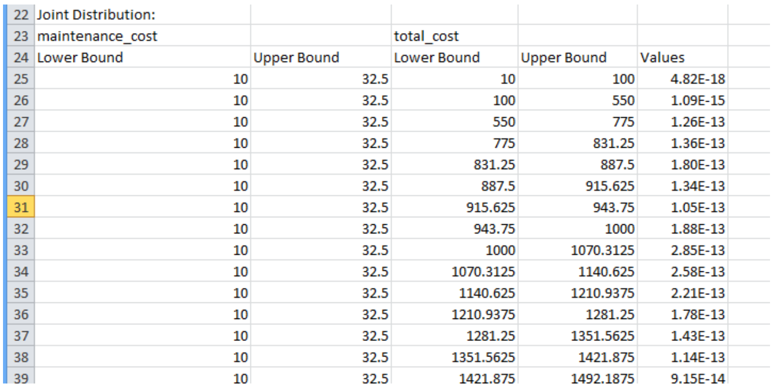
5. Click Run & Export then save the CSV file to a convenient location.

6. A snapshot of the contents of the resulting CSV file produces is shown in Figure 97 and Figure 98. Notice in this case that the nodes are weakly correlated.



	A	B	C	D	E
1	Risk Scenario Name:		honda :		
2	Risk Object Name:		Car Costs : .Car Cost Estimation		
3	Node Name:Total annual cost $		Node ID:	total_cost	
4	Node Name:Annual maintenance cost ($)		Node ID:	maintenance_cost	
5					
6	Summary Statistics:				
7		total_cost		maintenance_cost	
8	Mean	1346.50682		171.4973	
9	Median	1247.63393		191.4992	
10	Variance	292452.3698		4678.013	
11					
12	Pearson Coefficient :				
13		total_cost	maintenance_cost		
14	total_cost	1	0.12836		
15	maintenance_cost	0.12836	1		
16					
17	Spearman Coefficient :				
18		total_cost	maintenance_cost		
19	total_cost	1	0.195433		
20	maintenance_cost	0.195433	1		
21					

Figure 97 Multivariate analysis summary statistics



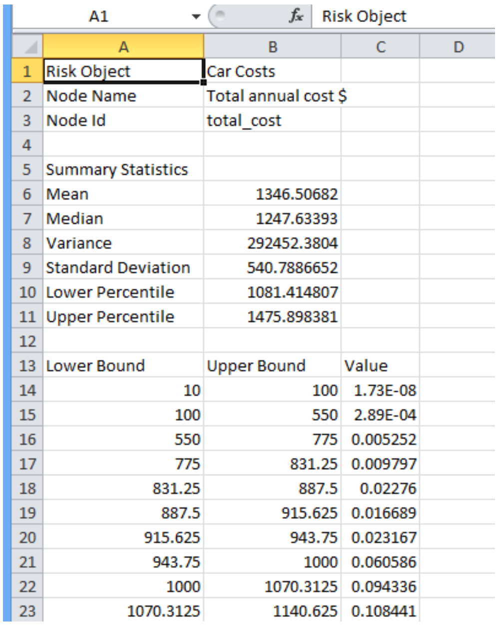
22	Joint Distribution:				
23	maintenance_cost		total_cost		
24	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Values
25	10	32.5	10	100	4.82E-18
26	10	32.5	100	550	1.09E-15
27	10	32.5	550	775	1.26E-13
28	10	32.5	775	831.25	1.36E-13
29	10	32.5	831.25	887.5	1.80E-13
30	10	32.5	887.5	915.625	1.34E-13
31	10	32.5	915.625	943.75	1.05E-13
32	10	32.5	943.75	1000	1.88E-13
33	10	32.5	1000	1070.3125	2.85E-13
34	10	32.5	1070.3125	1140.625	2.58E-13
35	10	32.5	1140.625	1210.9375	2.21E-13
36	10	32.5	1210.9375	1281.25	1.78E-13
37	10	32.5	1281.25	1351.5625	1.43E-13
38	10	32.5	1351.5625	1421.875	1.14E-13
39	10	32.5	1421.875	1492.1875	9.15E-14

Figure 98 Multivariate analysis bivariate distribution

You can export probability distributions results from your models to CSV (Comma Separated Value) files. Likewise with the multivariate analysis feature you can export bivariate distributions, involving pairs of nodes, along with their accompanying correlation (Pearson and Spearman) statistics and summary statistics. These can then be opened in any spreadsheet or application that supports CSV files.

Note that to ensure a fast computational response multivariate analysis uses a fixed discretisation for continuous nodes, regardless of the evidence.

For the probability distribution of individual nodes, simply right click the chart and choose the option Export chart values as CSV and enter a file name when prompted. The resulting probability distribution looks like that shown in Figure 99. Note that for continuous nodes the probabilities shown are listed as the total probability mass between the lower and upper bounds listed and for plotting purposes, say in Excel, you might want to transform these into densities.



A1 *fx* Risk Object

	A	B	C	D
1	Risk Object	Car Costs		
2	Node Name	Total annual cost $		
3	Node Id	total_cost		
4				
5	Summary Statistics			
6	Mean	1346.50682		
7	Median	1247.63393		
8	Variance	292452.3804		
9	Standard Deviation	540.7886652		
10	Lower Percentile	1081.414807		
11	Upper Percentile	1475.898381		
12				
13	Lower Bound	Upper Bound	Value	
14	10	100	1.73E-08	
15	100	550	2.89E-04	
16	550	775	0.005252	
17	775	831.25	0.009797	
18	831.25	887.5	0.02276	
19	887.5	915.625	0.016689	
20	915.625	943.75	0.023167	
21	943.75	1000	0.060586	
22	1000	1070.3125	0.094336	
23	1070.3125	1140.625	0.108441	

Figure 99 Exporting (univariate) probability distribution for single node

9. Learning from data

Automatically learning probability tables, or expressions, from data can be a useful way of defining the NPTs in your model. This tutorial describes how you can use this functionality using an example model for three different flavours of table learning:

- Learning from discrete data alone
- Learning from discrete data with expert judgement
- Learning from discrete data with expert judgement and custom settings
- Learning from continuous data and hybrid data

In all cases learning can be performed even when there are 'missing data'. But it is important to note that the time taken to learn an NPT is more a function of the amount of missing data than the amount of data, so be prepared to allow more processing time for large data files with lots of missing values. The learning process is performed using an Expectation-Maximisation (EM) algorithm and agena.ai modeller currently supports table learning for discrete data, both Boolean and Labelled nodes, and continuous interval nodes set as Normal distributions.

The example files, models and datasets, used in this tutorial are located in the default "*Model Library \Advanced\ Learning from Data*" folder.

Many options for learning are explained by the brief textual explanations that can be accessed by clicking this icon:

?

when you are in the data learning dialog.

9.1 Learning from discrete data alone

1. Open the example model *Asia.ast*.

2. There are two ways to launch the ***learning from data*** function, as shown in Figure 100.

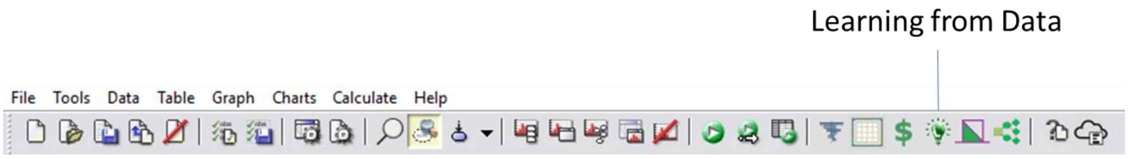
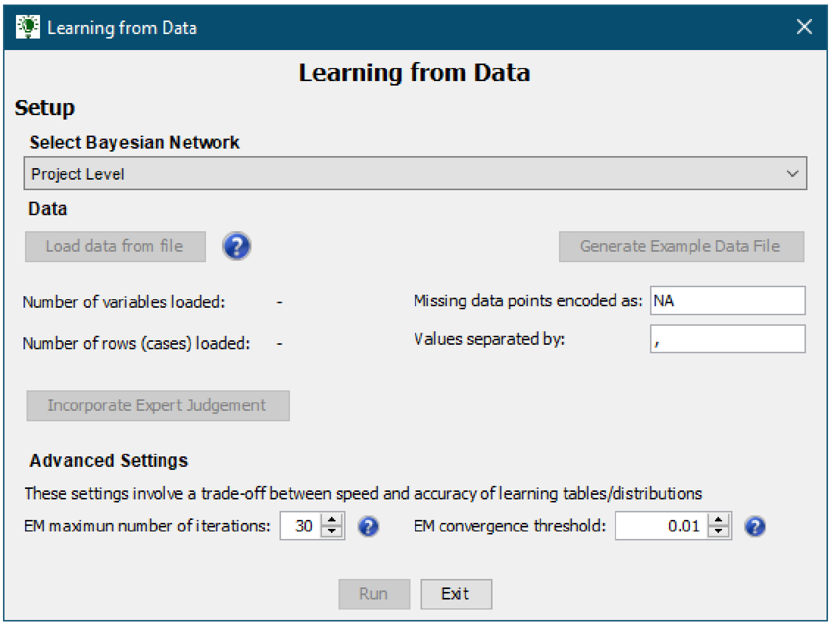


Figure 100 Accessing the learning from data tool

   Specifically, you can do so either via the Tools menu or, more quickly, click the toolbar menu icon:

   💡

3. You will then be presented with the ***learning from data*** main screen as shown in Figure 101.



Select Bayesian Network: Project Level

Number of variables loaded: -

Number of rows (cases) loaded: -

Missing data points encoded as: NA

Values separated by: ,

These settings involve a trade-off between speed and accuracy of learning tables/distributions

EM maximun number of iterations: 30

EM convergence threshold: 0.01

Figure 101 learning from data main screen

4. Learning node probability tables can be performed for nodes that are in a single Bayesian network. The active Bayesian network in the currently opened model is displayed as the default Bayesian network for which the learning process will be performed. The Bayesian network can be changed by choosing another one from the list in the section **Select Bayesian network** (Figure 102).

Figure 102 Selection of Bayesian network for table learning

5. Before running the learning process you have to prepare and load the data from a file. The dataset should be in “csv” format. If you are unsure what is the format in which the data should be prepared you can use a button **Generate Example Data File** (Figure 103).

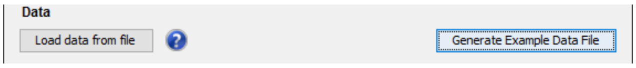


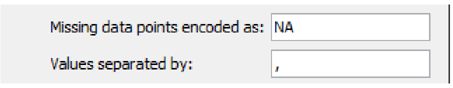
Figure 103 Location of Generate Example Data File button

This function will generate an example “csv” file at the location of your choice. For the currently selected Bayesian network ten rows of random values will be generated for each node. These random values are consistent with the possible states of each variable. Thus, for example, only “yes” and ”no” values are randomly generated for a Boolean node with these states defined. You can use this example file as a template and put your own values there (this can be done in any program, including Microsoft Excel® that supports editing csv files).

Note that you need to ensure that node names and states in the dataset match those in the model. Specifically, the first row of the dataset must contain a list of unique identifiers of model nodes.

The generated example data file for the Asia model can be found in the main folder for this tutorial as “Asia - example dataset.csv”.

6. To load a prepared dataset use the button **Load data from file** and navigate to the location where you have stored your dataset. If necessary, prior to loading the data, you may set a custom value separator (typically a comma, semicolon, space, etc.) and a string representing missing values (typically "NA", "– ", or an empty string) – see Figure 104.

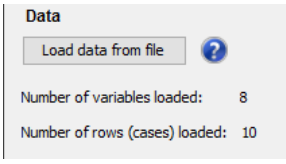


Missing data points encoded as: NA

Values separated by: ,

Figure 104 Parameters defining data formatting in a file

7. Upon completing the loading process, the window displays the number of variables and rows that have been loaded (Figure 105).

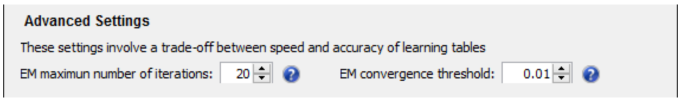


Number of variables loaded: 8

Number of rows (cases) loaded: 10

Figure 105 Data overview updated after loading data from file

8. Before starting the learning process, you may adjust advanced settings that involve a trade-off between speed and accuracy of learning tables (Figure 106).



These settings involve a trade-off between speed and accuracy of learning tables

EM maximun number of iterations: 20

EM convergence threshold: 0.01

Figure 106 Advanced settings for table learning window

The EM maximum number of iterations is the number of times the whole data file will be used to learn the model parameters. It is possible that the EM algorithm will run fewer iterations if the convergence threshold is satisfied earlier. The allowed range is from 1 to 50.

The EM convergence threshold is the difference between the expected log-likelihood of current and previous iterations. The algorithm continues until it either converges or the maximum number of iterations is reached. Note that for discrete only models the default converge will be set at 0.000001, but this can be overwritten.

9. Since in this first part of the tutorial we are using learning solely from data, there are no further settings to consider and so we perform the table learning process by simply clicking the **Run** button. This button is inactive when no data have been loaded (Figure 107).

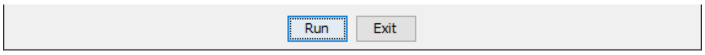
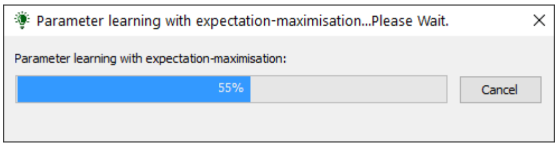


Figure 107 Location of Run button

10. During the calculation process a progress bar is displayed (Figure 108). If you wish to cancel the process, click the button **Cancel**.



Parameter learning with expectation-maximisation...Please Wait.

Parameter learning with expectation-maximisation:

Figure 108 The progress bar for the table learning process

11. Upon successful completion of the table learning process, the model in the main window of agena.ai modeller is updated and the table learning window is closed. In addition, the nodes in the Graph in the main window are marked with an icon of a red letter "D" to indicate that their node probability tables have been learnt from data (Figure 109). Note, that this icon is not preserved when you close and then reopen the model.

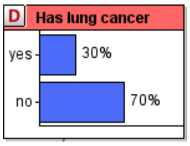


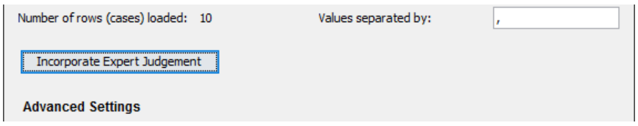
Figure 109 Additional icon indicating that node's probability table has been learnt from data

9.2 Learning from data with expert judgement

This advanced process is similar to the basic example discussed in Section 10.1. However, here additional settings allow us to incorporate expert judgement.

1. Follow the steps 1-8 as explained in earlier example in Section 9.1.

2. Click the ***Incorporate Expert Judgement*** button (Figure 110).


Number of rows (cases) loaded: 10

Values separated by: ,

Figure 110 Location of Incorporate Expert Judgement button

3. This opens a window with additional settings for the learning process (Figure 111). In this window you can set how the expert judgement will be incorporated during the table learning process. If you leave the window without changing any settings and run the learning process then this process will be performed as if you did not open this additional window, i.e., probability tables will be learnt solely from data.

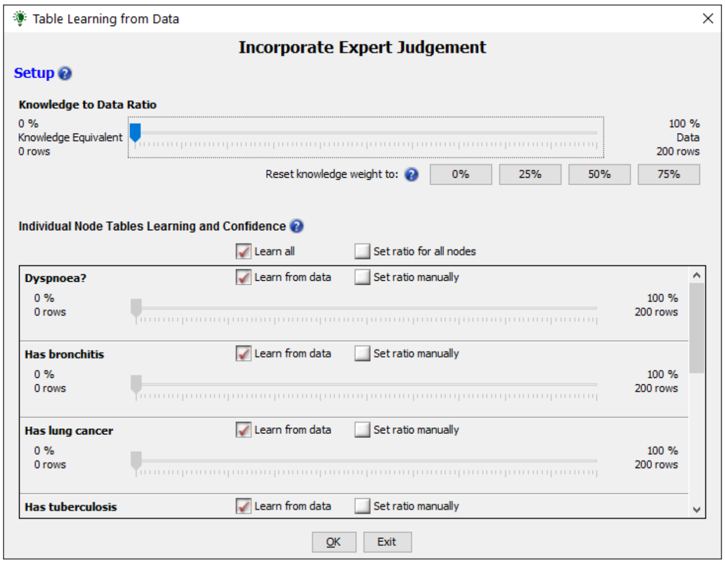


Table Learning from Data

Incorporate Expert Judgement

Setup

Knowledge to Data Ratio

0 % Knowledge Equivalent 0 rows — 100 % Data 200 rows

Reset knowledge weight to: 0% 25% 50% 75%

Individual Node Tables Learning and Confidence

☑ Learn all ☐ Set ratio for all nodes

Dyspnoea? ☑ Learn from data ☐ Set ratio manually
0 % 0 rows — 100 % 200 rows

Has bronchitis ☑ Learn from data ☐ Set ratio manually
0 % 0 rows — 100 % 200 rows

Has lung cancer ☑ Learn from data ☐ Set ratio manually
0 % 0 rows — 100 % 200 rows

Has tuberculosis ☑ Learn from data ☐ Set ratio manually

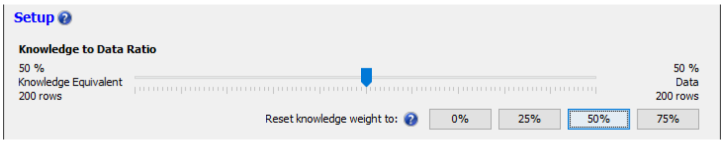
Figure 111 The window with additional settings for incorporating expert judgement

4. The topmost slider bar enables you to set the ratio of your confidence in knowledge (as represented by the existing node probability tables) compared to the data, from the loaded data file. Depending on the value of the slider the weight given to the model probability tables is a proportion of the number of rows in the loaded data file.

   With other settings at their defaults, changing this top slider sets the same ratio for each node probability table in the current Bayesian network. Changing individual settings for each node will be discussed in the next section.

   By default, this slider points at 0% (placed at far left). This indicates a situation where you have no confidence in the knowledge already encoded in the NPTs and you have 100% confidence in your data. In this case the existing NPT values will be ignored and the new NPTs will be learnt solely from data.

   If you set the ***knowledge to data ratio*** slider to 50% (which you can do either by moving the slider to the 50 mark or by clicking the **50%** button below the slider) it would mean that you have equal confidence about knowledge in the current probability tables and the data loaded. So if, for example, the dataset contains 200 rows, then this 50% setting means that your confidence in the current NPT is equivalent to it being based on 200 rows of data (Figure 112).



Setup

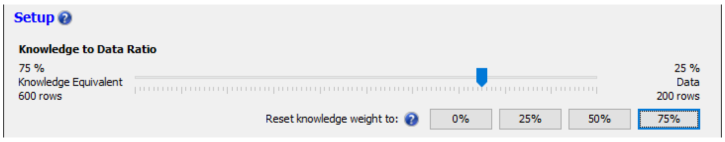
Knowledge to Data Ratio

50 %
Knowledge Equivalent
200 rows

50 %
Data
200 rows

Figure 112 Knowledge to data ratio set to 50%

   If you set the ***knowledge to data ratio*** slider to 75% (which you can do either by moving the slider to the 75% mark or by clicking the "Reset knowledge weight to" 75%) it would mean that your confidence about knowledge in the current probability tables is greater than your confidence in the data by a 75:25 ratio (i.e. you are three times more confident about the current probability table values than the new data). So if, for example, the dataset contains 200 rows, then this 75% setting means that your confidence in the current NPT is equivalent to it being based on 600 rows of data (Figure 113).



Setup

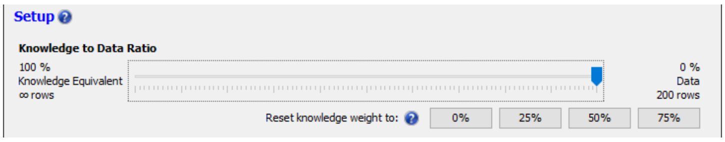
Knowledge to Data Ratio

75 %
Knowledge Equivalent
600 rows

25 %
Data
200 rows

Figure 113 Knowledge to data ratio set to 75%

   If you set the ***knowledge to data ratio*** slider to 100% it would mean that your confidence in the current NPT is 100% and thus your confidence in data is 0% (Figure 114). In this case the data is ignored and the NPT remains unchanged.



Setup

Knowledge to Data Ratio

100 %
Knowledge Equivalent
∞ rows

0 %
Data
200 rows

Figure 114 Knowledge to data ratio set to 100%

5. After setting the desired value for knowledge to data ratio, confirm this setting by clicking OK button.

6. Then, start the calculation process by following the steps 9-12 as explained in earlier example in Section 9.1.

   Depending on the setting for knowledge to data ratio, the nodes in the Bayesian network for which the table learning has been performed will be displayed in the Graph with an additional icon – a letter in red font (Figure 115).

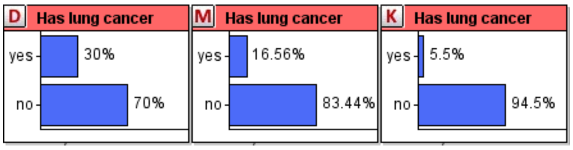


Figure 115 Icons indicating the source for probability table

The meaning of these icons is the following:

- “D” – the node’s probability table has been learnt solely from **data**, i.e. 0% for knowledge to data ratio,
- “M” – the node’s probability table has been learnt as a **mixture** – partially from data and partially from previous NPT, i.e. knowledge to data ratio higher than 0% and lower than 100%.
- “K” – the node’s probability table has been based solely on **knowledge** (previous NPT), i.e., 100% for knowledge to data ratio.

9.3 Learning from data with expert judgement with custom settings

This advanced process is similar to 10.2, except that here additional custom settings for incorporating expert judgement can be set, individually, on each node.

1. Follow the steps 1-3 as explained in earlier example in Section 9.2.

2. In the Incorporate Expert Judgement window, provide custom settings for individual nodes as in Figure 116.

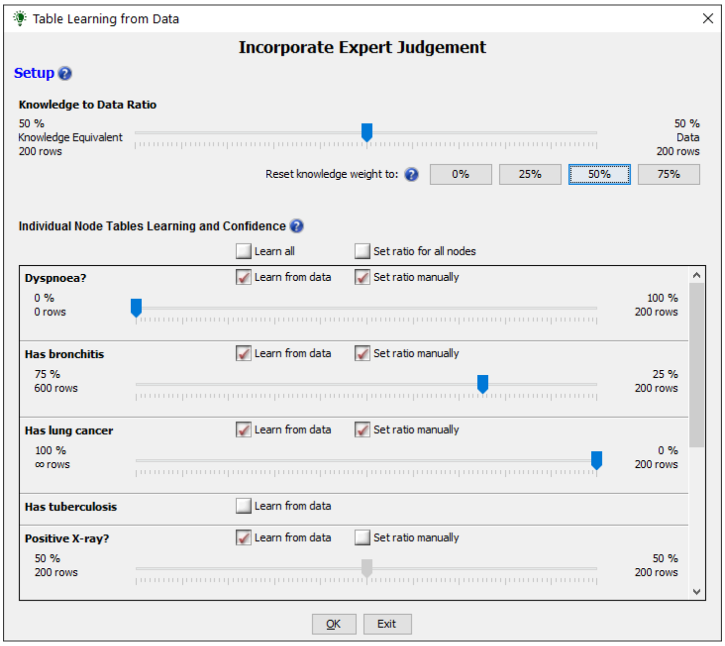


Table Learning from Data

Incorporate Expert Judgement

Setup

Knowledge to Data Ratio

50 % Knowledge Equivalent 200 rows — 50 % Data 200 rows

Individual Node Tables Learning and Confidence

☐ Learn all ☐ Set ratio for all nodes

Dyspnoea? ☑ Learn from data ☑ Set ratio manually — 0 % 0 rows — 100 % 200 rows

Has bronchitis ☑ Learn from data ☑ Set ratio manually — 75 % 600 rows — 25 % 200 rows

Has lung cancer ☑ Learn from data ☑ Set ratio manually — 100 % ∞ rows — 0 % 200 rows

Has tuberculosis ☐ Learn from data

Positive X-ray? ☑ Learn from data ☐ Set ratio manually — 50 % 200 rows — 50 % 200 rows

Figure 116 Custom settings for individual nodes in Incorporate Expert Judgement window

First, we set the value knowledge to data ratio to 50% (topmost slider). This will be the default value of knowledge to data ratio for nodes in the Bayesian network for which we plan to run table learning.

Then, we provide the following settings:

- For node “Dyspnoea?” tick the “Set ratio manually” box to activate the individual slider for this node. Set a custom value for knowledge to data ratio. If you set a value 0%, i.e. slider position is leftmost, his means table learning for this node solely from data.
- For node “Has bronchitis” tick the “Set ratio manually” box and provide a value 75% using the slider. This means table learning for this node as 75% knowledge from current NPT and 25% from data.
- For node “Has lung cancer” activate the slider and provide a value 100% .
- For node “Has tuberculosis” untick the “Learn from data” box. This means that the NPT for this node will not be learnt from data (it will stay unchanged). It is an equivalent of manually assigning 100% for knowledge to data ratio.
- Do not provide any individual settings for node “Positive X-ray?”. This means that the value of knowledge to data ratio for this node will be set based on the master setting in the topmost slider.

In fact, moving this master slider to a different value now also moves individual sliders for nodes for which the box “Set ratio manually” has been left unticked but does not move sliders for other nodes.

3. Complete steps 5-6 as explained in the earlier example in Section 9.2.

4. The model with learnt probability tables will now display icons indicating the sources for probability tables that may be different for different nodes (Figure 117).

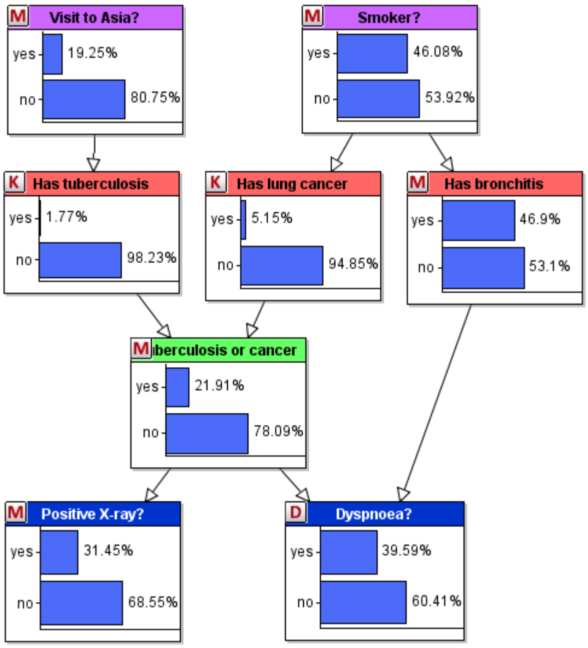


Figure 117 Different icons indicating the sources for probability tables

Specifically, because the master setting for the knowledge to data ratio has been set to 50%, all the nodes for which no custom setting has been provided are now marked with “M” (mixture). The node “Dyspnoea?” was the only one fully learnt from data, thus it has been marked with “D”. Two nodes, “Has lung cancer” and “Has tuberculosis” have kept their NPTs based on knowledge, thus they have been marked with “K”. The node “Has bronchitis”, for which a custom value of 75% for knowledge to data ratio has been provided, has also been marked with “M”.

5. Note that setting the data to knowledge ratio to 0% for a node is equivalent to simply unticking the “Learn from data” tick box for that variable. Likewise, setting it to 100% is equivalent to only learning from data and probabilities set on that node will be retained.

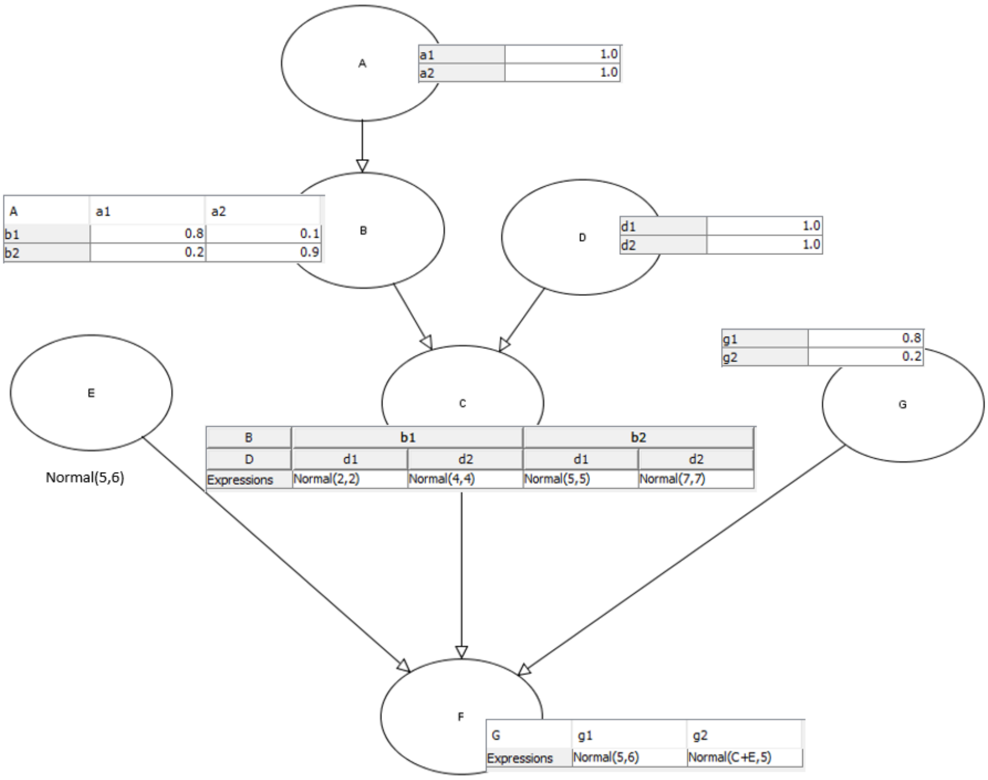
6. Also note that removing a column of data from your data file for a given node will not necessarily result in the same learned NPT values as learning from the full data set and choosing not to learn the NPT for the same given node. For example when learning P(B | A)P(A) the table P(B | A) will be learnt from all joint combinations of values for A and B in the data file, even if you choose not to learn the NPT for node A, P(A). So, removing a node from the analysis only prevents learning for that node and any nodes that depend on it will still use its’ data when learning their own NPT. If on the other hand you remove or do not have a data column in your data set for A and learn P(B | A) the values in the data set for B will be considered equally probable across all possible values for A, on the assumption that the node A is entirely composed of missing values. The latter situation gives different results to the former.

9.4 Learning from continuous data

When we have hybrid models, containing continuous and discrete data, or simply continuous data learning can be done as described in section 9.1-9.3 for discrete data but continuous data will be handled differently. Firstly, it is assumed that the continuous data is Normally distributed and secondly any continuous child node will either be a linear model, should it be a function of continuous parent nodes, or a mixture of linear Normal models if it is a function of discrete and continuous parent nodes. The output from learning will be new probability values for discrete nodes and new expressions from continuous nodes.

Note that particular continuous nodes can be excluded from the analysis in the same way as was done for discrete nodes - by clicking the ***Incorporate Expert Judgement*** button and de-selecting any continuous nodes you do not wish to include. Also note that when learning hybrid or continuous models the default convergence threshold will be set as 0.01.

To demonstrate the EM learning process this example learns from data on a hybrid model, with conditional distributions prior to learning, as shown in Figure 118. This model is located in *Model Library /Advanced/Learning from data/Hybrid EM*.ast.



a1	1.0
a2	1.0

A	a1	a2
b1	0.8	0.1
b2	0.2	0.9

d1	1.0
d2	1.0

g1	0.8
g2	0.2

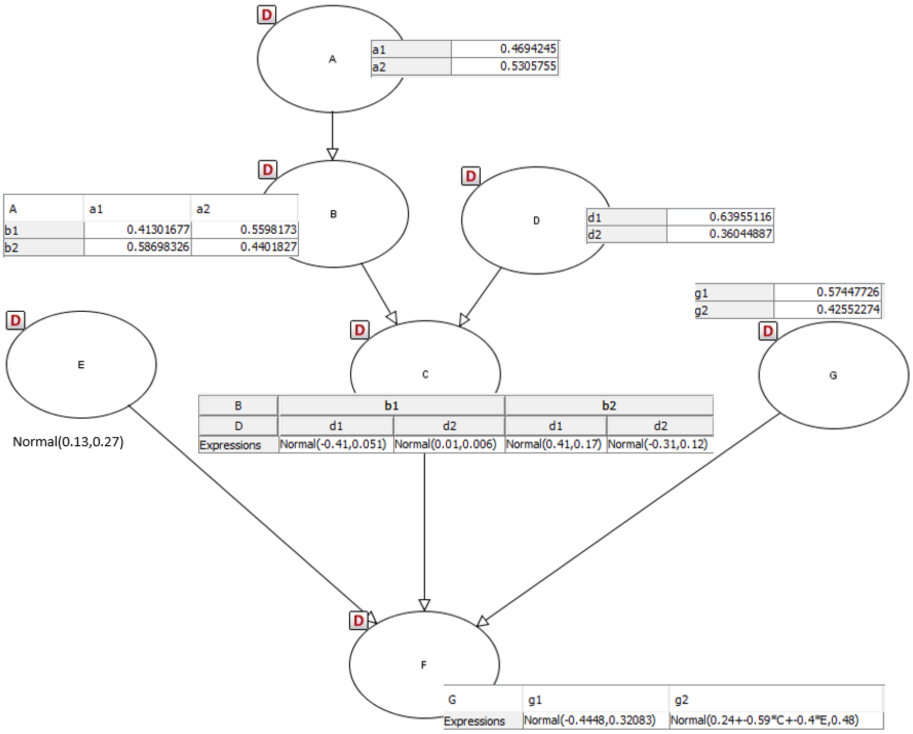
Normal(5,6)

B	b1		b2	
D	d1	d2	d1	d2
Expressions	Normal(2,2)	Normal(4,4)	Normal(5,5)	Normal(7,7)

G	g1	g2
Expressions	Normal(5,6)	Normal(C+E,5)

Figure 118 Hybrid model before data learning

To learn on this model select the data file Hybrid EM.csv. When we learn the from this data we achieve the revised BN NPTs and expressions as shown in Figure 119. Notice that the mixture model for node G is a simple Normal distribution and a linear regression equation dependent on nodes C and E.



a1	0.4694245
a2	0.5305755

A	a1	a2
b1	0.41301677	0.5598173
b2	0.58698326	0.4401827

d1	0.63955116
d2	0.36044887

g1	0.57447726
g2	0.42552274

Normal(0.13,0.27)

B	b1		b2	
D	d1	d2	d1	d2
Expressions	Normal(-0.41,0.051)	Normal(0.01,0.006)	Normal(0.41,0.17)	Normal(-0.31,0.12)

G	g1	g2
Expressions	Normal(-0.4448,0.32083)	Normal(0.24+-0.59*C+-0.4*E,0.48)

Figure 119 Hybrid model after data learning with new NPTs and expressions

The resulting model and charts superimposed the model are shown in Figure 120.

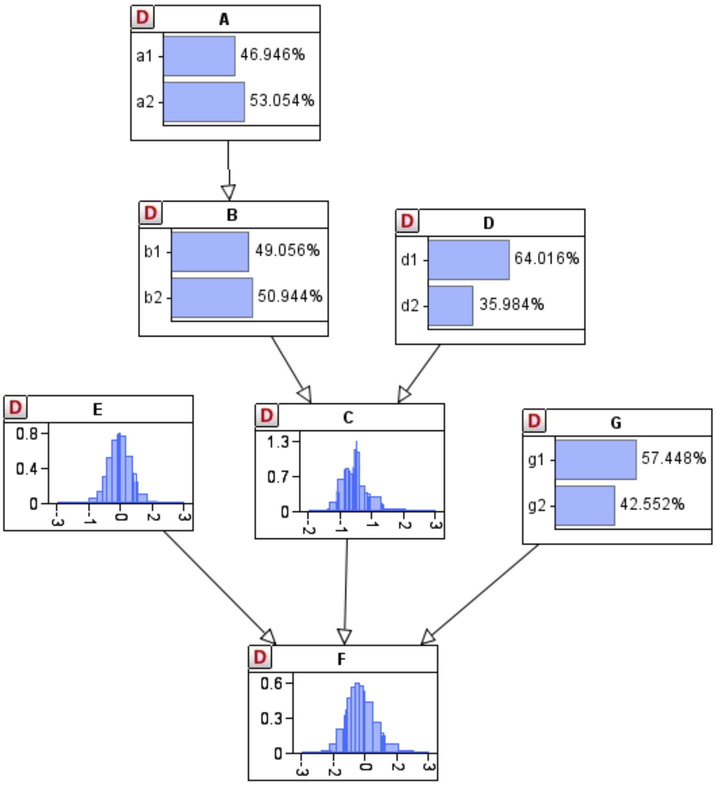


Figure 120 Hybrid model post-learning

10. Compound sum analysis

Compound sum analysis involves analysing the frequency distribution of events, where each of these events has an associated severity distribution and computing a compound total value from all of these. The severity distribution might represent an individual loss or profit event, the time to failure of a component, or some other variable. The frequency would be the number of times these events might be expected to occur over a defined period. For example, an insurer may hold a portfolio of insurance policies where any of these may involve a claim – the frequency of claims will be uncertain, and the severity of each claim will be uncertain too. What the insurer would like to know is the total claim amount as a function of the frequency of claims and the severity of each claim.

Mathematically this problem is solved using a process called convolution, which agena.ai modeller does automatically by performing arithmetical operations on simulation nodes. To perform a compound sum in agena.ai modeller you would simply add *n* nodes together. However, given *n* may itself be a variable quantity this compound problem cannot be directly solved in a straightforward way. Hence, the role of the compound sum analysis tool is to solve this more general problem. It is worth noting that this general problem also goes by the names *n*-fold and *N*-fold convolution, where *n* denotes a fixed frequency value and *N* denotes a variable frequency with its own distribution function.

Three nodes are used in the compound sum analysis:

- Frequency node – this must be a non-negative integer Interval type node
- Severity node – this can be any Continuous or Interval type
- Compound node – this is the target node for the outcome of the analysis and cannot have any parent nodes.

Computation proceeds by generating a series of Bayesian networks in the background for each frequency node value, composed of arithmetic additions of the severity variable *n* times. The time taken for this computation is proportional to the number of values in the frequency node.

The example file used in this tutorial is located in the default "*Model Library \Advanced\Compound Sum analysis*" folder.

1. Open the example model *Compound sum analysis.ast*.

2. There are two ways to launch the ***Compound sum analysis*** function, as shown in Figure 121.

   Specifically, you can do so either via the Tools menu or, more quickly, click the toolbar menu icon

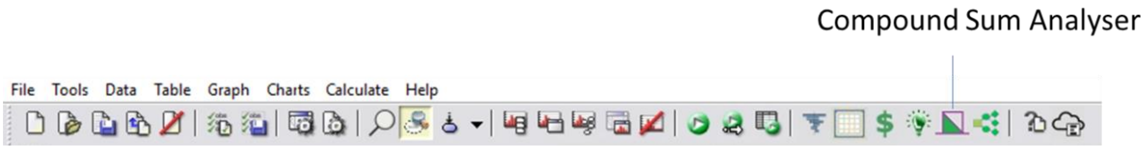


Figure 121 Accessing the Compound Sum Analysis tool

3. You will then be presented with the compound sum analysis main screen as shown in Figure 122.

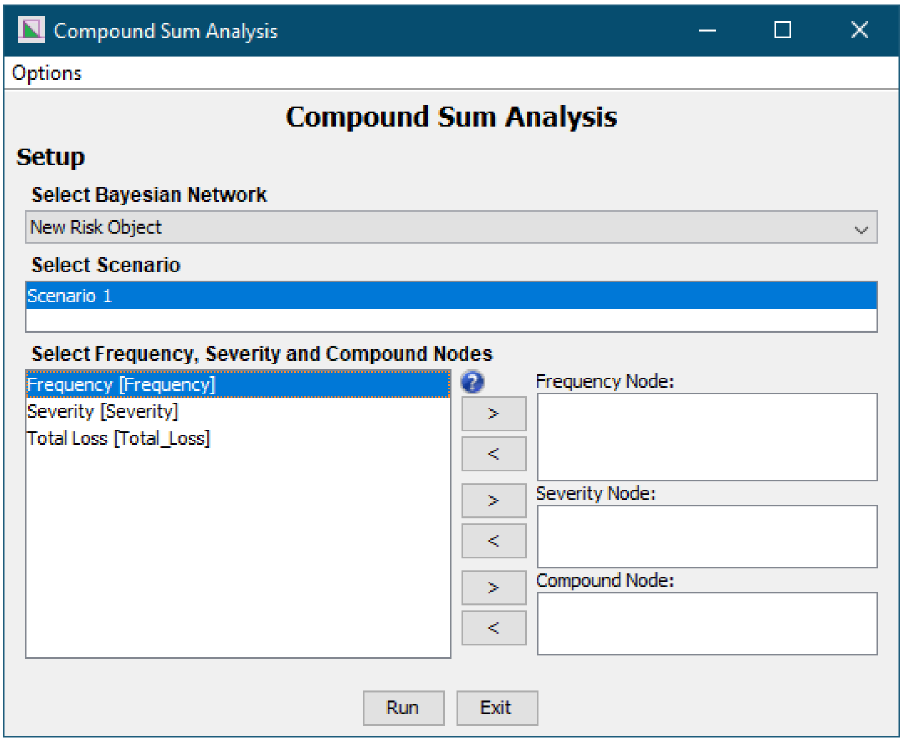
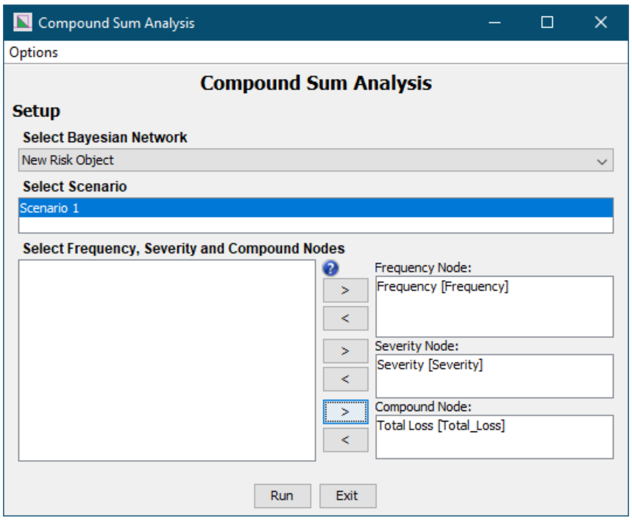


Figure 122 Compound Sum Analysis tool main screen

4. Now you simply select the Frequency, Severity and Compound nodes as shown in Figure 123 (in our example the Compound node is labelled “Total Profit”).



Frequency Node: Frequency [Frequency]

Severity Node: Severity [Severity]

Compound Node: Total Loss [Total_Loss]

Figure 123 Compound Sum Analysis tool main screen with Frequency, Severity and Compound Sum nodes entered

5. Press the run button to run the compound calculation.

6. Once the analysis has completed the computed compound sum will be inserted into the node you assigned for that purpose. If the node was a simulation node previously this will now be overwritten as a normal continuous node. In our example the resulting compound sum (i.e., the node “Total Loss”) distribution is shown in Figure 124.

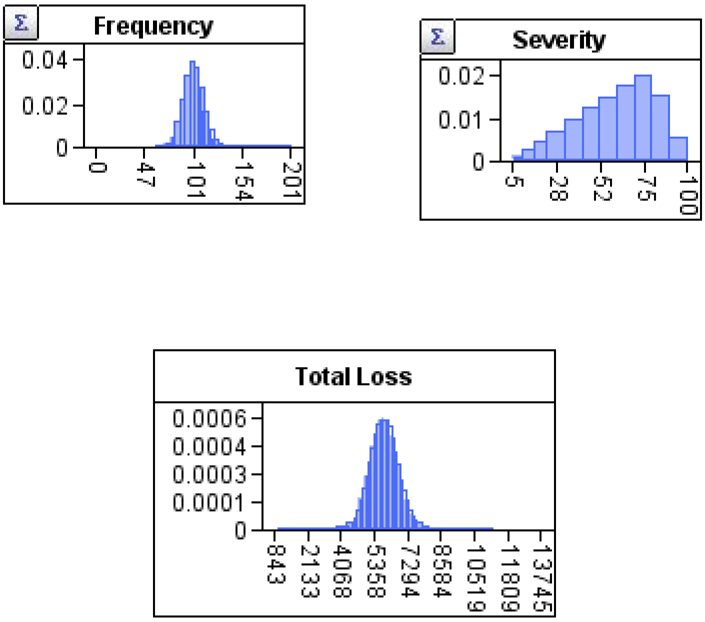


Figure 124 Compound Sum Analysis result for example

7. Be careful that each time the frequency and severity distributions change you will need to manually re-execute the compound sum analysis.
8. Since the compound sum analysis uses simulation, it follows that there are simulation settings specific to this analysis available from the Options menu on the screen. However, in contrast with the other options screens for sensitivity analysis etc., here there is an additional option to specify the entropy error target for the Frequency node to make the process faster.

11. Value of information analysis

Value of Information analysis enables users to calculate the amount a decision maker would be willing to pay for information prior to making a decision.

The example model used in this tutorial is in the default “*Model Library \Advanced\Value of Information*” folder. This model helps an investment advisor recommend the best type of investment for £10,000. There are three main unknown variables of interest: bank rate, stock market increase, and the customer’s risk aversion. For simplicity, there is a single decision node “investment decision” (which has three options) two utility nodes (“return on 10,000” and “emotional loss”). The “emotional loss” is there to capture the sentiments of risk averse people for whom the idea that an investment is disproportionate to the actual monetary loss. There is also an aggregated utility node named “total utility”. In this tutorial, we will focus on a decision variable named “Invest decision”, a utility variable named “Total utility”, and the uncertain variables named “Risk aversion” (discrete) and “Stock market increase” (continuous) respectively.

Value of information analysis can currently only be carried out for one decision node, one or several uncertain nodes and a single utility node.

11.1 Analysis with discrete uncertainty variable

1. Open the example model *Invest.cmp*.

2. There are two ways to launch the ***Value of Information Analyser***, as shown in Figure 125.

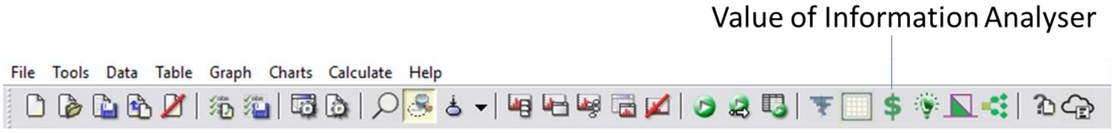
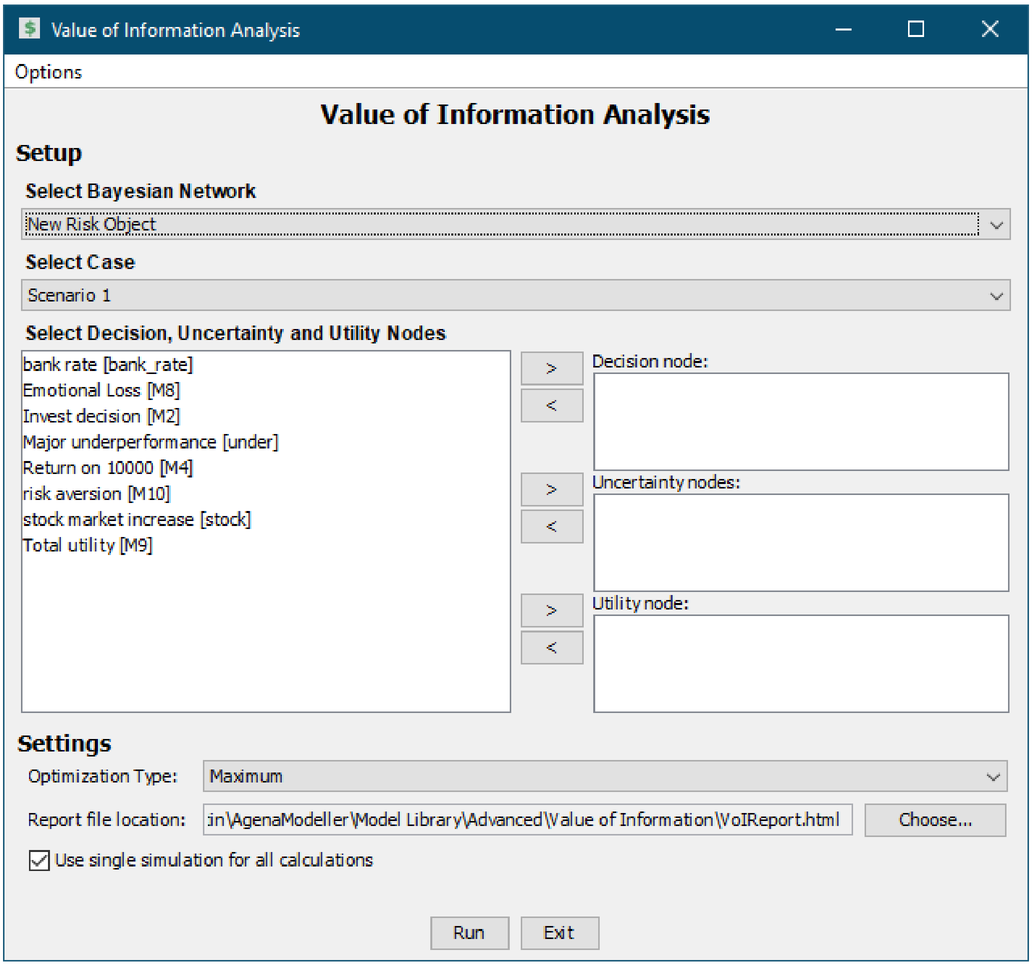


Figure 125 Accessing the value of information analyser

Specifically, you can do so either via the Tools menu or, more quickly, click the toolbar menu icon:

$

3. You will then be presented with the ***value of information*** main screen as shown in Figure 126.



Value of Information Analysis

Setup

Select Bayesian Network: New Risk Object

Select Case: Scenario 1

Select Decision, Uncertainty and Utility Nodes

bank rate [bank_rate]
Emotional Loss [M8]
Invest decision [M2]
Major underperformance [under]
Return on 10000 [M4]
risk aversion [M10]
stock market increase [stock]
Total utility [M9]

Settings

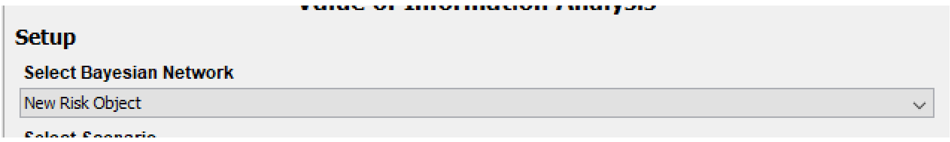
Optimization Type: Maximum

Report file location: in\AgenaModeller\Model Library\Advanced\Value of Information\VoIReport.html

☑ Use single simulation for all calculations

Figure 126 Value of information main screen

4. Value of information analysis can be performed when three main types of nodes (decision, uncertainty, utility) are in a single Bayesian network. The active Bayesian network in the currently opened model is displayed as the default Bayesian network for which the value of information analysis will be performed. The Bayesian network can be changed by choosing another one from the list in the section **Select Bayesian network** (Figure 127).

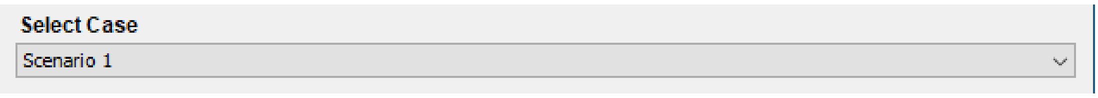


Setup

Select Bayesian Network: New Risk Object

Figure 127 Selection of Bayesian network for value of information analysis

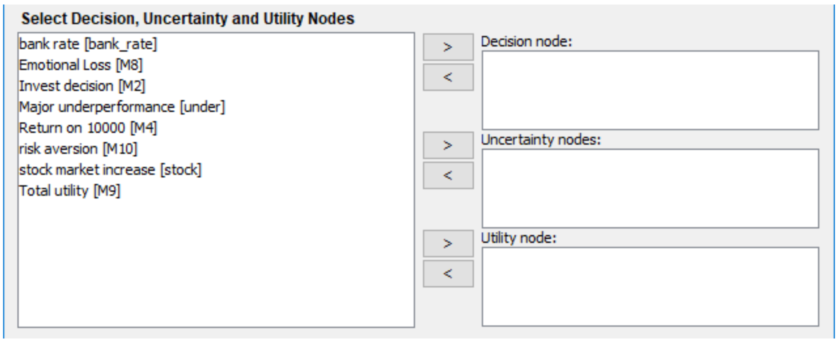
5. Value of information analysis preserves any possible evidence entered to nodes other than the main three. Any possible evidence entered to the main three nodes will be cleared upon the start of analysis. The window enables selecting a case that should be used in analysis by an option **Select Case** (Figure 128).



Select Case: Scenario 1

Figure 128 Selection of case for value of information analysis

6. The next step involves selecting three main categories of nodes: decision, uncertainty and utility nodes. Exactly one node has to be selected for decision and utility nodes. Several nodes can be selected as uncertainty nodes. In this example we assume a single uncertainty node is selected. This selection can be performed in the section **Select Decision, Uncertainty and Utility Nodes** (Figure 129).



Select Decision, Uncertainty and Utility Nodes

bank rate [bank_rate]
Emotional Loss [M8]
Invest decision [M2]
Major underperformance [under]
Return on 10000 [M4]
risk aversion [M10]
stock market increase [stock]
Total utility [M9]

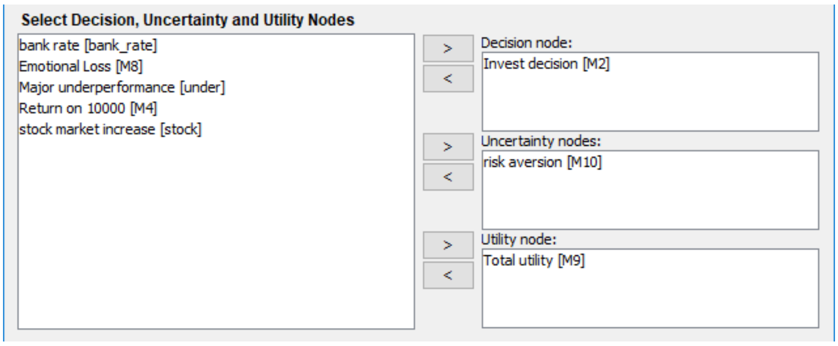
Decision node:

Uncertainty nodes:

Utility node:

Figure 129 Selection of Decision, Uncertainty and Utility Nodes

By highlighting specific node and using > and < buttons it is possible to move a node between selection boxes. Using them set these selections as shown in Figure 130.



Select Decision, Uncertainty and Utility Nodes

bank rate [bank_rate]
Emotional Loss [M8]
Major underperformance [under]
Return on 10000 [M4]
stock market increase [stock]

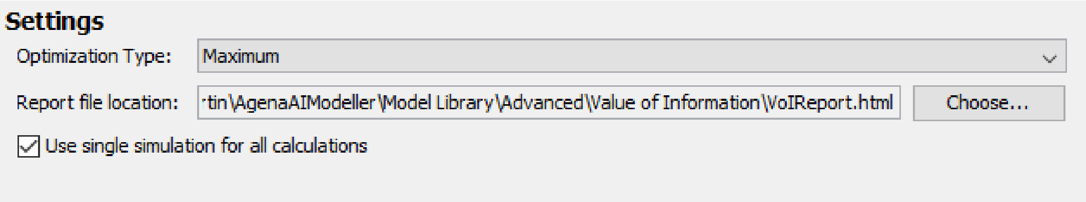
Decision node: Invest decision [M2]

Uncertainty nodes: risk aversion [M10]

Utility node: Total utility [M9]

Figure 130 Selected Decision, Uncertainty and Utility Nodes

7. There are three other settings that can be changed in the **Settings** section (Figure 131).



Settings

Optimization Type: Maximum

Report file location: tin\AgenaAIModeller\Model Library\Advanced\Value of Information\VoIReport.html

☑ Use single simulation for all calculations

Figure 131 Settings for value of information analysis

- **Optimization Type** – The default value “Maximum” means that the expected value of utility will be maximised. The other value “Minimum” means that the expected value of utility will be minimised – this can be useful e.g., when a utility represents costs that should be minimised. In this tutorial we set the value “Maximum” as we are interested in maximising the income.

- **Report file location** – The tool generates an HTML report file summarising the results of analysis. Using this setting, by clicking a button “Choose…” it is possible to specify the location and a file name of this report. By default, the report will be saved in the agena.ai modeller home directory, e.g. in “C:\Users\<user name>\Agena.ai modeller\” on Windows systems in a file “VoIReport.html”.

Note, that agena.ai modeller will attempt to overwrite a report file that might have been created earlier in the provided location and file name. In this tutorial we keep this default setting.

- **Use single simulation for all calculations** – With this setting it is possible to control the behaviour of simulation nodes during calculations.

    i. When this option is selected the tool will discretize all simulation nodes in the selected Bayesian network. Then, all iterations of calculations for value of information analysis will be performed using this discretization – i.e. the same in all iterations. This can significantly speed up calculations at the cost of reduced precision.

    ii. When this option is not selected agena.ai modeller will discretize all simulation nodes separately in each iteration of calculations for value of information analysis. Hence, simulation nodes can be discretized differently in each iteration. This option can significantly increase the calculation time for the analysis but may result in increased precision.

  Note that regardless of this setting, if decision and/or uncertainty nodes are simulation nodes, they will be discretized once because the iterations of calculations for value of information analysis are defined based on discretization for these two nodes.

  In this tutorial we keep this option selected.

8. In the next step start the analysis by clicking “Run” button. During calculations for analysis a progress window is displayed (Figure 132). If necessary, you may abort the calculation process by clicking “Cancel” button.

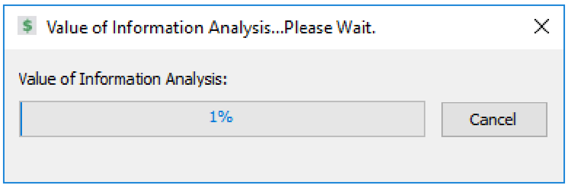
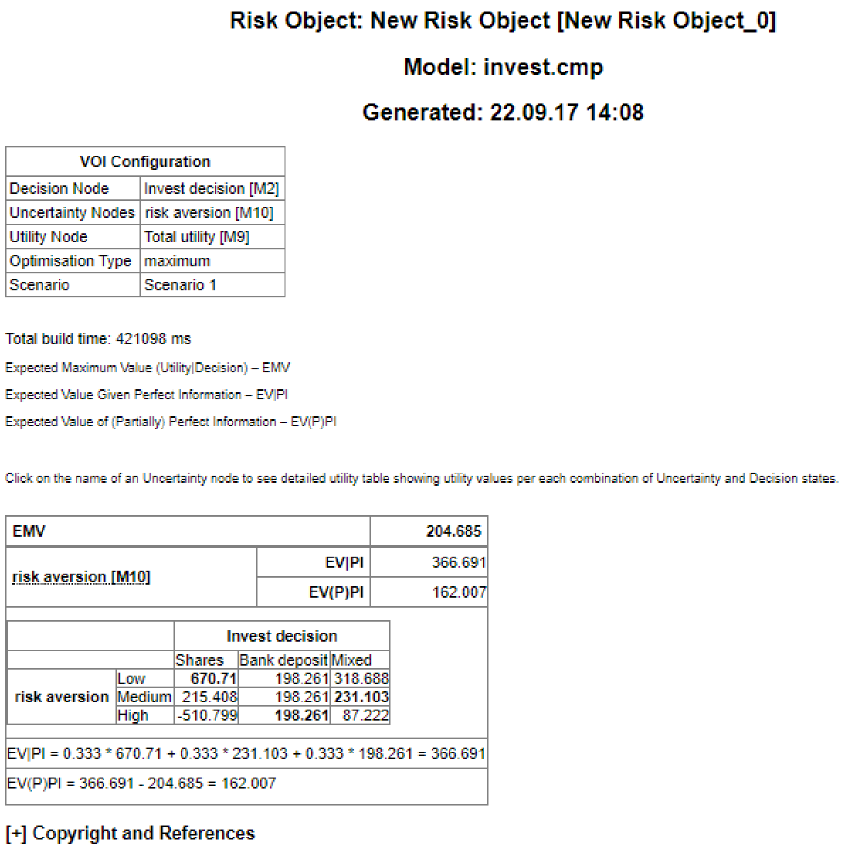


Figure 132 Progress window for value of information analysis

9. When the calculation process completes an HTML report file is saved in directory specified earlier. Then this report is opened in a default web browser. (Figure 133).

   After clicking the underlined name of uncertainty node in the table at the bottom of the report a more detailed table is displayed showing utility values per each combination of uncertainty and decision states. In this example, it is the table with column header “Invest decision” and with row header “risk aversion”. It contains some values marked in bold which are the optimum values for each state of uncertainty node.



Risk Object: New Risk Object [New Risk Object_0]

Model: invest.cmp

Generated: 22.09.17 14:08

VOI Configuration	
Decision Node	Invest decision [M2]
Uncertainty Nodes	risk aversion [M10]
Utility Node	Total utility [M9]
Optimisation Type	maximum
Scenario	Scenario 1

Total build time: 421098 ms

Expected Maximum Value (Utility|Decision) – EMV

Expected Value Given Perfect Information – EV|PI

Expected Value of (Partially) Perfect Information – EV(P)PI

Click on the name of an Uncertainty node to see detailed utility table showing utility values per each combination of Uncertainty and Decision states.

EMV		**204.685**
risk aversion [M10]	**EV|PI**	366.691
	EV(P)PI	162.007

		Invest decision		
		Shares	Bank deposit	Mixed
risk aversion	Low	**670.71**	198.261	318.688
	Medium	215.408	198.261	**231.103**
	High	-510.799	**198.261**	87.222

EV|PI = 0.333 * 670.71 + 0.333 * 231.103 + 0.333 * 198.261 = 366.691

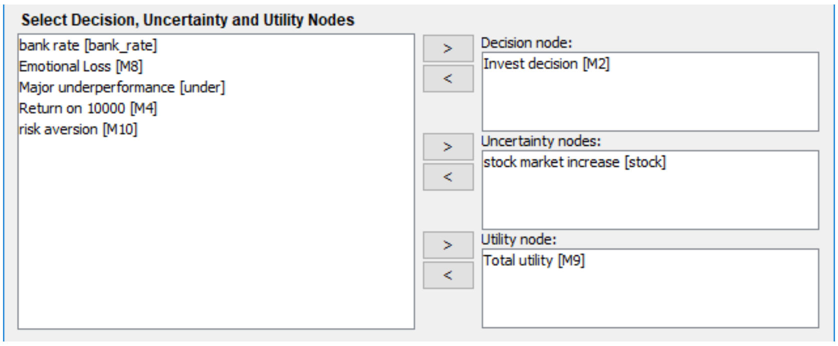
EV(P)PI = 366.691 - 204.685 = 162.007

[+] Copyright and References

Figure 133 HTML report with results of value of information analysis for discrete uncertainty node

11.2 Analysis with continuous uncertainty variable

1. Follow the steps explained in Section 11.1. As the only difference, set the uncertainty variable to “stock market increase” (Figure 134).



Select Decision, Uncertainty and Utility Nodes

bank rate [bank_rate]
Emotional Loss [M8]
Major underperformance [under]
Return on 10000 [M4]
risk aversion [M10]

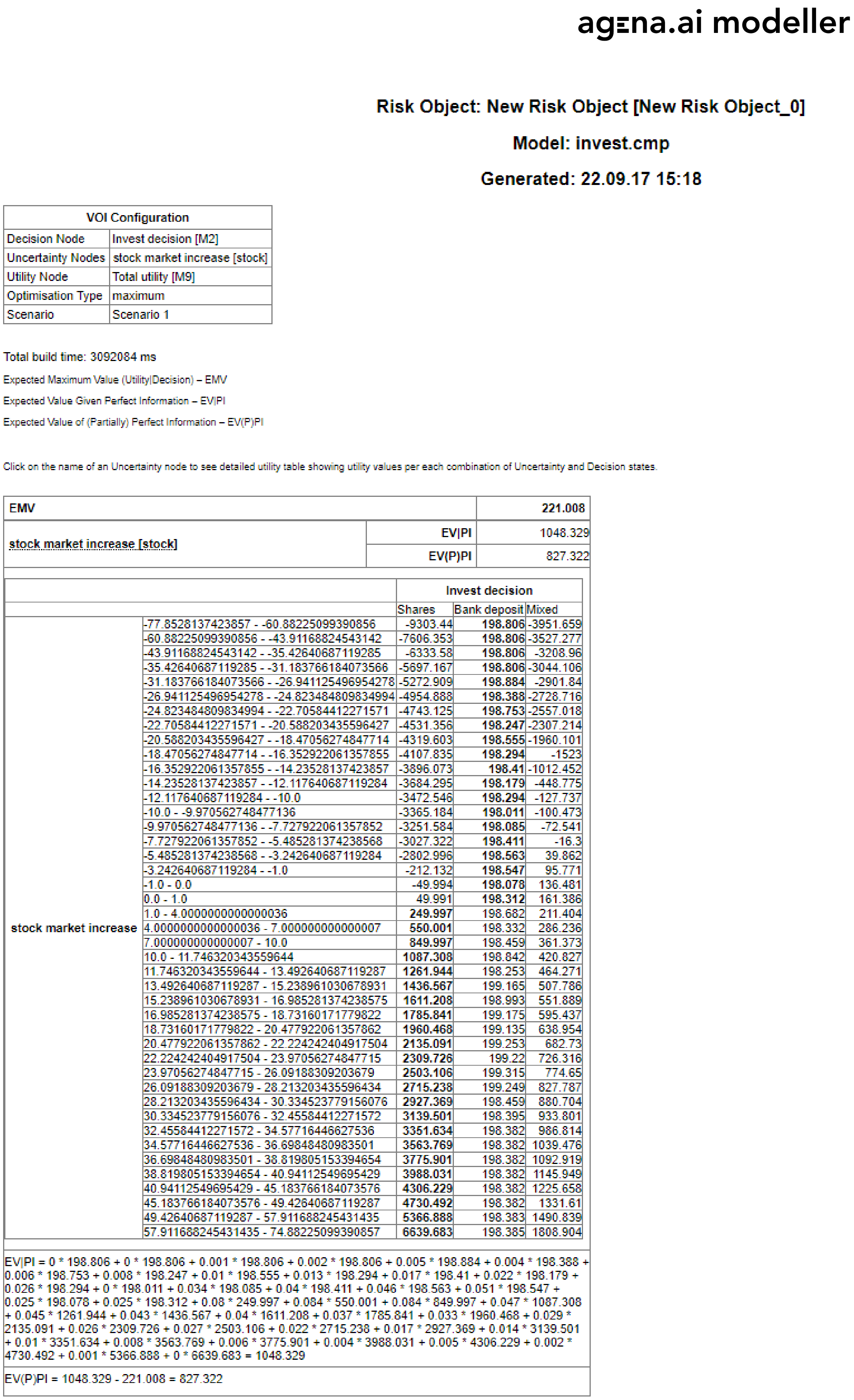
Decision node: Invest decision [M2]

Uncertainty nodes: stock market increase [stock]

Utility node: Total utility [M9]

Figure 134 Selected Decision, Uncertainty and Utility Nodes (continuous uncertainty)

2. Upon completion of calculations the generated HTML report shows that expected value given perfect information (EV|PI) has been calculated for a discretized uncertainty node (Figure 135).



agena.ai modeller

Risk Object: New Risk Object [New Risk Object_0]

Model: invest.cmp

Generated: 22.09.17 15:18

VOI Configuration	
Decision Node	Invest decision [M2]
Uncertainty Nodes	stock market increase [stock]
Utility Node	Total utility [M9]
Optimisation Type	maximum
Scenario	Scenario 1

Total build time: 3092084 ms

Expected Maximum Value (Utility|Decision) – EMV

Expected Value Given Perfect Information – EV|PI

Expected Value of (Partially) Perfect Information – EV(P)PI

Click on the name of an Uncertainty node to see detailed utility table showing utility values per each combination of Uncertainty and Decision states.

EMV		**221.008**
stock market increase [stock]	**EV|PI**	1048.329
	EV(P)PI	827.322

		Invest decision		
		Shares	Bank deposit	Mixed
stock market increase	-77.8528137423857 - -60.88225099390856	-9303.44	**198.806**	-3951.659
	-60.88225099390856 - -43.91168824543142	-7606.353	**198.806**	-3527.277
	-43.91168824543142 - -35.42640687119285	-6333.58	**198.806**	-3208.96
	-35.42640687119285 - -31.183766184073566	-5697.167	**198.806**	-3044.106
	-31.183766184073566 - -26.941125496954278	-5272.909	**198.884**	-2901.84
	-26.941125496954278 - -24.823484809834994	-4954.888	**198.388**	-2728.716
	-24.823484809834994 - -22.70584412271571	-4743.125	**198.753**	-2557.018
	-22.70584412271571 - -20.588203435596427	-4531.356	**198.247**	-2307.214
	-20.588203435596427 - -18.47056274847714	-4319.603	**198.555**	-1960.101
	-18.47056274847714 - -16.352922061357855	-4107.835	**198.294**	-1523
	-16.352922061357855 - -14.23528137423857	-3896.073	**198.41**	-1012.452
	-14.23528137423857 - -12.117640687119284	-3684.295	**198.179**	-448.775
	-12.117640687119284 - -10.0	-3472.546	**198.294**	-127.737
	-10.0 - -9.970562748477136	-3365.184	**198.011**	-100.473
	-9.970562748477136 - -7.727922061357852	-3251.584	**198.085**	-72.541
	-7.727922061357852 - -5.485281374238568	-3027.322	**198.411**	-16.3
	-5.485281374238568 - -3.242640687119284	-2802.996	**198.563**	39.862
	-3.242640687119284 - -1.0	-212.132	**198.547**	95.771
	-1.0 - 0.0	-49.994	**198.078**	136.481
	0.0 - 1.0	49.991	**198.312**	161.386
	1.0 - 4.0000000000000036	**249.997**	198.682	211.404
	4.0000000000000036 - 7.000000000000007	**550.001**	198.332	286.236
	7.000000000000007 - 10.0	**849.997**	198.459	361.373
	10.0 - 11.746320343559644	**1087.308**	198.842	420.827
	11.746320343559644 - 13.492640687119287	**1261.944**	198.253	464.271
	13.492640687119287 - 15.238961030678931	**1436.567**	199.165	507.786
	15.238961030678931 - 16.985281374238575	**1611.208**	198.993	551.889
	16.985281374238575 - 18.73160171779822	**1785.841**	199.175	595.437
	18.73160171779822 - 20.477922061357862	**1960.468**	199.135	638.954
	20.477922061357862 - 22.224242404917504	**2135.091**	199.253	682.73
	22.224242404917504 - 23.97056274847715	**2309.726**	199.22	726.316
	23.97056274847715 - 26.09188309203679	**2503.106**	199.315	774.65
	26.09188309203679 - 28.213203435596434	**2715.238**	199.249	827.787
	28.213203435596434 - 30.334523779156076	**2927.369**	198.459	880.704
	30.334523779156076 - 32.45584412271572	**3139.501**	198.395	933.801
	32.45584412271572 - 34.57716446627536	**3351.634**	198.382	986.814
	34.57716446627536 - 36.69848480983501	**3563.769**	198.382	1039.476
	36.69848480983501 - 38.819805153394654	**3775.901**	198.382	1092.919
	38.819805153394654 - 40.94112549695429	**3988.031**	198.382	1145.949
	40.94112549695429 - 45.183766184073576	**4306.229**	198.382	1225.658
	45.183766184073576 - 49.42640687119287	**4730.492**	198.382	1331.61
	49.42640687119287 - 57.911688245431435	**5366.888**	198.383	1490.839
	57.911688245431435 - 74.88225099390857	**6639.683**	198.385	1808.904

EV|PI = 0 * 198.806 + 0 * 198.806 + 0.001 * 198.806 + 0.002 * 198.806 + 0.005 * 198.884 + 0.004 * 198.388 + 0.006 * 198.753 + 0.008 * 198.247 + 0.01 * 198.555 + 0.013 * 198.294 + 0.017 * 198.41 + 0.022 * 198.179 + 0.026 * 198.294 + 0 * 198.011 + 0.034 * 198.085 + 0.04 * 198.411 + 0.046 * 198.563 + 0.051 * 198.547 + 0.025 * 198.078 + 0.025 * 198.312 + 0.08 * 249.997 + 0.084 * 550.001 + 0.084 * 849.997 + 0.047 * 1087.308 + 0.045 * 1261.944 + 0.043 * 1436.567 + 0.04 * 1611.208 + 0.037 * 1785.841 + 0.033 * 1960.468 + 0.029 * 2135.091 + 0.026 * 2309.726 + 0.027 * 2503.106 + 0.022 * 2715.238 + 0.017 * 2927.369 + 0.014 * 3139.501 + 0.01 * 3351.634 + 0.008 * 3563.769 + 0.006 * 3775.901 + 0.004 * 3988.031 + 0.005 * 4306.229 + 0.002 * 4730.492 + 0.001 * 5366.888 + 0 * 6639.683 = 1048.329

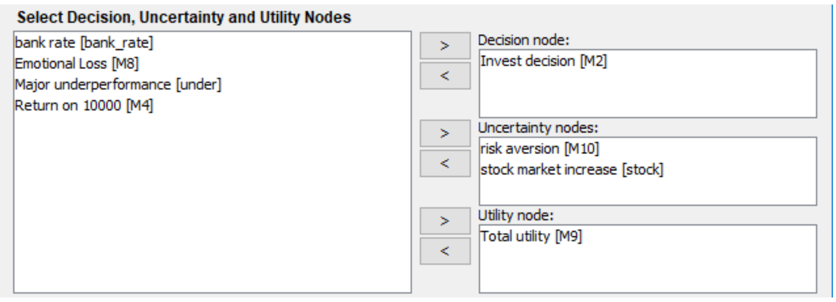
EV(P)PI = 1048.329 - 221.008 = 827.322

Figure 135 HTML report with results of value of information analysis for continuous uncertainty node

11.3 Value of information analysis for multiple uncertainty nodes

It is also possible to run value of information analysis for multiple uncertainty nodes at once. In this case the analysis is performed separately and sequentially for each uncertainty node, i.e. as if a single uncertainty node is selected but then a single report is created for all selected uncertainty nodes. Please note that optimum decisions are still selected for a single uncertainty node and not on their set at once.

1. Follow the steps explained in Section 11.1. As the only difference, set the uncertainty variables to “risk aversion” and “stock market increase” (Figure 136).



Select Decision, Uncertainty and Utility Nodes

bank rate [bank_rate]
Emotional Loss [M8]
Major underperformance [under]
Return on 10000 [M4]

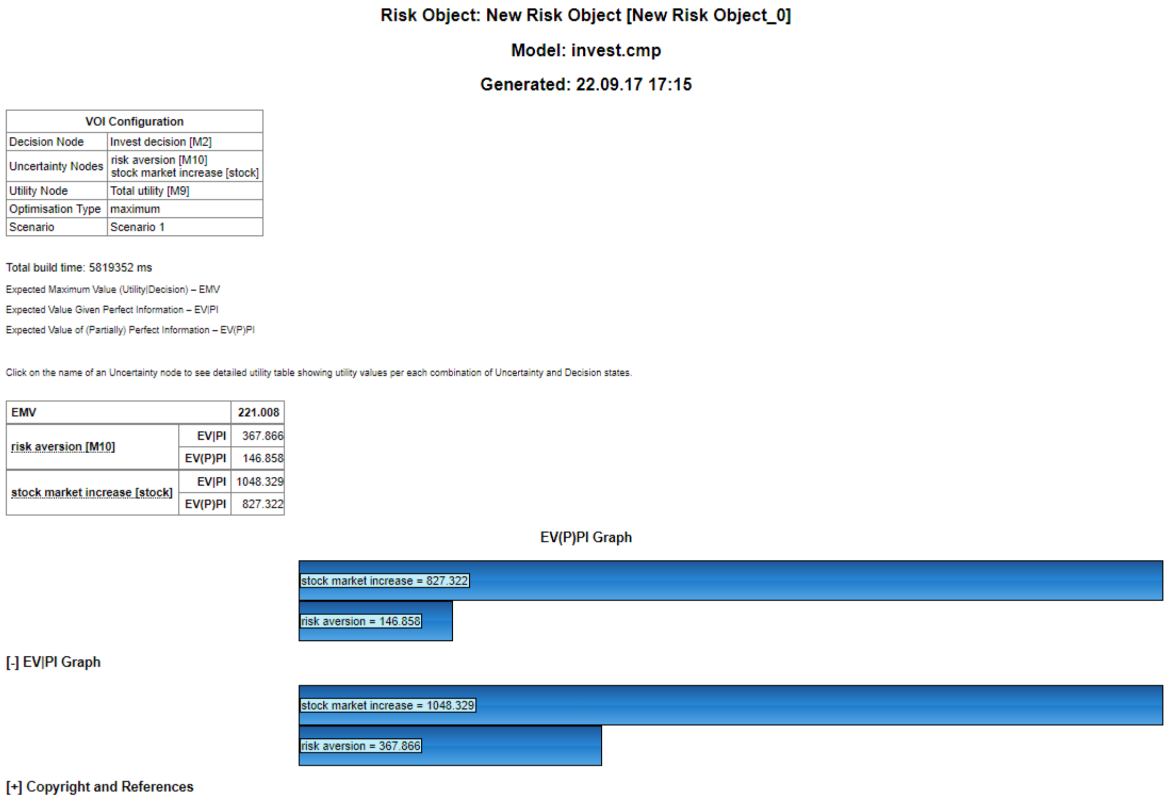
Decision node:
Invest decision [M2]

Uncertainty nodes:
risk aversion [M10]
stock market increase [stock]

Utility node:
Total utility [M9]

Figure 136 Selected Decision, Uncertainty and Utility Nodes (multiple uncertainty nodes)

2. Upon completion of calculations the generated HTML report shows that expected value given perfect information (EV|PI) has been calculated for a two uncertainty nodes (Figure 137). In addition, the report contains two bar graphs – illustrating values of EV(P)PI and EV|PI for all selected uncertainty nodes.



Risk Object: New Risk Object [New Risk Object_0]

Model: invest.cmp

Generated: 22.09.17 17:15

VOI Configuration	
Decision Node	Invest decision [M2]
Uncertainty Nodes	risk aversion [M10] stock market increase [stock]
Utility Node	Total utility [M9]
Optimisation Type	maximum
Scenario	Scenario 1

Total build time: 5819352 ms

Expected Maximum Value (Utility|Decision) – EMV

Expected Value Given Perfect Information – EV|PI

Expected Value of (Partially) Perfect Information – EV(P)PI

Click on the name of an Uncertainty node to see detailed utility table showing utility values per each combination of Uncertainty and Decision states.

EMV		221.008
risk aversion [M10]	EV|PI	367.866
	EV(P)PI	146.858
stock market increase [stock]	EV|PI	1048.329
	EV(P)PI	827.322

Figure 137 HTML report with results of value of information analysis for discrete uncertainty node

11.4 Notes on value of information analysis

1. The main window for value of information analysis contains a menu. Choosing “Options” followed by “Simulation settings” it is possible to change these settings locally for this analysis, and changing these settings will only affect the currently open analysis window and not the model in the main agena.ai modeller window. This occurs because value of information analysis is performed on a copy of model open from the main window. As a result, also any possible discretization performed on simulation nodes (see settings explained on page 110) will only affect this temporary model, not the model open in the main window.

2. The tool enables decision and uncertainty nodes to be of any type. The utility node may be Continuous Interval, Integer Interval or Ranked.

12. Hybrid influence diagram analyser

The Hybrid Influence Diagram (HID) analyser evaluates a set of interrelated decisions given an accompanying set of observed chance variables and computes the optimal strategy over the set of decisions given the state of unknown variables. The output of the HID analyser is a Decision Tree (DT). To use the analyser, make sure that your BN model is a valid HID and is composed of Decision, Chance and Utility nodes. We recommend you adopt the graphical notation as used in our example influence diagrams models which can be found and reused from this directory "*Model Library \Basic\Influence Diagrams*".

12.1 HID constraints

While building your HID, ensure that all important nodes that you will identify as Decision, Chance or Utility are connected (Figure 138b) rather than disjoint, i.e. they belong to the same BN rather than different subnetworks (Figure 138a).

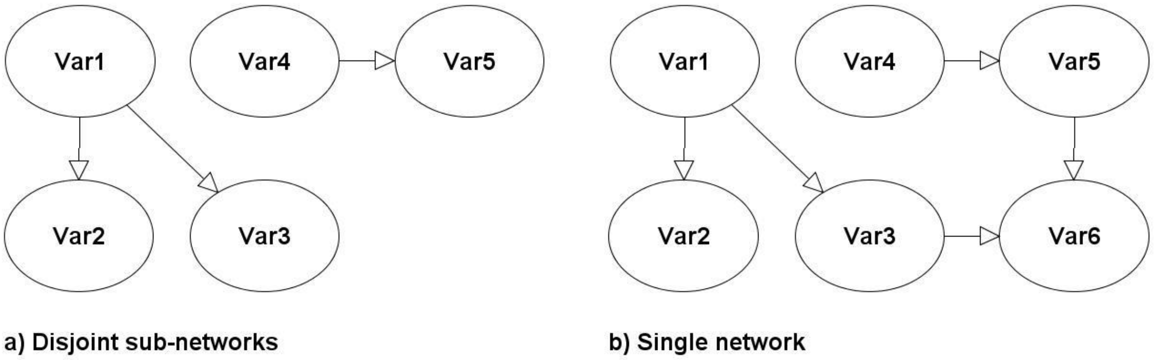


Figure 138 Disjoint sub-networks and a single network illustration

When you open the HID analyser, you will need to identify Decision and observed Chance nodes, as well as the Utility node. Note that your HID may have any number of intermediary utility nodes, but please make sure that your HID adheres to the following rules:

1) All scores and utilities are summed (or combined using an expression of your choice) in a single 'Utility' node.
2) This Utility node must be Continuous Interval or Integer Interval and must have no children.
3) Decision nodes must be of a discrete type.
4) All Decision, observed Chance and Utility nodes are in the same network and not in disjoint networks.
5) All Decision nodes must be connected by edges on the same path, that is each Decision node must be either an ancestor or a descendant of another Decision node (see Section 12.6 regarding asymmetric HIDs).
6) Only unobserved Chance nodes can be input nodes.
7) Except in the case where a Decision node is a descendant of another Decision node - and has to incorporate asymmetry (see Section 12.6) - its NPT must be defined as Uniform, i.e. consisting of all equal probabilities. This is the normal default setting for discrete nodes with discrete parents, which you can check by looking at the manual NPT. However, an HID may contain a discrete Decision node that has a continuous parent. Normally, this would mean it was impossible to define a manual NPT. However, there is a special function *DecisionUniform()* you should use in such cases that will automatically define the decision as uniform. A detailed example is provided in Section 12.5. Note that if you add a continuous parent to a Decision node that has no other parents, then it will be

automatically interpreted as uniform, but it is strongly recommended to define the Decision node with *DecisionUniform()* expression explicitly.

12.2 Basic steps

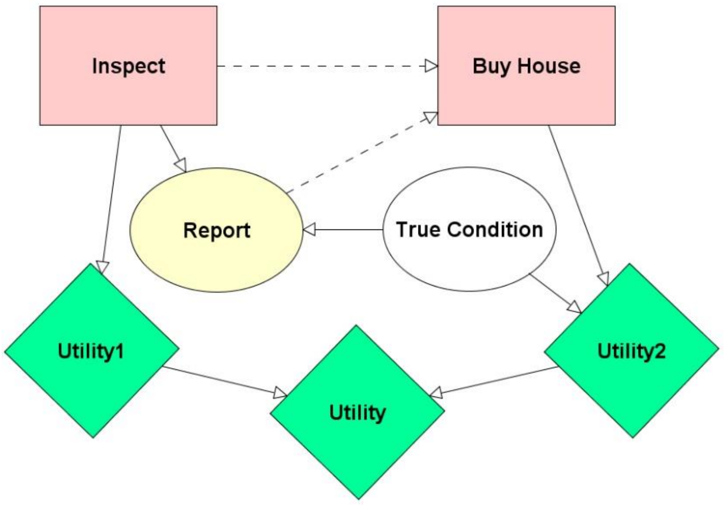
The example file used in this tutorial is in the default "*Model Library \Basic\Influence Diagrams*" folder.

1. Open the example model *House Buying.ast*.

   This model is about the decision to buy a house or not. Before you decide, you have an option of paying for a house inspection, which will produce a report about the condition of the house. If you decide to pay for this inspection, you will make a decision on whether to buy or not based on the report.

   The HID is shown in Figure 139. We suggest using a notation where Decision nodes are squares, observed Chance nodes are ellipses filled with some background colour, unobserved Chance nodes are ellipses with no (or white) background, and Utility nodes are diamonds. *Inspect* and *Report* are connected to *Buy House* by an information arc, which is essentially a normal arc with dashed line style (this is a cosmetic choice).

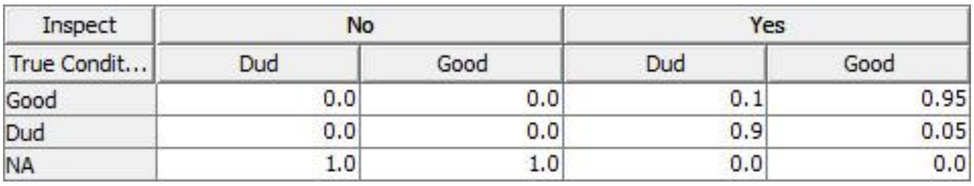
   Note that each decision in this model has an associated Utility node and total utility is calculated in the leaf Utility node.



This simple house buying example is based on one described in Korb and Nicholson "Bayesian Artificial Intelligence", CRC Press 2011

Figure 139 House Buying example HID

Also note that *True Condition* node refers to the actual unknown condition of the house and the *Report* may not accurately reflect this true condition. The accuracy of the report is encoded in its node probability table according to Figure 140.

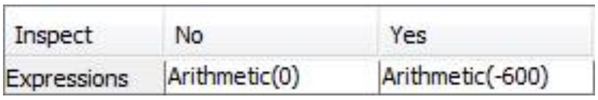


Inspect	No		Yes	
True Condit...	Dud	Good	Dud	Good
Good	0.0	0.0	0.1	0.95
Dud	0.0	0.0	0.9	0.05
NA	1.0	1.0	0.0	0.0

Figure 140 Report Accuracy

Note that we model the fact that there is no report if we decide not to pay for inspection by introducing an artificial state *NA* into the NPT of *Report*. Whenever the *Inspect* Decision node is in the state *No*, the *Report* is in the non-applicable state *NA*.

We specify that the inspection costs 600 in the node "Utility1", as shown in Figure 141.



Inspect	No	Yes
Expressions	Arithmetic(0)	Arithmetic(-600)

Figure 141 Utility of Inspection

The utility of buying the house is defined in the node *Utility2* as shown in Figure 142.

Buy House	False		True	
True Condit...	Dud	Good	Dud	Good
Expressions	Arithmetic(0)	Arithmetic(0)	Arithmetic(-3000)	Arithmetic(5000)

Figure 142 Utility of Buy House

We assume that buying a *Dud* house would costs us 3000 in lost value, while buying a *Good* house would yield 5000 in the potential resale value, while not buying does not cost us anything.

2. You can launch the **Hybrid Influence Diagram Analyser** dialog as shown in Figure 143.

   Specifically, you can do so either via the Tools menu or, more quickly, click the toolbar menu icon

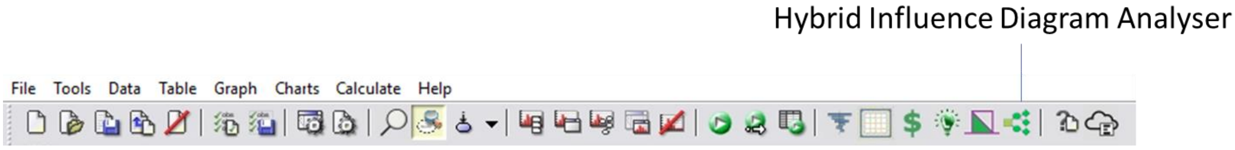
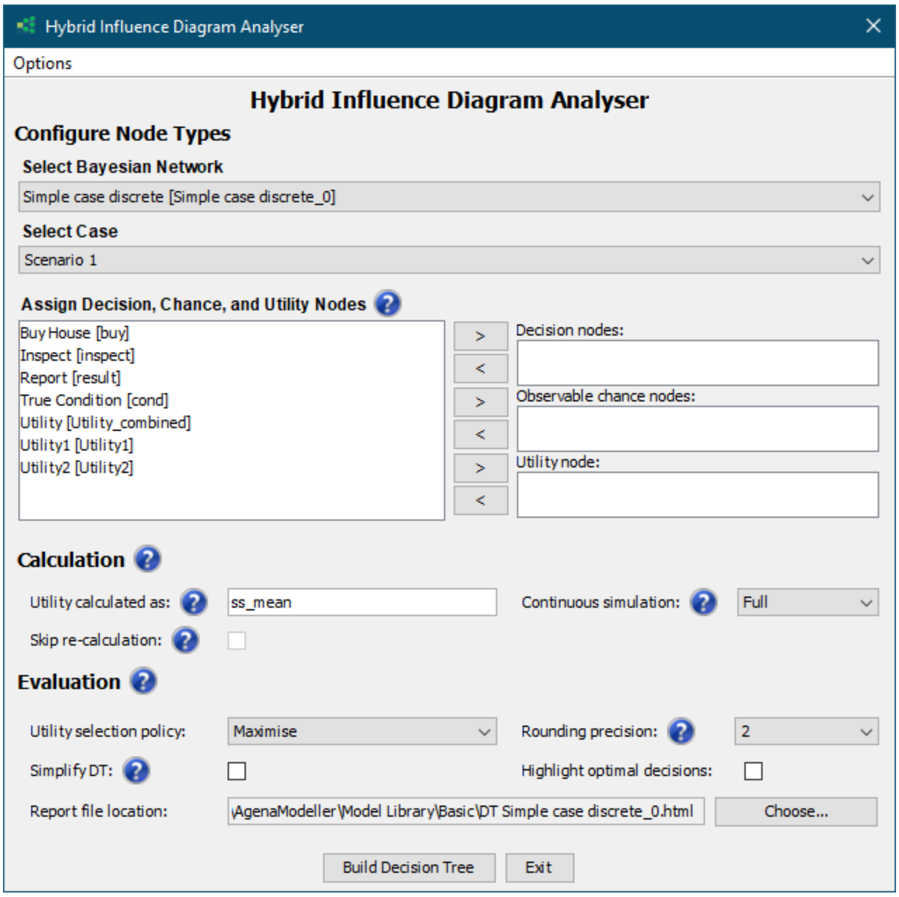


Figure 143 Accessing the HID Analyser

3. You will then be presented with HID Analyser main screen as shown in Figure 144.



Hybrid Influence Diagram Analyser

Configure Node Types

Select Bayesian Network: Simple case discrete [Simple case discrete_0]

Select Case: Scenario 1

Assign Decision, Chance, and Utility Nodes

Buy House [buy]
Inspect [inspect]
Report [result]
True Condition [cond]
Utility [Utility_combined]
Utility1 [Utility1]
Utility2 [Utility2]

Decision nodes:

Observable chance nodes:

Utility node:

Calculation

Utility calculated as: ss_mean
Continuous simulation: Full
Skip re-calculation: ☐

Evaluation

Utility selection policy: Maximise
Rounding precision: 2
Simplify DT: ☐
Highlight optimal decisions: ☐
Report file location: \AgenaModeller\Model Library\Basic\DT Simple case discrete_0.html

Figure 144 HID Analyser main screen

4. Normally you might need to select the correct Bayesian network, but in this model, there is only one, and it is selected by default.

5. Similarly, you may need to select a particular case (if there is more than one) or choose the option *N/A (Ignore data)*, which lets you run HID analysis without having to remove your data from the model. For this example, just leave the case selected by default.

   Note: if you have entered data into any nodes assigned as Decision, Chance, or Utility, you will be asked whether you would like to ignore this data when you click *Build Decision Tree*.

6. Next you need to assign nodes as Decision, observed Chance and Utility:

7. Select *Buy House* and *Inspect*.

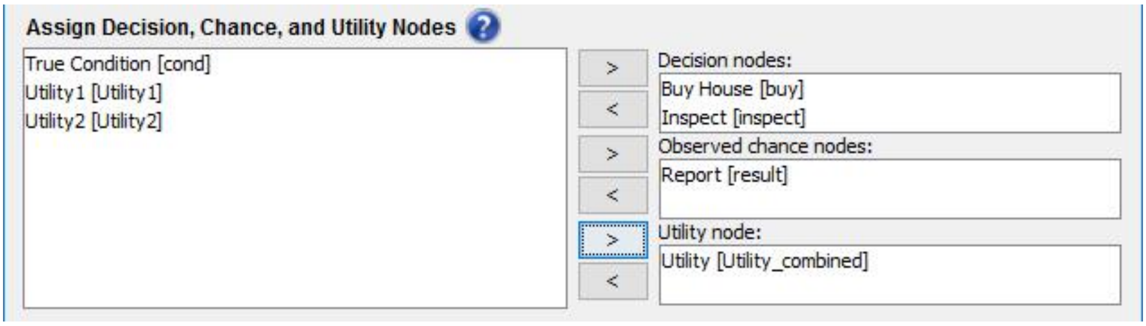
   Note: you can select multiple lines by holding CTRL key in Windows and clicking on all the nodes you want to select.

8. Click on the ">" button for *Decision nodes* to assign selected nodes as Decision.

   Note: you can select nodes on the right and click "<" to move them back to the list of unassigned nodes, which are treated as unobserved Chance nodes.

9. Similarly, assign *Report* as observed Chance node and *Utility* as Utility node as shown in Figure 145.

Note: you can only assign one node as Utility, so it should be the Utility node, which has as parents the intermediary utility nodes and has no children.



Assign Decision, Chance, and Utility Nodes

True Condition [cond]
Utility1 [Utility1]
Utility2 [Utility2]

Decision nodes:
Buy House [buy]
Inspect [inspect]

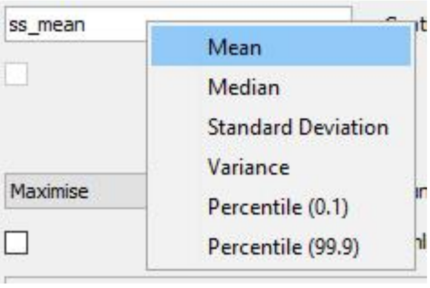
Observed chance nodes:
Report [result]

Utility node:
Utility [Utility_combined]

Figure 145 HID Nodes Assigned

10. The process of DT construction is separated into two further stages: calculation and evaluation. The calculation stage involves building the decision tree by exploring all possible state combinations for Decision and observed Chance nodes. The calculation settings only apply to this stage.

11. Utility is calculated for the leaf Utility node according to the function in the field *Utility calculated as*. You may use summary statistics accessible via right-clicking on the text field and basic arithmetic operators as shown in Figure 146.



ss_mean

Maximise

Figure 146 HID Utility Formula Context Menu

For this example, we will use simply the summary statistic mean, which in the formula is represented by the token “ss_mean”. See Table 7 for other summary statistic tokens.

Summary Statistic	Formula Value
Mean	ss_mean
Median	ss_median
Standard Deviation	ss_stdev
Variance	ss_variance
Entropy Error	ss_enterr
Percentile (0.1)	ss_percentile (0.1)
Percentile (99.9)	ss_percentile (99.9)

Table 7 Mapping of Summary Statistic Names and Formula Values

Note: you can use other arithmetic operators and common functions as available for continuous nodes with an arithmetic expression.

12. *Continuous simulation* has two options *Full* and *Semi-Static*. The *Semi-Static* option attempts to reduce computation time by making simulation nodes static after they have been calculated

dynamically initially. The *Full* option performs fully dynamic simulation for continuous nodes for every iteration of the model, and will typically take more time.

Note: while discrete nodes are unaffected by this setting, Utility nodes can be simulation nodes even in an otherwise discrete model.

13. *Skip re-calculation* allows you to re-use the calculation results, applying different evaluation settings. This check box only becomes active when these results are available either after building DT or importing configuration and calculation data.

14. In the evaluation stage optimal decisions are selected and values for Chance and Decision nodes are calculated. The decision tree is then pruned and simplified if the simplification option is chosen. Evaluation settings only apply to this stage.

15. *Utility selection policy* has two options *Maximise* and *Minimise*. When comparing two nodes, the optimal decision is selected based on this setting.

16. *Rounding precision* setting applies to utility values on nodes and probability values and ranges on arcs going out of Chance nodes.

    Note: due to various precision settings, arcs emanating from a continuous node may not have their values sum up to exactly one. If this happens, try to increase precision to make sure that values look accurate enough for your purposes.

17. You can choose to *Simplify DT* by eliminating nodes with only one incoming and one outgoing arc, and by grouping arcs going out from continuous observed Chance nodes by the next optimal policy.

    Note: this setting will normally only have an effect on HIDs with continuous Observed chance nodes or where there is decision asymmetry (see Section 12.6 for example).

18. If the *Highlight optimal decisions* check box is checked, edges corresponding to the optimal strategy (given local context such as outcome of observed Chance nodes) are shown in bold.

    For this example model you should tick this box.

19. You may choose file system location and file name for the resulting HTML report produced. This contains the DT and will open automatically in your browser when the DT is built.

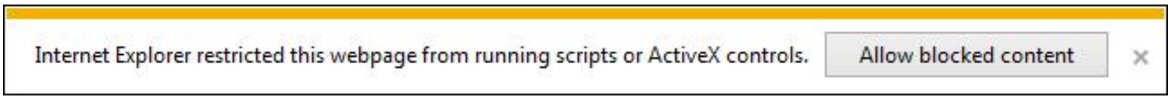
    Note: the HTML report uses JavaScript code that normally resides in the agena.ai modeller installation folder. If you simply send this report to somebody, it may not work for them if they do not have agena.ai modeller installed or their agena.ai modeller installation differs from yours.

    It is best to wait until the report is generated, and then either save and send the actual Decision Tree diagram. Alternatively press CTRL+S (or choose *File → Save as*) while in the browser page of the report, which will let you save the page with all necessary JavaScript files to a location of your choice, so you can send to somebody as an archive.

20. Once ready, press *Build Decision Tree*.

12.3 DT report web page

The DT report page will automatically open in your browser. Depending on your browser and its settings you may have to allow or enable JavaScript or ActiveX controls (e.g. Internet Explorer shows a message like in Figure 147, where you need to click *Allow blocked content*).



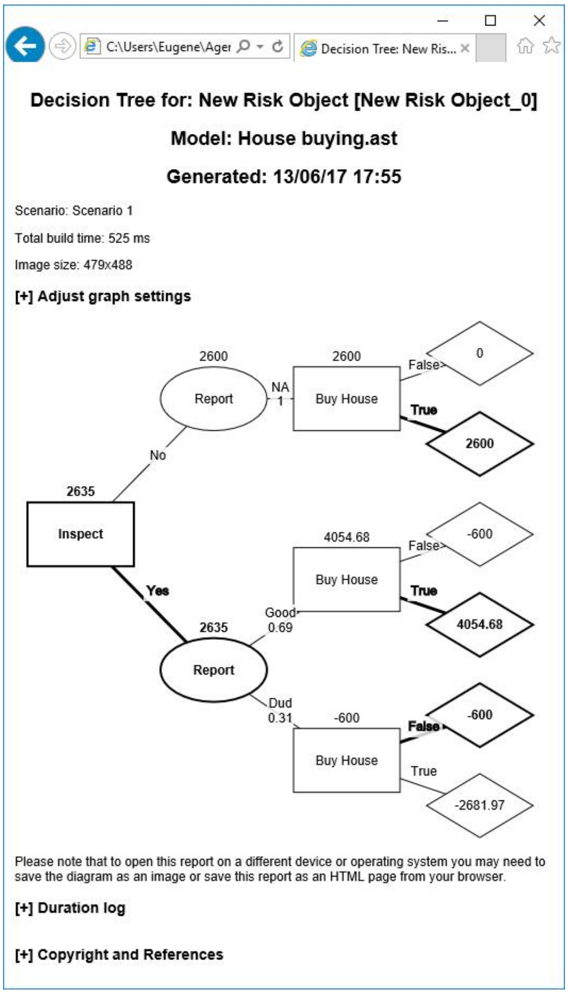
Internet Explorer restricted this webpage from running scripts or ActiveX controls.

Figure 147 Internet Explorer warning about scripts and ActiveX controls

The full page of the report looks like Figure 148. Note that it contains the model file name and names of the Bayesian network and case selected to produce this report. It also contains a time stamp of report creation and total time taken to produce this report. Note: total time taken will be much smaller if *Skip re-calculation* option was selected.

The report also contains applicable copyright notices in the bottom as well as prints locations of used libraries and relevant licenses.

The graph is first built as a vector graphics object and is then rasterised into a PNG image (except for Microsoft Internet Explorer 11 and Microsoft Edge). The “Image size” refers to the rasterised PNG image size. If you need to save the SVG version of the image, open the same report in MS IE11 or MS Edge. Alternatively, you can change the image resolution via graph settings. Larger image may look better when inserted into some text editors and zoomed in.



Decision Tree for: New Risk Object [New Risk Object_0]

Model: House buying.ast

Generated: 13/06/17 17:55

Scenario: Scenario 1

Total build time: 525 ms

Image size: 479x488

[+] Adjust graph settings

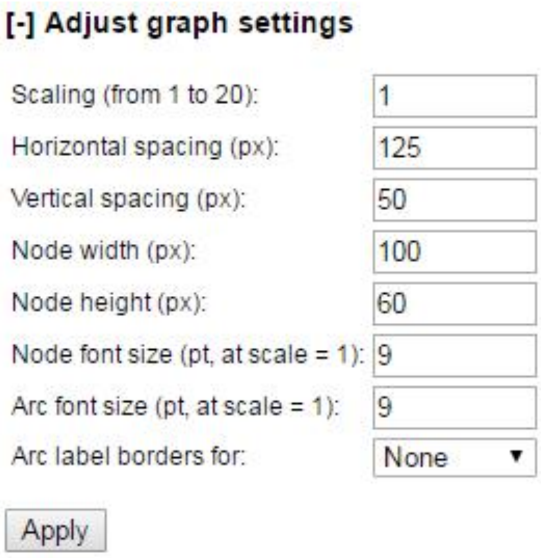
Please note that to open this report on a different device or operating system you may need to save the diagram as an image or save this report as an HTML page from your browser.

[+] Duration log

[+] Copyright and References

Figure 148 DT Web Page Report Example in Internet Explorer

Figure 149 shows the graph settings available when you expand these options by clicking *[+] Adjust graph settings*.



[-] Adjust graph settings

Scaling (from 1 to 20): 1
Horizontal spacing (px): 125
Vertical spacing (px): 50
Node width (px): 100
Node height (px): 60
Node font size (pt, at scale = 1): 9
Arc font size (pt, at scale = 1): 9
Arc label borders for: None

Figure 149 DT Graph Settings

- *Scaling* allows you to enlarge the rasterised version of the PNG image (or default scaling applied to its vector version);
- *Horizontal spacing* impacts horizontal distance between nodes.
- *Vertical spacing* impacts vertical distance between nodes (up to 0px between leaf nodes);
- *Node width* is the width of the node’s bounding rectangle in px;
- *Node height* is the height of the node’s bounding rectangle in px;
- *Node font size* is the font size in pt inside the nodes and directly above them.
- *Arc font size* if the font size in pt of all arc labels.
- *Arc label borders* lets you show a border around bold arc labels, all arc labels or none.

After changing these settings click *Apply* to redraw the diagram.

Figure 150 shows an edited version of the same DT and its graphical settings.

Finally, below the graph you can find the “Duration log” expandable section, which shows detailed calculation log where you can see how long each state combination took to calculate and total time. Note that when *Skip re-calculation* option is selected, the log may be empty as nothing was calculated.

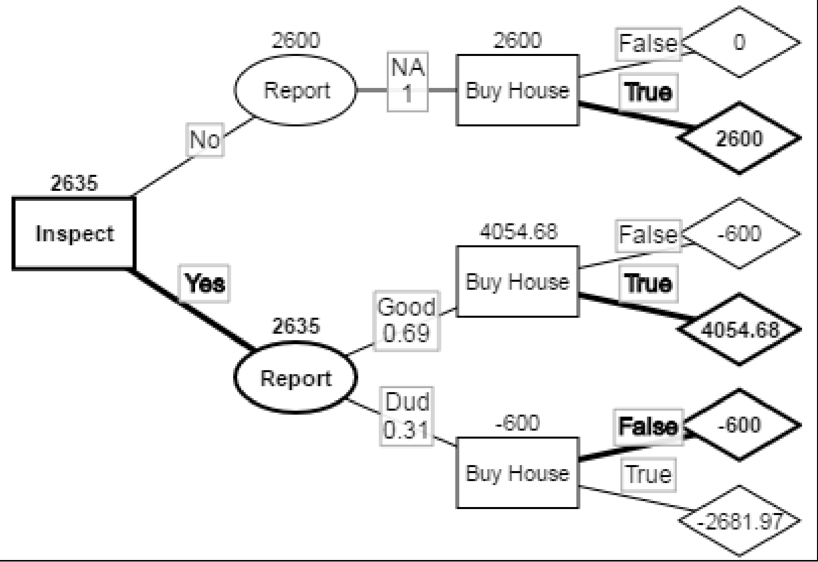
Setting	Value	Diagram
Scaling	1	
Horizontal spacing	110	
Vertical spacing	25	
Node width	70	
Node height	35	
Node font size	10	
Arc font size	9	
Arc label borders	All	

Figure 150 Modified DT Graph Settings and Graph

12.4 Settings and import/export

You can access model simulation settings from the *Options* menu of the HID Analyser window.

Additionally, you can export a HID configuration (node type assignment as well as calculation and evaluation settings) together with simulation settings and calculated data to a JSON file and import it later.

This is very useful if you are running the same model with different tweaks and don’t want to assign nodes every time or want to come back to calculated data at a later point without having to re-calculate very large models.

The option to export only becomes active after the DT diagram has been successfully built.

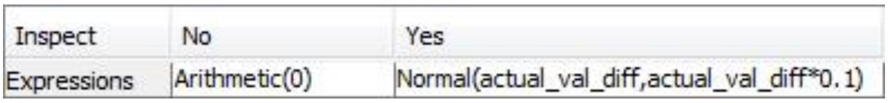
12.5 Modelling continuous observed variables

Imagine we wanted to extend this model so that the house inspection also gave an estimate of whether the house is overvalued or undervalued.

To do this, follow these steps:

1. Add a simulation unobserved Chance node *Actual Value Difference* with ID *actual_val_diff* and make it a Normal(0, 10000000).

2. Add a simulation observed Chance node *Report Valuation* with ID *V* and make it a child of *Actual Value Difference* and *Inspect*. Define it as a partitioned expression according to Figure 151. This formula implies that the report is 90% accurate.

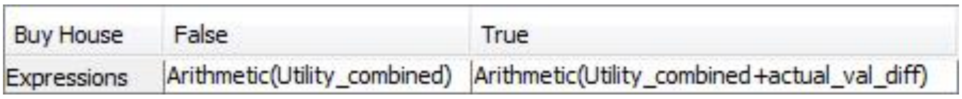


Inspect	No	Yes
Expressions	Arithmetic(0)	Normal(actual_val_diff,actual_val_diff*0.1)

Figure 151 Report Valuation [V] Expression

Then make this node a parent of *Buy House*.

3. Add a simulation Utility node *Utility Ext* with ID *Utility_Ext* with parents *Utility*, *Buy House* and *Actual Value Difference*. Define it as a partitioned expression according to Figure 152.



Buy House	False	True
Expressions	Arithmetic(Utility_combined)	Arithmetic(Utility_combined+actual_val_diff)

Figure 152 Utility Ext Expression

This formula ensures that our utility is affected by true value of the house only if we buy it. If the house is sold for more than it is actually worth, we lose value. If we pay less than the house is worth, we gain value.

4. You will notice that *Buy House* has an invalid NPT because it now has a continuous parent and can no longer have a manual NPT. In order to define it as uniform, set its NPT Editing Mode to "Expression" and type in the formula:

   DecisionUniform()

5. This formula will enforce a uniform distribution over the underlying NPT.

6. Now open the HID Analyser and configure it according to Table 8.

Field	**Value**
Continuous simulation	Semi-Static
Decision nodes	Buy House Inspect
Observed Chance nodes	Report Report Valuation
Utility node	Utility Ext
Simplify DT	Yes
Highlight optimal decisions	Yes

Table 8 HID Configuration for Extended House Buying Model

7. Click *Build Decision Tree*.

8. You will see that this version of a DT is larger and includes further analysis of optimal strategy, now accounting for the valuation difference.

12.6 Asymmetric HIDs

Agena.ai modeller supports asymmetric HIDs and provides a relevant example.

1. Open the example influence diagram model *Match ticket.ast*

2. This model presents a case with three potential decisions, where some decisions or their outcomes may render other decisions impossible. The premise is that you don't have a ticket for a football match. You may choose to still travel to the stadium and attempt to bribe a gateman to gain entry illegally or try to buy the ticket from a tout. The following situations are examples of asymmetry in this model:

   - Deciding not to go to the stadium makes the other two decisions inapplicable

- If you decide to try bribing the gateman and get arrested or are let through the gates, buying from a tout is no longer an option
- On the other hand, if you are turned away from the gate or decide not to try a bribe, buying from tout is a valid option.

3. Open the HID Analyser and configure according to Table 9.

Field	**Value**
Continuous simulation	Semi-Static
Decision nodes	Bribe gateman Buy from tout Go without ticket
Observed Chance nodes	Get into match Get ticket Outcome
Utility node	Utility
Simplify DT	Yes
Highlight optimal decisions	Yes

Table 9 HID Configuration for Match Ticket Model

Once configured, press *Build Decision Tree*.

4. You will now need to confirm that asymmetry is intentional.

5. Asymmetry in DT is very clear, and provides an easy way for following the optimal strategy.

6. You may untick *Simplify DT* and tick *Skip re-calculation* in order to see more details and visualise non-applicable event sequences.

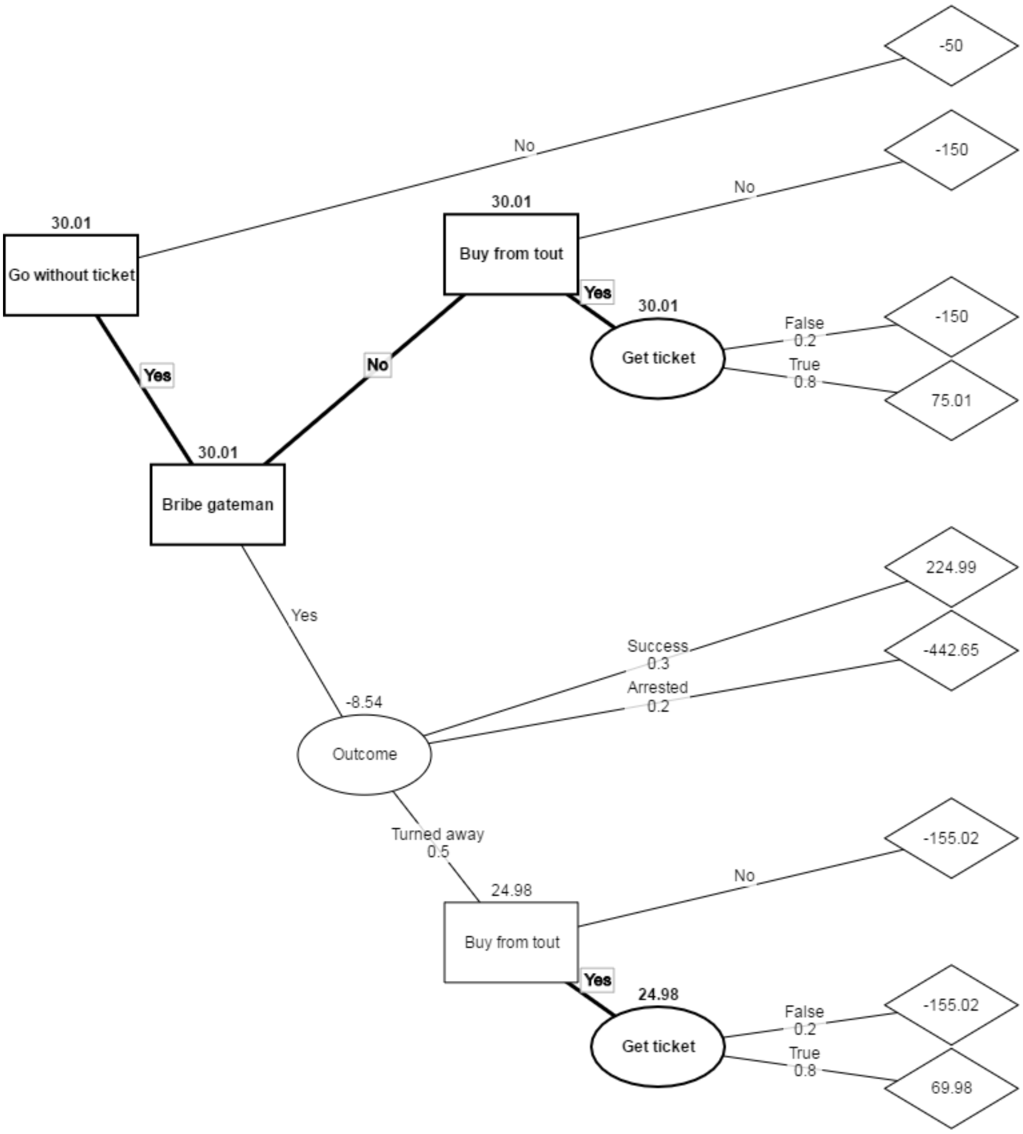


Figure 153 Asymmetrical DT for Match Ticket Model

PART B: Reference section

13. User interface

13.1 The Graph view

In addition to the Graph view commands described earlier these are also available on the Graph view.

- **Arrange** This menu allows grouping and depth positioning of the various Graph view view components.
    - **Arrange > Group** This will group all selected objects into a single object; from then on selecting any object within the group will select the whole group.
    - **Arrange > Ungroup** If the object selected was a previously created group then this function menu will ungroup the selected objects.
    - **Arrange > Move to back** This will move the selected item(s) to the back of the Graph vie.
    - **Arrange > Move to front** This will move the selected item(s) to the front of the Graph view.
    - **Arrange > Align Left** This brings all selected objects into line with the leftmost one.
    - **Arrange > Align Right** This brings all selected objects into line with the rightmost one.
    - **Arrange > Align Center** This brings all selected objects into line with the horizontal midpoint of the leftmost and rightmost selected objects.
    - **Arrange > Align Top** This brings all selected objects into line with the topmost one.
    - **Arrange > Align Bottom** This brings all selected objects into line with the bottommost one.
    - **Arrange > Align Middle** This brings all selected objects into line with the horizontal midpoint of the topmost and bottommost selected objects.
    - **Arrange > Distribute Horizontally** This distributes the selected objects horizontally
    - **Arrange > Distribute Vertically** This distributes the selected objects vertically
- **Selection** This menu provides two options for the quick selection and de-selection of Graph view view objects:
    - **Selection > All** This selects all objects on the Graph view
    - **Selection > None** This deselects all selected objects
- **Zoom** This menu provides functions for zooming in and out of the Graph view without having to be in the zoom mode.
    - **Zoom > To fit screen** This will fit the entire Graph view into the available screen space.
    - **Zoom > In by 30%** This will make the Graph view 30% bigger.

  - **Zoom > Out by 30%** This will make the Graph view 30% smaller.

- **Save as JPEG** This will save the Graph view to a JPEG. You will be prompted to provide a name for the JPEG file.

> When you select a Bayesian network or model and copy it using Ctrl-C or by right clicking and choosing Copy a JPEG file will be created that can be pasted into any other compatible application.

- **Shape Segment** This is only available when you have shown the extra control points on a node (by holding down Shift and Alt and clicking on it) and when you right-click on one of the control points.
  - **Insert Segment** This inserts a new line into the shape at the point where you have clicked.
  - **Delete Segment** This deletes the shape segment on which you have clicked.
- **Cut** This will tag the selected objects as having been ‘cut’. This will have no effect until the paste operation is invoked.
- **Copy** This will tag the selected objects as ‘copied’. This will have no effect until the paste operation is invoked.
- **Paste** The paste operation will take the last copied or cut objects and paste them onto the selected Graph view. (This does not have to be the same Graph view as the one that the objects were cut or copied from.)
- **Delete** This will remove the selected objects. Deleting nodes or edges is not reversible short of reloading your model.

> When a function is selected from the right-hand menu it will be applied to ALL selected objects. In the case of the properties dialog the settings will apply to all objects selected at the time it was opened. If a grouped object is selected, then the settings will apply to all objects within that group. If the shift key is held down during the right-click that brings up the popup menu, then only the object directly under the mouse pointer will be modified by the subsequent function or dialog. This is especially useful when you want to modify the settings of objects that reside within a grouping, rather than all objects in the group.

> You cannot copy and paste between two open agena.ai modeller windows. If you want to copy a previous model into a new model the best way to do this is by importing.

13.2 Display of data on the graph view

The Graph view can display labels anchored to node objects when data are present. Figure 154 displays an example of a node with an data in it.

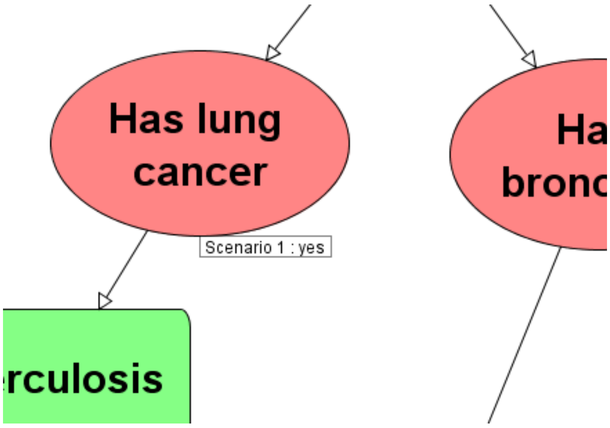


Figure 154 A node on the Graph view showing data

The data label will only be present if the case in which the data is present is active. Thus, if an data exists for a node in case X, and case X is active, then the data label will be shown on the Graph view; otherwise, it will not.

13.3 Graph view toolbar

The Graph view toolbar is in the top left-hand corner of the Graph view, as shown in Figure 155. It is used for changing the mode in which the mouse is currently operating.

Figure 155 The Graph view toolbar

The following modes are supported:

- **Pointer** In this mode, the Graph view objects can be manipulated. This is the default mode.

- **Zoom** The Graph view can be 'zoomed in' on in one of two ways. You can left-click anywhere and the map will zoom in by a factor of 1.3 and centre on the point clicked. Alternatively, you can 'pull out' an area in the same way that you area-select objects. 'Zooming out' is achieved by right-clicking in the Graph view. The Graph view will zoom out by a factor of 0.7, and centre on the point clicked. The Graph view will remain in Zoom mode until another mode is selected from the toolbar.

- **New Label** To create a new label, click on this icon then on an empty part of the Graph view. A new label will be created at the point where you click. After creating the label the Graph view will revert back to the Pointer mode. Labels are useful for annotating Graph views.

- **New Image** To create a new image, click on this icon then on an empty part of the Graph view. A dialog will open asking you to browse to a location containing the jpeg image you wish to use. After opening the image, the Graph view will revert back to the Pointer mode.

- **New Edge** To create a new edge, click on this icon then on the node from where the edge should start. A floating edge will appear, anchored to the start node. Complete the join by clicking on the target node. The edge will then be created, and the two nodes will appear linked. After creating the edge, the Graph view will revert back to the Pointer mode.

- **New Node** To create a new node, click on this icon and then click on an empty part of the Graph view. A new node will be created at the location of your mouse pointer. After creating the node the mouse icon will revert back to Pointer mode.

- **New Simulation Node** To create a new simulation node, click on this icon and then click on an empty part of the Graph view. A new node will be created at the location of your mouse pointer. After creating the node, the mouse icon will revert to Pointer mode. By default, a new simulation node is given a Normal Distribution.

Nodes and edges are deleted by right-clicking on the relevant object and selecting the delete function from the menu that appears. Alternatively, the delete key can be used.

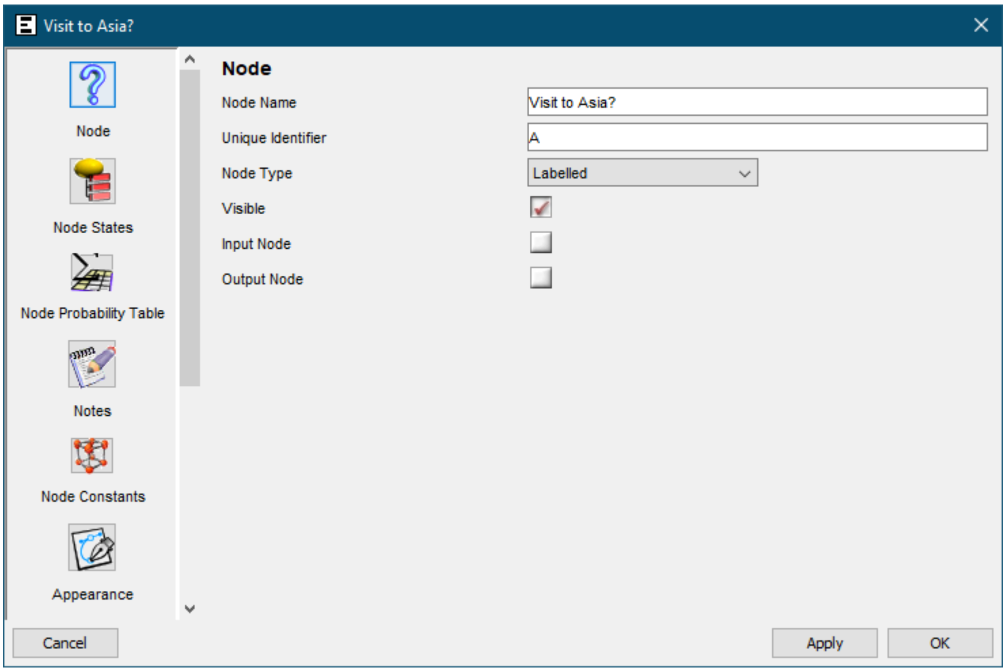
If you are in zoom mode and you want to quickly fit the contents of the Graph view to the available screen space, then double click anywhere on the Graph view.

Once you have finished in zoom mode it is advisable to select Pointer mode right away; if you don’t the you might forget you are in zoom mode.

13.4 The node and edge properties dialog

A properties dialog is available from the pop-up menu on the Graph view. It allows you to modify all available properties for the selected objects.

The properties dialog that appears when a node is selected is shown in Figure 156. The property categories are shown as a vertical list on the left-hand side of the dialog. (A reduced subset of the categories is available when you right-click on an edge or other object on the graph.) Selecting a category will display the details of the properties related to the selected category.



Node

Node Name: Visit to Asia?

Unique Identifier: A

Node Type: Labelled

Visible ☑

Input Node ☐

Output Node ☐

Figure 156 The node properties dialog

The categories are specified in more detail in the following sections:

13.4.1 Node properties

The following properties are available in this category (we refer only to nodes that have been selected in a particular Bayesian network rather than Bayesian networks that have been selected in the model view; for the latter see Chapter 21):

- **Node Name** This is the ‘display’ name of the node. It will appear on charts when they are loaded.

- **Unique Identifier** The name of the node when it is used in an expression. This name must be unique. If a name is added that duplicates another node’s unique identifier, you are warned, and the old name is restored.

- **Node Type** These are:

  - **Labelled** e.g., “Red”, “Blue”, “Green”

  - **Boolean** e.g., “True”, “False”

  - **Continuous Interval** e.g. [0-10], [10-20] (where each range contains all real values in that range)

  - **Ranked** e.g., “Low”, “Medium”, “High”

  - **Integer Interval** e.g. [0 – 10], [10 – 20] (where the range contains only the integers in that range)

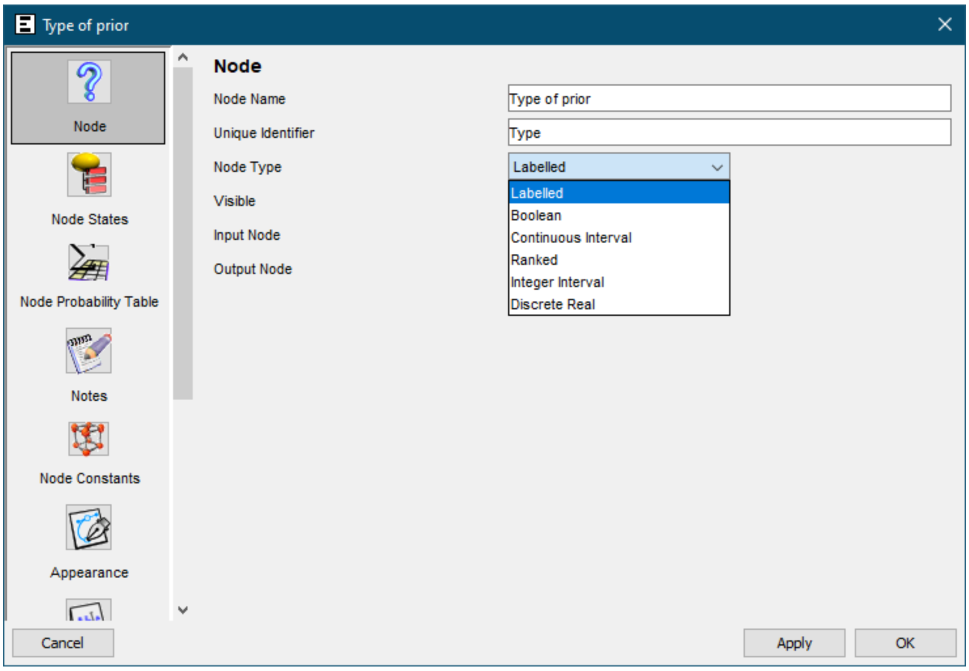
  - **Discrete Real** e.g., 0, .5, -2

- **Visible** This determines whether the node is visible or not.

- **Input** A node can be defined as an input node if it has no parents. This is relevant when joining models.
- **Output** A node can be defined as an output node if it has no children. This is relevant when joining models.
- **Simulation Node** Defines whether the node should be simulated.

  The “Simulation” check box is only available for Continuous Interval and Integer Interval nodes and only if they have expressions declared upon them.

Node type can be selected from the Node properties as shown in Figure 157.



Type of prior

Node

Node Name: Type of prior

Unique Identifier: Type

Node Type: Labelled

Visible

Input Node

Output Node

Figure 157 Choosing node type

13.4.2 Edge properties

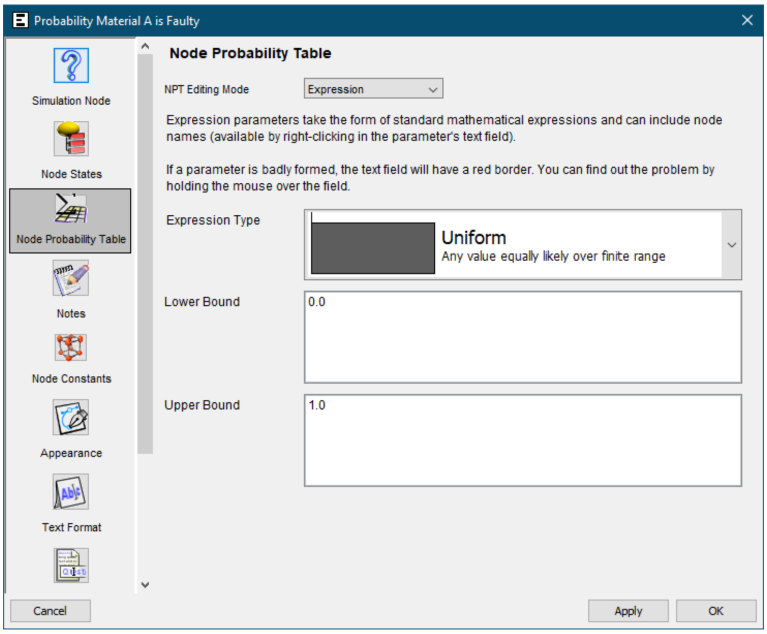
Edge formatting properties are available for edges and can be accessed by right clicking on the edge and selecting Properties. The main benefit of this is the ability to annotate edges with text and use this to better communicate the meaning of relationships between nodes.

13.4.3 Node probability table

There are three ways of specifying node probability tables in agena.ai modeller:

- Manual
- Expression
- Partitioned Expression

You can view the default NPT values of any node by selecting the node in the Graph view and selecting ‘Node Probability Table’ from the node properties dialog. The dialog displayed depends upon the selection of the NPT editing mode in the node properties dialog as shown in Figure 158.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: Uniform
Any value equally likely over finite range

Lower Bound: 0.0

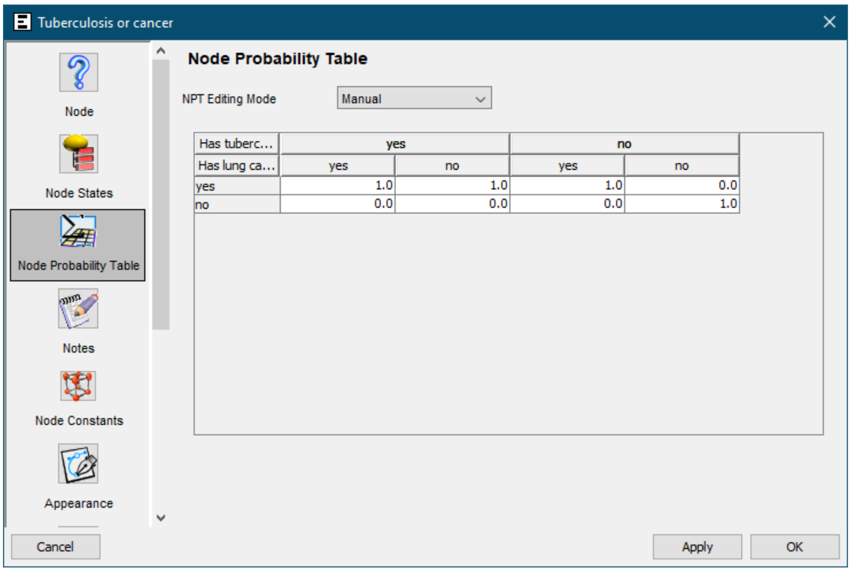
Upper Bound: 1.0

Figure 158 NPT editing mode

Figure 159 shows the NPT dialog for a node that is manually defined. Here you can simply enter the probability values directly. In addition, you can copy and paste between cells and from MS Excel using the Copy and Paste options on the menu available when you right-click on the table.

Note: For nodes with one or more parents, the order in which you add edges determines the structure of the NPT. Be aware of this when cutting and pasting to and from MS Excel.

Important: When you complete an NPT table manually, be sure to press Return or move the focus away from the last cell you edit before pressing Apply or OK, otherwise the value you type will not be registered.



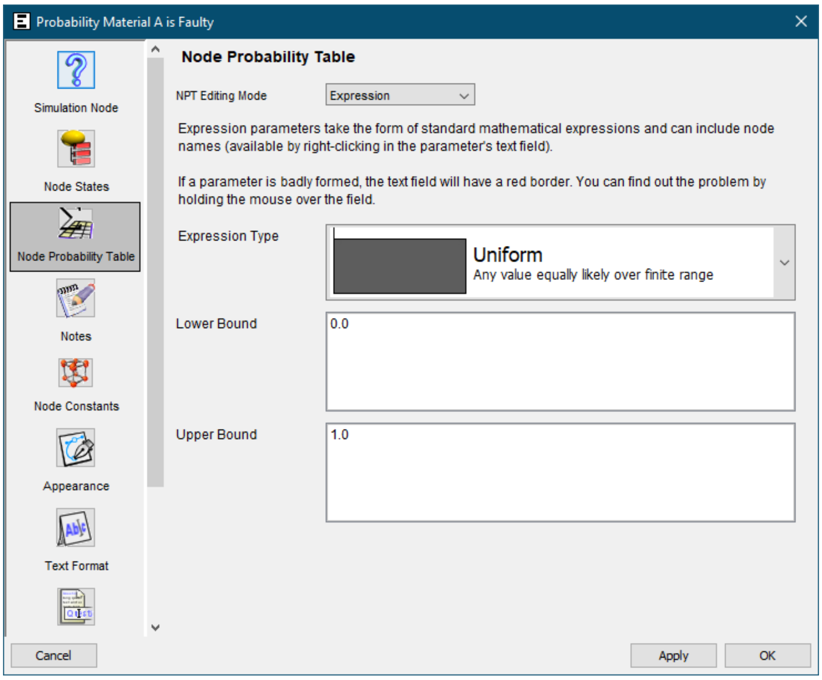
Node Probability Table

NPT Editing Mode: Manual

Has tuberc...	yes		no	
Has lung ca...	yes	no	yes	no
yes	1.0	1.0	1.0	0.0
no	0.0	0.0	0.0	1.0

Figure 159 NPT dialog for a node defined with NPT editing mode 'Manual'

Figure 160 shows the NPT dialog for a node whose NPT editing mode is 'Expression'.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

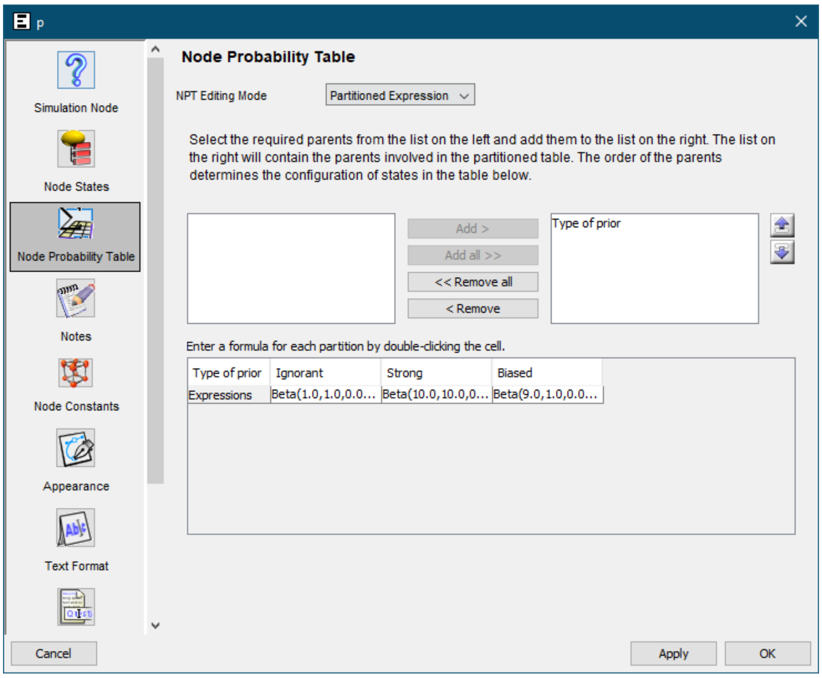
Expression Type: Uniform
Any value equally likely over finite range

Lower Bound: 0.0

Upper Bound: 1.0

Figure 160 NPT dialog for a node defined with NPT editing mode ‘Expression’

For nodes whose expressions are conditional on parent nodes the ‘Partitioned Expression’ mode is used as shown in Figure 161.



Node Probability Table

NPT Editing Mode: Partitioned Expression

Select the required parents from the list on the left and add them to the list on the right. The list on the right will contain the parents involved in the partitioned table. The order of the parents determines the configuration of states in the table below.

Type of prior

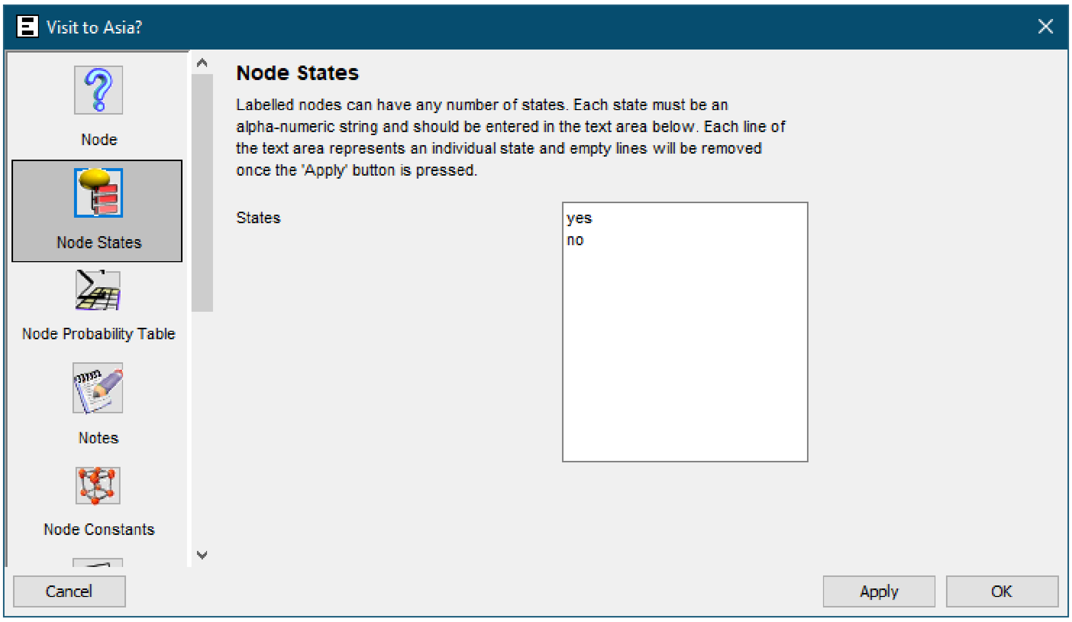
Enter a formula for each partition by double-clicking the cell.

Type of prior	Ignorant	Strong	Biased
Expressions	Beta(1.0,1.0,0.0...	Beta(10.0,10.0,0...	Beta(9.0,1.0,0.0...

Figure 161 NPT dialog for a node defined with NPT editing mode ‘Partitioned Expression’

13.4.4 Node states

The node states dialog allows you to add and remove states from a node. The dialog that appears depends on the node type defined for the node (Ranked, Continuous Interval, Discrete Real, etc.). Figure 162 and Figure 163 show the different dialogs available.



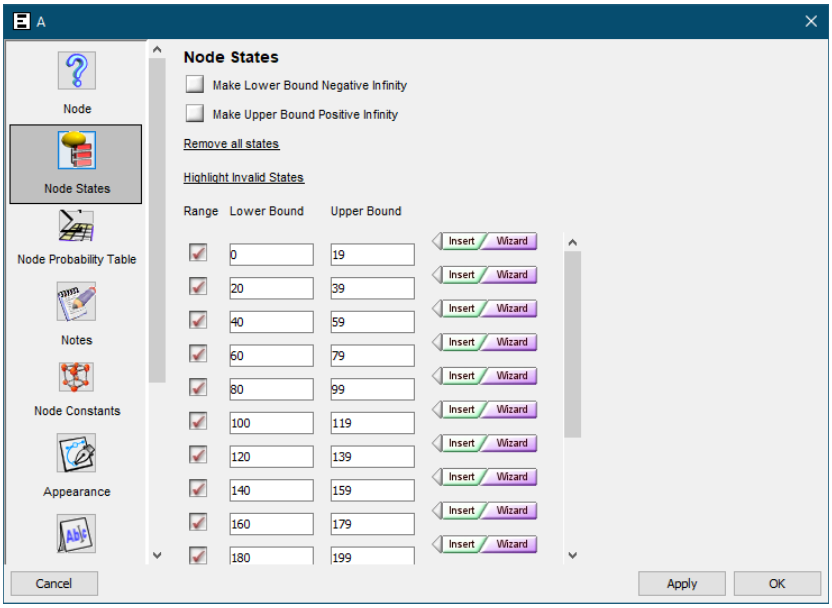
Node States

Labelled nodes can have any number of states. Each state must be an alpha-numeric string and should be entered in the text area below. Each line of the text area represents an individual state and empty lines will be removed once the 'Apply' button is pressed.

States
yes
no

Figure 162 The node states dialog for Ranked, Labelled and Discrete Real types

In the case of Ranked, Labelled and Discrete Real nodes, as shown in Figure 162, you type the states directly into the text area so that each state occupies a single line. When you click on OK or Apply a set of states corresponding to the list in the text area is created.



Node States

☐ Make Lower Bound Negative Infinity

☐ Make Upper Bound Positive Infinity

Remove all states

Highlight Invalid States

Range	Lower Bound	Upper Bound
☑	0	19
☑	20	39
☑	40	59
☑	60	79
☑	80	99
☑	100	119
☑	120	139
☑	140	159
☑	160	179
☑	180	199

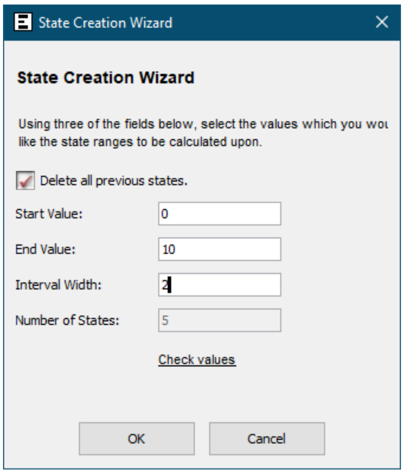
Figure 163 The node states dialog for Continuous and Integer Interval node types

In the case of Continuous Interval or Integer Interval nodes you define the upper and lower bounds of each state as shown in Figure 163. Validation of the numeric values you enter is performed when the OK or Apply buttons are pressed. If they are not valid then the values are reset to their starting values. The upper bounds of all but the last state are calculated automatically. In order to be valid the values must ascend in value from the first to the last state, while the upper bound of the last state must be equal to or exceed the lower bound of the last state.

For Continuous Interval and Integer Interval nodes we can insert states by pressing the "Insert" button and we can add multiple states using the state creation wizard by pressing the "Wizard" button. These buttons are located alongside each state.

The state creation wizard is shown in Figure 164. Here specify the start and end value and then either specify an interval width or the number of states required. agena.ai modeller then generates the appropriate evenly spaced state intervals. You can choose to delete any of the states that existed on the node beforehand, thereby ensuring that the states generated by the wizard completely define the node. In addition, if you have specified the interval width, you can click on the 'Check values' link to see how many states this choice will result in. Using the same link, you can check the interval width if you supply a fixed number of states.

Note that node states declared in this way for continuous nodes are static and do not change during the calculation process, thus running the risk of being inaccurate. Except in special circumstances, we recommend you use simulation nodes, whose states are dynamic and do not need to be changed. See Chapter 5 for more details.



State Creation Wizard

Using three of the fields below, select the values which you wou
like the state ranges to be calculated upon.

☑ Delete all previous states.

Start Value: 0
End Value: 10
Interval Width: 2
Number of States: 5

Figure 164 The state creation wizard

It is also possible to specify a single number for one or more states instead of a range. This is achieved by clicking the "Range" check box shown next to the appropriate state(s) when the dialog is launched and entering the required value.

13.4.5 Constants

This dialog allows you to define constants to be used in expressions on the specified node.

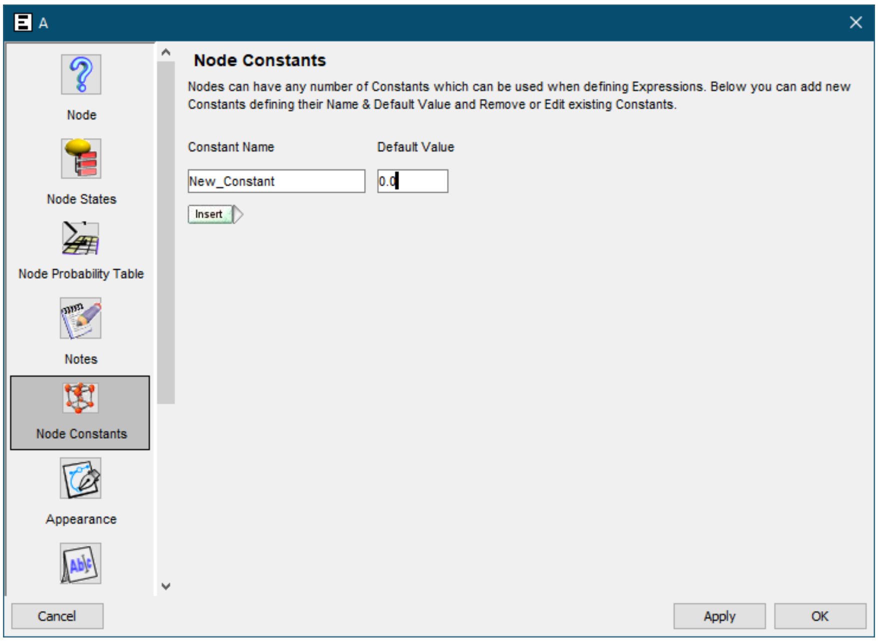
You might want to create constants for a number of reasons:

- As a value that can be overwritten at run time in different cases (for example, the constant might represent a 'weight' that changes in different cases)
- As local constant values used in an expression declared for that node

- As a global value accessible to other nodes

To create a constant simply access the Node Properties dialog, choose “insert” and agena.ai modeller will create a new constant and assign a default value of zero. Unless overwritten the default value will be used in all subsequent calculations.

Note that constant names need to be unique in the same way that nodes do. Figure 165 shows the dialog for declaring a constant and setting a default value.



Node Constants

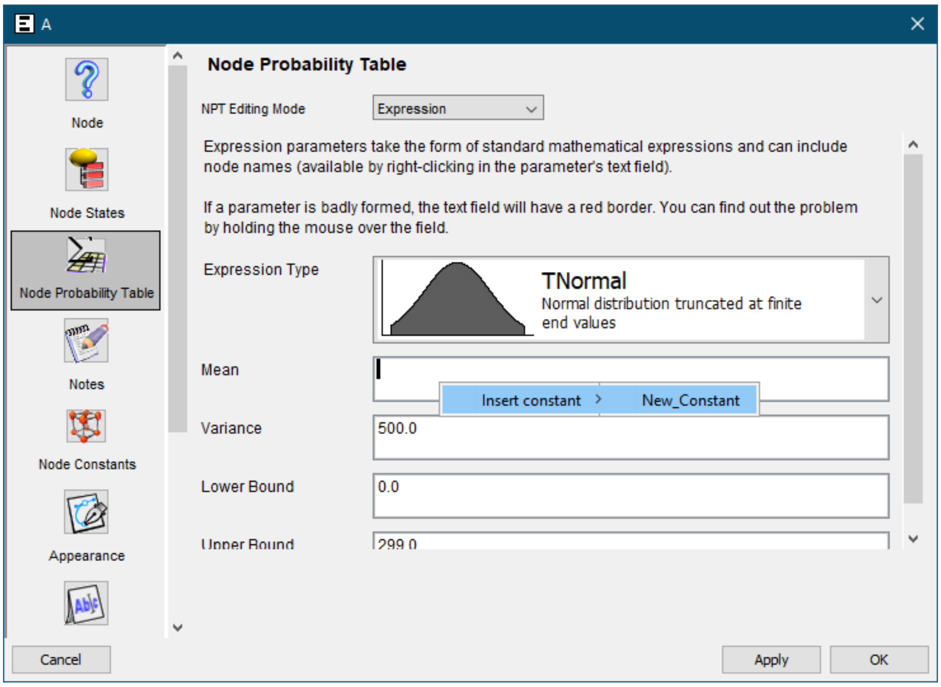
Nodes can have any number of Constants which can be used when defining Expressions. Below you can add new Constants defining their Name & Default Value and Remove or Edit existing Constants.

Constant Name: New_Constant

Default Value: 0.0

Figure 165 The node constant dialog

Once a constant has been declared it can be accessed in the Node Probability Table (the NPT editing mode needs to be set to “Expression” or “Partitioned Expression”) by right-clicking on the input parameters. This is shown in Figure 166 for a Beta distribution that we wish to set up with flexible parameters (this is available in the examples directory: *Model Library /Parameterised Distribution Using Constants*). Here two constants have been declared “alpha” and “beta” and these are inserted into the expression using right-click selection.



Node Probability Table

NPT Editing Mode: Expression

Expression parameters take the form of standard mathematical expressions and can include node names (available by right-clicking in the parameter's text field).

If a parameter is badly formed, the text field will have a red border. You can find out the problem by holding the mouse over the field.

Expression Type: TNormal – Normal distribution truncated at finite end values

Mean:

Insert constant › New_Constant

Variance: 500.0

Lower Bound: 0.0

Upper Bound: 299.0

Figure 166 Inserting a constant into an expression as a parameter

Should you wish to vary the constant values in different cases, say where you have different assumptions, you will need to do this by creating an associated entry in the Table view.

13.4.6 Notes

This dialog allows you to define notes for the specified node.

13.4.7 Appearance and text appearance properties

The attributes in these categories allow you to define how the node or edge appears on the screen.

13.4.8 Table entry

This will display the properties for the Table entry connected to the node. If there are many entries mapped to this node then there will be multiple ‘Entry’ property categories on the vertical list of the properties dialog, one for each entry.

An important difference between the Table entry dialog that appears via the Graph view and the one that is accessible from the Table view is the removal of the ‘Connect to node’ attributes. If you are editing the entry via the Graph view you cannot reconnect the entry to another node.

14. Charts

14.1 Chart overview

Charts (see Figure 167) display probability information for a selected node. They are created by selecting the relevant node in the Graph view and clicking ‘Display Chart’. They can also be displayed from the Table view by selecting the relevant entry and right-clicking. In the Graph view, you are presented with several places the chart can appear on creation. In the Table it will always appear on the chart panel.

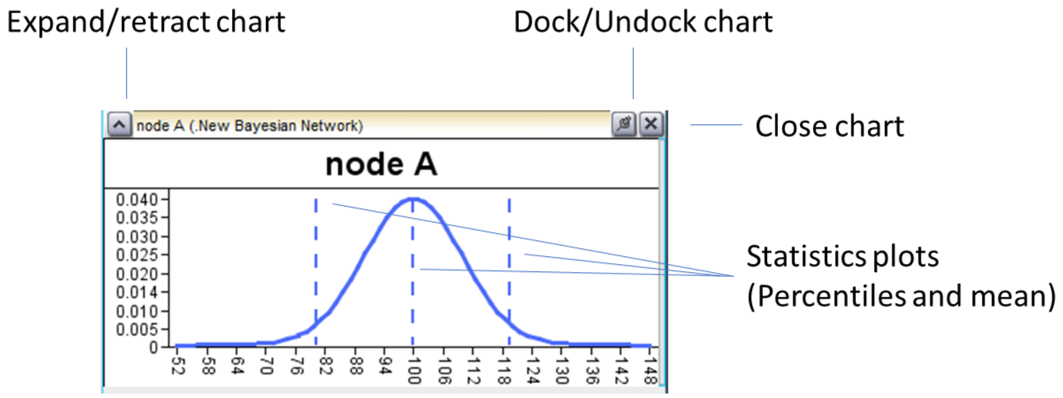


Figure 167 A Chart

You can also create charts by double-clicking on nodes in the Graph view. Note: If the node is selected, make sure you don’t double-click on any of the control points, or nothing will happen.

The newly created chart will display the data from the last calculation for the specified node. There will be one set of data per case.

When on the chart panel, charts can be expanded and retracted using the ‘Expand/Retract’ button. A minimised graph has only its title bar visible.

The legend is a ‘free-floating’ object that you can be move. Just click and hold the mouse over it, then drag it to the desired location.

Each data set plotted on the graph has its own defined graph type. In addition to the basic plots a defined set of statistical ‘values’ will also be drawn. These appear as vertical lines in the same colour as the dataset. By default, the mean (dashed line), median and lower and upper percentiles are plotted for all data sets. The visibility of these, and the properties of the data sets themselves, can be configured via the graph properties dialog.

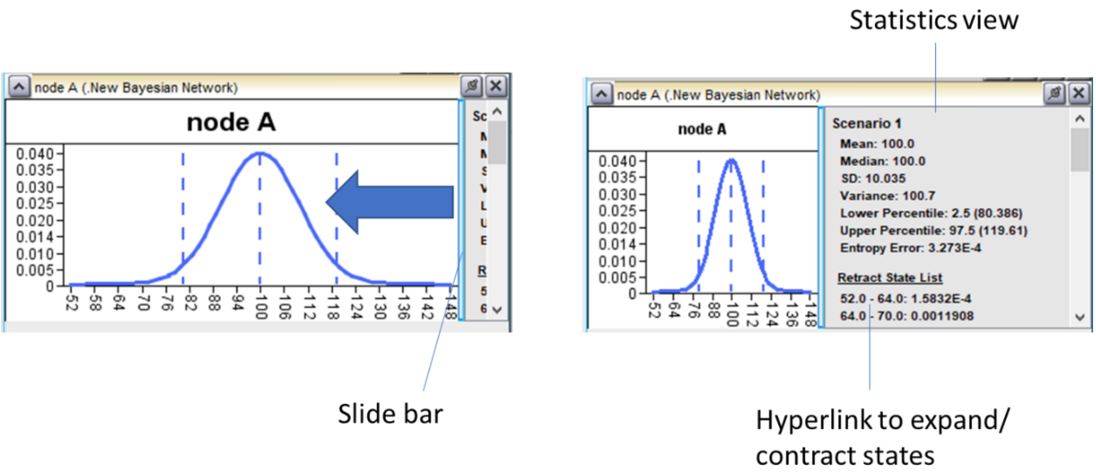
Right-clicking on the chart will open up a pop-up menu. From here the graph can be docked to a specified location (functionality which mirrors the ‘Dock / Undock’ button in the top right of the graph); it can also be moved to the top of the chart panel; it can be saved as a JPEG and it can be closed.

When you select a Bayesian network or model, showing charts, and copy it using Ctrl-C or by right clicking and choosing Copy a JPEG file will be created that can be pasted into any other compatible application and the charts will be pasted as shown in agena.ai modeller.

14.2 The statistics view

As well as the actual graph plot, each chart also has a statistical view. This contains the statistics for each case and the probability distribution displayed on the chart.

> To access the statistical view the chart must either be docked to the chart panel or docked to its own window. The statistical view cannot be accessed when the chart is docked in the Graph view.



Scenario 1
Mean: 100.0
Median: 100.0
SD: 10.035
Variance: 100.7
Lower Percentile: 2.5 (80.386)
Upper Percentile: 97.5 (119.61)
Entropy Error: 3.273E-4

Retract State List

52.0 - 64.0: 1.5832E-4
64.0 - 70.0: 0.0011908

Figure 168 The statistics view

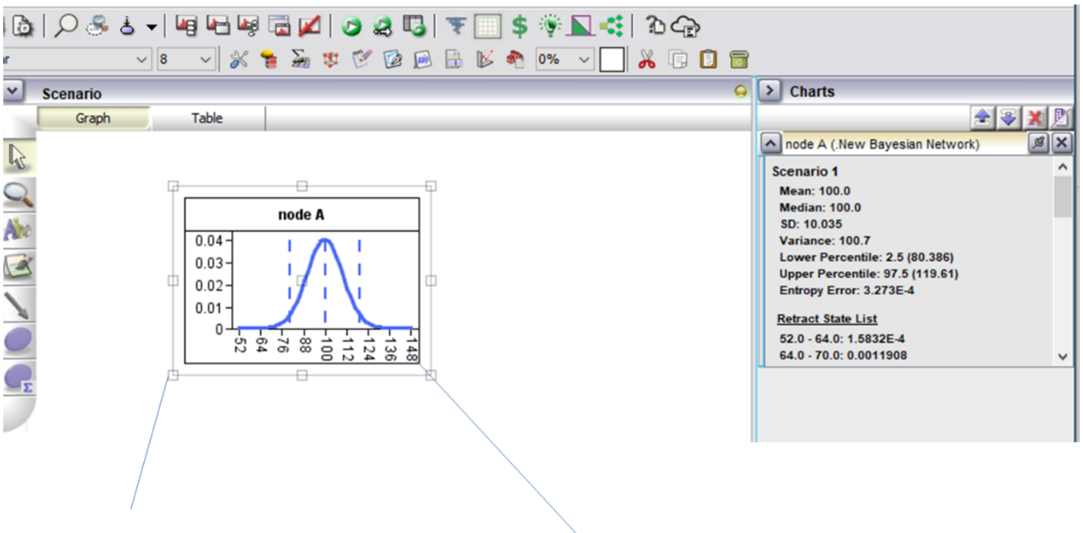
The statistics view can be accessed by using the vertical slider bar that, by default, is pushed all the way to the right of the chart. This vertical bar determines what percentage of the space available to the graph is taken by the statistical view and the actual graph plot. When the statistical view is first displayed, the state information for each case is retracted (not visible). To access it you must click on the ‘Expand State List’ hyperlink for the case for which you want full state information. Clicking a second time on this hyperlink will retract this list.

The default space taken by the statistics view when a chart is first loaded can be defined via the graph defaults dialog.

14.3 Docking charts

A chart can be docked in one of three places:

1. **On the chart panel** on the right of the application window.
2. **In its own floating window.**
3. **Docked to the graph view;** this will replace the graphic of the node in the Graph view with the chart. The statistics view will remain on the chart panel.



```
Scenario 1
  Mean: 100.0
  Median: 100.0
  SD: 10.035
  Variance: 100.7
  Lower Percentile: 2.5 (80.386)
  Upper Percentile: 97.5 (119.61)
  Entropy Error: 3.273E-4

 Retract State List
  52.0 - 64.0: 1.5832E-4
  64.0 - 70.0: 0.0011908
```

Chart can be resized using control points

Chart docked on
node in Graph view

Figure 169 Docking charts to the Graph view

To change the docked location of a graph, right click on the graph and select the docking menu; from here you can subsequently select the desired docking location. There is also a quick dock / undock button in the top right of the chart that allows you to toggle between docking on the chart panel and its own window. If the chart is docked on the Graph view, then clicking this button will immediately return the graph to the chart panel. This can be useful if you are viewing a different Graph view to the one the chart is docked on.

The Charts menu item at the top of the Agena.ai modeller application window also has functionality to dock *all* graphs to a specified docking location.

14.4 Properties of charts docked to the graph view

When charts are docked to the Graph view, they replace the node graphic that they represent. Like any other Graph view object, they can be resized and moved, but note that the node graphic will maintain the new position and size when the chart is closed or undocked from the Graph view. For this reason, if you want to make the chart larger it is worth using the zoom tool rather than resizing the chart directly.

All the menu items normally associated with a chart are available when the chart is docked to the Graph view, but are placed in a submenu called chart, available via the Chart view right click menu.

14.5 Chart sub menu

The chart sub menu is accessed by right clicking a node in the chart panel. The options available are:

- **Display Chart** This presents the options for the display of the chart
- **Enter Data** This either displays a selection or a popup dialog to enter a specific value.
- **Export Chart as CSV file** This exports the summary statistics and states for the node as a CSV file.

- **Display Summary Statistics** This displays the statistics and states for the node without showing the chart itself.
- **Save as JPEG** Saves a JPEG file of the chart.
- **Properties** Opens up the properties dialog
- **Arrange / Selection / Zoom**
- **Delete / Cut / Copy /Paste**

14.6 Defining chart defaults

Agena.ai modeller supports a powerful mechanism for defining the default configuration of charts. There are three levels at which chart properties can be configured:

- **Node** Each node can override selected model-level chart settings with its own attribute values.
- **Case** Once the chart has been created you can modify the defaults specific to each case. These modifications are persistent for the lifetime of the graph.
- **Model** This set of chart defaults are horizontal graph specific.

> ⚠ Making changes to node-level chart properties overwrites settings on case-level graph properties

In the model properties charts defaults the upper Y bound is defined as 1.0. This means that all charts created in the application will have an upper Y value of 1.0. In a particular model a user has created there is a node where the user expects the Y values to be very small, so they access the node chart defaults and override the upper Y value with a value of 0.1. From then on whenever a chart is created for that node the default upper Y value will be 0.1. The user subsequently displays the chart, and they decide that just for the lifetime of the chart they want the chart to automatically work out the upper Y. So, they open the chart properties dialog and remove the upper Y value.

14.6.1 The node properties chart defaults dialog

The node properties chart defaults dialog allows you to set node specific chart properties. Figure 170 shows this dialog.

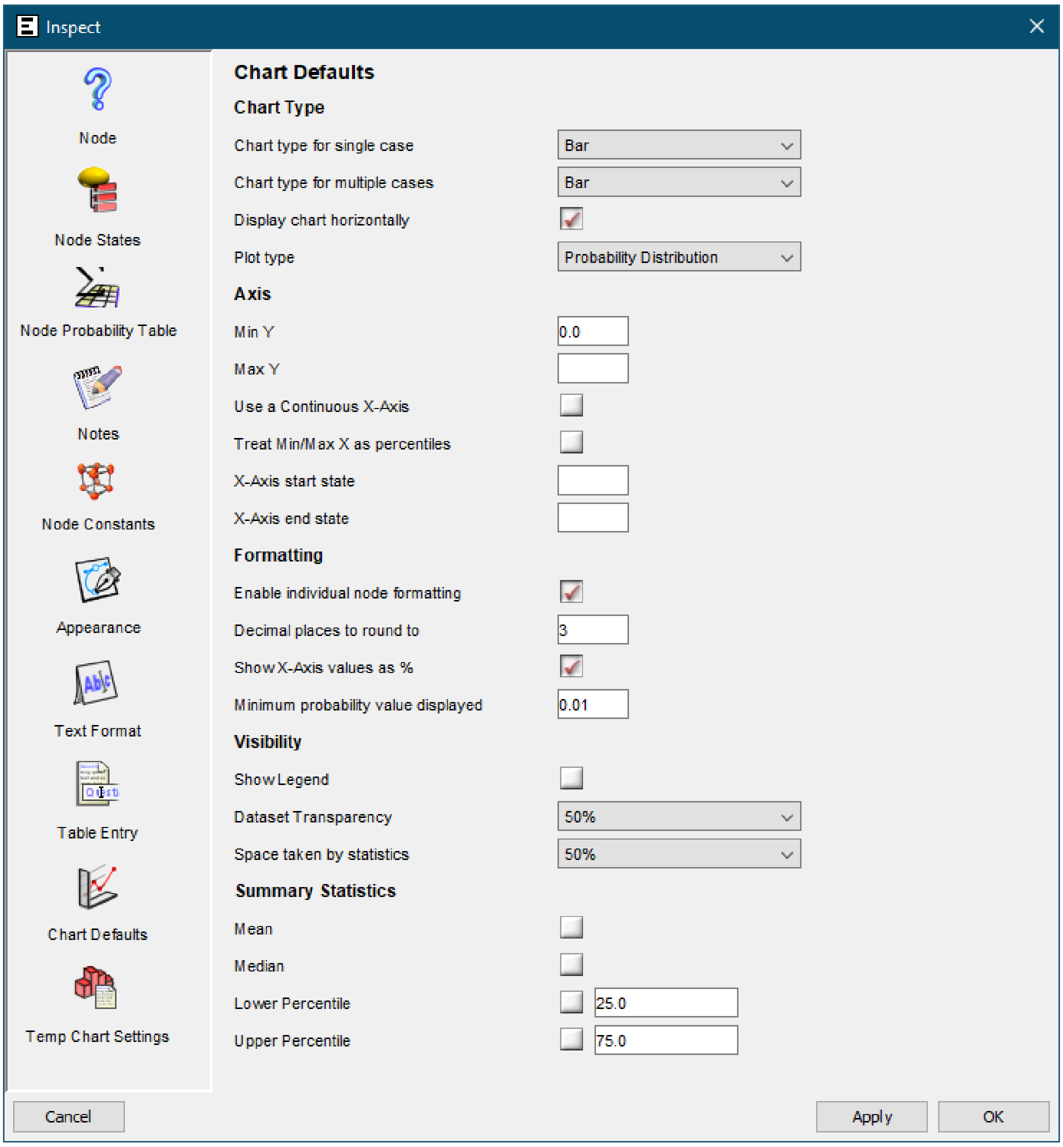


Chart Defaults

Chart Type

Chart type for single case: Bar
Chart type for multiple cases: Bar
Display chart horizontally ☑
Plot type: Probability Distribution

Axis

Min Y: 0.0
Max Y:
Use a Continuous X-Axis ☐
Treat Min/Max X as percentiles ☐
X-Axis start state:
X-Axis end state:

Formatting

Enable individual node formatting ☑
Decimal places to round to: 3
Show X-Axis values as % ☑
Minimum probability value displayed: 0.01

Visibility

Show Legend ☐
Dataset Transparency: 50%
Space taken by statistics: 50%

Summary Statistics

Mean ☐
Median ☐
Lower Percentile ☐ 25.0
Upper Percentile ☐ 75.0

Figure 170 The node properties chart defaults dialog

The node-level dialog contains the following fields:

- **Chart Type for single case** This defines the chart type used if there is only a single case
- **Chart Type for multiple cases** This defines the chart type used if there are multiple cases.
- **Display Chart Horizontally** This allows you to set the chart orientation to horizontal.
- **Plot Type** Whether the chart is plotted as a probability distribution or a cumulative distribution.
- **Min Y** Displays the lowest Y value at which the chart starts
- **Max Y** Displays the highest Y value at which the chart ends
- **Use a Continuous X-axis**
- **Treat Min/Max as percentiles**

- **X-Axis start state / Lower percentile X Bound**
- **X-Axis end state / Upper Percentile X Bound**
- **Show Legend** Displays a legend on the chart stating colours associated with cases.
- **Show Grid** Displays a grid on the work area.
- **Dataset Transparency** Defines how transparent the rendering of the data set should be on the chart.
- **Space taken by statistics**
- **Mean** Defines whether the mean should be visible on the plot. The actual value of the mean is shown in the un-editable text field next to the check box.
- **Median** Defines whether the median should be visible on the plot. The actual value of the median is shown in the un-editable text field next to the check box.
- **Lower Percentile** Defines whether the lower percentile should be visible on the plot. The value of the percentile can be defined in the text field next to the check box. It should be a value between 0 and 100.
- **Upper Percentile** Defines whether the upper percentile should be visible on the plot. The value of the percentile can be defined in the text field next to the check box. It should be a value between 0 and 100.

14.6.2 Defining the X bounds

The meaning of the values that can be specified in the lower and upper X-bound boxes on the chart-level dialog is subtly different depending on the other settings of the chart. Below is a summary of what these values represent given the other (relevant) settings:

- **The chart is treating min/max X as percentiles** You must specify the lower and upper percentiles for the bounds (which should be values in the range of 0-100).
- **The chart is plotted on a continuous X-axis but is treating the bounds as percentiles** The values you specify for the lower and upper X bounds are exact numeric values; that is, the scale will go from X to Y, where X = lower bound and Y = upper bound.
- **The chart is plotted on a non-continuous X-axis and the X bounds are not being treated as percentiles** The values provided in the lower and upper X bounds in this case are the start and end states (or data points). For example, if the bounds defined are state 2 for the lower bound and state 9 for the upper bound, then the chart will plot states 2-9. Note that even if the underlying dataset(s) that the chart is plotting have interval data points and numeric labels, the chart will plot using this mechanism when not plotting on a continuous X-axis.

14.6.3 Plotting on a continuous X-Axis

By default, charts are plotted on a non-continuous X-axis. That is, the chart plot data sets which consist of a set of data points for a given node during a given propagation. A data point consists of a label and a point. Each discrete data point occupies the same amount of space on the X-axis. Some nodes have interval information defined over their states. This interval information can be used to plot the X-axis on a continuous rather than discrete scale (in the same way as the Y-axis). Toggling the ‘Continuous X-axis’ graph setting will force the chart to attempt this type of plot. If the data set does not contain data points with interval information it can still plot the graph using a continuous X-Axis but assumes an absolute scale for the data points it plots i.e., 0-1, 1-2, 2-3 ….

When a continuous plot is chosen, the Y-axis value takes the discrete interval probability mass and divides it by the width of the interval. Note that this can result in the Y-axis values being greater than 1, but the total probability mass will equal 1 in all cases.

Note: If you have more than one case displayed on a chart for a continuous or integer interval node, you should only use a continuous X-axis. This is because each case will result in a different discretization of the node in question; the X-axis will be defined using only the first case's discretization; this, in turn, means that the plots for the other cases will not make sense with respect to the values on the X-axis.

14.6.4 Plotting on a log scale

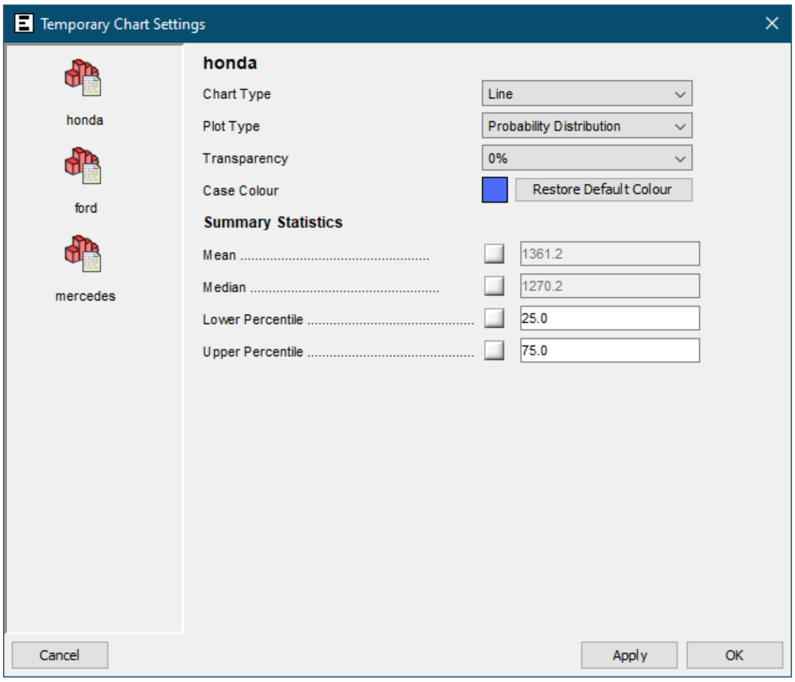
Any graph that is plotted on a continuous X-scale can also be plotted on a log scale. This setting is made available only when the continuous X-scale has been selected.

14.6.5 Treating the X bounds as percentiles

If you specify that the X bounds should be percentiles, then you are stating that rather than defining exact numeric beginning and end bounds for the X-axis you want to limit the X-plot to the defined lower and upper percentile of the data sets plotted on the graph. For example, if the 'treat as percentile' box is ticked and a lower bound of 5 and an upper bound of 95 are defined, then the X-axis will automatically scale, so that it starts at a value equal to the lowest 5th percentile of the plotted data sets and goes up the highest 95th percentile. This is useful when the plotted distributions have long tails that you are not interested in seeing.

14.6.6 Temporary chart settings dialog

The temporary chart settings dialog is available by clicking on temporary graph settings tab on the Node Properties dialog. Properties for a case are only available when a calculation has taken place on that case. This dialog allows you to configure properties for each case on a Node.



honda

Chart Type: Line

Plot Type: Probability Distribution

Transparency: 0%

Case Colour

Summary Statistics

Mean ☐ 1361.2

Median ☐ 1270.2

Lower Percentile ☐ 25.0

Upper Percentile ☐ 75.0

Figure 171 Temporary Graph Settings dialog

The case-level dialog contains many of the fields on the node-level chart properties. It also includes the following fields:

- **Chart Type** The type of chart used to plot the data set. This will override settings used on chart defaults.

- **Dataset Colour** The colour used to show the chart. There exists a button which restores the dataset colour to the default colour first generated. Please note that changing the colour for the case on this node will not change the case colour for every node.

- **Summary Statistics** For continuous nodes the summary statistics can be selected for display.

14.6.7 The model chart properties dialog

The model chart properties dialog contains properties relating to the horizontal graph view. You can get to this dialog going to Model Properties.

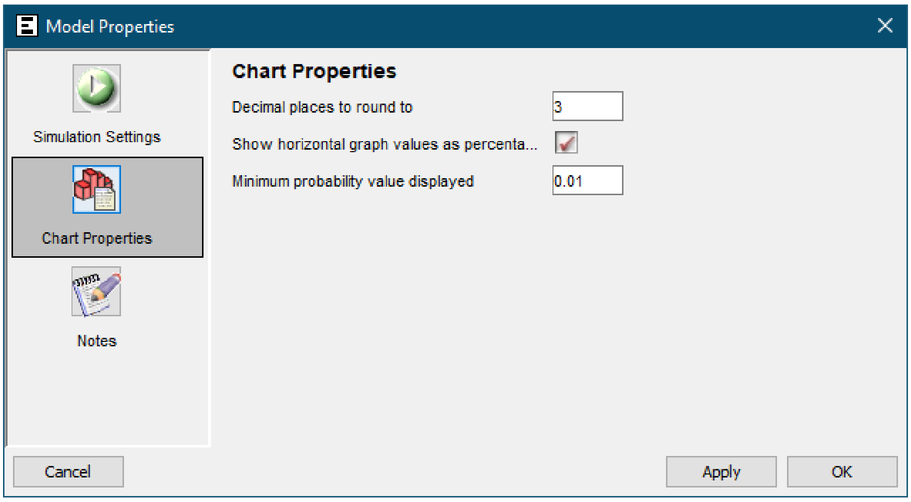


Chart Properties

Decimal places to round to: 3

Show horizontal graph values as percenta... ☑

Minimum probability value displayed: 0.01

Figure 172 Model Graph Properties

The model-level dialog contains the following fields:

- **Decimal places to round to** This is specific to the horizontal chart view. As horizontal chart view displays values on the chart area, the values are rounded according to the value entered here.

- **Show horizontal graph values as percentage** This option allows you to view values as percentage rather than absolute numbers.

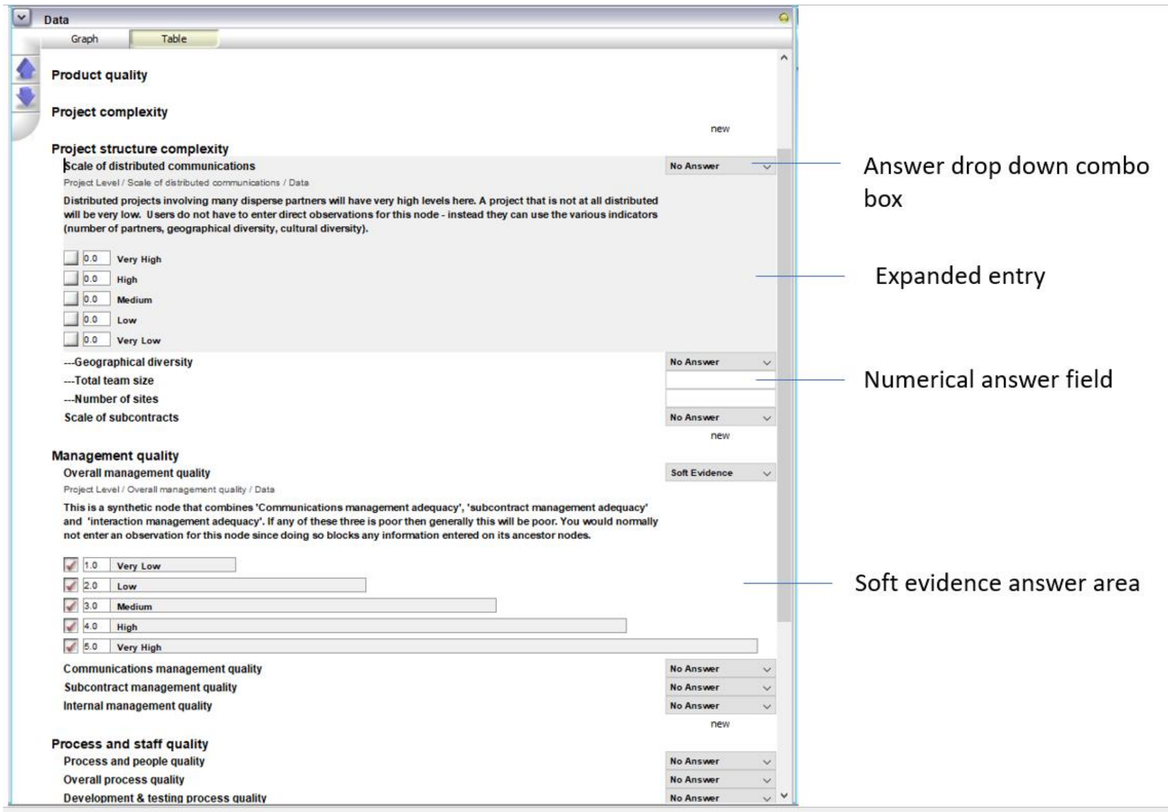
- **Minimum probability value displayed** This will be the minimum probability value from your result set displayed on a horizontal chart.

15. The Table view

If the Table view is selected via a tab and when selected it will display a Table for the Bayesian network selected in the explorer view. If the model item is selected, the Tables for all Bayesian networks will be shown.

Figure 173 The Table view

The components of the Table view are shown in Figure 174 below.



Product quality

Project complexity

new

Project structure complexity

Scale of distributed communications — No Answer

Project Level / Scale of distributed communications / Data

Distributed projects involving many disperse partners will have very high levels here. A project that is not at all distributed will be very low. Users do not have to enter direct observations for this node - instead they can use the various indicators (number of partners, geographical diversity, cultural diversity).

☐ 0.0 Very High
☐ 0.0 High
☐ 0.0 Medium
☐ 0.0 Low
☐ 0.0 Very Low

---Geographical diversity — No Answer
---Total team size
---Number of sites
Scale of subcontracts — No Answer

new

Management quality

Overall management quality — Soft Evidence

Project Level / Overall management quality / Data

This is a synthetic node that combines 'Communications management adequacy', 'subcontract management adequacy' and 'interaction management adequacy'. If any of these three is poor then generally this will be poor. You would normally not enter an observation for this node since doing so blocks any information entered on its ancestor nodes.

☑ 1.0 Very Low
☑ 2.0 Low
☑ 3.0 Medium
☑ 4.0 High
☑ 5.0 Very High

Communications management quality — No Answer
Subcontract management quality — No Answer
Internal management quality — No Answer

new

Process and staff quality

Process and people quality — No Answer
Overall process quality — No Answer
Development & testing process quality — No Answer

Figure 174 The Table view

15.1 The Table entry list

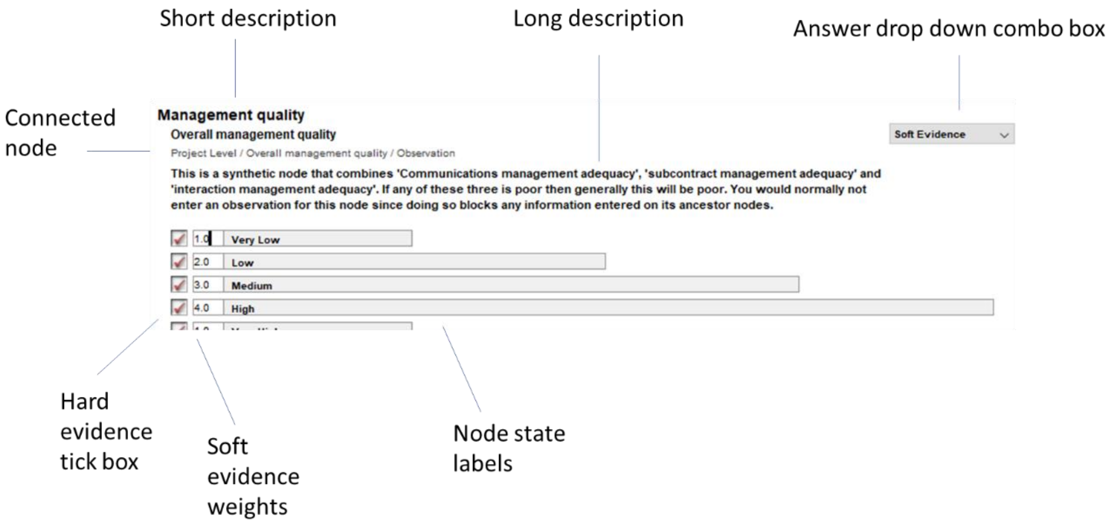
There is a list of entries that correspond to nodes of the selected Bayesian network. Entries are grouped under headings. The headings are shown in larger black text, while the entries are slightly indented and shown in blue.

The headings act simply as a convenient mechanism for grouping related entries together. These headings, like the entries themselves, are editable and users can define their own headings so that they can group entries in a way that they feel is suitable for a particular model or application.

Entries and headings are selected by clicking on their short description. When selected, their background colour will change to light blue. Clicking on the short description again will deselect the entry or heading. Multiple items can be selected if the control key is held down during selection, or if the right mouse button is used. Any functions (such as 'Show Chart' or 'Expand/Retract') that are applied when entries and headings are selected will then be applied to ALL selected items.

By default, entries appear in the entry list unexpanded, showing their short description and their answer box. For entries that do not require a numerical answer, the answer box is a drop-down box containing the set of possible answers. For entries that require a numeric answer, the answer box is a text box, which takes only numeric key presses.

The expanded version of an entry (see Figure 175) displays more details about the entry including guidelines on when, how, and if it should be answered. Non-numeric entries will also display a list of possible answers. The answers consist of a tick box (for answering the entry with a single answer), a text box (for soft evidence) and a short description of the answer. To expand an entry, double-click on its short description or right-click on it and select 'Expand / Retract' from the resulting menu. The expanded entry also contains details about which node it is connected to in the associated Graph view – this is shown directly under the short name of the entry in grey text.



Management quality

Overall management quality

Project Level / Overall management quality / Observation

This is a synthetic node that combines 'Communications management adequacy', 'subcontract management adequacy' and 'interaction management adequacy'. If any of these three is poor then generally this will be poor. You would normally not enter an observation for this node since doing so blocks any information entered on its ancestor nodes.

Soft Evidence

☑ 1.0 Very Low
☑ 2.0 Low
☑ 3.0 Medium
☑ 4.0 High

Figure 175 A Table entry

Note that both numeric and non-numeric entries will display a longer description when expanded.

- Entries can be answered via several mechanisms. In the case where the entry has a non-numeric answer:

- An answer can be selected from the drop-down box. The selection of ‘No answer’ will withdraw any current answer for that entry.

- In the expanded view one of the available tick boxes can be selected. Note that clicking again on the same tick box withdraws the answer.

Soft evidence can be entered by typing numeric values into the text boxes to the right of the answer tick boxes. Any positive number can be entered. When the soft evidence is applied to the underlying model it is normalised; the red bar directly under the answer description represents its normalised value.

Entries that require numeric answers can only be answered by typing into the associated text box.

15.2 The Table Menu

Right-clicking on an entry or a heading will display the Table menu.

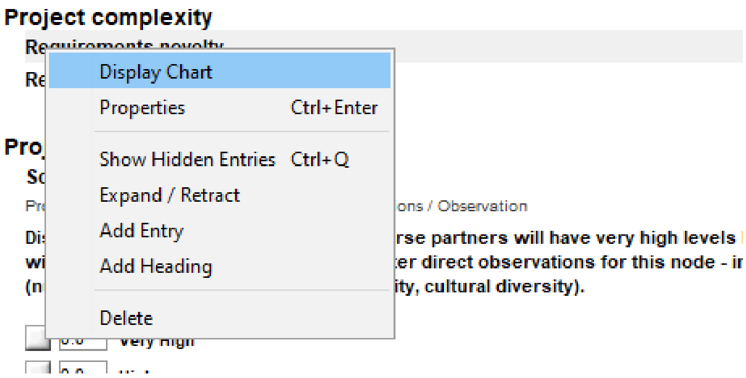


Figure 176 The Table menu

- **Display Chart** This creates a chart for all selected entries. The charts will be displayed on the chart panel.

- **Properties** This will load a configuration dialog for the first selected entry. In most cases this will be the entry you right-clicked on, since the action of right-clicking on an entry selects it.

- **Show Hidden Entries** This hides or shows all entries tagged as “hidden” (the value of which is configured on an entry-by-entry basis via the entry properties dialog). If the menu item is ticked then all entries will be visible.

- **Expand / Retract** This will toggle the expansion status of the selected items. If a heading is retracted, then all its connected entries will be hidden until expanded again.

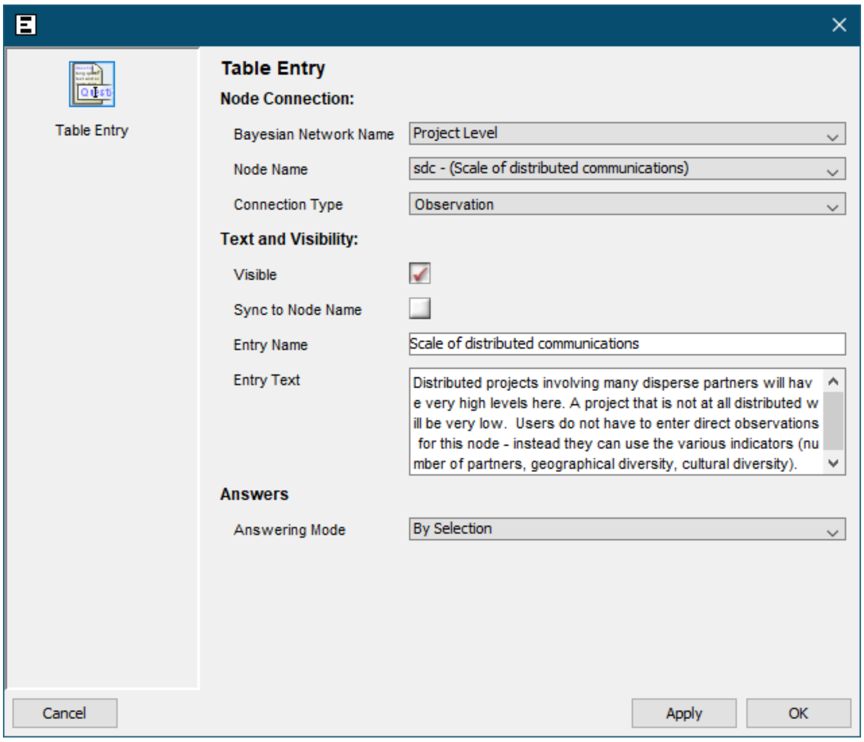
- **Add Entry** This will add a new entry to the Table. The new entry will be added to the Table in relation to the first selected item (which by default is the item you right clicked on to get the menu). If that item was an entry, then the new entry will be added directly above it. If it was a heading, then the entry will be added to the bottom (end) of the entries under that heading. If you add an entry, you should associate with it with a node in the associated Graph view.

- **Add Heading** When this button is pressed a new heading is added at the end of the Table. All selected items will be deselected, and the new heading selected.

- **Delete** This will delete all selected headings and entries. If a heading is deleted, then all of its entries will also be deleted. If you want to delete just the heading, then ensure that you move all entries from under that heading before deleting it.

15.3 The Table Entry Dialog

Right-clicking on an entry and selecting ‘Properties’ from the resulting menu will display the Table entry dialog shown in Figure 177.



Node Connection:

Bayesian Network Name: Project Level

Node Name: sdc - (Scale of distributed communications)

Connection Type: Observation

Text and Visibility:

Visible: ☑

Sync to Node Name: ☐

Entry Name: Scale of distributed communications

Entry Text: Distributed projects involving many disperse partners will have very high levels here. A project that is not at all distributed will be very low. Users do not have to enter direct observations for this node - instead they can use the various indicators (number of partners, geographical diversity, cultural diversity).

Answers

Answering Mode: By Selection

Figure 177 The Table entry properties dialog

The following fields are available when you open the Table entry properties dialog:

- **Bayesian network Name** This shows the name of the Bayesian network with which the Table entry is associated.

- **Node Name** This contains the name of the node in the currently selected Bayesian network with which the Table entry is associated.

- **Connection Type** This specifies the type of relationship the Table entry has with its corresponding node. Most often an answer to an entry will be treated as data on the connected node; accordingly, “Data” is the default value for this field. Occasionally, you may want to create Table entries for entering expression parameters. In this case, the “Parameter Value” field should be chosen. When you do so, an extra field, “Expression Variable”, is displayed and the answering mode is restricted to “Numerically”. See below.

- **Expression Variable (only shown when “Connection Type” is set to “Parameter Value”)** This allows the Table entry to be linked to a predefined expression variable on the connected node.

- **Visible** This defines whether the Table entry is actually shown in the Table. (Whether invisible entries are displayed in the Table or not depends on whether “Show Hidden Entries” in the “Table” menu is ticked.)

- **Sync to Node Name** This defines whether the name of the Table entry will always be the same as the connected node name. If this is checked, the “Entry Name” field below it is disabled.

- **Entry Name** This is the name of the entry as it appears in the Table. This cannot be changed if the “Sync to Node Name” box above is checked.

- **Entry Text** This is the long description that appears in the Table when the entry is expanded.

- **Answering Mode** This defines how the entry can be answered. If “Numerically” is chosen, an answer is supplied for the entry by typing a numerical value into a text box. If “By Selection” is chosen, the entry is answered by selecting a value from a drop-down box. If “Unanswerable” is chosen, the entry cannot be answered at all in the Table. If the “Connection Type” for this node is set to “Parameter Value”, the Answer Mode is automatically changed to “Numerically” and cannot be modified.

15.4 Moving Headings and Entries

The two arrow buttons in the top left of the Table are used to reorder selected entries and headings within the Table. To do this, select the relevant headings and/or entries and click on the ‘Move Up’ or ‘Move Down’ button. Multiple components can be selected by clicking while holding down the control key; they can then be moved together.

15.5 Accessing constants in the Table view.

We previously explained the role of constants where a constant is always associated with a particular node. If you want to vary the constant values in different cases, say where we have different assumptions, then you always need to do this via the Table view. Hence you need to set a new Table entry to refer to the constant value directly.

To do this create a new Table entry and right-click to bring up the dialog as shown in Figure 178.

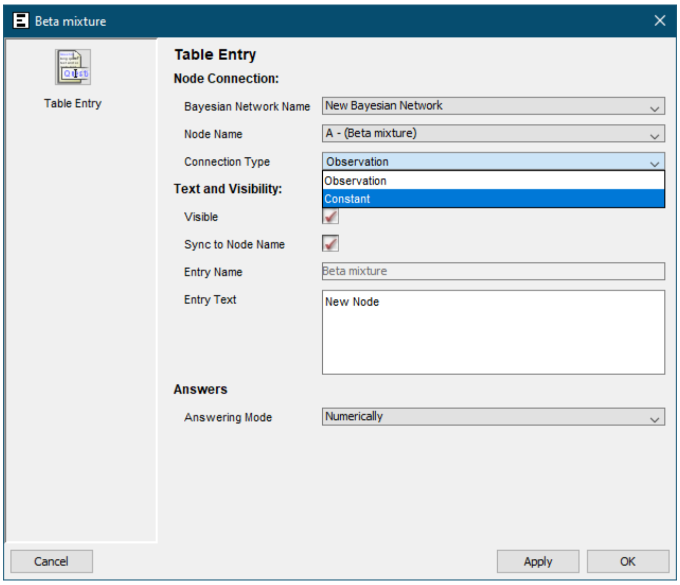


Table Entry

Node Connection:

Bayesian Network Name: New Bayesian Network

Node Name: A - (Beta mixture)

Connection Type: Observation

Observation
Constant

Text and Visibility:

Visible ☑

Sync to Node Name ☑

Entry Name: Beta mixture

Entry Text: New Node

Answers

Answering Mode: Numerically

Figure 178 Creating a Table entry for a constant

Having selected the current Bayesian network you will need to select (from the Node Name drop down list) the node that is associated with the constant. You must then select the option “Constant” in Connection Type. The dialog then changes and offers you a drop-down list to choose the constant you require, as shown in Figure 179.

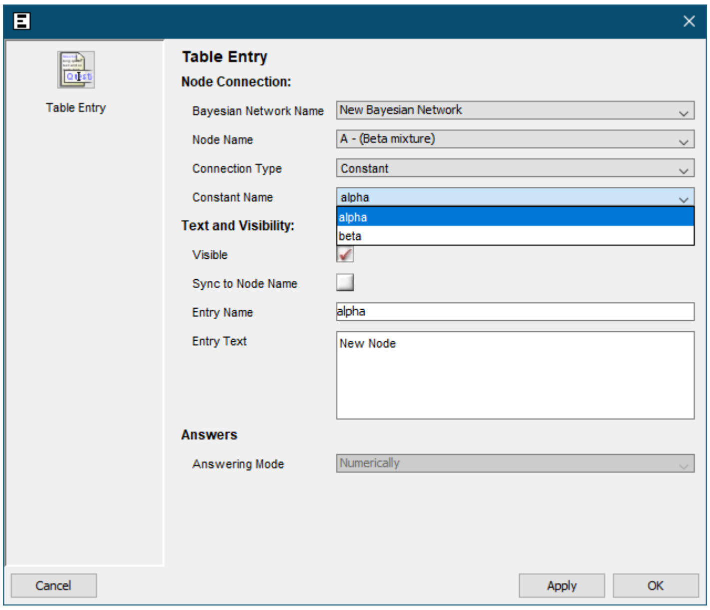


Table Entry

Node Connection:

Bayesian Network Name: New Bayesian Network
Node Name: A - (Beta mixture)
Connection Type: Constant
Constant Name: alpha
alpha
beta

Text and Visibility:

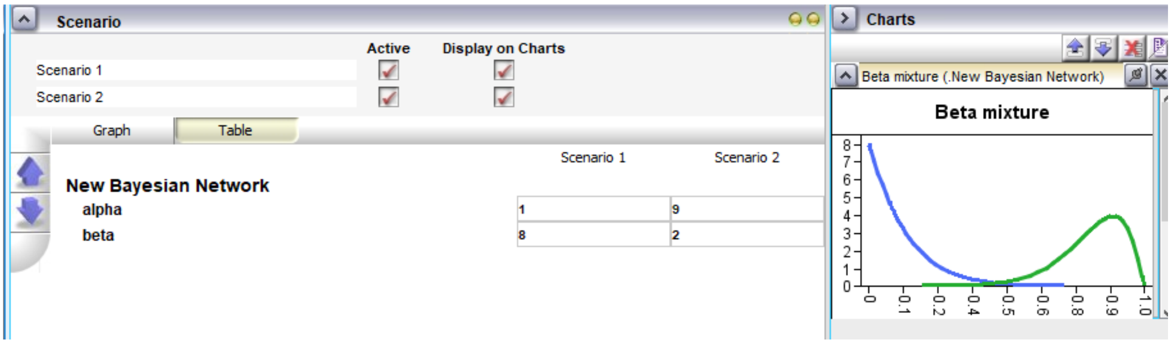
Visible: ☑
Sync to Node Name: ☐
Entry Name: alpha
Entry Text: New Node

Answers

Answering Mode: Numerically

Figure 179 Linking the Table entry to the constant

Now that you have declared your constants in the Table you can change their values by editing the values in any cases created. Using the same example as before shows the results from running two cases with two sets of values set for the constants we have created on the Beta distribution — Beta(3,7, 0, 1) and Beta(8,2,0,1).



	Active	Display on Charts
Scenario 1	☑	☑
Scenario 2	☑	☑

New Bayesian Network	Scenario 1	Scenario 2
alpha	1	9
beta	8	2

Figure 180 Results from editing the values declared on the constants

16. The Data panel

The data panel (shown in Figure 181) allows you to manipulate many cases at the same time, and to plot the results on the same charts for comparison purposes.

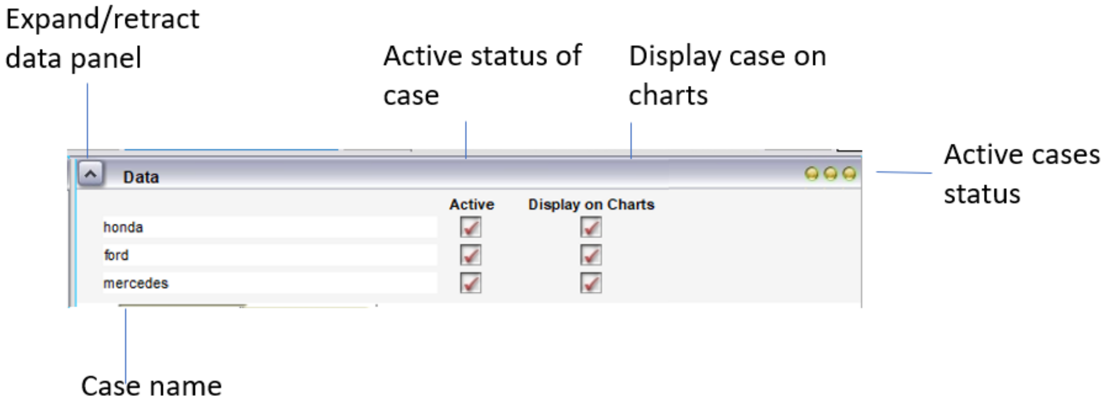
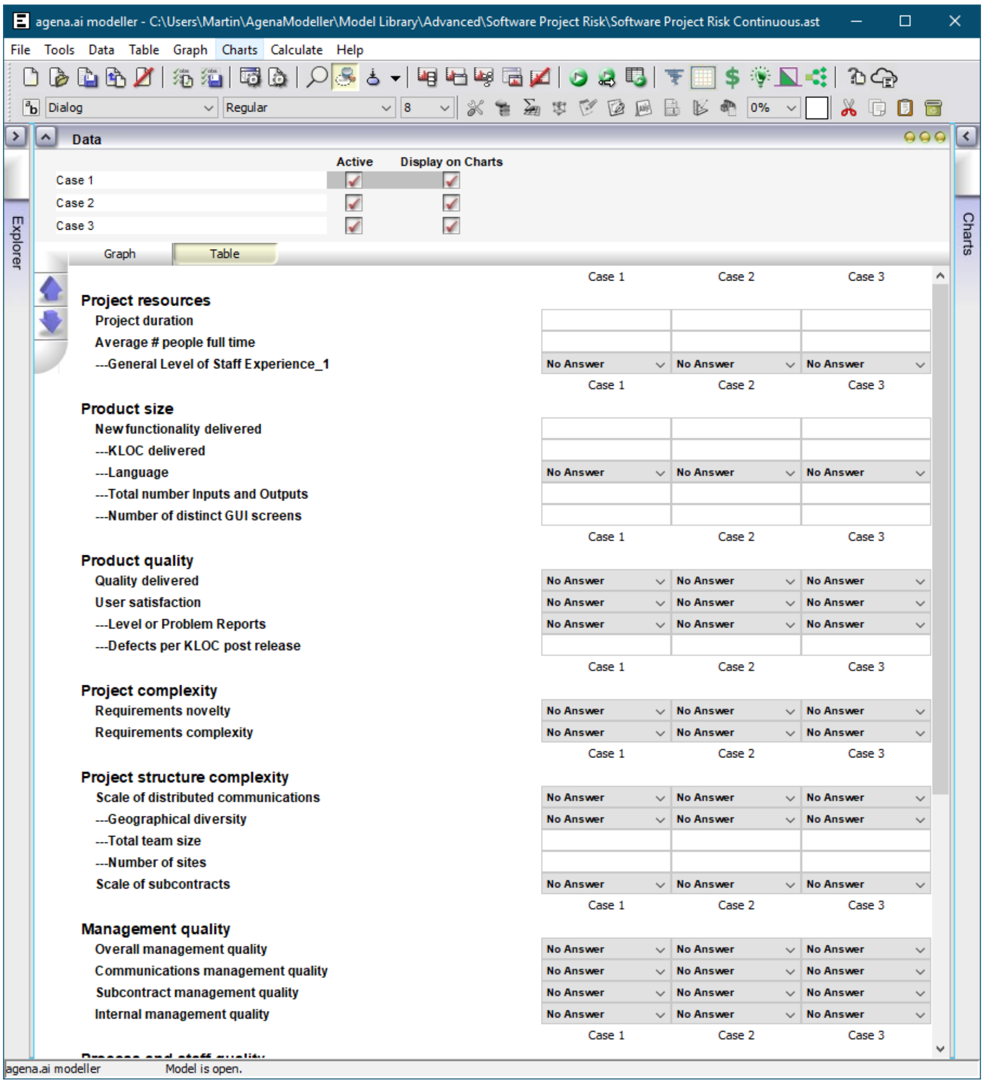


Figure 181 The data panel

The data panel is available for all objects selected in the explorer. It can be expanded and retracted via the ‘expand/retract’ button in its top left corner. The data panel displays a list of the available cases.

By default, when a model is opened it is configured with just a single case. Answers that are supplied for Table entries will be held in that case.

Cases are toggled active and inactive by clicking on their connected ‘Active’ check box. Active cases are then made available in the Table for answering. Inactive cases are not available in the Table. Similarly, cases can be shown or hidden in charts by toggling the ‘Display on Charts’ box. Figure 182 shows the Table and data panel configured with three cases, the first and last of which are active.



agena.ai modeller - C:\Users\Martin\AgenaModeller\Model Library\Advanced\Software Project Risk\Software Project Risk Continuous.ast

	Active	Display on Charts
Case 1	☑	☑
Case 2	☑	☑
Case 3	☑	☑

	Case 1	Case 2	Case 3
Project resources			
Project duration			
Average # people full time			
---General Level of Staff Experience_1	No Answer	No Answer	No Answer
Product size	Case 1	Case 2	Case 3
New functionality delivered			
---KLOC delivered			
---Language	No Answer	No Answer	No Answer
---Total number Inputs and Outputs			
---Number of distinct GUI screens			
Product quality	Case 1	Case 2	Case 3
Quality delivered	No Answer	No Answer	No Answer
User satisfaction	No Answer	No Answer	No Answer
---Level or Problem Reports	No Answer	No Answer	No Answer
---Defects per KLOC post release			
Project complexity	Case 1	Case 2	Case 3
Requirements novelty	No Answer	No Answer	No Answer
Requirements complexity	No Answer	No Answer	No Answer
Project structure complexity	Case 1	Case 2	Case 3
Scale of distributed communications	No Answer	No Answer	No Answer
---Geographical diversity	No Answer	No Answer	No Answer
---Total team size			
---Number of sites			
Scale of subcontracts	No Answer	No Answer	No Answer
Management quality	Case 1	Case 2	Case 3
Overall management quality	No Answer	No Answer	No Answer
Communications management quality	No Answer	No Answer	No Answer
Subcontract management quality	No Answer	No Answer	No Answer
Internal management quality	No Answer	No Answer	No Answer
	Case 1	Case 2	Case 3

agena.ai modeller Model is open.

Figure 182 The data panel with two three active cases

In the example in Figure 182, two columns of answer boxes are present: the first represents the first case, while the second is for the last cases. Thus, data placed into the first column will be applied to the first case, and data placed into the second column will be placed into the last case.

There are no restrictions on how many of the cases can be active in the data panel at any point (although screen space limits what is reasonable). It is useful to note that the order of the columns in the Table is equivalent to the order of the active cases. For example, if we made case 2 active in Figure 182 then there would be three answer columns in the Table, such that the middle column would now be connected to case 2.

The cases are independent of each other. Entering data into one of them in no way modifies the others.

Data supplied for Table entries via their expanded view are applicable only to the left-most active cas. This is highly relevant if you wish to enter soft evidence. Then you must ensure that the case into which you wish to enter the data is the only active cases.

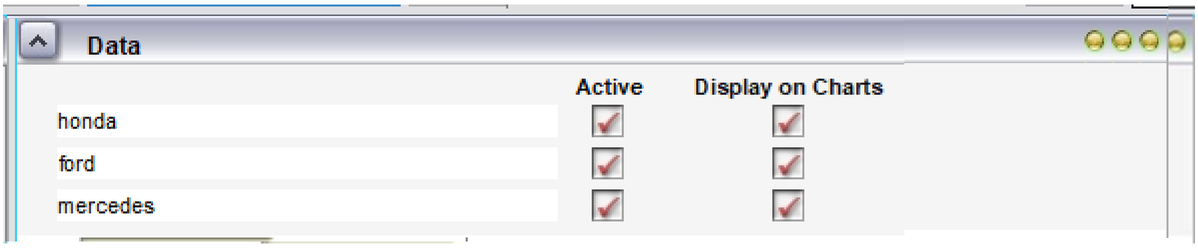
You can change the name of a case by typing the new name directly into the case name text box.

16.1 The Case Summary Display

In the top right-hand side of the data panel is the case summary display (see Figure 181). This consists of several small round buttons that allow quick access to the cases when the data panel is minimised (retracted). They can be used to activate and deactivate cases. When their connected case is active, they appear pressed (and are coloured green). Moving the mouse pointer over them reveals tool tips that display the case names. The right mouse button can be used on the case summary display to access a menu from which new cases can be added or cloned and existing ones deleted.

16.2 The Data Panel Menu

The data panel menu can be accessed via a right click on almost any part of the Data panel. It can also be accessed via the Data menu at the top of the application window.



	Active	Display on Charts
honda	☑	☑
ford	☑	☑
mercedes	☑	☑

Figure 183 The data panel menu

The menus perform the following functions:

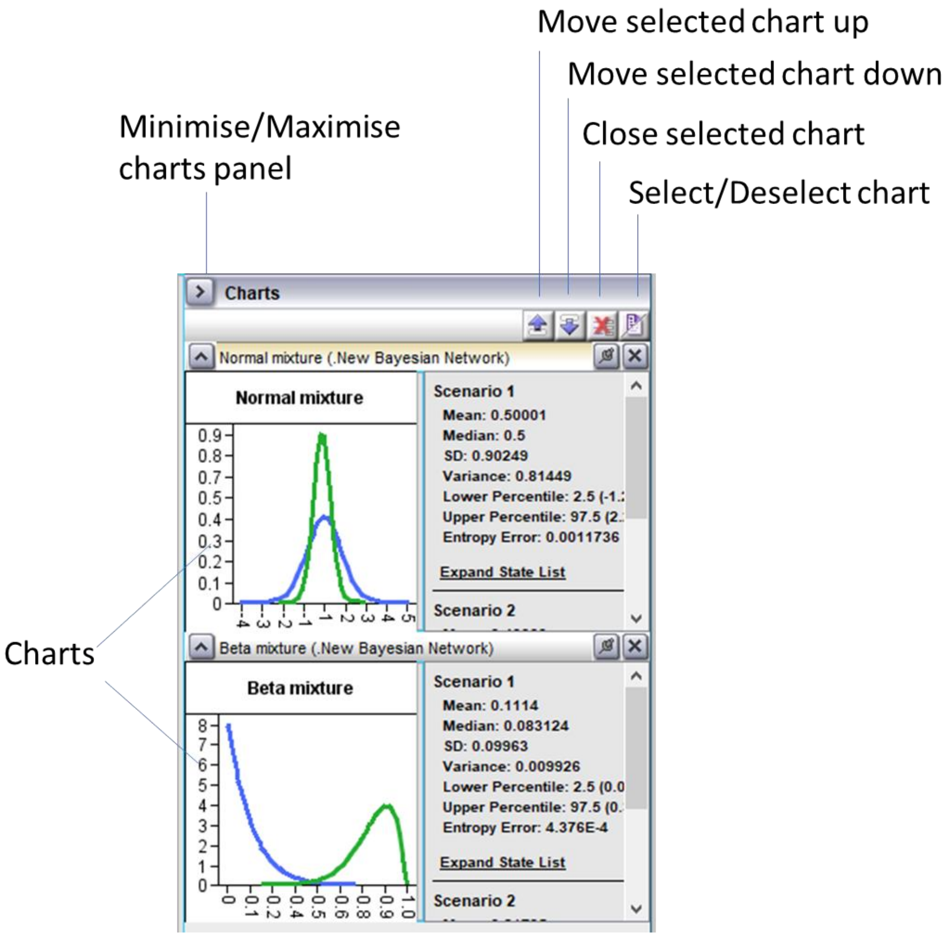
- **Add a New Case** This will add a new case, which will subsequently be added to the end of the case list. The new case will not be active by default; you must click on its ‘Active’ check box to make it active. It will, however, be displayed on charts by default.

- **Delete Selected Cases** This will delete all selected cases. Note that the act of right-clicking on a case selects it.

- **Clone Selected Cases** This will clone the selected cases, and place the newly cloned cases at the end of the cases list. A cloned case has the same set of data in it as the source case from which it was cloned.

17. The Chart Panel

The chart panel (Figure 184) is on the right-hand side of the application window. Users can choose to display charts on the chart panel, on the Graph view or in their own windows.



Normal mixture (.New Bayesian Network)

Scenario 1
Mean: 0.50001
Median: 0.5
SD: 0.90249
Variance: 0.81449
Lower Percentile: 2.5 (-1.
Upper Percentile: 97.5 (2.
Entropy Error: 0.0011736
Expand State List
Scenario 2

Beta mixture (.New Bayesian Network)

Scenario 1
Mean: 0.1114
Median: 0.083124
SD: 0.09963
Variance: 0.009926
Lower Percentile: 2.5 (0.0
Upper Percentile: 97.5 (0.
Entropy Error: 4.376E-4
Expand State List
Scenario 2

Figure 184 The chart panel

The chart panel can be minimised and maximised via the arrow button situated at the top of the panel. The width of the panel can also be changed by dragging the bar that separates the chart panel from the work area.

The chart panel toolbar contains functions that operate on selected charts. Charts are selected by clicking on the top of the relevant graph. They can be deselected by clicking on the graph title a second time.

- **Move selected charts up** This shifts all selected graphs one position upwards in the panel.
- **Move selected charts down** This shifts all selected graphs one position downwards in the panel.
- **Close selected charts** This closes all the selected graphs.
- **Select / deselect all charts** If one or more graphs are selected then this function will deselect them all. Conversely, if no graphs are selected then clicking this button will select them all.

Only a single chart can be open for a given node at any point. If a new graph is requested for a node that already has a graph open, the graph is selected and moved to the top of the list.

18. The Notes dialog

The notes dialog allows you to add notes to specific items in the model. Regardless of where the notes dialog is accessed it is used in the same way:

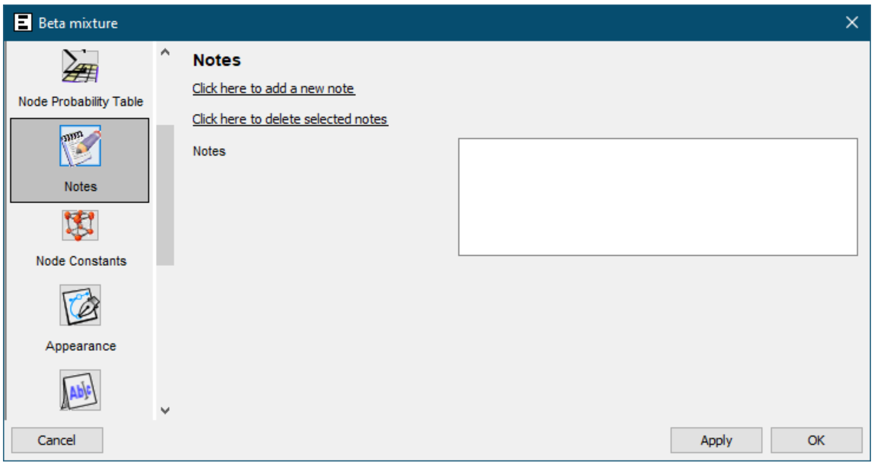


Figure 185 The notes dialog

New notes are added by clicking on the “new note” hyperlink at the top of the dialog. New notes (and existing ones) are listed in the notes list. A note can be selected and displayed by clicking on it in the list. Once displayed, its title and body can be edited in the lower half of the dialog.

Notes can be removed by selecting them in the list then clicking on the “remove selected notes” hyperlink at the top of the dialog. Multiple notes can be deleted at the same time by holding down the shift or control key during selection.

Clicking OK or Apply will save the notes back to the item the notes are relevant to.

19. The menu and model toolbar functions

19.1 The file menu

Agena.ai modeller handles two main types of files: one has the extension AST and the other has the extension CMPX. CMPX and AST files contain all the model information including entered data and cases. Because the tool allows users to work with multiple models and to link them appropriately, both AST and CMPX files can generally contain multiple models. AST files and CMP files are identical except the former are read-only. In this sense, AST files can be considered pre-defined model templates that should not be overwritten by users.

Normally when you are working with a model you will save files as CMPX files. However, if you are using the tool to generate your own application then you will typically open an existing AST file, edit it, save it within the tool as a CMPX file and then change its extension back to AST outside the tool.

ℹ	In addition to AST and CMPX files Agena.ai modeller is also able to open files with a NET extension. These are Bayesian Net model files created using the Hugin software.

The file menu contains the following functions:

- **File > Create New Model** The currently opened file will be closed, and a new empty model will be created.

  Associated tool bar icon:

- **File > Open Model** This opens a dialog listing the directories on the computer and showing any available AST, CMP and NET files. In contrast to the ‘Import Model’ function (see below), when this option is selected, any currently opened models will be closed and you will be prompted to save them.

  Associated tool bar icon:

- **File > Close Model** This closes the open model.

  Associated tool bar icon:

- **File > Save Model** This saves the current model (including all its links, cases, data, charts, and notes) as a CMPX file. If you have previously specified a file in which to save the model, the model will be saved in that same file. There will be no confirmation dialog.

  Associated tool bar icon:

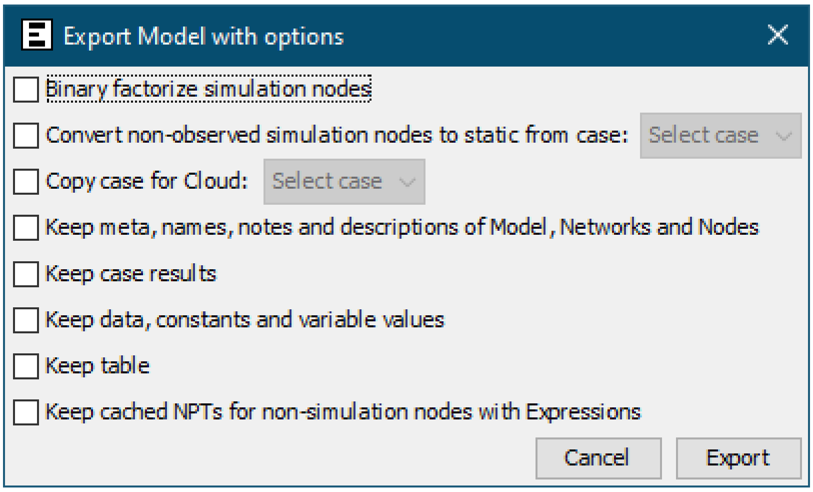
- **File > Save Model As** This saves the current model (including all its links, cases, data, and notes) as a CMPX file. A dialog will appear prompting you to specify a path and file name. If you wish to create your own AST files, then you can simply change the file extension outside the tool. We strongly recommend that you never overwrite the AST files provided with the application.

- **File > Import Model** This option, which requires users to have already opened an AST, CMPX or NET file, opens a dialog listing available AST, CMP and NET files. When a file is selected it is imported into the current model and all its component Graph views, Tables and cases are added to those in the current model. Importing does NOT close the existing model. Hence, this is the main mechanism for building up larger domain-dependent models.

  Associated tool bar icon:

- **File > Export Model** This option, which requires users to have already opened an AST CMPX or NET file, opens a dialog allowing you to save the file in XML format.

- **File > Export Model with options** This option, which requires users to have already opened an AST CMPX or NET file, opens a dialog allowing you to save the file in JSON format, for use in agena.ai cloud services. Choosing this opens a dialog with various options, as shown in Figure 186.



☐ Binary factorize simulation nodes

☐ Convert non-observed simulation nodes to static from case:

☐ Copy case for Cloud:

☐ Keep meta, names, notes and descriptions of Model, Networks and Nodes

☐ Keep case results

☐ Keep data, constants and variable values

☐ Keep table

☐ Keep cached NPTs for non-simulation nodes with Expressions

Figure 186 Export model with options dialog

- **File > Import Cases…** This opens a dialog that allows you to select a case file (with the extension SCN) to import. When you import cases, they are added to the model. Because the cases may have been created in a different version of the model or a different model entirely, you will be presented with a dialog that tells you the status of the cases. You can then select which cases and data you wish to import.

  Associated tool bar icon:

- **File > Export Cases** This saves the cases in the current model to an SCN file.

  Associated tool bar icon:

- **File > Model Properties** This opens a dialog that allows you to edit the properties that relate to the currently open model.

  Associated tool bar icon:

- **File > [recently opened file]** There will be up to four entries here containing the names of the files you most recently opened in Agena.ai modeller. Clicking on an entry will open that file, assuming that it hasn't been moved or deleted since you last opened it.

- **File > Exit** This closes the application. You will be asked to confirm that you want to exit the tool and you will be prompted to save any changes before exiting (The standard Windows Close button also has the same effect.)

19.2 The graph menu

The graph menu contains the following functions:

- **Graph > Find** This find nodes or Bayesian networks by name or identifier

  Associated tool bar icon:

- **Graph > Highlight nodes** This opens a drop-down list of nodes to highlight node types using icons. The icon displayed for this is dynamically selected.

  Associated tool bar icon:

- **Graph > Save chart as JPEG or PNG file** This saves the selected Bayesian network or model as a JPEG or PNG file

- **Graph > Show hidden nodes** This toggles the display of nodes that are hidden.

  Associated tool bar icon:

19.3 The tools menu

- **Tools > Clear Entered Data** This removes all the data in the specified bounds. The bounds are available in the sub-menus of this menu item, and consist of:
    - 'all' the data currently in the model.
    - all active cases; or
    - just those in a specified item.

  Note that you will not see the results of removing this data until the next calculation has been run.

- **Tools > Application Properties** This will display the application-level settings. These settings are not model specific.

  Associated tool bar icon:

- **Tools > Auto Calculation** When ticked, a new calculation will be run each time data is added to or removed from the model.

  Associated tool bar icon:

⚠	With large and complex models, calculations can take a long time, so you might find it easier to turn off the Auto Calculation option when working with large models.

- **Tools > Export Model to XML** This exports the model declaration to XML but without any graphical properties information.

- **Tools > Sensitivity analysis** This performs a sensitivity analysis on the model. The output is displayed as HTML in a browser.

  Associated tool bar icon:

- **Tools > Multivariate Analysis** This performs a multivariate analysis on the model. The output can be saved as a CSV file.

  Associated tool bar icon:

- **Tools > Value of Information analysis** This performs value of information analysis on the model. The output is displayed as HTML in a browser.

  Associated tool bar icon:

- **Tools > Table Learning** This performs table learning from data to update the NPTs in the model.

  Associated tool bar icon:

- **Tools > Compound Sum Analyser** This performs compound sum analysis (or convolution analysis) by combining frequency and severity variables.

  Associated tool bar icon:

- **Tools > Hybrid Influence Diagram Analyser** This translates a Bayesian network into a hybrid influence diagram for decision analysis. The output is displayed as HTML in a browser.

  Associated tool bar icon:

19.4 The data menu

Note that all the functions on the data menu are also available directly from the data panel:

- **Data > Add a New Case** This will add a new case to the Bayesian network currently selected in the explorer. The new case will appear in the data panel.

- **Data > Delete Selected Cases** This function will delete all cases currently selected in the data panel.

- **Data > Clone Selected Cases** This function will copy all cases currently selected in the data panel.

19.5 The table menu

Note that all the functions on the Table menu are also available directly from the Table view::

- **Table > Expand/Retract Selected Headings/Entries** This will toggle the expansion status of all selected headings or entries.

- **Table > Display Charts for Selected Headings/Entries** This will load the charts for all selected headings or entries.

- **Table > Add Entry above First Selected Heading/Entry** This will add a new entry above the first selected heading or entry. If no entry (or heading) is selected, then no new entry will be added.

- **Table > Add a Heading to Selected Heading** This adds a new heading underneath the currently selected heading.

- **Table > Delete Selected Headings/Entries** This deletes the selected headings or entries. Note that there is no dialog asking you whether, you are sure. If you delete a heading then all entries contained within that heading are also deleted.

- **Table > Show Hidden Entries** When toggled, this menu item will show or hide all Table entries that are marked as “hidden”.

19.6 The graph view menu

- **Graph view > Highlight Nodes** The sub-menus under this menu item will toggle icons on the nodes in the Graph view. These icons identify the nodes as having specific attributes. The attributes include whether the node:

    - is an input node

    - is an output node

    - has data entered

    - has an expression or expressions defined

    - requires its expression(s) to be re-applied

    - is a simulation node

    - is of a specific type (the different types are contained on a further sub-menu)

  Associated tool bar icon:

  The options that can be selected are shown in Figure 187.

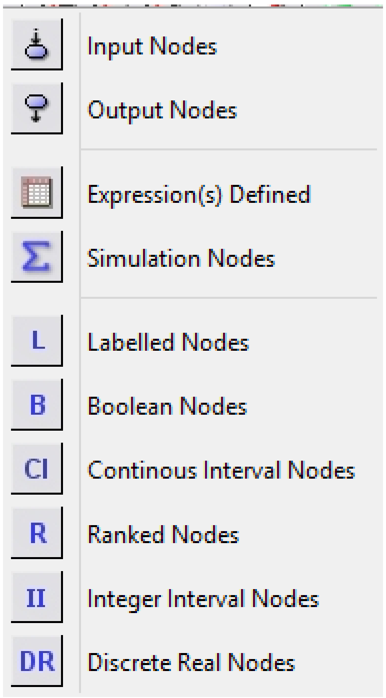


Figure 187 Show Selected Nodes Labels Options

Note: If you highlight NPT Expression Modes, when you hold your mouse over the icon for nodes with expressions defined on them, the expression(s) will appear in a tooltip. This saves opening up the properties dialog whenever you need to view a node's expression(s).

- **Graph view > Save Graph view as JPEG** This will save the current Graph view as a JPEG file.
- **Graph view > Show Hidden Nodes** When toggled, this menu item will show or hide all nodes that are marked as "hidden". See section 13.4.1 for more information on invisible nodes.

  Associated tool bar icon:

- **Graph view > Display Grid** This allows the Graph view grid to be turned on or off.

19.7 The charts menu

Some of these functions refer to 'selected' charts. Note that menu items in the chart panel sub-menu only act on charts docked on the chart panel.

- **Charts > Chart Panel > Move Selected Charts Up One** This function is used to reorder the charts in the chart panel. It will move all selected charts up towards the top of the chart panel. Note that if all the graphs are selected then they will all move up, thus no change will be perceived.
- **Charts > Chart Panel > Move Selected Charts Down One** This function is used to reorder the charts in the chart panel. It will move all selected charts downwards. Note that if all the graphs are selected then they will all move down and no change will be perceived.
- **Charts > Chart Panel > Close Selected Graphs** This will close (remove) all selected charts.

- **Charts > Chart Panel > Select / Deselect All Charts** If any graphs are selected in the charts panel then all graphs will be deselected, otherwise all graphs will be selected.

- **Charts > Close All Charts** This will close all charts, wherever they are docked.

  Associated tool bar icon

- **Charts > Dock All Charts** This will dock all graphs to the location selected from the sub-menus under this menu item, namely:

  - **to Chart Panel**

  Associated tool bar icon

  - **to Own Window**

  Associated tool bar icon

  - **to Graph view**

  Associated tool bar icon

- **Charts > Move Undocked Charts to Front** If you undock a chart, it becomes a window in its own right. Sometimes it can be difficult to keep track of where these undocked windows are. This function brings all undocked graph windows to the front so that they can be seen.

  Associated tool bar icon

19.8 The learning from data menu

- **Learning from Data > Table Learning from Data** This performs learning probability tables for nodes in the current model based on supplied dataset.

  Associated tool bar icon

19.9 The calculate menu

- **Calculate > Run Calculation** This function propagates the current model.

Associated tool bar icon

- **Calculate > Re-calculate NPTs for All Expressions Entered** This will regenerate all NPTs for all nodes in the model.

Associated tool bar icon:

When a calculation is run, all open charts will update.

19.10 The help menu

- **Help > About Agena.ai modeller** This displays information about the version of Agena.ai modeller that you are running.

- **Help > Agena Help on the web** this opens the Agena.ai modeller support webpages in your browser.

- **Help > Agena.ai modeller Help** This will launch a PDF version of the Agena.ai modeller user manual. To view it, you will need to have Adobe Acrobat Reader installed on your system.

  Associated tool bar icon:

19.11 About Agena.ai cloud services portal

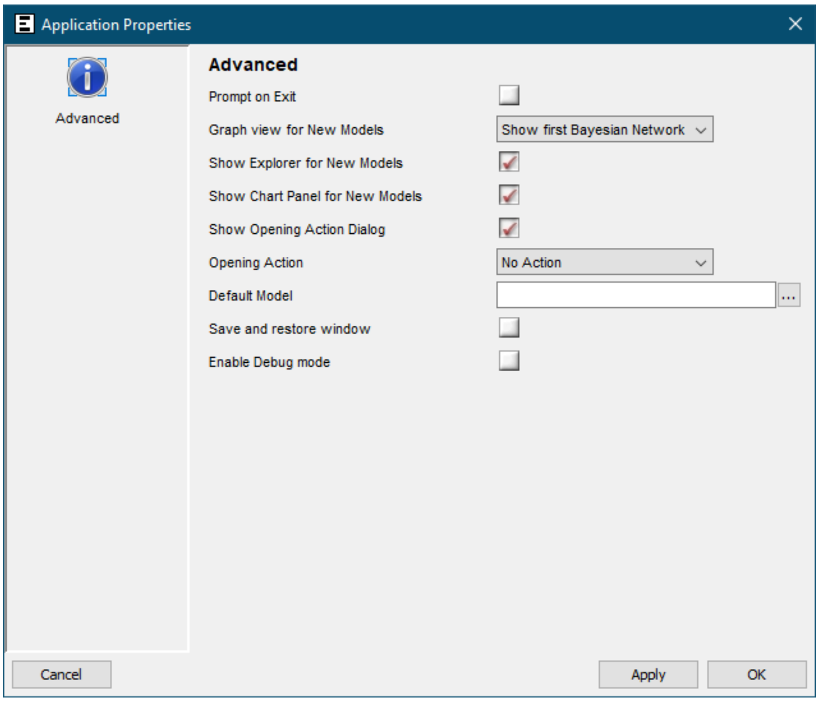
- **Help > About Agena.ai cloud services portal** This starts your web browser and opens the agena.ai cloud services portal where you can import your models for delivery of online services via cloud API and Web Applications (WebApp).

- Associated tool bar icon:

19.12 Application and model properties dialogs

The application properties dialog is available from the Tools menu and the menu bar button: . The dialog, which is shown in Figure 188, contains a single item called *File Handling*.

All settings defined in this dialog are model independent; that is, they will be applied to any model loaded into the application.



Advanced

Prompt on Exit ☐
Graph view for New Models: Show first Bayesian Network
Show Explorer for New Models ☑
Show Chart Panel for New Models ☑
Show Opening Action Dialog ☑
Opening Action: No Action
Default Model:
Save and restore window ☐
Enable Debug mode ☐

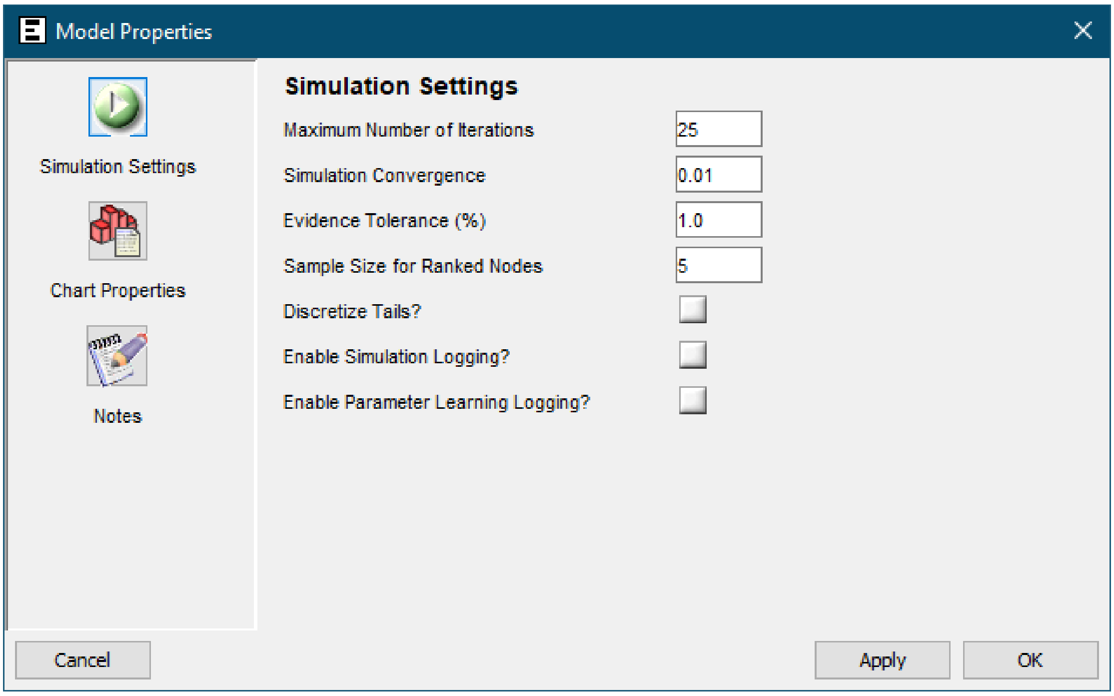
Figure 188 The application properties dialog

The options are:

- **Prompt on Exit** Defines whether a dialog asking whether you really want to exit should appear when the application is closed down.
- **Graph view for New Models** Defines what is shown on screen when new model is created. There are two options:
    - **Show entire Model** This shows the model containing Bayesian networks
    - **Show first Bayesian network** This shows the Graph view of the first Bayesian network in the model
- **Show Explorer for New Models** Defines whether the explorer will be expanded whenever a new model is created.
- **Show Chart Panel for New Models** Defines whether the chart panel will be expanded whenever a new model is created.
- **Show Opening Action Dialog** When this is checked, a dialog will be displayed each time you launch Agena.ai modeller asking you what you would like to do.
- **Default Model** The model file that is loaded as default when the application is loaded

The model properties dialog (shown in Figure 189) is available from the Tools menu and the menu bar button:

All settings defined in this dialog are saved with the model.



Simulation Settings

Maximum Number of Iterations 25

Simulation Convergence 0.01

Evidence Tolerance (%) 1.0

Sample Size for Ranked Nodes 5

Discretize Tails? ☐

Enable Simulation Logging? ☐

Enable Parameter Learning Logging? ☐

Figure 189 The model properties dialog

The model properties dialog covers:

- Simulation settings:
    - **Maximum Number of Iterations** This determines how long simulation goes on for. The default is 50 iterations, but most simulations stop well short of this number. If you want to force a simulation to stop early, perhaps to witness how simulation works, set the value to one or two.
    - **Evidence Tolerance (%)** This is percentage of the observed value used to mimic the true value in continuous interval nodes. Note that on occasion you may want to have as exact a value as possible and will choose tolerance values smaller than the default value which is 1%.
    - **Simulation Convergence** This determines the convergence of the entropy error values. Once a node's entropy value is less than this threshold it stops discretizing. Larger values E.g., 0.1 will stop simulation earlier and give less accurate results faster.
    - **Sample Size for Ranked Nodes** The number of samples taken when NPTs are generated for Ranked Nodes.
    - **Discretize tails.** Adds more states to tails in the probability distributions of all continuous and integer intervals nodes during simulation. This increase accuracy when calculating percentiles but increases calculation time.
    - **Enable Simulation Logging.** This generates an OutputLog.html file in your Agena.ai modeller directory giving details of the computations that have taken place during model calculation.
    - **Enable Parameter Learning Logging.** This generates an EMOutputLog.html file in your Agena.ai modeller directory giving details of the computations that have taken place during table learning from data.
- Model Graph Properties
    - **Decimal places to round to** This is the number of decimal places the graph will show for the probability values associated with each state of a node.
    - **Show horizontal graph properties as percentages** This is an on/off toggle. When ticked probabilities in nodes whose graphs are displayed horizontally (e.g. discrete nodes by default) are shown as percentages (e.g. 55%) rather than a probability value between 0 and 1 (e.g. 0.55). The default setting is ON.
    - **Minimum probability displayed** This is the minimum probability displayed on the graph as a number. The default setting is 0.01, so if a state has a probability value of less than this (say 0.005) its value will not be displayed. Obviously you should change the setting to a smaller number if you wish to see the values of states with very low probabilities.
- Notes

  In the notes area you can add any number of new notes, reminders and tips about your model.

19.13 Explorer dialog

The explorer pop-up dialog can be accessed by right clicking any Bayesian network in the explorer view. The options available are:

- **Delete** This deletes the selected Bayesian network and any links connected to them.
- **Rename** This renames the selected Bayesian network.
- **Locate in Graph view** this selects and shows the Bayesian network on the Graph view.
- **Local Calculation with Ancestors** This calculates this Bayesian network with all linked ancestors.
- **Sort** This sorts the list of Bayesian networks according by name and type
- **Export Data File (.CSV)** When chosen at the root of the explorer tree all Bayesian networks will be exported. Whilst selecting one Bayesian network, only that Bayesian network will be exported as a CSV file.
- **Import Data File (.CSV)** When chosen at the root of the explorer tree the CSV file imported will be imported to all Bayesian networks in the modal that match the Bayesian networks in the CSV file. When a specific Bayesian network is imported to only a CSV file previously exported from that file can be imported.

20. Node types

Nodes in Agena.ai modeller are either discrete or simulation. Discrete nodes are created in the Graph view by selecting the icon:

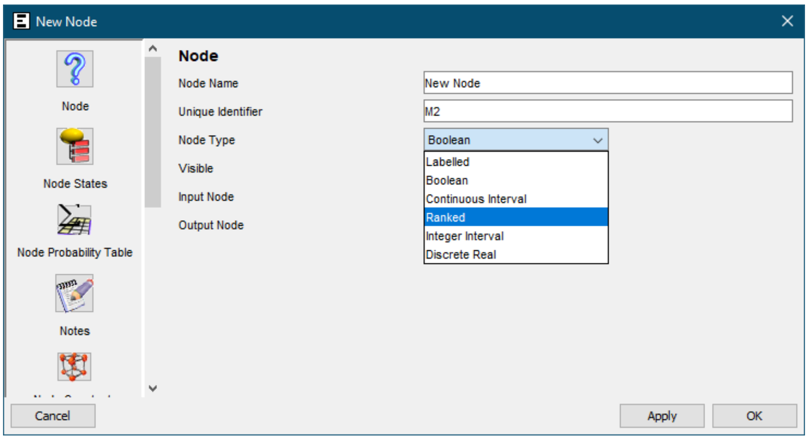
Simulation nodes (which are used to represent continuous numeric variables as described in Section 6) are created in the Graph view by selecting the icon

This section describes the various discrete node types supported in Agena.ai modeller. These are:

- Boolean: — "True", "False" for example
- Labelled — "Red", "Green", "Blue" for example
- Ranked — "Low", "Medium", "High" for example
- Integer Interval — 0, 1, [2, 3], [4], [5 – infinity] for example
- Continuous Interval — [0, 10], [10 – 20], [20 – infinity] for example
- Discrete Real — -2, 0, 2.5, 3.6, 10 for example (any unordered collection of values)

The expressions associated with each node type are covered later.

By default, when you create a new discrete node, its type is Boolean. To see this, create a new node, and bring up the properties dialog by right-clicking on the node. If you click on the drop-down box next to "Node Type", you will see the alternative node types available as shown in Figure 190.



New Node

Node

Node Name: New Node

Unique Identifier: M2

Node Type: Boolean

Figure 190 Available node types

To understand the different node types you should set each one in turn and, for each one, you should check:

1. what the default state values (by clicking on the ***Node States*** tab); and

2. what the available expressions are (by clicking on ***Node Probability Table*** tab, and setting the ***NPT Editing Mode*** to ***Expression***.

In the following subsections we summarise what you should find for each type.

20.1 Boolean type

The default states for Boolean nodes are “True” and “False”. If you click on the **Node States** tab and select **State Options,** you can see that you can also select “Yes” and “No” or “Customised”. The latter enables you to define any two labels of your choice.

The available expression types for Boolean nodes are comparative expressions. Examples of such valid expressions (see Section 21.2 for details) are:

> if (parent >0.5, “True”, “False”) meaning ‘if the parent value is greater than 0.5 then the value is “True”; otherwise it is “False”’.
>
> if (parent1 >0.5 && parent2==”False”, “True”, “False”)
>
> noisyor (parent1, 0.7, parent2, 0.8, parent3, 0.5, 0.1)

20.2 Labelled type

By default, when you change the type of a new node to Labelled, it will have two states with the text labels ‘False’ and ’True’. You can edit, add and delete states by typing directly into the text area in the Node States section of the Node Properties dialog.

You will normally use a labelled type if the variable represents some classification. For example, the node might be ‘programming language’ and have states C, C++, Java, Basic etc.

If the type is labelled, you would normally define its NPT manually. However, there is one Expression type you can use instead, namely Comparative. Suppose, for example, that you are trying to capture the relationship between programmer experience (a Ranked Node) and programming language used (Labelled). If the former is a parent of the latter, then you might want to express the NPT using a comparative expression like:

> if (pe >10, C, C++)

20.3 Ranked type

A node should be declared of Ranked type if it represents a variable on an ordinal scale. The default set of states for a new ranked node is a five-point scale from ‘Very Low’ to ‘Very High’). There are five alternative pre-defined scales that involve three, five or seven states and either ascend or descend. You may define your own scale altogether by editing the labels of any chosen pre-defined scale.

For detailed modelling purposes, the crucial thing you need to know about ranked nodes is that, no matter what the labels are or how many states a node has, the underlying mathematical scale goes from 0 to 1 in equal intervals. You need to know this in case you wish to use this information for defining certain types of expressions in any child nodes.

The only available expression for the Ranked node type is a version of the Normal called the Truncated Normal or TNormal, but you have great flexibility about how you define the mean and variance of the TNormal.

20.4 Continuous interval type

When you select Continuous Interval type you will get a warning message advising that it is more efficient to use a simulation node. However, there are many situations in which it does make sense to use a discrete node and, in such cases, you should simply ignore the warning (press “OK”). The default state values for a Continuous Interval node are the ranges –inf to -1.0, -1.0 to 1.0 and 1.0 to inf.

Suppose that you wish to define a completely different set of intervals, such as:

0-1, 1-2, 2-3, 3-4, 4-5, 5-10, 10-15, 15-20, 20-100, 100-inf

There are a number of ways you can achieve this. One way is as follows:

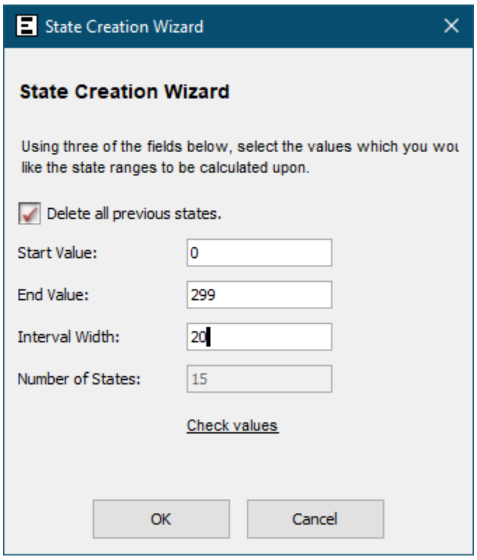
1. Deselect the option “Makes lower bound infinity”
2. Edit the lower bound text, replacing -1 with 0.
3. Press Apply.
4. Now click the last ‘insert’ button 4 times.

   Insert

5. The set of states has now been redefined as 0-1, 1-2, 2-3, 3-4, 4-5, 5-inf
6. Now click the add new state wizard button

   Wizard

   associated with the last state to bring up the state wizard. Enter the values shown in figure Figure 191 Using the state creation wizard to create a sequence of equal interval states and press “OK”. You will now have the states 5-10, 10-15 and 15-20 added.



State Creation Wizard

State Creation Wizard

Using three of the fields below, select the values which you wou
like the state ranges to be calculated upon.

☑ Delete all previous states.

Start Value: 0

End Value: 299

Interval Width: 20

Number of States: 15

Figure 191 Using the state creation wizard to create a sequence of equal interval states

7. Click the ‘insert new state button’ associated with the last state

This adds a new state 20-21. Simply edit the upper bound to 100 and press “Apply” to get the desired set of states.

An alternative approach would have been to start by clicking the “Remove all states” link and then creating the states 0-1, 1-2, 2-3, 3-4 and 4-5 using the wizard.

It is also possible to specify states that are single point values by deselecting the check box to the left of the state. For example, deselecting the check box next to state 1-2 will result in an additional ‘point value’ state of 1.0.

It is also important to note that you can delete any individual state by right-clicking on it and choosing “Remove Selected”. You can also multiple-select states (using the control key) and remove them in the same way.

For Continuous Interval nodes there is a wide range of functions and distributions available for use in expressions. These are described in detail in Section 21.

20.5 Integer interval type

Integer Interval types work exactly like the Continuous Interval type with the exception that the ranges are collections of consecutive integers, and the end points of any range must be an integer value (so whereas you could define an interval like 0.5 to 1.5 for a continuous interval type you cannot do this for integer interval type). So, the range 0-5 contains the integers 0, 1, 2, 3, 4 and 5 but not any of the real values in between. The default states for Integer Interval nodes are –inf – 1, 0 to 4, 4 to inf. Editing, adding and removing states is the same as for nodes of the Continuous Interval type. Like Continuous Interval nodes, Integer Interval nodes can have states that are a single value rather than a range.

The expressions available for Integer Interval nodes are the same as those available for Continuous Interval nodes.

20.6 Discrete real type

The Discrete Real type should be viewed as a special case of the Labelled type where the labels have meaningful numerical values. For example, suppose you are trying to calculate a sum payable S given different possible tax rates T. Suppose that the different tax rates are 5, 10, 17.5 and 25. Then you should make T a parent node of S and define its type as Discrete Real with states 5, 10, 17.5, 25. Like states for the Labelled type, Discrete Real states can be supplied by typing directly into the displayed text area in such a way that each state occupies a single line. Alternatively, you can define the states by clicking on the wizard button. The default states are 0 and 1.

20.7 Simulation node types

Note that for simulation nodes only two node types are supported Continuous interval and Integer interval.

21. Bayesian networks

You can build complex models in Agena.ai modeller by connecting simple models. With Bayesian networks we can modularise the model into chunks representing logical groupings of risks or time dependencies between objects.

When we do this, we call the individual model components *Bayesian networks*. Figure 192 shows a model composed of two Bayesian networks.

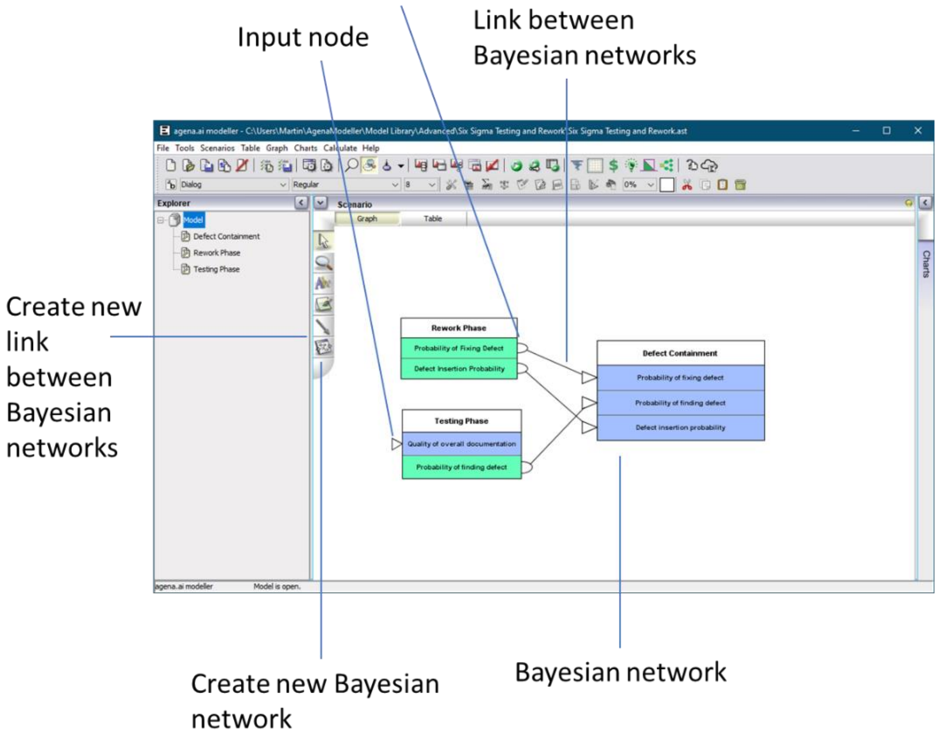
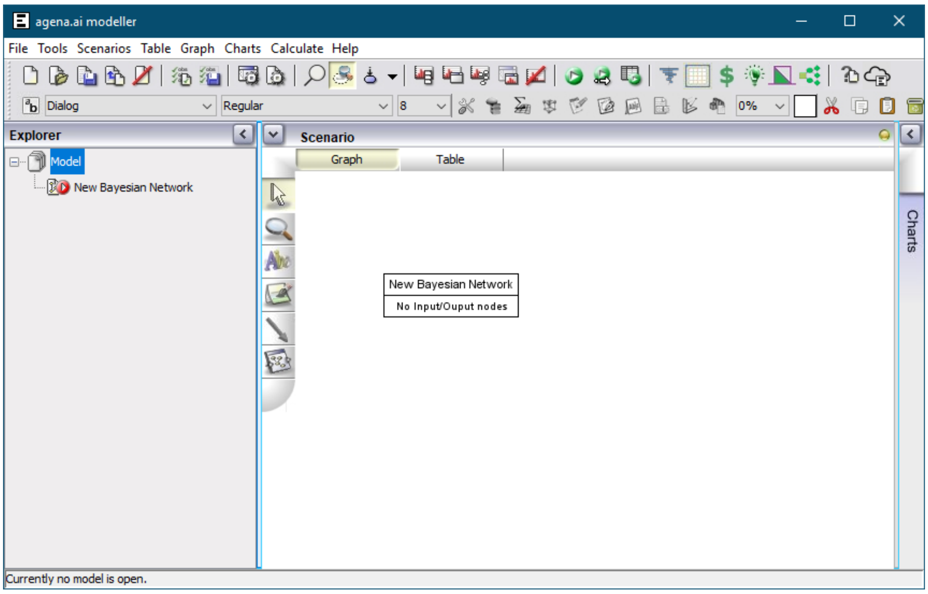


Figure 192 Bayesian networks

To create a model from Bayesian networks you should select the top-level item in the Explorer view. By default, when you create a new model there is a single Bayesian network as shown in Figure 193



Currently no model is open.

Figure 193 Default model and Bayesian network

To add a new Bayesian network you select the ‘create new Bayesian network’ icon highlighted in Figure 193. You can also import models as new Bayesian networks by using the icon on the tool bar or using the menu item.

21.1 Viewing and navigating Bayesian networks

New and imported Bayesian networks are displayed both in the Graph view and listed in the explorer view. Selecting a particular Bayesian network in explorer or double-clicking the Bayesian network in Graph view will result in that object being displayed in full in the Graph view as shown in Figure 194.

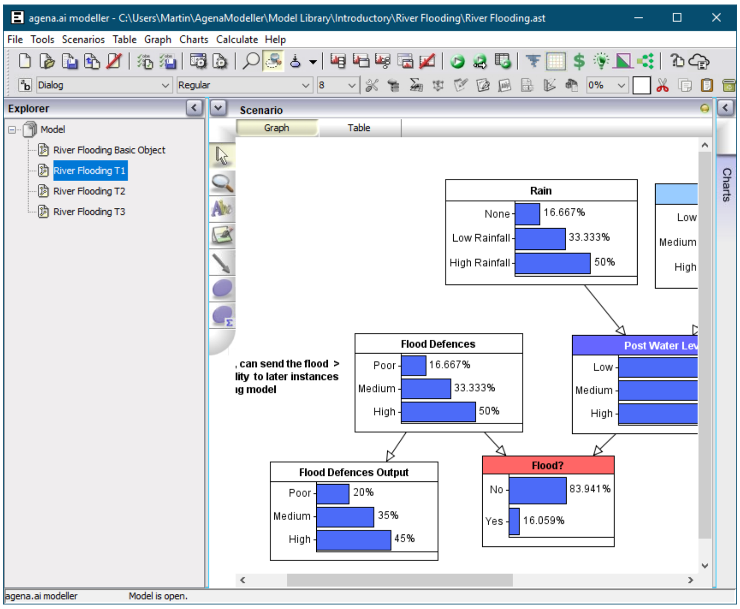


Figure 194 Displaying a model’s Bayesian network

To switch back to viewing the entire model as Bayesian networks simply select the top item in the explorer.

When you have a model containing many Bayesian networks it can be difficult to locate a particular Bayesian network that you might be interested in. To avoid this problem, you simply right-click on the Bayesian network in explorer view and select the “Locate in Model view” option as shown in Figure 195. After doing this the associated Bayesian network will be shown as selected in the Graph view.

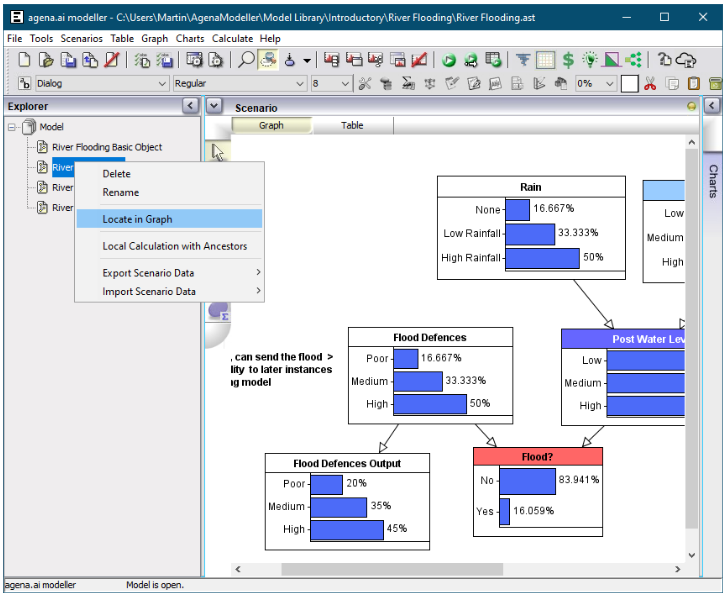


Figure 195 Locating a Bayesian network in model view

21.2 Connecting Bayesian networks

To connect two Bayesian networks together you must have appropriate input and output nodes defined in the Bayesian networks. As soon as you make a node an input or an output this will be displayed in the Entire Model view as shown in Figure 192

An output node from one Bayesian network can be connected to an input node in another by selecting the ‘create new link between objects’ icon. You then click once in the area of the output node name and then click once in the area of the input node name (this works like the edge join tool in Graph view). An edge connecting them will be created.

There are three types of connections that are allowed:

a. The default connection (prior to version 6.0 of Agena.ai modeller this was the only available type of connection). This is where the input and output nodes are exactly the same type with exactly the same set of state values. The result of this linking is to pass the entire set of probability values from the output node to the input node.

b. From a simulation node to a simulation node, you can either pass the full set of marginals (as was the previous default), or the value of a summary statistic as a constant. So, for example, the output node might represent a variable “height” and the input node might represent a variable “mean height”. In this case the link type you would select would be the summary statistic “Mean”.

c. From a non-simulation node to a simulation node, you can pass the value of a single state as a constant. For example, the node “Flood” in the above tutorial is a Boolean node. We could link this to a simulation node (with a range 0 to 1) called “Flood probability” in another Bayesian network and specify that the value passed is the value of “True”. If the value of the state “True” is 0.6 in the node “Flood” then the node “Flood probability” will have the value 0.6.

The new options **b** and **c** require you to specify what you wish to pass. There is a link type option in node properties which can be accessed by right-clicking on the **input** node and selecting properties (**note**: **the link type option is only displayed for an input simulation node for which a link has already been created**). You then select the link type option on the left hand side as shown in Figure 196.

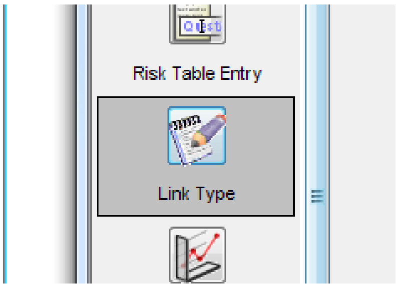
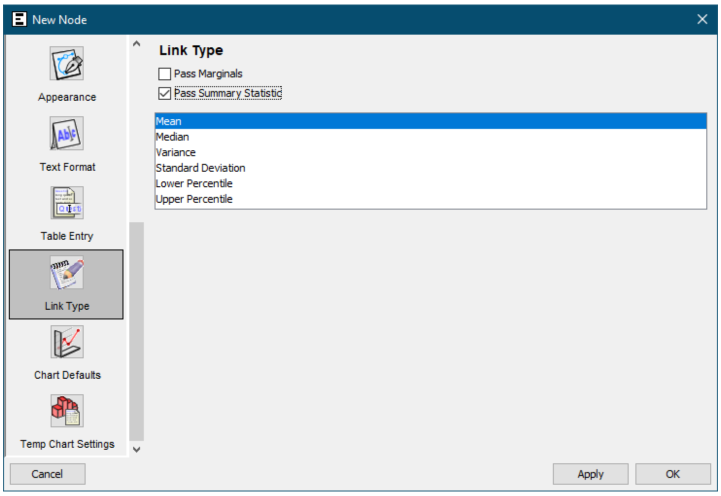


Figure 196 Selecting link type in node properties (only available for input simulation nodes where a link has already been created)

The options in link type vary based upon whether it is option **b** or **c**. For option **b** you will see the dialogue in Figure 198.



Link Type

☐ Pass Marginals

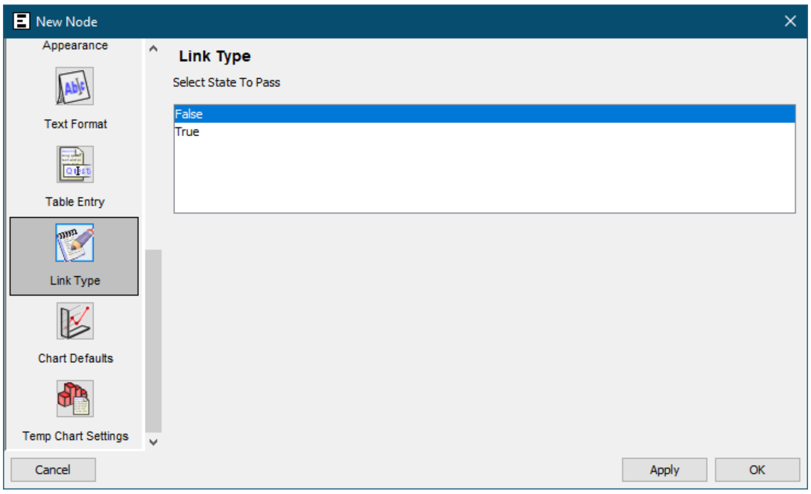
☑ Pass Summary Statistic

Mean
Median
Variance
Standard Deviation
Lower Percentile
Upper Percentile

Figure 197 Link types option for connecting continuous nodes

Here, in addition to passing full marginals, you have the option to select which summary statistic to pass.

For option **c**, you can select which state value to pass as shown in Figure 198



Link Type

Select State To Pass

False
True

Figure 198 Link type options for passing from non-continuous to continuous

Using a link type, passing a state or summary statistic, disables the node probability table and expression editing and turns off the constants declaration and use for that node.

An output node can be connected to input nodes in more than one Bayesian network. However, it is not possible to connect more than one output node to a single input node. If, for example you wanted to aggregate the values of two output nodes and use this aggregated value as the input to a new Bayesian network then the way to do this would be to ensure that the aggregation is done in the new Bayesian network (generally the new Bayesian network would have an aggregation node who two parents are the two input nodes). That way you can achieve the desired effect by connecting the two outputs to the two inputs nodes of the new Bayesian network.

Note: Loops are not allowed between Bayesian networks. Thus, Bayesian network A cannot link to object B if B connects to A.

The definitive list of input-output connection types is given in Table 10.

Table 10 Input-output connection types

Source	**Target**	**Types**
Boolean	Boolean	Marginal
Ranked	Ranked	Marginal
Labelled	Labelled	Marginal
Discrete Real	Discrete Real	Marginal
Continuous	Continuous	Marginal
Integer	Integer	Marginal
Continuous (Simulation)	Continuous	Marginal
Continuous (Simulation)	Continuous (Simulation)	Marginal, Summary Statistic Value
Integer (Simulation)	Integer	Marginal

Integer (Simulation)	Integer (Simulation)	Marginal, Summary Statistic Value
Boolean	Continuous (Simulation)	State
Labelled	Continuous (Simulation)	State
Boolean	Integer (Simulation)	State
Labelled	Integer (Simulation)	State

21.3 Deleting and renaming Bayesian networks

By right-clicking on any object in the explorer you will get the menu shown in Figure 199. By selecting the appropriate item, the Bayesian network can be deleted or renamed.

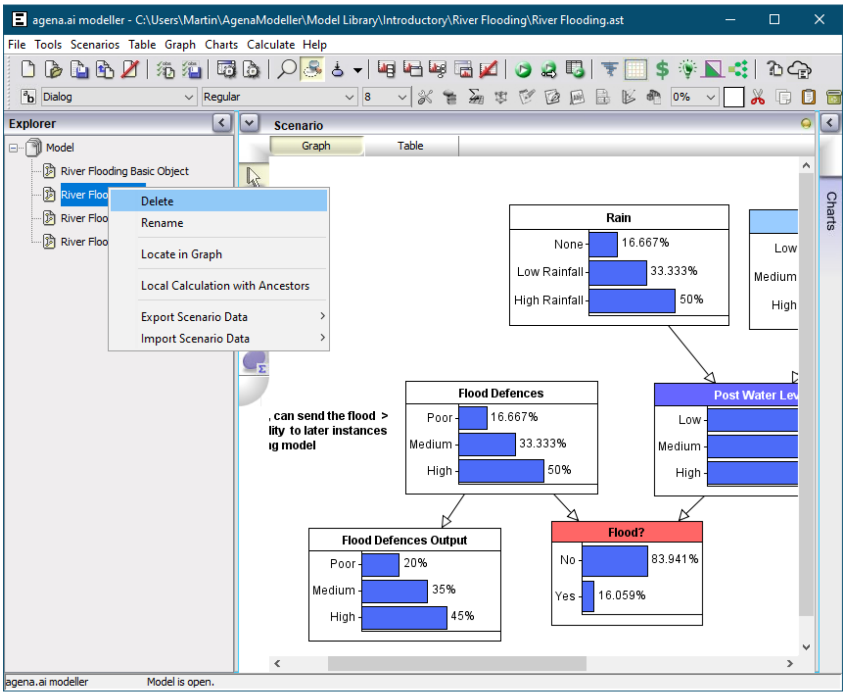


Figure 199 Deleting or renaming Bayesian networks

You can also delete Bayesian networks in Graph view by right-clicking on the object to bring up the menu shown in Figure 200 and choosing “Delete”.

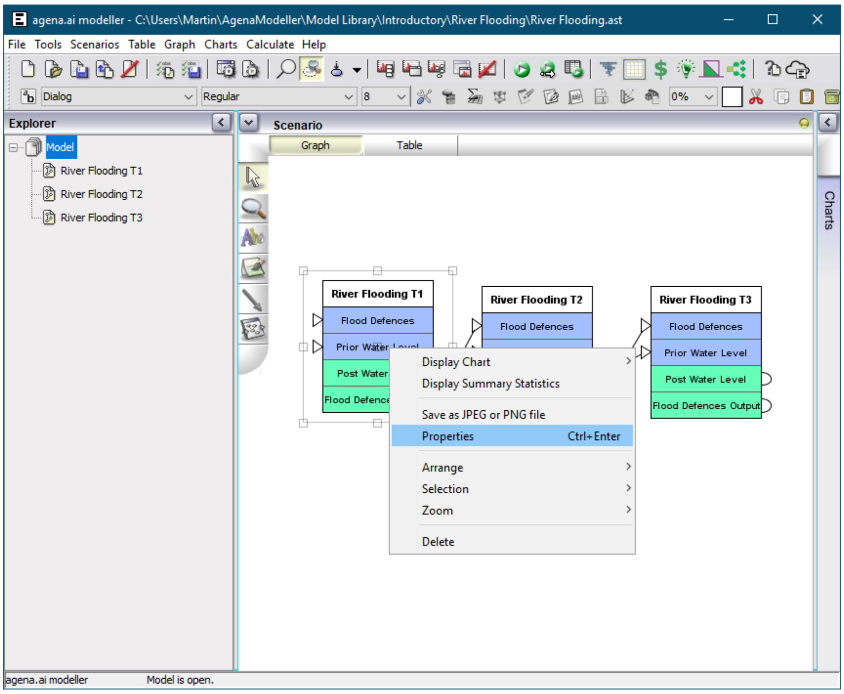
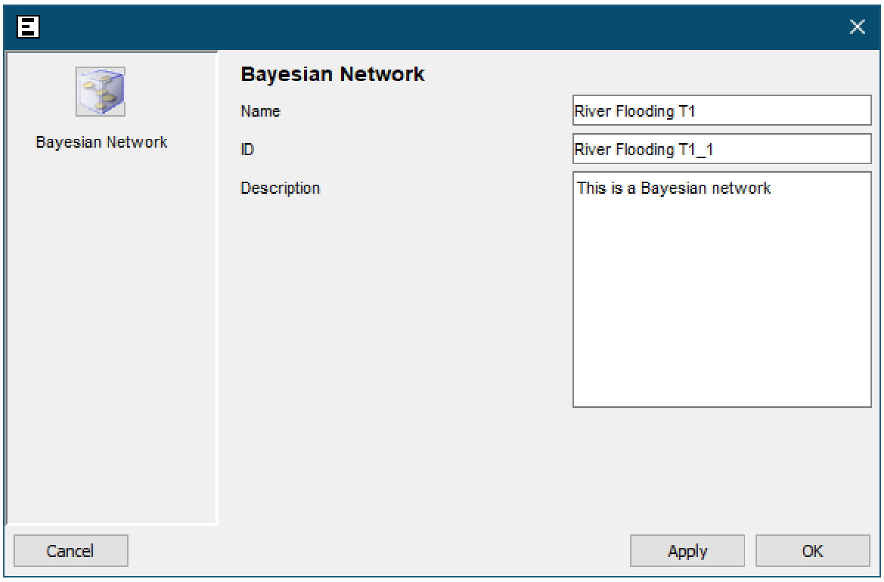


Figure 200 Editing Bayesian networks

If you select Properties from this menu you will get the dialog shown in Figure 201. To rename the Bayesian network simply change the Name property.



Bayesian Network

Name: River Flooding T1

ID: River Flooding T1_1

Description: This is a Bayesian network

Figure 201 Properties of a Bayesian network

21.4 Appearance of Bayesian networks

Just as you can change the appearance and layout of the graphs in model view, so you can in the top-level view for Bayesian networks. In particular, the following functions (available by right-clicking a Bayesian network or via an icon) work exactly the same here:

- Zoom (icon)

- Text (icon)
- Import picture (icon)
- Arrange (right-click properties)
- Save as jpeg (right-click properties)

21.5 Calculations using Bayesian networks

Within Bayesian networks full diagnostic calculations are supported. However, *between* Bayesian networks only predictive calculations are performed by sequentially passing the resulting marginal distributions from one (ancestor) Bayesian network to a connected descendant Bayesian network.

You must therefore be careful to only split a model into component Bayesian networks according to the purpose of the model. In purely diagnostic models ancestor objects will be completely unaffected by data entered in any or all ancestor Bayesian networks.

The flow of calculation is best illustrated by means of the River Flooding example (which you will find in the Agena.ai modeller model library) as shown in Figure 202. This example shows how water levels and flood defences might change over three time periods, where each time period is modelled by a separate Bayesian network {River Flooding T1, River Flooding T2, River Flooding T3}. Notice that when a calculation is run any marginal probabilities associated with output nodes are passed from T1 to T2 and then to T3. No information is passed from T3 back to its ancestors T2 and T1.

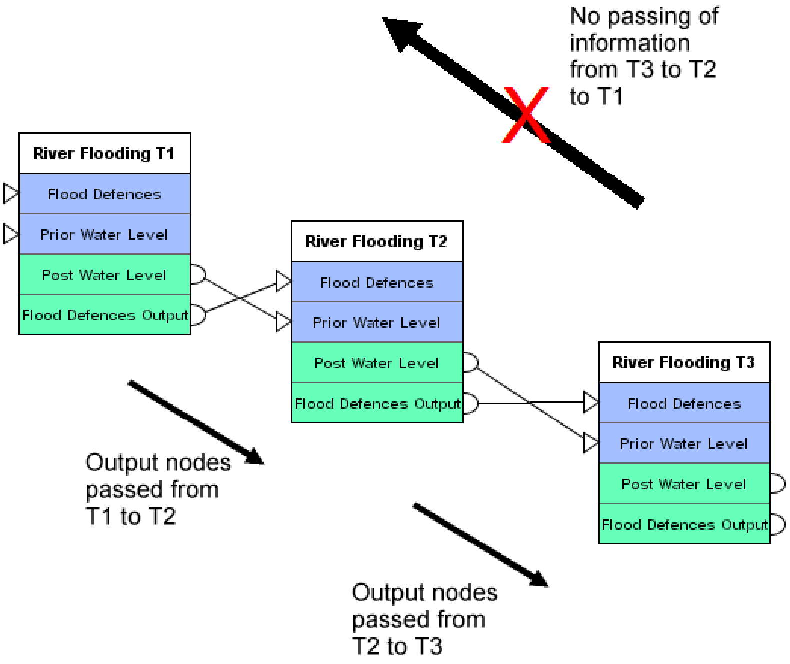


Figure 202 Sequential prediction *between* Bayesian networks

Each time you run a calculation with a model containing multiple Bayesian networks it will perform the calculation sequentially on all models. However, there is one other calculation options available that can be accessed via the explorer view:

- **Local calculation with ancestors** This allows you to run a Bayesian network plus all ancestor Bayesian networks. In this way a consistent result can be produced on the model but where the calculation stops short of running the whole model. In essence you can divide and conquer by running smaller calculations over parts of the overall model. This is especially useful where you want to reduce the memory load and execute a model a bit at a time and if you want to run different parts of the model at different levels of accuracy (iterations, convergence etc.)

To access these calculation options right-click on a Bayesian network on the Explorer view and select the option you require as shown in Figure 203.

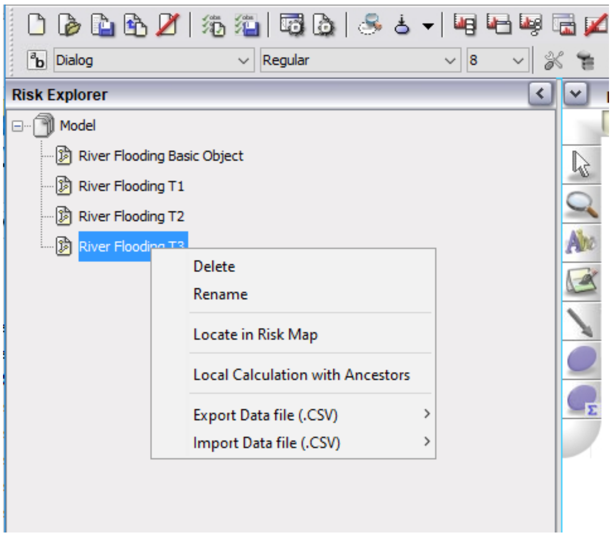


Figure 203 Accessing Calculation options on the Explorer

When you have performed a calculation the modification log will be updated as will the status of each Bayesian network. If a Bayesian network is inconsistent and has not been updated to take account of any modifications, the Bayesian network as listed in the explorer will appear with a "modified" icon as shown in Figure 204. Here the "River Flooding T1" Bayesian network has been updated with a new data and because the models T2 and T3 depend, sequentially, on T1 all are marked as modified meaning that next time a calculation is run all these models will be recalculated.

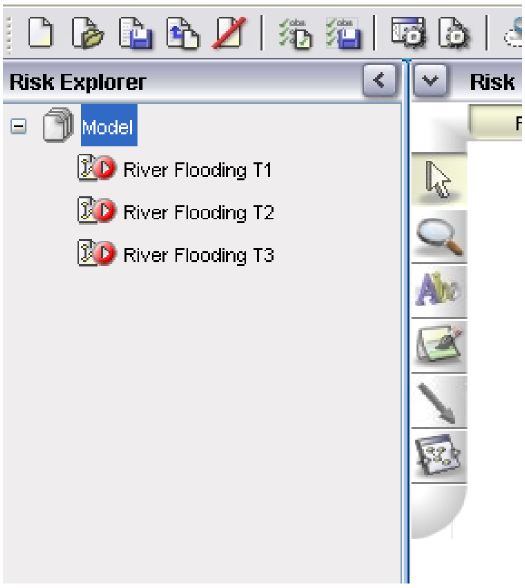


Figure 204 Modified Status icons on Bayesian networks in Explorer when model T1 modified

If, however, the only model to be modified was T3 this object alone in the explorer would be listed as modified, simply because it does not affect calculations at T1 and T2. This is shown in Figure 205.

Figure 205 Modified Status icons on Bayesian networks in Explorer when model T3 modified

Notes:

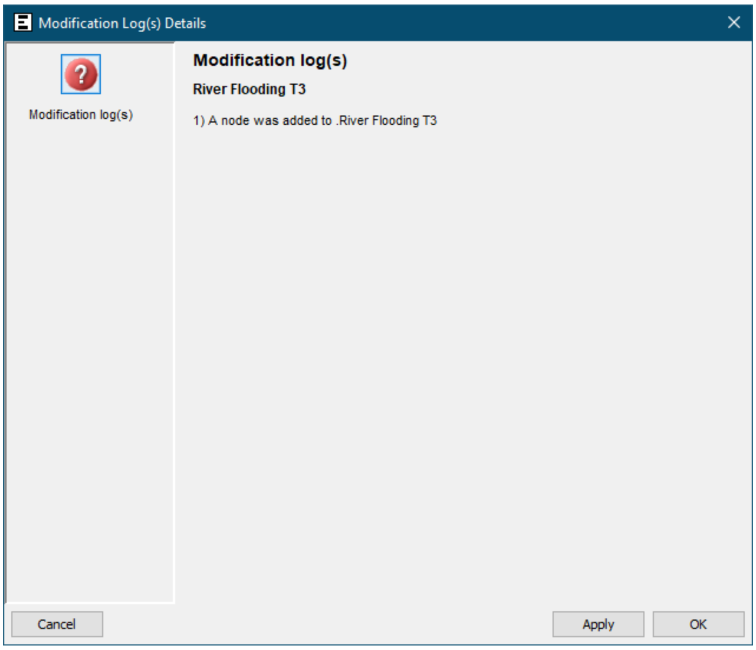
1. The sequential calculation scheme assumes that the input/output nodes are independent of each other. However, if they share a common parent node then they might not be. You must make sure that if there is very strong dependence between nodes it is important that they be placed within the same Bayesian network.

2. Any expressions or NPT values declared on an input node will be overwritten when full calculation or local calculation with ancestors is performed.

3. Input nodes cannot be set as simulation nodes and vice versa. This is a consequence of point 2.

21.6 The modification log and inconsistencies in the model

If a modification has been made the modification log appears as an option alongside the Local Calculation option when Bayesian networks are selected by right clicking in the explorer. The modification log lists the changes made to the model that require it to be recalculated, as shown in Figure 206.



Modification log(s)

River Flooding T3

1) A node was added to .River Flooding T3

Figure 206 Modification Log

The modification log tracks the status of the following changes to models in Agena.ai modeller:

- Expressions on nodes
- Simulation settings
- Bayesian networks
- Nodes in Graph views
- Node States
- Data and cases
- Data entered

22. Statistical distributions

The continuous and discrete (integer valued) statistical distributions that are supported in Agena.ai modeller are listed in Table 11.

Continuous
Beta
BetaPert
Chi-Square
Exponential
Extreme Value
Gamma
Log Normal
Logistic
Normal
Student-t
TNormal
Triangular
Uniform
Weibull

Discrete
Binomial
Exponential
Geometric
Hypergeometric
Negative Binomial
Normal
Poisson
TNormal
Triangular
Uniform

Table 11 Statistical distributions

Continuous distribution types are only usable on continuous interval node types, with the exception of Uniform, Triangular, Normal and TNormal which can also be applied to discrete (integer) node types.

Each distribution function is described more fully in the following subsections.

Note that:

1. Agena.ai modeller does not check for sensible bounds or parameters on any of these distributions. Be careful to ensure you set the node's interval bounds to match those required for the distribution. If you choose a value outside of the logical bounds for a parameter Agena.ai modeller will attempt to enforce the logical maximum or minimum value. For example, with the Chi-Squared distribution if you choose an invalid value of the parameter less than one Agena.ai modeller will set the value to the nearest valid value, which is one, automatically. However, despite this there remains a danger that you may generate a probability density function that does not lie in the state range defined for the node in question — when this happens an error message appears warning you that zero cells have been generated and this will lead to an inconsistency upon calculation of the model.

2. The Truncated Normal (TNormal) distribution will not necessarily have the mean and variance that you specify. However, if the variance is small and the mean you specify is not relatively close to zero for the range you specify, then the resulting distribution will have a mean close to what you specified (this is because the resulting TNormal will be "almost" the same as a Normal).

22.1 Continuous distributions

22.1.1 Beta distribution

Probability function: $p(X)=\dfrac{(1-x)^{\beta-1}x^{\alpha-1}}{B(\alpha,\beta)}=\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{\beta-1}x^{\alpha-1}$

Domain: $0\le X\le 1$

Parameter domain(s): $\alpha>0,\beta>0$

Mean: $E(X)=\dfrac{\alpha}{\alpha+\beta}$

Variance: $V(X)=\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Note: The domain of the Beta distribution can be extended to any finite range in the region $L\le X\le U$.

Example: Beta(3, 7, 0, 10)

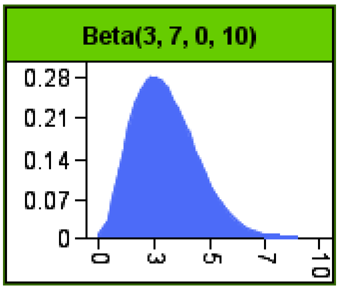


Figure 207 Beta(3,7,0,10) Distribution Example

22.1.2 **BetaPert distribution**

Probability function: $p(X)=\dfrac{(1-x)^{\beta-1}x^{\alpha-1}}{B(\alpha,\beta)}=\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(1-x)^{\beta-1}x^{\alpha-1}$

The probability distribution is the same as the Beta distribution except the paramters are defined by mode (*m*), confidence (*lambda*), lower (*a*) and upper(*b*) bounds, where:

$$\alpha=\frac{(\mu-a)}{(b-a)}\left[\frac{(\mu-a)(b-\mu)}{\sigma^2}-1\right]$$

$$\beta=\frac{(b-\mu)}{(b-a)}\left[\frac{(\mu-a)(b-\mu)}{\sigma^2}-1\right]$$

Domain: $a \le m \le b$

Parameter domain(s): $\alpha > 0, \beta > 0$

Mean: $E(X) = \mu = \dfrac{a + \lambda m + b}{\lambda + 2}$

Variance: $V(X) = \sigma^2 = \dfrac{(b-a)^2}{(\lambda+2)^2}$

Example: BetaPert(0.2, 4, 0.1, 1.0)

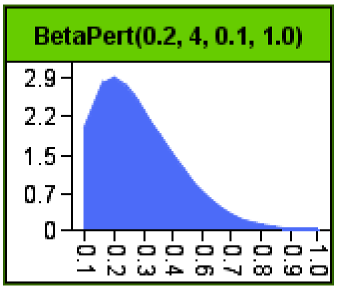


Figure 208 BetaPert(0.2, 4, 0.1, 1.0) Distribution Example

22.1.3 Chi-Square distribution

Probability function: $p(X) = \dfrac{2^{-v/2}}{\Gamma(v/2)} x^{(v/2)-1} e^{-v/2}$

Domain: $X > 0$

Parameter domain(s): $v > 1$ where v is the degrees of freedom.

Mean: $E(X) = v$

Variance: $V(X) = 2v$

Example: Chisquare(5)

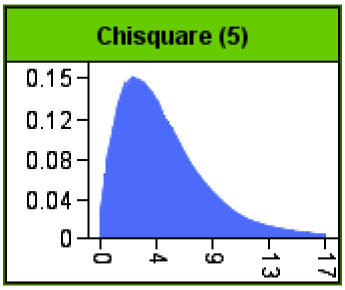


Figure 209 Chi-Square(5) Distribution Example

22.1.4 Exponential distribution

Probability function: $p(X) = e^{-\lambda x}$

Domain: $X > 0$

Parameter domain(s): $\lambda > 0$

Mean: $E(X) = \lambda$

Variance: $V(X) = \lambda^2$

Example: Exponential(2)

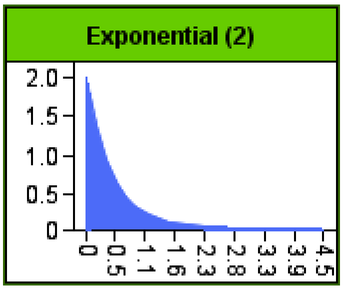


Figure 210 Exponential(2) Distribution Example

22.1.5 Extreme Value (Gumbel) distribution

Probability function: $p(X) = e^{-e^{v}}$

Domain: $-\infty < X < \infty$

Parameter domain(s): $-\infty < v < \infty$

Order: 0 for maxima, 1 for minima

Location: mean

Scale: standard deviation

Example: Extreme Value(0, 10, 10)

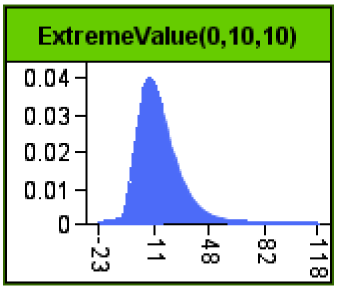


Figure 211 Extreme Value(0, 10, 10) Distribution Example

22.1.6 Gamma distribution

Probability function: $p(X) = x^{\alpha-1} \dfrac{\beta^{\alpha} e^{-\beta x}}{\Gamma(\alpha)}$

Domain: $X > 0$

Parameter domain(s): $\alpha > 0, \beta > 0$ where $\beta = \dfrac{1}{\lambda}$ and λ is a rate parameter.

Mean: $E(X) = \alpha\beta$

Variance: $V(X) = \alpha\beta^2$

Example: Gamma(3, 20)

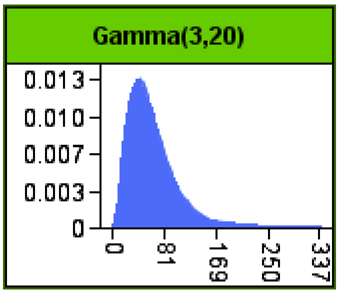


Figure 212 Gamma(3,20) Distribution Example

22.1.7 Log Normal distribution

Probability function: $p(X) = \dfrac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x - \mu)^2/(2\sigma^2)}$

Domain: $X > 0$

Parameter domain(s): $-\infty < \mu < \infty, \sigma^2 > 0$

Mean: $E(X) = e^{(\mu + (1/2\sigma^2))}$

Variance: $V(X) = e^{2\mu} e^{\sigma^2} (e^{\sigma^2} - 1)$

Example: Lognormal(1.5, 2)

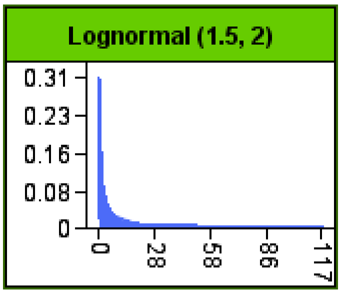


Figure 213 LogNormal(1.5, 2) Distribution Example

22.1.8 Normal distribution

Probability function: $p(X)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$

Domain: $-\infty < X < \infty$

Parameter domain(s): $-\infty < \mu < \infty, \sigma^2 > 0$

Mean: $E(X)=\mu$

Variance: $V(X)=\sigma^2$

Example: Normal(0, 100)

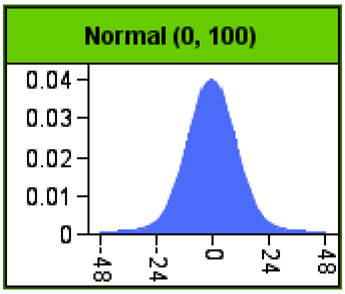


Figure 214 Normal(0,100) Distribution Example

22.1.9 Logistic distribution

Probability function: $p(X)=\frac{e^{-(x-\mu)/\beta}}{\beta[1+e^{-(x-\mu)/\beta}]^2}$

Domain: $-\infty < X < \infty$

Parameter domain(s): $v > 1$ where v is the degrees of freedom.

Mean: $E(X)=\beta$

Variance: $V(X) = \frac{1}{3}\pi^2\beta^2$

Note: In Agena.ai modeller the required parameters are "Mu" $= \mu$ and "Beta" $= \beta$.

Example: Logistic(2,3)

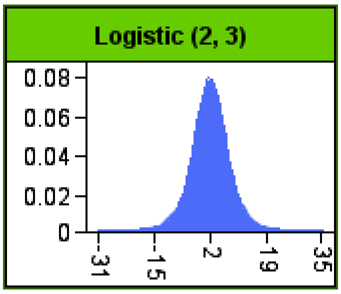


Figure 215 Logistic(2,3) Distribution Example

22.1.10 Student-t distribution

Probability function: $p(X) = \dfrac{\Gamma\left[\frac{1}{2}(r+1)\right]}{\sqrt{r\pi}\,\Gamma(r/2)(1+x^2/r)^{(r+1)/2}}$

Domain: $-\infty < X < \infty$

Parameter domain(s): $r \geq 0$ where $r \equiv n-1$ and n is the degrees of freedom.

Mean: $E(X) = 0$

Variance: $V(X) = \dfrac{r}{r-2}$

Example: Student(10)

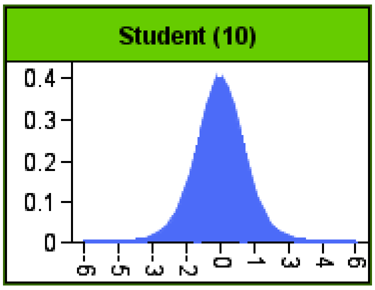


Figure 216 Student-t(10) Distribution Example

22.1.11 Truncated Normal (TNormal) distribution

Probability function: $p(X)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$

Domain: $L \le X \le U$

Parameter domain(s): $-\infty < \mu < \infty, \sigma^2 > 0$

Mean: $E(X) \approx \mu$

Variance: $V(X) \approx \sigma^2$

Note: The domain of the (doubly) truncated Normal distribution is restricted to the region $L \le X \le U$ and under these circumstances the mean and variance of the truncated distribution is only approximated by the mean and variance of the untruncated distribution. Depending on the truncation the true mean and variance may differ significantly from the supplied values.

Example: TNormal(0, 100, 0, 50)

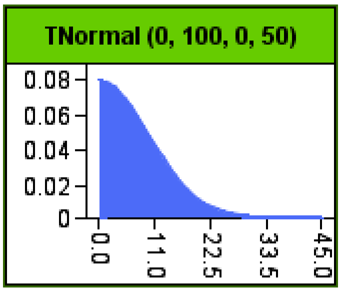


Figure 217 Truncated Normal (0,100,0,50) Distribution Example

22.1.12 Triangular distribution

Probability function: $p(X)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x \le c\\ \dfrac{2(b-x)}{(b-a)(c-a)} & \text{for } c < x \le b\end{cases}$

Domain: $a \le X \le b$

Parameter domain(s): $b > c > a$

Mean: $E(X)=\frac{a+b+c}{3}$

Variance: $V(X)=\frac{a^2+b^2+c^2-ac-ab-cb}{18}$

Note: In Agena.ai modeller a is “Left” c is “Middle” and b is “Right”.

Example: Triangular(5, 7, 10)

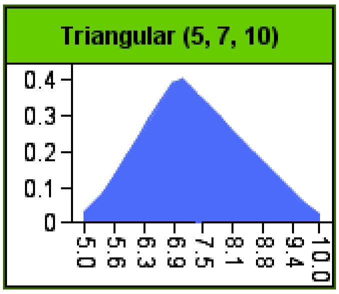


Figure 218 Triangular(5,7,10) Distribution Example

22.1.13 Uniform distribution

Probability function: $p(X) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{1}{(b-a)} & \text{for } a < x < b \\ 0 & \text{for } x > b \end{cases}$

Domain: $a < X < b$

Parameter domain(s): $b > a$

Mean: $E(X) = \dfrac{a+b}{2}$

Variance: $V(X) = \dfrac{(b-a)^2}{12}$

Note: In Agena.ai modeller a is “Lower Bound” and b is “Upper Bound”.

Example: Uniform(0, 50)

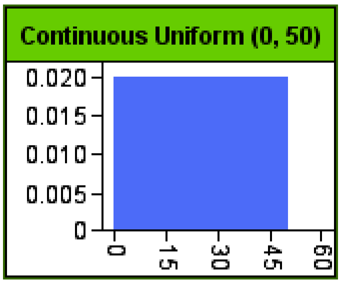


Figure 219 Uniform(0,50) Distribution Example

22.1.14 Weibull Distribution

Probability function: $p(X) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}}$

Domain: $X > 0$

Parameter domain(s): $\alpha > 0, \beta > 0$ [α is the shape parameter and β is the scale parameter]

Mean: $E(X) = \beta^{1/\alpha}\Gamma(1+1/\alpha)$

Variance: $V(X) = \beta^{2/\alpha}[\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)]$

Example: Weibull(5, 3)

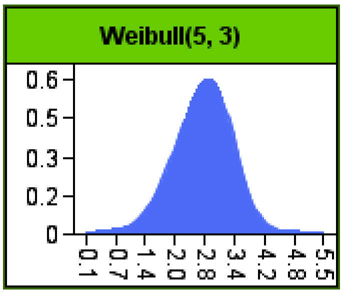


Figure 220 Weibull(5,3) Distribution Example

22.2 Discrete distributions

22.2.1 Binomial distribution

Probability function: $p(X = x) = \binom{N}{x} p^x (1-p)^{n-x}$

Domain: $x = 0, 1, \ldots, n$

Parameter domain(s): $n > 0$ where n is the number of trials. $0 < p < 1$ where p is the probability of success.

Mean: $E(X) = np$

Variance: $V(X) = np(1-p)$

Example: Binomial(0.2, 10)

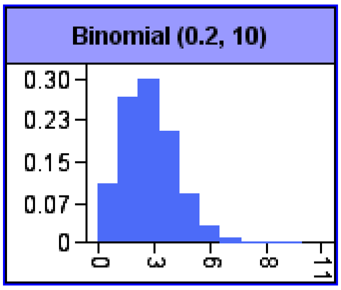


Figure 221 Binomial(0.2, 10) Distribution Example

22.2.2 Geometric distribution

Probability function: $p(X = x) = p(1-p)^{x-1}$

Domain: $x = 1, 2, 3, ...$ where x is the number of trials until the first success.

Parameter domain(s): $0 < p < 1$ where p is the probability of success.

Mean: $E(X) = \frac{1}{p}$

Variance: $V(X) = \frac{1-p}{p^2}$

Example: Geometric(0.2)

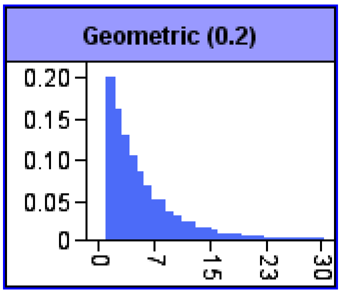


Figure 222 Geometric(0.2) Distribution Example

22.2.3 Hypergeometric distribution

Probability function: $p(X = x) = \frac{\binom{r}{x}\binom{m-r}{n-x}}{\binom{m}{n}}$

Domain: $x < m$ where x is the number of successful selections in a trial of size n from a total population of successes (type 1 observations), r , and failures, $m-r$, equal to a population of size m .

Parameter domain(s): $x < m$

Mean: $E(X) = \frac{nr}{m}$

Variance: $V(X) = n\frac{r(m-r)}{m^2}$

Example: Hypergeometric(100, 25, 10)

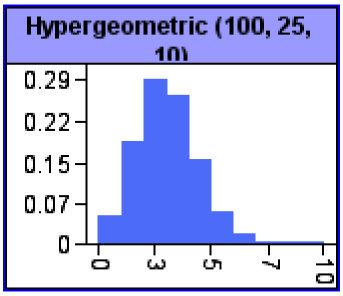


Figure 223 Hypergeometric(100, 25, 10) Distribution Example

22.2.4 Negative Binomial distribution

Probability function: $p(X = x) = \binom{x+r-1}{r-1} p^r (1-p)^x$

Domain: $x = 0,1,2,...$ where x is the number of successes in $(x+r-1)$ trials where r is the number of failed trials and p is the probability of failure.

Parameter domain(s): $r > 0$ and $0 < p < 1$

Mean: $E(X) = r\frac{(1-p)}{p}$

Variance: $V(X) = r\frac{(1-p)}{p^2}$

Example: Negative Binomial(10, 0.3)

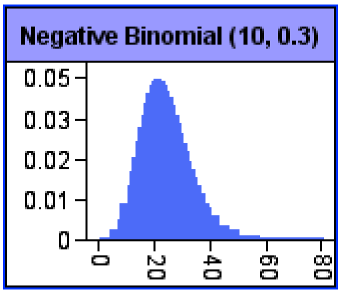


Figure 224 Negative Binomial(10,0.3) Distribution Example

22.2.5 Poisson distribution

Probability function: $p(X = x) = \frac{1}{x!}\lambda^x e^{-\lambda}$

Domain: $x = 0,1,2,3\ldots$

Parameter domain(s): $\lambda > 0$ where λ is the rate.

Mean: $E(X) = \lambda$

Variance: $V(X) = \lambda$

Example: Poisson(5)

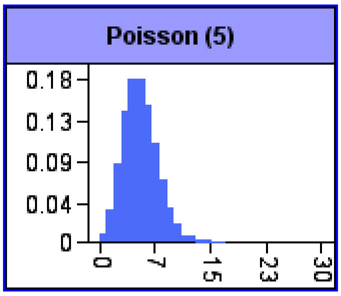


Figure 225 Poisson(5) Distribution Example

22.2.6 Uniform distribution

Probability function: $p(X = x) = \frac{1}{N}$

Domain: $x = 0,1,\ldots,N$

Parameter domain(s): $N > 0$

Mean: $E(X) = \frac{N+1}{2}$

Variance: $V(X) = \frac{(N-1)(N+1)}{12}$

Note: In Agena.ai modeller the required parameters are “Lower Bound”, *L*, and “Upper Bound”, *U*, and $N = (U - L) + 1$

Example: Uniform(0, 50)

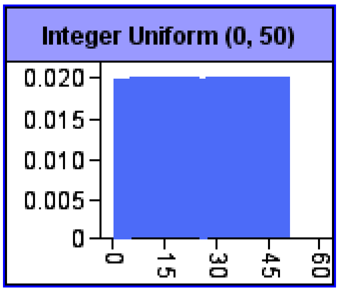


Figure 226 Uniform(0,50) Distribution Example

23. Expressions

This section deals with expressions that can be used for generating Node Probability Tables (NPTs) in Agena.ai modeller. Agena.ai modeller has a built-in set of supported mathematical constants and operators that can be used when defining expressions on nodes.

Agena.ai modeller expressions are case sensitive, so you have to enter them exactly as specified here otherwise you will get an error that may be hard to track. Thus, for example, typing in the expression Max(0, A+B) will result in an error since the correct syntax for the max function is max(0, A+B).

23.1 Continuous and Integer Nodes

Various constants and operators are available for use on Continuous and Integer Interval nodes.

The constants available for Continuous and Integer Interval nodes are listed in Table 12.

Table 12 Supported constants for Continuous and Integer Interval nodes

Constant	Syntax	Notes
Exponential	e	The base of the natural logarithm
Pi	pi	The constant π.

The operators available for Continuous and Integer Interval nodes are listed in Table 13. ***Unless otherwise stated the variable names (x, y, z) refer to either a number or a variable that corresponds to the name of a numeric parent variable***.

Table 13 Supported operators for Continuous and Integer Interval nodes

Operator	Syntax	Notes
Logarithm	log(x)	
Natural Logarithm	ln(x)	
Square Root	sqrt(x)	

Modulus	mod(x, y)	Returns the integer remainder after dividing x by y, so, e.g. mod(12,5)=2. This is then converted to a Uniform distribution accordingly.
Sum	sum(x, y, z,...)	This will return the sum of the supplied parameters.
Min	min(x, y, z,...)	This returns the minimum of the supplied parameters over a Uniform range.
Max	max(x, y, z,...)	This returns the maximum of the supplied parameters over a Uniform range.
Absolute	abs(x)	This returns the magnitude of the value of x without regard to its sign.
Addition	x + y	This is an alternative to sum(x, y).
Subtraction	x - y	
Multiplication	x * y	
Division	x / y	
Power	x ^ y	This raises x to the power of y.
Modulus	x % y	This is an alternative to mod(x, y)
Value	val(b)	Assume a Boolean variable parent named b. Returns 1 if b is “True” or “Yes” and 0 if b is “False” or “No”. This function can be used to convert Boolean nodes to integer 0, 1 values. The complement of Boolean expressions can be represented by using the val() function and prefixing the expression with “1 - “. For example, the complement of the Boolean value “b” would be “1 - val(b)”.
Sine	sin(x)	Where x is in radians. To use degrees you must convert to radians using pi/180.
SineH	sinh(x)	
Cosine	cos(x)	

CosineH	cosh(x)	
Tangent	tan(x)	
TangentH	tanh(x)	

Notes:

1. Multiple nesting is supported using () and [] parentheses.
2. When using a min, max or sum function take care to only do so when all variables and parameters are completely enclosed within the function. E.g. min(X, Y*Z) is find but min(X,Y)*Z is not. In this case create separate nodes, one for the min, max or sum function and another to process the rest of the expression you require.
3. Since Agena.ai modeller allows the use of two constants pi and e, it is best not to use either of these as unique names for nodes.
4. Arithmetic in Agena.ai modeller is performed using "interval arithmetic" which is executed using the Uniform distribution. Thus adding two intervals [0, 1] and [2, 3] will result in Uniform[2, 4].

23.2 Boolean, labelled and discrete real nodes

Table 14 shows the operators that are available for use on Boolean, Labelled and Discrete Real nodes.

Table 14 Supported operators for Boolean, Labelled and Discrete Real nodes

Operator	**Syntax**	**Notes**
IF THEN	If (<statement>, <value1>, <value2>)	If <statement> is true then the result is <value1> (which must be a valid state name for the node) else the result is <value2> (which must also be a valid state name for the node). For example, if the node is a Boolean node with states "True" and "False" and the node has numeric parents with names x and y, then the following comparative expressions can be used (where "==" means "equal to", "!=", means "not equal to", ">" means "greater than etc." if (x == y, "True", "False")

		if (x !=y, “True”, False”)
		if (x > y, “True”, False”)
		if (x >=y, “True”, False”)
		if (x >=y, “True”, False”)
		if (x >=y, “True”, False”)

IF	If (<statement>, <value1>)	Same as IF THEN except the result will only be set to <value1> if <statement> is true
AND	&&	For example, for a Boolean node with Boolean parents x and y: if(x ==”True” && y ==”True”, “True”, “False”) returns “True” if both x and y are “True”, otherwise returns “False”.
OR	||	For example, for a Boolean node with Boolean parents x and y: if(x ==”True” || y ==”True”, “True”, “False”)Returns “True” if either x or y are “True”, otherwise returns “False”.
XOR	xor	For example c for a Boolean node with Boolean parents x and y: if (xor(x == “True”, y == “True”), “True”, “False”) Returns “True” only if one of x and y are true, else returns “False”. Cannot be used on Labelled nodes.
M From N	mfromn (m, b1==”True”, b2==”True”,..,,bn==”True”)	This returns “True” if at least m of the following n statements are true and “False” otherwise. For example, for a Boolean node with 5 Boolean parents b1, b2, b3, b4, b5: mfromn(3, b1==”True”, b2==”True”, b3==”True”, b4==”True”, b5== “True”)

		returns “True” if at least 3 of the 5 parents are “True” and “False” otherwise. This is subject to constraint that m < number of arguments following m. Note that if alternative state names for a Boolean node are used (e.g. “Yes”/”No”, “Positive”/”Negative”, “On”/”Off”) then simply replace “True” with the corresponding ‘positive’ state name
NoisyOR	noisyor(x1, v1, x2, v2, …xn, vn, leak) where x1, x2, …,xn are names of Boolean parents and v1, v2, …,vn are the ‘weights’ (between 0 and 1) we associate respectively with each parent. The value ‘leak’ is a number between 0 and 1	For example Noisyor(x1, 0.7, x2, 0.4, x3, 0.1. 0.2) The NoisyOR function operates on Boolean type nodes by exploiting an assumption of independence between the effect of the parent nodes on the child node Y. So in the above example because x1 has the highest weight, Y is most likely to be true if x1 is true. The leak parameter models the “noise” in the child node when all of the parent nodes are false. So, if all the xi are false then the probability Y is true is equal to leak. Note that if all the weights are 1 and leak is 0 then noisyor is equivalent to the standard OR function Cannot be used on Labelled nodes.
NoisyAND	noisyand(x1, v1, x2, v2, …xn, vn, leak)	The NoisyAND function is simply the complement of the NoisyOR. Note that if all the weights are 1 and leak is 0 then noisyand is equivalent to the standard AND function Cannot be used on Labelled nodes.

Notes:

1. Multiple nesting is supported through the use of () and [] parentheses.
2. Literal state values for nodes are not checked at runtime so be careful to ensure that the values used exactly match the state values assigned to each node involved in the expression. So, if you have an expression with A == “true” where A is a Boolean node with default state values, this will be accepted at runtime despite the fact that the actual state value should be “True”.

3. Note that the equivalence operator can be declared at runtime directly on node names, i.e. A == B, but this will throw an error during calculation since it contains no literal values (E.g. == “False”); instead use the Equals operator as described below.

4. Comparative expressions which check for values within a fixed range have to be handled using two or more conditions. For example, to handle a range condition:

   15< parent < 20

   we would use the following expression:

   if(15< parent && parent < 20, “True”, “False”)

23.3 Ranked nodes

Table 15 shows the operators that are available for use on Ranked nodes.

Table 15: Ranked node operators

Operator	**Syntax**	**Notes**
Weighted Mean	wmean(wx, x, wy, y,...) (see notes below)	See Chapter 8 of Fenton & Neil's book for full explanation and examples. This returns the weighted mean of the values x, y, ... The weight for each value (e.g. wx) precedes the value that it weights. Formally this is the function $\frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$ where $w_i \geq 0$
Weighted Min	wmin(wx, x, wy, y,...) (see notes below)	See Chapter 8 of Fenton & Neil's book for full explanation and examples. This returns the weighted minimum of the values x, y, ... The weight for each value (e.g. wx) precedes the value that it weights. Formally this is the function: $\underset{i=1..n}{MIN} \left[\frac{w_i X_i + \sum_{i \neq j}^{n} X_j}{w_i + (n-1)} \right]$ where $w_i \geq 0$ and n is the number of parent nodes This function can be viewed as a generalised version of the MIN function. In fact, if all of the weights w_i are large then the WEIGHTED-MIN is close to the normal MIN. At the other extreme, **if all the weights w_i = 1 then f(X) is simply the average of the Xi's**. Mixing the magnitude of the weights gives a result between a MIN and an AVERAGE. The rationale for the WEIGHTED_MIN function is that some functions have only approximated a minimum of the parent values and have been some way between a min and an average. The distributions therefore contain a "tail" influenced by the presence of lower than higher values of the parents than returned by the max function.

Weighted Max	wmax(wx, x, wy, y,...) (see notes below)	See Chapter 8 of Fenton & Neil's book for full explanation and examples. This returns the weighted maximum of the values x, y, ... The weight for each value (e.g. wx) precedes the value that it weights. Formally this is the function $\underset{\forall i=1..n}{MAX}\left[\frac{w_i X_i + \sum_{i \neq j}^{n} X_j}{w_i + (n-1)}\right] \text{where } w_i \geq 0$ **This is analogous to the Weighted Min with regards to weights.**
Min-Max Mixture	mixminmax(wmin, wmax, x, y, z,...) (see notes below)	See Chapter 8 of Fenton & Neil's book for full explanation and examples. This returns the min-max mixture of the values x, y, z.... The weight assigned to the minimum (wmin) and the weight assigned to the maximum (wmax) are the first two parameters. Formally the function is: $\frac{w_{\min} MIN(X,Y,Z,...) + w_{\max} MAX(X,Y,Z,...)}{w_{\min} + w_{\max}}$

Notes:

1. In Agena.ai modeller, you can specify the parameters for the Ranked node operators via a dialog instead of typing them in by hand.

2. The last four operators can be used to supply the mean of a theoretical truncated normal distribution for Ranked nodes. In such cases, you must also supply the variance. The mean and the variance are then plugged into a TNormal function to generate the NPT of the node in question. This process is made easier by allowing you to supply the variance on the same dialog described in Note 1 above.

3. To use the above pre-defined operator expressions for Ranked nodes, all parents should normally be Ranked nodes; it is possible to use the expressions with numeric node parents but you will need to be sure that the it is possible to use the expressions with numeric node parents but you will need to be sure that the range of the resulting expression lies in the interval 0-1 (or errors will result).

4. Also note that for very small variance values the number of samples used should be increased. This setting can be found in the Application Properties dialog.

5. Ranked node operators must generate a mean value in the range [0 – 1], inclusive. It is left up to you to check whether their expression maintains this requirement.

23.4 Reserved keywords

Table 16 shows various keywords that may not be used as a unique identifier of a node. Note that these are case sensitive and you may use identifiers with a different capitalisation if required.

	Keywords	
A – E	**F – P**	**R – Z**
abs acos acosh arg asin asinh atan atan2 atanh avg Beta binom Binomial ceil ChiSquared cmod complex conj cos cosec cosh cot e exp Exponential ExtremeValue	false floor Gamma Geometric Hypergeometric i if im intcont internal_int_contDiscretePoint lg LIST ln log Logistic LogNormal max mfromn min mixminmax mod NegativeBinomial noisyand noisyor Normal pi Poisson polar pow	rand re rint round sec signum sin sinh sqrt str Student sum tan tanh TNormal Triangle true UDivide UMinus Uniform UPlus val vsum Weibull wmax wmean wmin xor

Table 16: Reserved Keywords

24. Keyboard shortcuts and common mouse actions

24.1 Keyboard Shortcuts

Function	Key Sequence
Open	CTRL+O
Save As	CTRL+S
Select All	CTRL+A
Copy	CTRL+C
Cut	CTRL+X
Paste	CTRL+V
Help	F1
Agena.ai modeller Tutorials	CTRL+F3
Agena.ai modeller Example Models	CTRL+F4
Import	CTRL+I
Create New Model	CTRL+N
Create New Model From Default	CTRL+D
Run Calculation	CTRL+R
Properties	CTRL+ENTER
Hide/Unhide hidden Node	CTRL+H
Hide/Unhide hidden Table Entry	CTRL+Q
Display Charts for Selected Nodes/Table Entry	CTRL+G
Expand/Retract Selected Table Entry	CTRL+E
Delete	Delete
Close Model	CTRL+W
Exit	CTRL+SHIFT+W

Table 17 Keyboard shortcuts

Notes:

1. If you are using CTRL + V to paste nodes from one Bayesian network into another, you need to ensure that you click in the target Graph view before pasting.

24.2 Common mouse actions

Function	Mouse Action
Open chart on Graph view	Double click
Access node properties dialog	Right click and choose Properties
Enter data on node	Right click and choose Enter Data
Expand/Contract Table entry	Double click on Table entry
Access temporary graph properties	Right click on chart and choose Chart Properties
Access permanent graph properties	Right click on node and choose Properties
Place all Charts on Graph view	This involves a three stage process: 1. Ctrl + A to Select All Nodes 2. Ctrl + G to Display Charts (places them on the chart panel) 3. Dock all Graphs to Graph view by pressing the 📊 button

Table 18 Common mouse actions

Last updated: 2024.10.10