GeNIe Bayes Net Model for

The GMU FAA Laser Strike Study 2021

# Definitions

A ***Bayes Network*** is[[1]](#footnote-1):

* A formal language for representing knowledge about uncertain quantities
  + nodes represent hypotheses
  + arcs represent direct dependency relationships among hypotheses
  + conditional probabilities encode strengths of dependencies
* A computational architecture for computing impact of evidence on beliefs
  + updates beliefs when new evidence is observed
  + exploits independence assumptions to make computation more efficient
* Representation language for probabilistic knowledge base
  + Domain ontology
  + Fragments of Bayesian networks
  + Composition operators for “gluing together” fragments
* Inference engine for probabilistic knowledge-based system
  + Retrieve appropriate fragments from knowledge base and “glue together”
  + Apply inference algorithm to draw appropriate inferences

***GeNIe*** Modeler[[2]](#footnote-2) is a “graphical user interface (GUI)”…that “allows for interactive model building and learning” and is used to create and model Bayes Networks.

# Preparation

Website to download GeNIe Software: <https://www.bayesfusion.com/downloads/>

* The website also contains quite a bit of introductory materials, documentation, and demos
* GeNIe doesn’t work well with Mac/Linux/Unix Operating Systems
  + Recommendation to do this in a windows Virtual Machine/Virtual Box instead (it is much easier and will reduce the frustration!)

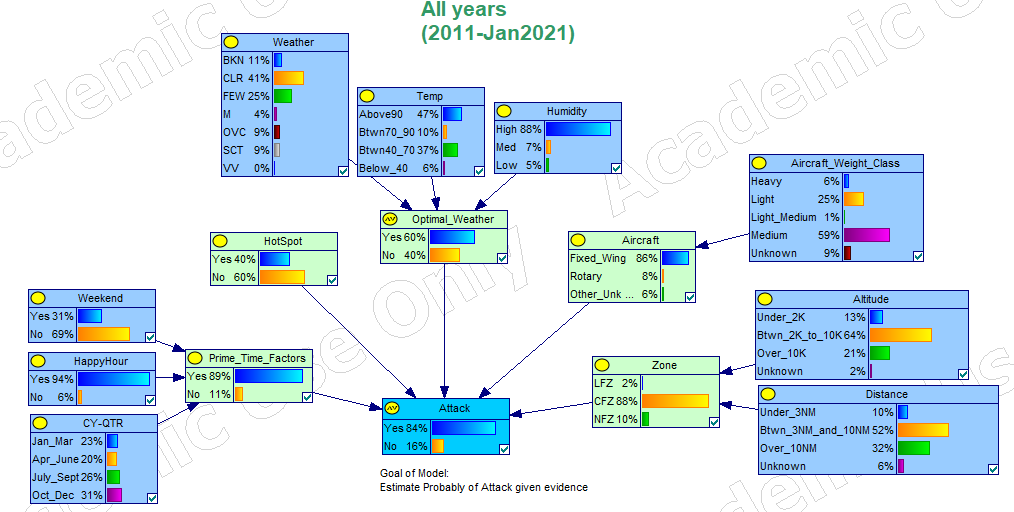
Some useful demos/tutorials of how to use GeNIe

* GeNIe Demo/Tutorial (20 min video) - <https://www.youtube.com/watch?v=aW3gxE6XB9E>
  + Note: some of the icons in the tutorial may look different from the current GeNIe tool; however most of the functionality is the same
* GeNIe Walkthrough - <https://support.bayesfusion.com/docs/GeNIe/>

# FAA Models

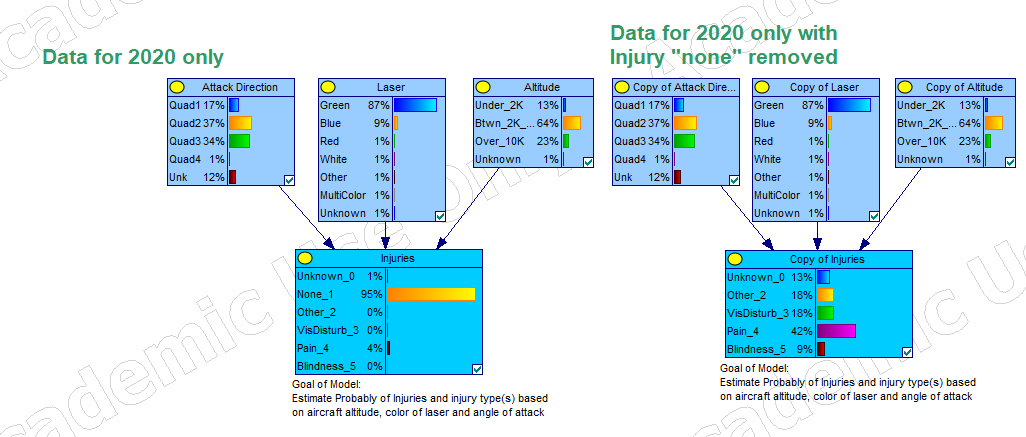
Two models were created for the GMU FAA Laser Strike Study – one focused on estimating the probability of attack and one estimating probability of injury type, each based on various factors reported by aircrew related to laser attacks:

* FAA\_Model\_Attack
  + 15 node Bayes Net model
    - 14 causal nodes and 1 outcome node
  + 2 versions created: one using data from 2011-2021, the other using 2020 data only



***Figure 1 – FAA Attack Model using data from 2011- January 2021***

* FAA\_Model\_Injuries
  + 4 node Bayes Net model
    - 3 causal nodes and 1 outcome node
  + 2 versions created: one using data from 2011-2021, the other using 2020 data only
  + Additionally, a third version was created that removed the outcome of “None” in the injury type. This was because 95% of attacks had an outcome of no injury reported and by removing this value from the model, we were better able to see the differences in probabilities across the other outcomes. In other words, when using this model, the user should keep in mind, there is still only a 5% probability of an injury and the outcomes estimated are within that 5%.



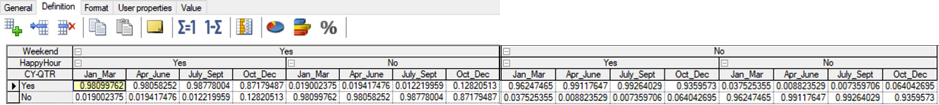
***Figure 2 – FAA Injury Model using data from Calendar Year 2020 only; with all injury types (left) and with “None” removed (right)***

## Nodes and Node Types

Each model contains “causal” nodes (variables that contribute to the probability of an event happening), some of these nodes feed into other causal nodes. This was done to minimize the number of variables feeding into the “Outcome” nodes and simplify the conditional probability tables (CPT). **Conditional probability** is the probability of one event occurring with some relationship to one or more other events.[[3]](#footnote-3)

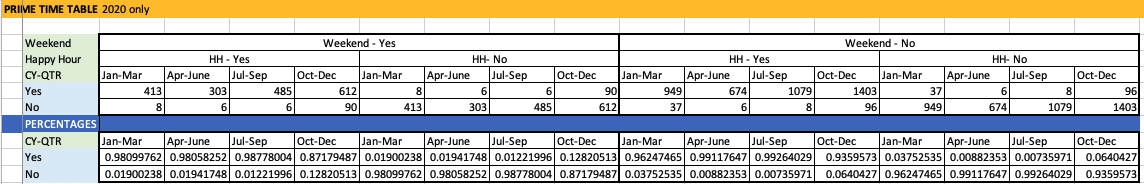
The CPTs in the models are based on the data from the GMU FAA Study. Models can later be updated using current data as variables are expected to change over time and may change the probability estimates.

Example CPT for the “Prime Time” node:



***Figure 3 – Prime Time Node Conditional Probability Table***

To make updating the models easier, some of the CPTs are provided in a separate excel file. The tables in the excel file have formulas to figure out the percentages, which can then be copied and pasted into the GeNIe CPTs. Prime Time CPT in excel:



***Figure 4 – FAA Prime Time Node Conditional Probability Table***

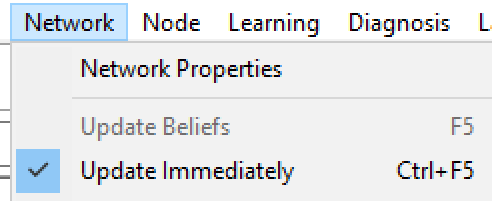
There is no right or wrong way to update the GeNIe models, it is a matter of preference. The team found that sometimes excel was easier to manipulate than the GeNIe tool.

## Opening a Model File

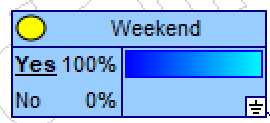
A model file can be opened two ways:

* Double-click on the file name from the explorer window (or from the file icon if you have it on your desktop)
* Open up the GeNIe tool, go to the file drop down menu and choose “open network”, and then navigate to where you have the file stored that you want to open. Select the file name and click “open”

## Using the Models to Estimate Probability

When a model is first opened, the probabilities presented are the values from the CPT tables, based on the data from the GMU study. If no values are shown for any of the items in the model, navigate to the “Network” dropdown menu at the top of the window and make sure “Update Immediately” is selected. This will ensure as values for conditional variables in the model are selected, the probabilities will immediately reflect the changes.

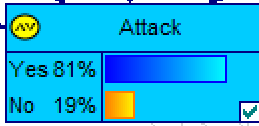
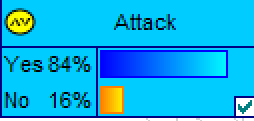
***Figure 5 – Network Menu in GeNIe***

Selecting values for variables in a model is also referred as entering “evidence” (i.e., this is something observed or known). To enter evidence, double-click on the desired value, which will change the probability of that value in that node to 100%. As more evidence is entered, you should see the probabilities change in nodes that the evidence directly relates to, and in turn any variables those nodes feed into (i.e., “causal chains”, think of a domino effect).

***Figure 6 – Weekend node with evidence selected***

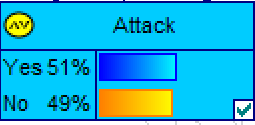
As more evidence is entered, the probabilities in the Outcome node should automatically update to reflect the estimated probability based on those variables.

For example, here are the probabilities of an attack before entering any evidence (note, yours may look different, as this screenshot was not on the final version of the model) and after entering evidence for Weekend (no), Happy Hour (yes), CY-QTR (Oct-Dec), HotSpot (yes), Optimal\_Weather (yes), Zone (CFZ), and Aircraft (Fixed Wing), the estimated probability of an attack changes to:



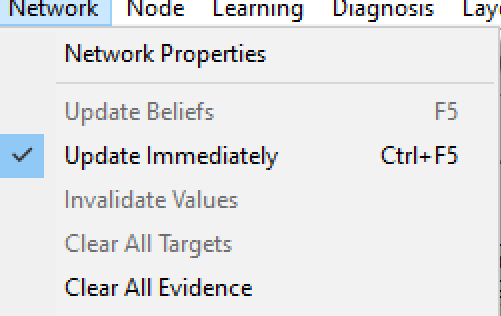
***Figure 7 – Probability of an attack before (left) and after (right) evidence is entered into the model***

If some values are changed (Aircraft – Rotary, Zone – LFZ), the probability of an attack changes to:



***Figure 8– Probability of an attack with evidence changed***

To clear evidence:

* Evidence in a node – right click on the node and select “clear evidence”
* All evidence in the model - navigate to the “Network” dropdown menu and select “Clear All Evidence”
  + This is basically resetting the model back the starting position.

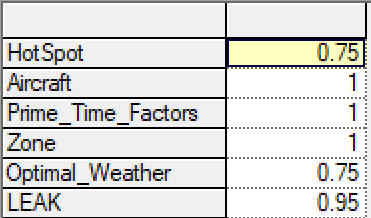
***Figure 9 – Network Menu – Clear all Evidence***

## Weighting Factors and Leak

Some of the nodes contain “weighting factors”, which attempt to consider the importance of different variables on the outcome.

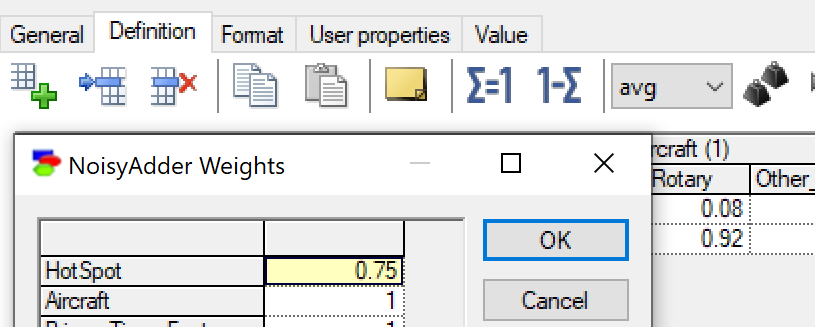
For example, in the attack model, it was estimated/assumed that aircraft (type, class), time, and zone (altitude + distance from an airport/airfield) were more important than HotSpot and Weather in estimating the probability of an attack. Additionally, the model should factor in error.

These were captured as “weighting factors” in the Outcome node:



***Figure 10– Attack model weighting factors***

In this example Aircraft, Prime\_Time\_Factors, and Zone are fully weighted (100%), HotSpot and Weather are weighted at 75%, and Leak shows that the model allows for 5% error.

To change/update these values:

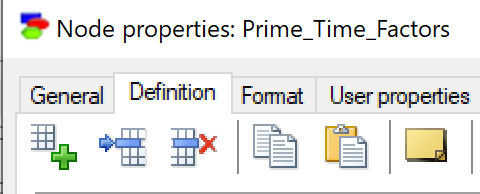
* Select the node by double clicking on it
* In the pop-up box, select the “Definition” tab
* Select the “weight” icon
* Enter the changes manually
* Click on “OK”

***Figure 11– Changing weighting factors***

Note, weighting factors are only available on the “Noisy Adder” nodes.

## Adding/Deleting variables in Nodes

Node variables should be kept to as few as possible, and binary nodes (e.g., True/False, Yes/No, Present/Not Present) are ideal because they result in simpler CPT tables (i.e., simpler calculations, easier to work with).

Should a new value be needed for one of the existing nodes, or a value is no longer needed, simply double click on the node, select the “Definition” tab.

* To add a variable, either click on the icon with a + sign or the icon to add a row
* To delete a variable, select the variable in the table and, click on the icon with a big red x

Delete

Add

***Figure 12– Adding variables to a node***

If values are added/deleted from a node, probability across the values and corresponding CPTs will need to be updated. [All probabilities for each column must sum to 1]

## Adding/Deleting Nodes

To delete a node that is no longer needed, select the node and press the “delete” key on your keyboard. A confirmation pop-up box will ask you to confirm, select “Yes” (note: this cannot be “undone”, if a node is inadvertently deleted, it will have to manually be created, “ctrl z” won’t bring it back).

To add a new node to the model,

* Select one of the node type icons (“Chance”, Deterministic, Noisy Max or Equation) in the menu (node type can be changed later, so don’t worry about selecting the wrong type)

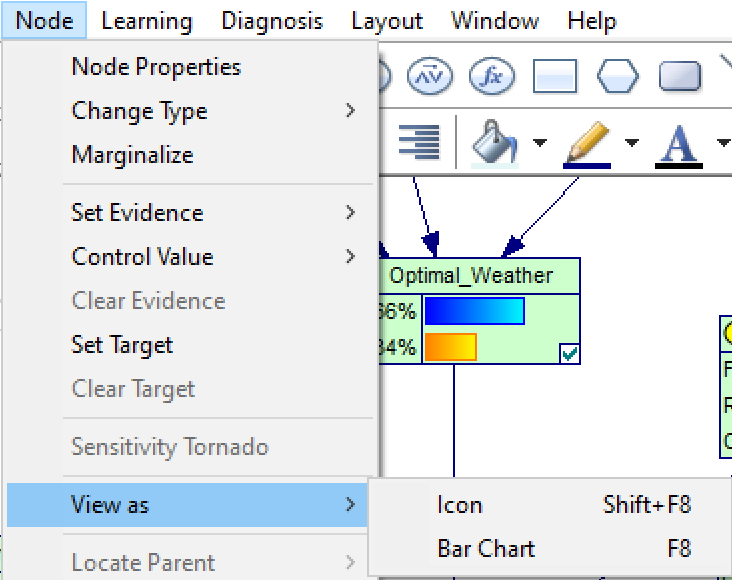
***Figure 13– Node icons***

* Click anywhere in the model window (note, while the node type is still selected, every time you click in the model window, a new node will be added; to unselect this, click on the arrow icon (“select objects”))

***Figure 14– “select objects” icon***

* The new node will come up as a default oval shape with “Node” and a node number as the node name
  + Node name can be changed by typing over the node name and entering the new name or later in the “Definition” tab of the node pop-up box. You can also change it by double clicking on the node and changing it via the “General” tab in the pop-up Node properties box.

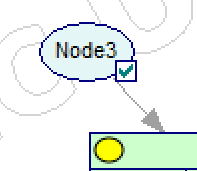
***Figure 15– New Node at creation***

* + Node shape can be changed by navigating to the “Node” dropdown menu and then to “View as”, and select “Bar Chart”
    - From this menu, Node type can also be changed via the “Change Type” sub-menu
  + To add values to the node, double click on the node to select it and then click on the “Definition” Tab.
  + Outcomes (or hypotheses) start out as “State0” and “State1”, with probability evenly distributed across the two (50/50).

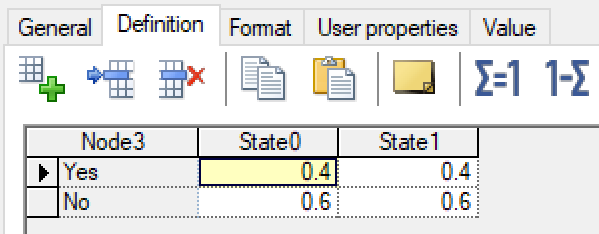
***Figure 16– Node menu***

* + - Outcome names and values can be changed by highlighting the name or value and manually typing in the new name/value
  + To associate it to another node, select the “Arc” icon in the menu and then “draw an arrow” from one node to another by left clicking on the node where you want the arrow to start and dragging it (still left clicked) to the node you want the arrow to end.

***Figure 17– Node Arc icon***

* + - A CPT table will automatically be created in the node the arrow points to. It will use the values from node and distribute evenly across the states (e.g., if the node has 2 values at 0.4 and 0.6, the CPT will show both states of the new node result in 0.4 and 0.6 in the outcome node).

## Updating CPT values

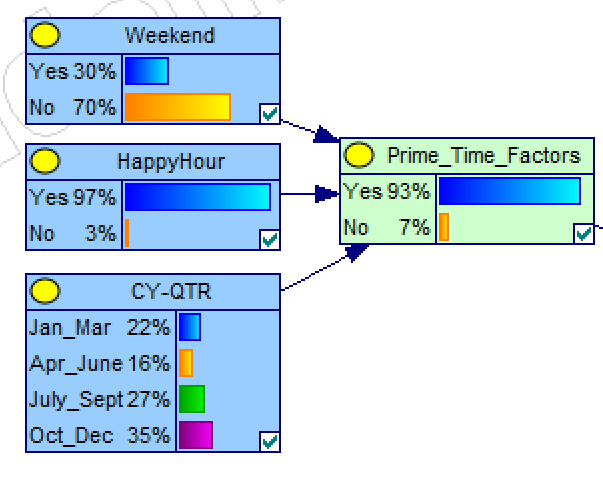
To change/update these values:

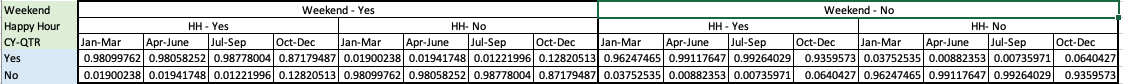
* Select the node by double clicking on it
* In the pop-up box, select the “Definition” tab
* Select the value you wish to update and manually change it

***Figure 18– Updating Conditional Probability Tables***

* All columns must sum to 1, if they do not, you will see an error flag at the top of the column; once you update one value, you can use the “” function to fill in the remaining value by selecting the remaining value in the column and then clicking on “”. It will automatically change the remaining value for you to sum the 2 values to 1.
* Note – because all outcomes in CPTs must sum to 1, in the excel files provided for the model, some values in the injury model were entered as a work-around when all values in the column were 0. These are highlighted yellow in the excel file and were all injury type “Unknown”.

CPTs with multiple nodes feeding in, can become large and complicated. Here is a simple walk-through using the Prime\_Time\_Factors node from the Attack model:

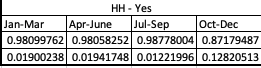




***Figure 19– Prime Time Node, Causal/Child nodes, and Conditional Probability Table***

The CPT table shows all possible outcomes for the Prime\_Time\_Factors node based on all possible outcomes of the conditions (causal nodes). While the table initially looks dauting, it can really be broken down to 4 smaller tables:

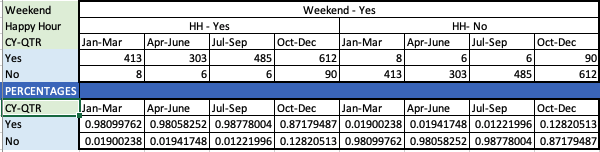
* Weekend Yes
  + Happy Hour Yes
  + Happy Hour No
* Weekend No
  + Happy Hour Yes
  + Happy Hour No



And from there, each “Happy Hour” table is broken down to the probability of an attack in each calendar-year quarter. Each column in the table must sum to 1.

***Figure 20– sub-table from the CPT***

The probability values in this table were determined based on actual number of attacks in each quarter to make the model more precise. We used pivot tables on excel files to pull the actual numbers and then formulas in excel to perform the calculations for the percentages, which were copied and pasted into GeNIe. If these values were not known, they could have been manually estimated and entered (e.g., .99/.01 or .95/.05, etc.) into the tool as described above based on best guess, subject matter expertise, etc.



***Figure 21– Excel file to calculate probabilities for Prime Time node, showing just data for strikes that occurred on a weekend (weekend-yes).***

## Summary

The model isn’t perfect. It uses the data that’s been entered and attempts to predict the probability of an event happening based on what’s known (evidence) and the data in the model. The interface is easy to use, so play around with it. Save a copy, change the numbers in the tables and see what future might hold.

For more information, there are a lot of sample models on the BayesFusion website. They have an informal, tutorial-like introduction in the [*Hello GeNIe!*](https://support.bayesfusion.com/docs/GeNIe/hello.html) section of the GeNIe Modeler manual that gives a nice walkthrough of the tool. They also have an introduction to  [Bayesian network](https://support.bayesfusion.com/docs/GeNIe/da_bns.html)s, that provides a nice introduction on the topic.

Enjoy!

## Questions?

Contact: Char Burrage, [cburrage@gmu.edu](mailto:cburrage@gmu.edu)

1. George Mason University (GMU) SYST 584 course – Paulo C. G. Costa, Ph.D. [↑](#footnote-ref-1)
2. BayesFusion website, <https://www.bayesfusion.com/genie/> , [accessed 24 April 2021] [↑](#footnote-ref-2)
3. Source: “Conditional Probability: Definitions & Examples”, Statistics How To website, <https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/conditional-probability-definition-examples/>, [Accessed 17 April 2021] [↑](#footnote-ref-3)