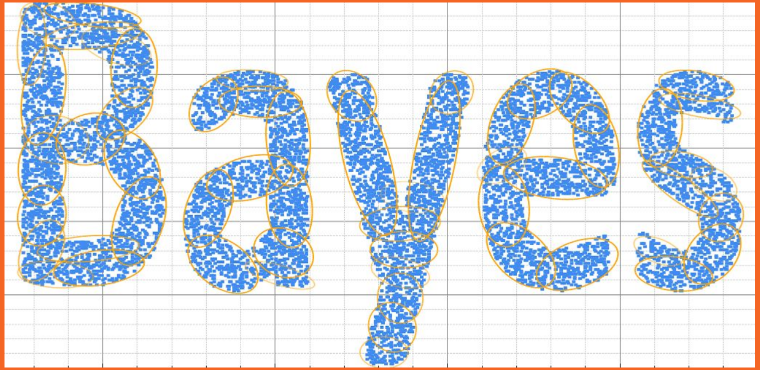
Bayesian Networks

CRNS
Centre for Robotics and Neural Systems

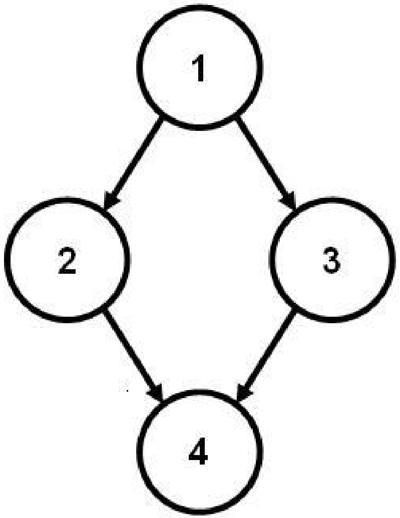
Massimiliano Patacchiola, PhD Student in Cognitive Robotics
Centre for Robotics and Neural Systems , Plymouth University

Introduction

What are Bayesian Networks?



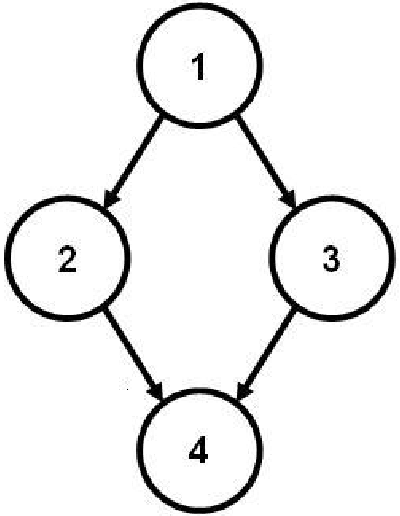
What is a Bayesian network?



"A Bayesian Network (BN) is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph." [Wikipedia]

It enormously simplified the representation of **probabilistic relationships** between random variables.

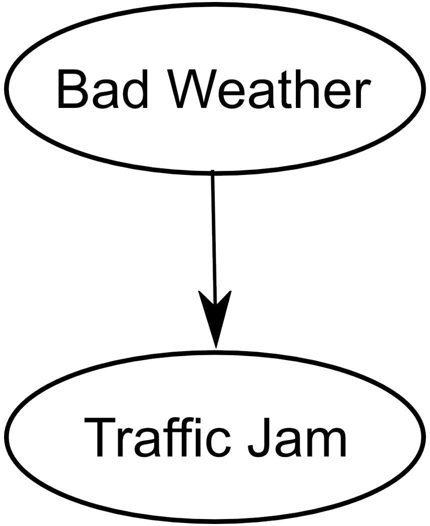
What is a Bayesian network?



The nodes represent **random variables.** The edges represent **causal relationships.**

Bad weather **cause** traffic jam (not the opposite).

Inference: given the presence of traffic jam it is possible to estimate the probability of bad weather

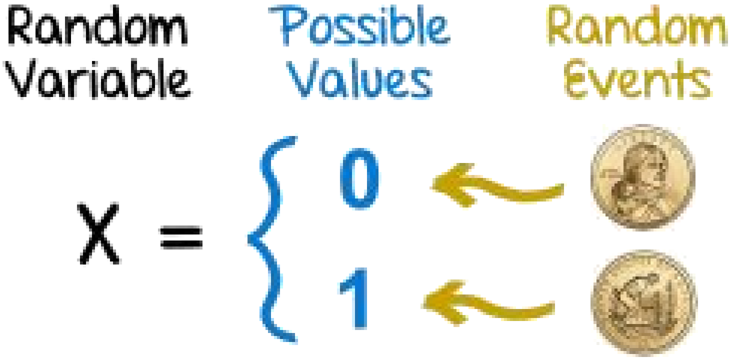


Probability Theory

The backbone of Bayesian Networks

Random Variable

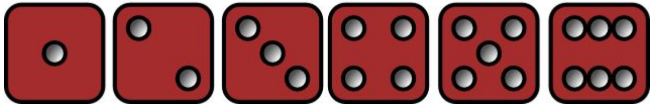
A random variable is defined as a function that maps probability to a physical outcome (labels). The possible outcome when tossing a **coin** are two, we can call these events [0, 1]. The probability of 0 is ½, the probability of 1 is ½.



Random Variable

What's the probability of obtaining 3 rolling a fair dice?

We have a **total of 6 possible outcomes**: [1, 2, 3, 4, 5, 6]

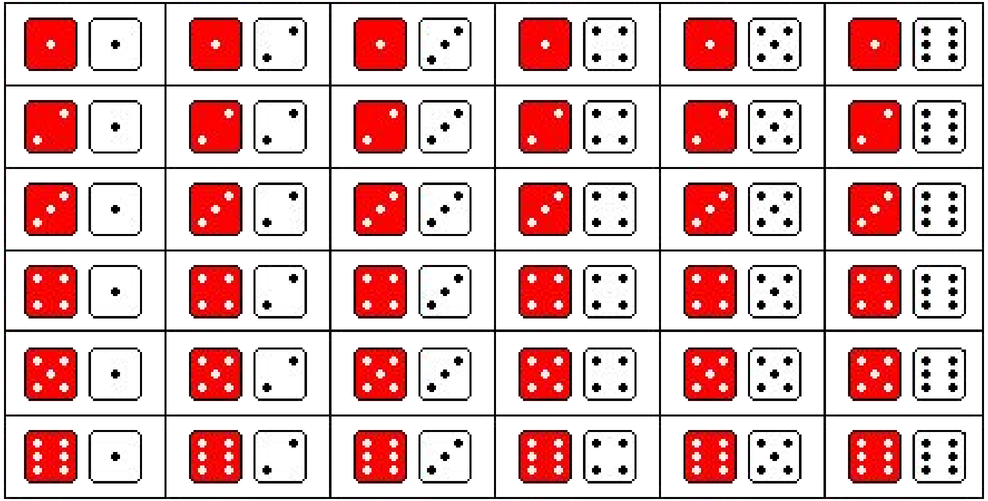


The possibility of having 3 is ⅙

If we call X our **random variable** representing all these outcomes, we can say that: $p_X(3) = 1/6$

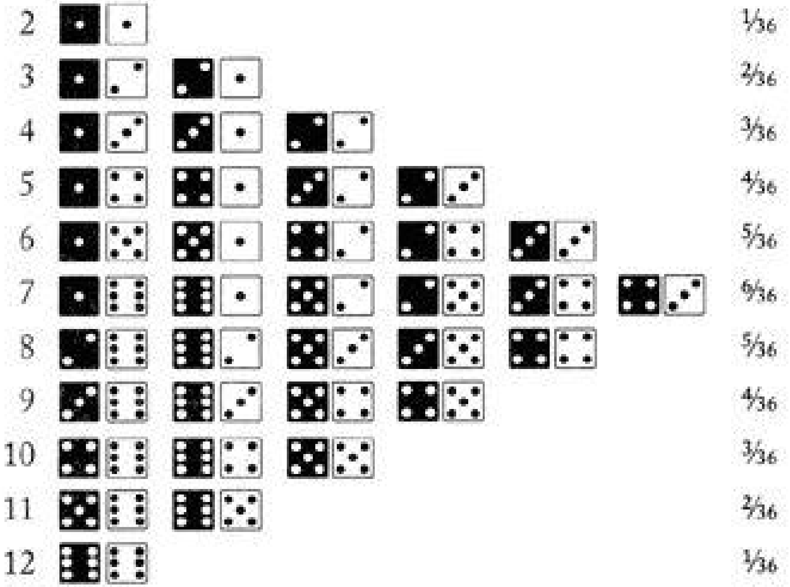
Random Variable

Rolling a pair of dices what's the probability of having 4 and 5? The probability is 2/36 because we have 36 possible outcomes when rolling two dices and the event we are looking for can happen in two way (4,5) and (5,4).



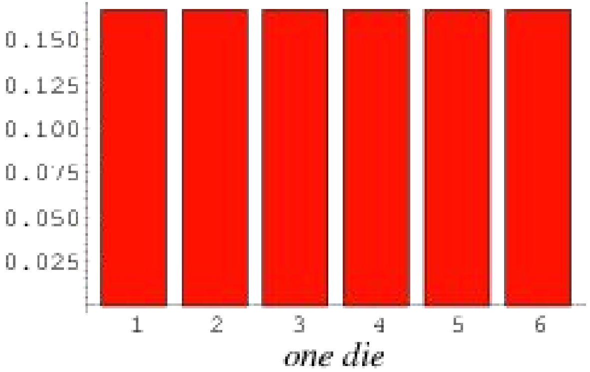
Random Variable

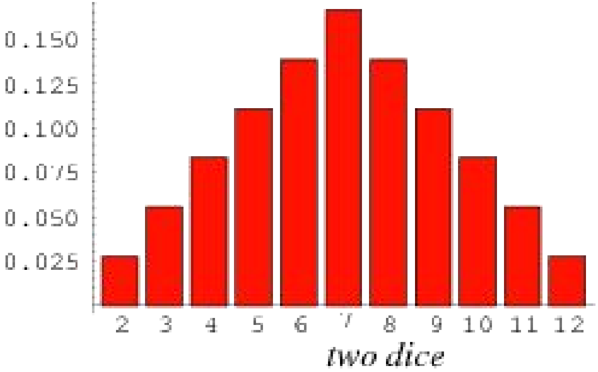
What's the probability that the sum of the dices is 4? The probability is **3**/36 because we can obtain that sum in **three** ways:



Probability Mass Function

Suppose that $X: S \to A$ is a discrete random variable defined on a sample space S. Then the **probability mass function** $f_X: A \to [0, 1]$ for X is defined as:





$$f_X(x) = \Pr(X = x) = \Pr(\{s \in S : X(s) = x\}).$$

$$\sum_{x \in A} f_X(x) = 1$$

Expected Value

Suppose random variable X can take value x_1 with probability p_1, value x_2 with probability p_2, and so on, up to value x_k with probability p_k. Then the expectation of this random variable X is defined as:

$$\mathrm{E}[X] = x_1p_1 + x_2p_2 + \cdots + x_kp_k$$

Let X represent the outcome of a roll of a fair six-sided die. The possible values for *X* are 1, 2, 3, 4, 5, and 6, all equally likely (each having the probability of 1/6). The expectation of *X* is

$$\mathrm{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Joint Probability

The probability that two (or more) events happen at the same time:

$$P(A, B) = P(A)P(B).$$

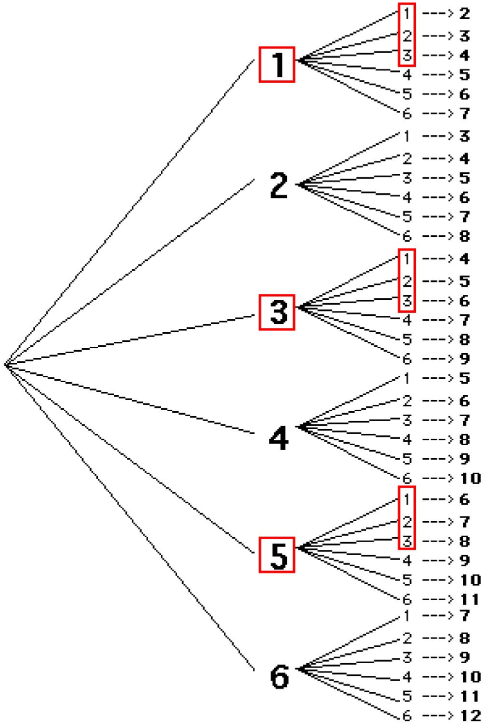
$$\mathrm{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathrm{P}(A_i).$$

Only if the events are independent.

Two events are **independent** if the occurrence of one does not affect the probability of occurrence of other.

Similarly, **two random variables** are **independent** if the realization of one does not affect the probability distribution of the other.

Joint Probability



Rolling a dice twice:

the **first value** is an odd number

AND

the **second value** is less than four

P(A) = 3/6

P(B) = 3/6

P(A,B) = P(A) * P(B) = **9/36**

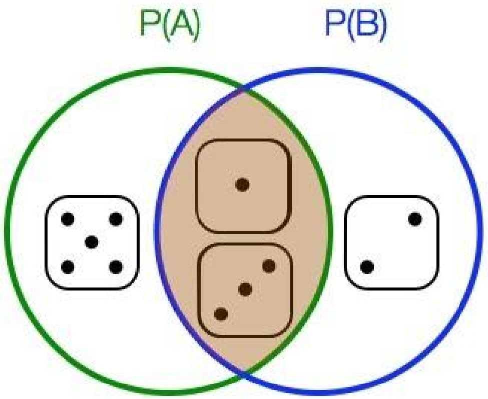
Conditional Probability

What is the Probability of

rolling a dice and it's value is less than 4

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

knowing that the value is an odd number



Bayes Theorem

The probability of an event, based on prior knowledge of conditions that might be related to the event:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)},$$

- **P(A)** and **P(B)** are the probabilities of observing **A** and **B** without regard to each other.
- **P(A | B)**, a conditional probability, is the probability of observing event **A** given that **B** is true.
- **P(B | A)** is the probability of observing event **B** given that **A** is true.

Bayes Theorem

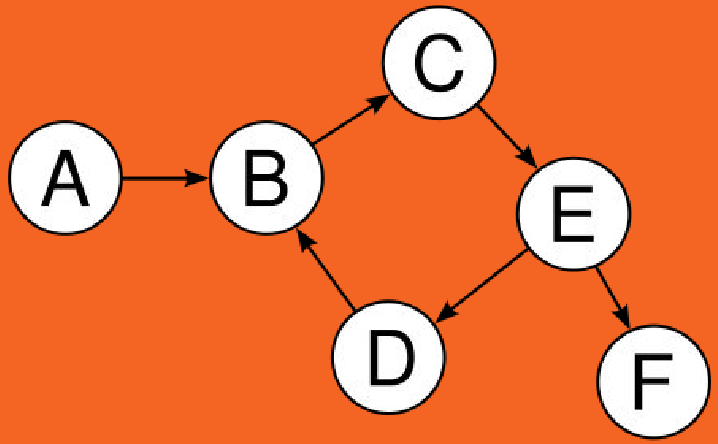
Relative size	Case B	Case B_	Total
Condition A	w	x	$w+x$
Condition Ā	y	z	$y+z$
Total	$w+y$	$x+z$	$w+x+y+z$

$$P(A|B) \times P(B) = \frac{w}{\cancel{w+y}} \times \frac{\cancel{w+y}}{w+x+y+z} = \frac{w}{w+x+y+z}$$

$$P(B|A) \times P(A) = \frac{w}{\cancel{w+x}} \times \frac{\cancel{w+x}}{w+x+y+z} = \frac{w}{w+x+y+z}$$

Bayesian Network Design

How to design a Bayesian Network?



Tools for Bayesian Networks

GeNIe	C++	Linux, Mac, Windows
JavaBayes	Java	Linux, Mac, Windows
Stan	R, Python, shell, MATLAB	Linux, Mac, Windows
Dlib	C++, Python	Linux, Mac, Windows

Tools for Bayesian Networks

BAYESFUSION, LLC
Data Analytics, Mathematical Modeling, Decision Support

GeNIe
Graphical interface for SMILE. Runs under Windows and (with Wine) on OSX and Linux.

genie_setup.exe
GeNIe Installer (32-bit)
Size: 10.3 MB, modified on Nov 4, 2016

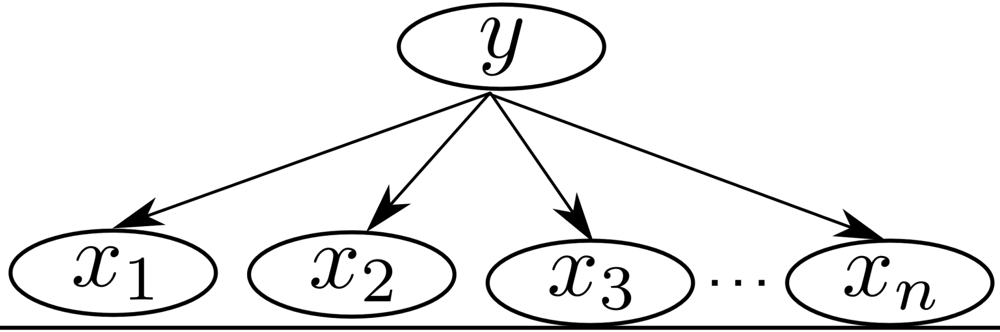
genie_setup_x64.exe
GeNIe Installer (64-bit)
Size: 11.0 MB, modified on Nov 4, 2016

Installing GeNIe: https://www.bayesfusion.com

Spam Classifier (Naive Bayes)

Why naive? We assume **independence** between each pair of **features**. This assumption could not be true, but for practical applications it works very well.

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y),$$



Spam Classifier (Naive Bayes)

Applying the Bayes theorem:

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) P(x_1, \dots x_n \mid y)}{P(x_1, \dots, x_n)}$$

Then applying the Naive assumption we get:

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \dots, x_n)}$$

The denominator is constant we can estimate the value of y as:

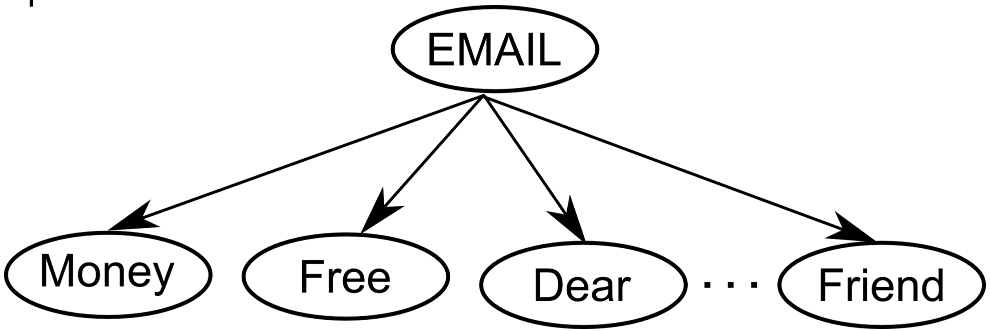
$$\hat{y} = \arg\max_y P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

Spam Classifier (Naive Bayes)

Feature Prior Distributions: using a corpus of emails classified as spam and ham estimating the frequency of each word. P(Money | Spam) = 0.8

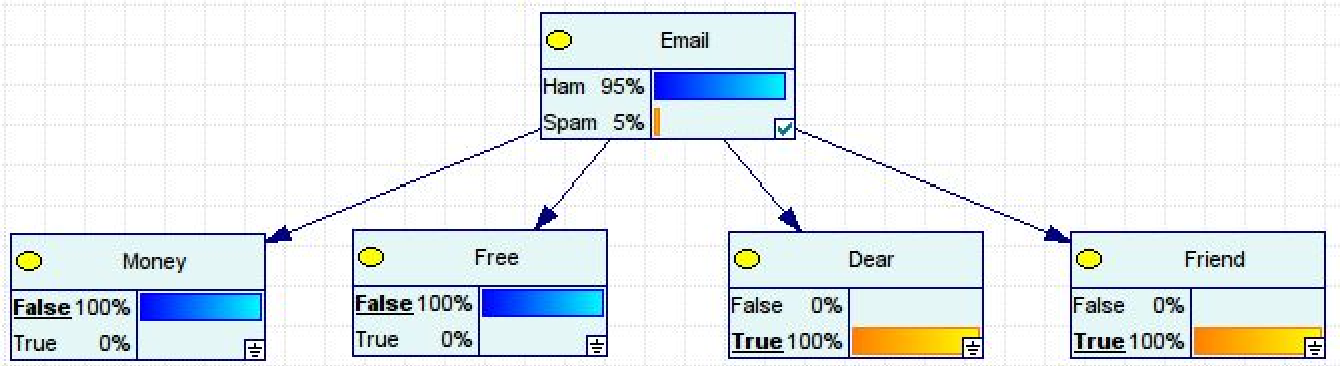
Parent Prior Distribution: we assume there is no a priori reason for any incoming message to be spam rather than ham (both cases have equal probability). P(Spam)=0.5; P(Ham)=0.5

Inference: Given the words present in the incoming email estimate if the email is spam or ham.



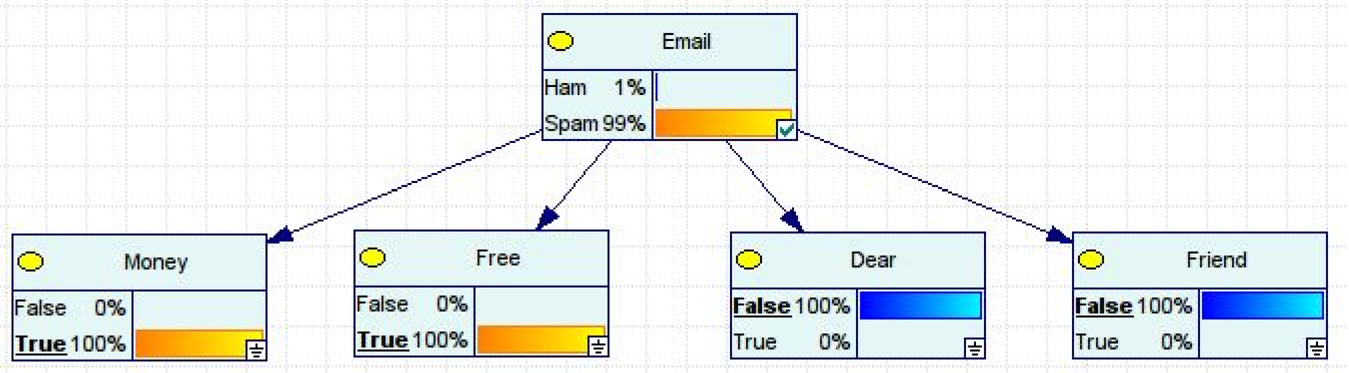
Spam Classifier (Naive Bayes)

“Dear Huda. Well, another year has passed. I just can't believe that it went so quickly. I've been studying for my exams and I finished high school. Now I start my studies in the College of Medicine and I really miss the times that we spent together. I am so sorry because I didn't send letters for you. Actually, I lost your address and I was lucky to find it again. And how are you? I hope that you and your family are all right. Did you join the Interior Design College as you always wished? I really want to know your latest news and know everything about your studies. I'm writing to tell you that I have been to Europe and it was great. Do you believe that I met our friend Yara in France? It was by accident. She has been there for five years. Well, that's all about me and we're all doing fine here. I hope to hear from you soon. Take care! Your best friend“



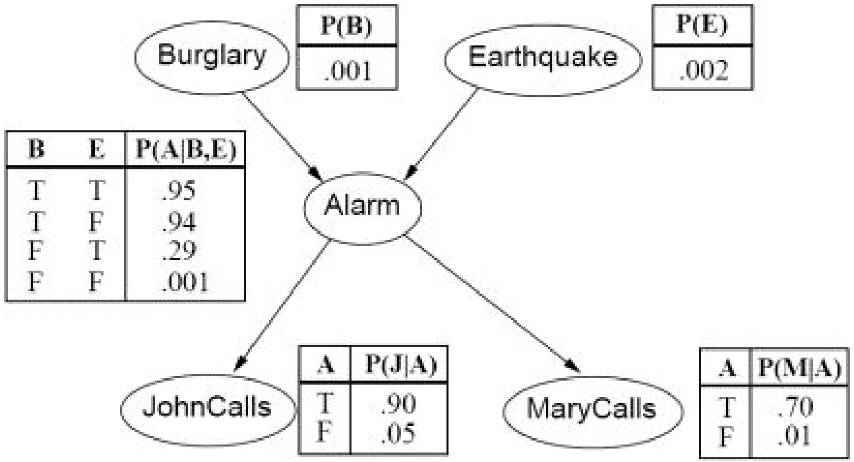
Spam Classifier (Naive Bayes)

“YES, GUARANTEED!!!! WE GUARANTEE THAT YOUR TEETH WILL BE BETWEEN 3 to 8 SHADES WHITER WITHIN THE NEXT 30 DAYS OR LESS OR YOUR MONEY BACK IN FULL. THIS IS HOW CONFIDENT WE ARE ABOUT THIS PRODUCT. NO ONE ELSE WILL OFFER YOU THIS. Take advantage as you have nothing to risk because of our money back guarantee. Buy for yourselves, your family and your friends. They'll thank you for this. Makes a wonderful Christmas gift! REMEMBER OUR SPECIAL: If you buy 2 kits, SHIPPING IS FREE. Buy 3 kits and get the 4th FREE! To order this professional-grade teeth whitening system for the superlow price of only $24.99, click on the link below to pay with a Visa or Mastercard.”



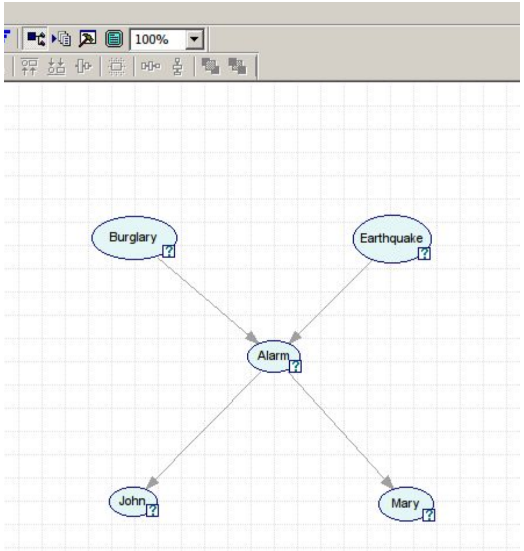
The alarm example

The alarm of my house can be activated by a burglary or by an earthquake.
When it is activated I can receive a call by John or by Mary.



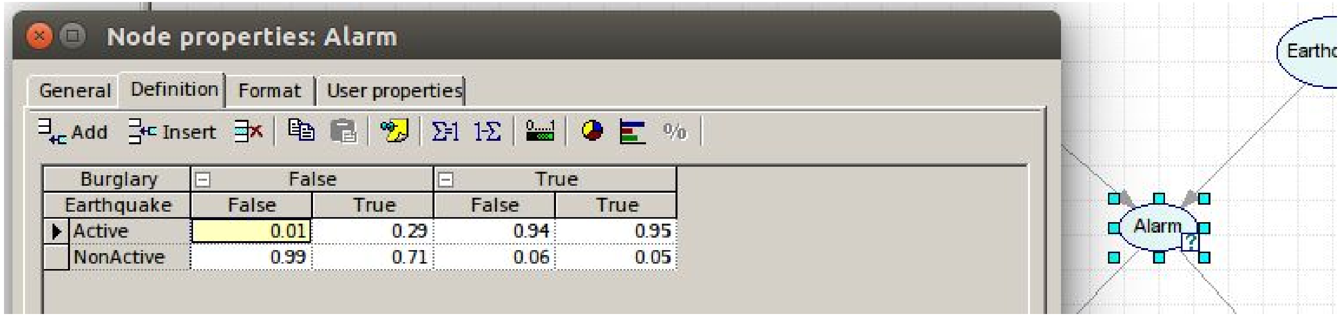
The alarm example

I will use GeNIe to replicate this example. First of all let's create the five nodes using the **yellow icon** in the toolbar.



The alarm example

Selecting a node it is possible to modify the probability distributions and the conditional probability tables. The **Alarm node** has a table with 8 entries, given by all the possible combinations of the parents.



Burglary	False		True	
Earthquake	False	True	False	True
Active	0.01	0.29	0.94	0.95
NonActive	0.99	0.71	0.06	0.05

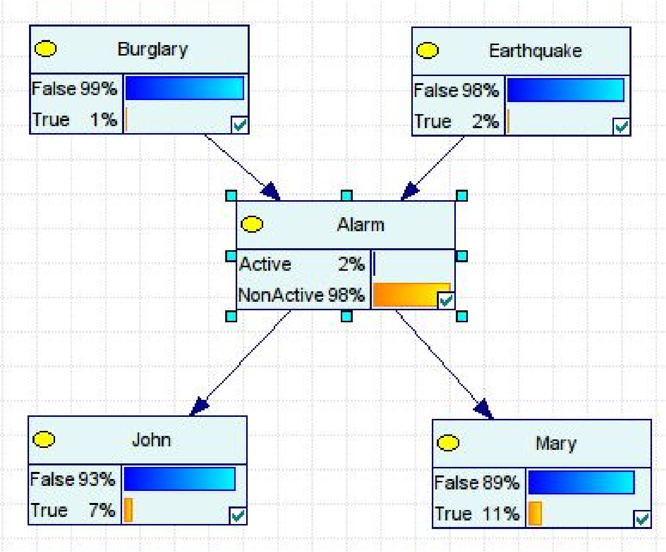
The alarm example

To enable bar **chart view**:

Right Click on Node > View As > Bar Chart

To estimate the **posterior distribution**:

press the update icon 

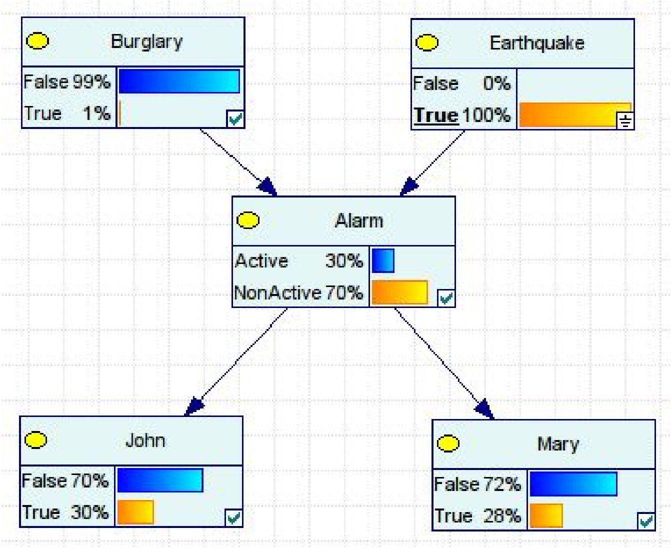


The alarm example

Let’s suppose we know there has been an **earthquake** while we were in the office. What’s the probability the alarm has been activated?

Right Click on Earthquake node > Set **evidence** > True

Press the update icon 



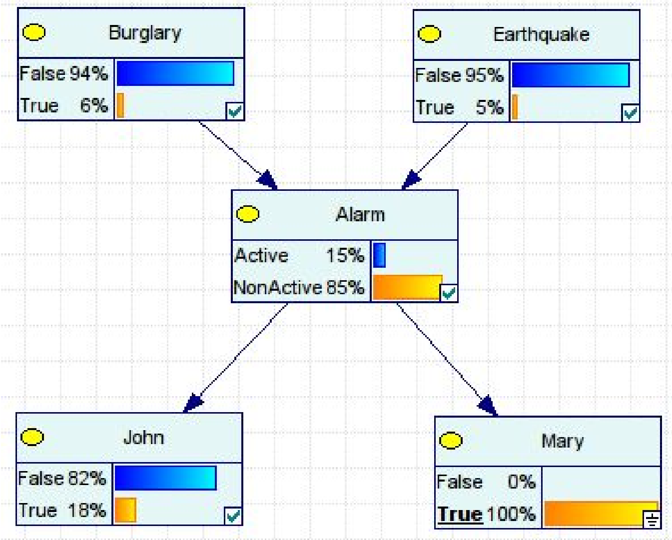
The alarm example

Let's suppose we receive a call from **Mary**. What's the probability the alarm is active?

Clear previous evidence

Right Click on Mary node > Set **evidence** > True

Press the update icon 



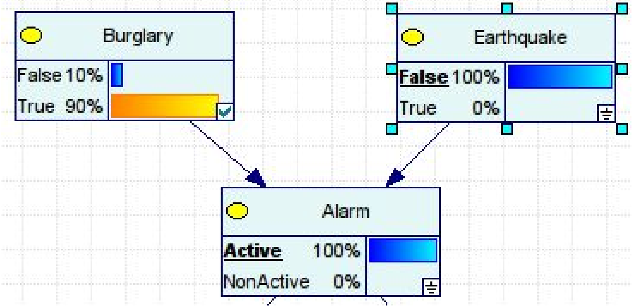
Inference (Bayes Theorem)

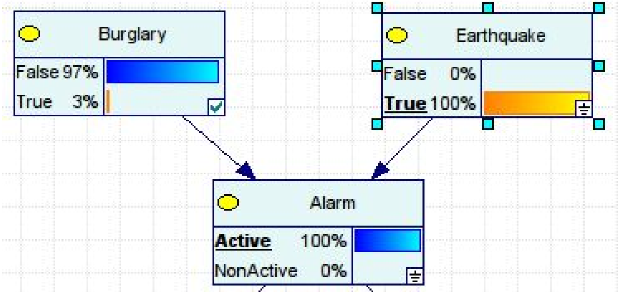
The alarm example

The parent nodes Burglary and Earthquake are **independent**.

The information that an earthquake occurred does not give us any hint about the occurrence of a burglary, but…

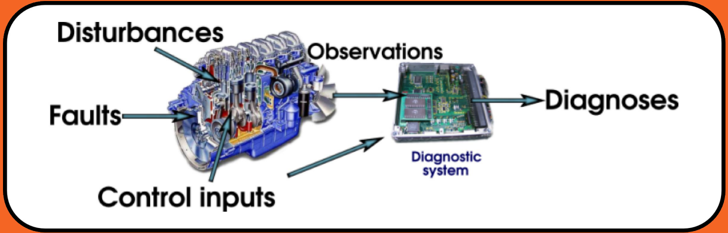
If the alarm is active (evidence) the two nodes become **dependent**. In this case the occurrence of an Earthquake change the probability of a burglary (and vice versa).





Case Studies

How are Bayesian Networks used?



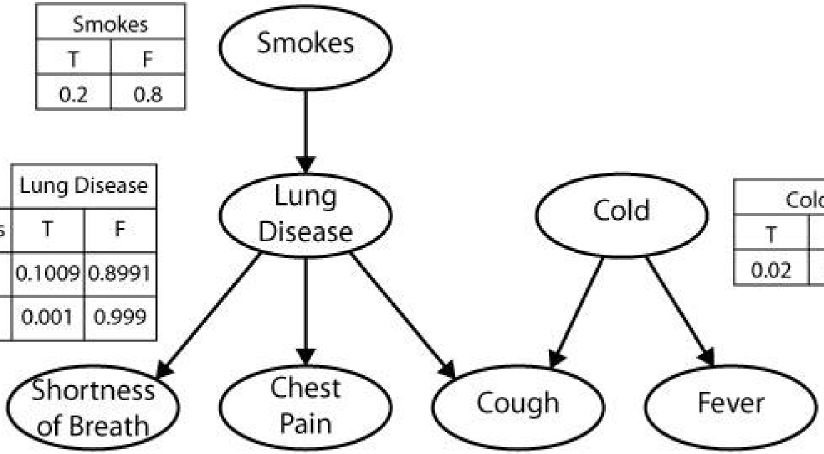
Case Study: Medical Diagnosis

Expert systems built with the help of doctors and with statistical record. Each **disease** has some **symptoms**. Using a Bayesian network we can infer the probability of a certain disease given the presence of specific symptoms.

Differential diagnostic tools, like **DXPlain**, use the Bayesian inference decision process to take into account symptom or lab result and then calculate the statistical probabilities of various diagnoses.

By combining attributes from the patient's file with clinical expertise, external data, **Watson for Oncology** identifies potential treatment plans for a patient.

Case Study: Medical Diagnosis



Smokes	
T	F
0.2	0.8

Smokes	Lung Disease T	Lung Disease F
T	0.1009	0.8991
F	0.001	0.999

Cold	
T	F
0.02	0.98

Lung Disease	Shortness of Breath T	Shortness of Breath F
T	0.208	0.792
F	0.01	0.99

Lung Disease	Chest Pain T	Chest Pain F
T	0.208	0.792
F	0.01	0.99

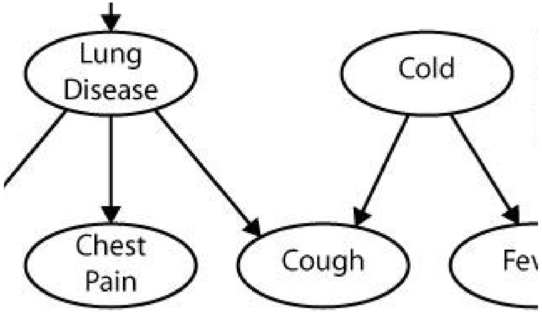
Lung Disease	Cold	Cough T	Cough F
T	T	0.7525	0.2475
T	F	0.505	0.495
F	T	0.505	0.495
F	F	0.01	0.99

Cold	Fever T	Fever F
T	0.307	0.693
F	0.01	0.99

Case Study: Medical Diagnosis

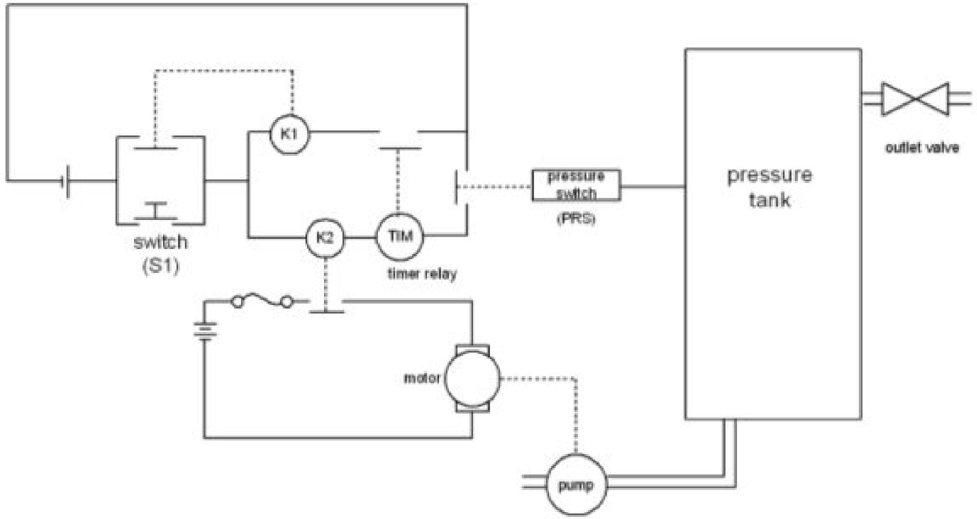
Quiz: Do you remember the definition of **independence** of two random variables?

Having a **cold** and having **lung disease** are a priori independent both causally and statistically. But because they are both causes of coughing if we observe **cough** (evidence) then cold and lung-disease become statistically dependent.



Case Study: Fault Diagnosis

It is possible to build a bayesian network which describes the correlations between a system behaviour and possible **faults**. For example, **pressure tank** system, a system that discharges fluid from a reservoir into a pressure tank with a control system that regulates the operation of the pump



Case Study: Fault Diagnosis

The top event considered is **Pressure Tank Overfilled** and the component failure modes are:

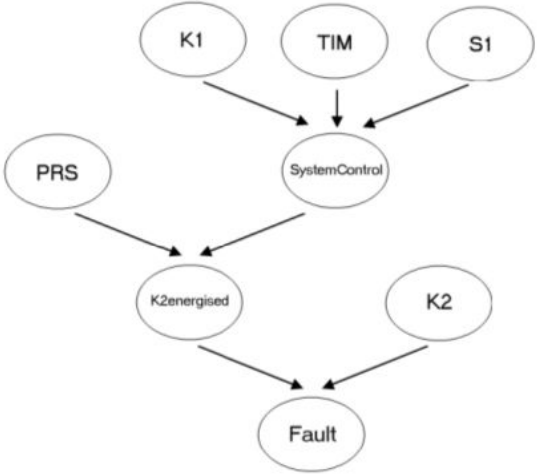
PRS Pressure switch fails to open

K2 Relay K2 contacts fail closed

K1 Relay K1 contacts fail closed

TIM Timer relay fails to timeout

S1 Switch contacts fails closed



Case Study: Cognitive Process Modelling

“Developmental Bayesian Model of Trust in Artificial Cognitive Systems.” Massimiliano Patacchiola and Angelo Cangelosi. Conference Epirob 2016.

Developmental Robotics: starting from children’s mental processes find mechanism to use in robotics. Replicating children’s mental processes it is possible to understand how human brain works.

Bayesian Networks: can describe the mental process of children dealing with the estimation of the reliability of informants.

Case Study: Cognitive Process Modelling

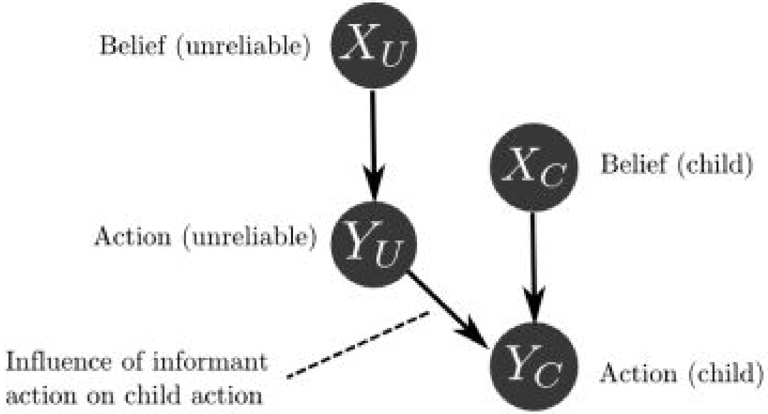
"Developmental Bayesian Model of Trust in Artificial Cognitive Systems." Massimiliano Patacchiola and Angelo Cangelosi. Conference Epirob 2016.

- Two informants, one reliable and one unreliable.
- The reliable informant suggests to the child the correct label for known objects. The unreliable suggests incorrect labels for known objects.
- Unknown objects are presented to the child. The reliable informant suggests a label, the unreliable suggests another label.
- Results: 3-years-old children do not differentiate between the two informant, whereas the 4-years-old can discriminate.

Case Study: Cognitive Process Modelling

"Developmental Bayesian Model of Trust in Artificial Cognitive Systems." Massimiliano Patacchiola and Angelo Cangelosi. Conference Epirob 2016.

The actions of the child are influenced by the child's belief and the informant's action.



Thank you

CRNS
Centre for Robotics and Neural Systems

Email: massimiliano.patacchiola@plymouth.ac.uk
Github: mpatacchiola