

The need for a framework and not only a bunch of benches came from 2 reasons : I had not enough time to implement all the benchmarks they could ever want, I could not foresee all the benchmarks they will need. They also wanted automated fetch of the benchmarks results and plotting. The aim for a Probayes engineer was that if he wanted to add a new benchmark : he would have minimum code to write and only one file (his new benchmark) to modify.

## 2.3 A note on Bayesian Programming

**“Probability is an alternative to logic to rationally reason with incomplete and uncertain knowledge.”** quote from [bayesian-programming.org](http://bayesian-programming.org)

The Bayes rule relates the conditional and marginal probabilities of events A and B :

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

with  $P(A|B)$  being the probability of having event A knowing events B occurred.

It can also be expressed on this form :

$$P(A \wedge B) = P(A)P(B|A)$$

Which means : “the probability of having event A and event B is the probability of having event A, and now that we have event A : the probability of having event B, knowing A”.

Bayesian programming is bound to Bayesian networks. A Bayesian network is a probabilistic graphical model. It reflects the states of some part of a world that is being modeled with variables and it describes how those states are related by probabilities. All the possible states of the model represent all the possible worlds that can exist, that is, all the possible ways that the parts or states can be configured.

### A quick example : naive Bayesian model

Think that you have to program an anti-spam filter, how would you do ? A simple and very efficient answer is to do it with a Bayesian model. It is so simple it is called “naive”. Take a sample of spam e-mails, find the words (or digrams, association of two words, or trigrams) that are significant in these mails. Note the probability that a message contains these words knowing that it is a spam.

$$P(W_i|Spam)$$

If you want to know (or have a good indication) if an incoming message is a spam, ask yourself : “What is the probability that the message is a spam and it contains  $W_1, \dots, \neg W_j, \dots, W_n$  ?” or “What is the probability to observe event *Spam* and events  $W_i$  ?” which can be expressed and decomposed in :

$$P(Spam \wedge W_1 \wedge \dots \wedge W_i \wedge \dots \wedge W_n) = P(Spam) \times \prod_{i=1}^n P(W_i|Spam)$$

And you just have to do the maths : the different  $P(W_i|Spam)$  have been observed (and can be learned) and  $P(Spam)$  is the *a priori* probability to have a spam mail (is 80% of your e-mails are *Spam* so let it be 0.8).

### Asia, a basic Bayesian network

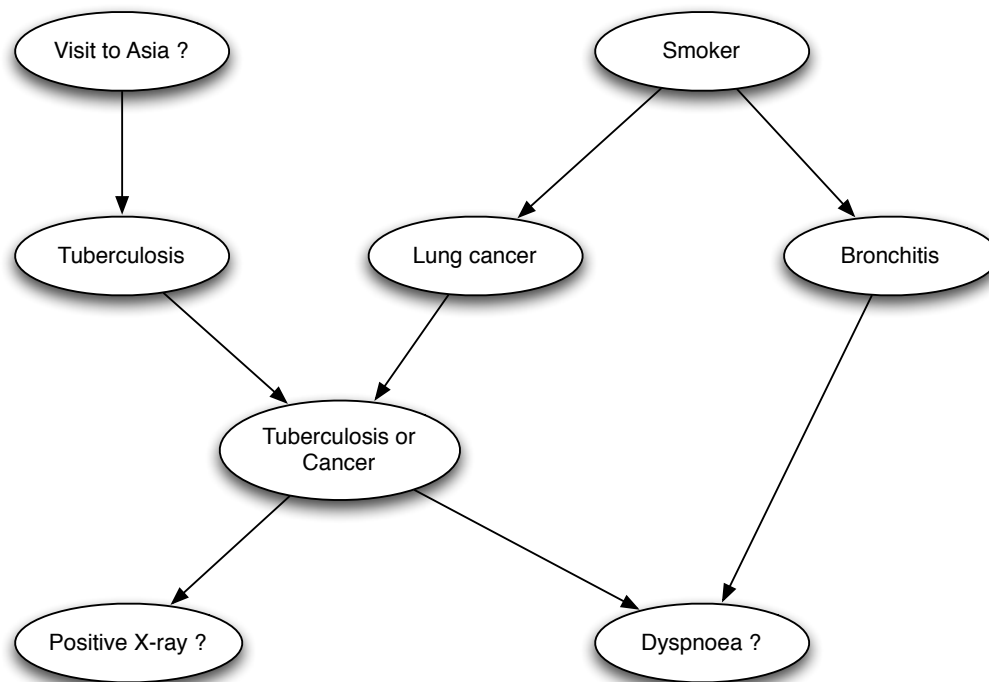


Figure 4: Asia Bayesian network

Asia is a small and well known Bayesian network (from [Lauritzen&Spiegelhalter88]) that calculates the probability of a patient having tuberculosis, lung cancer or bronchitis respectively based on different factors.

Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer, bronchitis, more than one of these diseases or none of them. A recent visit to Asia increases the risk of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea.

If we learn the fact that a patient is a smoker, we will adjust our beliefs (increased risks) regarding lung cancer and bronchitis. However, our beliefs regarding tuberculosis are unchanged (i.e., tuberculosis is **conditionally independent** of "smoking" given the empty set of variables). Now, suppose we get a positive X-ray result for the patient : This will affect our beliefs regarding tuberculosis and lung cancer, but not our beliefs regarding bronchitis (i.e., bronchitis is conditionally independent of X-ray given smoking). However, had we also known that the patient suffers from shortness-of-breath, the X-ray result would also have affected our beliefs regarding bronchitis (i.e., "bronchitis" is **not conditionally independent** of "X-ray" given "smoking" and "dyspnoea").