Bayesian Network

What is Bayesian Statistics

Bayes' theorem is the basis of Bayesian statistics. It enables the user to update the probabilities of unobserved events. Consider that you have a prior probability for the unobserved event. A related event has occurred. Now you can update the prior probability to get the posterior probability of the event.
Bayesian inference found the data that already occurred, and not on the data that could have occurred but did not. Bayes' theorem finds the posterior density of parameters for a given data. It combines information about the parameters from prior density with the observed data.

Bayes' Theorem – A theorem of probability theory states by the **Reverend Thomas Bayes**. A new piece of evidence affects the way of understanding how the probability that a theory is true. We use it in a wide variety of contexts from marine biology to the development of "**Bayesian**" spam blockers for email systems.
When applied, the probabilities involved in Bayes's theorem may have some probability interpretations. There is the use of the theorem as part of a particular approach to statistical inference.

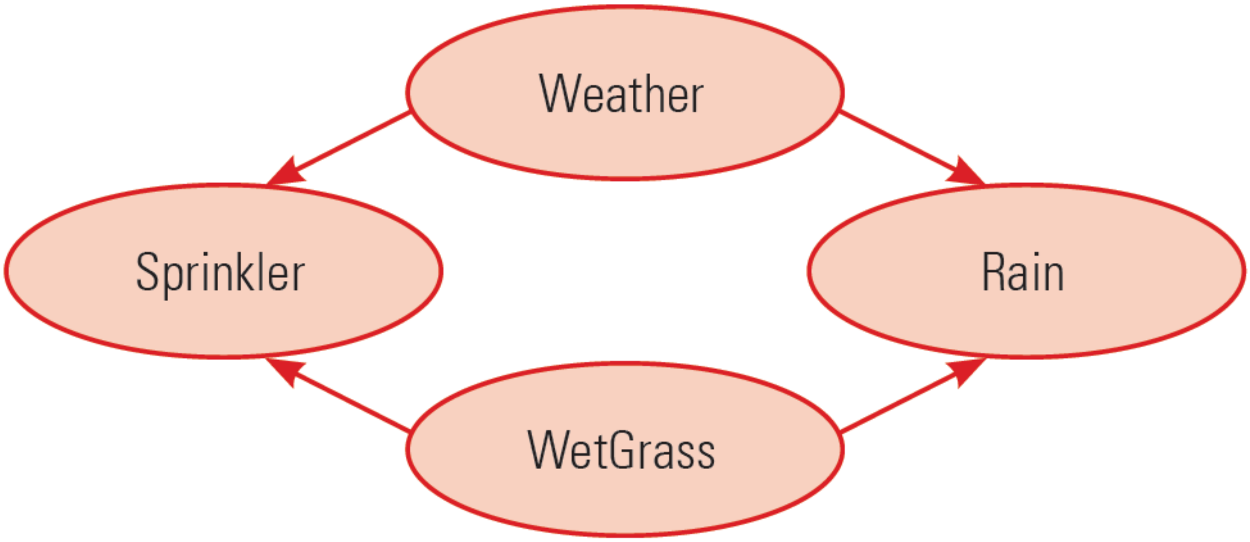
What is Bayesian Network?

● A simple, graphical notation for conditional independence assertions and hence for compact specification of full joint distributions

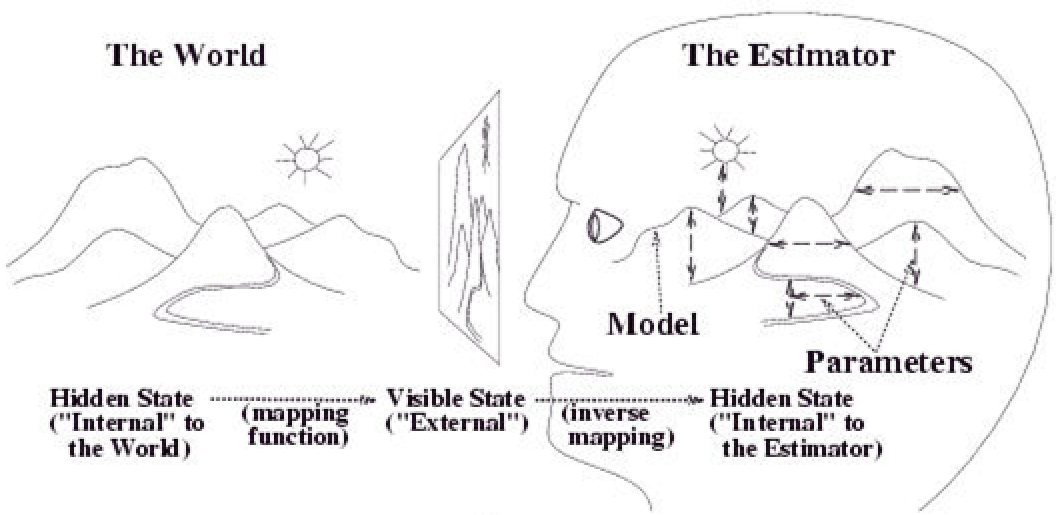
● Syntax – a set of nodes, one per variable – a directed, acyclic graph (link ≈ "directly influences") – a conditional distribution for each node given its parents: $P(X_i | Parents(X_i))$

Finnally, $P(x_1, \ldots, x_n) = \prod P(x_i | parents(X_i))$

Example1:

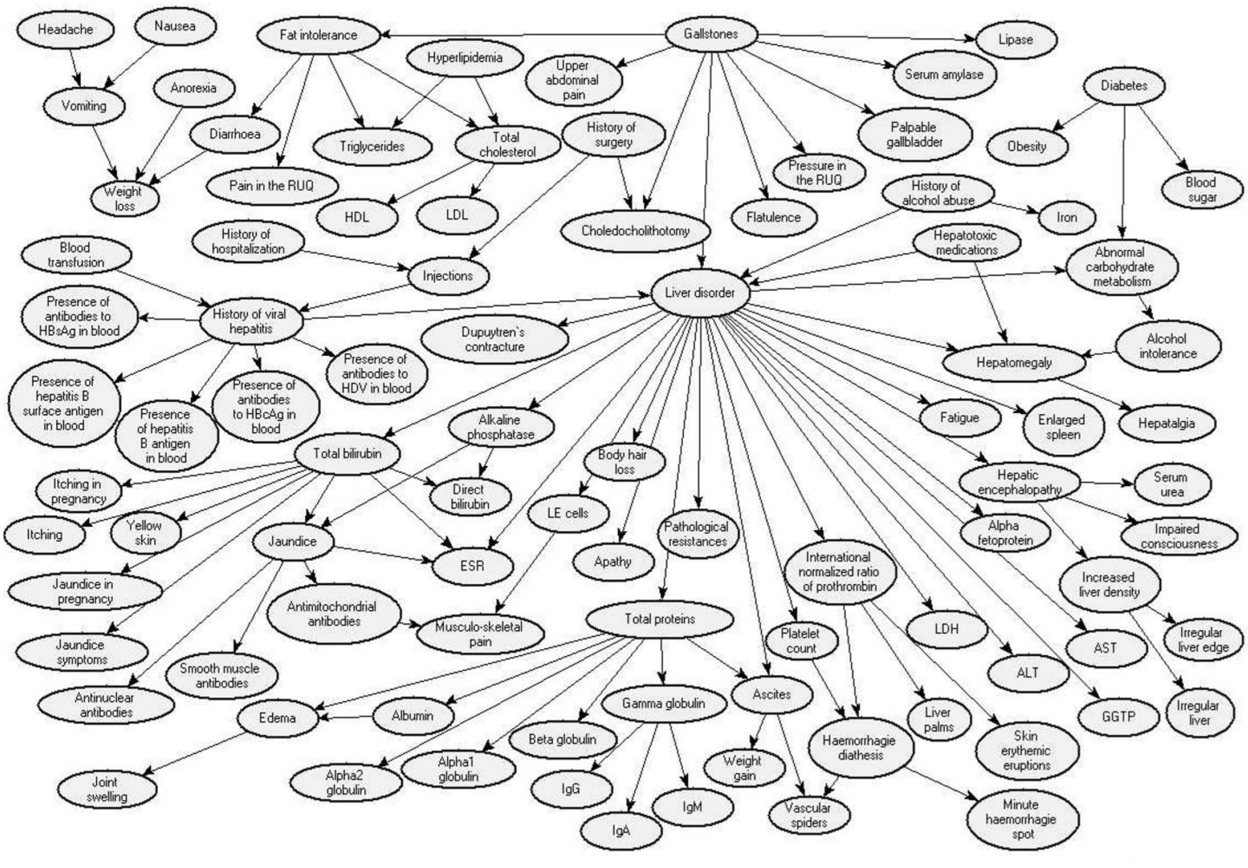


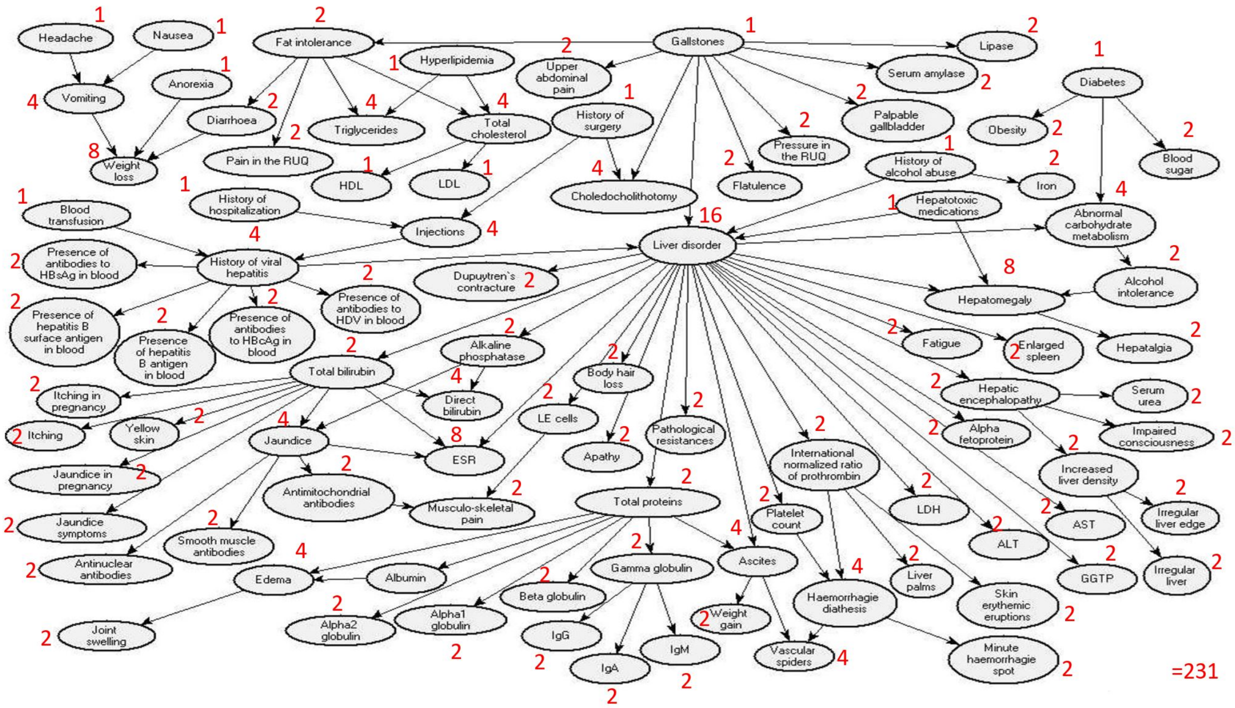
Bayesian network reflects the states of some part of the world that is being modeled and it describes how those states are related by probabilities. Consider the following figure



There is an agent which recognizes a world, and it perceives that the world has mountains, roads, sun and so on. he world’s model has been recognized with the help of knowledge representation. Now, how can an agent says that there will be a possibility for the occurrence of accidents on bend road. In that case probabilities comes in. Is there any belief for the statements delivered by an agent, in that case there is a need of Bayesian network which purely works based on conditional probability. Well, typically some states will tend to occur more frequently when other states are present. Thus, if you are sick, the chances of a runny nose are higher. If it is cloudy, the chances of rain are higher, and so on. Here is a simple Bayes net that illustrates these concepts. In this simple world, let us say the weather can have three states: sunny, cloudy, or rainy, also that the grass can be wet or dry, and that the sprinkler can be on or off. Now there are some causal links in this world. If it is rainy, then it will make the grass wet directly. But if it is sunny for a long time, that too can make the grass wet, indirectly, by causing us to turn on the sprinkler.

Example2:





Headache
Nausea
Vomiting
Anorexia
Diarrhoea
Weight loss
Fat intolerance
Hyperlipidemia
Pain in the RUQ
Triglycerides
Total cholesterol
HDL
LDL
Upper abdominal pain
History of surgery
Gallstones
Choledocholithotomy
Lipase
Serum amylase
Palpable gallbladder
Pressure in the RUQ
Flatulence
Diabetes
Obesity
Blood sugar
History of alcohol abuse
Iron
Hepatotoxic medications
Abnormal carbohydrate metabolism
Alcohol intolerance
Blood transfusion
History of hospitalization
Injections
Liver disorder
History of viral hepatitis
Presence of antibodies to HBsAg in blood
Presence of hepatitis B surface antigen in blood
Presence of hepatitis B antigen in blood
Presence of antibodies to HBcAg in blood
Presence of antibodies to HDV in blood
Dupuytren`s contracture
Hepatomegaly
Hepatalgia
Fatigue
Enlarged spleen
Total bilirubin
Alkaline phosphatase
Direct bilirubin
Body hair loss
Itching in pregnancy
Itching
Yellow skin
Jaundice
ESR
LE cells
Apathy
Pathological resistances
Hepatic encephalopathy
Serum urea
Alpha fetoprotein
Impaired consciousness
Jaundice in pregnancy
Antimitochondrial antibodies
Musculo-skeletal pain
International normalized ratio of prothrombin
Increased liver density
Jaundice symptoms
Smooth muscle antibodies
Total proteins
Platelet count
LDH
AST
Irregular liver edge
ALT
Antinuclear antibodies
Edema
Albumin
Gamma globulin
Ascites
Haemorrhagic diathesis
Liver palms
GGTP
Irregular liver
Beta globulin
Alpha2 globulin
Alpha1 globulin
IgG
IgA
IgM
Weight gain
Vascular spiders
Skin erythemic eruptions
Minute haemorrhagic spot
Joint swelling
=231

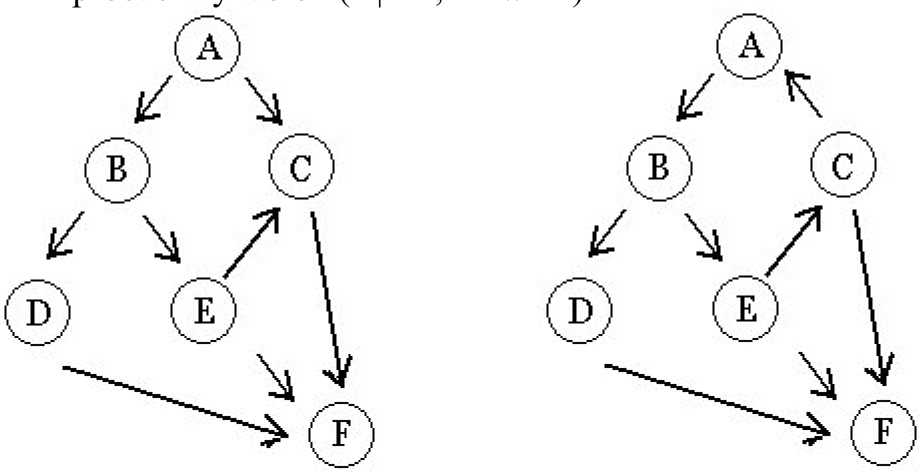
Bayesian Network Model

Definition of Bayesian Networks

A Bayesian Network consists of [Jensen, 1996]:

A set of variables and a set of direct edges between variables

- Each variables has a finite set of mutually exclusive states
- The variable and direct edge form a DAG (directed acyclic graph)
- To each variable A with parents B1, B2 ..Bn there is attached a conditional probability table P(A| B1, B2 .. Bn)



The graph on the left is a valid Bayesian netwrok. The probabilities to specifty are P(A), P(B|A), P(C|A,E),P(D|B), P(E|B) and P(F|C,D,E).
The one on the right is not a valid Bayesian network as the cycle ABEC exists.

In a Bayesian network, quoting directly from [Nilsson, 1998], each node of the graph are "conditionally independent of any subset of the nodes that are not descendants of itself given its parent". *(huh?)*

Let put this in another way: If we let V be a node in graph, and non(V) be any set of nodes that are non-descentant of V, and par(V) be set of the immediate parent of V, then non(V) is conditionally independent of V for par(V), or

$$\mathbf{P(V \mid non(V),par(V)) = P(V \mid par(V))}$$

Hence, for every node in the network (V_1, V_2 ... V_n) are we can write down the joint probability of all nodes as:

$$P(V_1, V_2, \ldots V_n) = \prod_{i=1}^{n} P(V_i \mid par(V_i))$$

After getting the joint probability the system, we can then extract the necessary probability by the use of marginalization.

From this rules, we can see that in order to calculate the value of joint probabilities, we just need to calculate the conditional probability between itself and its parent, for each nodes in the network. Then we can use the chain rule obtain the joint probability functions of the network.

In the left example above, we can obtain P(A,B,C,D,E,F) by the use of chain rules and theory on conditional independent(In this example (D,E) and (C,D) are conditionally independent of each other.):

$$\begin{aligned} P(A,B,C,D,E,F) &= P(F|C,D,E)P(A,B,C,D,E) \\ &= P(F|C,D,E)P(C|A,E)P(D|B)P(E|B)P(B,A) \\ &= P(F|C,D,E)P(C|A,E)P(D|B)P(E|B)P(B|A)P(A) \end{aligned}$$

For nodes that have no parents, their random variables are not conditioned on the other random variables, and are called *prior probabilities*. In the example above it will be A. Note that if a random variable is prior probabilities, then the table for its corresponding node reduces to unconditional probabilities table.

By using the Bayesian network, we have reduced number of joint probability required. For example, if A...F are all Boolean values (i.e. has two states), then we only need to store 1+2+4+2+2+8= 19 nodes conditional probabilities if we use Bayesian network formula, as oppose to 2^6-1 = 63 if we use joint probabilities directly.

NOTE: 1. Every node has no parent, given value is 1

(eg, suppose node A has no parent so node A value is 1)

2. Every node has k parent, given value is 2^k .

(eg, suppose node F has 3 parent so node F value is 2^3 is 8)