**Unit-III Part B**

| Introduction to Alternative techniques-Bayes Theorem |
| --- |
| Bayes theorem  and its usage |
| Naïve bayes classifier |
| Bayesian belief networks |

**🡪Introduction to Alternative techniques-Bayes Theorem:**

--Introduction : Bayesian Classifier

--Baye’s Theorem

**Introduction: Bayesian Classifier**

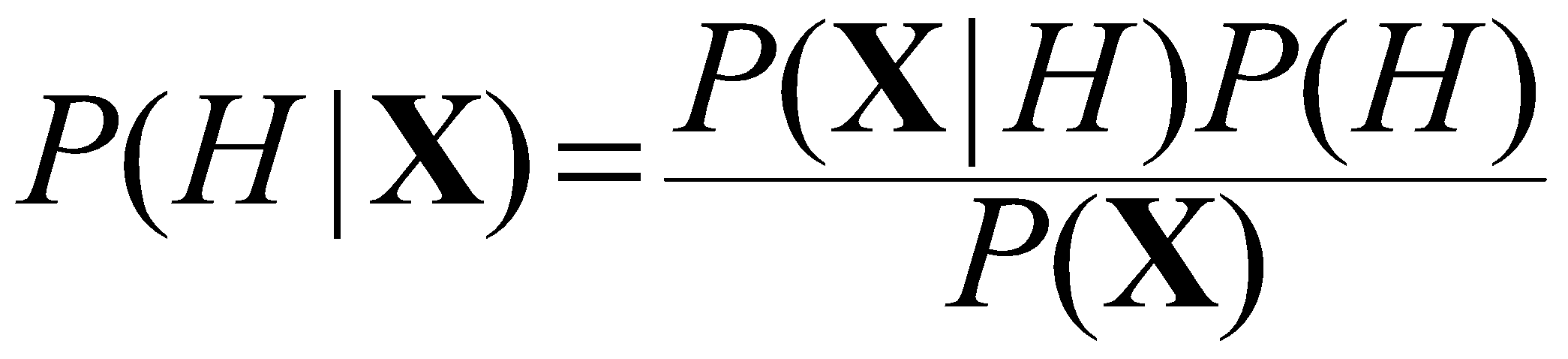
* It is a statistical classifier which performs *probabilistic prediction, i.e.,* predicts class membership probabilities
* The foundation of the Bayesian classifier is based on Bayes’ Theorem.
* The performance of the Bayesian classifier can be measured using a simple Bayesian classifier, *naïve Bayesian classifier.*

**Bayes Theorem:**

* Let X be a data sample
* Let H be a *hypothesis* that X belongs to class C
* Our Classification is to determine P(H|X), the probability that the hypothesis holds given the observed data sample X
  + Example: customer X will buy a computer given that know the customer’s age and income

In the above example:

* **P(H)** (*prior probability*) is the initial probability
  + E.g., X will buy computer, regardless of age, income, …
* P(X): is the probability that sample data is observed
* P(X|H) (*posteriori probability*), is the probability of observing the sample X, given that the hypothesis holds
  + E.g., Given that X will buy computer, the prob. that X is 31..40, medium income
* Given training data**X***, posteriori probability of a hypothesis* H*,* P(H|**X**)*,* follows the Bayes theorem



**Proof:**

Mathematically, Bayes' theorem gives the relationship between the probabilities of A and B, P(A) and P(B), and the conditional probabilities of A given B and B given A, P(A|B) and P(B|A).   
  
P(A|B) = {P(B|A)\*P(A)}/{P(B)}   
  
DERIVATION   
We can do it from set theory applied to conditional probabilities.   
  
P(A given B) = P(A and B)/P(B)   
  
Likewise, P(B given A) = P(A and B)/P(A)   
  
Rearrange to get the common term as follows:   
  
P(A and B) = P(A given B).P(B) = P(B given A).P(A)   
  
Divide the middle and right hand terms by P(B) on condition that it is not zero:   
  
P(A given B) = P(B given A).P(A)/P(B)   
  
That is a basic statement of Bayes' Theorem.

**Example:**

A simple example is as follows: There is a 40% chance of it raining on Sunday. If it rains on Sunday, there is a 10% chance it will rain on Monday. If it didn't rain on Sunday, there's an 80% chance it will rain on Monday.

"Raining on Sunday" is event A, and "Raining on Monday" is event B.

* P(*A*) = 0.40 = Probability of Raining on Sunday.
* P(*A`*) = 0.60 = Probability of not raining on Sunday.
* P(*B|A*) = 0.10 = Probability of it raining on Monday, if it rained on Sunday.
* P(*B`|A*) = 0.90 = Probability of it not raining on Monday, if it rained on Sunday.
* P(*B|A`*) = 0.80 = Probability of it raining on Monday, if it did not rain on Sunday.
* P(*B`|A`*) = 0.20 = Probability of it not raining on Monday, if it did not rain on Sunday.

The first thing we'd normally calculate is the probability of it raining on Monday: This would be the sum of the probability of "Raining on Sunday and raining on Monday" and "Not raining on Sunday and raining on Monday"

 chance

However, what if we said: "It rained on Monday. What is the probability it rained on Sunday?" That is where Bayes' theorem comes in. It allows us to calculate the probability of an earlier event, given the result of a later event.

The equation used is:



In our case, "Raining on Sunday" is event A, and "Raining on Monday" is event B.

* P(*B|A*) = 0.10 = Probability of it raining on Monday, if it rained on Sunday.
* P(*A*) = 0.40 = Probability of Raining on Sunday.
* P(*B*) = 0.52 = Probability of Raining on Monday.

So, to calculate the probability it rained on Sunday, given that it rained on Monday:



or:

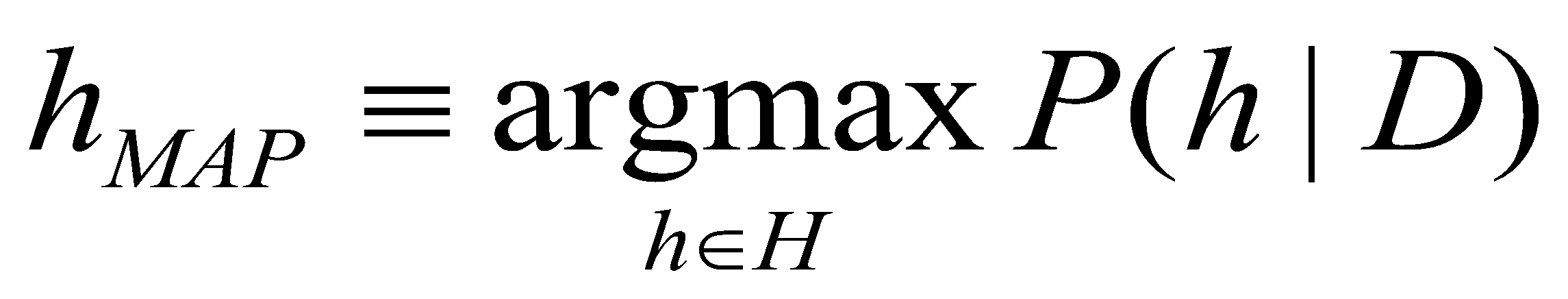


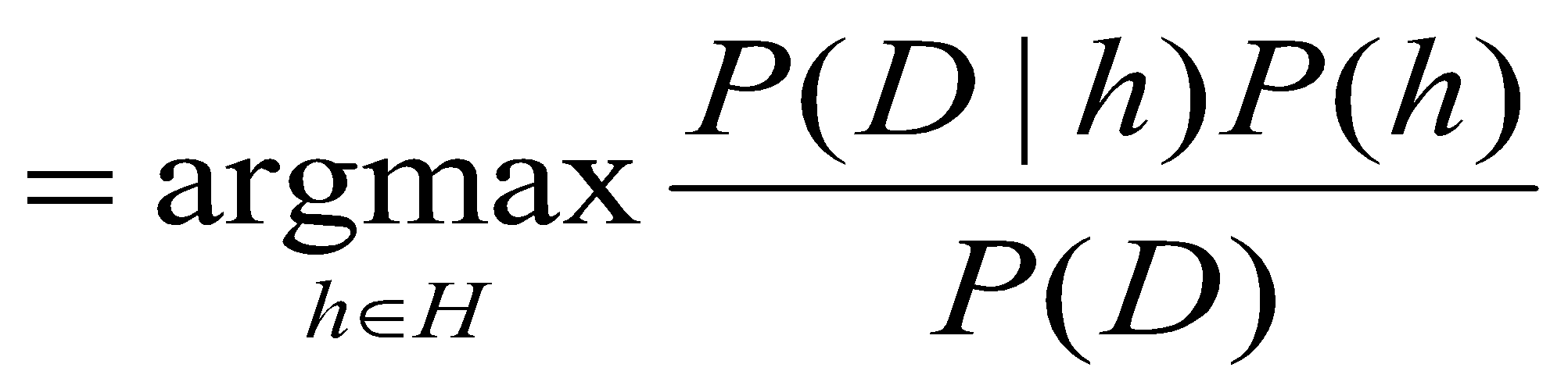
In other words, if it rained on Monday, there's a 7.69% chance it rained on Sunday.

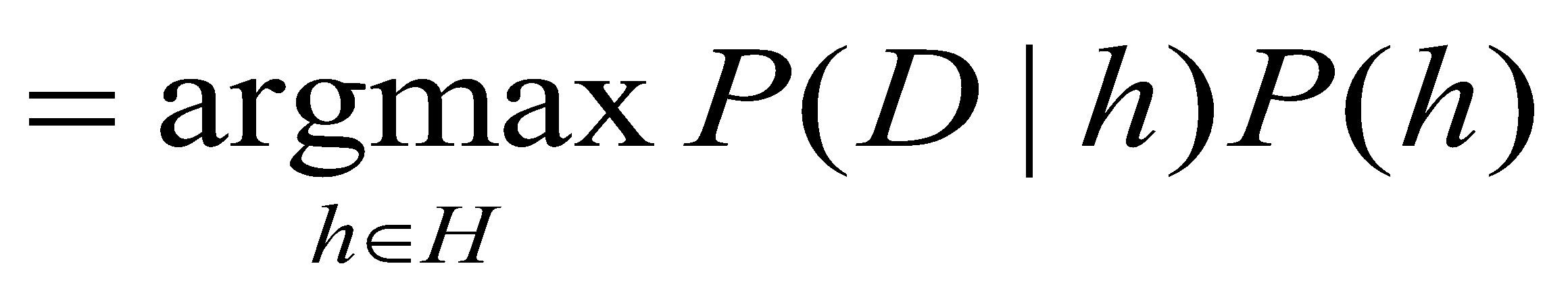
**🡪Bayes theorem  and its usage:**

**“For example refer to PPT”**

**🡪Naïve bayes classifier (NBC):**

* This classifier is also based on Bayes Theorem
* In the NBC we should compute the *Maximum A Posterior* (MAP) hypothesis for the data
* So, the MAP can be calculated using the hypothesis for some space H given observed training data D.





*H*: set of all hypothesis.

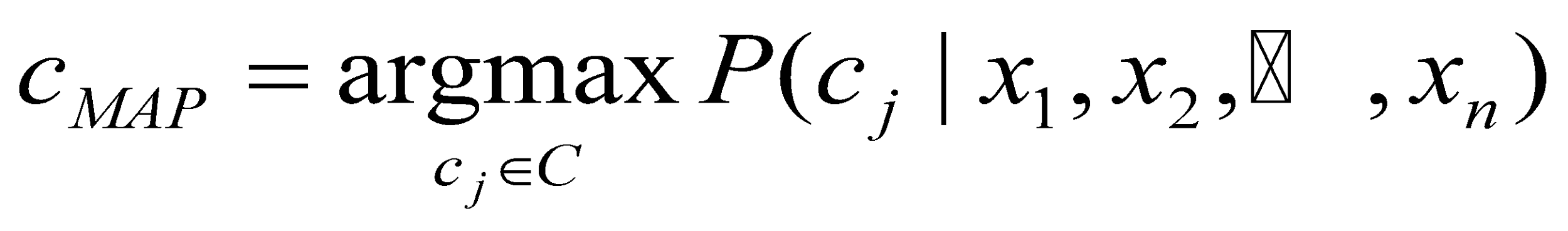
**Note** that we can drop *P(D)* as the probability of the data is constant (and independent of the hypothesis).

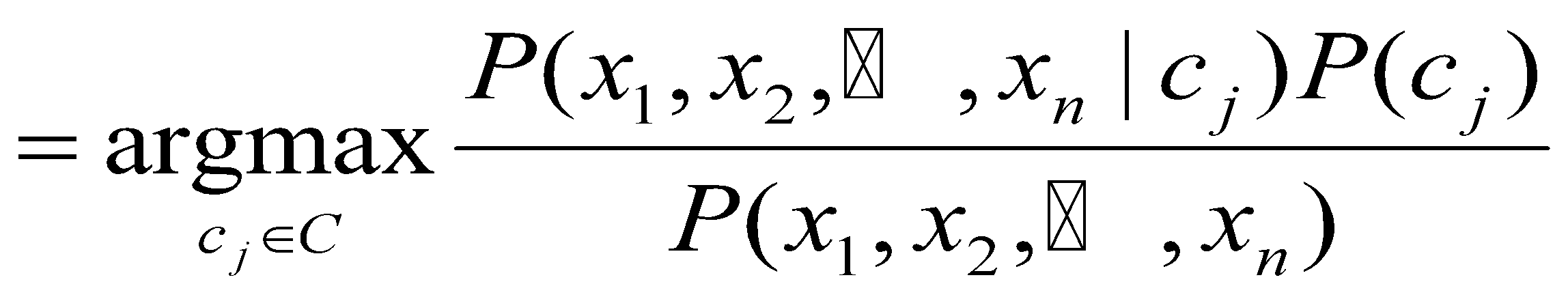
**Proof:**

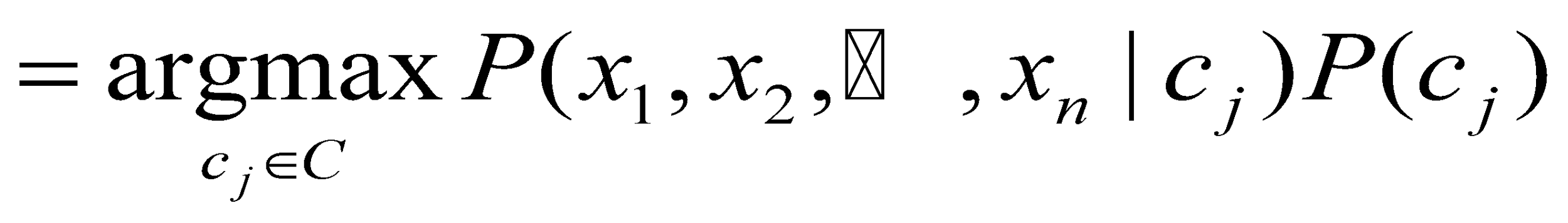
**Assumption:** training set consists of instances of different classes described *cj* as conjunctions of attributes values

**Task:** Classify a new instance *d* based on a tuple of attribute values into one of the classes *cj* ∈ *C*

**Key idea:** assign the most probable class using Bayes Theorem.







* *P*(*cj*)

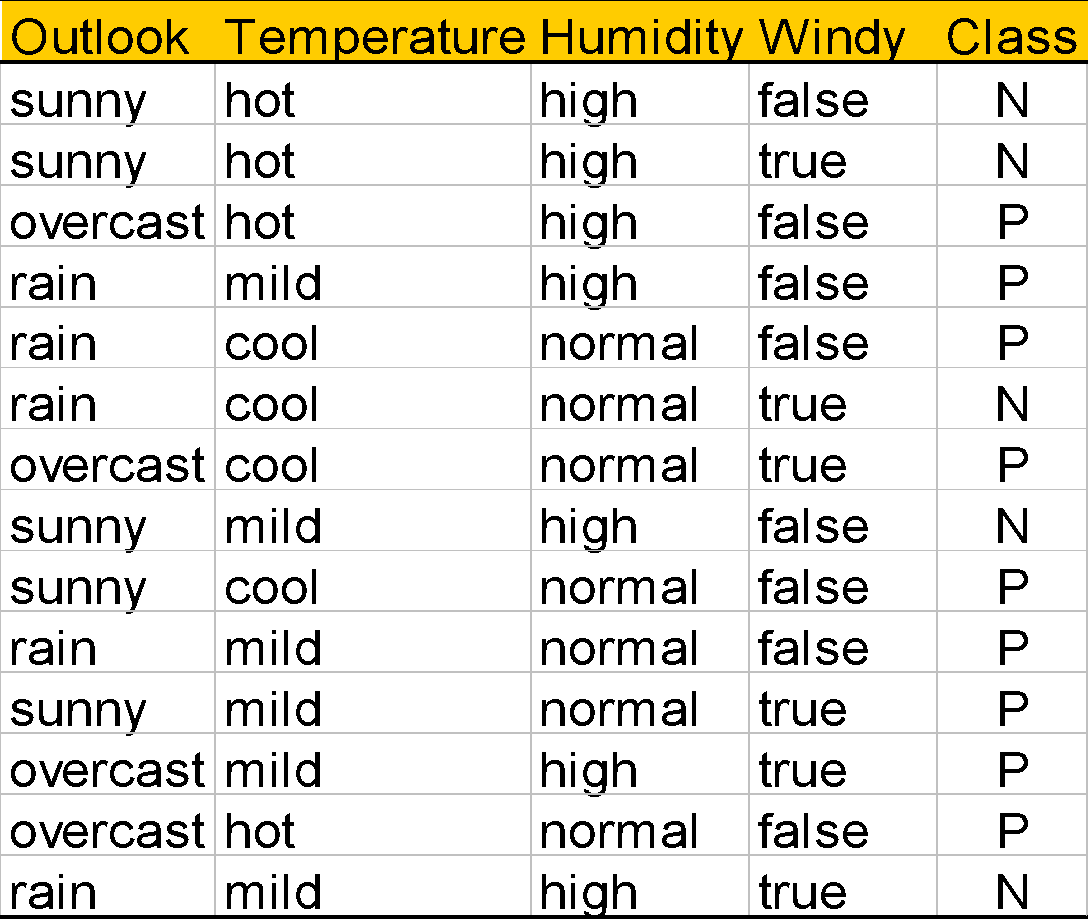
--Can be estimated from the frequency of classes in the training examples.

* *P*(*x1,x2,…,xn|cj*)

--Could only be estimated if a very, very large number of training examples was available.

* This Independent Assumption of attribute values is conditionally independent given the target value:
* **And this Independent Assumption is called as *naïve* *Bayes*.**

**Example:** Consider a tennis match schedule based on the following independent assumptions:



--Based on the above assumptions, the probabilities of Play **P( p)** and not play **P(n)** are:

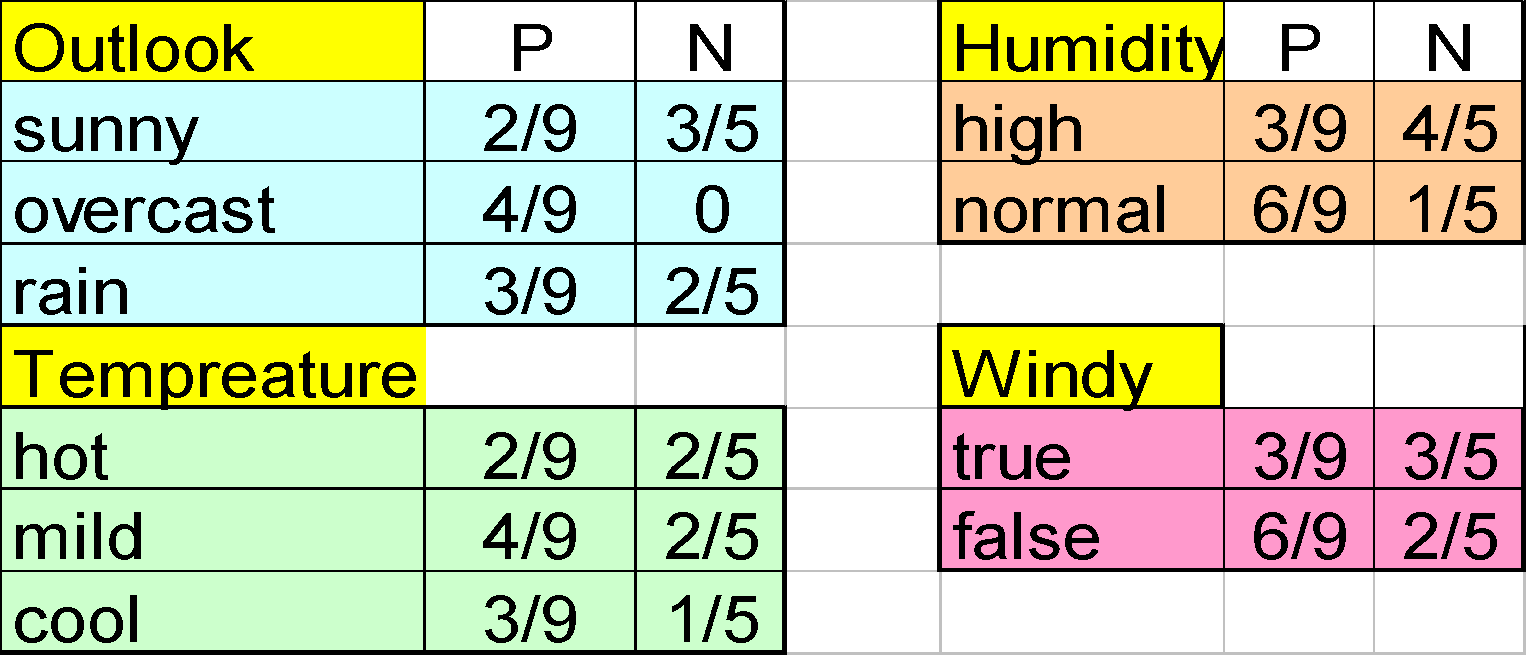




--And the all probabilities of the different attributes (sunny, overcast, hot ,….) are represented as:



--Given a training set, we can compute the probabilities:



--An unseen sample X = <rain, hot, high, false>

--P(X|p)·P(p) =   
 P(rain|p)·P(hot|p)·P(high|p)·P(false|p)·P(p) = 3/9·2/9·3/9·6/9·9/14 = 0.010582

--P(X|n)·P(n) =   
 P(rain|n)·P(hot|n)·P(high|n)·P(false|n)·P(n) = 2/5·2/5·4/5·2/5·5/14 = 0.018286

--Sample X is classified in class n (**don’t play**) as the sample indicates a high probality of P(n) then P(p).

**Features:**

--NBC makes computation possible

--NBC is seldom satisfied in practice, as attributes (variables) are often correlated.

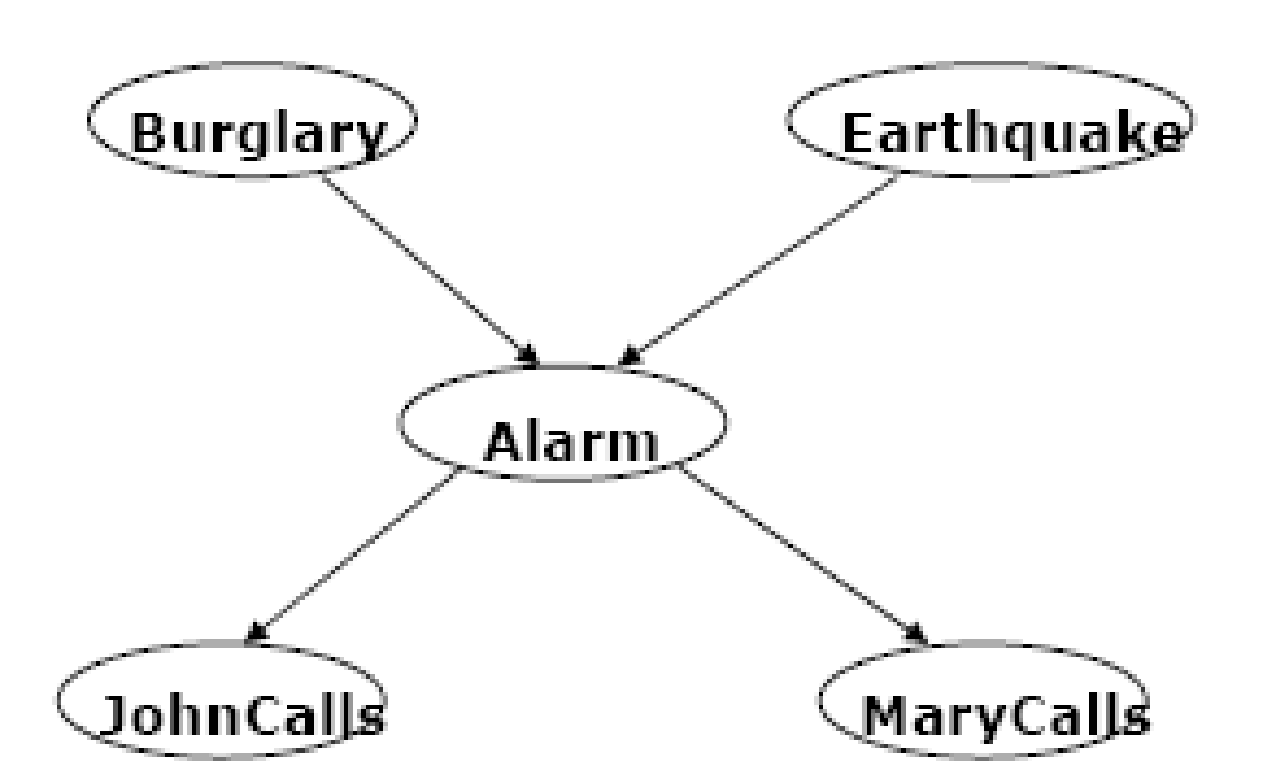
**🡪Bayesian belief networks: (BBN):**

* In Naïve Bayes Classifier we make the assumption of class conditional independence, that is given the class label of a sample, the value of the attributes are conditionally independent of one another.
* However, there can be dependences between value of attributes. To avoid this we use Bayesian Belief Network which provide joint conditional probability distribution.
* A Bayesian network is a form of probabilistic graphical model. Specifically, a Bayesian network is a directed acyclic graph of nodes representing variables and arcs representing dependence relations among the variables.
* DAG that represents the dependencies between variables and specifies the joint probability distribution
* Random variables make up the nodes
* Directed links represent causal direct influences
* Each node has a conditional probability table quantifying the effects from the parents
* No directed cycles

Assume your house has an **alarm system** against **burglary**. You live in the seismically active area and the alarm system can get occasionally set off by an **earthquake**. You have two neighbors, **Mary** and **John**, who do not know each other. If they hear the alarm they call you, but this is not guaranteed.

• We want to represent the probability distribution of events:

* Burglary, Earthquake, Alarm, Mary calls and John calls



**Directed acyclic graph**

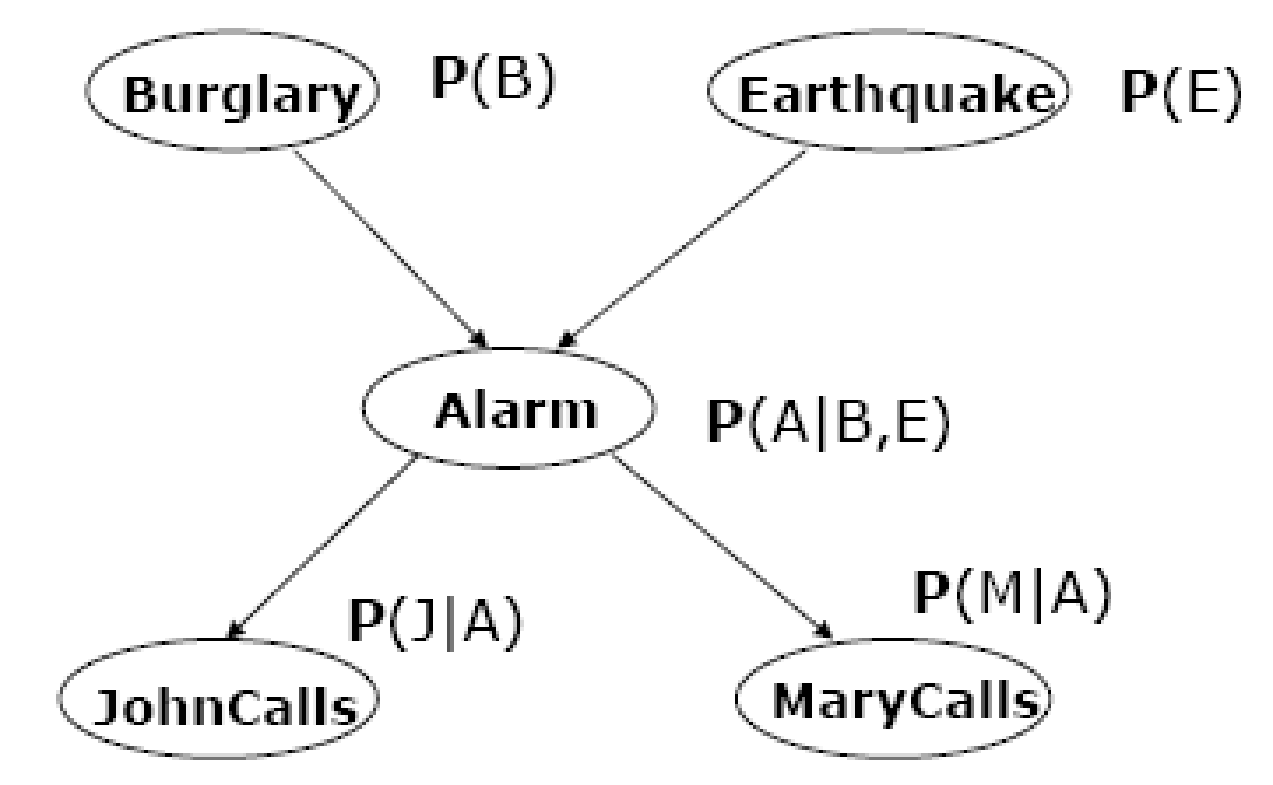
• **Nodes** = random variables

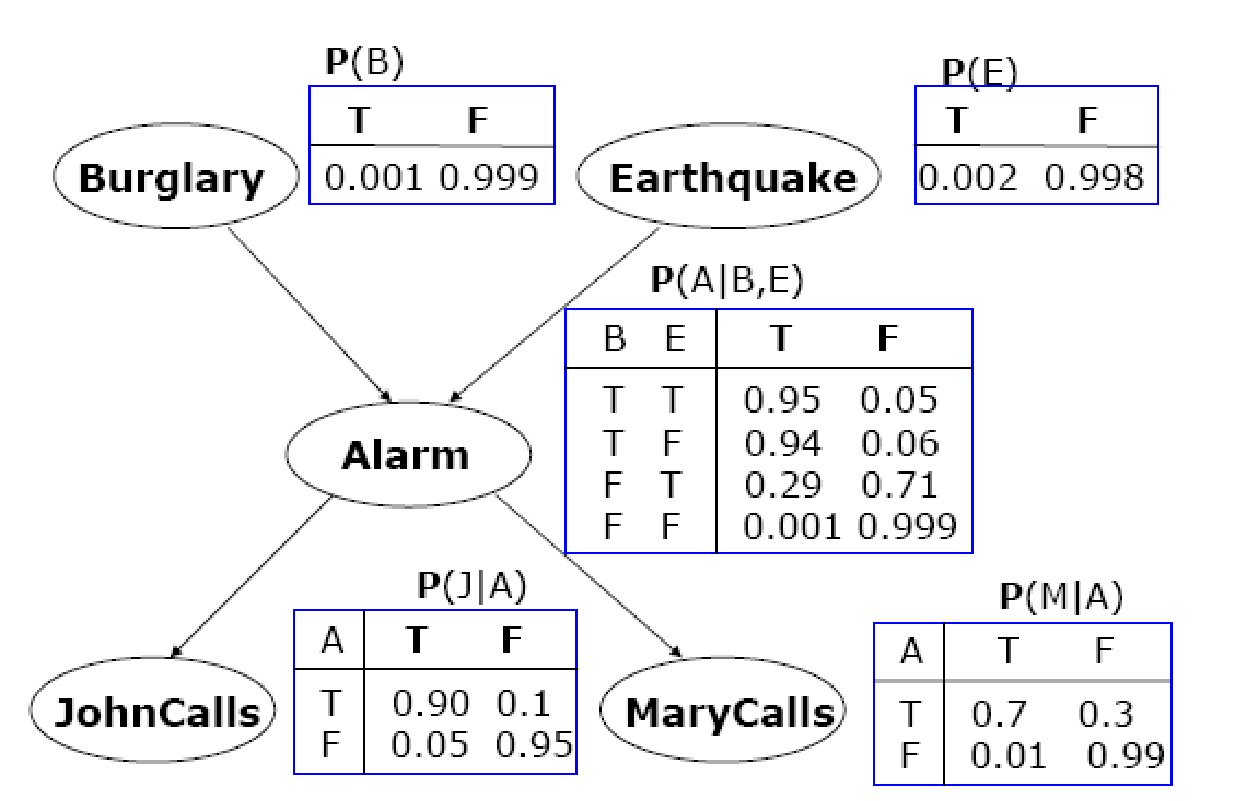
Burglary, Earthquake, Alarm, Mary calls and John calls

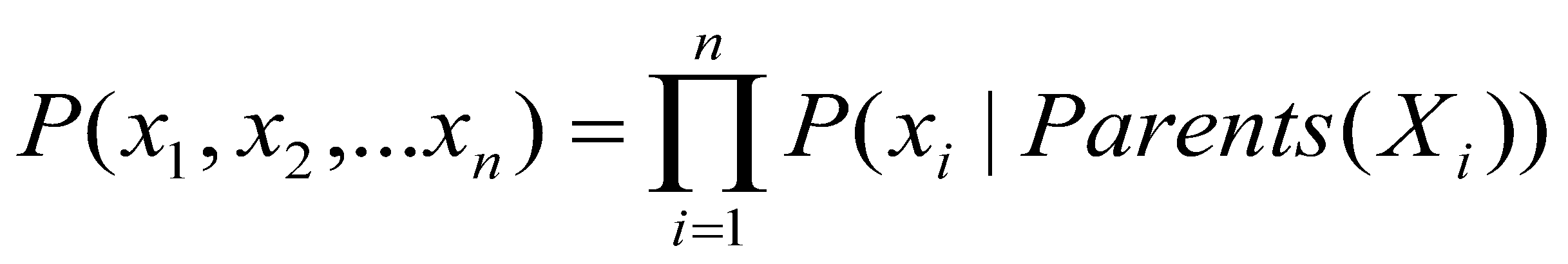
• **Links** = direct (causal) dependencies between variables.

The chance of Alarm is influenced by Earthquake, The

chance of John calling is affected by the Alarm



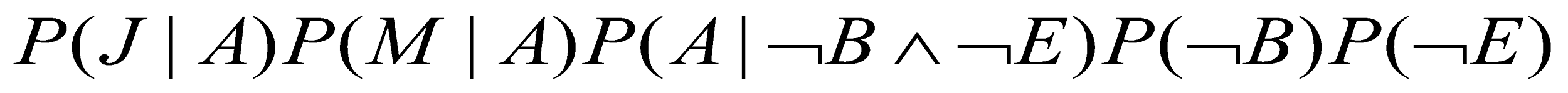


--The probability of the dependencies based on the events of their parent nodes can be represented as:

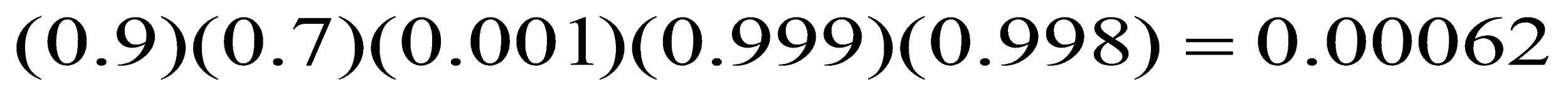
--Given a situation where:

* There is a Probability of alarm,
* With no burglary or earthquake,
* And both John and Mary call:

--This situation can be represented as:



--And the probability is:



--Therefore, the above situation has a probability of **0.00062**.