Fundamentals

**Measures of central tendency:**

• mode

• median   
• mean μ  
• x% trimmed mean

**Measures of spread:**

• ranges: crude range: highest, lowest

Extended/corrected range: adds one unit to the range (to account for a possible error in measurement)

trimmed ranges: drop x% of extreme points on both sides

* variance σ2 =Σi(xi-μ)2/n
* standard deviation σ =sqrt(σ2)
* average deviation Σi (xi-μ)/n

**Other measures of probability distribution:**

• kurtosis: a descriptor of the shape of a probability distribution

• skewness: a descriptor of asymmetry of the probability distribution

**Normal/Gaussain distribution ( σ>0)**

1. μ = median = mode

2. ***3-sigma rule***: 68% within 1σ, 95% within 2σ, 99.7% within 3σ

**Linear regression**

1. Correlation does not mean causation

**2. Least-squares regression:** minimize the sum of squares of the deviations of the data points from the line in the vertical direction

**3. Asymmetry of regression**

**4. Outliers,** typically removed manually.

Bayesian Probability Theory

1. describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

2. prior probability: P(A), P(B)

posterior probability: P(A|B), P(B|A)

3. P(A,B) = P(A|B) P(B) = P(B|A) P(A)

P(A|B) = P(B|A) / P(B) P(A)

4. example on P41

**Joint Probability Distribution** (over several random variables)

1. P(A,B) or P(A∩B)

1. Given the value of some of the variables in the join probability distribution, we can estimate the probability distributions over the remaining variables.

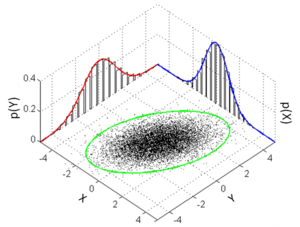
e.g. scatter plots give an idea of the joint probability distribution between two variables

3. Knowledge of the joint probability distribution is sufficient to derive any marginal and conditional probability over the model’s variables.

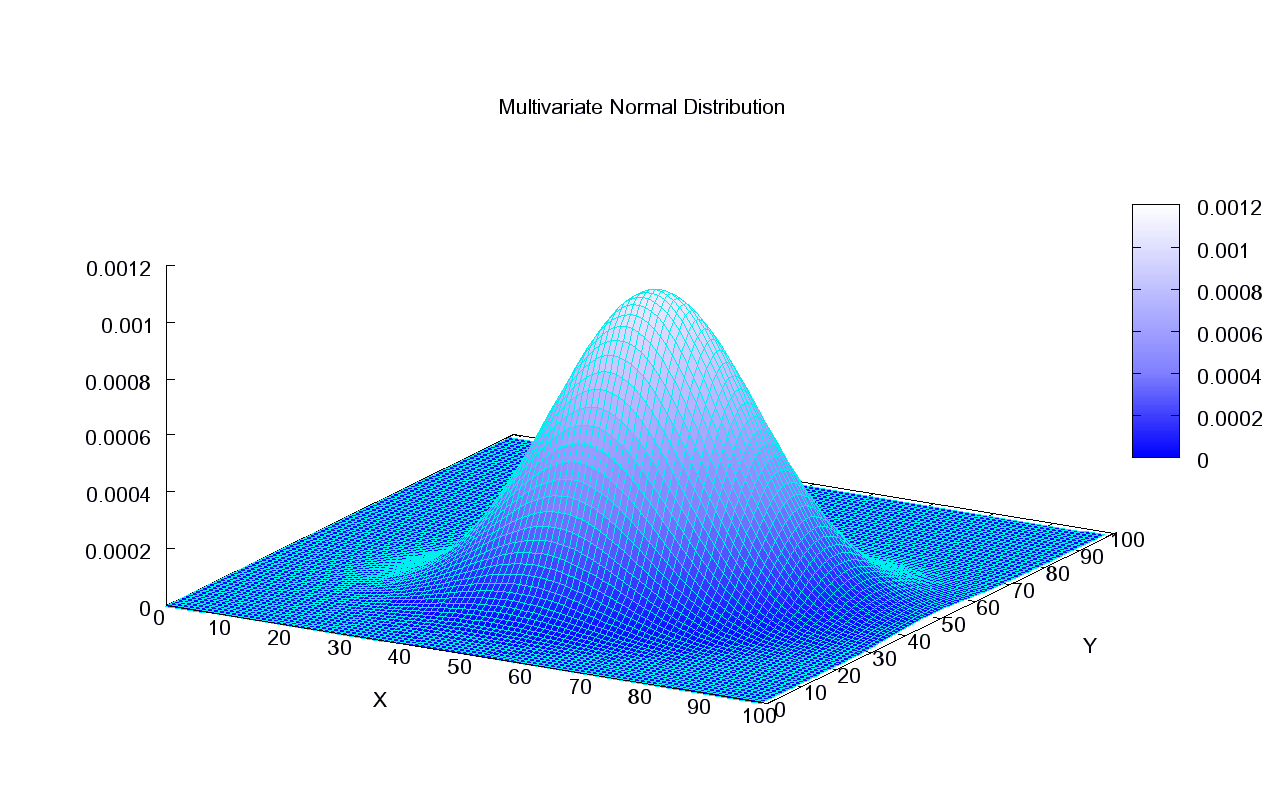
**Marginal Probability Distribution** (over a single variable)

1. P(A)

2. contrasts with a conditional distribution, which gives the probabilities contingent upon the values of the other variables



**Conditional Probability Distribution**

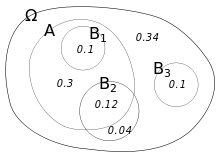


1. P(A|B)

2. Once we know the value of one of the variables, we can make a statement about the probability distribution over the other variable

3. The unconditional [probability](https://en.wikipedia.org/wiki/Probability) *P*(*A*) = 0.30 + 0.10 + 0.12 = 0.52.

The conditional probability *P*(*A*|*B*1) = 1, *P*(*A*|*B*2) = 0.12 ÷ (0.12 + 0.04) = 0.75, *P*(*A*|*B*3) = 0.

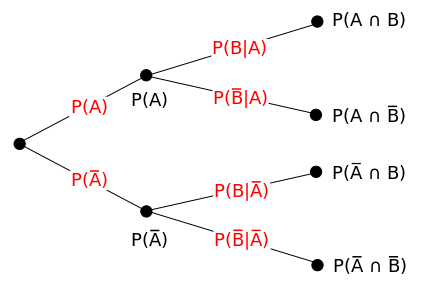


**Independence**

1. Definition: A⊥B ⇒ P(A,B) = P(A) P(B)

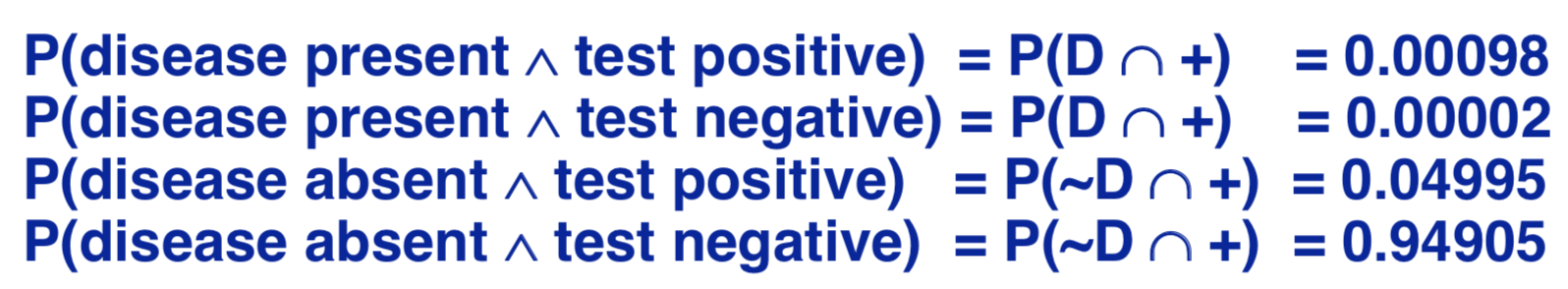
Representations of the Joint Probability Distribution

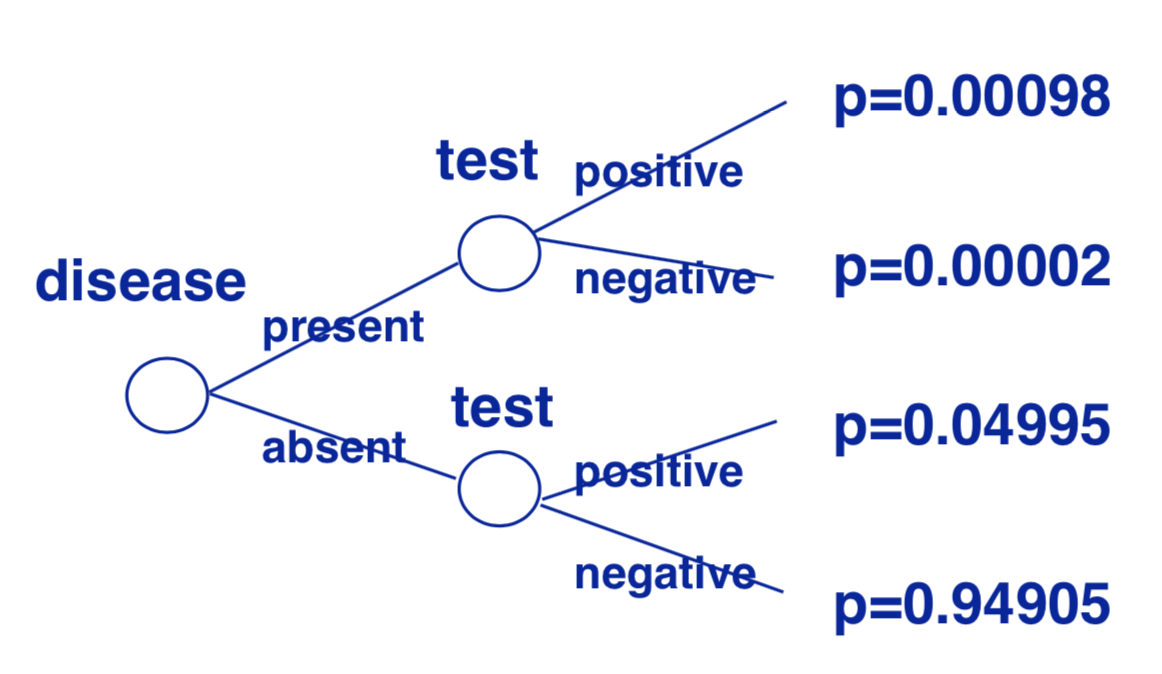
**Probability trees**



On a [tree diagram](https://en.wikipedia.org/wiki/Tree_diagram_(probability_theory)), branch probabilities are conditional on the event associated with the parent node. (Here the overbars indicate that the event does not occur.)

1. example:





What is the probability of the disease present given a positive test result?

P(D|+) = P(D∩+)/P(+) = 0.00098/(0.00098+0.04995) ≈ 0.01924

**\**Sensitivity* and *specificity* are statistical measures of the performance of a binary classification test.**

***Sensitivity*: measures the proportion of actual positives that are correctly identified as such**

***Specificity*: measures the proportion of actual negatives that are correctly identified as such**

2. Trees grow exponentially with the number of variables when branches increases.

Use Use independences among variables in the joint probability distribution to reduce the number of parameters in its representation.

**Factorability**

1. f(X1, X2, ..., Xn) = f(X1 | X2, X3, ..., Xn) f(X2 | X3, ..., Xn) ... f(Xn-2 | Xn-1, Xn) f(Xn-1 | Xn) f(Xn)

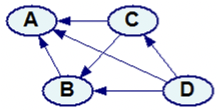
2. Simplified if we consider independencies among variables:

Suppose B⊥D|C, D⊥A|C, and A⊥C

P(A,B,C,D)=P(B|A,C,D) P(D|A,C) P(A|C) P(C) ⇒ P(A,B,C,D)=P(B|A,C) P(D|C) P(A) P(C)

**Bayesian networks**

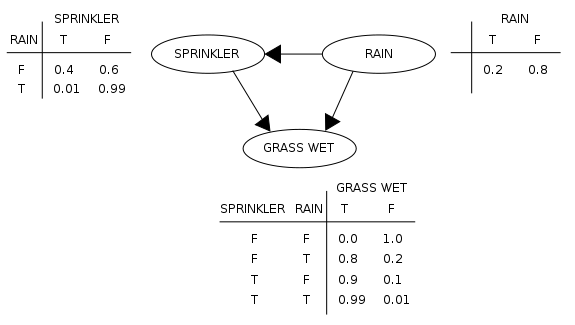
1. P(A,B,C,D)=P(A|B,C,D) P(B|C,D) P(C|D) P(D)



Arc from the conditioning variables to the variables in the factorization

2. Prior probability distribution tables for nodes without predecessors

Conditional probability distributions tables for nodes with predecessors



3. when there are independences in the domain, Bayesian networks are much more efficient than probability tree.

4. Causal Markov condition

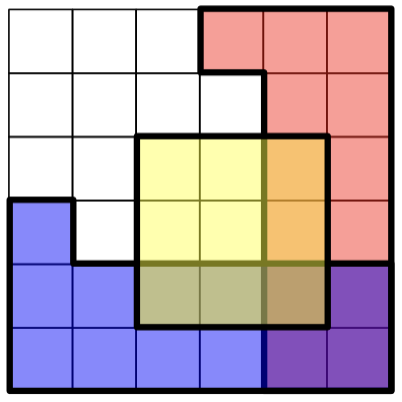
For a Bayesian network states that every node in a Bayesian network is conditionally independent\* of its nondescendents, given its parents.

當一個隨機過程在給定現在狀態及所有過去狀態情況下，其未來狀態的條件機率分布僅依賴於當前狀態；換句話說，在給定現在狀態時，它與過去狀態（即該過程的歷史路徑）是條件獨立的，那麼此隨機過程即具有馬可夫性質(property)。

\*In the standard notation of probability theory, ***R* and *B* are conditionally independent** given *Y* if and only if

P(R∩B|Y) = P(R|Y) P(B|Y)

R, B and Y are represented by the areas shaded red, blue and yellow respectively.

{\displaystyle \Pr(R\mid B\cap Y)=\Pr(R\mid Y).\,} 

5. Directed graphs

Advantages: reflect the causal structure, accommodate representation of uncertainty, can be reconfigured as needed

P70