# Probability and Machine learning

Probability is a field of mathematics that quantifies uncertainty.  
Machine learning is about developing predictive modeling from uncertain data. Uncertainty means working with imperfect or incomplete information.  
There are three main sources of uncertainty in machine learning; they are:

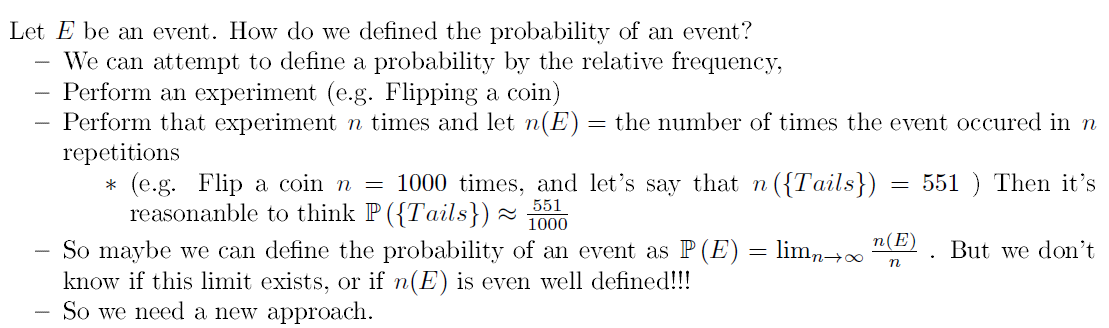
* **Noise in observations**, e.g. measurement errors and random noise.
* **Incomplete coverage of the domain**, e.g. you can never observe all data.
* **Imperfect model of the problem**, e.g. all models have errors, some are useful.

# Probability space (S, F, P) = (Sample space S, events F, probability P)

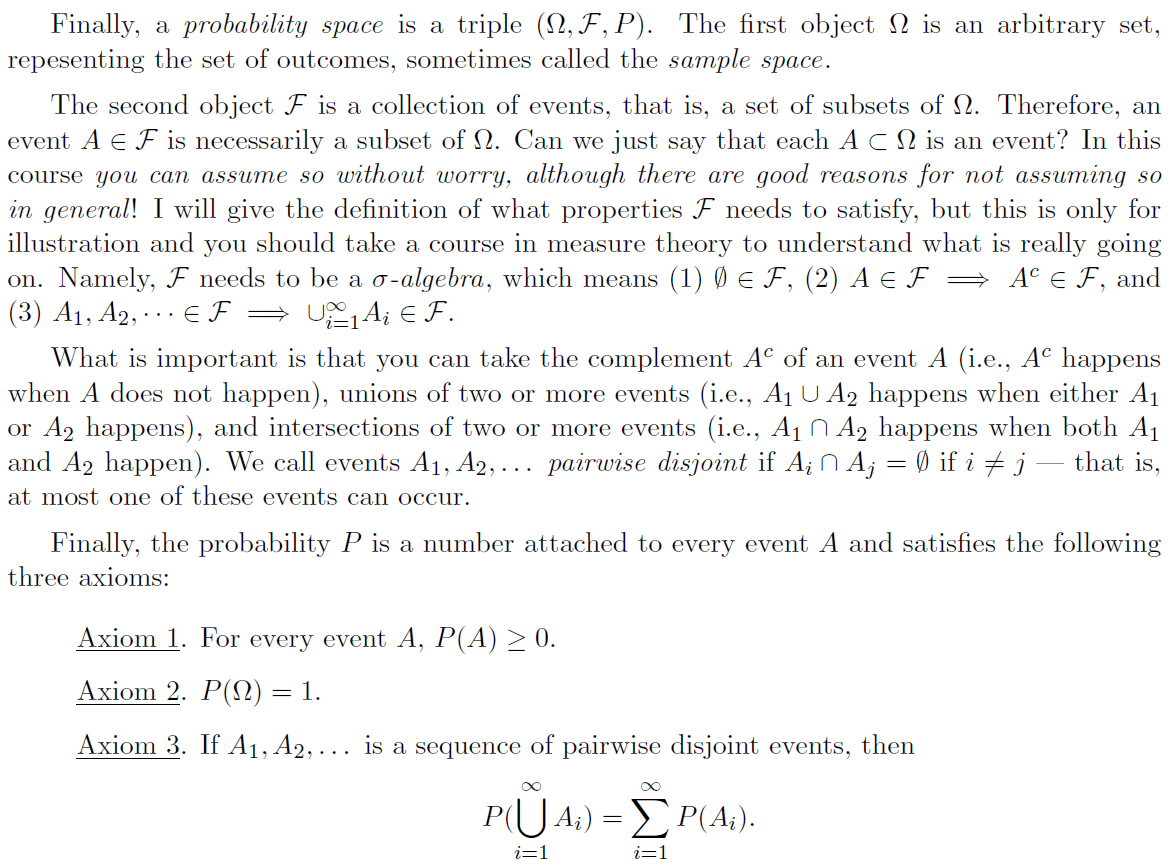
A sample space, denoted by S in the followings, is a set that consists of all possible outcomes from an experiment.

An event A is a subset of S: A ⊂ S. The collection of subsets of S must be a sigma-algebra, which is a collection of subsets of S that is closed under complement and countable union.

How to define the probability of an event?



And now:



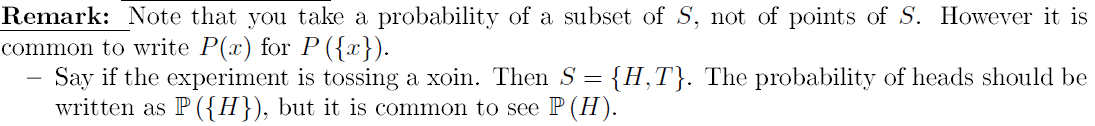
In Probability we say that events A and B are mutually exclusive if they are disjoint: A ∩ B = ∅

# Probability model

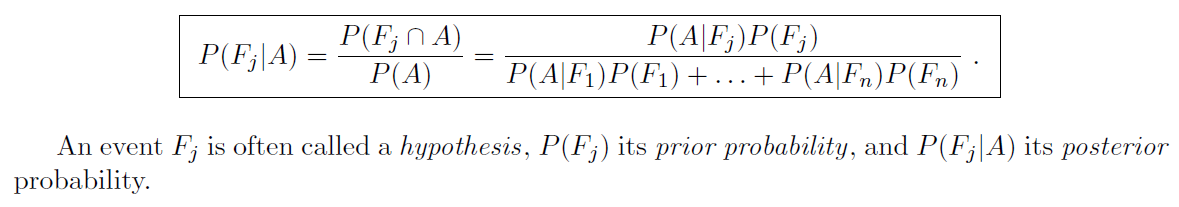
A **probability model** is a mathematical description of an experiment listing all possible outcomes and their associated probabilities.

A ***probability model*** is a mathematical representation of a random phenomenon. It is defined by its ***sample space***, ***events*** within the sample space, and ***probabilities*** associated with each event.

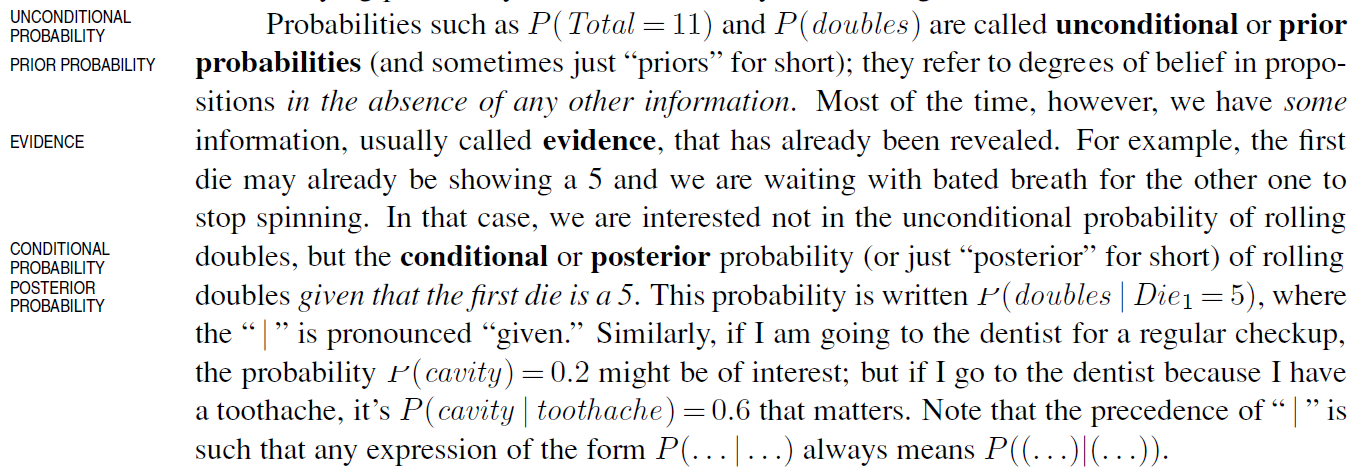
# Notation: probability is a function on events (subset of sample space), not on outcomes (element of sample space)



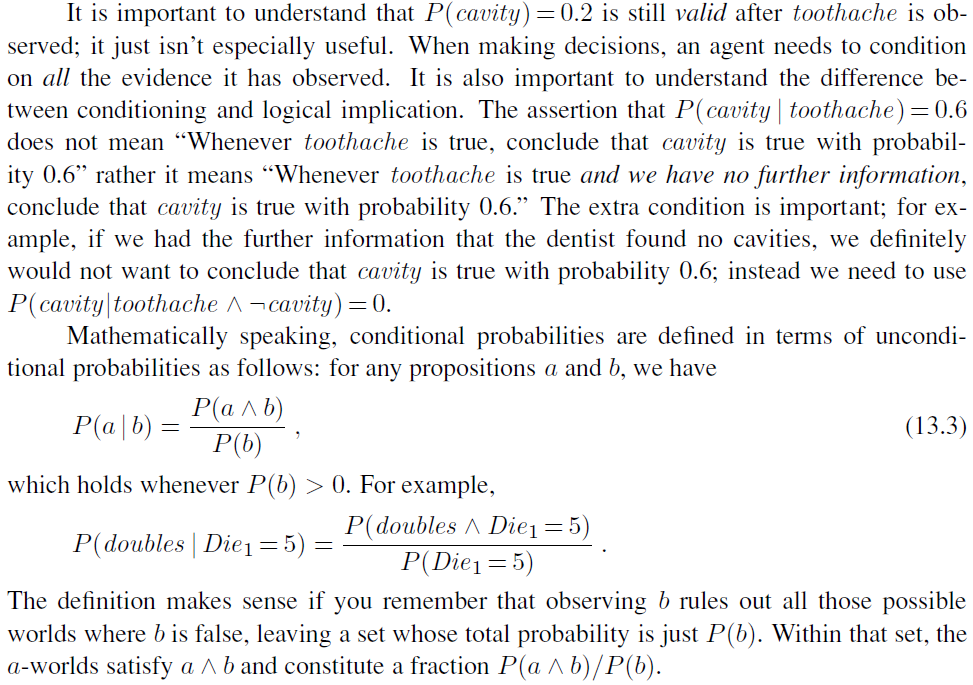
# prior (unconditional prob), evidence, posterior (conditional prob)



the followings discusses tossing two dices



# Conditional probability



# Joint, marginal, conditional probability

* Joint probability is the probability of two events occurring simultaneously. Exp P(AB)
* Marginal probability is the probability of an event irrespective of the outcome of another variable.

Exp: P(A) = P(A|B)P(B) + P(A|notB).P(notB)

* Conditional probability is the probability of one event occurring in the presence of a second event.

Exp: P(A|B) = P(AB)/P(B) = P(B|A).P(A)/P(B)

# Joint Probability Distribution (Discrete Rand Var), Probabilistic Inference

Consider Discrete Random Variable X, Y on a same sample space

Probability Distribution: P(X = ai) for all

Joint Probability Distribution: for all

From this (Joint) Probability Distribution, one can calculate other conditional, marginal probabilities; this process is called probabilistic inference.

# Normalization constant for conditional probability

Assume

is called the normalization constant for condition probability:

Since

# P(X|Y) is different from P(XY), which should be written as P(X⋀Y) or P(X⋂Y)

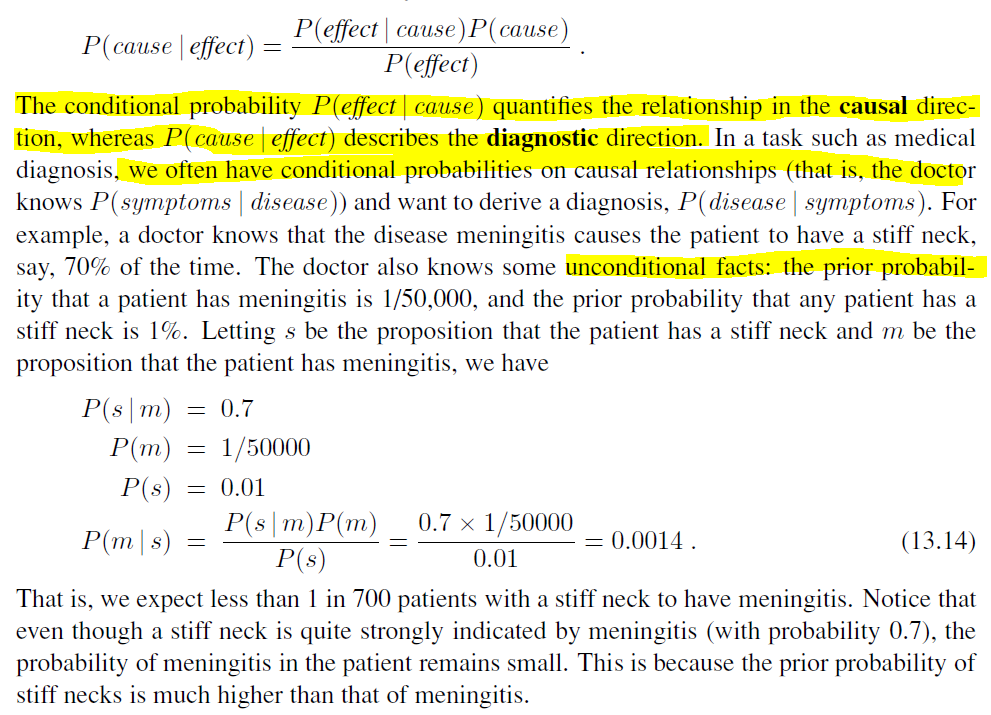
P(XY) probability for both X and Y happen so P(XY) = P(YX)

P(X|Y) probability for X happens assuming that Y already happens

P(XY) = P(X|Y).P(Y)

(from here: Bayes rule: P(X|Y).P(Y) = P(Y|X).P(X))

# Bayes rule: to calculate posterior (conditional prob of hypothesis given evidence) from prior (unconditional probability)



# P(X|Y) = 1 – P(⌐X| Y) but P(X|Y) != 1 – P(X| ⌐Y); also P(X) != P(X|Y) + P(X|⌐Y)

P(X|Y) = 1 – P(⌐X| Y) but P(X|Y) != 1 – P(X| ⌐Y)

P(X) != P(X|Y) + P(X|⌐Y)

But P(X) = P(XY) + P(X⌐Y) and hence P(X) = P(X|Y).P(Y) + P(X|⌐Y).P(⌐Y)

So **P(X) != P(X|Y) + P(X|⌐Y) but P(X) = P(XY) + P(X⌐Y)**

# Conditional independence vs Absolute independence

Notation: C is having cancer, T1, T2 are tests for cancer

**Conditional independence**

Test T1, T2 are said to be conditionally independent given C (having cancer), denoted T1 ⊥ T2 | C if

T1 ⊥ T2 | C ⬄ P(T1⋂T2 | C) = P(T1|C).P(T2|C)

Or equivalently:

T1 ⊥ T2 | C ⬄ P(T1|C) = P(T1|T2⋂C) (i.e. knowing test result T2 doesn’t affect T1) or P(T2|C) = 1

(note that P(T2|C) = 1 then P(T1|T2⋂C) = P(T1|C))

**(Absolute) independence**

Test T1, T2 are said to be independent (or marginally independent, absolutely independent), denoted T1 ⊥ T2 if

T1 ⊥ T2 ⬄ P(T1⋂T2) = P(T1).P(T2)

Note that: A ⊥ B ⇏ A ⊥ B |C and also A ⊥ B |C ⇏ A ⊥ B

# Conditional probability with 3 events: P(A|BC).P(B|C) = P(AB|C)

P(A|BC).P(B|C) = P(AB|C)

P(A|B) = P(A|BC).P(C) + P(A|B⌐C).P(⌐C)

<https://www.edureka.co/blog/bayesian-networks/>

<https://pomegranate.readthedocs.io/en/latest/BayesianNetwork.html>

<https://ermongroup.github.io/cs228-notes/representation/directed/>

# 4 probability distributions: (discrete) Empirical, Poisson, Bernoulli (binomial), (continuous) Gauss

# Bayes network is a graph of conditional probabilities that can calculate any joint probability

Bayes network is a graph that represent conditional probabilities; from a Bayes network, one can calculate any joint probabilities.

Bayes network is a **directed** **ACYCLIC** graph:

* Nodes are random variables
* Edge from to