<http://www.bioinformatics.org/wiki/Hidden_Markov_Model>

Hidden Markov Model

Markov chains are named for Russian mathematician [Andrei Markov](http://en.wikipedia.org/wiki/Andrei_Markov) (1856-1922), and they are defined as observed sequences. A Markov model is a system that produces a Markov chain, and a hidden Markov model is one where the rules for producing the chain are unknown or "hidden." The rules include two probabilities: (i) that there will be a certain observation and (ii) that there will be a certain state transition, given the state of the model at a certain time.[[1]](http://www.bioinformatics.org/wiki/Hidden_Markov_Model#cite_note-dugad-0)

The Hidden Markov Model (HMM) method is a mathematical approach to solving certain types of problems: (i) given the model, find the probability of the observations; (ii) given the model and the observations, find the most likely state transition trajectory; and (iii) maximize either *i* or *ii* by adjusting the model's parameters. For each of these problems, algorithms have been developed: (i) [Forward-Backward](http://www.bioinformatics.org/w/index.php?title=Forward-Backward&action=edit&redlink=1), (ii) [Viterbi](http://www.bioinformatics.org/w/index.php?title=Viterbi&action=edit&redlink=1" \o "Viterbi (page does not exist)), and (iii) [Baum-Welch](http://www.bioinformatics.org/w/index.php?title=Baum-Welch&action=edit&redlink=1) (and the [Segmental K-means](http://www.bioinformatics.org/w/index.php?title=Segmental_K-means&action=edit&redlink=1) alternative).[[1]](http://www.bioinformatics.org/wiki/Hidden_Markov_Model#cite_note-dugad-0)[[2]](http://www.bioinformatics.org/wiki/Hidden_Markov_Model#cite_note-1)

The HMM method has been traditionally used in signal processing, speech recognition, and, more recently, [bioinformatics](http://www.bioinformatics.org/wiki/Bioinformatics). It may generally be used in [pattern recognition](http://www.bioinformatics.org/w/index.php?title=Pattern_recognition&action=edit&redlink=1) problems, anywhere there may be a model producing a [sequence of observations](http://www.bioinformatics.org/w/index.php?title=Sequence_of_observations&action=edit&redlink=1). In bioinformatics, it has been used in [sequence alignment](http://www.bioinformatics.org/wiki/Sequence_alignment), [*in silico* gene detection](http://www.bioinformatics.org/w/index.php?title=In_silico_gene_detection&action=edit&redlink=1), [structure prediction](http://www.bioinformatics.org/wiki/Structure_prediction), [data-mining literature](http://www.bioinformatics.org/w/index.php?title=Data-mining_literature&action=edit&redlink=1), and so on.

Here is a simple example of the use of the HMM method in *in silico* gene detection:

* [Codons](http://www.bioinformatics.org/w/index.php?title=Codons&action=edit&redlink=1) (or [DNA triplets](http://www.bioinformatics.org/w/index.php?title=DNA_triplets&action=edit&redlink=1)) are the observations.
  + 64 codons equals a probability of 1/64.
* The DNA sequence is the Markov chain (set of observations).
* Switches from one genomic region to another are the state transitions.
  + 5'[UTR](http://www.bioinformatics.org/w/index.php?title=UTR&action=edit&redlink=1) to [CDS](http://www.bioinformatics.org/w/index.php?title=CDS&action=edit&redlink=1)
  + CDS to [intron](http://www.bioinformatics.org/w/index.php?title=Intron&action=edit&redlink=1)
  + intron to CDS
  + CDS to 3'UTR
  + 4 state transitions equals a probability of ¼.

Difficulties with the HMM method include the need for accurate, applicable, and sufficiently sized [training sets](http://www.bioinformatics.org/w/index.php?title=Training_set&action=edit&redlink=1) of data. As for the example of gene detection, in order to accurately predict genes in the [human genome](http://www.bioinformatics.org/w/index.php?title=Human_genome&action=edit&redlink=1), many genes in the genome must be accurately known.