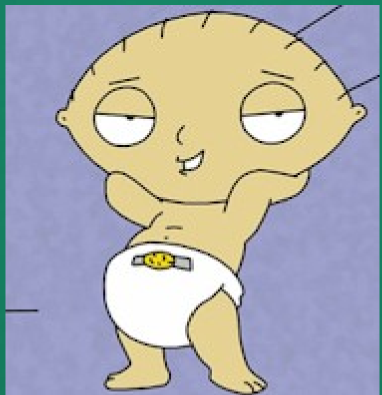
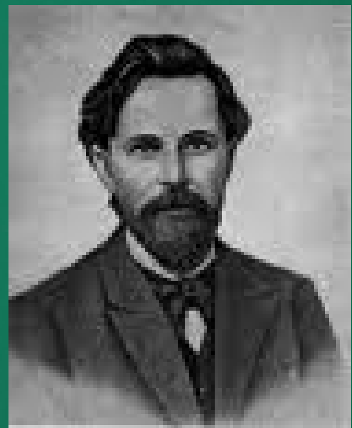
Journey to Markovian Lands (visiting some Bayesian learning realms too): Jukka Corander - The journey Guide

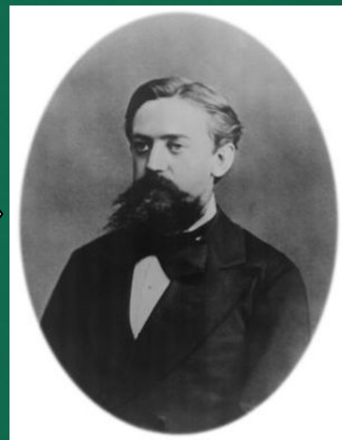
The two key elements in Markov models are:

- Transitions between states of random quantities
- Separation of random quantities in terms of probabilistic independence

TRANSITION...







...and the PROBABILITIES that govern it!

SEPARATION...

...and the STRUCTURES that govern it!

The key element in Bayesian learning is:
Use laws of probability to handle uncertainty!
Uncertainty may be about current data, future data, model structures etc – even about deterministic things!

We will scoop models like:

- Discrete time Markov chains
- Continuous time Markov chains
- Hidden Markov models
- Mixtures of models
- Variable order Markov chains
- Graphical Markov models

In general, mathematical graphs are invaluable tools for understanding various types of Markovian models!

Hence, we will use graphs!

Let's consider some processes,
right here, right now...

- Consider the seats on each row of the classroom.
- We ignore the unoccupied seats.
- Starting from the left end of the rows, we consider each (occupied) seat sequentially.
- If you are wearing glasses, please stand up.
- The observation we record for each case is whether the student in question is sitting (0) or standing (1).

- This gives us sequences of binary quantities $x_i \in \{0,1\}$, with prior uncertainty about the value each x_i is going to take.
- We notice that x_{i-1} does not directly influence the value x_i is going to take.
- The general tendency to wear glasses is relevant for the outcome of the process (HOW??).
- Does it matter if we just see the outcomes, but don't know the generating mechanism?

Let's modify the process a bit…

- Consider again the seats on each row.
- Starting from the left end of the rows, we consider each (occupied) seat sequentially.
- If you are wearing glasses, please stand up.
- If you are not already standing up, stand up if the person to your left is standing AND you are taller than she/he is.
- We record for each case again whether you are sitting (0) or standing (1).

- What has changed in the process?
- Now the state of x_{i-1} influences the value x_i will take (in terms of probability distribution).
- Note that state of x_{i-2} does not directly influence the value of x_i, but the effect is mediated by x_{i-1} (recall the separation in the NHL snapshot).
- Separation like this lies at the very heart of Markov chains.

- Memory of the process could also be longer.
- For instance, we could use a definition like this:
- If you are not already standing up, stand up if the person to your left OR next to her/him is standing AND you are taller than either of them.
- Now the state of x_{i-2} enters the picture together with x_{i-1} when the value of x_i becomes realized.

We can still generalize the process...

- Consider again the seats on one particular row, say the front row.
- Starting from the left end of the rows, we consider each (occupied) seat sequentially.
- If you are wearing glasses, please stand up.
- If you are not already standing up, stand up if the person to your left is standing AND you are taller than she/he is.

- Now, every person in the row behind the front row will either raise one or two arms.
- If the person in front of you is sitting AND you are taller than that person, raise one arm (you can ask for the height).
- If NOT taller or the person in front of you is standing, raise two arms.
- We record a sequence $\{y_i\}$ of binary quantities $y_i \in \{1,2\}$, which correspond to the number of arms raised.

- Does the state of y_i depend on the value of y_{i-1}?
- A graphical characterization of the observation process reveals that there is no direct dependence between y_i and y_{i-1}.
- The observed states are independent GIVEN the hidden, unobserved states $\{x_i\}$.
- A two-stage process such as this one opens the realm of Hidden Markov Models (HMM) to us.

Things seem to be rolling well, so let's have look at even weirder processes...

- Consider again the seats on each row of the classroom.
- Starting from the left end of the rows, we consider each (occupied) seat sequentially.
- To consider whether you (at seat i) should stand up or stay down, we first need to flip a coin, which can be done virtually at http://www.random.org/coins/, and let's store a sequence of flips before continuing.

- Consider now the person at seat i.
- If 'your' coin flip gave HEADS, then stand up if you are wearing glasses.
- If 'your' coin flip gave TAILS, then stand up if your hair is fair.
- What are these sequences of observations?
- They are generated by a mixture of two distinct mechanisms, with a random switch between them.
- Believe it or not, even weirder things happen in genomes...

- Mixture models can be put also in the context of sequences where the state of x_{i-1} influences the value x_i will take.
- Let's have a look...

- If 'your' coin flip gave HEADS, then stand up if you are wearing glasses.
- Further, if you are not already standing up, stand up if the person to your left is standing AND you are taller than she/he is.
- If 'your' coin flip gave TAILS, then stand up if your hair is fair.
- Further, if you are not already standing up, stand up if the person to your left is sitting AND you are shorter than she/he is.

We see that things can get really hairy in this context...

Finally, some words about Bayesian learning...

- *Learning* refers generally to the procedure where we try to infer something from a batch of *training* information.
- We can learn:
- about *models* (in structural meaning),
- about *model parameters* (the quantitative specification of a model),
- about what's to happen in the future (predictions)

- *Bayesian learning* refers generally to a procedure where we learn by using formally the laws of probability.
- This necessitates that we specify probabilities for things that are uncertain to us, even if they in principle would be deterministic in nature.
- It is important to emphasize that learning SERVES A PURPOSE, we don't do it for just having fun.

- Models for learning will in most cases be just approximations to complex reality.
- It is not particularly useful to think models being 'TRUE' data generating mechanisms, but that they help us to solve problems of various type (i.e. they serve a purpose).
- In Bayesian learning we can often utilize the purpose to set up our models and probability distributions for the unknown things.

An example of learning in the Markov chain context

- Data is a DNA sequence.
- The background variation is described by a higher order Markov chain model.
- We try to detect multiple noisy copies of a WORD in the sequence, i.e. partially conserved patterns of nucleotides.
- The rationale is that probability of such a word occurring many times is low under the background model.

Could look like this:

GCAGCGTATGCAGTTGGATCAATTAGTGGGGCACATTTGAATCCGGCTTTAACGATAG
GACTTGCG(TTTAAG)GGAGCGTTCCCATGGAGTGATGTACCTATGTATATCGCAGCA
CAAATGATTGGGGCAATTATCGGGGCAGTTCTTGTATATTTACATTACTTACCACACT
GGAAAGAAACAGAAGATCCAGGAACAAAGTTAGGTGTATTTGCAACAGGTCCAGCAAT
TCCGAACACA(TTTACA)AACCTTTTAAGTGAAATGATTGGAACATTCGTTTTAGTATT
TGGTATATTAGCAATTGGAGCAAATAAATTTGCAGATGG(TTTAAA)TCCATTTATCGT
AGGTTTCTTAATTGTAAGTATTGGTTTGCAGCGTATGCAGTTGGATCAATTAGTGGGG
CACATTTGAATCCGGC(TTTAAC)GATAGGACTTGCGTTTAAGGGAGCGTTCCCATGG
AGTGATGTACCTATGTATATCGCAGCACAAATGATTGGGGCAATTATCGGGGCAGTTC
TTGTATATTTACATTACTTACCACACTGGAAAGAAACAGAAGATCCAGGAACAAAGTT
AGGTGTATTTGCAACAGGTCCAGCAATTCCGAACACA(TTTACA)AACCTTTTAAGTG
AAATGATTGGAACATTCGTTTTAGTATTTGGTATATTAGCAATTGGAGCAAATAAATTT
GCAGATGGTTTAAATCCATTTATCGTAGGTTTCTTAATTGTAAGTATTGGTTTGCAGC
GTATGCAGTTGGATCAATTAGTGGGGCACATTTGAATCCGGCTTTAACGATAGGACTT
GCTTTTAAGGGAGCGTTCCCATGGAGTGATGTACCTATGTATATCGCAGCACAAATGA
TTGGGGCAATTATAGGGGCAGTTCTTGTATATTTACATTACTTACCACACTGGAAAGA
AACAGAAGATCCAGGAACAAAGTTAGGTGTGTTTGCAACAGGTCCAGCAATTCCGAAC
ACATTTACAAACCT(TTTAAG)TGAAATGATTGGAACATTCGTTTTAGTATTTGGTATA

End of teaser trailer...