Introduction to Bioinformatics

JTMS-19

Marc-Thorsten Hütt mhuett@constructor.university
Felix Jonas fjonas@constructor.university

Hidden Markov models II

What is this session about?

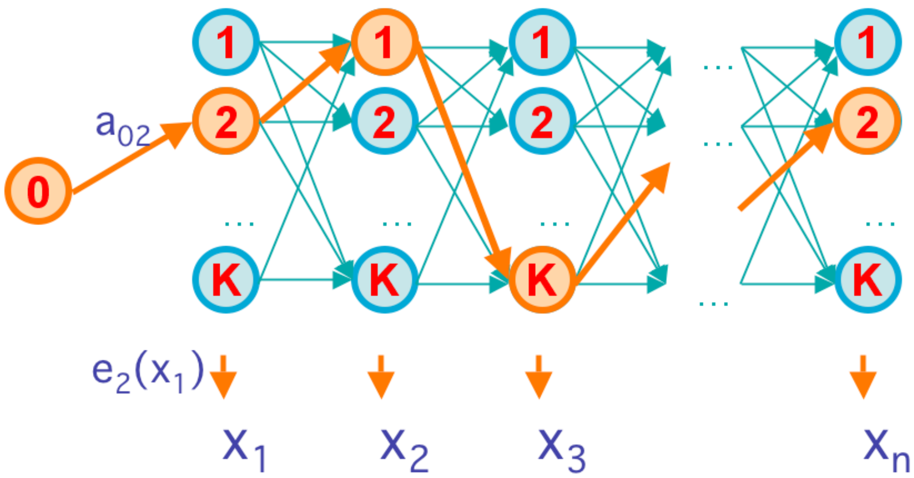
Posterior decoding for HMMs is introduced. First applications of HMMs are discussed.

How can you revise the material after the session?

Read Durbin et al. chapters 3.3, 3.4
Read Baxevanis/Oullette pages 208 – 210
Look at the HMM in Kundaje, et al. (2015). Nature, 518, 317-330.
alternative reading: Hütt/Dehnert chapters 2.8.3 – 2.8.4, 2.9

general properties of HMM

simulation of a sequence



internal path

observed sequence x

joint probability of the path and the sequence

$$P(x, \pi) = a_{\pi_0 \pi_1} \prod_{i=1}^{L} e_{\pi_i}(x_i)\, a_{\pi_i \pi_{i+1}}$$

Summary of the Viterbi algorithm

initialization ⟶ recursion ⟶ termination ⟶ traceback

$$v_0(0) = 1\,,\;\; v_k(0) = 0 \quad \forall\, k \in \Sigma_{HMM}$$

initialization

$$v_l(i+1) = e_l(x_{i+1})\, \max_k\{v_k(i)\, a_{kl}\}$$

recursion

$$\max_k\{v_k(L)\, a_{k0}\} = P(x, \pi^*)$$

termination

$$a_{k0} = \frac{1}{|\Sigma_{HMM}|} \quad \forall\, k \in \Sigma_{HMM}$$

$$\pi^*_L$$

$$\pi^*_{L-1} = Z_{L-1}(\pi^*_L)$$

$$\vdots$$

$$\pi^*_{i-1} = Z_{i-1}(\pi^*_i)$$

$$\vdots$$

traceback

A small numerical example

$x = 3, 1, 5, 6, 6$ $a_{FF} = a_{UU} = 0.8$
$a_{UF} = a_{FU} = 0.2$

Summary of the Viterbi algorithm

initialization → recursion → termination → traceback

$$v_0(0) = 1\ ,\ \ v_k(0) = 0 \quad \forall\, k \in \Sigma_{HMM}$$ initialization

$$v_l(i+1) = e_l(x_{i+1})\ \max_k\{v_k(i)\, a_{kl}\}$$ recursion

$$\max_k\{v_k(L)\, a_{k0}\} = P(x, \pi^*)$$ termination

$$a_{k0} = \frac{1}{|\Sigma_{HMM}|} \quad \forall\, k \in \Sigma_{HMM}$$

$$\pi_L^*$$
$$\pi_{L-1}^* = Z_{L-1}(\pi_L^*)$$
$$\vdots$$
$$\pi_{i-1}^* = Z_{i-1}(\pi_i^*)$$
$$\vdots$$

traceback

A small numerical example

$x = 3, 1, 5, 6, 6$ $\quad a_{FF} = a_{UU} = 0.8$
$a_{UF} = a_{FU} = 0.2$

$$\begin{aligned} v_F(1) &= e_F(3)\ \max\left\{v_0(0)\,a_{0F},\, v_F(0)\,a_{FF},\ v_U(0)\,a_{UF}\right\} \\ &= \frac{1}{6}\ \max\{1\cdot 0.5,\ 0\cdot 0.8,\ 0\cdot 0.2\} = \frac{1}{12} \approx 0.0833 \end{aligned}$$

$$\begin{aligned} v_U(1) &= e_U(3)\ \max\left\{v_0(0)\,a_{0U},\, v_F(0)\,a_{FU},\ v_U(0)\,a_{UU}\right\} \\ &= \frac{1}{10}\ \max\{1\cdot 0.5,\ 0\cdot 0.2,\ 0\cdot 0.8\} = \frac{1}{20} = 0.05\ . \end{aligned}$$

$$\begin{aligned} v_F(2) &= e_F(1)\ \max\left\{v_F(1)\,a_{FF},\ v_U(1)\,a_{UF}\right\} \\ &= \frac{1}{6}\ \max\left\{\frac{1}{12}\cdot 0.8,\ \frac{1}{20}\cdot 0.2\right\} \\ &\approx \frac{1}{6}\ \max\{0.067,\ 0.01\} = 0.011 \end{aligned}$$

$$\begin{aligned} v_U(2) &= e_U(1)\ \max\left\{v_F(1)\,a_{FU},\ v_U(1)\,a_{UU}\right\} \\ &= \frac{1}{10}\ \max\left\{\frac{1}{12}\cdot 0.2,\ \frac{1}{20}\cdot 0.8\right\} \\ &\approx \frac{1}{10}\ \max\{0.0167,\ 0.04\} = 0.004\ . \end{aligned}$$

A small numerical example

$x = 3, 1, 5, 6, 6$

$a_{FF} = a_{UU} = 0.8$
$a_{UF} = a_{FU} = 0.2$

$x = 3, 1, 5, 6$

i	$v_F(i)$	$v_U(i)$	$\pi^*(i)$
1	0.0833	0.05	F
2	0.011	0.004	F
3	0.0015	0.00032	F
4	0.0002	0.00015	→ F

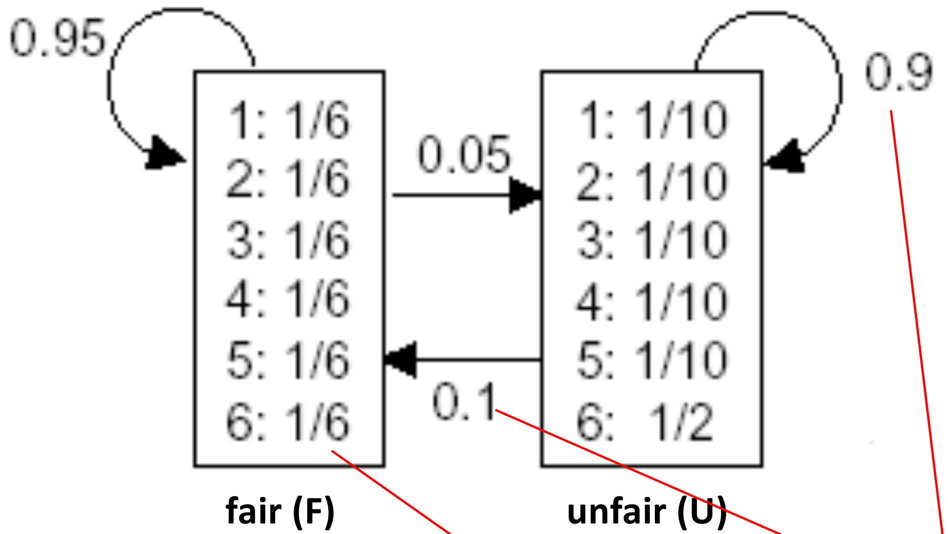
3	1	5	6
F	F	F	F

$x = 3, 1, 5, 6, 6$

i	$v_F(i)$	$v_U(i)$	$\pi^*(i)$
1	0.0833	0.05	F
2	0.011	0.004	F
3	0.0015	0.00032	F
4	0.0002	0.00015	U
5	0.000026	0.000059	→ U

3	1	5	6	6
F	F	F	U	U

elementary example: casino with two dice



emission probabilities

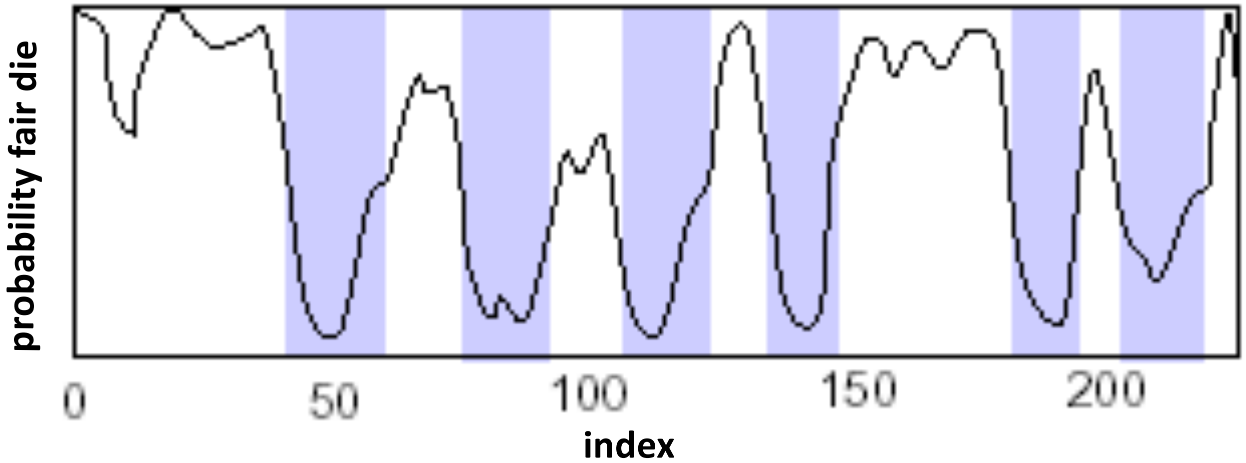
transition probabilities

$\boxed{S_{HMM} = \{ F, U \}}$

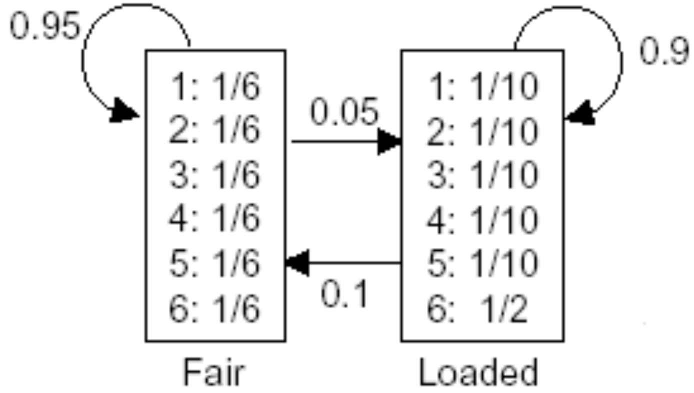
Casino: results

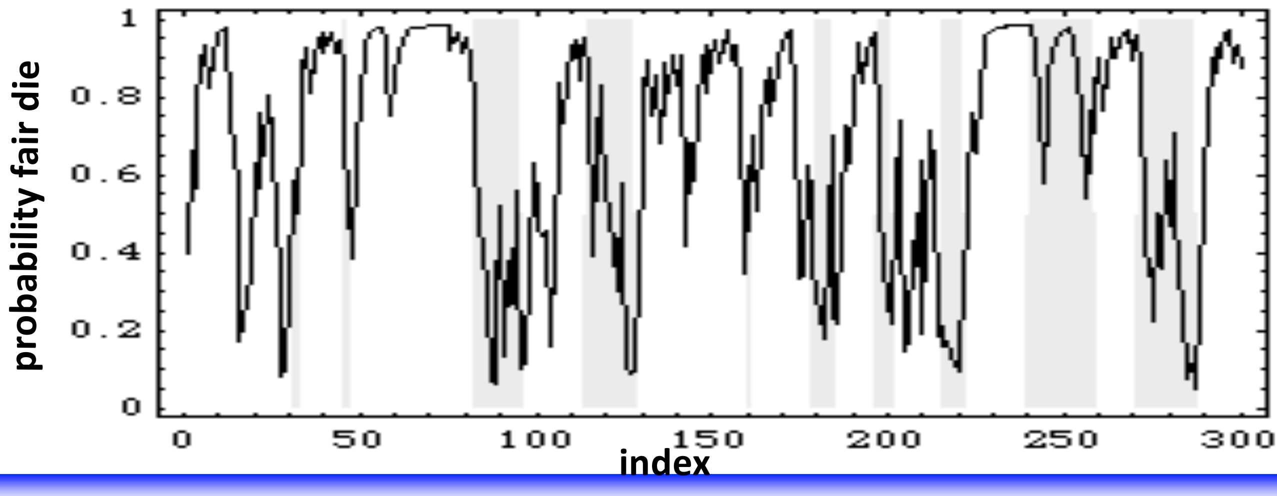
Rolls	315116246446644245321131631164152133625144543631656626566666
Die	FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFLLLLLLLLLLLLLLL
Viterbi	FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFLLLLLLLLLLLL
Rolls	651166453132651245636664631636663162326455235266666625151631
Die	LLLLLLFFFFFFFFFFFFLLLLLLLLLLLLLLLLFFFLLLLLLLLLLLLLLFFFFFFFFF
Viterbi	LLLLLLFFFFFFFFFFFFLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLFFFFFFFF
Rolls	222555441666566563564324364131513465146353411126414626253356
Die	FFFFFFFFLLLLLLLLLLLLLFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFL
Viterbi	FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFL
Rolls	366163666466232534413661661163252562462255265252266435353336
Die	LLLLLLLLFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
Viterbi	LLLLLLLLLLLLFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
Rolls	233121625364414432335163243633665562466662632666612355245242
Die	FFFFFFFFFFFFFFFFFFFFFFFFFFFLLLLLLLLLLLLLLLLLLLLLLFFFFFFFFFFF
Viterbi	FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFLLLLLLLLLLLLLLLLLLLFFFFFFFFFFF

Casino: results

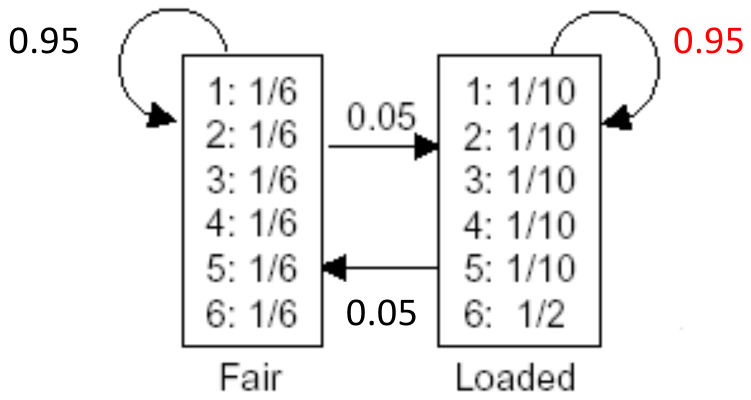


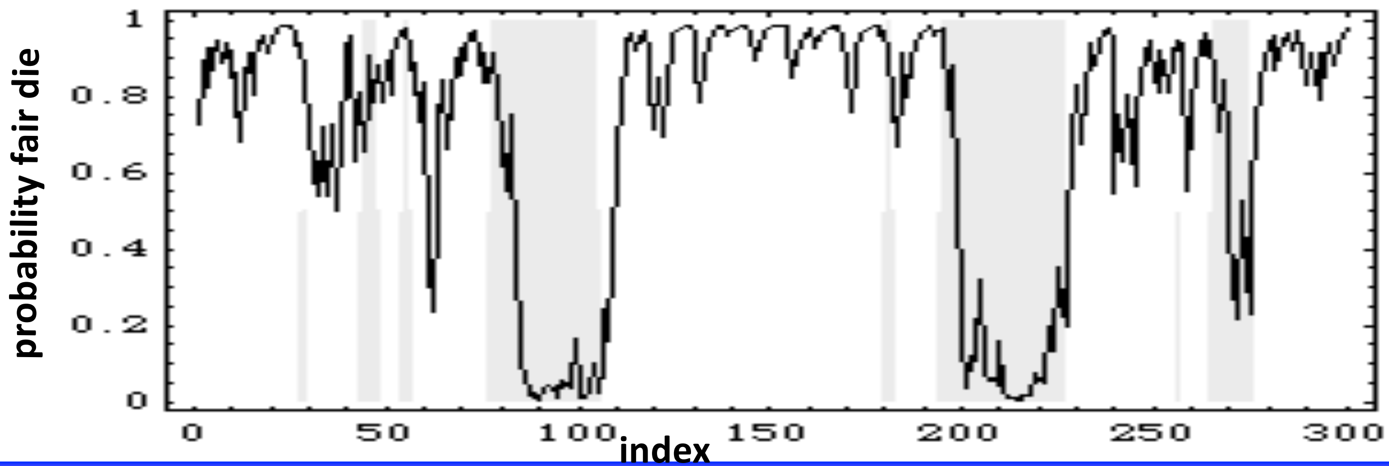
Casino: results



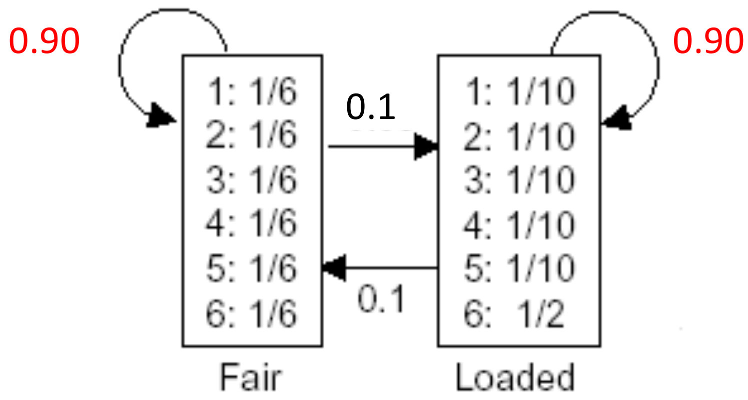


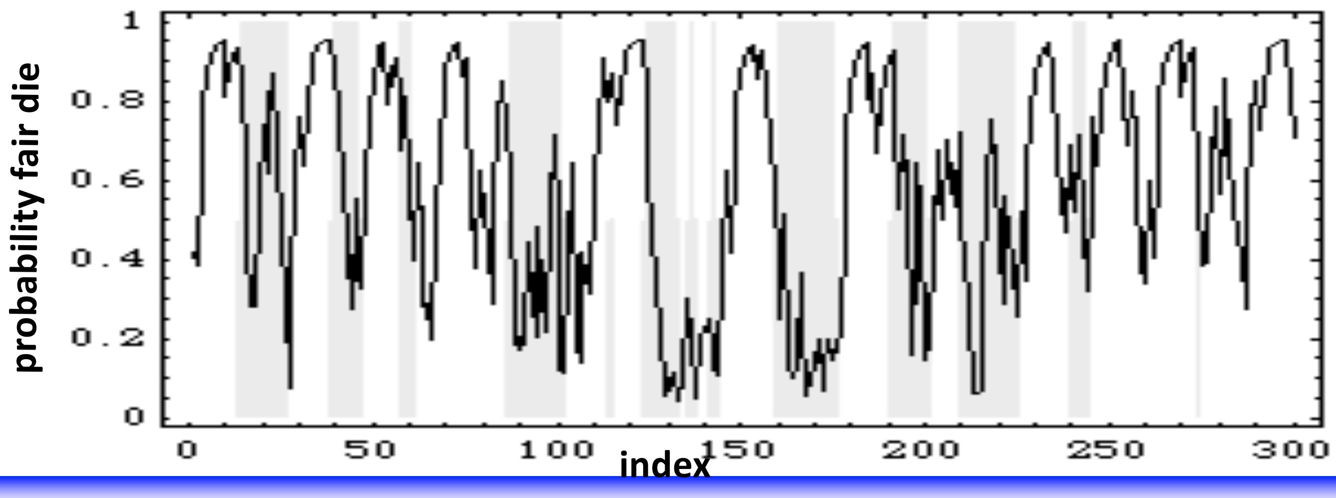
Casino: results





Casino: results





Summary of the forward and backward algorithms

$$P(x, \pi_i = k) = \underbrace{P(x_1, \ldots, x_i, \pi_i = k)}_{\equiv f_k(i)} \; \underbrace{P(x_{i+1}, \ldots, x_L \,|\, x_1, \ldots, x_i, \pi_i = k)}_{= P(x_{i+1}, \ldots, x_L \,|\, \pi_i = k) \equiv b_k(i)}$$

$$f_0(0) = 1 \;, f_k(0) = 0 \;\; \forall k \in \Sigma_{HMM}$$

initialization

$$f_l(i+1) = e_l(x_{i+1}) \sum_k f_k(i) a_{kl}$$

recursion

$$P(x) = \sum_{k \in \Sigma_{HMM}} f_k(L)\, a_{k0}$$

termination

$$a_{k0} = \frac{1}{|\Sigma_{HMM}|} \quad \forall\, k \in \Sigma_{HMM}$$

$$P(x) = \sum_{\pi} P(x, \pi)$$

marginal probability

Summary of the forward and backward algorithms

$$P(x, \pi_i = k) = \underbrace{P(x_1, \ldots, x_i, \pi_i = k)}_{\equiv f_k(i)} \underbrace{P(x_{i+1}, \ldots, x_L \,|\, x_1, \ldots, x_i, \pi_i = k)}_{\substack{= P(x_{i+1}, \ldots, x_L \,|\, \pi_i = k) \\ \equiv b_k(i)}}$$

$$b_k(L) = a_{k0} \quad k \in \Sigma_{HMM}$$ initialization

$$b_k(i) = \sum_{l \in \Sigma_{HMM}} a_{kl} e_l(x_{i+1}) \, b_l(i+1)$$ recursion

$$P(x) = \sum_{l \in \Sigma_{HMM}} a_{0l} e_l(x_1) \, b_l(1)$$ termination

Summary of the forward and backward algorithms

$$P(x, \pi_i = k) = \underbrace{P(x_1, \ldots, x_i, \pi_i = k)}_{\equiv f_k(i)} \, \underbrace{P(x_{i+1}, \ldots, x_L \,|\, x_1, \ldots, x_i, \pi_i = k)}_{= P(x_{i+1}, \ldots, x_L \,|\, \pi_i = k) \equiv b_k(i)}$$

$$P(x,\ \pi_i = k) = f_k(i)\, b_k(i)$$

intermediate result

$$P(x,\ \pi_i = k) = P(\pi_i = k \,|\, x) P(x)$$

definition of the conditional probability

$$P(\pi_i = k \,|\, x) = \frac{f_k(i)\, b_k(i)}{P(x)}$$

final result:
posterior probability of a HMM state *k* at position i given the sequence *x.*

'posterior decoding'