

BIOINFORMATIKA – CV. 2

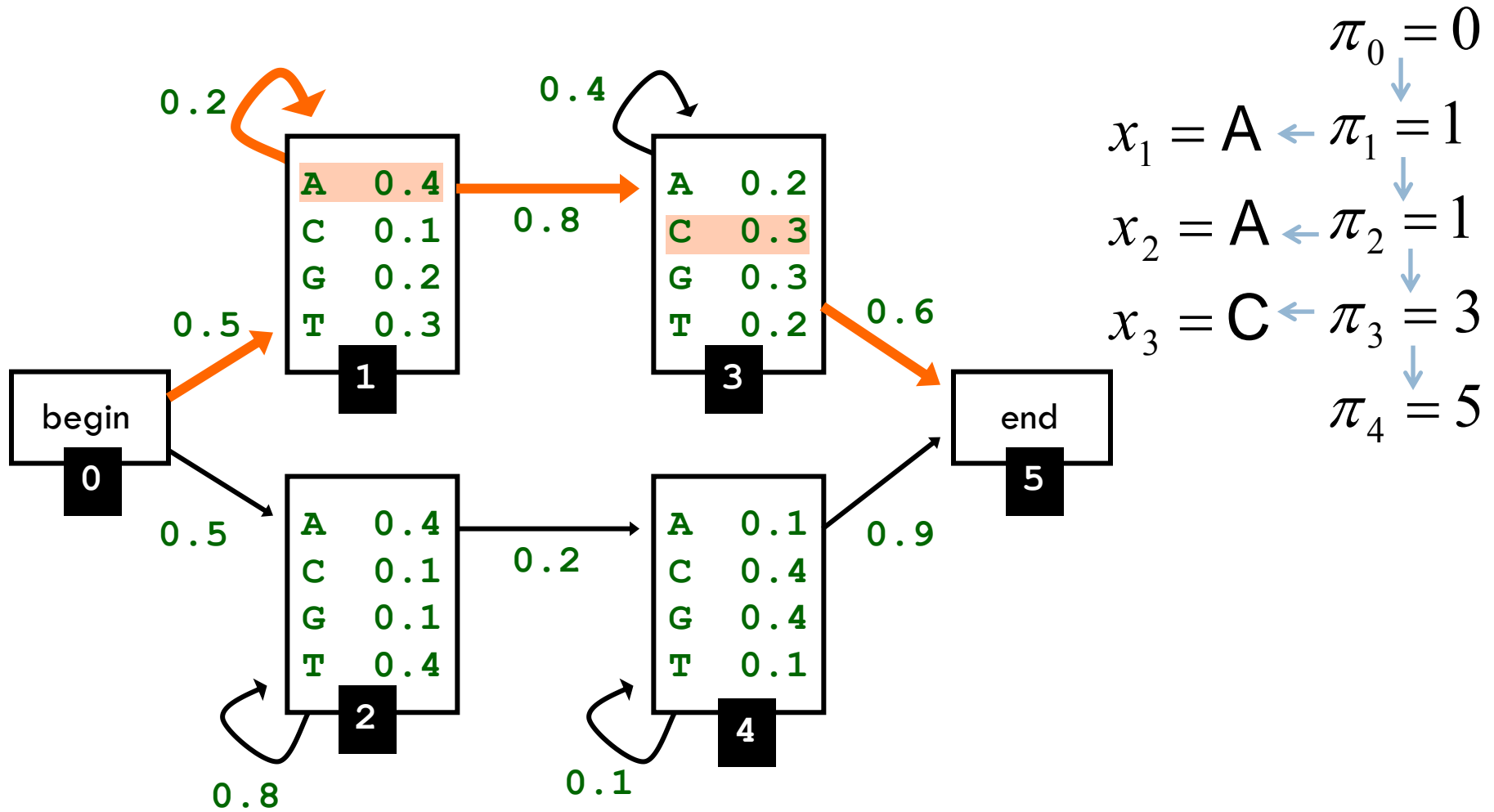
Skryté markovské modely

(Some slides are courtesy of Mark Craven, U. of Wisconsin)

Motivace

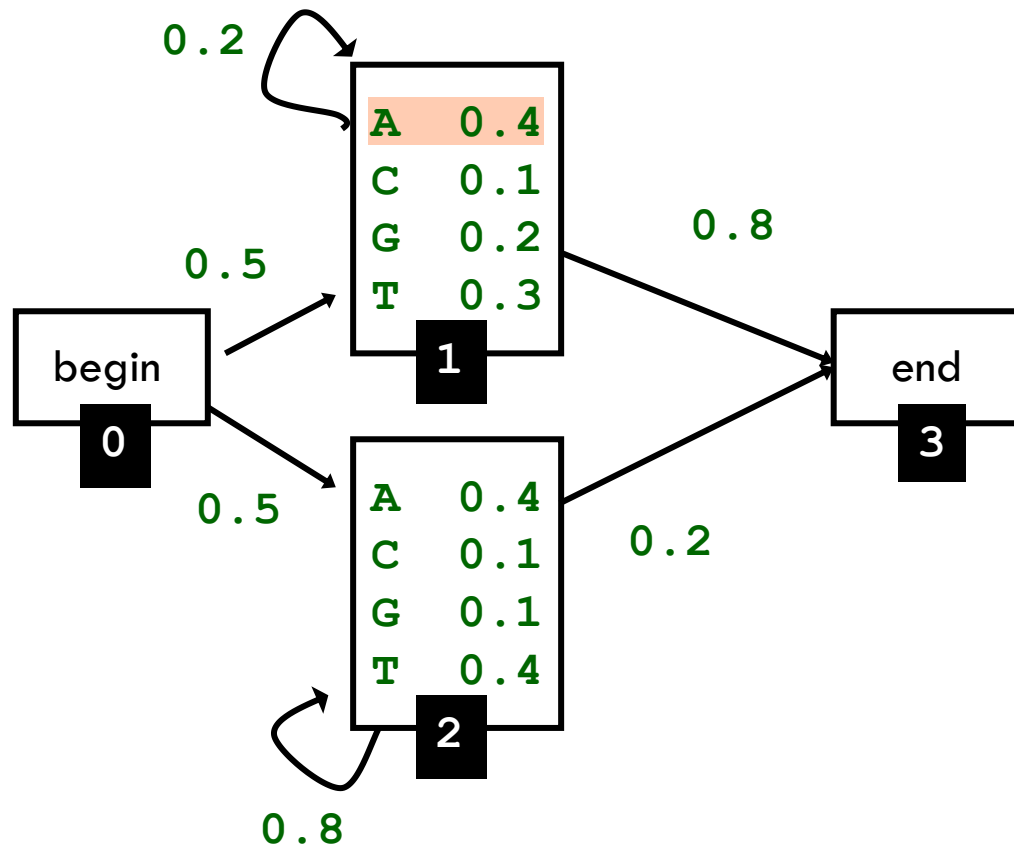
- Detekce CpG ostrovy (viz tabule)
- Detekce genů v DNA
- Detekce „binding-regions“ proteinů

Notace



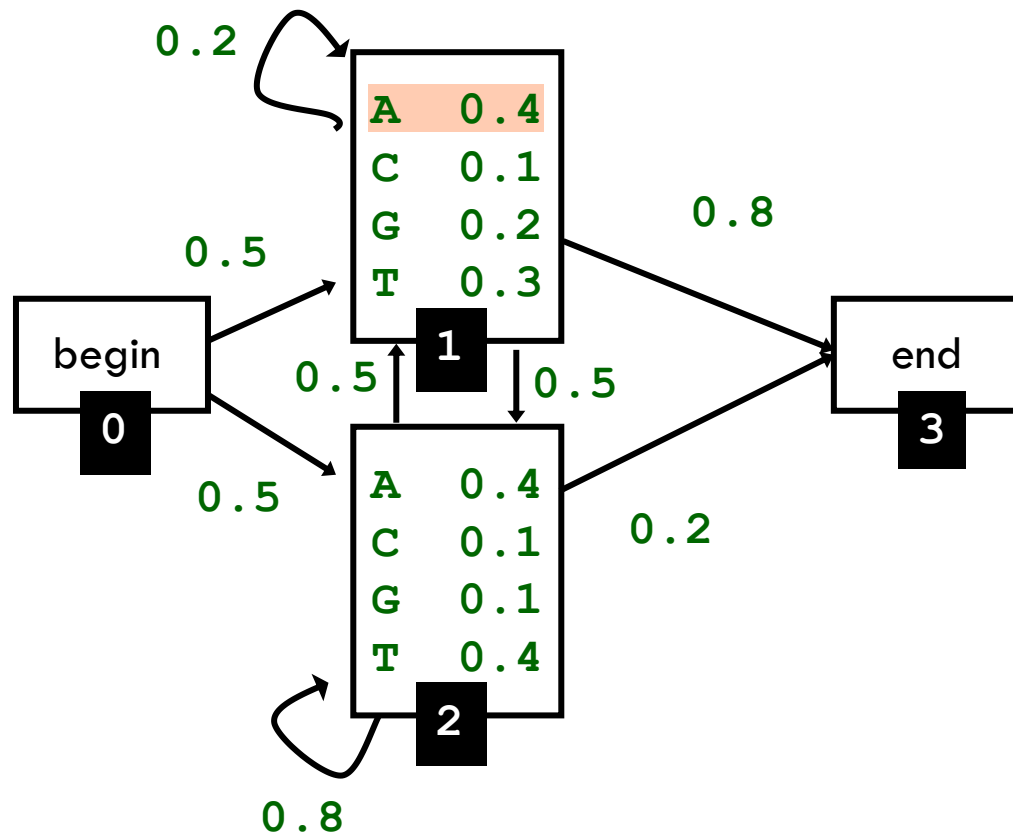
Jednoduchý příklad

- Jaká je pravděpodobnost sekvence AACCC? Vypočítejte ji pomocí výčtu všech relevantních cest (= naivní přístup).



Trošku těžší příklad

- Dokázali byste spočítat pravděpodobnost sekvence AACCC naivním způsobem pro následující model? Kolik cest byste museli vzít v úvahu? Vypište je všechny – pravděpodobnost počítat nemusíte.



Připomínka: Dynamického programování

Fibonacciho čísla:

$$F(0) = 0$$

$$F(1) = 1$$

$$F(n) = F(n - 1) + F(n - 2)$$

1. Vypočítejte $F(8)$.

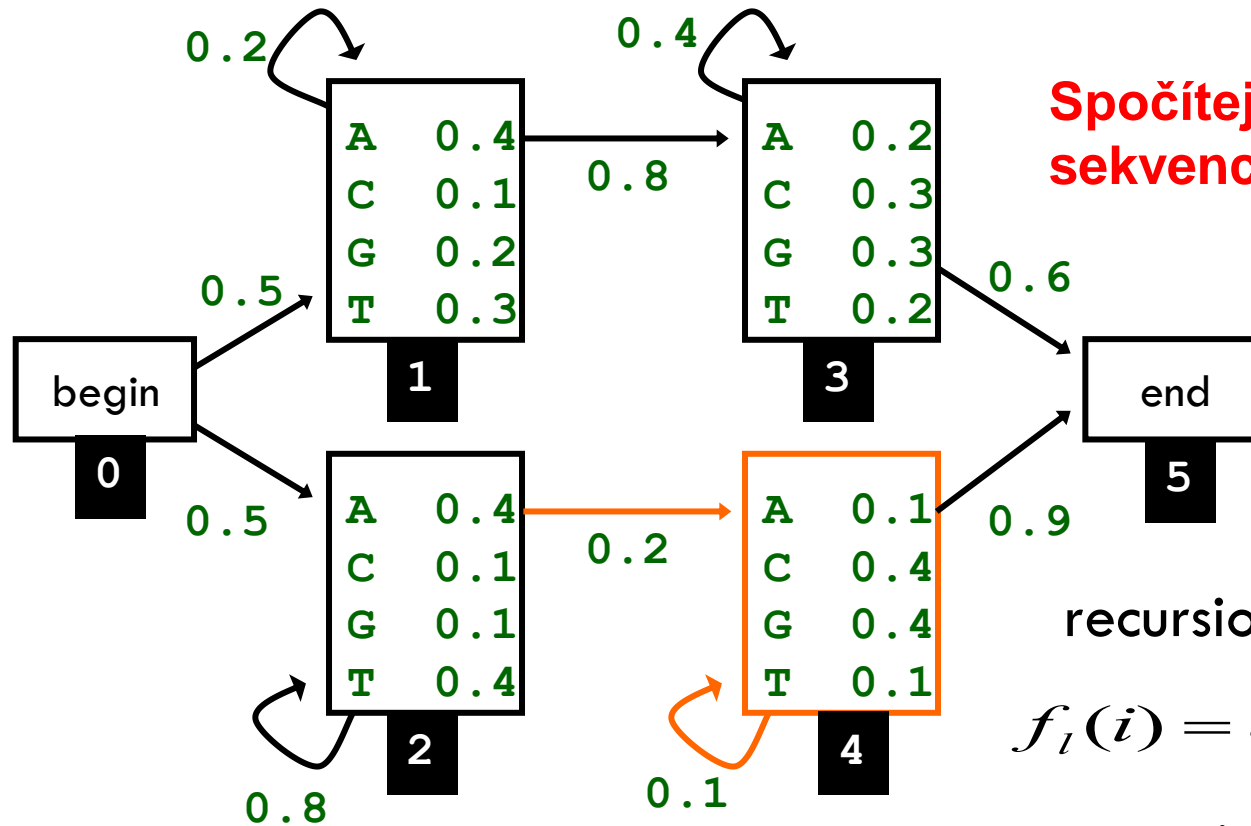
2. Co by se stalo, kdybyste výpočet provedli přesně podle následujícího pseudokódu?

function fib(n)

if $n == 0$ or $n == 1$ **then return** n ;

else return fib($n-1$)+fib($n-2$)

Ještě těžší příklad – Forward algorithm (1)



Spočítejte pravděpodobnost sekvence TAG

recursion for emitting states:

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

recursion for silent states:

$$f_l(i) = \sum_k f_k(i) a_{kl}$$

Ještě těžší příklad – Forward algorithmus (2)

□ Sekvence: **TAG**

□ Inicializace:

$$f_0(0) = 1 \quad f_1(0) = 0 \quad \dots \quad f_5(0) = 0$$

• Rekurzivní výpočet:

$$f_1(1) = e_1(T) \times (f_0(0)a_{01} + f_1(0)a_{11}) = \\ 0.3 \times (1 \times 0.5 + 0 \times 0.2) = 0.15$$

$$f_2(1) = 0.4 \times (1 \times 0.5 + 0 \times 0.8)$$

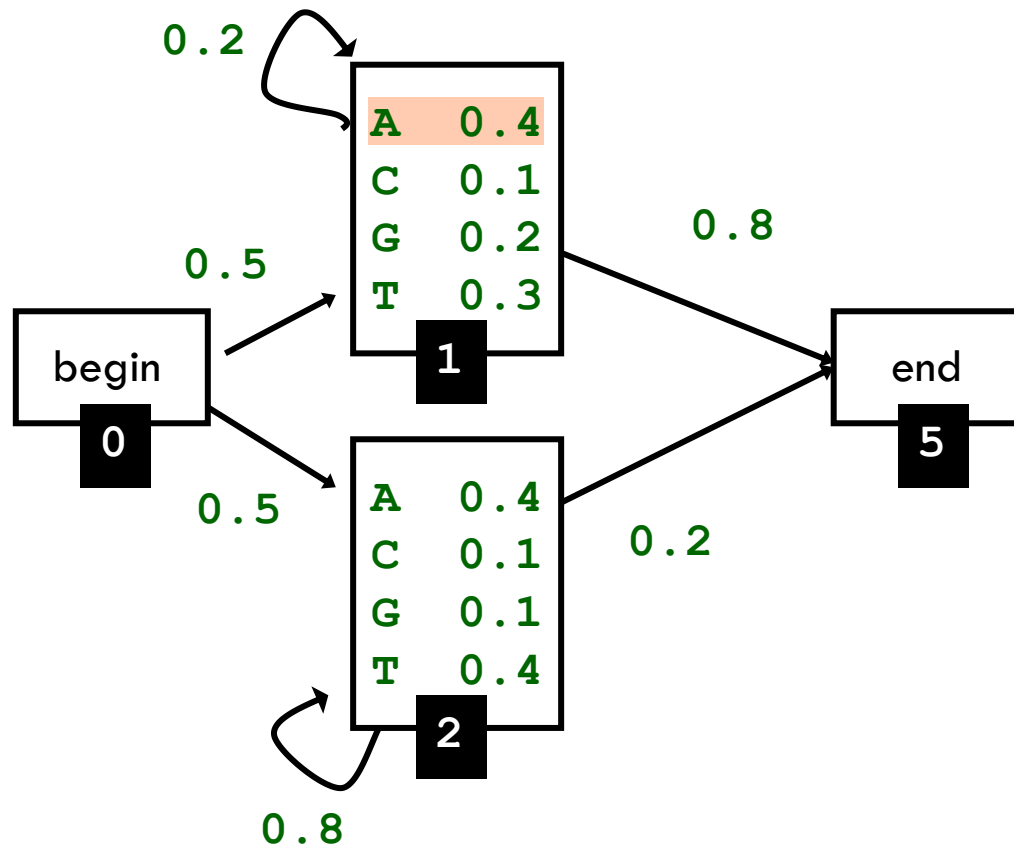
$$f_1(2) = e_1(A) \times (f_0(1)a_{01} + f_1(1)a_{11}) = \\ 0.4 \times (0 \times 0.5 + 0.15 \times 0.2)$$

•••

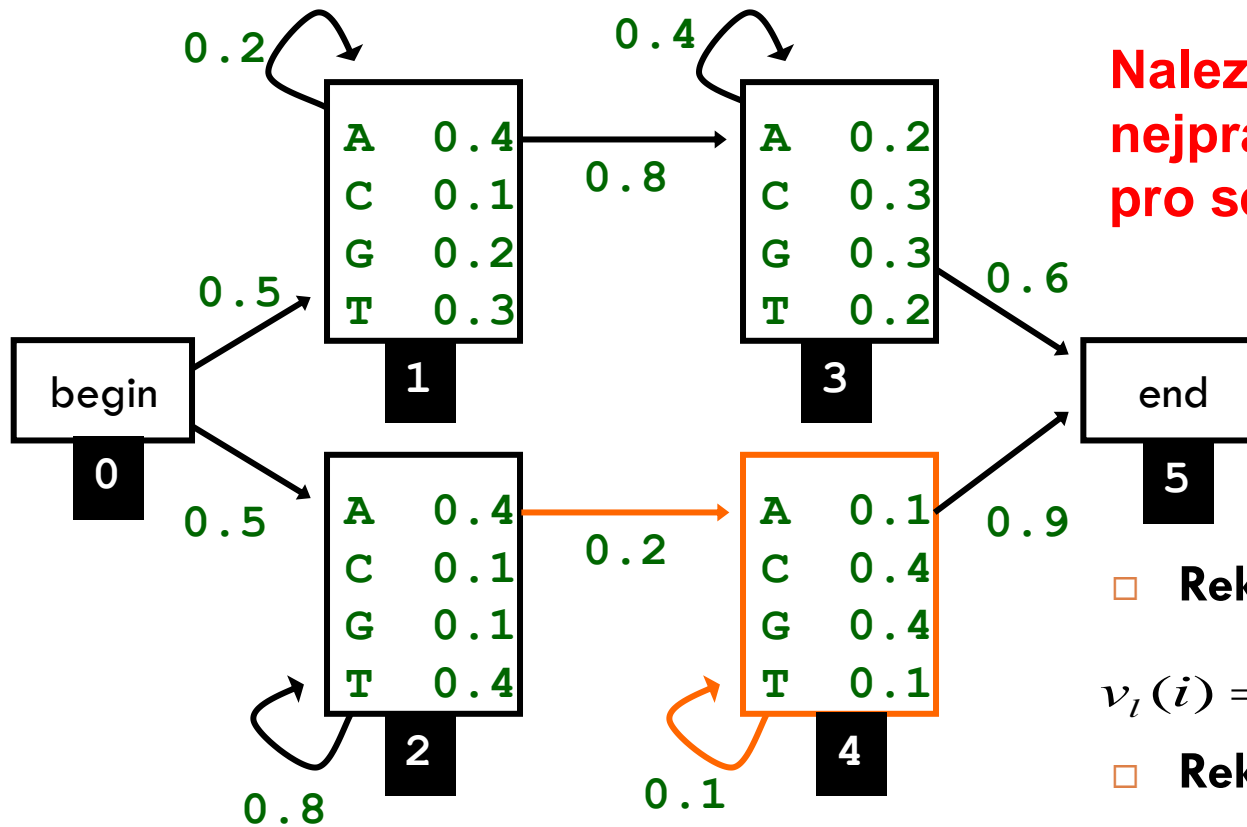
$$P(TAG) = f_5(3) = (f_3(3)a_{35} + f_4(3)a_{45})$$

Nejpravděpodobnější sled

- Jaký je nejpravděpodobnější sled pro sekvenci AACCC? Určete pomocí výčtu všech relevantních cest (= naivní přístup).



Viterbiho algoritmus



**Nalezněte
nejpravděpodobnější sled
pro sekvenci TAGA**

□ **Rekurze pro emitující stavy:**

$$v_l(i) = e_l(x_i) \max_k [v_k(i-1)a_{kl}]$$

□ **Rekurze pro tiché stavy:**

$$v_l(i) = \max_k [v_k(i)a_{kl}]$$

□ **Nezapomeňte udržovat pointery
během výpočtu.**

Příště: Učení se parametrů HMM

- Baum-Welch Algorithm