Machine Learning Journal Club

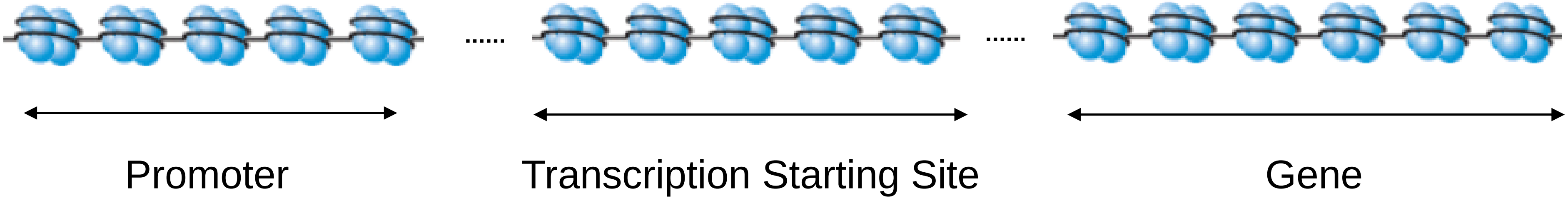
- Introduction to Hidden Markov Models -

Wednesday, June 7th

Baptiste AMELINE, Dyogen team

Markov Models, tools to predict

In biology, the markov models are used to segregrate the genome into small fragments sharing common biological features.



Promoter

Transcription Starting Site

Gene

As a consequence, Markov models are commonly used to perform genomic annotations from whole genome NGS sequencing (ChIP-seq, Bis-Seq...)

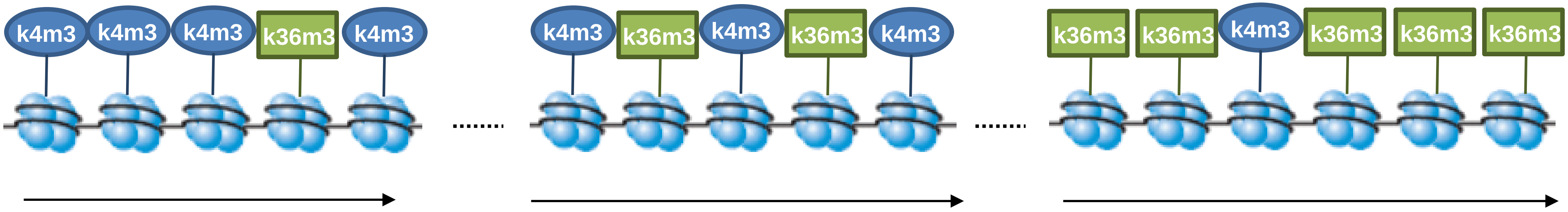
↳ Encode Project, 2012. *Nature*. «An integrated encyclopedia of DNA elements in the human genome»

Two main types of Markov Models

- Markov chain -

Definition : It models the state of a system with a random variable (represented by a probability to observe a pattern) that changes through time. Distribution for this variable depends only on the distribution of the previous state.

You split your DNA sequence into different states (promoter, TSS,...), the transition to a new state depends on the current state, and inside a specific state you have a probability to observe your pattern (ex: histon methylation) a little bit different compared to the observation probability in the previous state.

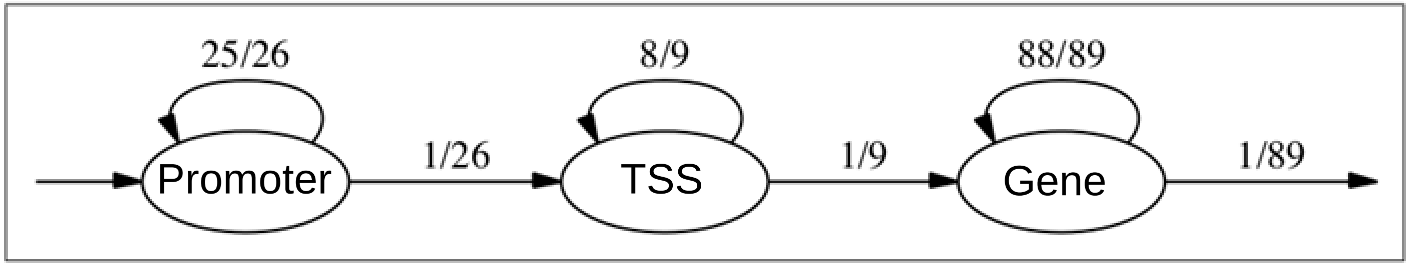


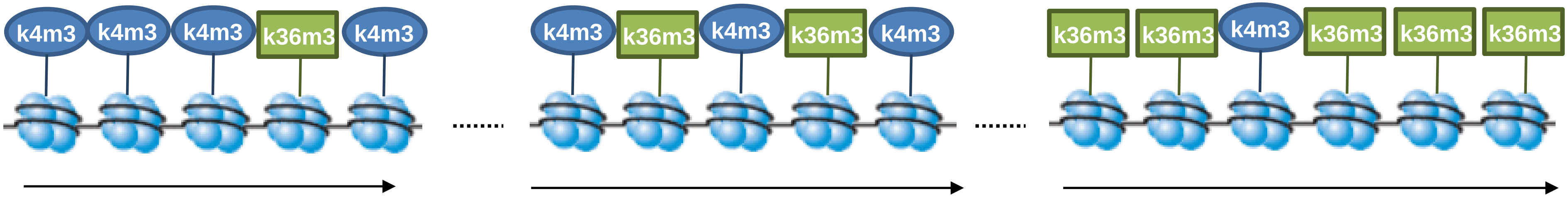
How to rightly describe a model using MM ?

Observations Probability

	Probability to observe H3K4me3	Probability to observe H3K36me3
Promoter	0.85	0.15
TSS	0.75	0.25
Gene	0.10	0.90

Transitions Probabilities





Promoter

Transcription Starting Site

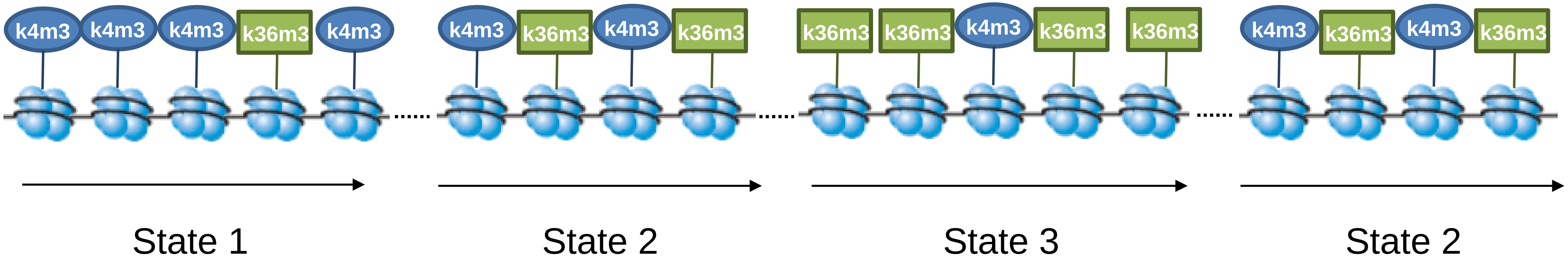
Gene

Two main types of Markov Models

- Hidden Markov Models -

Definition : Hidden Markov Model (HMM) is a Markov chain for which the state is only partially observable. In consequence, observations are related to the state of the system, but they are typically insufficient to precisely determine the state.

The model is splitted into different states (number defined by user) according to the distribution of the random variable (observations).
All transitions from one state to an other are possible and associated to a transition probability.

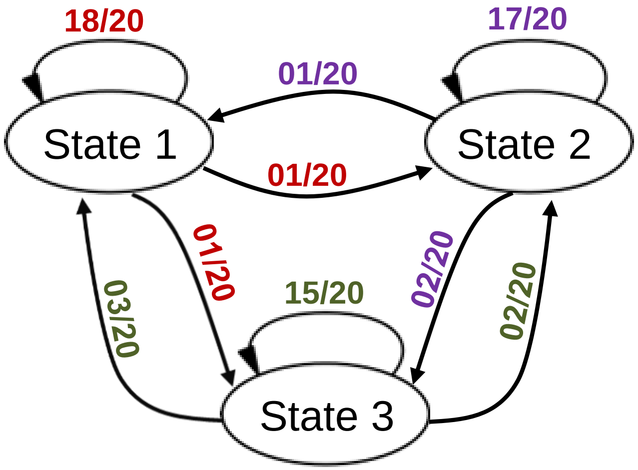


How to rightly describe a model using HMM ?

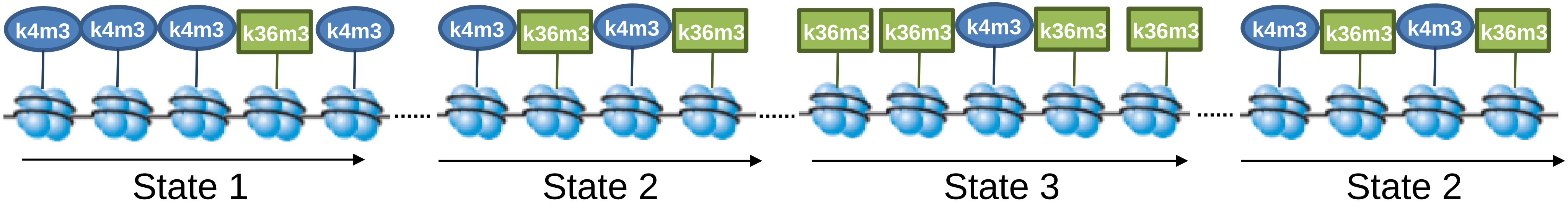
Observations Probability

	Probability to observe H3K4me3	Probability to observe H3K36me3
State 1	0.85	0.15
State 2	0.75	0.25
State 3	0.10	0.90

Transitions Probabilities



From \ To	State 1	State 2	State 3
State 1	18/20	01/20	01/20
State 2	01/20	17/20	02/20
State 3	03/20	02/20	15/20



Hidden Markov Models

General description

A « basic » hidden markov model can be described as follows:

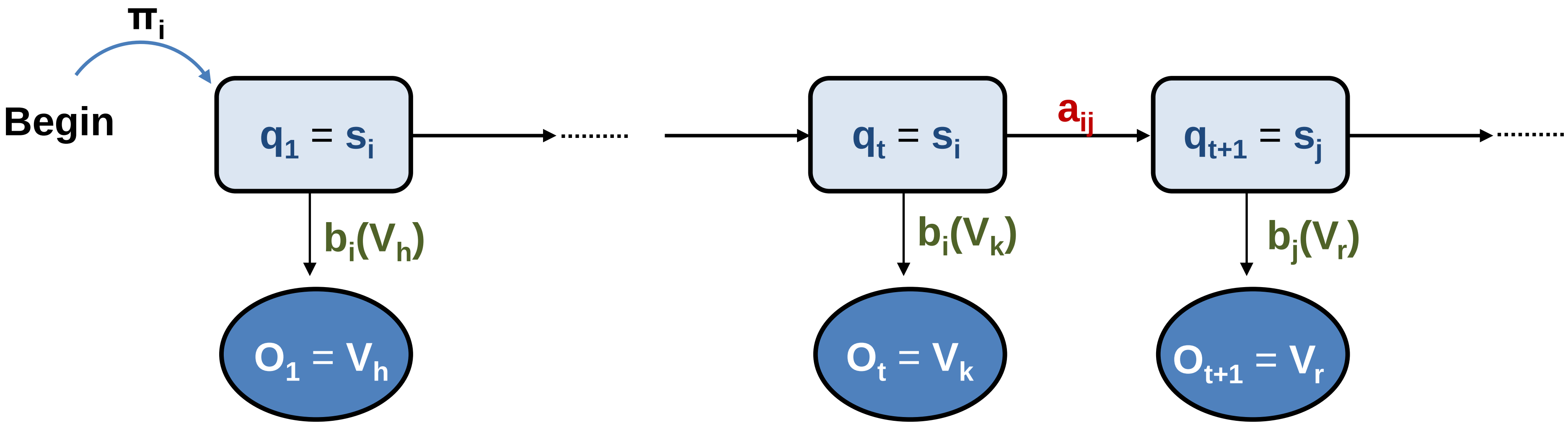
n = number of states
$S = \{s_1, s_2, \dots s_n\}$ = the different states
$V = \{v_1, v_2, \dots, v_M\}$ = alphabet of observations
A = matrix of probabilites of transition
$a_{ij},\ i, j \in [1, n]$ = probability of transition from state i to state j
B = matrix of probabilities of observation
$b_j(k),\ j \in [1, n], k \in [1, M]$ = probability of observation of k, in the state j
π = vector of initial probability
$\Lambda = (A, B, \pi)$ = minimal description of a HMM model

T = Length of an observed sequence
$O = O_1 \dots O_t \dots O_T$ = an observed sequence
$O_t \in V$
$O(i : j) = O_i \dots O_j$ = sub-sequence
$q_1 \dots q_t \dots q_T \quad q_t \in S$ = the series of states that have emitted the sequence
$\mathsf{P}(O \mid \Lambda)$ = probability to observed a sequence O given the HMM Λ
$\mathcal{O} = O^1 \dots O^m$ = a series of m sequences of observations
$\mathsf{P}(\Lambda \mid \mathcal{O})$ = probability the HMM Λ emitted the serie of sequences Θ

	Obs1	Obs2
S1	$B_1(O_1)=0.85$	$B_1(O_2)=0.15$
S2	$B_2(O_1)=0.75$	$B_2(O_2)=0.25$
S3	$B_3(O_1)=0.10$	$B_3(O_2)=0.90$

To From	S1	S2	S3
S1	**a_{11}=18/20**	**a_{12}=01/20**	**a_{13}=01/20**
S2	**a_{21}=01/20**	**a_{22}=17/20**	**a_{23}=02/20**
S3	**a_{31}=03/20**	**a_{32}=02/20**	**a_{33}=15/20**

Generate a sequence with a HMM



```
Begin
    $t \leftarrow 1$
    Choose an initial state $q_1 = s_i$ with a probability $\pi_i$
    While $t \leq T$
        Choose observation $o_t = v_k$ with a probability $b_i(k)$
        Switch to next state $q_{t+1} = s_j$ with a probability $a_{ij}$
        $t \leftarrow t + 1$
    End while
End
```

Hidden Markov Models

- Three main applications of Hidden Markov Models -

Training of the model.

Considering a series of sequences Θ, you adjust the parameters of the model Λ (A, B, π), to maximize : $\mathsf{P}(\mathcal{O} \mid \Lambda) = \prod_{O \in \mathcal{O}} \mathsf{P}(O \mid \Lambda)$

Evaluation of the probability of observation of a sequence.

Given an observed sequence O and an HMM Λ (A, B, π), you apply a maximum likelyhood strategy to estimate the probability of observation.

Searching for the most likely sequence of hidden states.

Given an observation sequence O and an HMM Λ (A, B, π), you apply a maximum likelyhood strategy to estimate the probability of observation.

Hidden Markov Models

- Evaluation of the probability of observation of a sequence -

→ A sequence of observation O can be built using Q different pathways of a model Λ, whose the likelyhoods greatly varies. Thus the probability to observe the sequence O is the sum of probabilities of all the Q differents pathways.

$$\mathbf{P}(O \mid \Lambda) = \sum_{Q} \mathbf{P}(O \mid Q, \Lambda)\mathbf{P}(Q \mid \Lambda)$$

where

$$\mathbf{P}(Q \mid \Lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T}$$

$$\mathbf{P}(O \mid Q, \Lambda) = b_{q_1}(O_1) b_{q_2}(O_2) \ldots b_{q_T}(O_T)$$

→ transition sequence of the pathway q

→ observation sequence of the pathway q

↓

$$\mathbf{P}(O \mid \Lambda) = \sum_{Q = q_1, q_2, \ldots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \ldots a_{q_{T-1} q_T} b_{q_T}(O_T)$$

Considering the Q pathways of length T, the complexity is o(Q^T)

Hidden Markov Models

- Evaluation of the probability of observation of a sequence -

→ The « Forward – backward algorithm »reduces the complexicity to o(n²T)

$$\mathrm{P}(O \mid \Lambda) = \sum_{i=1}^{n} \alpha_t(i)\beta_t(i)$$

Probability to observe the beginning (1 : t)

$$\alpha_t(i) = \mathrm{P}(O_1 O_2 \ldots O_t,\ q_t = s_i \mid \Lambda)$$

Probability to observe the end (t+1 : T)

$$\beta_t(i) = \mathrm{P}(O(t+1:T),\, q_{t+1} = s_i \mid \Lambda)$$

Forward

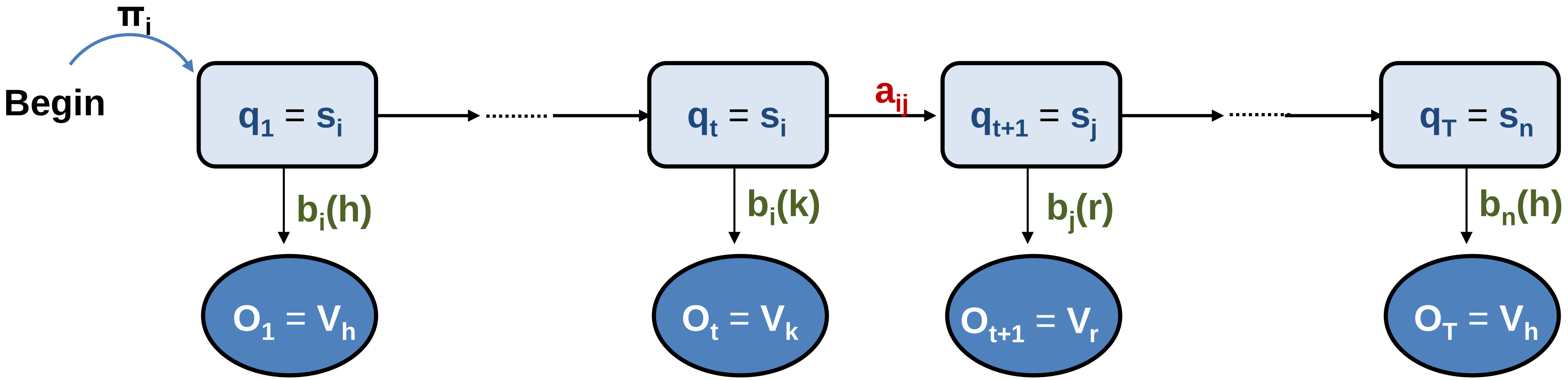
```
Begin
    For i = 1 : n; Do $\alpha_1(i) \leftarrow \pi_i b_i(O_1)$
    t ← 1
    While t ≤ T
        j ← 1
        While j ≤ n
            $\alpha_{t+1}(j) \leftarrow [\sum_{i=1}^{n} \alpha_t(i) a_{ij}] b_j(O_{t+1})$
            j ← j + 1
        End while
        t ← t + 1
    End while
    $\mathrm{P}(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \alpha_T(i)$
End
```

Backward

```
Begin
    For i = 1 : n; Do $\beta_T(i) \leftarrow 1$
    t ← T - 1
    While t ≥ 1
        i ← 1
        While i ≤ n
            $\beta_t(i) \leftarrow \sum_{j=1}^{n} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$
            i ← i + 1
        End while
        t ← t - 1
    End while
    $\mathrm{P}(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \beta_1(i)$
End
```

Hidden Markov Models

- Evaluation of the probability of observation of a sequence -



Forward

```
Begin
    For i = 1 : n; Do $\alpha_1(i) \leftarrow \pi_i b_i(O_1)$
    t ← 1
    While t ≤ T
        j ← 1
        While j ≤ n
            $\alpha_{t+1}(j) \leftarrow [\sum_{i=1}^{n} \alpha_t(i) a_{ij}] b_j(O_{t+1})$
            j ← j + 1
        End while
        t ← t + 1
    End while
    P$(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \alpha_T(i)$
End
```

Backward

```
Begin
    For i = 1 : n; Do $\beta_T(i) \leftarrow 1$
    t ← T - 1
    While t ≥ 1
        i ← 1
        While i ≤ n
            $\beta_t(i) \leftarrow \sum_{j=1}^{n} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$
            i ← i + 1
        End while
        t ← t - 1
    End while
    P$(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \beta_1(i)$
End
```

Hidden Markov Models

- Evaluation of the probability of observation of a sequence -

→ Reduction of the complexicity is based on the assumption that we only need the information of the previous state to predict the next one.

Forward

```
Begin
  For i = 1 : n; Do $\alpha_1(i) \leftarrow \pi_i b_i(O_1)$
  t ← 1
  While t ≤ T
      j ← 1
      While j ≤ n
          $\alpha_{t+1}(j) \leftarrow [\sum_{i=1}^{n} \alpha_t(i) a_{ij}] b_j(O_{t+1})$
          j ← j + 1
      End while
      t ← t + 1
  End while
  P$(O \mid \Lambda) \leftarrow \sum_{i=1}^{n} \alpha_T(i)$
End
```

Complexicity : linear

Complexicity : quadratic

Complexicity : linear

Hidden Markov Models

- Searching for the most likely sequence of hidden states -

→ A sequence of observation O can be built using Q different pathways of a model Λ, whose the likelyhoods greatly vary. Viterbi algorithm gives the state sequence **q** which maximize the probability of observation P(**Q,O** | **Λ**).

Let's define $\delta_t(i)$ as : $\delta_t(i) = \underset{q_1,\ldots,q_{t-1}}{Max} \mathsf{P}(q_1, q_2, \ldots, q_t = s_i, O_1, O_2, \ldots, O_t \mid \Lambda)$


Maths > Bio

→ Best pathway to reach the q_t element

By recurrence, we can set : $\delta_{t+1}(j) = [\underset{i}{\mathrm{Max}}\ \delta_t(i)\, a_{ij}]\, b_j(O_{t+1})$

Hidden Markov Models

- Evaluation of the probability of observation of a sequence -

Viterbi

```
Begin
    For i = 1 : n
        $\delta_1(i) \leftarrow \pi_i b_i(O_1)$
        $\psi_1(i) \leftarrow 0$          = recording of the current state
    End for
    t ← 2
    While t ≤ T - 1
        j ← 1
        While j ≤ n
            $\delta_t(j) \leftarrow \mathrm{Max}_{1\leq i\leq n}\,[\delta_{t-1}(i)\,a_{ij}]\,b_j(O_t)$
            $\psi_t(j) \leftarrow \mathrm{ArgMax}_{1\leq i\leq n}\,[\delta_{t-1}(i)\,a_{ij}]$
            j ← j + 1
        End while
        t ← t + 1
    End while
    P* ← $\mathrm{Max}_{1\leq i\leq n}\,[\delta_T(i)]$
    $q_T^* \leftarrow \mathrm{ArgMax}_{1\leq i\leq n}\,[\delta_T(i)]$
    t ← T
    While t ≥ T - 1
        $q_t^* \leftarrow \psi_{t+1}(q_{t+1}^*)$
        $t \leftarrow t - 1$
    End while
End
```

Complexicity : linear

Complexicity : quadratic

Complexicity : linear

Hidden Markov Models

- Training of the model -

→ In absence of information, initialization of the model Λ^0 (A^0, B^0, π^0) starts with:

- Equiprobability to switch from a state to an other :

π^0

To From	S1	S2	S3
π	**π_1=1/3**	**π_2=1/3**	**π_3=1/3**

A^0

To From	S1	S2	S3
S1	**a_{11}=1/3**	**a_{12}=1/3**	**a_{13}=1/3**
S2	**a_{21}=1/3**	**a_{22}=1/3**	**a_{23}=1/3**
S3	**a_{31}=1/3**	**a_{32}=1/3**	**a_{33}=1/3**

- Probabilities of observation set according to the frequence of observations of the pattern in the dataset.

B^0

	Obs1	Obs2
S1	$B_1(O_1)$=obs1 / (obs1+obs2)	$B_1(O_2)$=obs2 / (obs1+obs2)
S2	$B_2(O_1)$=obs1 / (obs1+obs2)	$B_2(O_2)$=obs2 / (obs1+obs2)
S3	$B_3(O_1)$=obs1 / (obs1+obs2)	$B_3(O_2)$=obs2 / (obs1+obs2)

Hidden Markov Models

- Training of the model -

Training of the model.

Considering a series of sequences Θ, you adjust the parameters of the model Λ (A, B, π), to maximize : $\mathsf{P}(\mathcal{O} \mid \Lambda) = \prod_{O \in \mathcal{O}} \mathsf{P}(O \mid \Lambda)$

Viterbi algorithm on Λ^0 (A^0, B^0, π^0)

$$O^1 = O^1_1, O^1_2 \ldots\ldots O^1_t \ldots\ldots O^1_T$$

$$\vdots$$

$$O^k = O^k_1, O^k_2 \ldots\ldots O^k_t \ldots\ldots O^k_T$$

$$\vdots$$

$$O^m = O^m_1, O^m_2 \ldots\ldots O^m_t \ldots\ldots O^m_T$$

$$P(O^1) = \pi_i b_i(O^1_1),\ a_{ij} b_j(O^1_2) \ldots\ a_{ji} b_i(O^1_t) \ldots\ldots a_{ii} b_i(O^1_T)$$

$$\vdots$$

$$P(O^k) = \pi_j b_j(O^k_1),\ a_{jj} b_j(O^k_2) \ldots\ a_{ji} b_i(O^k_t) \ldots\ldots a_{ii} b_i(O^k_T)$$

$$\vdots$$

$$P(O^m) = \pi_n b_n(O^m_1),\ a_{ni} b_i(O^m_2) \ldots\ a_{ii} b_i(O^m_t) \ldots\ldots a_{ij} b_j(O^m_T)$$

Update

$$\pi^0_i \longrightarrow \pi^1_i = \frac{\text{number of times the model is in the state i to emit } O_1}{\text{number of times the model emits } O_1}$$

Hidden Markov Models

- Training of the model -

$\Lambda^0 (A^0, B^0, \pi^0)$

$O^1 = O^1_1, O^1_2 \dots O^1_t \dots O^1_T$

$\vdots$

$O^k = O^k_1, O^k_2 \dots O^k_t \dots O^k_T$

$\vdots$

$O^m = O^m_1, O^m_2 \dots O^m_t \dots O^m_T$

$P(O^1) = \pi_i b_i(O^1_1),\ a_{ij} b_j(O^1_2) \dots a_{ji} b_i(O^1_t) \dots a_{ii} b_i(O^1_T)$

$\vdots$

$P(O^k) = \pi_j b_j(O^k_1),\ a_{jj} b_j(O^k_2) \dots a_{ji} b_i(O^k_t) \dots a_{ii} b_i(O^k_T)$

$\vdots$

$P(O^m) = \pi_n b_n(O^m_1), a_{ni} b_i(O^m_2) \dots a_{ii} b_i(O^m_t) \dots a_{ij} b_j(O^m_T)$

Update

$$\mathbf{B^0_j(O_k)} \qquad \mathbf{B^1_j(O_k)} = \frac{\text{number of times the model is in the state j to emit } O_k}{\text{number of times the modelis in the state j}}$$

Hidden Markov Models

- Training of the model -

$$\Lambda^0\ (A^0, B^0, \pi^0)$$

$$\begin{aligned}
&O^1 = O^1_1, O^1_2 \ldots O^1_t \ldots O^1_T && P(O^1) = \pi_i b_i(O^1_1),\ a_{ij} b_j(O^1_2) \ldots a_{ji} b_i(O^1_t) \ldots a_{ii} b_i(O^1_T) \\
&\vdots && \vdots \\
&O^k = O^k_1, O^k_2 \ldots O^k_t \ldots O^k_T && P(O^k) = \pi_j b_j(O^k_1),\ a_{jj} b_j(O^k_2) \ldots a_{ji} b_i(O^k_t) \ldots a_{ii} b_i(O^k_T) \\
&\vdots && \vdots \\
&O^m = O^m_1, O^m_2 \ldots O^m_t \ldots O^m_T && P(O^m) = \pi_n b_n(O^m_1)\ a_{ni} b_i(O^m_2) \ldots a_{ij} b_j(O^m_t) \ldots a_{ij} b_j(O^m_T)
\end{aligned}$$

Update

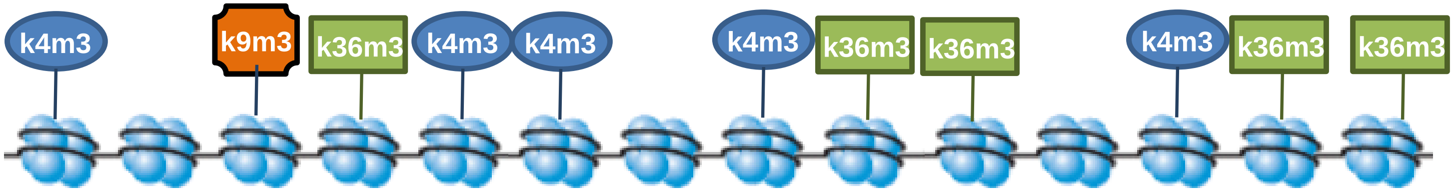
$$\mathbf{A^0_{ij}} \xrightarrow{\text{Update}} \mathbf{A^1_{ij}} = \frac{\text{number of times the transition } i \rightarrow j \text{ is used}}{\text{number of transitions performed from the state } i}$$

Implementation of this strategy : Algorithm of Baum-Welch

(special case of the Expectation Maxizimation algorithm for the adjustement of the HMM parameters)

Hidden Markov Models

- ChromHMM : user (rather) friendly package to apply HMM strategy -



- Considering you have a dataset composed of different ChIP-seq covering (a part of) the genome, by setting a number of expected states, you can :

1. **Train your model and define the parameters (A, B, π) of your HMM Λ using a first replicate.**

π

To / From	S1	S2	S3
π	**π_1**	**π_2**	**π_3**

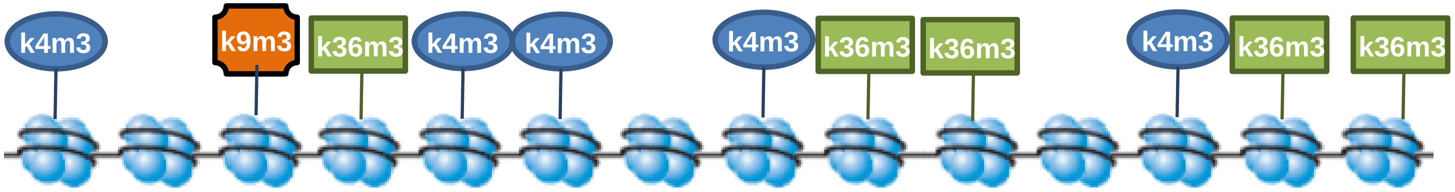
A

To / From	S1	S2	S3
S1	**a_{11}**	**a_{12}**	**a_{13}**
S2	**a_{21}**	**a_{22}**	**a_{23}**
S3	**a_{31}**	**a_{32}**	**a_{33}**

B

	Obs1	Obs2
S1	$B_1(O_1)$	$B_1(O_2)$
S2	$B_2(O_1)$	$B_2(O_2)$
S3	$B_3(O_1)$	$B_3(O_2)$

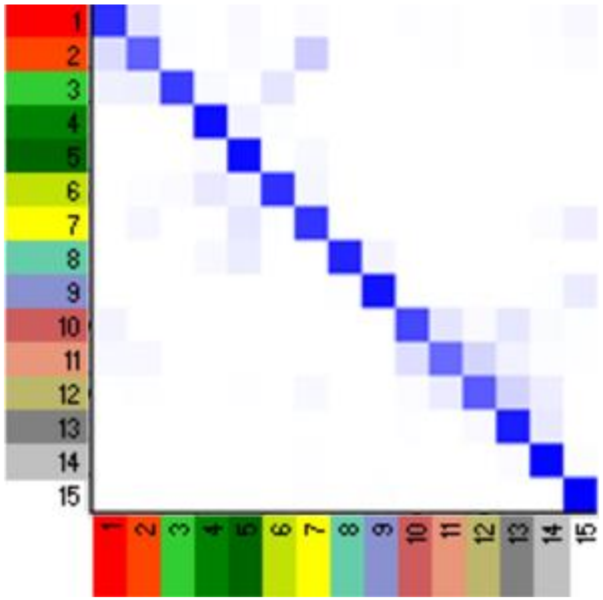
Hidden Markov Models



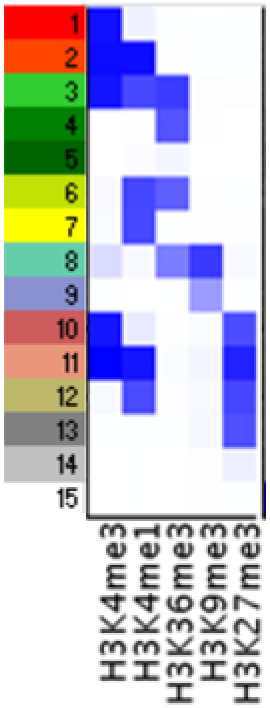
2. Based on a second replicate, search for the sequence of hidden states.

- 15 -states model:

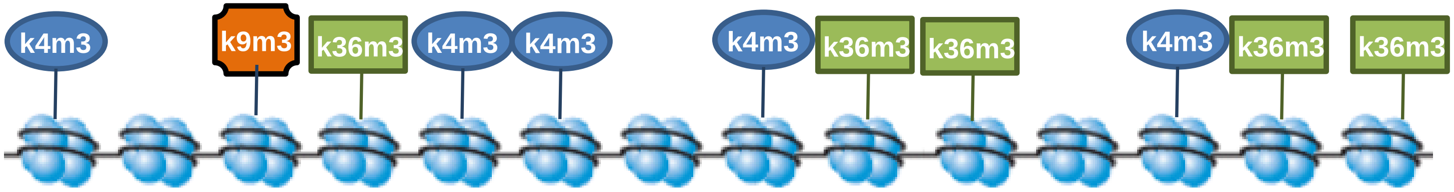
State transition matrix



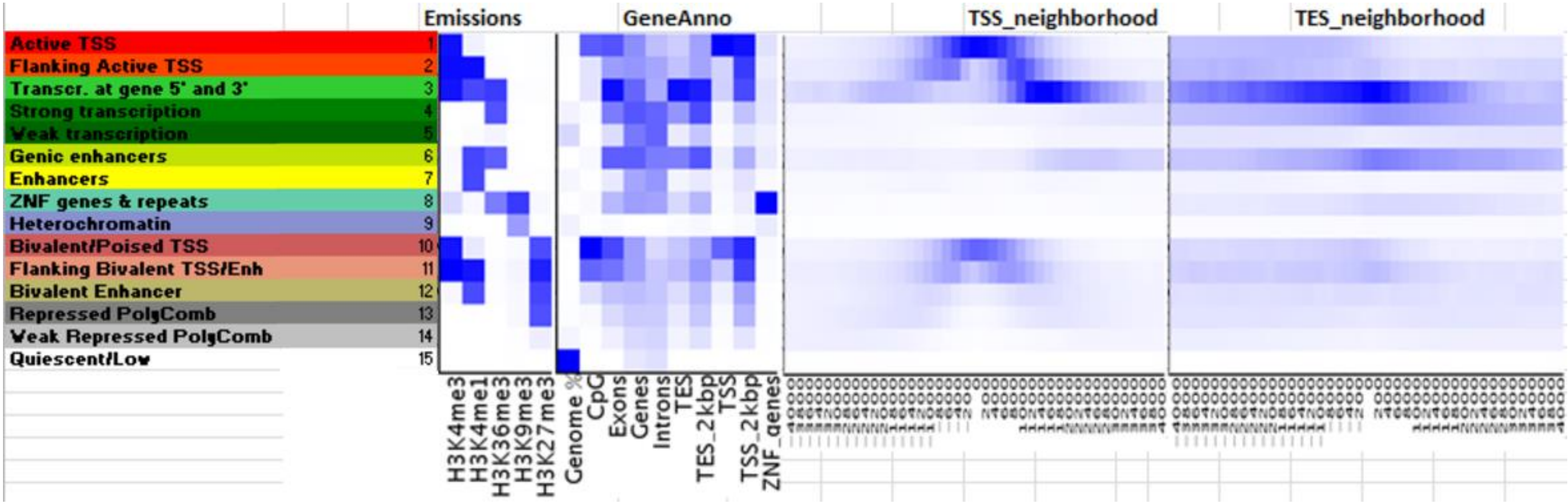
probability obs matrix



Hidden Markov Models



3. Associate your probability observation matrix to genomic annotations to estimate the « biological effect » behind the hidden states .



Do you want more ?

Bioinfo seminar on Thurday, June 22th

« Toward a functionnal map of the human germline genome »

Thank you to your attention