Chaînes de Markov cachées

Chaînes de Markov cachées

Introduction

Origine historique : Markov et les textes

En 1913, Andrei Andreevich Markov a considéré une suite de 20 000 caractères pris dans Eugène Oneguine d'Alexandre Pouchkine, et a distingué les voyelles des consonnes

Не мысля гордый свет забавить,
Вниманье дружбы возлюбя,
Хотел бы я тебе представить
Залог достойнее тебя,
Достойнее души прекрасной,
Святой исполненной мечты,
Поэзии живой и ясной,
Высоких дум и простоты;
Но так и быть — рукой пристрастной

Résultat :

- Consonnes : 56,8%
- Voyelles : 43,2%

→ Peut-on modéliser la langue (russe) par une loi de probabilité ? Cela pourrait être :

$$\pi = \{0.568, 0.432\}$$

Origine historique : Markov et les textes (II)

On peut déjà essayer de simuler selon une loi indépendante et identiquement distribuée (chaque symbole est équiprobable) :

```
lœ2zz. . .gUJCâ ?n)QdKîôê. . .q-        eSœ2,ûP11y).n. . .].tév(5àR8ùùJ«
DxSÀw ?t[nf](mQ,ùy«âw5»F»œbuaAMq ?vSB
Eûef(2eùçJ8Ebw                              ôULù–’ ?vj !G(z,Qqdçoyf ?nJ–’m !3-
.WùD1QwhOôH      : !bBcê1êViaz ;yÀzmô5h0      :PUî-O ;gI.CtBTh       :
SfkcWSmD[l  :z,bu( ;e’.m–WKR !ogv  «3Qq«Pz(). . .LfvbQç  bt. . .–oN
T) ;l88IJù58e«-Gè ;[d(cJEîz(HPIkx !       eâli))5rHxCTk’gûfQULLRd ?ù
s0J      :sz3èh.1êTB ?dà25Jî.wyRq,»rg      pê ;BùèSS«mÀ)NVéaû3k83H
« ?’àkNnC«Ej-V( ?ûLIBCD)èBp-rSè    FdTiO0hv    :ÀûHWîq]voPI«3p
. . .WùgJdqbwnqJ»-œ–’stqC. . .O(guk  é’bâ ?[ù)  : ?JeFbcOIR»V3L«el
pêpzzTgù HLgpUè’Qù3DégVA.Clj] çNÀ]x(iqN]SJGÀôH(LàVruS3Q-
```

Origine historique : Markov et les textes (II)

On peut ensuite distinguer consonnes et voyelles selon $\pi = \{0.568, 0.432\}$

iimu lewiryseeyxupem anaaymajmomaifunuoitl uyymjohioqvwaeaoldini-nugsyieqjiikuiximeayagoiiuiioeaedybmatsqepuh oeahtazeiioyq o nioiui-rhx stnaihxaiu zpulitcma hkzdijjgnsppoxaxvzuuiyjit ffisuybayjvfrvyo-liuidofqyrihozzxusjynpuiqa gizroee qh yil iaere agrieeuiaokoiusxo i deyn ueepyyo qotoqubbiyiizaubi aih zudv yaieuhypyfueheaeosaobajfeeub eekxzadhmixl ijus qyoiez dpd eoorkyfodeyp cy uvoe ehieneanyeezsba-joiacoypxrhrdeiesr pplm yo kjocmyae gaosutgydoaq zaaqyqyodoivrbuy-luypzyuyeroyeeoqyr iegmv oibyyywwsk

Origine historique : Markov et les textes (III)

Affinons le modèle avec des V.A. indépendantes de loi non uniforme.

On tire un caractère d'après sa fréquence d'apparition dans un texte de référence.

aisacscédteaau eeede'e s ua.re'a sssuntcn nnm otac, m nrlr n pte e dldsi r v o. nara àuon irpj ebnridràsàm-fenveteheniIdscs vnrsc vasànncétcéen uig e s eeat n'iede ti a p emoun shqusaGiotmpuilyr contp n'aé,mnr iotOeserep-nmuirma mctlrje a itap ltuziu ajdrr unerj IdeodxAueseres o ,llst epam eddueo tx i e nrcaeir vr'nnrtpe neertjrtp' e avhds caanr tdpua'e'itunnoisq ed'se'-àeue st cre dHt ptu ue'oks c rmr sd,t sl, errnr-nacari.teu koe,duv,r Sqn ierttiq fveqluseuindg' r I stédarsétar'trE

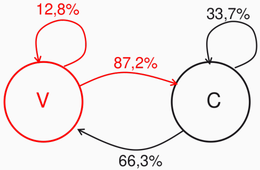
Texte de référence : Guerre et Paix, Tolstoï

A. Markov s'intéresse alors aux couples de lettre.

En revenant au problème voyelle/consonne :

Не мысля гордый свет забавить,
Вниманье дружбы возлюбя,
Хотел бы я тебе представить
Залог достойнее тебя,

Не|е | м|мы|ыс|сл|ля|я | г|го|ор|рд| ды|ый|й | с|св|ве|ет|т | з|
за|аб|ба|ав|ви|ит|ть|ь,|
Вн|ни|им|ма|ан|нь|ье|е | д| др|ру|уж|жб|бы|ы в| во|оз|зл|лю|юб|бя|я,|
Хо|от|те|ел|л | б|бы|ы | я|я | т|те|еб|бе|е | п| пр|ре|ед|дс|ст|та|ав|ви|ит|ть|ь |
За|ал|ло|ог|г | д|до|ос|ст|то|ой|йн|не|ее|е | т|те|еб|бя|я, |



Origine historique : Markov et les textes (V)

On peut généraliser cela à tous les couples de lettres :

aiasoité, erie ppene, le s deprexppanauntre ss mona.Al'antror s qui é de, i Prt Il sari à Ilourri L'e idananvemes e densst-itable ; monanssou-trêti degre jouila stie, ble a r e ces Manous'es fffést vrare qu sttégre bitminaite de de ri ravarie a nta mavesie andar béle hâcuit-veuise us coulant Conoye, de pocome e pre lies mpronss cou sint hon l e aride lass maurétrexpouns ell'é, hex, Cen somare dit que dex vi bl vow bri, foifu e vioilas soneu vle, ds, pépa ne cle quve Quseuns, l t lonté, Pait. lars

Texte de référence : Guerre et Paix, Tolstoï

Origine historique : Markov et les textes (V)

On peut généraliser cela à tous les couples de lettres :

anisosoug sthe m nde, wive Anctis) asad. fa as Mover, omitar yo ooire d dre t t t. he l auon hayo n hang ascoutintileceanadoth thaso theche-rer.“Yo Yo, iol innconan,”“Fre I o s tes vecousthepoy Anary f hay pache, ar s I nee s Shtes of otof an I haly wading asury buss a weexitedele in t d wnt s wa iguesppor pi s m hathes k trcinoome henoowat w s lowel fofar ce etaishainut ; Pred anad sen at f imuto ote Anatyo ba He k yougr aw, s Pávevis hifed Pávede g y’th he I hay titheleding, mopo ated thers wara

Texte de référence : Guerre et Paix, Tolstoï - Traduction en anglais

On peut généraliser cela à tous les couples de lettres :

iöte un, ichichrbeicke d uschn Dierzunk. zuben sa Waglort Haller icherwigem Er Van An ericht jerirert Schien wich dichene witer u hän Wiene f zußt weg d hreie be Scker Sier velaftr Ih wienesin we Frsine ssotte.Dat veo Bendausch wes er ze End ere Läninkemeramu Rorchestonge ksckendachr sf en seralleffde belllerersot zuch sspr An ind Auesiesinarerast, vontes iklde dihn, Alen se, Be wicherzer ndlch von, we m hrber Frorst, ine gles u unnt Küreur mekdeskor, Aret imelsindanktedengk ber stzugeste esit e

Texte de référence : Guerre et Paix, Tolstoï - Traduction en allemand

Origine historique : Markov et les textes (V)

En résumé :

- la fréquence des lettres seules (cas indépendant) est peu informative
- considérer les couples de lettre est plus informatif
- le texte généré n'a pas de sens...
- ...mais contient des caractéristiques propres à chaque langue

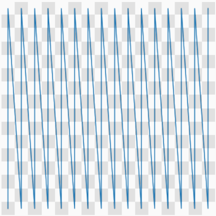
Dans le contexte des images :

- Précédemment : le cas indépendant
- Ensuite : le cas avec dépendances de proche en proche

Une chaîne pour segmenter une image ? la question du parcours

Il faut choisir un parcours pour transformer une image (pixels sur une grille) en chaîne (pixels en séquence).

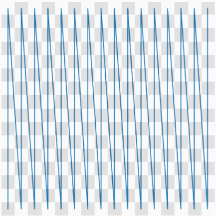
Parcours « informatique » :



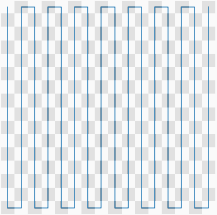
Une chaîne pour segmenter une image ? la question du parcours

Il faut choisir un parcours pour transformer une image (pixels sur une grille) en chaîne (pixels en séquence).

Parcours « informatique » :



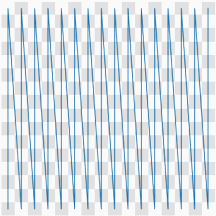
Parcours alterné :



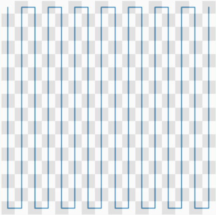
Une chaîne pour segmenter une image ? la question du parcours

Il faut choisir un parcours pour transformer une image (pixels sur une grille) en chaîne (pixels en séquence).

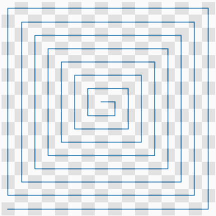
Parcours « informatique » :



Parcours alterné :



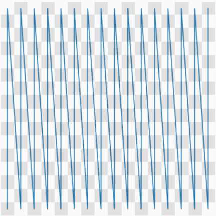
Parcours spiral :



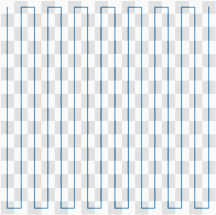
Une chaîne pour segmenter une image ? la question du parcours

Il faut choisir un parcours pour transformer une image (pixels sur une grille) en chaîne (pixels en séquence).

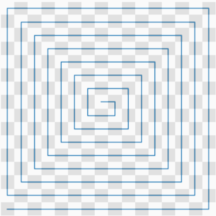
Parcours « informatique » :



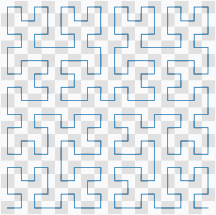
Parcours alterné :



Parcours spiral :



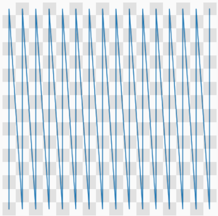
Parcours de Peano :



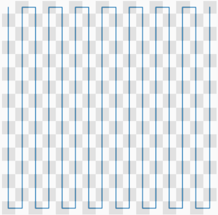
Une chaîne pour segmenter une image ? la question du parcours

Il faut choisir un parcours pour transformer une image (pixels sur une grille) en chaîne (pixels en séquence).

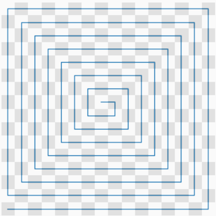
Parcours « informatique » :



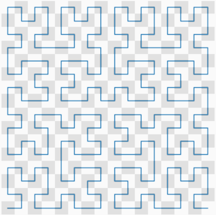
Parcours alterné :



Parcours spiral :



Parcours de Peano :



Critères :

- éviter les discontinuités : voisins dans la chaine $\Rightarrow$ voisins dans l'image
- éviter les motifs artificiels
- favoriser les directions « aléatoires »

Chaînes de Markov cachées

Lois de probabilité

Chaîne de Markov : définition

On se donne une grille $\mathcal{S}$ contenant N pixels ; que l'on « lit » par un parcours donné.

$\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$ est une **chaîne de Markov** si $\forall n$:

$$p(x_n | x_1, x_2, \ldots x_{n-1}) = p(x_n | x_{n-1})$$

Chaîne de Markov : définition

On se donne une grille $\mathcal{S}$ contenant N pixels ; que l'on « lit » par un parcours donné.

$\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$ est une **chaîne de Markov** si $\forall n$:

$$p(x_n | x_1, x_2, \ldots x_{n-1}) = p(x_n | x_{n-1})$$

La loi de $\boldsymbol{x}$ est alors :

$$\begin{aligned} p(\boldsymbol{x}) &= p(x_1) p(x_2 | x_1) p(x_3 | x_2) \ldots p(x_N | x_{N-1}) \\ &= p(x_1) \prod_{n=2}^{N} p(x_n | x_{n-1}) \end{aligned}$$

Chaîne de Markov : homogénéité

$\boldsymbol{X}$ est une chaîne de Markov **homogène** si $p(x_{n+1}|x_n)$ ne dépend pas de n.

Une chaîne de Markov homogène est caractérisée par les paramètres :

- $\pi_i = p(x_1 = \omega_i)$ « probabilité de commencer par la classe i »
- $a_{ij} = p(x_n = \omega_j | x_{n-1} = \omega_i)$ « probabilité d'aller de i vers j »

Modèle d'observation, lois jointe & postérieure

À partir d'une réalisation de chaîne (transformée) $\boldsymbol{x}$; on observe une image $\boldsymbol{y}$.

Rappel : hypothèse de bruit indépendant

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{n=1}^{N} p(y_n|x_n)$$

Modèle d'observation, lois jointe & postérieure

À partir d'une réalisation de chaîne (transformée) $\boldsymbol{x}$; on observe une image $\boldsymbol{y}$.

Rappel : hypothèse de bruit indépendant

$$p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{n=1}^{N} p(y_n|x_n)$$

Donc la postérieure s'obtient comme suit :

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \\
&= p(x_1)p(y_1|x_1) \times \prod_{n=2}^{N} p(y_n|x_n)p(x_n|x_{n-1}) \\
&= \underbrace{p(X_1 = \omega_i)}_{\pi_i}\underbrace{p(y_1|X_1 = \omega_i)}_{\mathcal{N}(y_1;\mu_i,\sigma_i)} \times \prod_{n=2}^{N} \underbrace{p(y_n|X_n = \omega_i)}_{\mathcal{N}(y_n;\mu_i,\sigma_i)}\underbrace{p(X_n = \omega_j|X_{n-1} = \omega_j)}_{a_{ij}}
\end{aligned}$$

Modèle d'observation, lois jointe & postérieure (II)

$$p(\boldsymbol{x}|\boldsymbol{y}) = \underbrace{p(X_1 = \omega_i)}_{\pi_k}\underbrace{p(y_1|X_1 = \omega_i)}_{\mathcal{N}(y_1;\mu_i,\sigma_i)} \times \prod_{n=2}^{N} \underbrace{p(y_n|X_n = \omega_i)}_{\mathcal{N}(y_n;\mu_i,\sigma_i)}\underbrace{p(X_n = \omega_j|X_{n-1} = \omega_j)}_{a_{ij}}$$

En supposant K classes, les paramètres du modèle sont donc :

$$\boldsymbol{\Theta} = \left\{\{\pi_i\}_{1\leq i\leq K}, \{\mu_i\}_{1\leq i\leq K}, \{\sigma_i\}_{1\leq i\leq K}, \{a_{ij}\}_{1\leq i,j\leq K}\right\}$$

▷ Segmentation non supervisée

Représentation graphique : graphe de dépendance

Graphe de dépendance :

- un sommet = une variable
- pas de lien entre A et B $\rightarrow$ indépendance de A et B conditionnellement aux autres variables

Graphe pour une chaîne de Markov :

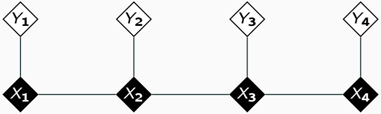
Représentation graphique : graphe de dépendance

Graphe de dépendance :

- un sommet = une variable
- pas de lien entre A et B $\rightarrow$ indépendance de A et B conditionnellement aux autres variables

Graphe pour une chaîne de Markov :

Chaîne de Markov cachée :



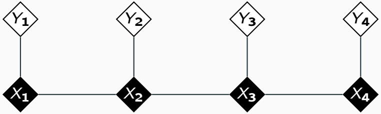
Représentation graphique : graphe de dépendance

Graphe de dépendance :

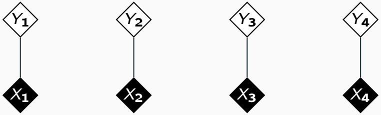
- un sommet = une variable
- pas de lien entre A et B $\rightarrow$ indépendance de A et B conditionnellement aux autres variables

Graphe pour une chaîne de Markov :

Chaîne de Markov cachée :



Modèle indépendant :



Simuler une chaîne de Markov cachée

Pour simuler $(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$ il faut :

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{x})$
2. Simuler $\boldsymbol{y}$ selon $p(\boldsymbol{y}|\boldsymbol{x})$

Simuler une chaîne de Markov cachée

Pour simuler $(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$ il faut :

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{x})$
2. Simuler $\boldsymbol{y}$ selon $p(\boldsymbol{y}|\boldsymbol{x})$

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{X} = \boldsymbol{x})$, c'est :
   - tirer $X_1 = x_1$ d'après $p(X_1)$ $\triangleleft$ paramètres π_i

Simuler une chaîne de Markov cachée

Pour simuler $(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$ il faut :

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{x})$
2. Simuler $\boldsymbol{y}$ selon $p(\boldsymbol{y}|\boldsymbol{x})$

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{X} = \boldsymbol{x})$, c'est :
   - tirer $X_1 = x_1$ d'après $p(X_1)$ $\triangleleft$ paramètres π_i
   - tirer $X_2 = x_2$ d'après $p(X_2|X_1 = x_1)$ $\triangleleft$ paramètres a_{ij}

Simuler une chaîne de Markov cachée

Pour simuler $(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$ il faut :

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{x})$
2. Simuler $\boldsymbol{y}$ selon $p(\boldsymbol{y}|\boldsymbol{x})$

1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{X} = \boldsymbol{x})$, c'est :

- tirer $X_1 = x_1$ d'après $p(X_1)$ ◁ paramètres π_i
- tirer $X_2 = x_2$ d'après $p(X_2|X_1 = x_1)$ ◁ paramètres a_{ij}
- tirer $X_3 = x_3$ d'après $p(X_3|X_2 = x_2)$

  (...)
- tirer X_N d'après $p(X_N|X_{N-1} = x_{N-1})$

Simuler une chaîne de Markov cachée

Pour simuler $(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{Y} = \boldsymbol{y})$ il faut :

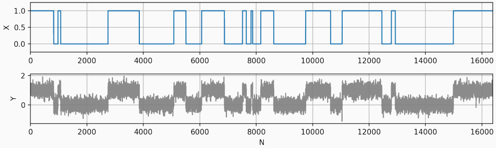
1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{x})$
2. Simuler $\boldsymbol{y}$ selon $p(\boldsymbol{y}|\boldsymbol{x})$

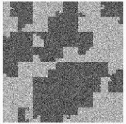
1. Simuler $\boldsymbol{x}$ d'après $p(\boldsymbol{X} = \boldsymbol{x})$, c'est :
   - tirer $X_1 = x_1$ d'après $p(X_1)$ ◁ paramètres π_i
   - tirer $X_2 = x_2$ d'après $p(X_2|X_1 = x_1)$ ◁ paramètres a_{ij}
   - tirer $X_3 = x_3$ d'après $p(X_3|X_2 = x_2)$

     (...)
   - tirer X_N d'après $p(X_N|X_{N-1} = x_{N-1})$

2. Simuler $\boldsymbol{y}$ selon $p(\boldsymbol{y}|\boldsymbol{x})$, c'est « bruiter » l'image avec les $\mathcal{N}(\mu_i, \sigma_i)$.

Quelques exemples



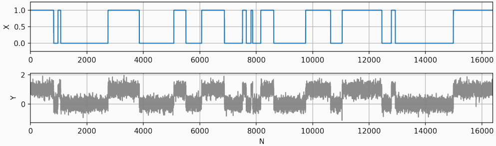


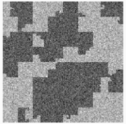
$$\mu_0 = 0,\ \mu_1 = 1$$

$$\sigma_0 = \sigma_1 = 0.25$$

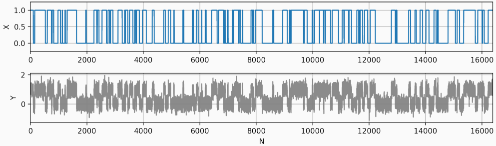
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$

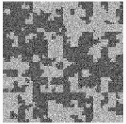
Quelques exemples



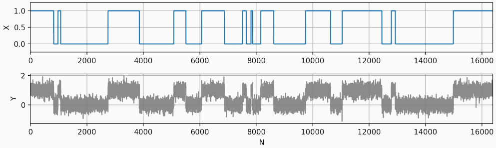


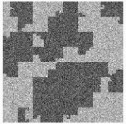
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$



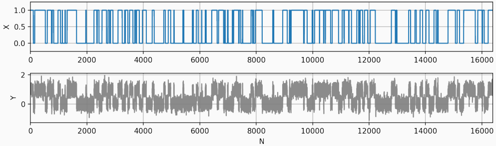


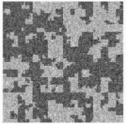
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$$



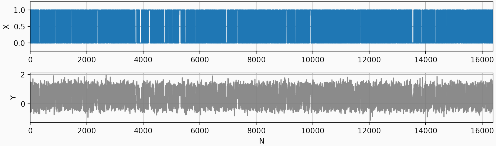


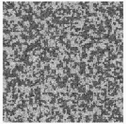
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$





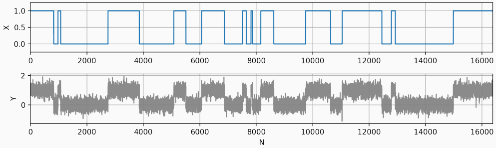
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$$

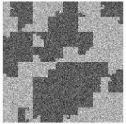




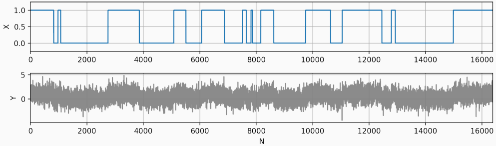
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$$

Quelques exemples (II)



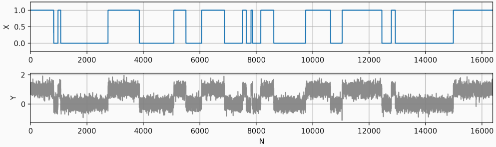


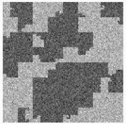
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$



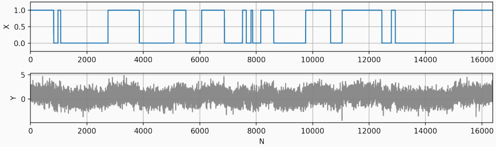
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 1$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$

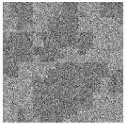
Quelques exemples (II)



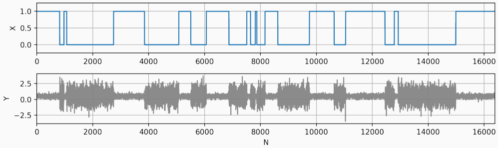


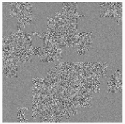
$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 0.25$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$





$$\mu_0 = 0,\ \mu_1 = 1$$
$$\sigma_0 = \sigma_1 = 1$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$





$$\mu_0 = 0.5,\ \mu_1 = 0.5$$
$$\sigma_0 = 0.9, \sigma_1 = 0.2$$
$$A = \begin{bmatrix} 0.999 & 0.001 \\ 0.001 & 0.999 \end{bmatrix}$$

Chaînes de Markov cachées

Segmentation

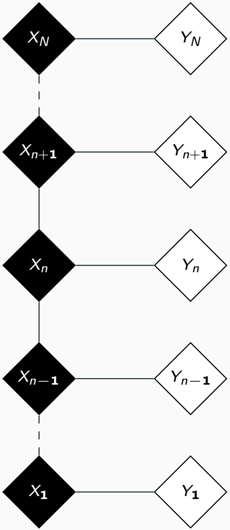
MPM pour les chaînes de Markov cachées

Pour rappel, le MPM est, $\forall 1 \leq n \leq N$:

$$\hat{x}_n^{\text{MPM}} = \arg\max_{\omega \in \Omega} p(X_n = \omega | \boldsymbol{Y} = \boldsymbol{y})$$

On introduit la notation de la postérieure :

$$\xi(n, i) = p(X_n = \omega_i | \boldsymbol{Y} = \boldsymbol{y})$$



MPM pour les chaînes de Markov cachées

Pour rappel, le MPM est, $\forall 1 \leq n \leq N$:

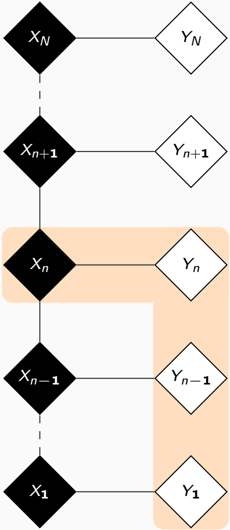
$$\hat{x}_n^{\text{MPM}} = \arg\max_{\omega \in \Omega} p(X_n = \omega | \boldsymbol{Y} = \boldsymbol{y})$$

On introduit la notation de la postérieure :

$$\xi(n, i) = p(X_n = \omega_i | \boldsymbol{Y} = \boldsymbol{y})$$

Ainsi qu'une probabilité « ascendante » ou « forward » :

$$\alpha(n, i) = p(X_n = \omega_i, Y_1 = y_1, \ldots, Y_n = y_n)$$



MPM pour les chaînes de Markov cachées

Pour rappel, le MPM est, $\forall 1 \leq n \leq N$:

$$\hat{x}_n^{\text{MPM}} = \arg\max_{\omega \in \Omega} p(X_n = \omega | \boldsymbol{Y} = \boldsymbol{y})$$

On introduit la notation de la postérieure :

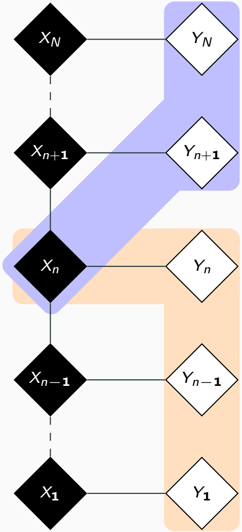
$$\xi(n, i) = p(X_n = \omega_i | \boldsymbol{Y} = \boldsymbol{y})$$

Ainsi qu'une probabilité « ascendante » ou « forward » :

$$\alpha(n, i) = p(X_n = \omega_i, Y_1 = y_1, \dots, Y_n = y_n)$$

et une probabilité « descendante » ou « backward » :

$$\beta(n, i) = p(Y_{n+1} = y_{n+1}, \dots, Y_N = y_N | X_n = \omega_i)$$



MPM pour les chaînes de Markov cachées

$$\beta(n,i) = p(Y_{n+1} = y_{n+1}, \dots, Y_N = y_n | \, X_n = \omega_i)$$
$$\alpha(n,i) = p(X_n = \omega_i, Y_1 = y_1, \dots, Y_n = y_n)$$

On a :

$$\xi(n,i) = \frac{\alpha(n,i)\beta(n,i)}{\sum_{j=1}^{K} \alpha(n,j)\beta(n,j)}$$

Donc :

$$\arg\max_{\omega\in\Omega} p(X_n = \omega | \, \boldsymbol{Y} = \boldsymbol{y}) = \arg\max_{\omega\in\Omega} \xi(n,i) = \arg\max_{\omega\in\Omega} \alpha(n,i)\beta(n,i)$$

MPM pour les chaînes de Markov cachées

α et β sont calculables par récursion !

Pour α :

$$
\left|
\begin{array}{ll}
\text{Initialisation :} & \alpha(1, i) = \pi_i p(y_1 | x_1 = \omega_i) \\
\text{Récursion :} & \alpha(n+1, i) = \left(\sum_{j=1}^{K} \alpha(n, j) a_{ji} \right) p(y_{n+1} | x_{n+1} = \omega_i)
\end{array}
\right.
$$

MPM pour les chaînes de Markov cachées

α et β sont calculables par récursion !

Pour α :

$$\left|\begin{array}{ll} \text{Initialisation :} & \alpha(1,i) = \pi_i p(y_1|x_1=\omega_i) \\ \text{Récursion :} & \alpha(n+1,i) = \left(\sum_{j=1}^{K} \alpha(n,j) a_{ji}\right) p(y_{n+1}|x_{n+1}=\omega_i) \end{array}\right.$$

Pour β :

$$\left|\begin{array}{ll} \text{Initialisation :} & \beta(N,i) = 1 \\ \text{Récursion :} & \beta(n,i) = \sum_{j=1}^{K} \beta(n+1,j) a_{ij} p(y_{n+1}|x_{n+1}=\omega_j) \end{array}\right.$$

MPM pour les chaînes de Markov cachées

Bilan :

- On peut calculer les α et β connaissant $\boldsymbol{Y} = \boldsymbol{y}$, et $\boldsymbol{\Theta}$
- L'algorithme « forward-backward » est déterministe, non itératif
- Malgré la « structure » imposée par le modèle en chaîne, toutes les probabilités sont calculables exactement.

MPM pour les chaînes de Markov cachées

Bilan :

- On peut calculer les α et β connaissant $\boldsymbol{Y} = \boldsymbol{y}$, et $\boldsymbol{\Theta}$
- L'algorithme « forward-backward » est déterministe, non itératif
- Malgré la « structure » imposée par le modèle en chaîne, toutes les probabilités sont calculables exactement.

Mais aussi :

- On peut également formuler (et appliquer) le MAP
- Il faut alors calculer $p(\boldsymbol{X} = \boldsymbol{x} | \boldsymbol{Y} = \boldsymbol{y})$, récursivement

$\rightarrow$ Algorithme de Viterbi. (non vu ici)

Chaînes de Markov cachées

Non supervisée

Estimation non supervisée avec SEM

La segmentation MPM suppose $\mathbf{\Theta}$ connu ; ce qui n'est pas toujours le cas.

On peut alors (comme précédemment) adopter l'approche SEM. À l'itération $q+1$:

1. Simuler $\boldsymbol{x}^{(q+1)}$ selon $p_{\mathbf{\Theta}^{(q)}}(\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y})$
2. Estimer $\mathbf{\Theta}^{(q+1)}$ par les EMV avec $(\boldsymbol{x}^{(q+1)}, \boldsymbol{y})$

Estimation non supervisée avec SEM

La segmentation MPM suppose $\boldsymbol{\Theta}$ connu ; ce qui n'est pas toujours le cas.

On peut alors (comme précédemment) adopter l'approche SEM. À l'itération $q+1$:

1. Simuler $\boldsymbol{x}^{(q+1)}$ selon $p_{\boldsymbol{\Theta}^{(q)}}(\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y})$
2. Estimer $\boldsymbol{\Theta}^{(q+1)}$ par les EMV avec $(\boldsymbol{x}^{(q+1)}, \boldsymbol{y})$

Pour (2) on utilise les EMV précédents, ainsi que :

$$\hat{a}_{ij} = \frac{1}{N-1} \sum_{n=1}^{N-1} \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}$$

Estimation non supervisée avec SEM

La segmentation MPM suppose $\mathbf{\Theta}$ connu ; ce qui n'est pas toujours le cas.

On peut alors (comme précédemment) adopter l'approche SEM. À l'itération $q+1$:

1. Simuler $\boldsymbol{x}^{(q+1)}$ selon $p_{\mathbf{\Theta}^{(q)}}(\boldsymbol{x}|\boldsymbol{Y}=\boldsymbol{y})$
2. Estimer $\mathbf{\Theta}^{(q+1)}$ par les EMV avec $(\boldsymbol{x}^{(q+1)}, \boldsymbol{y})$

Pour (2) on utilise les EMV précédents, ainsi que :

$$\hat{a}_{ij} = \frac{1}{N-1}\sum_{n=1}^{N-1} \mathbb{1}_{\{x_n=\omega_i, x_{n+1}=\omega_j\}}$$

Pour (1) il faut connaître les **probabilités conditionnelles *a posteriori*** :
$p(x_{n+1}=\omega_j|x_n=\omega_i, \boldsymbol{y})$

Probabilités conditionnelles a posteriori

Notons $\psi(i, j, n) = p(x_{n+1} = \omega_j, x_n = \omega_i | \boldsymbol{y})$.

Alors

$$p(x_{n+1} = \omega_j | x_n = \omega_i, \boldsymbol{y}) = \frac{\psi(i, j, n)}{\sum_{j=1}^{k} \psi(i, j, n)} = \frac{\psi(i, j, n)}{\xi(i, n)}$$

Les ψ sont par ailleurs calculables à partir des α, β déjà vus.

Probabilités conditionnelles a posteriori

Notons $\psi(i,j,n) = p(x_{n+1} = \omega_j, x_n = \omega_i | \mathbf{y})$.

Alors

$$p(x_{n+1} = \omega_j | x_n = \omega_i, \mathbf{y}) = \frac{\psi(i,j,n)}{\sum_{j=1}^{k} \psi(i,j,n)} = \frac{\psi(i,j,n)}{\xi(i,n)}$$

Les ψ sont par ailleurs calculables à partir des α, β déjà vus.

$$\psi(i,j,n) = \frac{a_{ij}\alpha(i,n)\beta(j,n+1)p(y_{n+1}|x_{n+1} = \omega_j)}{\sum_u \left(\sum_v a_{vu}\alpha(v,n)\right)\beta(u,n+1)p(y_{n+1}|x_{n+1} = \omega_u)}$$

Interprétation : numérateur normalisé pour tout i, j

Chaînes de Markov cachées : résumé

Démarche pour la segmentation non supervisée d'image :

- transformer l'image en chaîne
- l'algorithme SEM donne une estimation de $\boldsymbol{\Theta}$ sachant $\boldsymbol{y}$.
- l'algorithme Forward-Backward donne une estimation de $\boldsymbol{x}$ sachant $\boldsymbol{\Theta}, \boldsymbol{y}$
- transformer le résultat en image

Chaînes de Markov cachées : résumé

Démarche pour la segmentation non supervisée d'image :

- transformer l'image en chaîne
- l'algorithme SEM donne une estimation de $\boldsymbol{\Theta}$ sachant $\boldsymbol{y}$.
- l'algorithme Forward-Backward donne une estimation de $\boldsymbol{x}$ sachant $\boldsymbol{\Theta}, \boldsymbol{y}$
- transformer le résultat en image

Ce que nous n'avons pas vu :

- le MAP est également réalisable : algorithme de Viterbi
- EM est aussi possible : estimateurs pondérés par les ξ