

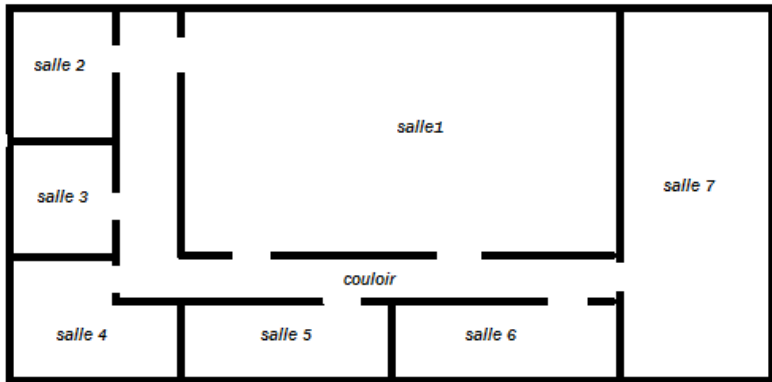
# Estimation de l'arbre de contexte dans les VLHMM

Thierry Dumont

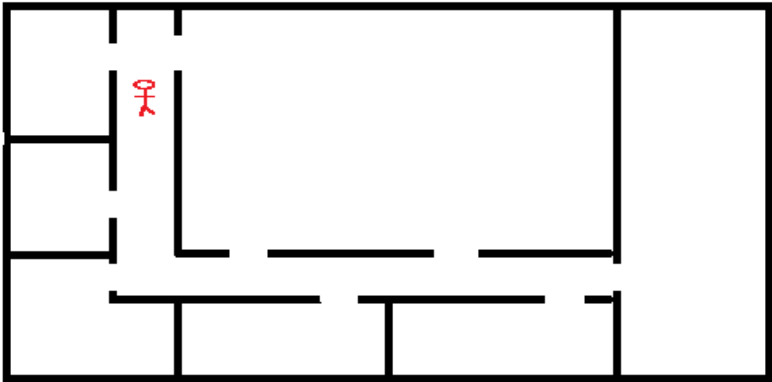
ID Services - Université de Paris sud XI

November 16, 2011

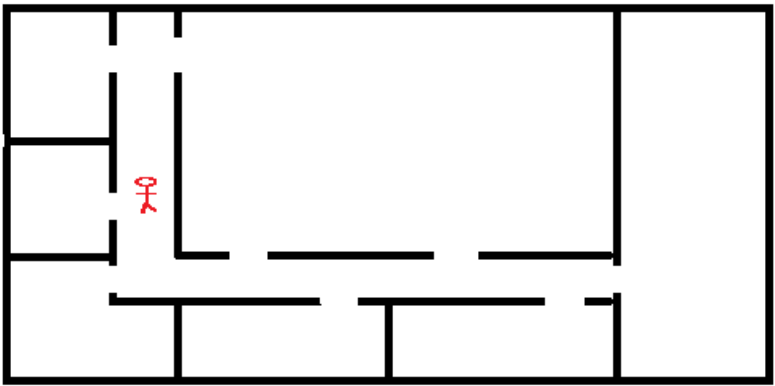
## Premier Exemple : La localisation "Indoor".c



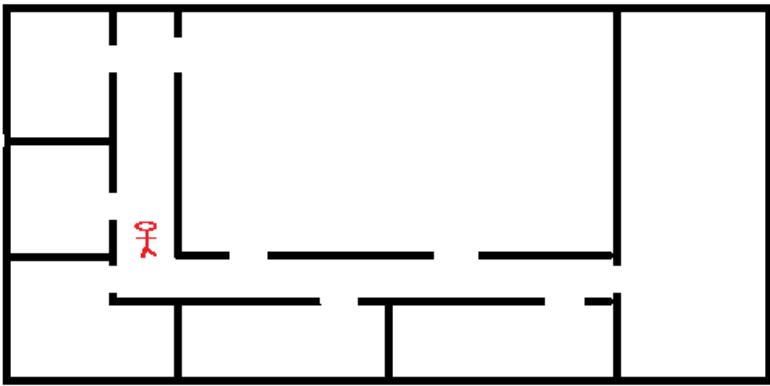
## Premier Exemple : La localisation "Indoor".



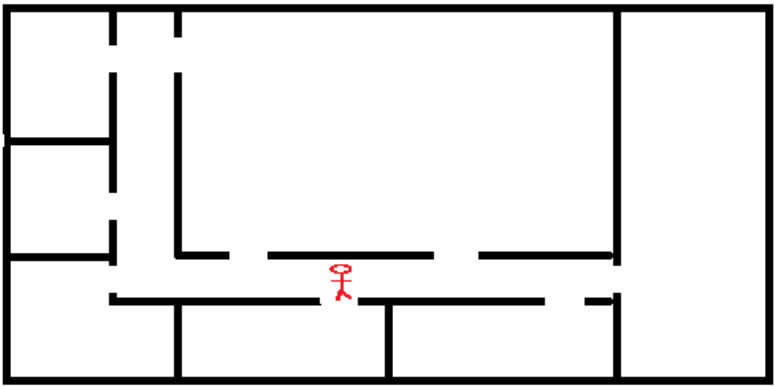
# Premier Exemple : La localisation "Indoor".



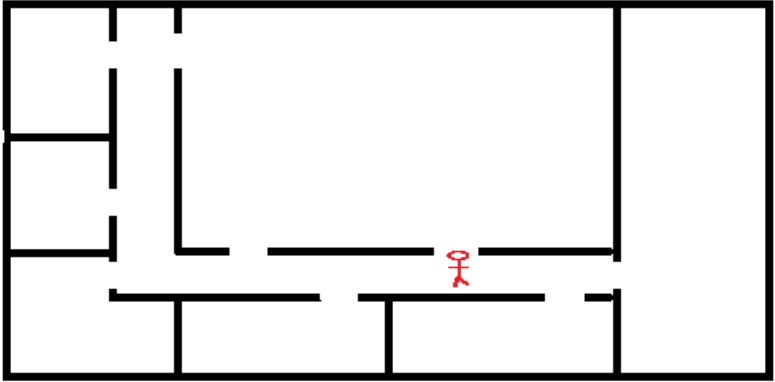
# Premier Exemple : La localisation "Indoor".



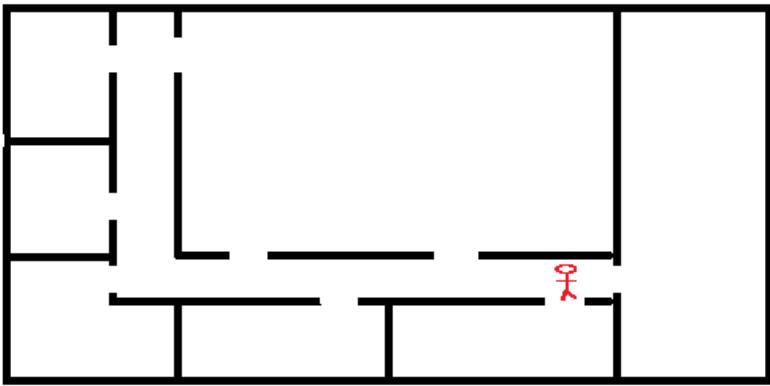
# Premier Exemple : La localisation "Indoor".



# Premier Exemple : La localisation "Indoor".

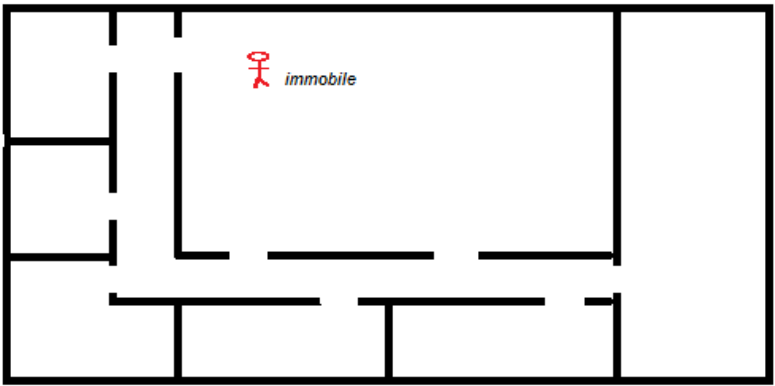


# Premier Exemple : La localisation "Indoor".

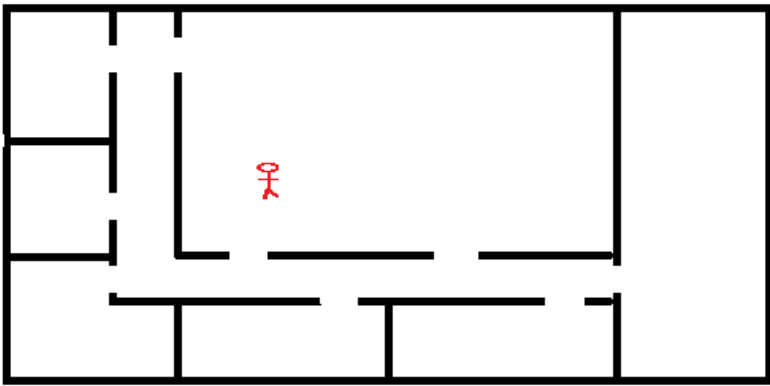




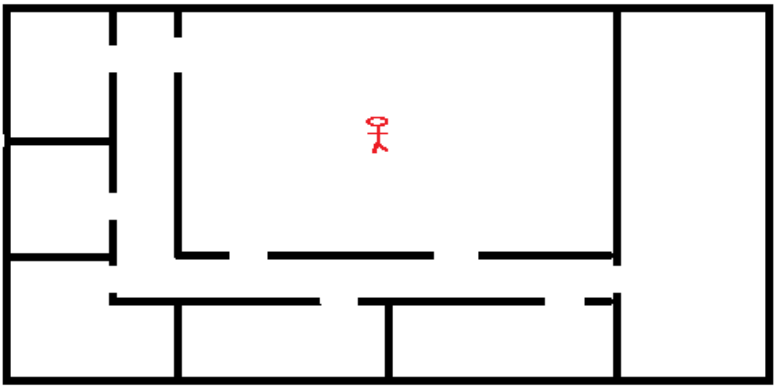
# Premier Exemple : La localisation "Indoor".



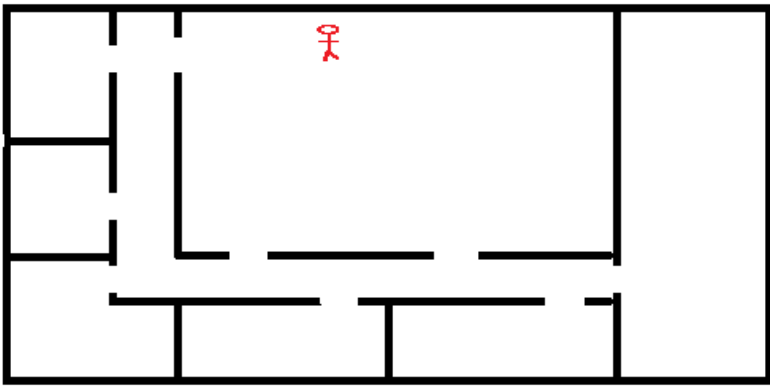
# Premier Exemple : La localisation "Indoor".



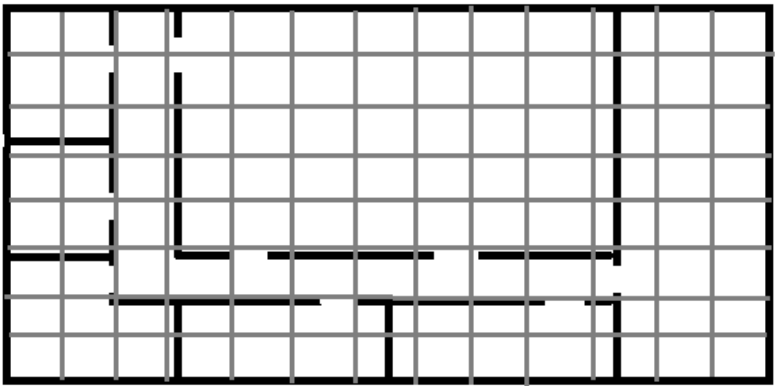
# Premier Exemple : La localisation "Indoor".



# Premier Exemple : La localisation "Indoor".



Premier Exemple : La localisation "Indoor". Discrétisation de l'environnement.



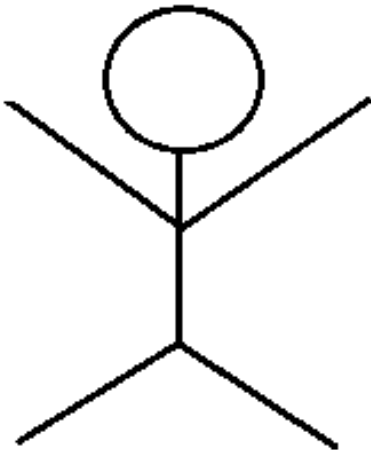
Cette discrétisation va nous permettre de modéliser le déplacement du bonhomme par une chaîne de Markov à espace d'états finis.

### Définition

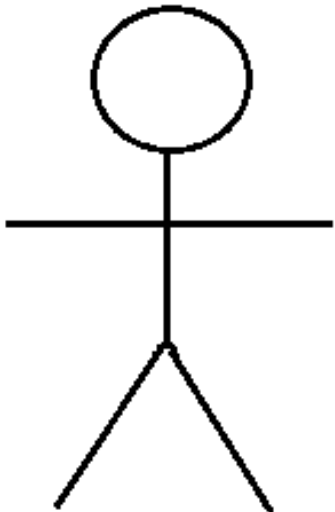
*On appelle chaîne de Markov une suite de variables aléatoires  $(X_n)_{n \geq 1}$  à valeurs dans un espace probabilisé  $(E, \mathcal{B}, \mathbb{P})$  telle que, pour chaque  $n$ , connaissant la valeur de  $X_n$ ,  $X_{n+1}$  soit indépendante de  $X_k$ , pour  $k$  inférieur ou égal à  $n - 1$ . Autrement dit, pour tout  $x$  de  $E$  :*

$$\mathbb{P}(X_n = x | X_{-\infty}, \dots, X_{n-1}) = \mathbb{P}(X_n = x | X_{n-1})$$

## Deuxième exemple : L'analyse des mouvements.



# Deuxième exemple : L'analyse des mouvements.

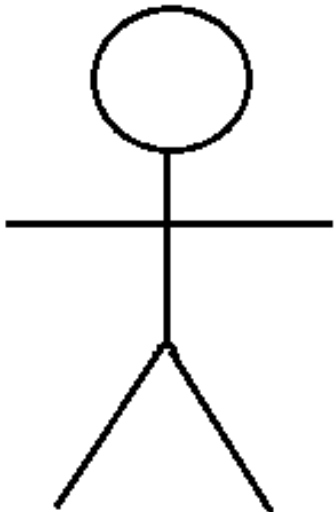




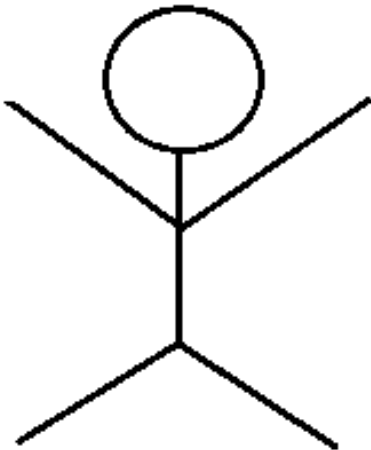
# Deuxième exemple : L'analyse des mouvements.



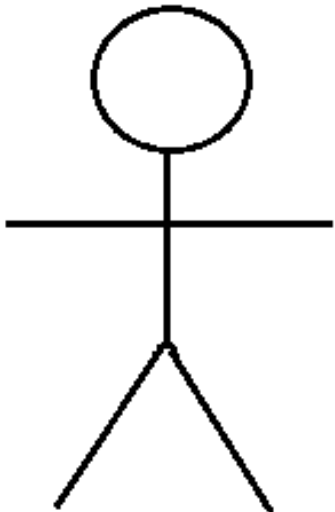
# Deuxième exemple : L'analyse des mouvements.



## Deuxième exemple : L'analyse des mouvements.



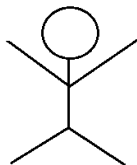
# Deuxième exemple : L'analyse des mouvements.



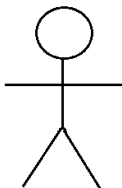
## Deuxième exemple : L'analyse des mouvements.



Deuxième exemple : L'analyse des mouvements. La discrétisation.



Position 1



Position 2



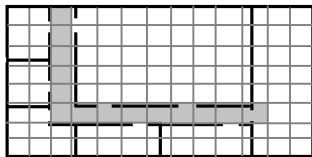
Position 3

Notons  $X_n$  position du corps à l'instant  $n$ .  $X_n$  peut être modélisé par une chaîne de Markov.

$$\mathbb{P}(X_n = x | X_{-\infty}, \dots, X_{n-1}) = \mathbb{P}(X_n = x | X_{n-1})$$

Ces deux modélisations vous paraissent-elles correctes?????

Retour sur l'exemple 1.



Si  $X_{n-1}$  est une position du couloir, Le bonhomme a une dynamique (une "vitesse") qu'il faut prendre en compte si l'on souhaite décrire au mieux la loi de  $X_n$ !

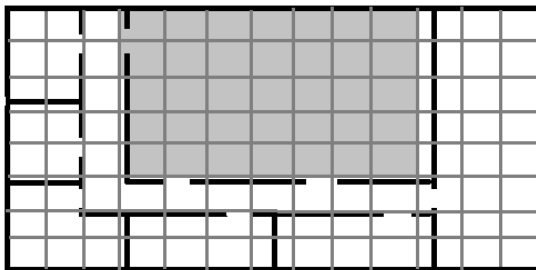
$$\mathbb{P}(X_n = x | X_{-\infty}, \dots, X_{n-1}) \neq \mathbb{P}(X_n = x | X_{n-1})$$

On peut, par exemple, modéliser notre processus  $(X_n)_n$  par une chaîne de Markov d'ordre 2 :

$$\mathbb{P}(X_n = x | X_{-\infty}, \dots, X_{n-1}) = \mathbb{P}(X_n = x | X_{n-1}, X_{n-2})$$



Si le bonhomme est dans une salle :



Le comportement du bonhomme peut sembler plus anarchique et peut très bien être modélisé par une chaîne de Markov classique (d'ordre 1).

## Définition

Le processus stationnaire  $(X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov de longueur variable s'il existe une fonction  $c$  telle que :

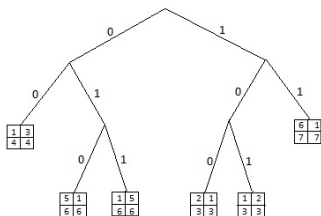
$$c(x_{-\infty:0}) = x_{-\ell(x_{-\infty:0})+1:0}$$

où :

$$\ell(x_{-\infty:0}) = \min\{k \mid \forall x_1 \in E, \\ \mathbb{P}(X_1 = x_1 \mid X_{-\infty:0} = x_{-\infty:0}) = \\ \mathbb{P}(X_1 = x_1 \mid X_{-k+1:0} = x_{-k+1:0})\}$$

Les éléments de l'image de  $c$  sont appelés contextes et possèdent une représentation d'arbre. La fonction  $c$  est par conséquent appelée fonction contexte

## Représentation en arbre :

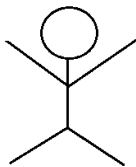


Ici  $E = \{0, 1\}$ , exemple d'interprétation de cet arbre :

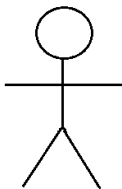
$$\mathbb{P}(X_1 = 0 | 0, 0, x_{-2}, x_{-3}, \dots) = \mathbb{P}(X_1 = 0 | 0, 0) = \frac{1}{4}$$

$$\mathbb{P}(X_1 = 1 | 0, 0, x_{-2}, x_{-3}, \dots) = \mathbb{P}(X_1 = 1 | 0, 0) = \frac{3}{4}$$

## Retour sur l'exemple 2.



Position 1



Position 2



Position 3

$E = \{1, 2, 3\}$ ,  
pour tout  $x \in E$ ,

$$P(X_1 = x | X_0 = 1, X_{-1} = 2, \dots) = P(X_1 = x | X_0 = 1)$$

$$P(X_1 = x | X_0 = 2, X_{-1} = 1, \dots) = P(X_1 = x | X_0 = 2, X_{-1} = 1)$$

...

...

## Remarque

*Connaitre l'arbre de contexte "minimal" d'un processus permet de connaitre le nombre minimal de paramètres à estimer nécessaires à l'analyse du processus.*

*L'utilisation la plus répandue des modèles VLMC est dans la théorie du codage (théorie de l'information) ou pour la modélisation de séquences biologiques.*

## Remarque

*On est tenté de prendre un arbre de contexte le plus grand possible de manière à capturer la plus grande quantité d'information possible. Cependant le nombre de paramètres à estimer croît exponentiellement avec l'ordre (profondeur de l'arbre)...*

*Prendre un arbre trop petit éloigne le modèle de la réalité.*

*Un compromis doit être trouvé via de la sélection de modèle :*

*l' algorithme "context" de Rissanen (1983) [1] permet de trouver l'arbre de contexte d'un processus grâce à une réalisation finie de celui-ci.*

*Imre Csiszár et Zsolt Talata [2] démontrent la consistance de l'estimateur BIC sans supposer de borne a priori sur la profondeur de l'arbre de contextes réel du processus.*

## Définition

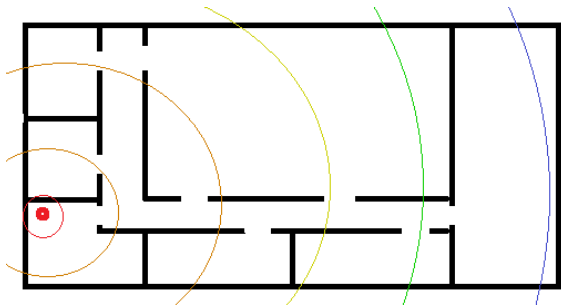
$(X_n, Y_n)_{n \in \mathbb{N}}$  est un modèle de Markov caché (HMM pour Hidden Markov Model) si  $(X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov **non-observée** et  $(Y_n)_{n \in \mathbb{N}}$ , appelées **observations** soient indépendantes relativement à  $(X_n)_{n \in \mathbb{N}}$  :

$$\mathbb{P}(Y_1 \in A_1, \dots, Y_k \in A_k | X_1, \dots, X_k) = \prod_{i=1}^k \mathbb{P}(Y_i \in A_i | X_i)$$

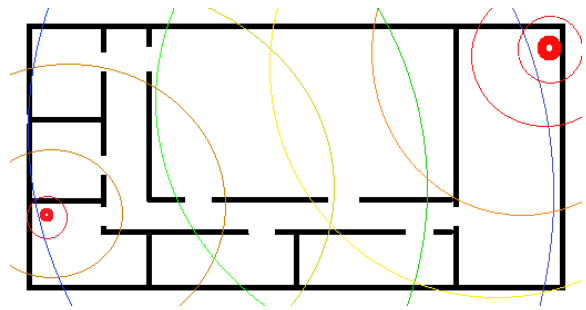
Retour sur le Premier Exemple : La localisation "Indoor".  
On ne connaît pas la suite des positions! (c'est justement ce que l'on cherche en localisation "indoor") On a accès a des observations et c'est grâce à ces observations que nous allons estimer ces positions.



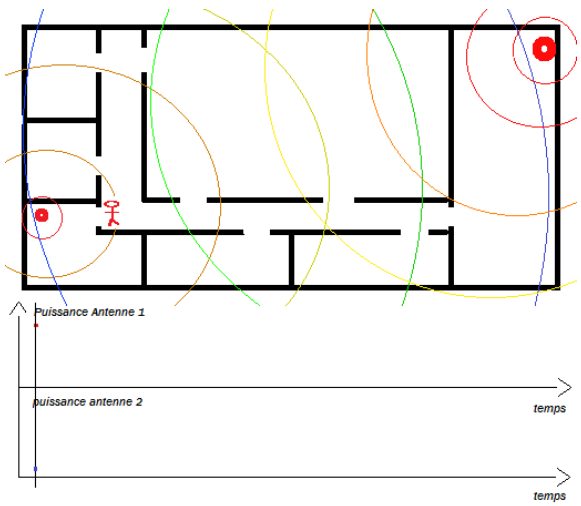
La localisation "Indoor" par ondes Radio :



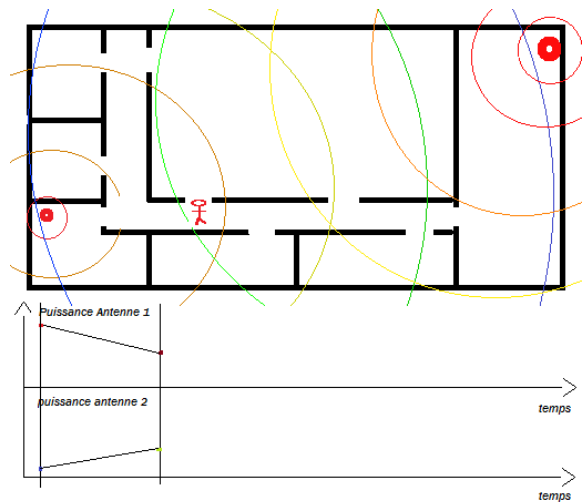
# La localisation "Indoor" par ondes Radio :



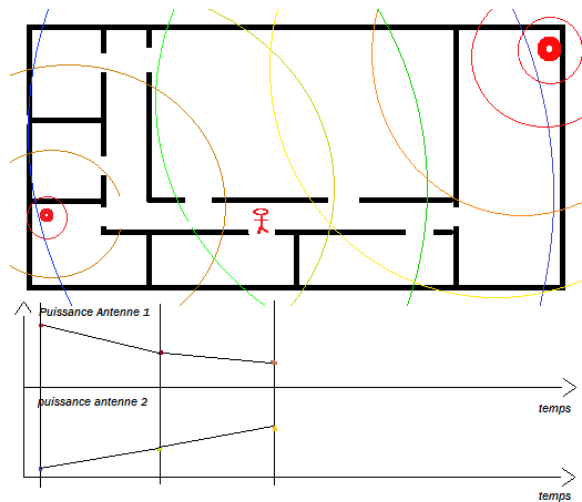
# La localisation "Indoor" par ondes Radio :



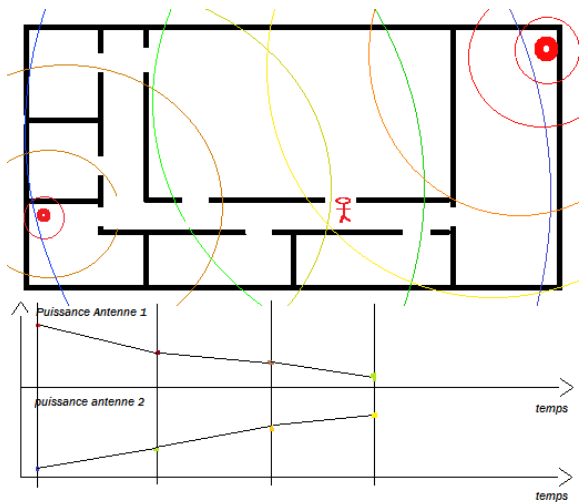
## La localisation "Indoor" par ondes Radio :



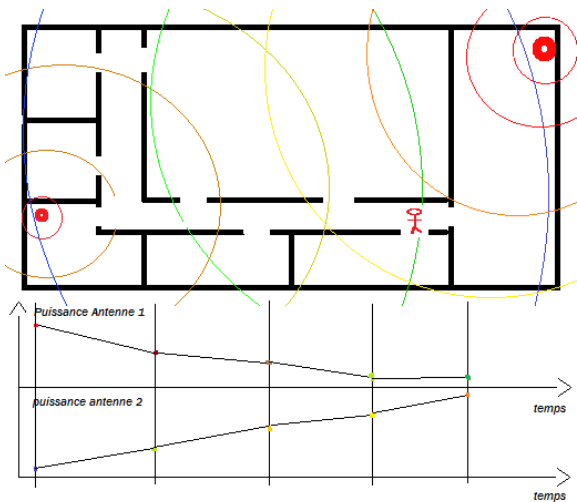
## La localisation "Indoor" par ondes Radio :



## La localisation "Indoor" par ondes Radio :

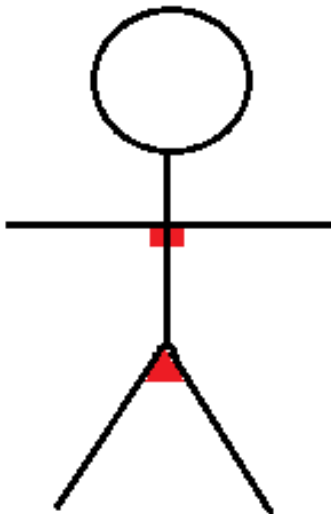


## La localisation "Indoor" par ondes Radio :



Comme pour l'exemple précédent, les positions prises par le corps dans l'analyse de mouvements ne peuvent être observées directement (capteurs d'angles placés sur les articulations du sujet, observation du mouvement par vidéo...)





## Définition

$(X_n, Y_n)_{n \in \mathbb{N}}$  est un modèle de Markov caché à longueur variable (VLHMM pour Variable Length Hidden Markov Model) si  $(X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov à longueur variable **non-observée** et  $(Y_n)_{n \in \mathbb{N}}$ , appelées **observations** sont indépendantes relativement à  $(X_n)_{n \in \mathbb{N}}$  :

$$\mathbb{P}(Y_1 \in A_1, \dots, Y_k \in A_k | X_1, \dots, X_k) = \prod_{i=1}^k \mathbb{P}(Y_i \in A_i | X_i)$$

Cette définition apparaît pour la première fois dans [3] et [4] dans le cadre de l'analyse de mouvements humains.

## Notations

Supposons  $(X_n, Y_n)_{n \in \mathbb{N}}$  VLHMM,  $X_n$  à valeur dans  $\mathbb{X}$  fini et  $Y_n$  à valeurs dans  $\mathbb{R}^B$ .

- notons  $\tau^*$  l'arbre de contextes du processus  $(X_n)_n$  (son vrai arbre de contextes),
- pour tout  $\tau$  arbre de contextes notons

$$\Theta_\tau = \left\{ (P_{s,i})_{s \in \tau, i \in \mathbb{X}}, P_{s,x} \geq 0 \forall (s,x) \in \tau \times \mathbb{X}, \right. \\ \left. \sum_{x \in \mathbb{X}} P_{s,x} = 1, \forall s \in \tau \right\}$$

Espace des paramètres de transition.

## Notations

- $\mathbb{P}(Y_1 \in A_1, \dots, Y_k \in A_k | X_1 = x_1, \dots, X_k = x_k) = \prod_{i=1}^k \left[ \int_{A_i} g_{\theta_{e,x_i}, \eta}(y) d\mu_y \right]$

- 

$$\Theta_e = \left\{ ((\theta_{e,x})_{x \in \mathbb{X}}, \eta) \in \mathbb{R}^{|\mathbb{X}|} \times \mathbb{R} \right\}$$

*Espace des paramètres d'émission.*

- $\Theta_\tau = \Theta_{t,\tau} \times \Theta_e$

## Définition

### La Vraisemblance

Pour tout  $\tau$  arbre de contextes, pour tout  $\theta = (\theta_t, \theta_e) \in \Theta_\tau$ , nous définissons ce que nous appelons la vraisemblance par :

$$\forall y_{1:n} \in (\mathbb{R}^B)^n, g_\theta(y_{1:n}) = \sum_{x_{1:n} \in \mathbb{X}^n} \left[ \prod_{i=1}^n g_{\theta_{e, x_i, \eta}}(y_i) \right] g_{\theta_t}(x_{1:n})$$

Où :

$$g_{\theta_t}(x_{1:n}) = \sum_{x_{-d(\tau)+1:0} \in \mathbb{X}^{d(\tau)}} \nu_{d(\tau)}(x_{-d(\tau)+1:0}) \prod_{i=1}^n P_{c(x_{-d(\tau)+i:i-1}), x_i}$$

## Définition

Si nous observons  $Y_{1:n}$ , nous définissons l'estimateur de l'arbre de contexte pour la chaîne de Markov cachée à longueur variable  $X$  par

$$\hat{\tau}_n = \underset{\tau \text{ arbre complet}}{\operatorname{argmin}} \left[ - \sup_{\theta \in \Theta_\tau} g_\theta(Y_{1:n}) + \operatorname{pen}(n, \tau) \right]$$

Où  $\operatorname{pen}(n, \tau) = \operatorname{pen}_\alpha(n, \tau) = \left[ \sum_{t=1}^{|\tau|} \frac{(|\mathbb{X}|-1)t + \alpha}{2} \right] \log n$

## Theorem

*Sous certaines conditions classiques sur le modèle (irréductibilité de la VLHC,  $\theta_{e,x} \neq \theta_{e,x'}$  si  $x \neq x'$ , des conditions d'intégrabilité et de continuité de  $g_{\theta_e \dots}$ ),*

*Si il existe  $b \in \mathbb{R}$  tel que :*

$$(*) \sup_{\theta_e \in \Theta_e} \sup_{x_{1:n}} \left[ \log \prod_{i=1}^n g_{\theta_e, x_i, \eta}(Y_i) - \log \mathbb{KT}_e^n(Y_{1:n} | x_{1:n}) \right] \leq b \log n \text{ p.s.}$$

où

$$\mathbb{KT}_e^n(Y_{1:n} | x_{1:n}) = \int_{\Theta_e} \left[ \prod_{i=1}^n g_{\theta_e, x_i, \eta}(y_i) \right] \pi_e^n(d\theta_e)$$

*Si  $\alpha > 2(b + 1)$ , alors  $\hat{\tau}_n \sim \tau^*$  pour  $n$  assez grand p.s.*

**Cas particulier important où les lois d'émission sont des gaussiennes de moyennes inconnues et de variance commune à tous les  $x \in \mathbb{X}$  mais inconnue**

On peut trouver une loi mélangeante  $\pi_e$  sur  $\Theta_e$  telle que (\*) soit vérifiée pour  $b = \frac{k+1}{2}$ , on a alors pour ce cas particulier :

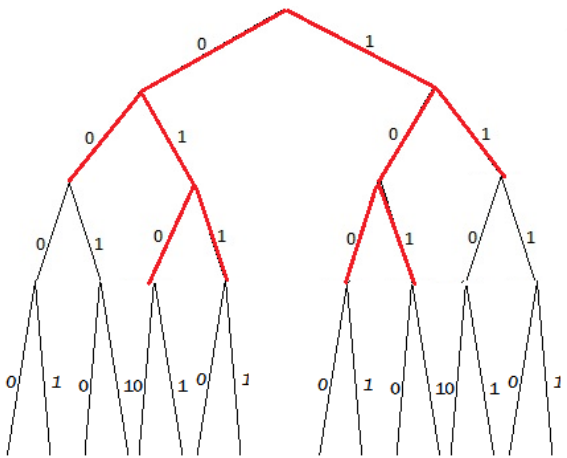
**Theorem**

*Si  $\alpha > k + 3$ , alors  $\hat{\tau}_n \sim \tau^*$  p.s.*

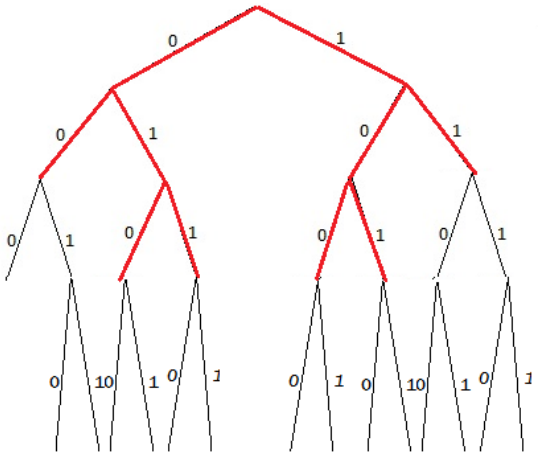


Un algorithme d'élagage pour le cas général est présenté dans [5].  
Il permet, se donnant une suite d'observations  $Y_{1:n}$  de calculer  $\hat{\tau}_n$   
en "élaguant" un arbre *couvrant* notre  $\tau^*$ .

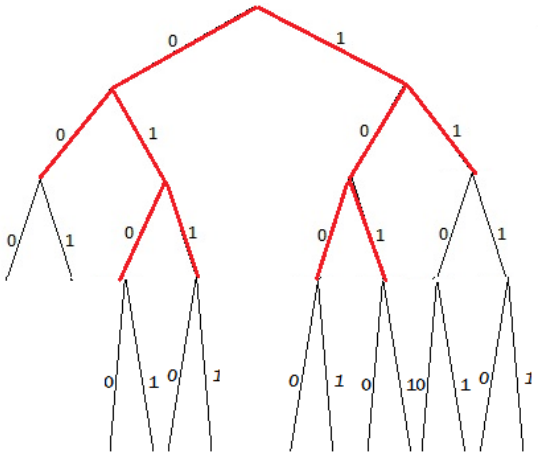
### Arbre couvrant *maximal*.



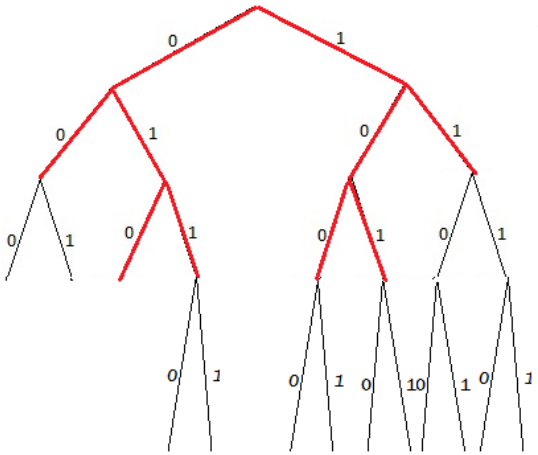
# Processus d'élagage



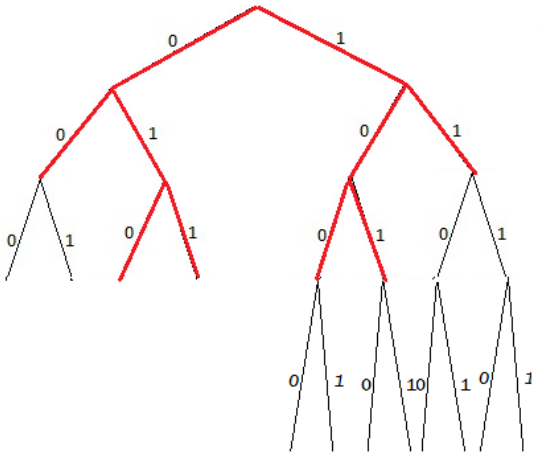
# Processus d'élagage



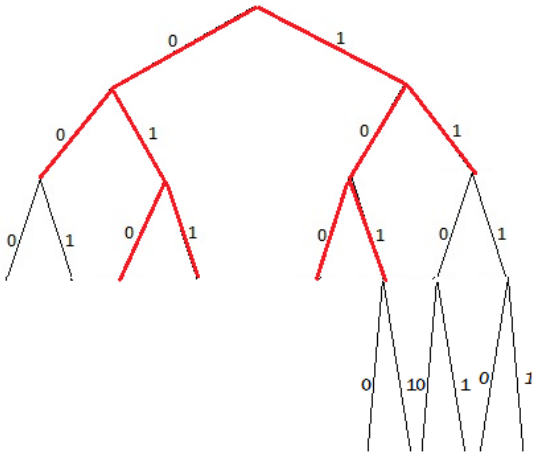
# Processus d'élagage



# Processus d'élagage



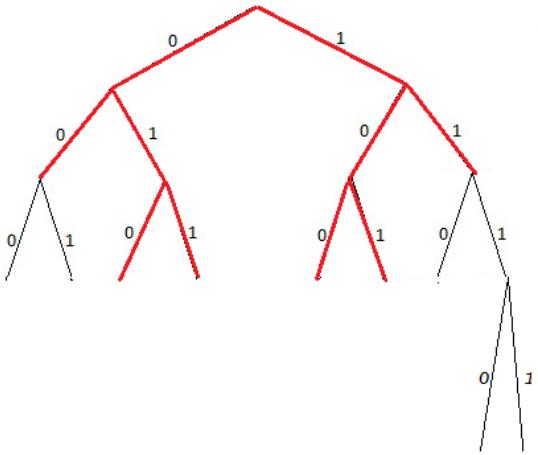
# Processus d'élagage



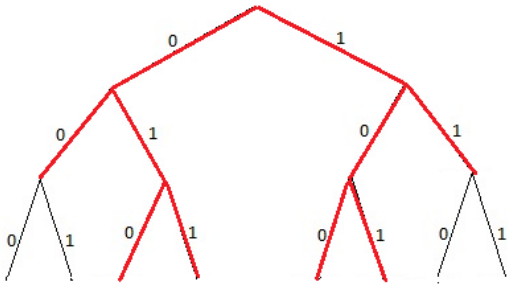




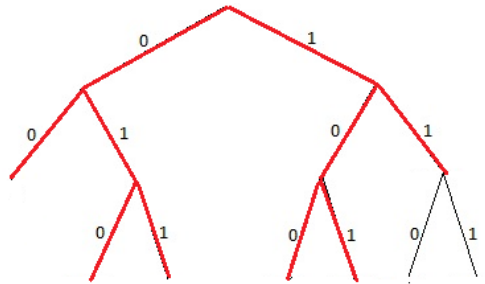
# Processus d'élagage



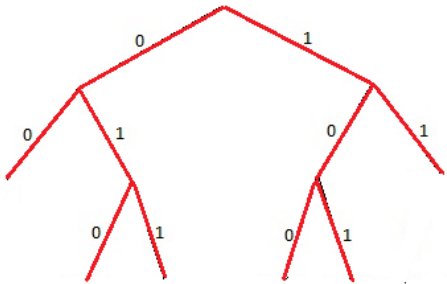
# Processus d'élagage



## Processus d'élagage



# Processus d'élagage



Pour tester nos résultats, nous simulons une VLHMM

$(X_k, Y_k)_{k=1..n}$  où

- $\mathbb{X} = \{0, 1\}$
- l'arbre de contextes minimal est celui représenté précédemment.
- le vrai paramètre d'émission est donné par  $m_0^* = 0$ ,  $m_1^*$  valant 2,3 ou 4 (cf tableau suivant), et  $\sigma^{2,*} = 1$ .
- $n$  varie entre 100 et 50000

Nous comparons notre estimateur  $\hat{\tau}_n$  avec l'estimateur BIC, qui vérifie

$$\text{pen}_{BIC}(n, \tau) = \frac{|\mathbb{X}| - 1}{2} |\tau| \log n$$






qui est beaucoup moins "lourde" que  $\text{pen}_\alpha(n, \tau)$ .

## Résultat de simulations :

$\tau^* = \tau_1^*,  \tau^*  = 6$						
$n/m_1^*$	Penalty (5)			BIC penalty		
	2	3	4	2	3	4
100	2	2	2	2	3	3
1000	2	2	2	7	6	6
2000	2	2	4	6	6	6
5000	2	4	4	7	6	6
10000	4	6	6	7	6	6
20000	5	6	6	6	6	6
30000	5	6	6	6	6	6
40000	6	6	6	7	6	6
50000	6	6	6	7	6	6

Table I: Case  $\tau^* = \tau_1^*$ . Comparison of  $|\hat{\tau}_n|$  between our estimator and the BIC estimator for different values of  $n$  and  $m_1^*$ .

Merci beaucoup pour votre attention !!!

-  J. Rissanen, “A universal data compression system,” *IEEE Trans. Inform. Theory*, vol. 29, pp. 656 – 664, 1983.
-  I. Csiszár and Z. Talata, “Context tree estimation for not necessarily finite memory processes, via bic and mdl,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 1007–1016, 2006.
-  Z. L. W. J. Wang, Y. and Z. Liu, “Mining complex time-series by learning markovian models,” in *Proceedings ICDM'06, sixth international conference on data mining, China, 2005*.
-  Y. Wang, “The variable-length hidden markov model and its applications on sequential data mining,” Département of computer science, Tech. Rep., 2005.
-  T. Dumont, “Context tree estimation in variable length hidden markov models,” *Soumis*, 2011.