Un formalisme assez général pour l'étude des systèmes complexes

Gregor Chliamovitch & Bastien Chopard
Université de Genève

8 novembre 2013

Incipit

"Un système est 'complexe' lorsqu'il est plus que la somme de ses parties."

Les chaînes de Markov comme formulation universelle

- La plupart des systèmes dynamiques peuvent être exprimés comme chaîne de Markov
- N agents booléens $\Rightarrow 2^N$ états possibles
- Petit espace des états non-markovien $\Leftrightarrow$ espace des états markovien mais moins petit
- Règle d'évolution usuelle
  $p(s_j, t+k; s_i, t) = p(s_i, t)(W^k)_{ij}$
  $p(s_j, t+k) = \sum_i p(s_j, t+k; s_i, t)$
- **W** est dans un sens prépondérant par rapport aux détails de la dynamique
- Vorsicht : état $\neq$ agent !

La théorie de l'information en 30 secondes

- *Entropie de Shannon* : $H(X) := \sum_i p(x_i) \ln p(x_i)$
- *Information mutuelle* : $I(X,Y) := H(X) - H(X|Y)$
- On a aussi

$$\begin{aligned} I(X,Y) &= H(X) + H(Y) - H(X,Y) \\ &= \sum_{i,j} p(x_i, y_j) \ln \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \end{aligned}$$

- *Distance de Kullback-Leibler* $\mathcal{D}(p||q) := \sum p \ln \frac{p}{q}$

Multi-information

- $\{X\}_N$ un ensemble de N variables
- On le découpe en $\{X\}_N = \{X\}_L \cup \{X\}_M$
- On définit *l'information bipartite*

$$M_{L,M} = H(\{X\}_L) + H(\{X\}_M) - H(\{X\}_N) \geq 0$$

- On redécoupe $\{X\}_L = \{X\}_{L_1} \cup \{X\}_{L_2}$
- On définit *l'information tripartite*

$$M_{L_1,L_2,M} = H(\{X\}_{L_1}) + H(\{X\}_{L_2}) + H(\{X\}_M) - H(\{X\}_N)$$

- Découpage plus fin $\Rightarrow$ M croissant

Multi-information (suite)

- On définit de façon générale *l'information n-partite*

$$M_{S_1,\dots,S_n} = \sum_{i=1}^{n} H(\{X\}_{S_i}) - H(\{X\}_N)$$

- ... dont la *multi-information* est l'exemple le plus raffiné

$$M_{X_1,\dots,X_N} = \sum_{i=1}^{N} H(X_i) - H(\{X\}_N)$$

- Notez que

$$M_{X_1,\dots,X_N} = \sum p(\{x\}_N) \ln \frac{p(\{x\}_N)}{\prod_i p(x_i)} = \mathcal{D}(p(\{x\}_N) || \prod_i p(x_i))$$

Quelques remarques

La multi-information...

- fournit une caractérisation "en un coup d'oeil" du degré d'interdépendance dans un ensemble
- peut être envisagé comme une (pseudo-)distance à l'indépendance dans l'espace des distributions
- est égale à $\sum_{i=1}^{N} H(X_i) - H(\{X\}_N)$
  $\Rightarrow$ Qu'en est-il d'un théorème M ?
- peut être normalisée : $\mu \equiv \frac{M}{(\prod_{i=1}^{N} G_i)^{1/N}}$ où $G_i \equiv \sum_{j \neq i} H(X_j)$
- peut être utilisée à différents niveaux de structure
- généralise des mesures plus simples
- permet une analyse temporelle lorsque utilisée dans un cadre markovien

Cas d'école : le modèle de vote sur un anneau

Modèle de vote :

- N agents booléens sur un réseau
- Calculer le pourcentage r de voisins optimistes
- Soit $f(r)$ la probabilité pour un agent d'être optimiste à la prochaine itération
- Prenons $f(r) = r$
  $\Rightarrow$ présence d'états absorbants
- On peut déduire la matrice de transition $\mathbf{W}$

Cas d'école (suite)

Calculons la valeur asymptotique de l'*information mutuelle retardée* entre deux agents X_0 et X_3, à savoir

$$\lim_{t\to\infty} \sum p(X_3, t; X_0, 0) \ln \frac{p(X_3, t; X_0, 0)}{p(X_3, t)p(X_0, 0)}$$

- On a besoin de

$$W^\infty = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 \\
3/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
3/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/4 \\
3/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/4 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/4 \\
1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}$$

- Distribution initiale telle que $p(s_i, 0) \equiv \rho^{\#1_i}(1-\rho)^{\#0_i}$

Cas d'école (suite)

- $p(s_j, \infty; s_i, 0) = p(s_i, 0)(W^\infty)_{ij}$
- Seuls survivent $p(s_0, \infty; s_i, 0) = p(s_i, 0)(W^\infty)_{i0}$ et $p(s_{15}, \infty; s_i, 0) = p(s_i, 0)(W^\infty)_{i,15}$
- Ainsi, par exemple,

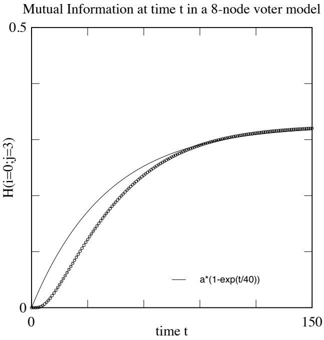
$$\begin{aligned} p(X_3 = 0, \infty; X_0 = 1, 0) &= \sum_{i=1,3,5,7,9,11,13,15} p(s_0, \infty; s_i, 0) \\ &= \sum_{i=1,3,5,7,9,11,13,15} p(s_i, 0)(W^\infty)_{i0} \\ &= \ldots \\ &= \frac{3}{4}\rho - \frac{3}{4}\rho^2 \end{aligned}$$

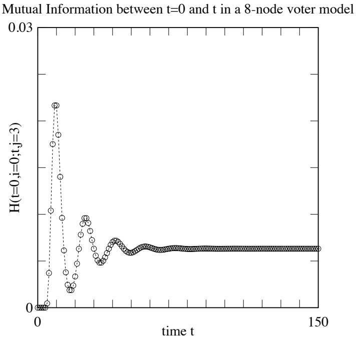
- $p(0, \infty; 0, 0)$, $p(1, \infty; 0, 0)$ et $p(1, \infty; 1, 0)$ similaires

Cas d'école (fin)

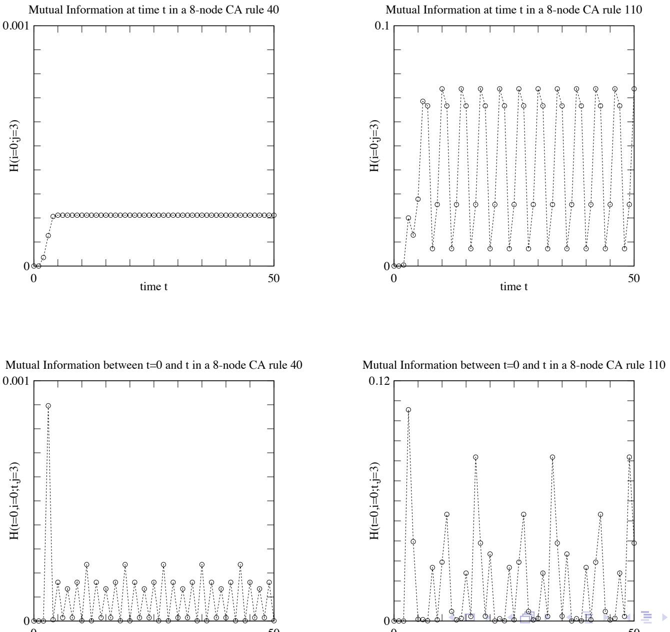
- Les probabilités pour un seul agent sont determinées par sommation, p. ex.
  $p(X_3 = 0, \infty) = p(X_3 = 0, \infty; X_0 = 0, 0) + p(X_3 = 0, \infty; X_0 = 1, 0)$
- On peut calculer les entropies
- $I(X_3, \infty; X_0, 0)$

Quelques illustrations : modèle de vote

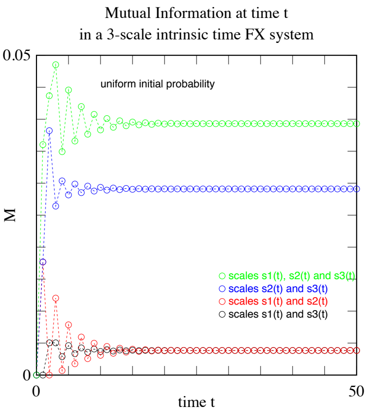


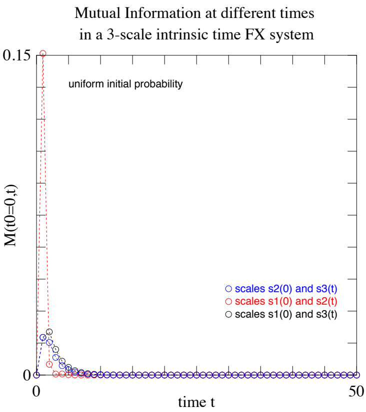


Quelques illustrations : automates cellulaires 1D

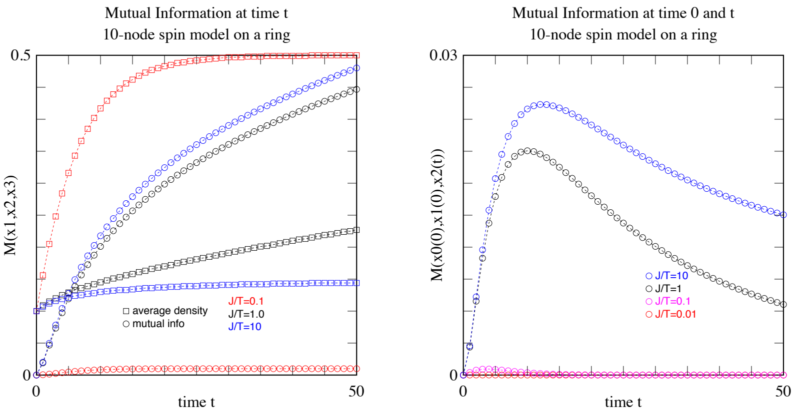


Quelques illustrations : réseau intrinsèque





Quelques illustrations : modèle d'Ising



Une autre approche

Est-il possible de décomposer M en une somme de termes non-négatifs M_r dont chacun caractérise la contribution d'un ordre d'interaction donné ?

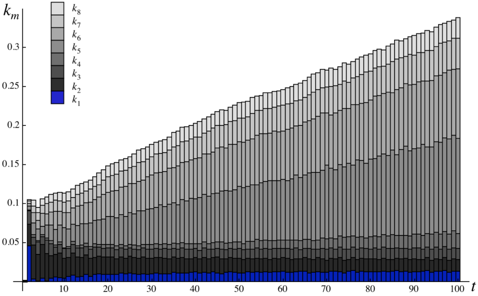


Figure: Règle 110

Cas bête : $k_r = H(X_r | X_2, ..., X_{r-1}) - H(X_r | X_1, ..., X_{r-1})$

Une autre approche (suite)

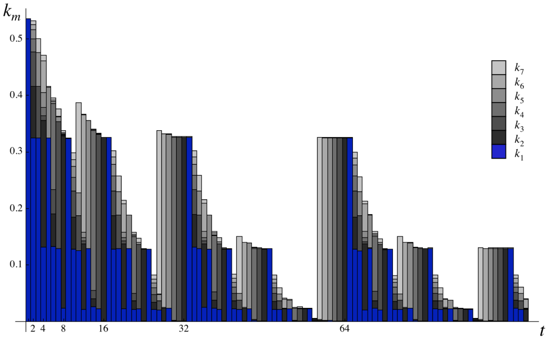


Figure: Règle 60

Une autre approche : entropie maximum

Quelle est la distribution minimalement biaisée p_2^{ME} possédant des 2-marginales données ?

- On maximise $H(p_2^{ME})$ sous les contraintes

$$\sum p_2^{ME}(x_1, ..., x_N) = 1$$

$$\sum_{x_i \neq x_m, x_n} p_2^{ME}(x_1, ..., x_N) = p(x_m, x_n)$$

- Réponse :

$$p_2^{ME}(x_1, ..., x_N) = \prod_{m,n=1}^{N} f_{m,n}(x_m, x_n)$$

- Malheureusement $f_{m,n}(x_m, x_n) \neq p(x_m, x_n)$

Reformuler M

- On peut récrire

$$M(x_1, ..., x_N) = \mathcal{D}(p||p_1^{ME})$$

- Quid de $\mathcal{D}(p||p_2^{ME})$, $\mathcal{D}(p||p_3^{ME})$, etc. ?
- p_k^{ME} intègre les interactions *jusqu'à* l'ordre k
- On définit

$$M_k := \mathcal{D}(p||p_{k-1}^{ME}) - \mathcal{D}(p||p_k^{ME})$$

- Par sommation télescopique

$$\sum_{k=2}^{N} M_k = M(X_1, ..., X_N)$$

- On est content

Conclusions

- Information multi-partite et chaînes de Markov sont deux outils très généraux et adaptables
- Mélanger les deux semble pertinent
- Permet d'investiguer au niveau du $\mathbf{W}$ ou au niveau de la dynamique entre les agents
- Angle d'attaque pour les *tipping points* ?
- Permet *quelques* calculs analytiques
- Numériquement exponentiellement lourd
- Approche via ME plus élégante mais moins générale ?
- *Quelle* entropie ?