

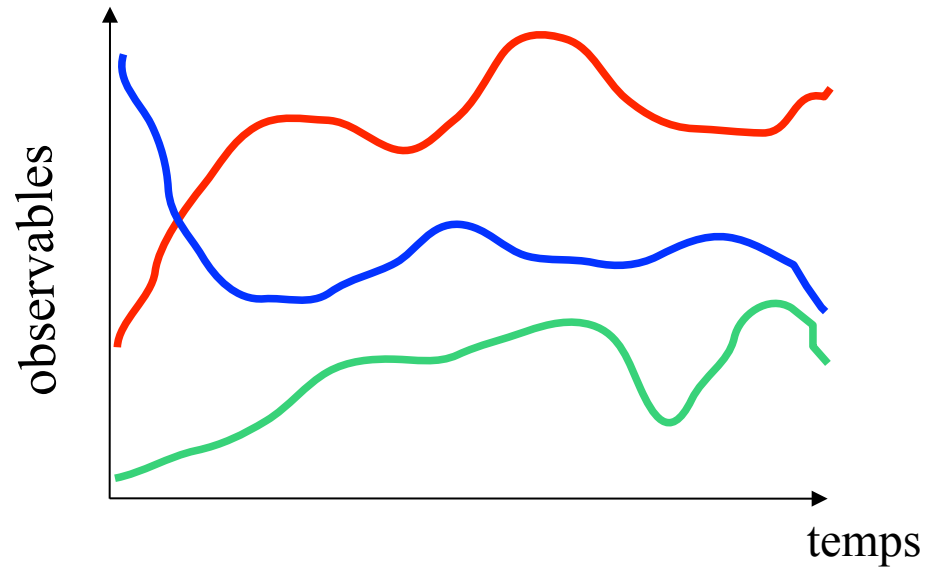
Inférence, méthodes « physiques » et applications

Simona Cocco - Rémi Monasson
cocco@lps.ens.fr
monasson@lpt.ens.fr

1. Introduction : à quoi sert l'inférence?
2. Méthodes : inférence bayésienne et algorithmes
3. Application aux réseaux de neurones (représentation de l'espace)
4. Application aux protéines (structure, paysage de fitness)

1- Introduction et exemples

Inférence de modèle à partir d'observations



dynamique (corrélations) $\longrightarrow$ mécanismes sous-jacents (interactions) $\longrightarrow$ modèles prédictifs

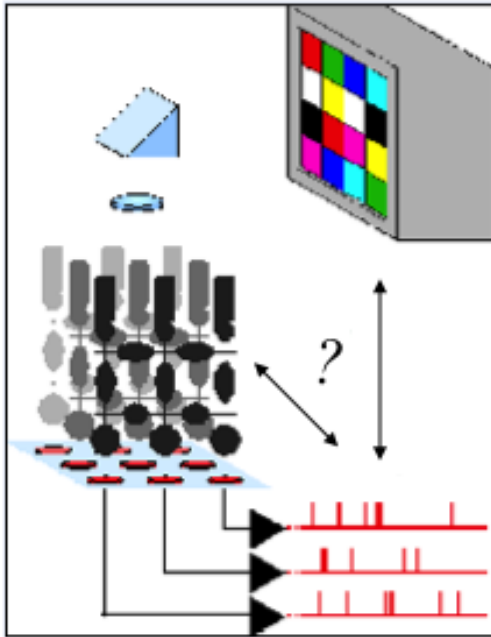
Beaucoup de problèmes/difficultés :

limitations temporelles et spatiales sur l'échantillonnage,
bruit (de mesure, dynamique),
classes de modèles (nombre de paramètres),
coût computationnel, ...

} signal/bruit < 1

$\neq$ physique (interactions uniformes $\longrightarrow$ modèles avec peu de paramètres,
reproductibilité $\longrightarrow$ bon échantillonnage, équilibre thermique)

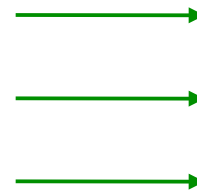
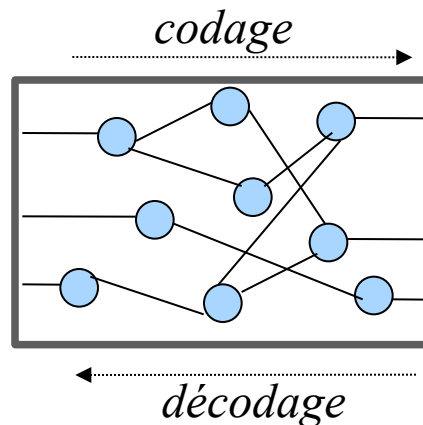
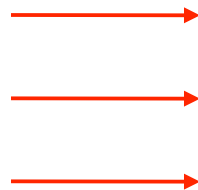
In vitro recording on retina



Schnitzer, Meister (2003)
Schneidman et al. (2006)

Exemple 1 : **Activité d'une population neuronale**

**Stimulus
(entrée)**

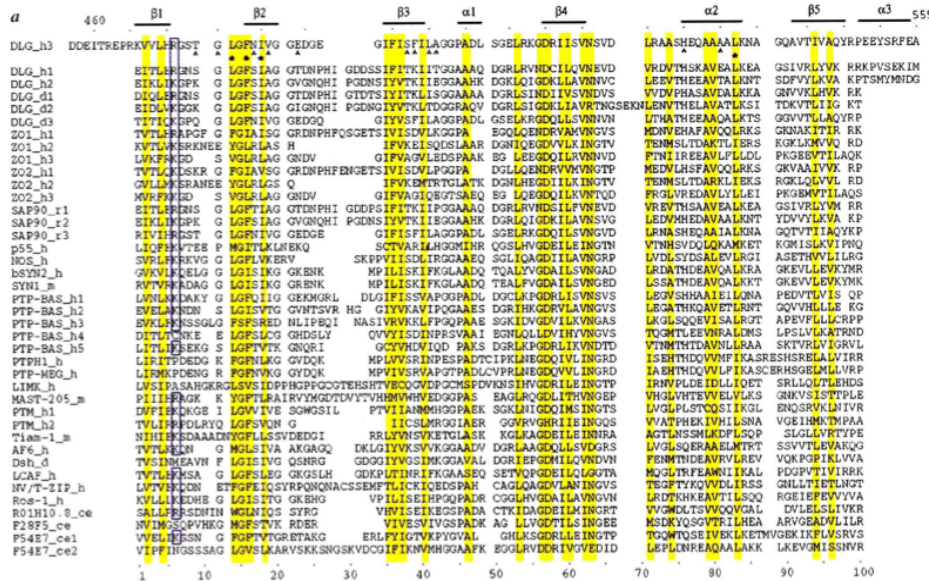


**Activité
(sortie)**

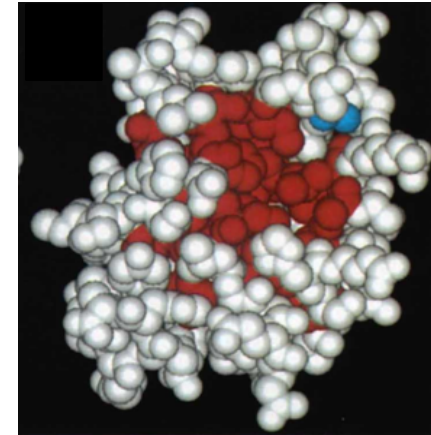
Compréhension des principes de codage? Représentation et traitement de l'information?

Exemple 2:

Coévolution de résidus dans les familles de protéines

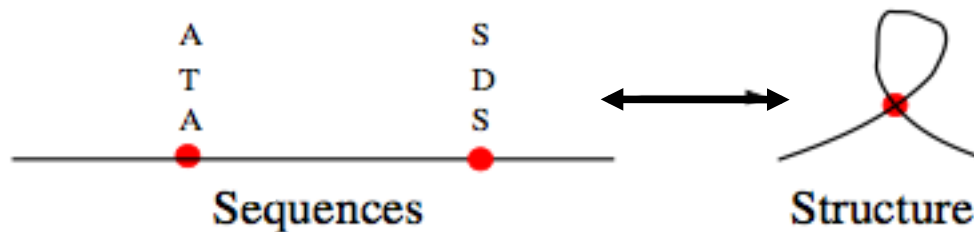


PDZ domains



Morais Cabral et al. (1996)

- Conservation des résidus (utilisée pour la reconstruction phylogénétique ...)
- Corrélations de paires? Reliées aux contraintes structurales et fonctionnelles?



Gobel et al. (1994)

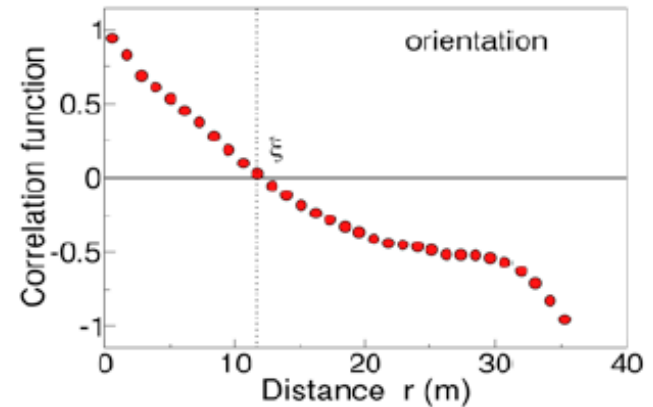
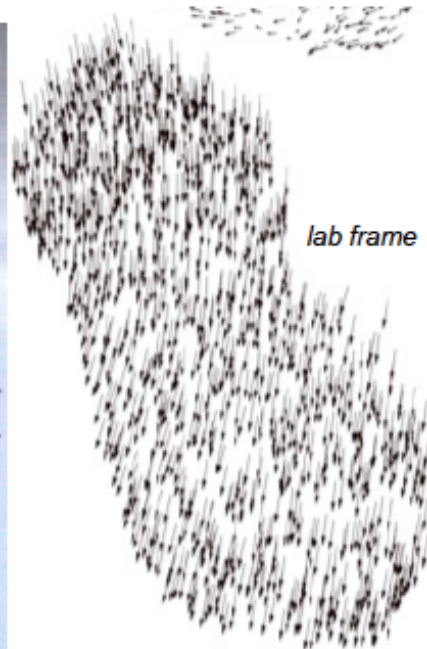
Exemple 3:

Ordre et organisation dans les volées d'oiseau

can one get global order from local communication ?



amazing coordination
in flight direction



ordering over large distances across flock

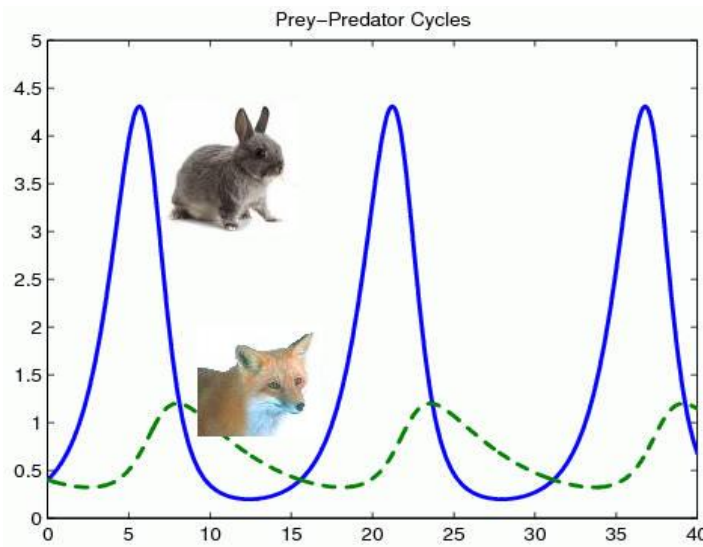
statistical mechanics analysis shows that
local interactions are enough

birds mimic a fixed number of neighbors,
regardless of their distance

Exemple 4:

Dynamique couplée d'espèces dans un système écologique

Ecologie des populations : interactions entre espèces (existence, additivité, ...)



Equations de Lotka-Volterra :

$$\frac{dN_i}{dt} = N_i \left(r_i - \sum_j A_{ij} N_j \right)$$

Questions : Bruit de mesure (échantillonnage très incomplet),
Espèces (interactions) inconnues, ...

Est-ce qu'il y a un signal fiable dans ces interactions? (additivité?)
Exploiter les interactions pour prédire la dynamique, **extinction** ?

2 - Méthodes d'inférence :

a. approche bayésienne

b. algorithmes

Une illustration

Boîte A

20 biscuits ordinaires
+
20 biscuits au chocolat

Boîte B

10 biscuits ordinaires
+
30 biscuits au chocolat

- On retire un biscuit au hasard d'une des deux boîtes (sans savoir laquelle)
- Il se trouve que ce biscuit est au chocolat
- Quelle est la probabilité qu'il a été retiré de la boîte A? de B?

Approche bayésienne

1. Connaissance a priori sur les probabilités des réponses possibles (sans connaître les observations, ici = biscuit au chocolat) :

$$P_{\text{prior}}(A) = P_{\text{prior}}(B) = \frac{1}{2}$$

2. Connaissance sur les probabilités des observations si l'on connaît la bonne réponse :

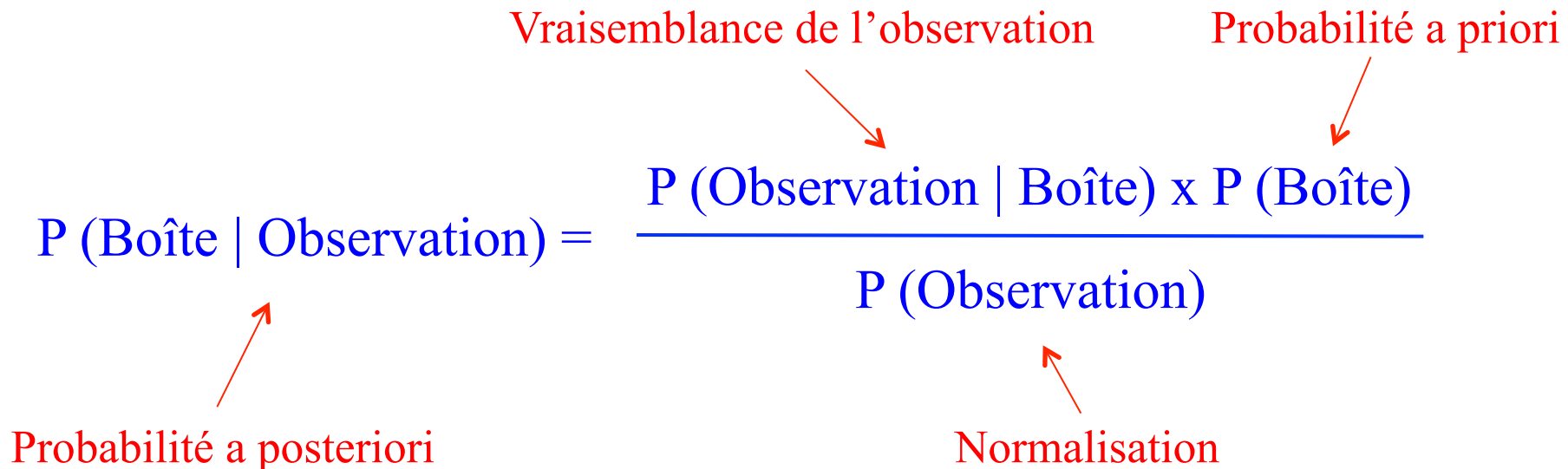
$$P(\text{biscuit} = \text{ordinaire} \mid A) = \frac{1}{2}, P(\text{biscuit} = \text{chocolat} \mid A) = \frac{1}{2}$$

$$P(\text{biscuit} = \text{ordinaire} \mid B) = \frac{1}{4}, P(\text{biscuit} = \text{chocolat} \mid B) = \frac{3}{4}$$

Approche bayésienne

Théorème de Bayes :

$$\begin{aligned}\text{Proba (Observation , Boîte)} &= \text{Proba (Observation | Boîte)} \times \text{Proba(Boîte)} \\ &= \text{Proba (Boîte | Observation)} \times \text{Proba(Observation)}\end{aligned}$$



The diagram illustrates Bayes' Theorem with the following equation and annotations:

$$P(\text{Boîte} | \text{Observation}) = \frac{P(\text{Observation} | \text{Boîte}) \times P(\text{Boîte})}{P(\text{Observation})}$$

Annotations (in red):

- Vraisemblance de l'observation** (Likelihood) points to $P(\text{Observation} | \text{Boîte})$.
- Probabilité a priori** (Prior probability) points to $P(\text{Boîte})$.
- Probabilité a posteriori** (Posterior probability) points to $P(\text{Boîte} | \text{Observation})$.
- Normalisation** points to $P(\text{Observation})$.

Une illustration

Application aux boîtes de biscuits :

$$P(A | \text{Chocolat}) = \frac{P(\text{Chocolat} | A) \times P_{\text{prior}}(A)}{P(\text{Chocolat})} = \frac{\frac{1}{2} \times \frac{1}{2}}{P(\text{Chocolat})}$$

$$P(B | \text{Chocolat}) = \frac{P(\text{Chocolat} | B) \times P_{\text{prior}}(B)}{P(\text{Chocolat})} = \frac{\frac{3}{4} \times \frac{1}{2}}{P(\text{Chocolat})}$$

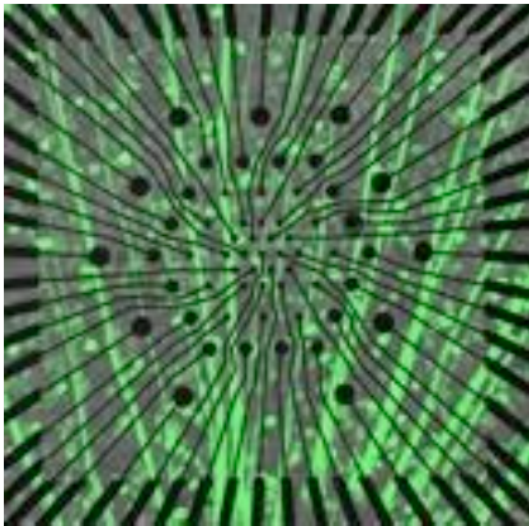
puis : $P(\text{Chocolat}) = 1/4 + 3/8 = 5/8$

$$P(A | \text{Chocolat}) = (1/4) / (5/8) = 2/5$$

$$P(B | \text{Chocolat}) = (3/8) / (5/8) = 3/5$$

Et si on avait retiré un biscuit ordinaire?

Application à des données neuronales



stimulus visuel



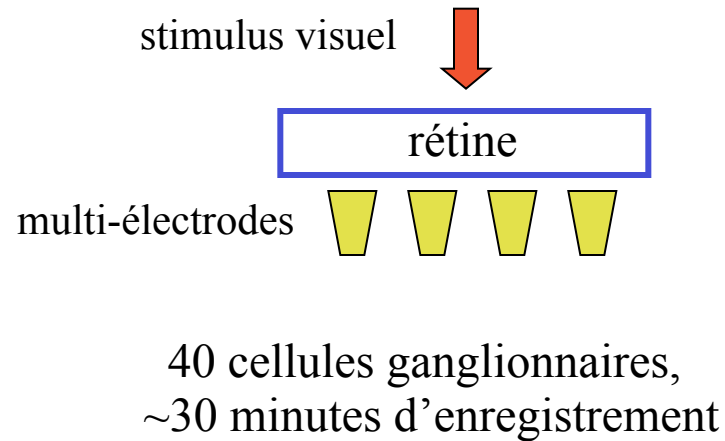
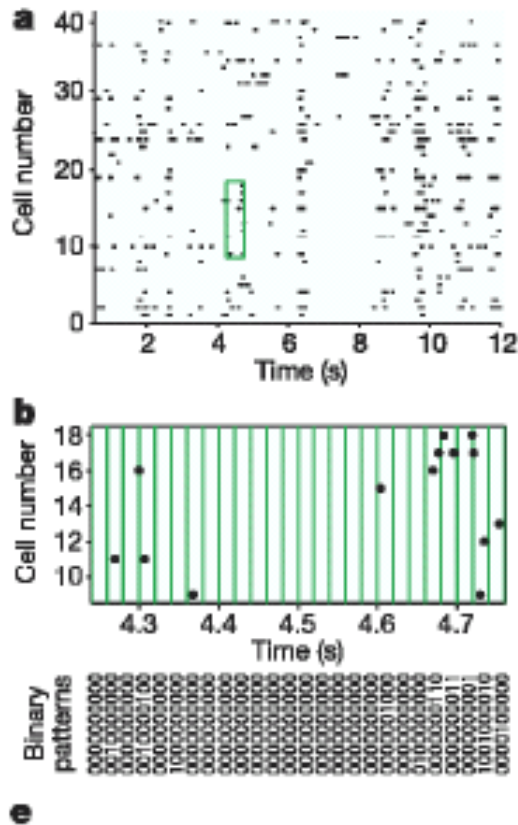
rétine



multi-électrodes

dizaines-centaines de cellules ganglionnaires,
durée d'enregistrement = heures

Application à des données neuronales



Modèle pour la distribution d'activité
 $p(\sigma_1, \sigma_2, \dots, \sigma_N)$
(mots binaires de N symboles 0,1) ?

MAIS : Nombre de configurations possibles pour $N=40$: $2^{40} \sim 10^{12}$
Nombre de fenêtres temporelles (20 ms) : 10^5

Application à des données neuronales

Modèle de cellules indépendantes :

Vraisemblance d'une configuration : $p(\sigma_1, \sigma_2, \dots, \sigma_N | \{h_i\}) = \prod_i \frac{e^{h_i \sigma_i}}{1 + e^{h_i}}$

Vraisemblance de toutes les configurations enregistrées :

$$p(\{\sigma_i^t\} | \{h_i\}) = \prod_t p(\sigma_1^t, \sigma_2^t, \dots, \sigma_N^t | \{h_i\})$$

Connaissance a priori sur les « champs » : aucune, $P_{\text{prior}}(\{h_i\})$ est uniforme

Probabilité a posteriori des « champs » (Bayes) :

$$\begin{aligned} p(\{h_i\} | \{\sigma_i^t\}) &\propto \prod_t p(\sigma_1^t, \sigma_2^t, \dots, \sigma_N^t | \{h_i\}) \\ &\propto \prod_i \frac{e^{\sum_t h_i \sigma_i^t}}{(1 + e^{h_i})^T} = \prod_i \frac{e^{h_i N_i}}{(1 + e^{h_i})^T} \end{aligned}$$

(N_i = nombre de spikes du neurone i)

Application à des données neuronales

Probabilité a posteriori
des « champs » (Bayes) :

$$\propto \prod_i \frac{e^{h_i N_i}}{(1 + e^{h_i})^T} \quad (N_i = \text{nombre de spikes du neurone } i)$$

Maximisation de la probabilité a posteriori par rapport à h_i :

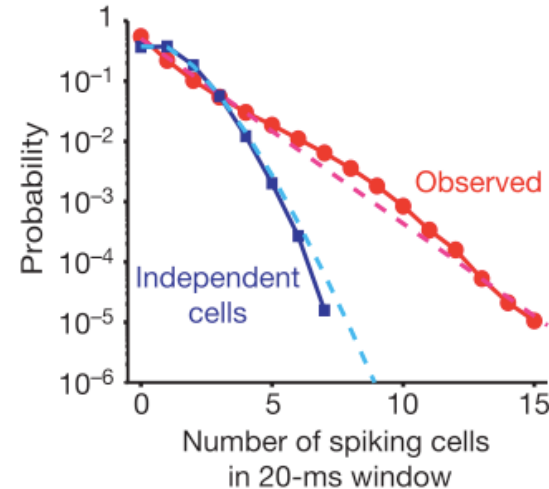
$$0 = \frac{\partial}{\partial h_i} \log \prod_i \frac{e^{h_i N_i}}{(1 + e^{h_i})^T} = N_i - T \frac{e^{h_i}}{1 + e^{h_i}} \Rightarrow h_i = \log \left(\frac{f_i}{1 - f_i} \right)$$

$$\text{avec} \quad f_i = \frac{N_i}{T} \quad (\text{taux de décharge})$$

On a maintenant complètement défini la distribution de l'activité neuronale ...
Comparaison avec les données?

Activité neuronale dans la rétine

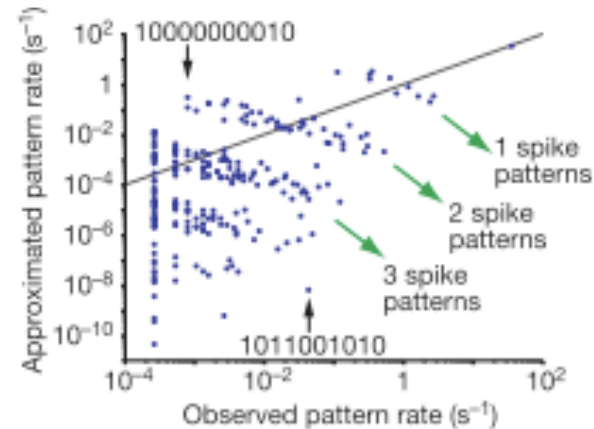
Capacité à prédire les moments
d'ordres supérieurs :



Schneidmann et al.

Nature (2006)

Capacité à prédire les fréquences des
configurations observées :



Mauvais modèle : les corrélations entre neurones sont importantes ...

Mais en ce qui concerne l'information mutuelle ...

Retinal ganglion cells act largely as independent encoders

S. Nirenberg, S. M. Carcieri, A. L. Jacobs & P. E. Latham

*Department of Neurobiology, University of California Los Angeles,
10833 Le Conte Avenue, Los Angeles, California 90095-1763, USA*

Correlated firing among neurons is widespread in the visual system. Neighbouring neurons, in areas from retina to cortex, tend to fire together more often than would be expected by chance. The importance of this correlated firing for encoding visual information is unclear and controversial¹⁻⁵. Here we examine its importance in the retina. We present the retina with natural stimuli and record the responses of its output cells, the ganglion cells. We then use information theoretic techniques to measure the amount of information about the stimuli that can be obtained from the cells under two conditions: when their correlated firing is taken into account, and when their correlated firing is ignored. We find that more than 90% of the information about the stimuli can be obtained from the cells when their correlated firing is ignored.

This indicates that ganglion cells act largely independently to encode information, which greatly simplifies the problem of decoding their activity.

Modèles avec interactions

Données

(1,0,0,0,1,0,1,1, ..., 1,0,0,1,1,0)
(0,1,0,0,0,1,1,1, ..., 0,1,1,0,0,0)
(1,1,0,1,0,1,1,0, ..., 1,1,0,0,1,1)
...
(0,1,0,0,1,0,0,0, ..., 0,0,0,1,0,0)

$\Rightarrow$ Probability $p(\sigma_1, \sigma_2, \dots, \sigma_N)$?

Modèle à cellules non indépendantes

$$p(\sigma_1, \sigma_2, \dots, \sigma_N | \{h_i\}, \{J_{ij}\}) = \frac{e^{\sum_i h_i \sigma_i + \sum_{i < j} J_{ij} \sigma_i \sigma_j}}{Z[\{h_i\}, \{J_{ij}\}]}$$

- Généralisation du modèle à cellules indépendantes ($J=0$)
- Correspond au modèle d'Ising de la physique statistique ...
- Comment déterminer les champs et couplages?

Modèles avec interactions

Probabilité a posteriori des paramètres

$$\propto \prod_t p(\sigma_1^t, \sigma_2^t, \dots, \sigma_N^t \mid \{h_i\}, \{J_{ij}\}) = \prod_t \frac{e^{\sum_i h_i \sigma_i^t + \sum_{i < j} J_{ij} \sigma_i^t \sigma_j^t}}{Z[\{h_i\}, \{J_{ij}\}]}$$

$$Z[\{h_i\}, \{J_{ij}\}] = \sum_{\{\sigma_i\}} e^{\sum_i h_i \sigma_i + \sum_{i < j} J_{ij} \sigma_i \sigma_j}$$

est la fonction de partition
(très compliquée à calculer ...)

On veut minimiser (sur les champs et les couplages) :

$$S = - \sum_t \log p(\sigma_1^t, \sigma_2^t, \dots, \sigma_N^t \mid \{h_i\}, \{J_{ij}\}) = - \sum_i h_i N_i - \sum_{i < j} J_{ij} N_{ij} + T \log Z[\{h_i\}, \{J_{ij}\}]$$

 **cross-entropie**

avec N_i = nombre de spikes du neurone i ,
 N_{ij} = nombre de spikes « simultanés » de i et j

Modèles avec interactions

Equations de minimisation :

$$S = - \sum_t \log p(\sigma_1^t, \sigma_2^t, \dots, \sigma_N^t \mid \{h_i\}, \{J_{ij}\}) = - \sum_i h_i N_i - \sum_{i < j} J_{ij} N_{ij} + T \log Z[\{h_i\}, \{J_{ij}\}]$$

$$\begin{aligned} \frac{\partial S}{\partial h_i} = 0 &\Rightarrow \boxed{\frac{N_i}{T}} = \frac{\partial}{\partial h_i} \log Z[\{h_i\}, \{J_{ij}\}] = \boxed{\langle \sigma_i \rangle} \\ \frac{\partial S}{\partial J_{ij}} = 0 &\Rightarrow \boxed{\frac{N_{ij}}{T}} = \frac{\partial}{\partial J_{ij}} \log Z[\{h_i\}, \{J_{ij}\}] = \boxed{\langle \sigma_i \sigma_j \rangle} \end{aligned}$$

moyennes sur
les données

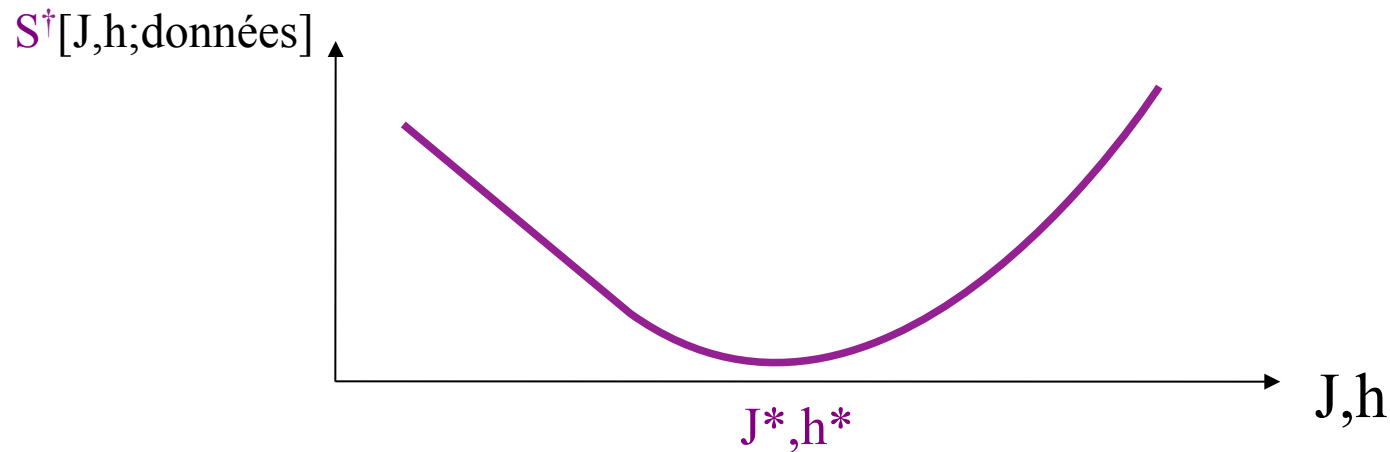
moyennes sur
la distribution d'Ising

Il faut trouver les couplages et champs du modèle d'Ising telles que les valeurs moyennes des observables σ_i , $\sigma_i \sigma_j$ calculées avec ce modèle coïncident avec les données

Modèle avec interactions

Cross-entropie des données ($= \sigma^1, \dots, \sigma^T$)

$$S^\dagger [J, h] = \sum_{t=1}^T -\log p(\sigma^t; J, h)$$
$$= T \log Z[J, h] - \sum_{i < j} J_{ij} N_{ij} - \sum_i h_i N_i$$



- La fonction S est convexe donc le minimum est unique
- Lien avec le principe d'entropie maximale (contraintes = les N_i et les N_{ij})

$$\begin{array}{c} S[p] \longrightarrow \hspace{10em} \longleftarrow S^\dagger [J, h] \\ \hline S[p^{\text{MAXENT}}] = S^\dagger [J^*, h^*] \end{array}$$

Modèle avec interactions

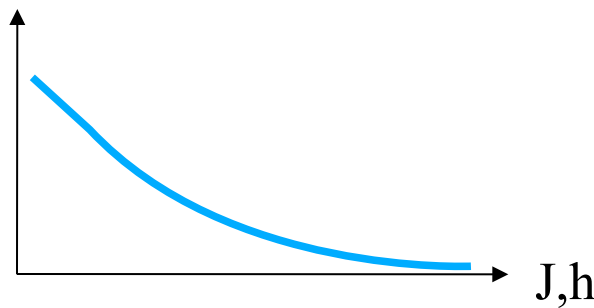
Hessienne de la cross-entropie

$$\frac{\partial^2 S(J,h;\text{données})}{\partial (J,h) \partial (J,h)} = \begin{pmatrix} \langle \sigma_i \sigma_j \sigma_k \sigma_l \rangle - \langle \sigma_i \sigma_j \rangle \langle \sigma_k \sigma_l \rangle & \langle \sigma_i \sigma_j \sigma_k \rangle - \langle \sigma_i \rangle \langle \sigma_j \sigma_k \rangle \\ \langle \sigma_i \sigma_j \sigma_k \rangle - \langle \sigma_i \sigma_j \rangle \langle \sigma_k \rangle & \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \end{pmatrix}$$

où $\langle \rangle =$ moyenne de Gibbs avec le modèle d'Ising

$$X = \begin{pmatrix} x_{ij} \\ x_k \end{pmatrix} \rightarrow X^T (+) X = \langle \left(\sum_{i < j} x_{ij} (\sigma_i \sigma_j - \langle \sigma_i \sigma_j \rangle) + \sum_k x_k (\sigma_k - \langle \sigma_k \rangle) \right)^2 \rangle$$

Modes zéros ?



Prior sur les paramètres

Méthode 1 : Apprentissage machine de Boltzmann

Ackley, Hinton, Sejnowski (1985)

- On part de J_{ij} et h_i quelconques
- On calcule $\langle \sigma_i \sigma_j \rangle$ et $\langle \sigma_k \rangle$ avec des simulations de Monte Carlo
- On compare avec les valeurs moyennes sur les données (c_{ij} , m_k) et on met à jour

$$\begin{aligned} J_{ij} &\rightarrow J_{ij} - a (\langle \sigma_i \sigma_j \rangle - c_{ij}) \\ h_k &\rightarrow h_k - a (\langle \sigma_k \rangle - m_k) \end{aligned}$$

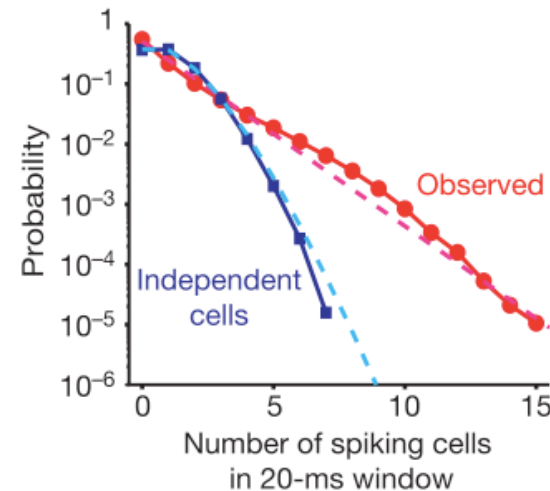
Problème : peut être très lent (exponentiel en N ...)

Application aux enregistrements de cellules ganglionnaires

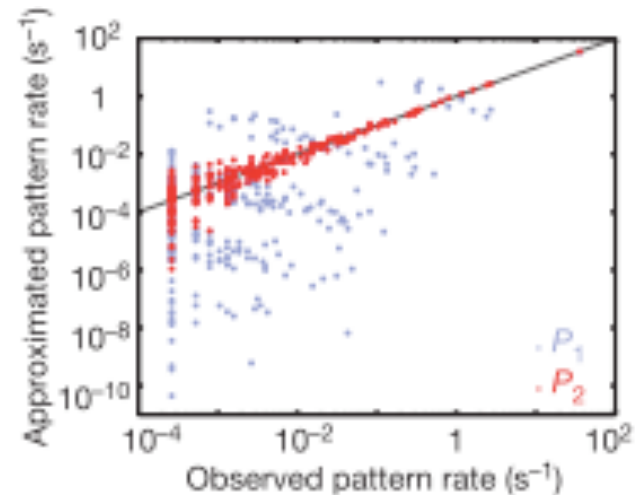
Schneidmann et al.

Nature (2006)

Capacité à prédire les moments d'ordres supérieurs :



Capacité à prédire les fréquences des configurations observées :



**Très bon modèle : les corrélations entre neurones sont importantes ...
(99% de l'information mutuelle)**

Méthode 2 : Champ moyen

Résultat de l'approximation de champ moyen

Matrice de corrélations connexes : $C_{ij} = \langle \sigma_i \sigma_j \rangle_{\text{données}} - \langle \sigma_i \rangle_{\text{données}} \langle \sigma_j \rangle_{\text{données}}$

Matrice de couplages approchées : $J_{ij} \approx -(C^{-1})_{ij}$

Propriétés :

- Rapide (temps N^3)
- Matrice c affectée par le bruit, ce qui rend l'inversion peu fiable

Matrice empirique $\hat{c} = \left(c_{ij} \pm \underset{\substack{\uparrow \\ \text{nb. of data}}}{B^{-1/2}}} \right) \overset{\substack{\updownarrow \\ N}}{\quad} \Rightarrow \text{erreur d'ordre } (N/B)^{1/2} \dots$