From Variable neighborhood random fields to the estimation of interaction graphs.

Joint work with Enza Orlandi (first part) and with A. Duarte, A. Galves, G. Ost (2nd part)

Rouen, September 2016 - Conference dedicated to Enza's memory

• I started working with Enza in 2008 : we discussed, together with Antonio Galves, about the *perfect simulation of infinite range Gibbs measures.* Main technical ingredient : Kalikow-type decomposition.
$\Rightarrow$ *Perfect simulation of infinite range Gibbs measures and coupling with their finite range approximations*, JSP 2010.

• I started working with Enza in 2008 : we discussed, together with Antonio Galves, about the *perfect simulation of infinite range Gibbs measures.* Main technical ingredient : Kalikow-type decomposition.
$\Rightarrow$ *Perfect simulation of infinite range Gibbs measures and coupling with their finite range approximations*, JSP 2010.
Followed : Lot of visits of Enza to Paris and of me to Rome....

• Continued in : *Kalikow-type decomposition for multicolor infinite range particle systems.* AAP 2013 (with N. Garcia, A. Galves)

• What I am going to talk about today : Neighborhood radius estimation in Variable-neighborhood random fields (VNRF) SPA, 2011.

What are VNRF ? ? ?

Observe a random field μ on $\mathbb{Z}^d$ where

- every **site** : $i \in \mathbb{Z}^d$
- might have a finite number of **colors** : $a \in A$: finite alphabet,

and where the color of a given site i depends on a finite neighborhood of the site which depends on the total configuration of the field.

Typical Example

• Consider a Markov random field of order 1, taking values ± 1 at each site.

• Report each site with probability p, and with $1-p$, put the value -1.

• Yields a VNRF : contexts are the interiors of self-avoiding paths having $+1$ on its boundary. I do not speak about wheather these contexts are finite or not here...

Typical Example

• Consider a Markov random field of order 1, taking values ± 1 at each site.

• Report each site with probability p, and with $1-p$, put the value -1.

• Yields a VNRF : contexts are the interiors of self-avoiding paths having $+1$ on its boundary. I do not speak about wheather these contexts are finite or not here...

• See : Cassandro, Galves, L. 2012, see also the literature about **factor maps** in one-dimensional frame (Chazotte and Ugalde 2011, Verbitskiy 2011, ...)

To come back to the general concept of Variable neighborhood random fields :

- **No Markovian** assumption.
- Call the relevant neighborhood needed in order to determine the color of site i an $i-$**context.**
- Hence if the total configuration is $\omega \in A^{\mathbb{Z}^d \setminus \{i\}}$, the $i-$context is

$$\text{a finite configuration } c_i(\omega) \in A^{\mathbb{Z}^d \setminus \{i\}}.$$

It is the smallest configuration needed in order to determine the value of site i, given the outside ω.
Write $\ell_i(\omega)$ for its radius (radius of the smallest ball containing it)

Context-trees

- The set of all possible contexts forms a tree : **the context-tree.**
- That means : No context $c_i(\omega)$ can be shortened. And an extension is not needed.
- The set of all contexts $c_i(\omega)$ defines a partition of all configurations $\eta \in A^{\mathbb{Z}^d \setminus \{i\}}$.

In dimension $d = 1$: Variable-length Markov chains : Rissanen (1983) (data compression), Bühlman and Wyner (1999), Galves and Leonardi (2008) (bio-informatics, protein-expression), Cénac, Chauvin, Paccaut and Pouyanne (2015) and many others.

Random Fields with variable length interactions

- $i-$**context-function** : this is a function $f : \Omega = A^{\mathbb{Z}^d} \to \mathbb{R}$ such that $f(\omega) = f(\eta)$ whenever $c_i(\omega) = c_i(\eta)$.
- **Specification :** μ defined though its specification $\{\gamma_\Lambda\}_{\Lambda \subset \mathbb{Z}^d}$.

Random Fields with variable length interactions

- $i-$**context-function** : this is a function $f : \Omega = A^{\mathbb{Z}^d} \to \mathbb{R}$ such that $f(\omega) = f(\eta)$ whenever $c_i(\omega) = c_i(\eta)$.
- **Specification :** μ defined though its specification $\{\gamma_\Lambda\}_{\Lambda \subset \mathbb{Z}^d}$.
- For simplicity : Consider only the one-point specification $\gamma_i(a|\omega)$.

Definition

A RF μ consistent with the above specification is called a VNRF if for any fixed a, $\gamma_i(a|\cdot)$ is an $i-$context function.

Context-Estimation

Statistical question : Sampling μ over an increasing sequence of finite regions $\Lambda_n \subset \mathbb{Z}^d$, is it possible to estimate $\ell_i(\omega)$, the length of the context of i given the realization ω ?

Context-Estimation

Statistical question : Sampling μ over an increasing sequence of finite regions $\Lambda_n \subset \mathbb{Z}^d$, is it possible to estimate $\ell_i(\omega)$, the length of the context of i given the realization ω ?

Translation to the framework of random fields of the algorithm *Context* introduced by Rissanen (1983).

Given the observation $\omega(\Lambda_n)$:

- Fix a site i. Start with a candidate context

$$\omega_i^{k(n)} := \{\omega_j : 0 < ||j - i|| \leq k(n)\},$$

  where $k(n) = (\log|\Lambda_n|)^{\frac{1}{2d}}$.

- Then decide to shorten or not this candidate context by using some **gain function**, for example the log-likelihood ratio statistics.

Let for any $\ell \leq k(n)$

$$N_n(\omega_i^\ell) = \sum_j 1_{\{X_j^\ell = \omega_i^\ell\}}$$

be the *total number of occurrences* of the observed pattern $\omega_i^\ell = \{\omega_j : 0 < ||j - i|| \leq \ell\}$.

Let for any $\ell \leq k(n)$

$$N_n(\omega_i^\ell) = \sum_j 1_{\{X_j^\ell = \omega_i^\ell\}}$$

be the *total number of occurrences* of the observed pattern $\omega_i^\ell = \{\omega_j : 0 < ||j - i|| \leq \ell\}$. In the same way :

$$N_n(\omega_i^\ell, a) = \sum_j 1_{\{X_j^\ell = \omega_i^\ell, X_j = a\}}.$$

The **estimator of the one-point specification** – supposing that the true context is of length at most ℓ – is then defined by

$$\hat{p}_n(a|\omega_i^\ell) := \frac{N_n(\omega_i^\ell, a)}{N_n(\omega_i^\ell)}.$$

Finally define

$$\log L_n(i,\ell) = \sum_{v \in A^{\partial B_l(i)}} \sum_a N_n((\omega_i^{\ell-1} v, a) \log \left(\frac{\hat{p}_n(a|\omega_i^{\ell-1} v)}{\hat{p}_n(a|\omega_i^{\ell-1}} \right) :$$

the **log-likelihood ratio statistics** for testing the consistency of the sample with a context of length $\ell - 1$ against length ℓ.

Finally define

$$\log L_n(i,\ell) = \sum_{v \in A^{\partial B_l(i)}} \sum_a N_n((\omega_i^{\ell-1} v, a) \log \left(\frac{\hat{p}_n(a|\omega_i^{\ell-1} v)}{\hat{p}_n(a|\omega_i^{\ell-1}} \right) :$$

the **log-likelihood ratio statistics** for testing the consistency of the sample with a context of length $\ell - 1$ against length ℓ.

Definition

$$\hat{\ell}_n(i) := \max\{\ell \leq k(n) : \log L_n(i,\ell) > pen(\ell, n)\}$$

where $k(n) = (\log|\Lambda_n|)^{\frac{1}{2d}}$ *and where the penalty term is chosen by*

$$pen(\ell, n) = C|A|^{|\partial B_\ell(i)|} \log|\Lambda_n|.$$

$|A|^{|\partial B_\ell(i)|}$: degree of freedom when comparing a context of length $\ell - 1$ to all possible contexts of length ℓ.

Theorem

1. The probability of overestimation can be bounded by

$$\mu(\hat{\ell}_n(i) > \ell_i(\omega)) \leq C_1 \exp\left(-C_2 q_{min} (\log|\Lambda_n|)^{1/2}\right) + Rem_n.$$

Here,

$$q_{min} = \inf_a \inf_\omega \gamma_i(a|\omega).$$

2. In the case of bounded trees and of Dobrushin's uniqueness condition : The probability of underestimation can be bounded by

$$\mu(\hat{\ell}_n(i) < \ell_i(\omega)) \leq C_1 \exp\left(-C_2 |\Lambda_n|^{1/2}\right) + Rem_n.$$

1. can be improved for bounded trees. To have consistency we need rapid convergence of $|\Lambda_n| \to \infty$: $|\Lambda_n| = e^{(1+\varepsilon)(\log n)^2}$!!!!

Ideas of the proof

Two main ingredients :

- For the underestimation : *Deviation inequality for the ergodic theorem* at exponential rate (following Dedecker 2001). The field is $\Phi-$mixing !
- For the overestimation : Easier ! *Typicality results obtained by Csiszàr and Talata 2006.*

On the test statistics

Suppose $l \leq \ell_i(\omega)$. Then on a good set :

$$\frac{1}{|\Lambda_n|}\log L_n(i,l) \sim \sum_{v \in A^{\partial B_l(i)}} \sum_a p(\omega_i^{l-1}v, a) \log \left(\frac{p(a|\omega_i^{l-1}v)}{p(a|\omega_i^{l-1})} \right)$$

$$= \sum_{v \in A^{\partial B_l(i)}} p(\omega_i^{l-1}v) H(p(\cdot|\omega_i^{l-1}v), p(\cdot|\omega_i^{l-1})),$$

where H is relative entropy.

Well-known : if there exists at least one $a \in A$ such that $p(a|\omega_i^{l-1}v) \neq p(a|\omega_i^{l-1})$, then relative entropy is strictly positive :

On the test statistics

Suppose $l \leq \ell_i(\omega)$. Then on a good set :

$$\begin{aligned}
\frac{1}{|\Lambda_n|} \log L_n(i,l) \quad &\sim \quad \sum_{v \in A^{\partial B_l(i)}} \sum_a p(\omega_i^{l-1} v, a) \log \left(\frac{p(a|\omega_i^{l-1} v)}{p(a|\omega_i^{l-1})} \right) \\
&= \quad \sum_{v \in A^{\partial B_l(i)}} p(\omega_i^{l-1} v) H(p(\cdot|\omega_i^{l-1} v), p(\cdot|\omega_i^{l-1})),
\end{aligned}$$

where H is relative entropy.

Well-known : if there exists at least one $a \in A$ such that $p(a|\omega_i^{l-1} v) \neq p(a|\omega_i^{l-1})$, then relative entropy is strictly positive :

We gain as long as $l \leq l_i(\omega)$, and the order of gain should be $|\Lambda_n|$!

Systems of interacting neurons

• Recently, in a joint paper with A. Duarte, A. Galves and G. Ost, we have extended these ideas to systems of interacting neurons.

• Huge or infinite system of neurons that interact.

• Spike train : for each neuron i we indicate if there is a spike or not at time t, $t \in \mathbb{Z}$.

$$X_t(i) \in \{0,1\}, X_t(i) = 1 \Leftrightarrow \text{ neuron } i \text{ has a spike at time } t \ .$$

• t is an index of the time window in which we observe the neuron. In the data we considered, the width of this window is typically 3 ms.

Background

• Integrate and fire models : the membrane potential process of one neuron accumulates the stimulus coming from the other neurons. It spikes depending on the height of the accumulated potential.

• Then : reset to a resting potential (here : $= 0$). Restart accumulating potentials coming from other neurons.

Background

• Integrate and fire models : the membrane potential process of one neuron accumulates the stimulus coming from the other neurons. It spikes depending on the height of the accumulated potential.

• Then : reset to a resting potential (here : $= 0$). Restart accumulating potentials coming from other neurons.

• Hence : **Variable length memory :** the memory of the neuron goes back up to its last spike – at least at a first glance.

Background

• Integrate and fire models : the membrane potential process of one neuron accumulates the stimulus coming from the other neurons. It spikes depending on the height of the accumulated potential.

• Then : reset to a resting potential (here : $= 0$). Restart accumulating potentials coming from other neurons.

• Hence : **Variable length memory :** the memory of the neuron goes back up to its last spike – at least at a first glance.

• This is the framework considered e.g. by Cessac (2011) - but only for a **finite** number of neurons.

The model

Chain $X_t \in \{0,1\}^{\mathcal{I}}$,

$$X_t = (X_t(i), i \in \mathcal{I}), X_t(i) \in \{0,1\}, t \in \mathbb{Z},$$

$\mathcal{I}$ countable is the set of neurons. **We will work in the case where $\mathcal{I}$ is infinite**.

Time evolution : At each time step, neurons update independently from each other :

The model

Chain $X_t \in \{0,1\}^{\mathcal{I}}$,

$$X_t = (X_t(i), i \in \mathcal{I}), X_t(i) \in \{0,1\}, t \in \mathbb{Z},$$

$\mathcal{I}$ countable is the set of neurons. **We will work in the case where $\mathcal{I}$ is infinite**.

Time evolution : At each time step, neurons update independently from each other : For any finite subset J of neurons,

$$P(X_t(i) = a_i, i \in J | \mathcal{F}_{t-1}) = \prod_{i \in J} P(X_t(i) = a_i | \mathcal{F}_{t-1}),$$

where

$$\mathcal{F}_{t-1} \text{ is the past history up to time } t-1 \ .$$

The model II

$$P(X_t(i) = 1 | \mathcal{F}_{t-1}) = \phi\left(\sum_j W_{j\to i} \sum_{s=L_t^i+1}^{t-1} g(t-s) X_s(j)\right).$$

Here :

- $W_{j\to i} \in \mathbb{R}$: **synaptic weight** of neuron j on i.
- $L_t^i = \sup\{s < t : X_s(i) = 1\}$ last spike strictly before time t in neuron i.
- $g : \mathbb{N} \to \mathbb{R}_+$ describes a leak effect.

Excitatory versus inhibitory influence

Neurons who have a direct influence on i are those belonging to

$$\mathcal{V}_i := \{j : W_{j\to i} \neq 0\} :$$

Excitatory versus inhibitory influence

Neurons who have a direct influence on i are those belonging to

$$\mathcal{V}_i := \{j : W_{j\to i} \neq 0\} :$$

Either excitatory : $W_{j\to i} > 0$.
Or inhibitory : $W_{j\to i} < 0$.

Excitatory versus inhibitory influence

Neurons who have a direct influence on i are those belonging to

$$\mathcal{V}_i := \{j : W_{j\to i} \neq 0\} :$$

Either excitatory : $W_{j\to i} > 0$.
Or inhibitory : $W_{j\to i} < 0$.

Goal : Estimate the Interaction neighborhood $\mathcal{V}_i$ of a fixed neuron i.

Conditions

1) Spiking rate function ϕ *is strictly increasing and Lipschitz :*

$$|\phi(z) - \phi(z')| \leq \gamma |z - z'|.$$

2) Uniform summability of the synaptic weights

$$r := \sup_i \sum_j |W_{j \to i}| < \infty.$$

3) Put $\varrho = \sum_{s=1}^{\infty} g(s)$. *Then we have* $\gamma r \varrho < 1$.

Theorem

Under the above hypotheses : there exists a unique stationary chain $X_t(i), t \in \mathbb{Z}, i \in \mathcal{I}$, consistent with the dynamics.

AIM :

Put $V_i^{\geq\delta} = \{j \in V_i : |W_{j\to i}| \geq \delta\}$ and try to estimate it ! ! !

Estimation procedure

- Growing sequence of finite windows F_n - centered around site i.
- For a test-configuration $w \in \{0,1\}^{\{-\ell,\ldots,-1\}\times F_n\setminus\{i\}}$:

$N_{(i,n)}(w,1)$ counts the number of occurrences of w followed by a spike of neuron i in the sample $X_1(F_n),\ldots,X_n(F_n)$, when the last spike of neuron i has occurred $\ell+1$ time steps before in the past.

Estimation procedure

- Growing sequence of finite windows F_n - centered around site i.
- For a test-configuration $w \in \{0,1\}^{\{-\ell,\ldots,-1\}\times F_n\setminus\{i\}}$:

$N_{(i,n)}(w,1)$ counts the number of occurrences of w followed by a spike of neuron i in the sample $X_1(F_n),\ldots,X_n(F_n)$, when the last spike of neuron i has occurred $\ell+1$ time steps before in the past.

- Estimated spiking probability $\hat{p}_{(i,n)}(1|w) = \frac{N_{(i,n)}(w,1)}{N_{(i,n)}(w)}$.

Estimation procedure

- Growing sequence of finite windows F_n - centered around site i.
- For a test-configuration $w \in \{0,1\}^{\{-\ell,\dots,-1\}\times F_n\setminus\{i\}}$:

$N_{(i,n)}(w,1)$ counts the number of occurrences of w followed by a spike of neuron i in the sample $X_1(F_n),\dots,X_n(F_n)$, when the last spike of neuron i has occurred $\ell+1$ time steps before in the past.

- Estimated spiking probability $\hat{p}_{(i,n)}(1|w) = \frac{N_{(i,n)}(w,1)}{N_{(i,n)}(w)}$.
- Test statistics to test the influence of neuron j on neuron i :

$$\Delta_{(i,n)}(j) = \max_{w,v:v_{\{j\}^c}=w_{\{j\}^c}} |\hat{p}_{(i,n)}(1|w) - \hat{p}_{(i,n)}(1|v)|.$$

Definition

For any positive threshold parameter $\epsilon > 0$, the estimated interaction neighborhood of neuron $i \in F_n$, at accuracy ϵ, given the sample $X_1(F_n), \dots, X_n(F_n)$, is defined as

$$\hat{V}^{(\epsilon)}_{(i,n)} = \{j \in F_n \setminus \{i\} : \Delta_{(i,n)}(j) > \epsilon\}.$$

Remark

1) Spiking probability of each neuron depends on spatio-temporal portions of the past : $X_t(i)$ *depends on all random variables* $X_s(j)$ *for* $j \in V_i$ *and* $L_t^i + 1 \leq s \leq t - 1$.

Remark

1) Spiking probability of each neuron depends on spatio-temporal portions of the past : $X_t(i)$ depends on all random variables $X_s(j)$ for $j \in V_i$ and $L_t^i + 1 \leq s \leq t - 1$. In particular, since L_t^i is known, temporal dependencies do not need to be estimated.

Remark

1) Spiking probability of each neuron depends on spatio-temporal portions of the past : $X_t(i)$ depends on all random variables $X_s(j)$ for $j \in V_i$ and $L_t^i + 1 \leq s \leq t - 1$. In particular, since L_t^i is known, temporal dependencies do not need to be estimated.

2) One could modify the definition of $\hat{V}_{(i,n)}^{(\epsilon)}$ by considering a sequential pruning procedure.... Our procedure is more robust with respect to the control of the underestimation.

Conditions

1) Transition probabilities are uniformly positive (on the compact where observations take place).

Conditions

1) Transition probabilities are uniformly positive (on the compact where observations take place).

2) If $\mathrm{card}(V_i) = \infty$: $|F_n| \leq D \log n$. *Moreover, there are constants* $C > 0$ *and* $\alpha > 2$ *such that for* $V_i(n) = V_i \cap F_n$,

$$\sum_{j \notin V_i(n)} |W_{j \to i}| \leq C n^{-\alpha},$$

for all $n \in \mathbb{N}$.

Theorem

Let $X_1(F_n), \ldots, X_n(F_n)$ be a sample produced by a stationary stochastic chain $(X_t)_{t\in\mathbb{Z}}$ satisfying our assumptions. Then for $\epsilon_n = O(n^{-\xi/2})$, for some $\xi > 0$,

$$V_i^{\geq\delta} \subseteq \hat{V}_{(i,n)}^{(\epsilon_n)} \subseteq V_i(n) \text{ almost surely as } n \to \infty.$$

In particular, for any subset $F \subset I$ finite, it holds that

$$\hat{V}_{(i,n)}^{(\epsilon_n)} \cap F = V_i \cap F \text{ almost surely as } n \to \infty.$$

Ingredients of the proof

• Cut the spatial dependencies ! ! ! ! Coupling of the stationary process $X = (X_t)_{t\in\mathbb{Z}}$ with a finite range approximation $X^{[R]} = (X_t^{[R]})_{t\in\mathbb{Z}}$, for some fixed $R \geq 1$:

$$\sup_{j\in I, s\leq t} P(X_s(j) \neq X_s^{[R]}(j)) \leq \frac{\gamma\varrho}{1-(\gamma r \varrho)} \sum_{k\notin V_i(R)} |W_{k\to i}|.$$

New Hoeffding-type inequality

$w \in \{0,1\}^{\{-\ell,\dots,-1\}\times I\setminus\{i\}} \Rightarrow$

$$p_i(1|w) = \phi\left(\sum_{j\in V_i} W_{j\to i}\sum_{s=-\ell}^{-1} g(-s)w_j(s)\right).$$

Proposition

For any $w \in \{0,1\}^{\{-\ell,\dots,-1\}\times I\setminus\{i\}}$, *any* $\lambda > 0$ *and all* $n > \ell+1$,

$$P(|M_{(i,n)}(w)| > \lambda) \le 2\exp\left\{-\frac{2\lambda^2}{n-\ell+1}\right\} P(N_{(i,n)}(w) > 0),$$

where $M_{(i,n)}(w) := N_{(i,n)}(w,1) - p_i(1|w)N_{(i,n)}(w)$.

Perspectives

• Extension to non-uniqueness frame seems possible. Allows to consider situations where the stationary regime depends on the initial conditions.

• Add some external stimulus. Therefore, work in the non-stationary and even non-time-homogeneous case ?

Some literature

• ORLANDI, E., L.E. Neighborhood radius estimation in VNRF *SPA* 2011.

• DUARTE, A., GALVES, A., L.E., OST, G., Estimating the interaction graph of stochastic neural dynamics. 2016, arXiv.

Thank you for your attention.