The social network

Soutenance du projet de Simulations et MonteCarlo

Pierre Delanoue - Clément Guillo - Hugues Gallier

ENSAE ParisTech

21 mai 2019

Introduction

- Objectif 1 : Simuler un réseau modélisé par un Exponential Random Graph Model (ERGM)
- Objectif 2 : À partir d'un réseau donné,le caractériser à partir des coefficients estimés du modèle

1. Simulation par MCMC
   - Gibbs sampling
   - Convergence et indépendance
2. Estimation des paramètres
   - Important sampling
   - EMV : descente de gradient
3. Perpectives

Gibbs Sampling

- Réseaux de n noeuds, on modélise ses liens par :
  $x_{ij} = 1$ si i et j sont reliés, 0 sinon
- Représentation matricielle $X \in \mathcal{M}(\{0,1\}^{n \times n})$
- Loi jointe des $X_{i,j}$ est $p(x|\theta) = \frac{1}{Z(\theta)} \exp^{\theta^T S(x)}$
- $S(x) = \left[\sum_{i<j} x_{ij}, \sum_{i<j<k} x_{ij}x_{jk}x_{ki}\right]$ et $Z(.)$ inconnu.

Gibbs Sampling

- On remplace tour à tour x_{ij} par la réalisation d'une Bernoulli de paramètre :

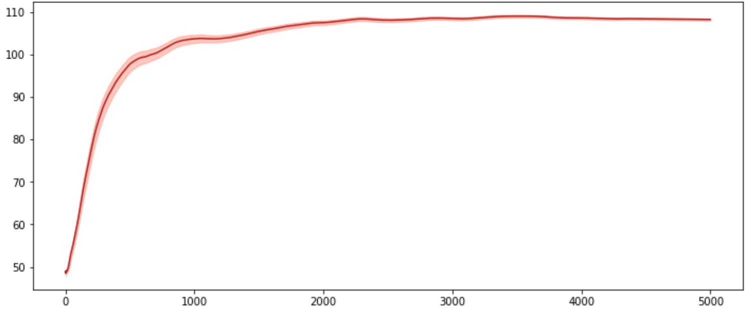
$$\mathbf{P}(X_{ij} = 1 | \theta, x_{-ij}) = \frac{1}{1 + \exp^{\theta^T (S_0(x) - S_1(x))}}$$

- $S_1(x)$, statistique de la matrice x où x_{ij} vaut 1, $S_0(x)$ la même chose pour x_{ij} vallant 0.
- Seulement besoin de calculer les statistiques d'une matrice

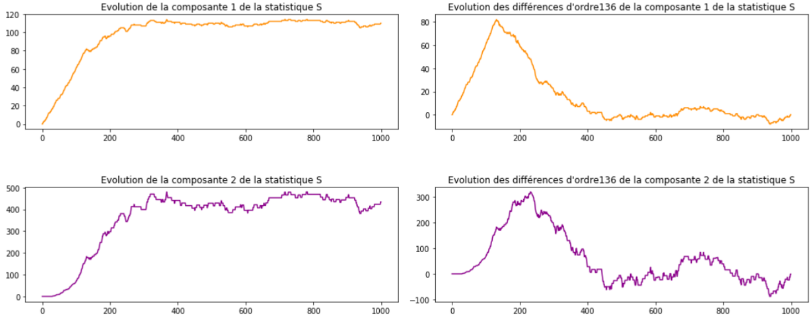
Convergence

- $M_1 \dots M_n$, générations issues du Gibbs sampling
- Chaîne de Markov à espace d'états fini, irréductible : récurrente positive
- Théorème ergodique : $\sum_{i=1}^{n} S(M_i) \longrightarrow \mathbb{E}_{p(x|\theta)}(S(X))$
- Combien d'itérations suffisent pour converger ?

Théorème ergodique



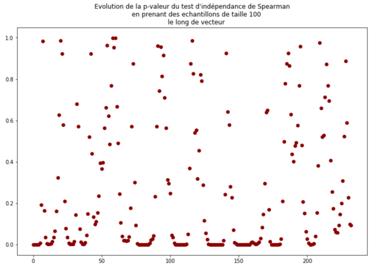
Convergence de la loi

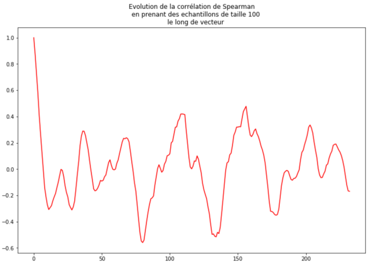


Pseudo Indépendance

- Calcul des autocorrélations $\gamma(h)$ à différent ordres h
- Recherche de l'ordre h tel que $\gamma(h) \approx 0$

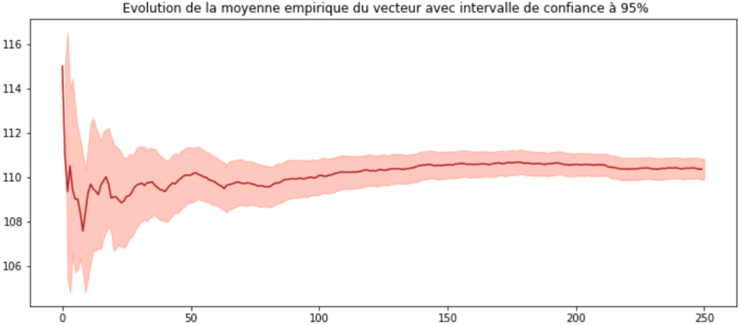
Pseudo Indépendance





- Test et Corrélation de Spearman

Estimateur de Monte Carlo



1. Simulation par MCMC
   - Gibbs sampling
   - Convergence et indépendance
2. Estimation des paramètres
   - Important sampling
   - EMV : descente de gradient
3. Perpectives

Importance sampling

On part de notre équation :

$$1 = \mathbb{E}_{\theta_0}\left(\frac{f_\theta(X)}{f_{\theta_0}(X)}\right) = \mathbb{E}_{\theta_0}\left(\frac{Z(\theta_0)}{Z(\theta)} \exp^{(\theta-\theta_0)S(X)}\right)$$

Donc :

$$Z(\theta) = Z(\theta_0)\mathbb{E}_{\theta_0} \exp^{(\theta-\theta_0)S(X)}$$

Ainsi :

$$L(x,\theta) \propto \frac{1}{\mathbb{E}_{\theta_0}(\exp^{(\theta-\theta_0)S(X)})} \exp^{\theta S(x)}$$

Important sampling

On maximise donc :

$$
\begin{aligned}
l(x,\theta) &\propto -\log(\mathbb{E}_{\theta_0}(\exp^{(\theta-\theta_0)S(X)}) + \theta S(x) \\
&\propto \theta S(x) - \log(\frac{1}{N}\sum_{i=1}^{N} \exp^{(\theta-\theta_0)S(X_i)}) \\
&\propto \theta S(x) - \log\left(\exp^{(\theta-\theta_0)S(X_m)} \frac{1}{N}\sum_{i=1}^{N} \exp^{(\theta-\theta_0)*(S(X_i)-S(X_m)} \right) \\
&\propto \theta S(x) - (\theta-\theta_0)S(X_m) - \log\left(\frac{1}{N}\sum_{i=1}^{N} \exp^{(\theta-\theta_0)*(S(X_i)-S(X_m))} \right)
\end{aligned}
$$

Descente de gradient

- $ln(\frac{Z(\theta)}{Z(\theta_0})$ estimé par importance sampling
- Recherche du bon estimateur par descente de gradient, avec :

$$\frac{\partial l(x,\theta)}{\partial\theta} = S(x) - S(X_m) - \frac{\frac{1}{N}\sum_{i=1}^{N}(S(X_i) - S(X_m)) \exp^{(\theta-\theta_0)*(S(X_i)-S(X_m)}}{\frac{1}{N}\sum_{i=1}^{N} \exp^{(\theta-\theta_0)*(S(X_i)-S(X_m))}}$$

1. Simulation par MCMC
   - Gibbs sampling
   - Convergence et indépendance

2. Estimation des paramètres
   - Important sampling
   - EMV : descente de gradient

3. Perpectives

Perspectives

- Travail généralisable à des réseaux orientés
- Développer les statistiques associées au réseau
- algorithme tie-no-tie