Maximum likelihood inference for large & sparse hidden random graphs

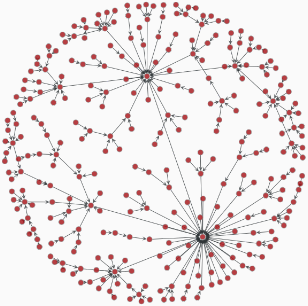
Sylvain Le Corff
based on joint works with É. Vernet, M. Lerasle and R. Diel

Télécom SudParis

Pairwise comparisons based graphs

Zermelo, Math. Z. (1929), Bradley & Terry, Biometrika (1952)...

Weighted edges and an *influence* parameter attached to each node.



Motivations - what to do with pairwise comparison based graphs ?

Maximum likelihood - what is known with many observations ?

Bayesian setting - comparison graphs in random environment

A few algorithms

Pairwise comparisons based graphs

Zermelo, Math. Z. (1929), Bradley & Terry, Biometrika (1952)...

Nodes - Latent data

n individuals characterized by their abilities $(V_i)_{1\leqslant i\leqslant n}$.

Edges - Observations

Abilities indirectly observed through discrete valued scores $(X_{i,j})_{(i,j)\in E_n}$ describing the results of the *comparison* between individuals i and j.

Conditionally on $V_{1:n}$, the random variables $(X_{i,j})_{(i,j)\in E_n}$ are independent with:

$$\mathbb{P}(X_{i,j} = x | V_{1:n}) = k(x, V_i, V_j) \; .$$

Caron & Doucet, JCGS (2012) $\quad k(x, v_i, v_j) = \dfrac{(\theta v_i)^x v_j^{1-x}}{\theta v_i + v_j} \; .$

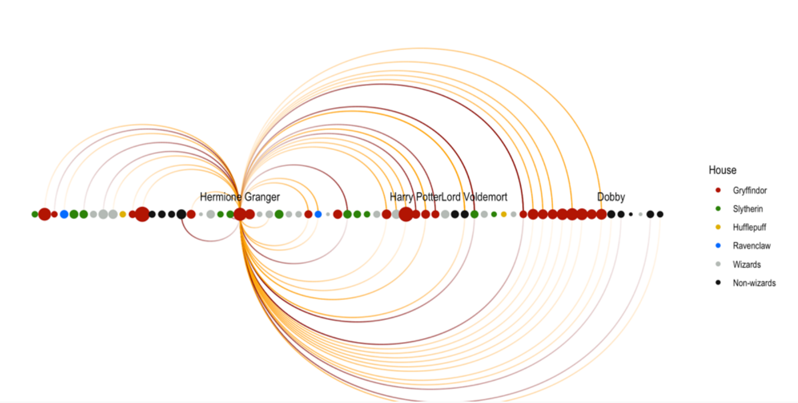
Chatterjee, Diaconis & Sly, AoAP (2011) $\quad k(x, v_i, v_j) = \left(\dfrac{v_i v_j}{1 + v_i v_j}\right)^x \left(\dfrac{1}{1 + v_i v_j}\right)^{1-x} \; .$

Motivations (i) - "social" networks

Bonato et al. (2016)

(Latent) nodes n individuals Harry P., Hermione G., Dobby, etc.

(Observed) edges $X_{i,j}$ is the number of occurrences of two key names i and j both within a specified distance in the text.

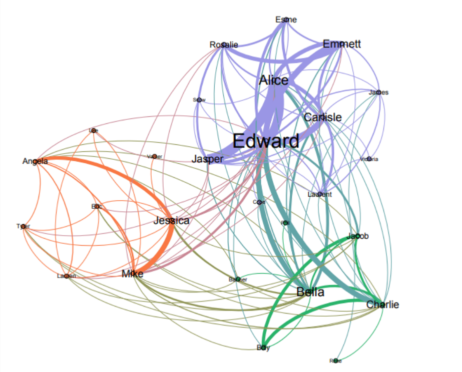


Motivations (i) - "social" networks

Bonato et al. (2016)

(Latent) nodes n individuals Charlie, Bella, etc.

(Observed) edges $X_{i,j}$ is the number of occurrences of two key names i and j both within a specified distance in the text.



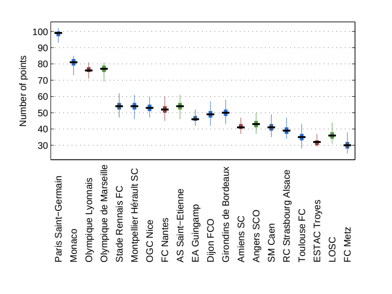
Open problem predict the degree of a given node, the links between clusters.

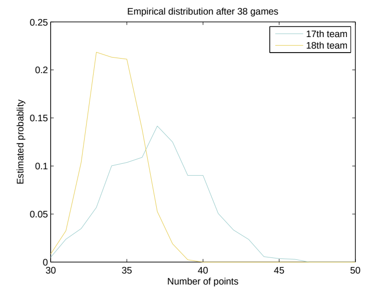
Motivations (ii) - sport analytics

Sire & Redner, Eur. Phys. J. B (2009)

(Latent) nodes n teams FC Nantes, EA Guinguamp, Stade Rennais, etc.

(Observed) edges $X_{i,j}$ is 3 (V_i beats V_j), 1 (tie) or 0 (V_j beats V_i).





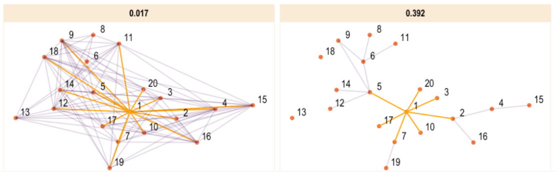
Open problem predict the final ranking of a championship, the minimum number of points to reach a certain goals, detect outliers.

Motivations (iii) - animal behaviour: fighting ability during male contests

Stapley et al., Biol. Lett. (2006), Firth et al. (2010)

(Latent) nodes $n = 77$ Platysaurus broadleyi.

(Observed) edges $X_{i,j}$ is 1 (V_i beats V_j), 0 (V_j beats V_i).



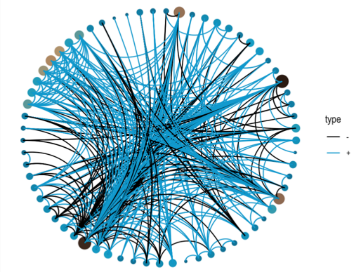
Open problem predict the ecosystem mapping, the roles of covariates, ...(u.v. signals) in female choice and in fighting ability.

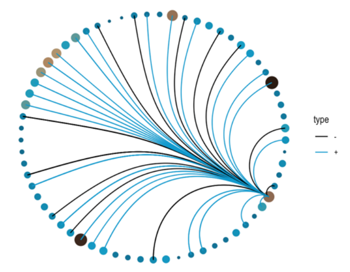
Motivations (iv) - online matchmaking

Herbrich, R. et al., NIPS (2007)

(Latent) nodes n (very large) players .

(Observed) edges $X_{i,j}$ is 1 (V_i beats V_j), 0 (V_j beats V_i).





Open problem online estimation of the abilities and choose optimal matchmaking to decide future matchings simultaneously.

MLE of $V_{1:n}$ for Pairwise comparisons

Zermelo, Math. Z. (1929), Bradley & Terry, Biometrika (1952), David, AMS, (1964)
n fixed, number of observed edges $X_{i,j}$ for each pair (i, j) grows to infinity.
Consistency and asymptotic normality of the MLE.

Simons & Yao, AoS (1999)
At least one weighted edge $X_{i,j}$ for each pair (i, j) when n goes to infinity.
Consistency and asymptotic normality of the MLE.

Chatterjee, Diaconis & Sly, AoAP (2011)
One weighted edge $X_{i,j}$ for each pair (i, j).
Probability larger than $1 - 1/n^2$, there exists a unique MLE.
Supremum norm of the estimation error upper bounded by $\sqrt{\log n/n}$.

Chétrite, Diel & Lerasle, AoAP (2017)
The abilities are realizations of independent and identically distributed random variables with common distribution $\pi_\star$.

$\pi_\star$ is relevant to make predictions... what about MLE for $\pi_\star$?

MLE with "sufficently enough" observations

$n = 100$ players, one strength parameter $\lambda_i \in (0, 1)$ for each.

All players are involved in a game at each time step.

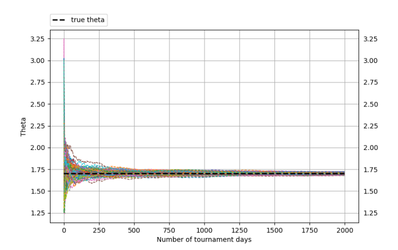
Prior $(\lambda_i)_{1 \leqslant i \leqslant n}$ i.i.d. $\mathcal{G}(a, b)$.

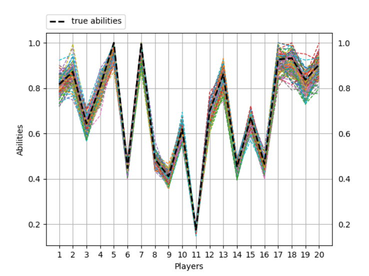
$$k(3, \lambda_i, \lambda_j) = \frac{\lambda_i}{\lambda_i + \theta\lambda_j} \quad \text{and} \quad k(1, \lambda_i, \lambda_j) = \frac{(\theta^2 - 1)\lambda_i}{(\lambda_i + \theta\lambda_j)(\theta\lambda_i + \lambda_j)} \,.$$

Unique solution for the maximum likelihood equations and MLE approximately computed using an Expectation Maximization algorithm.

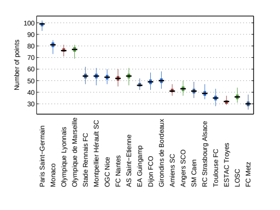
100 independent Monte Carlo runs with 2000 time steps.

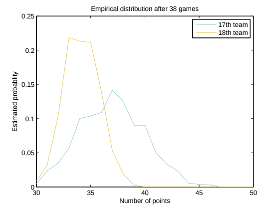
MLE with "sufficently enough" observations

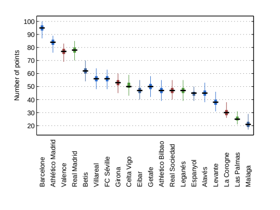


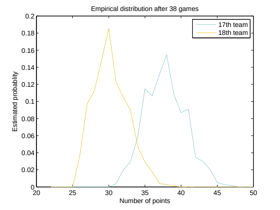


MLE with "sufficently enough" observations









Likelihood and estimator - random environment

The log-likelihood is given, for all $\pi \in \Pi$, by

$$\log \mathbb{P}_\pi((X)_{(i,j)\in E_n}) = \log \left(\int \prod_{(i,j)\in E_n} k(X_{i,j}, v_i, v_j) \pi^{\otimes n}(\mathrm{d}v_{1:n}) \right) .$$

True distribution $\pi_\star$ estimated by the standard maximum likelihood estimator:

$$\hat{\pi}^{E_n} \in \mathrm{argmax}_{\pi\in\Pi} \log \mathbb{P}_\pi((X)_{(i,j)\in E_n}) .$$

Analysis of $\hat{\pi}^{E_n}$...

... relies on the asymptotic behavior of $\log \mathbb{P}_\pi((X)_{(i,j)\in E_n})$ when n grows to ∞.

Risk bounds for $\hat{\pi}^{E_n}$? Practical computation of an approximation of $\hat{\pi}^{E_n}$?

Bayesian setting - posterior consistency

Prior distribution μ on Π.

The posterior distribution given the observations:

$$\mu\left(\pi \in A | X_{1:n}\right) = \frac{\int \prod_{i=1}^{n} k(X_i, v_i, v_{i+1}) \mathbb{1}_A(\pi) \pi^{\otimes n+1}(\mathrm{d}v_{1:n+1}) \mu(\mathrm{d}\pi)}{\int \prod_{i=1}^{n} k(X_i, v_i, v_{i+1}) \pi^{\otimes n+1}(\mathrm{d}v_{1:n+1}) \mu(\mathrm{d}\pi)} \; .$$

Posterior consistency: the posterior distribution concentrates around the true parameter $\pi_\star$ w.r.t. a loss function d. For all $\varepsilon > 0$,

$$\mu\left(\mathrm{d}(\pi, \pi_\star) < \varepsilon | X_{1:n}\right) \longrightarrow_{n \to \infty} 1 \; , \quad \mathbb{P}_{\pi_\star} - a.s.$$

Minimal requirement, in particular in the context of large dimensional models where it is not possible to construct fully subjective priors...

Bayesian setting - assumptions

Barron, A., TR (1988)

Posterior consistency is established w.r.t. a loss function d if, for all $\varepsilon > 0$,

Limiting loglikelihood:

$$\frac{1}{n}\log \mathbb{P}_{\pi}(X_{1:n}) - \frac{1}{n}\log \mathbb{P}_{\pi_\star}(X_{1:n}) \longrightarrow_{n\to\infty} \mathrm{d}(\pi,\pi_\star)\ , \quad \mathbb{P}_{\pi_\star} - a.s.$$

Kullback-Leibler condition: for all $\varepsilon > 0$, $\mu(\{\pi\,;\,\mathsf{d}(\pi,\pi_\star) < \varepsilon\}) > 0$.

Test condition: there exists a sequence of tests (ϕ_n) and a sequence of space (Π_n) such that

$$\mu(\Pi_n^c) \leqslant \mathrm{e}^{-nr}\ , \quad \mathbb{E}_{\pi_\star}[\phi_n] \leqslant \mathrm{e}^{-n\delta}\ , \quad \sup_{\pi\in\Pi_n\ ,\mathsf{d}(\pi,\pi_\star)>\varepsilon} \mathbb{E}_{\pi}[1-\phi_n] \leqslant \mathrm{e}^{-n\delta}\ .$$

Bayesian setting - i.i.d. case

Ghosal et al. AoS (2007a, 2007b), Kruijer et al., EJS (2010), Scricciolo, Bayesian analysis (2014)

$(X_1,\dots,X_n)$ are i.i.d. with unknown density $\pi_\star$ on $\mathbb{R}^d$.

Limiting loglikelihood:

$$\frac{1}{n}\log \mathbb{P}_\pi(X_{1:n}) - \frac{1}{n}\log \mathbb{P}_{\pi_\star}(X_{1:n}) \longrightarrow_{n\to\infty} \mathrm{KL}(\pi_\star,\pi)\;, \quad \pi_\star - a.s.$$

Prior distribution - Dirichlet process

$$G = \sum_{j\geqslant 0} p_j \delta_{\theta_j}\;, \quad \theta_j \sim_{i.i.d.} G_0\;, \quad p_j = \vartheta_j \prod_{\ell<j}(1-\vartheta_\ell)\;, \quad \vartheta_j \sim_{i.i.d.} \mathcal{B}(1,\alpha)\;.$$

Sample independently $\theta \sim G$ and $\sigma \sim \eta$ and set $\pi = \varphi_{\theta,\sigma I_d}$.

Kullback-Leibler condition: for regular $\pi_\star$ (β Holder), $\pi_\star$ can be approximated by finite mixtures of Gaussian distributions so that $\mu(\{\pi\,;\, \mathrm{KL}(\pi,\pi_\star) < \varepsilon_n\}) > 0$, for $\varepsilon_n^2 \propto n^{-2\beta/(2\beta+1)}$.

Test condition: upper bound for the entropy (i.e. the number of balls needed to cover finite location mixtures of Gaussian distributions) with at most $n^{1/(2\beta+1)}$ components.

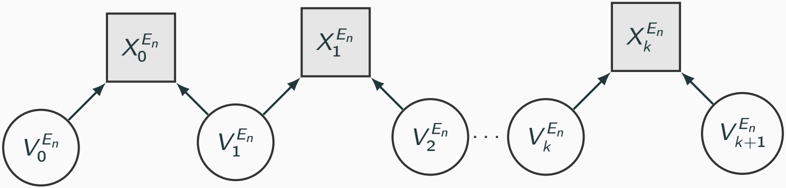
Graphical model

$d_0^{E_n}$ graph distance in $(\{1, \dots, n\}, E_n)$.

$d_0^{E_n}(i, j)$ is the minimal length of a path between nodes i and j.

Random graph decomposed as $\{V_1, \dots, V_n\} = \cup_{q=0}^{n} V_q^{E_n}$, where:

- $V_0^{E_n} = \{V_1\}$;
- for all $q \geqslant 1$, $V_q^{E_n}$ is the set of V_i such that $d_0^{E_n}(1, i) = q$.



$(V_{q+1}^{E_n}, X_q^{E_n})_{q \geqslant 0}$ is a Markov chain...

The journey is paved...

$$\frac{1}{n}\log \mathbb{P}_\pi\left(X_{1:n}\right) = \frac{1}{n}\sum_{q=1}^{n}\log \mathbb{P}_\pi\left(X_q \middle| X_{1:q-1}\right)$$

Forgetting properties

$\rho \in (0,1)$ such that for each term and all $p' < p < q$,

$$\sup_{\pi\in\Pi}\left|\log \mathbb{P}_\pi\left(X_q \middle| X_{p:q-1}\right) - \log \mathbb{P}_\pi\left(X_q \middle| X_{p':q-1}\right)\right| \leqslant \rho^{q-p}\ .$$

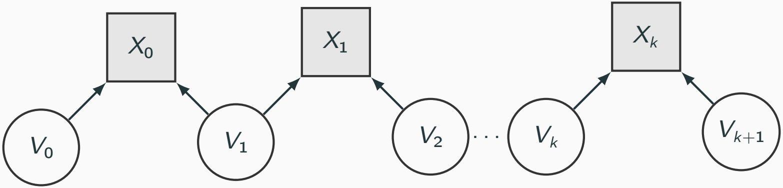
$\log \mathbb{P}_\pi\left(X_q \middle| X_{p':q-1}\right)$ converges a.s.

Ergodicity

Approximate $\log \mathbb{P}_\pi\left(X_{1:n}\right)$ by the sum of these limits.

Normalized loglikelihood converges by the ergodic theorem.

Asymptotic behavior of the loglikelihood



$(V_{q+1}, X_q)_{q \geqslant 0}$ extended to a stationary sequence indexed by $\mathbb{Z}$ with law $\mathbf{P}_{\pi_\star}$.

Assumption

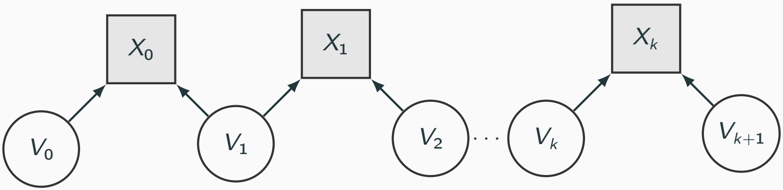
For all $x \in X, \pi \in \Pi \cup \{\pi_\star\}$ and $v_1, v_2 \in \mathrm{supp}(\pi),\ k(x, v_1, v_2) \geq \varepsilon$.

The transition kernel of the Markov chain is uniformly lower bounded

$$P_\pi(X_{i-1}, V_i; X_i, A) = \int \mathbb{1}_A(v_{i+1})\pi(\mathrm{d}v_{i+1})k(X_i, V_i, v_{i+1}) \geqslant \nu\pi(A)\ .$$

The joint Markov chain $(V_{i+1}, X_i)_{i \in \mathbb{Z}}$ is unirformly ergodic.

Asymptotic behavior of the loglikelihood



For all $p' < p < q$ in $\mathbb{Z}$,

$$\sup_{\pi \in \Pi} \left| \log \mathbf{P}_\pi \left(X_q | X_{p:q-1} \right) - \log \mathbf{P}_\pi \left(X_q | X_{p':q-1} \right) \right| \leqslant \varepsilon^{-1} \left(1 - \varepsilon\right)^{q-p} .$$

There exists a function ℓ_π such that for all q in $\mathbb{Z}$,

$$\sup_{\pi \in \Pi} \left| \log \mathbf{P}_\pi \left(X_q | X_{p:q-1} \right) - \ell_\pi(\vartheta^q X) \right| \underset{\mathbf{p} \to \infty}{\longrightarrow} 0, \qquad \mathbf{P}_{\pi_\star}\text{-a.s} .$$

For all $\pi \in \Pi$, $\mathbf{P}_{\pi_\star}$-a.s. and in $\mathrm{L}^1(\mathbf{P}_{\pi_\star})$,

$$\frac{1}{n} \log \mathbf{P}_\pi \left(X_{0:n} \right) \underset{n \to \infty}{\longrightarrow} \mathsf{L}_{\pi_\star}(\pi) = \mathbb{E}_{\pi_\star} \left[\ell_\pi(X) \right] .$$

Risk bounds - Vapnik

$X_1, \ldots, X_n$ i.i.d. observations and ℓ a loss function.

Empirical risk minimizer

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} \ell(\theta, X_i)\ .$$

The risk of any θ is measured by the *excess risk*

$$R(\theta) = \mathbb{E}\left[\ell(\theta, X_1)\right] - \mathbb{E}\left[\ell(\theta_\star, X_1)\right]\ ,$$

where $\theta_\star$ is the minimizer of $\theta \mapsto \mathbb{E}[\ell(\theta, X_1)]$.

The normalized empirical criterion satisfies almost surely,

$$\frac{1}{n} \sum_{i=1}^{n} \ell(\theta, X_i) \to \mathbb{E}[\ell(\theta, X_1)]\ .$$

The excess risk $R(\theta)$ is the difference between the asymptotic normalized empirical loss and the minimizer of this quantity.

Risk bounds - Random graphs

$\pi_\star$ is actually a minimizer of $-\mathsf{L}_{\pi_\star}(\pi)$ over $\Pi \cup \{\pi_\star\}$.

$$\frac{1}{n}\log \mathbf{P}_\pi\left(X_{0:n}\right) \underset{n\to\infty}{\longrightarrow} \mathsf{L}_{\pi_\star}(\pi) = \mathbb{E}_{\pi_\star}\left[\ell_\pi(X)\right] \ .$$

Suggests to use $-\mathsf{L}_{\pi_\star}(\pi)$ as a proxy for the asymptotic normalized empirical loss:

$$R_{\pi_\star}(\pi) = \mathsf{L}_{\pi_\star}(\pi_\star) - \mathsf{L}_{\pi_\star}(\pi) \ .$$

$$\mathbf{P}_{\pi_\star}\left(R_{\pi_\star}(\widehat{\pi}) > \frac{c}{\sqrt{n}}\left(\log^{1/2}\mathsf{N}(\Pi, \|\cdot\|_{\mathsf{tv}}, \varepsilon) + t\right)\right) \leq e^{-t^2} \ .$$

Usual setting, $t_n \propto \sqrt{n}\varepsilon_n$ and entropy of order $\sqrt{n}\varepsilon_n$ with $(\sqrt{n}\varepsilon_n)^{-1} = o(1)$.

Entropy controlled for bounded in variation functions, Sobolev classes...

Risk bounds - Random graphs

Using the forgetting properties,

$$R_{\pi_\star}(\widehat{\pi}) = O(n^{-1}) + G_{\pi\star}(X_{1:n})\,.$$

with

$$G_{\pi\star}(X_{0:n}) = \sup_\pi \left| \frac{1}{n}\log \mathbb{P}_\pi(X_{0:n}) - \mathbb{E}_{\pi_\star}\left[\frac{1}{n}\log \mathbb{P}_\pi(X_{0:n})\right]\right|\,.$$

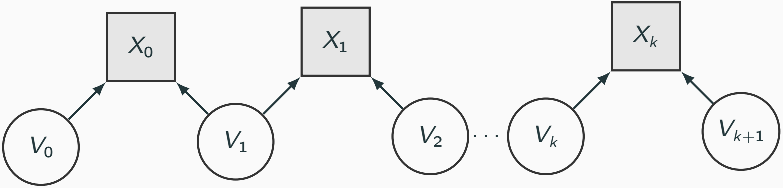
$G_\pi(X_{0:n})$ is a function of the strong mixing Markov chain $(X_q, V_{q+1})_{q\geqslant 0}$.

By Dedecker & Guezel (2015), Douc et al. (201?), concentration inequality for this term using bounded difference inequalities: for all $t > 0$,

$$\mathbb{P}_{\pi_\star}\left(|G_{\pi_\star}(X_{0:n}) - \mathbb{E}_{\pi_\star}[G_{\pi_\star}(X_{0:n})]| > t\right) \leqslant \exp\left(-cnt^2\right)\,.$$

The expectation is then controlled with Dudley's entropy bound.

Nonparametric Expectation Maximization (i)



The Expectation Maximization produces a sequence of parameter estimates $(\widehat{\pi}_p)_{p\geq 0}$ following two steps:

Compute the intermediate quantity:

$$\pi \mapsto Q(\widehat{\pi}_p, \pi) = n^{-1}\, \mathbb{E}_{\widehat{\pi}_p}\left[\sum_{k=0}^{n+1} \log \pi(V_k) \middle| X_{0:n}\right] ,$$

Define $\widehat{\pi}_{p+1}$ as a maximizer of this intermediate quantity:

$$\widehat{\pi}_{p+1} \in \mathrm{argmax}_{\pi\in\Pi} Q(\widehat{\pi}_p, \pi) .$$

Nonparametric Expectation Maximization (ii)

E-step not available explicitly

$$\pi \mapsto Q(\widehat{\pi}_p, \pi) = n^{-1}\, \mathbb{E}_{\widehat{\pi}_p}\left[\sum_{k=0}^{n+1} \log \pi(V_k) \middle| X_{0:n}\right] ,$$

Conditional law of V_k given $X_{0:n}$ approximated by a random empirical measure with uniformly weighted particles ξ_k^ℓ, $1 \leqslant \ell \leqslant M$.

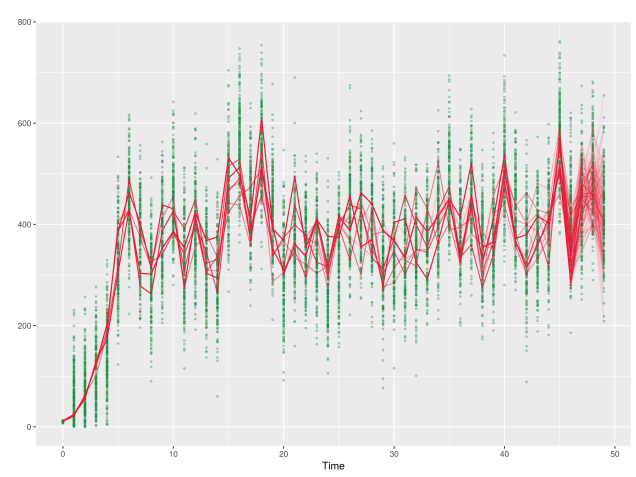
Forward (filt.) pass to approximate the conditional law of V_k given $X_{0:k}$.

$\Rightarrow$ approximate the filtering distribution by $\sum_{\ell=1}^M \omega_k^\ell \delta_{\tilde{\xi}_k^\ell}$.

Backward (smooth.) pass to approximate the conditional law of V_k given $X_{0:n}$.

$\Rightarrow$ approximate the smoothing distribution by $M^{-1} \sum_{\ell=1}^M \delta_{\xi_k^\ell}$.

Nonparametric Expectation Maximization (ii)



Courtesy of the famous team allegedly responsible for RHabits package.

Nonparametric Expectation Maximization (iii)

E-step not available explicitly

$$\pi \mapsto Q(\widehat{\pi}_p, \pi) = n^{-1}\,\mathbb{E}_{\widehat{\pi}_p}\left[\sum_{k=0}^{n+1} \log \pi(V_k) \middle| X_{0:n}\right] ,$$

$$\sim (nM)^{-1} \sum_{k=0}^{n+1} \sum_{\ell=1}^{M} \log \pi(\xi_i^\ell) \;.$$

Kernel regularization and EM update

Let W be a positive kernel and $h > 0$.

The quantity to be maximized is approximated by:

$$\widehat{\mathrm{Q}}^M(\widehat{\pi}_p, \pi) = \int \widetilde{\pi}_p^M(v) \log \pi(v) \mathrm{d}v \;,$$

where

$$\widetilde{\pi}_p^M : v \mapsto (nM)^{-1} \sum_{\ell=1}^{M} \sum_{k=0}^{n+1} W_h\left(v - \xi_k^\ell\right) \;.$$

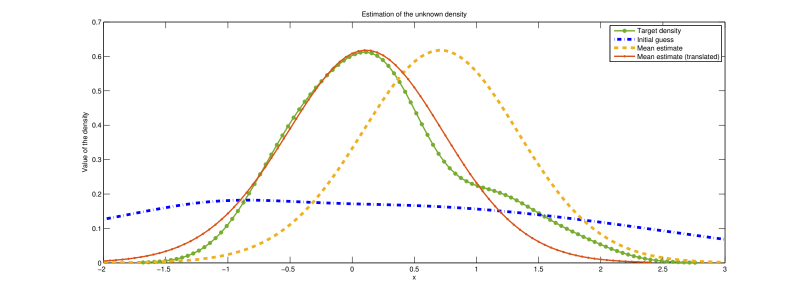


Figure 1: Density estimates from the using Monte Carlo nonparametric EM.

Back to Bayesian setting (i)

Prior distribution μ on Π.

The posterior distribution given the observations:

$$\mu\left(\pi \in A | X_{1:n}\right) = \frac{\int \prod_{i=1}^{n} k(X_i, v_i, v_{i+1}) \mathbb{1}_A(\pi) \pi^{\otimes n+1}(\mathrm{d}v_{1:n+1}) \mu(\mathrm{d}\pi)}{\int \prod_{i=1}^{n} k(X_i, v_i, v_{i+1}) \pi^{\otimes n+1}(\mathrm{d}v_{1:n+1}) \mu(\mathrm{d}\pi)} \, .$$

The aim is to obtain Bayesian posterior concentration rates around $\pi_\star$.

For all $\varepsilon > 0$, define

$$\mathrm{B}_\star(\varepsilon) = \{\pi \in \Pi \ : \ R_{\pi_\star}(\pi) \leqslant \varepsilon\} \, .$$

Find sequences (ε_n) and (α_n) such that:

$$\mathbf{P}_{\pi_\star}\left(\mu\left(\mathrm{B}^c_\star(\varepsilon_n) | X_{1:n}\right) > \alpha_n\right) = o(1) \quad \text{or} \quad \mathbb{E}_{\pi_\star}\left[\mu\left(\mathrm{B}^c_\star(\varepsilon_n) | X_{1:n}\right)\right] = o(1) \, .$$

Back to Bayesian setting (ii)

For all $\varepsilon > 0$,

$$\mathrm{B}_\star(\varepsilon) = \{\pi \in \Pi \ : \ R_{\pi_\star}(\pi) \leqslant \varepsilon\} \ .$$

Requires to build a sequence of spaces (Π_n) such that

$$\mu\left(\mathrm{B}_\star(\varepsilon_n)\right) \geqslant \mathrm{e}^{-c_1 n\varepsilon_n^2} \quad \text{and} \quad \mu\left(\Pi_n^c\right) \leqslant \mathrm{e}^{-c_2 n\varepsilon_n^2} \ .$$

and

$$\sqrt{\log \mathrm{N}(\Pi_n, \|\cdot\|_{\mathrm{tv}}, \varepsilon_n)} \leqslant \sqrt{n}\varepsilon_n \ .$$

Choosing $\varepsilon_n \propto n^{-\beta/(2\beta+1)}(\log n)^{5\beta/(4\beta+2)}$ with $\alpha_n \propto \exp(-cn\varepsilon_n^2)$,

$$\mathbf{P}_{\pi_\star}\left(\mu\left(\mathrm{B}_\star^c(\varepsilon_n)|X_{1:n}\right) > \alpha_n\right) = o(1) \ .$$

Markov Chain Monte Carlo implementation

The unknown distribution π is specified as a mixture model in which some probability density φ_z is mixed with respect to a discrete probability measure P.

The mixture of Dirichlet processes is given by:

$$(\vartheta_j)_{j\geqslant 1} \underset{\text{i.i.d.}}{\sim} \mathrm{Beta}(1,\alpha) \; ,$$

$$\omega_1 = \vartheta_1 \quad \text{and} \quad \text{for } \; j \geqslant 2 \; , \;\; \omega_j = \vartheta_j \prod_{i=1}^{j-1} (1-\vartheta_i) \; ,$$

$$(z_j)_{j\geqslant 1} \underset{\text{i.i.d.}}{\sim} Q \; ,$$

$$(\kappa_i, u_i)_{1\leqslant i \leqslant n+1} \underset{\text{i.i.d.}}{\sim} \sum_{j\geqslant 1} \mathbb{1}_{u_i < \omega_j} \delta_j(\kappa_i) \; ,$$

$$V_i \sim \varphi_{z_{\kappa_i}} \quad \text{for } \; 1 \leqslant i \leqslant n+1 \; ,$$

where Q is a base measure for the parameters $(z_j)_{j\geqslant 1}$ of the density φ_z.

Practical implementation

Yau et al. JRSSB (2011)

Target

Joint posterior distribution of $(V_{1:n+1}, u_{1:n+1}, \kappa_{1:n+1}, z, \vartheta)$.

Solution Block Gibbs sampling.

Posterior distribution of $V_{1:n+1}$ given $(X_{1:n}, u_{1:n+1}, z, \vartheta)$... (Tricky).

Posterior distribution of κ_i given $(X_{1:n}, V_{1:n+1}, u_{1:n+1}, \vartheta, z)$... (Explicit).

Posterior distribution of $(\vartheta, u_{1:n+1})$ given $(\kappa_{1:n+1}, \alpha)$... (Explicit).

Posterior distribution of z given $(X_{1:n}, \kappa_{1:n}, V_{1:n+1})$:

$$z_j \sim Q(z_j) \prod_{i=1,\, \kappa_i=j}^{n+1} \varphi_{z_j}(V_i)\ .$$

Allegedly computationally efficient, with good MCMC mixing properties and robustness to the length of the time series being investigated. Easy to implement and requiring little or no user-interaction..

Practical implementation - Block Gibbs HMM SMC sampler *where is ABC*

Integrating out $\kappa_{1:n+1}$, the conditional distribution of $(X_{1:n}, V_{1:n+1})$ given $(u_{1:n+1}, z, \vartheta)$ is given by

$$p_n(V_{1:n+1}, X_{1:n}) \propto \prod_{i=1}^{n+1} \left(\sum_{j,\, u_i < \omega_j} \varphi_{z_j}(V_i) \right) \prod_{i=1}^{n} k(X_i, V_i, V_{i+1}) \,.$$

The posterior distribution of $V_{1:n+1}$ given the random variables $(X_{1:n}, u_{1:n+1}, z, \vartheta)$ is the joint smoothing distribution of $V_{1:n+1}$ given $X_{1:n}$ when $(V_i)_{1 \leqslant i \leqslant n+1}$ are independent with $V_i \sim \sum_{j,\, u_i < \omega_j} \varphi_{z_j}$ for all $1 \leqslant i \leqslant n+1$.

Cannot be done explicitly and a Sequential Monte Carlo smoother is used instead.

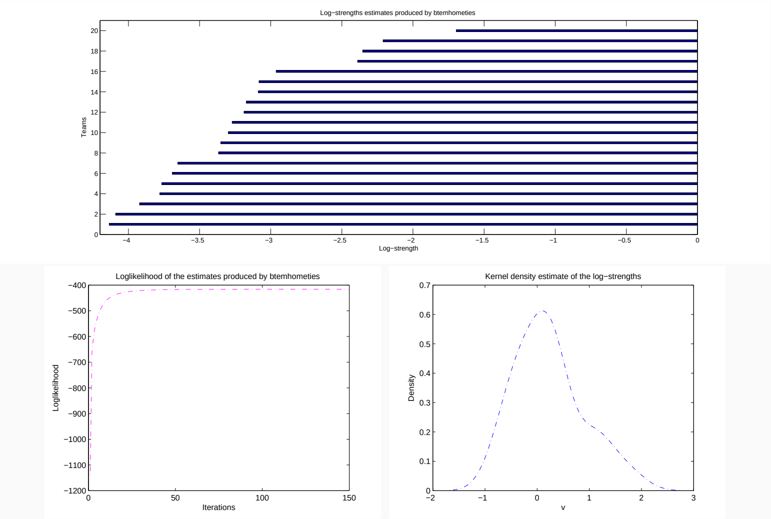


Figure 2: Results of the EM algorithm btemhometies with 2017-2018 Ligue 1 results.

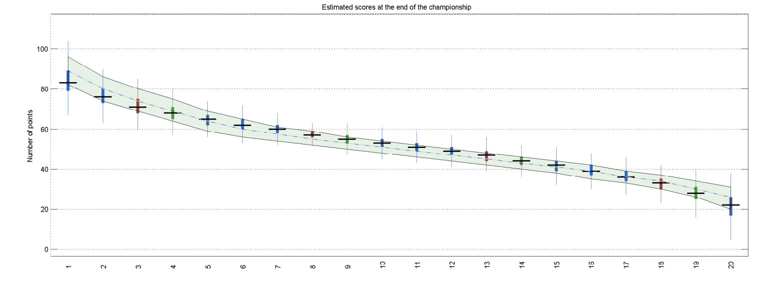


Figure 3: Estimated scores at the end of the championship with *btemhometies* parameter estimates: median (dotted line) and first and last deciles (grey area). Boxplots of the scores obtained with the Block Gibbs Sampler.

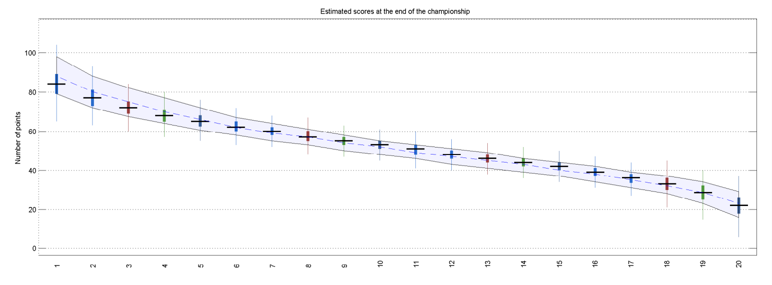


Figure 4: Estimated scores at the end of the championship with parameter sampled with the target distribution: median (dotted line) and first and last deciles (grey area). Boxplots of the scores obtained with the Block Gibbs Sampler.

Perspectives

Some extensions

Numerical results for the nonparametric kernel regularized SMC EM algorithm.

Numerical results for the Bayesian posterior - MCMC.

Regression with covariates for each V_i.

Challenging issues

Generic assumptions to extract n-regular graphs form general random graphs with the same limiting loglikelihood.

Lower bounds for $\mathrm{L}_{\pi_\star}(\pi_\star) - \mathrm{L}_{\pi_\star}(\pi)$ as a function of $\|\pi_\star - \pi\|_{\mathsf{tv}}$.