

SMC sampling from deterministic approximations: Application to the Poisson stochastic block-model

S. Robin

Joint work with S. Donnet

Sorbonne université

Laboratoire de Probabilités, Statistique et Modélisation (LPSM)

CMStatistics, December 2022, London

Outline

Motivating example

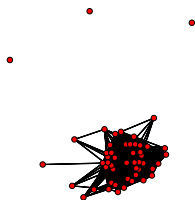
Variational EM inference

SMC sampling

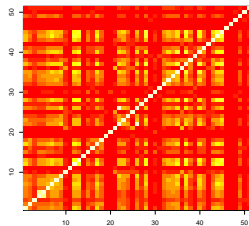
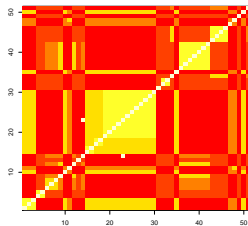
Illustrations

Motivating example

Interaction network



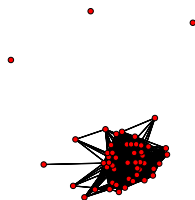
Edge covariates



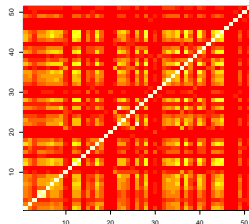
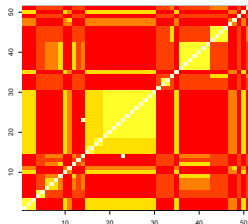
Y_{ij} = number of interactions between nodes i and j (count)

Motivating example

Interaction network



Edge covariates



Y_{ij} = number of interactions between nodes i and j (count)

Questions.

- ▶ Is there some structure in the network?
- ▶ Do the covariates contribute to explain it?
- ▶ Do they explain all of the structure? Is there some 'residual' structure?

Stochastic block-model (SBM)

Proposed model. Poisson SBM, including covariates [MRV10]

Stochastic block-model (SBM)

Proposed model. Poisson SBM, including covariates [MRV10]

Frequentist version.

n nodes ($1 \leq i, j \leq n$)

$\{Z_i\}_i$ iid $\sim \mathcal{M}_K(\mathbf{1}, \pi)$

$\{Y_{ij}\}_{i < j}$ independent $\mid \{Z_i\}$

$Y_{ij} \mid (Z_i = k, Z_j = \ell) \sim \mathcal{P}(\exp(\alpha_{k\ell} + \mathbf{x}_{ij}^\top \beta))$

Latent variables Z , parameter $\theta = (\pi, \alpha, \beta)$.

$Z = \{Z_i\} =$ node memberships
 $\alpha =$ between group interactions

$\pi =$ group proportions
 $\beta =$ effects of the covariates

Stochastic block-model (SBM)

Proposed model. Poisson SBM, including covariates [MRV10]

Frequentist version.

n nodes ($1 \leq i, j \leq n$)

$\{Z_i\}_i$ iid $\sim \mathcal{M}_K(\mathbf{1}, \pi)$

$\{Y_{ij}\}_{i < j}$ independent $\mid \{Z_i\}$

$Y_{ij} \mid (Z_i = k, Z_j = \ell) \sim \mathcal{P}(\exp(\alpha_{k\ell} + \mathbf{x}_{ij}^\top \beta))$

Bayesian version.

$\pi \sim \mathcal{D}_K(\mathbf{a})$

$\gamma = (\alpha, \beta) \sim \mathcal{N}(\gamma_0, \mathbf{V}_0)$

Latent variables Z , parameter $\theta = (\pi, \alpha, \beta)$.

$Z = \{Z_i\} =$ node memberships

$\alpha =$ between group interactions

$\pi =$ group proportions

$\beta =$ effects of the covariates

Inference of SBM

- ▶ Bayesian inference using MCMC: time consuming + convergence issues
- ▶ Frequentist inference via maximum likelihood (ML): intractable
- ▶ Variational approximation of ML (VEM): efficient, but with no statistical guaranty
- ▶ No easy-to-handle variational Bayes approximation (no conjugacy)

Inference of SBM

- ▶ Bayesian inference using MCMC: time consuming + convergence issues
- ▶ Frequentist inference via maximum likelihood (ML): intractable
- ▶ Variational approximation of ML (VEM): efficient, but with no statistical guaranty
- ▶ No easy-to-handle variational Bayes approximation (no conjugacy)

Aim.

- ▶ Design an efficient posterior sampling algorithm taking advantage of the efficiency of (frequentist) VEM inference

Outline

Motivating example

Variational EM inference

SMC sampling

Illustrations

EM and VEM

SBM = incomplete data model

Maximum likelihood. Most popular way: EM

$$\log p_{\theta}(Y) = \mathbb{E}((\log p_{\theta}(Y, Z) | Y) - \mathbb{E}(\log p_{\theta}(Z | Y) | Y))$$

→ Requires to determine (some moments of) $p_{\theta}(Z | Y)$, which is intractable.

EM and VEM

SBM = incomplete data model

Maximum likelihood. Most popular way: EM

$$\log p_{\theta}(Y) = \mathbb{E}((\log p_{\theta}(Y, Z) | Y) - \mathbb{E}(\log p_{\theta}(Z | Y) | Y))$$

→ Requires to determine (some moments of) $p_{\theta}(Z | Y)$, which is intractable.

Variational approximation. When $p_{\theta}(Z | Y)$ is intractable, rather maximize the ELBO

$$\begin{aligned} J(\theta, q) &= \log p_{\theta}(Y) - KL(q(Z) \| p_{\theta}(Z | Y)) \\ &= \mathbb{E}_q \log p_{\theta}(Y, Z) - \mathbb{E}_q \log q(Z) \leq \log p_{\theta}(Y) \end{aligned}$$

taking $q \in \mathcal{Q}$.

Mean field. Typical choice for SBM: $\mathcal{Q} = \{q : q(Z) = \prod_i q_i(Z_i)\}$ (Blockmodels [Lég16]).

Approximate posterior

Taylor expansion. Denote $(\tilde{\theta}, \tilde{q}) = \arg \max_{\theta, q \in \mathcal{Q}} J(\theta, q)$ and approximate

$$\begin{aligned} \log p(\theta | Y) &\propto \exp(\log \pi(\theta) + \log p_{\theta}(Y)) \simeq \exp(\log \pi(\theta) + J(\theta, \tilde{q})) \\ &\simeq \exp\left(\log \pi(\theta) + J(\tilde{\theta}, \tilde{q}) + \frac{1}{2}(\theta - \tilde{\theta})^{\top} \partial_{\theta^2} J(\theta, \tilde{q})(\theta - \tilde{\theta})\right) \end{aligned}$$

Approximate posterior

Taylor expansion. Denote $(\tilde{\theta}, \tilde{q}) = \arg \max_{\theta, q \in \mathcal{Q}} J(\theta, q)$ and approximate

$$\begin{aligned} \log p(\theta | Y) &\propto \exp(\log \pi(\theta) + \log p_{\theta}(Y)) \simeq \exp(\log \pi(\theta) + J(\theta, \tilde{q})) \\ &\simeq \exp\left(\log \pi(\theta) + J(\tilde{\theta}, \tilde{q}) + \frac{1}{2}(\theta - \tilde{\theta})^{\top} \partial_{\theta^2} J(\theta, \tilde{q})(\theta - \tilde{\theta})\right) \end{aligned}$$

Variance proxy for VEM estimates. Set $\tilde{V}_{\gamma} := -(\partial_{\gamma^2} J(\theta, \tilde{q}))^{-1}$ and use conjugacy rules to get

$$\tilde{V}(\gamma) = (V_0^{-1} + \tilde{V}_{\gamma}^{-1})^{-1}, \quad \tilde{\mathbb{E}}(\gamma) = \tilde{V}(\gamma)^{-1} (V_0^{-1} \gamma_0 + \tilde{V}_{\gamma}^{-1} \tilde{\gamma})^{-1}$$

and define

$$\tilde{p}(\gamma) := \mathcal{N}(\tilde{\mathbb{E}}(\gamma), \tilde{V}(\gamma)) \quad \simeq p(\gamma | Y).$$

Approximate posterior

Taylor expansion. Denote $(\tilde{\theta}, \tilde{q}) = \arg \max_{\theta, q \in \mathcal{Q}} J(\theta, q)$ and approximate

$$\begin{aligned} \log p(\theta | Y) &\propto \exp(\log \pi(\theta) + \log p_{\theta}(Y)) \simeq \exp(\log \pi(\theta) + J(\theta, \tilde{q})) \\ &\simeq \exp\left(\log \pi(\theta) + J(\tilde{\theta}, \tilde{q}) + \frac{1}{2}(\theta - \tilde{\theta})^{\top} \partial_{\theta^2} J(\theta, \tilde{q})(\theta - \tilde{\theta})\right) \end{aligned}$$

Variance proxy for VEM estimates. Set $\tilde{V}_{\gamma} := -(\partial_{\gamma^2} J(\theta, \tilde{q}))^{-1}$ and use conjugacy rules to get

$$\tilde{V}(\gamma) = (V_0^{-1} + \tilde{V}_{\gamma}^{-1})^{-1}, \quad \tilde{\mathbb{E}}(\gamma) = \tilde{V}(\gamma)^{-1} (V_0^{-1} \gamma_0 + \tilde{V}_{\gamma}^{-1} \tilde{\gamma})^{-1}$$

and define

$$\tilde{p}(\gamma) := \mathcal{N}(\tilde{\mathbb{E}}(\gamma), \tilde{V}(\gamma)) \quad \simeq p(\gamma | Y).$$

Approximate posterior. Proceed similarly to define \tilde{a} and set

$$\tilde{p}(\pi) := \mathcal{D}(\tilde{a}) \quad \simeq p(\pi | Y),$$

then combine the two

$$\tilde{p}(\theta) := \tilde{p}(\pi) \tilde{p}(\gamma) \quad \simeq p(\theta | Y).$$

Outline

Motivating example

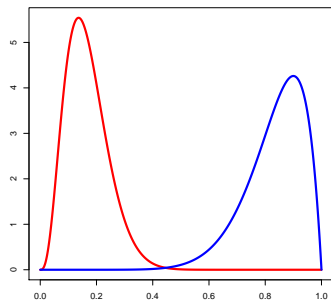
Variational EM inference

SMC sampling

Illustrations

Sampling principle

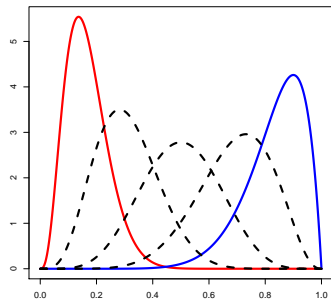
- ▶ p_0 = proposal, p^* = target



Sampling principle

- ▶ p_0 = proposal, p^* = target
- ▶ Intermediate distributions

$$p_0, p_1, \dots, p_H = p^*$$



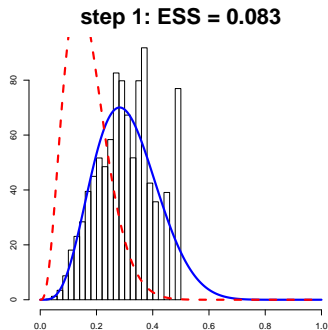
Sampling principle

▶ p_0 = proposal, p^* = target

▶ Intermediate distributions

$$p_0, p_1, \dots, p_H = p^*$$

▶ Iteratively:
use p_h to get a sample from p_{h+1}



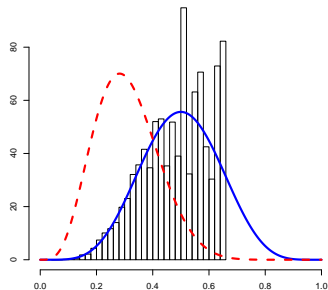
Sampling principle

- ▶ p_0 = proposal, p^* = target
- ▶ Intermediate distributions

$$p_0, p_1, \dots, p_H = p^*$$

- ▶ Iteratively:
use p_h to get a sample from p_{h+1}

step 2: ESS = 0.14



Sampling principle

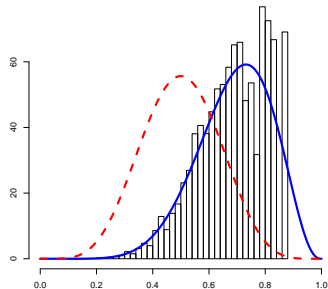
▶ p_0 = proposal, p^* = target

▶ Intermediate distributions

$$p_0, p_1, \dots, p_H = p^*$$

▶ Iteratively:
use p_h to get a sample from p_{h+1}

step 3: ESS = 0.16



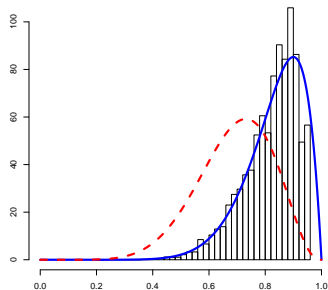
Sampling principle

- ▶ p_0 = proposal, p^* = target
- ▶ Intermediate distributions

$$p_0, p_1, \dots, p_H = p^*$$

- ▶ Iteratively:
use p_h to get a sample from p_{h+1}

step 4: ESS = 0.31



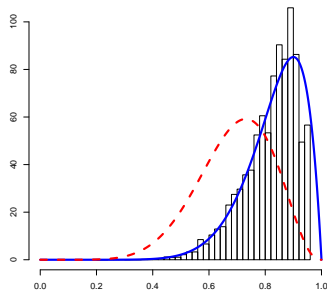
Sampling principle

- ▶ p_0 = proposal, p^* = target
- ▶ Intermediate distributions

$$p_0, p_1, \dots, p_H = p^*$$

- ▶ Iteratively:
use p_h to get a sample from p_{h+1}

step 4: ESS = 0.31



Here. Take $p_0 = \tilde{p}$ (rather than $p_0 = \pi = \text{prior}$), $p^* = p(\cdot | Y)$

Sequential importance sampling scheme

Denote

$$U = (\theta, Z), \quad \pi = \text{prior}, \quad \ell = \text{likelihood}$$

Distribution path: set $0 = \rho_0 < \rho_1 < \dots < \rho_{H-1} < \rho_H = 1$,

$$p_h(U) \propto \tilde{p}(U)^{1-\rho_h} \times p(U|Y)^{\rho_h}$$

$$\propto \tilde{p}(U) \times r(U)^{\rho_h},$$

$$r(U) = \frac{\pi(U)\ell(Y|U)}{\tilde{p}(U)}$$

Sequential importance sampling scheme

Denote

$$U = (\theta, Z), \quad \pi = \text{prior}, \quad \ell = \text{likelihood}$$

Distribution path: set $0 = \rho_0 < \rho_1 < \dots < \rho_{H-1} < \rho_H = 1$,

$$p_h(U) \propto \tilde{p}(U)^{1-\rho_h} \times p(U|Y)^{\rho_h}$$

$$\propto \tilde{p}(U) \times r(U)^{\rho_h},$$

$$r(U) = \frac{\pi(U)\ell(Y|U)}{\tilde{p}(U)}$$

Sequential sampling. At each step h , provides

$$\mathcal{E}_h = \{(U_h^m, w_h^m)\}_m = \text{weighted sample of } p_h$$

Sequential importance sampling scheme

Denote

$$U = (\theta, Z), \quad \pi = \text{prior}, \quad \ell = \text{likelihood}$$

Distribution path: set $0 = \rho_0 < \rho_1 < \dots < \rho_{H-1} < \rho_H = 1$,

$$p_h(U) \propto \tilde{p}(U)^{1-\rho_h} \times p(U|Y)^{\rho_h}$$

$$\propto \tilde{p}(U) \times r(U)^{\rho_h},$$

$$r(U) = \frac{\pi(U)\ell(Y|U)}{\tilde{p}(U)}$$

Sequential sampling. At each step h , provides

$$\mathcal{E}_h = \{(U_h^m, w_h^m)\}_m = \text{weighted sample of } p_h$$

Question. How to tune $\{\rho_h\}$ or H to keep each sampling step efficient?

Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \tilde{p}$, $w_0^m = 1$

¹To avoid degeneracy. Weights set to 1 after it.

² K_h has stationary distribution p_h (e.g. Gibbs sampler). Only propagation: no convergence needed

Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \tilde{p}$, $w_0^m = 1$

Step h : Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

¹To avoid degeneracy. Weights set to 1 after it.

² K_h has stationary distribution p_h (e.g. Gibbs sampler). Only propagation: no convergence needed

Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \tilde{p}$, $w_0^m = 1$

Step h : Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. compute $w_h^m = w_{h-1}^m \times (r_{h-1}^m)^{\rho_h - \rho_{h-1}}$
tuning ρ_h so that $cESS(\mathcal{E}_{h-1}; \rho_{h-1}, \rho_h) = \tau_1$

¹To avoid degeneracy. Weights set to 1 after it.

² K_h has stationary distribution ρ_h (e.g. Gibbs sampler). Only propagation: no convergence needed

Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \tilde{p}$, $w_0^m = 1$

Step h : Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. compute $w_h^m = w_{h-1}^m \times (r_{h-1}^m)^{\rho_h - \rho_{h-1}}$
tuning ρ_h so that $cESS(\mathcal{E}_{h-1}; \rho_{h-1}, \rho_h) = \tau_1$
2. ⁽¹⁾ if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles

¹To avoid degeneracy. Weights set to 1 after it.

² K_h has stationary distribution ρ_h (e.g. Gibbs sampler). Only propagation: no convergence needed

Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \tilde{p}$, $w_0^m = 1$

Step h : Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. compute $w_h^m = w_{h-1}^m \times (r_{h-1}^m)^{\rho_h - \rho_{h-1}}$
tuning ρ_h so that $cESS(\mathcal{E}_{h-1}; \rho_{h-1}, \rho_h) = \tau_1$
2. ⁽¹⁾ if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles
3. ⁽²⁾ propagate the particles $U_h^m \sim K_h(U_h^m | U_{h-1}^m)$

¹To avoid degeneracy. Weights set to 1 after it.

² K_h has stationary distribution ρ_h (e.g. Gibbs sampler). Only propagation: no convergence needed

Proposed algorithm

Init.: Sample $(U_0^m)_m$ iid $\sim \tilde{p}$, $w_0^m = 1$

Step h : Using the previous sample $\mathcal{E}_{h-1} = \{(U_{h-1}^m, w_{h-1}^m)\}$

1. compute $w_h^m = w_{h-1}^m \times (r_{h-1}^m)^{\rho_h - \rho_{h-1}}$
tuning ρ_h so that $cESS(\mathcal{E}_{h-1}; \rho_{h-1}, \rho_h) = \tau_1$
2. ⁽¹⁾ if $ESS_h = \overline{w_h^2} / \overline{w_h^2} < \tau_2$, resample the particles
3. ⁽²⁾ propagate the particles $U_h^m \sim K_h(U_h^m | U_{h-1}^m)$

Stop: When ρ_h reaches 1.

¹To avoid degeneracy. Weights set to 1 after it.

² K_h has stationary distribution ρ_h (e.g. Gibbs sampler). Only propagation: no convergence needed

Some comments

Justification of the algorithm [DDJ06]. At each step h , construct a distribution for the whole particle path with marginal ρ_h .

Some comments

Justification of the algorithm [DDJ06]. At each step h , construct a distribution for the whole particle path with marginal p_h .

Conditional ESS. Efficiency of sample \mathcal{E} from p_{h-1} for distribution p_h

$$cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \frac{M[\sum_m W_{h-1}^m (r_{h-1}^m)^{\rho_h - \rho_{h-1}}]^2}{\sum_m W_{h-1}^m (r_{h-1}^m)^{2\rho_h - 2\rho_{h-1}}}$$

- ▶ Can be computed for any ρ_h **before sampling**.
- ▶ ρ_h tuned to meet τ_1 , which controls the step size $\rho_h - \rho_{h-1}$ (and H)

Some comments

Justification of the algorithm [DDJ06]. At each step h , construct a distribution for the whole particle path with marginal ρ_h .

Conditional ESS. Efficiency of sample \mathcal{E} from p_{h-1} for distribution p_h

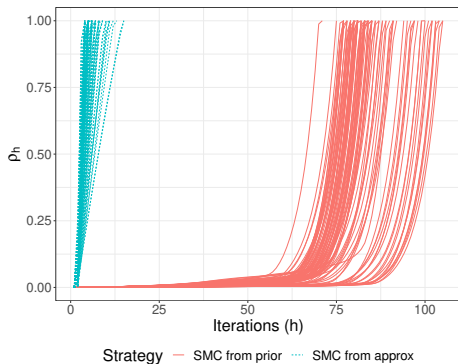
$$cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \frac{M[\sum_m W_{h-1}^m (r_{h-1}^m)^{\rho_h - \rho_{h-1}}]^2}{\sum_m W_{h-1}^m (r_{h-1}^m)^{2\rho_h - 2\rho_{h-1}}}$$

- ▶ Can be computed for any ρ_h **before sampling**.
- ▶ ρ_h tuned to meet τ_1 , which controls the step size $\rho_h - \rho_{h-1}$ (and H)

Marginal likelihood. An estimate of the marginal likelihood $p(Y)$ is also available as a side product.

Variational approximation vs prior

Starting from $\rho_0 = \tilde{\rho}$ reduces the number of SMC steps wrt starting from $\rho_0 = \pi$.



(synthetic data)

Outline

Motivating example

Variational EM inference

SMC sampling

Illustrations

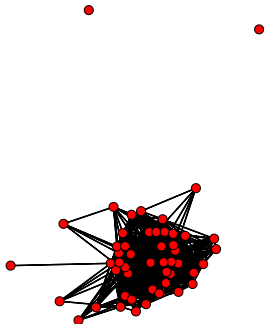
Tree network

From [VPDL08].

$n = 51$ tree species

3 covariates (distances):
taxonomy, geography, genetics

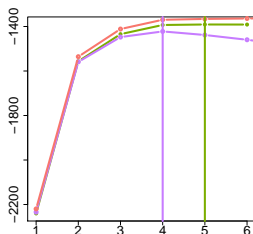
Y_{ij} = number of shared fungal parasites



Sampling path & choice of K

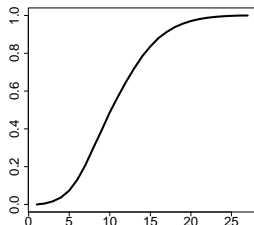
Full model. All covariates

Model selection



J_K , $\log(Y|K)$, $\widehat{ICL}(K)$

Sampling path: ρ_h

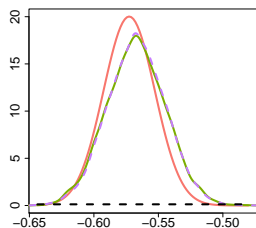


Choosing the number of groups: $\widehat{K} = \arg \max_K \widehat{p}(K|Y)$

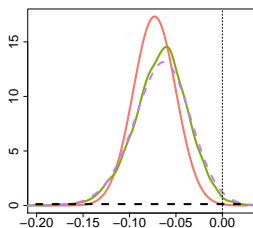
- Different from $\arg \max_K \widetilde{ICL}(K)$ here.

Posterior distribution of β

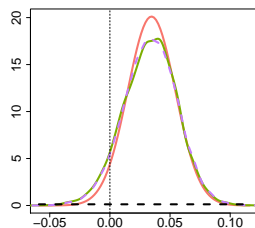
taxonomy



geography



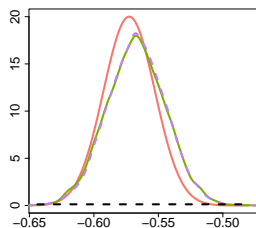
genetics



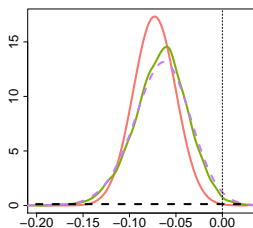
$$\tilde{p}(\beta | \hat{K}), \quad \hat{p}(\beta | Y, \hat{K}), \quad \hat{p}(\beta | Y) = \sum_K \hat{p}(K | Y) \hat{p}(\beta | Y, K)$$

Posterior distribution of β

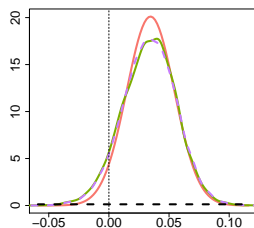
taxonomy



geography



genetics



$$\tilde{p}(\beta | \hat{K}), \quad \hat{p}(\beta | Y, \hat{K}), \quad \hat{p}(\beta | Y) = \sum_K \hat{p}(K | Y) \hat{p}(\beta | Y, K)$$

Correlation between estimates.

	(β_1, β_2)	(β_1, β_3)	(β_2, β_3)
$\tilde{p}(\beta)$	-0.012	0.021	0.318
$\hat{p}(\beta Y)$	-0.274	-0.079	-0.088

Model selection. $\hat{P}\{x = (\text{taxo.}, \text{geo.}) | Y\} \simeq 70\%$, $\hat{P}\{x = (\text{taxo.}) | Y\} \simeq 30\%$

Residual structure

Between group interactions $(\alpha_{k\ell}) =$ 'residuals' = not explained by the covariates.

³with increasing marginal $\bar{\phi}(u) = \int \phi(u, v) dv$ to ensure identifiability.

Residual structure

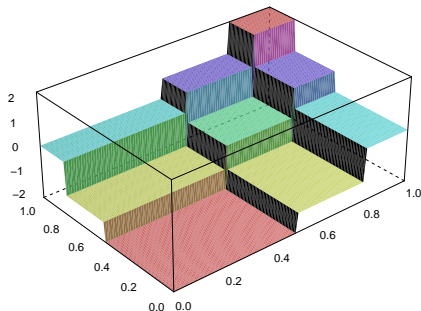
Between group interactions ($\alpha_{k\ell}$) = 'residuals' = not explained by the covariates.

'Graphon' representation. [LRO17]

Group interactions encoded as

$$\phi : [0, 1]^2 \mapsto \mathbb{R}$$

- ▶ symmetric³,
- ▶ block-wise constant,
- ▶ block width = π_k
- ▶ block height = $\alpha_{k\ell}$



³with increasing marginal $\bar{\phi}(u) = \int \phi(u, v) dv$ to ensure identifiability.

Residual structure

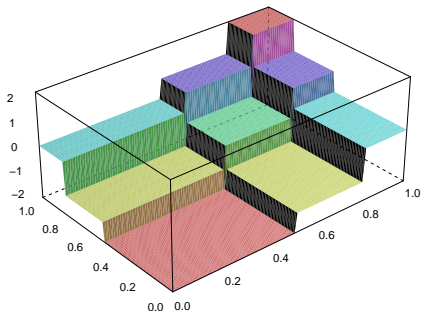
Between group interactions ($\alpha_{k\ell}$) = 'residuals' = not explained by the covariates.

'Graphon' representation. [LRO17]

Group interactions encoded as

$$\phi : [0, 1]^2 \mapsto \mathbb{R}$$

- ▶ symmetric³,
- ▶ block-wise constant,
- ▶ block width = π_k
- ▶ block height = $\alpha_{k\ell}$



Same representation for all K . $Y_{ij}|(U_i, U_j) \sim \mathcal{P}\left(\exp(\phi(U_i, U_j) + \mathbf{x}_{ij}^T \beta)\right)$

³with increasing marginal $\bar{\phi}(u) = \int \phi(u, v) dv$ to ensure identifiability.

Tree network residual structure

Residual graphon.

Each particle θ^m provides an estimate of $\phi^m(u, v)$

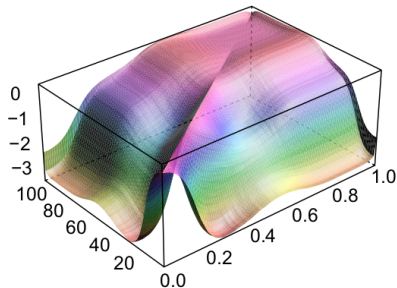
All estimates can be averaged (over both m and K)

Tree network residual structure

Residual graphon.

Each particle θ^m provides an estimate of $\phi^m(u, v)$

All estimates can be averaged (over both m and K)

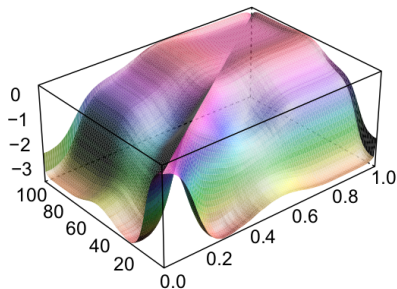


Tree network residual structure

Residual graphon.

Each particle θ^m provides an estimate of $\phi^m(u, v)$

All estimates can be averaged (over both m and K)



Interpretation.

- ▶ A remaining individual effect (some species interact more than other in average)
- ▶ A small fraction of species interact much less than expected.

Social network of equid species

2 datasets [RSF⁺15].

- ▶ $n = 28$ zebras, $n = 29$ onagers
- ▶ sex and age (juvenile / adult) recorded

Social network of equid species

2 datasets [RSF⁺15].

- ▶ $n = 28$ zebras, $n = 29$ onagers
- ▶ sex and age (juvenile / adult) recorded

Model comparison.

Zebras:

$$\widehat{P}(x = (\text{sex}) \mid Y) \simeq 1$$

Onagers:

$$\widehat{P}(x = (\text{sex} \times \text{age}) \mid Y) \simeq 1$$

Social network of equid species

2 datasets [RSF⁺15].

- ▶ $n = 28$ zebras, $n = 29$ onagers
- ▶ sex and age (juvenile / adult) recorded

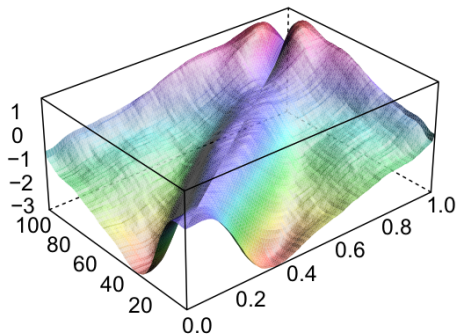
Model comparison.

Zebras:

$$\widehat{P}(x = (\text{sex}) \mid Y) \simeq 1$$

Onagers:

$$\widehat{P}(x = (\text{sex} \times \text{age}) \mid Y) \simeq 1$$



Onager network: residual structure

Discussion

Rational.

- ▶ Frequentist VEM side-product can be used to define an approximate posterior
- ▶ SMC sampling can start from there to the sample from the posterior

Discussion








Rational.

- ▶ Frequentist VEM side-product can be used to define an approximate posterior
- ▶ SMC sampling can start from there to the sample from the posterior

Open problems. (About dig data...)

- ▶ Louis approximate prior \tilde{p} is not that bad. Still, numerous steps are needed to reach the posterior
... because of the large dimension of $U = (\theta, Z)$
- ▶ Especially true for (uselessly) large K
... but VEM inference can not be trusted to choose it

References

-  F. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436, 2006.
-  S. Bonnnet and S. Robin. Accelerating Bayesian estimation for network poisson models using frequentist variational estimates. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2021.
-  J.-B. Léger. Blockmodels: A R-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates. Technical report, arXiv:1602.07587, 2016.
-  P. Latouche, S. Robin, and S. Ouadah. Goodness of fit of logistic regression models for random graphs. *Journal of Computational and Graphical Statistics*, (just-accepted), 2017.
-  M. Mariadassou, S. Robin, and C. Vacher. Uncovering latent structure in valued graphs: a variational approach. *The Annals of Applied Statistics*, pages 715–742, 2010.
-  D. Rubenstein, S. R Sundaresan, I. R Fischhoff, C. Tantipathananandh, and T. Y Berger-Wolf. Similar but different: dynamic social network analysis highlights fundamental differences between the fission-fusion societies of two equid species, the onager and Grevy's zebra. *PLoS one*, 10(10):e0138645, 2015.
-  C. Vacher, D. Piou, and M.-L. Desprez-Loustau. Architecture of an antagonistic tree/fungus network: The asymmetric influence of past evolutionary history. *PLoS ONE*, 3(3):1740, 2008.

Theoretical justification

At each step h , [DDJ06] construct a distribution for the whole particle path with marginal p_h .

- ▶ $\bar{p}_h(\theta_{0:h})$ distribution of the particle path

$$\bar{p}_h(\theta_{0:h}) \propto p_h(\theta_h) \prod_{k=1}^h L_k(\theta_{k-1}|\theta_k)$$

- ▶ $L_h =$ backward kernel

$$L_h(\theta_{h-1}|\theta_h) = K_h(\theta_h|\theta_{h-1})p_h(\theta_{h-1})/p_h(\theta_h)$$

- ▶ Update for the weights

$$w_h(\theta_{0:h}) = w_{h-1}(\theta_{0:h-1})\alpha(\theta_h)^{\rho_h - \rho_{h-1}}$$

Some comments

Resampling (optional step 3).

- ▶ avoids degeneracy
- ▶ set weights $w_h^m = 1$ after resampling

Propagation kernel K_h (step 4).

- ▶ with stationary distribution p_h (e.g. Gibbs sampler)
- ▶ just propagation: does not change the distribution \rightarrow no convergence needed

Some comments

Resampling (optional step 3).

- ▶ avoids degeneracy
- ▶ set weights $w_h^m = 1$ after resampling

Propagation kernel K_h (step 4).

- ▶ with stationary distribution p_h (e.g. Gibbs sampler)
- ▶ just propagation: does not change the distribution \rightarrow no convergence needed

Theoretical justification: [DDJ06]. At each step h , construct a distribution for the whole particle path with marginal p_h .

Marginal likelihood

Denote

$$\gamma_h(U) = \tilde{p}(U)\alpha(U)^{\rho_h}, \quad Z_h = \int \gamma_h(U) \, dU, \quad \rho_h = \gamma_h(U)/Z_h$$

The marginal likelihood is given by

$$p(Y) = \int \pi(U)\ell(Y|U) \, dU = \int \gamma_H(U) \, dU = Z_H$$

which can be estimated with

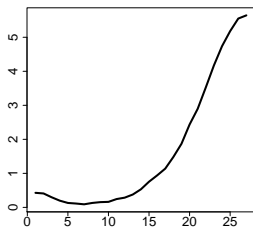
$$\widehat{\left(\frac{Z_H}{Z_0}\right)} = \prod_{h=1}^H \widehat{\left(\frac{Z_h}{Z_{h-1}}\right)} \quad \text{where} \quad \widehat{\left(\frac{Z_h}{Z_{h-1}}\right)} = \sum_m W_h^m (\alpha_h^m)^{\rho_h - \rho_{h-1}}$$

Conditional dependence between the Z_i

The conditional dependency between the latent Z_i can be measured at each sampling step by their mutual information

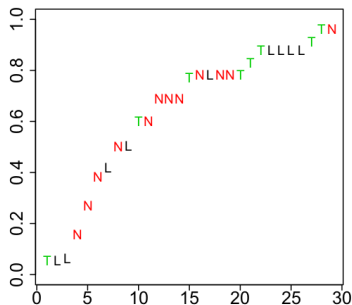
$$MI = KL\left(\prod_i p_h(Z_i) \mid p_h(Z)\right).$$

Part of the effort of the algorithm is dedicated to the recovery of this conditional dependency structure.



Onager residual structure

Estimated latent coordinate $U_i \in [0, 1]$ are uncorrelated with covariates



Individual's status: T = territorial male, N = non-lactating, L = lactating