From variational Bayes to exact posterior: a bridge sampling scheme

S. Robin

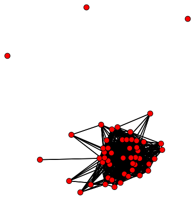
Joint work with S. Donnet

AppliBugs, Jun. 2017, Paris

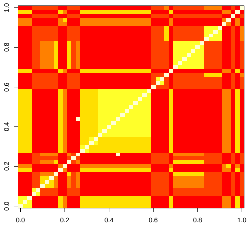
Outline

Ecological network

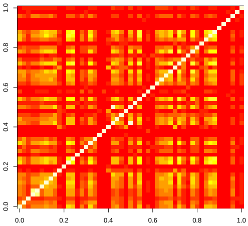
Network



Taxonomic dist.



Geographic dis.



Data.

- $n = 51$ tree species,
- $Y_{ij} = 1$ if species i and j share some (fungal) parasite, 0 otherwise
- $Y = (Y_{ij})$ = adjacency matrix of the network
- $x_{ij} = (x_{ij}^1, \dots, x_{ij}^d)$ = vector of covariates for species pair (i, j).

Problem

Questions.

1. Do covariates contribute to explain the links between the species?
2. Is there a remaining structure in the network?
3. If so, how is it organized?

Problem

Questions.

1. Do covariates contribute to explain the links between the species?
2. Is there a remaining structure in the network?
3. If so, how is it organized?

'SBM-reg' model. [LR15,LRO15]

- Each species i belongs to a (hidden) class $Z_i \in \{1, \dots K\}$
- $\pi = (\pi_1, \dots, \pi_K)$ = class proportions
- $\alpha = (\alpha_{k\ell})$ = between-classes interaction coefficients
- β = regression coefficients

Edges (Y_{ij}) are independent conditional on the classes (Z_i):

$$\text{logit } P(Y_{ij} = 1 | Z_i, Z_j) = x_{ij}^\top \beta + \alpha_{Z_i, Z_j}.$$

Bayesian inference

Priors. Parameter $\theta = (\pi, \alpha, \beta)$

$$\pi \sim \mathcal{D}(\cdot), \qquad \alpha_{kl} \sim \mathcal{N}(\cdot, \cdot), \qquad \beta \sim \mathcal{N}(\cdot, \cdot), \qquad K \sim \mathcal{U}[\dots]$$

Bayesian inference

Priors. Parameter $\theta = (\pi, \alpha, \beta)$

$$\pi \sim \mathcal{D}(\cdot), \qquad \alpha_{kl} \sim \mathcal{N}(\cdot, \cdot), \qquad \beta \sim \mathcal{N}(\cdot, \cdot), \qquad K \sim \mathcal{U}[\dots]$$

Questions.

1. Effect of the covariates: $p(\beta|Y)$
2. Remaining structure: $P(K = 1|Y)$
3. Classification $p(Z_i = k|Y)$

Bayesian inference

Priors. Parameter $\theta = (\pi, \alpha, \beta)$

$$\pi \sim \mathcal{D}(\cdot), \qquad \alpha_{kl} \sim \mathcal{N}(\cdot,\cdot), \qquad \beta \sim \mathcal{N}(\cdot,\cdot), \qquad K \sim \mathcal{U}[\dots]$$

Questions.

1. Effect of the covariates: $p(\beta|Y)$
2. Remaining structure: $P(K=1|Y)$
3. Classification $p(Z_i = k|Y)$

Problem. No conjugacy holds → intractable posterior and conditional

$$p(\theta|Y), \qquad p(Z|Y), \qquad p(\theta, Z|Y)$$

Variational Bayes inference

General principle. Approximate $p(\theta, Z|Y)$ with

$$\widetilde{p}_Y(\theta, Z) = \arg\min_{q \in \mathcal{Q}} KL\left[q(\cdot,\cdot)||p(\cdot,\cdot|Y)\right]$$

$\widetilde{p}_Y$ can be retrieved using a VB-EM algorithm [BG03]

Variational Bayes inference

General principle. Approximate $p(\theta, Z|Y)$ with

$$\widetilde{p}_Y(\theta, Z) = \arg\min_{q \in \mathcal{Q}} KL\left[q(\cdot,\cdot) || p(\cdot,\cdot|Y)\right]$$

$\widetilde{p}_Y$ can be retrieved using a VB-EM algorithm [BG03]

Here we take

$$q(\theta, Z) = \mathcal{D}(\pi; \cdot) \times \prod_{k \leq \ell} \mathcal{N}(\alpha_{kl}; \cdot, \cdot) \times \mathcal{N}(\beta; \cdot, \cdot) \times \prod_i \mathcal{M}(Z_i; \cdot)$$

- [LR15] + R package Mixer

So far so good, but

Properties of VB inference.

- No general theoretical guaranties
- Some specific favorable cases (SBM)
- In most cases

$$\widetilde{\mathbb{E}}(\theta) \simeq \mathbb{E}(\theta|Y), \qquad \widetilde{\mathbb{V}}(\theta) \ll \mathbb{V}(\theta|Y),$$

KL-minimization captures the mode but underestimates the variability.

→ Credibility intervals and model selection may not be valid

Outline

General problem

Bayesian inference: θ = parameter, Y = observed data

$$\begin{aligned} \text{prior distribution:} \quad & \theta & \sim \pi(\cdot) \\ \text{likelihood:} \quad & Y|\theta & \sim \ell(\cdot|\theta) \\ \text{posterior distribution:} \quad & \theta|Y & \sim p(\cdot|Y) \end{aligned}$$

→ Goal: evaluate $p(\theta|Y)$

General problem

Bayesian inference: θ = parameter, Y = observed data

$$\begin{aligned} \text{prior distribution:} \quad & \theta & \sim \pi(\cdot) \\ \text{likelihood:} \quad & Y|\theta & \sim \ell(\cdot|\theta) \\ \text{posterior distribution:} \quad & \theta|Y & \sim p(\cdot|Y) \end{aligned}$$

→ Goal: evaluate $p(\theta|Y)$

With latent variables: same goal for

$$p(\theta, Z|Y)$$

Now focusing on $p(\theta|Y)$

Our goal

Three main approaches to get $p(\theta|Y)$

	Method	Pros.	Cons.
Exact	conjugacy, algebra	exact	often intractable
Approximate	VB	fast	approximate
Sampling	MCMC, Gibbs, SMC	exact (if convergence)	computational time

Our goal

Three main approaches to get $p(\theta|Y)$

	Method	Pros.	Cons.
Exact	conjugacy, algebra	exact	often intractable
Approximate	VB	fast	approximate
Sampling	MCMC, Gibbs, SMC	exact (if convergence)	computational time

Our goal

- Get an exact sample from $p(\theta|Y)$
- Avoiding MCMC's convergence issues
- Taking advantage of a quickly available approximation $\widetilde{p}_Y(\theta)$

First idea: Importance sampling

Goal. Estimate

$$\mathbb{E}[f(\theta)|Y] = \int f(\theta)p(\theta|Y)\ \mathrm{d}\theta$$

First idea: Importance sampling

Goal. Estimate

$$\mathbb{E}[f(\theta)|Y] = \int f(\theta)p(\theta|Y) \, \mathrm{d}\theta$$

Monte-Carlo: (θ^m) = iid sample from $p(\cdot|Y)$

$$\widehat{\mathbb{E}}[f(\theta)|Y] = \frac{1}{M}\sum_m f(\theta^m)$$

First idea: Importance sampling

Goal. Estimate

$$\mathbb{E}[f(\theta)|Y] = \int f(\theta)p(\theta|Y)\,\mathrm{d}\theta$$

Monte-Carlo: (θ^m) = iid sample from $p(\cdot|Y)$

$$\widehat{\mathbb{E}}[f(\theta)|Y] = \frac{1}{M}\sum_m f(\theta^m)$$

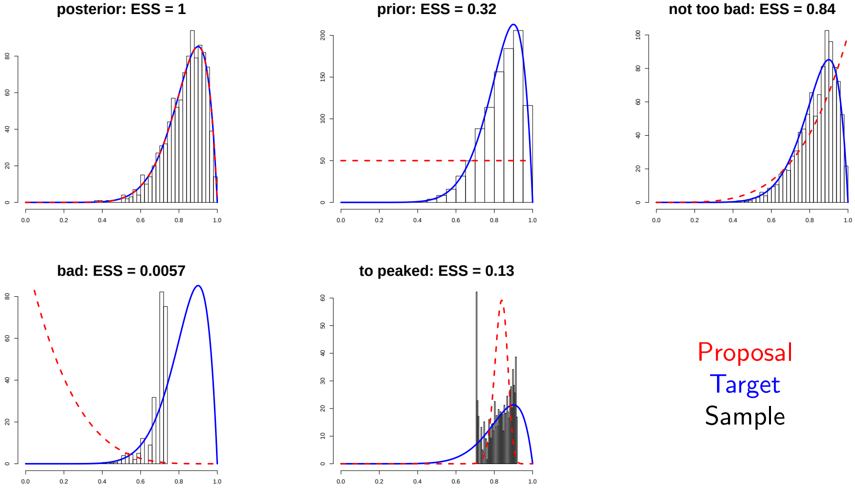
Importance sampling: (θ^m) = iid sample from a proposal q

$$\widehat{\mathbb{E}}[f(\theta)|Y] = \sum_m W^m f(\theta^m), \qquad w^m = \frac{p(\theta^m|Y)}{q(\theta^m)}, \qquad W^m = \frac{w^m}{\sum_m w^m}$$

$\mathcal{E} = \{(\theta^m, 1)\}$ = q-sample $\quad \rightarrow \quad \mathcal{E}' = \{(\theta^m, w^m)\}$ = p^*-sample,

Importance of the proposal

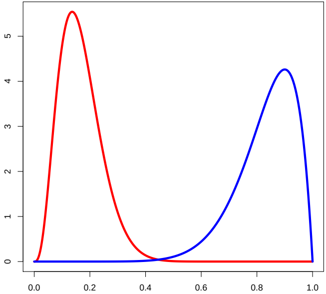
Effective sample size = $ESS := \overline{w}^2 / \overline{w^2}$.



Outline

Bridge sampling[1] principle

- q = proposal, p^* = target



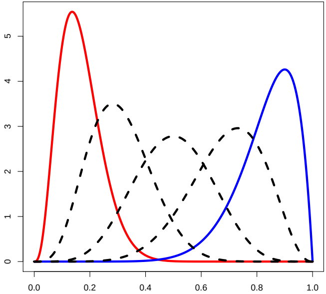
[1]'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

Bridge sampling[1] principle

- q = proposal, p^* = target
- Define intermediate distributions

$$p_0, p_1, \dots, p_H$$

with $p_0 = q$, $p_H = p^*$



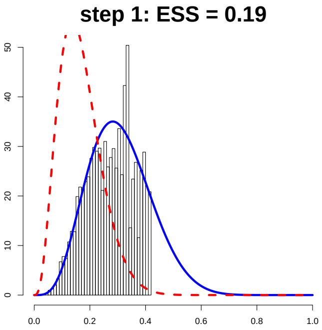
[1] 'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

Bridge sampling[1] principle

- q = proposal, p^* = target
- Define intermediate distributions

$$p_0, p_1, \dots, p_H$$

  with $p_0 = q$, $p_H = p^*$
- Iteratively:
  use p_h to get a sample from p_{h+1}



[1]'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

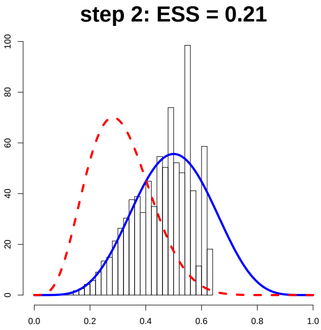
Bridge sampling[1] principle

- q = proposal, p^* = target
- Define intermediate distributions

$$p_0, p_1, \dots, p_H$$

  with $p_0 = q$, $p_H = p^*$

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



[1]'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

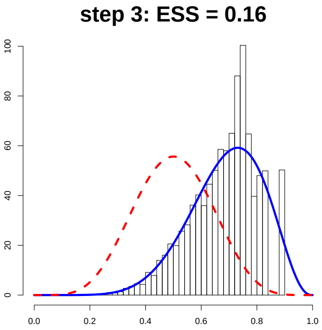
Bridge sampling[1] principle

- q = proposal, p^* = target
- Define intermediate distributions

$$p_0, p_1, \dots, p_H$$

with $p_0 = q$, $p_H = p^*$

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



[1] 'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

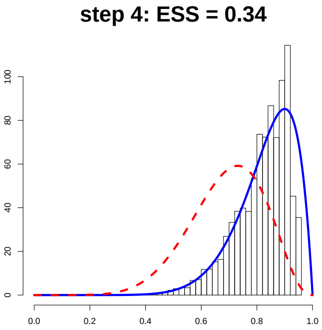
Bridge sampling[1] principle

- q = proposal, p^* = target
- Define intermediate distributions

$$p_0, p_1, \dots, p_H$$

with $p_0 = q$, $p_H = p^*$

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



[1]'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

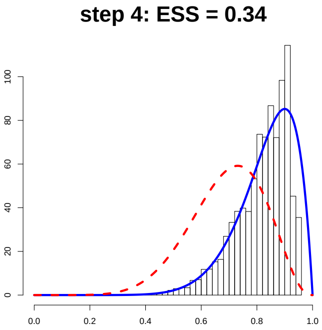
Bridge sampling[1] principle

- q = proposal, p^* = target
- Define intermediate distributions

$$p_0, p_1, \dots, p_H$$

  with $p_0 = q$, $p_H = p^*$

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



Application:

$$q = \widetilde{p}_Y, \qquad p^* = p(\cdot|Y)$$

[1] 'Bridge sampling' = 'Sequential importance sampling' (= 'SMC' ?)

Path sampling

Distribution path[2]: set $0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1$,

$$\begin{aligned} p_h(\theta) &\propto \widetilde{p}_Y(\theta)^{1-\rho_h} \times p(\theta|Y)^{\rho_h} \\ &\propto \widetilde{p}_Y(\theta) \times \alpha(\theta)^{\rho_h}, \qquad \alpha(\theta) = \frac{\pi(\theta)\ell(Y|\theta)}{\widetilde{p}_Y(\theta)} \end{aligned}$$

[2] [Nea01]: $p_h(\theta) \propto \pi(\theta)\ell(Y|\theta)^{\rho_h}$, i.e. $\widetilde{p}_Y = \pi$

Path sampling

Distribution path[2]: set $0 = \rho_0 < \rho_1 < \cdots < \rho_{H-1} < \rho_H = 1$,

$$\begin{aligned} p_h(\theta) & \propto \widetilde{p}_Y(\theta)^{1-\rho_h} \times p(\theta|Y)^{\rho_h} \\ & \propto \widetilde{p}_Y(\theta) \times \alpha(\theta)^{\rho_h}, \qquad \alpha(\theta) = \frac{\pi(\theta)\ell(Y|\theta)}{\widetilde{p}_Y(\theta)} \end{aligned}$$

Aim of bridge sampling: at each step h, provide

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

[2] [Nea01]: $p_h(\theta) \propto \pi(\theta)\ell(Y|\theta)^{\rho_h}$, i.e. $\widetilde{p}_Y = \pi$

Path sampling

Distribution path[2]: set $0 = \rho_0 < \rho_1 < \dots < \rho_{H-1} < \rho_H = 1$,

$$\begin{aligned} p_h(\theta) &\propto \widetilde{p}_Y(\theta)^{1-\rho_h} \times p(\theta|Y)^{\rho_h} \\ &\propto \widetilde{p}_Y(\theta) \times \alpha(\theta)^{\rho_h}, \qquad \alpha(\theta) = \frac{\pi(\theta)\ell(Y|\theta)}{\widetilde{p}_Y(\theta)} \end{aligned}$$

Aim of bridge sampling: at each step h, provide

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

Questions

- Step number H ?
- Step size $\rho_h - \rho_{h-1}$?
- How to actually sample p_h from the sample $\mathcal{E}_{h-1}$?

[2] [Nea01]: $p_h(\theta) \propto \pi(\theta)\ell(Y|\theta)^{\rho_h}$, i.e. $\widetilde{p}_Y = \pi$

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

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

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

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

1. set ρ_h such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

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

1. set ρ_h such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (\alpha_h^m)^{\rho_h - \rho_{h-1}}$

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

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

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

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

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

1. set ρ_h such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (\alpha_h^m)^{\rho_h - \rho_{h-1}}$
3. if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles
4. propagate the particles $\theta_h^m \sim K_h(\theta_h^m | \theta_{h-1}^m)$

Proposed 'SMC' algorithm

Init.: Sample $(\theta_0^m)_m$ iid $\sim \widetilde{p}_Y$, $w_0^m = 1$

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

1. set ρ_h such that $cESS(\mathcal{E}_{h-1}; p_{h-1}, p_h) = \tau_1$
2. compute $w_h^m = w_{h-1}^m \times (\alpha_h^m)^{\rho_h - \rho_{h-1}}$
3. if $ESS_h = \overline{w}_h^2 / \overline{w_h^2} < \tau_2$, resample the particles
4. propagate the particles $\theta_h^m \sim K_h(\theta_h^m | \theta_{h-1}^m)$

Stop: When ρ_h reaches 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 → 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 → no convergence needed

Main property. The last sample $\mathcal{E}_H = \{(\theta_H^m, w_H^m)\}$ is a weighted sample of the target distribution $p^*(\theta) = p(\theta|Y)$.

Theoretical justification

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

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

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

- 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

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

Adaptive step size

Conditional ESS: efficiency of sample $\mathcal{E}$ from q for distribution p

$$cESS(\mathcal{E}; q, p) = \frac{M\left(\sum_m W^m a^m\right)^2}{\sum_m W^m (a^m)^2}, \qquad a^m = \frac{p(\theta^m)}{q(\theta^m)}$$

→ Step 1: find next p_h s.t. sample $\mathcal{E}_{h-1}$ is reasonably efficient.

Adaptive step size

Conditional ESS: efficiency of sample $\mathcal{E}$ from q for distribution p

$$cESS(\mathcal{E}; q, p) = \frac{M \left(\sum_m W^m a^m\right)^2}{\sum_m W^m (a^m)^2}, \qquad a^m = \frac{p(\theta^m)}{q(\theta^m)}$$

→ Step 1: find next p_h s.t. sample $\mathcal{E}_{h-1}$ is reasonably efficient.

Thanks to the update formula of the weights

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

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

Denote

$$\gamma_h(\theta) = \widetilde{p}_Y(\theta)\alpha(\theta)^{\rho_h}, \qquad Z_h = \int \gamma_h(\theta) \, \mathrm{d}\theta, \qquad p_h = \gamma_h(\theta)/Z_h$$

The marginal likelihood is given by

$$p(Y) = \int \pi(\theta)\ell(Y|\theta) \, \mathrm{d}\theta = \int \gamma_H(\theta) \, \mathrm{d}\theta = Z_H$$

which can be estimated without bias with

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

Outline

Logistic regression

Model. x_i = covariates, β = regression coefficients, Y_i = binary outcome

$$\beta \sim \mathcal{N}, \qquad (Y_i) \text{ indep.} \mid \beta \sim \mathcal{B}(p_i), \qquad \text{logit}(p_i) = x_i^\intercal \beta$$

[JJ00] VBEM with approximate Gaussian posterior for β.

Logistic regression

Model. x_i = covariates, β = regression coefficients, Y_i = binary outcome

$$\beta \sim \mathcal{N}, \qquad (Y_i) \text{ indep.} \mid \beta \sim \mathcal{B}(p_i), \qquad \text{logit}(p_i) = x_i^\intercal \beta$$

[JJ00] VBEM with approximate Gaussian posterior for β.

Simulation study.

- $n = 200$, $d = 4$ covariates
- Aim: compare initial proposals

Logistic regression: Sampling path

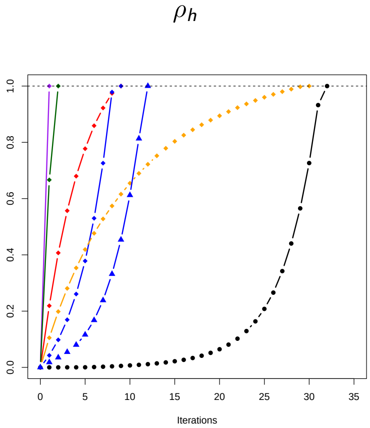
SMC: $(\Delta_{VB} = \text{diag}(\Sigma_{VB}))$

- ◆ : $\widetilde{p}_Y = \widetilde{p}_{VB}$
- ◆ : $\widetilde{p}_Y = \widetilde{p}_{ML}$
- ◆ : variance $\widetilde{p}_Y = \Delta_{VB}/5$
- ◆ : variance $\widetilde{p}_Y = 10\Delta_{VB}$
- ◆ : $\widetilde{p}_Y = \mathcal{N}(\mu_{VB} + .5, \Delta_{VB}/5)$

[Nea01]:

- ● : $\widetilde{p}_Y = \pi$

▲ = hybrid



SBM-regmodel

Model. Z_i node class, Y_{ij} links, x_{ij} edge covariates.

$$(Z_i) \text{ iid } \mathcal{M}(1; \pi), \quad Y_{ij} | Z_i, Z_j \sim \mathcal{B}(p_{ij}), \quad \text{logit}(p_{ij}) = x_{ij}^\top \beta + \alpha_{Z_i, Z_j}$$

Parameter $\theta = (\pi, \alpha, \beta)$.

SBM-regmodel

Model. Z_i node class, Y_{ij} links, x_{ij} edge covariates.

$$(Z_i) \text{ iid } \mathcal{M}(1;\pi), \quad Y_{ij}|Z_i, Z_j \sim \mathcal{B}(p_{ij}), \quad \text{logit}(p_{ij}) = x_{ij}^\top \beta + \alpha_{Z_i,Z_j}$$

Parameter $\theta = (\pi, \alpha, \beta)$.

Simulation design.

- $n = 20, 50$ nodes, $K^* = 1, 2$ classes, $d = 3$ covariates,
- $M = 1000$ particles, $B = 100$ samples.
- Parameters sampled from the prior.

SBM-regmodel

Model. Z_i node class, Y_{ij} links, x_{ij} edge covariates.

$$(Z_i) \text{ iid } \mathcal{M}(1;\pi), \quad Y_{ij}|Z_i, Z_j \sim \mathcal{B}(p_{ij}), \quad \text{logit}(p_{ij}) = x_{ij}^\top \beta + \alpha_{Z_i,Z_j}$$

Parameter $\theta = (\pi, \alpha, \beta)$.

Simulation design.

- $n = 20, 50$ nodes, $K^* = 1, 2$ classes, $d = 3$ covariates,
- $M = 1000$ particles, $B = 100$ samples.
- Parameters sampled from the prior.

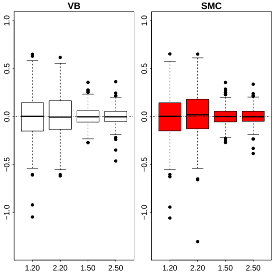
Property check. $\theta^* \sim \pi$, $Y \sim \ell(Y|\theta^*)$ and $\{(\theta^m, w^m)\}$ a sample from $q(\theta)$:

$$q(\theta) = p(\theta|Y) \qquad \Rightarrow \qquad \sum_m W_m \mathbb{I}\{\theta^m \leq \theta^*\} \sim \mathcal{U}[0,1]$$

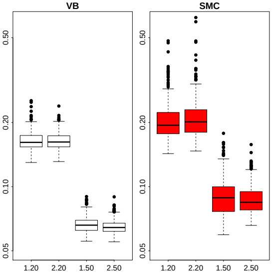
SBM-reg: K^* known

Posterior distribution of the regression coefficients β_ℓ

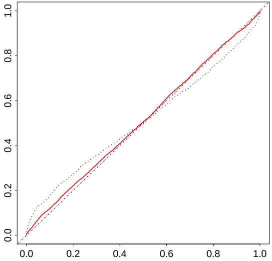
posterior mean



posterior sd

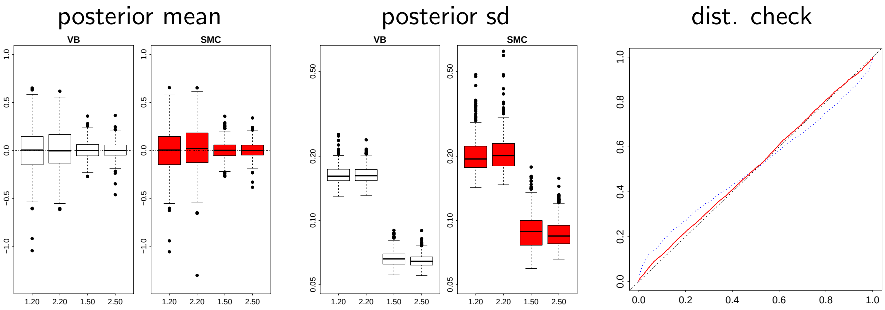


dist. check



SBM-reg: K^* known

Posterior distribution of the regression coefficients β_ℓ



Empirical level of 95%-credibility intervals (CI):

VB: 84.75%, SMC: 93.75%

SBM-reg: Model selection

For each sample, compute

$$p_{SMC}(K|Y) = \widehat{Z}_H, \qquad p_{VB}(K|Y) = \widetilde{p}_Y(K)$$

$\widehat{K}_{SMC} = \arg\max_K p_{SMC}(K|Y)$, idem $\widehat{K}_{VB}$

SBM-reg: Model selection

For each sample, compute

$$p_{SMC}(K|Y) = \widehat{Z}_H, \qquad p_{VB}(K|Y) = \widetilde{p}_Y(K)$$

$\widehat{K}_{SMC} = \arg\max_K p_{SMC}(K|Y)$, idem $\widehat{K}_{VB}$

Results.

		$\widehat{K} = K^*$		mean $p(K^*|Y)$	
n	g^*	VB	SMC	VB	SMC
20	1	1.00	0.46	0.947	0.435
20	2	0.10	0.23	0.138	0.257
50	1	1.00	0.60	0.982	0.562
50	2	0.42	0.36	0.410	0.387

→ Better performances for VB...

SBM-reg: Model averaging

Account for model uncertainty [HMRV99]: Rather than choosing $\widehat{K}$, consider

$$
\begin{aligned}
p(\theta|Y) &= \sum_K p(K|Y) p(\theta|Y, K) \\
\Rightarrow \qquad \mathbb{V}(\theta|Y) &= \underbrace{\mathbb{E}_{K|Y}\left[\mathbb{V}(\theta|Y, K)\right]}_{\text{within models}} + \underbrace{\mathbb{V}_{K|Y}\left[\mathbb{E}(\theta|Y, K)\right]}_{\text{between models}}
\end{aligned}
$$

SBM-reg: Model averaging

Account for model uncertainty [HMRV99]: Rather than choosing $\widehat{K}$, consider

$$p(\theta|Y) = \sum_K p(K|Y) p(\theta|Y, K)$$

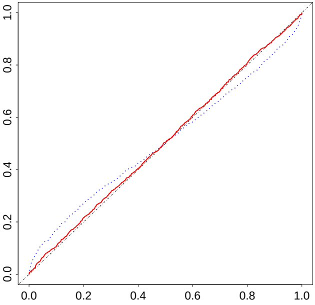
$$\Rightarrow \quad \mathbb{V}(\theta|Y) = \underbrace{\mathbb{E}_{K|Y}\left[\mathbb{V}(\theta|Y, K)\right]}_{\text{within models}} + \underbrace{\mathbb{V}_{K|Y}\left[\mathbb{E}(\theta|Y, K)\right]}_{\text{between models}}$$

Results.

Empirical level of 95%-CI:

VB: 85.8%

SMC: 93.25%



Outline

Tree network

Covariates: x_{ij} = genetic, geographic and taxonomic distances

Posterior distribution of the regression coefficients

	VB			SMC		
	genet.	geo.	taxo.	genet.	geo.	taxo.
mean	$4.6\ 10^{-5}$	$2.3\ 10^{-1}$	$-9.0\ 10^{-1}$	$4.1\ 10^{-5}$	$3.6\ 10^{-1}$	$-9.1\ 10^{-1}$
within var.	$2.2\ 10^{-10}$	$4.3\ 10^{-2}$	$1.7\ 10^{-3}$	$1.1\ 10^{-9}$	$2.2\ 10^{-1}$	$8.9\ 10^{-3}$
between var.	$5.6\ 10^{-17}$	$1.2\ 10^{-6}$	$2.4\ 10^{-7}$	$4.0\ 10^{-12}$	$1.9\ 10^{-3}$	$2.8\ 10^{-3}$
st. dev.	$1.5\ 10^{-5}$	$2.1\ 10^{-1}$	$4.2\ 10^{-2}$	$3.3\ 10^{-5}$	$4.7\ 10^{-1}$	$1.1\ 10^{-1}$
ratio	3.1	1.1	−21	1.2	$7.6\ 10^{-1}$	−8.4

- Smaller posterior between-model variance with VB
- Smaller posterior variance with VB
- Can affect the conclusions in terms of significance

Residual structure

Following [LRO15],

$$P(K = 1|Y) = P(\text{no residual structure}|Y)$$

measures the goodness-of-fit of the regression model

Some examples.

Network	Marriage	Business	Karate	Tree	Blog
n	16	16	34	51	196
d	3	3	8	3	3
$p_{VB}(K = 1\|Y)$	$9.54\ 10^{-1}$	$7.04\ 10^{-1}$	$2.56\ 10^{-1}$	$4.83\ 10^{-153}$	$8.63\ 10^{-174}$
$p_{SMC}(K = 1\|Y)$	1.00	1.00	$7.07\ 10^{-3}$	$1.06\ 10^{-161}$	$4.04\ 10^{-290}$

- Similar conclusions with VB and SMC
- But the estimated residual graphon may be different

Outline

Conclusion

Summary.

- A generic framework to get an exact sample from the posterior
- Taking advantage of fast preliminary inference (VB, ML, ...)
- No convergence issue (as opposed to MCMC)

Conclusion

Summary.

- A generic framework to get an exact sample from the posterior
- Taking advantage of fast preliminary inference (VB, ML, ...)
- No convergence issue (as opposed to MCMC)

Some limitations.

- Large number of iterations when starting far from the target
- Requires a model-specific Gibbs sampler
- Suffer general issues in Bayesian inference (e.g. label switching)

References

J. Beal, M. and Z. Ghahramani. The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayes. Statist.*, 7:543–52, 2003.

P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 68(3):411–436, 2006.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.

T. S. Jaakkola and M. I. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10(1):25–37, 2000.

P. Latouche and S. Robin. Variational bayes model averaging for graphon functions and motif frequencies inference in W-graph models. *Statistics and Computing*, pages 1–13, 2015.

P. Latouche, S. Robin, and S. Ouadah. Goodness of fit of logistic models for random graphs. Technical report, arXiv:1508.00286, 2015.

R. M Neal. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139, 2001.