High-dimensional Bayesian asymptotics and computation

Aguemon Yves Atchadé

University of Michigan

Graphical models

- Graphs useful to represent dependencies between random variables.
- Two main types of graphical models
  - Directed acylic graph (DAG); a.k.a. Bayesian networks
  - Undirected graph; known as Markov networks. Main topic.

Graphical models

- Graphs useful to represent dependencies between random variables.
- Two main types of graphical models
  - Directed acylic graph (DAG); a.k.a. Bayesian networks
  - Undirected graph; known as Markov networks. Main topic.
- Useful in many applications: speech recognition, biological networks modeling, protein folding problems, etc...
- Some notation: $\mathcal{M}_p$ space of $p \times p$ symmetric matrices. $\mathcal{M}_p^+$ its cone of spd elements,

$$\langle A, B \rangle_F \stackrel{\text{def}}{=} \sum_{i \le j} A_{ij} B_{ij}, \quad A,\ B \in \mathcal{M}_p.$$

Graphical models

- A parametric graphical model:
- p nodes. A set $\mathsf{Y} \subset \mathbb{R}$.
- Non-zero functions $B_0 : \mathsf{Y} \to \mathbb{R}$, and $B : \mathsf{Y} \times \mathsf{Y} \to \mathbb{R}$ symmetric.
- Then define $\{f_\theta, \theta \in \Omega\}$,

$$f_\theta(y) = \frac{1}{Z(\theta)} \exp\left(\sum_{j=1}^{p} \theta_{jj} B_0(y_j) + \sum_{i<j} \theta_{ij} B(y_i, y_j)\right),$$

$$\Omega \stackrel{\text{def}}{=} \left\{\theta \in \mathcal{M}_p : \; Z(\theta) \stackrel{\text{def}}{=} \int e^{-\left\langle \theta, \bar{B}(y)\right\rangle_{\mathsf{F}}} \mathrm{d}y < \infty\right\}.$$

Graphical models

- Parametric model $\{f_\theta, \theta \in \Omega\}$.
- The parameter $\theta \in \Omega$ modulates the interaction. Importantly, $\theta_{ij} = 0$ implies conditional independence of y_i, y_j given remaining variables.

Graphical models

- Parametric model $\{f_\theta, \theta \in \Omega\}$.
- The parameter $\theta \in \Omega$ modulates the interaction. Importantly, $\theta_{ij} = 0$ implies conditional independence of y_i, y_j given remaining variables.
- It is often very appealing to assume that θ is sparse, particularly when p is large.
- Goal: estimate $\theta \in \Omega$ from multiple (n) samples from $f_{\theta_\star}$ arranged in a data matrix $Z \in \mathbb{R}^{n \times p}$.

Graphical models

- Given a prior Π on Ω. Main object of interest:

$$\Pi(\mathrm{d}\theta|Z) \propto \Pi(\mathrm{d}\theta) \prod_{i=1}^{n} f_\theta(Z_{i\cdot}).$$

- Set Δ the set of graph-skeletons (symmetric $0-1$ matrices with diagonal 1). For sparse estimation, we consider priors of the form

$$\Pi(\mathrm{d}\theta) = \sum_{\delta\in\Delta} \pi_\delta \Pi(\mathrm{d}\theta|\delta),$$

- where $\Pi(\mathrm{d}\theta|\delta)$ has support $\Omega(\delta)$.

Graphical models

- Given a prior Π on Ω. Main object of interest:

$$\Pi(\mathrm{d}\theta|Z) \propto \Pi(\mathrm{d}\theta) \prod_{i=1}^{n} f_\theta(Z_{i\cdot}).$$

- Set Δ the set of graph-skeletons (symmetric $0-1$ matrices with diagonal 1). For sparse estimation, we consider priors of the form

$$\Pi(\mathrm{d}\theta) = \sum_{\delta \in \Delta} \pi_\delta \Pi(\mathrm{d}\theta|\delta),$$

- where $\Pi(\mathrm{d}\theta|\delta)$ has support $\Omega(\delta)$.
- Difficulty with $\Pi(\cdot|Z)$: Either the likelihood is intractable,
- Or $\Omega(\delta)$ is a complicated space and prior is intractable.

Quasi-Bayesian inference

- In large applications, it may be worth exploring less accurate but faster alternatives.
- Quasi-Bayesian inference is a framework to formulate these trade-offs.
- Think of Quasi-Bayesian inference as the Bayesian analog of M-estimation.
- General idea: instead of the model $\{f_\theta,\ \theta \in \Omega\}$, we consider a "larger pseudo-model" $\{\check{f}_\theta,\ \theta \in \check{\Omega}\}$.

Quasi-Bayesian inference

- Pseudo-model: $z \mapsto \check{f}_\theta(z)$ needs not be a density. Chosen for computational convenience.
- Larger pseudo-model: $\Omega \subseteq \check{\Omega}$. Very useful to build interesting priors on $\check{\Omega}(\delta)$.

Quasi-Bayesian inference

- Pseudo-model: $z \mapsto \check{f}_\theta(z)$ needs not be a density. Chosen for computational convenience.
- Larger pseudo-model: $\Omega \subseteq \check{\Omega}$. Very useful to build interesting priors on $\check{\Omega}(\delta)$.
- Quasi-posterior distributions have been used extensively in the PAC-Bayesian literature (Catoni 2004).
- ABC is a form of quasi-Bayesian inference.
- Chernozukhov-Hong (J. Econ. 2003). Also popular in Bayesian semi-parametric inference (Yang & He (AoS 2012), Kato (AoS 2013).

Asymptotics of quasi-posterior distributions

Consider the quasi-posterior distribution

$$\check{\Pi}(\mathrm{d}\theta|Z) \propto q_\theta(Z)\Pi(\mathrm{d}\theta).$$

Asymptotics of quasi-posterior distributions

Consider the quasi-posterior distribution

$$\check{\Pi}(\mathrm{d}\theta|Z) \propto q_\theta(Z)\Pi(\mathrm{d}\theta).$$

Theorem

$\check{\Pi}(\cdot|Z)$ is a solution to the problem

$$\min_{\mu \ll \Pi} \left[-\int_{\mathbb{R}^d} \log q_\theta(Z)\mu(\mathrm{d}\theta) + KL(\mu|\Pi) \right],$$

where $KL(\mu|\Pi) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} \log(\mathrm{d}\mu/\mathrm{d}\Pi)\mathrm{d}\mu$ is the KL-divergence of Π from μ.

- Proof is Easy. See e.g. T. Zhang (AoS 2006).
- If q_θ is good enough for a frequentist M-estimation inference, it is good enough for a quasi-Bayesian inference– upto the prior.

Example: binary graphical models

- Binary graphical model. Y $= \{0, 1\}$. $B(x, y) = xy$. Here $\Omega = \mathcal{M}_p$ and

$$Z(\theta) \quad \text{is typically intractable} \; .$$

Example: binary graphical models

- Binary graphical model. $\mathsf{Y} = \{0,1\}$. $B(x,y) = xy$. Here $\Omega = \mathcal{M}_p$ and

$$Z(\theta) \quad \text{is typically intractable} \;.$$

- There is a very commonly used pseudo-likelihood function to circumvent the intractable normalizing constant.

$$\begin{aligned} q_\theta(Z) &= \prod_{i=1}^n \prod_{j=1}^p \frac{\exp\left(Z_{ij}\left(\theta_{jj} + \sum_{k\neq j} \theta_{jk} Z_{ik}\right)\right)}{1+\exp\left(\theta_{jj} + \sum_{k\neq j} \theta_{jk} Z_{ik}\right)}, \; \theta \in \mathcal{M}_p, \\ &= \prod_{i=1}^n \prod_{j=1}^p f_{\theta_{\cdot j}}^{(j)}(Z_{ij} | Z_{i,-j}) \;\; \theta \in \mathcal{M}_p, \end{aligned}$$

- Note: $f_{\theta_{\cdot j}}^{(j)}(Z_{ij}|Z_{i,-j})$ depends only on the j-th column of θ.

Example: binary graphical models

- Then very easy to set up prior of $\mathcal{M}_p(\delta)$.
- However, dimension of $\mathcal{M}_p$ grows fast. Larger than 10^5, for $p \approx 500$.

Example: binary graphical models

- Then very easy to set up prior of $\mathcal{M}_p(\delta)$.
- However, dimension of $\mathcal{M}_p$ grows fast. Larger than 10^5, for $p \approx 500$.
- We can further simplify the problem by enlarging the parameter space from $\mathcal{M}_p$ to $\mathbb{R}^{p\times p}$:

$$\begin{aligned} q_\theta(Z) &= \prod_{i=1}^{n}\prod_{j=1}^{p} f^{(j)}_{\theta_{\cdot j}}(Z_{ij}|Z_{i,-j}) \;\; \theta \in \mathbb{R}^{p\times p}, \\ &= \prod_{j=1}^{p}\left(\prod_{i=1}^{n} f^{(j)}_{\theta_{\cdot j}}(Z_{ij}|Z_{i,-j})\right), \;\; \theta \in \mathbb{R}^{p\times p}. \end{aligned}$$

- In that case $q_\theta(Z)$ factorizes along the columns of θ.

Example: binary graphical models

- Take p independent sparsity inducing priors on $\mathbb{R}^p$, and we get a posterior on $\mathbb{R}^{p\times p}$:

$$\check{\Pi}(\mathrm{d}\theta|Z) = \prod_{j=1}^{p} \check{\Pi}_j(\mathrm{d}\theta_{\cdot j}|Z),$$

  where

$$\check{\Pi}_j(\mathrm{d}u|Z) = \prod_{i=1}^{n} f_{\theta_{\cdot j}}^{(j)}(Z_{ij}|Z_{i,-j}) \sum_{\delta\in\Delta_p} \pi_\delta \Pi(\mathrm{d}\theta|\delta).$$

- We can sample from the distribution $\check{\Pi}_j(\mathrm{d}\theta|Z)$ in parallel. Potentially huge computing gain.

Example: binary graphical models

- Very popular method for fitting large graphical models in frequentist inference.
- Initially introduced by Meinhausen & Buhlmann (AoS 2006), for Gaussian graphical models.
- See also Ravikumar et al. (AoS 2010) for binary graphical models. Sun & Zhang (JMLR, 2013) for a scaled-Lasso version.
- Very efficient (divide and conquer). We can fit $p = 1000$ nodes in few minutes on large clusters.

Example: binary graphical models

- Very popular method for fitting large graphical models in frequentist inference.
- Initially introduced by Meinhausen & Buhlmann (AoS 2006), for Gaussian graphical models.
- See also Ravikumar et al. (AoS 2010) for binary graphical models. Sun & Zhang (JMLR, 2013) for a scaled-Lasso version.
- Very efficient (divide and conquer). We can fit $p = 1000$ nodes in few minutes on large clusters.
- Loss of symmetry.
- Should we worry about all the simplification involved?

Example: binary graphical models

- Assume we build the prior Π $\mathbb{R}^p$ as follows.

$$\Pi(\mathrm{d}\theta) = \sum_{\delta \in \Delta_p} \pi_\delta \Pi(\mathrm{d}\theta|\delta). \qquad (1)$$

  - $\pi_\delta = \prod_{j=1}^p q^{\delta_j}(1-q)^{1-\delta_j}$, $q = p^{-u}$, $u > 1$.
  - $$\theta_j|\delta \sim \begin{cases} \text{Dirac}(0) & \text{if } \delta_j = 0 \\ \text{Laplace}(\rho) & \text{if } \delta_j = 1 \end{cases}, \qquad (2)$$
  - $$\rho = 24\sqrt{n\log(p)}.$$
  - See Castillo et al. (AoS 2015).

Example: binary graphical models

H

H1: *There exists* $\theta_\star \in \mathcal{M}_p$ *such that the rows of* Z *are i.i.d.* $f_{\theta_\star}$.

- Set

$$s_\star \stackrel{\text{def}}{=} \max_{1\leq j\leq p} \sum_{i=1}^{p} \mathbf{1}_{\{|\theta_{ij}|>0\}},$$

  the max. degree of $\theta_\star$.

- For $\theta \in \mathbb{R}^{p\times p}$, define the norm

$$|||\theta||| \stackrel{\text{def}}{=} \sup_{1\leq j\leq p} \|\theta_{\cdot j}\|_2.$$

Example: binary graphical models

Theorem (A.A.(2015))

With prior and assumption above, and under some regularity conditions, define

$$r_{n,d} = \frac{1}{\underline{\kappa}(s_\star)} \sqrt{\frac{s_\star \log(p)}{n}}.$$

There exists universal constants $M > 2, A_1 > 0, A_2 > 0$ *such that for* p *large enough, and*

$$n \geq A_1 \left(\frac{s_\star}{\underline{\kappa}(s_\star)}\right)^2 \log(p),$$

$$\mathbb{E}\left[\check{\Pi}\left(\left\{\theta \in \mathbb{R}^{p\times p} : \|\theta - \theta_\star\| > M_0 r_{n,d}\right\} |Z\right)\right] \leq \frac{2}{e^{A_2 n}} + \frac{12}{d}.$$

Example: binary graphical models

- Gives some guarantee that the method is not completely silly.
- Regularity conditions: restricted smallest eigenvalues of Fisher information matrix bounded away from 0.
- Minimax rate. Even in full likelihood inference cannot do better in term of convergence rate.
- Extension to more general class of prior is possible.
- Similar results hold for Gaussian graphical models, and more general models.

Approximate Computations

- How to sample from

$$\check{\Pi}(\mathrm{d}\theta|Z) = q_\theta(Z) \sum_{\delta\in\Delta_p} \pi_\delta \prod_{j:\ \delta_j=1} \phi(\theta_j)\mu_{p,\delta}(\mathrm{d}\theta) \ \ ?$$

- Rather we consider:

$$\check{\Pi}(\delta, \mathrm{d}\theta|Z) = \pi_\delta \exp\left(\log q_\theta(Z) + \sum_{j=1}^{p} \delta_j \log \phi(\theta_j)\right) \mu_{p,\delta}(\mathrm{d}\theta).$$

Approximate Computations

- How to sample from

$$\check{\Pi}(\mathrm{d}\theta|Z) = q_\theta(Z) \sum_{\delta \in \Delta_p} \pi_\delta \prod_{j:\ \delta_j = 1} \phi(\theta_j) \mu_{p,\delta}(\mathrm{d}\theta) \ \ ?$$

- Rather we consider:

$$\check{\Pi}(\delta, \mathrm{d}\theta|Z) = \pi_\delta \exp\left(\log q_\theta(Z) + \sum_{j=1}^{p} \delta_j \log \phi(\theta_j) \right) \mu_{p,\delta}(\mathrm{d}\theta).$$

- Issue: for $\delta \neq \delta'$, $\check{\Pi}(\mathrm{d}\theta|\delta, Z)$ and $\check{\Pi}(\mathrm{d}\theta|\delta', Z)$ are singular measures.
- We want to avoid transdimensional MCMC techniques (reversible-jump style MCMC). Poor mixing.
- We propose an approximation method using the Moreau envelops.

Approximate Computations

- Suppose $h : \mathbb{R}^p \to (-\infty, +\infty]$ is convex (possibly not smooth).
- For $\gamma > 0$, the Moreau-Yosida approximation of h is:

$$h_\gamma(\theta) = \min_{u \in \mathbb{R}^p} \left[h(u) + \frac{1}{2\gamma} \|u - \theta\|^2 \right].$$

- h_γ is convex, class $\mathcal{C}^1$ with Lip. gradient, and $h_\gamma \uparrow h$ pointwise as $\gamma \to 0$.
- Well-studied approximation method.
- Leads to the proximal algorithm.

Approximate Computations

- In many cases, h_γ cannot be computed/evaluated.
- If $h = f + g$, and f is smooth, one can use the forward-backward approximation

$$\tilde{h}_\gamma(x) = \min_{u \in \mathbb{R}^d} \left[f(x) + \langle \nabla f(x), u - x \rangle + g(u) + \frac{1}{2\gamma} \|u - x\|^2 \right].$$

- $\tilde{h}_\gamma \leq h_\gamma \leq h$, and has similar properties as h_γ.
- h_γ is easy to compute when g is simple enough.
- Explored by (Pereyra (Stat. Comp. (2015), Schrek et al. (2014)) as proposal mechanism in MCMC.

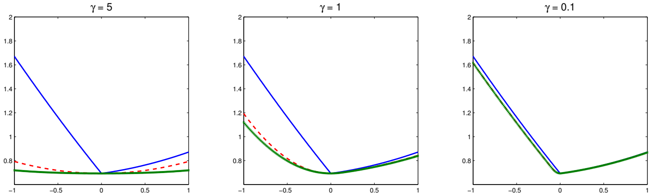

Figure: Figure showing the function $h(x) = -ax + \log(1 + e^{ax}) + b|x|$ for $a = 0.8$, $b = 0.5$ (blue/solid line), and the approximations h_γ and $\tilde{h}_\gamma$ ($h_\gamma \leq \tilde{h}_\gamma$), for $\gamma \in \{5, 1, 0.1\}$.

Approximate Computations

- For $\gamma > 0$, the Moreau-Yosida approximation of h is:

$$h_\gamma(\theta) = \min_{u\in\mathbb{R}^p}\left[h(u) + \frac{1}{2\gamma}\|u-\theta\|^2\right].$$

- Notice that even if $\mathrm{dom}(h) \neq \mathbb{R}^p$, h_γ is still finite everywhere.
- Hence if $h(x) = -\log \pi(x)$ for some log-concave density π

$$\pi_\gamma(x) = \frac{1}{Z_\gamma} e^{-h_\gamma(x)},\ x \in \mathbb{R}^p,$$

  is an approximation of π (assume $Z_\gamma < \infty$), and $\pi_\gamma \ll \mathrm{Leb}_{\mathbb{R}^d}$.
- We show that π_γ converges weakly to π as $\gamma \to 0$.

Approximate Computations

- Back to $\check{\Pi}(\cdot|Z)$.

$$\check{\Pi}(\delta, \mathrm{d}\theta|Z) \propto \pi_\delta \exp\left(\log q_\theta(Z) + \sum_{j=1}^{p} \delta_j \log \phi(\theta_j)\right) \mu_{p,\delta}(\mathrm{d}u),$$

$$\propto \pi_\delta \exp\left[\underbrace{\log q_\theta(Z) + \sum_{j=1}^{d} \delta_j \log \phi(\theta_j) - \iota_{\Theta_\delta}(\theta)}_{-h(\theta|\delta)}\right] \mu_{p,\delta}(\mathrm{d}u).$$

- Leads to

$$\check{\Pi}_\gamma(\delta, \mathrm{d}\theta) \propto \pi_\delta \left(2\pi\gamma\right)^{\|\delta\|_1/2} e^{-h_\gamma(\theta|\delta)} \mathrm{d}\theta,$$

where $h_\gamma(\cdot|\delta)$ is the forward-backward approx. of h.

Approximate Computations

$$\check{\Pi}_\gamma(\delta, \mathrm{d}\theta) \propto \pi_\delta \, (2\pi\gamma)^{\frac{\|\delta\|_1}{2}} \, e^{-h_\gamma(\theta|\delta)} \mathrm{d}\theta.$$

Approximate Computations

$$\check{\Pi}_\gamma(\delta, \mathrm{d}\theta) \propto \pi_\delta \,(2\pi\gamma)^{\frac{\|\delta\|_1}{2}}\, e^{-h_\gamma(\theta|\delta)}\mathrm{d}\theta.$$

- Assume: $-\log q_\theta(Z)$ is convex, has L-Lip. gradient, and

$$-\log q_\theta(Z) \geq \frac{1}{2L}\|\nabla \log q_\theta(Z)\|^2.$$

- Assume: $-\log \phi$ is convex.

Theorem

Take $\gamma = \gamma_0/L$, $\gamma_0 \in (0, 1/4]$. Then Π_γ is a well-defined p.m. on $\Delta_p \times \mathbb{R}^p$, and there exists a finite constant (in p) C such that

$$\beta\left(\check{\Pi}_\gamma, \check{\Pi}\right) \leq \sqrt{\gamma_0} + C\gamma_0 p,$$

where $\beta(\cdot,\cdot)$ is the β-metric between p.m. (metricizes weak convergence).

Approximate Computations

- In theory, we get better bound by taking for e.g.

$$\gamma = \frac{\gamma_0}{Lp}.$$

- However as $\check{\Pi}_\gamma$ gets very close to $\check{\Pi}$, sampling from $\check{\Pi}_\gamma$ becomes hard.
- The theorem above is a worst case analysis. What is the behavior for typical data realizations?

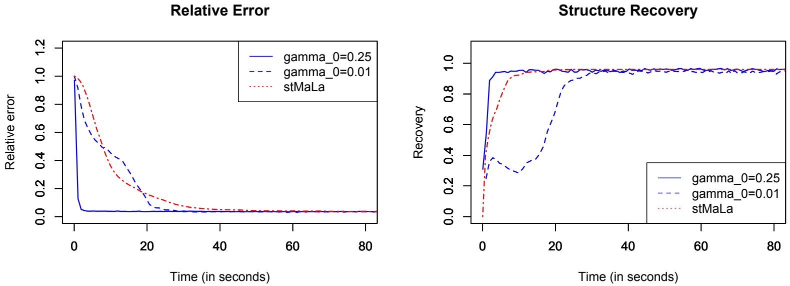

Figure: Sparse Bayesian linear regression example. $p = 500$, $n = 200$.

Approximate Computations

$$\check{\Pi}_\gamma(\delta, \mathrm{d}\theta) \propto \pi_\delta \, (2\pi\gamma)^{\frac{\|\delta\|_1}{2}} \, e^{-h_\gamma(\theta|\delta)} \mathrm{d}\theta.$$

- Linear regression: $-\log q_\theta(Z) = \|Z - X\theta\|^2/2\sigma^2$.
- Assume $Z \sim \mathbf{N}(X\theta_\star, \sigma^2 I_n)$.
- Assume: the sparse prior assumption in Theorem 1.

Theorem

Take $\gamma = \gamma_0/L$, $\gamma_0 \in (0, 1/4]$. There exists a finite constant (in p) C such that

$$\mathbb{E}\left[\beta\left(\check{\Pi}_\gamma, \check{\Pi}\right)\right] \leq \sqrt{\gamma_0} + C\left(1 + \gamma_0 \log(p)\right).$$

Approximate Computations

$$\check{\Pi}_\gamma(\delta, \mathrm{d}\theta) \propto \pi_\delta \, (2\pi\gamma)^{\frac{\|\delta\|_1}{2}} \, e^{-h_\gamma(\theta|\delta)}\mathrm{d}\theta.$$

- We can sample from $\check{\Pi}$ using "standard" MCMC methods.
- Key advantage: given θ, the comp. of δ are conditionally indep. Bernoulli.
- Given δ, do a Metropolis-Langevin approach that takes adv. of the smoothness of h_γ.
- The gradient of $\theta \mapsto h_\gamma(\theta|\delta)$ is related to the proximal map of h.

A Gaussian graphical example

- Example: sparse estimation of large Gaussian graphical models.
- We compare the quasi-posterior mean and g-lasso estimator

$$\hat{\vartheta}_{\mathsf{glasso}} = \mathsf{Argmin}_{\theta \in \mathcal{M}_p^+} \left[-\log\det\theta + \mathsf{Tr}(\theta S) + \lambda \sum_{i,j} \left(\alpha|\theta_{ij}| + \frac{(1-\alpha)}{2}\theta_{ij}^2 \right) \right],$$

where $S = (1/n)Z'Z$.

- We do comparison along:

$$\mathcal{E} = \frac{\|\hat{\vartheta} - \vartheta\|_{\mathsf{F}}}{\|\vartheta\|_{\mathsf{F}}}, \quad \mathsf{SEN} = \frac{\sum_{i<j} \mathbf{1}_{\{|\vartheta_{ij}|>0\}} \mathbf{1}_{\{\mathsf{sign}(\hat{\vartheta}_{ij}) = \mathsf{sign}(\vartheta_{ij})\}}}{\sum_{i<j} \mathbf{1}_{\{|\vartheta_{ij}|>0\}}};$$

$$\text{and} \quad \mathsf{PREC} = \frac{\sum_{i<j} \mathbf{1}_{\{|\hat{\vartheta}_{ij}|>0\}} \mathbf{1}_{\{\mathsf{sign}(\hat{\vartheta}_{ij}) = \mathsf{sign}(\vartheta_{ij})\}}}{\sum_{i<j} \mathbf{1}_{\{|\hat{\vartheta}_{ij}|>0\}}}. \quad (3)$$

A Gaussian graphical example

	ϑ_{jj}^2 known	Empirical Bayes	Glasso
Relative Error ($\mathcal{E}$ in %)	19.2	21.6	63.1
Sensitivity (SEN in %)	68.4	69.0	40.5
Precision (PREC in %)	100.0	100.0	74.9

Table: Table showing the relative error, sensitivity and precision (as defined in (3)) for Setting (a), with $p = 100$ nodes. Based on 20 simulation replications. Each MCMC run is 5×10^4 iterations.

A Gaussian graphical example

	ϑ_{jj}^2 known	Empirical Bayes	Glasso
Relative Error ($\mathcal{E}$ in %)	23.1	26.2	45.2
Sensitivity (SEN in %)	44.6	45.4	87.9
Precision (PREC in %)	100	99.9	56.1

Table: Table showing the relative error, sensitivity and precision (as defined in (3)) for Setting (b), with $p = 500$ nodes. Based on 20 simulation replications. Each MCMC run is 5×10^4 iterations.

A Gaussian graphical example

	ϑ_{jj}^2 known	Empirical Bayes	Glasso
Relative Error ($\mathcal{E}$ in %)	30.8	35.2	66.9
Sensitivity (SEN in %)	16.3	16.4	6.6
Precision (PREC in %)	99.9	99.8	94.7

Table: Table showing the relative error, sensitivity and precision (as defined in (3)) for Setting (c), with $p = 1,000$ nodes. Based on 20 simulation replications. Each MCMC run is 5×10^4 iterations.

A Gaussian graphical example

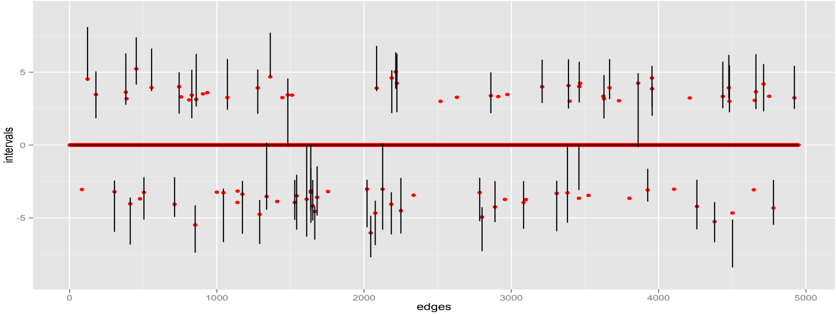

Figure: Figure showing the confidence interval bars (obtained from one MCMC run), for the non-diagonal entries of ϑ in Setting (a). The dots represent the true values.

Conclusion

- Quasi-posterior inference is consistent in high-dimensional setting.
- On the approx. computation, how to formalize the trade-off between good approx. and fast MCMC computation.
- Joint statistical and computational asymptotics.
- Matlab code available from website.

Conclusion

- Quasi-posterior inference is consistent in high-dimensional setting.
- On the approx. computation, how to formalize the trade-off between good approx. and fast MCMC computation.
- Joint statistical and computational asymptotics.
- Matlab code available from website.
- Postdoc opening available at the University of Michigan:

  `www.stat.lsa.umich.edu/~yvesa`

Conclusion

- Quasi-posterior inference is consistent in high-dimensional setting.
- On the approx. computation, how to formalize the trade-off between good approx. and fast MCMC computation.
- Joint statistical and computational asymptotics.
- Matlab code available from website.
- Postdoc opening available at the University of Michigan:

  `www.stat.lsa.umich.edu/~yvesa`
- Thanks for your attention... and patience !