Identifying Summary and Parameter Structures in ABC: A Gaussian Graphical Model Approach

Yangqi ZHANG

UNSW

Bayesian Nonparametrics Networking Workshop 2023

December 3, 2023

Agenda

Table of Contents

Likelihood-free inference using summary statistics

- We study the problem of estimating complex models with an intractable likelihood.
- Instead, the evaluation of likelihood is approximated by auxiliary statistics/ moments.
- Approximate Bayesian Computation (ABC) and Simulated Method of Moments (SMM).
- Examples: DSGE models, network formation models and time series.

High-dimensional problems

In complex models, we often have some parameters of interest with a large vector of summary statistics.

- Curse of dimensionality
- Univariate/ Bivariate marginal ABC might neglect significant multivariate dependency with > 2 parameters
- Choosing summary statistics for a subset of parameters.
- Parameters identification and non-informative summary statistics

Can we improve on this likelihood-free inference method?

ABC-network method

We propose a method called 'ABC-network' to learn the relationship between parameters and summaries in ABC inference. We utilise some statistical concepts called the *Gaussian graphical model* and *graph Lasso*.

Contribution:

- Factorisation of high-dimensional models with complex dependency
- Learn the relationship between parameters and summary statistics
- Heuristic of selecting informative subset of summary statistics for each parameter/ groups of parameters.

Table of Contents

Inference problem

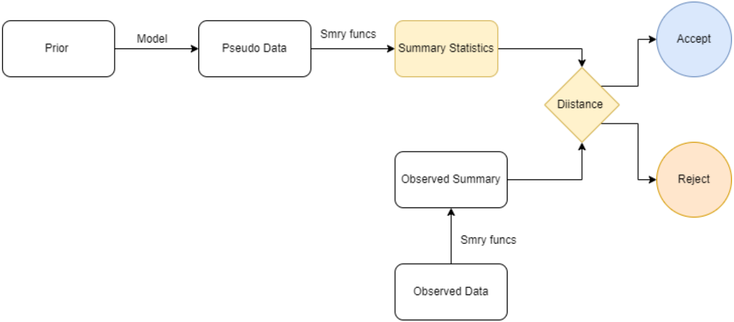
- We observe data y_{obs} and conduct inferences on a parametric model $\{p(.|\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\}$.
- The model is generative: we can simulate pseudo data y from $p(.|\theta)$. However, we cannot directly compute the likelihood of the observed data.
- The posterior distribution of parameters

$$\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta),$$

is intractable.

ABC method - procedure

Repeat the following process many times to generate samples $(\theta^{(i)}, s^{(i)}), i = 1, \cdots, N$.



The accepted samples will be an approximation of the posterior of interest $\pi(\theta|y_{obs})$.

ABC method - procedure

In addition to the model $p(.|\theta)$, other ingredients of ABC:

- Prior distribution of parameter $\pi(\theta)$
- Summary statistics function $s(.) : y \rightarrow \mathbb{R}^q$
- Distance threshold $\epsilon > 0$

ABC method - procedure

The ABC procedure can be described as follows.

1. Draw a parameter $\theta^{(i)}$ from the prior π and generate a pseudo data $y^{(i)} \sim p(.|\theta^{(i)})$.
2. Compute the distance $\rho(s_{obs}, s(y^{(i)}))$ and assign a weight $K_\epsilon\{\rho(s_{obs}, s(y^{(i)}))\}$ to the samples $(\theta^{(i)}, s^{(i)})$ via standard smoothing kernel function K_ϵ with with a scale parameter ϵ.

Repeat this procedure N times; the samples are drawn from the ABC posterior distribution

$$\pi_\epsilon^{ABC}(\theta, s|s_{obs}) \propto K_\epsilon\{\rho(s_{obs}, s)\}p(s|\theta)\pi(\theta).$$

Existing remedies for high-dimensional ABC problems

Researchers have considered various methods to deal with the curse of dimensionality (see Li, Nott, Fan, and Sisson (2017) and references therein):

- Regression - imposing a linear relationship between the summary statistics and parameters
- Marginalization - 1) estimate the joint ABC posterior and obtain the order copula 2) estimate univariate marginal using a subset of the summary statistics and joining the marginals by order copula
- Gaussian copula - estimating each parameter using a subset of summary statistics and joining the marginals by Gaussian copulas
- Conditional Gibbs - ABC-Gibbs sampling on each parameter (Clarté, Robert, Ryder, and Stoehr, 2021)

Problems and research gaps

There are two major limitations to these methods:

- The choice of summary statistics for each marginal often relies on prior knowledge or intuition.
- Most of the methods focus on univariate marginal estimation $\pi(\theta_1|S_1)$ and hence neglect possible multivariate dependencies.

Motivation example - twisted normal

Consider a p-dimensional multivariate normal model:

- $y = (y_1, \ldots, y_p)^T \sim N_p(\theta, \Sigma)$, where $p \geq 3$. Here, $\theta = (\theta_1, \ldots, \theta_p)$ is the parameter of interest, and $\Sigma = \mathrm{diag}(1, \ldots, 1)$.
- We 'twist' the prior distribution so that θ_1, θ_2, and θ_3 are dependent.

$$\pi(\theta) \propto \exp\left\{ -\frac{\theta_1^2}{200} - \frac{\left(\theta_2 - b\theta_1^2 + 100b\right)^2}{2} - \frac{\left(\theta_3 - c\theta_1^2 + 100c\right)^2}{2} - \sum_{j=4}^{p} \frac{\theta_j^2}{2} \right\}.$$

Twisted-normal (cont'd)

We compare the Kullback–Leibler divergence (KL divergence) between the ABC posterior $\pi^{ABC}(\theta_1, \theta_2, \theta_3|y)$ and the true posterior $\pi(\theta_1, \theta_2, \theta_3|y)$ with different values of p.

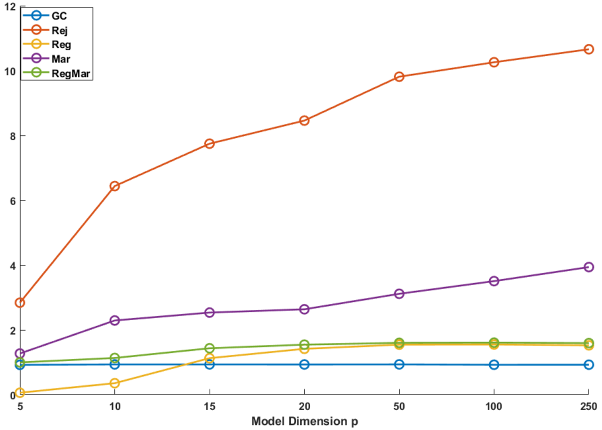


Table of Contents

ABC-Network method summary

The method consists of the following steps:

1. Implement standard ABC as a pilot search to obtain samples $(\boldsymbol{\theta}^{(i)}, \mathbf{S}^{(i)})$ around the observed summary statistics $\mathbf{S}_{obs}$.
2. Use the samples to estimate the inverse covariance matrix $\hat{\boldsymbol{\Sigma}}^{-1}$ using the graphical lasso algorithm.
3. Construct the ABC-Network from $\hat{\boldsymbol{\Sigma}}^{-1}$ and obtain clusters $(\boldsymbol{\theta}, \mathbf{S})_c$.
4. Finally, estimate the posterior for each cluster using ABC to obtain $\pi^{ABC}(\boldsymbol{\theta}_c|\mathbf{S}_c)$.

Step 1: Pilot search

We set up a relatively large tolerance threshold ϵ_0 for the ABC algorithm to generate samples $(\theta^{(i)}, s^{(i)}) \sim \pi_{\epsilon_0}^{ABC}$,

$$\pi_{\epsilon_0}^{ABC}(\theta, s|s_{obs}) \propto K_{\epsilon_0}\{\rho(s_{obs}, s)\}p(s|\theta)\pi(\theta),$$

- The restricted samples estimate the structure around s_{obs} rather than global relationship.
- Similar approach are considered in Fearnhead and Prangle (2012); Drovandi, Nott, and Frazier (2022)
- Typical choice of ϵ_0 10% or 5% quantile of the distance.

Step 2: Estimate the inverse covariance matrix

We consider a Gaussian Graphical Model (GGM) with parameters and summary statistics, so $V = \{\theta_1, \ldots, \theta_p, S_1, \cdots, S_q\}$ represents the nodes of the graph.
GGM is widely used to study the conditional dependency between variables, based on the inverse covariance matrix $\mathbf{\Sigma}^{-1}$.
Two variables i and j are independent conditional on other variables if and only if $\mathbf{\Sigma}^{-1}_{i,j} = 0$.

Step 2 (Cont): Graphical lasso estimation

We adopt the graphical lasso method by Friedman, Hastie, and Tibshirani (2008), which finds the inverse covariance matrix $\boldsymbol{\Sigma}^{-1}$ that maximizes

$$\log\det\boldsymbol{\Sigma}^{-1} - \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \lambda\sum_{i\neq j}\left|\boldsymbol{\Sigma}_{ij}^{-1}\right|,$$

where $\boldsymbol{\Sigma}^{-1}$ is the inverse covariance matrix, $\mathbf{S}$ is the sample covariance, and λ is the lasso penalty parameter.

Step 3: Obtain the cluster from the ABC-Network (Cont)

There are two types of nodes in the ABC-Network, thus there are three types of edges.

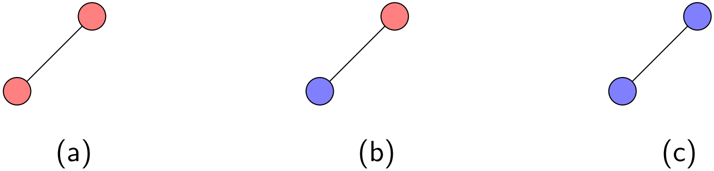


Figure: Nodes in red refer to summary statistics. Nodes in blue refer to the parameters.

The cluster is defined by parameters that are in the same components and summary statistics that are **directly** connected to the parameters.

Step 3: Obtain the cluster from the ABC-Network (Cont)

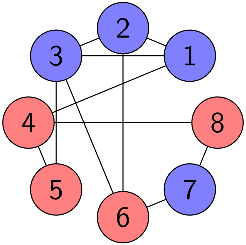
Definition

Given a ABC-Network $G = (V, E)$, a cluster is a subset of nodes $\{\boldsymbol{\theta}_c, \mathbf{S}_c\} \subseteq V$, where the parameters nodes $\boldsymbol{\theta}_c$ are nodes in the same network components:

- $\forall \theta_i \in \boldsymbol{\theta}_c, \exists \theta_j \in \boldsymbol{\theta}_c,$ *if* $(\theta_i, \theta_j) \in E$; *or*
- $\boldsymbol{\theta}_c = \theta_i$ *if* $\forall \theta \in \boldsymbol{\theta} \setminus \theta_i, (\theta_i, \theta_j) \notin E$

and the summary statistics nodes in one cluster are those **directly** *connected to parameter nodes in that cluster:*

$$\forall s \in \mathbf{S}_c, \exists \theta \in \boldsymbol{\theta}_c, \text{ such that } (s, \theta) \in E.$$



$$C_1 = \{(1, 2, 3), (4, 5, 6)\}$$

$$C_2 = \{(7), (6, 8)\}$$

Step 4: Estimate cluster-based posterior

For each cluster, we obtain $\pi^{ABC}(\boldsymbol{\theta_c}|\mathbf{S_c})$ from

$$\pi_{\epsilon}^{ABC}(\theta_c, s_c|s_{c,obs}) \propto K_\epsilon\{\rho(s_{c,obs}, s_c)\}p(s|\theta)\pi(\theta).$$

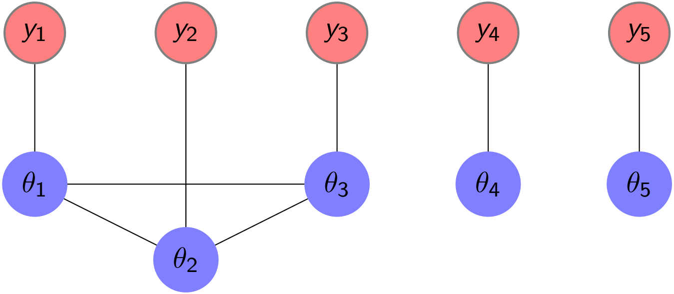
Note that we don't need to re-sample $(\theta^{(i)}, s^{(i)})$ and we can use the samples $(\theta^{(i)}, s^{(i)}) \sim \pi_{\epsilon_0}^{ABC}$ from step 1.
The final posterior distribution can be obtained via a copula $\mathcal{C}$

$$\pi^{ABC}(\boldsymbol{\theta}|\mathbf{S}) = \mathcal{C}[\pi^{ABC}(\boldsymbol{\theta}_1|\mathbf{S}_1), \cdots, \pi^{ABC}(\boldsymbol{\theta}_C|\mathbf{S}_C)] \times \prod_{c=1}^{C} \pi^{ABC}(\boldsymbol{\theta}_c|\mathbf{S}_c).$$

Table of Contents

Motivation example - revisit

We obtain the following ABC Network for the motivation example -



Our method successfully identifies the correct dependency structure among parameters and their corresponding informative summary statistics using the ABC-Network method.

Motivation example - revisit

KL Divergence between ABC Posterior and True Posterior $D_{KL}(\pi_{ABC}(\theta|y_{obs})||\pi(\theta|y_{obs}))$ over 50 replicates with standard errors given in parentheses.

p	5	10	15	20	50	100	250
Network	0.125	0.124	0.124	0.124	0.124	0.124	0.127
	(0.009)	(0.011)	(0.016)	(0.008)	(0.012)	(0.010)	(0.011)
GC	0.922	0.935	0.935	0.933	0.936	0.925	0.926
	(0.022)	(0.032)	(0.036)	(0.023)	(0.032)	(0.028)	(0.042)

The table shows the KL divergence values and standard errors for various values of p, indicating that the ABC network has a relatively low divergence even when p increases.

DSGE application

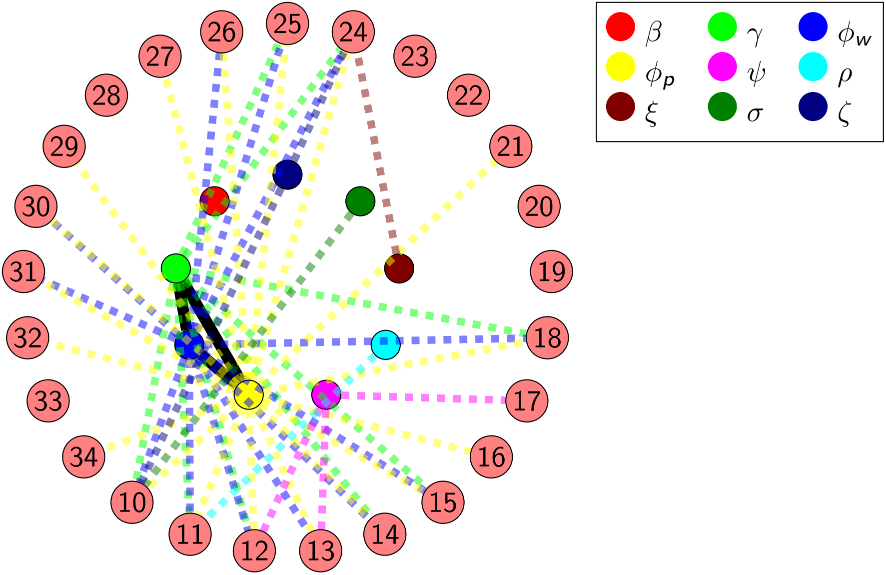
We apply our ABC-network method to a New Keynesian model with moderate dimensions of parameters and summary statistics and contrast its performance with the simulated method of moments (SMM) (Ruge-Murcia, 2012).

- Model introduced by Kim and Ruge-Murcia (2009) aims at investigating Tobin's proposition that positive inflation may be socially beneficial. The conclusion is that 0.35% inflation is optimal when there exits asymmetry wage adjustment cost.
- ABC and SMM are asymptotically equivalent under some high-level conditions (Forneron and Ng, 2018).

Model's details

- The key parameter of interest is wage asymmetry (ψ).
- Other parameters include the discount rate (β), consumption curvature (γ), adjustment cost parameters (ϕ_w, ϕ_p) and parameters of the shock process (ρ, ξ, σ, ζ).
- The authors utilize five data series: consumption, hours worked, price inflation, wage inflation, and the nominal interest rate. The moments to match include the variance, covariance, and first- and second-order autocovariance.

ABC Network


26
25
24
27
23
28
22
29
21
30
20
31
19
32
18
33
17
34
16
10
15
11
12
13
14
β
γ
ϕ_w
ϕ_p
ψ
ρ
ξ
σ
ζ

Summary statistics

Covariance Matrix	C	H	PI	WI	I
Consumption	10	11	12	13	14
Hours worked		15	16	17	18
Price inflation			19	20	21
Wage inflation				22	23
Interest rate					24

	First-order AutoCoV	Second-order AutoCoV
Consumption	25	30
Hours worked	26	31
Price inflation	27	32
Wage inflation	28	33
Interest rate	29	34

Table: Summary Nodes' Index for Parameter-Summary Network

Relationship between parameters and summary

Our method provides a heuristic to study the possible identification/informativeness between parameters and summary statistics.

- The wage asymmetry is mainly identified by the following three summaries: a) covariance of consumption and price inflation; b) covariance of consumption and wage inflation; and c) covariance of hours and wage inflation. - This is consistent with the identification study in the original paper.
- Some moments provide weak information in this identification problem
- Discount rate β are not well identified by the moments. It would be better if it is calibrated from the historical data directly.

Simulation results - optimal inflation

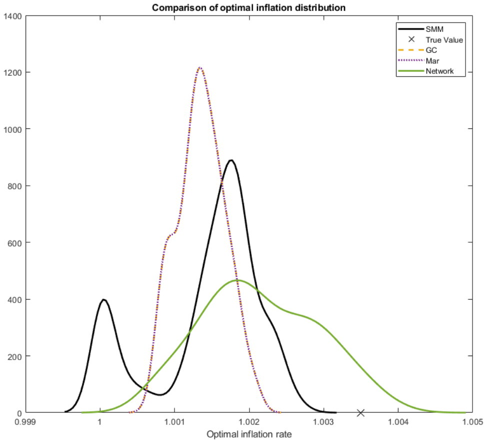
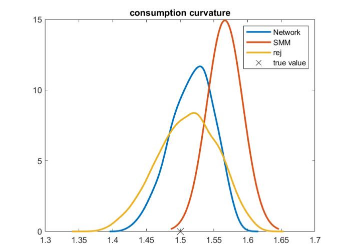
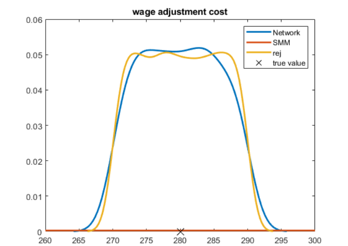


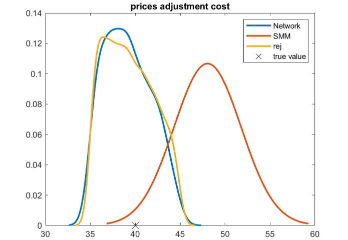
Figure: Distribution of optimal inflation with various inference methods.



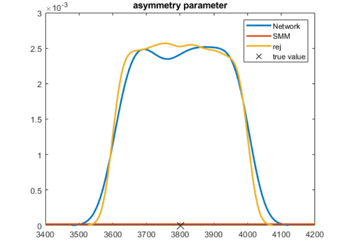
(a) Consumption Curvature



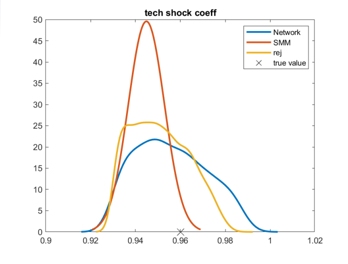
(b) Wage Adjustment Cost



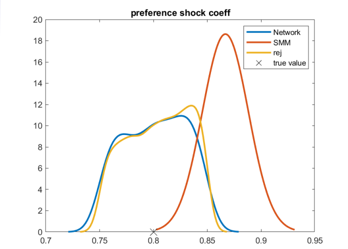
(c) Price Adjustment Cost



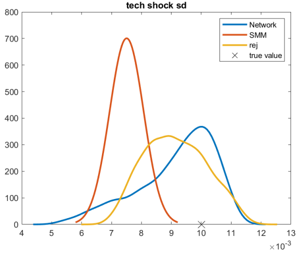
(d) Asymmetry Parameter



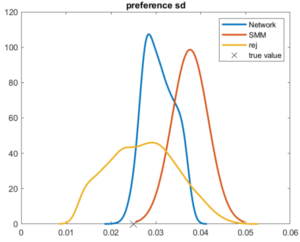
(a) Tech Shock Coeff



(b) Preferenece Coeff



(c) Tech Shock SD



(d) Preference SD

Empirical result - How much inflation is optimal?

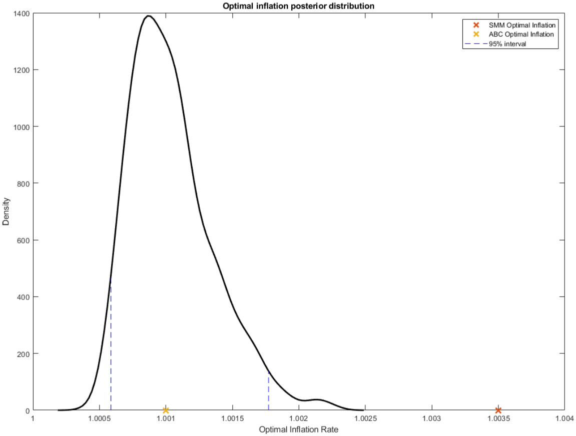


Figure: Optimal inflation rate

Table of Contents

Conclusion

ABC-Network method.

- The initial idea of this method was only to deal with high-dimensional ABC problems when there exists multivariate dependency.
- We also found the network provides extra information about the identification and selection of informative summary statistics.
- The application of the method to a DSGE model has demonstrated our contributions.
- Future extension: ABC-Gibbs, model cut, and sensitivity measure.

Thanks

- Questions?
- Email: yangqi.zhang@unsw.edu.au

References I

Clarté, G., C. P. Robert, R. J. Ryder, and J. Stoehr (2021). Componentwise approximate bayesian computation via gibbs-like steps. *Biometrika 108*(3), 591–607.

Drovandi, C., D. J. Nott, and D. T. Frazier (2022). Improving the accuracy of marginal approximations in likelihood-free inference via localisation. *arXiv preprint arXiv:2207.06655*.

Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate bayesian computation: Semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(3), 419–474.

Forneron, J.-J. and S. Ng (2018). The abc of simulation estimation with auxiliary statistics. *Journal of Econometrics 205*(1), 112–139.

References II

Friedman, J., T. Hastie, and R. Tibshirani (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics 9*(3), 432–441.

Kim, J. and F. J. Ruge-Murcia (2009). How much inflation is necessary to grease the wheels? *Journal of Monetary Economics 56*(3), 365–377.

Li, J., D. J. Nott, Y. Fan, and S. A. Sisson (2017). Extending approximate bayesian computation methods to high dimensions via a gaussian copula model. *Computational Statistics & Data Analysis 106*, 77–89.

Ruge-Murcia, F. (2012). Estimating nonlinear dsge models by the simulated method of moments: With an application to business cycles. *Journal of Economic Dynamics and Control 36*(6), 914–938.