Lecture 7. June 17 2019

Recap

The last lecture gave an overview methods for choosing summary statistics

Recap

The last lecture gave an overview methods for choosing summary statistics

Reminder: the (updated) slides, and handout versions, are available at `https://cancerdynamics.columbia.edu/content/summer-program`

Recap

The last lecture gave an overview methods for choosing summary statistics

Reminder: the (updated) slides, and handout versions, are available at `https://cancerdynamics.columbia.edu/content/summer-program`

Today, we will look at

- Indirect Inference (II), and ABC II
- A summary of the book
- A list of algorithms in case you need to do ABC in anger
- A list of available generic ABC code

Indirect Inference

Reference: Drovandi (2018) Chapter 7 in *Handbook of Approximate Bayesian Computation*, CRC Press

Went mainstream after Gourieroux, Montfort and Renault (1993) *J Appl Econometrics*, **8**, S85–S118

Indirect Inference

Reference: Drovandi (2018) Chapter 7 in *Handbook of Approximate Bayesian Computation*, CRC Press

Went mainstream after Gourieroux, Montfort and Renault (1993) *J Appl Econometrics*, **8**, S85–S118

Aim: estimate parameters of a generative model on basis of indirect (or auxiliary) summary of observed data

Indirect Inference

Reference: Drovandi (2018) Chapter 7 in *Handbook of Approximate Bayesian Computation*, CRC Press

Went mainstream after Gourieroux, Montfort and Renault (1993) *J Appl Econometrics*, **8**, S85–S118

Aim: estimate parameters of a generative model on basis of indirect (or auxiliary) summary of observed data

Special cases:

- Generalized method of moments (GMM) – Hansen (1982)
- Simulated method of moments (SMM) – McFadden (1989)

GMM

- Start with series of moment conditions – functions of model parameters and data – with expectation 0 at true parameter value

GMM

- Start with series of moment conditions – functions of model parameters and data – with expectation 0 at true parameter value

- Minimize norm of sample averages of moment conditions

GMM

- Start with series of moment conditions – functions of model parameters and data – with expectation 0 at true parameter value

- Minimize norm of sample averages of moment conditions

- GMM estimators are known to be consistent, asymptotically Normal and efficient (among class of estimators that just the information in the moment conditions)

GMM

- Start with series of moment conditions – functions of model parameters and data – with expectation 0 at true parameter value

- Minimize norm of sample averages of moment conditions

- GMM estimators are known to be consistent, asymptotically Normal and efficient (among class of estimators that just the information in the moment conditions)

We have a vector-valued function $g(\mathcal{D}, \theta)$ such that for true parameter θ_0,

$$m(\theta_0) := \mathbb{E}g(\mathcal{D}, \theta_0) = 0$$

and $m(\theta) \neq 0$ if $\theta \neq \theta_0$

GMM

Estimate $m(\theta)$ via

$$\hat{m}(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(\mathcal{D}_i, \theta)$$

for observed datasets $\mathcal{D}_i$

GMM

Estimate $m(\theta)$ via

$$\hat{m}(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(\mathcal{D}_i, \theta)$$

for observed datasets $\mathcal{D}_i$

Choose good θ by minimizing the norm

$$||\hat{m}||_W^2 = \hat{m}(\theta)^T \, W \, \hat{m}(\theta),$$

for a positive-definite weight matrix W, to get

$$\hat{\theta} = \arg\min_{\theta} \, \hat{m}(\theta)^T \, \hat{W} \, \hat{m}(\theta),$$

where $\hat{W}$ is an estimator of W.

GMM

Estimate $m(\theta)$ via

$$\hat{m}(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(\mathcal{D}_i, \theta)$$

for observed datasets $\mathcal{D}_i$

Choose good θ by minimizing the norm

$$||\hat{m}||_W^2 = \hat{m}(\theta)^T \, W \, \hat{m}(\theta),$$

for a positive-definite weight matrix W, to get

$$\hat{\theta} = \arg\min_{\theta} \, \hat{m}(\theta)^T \, \hat{W} \, \hat{m}(\theta),$$

where $\hat{W}$ is an estimator of W.

[Lots of work on choice of W]

SMM

Sometimes the function $m(\theta)$ cannot be calculated (e.g., because it involved lots of integration)

SMM

Sometimes the function $m(\theta)$ cannot be calculated (e.g., because it involved lots of integration)

McFadden proposed SMM to get around this, the first of the II methods

SMM

Sometimes the function $m(\theta)$ cannot be calculated (e.g., because it involved lots of integration)

McFadden proposed SMM to get around this, the first of the II methods

The idea is to simulate $m(\theta)$, and compare the simulated moments with those of the observed data

- This is GMM without closed form for $m(\theta)$
- Can be extremely computationally expensive compared to GMM

Indirect Inference

Indirect Inference

We focus on methods that use an auxiliary model, starting with II

- Assume a parametric auxiliary model with likelihood $p_A(\mathcal{D}|\phi)$, where ϕ is parameter of the model. This could be
  - a simplified version of generative model (which has parameter θ)
  - a data-analytical model designed to capture essential features of the data

Indirect Inference

We focus on methods that use an auxiliary model, starting with II

- Assume a parametric auxiliary model with likelihood $p_A(\mathcal{D}|\phi)$, where ϕ is parameter of the model. This could be
  - a simplified version of generative model (which has parameter θ)
  - a data-analytical model designed to capture essential features of the data
- What is relationship between θ and ϕ?
  - Write as $\phi(\theta)$, the *mapping* or *binding* function
  - If ϕ is known, and 1:1, then the II estimate is $\theta_{\text{obs}} = \theta(\phi_{\text{obs}})$, where $\theta()$ is the inverse function

- Thus we find θ that could have produced the auxiliary estimate ϕ_{obs}

- Thus we find θ that could have produced the auxiliary estimate ϕ_{obs}
- We do not know $\phi(\theta)$, but we can estimate from simulation:
  - define estimating function $Q(\mathcal{D}; \phi)$ for auxiliary model (e.g. log-likelihood function)
  - fit auxiliary model to data:

$$\phi_{\text{obs}} = \arg\max_{\phi} Q(\mathcal{D}_{\text{obs}}; \phi)$$

- Thus we find θ that could have produced the auxiliary estimate ϕ_{obs}
- We do not know $\phi(\theta)$, but we can estimate from simulation:
  - define estimating function $Q(\mathcal{D}; \phi)$ for auxiliary model (e.g. log-likelihood function)
  - fit auxiliary model to data:

$$\phi_{\text{obs}} = \arg\max_{\phi} Q(\mathcal{D}_{\text{obs}}; \phi)$$

Now the II step:

- for a particular θ value, simulate n iid datasets from generative model, to obtain $\mathcal{D}_1, \ldots, \mathcal{D}_n$. Each replicate set has same dimension as $\mathcal{D}_{\text{obs}}$
- Auxiliary model is fitted to simulated data to get an estimate of the binding function, $\phi_n(\theta)$

- That is,

$$\phi_n(\theta) = \arg\max_{\phi} Q((\mathcal{D}_1, \ldots, \mathcal{D}_n); \phi)$$

  where

$$Q((\mathcal{D}_1, \ldots, \mathcal{D}_n); \phi) = \sum_{i=1}^{n} Q(\mathcal{D}_i; \phi)$$

- The binding function is $\phi(\theta) = \lim_{n\to\infty} \phi_n(\theta)$

- That is,

$$\phi_n(\theta) = \arg\max_{\phi} Q((\mathcal{D}_1, \ldots, \mathcal{D}_n); \phi)$$

  where

$$Q((\mathcal{D}_1, \ldots, \mathcal{D}_n); \phi) = \sum_{i=1}^{n} Q(\mathcal{D}_i; \phi)$$

- The binding function is $\phi(\theta) = \lim_{n\to\infty} \phi_n(\theta)$

Here is an alternative view of the problem:

- estimated auxiliary parameter based on ith set is

$$\phi(\theta; \mathcal{D}_i) = \arg\max_{\phi} Q(\mathcal{D}_i; \phi)$$

- set $\phi_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \phi(\theta; \mathcal{D}_i)$
- With this definition, the binding function is

$$\mathbb{E}\,\phi(\theta, \mathcal{D}), \quad \mathcal{D} \sim f(\cdot|\theta)$$

The II procedure then involves solving an optimization problem to find the θ that generates a $\phi_n(\theta)$ closest to ϕ_{obs}:

$$\theta_{\text{obs},n} = \arg\min \theta \left\{ \left(\phi_n(\theta) - \phi_{\text{obs}}\right)^T W \left(\phi_n(\theta) - \phi_{\text{obs}}\right) \right\}, \tag{6}$$

where W is a positive definite weight matrix

Notes:

- This assumes that auxiliary estimates of $\phi_n(\theta)$ and ϕ_{obs} are unique

Notes:

- This assumes that auxiliary estimates of $\phi_n(\theta)$ and ϕ_{obs} are unique
- The II estimator should have lower variance with increasing n, but at greater computational cost

Notes:

- This assumes that auxiliary estimates of $\phi_n(\theta)$ and ϕ_{obs} are unique
- The II estimator should have lower variance with increasing n, but at greater computational cost
- Result depends on choice of W, which allows for efficient comparison when components of the auxiliary estimator have different variances, and when there is correlation among components of auxiliary estimator

Notes:

- This assumes that auxiliary estimates of $\phi_n(\theta)$ and ϕ_{obs} are unique
- The II estimator should have lower variance with increasing n, but at greater computational cost
- Result depends on choice of W, which allows for efficient comparison when components of the auxiliary estimator have different variances, and when there is correlation among components of auxiliary estimator
- One choice is $W = I$, another the observed information matrix $J(\phi_{\text{obs}})$, which approximate the inverse of the asymptotic variance of ϕ_{obs}

Alternative approaches

1. *Simulated quasi-maximum likelihood (SQML) estimator*

- Maximize the auxiliary estimating function using observed data, and estimated mapping

$$\theta_{\mathrm{obs},n} = \arg\max_{\theta} Q(\mathcal{D}_{\mathrm{obs}}; \phi_n(\theta)), \tag{7}$$

  for example using the log-likelihood of the auxiliary model as estimating function

- For optimal choice of W, the method in (6) is asymptotically more efficient than that in (7)

2. *Efficient method-of-moments* (Gallant and Tauchen, 1996)

This uses derivative of the estimating function of the auxiliary model:

$$S_A(\mathcal{D}, \phi) = \left(\frac{\partial Q(\mathcal{D}, \phi)}{\partial \phi_1}, \ldots, \frac{\partial Q(\mathcal{D}, \phi)}{\partial \phi_p} \right)^T,$$

where $p = \dim(\phi)$. Then

$$\theta_{\text{obs},n} = \arg\min_{\theta} \left\{ \left(\frac{1}{n} \sum_{i=1}^{n} S_A(\mathcal{D}_i, \phi_{\text{obs}}) \right)^T \sum \left(\frac{1}{n} \sum_{i=1}^{n} S_A(\mathcal{D}_i, \phi_{\text{obs}}) \right) \right\} \tag{8}$$

Note that $S_A(\mathcal{D}_{\text{obs}}, \phi_{\text{obs}})$ does not appear in (8), as it can be assumed to be 0 by definition

Notes:

- Called EMM when Q is the auxiliary log-likelihood
- Can be very efficient if can calculate S_A explicitly, as it only requires fitting model once to find ϕ_{obs} before doing the II optimization in (8)
- Could estimate derivatives, which can still be faster than continually fitting auxiliary model to simulated data
- For Σ, could choose $J(\phi_{\text{obs}})^{-1}$, since $J(\phi_{\text{obs}})$ can estimate variance of the score.

3. *A common estimating function*
is the auxiliary log-likelihood,

$$Q((\mathcal{D}_1, \ldots, \mathcal{D}_n), \phi) = \sum_{i=1}^{n} \log p_A(\mathcal{D}_i|\phi)$$

Can use others, such as sample moments. McFadden's method involves finding a generative parameter value that generates sample moments closest to observed sample moments.

3. *A common estimating function*
is the auxiliary log-likelihood,

$$Q((\mathcal{D}_1, \ldots, \mathcal{D}_n), \phi) = \sum_{i=1}^{n} \log p_A(\mathcal{D}_i|\phi)$$

Can use others, such as sample moments. McFadden's method involves finding a generative parameter value that generates sample moments closest to observed sample moments.

Notes:

- Auxiliary statistic should be sensitive to changes in θ, but robust to different independent, simulated datasets based on a fixed θ
- Heggland and Frigessi (2004) show that when the auxiliary statistic is sufficient for θ and has same dimension as θ, then the II estimator has the same asymptotic efficiency as the MLE, except for a multiplicative factor or $1 + 1/n$

ABC II methods

ABC II acronyms

TABLE 7.1

A List of Acronyms Together with Their Expansions for the Bayesian Likelihood-Free Methods Surveyed in this Chapter. A Description of Each Method is Shown Together with Some Key References

Acronym	Expansion	Description	Key References
ABC II	ABC indirect inference	ABC that uses an auxiliary model to form a summary statistic	Gleim and Pigorsch (2013); Drovandi et al. (2015)
ABC IP	ABC indirect parameter	ABC that uses the parameter estimate of an auxiliary model as a summary statistic	Drovandi et al. (2011); Drovandi et al. (2015)
ABC IL	ABC indirect likelihood	ABC that uses the likelihood of an auxiliary model to form a discrepancy function	Gleim and Pigorsch (2013); Drovandi et al. (2015)
ABC IS	ABC indirect score	ABC that uses the score of an auxiliary model to form a summary statistic	Gleim and Pigorsch (2013); Martin et al. (2014); Drovandi et al. (2015)
BIL	Bayesian indirect likelihood	A general approach that replaces an intractable likelihood with a tractable likelihood within a Bayesian algorithm	Drovandi et al. (2015)
pdBIL	Parametric BIL on the full data level	BIL method that uses the likelihood of a parametric auxiliary model on the full data level to replace the intractable likelihood	Reeves and Pettitt (2005); Gallant and McCulloch (2009); Drovandi et al. (2015)
psBIL	Parametric BIL on the summary statistic level	BIL method that uses the likelihood of a parametric auxiliary model on the summary statistic level to replace the intractable likelihood of the summary statistic	Drovandi et al. (2015)
ABC-cp	ABC composite parameter	ABC that uses the parameter of a composite likelihood to form a summary statistic	This chapter, but see Ruli et al. (2016) for a composite score approach
ABC-ec	ABC extremal coefficients	ABC approach of Erhardt and Smith (2012) for spatial extremes models	Erhardt and Smith (2012)

ABC IP

1. *ABC Indirect Parameter* – Drovandi et al. (2011)

- Uses parameter of auxiliary model as summary statistic
- Simulate $\mathcal{D} \sim f(\cdot|\theta)$, and fit auxiliary model to produce simulated summary statistics $S = \phi_{\mathcal{D}}$
- Compare this summary statistic to $S_{\text{obs}} = \phi_{\text{obs}}$, found by fitting the auxiliary model to the observed data
- The discrepancy is

$$||S - S_{\text{obs}}|| = \sqrt{(\phi_{\mathcal{D}} - \phi_{\text{obs}})^T J(\phi_{\text{obs}})(\phi_{\mathcal{D}} - \phi_{\text{obs}})} \tag{9}$$

- uses observed information matrix of auxiliary model evaluated at the observed summary statistic, $J(\phi_{\text{obs}})$

ABC IL

2. *ABC Indirect Likelihood* – Gleim and Pigorsch (2013)

- Here the discrepancy is

$$||S - S_{\text{obs}}|| = \log p_A(\mathcal{D}_{\text{obs}}|\phi_{\text{obs}}) - \log p_A(\mathcal{D}_{\text{obs}}|\phi_{\mathcal{D}})$$

Notes:

- Both use auxiliary parameter estimate as a summary statistic, so have to fit auxiliary model to every dataset simulated during the ABC algorithm

  This suggests that want as much information out of the auxiliary model as possible. So try

- MLE $\phi_{\text{obs}} = \arg\max_\phi p_A(\mathcal{D}_{\text{obs}}|\phi)$

- Remember the Nelder-Mead simplex method for derivative-free optimization when we do not have exact expression (e.g. *neldermead* in R)
- Can then have numerical issues to contend with; one suggestion is to start optimizer at ϕ_{obs}
- There is a trade-off between computational cost of obtaining and the statistical efficiency of chosen auxiliary estimator
- These are ABC version of classical II approaches

ABC IS

3. *ABC Indirect Score* – Gleim and Pigorsch (2013)

- Uses the score statistic of the auxiliary model as the summary statistic

ABC IS

3. *ABC Indirect Score* – Gleim and Pigorsch (2013)

- Uses the score statistic of the auxiliary model as the summary statistic

Notes:

- Always evaluate score at observed auxiliary statistic, ϕ_{obs}
- So any simulated $\mathcal{D}$ obtained during the ABC step can be substituted into auxiliary score to get simulated summary statistic, without fitting auxiliary model to simulated data
- If analytical expression for score is available, the summary statistic can be very fast to compute. Saves a lot over LI and IL
- If score not known analytically, need to estimate derivatives numerically. Adds overhead, but still likely less than that needed to determine MLE

When MLE is chosen as auxiliary estimator, the observed score (here, S_{obs}) is ≈ 0. So the natural discrepancy function is

$$||S - S_{\text{obs}}|| = \sqrt{S_A(\mathcal{D}, \phi_{\text{obs}})^T \, J(\phi_{\text{obs}}) \, S_A(\mathcal{D}, \phi_{\text{obs}})}$$

with $S = S_A(\mathcal{D}, \phi_{\text{obs}})$

- ABC IS is really ABC version of EMM

When does ABC II work?

Drovandi et al. (2015) *Stat Sci*, **30**, 72–95

- ABC IP: needs unique auxiliary parameter estimator, so that each simulated dataset gives a unique value of the ABC discrepancy
- ABC IL: needs unique MLE value
- ABC IS: needs unique score

When does ABC II work?

Drovandi et al. (2015) *Stat Sci*, **30**, 72–95

- ABC IP: needs unique auxiliary parameter estimator, so that each simulated dataset gives a unique value of the ABC discrepancy
- ABC IL: needs unique MLE value
- ABC IS: needs unique score

Which auxiliary model?

- Desirable for auxiliary model to give good fit to observed data, so that quantities derived from auxiliary model capture most of the information in the observed data
- Might be convenient to select auxiliary model as a simplified version of the generative model. The auxiliary parameter then has same interpretation as the generative parameter. Of course, simplified model might not provide good fit to observed data

- Well known that basing inferences on wrong model may result in biased parameter estimates. In Bayesian setting, mode and concentration of posterior may not be estimated correctly when using approximate model. II tries to correct this
- Martin et al. (2014) *arXiv 1409.8363* give conditions under which get Bayes consistency for ABC II. In (9), $\phi_{\text{obs}} \to \phi(\theta_0)$ where θ_0 is the true value, and $\phi_{\mathcal{D}} \to \phi(\theta)$ when $\mathcal{D} \sim f(\cdot|\theta)$, as $n \to \infty$. Thus ABC only keeps $\theta = \theta_0$ as $\epsilon \to 0$.

pdBIL and psBIL

Bayesian Indirect Likelihood with parametric auxiliary model – Reeves and Pettitt (2005), Gallant and McCullough (2009)

Idea: use likelihood of an auxiliary parametric model as a replacement for that of the generative model, assuming that $\phi(\theta)$ has been estimated between the parameter sets

pdBIL and psBIL

Bayesian Indirect Likelihood with parametric auxiliary model – Reeves and Pettitt (2005), Gallant and McCullough (2009)

Idea: use likelihood of an auxiliary parametric model as a replacement for that of the generative model, assuming that $\phi(\theta)$ has been estimated between the parameter sets

If the binding function is known, then approximate posterior of pdBIL is

$$f_A(\theta|\mathcal{D}_{\text{obs}}) \propto p_A(\mathcal{D}_{\text{obs}}|\phi(\theta))\pi(\theta),$$

where $f_A(\theta|\mathcal{D}_{\text{obs}})$ is the pdBIL approximation to the true posterior, and $p_A(\mathcal{D}_{\text{obs}}|\phi(\theta))$ is likelihood of auxiliary model evaluated at $\phi(\theta)$

pdBIL and psBIL

Bayesian Indirect Likelihood with parametric auxiliary model – Reeves and Pettitt (2005), Gallant and McCullough (2009)

Idea: use likelihood of an auxiliary parametric model as a replacement for that of the generative model, assuming that $\phi(\theta)$ has been estimated between the parameter sets

If the binding function is known, then approximate posterior of pdBIL is

$$f_A(\theta|\mathcal{D}_{\text{obs}}) \propto p_A(\mathcal{D}_{\text{obs}}|\phi(\theta))\pi(\theta),$$

where $f_A(\theta|\mathcal{D}_{\text{obs}})$ is the pdBIL approximation to the true posterior, and $p_A(\mathcal{D}_{\text{obs}}|\phi(\theta))$ is likelihood of auxiliary model evaluated at $\phi(\theta)$

The problem is that the relationship between ϕ and θ is not usually known. Could estimate via simulation of n iid datasets based on generative model with parameter θ.

Then auxiliary model is fitted to large dataset to get $\phi_n(\theta)$, for example by MLE:

$$\phi_n(\theta) = \arg\max_{\phi} \prod_{i=1}^{n} p_A(\mathcal{D}_i|\phi)$$

Then auxiliary model is fitted to large dataset to get $\phi_n(\theta)$, for example by MLE:

$$\phi_n(\theta) = \arg\max_{\phi} \prod_{i=1}^{n} p_A(\mathcal{D}_i|\phi)$$

The target of the resulting method is given by

$$f_{A,n}(\theta|\mathcal{D}_{\text{obs}}) \propto \pi(\theta) p_{A,n}(\mathcal{D}_{\text{obs}}|\theta), \tag{10}$$

where

$$p_{A,n}(\mathcal{D}_{\text{obs}}|\theta) = \int p_{A,n}(\mathcal{D}_{\text{obs}}|\phi_n(\theta)) \prod_{i=1}^{n} f(\mathcal{D}_i|\theta),$$

which can be estimated unbiasedly using a single draw of n iid datasets with distribution $f(\cdot|\theta)$

Then auxiliary model is fitted to large dataset to get $\phi_n(\theta)$, for example by MLE:

$$\phi_n(\theta) = \arg\max_{\phi} \prod_{i=1}^{n} p_A(\mathcal{D}_i|\phi)$$

The target of the resulting method is given by

$$f_{A,n}(\theta|\mathcal{D}_{\text{obs}}) \propto \pi(\theta) p_{A,n}(\mathcal{D}_{\text{obs}}|\theta), \tag{10}$$

where

$$p_{A,n}(\mathcal{D}_{\text{obs}}|\theta) = \int p_{A,n}(\mathcal{D}_{\text{obs}}|\phi_n(\theta)) \prod_{i=1}^{n} f(\mathcal{D}_i|\theta),$$

which can be estimated unbiasedly using a single draw of n iid datasets with distribution $f(\cdot|\theta)$
[The extra subscript n in (10) highlights the extra layer of approximation that is introduced in selection of a finite value of n to estimate the mapping]

Notes:

- Like ABC IP/IL, pdBIL requires optimization step for every simulated dataset. But can try preprocessing to get estimate $\hat{\phi}(\theta)$ that can be reused. Not clear how to do this for high dimensional θ
- To get good approximations, the auxiliary likelihood should be a useful replacement likelihood across the parameter space with non-negligible posterior support, and should reduce further in areas of low posterior support
- In contrast, ABC II requires that summary statistics from the auxiliary model are informative, and there is an efficient algorithm to match observed and simulated summary statistics
- Roughly speaking, pdBIL will target true posterior as $n \to \infty$ if generative model is nested within auxiliary model

- In this ideal scenario, ABC II methods will not be exact as $\epsilon \to 0$, since quantities drawn from auxiliary model will not produce a sufficient statistic.
  - It seems infeasible to find a tractable auxiliary model that contains an intractable model as a special case! (Perhaps in hierarchical model setting?)
- Parametric auxiliary model can also be applied as summary statistic level, e.g. when some data reduction has been done. This is psBIL.
- One such is to assume a multivariate Normal auxiliary model. The likelihood is from the multivariate normal distribution with mean and variance dependent on θ, and used as a replacement for intractable summary statistic likelihood. This is sometimes called the *synthetic likelihood* method.

Composite likelihood methods

Ruli et al. (2016) *Stats and Computing*, **26**, 679–692

- They propose using the score of a composite likelihood approximation of the full likelihood as a summary statistic for ABC
- Useful when likelihoods of certain subsets of $\mathcal{D}$ can be evaluated easily
- For a review of composite likelihood methods, see e.g. Varin et al. (2011) *Stat Sinica*, **21**, 5–42
- Composite likelihood method is not associated with a parametric auxiliary model, but rather is used as an approximation to the likelihood by simplifying dependence structure of generative model
- Could use the composite likelihood estimate as a summary statistic, and obtain approached similar to ABC IP/IL

Handbook of ABC

Handbook of ABC

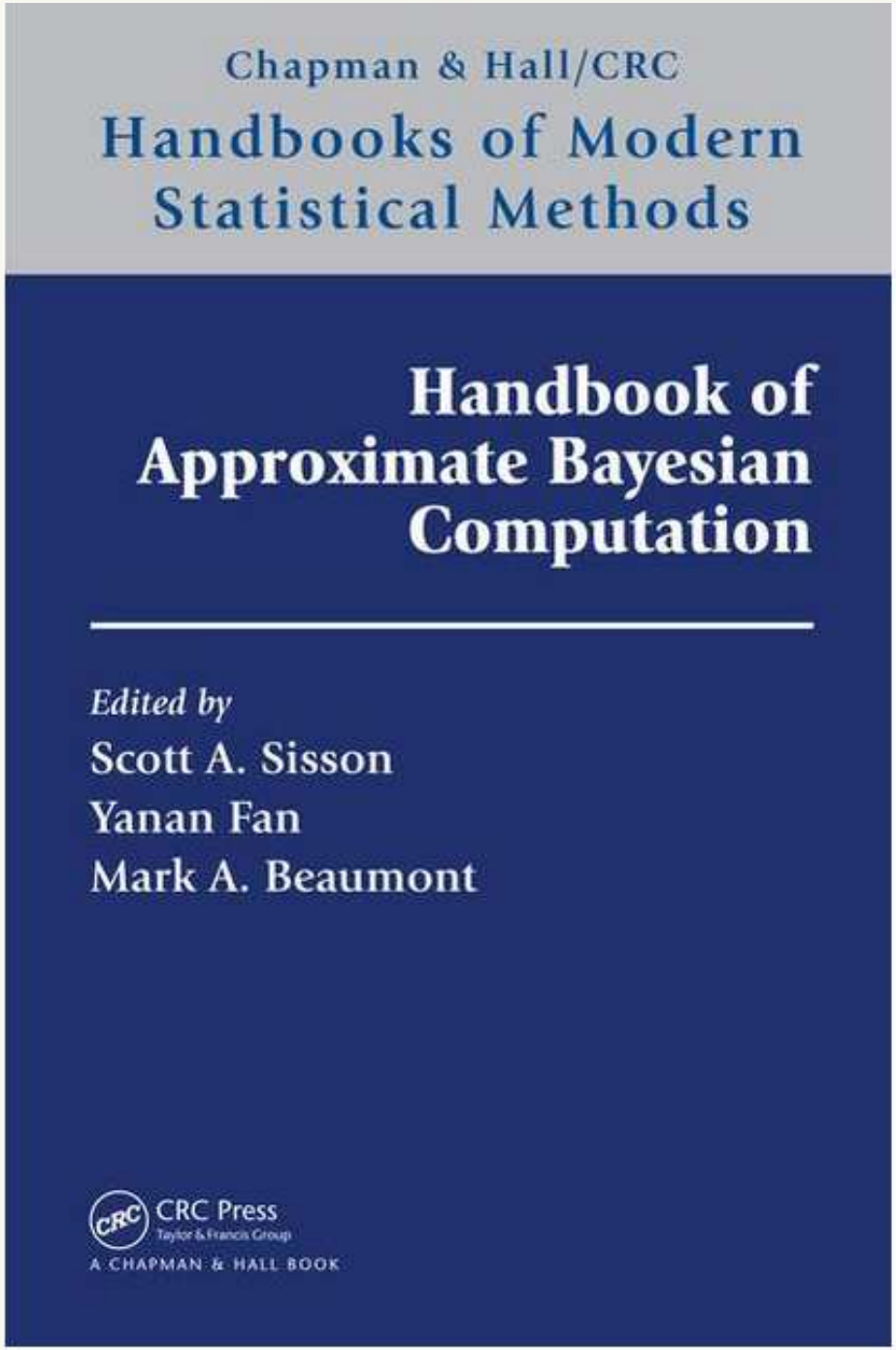

Table of Contents: Methods

Table of Contents

Table of Contents: Applications

ABC Samplers

ABC Samplers

Reference: Sisson and Fan (2018) Chapter 4 in *Handbook of Approximate Bayesian Computation*, CRC Press

As a handy reference, here are eight useful algorithms from the chapter

ABC Rejection Sampling

Algorithm 4.1: ABC Rejection Sampling

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- A proposal density $g(\theta)$, with $g(\theta) > 0$ if $\pi(\theta|y_{obs}) > 0$.
- An integer $N > 0$.
- A kernel function $K_h(u)$ and scale parameter $h > 0$.
- A low-dimensional vector of summary statistics $s = S(y)$.

Sampling:
For $i = 1, \ldots, N$:

1. Generate $\theta^{(i)} \sim g(\theta)$ from sampling density g.
2. Generate $y^{(i)} \sim p(y|\theta^{(i)})$ from the model.
3. Compute summary statistic $s^{(i)} = S(y^{(i)})$.
4. Accept $\theta^{(i)}$ with probability $\frac{K_h(\|s^{(i)} - s_{obs}\|)\pi(\theta^{(i)})}{Mg(\theta^{(i)})}$ where $M \geq K_h(0) \max_\theta \frac{\pi(\theta)}{g(\theta)}$.
   Else go to 1.

Output:
A set of parameter vectors $\theta^{(1)}, \ldots, \theta^{(N)} \sim \pi_{ABC}(\theta|s_{obs})$.

ABC Importance Sampling

Algorithm 4.2: ABC Importance Sampling

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$, consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- An importance sampling density $g(\theta)$, with $g(\theta) > 0$ if $\pi(\theta|y_{obs}) > 0$.
- An integer $N > 0$.
- A kernel function $K_h(u)$ and scale parameter $h > 0$.
- A low-dimensional vector of summary statistics $s = S(y)$.

Sampling:
For $i = 1, \ldots, N$:

1. Generate $\theta^{(i)} \sim g(\theta)$ from importance sampling density g.
2. Generate $y^{(i)} \sim p(y|\theta^{(i)})$ from the model.
3. Compute summary statistic $s^{(i)} = S(y^{(i)})$.
4. Compute weight $\tilde{w}^{(i)} = K_h(\|s^{(i)} - s_{obs}\|)\pi(\theta^{(i)})/g(\theta^{(i)})$.

Output:
A set of weighted parameter vectors $(\theta^{(1)}, \tilde{w}^{(1)}), \ldots, (\theta^{(N)}, \tilde{w}^{(N)}) \sim \pi_{ABC}(\theta|s_{obs})$.

ABC Importance/Rejection Sampling

Algorithm 4.3: ABC Importance/Rejection Sampling

Same as Algorithm 4.2, but replacing step 4 of Sampling with:

4. With probability $K_h(\|s^{(i)} - s_{obs}\|)/K_h(0)$ set $\tilde{w}^{(i)} = \pi(\theta^{(i)})/g(\theta^{(i)})$, else go to 1.

ABC Rejection Control Importance Sampling

Algorithm 4.4: ABC Rejection Control Importance Sampling

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- An importance sampling density $g(\theta)$, with $g(\theta) > 0$ if $\pi(\theta|y_{obs}) > 0$.
- An integer $N > 0$.
- A kernel function $K_h(u)$ and scale parameter $h > 0$.
- A low-dimensional vector of summary statistics $s = S(y)$.
- A rejection control threshold $c > 0$.

Sampling:
For $i = 1, \ldots, N$:

1. Generate $\theta^{(i)} \sim g(\theta)$ from importance sampling density g.
2. Generate $y^{(i)} \sim p(y|\theta^{(i)})$ from the model and compute summary statistic $s^{(i)} = S(y^{(i)})$.
3. Compute weight $\tilde{w}^{(i)} = K_h(\|s^{(i)} - s_{obs}\|)\pi(\theta^{(i)})/g(\theta^{(i)})$.
4. Reject $\theta^{(i)}$ with probability $1 - r^{(i)} = 1 - \min\{1, \frac{\tilde{w}^{(i)}}{c}\}$, and go to Step 1.
5. Otherwise, accept $\theta^{(i)}$ and set modified weight $\tilde{w}^{*(i)} = \tilde{w}^{(i)}/r^{(i)}$.

Output:
A set of weighted parameter vectors $(\theta^{(1)}, \tilde{w}^{*(1)}), \ldots, (\theta^{(N)}, \tilde{w}^{*(N)}) \sim \pi_{ABC}(\theta|s_{obs})$.

ABC $k$nn Importance Sampling

Algorithm 4.5: ABC *k*-nn Importance Sampling

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$, and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- An importance sampling density $g(\theta)$, with $g(\theta) > 0$ if $\pi(\theta|y_{obs}) > 0$.
- Integers $N' \gg N > 0$.
- A kernel function $K_h(u)$ with compact support.
- A low-dimensional vector of summary statistics $s = S(y)$.

Sampling:
For $i = 1, \ldots, N'$:

1. Generate $\theta^{(i)} \sim g(\theta)$ from importance sampling density g.
2. Generate $y^{(i)} \sim p(y|\theta^{(i)})$ from the model.
3. Compute summary statistic $s^{(i)} = S(y^{(i)})$.

- Identify the N-nearest neighbours of s_{obs} as measured by $\|s^{(i)} - s_{obs}\|$.
- Index these nearest neighbours by $[1], \ldots, [N]$.
- Set h to be the largest possible value, such that $K_h(\max_i\{\|s^{([i])} - s_{obs}\|\}) = 0$.
- Compute weights $\tilde{w}^{([i])} = K_h(\|s^{([i])} - s_{obs}\|)\pi(\theta^{([i])})/g(\theta^{([i])})$ for $i = 1, \ldots, N$.

Output:
A set of weighted parameter vectors $(\theta^{([1])}, \tilde{w}^{([1])}), \ldots, (\theta^{([N])}, \tilde{w}^{([N])}) \sim \pi_{ABC}(\theta|s_{obs})$.

ABC Markov Chain Monte Carlo

Algorithm 4.6: ABC Markov Chain Monte Carlo Algorithm

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$, and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- A Markov proposal density $g(\theta, \theta') = g(\theta'|\theta)$.
- An integer $N > 0$.
- A kernel function $K_h(u)$ and scale parameter $h > 0$.
- A low-dimensional vector of summary statistics $s = S(y)$.

Initialise:
Repeat:

1. Choose an initial parameter vector $\theta^{(0)}$ from the support of $\pi(\theta)$.
2. Generate $y^{(0)} \sim p(y|\theta^{(0)})$ from the model and compute summary statistics $s^{(0)} = S(y^{(0)})$,

until $K_h(\|s^{(0)} - s_{obs}\|) > 0$.

Sampling:
For $i = 1, \ldots, N$:

1. Generate candidate vector $\theta' \sim g(\theta^{(i-1)}, \theta)$ from the proposal density g.
2. Generate $y' \sim p(y|\theta')$ from the model and compute summary statistics $s' = S(y')$.
3. With probability:0

$$\min\left\{1, \frac{K_h(\|s' - s_{obs}\|)\pi(\theta')g(\theta', \theta^{(i-1)})}{K_h(\|s^{(i-1)} - s_{obs}\|)\pi(\theta^{(i-1)})g(\theta^{(i-1)}, \theta')}\right\}$$

set $(\theta^{(i)}, s^{(i)}) = (\theta', s')$. Otherwise set $(\theta^{(i)}, s^{(i)}) = (\theta^{(i-1)}, s^{(i-1)})$.

Output:
A set of correlated parameter vectors $\theta^{(1)}, \ldots, \theta^{(N)}$ from a Markov chain with stationary distribution $\pi_{ABC}(\theta|s_{obs})$.

ABC Sequential Rejection Control Importance Sampling

Algorithm 4.7: ABC Sequential Rejection Control Importance Sampling Algorithm

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$, consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- A kernel function $K_h(u)$ and a sequence of scale parameters $h_0 \geq h_1 \geq \ldots \geq h_M$.
- An initial sampling distribution $g(\theta)$, and a method of constructing subsequent sampling distributions $g_m(\theta)$, $m = 1, \ldots, M$.
- An integer $N > 0$.
- A sequence of rejection control thresholds values c_m, $m = 1, \ldots, M$.
- A low dimensional vector of summary statistics $s = S(y)$.

Initialise:
For $i = 1, \ldots, N$:

- Generate $\theta_0^{(i)} \sim g(\theta)$ from initial sampling distribution g.
- Generate $y_0^{(i)} \sim p(y|\theta_0^{(i)})$ and compute summary statistics $s_0^{(i)} = S(y_0^{(i)})$.
- Compute weights $w_0^{(i)} = K_{h_0}(\|s_0^{(i)} - s_{obs}\|)\pi(\theta_0^{(i)})/g(\theta_0^{(i)})$.

Sampling:
For $m = 1, \ldots, M$:

1. Construct sampling distribution $g_m(\theta)$.
2. For $i = 1, \ldots, N$:
   (a) Generate $\theta_m^{(i)} \sim g_m(\theta)$, $y_m^{(i)} \sim p(y|\theta_m^{(i)})$ and compute $s_m^{(i)} = S(y_m^{(i)})$.
   (b) Compute weight $w_m^{(i)} = K_{h_m}(\|s_m^{(i)} - s_{obs}\|)\pi(\theta_m^{(i)})/g_m(\theta_m^{(i)})$.
   (c) Reject $\theta_m^{(i)}$ with probability $1 - r_m^{(i)} = 1 - \min\{1, \frac{w_m^{(i)}}{c_m}\}$, and go to step 2a.
   (d) Otherwise, accept $\theta_m^{(i)}$ and set modified weight $w_m^{*(i)} = w_m^{(i)}/r_m^{(i)}$.

Output:
A set of weighted parameter vectors $(\theta_M^{(1)}, w_M^{*(1)}), \ldots, (\theta_M^{(N)}, w_M^{*(N)})$ drawn from $\pi_{ABC}(\theta|s_{obs}) \propto \int K_{h_M}(\|s - s_{obs}\|)p(s|\theta)\pi(\theta)ds$.

ABC Sequential Monte Carlo

Algorithm 4.8: ABC Sequential Monte Carlo Algorithm

Inputs:

- A target posterior density $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ consisting of a prior distribution $\pi(\theta)$ and a procedure for generating data under the model $p(y_{obs}|\theta)$.
- A kernel function $K_h(u)$, and an integer $N > 0$.
- An initial sampling density $g(\theta)$ and sequence of proposal densities $g_m(\theta, \theta')$, $m = 1, \ldots, M$.
- A value $\alpha \in [0, 1]$ to control the effective sample size.
- A low dimensional vector of summary statistics $s = S(y)$.

Initialise:
For $i = 1, \ldots, N$:

- Generate $\theta_0^{(i)} \sim g(\theta)$ from initial sampling distribution g.
- Generate $y_0^{(i)}(t) \sim p(y|\theta_0^{(i)})$ and compute summary statistics $s_0^{(i)}(t) = S(y_0^{(i)})$ for $t = 1, \ldots, T$.
- Compute weights $w_0^{(i)} = \pi(\theta_i^0)/g(\theta_0^{(i)})$, and set $m = 1$.

Sampling:

1. Reweight: Determine h_m such that $ESS(w_m^{(1)}, \ldots, w_m^{(N)}) = \alpha ESS(w_{m-1}^{(1)}, \ldots, w_{m-1}^{(N)})$ where

$$w_m^{(i)} = w_{m-1}^{(i)} \frac{\sum_{t=1}^{T} K_{h_m}(\|s_{m-1}^{(i)}(t) - s_{obs}\|)\pi(\theta_m^{(i)})}{\sum_{t=1}^{T} K_{h_{m-1}}(\|s_{m-1}^{(i)}(t) - s_{obs}\|)\pi(\theta_{m-1}^{(i)})},$$

   and then compute new particle weights and set $\theta_m^{(i)} = \theta_{m-1}^{(i)}$ and $s_m^{(i)}(t) = s_{m-1}^{(i)}(t)$ for $i = 1, \ldots, N$, and $t = 1, \ldots, T$.

2. Resample: If $ESS(w_m^{(1)}, \ldots, w_m^{(N)}) < E$ then resample N particles from the empirical distribution function $\{\theta_m^{(i)}, s_m^{(i)}(1), \ldots, s_m^{(i)}(T), W_m^{(i)}\}$ where $W_m^{(i)} = w_m^{(i)} / \sum_{j=1}^{N} w_m^{(j)}$ and set $w_m^{(i)} = 1/N$.

3. Move: For $i = 1, \ldots, N$: If $w_m^{(i)} > 0$:

   - Generate $\theta' \sim g_m(\theta_m^{(i)}, \theta)$, $y'(t) \sim p(y|\theta_m^{(i)})$ and compute $s'(t) = S(y'(t))$ for $t = 1, \ldots, T$.

   - Accept θ' with probability

$$\min\left\{1, \frac{\sum_{t=1}^{T} K_{h_m}(\|s'(t) - s_{obs}\|)\pi(\theta')g(\theta', \theta_m^{(i)})}{\sum_{t=1}^{T} K_{h_m}(\|s_m^{(i)}(t) - s_{obs}\|)\pi(\theta_m^{(i)})g(\theta_m^{(i)}, \theta')}\right\}$$

   and set $\theta_m^{(i)} = \theta'$, $s_m^{(i)}(t) = s'(t)$ for $t = 1, \ldots, T$.

4. Increment $m = m + 1$. If stopping rule is not satisfied, go to 1.

Output:
A set of weighted parameter vectors $(\theta_M^{(1)}, w_M^{(1)}), \ldots, (\theta_M^{(N)}, w_M^{(N)})$ drawn from $\pi_{ABC}(\theta|s_{obs}) \propto \int K_{h_M}(\|s - s_{obs}\|)p(s|\theta)\pi(\theta)ds$.

General Purpose ABC Software

ABC Samplers

Reference: Kousathanas et al. (2018) Chapter 13 in *Handbook of Approximate Bayesian Computation*, CRC Press

As another handy reference, here are some useful packages from the chapter

ABC Rejection Sampling

TABLE 13.1

General and Specific Purpose ABC Software

Software	Purpose	Reference
ABCtoolbox	General	Wegmann et al. (2010)
abc	General	Csillery et al. (2012)
ABC-EP	General	Barthelmé and Chopin (2011)
ABCdistrib	General	Beaumont et al. (2002)
ABCreg	General	Thornton (2009)
EasyABC	General	Jabot et al. (2013)
DIYABC	Population genetics	Cornuet et al. (2008)
msABC	Population genetics	Pavlidis et al. (2010)
ONeSAMP	Population genetics	Tallmon et al. (2008)
PopABC	Population genetics	Lopes et al. (2009)
2BAD	Population genetics	Bray et al. (2010)
ABCĄF	Population genetics	Foll et al. (2008)
msBayes	Phylogeography	Hickerson et al. (2007)
ABC-SysBio	Systems Biology	Liepe et al. (2010)
abc-sde	Stochastic differential equations	Picchini (2014)