Lecture 4. June 11 2019

Recap

Last time we looked at

- Regression adjustment
- MCMC-ABC
- SMC-ABC

Today we will look at an example of ABC-Gibbs, and model choice.

The first topic is motivated by a recent article of Clarté et al. (2019) Component-wise Approximate Bayesian Computation via Gibbs-like steps. *arXiv:1905.13599v1*

Estimating the divergence time of primates

Model choice

Introduction

Reference: Marin JM et al. (2018) Chapter 6 in *Handbook of Approximate Bayesian Computation*, CRC Press

- Model choice is a difficult issue for ABC

Introduction

Reference: Marin JM et al. (2018) Chapter 6 in *Handbook of Approximate Bayesian Computation*, CRC Press

- Model choice is a difficult issue for ABC
- Obvious scientific drawback for hypothesis testing

Introduction

Reference: Marin JM et al. (2018) Chapter 6 in *Handbook of Approximate Bayesian Computation*, CRC Press

- Model choice is a difficult issue for ABC
- Obvious scientific drawback for hypothesis testing

ABC model choice is a conceptual problem: wrong vector of summary statistics may produce inconsistent inference

Two issues:

- not easy to select good summary statistics

Introduction

Reference: Marin JM et al. (2018) Chapter 6 in *Handbook of Approximate Bayesian Computation*, CRC Press

- Model choice is a difficult issue for ABC
- Obvious scientific drawback for hypothesis testing

ABC model choice is a conceptual problem: wrong vector of summary statistics may produce inconsistent inference

Two issues:

- not easy to select good summary statistics
- even getting a set which give convergent Bayes factors may give poor approximation at practical level

Standard ABC model choice, given $\mathcal{D}, S(\mathcal{D})$

First example adds the model index as an extra parameter, $m \in \mathcal{M}$.

- For $i = 1, \ldots, n$ do
  - Generate m from $\pi(\mathcal{M})$

Standard ABC model choice, given $\mathcal{D}, S(\mathcal{D})$

First example adds the model index as an extra parameter, $m \in \mathcal{M}$.

- For $i = 1, \ldots, n$ do
  - Generate m from $\pi(\mathcal{M})$
  - Generate θ from prior $\pi_m(\theta)$

Standard ABC model choice, given $\mathcal{D}, S(\mathcal{D})$

First example adds the model index as an extra parameter, $m \in \mathcal{M}$.

- For $i = 1, \ldots, n$ do
  - Generate m from $\pi(\mathcal{M})$
  - Generate θ from prior $\pi_m(\theta)$
  - Generate $\mathcal{D}'$ from model $f_m(\mathcal{D}|\theta)$

Standard ABC model choice, given $\mathcal{D}, S(\mathcal{D})$

First example adds the model index as an extra parameter, $m \in \mathcal{M}$.

- For $i = 1, \ldots, n$ do
  - Generate m from $\pi(\mathcal{M})$
  - Generate θ from prior $\pi_m(\theta)$
  - Generate $\mathcal{D}'$ from model $f_m(\mathcal{D}|\theta)$
  - Set $m_i = m, \theta_i = \theta, s'_i = S(\mathcal{D}')$

Standard ABC model choice, given $\mathcal{D}, S(\mathcal{D})$

First example adds the model index as an extra parameter, $m \in \mathcal{M}$.

- For $i = 1, \ldots, n$ do
  - Generate m from $\pi(\mathcal{M})$
  - Generate θ from prior $\pi_m(\theta)$
  - Generate $\mathcal{D}'$ from model $f_m(\mathcal{D}|\theta)$
  - Set $m_i = m, \theta_i = \theta, s'_i = S(\mathcal{D}')$
- Return the values m_i with k smallest distances $\rho(s'_i, s)$

Standard ABC model choice, given $\mathcal{D}, S(\mathcal{D})$

First example adds the model index as an extra parameter, $m \in \mathcal{M}$.

- For $i = 1, \ldots, n$ do
  - Generate m from $\pi(\mathcal{M})$
  - Generate θ from prior $\pi_m(\theta)$
  - Generate $\mathcal{D}'$ from model $f_m(\mathcal{D}|\theta)$
  - Set $m_i = m, \theta_i = \theta, s'_i = S(\mathcal{D}')$
- Return the values m_i with k smallest distances $\rho(s'_i, s)$

The chosen model indices are a sample from $\pi(m|S)$

Local logistic regression model choice

Note that

$$\mathbb{P}(M = m|S = s) = \mathbb{E}(\mathbb{1}_{M=m}|S = s)$$

so we can treat the analysis as a regression problem: iid draws from law of (m, s), the response being indicator of whether simulation comes from model m, covariates being the summary statistics. For example,

- Generate N samples (m_i, s_i)

Local logistic regression model choice

Note that

$$\mathbb{P}(M = m | S = s) = \mathbb{E}(\mathbb{1}_{M=m} | S = s)$$

so we can treat the analysis as a regression problem: iid draws from law of (m, s), the response being indicator of whether simulation comes from model m, covariates being the summary statistics. For example,

- Generate N samples (m_i, s_i)
- Compute weights $w_i = K_h(s_i - s_0)$ where K is a kernel density and h the bandwidth estimated from the s_i

Local logistic regression model choice

Note that

$$\mathbb{P}(M = m|S = s) = \mathbb{E}(\mathbb{1}_{M=m}|S = s)$$

so we can treat the analysis as a regression problem: iid draws from law of (m, s), the response being indicator of whether simulation comes from model m, covariates being the summary statistics. For example,

- Generate N samples (m_i, s_i)
- Compute weights $w_i = K_h(s_i - s_0)$ where K is a kernel density and h the bandwidth estimated from the s_i
- Estimate probabilities $\mathbb{P}(M = m|s_0)$ using logistic link (e.g. in *vgam* in R) from the weighted data (m_i, s_i, w_i)

Local logistic regression model choice

Note that

$$\mathbb{P}(M = m|S = s) = \mathbb{E}(\mathbb{1}_{M=m}|S = s)$$

so we can treat the analysis as a regression problem: iid draws from law of (m, s), the response being indicator of whether simulation comes from model m, covariates being the summary statistics. For example,

- Generate N samples (m_i, s_i)
- Compute weights $w_i = K_h(s_i - s_0)$ where K is a kernel density and h the bandwidth estimated from the s_i
- Estimate probabilities $\mathbb{P}(M = m|s_0)$ using logistic link (e.g. in *vgam* in R) from the weighted data (m_i, s_i, w_i)

Warning: $\mathbb{P}(M = m|S = s_0)$ is a surrogate for $\mathbb{P}(M = m|\mathcal{D})$ ...

Bayes factors

ABC posterior probability estimators are (i) *imprecise* and (ii) *inconsistent*, so may not converge to a point mass on the true model.

Bayes factors

ABC posterior probability estimators are (i) *imprecise* and (ii) *inconsistent*, so may not converge to a point mass on the true model.

We want to estimate posterior ratios of the model probabilities, for example

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})}$$

Note that

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_2)} \tag{5}$$

The term

$$B_{12} := \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)}$$

is known as the *Bayes factor*.

Bayes factors

ABC posterior probability estimators are (i) *imprecise* and (ii) *inconsistent*, so may not converge to a point mass on the true model.

We want to estimate posterior ratios of the model probabilities, for example

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})}$$

Note that

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_2)} \tag{5}$$

The term

$$B_{12} := \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)}$$

is known as the *Bayes factor*.

- If we can estimate this ratio, then we can compute the posterior ratio on left of (5)

- If we can estimate this ratio, then we can compute the posterior ratio on left of (5)
- We can approximate B_{12} using the relative acceptance rate of the rejection method

But what happens with summary statistics?
It turns out that things can get complicated, because

$$B_{12} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)}$$

- If we can estimate this ratio, then we can compute the posterior ratio on left of (5)
- We can approximate B_{12} using the relative acceptance rate of the rejection method

But what happens with summary statistics?
It turns out that things can get complicated, because

$$
\begin{aligned}
B_{12} &= \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} \\
&= \frac{\mathbb{P}(\mathcal{D}|S(\mathcal{D}),\mathcal{M}_1)\,\mathbb{P}(S(\mathcal{D})|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|S(\mathcal{D}),\mathcal{M}_2)\,\mathbb{P}(S(\mathcal{D})|\mathcal{M}_2)}
\end{aligned}
$$

- If we can estimate this ratio, then we can compute the posterior ratio on left of (5)
- We can approximate B_{12} using the relative acceptance rate of the rejection method

But what happens with summary statistics?
It turns out that things can get complicated, because

$$\begin{aligned} B_{12} &= \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} \\ &= \frac{\mathbb{P}(\mathcal{D}|S(\mathcal{D}),\mathcal{M}_1)\,\mathbb{P}(S(\mathcal{D})|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|S(\mathcal{D}),\mathcal{M}_2)\,\mathbb{P}(S(\mathcal{D})|\mathcal{M}_2)} \\ &= \frac{\mathbb{P}(\mathcal{D}|S(\mathcal{D}),\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|S(\mathcal{D}),\mathcal{M}_2)}\,B_{12}^S \end{aligned}$$

Thus a summary statistic is sufficient for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ if, and only if,

$$\mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_1) = \mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_2)$$

Note that

- Sufficiency for $\mathcal{M}_1$ or $\mathcal{M}_2$ alone, or for both models, does not guarantee sufficiency for ranking the models

Thus a summary statistic is sufficient for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ if, and only if,

$$\mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_1) = \mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_2)$$

Note that

- Sufficiency for $\mathcal{M}_1$ or $\mathcal{M}_2$ alone, or for both models, does not guarantee sufficiency for ranking the models
- If the summary statistic is sufficient for a model $\mathcal{M}$ in which both $\mathcal{M}_1$ and $\mathcal{M}_2$ are both nested, then models can be ranked

Thus a summary statistic is sufficient for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ if, and only if,

$$\mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_1) = \mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_2)$$

Note that

- Sufficiency for $\mathcal{M}_1$ or $\mathcal{M}_2$ alone, or for both models, does not guarantee sufficiency for ranking the models
- If the summary statistic is sufficient for a model $\mathcal{M}$ in which both $\mathcal{M}_1$ and $\mathcal{M}_2$ are both nested, then models can be ranked
- What are we to do?

Thus a summary statistic is sufficient for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ if, and only if,

$$\mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_1) = \mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_2)$$

Note that

- Sufficiency for $\mathcal{M}_1$ or $\mathcal{M}_2$ alone, or for both models, does not guarantee sufficiency for ranking the models
- If the summary statistic is sufficient for a model $\mathcal{M}$ in which both $\mathcal{M}_1$ and $\mathcal{M}_2$ are both nested, then models can be ranked
- What are we to do?
- Choose maximum a posteriori model via machine learning (e.g. random forests)

Thus a summary statistic is sufficient for comparing $\mathcal{M}_1$ and $\mathcal{M}_2$ if, and only if,

$$\mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_1) = \mathbb{P}(\mathcal{D}|S(\mathcal{D}), \mathcal{M}_2)$$

Note that

- Sufficiency for $\mathcal{M}_1$ or $\mathcal{M}_2$ alone, or for both models, does not guarantee sufficiency for ranking the models
- If the summary statistic is sufficient for a model $\mathcal{M}$ in which both $\mathcal{M}_1$ and $\mathcal{M}_2$ are both nested, then models can be ranked
- What are we to do?
- Choose maximum a posteriori model via machine learning (e.g. random forests)

See Didelot et al (*Bayesian Analysis*, 2011), Robert et al (*PNAS*, 2011) and the Marin chapter (2018) for further details