

What we will do today

- Choosing summary statistics
- Model comparison and ABC
- MCMC
- ABC-MCMC
- An example: the divergence time of primates

49

Choosing summary statistics

There are now several methods for

- choosing sets of summary statistics (Fearnhead and Prangle *JRSSB*, 2012)
- choosing from a set of summary statistics (Joyce and Marjoram *SAGMB*, 2007)

For a comparative review of dimension-reduction methods in ABC, see Blum et al. (*Statistical Science*, 2013)

The package `abctools` in R implements these methods. See Prangle et al (*Aust N Z J Stat*, 2014) for further details

50

Model comparison – 1

How do we choose between different models?

Suppose we have p models that might have generated the data.

Assume a prior probability $\pi(\mathcal{M}_i)$ for model i ; $\sum_{i=1}^p \pi(\mathcal{M}_i) = 1$

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

$$\frac{\mathbb{P}(\mathcal{M}_1|\mathcal{D})}{\mathbb{P}(\mathcal{M}_2|\mathcal{D})} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)} \frac{\pi(\mathcal{M}_1)}{\pi(\mathcal{M}_2)} \quad (4)$$

51

Model comparison – 2

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 (4)

We can approximate B_{12} using the relative acceptance rate of the rejection method

52

Model comparison – 3

Here is a hierarchical implementation:

- Sample a model from the prior for models
- Given the model, sample parameters from the prior for that model
- Simulate data from that model and prior
- Accept the model and parameter if $\rho(\mathcal{D}, \mathcal{D}') \leq \epsilon$; repeat
- Keep track of the acceptance probability for each model

The ratio of acceptance rates is an estimate of B_{12}

53

Model comparison – 4

Next, we consider the effects of the summary statistic aspect.

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,

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

54

Model comparison – 5

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

See Didelot et al (*Bayesian Analysis*, 2011), Robert et al (*PNAS*, 2011) for further details

55

Markov Chain Monte Carlo methods

56

MCMC – 1

The idea is to construct an ergodic Markov chain that has $f(\theta|\mathcal{D})$ as its stationary distribution, in the case that normalising constants cannot be computed. Here is Hastings' (*Biometrika*, 1970) classic method:

1. Now at θ
2. Propose move to θ' according to $q(\theta \rightarrow \theta')$
3. Calculate the Hastings ratio

$$h = \min \left(1, \frac{\mathbb{P}(\mathcal{D} | \theta') \pi(\theta') q(\theta' \rightarrow \theta)}{\mathbb{P}(\mathcal{D} | \theta) \pi(\theta) q(\theta \rightarrow \theta')} \right)$$

4. Accept θ' with probability h , else return θ

57

MCMC – 2

There are more things to check:

- Is the chain ergodic?
- Does it mix well?
- Is the chain stationary?
- Burn in?
- Diagnostics of the run (no free lunches) – see coda package in R for example

58

MCMC – 3

MCMC in evolutionary genetics setting



- Small tweaks in the biology often translate into huge changes in algorithm
- Long development time
- All the usual problems with convergence
- *Almost all the effort goes into evaluation of likelihood*

59

ABC-MCMC – 1

Here is an ABC version (Marjoram et al, *PNAS*, 2003)

1. Now at θ
2. Propose a move to θ' according to $q(\theta \rightarrow \theta')$
3. Generate \mathcal{D}' using θ'
4. If $\mathcal{D}' = \mathcal{D}$, go to next step, else return θ
5. Calculate

$$h = h(\theta, \theta') = \min \left(1, \frac{\pi(\theta')q(\theta' \rightarrow \theta)}{\pi(\theta)q(\theta \rightarrow \theta')} \right)$$

6. Accept θ' with probability h , else return θ

60

ABC-MCMC – 2

Lemma: The stationary distribution of the chain is, indeed, $f(\theta|\mathcal{D})$.

Proof: In class ...

61

ABC-MCMC – 3

Here is the practical version, for data \mathcal{D} , summary statistics S

4' If $\rho(\mathcal{D}', \mathcal{D}) \leq \epsilon$, go to next step, otherwise return θ

4'' If $\rho(S', S) \leq \epsilon$, go to next step, otherwise return θ

for some suitable metric ρ and approximation level ϵ

Observations now from $f(\theta \mid \rho(\mathcal{D}', \mathcal{D}) \leq \epsilon)$ or $f(\theta \mid \rho(S', S) \leq \epsilon)$

62

Variations on a theme – 1

There have been many variants on the theme. For example, one might use multiple simulations from a given θ to get a better estimate of the likelihood. This is known as the pseudo-marginal method (Beaumont, *Genetics*, 2003; Tavaré et al. *PNAS*, 2003; Andrieu & Roberts *Ann Statist*, 2009). The idea is to simulate pairs of data points $(\theta, \hat{\mathbb{P}}(\mathcal{D}|\theta))$:

2' If at θ' simulate B values of \mathcal{D}' , and use these to estimate $\mathbb{P}(\mathcal{D}|\theta')$ via

$$\hat{\mathbb{P}}(\mathcal{D}|\theta') = \frac{1}{B} \sum_{j=1}^B \mathbb{1}(\mathcal{D}' = \mathcal{D})$$

3' If this is 0, stay at θ ; else

63

Variations on a theme – 2

4' Accept θ' and $\hat{\mathbb{P}}(\mathcal{D}|\theta')$ with probability

$$h = \min \left(1, \frac{\hat{\mathbb{P}}(\mathcal{D} | \theta') \pi(\theta') q(\theta' \rightarrow \theta)}{\hat{\mathbb{P}}(\mathcal{D} | \theta) \pi(\theta) q(\theta \rightarrow \theta')} \right)$$

else stay at θ .

64

Variations on a theme – 3

- Convergence an issue?
- These methods can often be started at stationarity, so no burn-in
- If the underlying probability model is complex, simulating data will often not lead to acceptance. Thus need update for parts of the probability model (data augmentation)
- There are versions with varying ϵ ; see Bortot et al (*JASA*, 2007) for example
- There are now many hybrid versions of these approaches (e.g. ABC-within-Gibbs)

65