Bayesian model choice as a classification problem

Jean-Michel Marin

University of Montpellier, CNRS
Alexander Grothendieck Montpellier Institute

22 May 2019

Thanks

Numerous colleagues participated to parts of this work

- Christian Robert (bayesian model selection and Approximate Bayesian Computation)
- Pierre Pudlo (Approximate bayesian Computation)
- Arnaud Estoup (population genetics applications)
- Judith, Louis, Mohammed, Natesh ...

Thanks

Numerous colleagues participated to parts of this work

- Christian Robert (bayesian model selection and Approximate Bayesian Computation)
- Pierre Pudlo (Approximate bayesian Computation)
- Arnaud Estoup (population genetics applications)
- Judith, Louis, Mohammed, Natesh ...

Thanks

Numerous colleagues participated to parts of this work

- Christian Robert (bayesian model selection and Approximate Bayesian Computation)
- Pierre Pudlo (Approximate bayesian Computation)
- Arnaud Estoup (population genetics applications)
- Judith, Louis, Mohammed, Natesh ...

Thanks

Numerous colleagues participated to parts of this work

- Christian Robert (bayesian model selection and Approximate Bayesian Computation)
- Pierre Pudlo (Approximate bayesian Computation)
- Arnaud Estoup (population genetics applications)
- Judith, Louis, Mohammed, Natesh ...

Thanks

Numerous colleagues participated to parts of this work

- Christian Robert (bayesian model selection and Approximate Bayesian Computation)
- Pierre Pudlo (Approximate bayesian Computation)
- Arnaud Estoup (population genetics applications)
- Judith, Louis, Mohammed, Natesh ...

Introduction

M Bayesian parametric models in competition

$$f_m(\mathbf{y}|\theta_m) \quad \pi_m(\theta_m) \quad m = 1, \ldots, M$$

Prior probabilities in the model space $\mathbb{P}(\mathscr{M} = m)$

Target: the model's posterior probabilities

$$\mathbb{P}(\mathscr{M} = m|\mathbf{y}) \propto \mathbb{P}(\mathscr{M} = m) \int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Introduction

M Bayesian parametric models in competition

$$f_m(\mathbf{y}|\theta_m) \quad \pi_m(\theta_m) \quad m = 1, \ldots, M$$

Prior probabilities in the model space $\quad \mathbb{P}(\mathscr{M} = m)$

Target: the model's posterior probabilities

$$\mathbb{P}(\mathscr{M} = m|\mathbf{y}) \propto \mathbb{P}(\mathscr{M} = m) \int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Introduction

M Bayesian parametric models in competition

$$f_m(\mathbf{y}|\theta_m) \quad \pi_m(\theta_m) \quad m = 1, \ldots, M$$

Prior probabilities in the model space $\quad \mathbb{P}(\mathscr{M} = m)$

Target: the model's posterior probabilities

$$\mathbb{P}(\mathscr{M} = m|\mathbf{y}) \propto \mathbb{P}(\mathscr{M} = m) \int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Introduction

M Bayesian parametric models in competition

$$f_m(\mathbf{y}|\theta_m) \quad \pi_m(\theta_m) \quad m = 1, \ldots, M$$

Prior probabilities in the model space $\quad \mathbb{P}(\mathscr{M} = m)$

Target: the model's posterior probabilities

$$\mathbb{P}(\mathscr{M} = m|\mathbf{y}) \propto \mathbb{P}(\mathscr{M} = m) \int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Introduction

A key quantity the marginal likelihood (the evidence)

$$\int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Bayesian inference embodies Occam's razor

A simple model, like Model 0, makes only a limited range of predictions; a more powerful model, like Model 1, is able to predict a greater variety of data sets

If the data set falls in region R, the less powerful model will be the more probable model

Introduction

A key quantity the marginal likelihood (the evidence)

$$\int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Bayesian inference embodies Occam's razor

A simple model, like Model 0, makes only a limited range of predictions; a more powerful model, like Model 1, is able to predict a greater variety of data sets

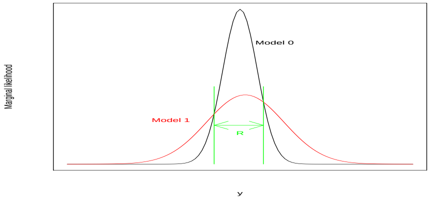
If the data set falls in region R, the less powerful model will be the more probable model

Introduction

A key quantity the marginal likelihood (the evidence)

$$\int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Bayesian inference embodies Occam's razor



A simple model, like Model 0, makes only a limited range of predictions; a more powerful model, like Model 1, is able to predict a greater variety of data sets

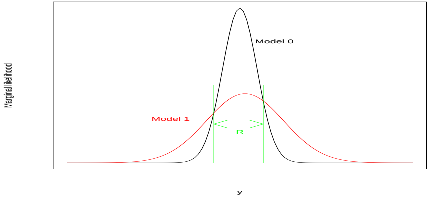
If the data set falls in region R, the less powerful model will be the more probable model

Introduction

A key quantity the marginal likelihood (the evidence)

$$\int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Bayesian inference embodies Occam's razor



A simple model, like Model 0, makes only a limited range of predictions; a more powerful model, like Model 1, is able to predict a greater variety of data sets

If the data set falls in region R, the less powerful model will be the more probable model

Introduction

The marginal likelihood corresponds to a **penalized** likelihood

The BIC information criterium comes from an asymptotic Laplace approximation of the marginal likelihood

Drton and Plummer (2017) Very nice extensions for singular model selection problems

Bayes factor for models M_1 and M_0

$$B_{10} = \frac{\int f_1(\mathbf{y}|\theta_1)\pi_1(\theta_1)d\theta_1}{\int f_0(\mathbf{y}|\theta_0)\pi_0(\theta_0)d\theta_0}$$

Introduction

The marginal likelihood corresponds to a **penalized** likelihood

The BIC information criterium comes from an asymptotic Laplace approximation of the marginal likelihood

Drton and Plummer (2017) Very nice extensions for singular model selection problems

Bayes factor for models M_1 and M_0

$$B_{10} = \frac{\int f_1(\mathbf{y}|\theta_1)\pi_1(\theta_1)d\theta_1}{\int f_0(\mathbf{y}|\theta_0)\pi_0(\theta_0)d\theta_0}$$

Introduction

The marginal likelihood corresponds to a **penalized** likelihood

The BIC information criterium comes from an asymptotic Laplace approximation of the marginal likelihood

Drton and Plummer (2017) Very nice extensions for singular model selection problems

Bayes factor for models M_1 and M_0

$$B_{10} = \frac{\int f_1(\mathbf{y}|\theta_1)\pi_1(\theta_1)d\theta_1}{\int f_0(\mathbf{y}|\theta_0)\pi_0(\theta_0)d\theta_0}$$

Introduction

The marginal likelihood corresponds to a **penalized** likelihood

The BIC information criterium comes from an asymptotic Laplace approximation of the marginal likelihood

Drton and Plummer (2017) Very nice extensions for singular model selection problems

Bayes factor for models $\mathfrak{M}_1$ and $\mathfrak{M}_0$

$$B_{10} = \frac{\int f_1(\mathbf{y}|\theta_1)\pi_1(\theta_1)d\theta_1}{\int f_0(\mathbf{y}|\theta_0)\pi_0(\theta_0)d\theta_0}$$

Difficulties with the Bayesian model choice paradigm

Prior difficulties

- When we have prior informations, how to choose the prior distributions on the parameters of each model in a compatible way?
- When we do not have any prior information, we can not use easily improper prior distributions
- What about the prior distribution on the models's space

We do not address these crucial questions in this talk

Difficulties with the Bayesian model choice paradigm

Prior difficulties

- When we have prior informations, how to choose the prior distributions on the parameters of each model in a compatible way?
- When we do not have any prior information, we can not use easily improper prior distributions
- What about the prior distribution in the models's space?

We do not address these crucial questions in this talk

Difficulties with the Bayesian model choice paradigm

Prior difficulties

- When we have prior informations, how to choose the prior distributions on the parameters of each model in a compatible way?
- When we do not have any prior information, we can not use easily improper prior distributions
- What about the prior distribution in the models's space?

We do not address these crucial questions in this talk

Introduction

Difficulties with the Bayesian model choice paradigm

Prior difficulties

- When we have prior informations, how to choose the prior distributions on the parameters of each model in a compatible way?
- When we do not have any prior information, we can not use easily improper prior distributions
- What about the prior distribution in the models's space?

We do not address these crucial questions in this talk

Difficulties with the Bayesian model choice paradigm

Prior difficulties

- When we have prior informations, how to choose the prior distributions on the parameters of each model in a compatible way?
- When we do not have any prior information, we can not use easily improper prior distributions
- What about the prior distribution in the models's space?

We do not address these crucial questions in this talk

Difficulties with the Bayesian model choice paradigm

Prior difficulties

- When we have prior informations, how to choose the prior distributions on the parameters of each model in a compatible way?
- When we do not have any prior information, we can not use easily improper prior distributions
- What about the prior distribution in the models's space?

We do not address these crucial questions in this talk

Introduction

Computational difficulties

- How to approximate the marginal likelihoods?
- When the number of models in consideration is huge, how to explore the models's space?

We consider the case of a limited number of models and not address trans-dimensional sampling solutions, like the reversible jump algorithm

Introduction

Computational difficulties

- How to approximate the marginal likelihoods?
- When the number of models in consideration is huge, how to explore the models's space?

We consider the case of a limited number of models and not address trans-dimensional sampling solutions, like the reversible jump algorithm

Introduction

Computational difficulties

- How to approximate the marginal likelihoods?
- When the number of models in consideration is huge, how to explore the models's space?

We consider the case of a limited number of models and not address trans-dimensional sampling solutions, like the reversible jump algorithm

Introduction

Computational difficulties

- How to approximate the marginal likelihoods?
- When the number of models in consideration is huge, how to explore the models's space?

We consider the case of a limited number of models and not address trans-dimensional sampling solutions, like the reversible jump algorithm

Introduction

We concentrate on the crucial question: how to approximate the marginal likelihood or find the model that maximises it

Two cases: the calculating of the likelihood is tractable or not

Goal of this talk: show you that, in each case, these problems can be re-written as classification problems and that the associated estimation methods can be very effective

Introduction

We concentrate on the crucial question: how to approximate the marginal likelihood or find the model that maximises it

Two cases: the calculating of the likelihood is tractable or not

Goal of this talk: show you that, in each case, these problems can be re-written as classification problems and that the associated estimation methods can be very effective

Introduction

We concentrate on the crucial question: how to approximate the marginal likelihood or find the model that maximises it

Two cases: the calculating of the likelihood is tractable or not

Goal of this talk: show you that, in each case, these problems can be re-written as classification problems and that the associated estimation methods can be very effective

Introduction

- Tractable likelihood: use of a logistic regression to estimate the marginal likelihood
- Intractable likelihood: use of Approximate Bayesian Model choice using random forests

Introduction

- Tractable likelihood: use of a logistic regression to estimate the marginal likelihood
- Intractable likelihood: use of Approximate Bayesian Model choice using random forests

Tractable likelihood
Standard Monte Carlo approximation

The standard Monte Carlo approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_\pi\left[f(\mathbf{y}|\theta)\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot)$

When the prior is far from the posterior, very high variance

Tractable likelihood
Standard Monte Carlo approximation

The standard Monte Carlo approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_\pi\left[f(\mathbf{y}|\theta)\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot)$

When the prior is far from the posterior, very high variance

Tractable likelihood
Standard Monte Carlo approximation

The standard Monte Carlo approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_\pi\left[f(\mathbf{y}|\theta)\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot)$

When the prior is far from the posterior, very high variance

Tractable likelihood
Standard Monte Carlo approximation

The standard Monte Carlo approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_\pi\left[f(\mathbf{y}|\theta)\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot)$

When the prior is far from the posterior, very high variance

Tractable likelihood
Importance sampling approximation

Let $g(\cdot)$ be a distribution such that $g(\theta) > 0$ when $f(\mathbf{y}|\theta)\pi(\theta) > 0$

The importance sampling approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)\,d\theta = \mathbb{E}_g\left[f(\mathbf{y}|\theta)\frac{\pi(\theta)}{g(\theta)}\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)\frac{\pi(\theta^i)}{g(\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $g(\cdot)$

Problem specific and curse of dimensionality

Tractable likelihood
Importance sampling approximation

Let $g(\cdot)$ be a distribution such that $g(\theta) > 0$ when $f(\mathbf{y}|\theta)\pi(\theta) > 0$

The importance sampling approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_g\left[f(\mathbf{y}|\theta)\frac{\pi(\theta)}{g(\theta)}\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)\frac{\pi(\theta^i)}{g(\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $g(\cdot)$

Problem specific and curse of dimensionality

Tractable likelihood
Importance sampling approximation

Let $g(\cdot)$ be a distribution such that $g(\theta) > 0$ when $f(\mathbf{y}|\theta)\pi(\theta) > 0$

The importance sampling approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_g\left[f(\mathbf{y}|\theta)\frac{\pi(\theta)}{g(\theta)}\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)\frac{\pi(\theta^i)}{g(\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $g(\cdot)$

Problem specific and curse of dimensionality

Tractable likelihood
Importance sampling approximation

Let $g(\cdot)$ be a distribution such that $g(\theta) > 0$ when $f(\mathbf{y}|\theta)\pi(\theta) > 0$

The importance sampling approximation of

$$m(\mathbf{y}) = \int f(\mathbf{y}|\theta)\pi(\theta)d\theta = \mathbb{E}_g\left[f(\mathbf{y}|\theta)\frac{\pi(\theta)}{g(\theta)}\right]$$

is given by

$$\frac{1}{N}\sum_{i=1}^{N} f(\mathbf{y}|\theta^i)\frac{\pi(\theta^i)}{g(\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $g(\cdot)$

Problem specific and curse of dimensionality

Tractable likelihood
The harmonic mean estimator

Let $\varphi(\cdot)$ be a distribution such that $\varphi(\theta) = 0$ when $\pi(\theta) f(\mathbf{y}|\theta) = 0$

$$\mathbb{E}_\pi \left[\frac{\varphi(\theta)}{\pi(\theta) f(\mathbf{y}|\theta)} \middle| \mathbf{y} \right] = \int \frac{\varphi(\theta)}{\pi(\theta) f(\mathbf{y}|\theta)} \frac{\pi(\theta) f(\mathbf{y}|\theta)}{m(\mathbf{y})} \, d\theta = \frac{1}{m(\mathbf{y})}$$

The harmonic mean approximation Newton and Raftery (1994) of $m(\mathbf{y})$ is given by

$$1 \Bigg/ N^{-1} \sum_{i=1}^{N} \frac{\varphi(\theta^i)}{\pi(\theta^i) f(\mathbf{y}|\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$

Tractable likelihood
The harmonic mean estimator

Let $\varphi(\cdot)$ be a distribution such that $\varphi(\theta) = 0$ when $\pi(\theta)f(\mathbf{y}|\theta) = 0$

$$\mathbb{E}_\pi\left[\frac{\varphi(\theta)}{\pi(\theta)f(\mathbf{y}|\theta)}\middle|\mathbf{y}\right] = \int \frac{\varphi(\theta)}{\pi(\theta)f(\mathbf{y}|\theta)}\frac{\pi(\theta)f(\mathbf{y}|\theta)}{m(\mathbf{y})}\,d\theta = \frac{1}{m(\mathbf{y})}$$

The harmonic mean approximation Newton and Raftery (1994) of $m(\mathbf{y})$ is given by

$$1 \Big/ N^{-1}\sum_{i=1}^{N}\frac{\varphi(\theta^i)}{\pi(\theta^i)f(\mathbf{y}|\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$

Tractable likelihood
The harmonic mean estimator

Let $\varphi(\cdot)$ be a distribution such that $\varphi(\theta) = 0$ when $\pi(\theta)f(\mathbf{y}|\theta) = 0$

$$\mathbb{E}_\pi\left[\left.\frac{\varphi(\theta)}{\pi(\theta)f(\mathbf{y}|\theta)}\right|\mathbf{y}\right] = \int \frac{\varphi(\theta)}{\pi(\theta)f(\mathbf{y}|\theta)}\frac{\pi(\theta)f(\mathbf{y}|\theta)}{m(\mathbf{y})}\,d\theta = \frac{1}{m(\mathbf{y})}$$

The harmonic mean approximation Newton and Raftery (1994) of $m(\mathbf{y})$ is given by

$$1 \Big/ N^{-1}\sum_{i=1}^{N}\frac{\varphi(\theta^i)}{\pi(\theta^i)f(\mathbf{y}|\theta^i)}$$

where $\theta^1,\ldots,\theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$

Tractable likelihood
The harmonic mean estimator

Let $\varphi(\cdot)$ be a distribution such that $\varphi(\theta) = 0$ when $\pi(\theta) f(\mathbf{y}|\theta) = 0$

$$\mathbb{E}_\pi \left[\left. \frac{\varphi(\theta)}{\pi(\theta) f(\mathbf{y}|\theta)} \right| \mathbf{y} \right] = \int \frac{\varphi(\theta)}{\pi(\theta) f(\mathbf{y}|\theta)} \frac{\pi(\theta) f(\mathbf{y}|\theta)}{m(\mathbf{y})} \, d\theta = \frac{1}{m(\mathbf{y})}$$

The harmonic mean approximation **Newton and Raftery (1994)** of $m(\mathbf{y})$ is given by

$$1 \Bigg/ N^{-1} \sum_{i=1}^{N} \frac{\varphi(\theta^i)}{\pi(\theta^i) f(\mathbf{y}|\theta^i)}$$

where $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$

Tractable likelihood
The harmonic mean estimator

As opposed to usual importance sampling constraints, the density $\varphi(\theta)$ must have lighter—rather than fatter—tails than $\pi(\theta)f(\mathbf{y}|\theta)$ for the approximation of the marginal likelihood to enjoy finite variance

Using $\varphi(\theta) = \pi(\theta)$ as in the original harmonic mean approximation will most usually result in an infinite variance estimator

Very high variance

Tractable likelihood
The harmonic mean estimator

As opposed to usual importance sampling constraints, the density $\varphi(\theta)$ must have lighter—rather than fatter—tails than $\pi(\theta)f(\mathbf{y}|\theta)$ for the approximation of the marginal likelihood to enjoy finite variance

Using $\varphi(\theta) = \pi(\theta)$ as in the original harmonic mean approximation will most usually result in an infinite variance estimator

Very high variance

Tractable likelihood
The harmonic mean estimator

As opposed to usual importance sampling constraints, the density $\varphi(\theta)$ must have lighter—rather than fatter—tails than $\pi(\theta)f(\mathbf{y}|\theta)$ for the approximation of the marginal likelihood to enjoy finite variance

Using $\varphi(\theta) = \pi(\theta)$ as in the original harmonic mean approximation will most usually result in an infinite variance estimator

Very high variance

Tractable likelihood
Chib's solution

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\,\pi(\theta)}{\pi(\theta|\mathbf{y})}\,, \forall\theta$$

For an arbitrary value θ^* of θ, the Chib's Chib (1995) approximation to the marginal likelihood is

$$\hat{m}(\mathbf{y}) = \frac{f(\mathbf{y}|\theta^*)\,\pi(\theta^*)}{\hat{\pi}(\theta^*|\mathbf{y})}$$

$\hat{\pi}(\theta|\mathbf{y})$ may be the Gaussian approximation based on the MLE

Tractable likelihood
Chib's solution

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\,\pi(\theta)}{\pi(\theta|\mathbf{y})}\,, \forall\theta$$

For an arbitrary value θ^* of θ, the Chib's Chib (1995) approximation to the marginal likelihood is

$$\hat{m}(\mathbf{y}) = \frac{f(\mathbf{y}|\theta^*)\,\pi(\theta^*)}{\hat{\pi}(\theta^*|\mathbf{y})}$$

$\hat{\pi}(\theta|\mathbf{y})$ may be the Gaussian approximation based on the MLE

Tractable likelihood
Chib's solution

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\,\pi(\theta)}{\pi(\theta|\mathbf{y})}\,, \forall\theta$$

For an arbitrary value θ^* of θ, the Chib's **Chib (1995)** approximation to the marginal likelihood is

$$\hat{m}(\mathbf{y}) = \frac{f(\mathbf{y}|\theta^*)\,\pi(\theta^*)}{\hat{\pi}(\theta^*|\mathbf{y})}$$

$\hat{\pi}(\theta|\mathbf{y})$ may be the Gaussian approximation based on the MLE

Tractable likelihood
Chib's solution

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\,\pi(\theta)}{\pi(\theta|\mathbf{y})}\,, \forall\theta$$

For an arbitrary value θ^* of θ, the Chib's **Chib (1995)** approximation to the marginal likelihood is

$$\hat{m}(\mathbf{y}) = \frac{f(\mathbf{y}|\theta^*)\,\pi(\theta^*)}{\hat{\pi}(\theta^*|\mathbf{y})}$$

$\hat{\pi}(\theta|\mathbf{y})$ may be the Gaussian approximation based on the MLE

Tractable likelihood
Chib's solution

A second solution is to use a nonparametric approximation based on a preliminary MCMC sample

In the setting of latent variables models, Chib's approximation can be attractive as there exists a natural approximation to $\pi_k(\theta^*|\mathbf{y})$

$$\hat{\pi}(\theta^*|\mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T} \pi(\theta^*|\mathbf{y}, \mathbf{z}^{(t)})$$

where the $\mathbf{z}^{(t)}$'s are the latent variables simulated by the MCMC sampler

High variance and curse of dimensionality

Tractable likelihood
Chib's solution

A second solution is to use a nonparametric approximation based on a preliminary MCMC sample

In the setting of latent variables models, Chib's approximation can be attractive as there exists a natural approximation to $\pi_k(\theta^*|\mathbf{y})$

$$\hat{\pi}(\theta^*|\mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T}\pi(\theta^*|\mathbf{y}, \mathbf{z}^{(t)})$$

where the $\mathbf{z}^{(t)}$'s are the latent variables simulated by the MCMC sampler

High variance and curse of dimensionality

Tractable likelihood
Chib's solution

A second solution is to use a nonparametric approximation based on a preliminary MCMC sample

In the setting of latent variables models, Chib's approximation can be attractive as there exists a natural approximation to $\pi_k(\theta^*|\mathbf{y})$

$$\hat{\pi}(\theta^*|\mathbf{y}) = \frac{1}{T}\sum_{t=1}^{T}\pi(\theta^*|\mathbf{y},\mathbf{z}^{(t)})$$

where the $\mathbf{z}^{(t)}$'s are the latent variables simulated by the MCMC sampler

High variance and curse of dimensionality

Tractable likelihood
Some others alternatives

Large set of approximations for marginal likelihood or Bayes factors

- Annealed Importance Sampling by Neal (2001)
- Bridge sampling techniques Meng and Wong (1996), Meng and Schilling 2002
  Nice R library bridgesampling (Gronau, Singmann, Wagenmakers)
- The Savage-Dickey ratio Verdinelli and Wasserman (1995), Marin and Robert (2010)

Tractable likelihood
Some others alternatives

Large set of approximations for marginal likelihood or Bayes factors

- Annealed Importance Sampling by Neal [2001]
- Bridge sampling techniques Meng and Wong 1996; Meng and Schilling 2002
  Nice R library bridgesampling (Gronau, Singmann, Wagenmakers)
- The Savage–Dickey ratio Verdinelli and Wasserman (1995), Marin and Robert (2010)
- ...

Tractable likelihood
Some others alternatives

Large set of approximations for marginal likelihood or Bayes factors

- Annealed Importance Sampling by **Neal [2001]**
- Bridge sampling techniques **Meng and Wong 1996; Meng and Schilling 2002**
  Nice R library bridgesampling (Gronau, Singmann, Wagenmakers)
- The Savage–Dickey ratio **Verdinelli and Wasserman (1995), Marin and Robert (2010)**
- ...

Tractable likelihood
Some others alternatives

Large set of approximations for marginal likelihood or Bayes factors

- Annealed Importance Sampling by **Neal [2001]**
- Bridge sampling techniques **Meng and Wong 1996; Meng and Schilling 2002**
  Nice R library bridgesampling (Gronau, Singmann, Wagenmakers)
- The Savage–Dickey ratio **Verdinelli and Wasserman (1995), Marin and Robert (2010)**
- ...

Tractable likelihood
Some others alternatives

Large set of approximations for marginal likelihood
or Bayes factors

- Annealed Importance Sampling by **Neal [2001]**
- Bridge sampling techniques **Meng and Wong 1996; Meng and Schilling 2002**
  Nice R library bridgesampling (Gronau, Singmann, Wagenmakers)
- The Savage–Dickey ratio **Verdinelli and Wasserman (1995), Marin and Robert (2010)**
- ...

Tractable likelihood
Logistic regression approximation

Idea: reduce an estimation problem to a classification problem
Several versions:

- Logistic regression for density estimation: Hastie et al. (2003)
- Intensity estimation: Baddeley et al. (2010)
- Logistic regression for estimation in unnormalised models: Geyer (1994) and Gutmann and Hyvärinen (2012)

The last one is called noise-contrastive estimation by the authors

Tractable likelihood
Logistic regression approximation

Idea: reduce an estimation problem to a classification problem
Several versions:

- Logistic regression for density estimation: **Hastie et al. (2003)**
- Intensity estimation: **Baddeley et al. (2010)**
- Logistic regression for estimation in unnormalised models: **Geyer (1994) and Gutmann and Hyvärinen (2012)**

The last one is called noise-contrastive estimation by the authors

Idea: reduce an estimation problem to a classification problem
Several versions:

- Logistic regression for density estimation: **Hastie et al. (2003)**
- Intensity estimation: **Baddeley et al. (2010)**
- Logistic regression for estimation in unnormalised models: **Geyer (1994) and Gutmann and Hyvärinen (2012)**

The last one is called noise-contrastive estimation by the authors

Tractable likelihood
Logistic regression approximation

Idea: reduce an estimation problem to a classification problem
Several versions:

- Logistic regression for density estimation: **Hastie et al. (2003)**
- Intensity estimation: **Baddeley et al. (2010)**
- Logistic regression for estimation in unnormalised models: **Geyer (1994) and Gutmann and Hyvärinen (2012)**

The last one is called noise-contrastive estimation by the authors

Tractable likelihood
Logistic regression approximation

Idea: reduce an estimation problem to a classification problem
Several versions:

- Logistic regression for density estimation: **Hastie et al. (2003)**
- Intensity estimation: **Baddeley et al. (2010)**
- Logistic regression for estimation in unnormalised models: **Geyer (1994) and Gutmann and Hyvärinen (2012)**

The last one is called noise-contrastive estimation by the authors

Tractable likelihood
Logistic regression approximation

Suppose that

- $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$
- $u_1, \ldots, u_N$ is an N-sample from $\pi(\cdot)$

Let $\zeta = (\theta^1, \ldots, \theta^N, u_1, \ldots, u_N)$

We note $z_i = 1$ if the ζ_i comes from $\pi(\cdot|\mathbf{y})$ and $z_i = 0$ if $\zeta_i = 0$ comes from $\pi(\cdot)$:

$$f(\theta|z=1) = \pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{m(\mathbf{y})}$$

$$f(\theta|z=0) = \pi(\theta)$$

Tractable likelihood
Logistic regression approximation

Suppose that

- $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$
- $u_1, \ldots, u_N$ is an N-sample from $\pi(\cdot)$

Let $\zeta = (\theta^1, \ldots, \theta^N, u_1, \ldots, u_N)$

We note $z_i = 1$ if the ζ_i comes from $\pi(\cdot|\mathbf{y})$ and $z_i = 0$ if $\zeta_i = 0$ comes from $\pi(\cdot)$:

$$f(\theta|z=1) = \pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{m(\mathbf{y})}$$

$$f(\theta|z=0) = \pi(\theta)$$

Tractable likelihood
Logistic regression approximation

Suppose that

- $\theta^1, \ldots, \theta^N$ is an N-sample from $\pi(\cdot|\mathbf{y})$
- $u_1, \ldots, u_N$ is an N-sample from $\pi(\cdot)$

Let $\zeta = (\theta^1, \ldots, \theta^N, u_1, \ldots, u_N)$

We note $z_i = 1$ if the ζ_i comes from $\pi(\cdot|\mathbf{y})$ and $z_i = 0$ if $\zeta_i = 0$ comes from $\pi(\cdot)$:

$$f(\theta|z=1) = \pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{m(\mathbf{y})}$$

$$f(\theta|z=0) = \pi(\theta)$$

Tractable likelihood
Logistic regression approximation

The log-odds ratio is

$$\eta(\theta) = \log\left[\frac{\mathbb{P}(z=1|\theta)}{\mathbb{P}(z=0|\theta)}\right]$$

From an estimate of $\eta(\theta)$, we can deduce an estimate of $m(\mathbf{y})$

$$\eta(\theta) = \log(f(\mathbf{y}|\theta)\pi(\theta)) - \log(m(\mathbf{y})) - \log(\pi(\theta))$$

$$\eta(\theta) = c + \log(f(\mathbf{y}|\theta))$$

where

$$c = -\log(m(\mathbf{y}))$$

Tractable likelihood
Logistic regression approximation

The log-odds ratio is

$$\eta(\theta) = \log\left[\frac{\mathbb{P}(z=1|\theta)}{\mathbb{P}(z=0|\theta)}\right]$$

From an estimate of $\eta(\theta)$, we can deduce an estimate of $m(\mathbf{y})$

$$\eta(\theta) = \log(f(\mathbf{y}|\theta)\pi(\theta)) - \log(m(\mathbf{y})) - \log(\pi(\theta))$$

$$\eta(\theta) = c + \log(f(\mathbf{y}|\theta))$$

where

$$c = -\log(m(\mathbf{y}))$$

Tractable likelihood
Logistic regression approximation

The log-odds ratio is

$$\eta(\theta) = \log\left[\frac{\mathbb{P}(z=1|\theta)}{\mathbb{P}(z=0|\theta)}\right]$$

From an estimate of $\eta(\theta)$, we can deduce an estimate of $m(\mathbf{y})$

$$\eta(\theta) = \log(f(\mathbf{y}|\theta)\pi(\theta)) - \log(m(\mathbf{y})) - \log(\pi(\theta))$$

$$\eta(\theta) = c + \log(f(\mathbf{y}|\theta))$$

where

$$c = -\log(m(\mathbf{y}))$$

Tractable likelihood
Logistic regression approximation

The log-odds ratio is

$$\eta(\theta) = \log \left[\frac{\mathbb{P}(z = 1|\theta)}{\mathbb{P}(z = 0|\theta)} \right]$$

From an estimate of $\eta(\theta)$, we can deduce an estimate of $m(\mathbf{y})$

$$\eta(\theta) = \log(f(\mathbf{y}|\theta)\pi(\theta)) - \log(m(\mathbf{y})) - \log(\pi(\theta))$$

$$\eta(\theta) = c + \log(f(\mathbf{y}|\theta))$$

where

$$c = -\log(m(\mathbf{y}))$$

Tractable likelihood
Logistic regression approximation

c is the intercept of a logistic regression model

c can be estimate using our two simulated datasets and the maximum likelihood estimator (not explicit)

$$\hat{c} \in \arg\max_{c} \left(\sum_{i=1}^{n} [c + \log(f(\mathbf{y}|\theta_i))] - \sum_{i=1}^{n} \log(1 + \exp(c) \log(f(\mathbf{y}|\theta_i))) - \sum_{i=1}^{n} \log(1 + \exp(c) \log(f(\mathbf{y}|u_i))) \right)$$

Tractable likelihood
Logistic regression approximation

c is the intercept of a logistic regression model

c can be estimate using our two simulated datasets and the maximum likelihood estimator (not explicit)

$$\hat{c} \in \arg\max_{c} \left(\sum_{i=1}^{n} [c + \log(f(\mathbf{y}|\theta_i))] - \sum_{i=1}^{n} \log(1 + \exp(c) \log(f(\mathbf{y}|\theta_i))) - \sum_{i=1}^{n} \log(1 + \exp(c) \log(f(\mathbf{y}|u_i))) \right)$$

Tractable likelihood
Logistic regression approximation

c is the intercept of a logistic regression model

c can be estimate using our two simulated datasets and the maximum likelihood estimator (not explicit)

$$\hat{c} \in \arg\max_{c} \left(\sum_{i=1}^{n} [c + \log(f(\mathbf{y}|\theta_i))] \right.$$
$$- \sum_{i=1}^{n} \log(1 + \exp(c) \log(f(\mathbf{y}|\theta_i)))$$
$$\left. - \sum_{i=1}^{n} \log(1 + \exp(c) \log(f(\mathbf{y}|u_i))) \right)$$

Tractable likelihood
Logistic regression approximation

A very toy example

$$y|\theta \sim \mathcal{N}(\theta, 1)$$

$$\theta \sim \mathcal{N}(0, 1)$$

In such a case,

$$\theta|y \sim \mathcal{N}\left(\frac{y}{2}, \frac{1}{2}\right)$$

$$m(y) = \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left(-\frac{y^2}{4}\right)$$

Tractable likelihood
Logistic regression approximation

A very toy example

$$y|\theta \sim \mathcal{N}(\theta, 1)$$

$$\theta \sim \mathcal{N}(0, 1)$$

In such a case,

$$\theta|y \sim \mathcal{N}\left(\frac{y}{2}, \frac{1}{2}\right)$$

$$m(y) = \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left(-\frac{y^2}{4}\right)$$

Tractable likelihood
Logistic regression approximation

A very toy example

$$y|\theta \sim \mathcal{N}(\theta, 1)$$

$$\theta \sim \mathcal{N}(0, 1)$$

In such a case,

$$\theta|y \sim \mathcal{N}\left(\frac{y}{2}, \frac{1}{2}\right)$$

$$m(y) = \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left(-\frac{y^2}{4}\right)$$

```
# y <- rnorm(1,rnorm(1),1)
y <- -0.5
target <- dnorm(y,0,sqrt(2))

thetaprior <- rnorm(1000)
thetapost <- rnorm(1000,y/2,sqrt(1/2))

zeta <- c(thetapost,thetaprior)

z <- c(rep(1,1000),rep(0,1000))
x <- log(dnorm(y,zeta,1))
df <- data.frame(z=z,x=x)
model <- glm(z~offset(x),data=df,family=binomial)

1/exp(as.numeric(model$coefficients))
```

Tractable likelihood
Logistic regression approximation

With Christian Robert, we are testing this strategy, the first results are impressive (work in progress)

We will work also on promising extension to

- estimate ratio of normalizing contants
- replace the prior by a distribution
- adapt to MCMC samples from the posterior
- adapt to latent variable models

Tractable likelihood
Logistic regression approximation

With Christian Robert, we are testing this strategy, the first results are impressive (work in progress)

We will work also on promising extension to

- estimate ratio of normalizing contants
- replace the prior by a distribution
- adapt to MCMC samples from the posterior
- adapt to latent variable models

Tractable likelihood
Logistic regression approximation

With Christian Robert, we are testing this strategy, the first results are impressive (work in progress)

We will work also on promising extension to

- estimate ratio of normalizing contants
- replace the prior by a distribution
- adapt to MCMC samples from the posterior
- adapt to latent variable models

With Christian Robert, we are testing this strategy, the first results are impressive (work in progress)

We will work also on promising extension to

- estimate ratio of normalizing contants
- replace the prior by a distribution
- adapt to MCMC samples from the posterior
- adapt to latent variable models

Tractable likelihood
Logistic regression approximation

With Christian Robert, we are testing this strategy, the first results are impressive (work in progress)

We will work also on promising extension to

- estimate ratio of normalizing contants
- replace the prior by a distribution
- adapt to MCMC samples from the posterior
- adapt to latent variable models

Tractable likelihood
Logistic regression approximation

With Christian Robert, we are testing this strategy, the first results are impressive (work in progress)

We will work also on promising extension to

- estimate ratio of normalizing contants
- replace the prior by a distribution
- adapt to MCMC samples from the posterior
- adapt to latent variable models

Intractable likelihood
Context

When the likelihood function $f(\mathbf{y}|\theta)$ is expensive or impossible to calculate, it is extremely difficult to sample from the posterior distribution

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta)f(\mathbf{y}|\theta)$$

Two typical situations:

$f(\mathbf{y}|\theta) = \int f(\mathbf{y}, \mathbf{u}|\theta)\mu(d\mathbf{u})$, the calculation of this integral is intractable and the latent vector $\mathbf{u}$ takes values in a high dimensional space (e.g. population genetics models)

$f(\mathbf{y}|\theta) = g(\mathbf{y}, \theta)/Z(\theta)$ and the calculation of $Z(\theta)$ is intractable (e.g. for Markov random fields)

Intractable likelihood
Context

When the likelihood function $f(\mathbf{y}|\theta)$ is expensive or impossible to calculate, it is extremely difficult to sample from the posterior distribution

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) f(\mathbf{y}|\theta)$$

Two typical situations:

$f(\mathbf{y}|\theta) = \int f(\mathbf{y}, \mathbf{u}|\theta)\mu(d\mathbf{u})$, the calculation of this integral is intractable and the latent vector $\mathbf{u}$ takes values in a high dimensional space (e.g. population genetics models)

$f(\mathbf{y}|\theta) = g(\mathbf{y}, \theta)/Z(\theta)$ and the calculation of $Z(\theta)$ is intractable (e.g. for Markov random fields)

Intractable likelihood
Context

When the likelihood function $f(\mathbf{y}|\theta)$ is expensive or impossible to calculate, it is extremely difficult to sample from the posterior distribution

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) f(\mathbf{y}|\theta)$$

Two typical situations:

$f(\mathbf{y}|\theta) = \int f(\mathbf{y}, \mathbf{u}|\theta)\mu(d\mathbf{u})$, the calculation of this integral is intractable and the latent vector $\mathbf{u}$ takes values in a high dimensional space (e.g. population genetics models)

$f(\mathbf{y}|\theta) = g(\mathbf{y}, \theta)/Z(\theta)$ and the calculation of $Z(\theta)$ is intractable (e.g. for Markov random fields)

Intractable likelihood
Context

When the likelihood function $f(\mathbf{y}|\theta)$ is expensive or impossible to calculate, it is extremely difficult to sample from the posterior distribution

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) f(\mathbf{y}|\theta)$$

Two typical situations:

$f(\mathbf{y}|\theta) = \int f(\mathbf{y}, \mathbf{u}|\theta)\mu(d\mathbf{u})$, the calculation of this integral is intractable and the latent vector $\mathbf{u}$ takes values in a high dimensional space (e.g. population genetics models)

$f(\mathbf{y}|\theta) = g(\mathbf{y}, \theta)/Z(\theta)$ and the calculation of $Z(\theta)$ is intractable (e.g. for Markov random fields)

When the likelihood function $f(\mathbf{y}|\theta)$ is expensive or impossible to calculate, it is extremely difficult to sample from the posterior distribution

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) f(\mathbf{y}|\theta)$$

Two typical situations:

$f(\mathbf{y}|\theta) = \int f(\mathbf{y}, \mathbf{u}|\theta)\mu(d\mathbf{u})$, the calculation of this integral is intractable and the latent vector $\mathbf{u}$ takes values in a high dimensional space (e.g. population genetics models)

$f(\mathbf{y}|\theta) = g(\mathbf{y}, \theta)/Z(\theta)$ and the calculation of $Z(\theta)$ is intractable (e.g. for Markov random fields)

Intractable likelihood
Context

ABC is a technique that only requires being able to sample from the likelihood $f(\cdot|\theta)$

This technique stemmed from population genetics models, about 20 years ago, and population geneticists still significantly contribute to methodological developments of ABC

If, with Christian Robert, we work on ABC methods, we can be very grateful to our biologist colleagues!

When the likelihood functions $f_m(y|\theta_m)$ are intractable, it is very challenging to estimate the marginal likelihoods

$$\int f_m(y|\theta_m)\pi_m(\theta_m)d\theta_m$$

ABC is a technique that only requires being able to sample from the likelihood $f(\cdot|\theta)$

This technique stemmed from population genetics models, about 20 years ago, and population geneticists still significantly contribute to methodological developments of ABC

If, with Christian Robert, we work on ABC methods, we can be very grateful to our biologist colleagues!

When the likelihood functions $f_m(y|\theta_m)$ are intractable, it is very challenging to estimate the marginal likelihoods

$$\int f_m(y|\theta_m)\pi_m(\theta_m)d\theta_m$$

ABC is a technique that only requires being able to sample from the likelihood $f(\cdot|\theta)$

This technique stemmed from population genetics models, about 20 years ago, and population geneticists still significantly contribute to methodological developments of ABC

If, with Christian Robert, we work on ABC methods, we can be very grateful to our biologist colleagues!

When the likelihood functions $f_m(y|\theta_m)$ are intractable, it is very challenging to estimate the marginal likelihoods

$$\int f_m(y|\theta_m)\pi_m(\theta_m)d\theta_m$$

Intractable likelihood
Context

ABC is a technique that only requires being able to sample from the likelihood $f(\cdot|\theta)$

This technique stemmed from population genetics models, about 20 years ago, and population geneticists still significantly contribute to methodological developments of ABC

If, with Christian Robert, we work on ABC methods, we can be very grateful to our biologist colleagues!

When the likelihood functions $f_m(\mathbf{y}|\theta_m)$ are intractable, it is very challenging to estimate the marginal likelihoods

$$\int f_m(\mathbf{y}|\theta_m)\pi_m(\theta_m)d\theta_m$$

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis

1) For $i = 1, \ldots, N$

   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate z from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(z), \eta(y))$

2) Order the distances $d_{(1)}, \ldots, d_{(N)}$

3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis

1) For $i = 1, \ldots, N$

   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate z from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(z), \eta(y))$

2) Order the distances $d_{(1)}, \ldots, d_{(N)}$

3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields **Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis**

1) For $i = 1, \ldots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate z from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(z), \eta(y))$
2) Order the distances $d_{(1)}, \ldots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields **Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis**

1) For $i = 1, \ldots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate $\mathbf{z}$ from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(\mathbf{z}), \eta(\mathbf{y}))$
2) Order the distances $d_{(1)}, \ldots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields **Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis**

1) For $i = 1, \dots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate $\mathbf{z}$ from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(\mathbf{z}), \eta(\mathbf{y}))$
2) Order the distances $d_{(1)}, \dots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis

1) For $i = 1, \ldots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate $\mathbf{z}$ from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(\mathbf{z}), \eta(\mathbf{y}))$
2) Order the distances $d_{(1)}, \ldots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis

1) For $i = 1, \dots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate $\mathbf{z}$ from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(\mathbf{z}), \eta(\mathbf{y}))$
2) Order the distances $d_{(1)}, \dots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

ABC likelihood-free methods for model choice in Gibbs random fields Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis

1) For $i = 1, \ldots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate $\mathbf{z}$ from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(\mathbf{z}), \eta(\mathbf{y}))$
2) Order the distances $d_{(1)}, \ldots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

ABC likelihood-free methods for model choice in Gibbs random fields Grelaud, Robert, Marin, Rodolphe and Taly (2009) Bayesian Analysis

1) For $i = 1, \ldots, N$
   a) Generate m_i from the prior $\mathbb{P}(\mathscr{M} = m)$
   b) Generate θ'_{m_i} from the prior $\pi_{m_i}(\cdot)$
   c) Generate $\mathbf{z}$ from the model $f_{m_i}(\cdot|\theta'_{m_i})$
   d) Calculate $d_i = d(\eta(\mathbf{z}), \eta(\mathbf{y}))$
2) Order the distances $d_{(1)}, \ldots, d_{(N)}$
3) Select the model using the majority rule among the k-smallest distances index set

A standard K-Nearest Neighbor classifier

Intractable likelihood
Classic ABC model choice procedure

If $\eta(\mathbf{y})$ is a sufficient statistics for the model choice problem, this can work pretty well

If not...

Lack of confidence in approximate Bayesian computation model choice
Robert, Cornuet, Marin, Pillai (2011) PNAS

Relevant statistics for Bayesian model choice
Marin, Pillai, Robert, Rousseau (2014) JRSS B

If $\eta(\mathbf{y})$ is a sufficient statistics for the model choice problem, this can work pretty well

If not...

Lack of confidence in approximate Bayesian computation model choice
Robert, Cornuet, Marin, Pillai (2011) PNAS

Relevant statistics for Bayesian model choice
Marin, Pillai, Robert, Rousseau (2014) JRSS B

Intractable likelihood
Classic ABC model choice procedure

- intuitive
- simple to implement
- embarrassingly parallelisable
- BUT curse of dimensionality: most of the simulations are at the boundary of the space as the number of summary statistics increases

Intractable likelihood
Classic ABC model choice procedure

- intuitive
- simple to implement
- embarrassingly parallelisable
- BUT curse of dimensionality: most of the simulations are at the boundary of the space as the number of summary statistics increases

Intractable likelihood
Classic ABC model choice procedure

- intuitive
- simple to implement
- embarrassingly parallelisable
- BUT curse of dimensionality: most of the simulations are at the boundary of the space as the number of summary statistics increases

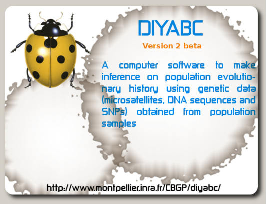
Intractable likelihood
Classic ABC model choice procedure

- intuitive
- simple to implement
- embarrassingly parallelisable
- BUT curse of dimensionality: most of the simulations are at the boundary of the space as the number of summary statistics increases

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation **Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics**

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data **Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics**

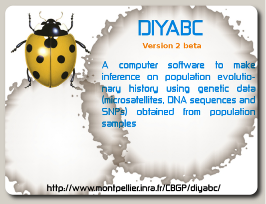


Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics

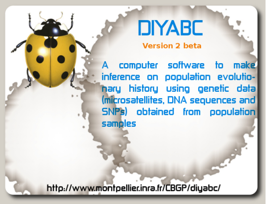


Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics

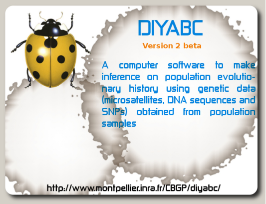


Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation **Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics**

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data **Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics**

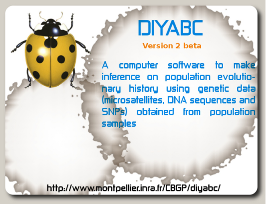


Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation **Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics**

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data **Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics**

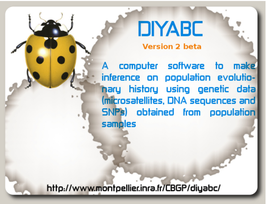


Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation **Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics**

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data **Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics**

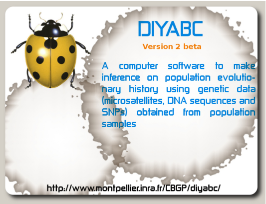


Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
DIY-ABC software

Infering population history with DIY ABC: a user-friedly approach Approximate Bayesian Computation **Cornuet, Santos, Beaumont, Robert, Marin, Balding, Guillemaud, Estoup (2008) Bioinformatics**

DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data **Cornuet, Pudlo, Veyssier, Dehne-Garcia, Gautier, Leblois, Marin, Estoup (2014) Bioinformatics**



Asian ladybug
European honey bee
Drosophila suzukii
Four human populations, to study the out-of-Africa colonization
Pigmy human populations

Intractable likelihood
Frontline news from population geneticists country

DIYABC (2014) paper has now around 500 citations

- simulate from the model can be very computationally intensive, parallelizable algorithms are necessary
- sequential methods are difficult to calibrate and do not give reproducible results
- available techniques to select the summary statistics do not give reproducible results

Intractable likelihood
Frontline news from population geneticists country

DIYABC (2014) paper has now around 500 citations

- simulate from the model can be very computationally intensive, parallelizable algorithms are necessary
- sequential methods are difficult to calibrate and do not give reproducible results
- available techniques to select the summary statistics do not give reproducible results

Intractable likelihood
Frontline news from population geneticists country

DIYABC (2014) paper has now around 500 citations

- simulate from the model can be very computationally intensive, parallelizable algorithms are necessary
- sequential methods are difficult to calibrate and do not give reproducible results
- available techniques to select the summary statistics do not give reproducible results

Intractable likelihood
Frontline news from population geneticists country

DIYABC (2014) paper has now around 500 citations

- simulate from the model can be very computationally intensive, parallelizable algorithms are necessary
- sequential methods are difficult to calibrate and do not give reproducible results
- available techniques to select the summary statistics do not give reproducible results

Intractable likelihood
Frontline news from population geneticists country

DIYABC (2014) paper has now around 500 citations

- simulate from the model can be very computationally intensive, parallelizable algorithms are necessary
- sequential methods are difficult to calibrate and do not give reproducible results
- available techniques to select the summary statistics do not give reproducible results

Intractable likelihood
Frontline news from population geneticists country

Two major difficulties

- to ensure reliability of the method, the number of simulations should be large
- choice of the summary statistics is still a problem

Intractable likelihood
Frontline news from population geneticists country

Two major difficulties

- to ensure reliability of the method, the number of simulations should be large
- choice of the summary statistics is still a problem

Intractable likelihood
Frontline news from population geneticists country

Two major difficulties

- to ensure reliability of the method, the number of simulations should be large
- choice of the summary statistics is still a problem

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
Use modern machine learning tools

Exploiting a large number of summary statistics is not an issue for some machine learning methods

Idea: learn on a huge reference table using random forests

Some theoretical guarantees for sparse problems

Analysis of a random forest model
Biau (2012) JMLR

Consistency of random forests
Scornet, Biau, Vert (2015) The Annals of Statistics

This work stands at the interface between Bayesian inference and machine learning techniques

Intractable likelihood
ABC random forests

Reliable ABC model choice via random forests **Pudlo, Marin, Estoup, Cornuet, Gauthier and Robert (2016) Bioinformatics**

Input ABC reference table involving model index and summ. statistics

$$\begin{bmatrix} m^{(1)} & \eta_1(\mathbf{z}^{(1)}) & \eta_2(\mathbf{z}^{(1)}) & \dots & \eta_d(\mathbf{z}^{(1)}) \\ m^{(2)} & \eta_1(\mathbf{z}^{(2)}) & \eta_2(\mathbf{z}^{(2)}) & \dots & \eta_d(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m^{(N)} & \eta_1(\mathbf{z}^{(N)}) & \eta_2(\mathbf{z}^{(N)}) & \dots & \eta_d(\mathbf{z}^{(N)}) \end{bmatrix}$$

possibly large collection of summary statistics: from scientific theory input to machine-learning alternatives

Output a random forest classifier to infer model indexes $\widehat{m(\eta(\mathbf{y}))}$

abcrf R library

Intractable likelihood
ABC random forests

Reliable ABC model choice via random forests **Pudlo, Marin, Estoup, Cornuet, Gauthier and Robert (2016) Bioinformatics**

Input ABC reference table involving model index and summ. statistics

$$\begin{bmatrix} m^{(1)} & \eta_1(z^{(1)}) & \eta_2(z^{(1)}) & \dots & \eta_d(z^{(1)}) \\ m^{(2)} & \eta_1(z^{(2)}) & \eta_2(z^{(2)}) & \dots & \eta_d(z^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m^{(N)} & \eta_1(z^{(N)}) & \eta_2(z^{(N)}) & \dots & \eta_d(z^{(N)}) \end{bmatrix}$$

possibly large collection of summary statistics: from scientific theory input to machine-learning alternatives

Output a random forest classifier to infer model indexes $\widehat{m(\eta(y))}$

abcrf R library

Intractable likelihood
ABC random forests

Reliable ABC model choice via random forests **Pudlo, Marin, Estoup, Cornuet, Gauthier and Robert (2016) Bioinformatics**

Input ABC reference table involving model index and summ. statistics

$$\begin{bmatrix} m^{(1)} & \eta_1(\mathbf{z}^{(1)}) & \eta_2(\mathbf{z}^{(1)}) & \dots & \eta_d(\mathbf{z}^{(1)}) \\ m^{(2)} & \eta_1(\mathbf{z}^{(2)}) & \eta_2(\mathbf{z}^{(2)}) & \dots & \eta_d(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m^{(N)} & \eta_1(\mathbf{z}^{(N)}) & \eta_2(\mathbf{z}^{(N)}) & \dots & \eta_d(\mathbf{z}^{(N)}) \end{bmatrix}$$

possibly large collection of summary statistics: from scientific theory input to machine-learning alternatives

Output a random forest classifier to infer model indexes $\widehat{m(\eta(y))}$

abcrf R library

Intractable likelihood
ABC random forests

Reliable ABC model choice via random forests **Pudlo, Marin, Estoup, Cornuet, Gauthier and Robert (2016) Bioinformatics**

Input ABC reference table involving model index and summ. statistics

$$\begin{bmatrix} m^{(1)} & \eta_1(\mathbf{z}^{(1)}) & \eta_2(\mathbf{z}^{(1)}) & \dots & \eta_d(\mathbf{z}^{(1)}) \\ m^{(2)} & \eta_1(\mathbf{z}^{(2)}) & \eta_2(\mathbf{z}^{(2)}) & \dots & \eta_d(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m^{(N)} & \eta_1(\mathbf{z}^{(N)}) & \eta_2(\mathbf{z}^{(N)}) & \dots & \eta_d(\mathbf{z}^{(N)}) \end{bmatrix}$$

possibly large collection of summary statistics: from scientific theory input to machine-learning alternatives

Output a random forest classifier to infer model indexes $\widehat{m(\eta(y))}$

abcrf R library

Intractable likelihood
ABC random forests

Reliable ABC model choice via random forests **Pudlo, Marin, Estoup, Cornuet, Gauthier and Robert (2016) Bioinformatics**

Input ABC reference table involving model index and summ. statistics

$$\begin{bmatrix} m^{(1)} & \eta_1(\mathbf{z}^{(1)}) & \eta_2(\mathbf{z}^{(1)}) & \dots & \eta_d(\mathbf{z}^{(1)}) \\ m^{(2)} & \eta_1(\mathbf{z}^{(2)}) & \eta_2(\mathbf{z}^{(2)}) & \dots & \eta_d(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m^{(N)} & \eta_1(\mathbf{z}^{(N)}) & \eta_2(\mathbf{z}^{(N)}) & \dots & \eta_d(\mathbf{z}^{(N)}) \end{bmatrix}$$

possibly large collection of summary statistics: from scientific theory input to machine-learning alternatives

Output a random forest classifier to infer model indexes $\widehat{m(\eta(\mathbf{y}))}$

abcrf R library

Intractable likelihood
ABC random forests

Reliable ABC model choice via random forests **Pudlo, Marin, Estoup, Cornuet, Gauthier and Robert (2016) Bioinformatics**

Input ABC reference table involving model index and summ. statistics

$$\begin{bmatrix} m^{(1)} & \eta_1(\mathbf{z}^{(1)}) & \eta_2(\mathbf{z}^{(1)}) & \dots & \eta_d(\mathbf{z}^{(1)}) \\ m^{(2)} & \eta_1(\mathbf{z}^{(2)}) & \eta_2(\mathbf{z}^{(2)}) & \dots & \eta_d(\mathbf{z}^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m^{(N)} & \eta_1(\mathbf{z}^{(N)}) & \eta_2(\mathbf{z}^{(N)}) & \dots & \eta_d(\mathbf{z}^{(N)}) \end{bmatrix}$$

possibly large collection of summary statistics: from scientific theory input to machine-learning alternatives

Output a random forest classifier to infer model indexes $\widehat{m(\eta(\mathbf{y}))}$

abcrf R library

Random forest predicts a MAP model index, from the observed dataset: the predictor provided by the forest is good enough to select the most likely model but not to derive directly the associated posterior probability

frequency of trees associated with majority model is no proper substitute to the true posterior probability

Intractable likelihood
ABC random forests

Estimate of the posterior probability of the selected model

$$\mathbb{P}[\mathscr{M} = \widehat{m}(\eta(\mathbf{y}))|\eta(\mathbf{y})]$$

random comes from $\mathscr{M}$ (bayesian)!

$$\mathbb{P}[\mathscr{M} = \widehat{m}(\eta(\mathbf{y}))|\eta(\mathbf{y})] = 1 - \mathbb{E}\left[\mathbb{I}(\mathscr{M} \neq \widehat{m}(\eta(\mathbf{y})))|\eta(\mathbf{y})\right]$$

A second random forest in regression

1) compute the value of $\mathbb{I}(\mathscr{M} \neq \widehat{m(\eta(\mathbf{z}))})$ for the trained random forest and for all terms in the ABC reference table using the out-of-bag classifiers;
2) train a RF regression and get
$\rho(\eta(\mathbf{z})) = \widehat{\mathbb{E}}\left[\mathbb{I}(\mathscr{M} \neq \widehat{m(\eta(\mathbf{z}))})|\eta(\mathbf{z})\right]$;
3) return $\widehat{\mathbb{P}}[\mathscr{M} = \widehat{m(\eta(\mathbf{y}))}|\eta(\mathbf{y})] = 1 - \rho(\eta(\mathbf{y}))$.

on same reference table out-of-bag magic trick avoid over-fitting!