Can ABC be Used for Model Selection?

Xavier Didelot (University of Oxford)
Rich Everitt (University of Oxford)
Adam Johansen (University of Warwick)
Dan Lawson (University of Bristol)

ABC in London Workshop
Imperial College London
5th May 2011

ABC

- Pritchard *et al* (1999) proposed the ABC algorithm to approximately sample from $p(\theta|x_{\text{obs}})$:

ABC

- Pritchard *et al* (1999) proposed the ABC algorithm to approximately sample from $p(\theta|x_{\mathrm{obs}})$:
  1. Generate $\theta^* \sim p(\theta)$
  2. Simulate $x^* \sim p(x|\theta^*)$
  3. Accept θ^* if $D(s(x^*), s(x_{\mathrm{obs}})) < \epsilon$

ABC

- Pritchard *et al* (1999) proposed the ABC algorithm to approximately sample from $p(\theta|x_{\text{obs}})$:
  1. Generate $\theta^* \sim p(\theta)$
  2. Simulate $x^* \sim p(x|\theta^*)$
  3. Accept θ^* if $D(s(x^*), s(x_{\text{obs}})) < \epsilon$
- Also proposed ABC approach to model selection when M_1 nested in M_2: infer under M_2 with prior such that M_1 gets half of the weight

ABC for model selection

- Nested models (Pritchard *et al* 1999)

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):
  1. With probability 0.5 set $M^* = M_1$ otherwise set $M^* = M_2$
  2. Generate $\theta^* \sim p(\theta|M^*)$
  3. Simulate $x^* \sim p(x|\theta^*, M^*)$
  4. Accept (M^*, θ^*) if $D(s(x^*), s(x_{\text{obs}})) < \epsilon$

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):
  1. With probability 0.5 set $M^* = M_1$ otherwise set $M^* = M_2$
  2. Generate $\theta^* \sim p(\theta | M^*)$
  3. Simulate $x^* \sim p(x | \theta^*, M^*)$
  4. Accept (M^*, θ^*) if $D(s(x^*), s(x_{\text{obs}})) < \epsilon$

  Ratio of acceptance of M_1 and M_2 approximates Bayes Factor

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):
  1. With probability 0.5 set $M^* = M_1$ otherwise set $M^* = M_2$
  2. Generate $\theta^* \sim p(\theta|M^*)$
  3. Simulate $x^* \sim p(x|\theta^*, M^*)$
  4. Accept (M^*, θ^*) if $D(s(x^*), s(x_{\text{obs}})) < \epsilon$

  Ratio of acceptance of M_1 and M_2 approximates Bayes Factor
- ABC-SMC technique (Toni *et al* 2009)

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):
  1. With probability 0.5 set $M^* = M_1$ otherwise set $M^* = M_2$
  2. Generate $\theta^* \sim p(\theta|M^*)$
  3. Simulate $x^* \sim p(x|\theta^*, M^*)$
  4. Accept (M^*, θ^*) if $D(s(x^*), s(x_{\mathrm{obs}})) < \epsilon$

  Ratio of acceptance of M_1 and M_2 approximates Bayes Factor
- ABC-SMC technique (Toni *et al* 2009)
- Calculate evidence of each model separately, using ABC, ABC-MCMC or ABC-SMC (Didelot *et al* 2011)

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):
  1. With probability 0.5 set $M^* = M_1$ otherwise set $M^* = M_2$
  2. Generate $\theta^* \sim p(\theta|M^*)$
  3. Simulate $x^* \sim p(x|\theta^*, M^*)$
  4. Accept (M^*, θ^*) if $D(s(x^*), s(x_{\text{obs}})) < \epsilon$

  Ratio of acceptance of M_1 and M_2 approximates Bayes Factor
- ABC-SMC technique (Toni *et al* 2009)
- Calculate evidence of each model separately, using ABC, ABC-MCMC or ABC-SMC (Didelot *et al* 2011)
- In fact, all these algorithms estimate $\frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)}$

ABC for model selection

- Nested models (Pritchard *et al* 1999)
- ABC-MC algorithm (Grelaud *et al* 2009):
  1. With probability 0.5 set $M^* = M_1$ otherwise set $M^* = M_2$
  2. Generate $\theta^* \sim p(\theta|M^*)$
  3. Simulate $x^* \sim p(x|\theta^*, M^*)$
  4. Accept (M^*, θ^*) if $D(s(x^*), s(x_{\text{obs}})) < \epsilon$

  Ratio of acceptance of M_1 and M_2 approximates Bayes Factor
- ABC-SMC technique (Toni *et al* 2009)
- Calculate evidence of each model separately, using ABC, ABC-MCMC or ABC-SMC (Didelot *et al* 2011)
- In fact, all these algorithms estimate $\frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)}$
- Question: which summary statistic s should be used?

Sufficient statistics

- We have:

$$B_{1,2} = \frac{p(x_{\mathrm{obs}}|M_1)}{p(x_{\mathrm{obs}}|M_2)} = \frac{p(s(x_{\mathrm{obs}})|M_1)}{p(s(x_{\mathrm{obs}})|M_2)} \frac{p(x_{\mathrm{obs}}|s(x_{\mathrm{obs}}), M_1)}{p(x_{\mathrm{obs}}|s(x_{\mathrm{obs}}), M_2)}$$

Sufficient statistics

- We have:

$$B_{1,2} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)} = \frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} \frac{p(x_{\text{obs}}|s(x_{\text{obs}}), M_1)}{p(x_{\text{obs}}|s(x_{\text{obs}}), M_2)}$$

- We say that a summary statistic s is sufficient *for comparing* M_1 and M_2 if and only if:

$$\frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)}$$

Sufficient statistics

- We have:

$$B_{1,2} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)} = \frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} \frac{p(x_{\text{obs}}|s(x_{\text{obs}}), M_1)}{p(x_{\text{obs}}|s(x_{\text{obs}}), M_2)}$$

- We say that a summary statistic s is sufficient *for comparing* M_1 and M_2 if and only if:

$$\frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)}$$

- Sufficiency under a model:

$$p(x|s(x), \theta, M) = p(x|s(x), M)$$

Sufficient statistics

- We have:

$$B_{1,2} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)} = \frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} \frac{p(x_{\text{obs}}|s(x_{\text{obs}}), M_1)}{p(x_{\text{obs}}|s(x_{\text{obs}}), M_2)}$$

- We say that a summary statistic s is sufficient *for comparing* M_1 and M_2 if and only if:

$$\frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)}$$

- Sufficiency under a model:

$$p(x|s(x), \theta, M) = p(x|s(x), M)$$

- Grelaud *et al* (2009): the concatenation of the sufficient statistic of two Gibbs Random models is sufficient *for comparing* them

Sufficient statistics

- We have:

$$B_{1,2} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)} = \frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} \frac{p(x_{\text{obs}}|s(x_{\text{obs}}), M_1)}{p(x_{\text{obs}}|s(x_{\text{obs}}), M_2)}$$

- We say that a summary statistic s is sufficient *for comparing* M_1 and M_2 if and only if:

$$\frac{p(s(x_{\text{obs}})|M_1)}{p(s(x_{\text{obs}})|M_2)} = \frac{p(x_{\text{obs}}|M_1)}{p(x_{\text{obs}}|M_2)}$$

- Sufficiency under a model:

$$p(x|s(x), \theta, M) = p(x|s(x), M)$$

- Grelaud *et al* (2009): the concatenation of the sufficient statistic of two Gibbs Random models is sufficient *for comparing* them
- But as they point out, this does not hold for all types of models!

Poisson/Geometric example

- Data: $x_{\rm obs} = \{x_i\}_{i=1..n}$

Poisson/Geometric example

- Data: $x_{\text{obs}} = \{x_i\}_{i=1..n}$
- Model M_1: $x_i \sim \text{Poisson}(\lambda)$ with $\lambda \sim \text{Exp}(1)$

Poisson/Geometric example

- Data: $x_{\mathrm{obs}} = \{x_i\}_{i=1..n}$
- Model M_1: $x_i \sim \mathrm{Poisson}(\lambda)$ with $\lambda \sim \mathrm{Exp}(1)$
- Model M_2: $x_i \sim \mathrm{Geom}(\mu)$ with $\mu \sim \mathrm{Unif}([0, 1])$

Poisson/Geometric example

- Data: $x_{\mathrm{obs}} = \{x_i\}_{i=1..n}$
- Model M_1: $x_i \sim \mathrm{Poisson}(\lambda)$ with $\lambda \sim \mathrm{Exp}(1)$
- Model M_2: $x_i \sim \mathrm{Geom}(\mu)$ with $\mu \sim \mathrm{Unif}([0,1])$
- The Bayes Factor can be calculated analytically

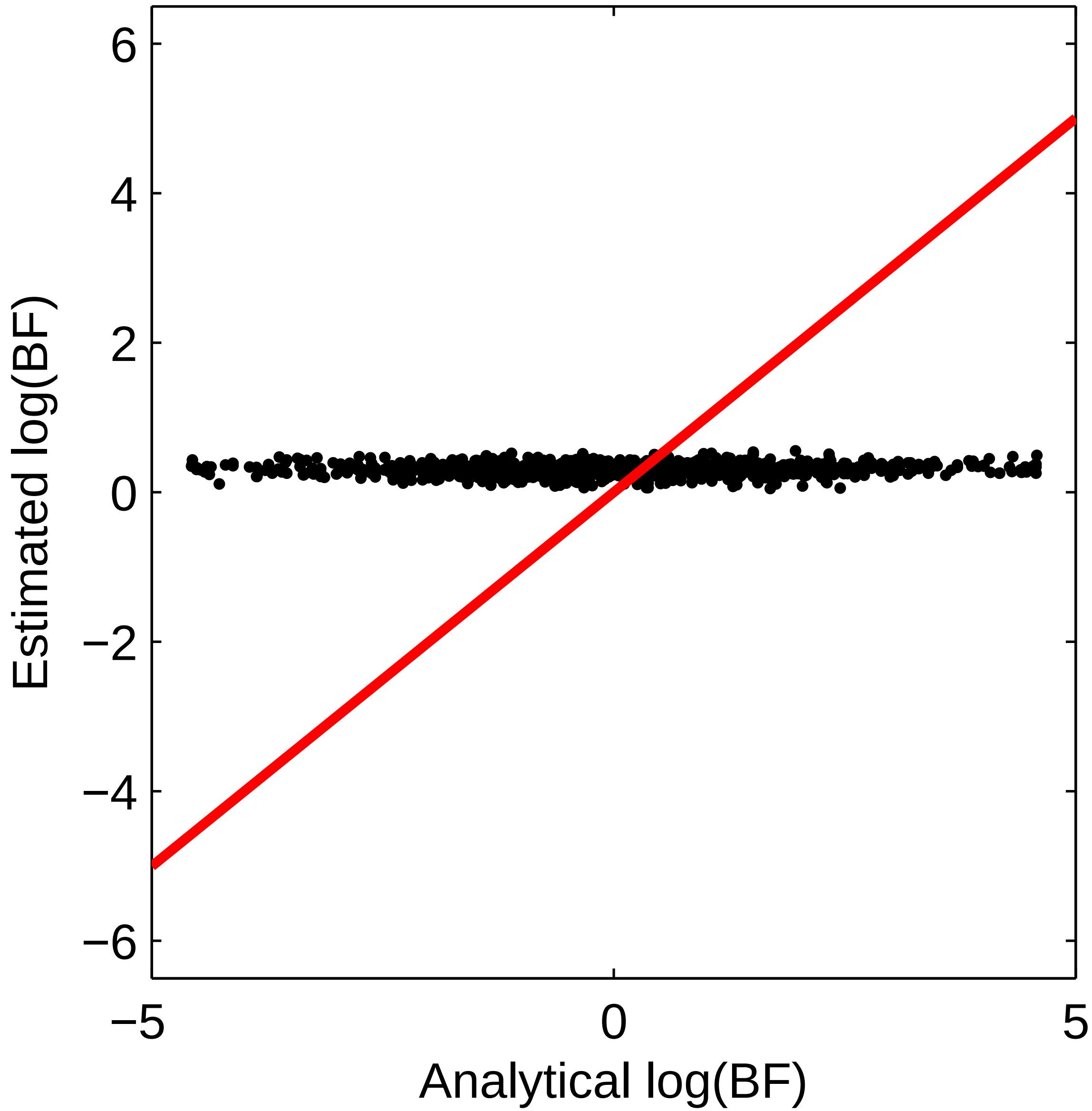
Poisson/Geometric example

- Data: $x_{\text{obs}} = \{x_i\}_{i=1..n}$
- Model M_1: $x_i \sim \text{Poisson}(\lambda)$ with $\lambda \sim \text{Exp}(1)$
- Model M_2: $x_i \sim \text{Geom}(\mu)$ with $\mu \sim \text{Unif}([0,1])$
- The Bayes Factor can be calculated analytically
- We have $s_1 = \sum_{i=1}^{n} x_i$ sufficient for both M_1 and M_2

Poisson/Geometric example

- Data: $x_{\text{obs}} = \{x_i\}_{i=1..n}$
- Model M_1: $x_i \sim \text{Poisson}(\lambda)$ with $\lambda \sim \text{Exp}(1)$
- Model M_2: $x_i \sim \text{Geom}(\mu)$ with $\mu \sim \text{Unif}([0,1])$
- The Bayes Factor can be calculated analytically
- We have $s_1 = \sum_{i=1}^{n} x_i$ sufficient for both M_1 and M_2
- But s_1 is not sufficient *for comparing* M_1 and M_2

Example using s_1



Theorem

If M_1 and M_2 are both nested in model M, then any summary statistic sufficient for model M is sufficient *for comparing* M_1 and M_2

Proof

$$p(x|M_1) = \int_\theta p(x|\theta, M_1)p(\theta|M_1)d\theta$$

Proof

$$
\begin{aligned}
p(x|M_1) &= \int_\theta p(x|\theta, M_1) p(\theta|M_1) d\theta \\
&= \int_\theta p(x|\theta, M) p(\theta|M_1) d\theta
\end{aligned}
$$

Proof

$$
\begin{aligned}
p(x|M_1) &= \int_\theta p(x|\theta, M_1)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|\theta, M)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|s(x), \theta, M)p(s(x)|\theta, M)p(\theta|M_1)d\theta
\end{aligned}
$$

Proof

$$
\begin{aligned}
p(x|M_1) &= \int_\theta p(x|\theta, M_1)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|\theta, M)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|s(x), \theta, M)p(s(x)|\theta, M)p(\theta|M_1)d\theta \\
&= p(x|s(x), M) \int_\theta p(s(x)|\theta, M_1)p(\theta|M_1)d\theta
\end{aligned}
$$

Proof

$$\begin{aligned}
p(x|M_1) &= \int_\theta p(x|\theta, M_1)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|\theta, M)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|s(x), \theta, M)p(s(x)|\theta, M)p(\theta|M_1)d\theta \\
&= p(x|s(x), M) \int_\theta p(s(x)|\theta, M_1)p(\theta|M_1)d\theta \\
&= p(x|s(x), M)p(s(x)|M_1)
\end{aligned}$$

Proof

$$
\begin{aligned}
p(x|M_1) &= \int_\theta p(x|\theta, M_1)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|\theta, M)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|s(x), \theta, M)p(s(x)|\theta, M)p(\theta|M_1)d\theta \\
&= p(x|s(x), M) \int_\theta p(s(x)|\theta, M_1)p(\theta|M_1)d\theta \\
&= p(x|s(x), M)p(s(x)|M_1)
\end{aligned}
$$

Similarly $p(x|M_2) = p(x|s(x), M)p(s(x)|M_2)$ and therefore:

Proof

$$
\begin{aligned}
p(x|M_1) &= \int_\theta p(x|\theta, M_1)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|\theta, M)p(\theta|M_1)d\theta \\
&= \int_\theta p(x|s(x), \theta, M)p(s(x)|\theta, M)p(\theta|M_1)d\theta \\
&= p(x|s(x), M) \int_\theta p(s(x)|\theta, M_1)p(\theta|M_1)d\theta \\
&= p(x|s(x), M)p(s(x)|M_1)
\end{aligned}
$$

Similarly $p(x|M_2) = p(x|s(x), M)p(s(x)|M_2)$ and therefore:

$$
\frac{p(x|M_1)}{p(x|M_2)} = \frac{p(x|s(x), M)p(s(x)|M_1)}{p(x|s(x), M)p(s(x)|M_2)} = \frac{p(s(x)|M_1)}{p(s(x)|M_2)}
$$

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2) \end{aligned}$$

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2\right)$$

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

$$\begin{aligned} p(x|M_1, \theta_1) &\propto \exp(s_1(x)\theta_1) \\ p(x|M_2, \theta_2) &\propto \exp(s_2(x)\theta_2) \end{aligned}$$

- Let M be the model such that:

$$p(x|M, \theta_1, \theta_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0$

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

$$p(x|M_1, \theta_1) \quad \propto \quad \exp(s_1(x)\theta_1)$$
$$p(x|M_2, \theta_2) \quad \propto \quad \exp(s_2(x)\theta_2)$$

- Let M be the model such that:

$$p(x|M, \theta_1, \theta_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0$
- M reduces to M_2 if we take $\theta_1 = 0$

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

$$\begin{aligned} p(x|M_1, \theta_1) &\propto \exp(s_1(x)\theta_1) \\ p(x|M_2, \theta_2) &\propto \exp(s_2(x)\theta_2) \end{aligned}$$

- Let M be the model such that:

$$p(x|M, \theta_1, \theta_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0$
- M reduces to M_2 if we take $\theta_1 = 0$
- Both M_1 and M_2 are nested within M

Gibbs Random Fields

- If M_1 and M_2 are members of the GRF family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0$
- M reduces to M_2 if we take $\theta_1 = 0$
- Both M_1 and M_2 are nested within M
- $[s_1(x), s_2(x)]$ is sufficient for M and therefore sufficient *for comparing* M_1 and M_2.

Exponential family

- If M_1 and M_2 are members of the exponential family:

Exponential family

- If M_1 and M_2 are members of the exponential family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1 + t_1(x)) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2 + t_2(x)) \end{aligned}$$

Exponential family

- If M_1 and M_2 are members of the exponential family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1 + t_1(x)) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2 + t_2(x)) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2 + t_1(x)\alpha_1 + t_2(x)\alpha_2\right)$$

Exponential family

- If M_1 and M_2 are members of the exponential family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1 + t_1(x)) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2 + t_2(x)) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2 + t_1(x)\alpha_1 + t_2(x)\alpha_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0, \alpha_1 = 1, \alpha_2 = 0$

Exponential family

- If M_1 and M_2 are members of the exponential family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1 + t_1(x)) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2 + t_2(x)) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2 + t_1(x)\alpha_1 + t_2(x)\alpha_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0, \alpha_1 = 1, \alpha_2 = 0$
- M reduces to M_2 if we take $\theta_1 = 0, \alpha_1 = 0, \alpha_2 = 1$

Exponential family

- If M_1 and M_2 are members of the exponential family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1 + t_1(x)) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2 + t_2(x)) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2 + t_1(x)\alpha_1 + t_2(x)\alpha_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0, \alpha_1 = 1, \alpha_2 = 0$
- M reduces to M_2 if we take $\theta_1 = 0, \alpha_1 = 0, \alpha_2 = 1$
- Both M_1 and M_2 are nested within M

Exponential family

- If M_1 and M_2 are members of the exponential family:

$$\begin{aligned} p(x|M_1,\theta_1) &\propto \exp(s_1(x)\theta_1 + t_1(x)) \\ p(x|M_2,\theta_2) &\propto \exp(s_2(x)\theta_2 + t_2(x)) \end{aligned}$$

- Let M be the model such that:

$$p(x|M,\theta_1,\theta_2,\alpha_1,\alpha_2) \propto \exp\left(s_1(x)\theta_1 + s_2(x)\theta_2 + t_1(x)\alpha_1 + t_2(x)\alpha_2\right)$$

- M reduces to M_1 if we take $\theta_2 = 0, \alpha_1 = 1, \alpha_2 = 0$
- M reduces to M_2 if we take $\theta_1 = 0, \alpha_1 = 0, \alpha_2 = 1$
- Both M_1 and M_2 are nested within M
- M is an exponential family model for which $[s_1(x), s_2(x), t_1(x), t_2(x)]$ is sufficient.

Poisson/Geometric example

- Canonical forms of M_1 and M_2:

Poisson/Geometric example

- Canonical forms of M_1 and M_2:

$$p(x|\theta_1, M_1) \propto \exp\left(\sum_{i=1}^{n} x_i\theta_1 - \sum_{i=1}^{n} \log x_i!\right)$$

$$p(x|\theta_2, M_2) \propto \exp\left(\sum_{i=1}^{n} x_i\theta_2\right)$$

Poisson/Geometric example

- Canonical forms of M_1 and M_2:

$$
\begin{aligned}
p(x|\theta_1, M_1) &\propto \exp\left(\sum_{i=1}^{n} x_i\theta_1 - \sum_{i=1}^{n} \log x_i!\right) \\
p(x|\theta_2, M_2) &\propto \exp\left(\sum_{i=1}^{n} x_i\theta_2\right)
\end{aligned}
$$

- We can incorporate both in a model M of the form:

$$
p(x|\theta, \alpha, M) \propto \exp\left((\theta_1 + \theta_2)\sum_{i=1}^{n} x_i + \alpha \sum_{i=1}^{n} \log x_i!\right)
$$

Poisson/Geometric example

- Canonical forms of M_1 and M_2:

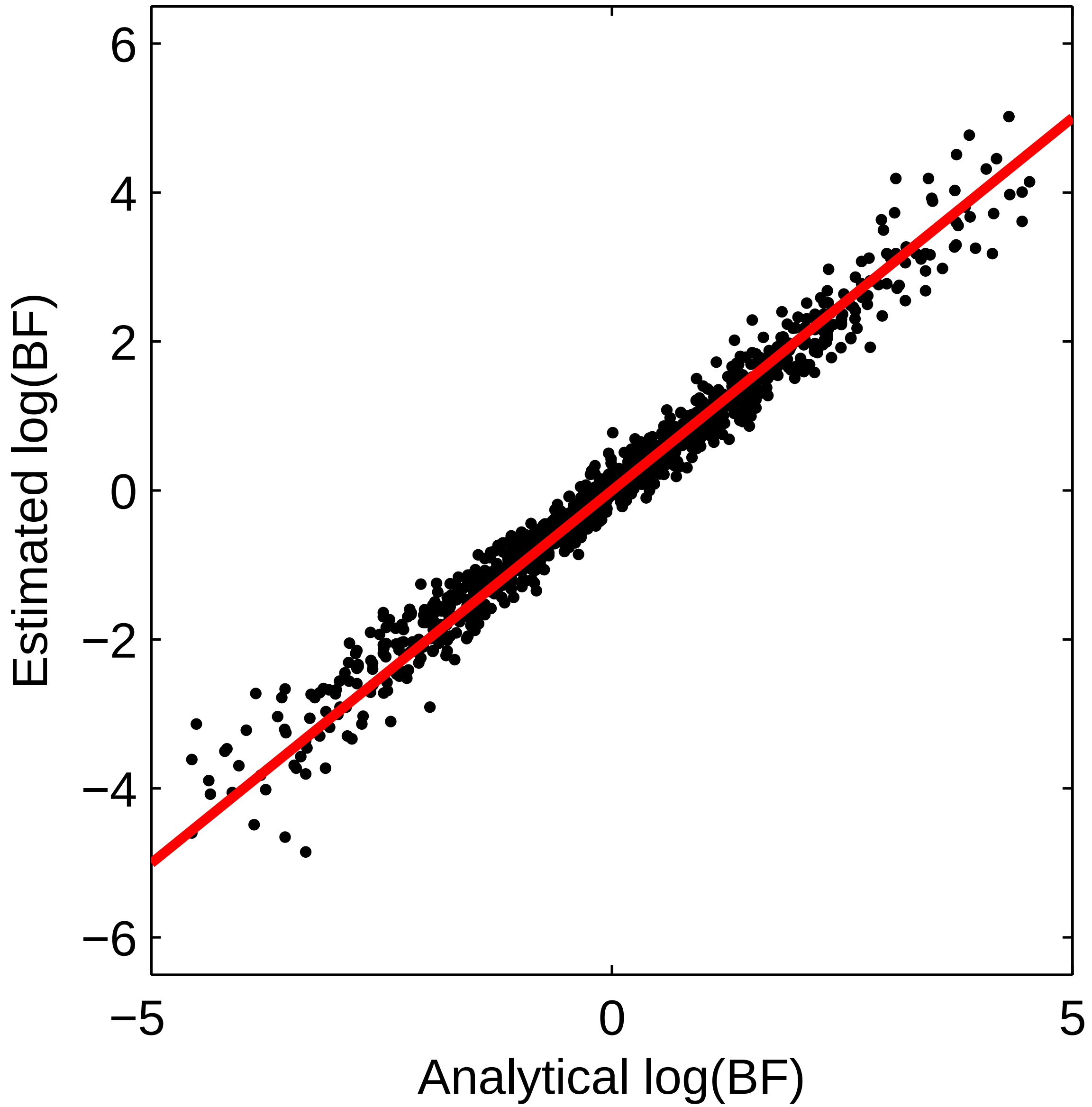
$$\begin{aligned} p(x|\theta_1, M_1) &\propto \exp\left(\sum_{i=1}^{n} x_i\theta_1 - \sum_{i=1}^{n} \log x_i!\right) \\ p(x|\theta_2, M_2) &\propto \exp\left(\sum_{i=1}^{n} x_i\theta_2\right) \end{aligned}$$

- We can incorporate both in a model M of the form:

$$p(x|\theta, \alpha, M) \propto \exp\left((\theta_1 + \theta_2)\sum_{i=1}^{n} x_i + \alpha \sum_{i=1}^{n} \log x_i!\right)$$

- $(s_1, s_2) = (\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} \log x_i!)$ is sufficient for model M and therefore sufficient *for comparing* M_1 and M_2.

Example using (s_1, s_2)



Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection
- But which summary statistic should be used?

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection
- But which summary statistic should be used?
- We have described a method for constructing a summary statistic sufficient for comparing models

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection
- But which summary statistic should be used?
- We have described a method for constructing a summary statistic sufficient for comparing models
- This works well for models of the exponential family

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection
- But which summary statistic should be used?
- We have described a method for constructing a summary statistic sufficient for comparing models
- This works well for models of the exponential family
- For more complex models this approach is often unfeasible, and non-sufficient statistics need to be used that are thought to be informative about the model selection problem (eg Pritchard *et al* 1999)

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection
- But which summary statistic should be used?
- We have described a method for constructing a summary statistic sufficient for comparing models
- This works well for models of the exponential family
- For more complex models this approach is often unfeasible, and non-sufficient statistics need to be used that are thought to be informative about the model selection problem (eg Pritchard *et al* 1999)
- This is analogous to to the necessity to use non-sufficient statistic in standard ABC when complex models are involved

Conclusions

- There are many applications where the likelihood is not computable in a reasonable amount of time
- Interest in using ABC to perform Bayesian model selection
- But which summary statistic should be used?
- We have described a method for constructing a summary statistic sufficient for comparing models
- This works well for models of the exponential family
- For more complex models this approach is often unfeasible, and non-sufficient statistics need to be used that are thought to be informative about the model selection problem (eg Pritchard *et al* 1999)
- This is analogous to to the necessity to use non-sufficient statistic in standard ABC when complex models are involved
- Recent advances in this setting (cf previous talks!) may be useful guides to choosing statistics for the model selection problem