An Overview of Structural Model Uncertainty

Jeremy Oakley

Department of Probability and Statistics, University of Sheffield

25/03/2009
The world of structural model uncertainty

- model averaging
- deterministic model
- statistical model
- model discrepancy
- value of information
- reification
- decision-making without data
- data-driven model structure
What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$

Probabilistic Sensitivity Analysis (PSA): sample $x_1, \ldots, x_n$ from $p(X)$, evaluate $\eta(x_1), \ldots, \eta(x_n)$ to get sample from $p(Y)$. Quantifies uncertainty about $Y$, not $Y^*$. 

Some authors include uncertainty about $p(X)$. 

What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$
- $X$: true values of the model inputs, also observable, with $Y = \eta(X)$
What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$
- $X$: true values of the model inputs, also observable, with $Y = \eta(X)$
- Model structure error if $Y = \eta(X) \neq Y^*$

Probabilistic Sensitivity Analysis (PSA): sample $x_1, \ldots, x_n$ from $p(X)$, evaluate $\eta(x_1), \ldots, \eta(x_n)$ to get sample from $p(Y)$

Quantifies uncertainty about $Y$, not $Y^*$
What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$
- $X$: true values of the model inputs, also observable, with $Y = \eta(X)$
- Model structure error if $Y = \eta(X) \neq Y^*$
- Structural model uncertainty: uncertainty about $Y^*$ given uncertainty about $\eta(\cdot)$
What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$
- $X$: true values of the model inputs, also observable, with $Y = \eta(X)$
- Model structure error if $Y = \eta(X) \neq Y^*$
- Structural model uncertainty: uncertainty about $Y^*$ given uncertainty about $\eta(.)$
  - Some authors include uncertainty about $p(X)$
What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$
- $X$: true values of the model inputs, also observable, with $Y = \eta(X)$
- Model structure error if $Y = \eta(X) \neq Y^*$
- Structural model uncertainty: uncertainty about $Y^*$ given uncertainty about $\eta(.)$
  - Some authors include uncertainty about $p(X)$
- Probabilistic Sensitivity Analysis (PSA): sample $x_1, \ldots, x_n$ from $p(X)$, evaluate $\eta(x_1), \ldots, \eta(x_n)$ to get sample from $p(Y)$
What is structural model uncertainty?

- Deterministic model $y = \eta(x)$, designed to predict observable quantity $Y^*$
- $X$: true values of the model inputs, also observable, with $Y = \eta(X)$
- Model structure error if $Y = \eta(X) \neq Y^*$
- Structural model uncertainty: uncertainty about $Y^*$ given uncertainty about $\eta(.)$
  - Some authors include uncertainty about $p(X)$
- Probabilistic Sensitivity Analysis (PSA): sample $x_1, \ldots, x_n$ from $p(X)$, evaluate $\eta(x_1), \ldots, \eta(x_n)$ to get sample from $p(Y)$
  - Quantifies uncertainty about $Y$, not $Y^*$
Structural model uncertainty: an example

![Diagram showing the states of Respond, Stable, Progressive, and Death.]
From Bernardo & Smith (1994). We have set of models \( \{ M_i, i \in I \} \), with \( M_i = \{ \eta_i(x_{(i)}), p_i(X_{(i)}) \} \).
From Bernardo & Smith (1994). We have set of models \( \{ M_i, i \in I \} \), with \( M_i = \{ \eta_i(x(i)), p_i(X(i)) \} \).

1. The \( \mathcal{M} \) – closed view:
   One of the models in \( \{ M_i, i \in I \} \) is “true”.

No data: an expert weighting problem?
Suitable data: (Bayesian) model averaging

The \( \mathcal{M} \) – open view:
None of the models in \( \{ M_i, i \in I \} \) are correct. Not meaningful to consider \( p(M_i) \).

Perspectives on model uncertainty

From Bernardo & Smith (1994). We have set of models \( \{M_i, i \in I\} \), with \( M_i = \{\eta_i(x_{(i)}), p_i(X_{(i)})\} \).

The \( M \)-closed view:
- One of the models in \( \{M_i, i \in I\} \) is “true”.

\[
p(Y^*) = \sum_{i \in I} p(Y^* | M_i) p(M_i).
\]
Perspectives on model uncertainty

From Bernardo & Smith (1994). We have set of models \( \{M_i, i \in I\} \), with \( M_i = \{\eta_i(x_{(i)}), p_i(X_{(i)})\} \).

1. The \( M \) – closed view:
   One of the models in \( \{M_i, i \in I\} \) is “true”.

   \[
   p(Y^*) = \sum_{i \in I} p(Y^* | M_i) p(M_i).
   \]

   - No data: an expert weighting problem?
Perspectives on model uncertainty

From Bernardo & Smith (1994). We have set of models \( \{ M_i, i \in I \} \), with \( M_i = \{ \eta_i(x(i)), p_i(X(i)) \} \).

1. The \( M \) – closed view:
   One of the models in \( \{ M_i, i \in I \} \) is “true”.

\[
p(Y^*) = \sum_{i \in I} p(Y^* | M_i) p(M_i).
\]

- No data: an expert weighting problem?
- Suitable data: (Bayesian) model averaging
Perspectives on model uncertainty

From Bernardo & Smith (1994). We have set of models \( \{M_i, i \in I\} \), with 
\[
M_i = \{\eta_i(x_{(i)}), p_i(X_{(i)})\}.
\]

1. **The \( \mathcal{M} \) – closed view:**
   One of the models in \( \{M_i, i \in I\} \) is “true”.
   
   \[
p(Y^*) = \sum_{i \in I} p(Y^* | M_i) p(M_i).
   \]
   - No data: an expert weighting problem?
   - Suitable data: (Bayesian) model averaging

2. **The \( \mathcal{M} \) – open view:**
   None of the models in \( \{M_i, i \in I\} \) are correct. Not meaningful to consider \( p(M_i) \)
Bayesian Model Averaging

- Given data $D$, need a likelihood $p(D|M_i)$
Bayesian Model Averaging

- Given data $D$, need a likelihood $p(D|M_i)$
- Weight models using posterior model probabilities

$$p(Y^*|D) = \sum_{i \in I} p(Y^*|M_i, D)p(M_i|D)$$
Bayesian Model Averaging

- Given data $D$, need a likelihood $p(D|M_i)$
- Weight models using posterior model probabilities

\[
p(Y^*|D) = \sum_{i \in I} p(Y^*|M_i, D)p(M_i|D)
\]

- Different models, or one model with particular prior structure?
Bayesian Model Averaging

- Given data \(D\), need a likelihood \(p(D|M_i)\)
- Weight models using posterior model probabilities

\[
p(Y^*|D) = \sum_{i \in I} p(Y^*|M_i, D)p(M_i|D)
\]

- Different models, or one model with particular prior structure?

\[M_1 : \text{response} = \alpha + \beta \text{age} + \varepsilon,\]
\[M_2 : \text{response} = \alpha + \varepsilon,\]
Bayesian Model Averaging

- Given data $D$, need a likelihood $p(D|M_i)$
- Weight models using posterior model probabilities

$$p(Y^*|D) = \sum_{i \in I} p(Y^*|M_i, D)p(M_i|D)$$

- Different models, or one model with particular prior structure?

  $M_1 :$ \textit{response} = $\alpha + \beta \text{age} + \varepsilon$,
  $M_2 :$ \textit{response} = $\alpha + \varepsilon$

  or just

  $M_0 :$ \textit{response} = $\alpha + \beta \text{age} + \varepsilon$,

  with $p(\beta = 0) \neq 0$?
Bayesian Model Averaging

- $p(D|M_i)$ sensitive to choice of prior
- Long running debate in the Bayesian literature
  - See Jackson et al (2009)
Bayesian Model Averaging

- $p(D|M_i)$ sensitive to choice of prior
- Long running debate in the Bayesian literature
  - See Jackson et al (2009)
- Proper prior specification hard (impossible?)

\[
M_1 : \quad \text{log costs} | \mu, \sigma^2 \sim N(\mu, \sigma^2) \\
M_2 : \quad \text{costs} | \alpha, \beta \sim \text{Gamma}(\alpha, \beta),
\]
Bayesian Model Averaging

- $p(D|M_i)$ sensitive to choice of prior
- Long running debate in the Bayesian literature
  - See Jackson et al (2009)
- Proper prior specification hard (impossible?)

\[ M_1 : \text{log costs}|\mu, \sigma^2 \sim N(\mu, \sigma^2) \]
\[ M_2 : \text{costs}|\alpha, \beta \sim \text{Gamma}(\alpha, \beta), \]

Bayesian Model Averaging

- $p(D|M_i)$ sensitive to choice of prior
- Long running debate in the Bayesian literature
  - See Jackson et al (2009)
- Proper prior specification hard (impossible?)

$$
M_1 : \text{log costs}|\mu, \sigma^2 \sim N(\mu, \sigma^2)
$$

$$
M_2 : \text{costs}|\alpha, \beta \sim Gamma(\alpha, \beta),
$$

Bojke et al (2006) propose explicitly parameterising model structure uncertainty
- Can consider value of reducing model structure uncertainty with EVPI
Bayesian Model Averaging

- \( p(D|M_i) \) sensitive to choice of prior
- Long running debate in the Bayesian literature
  - See Jackson et al (2009)
- Proper prior specification hard (impossible?)

\[
\begin{align*}
M_1 & : \log \text{costs} | \mu, \sigma^2 \sim N(\mu, \sigma^2) \\
M_2 & : \text{costs} | \alpha, \beta \sim \text{Gamma}(\alpha, \beta),
\end{align*}
\]

  - Can consider value of reducing model structure uncertainty with EVPI
  - But statistical formulation equivalent to model averaging (with associated pitfalls)?
In the computer experiments literature, we consider *model discrepancy* (e.g. Kennedy and O’Hagan, 2001)

\[ Y^* = \eta(X) + \delta. \]
In the computer experiments literature, we consider *model discrepancy* (e.g. Kennedy and O’Hagan, 2001)

\[ Y^* = \eta(X) + \delta. \]

Can we (usefully) specify \( p(\delta) \)?
In the computer experiments literature, we consider *model discrepancy* (e.g. Kennedy and O’Hagan, 2001)

\[ Y^* = \eta(X) + \delta. \]

Can we (usefully) specify \( p(\delta) \)?

Yes, but need suitable data. Example: observations of treatment outcomes at times \( t = 1, 2 \), wish to predict outcomes at times \( t = 3, 4, \ldots \)

\[ Y^*(t) = \eta(X, t) + \delta(t). \]
In the computer experiments literature, we consider *model discrepancy* (e.g. Kennedy and O’Hagan, 2001)

\[ Y^* = \eta(X) + \delta. \]

Can we (usefully) specify \( p(\delta) \)?

Yes, but need suitable data. Example: observations of treatment outcomes at times \( t = 1, 2 \), wish to predict outcomes at times \( t = 3, 4, \ldots \)

\[ Y^*(t) = \eta(X, t) + \delta(t). \]

Goldstein and Rougier (2009) propose *reified modelling* for physical systems
In the computer experiments literature, we consider *model discrepancy* (e.g. Kennedy and O’Hagan, 2001)

\[ Y^* = \eta(X) + \delta. \]

Can we (usefully) specify \( p(\delta) \)?

Yes, but need suitable data. Example: observations of treatment outcomes at times \( t = 1, 2 \), wish to predict outcomes at times \( t = 3, 4, \ldots \)

\[ Y^*(t) = \eta(X, t) + \delta(t). \]

Goldstein and Rougier (2009) propose *reified modelling* for physical systems

- Involves notion of model discrepancy, potential for dealing with multiple (conflicting) models.
1. $\mathcal{M}$ – closed view
   - Act as if one of the models is true
Summary

1. $\mathcal{M}$ – closed view
   - Act as if one of the models is true
   - Model averaging type methods
$\mathcal{M}$ – *closed* view

- Act as if one of the models is true
- Model averaging type methods
- Useful, but cannot *fully* account for structural model uncertainty
1. $M$ – *closed* view
   - Act as if one of the models is true
   - Model averaging type methods
   - Useful, but cannot *fully* account for structural model uncertainty

2. $M$ – *open* view
   - Acknowledges that none of the models are true
   - Methods developed in computer experiments literature
1. $M$ – *closed* view
   - Act as if one of the models is true
   - Model averaging type methods
   - Useful, but cannot *fully* account for structural model uncertainty

2. $M$ – *open* view
   - Acknowledges that none of the models are true
   - Methods developed in computer experiments literature
   - Can *fully* account for structural model uncertainty, even with only one model...
Summary

1. \( M - \text{closed view} \)
   - Act as if one of the models is true
   - Model averaging type methods
   - Useful, but cannot \textit{fully} account for structural model uncertainty

2. \( M - \text{open view} \)
   - Acknowledges that none of the models are true
   - Methods developed in computer experiments literature
   - Can \textit{fully} account for structural model uncertainty, even with only one model...
   - ...probably less practical here, given data requirements
References