Bayesian Model Specification (Or At Least Some of What Can Be Said About This Topic in 25 Minutes)

David Draper

Department of Applied Mathematics and Statistics
University of California, Santa Cruz

draper@ucsc.edu

OPRE: Bayesian Methods for Social Policy Research and Evaluation

19 Oct 2017

An Example

Draper et al. (1993) reports the results of a **meta-analysis** of $k = 6$ **randomized controlled trials** comparing the use of **low-dose aspirin** (treatment (T)) and placebo (control (C)) by patients following one or more **heart attacks**.

An Example

Draper et al. (1993) reports the results of a **meta-analysis** of $k = 6$ **randomized controlled trials** comparing the use of **low-dose aspirin** (treatment (T)) and placebo (control (C)) by patients following one or more **heart attacks**.

Sample sizes in the table below are (n_i^T, n_i^C), and the **one-year all-cause mortality rates** (%) are $(\hat{p}_i^T, \hat{p}_i^C)$.

An Example

Draper et al. (1993) reports the results of a **meta-analysis** of $k = 6$ **randomized controlled trials** comparing the use of **low-dose aspirin** (treatment (T)) and placebo (control (C)) by patients following one or more **heart attacks**.

Sample sizes in the table below are (n_i^T, n_i^C), and the **one-year all-cause mortality rates** (%) are $(\hat{p}_i^T, \hat{p}_i^C)$.

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	**8.30**	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	**14.82**	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

An Example

Draper et al. (1993) reports the results of a **meta-analysis** of $k = 6$ **randomized controlled trials** comparing the use of **low-dose aspirin** (treatment (T)) and placebo (control (C)) by patients following one or more **heart attacks**.

Sample sizes in the table below are (n_i^T, n_i^C), and the **one-year all-cause mortality rates** (%) are $(\hat{p}_i^T, \hat{p}_i^C)$.

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	**8.30**	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	**14.82**	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

Treatment protocols were similar in the 6 studies, but the patient cohorts exhibited **(strong) between-study heterogeneity**.

An Example

Draper et al. (1993) reports the results of a **meta-analysis** of $k = 6$ **randomized controlled trials** comparing the use of **low-dose aspirin** (treatment (T)) and placebo (control (C)) by patients following one or more **heart attacks**.

Sample sizes in the table below are (n_i^T, n_i^C), and the **one-year all-cause mortality rates** (%) are $(\hat{p}_i^T, \hat{p}_i^C)$.

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	**8.30**	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	**14.82**	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

Treatment protocols were similar in the 6 studies, but the patient cohorts exhibited **(strong) between-study heterogeneity**.

$\mathbb{Q}$: Estimate the **effect** θ of **aspirin** for **similar future patients**.

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest;

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered,

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered, and $\mathbb{C}$ represents the (real-world) **context** in which the questions are raised; let Θ be the set of possible values of $\boldsymbol{\theta}$.

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered, and $\mathbb{C}$ represents the (real-world) **context** in which the questions are raised; let Θ be the set of possible values of $\boldsymbol{\theta}$.

You examine Your resources and find that it's possible to obtain a new **data set** $\boldsymbol{D}$ to decrease Your uncertainty about $\boldsymbol{\theta}$.

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered, and $\mathbb{C}$ represents the (real-world) **context** in which the questions are raised; let Θ be the set of possible values of $\boldsymbol{\theta}$.

You examine Your resources and find that it's possible to obtain a new **data set** $\boldsymbol{D}$ to decrease Your uncertainty about $\boldsymbol{\theta}$.

Since context is crucial to **accurate uncertainty quantification**, let $\mathcal{B} = (B_1, \ldots, B_b)$ be a finite set of (true/false) propositions, all rendered true by the **design of the data-gathering process** and $\mathbb{C}$.

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered, and $\mathbb{C}$ represents the (real-world) **context** in which the questions are raised; let Θ be the set of possible values of $\boldsymbol{\theta}$.

You examine Your resources and find that it's possible to obtain a new **data set** $\boldsymbol{D}$ to decrease Your uncertainty about $\boldsymbol{\theta}$.

Since context is crucial to **accurate uncertainty quantification**, let $\mathcal{B} = (B_1, \ldots, B_b)$ be a finite set of (true/false) propositions, all rendered true by the **design of the data-gathering process** and $\mathbb{C}$.

The Bayesian paradigm requires You to formulate a **probabilistic model** $\mathbb{M}$ that relates **knowns** $(\boldsymbol{D}, \mathcal{B})$ to **unknowns** (inference: $\boldsymbol{\theta}$, and prediction: new data $\boldsymbol{D}^*$).

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered, and $\mathbb{C}$ represents the (real-world) **context** in which the questions are raised; let Θ be the set of possible values of $\boldsymbol{\theta}$.

You examine Your resources and find that it's possible to obtain a new **data set $\boldsymbol{D}$** to decrease Your uncertainty about $\boldsymbol{\theta}$.

Since context is crucial to **accurate uncertainty quantification**, let $\mathcal{B} = (B_1, \ldots, B_b)$ be a finite set of (true/false) propositions, all rendered true by the **design of the data-gathering process** and $\mathbb{C}$.

The Bayesian paradigm requires You to formulate a **probabilistic model** $\mathbb{M}$ that relates **knowns** $(\boldsymbol{D}, \mathcal{B})$ to **unknowns** (inference: $\boldsymbol{\theta}$, and prediction: new data $\boldsymbol{D}^*$).

There's a **Theorem** (de Finetti, RT Cox) that says that, for **optimal information-processing**, this model must be of the form

A General Framework for Inference and Prediction

You're given a **problem** $\mathbb{P} = (\mathbb{Q}, \mathbb{C})$ involving **uncertainty** about $\boldsymbol{\theta}$, the unknown aspect of $\mathbb{P}$ of principal interest; here $\mathbb{Q}$ identifies the main **questions** to be answered, and $\mathbb{C}$ represents the (real-world) **context** in which the questions are raised; let Θ be the set of possible values of $\boldsymbol{\theta}$.

You examine Your resources and find that it's possible to obtain a new **data set** $\boldsymbol{D}$ to decrease Your uncertainty about $\boldsymbol{\theta}$.

Since context is crucial to **accurate uncertainty quantification**, let $\mathcal{B} = (B_1, \ldots, B_b)$ be a finite set of (true/false) propositions, all rendered true by the **design of the data-gathering process** and $\mathbb{C}$.

The Bayesian paradigm requires You to formulate a **probabilistic model** $\mathbb{M}$ that relates **knowns** $(\boldsymbol{D}, \mathcal{B})$ to **unknowns** (inference: $\boldsymbol{\theta}$, and prediction: new data $\boldsymbol{D}^*$).

There's a **Theorem** (de Finetti, RT Cox) that says that, for **optimal information-processing**, this model must be of the form

$$\mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta} \, \mathcal{B}) \} \, .$$

Bayesian Inference and Prediction

$$\mathbb{M} = \{p(\boldsymbol{\theta} \mid \mathcal{B}), p(\boldsymbol{D} \mid \boldsymbol{\theta}\, \mathcal{B})\}\,.$$

Bayesian Inference and Prediction

$$\mathbb{M} = \{p(\boldsymbol{\theta} \mid \mathcal{B}), p(\boldsymbol{D} \mid \boldsymbol{\theta}\, \mathcal{B})\}\,.$$

$\mathbb{M}$ consists of **two probability distributions**:

Bayesian Inference and Prediction

$$\mathbb{M} = \{p(\boldsymbol{\theta} \mid \mathcal{B}), p(\boldsymbol{D} \mid \boldsymbol{\theta}\, \mathcal{B})\}\,.$$

$\mathbb{M}$ consists of **two probability distributions**:

- $p(\boldsymbol{\theta} \mid \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **external** to $\boldsymbol{D}$ (this is usually referred to as Your **prior distribution)**, and

Bayesian Inference and Prediction

$$\mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B}) \} \, .$$

$\mathbb{M}$ consists of **two probability distributions**:

- $p(\boldsymbol{\theta} \,|\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **external** to $\boldsymbol{D}$ (this is usually referred to as Your **prior distribution)**, and

- $p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **internal** to $\boldsymbol{D}$ (this is Your **sampling distribution**), once it has been transformed into Your **likelihood distribution** $\ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \propto p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$.

Bayesian Inference and Prediction

$$\mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B}) \} \,.$$

$\mathbb{M}$ consists of **two probability distributions**:

- $p(\boldsymbol{\theta} \,|\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **external** to $\boldsymbol{D}$ (this is usually referred to as Your **prior distribution)**, and

- $p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **internal** to $\boldsymbol{D}$ (this is Your **sampling distribution**), once it has been transformed into Your **likelihood distribution** $\ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \propto p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$.

Optimal inference — **conditional on $\mathbb{M}$** — now proceeds via Bayes's Theorem:

$$\mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\}\,.$$

$\mathbb{M}$ consists of **two probability distributions**:

- $p(\boldsymbol{\theta} \,|\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **external** to $\boldsymbol{D}$ (this is usually referred to as Your **prior distribution)**, and
- $p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **internal** to $\boldsymbol{D}$ (this is Your **sampling distribution**), once it has been transformed into Your **likelihood distribution** $\ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \propto p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$.

Optimal inference — **conditional on $\mathbb{M}$** — now proceeds via Bayes's Theorem:

$$\begin{array}{ccccc} p(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) & \propto & p(\boldsymbol{\theta} \,|\, \mathcal{B}) & \cdot & \ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \\ \left(\begin{array}{c}\text{posterior}\\ \text{distribution}\end{array}\right) & \propto & \left(\begin{array}{c}\text{prior}\\ \text{distribution}\end{array}\right) & \cdot & \left(\begin{array}{c}\text{likelihood}\\ \text{distribution}\end{array}\right) \end{array} \quad (1)$$

Bayesian Inference and Prediction

$$\mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B}) \} \, .$$

$\mathbb{M}$ consists of **two probability distributions**:

- $p(\boldsymbol{\theta} \,|\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **external** to $\boldsymbol{D}$ (this is usually referred to as Your **prior distribution)**, and
- $p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **internal** to $\boldsymbol{D}$ (this is Your **sampling distribution**), once it has been transformed into Your **likelihood distribution** $\ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \propto p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$.

Optimal inference — **conditional on $\mathbb{M}$** — now proceeds via Bayes's Theorem:

$$\begin{array}{ccccc} p(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) & \propto & p(\boldsymbol{\theta} \,|\, \mathcal{B}) & \cdot & \ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \\ \left(\begin{array}{c} \text{posterior} \\ \text{distribution} \end{array} \right) & \propto & \left(\begin{array}{c} \text{prior} \\ \text{distribution} \end{array} \right) & \cdot & \left(\begin{array}{c} \text{likelihood} \\ \text{distribution} \end{array} \right) \end{array} \qquad (1)$$

and **optimal prediction** of new data $\boldsymbol{D}^*$ — **conditional on $\mathbb{M}$** — is based on

$$\mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B}) \} \, .$$

$\mathbb{M}$ consists of **two probability distributions**:

- $p(\boldsymbol{\theta} \,|\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **external** to $\boldsymbol{D}$ (this is usually referred to as Your **prior distribution)**, and
- $p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$ quantifies Your information about $\boldsymbol{\theta}$ **internal** to $\boldsymbol{D}$ (this is Your **sampling distribution**), once it has been transformed into Your **likelihood distribution** $\ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \propto p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})$.

Optimal inference — **conditional on $\mathbb{M}$** — now proceeds via Bayes's Theorem:

$$\begin{array}{ccccc} p(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) & \propto & p(\boldsymbol{\theta} \,|\, \mathcal{B}) & \cdot & \ell(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \\ \left(\begin{array}{c} \text{posterior} \\ \text{distribution} \end{array} \right) & \propto & \left(\begin{array}{c} \text{prior} \\ \text{distribution} \end{array} \right) & \cdot & \left(\begin{array}{c} \text{likelihood} \\ \text{distribution} \end{array} \right) \end{array} \qquad (1)$$

and **optimal prediction** of new data $\boldsymbol{D}^*$ — **conditional on $\mathbb{M}$** — is based on

$$p(\boldsymbol{D}^* \,|\, \boldsymbol{D}\, \mathcal{B}) = \int_\Theta p(\boldsymbol{D}^* \,|\, \boldsymbol{\theta}\, \boldsymbol{D}\, \mathcal{B}) \, p(\boldsymbol{\theta} \,|\, \boldsymbol{D}\, \mathcal{B}) \, d\boldsymbol{\theta} \, . \qquad (2)$$

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{\textit{unique}}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \tag{3}$$

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{\textit{unique}}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{\textit{unique}}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{\textit{unique}}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \overset{\textit{unique}}{\rightarrow} \mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\}\,. \tag{4}$$

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \stackrel{unique}{\rightarrow} \mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\} \,. \tag{4}$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \overset{unique}{\rightarrow} \mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta} \, \mathcal{B}) \} \,. \tag{4}$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true,

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \overset{unique}{\rightarrow} \mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\}\,. \tag{4}$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true, but **in most problems** $\mathbb{M}$ **is not uniquely specified by context**:

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \overset{unique}{\rightarrow} \mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\}\,. \tag{4}$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true, but **in most problems** $\mathbb{M}$ **is not uniquely specified by context**:

You're **uncertain** about $\boldsymbol{\theta}$, but You're **also uncertain** about

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \,. \qquad (3)$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \stackrel{unique}{\rightarrow} \mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B}) \} \,. \qquad (4)$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true, but **in most problems** $\mathbb{M}$ **is not uniquely specified by context**:

You're **uncertain** about $\boldsymbol{\theta}$, but You're **also uncertain** about

how to specify Your uncertainty about $\boldsymbol{\theta}$ —

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \qquad (3)$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \overset{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \overset{unique}{\rightarrow} \mathbb{M} = \{ p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B}) \}\,. \qquad (4)$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true, but **in most problems** $\mathbb{M}$ **is not uniquely specified by context**:

You're **uncertain** about $\boldsymbol{\theta}$, but You're **also uncertain** about

how to specify Your uncertainty about $\boldsymbol{\theta}$ — this is **model uncertainty**, the **fundamental problem of applied statistics**.

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \tag{3}$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \stackrel{unique}{\rightarrow} \mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\}\,. \tag{4}$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true, but **in most problems** $\mathbb{M}$ **is not uniquely specified by context**:

You're **uncertain** about $\boldsymbol{\theta}$, but You're **also uncertain** about

how to specify Your uncertainty about $\boldsymbol{\theta}$ — this is **model uncertainty**, the **fundamental problem of applied statistics**.

At present we have **relatively few Theorems** telling us uniquely what to do,

The Fundamental Problem of Applied Statistics

So far we have

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B})\,. \qquad (3)$$

It would be **wonderful** if, for all problems $\mathbb{P}$,

$$\mathbb{P} = (\mathbb{Q}, \mathbb{C}) \stackrel{unique}{\rightarrow} (\boldsymbol{\theta}, \boldsymbol{D}, \mathcal{B}) \stackrel{unique}{\rightarrow} \mathbb{M} = \{p(\boldsymbol{\theta} \,|\, \mathcal{B}), p(\boldsymbol{D} \,|\, \boldsymbol{\theta}\, \mathcal{B})\}\,. \qquad (4)$$

In other words, it would be great if the context $\mathbb{C}$ of the problem $\mathbb{P}$ **always uniquely specified a single Bayesian model** $\mathbb{M}$.

There are **some** problems in which this is (partially or completely) true, but **in most problems** $\mathbb{M}$ **is not uniquely specified by context**:

You're **uncertain** about $\boldsymbol{\theta}$, but You're **also uncertain** about

how to specify Your uncertainty about $\boldsymbol{\theta}$ — this is **model uncertainty**, the **fundamental problem of applied statistics**.

At present we have **relatively few Theorems** telling us uniquely what to do, to get the **"best" answers when model uncertainty is present**.

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

These questions implicitly acknowledge the strong possibility of **model uncertainty**, in specifying **both** the prior and the likelihood;

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

These questions implicitly acknowledge the strong possibility of **model uncertainty**, in specifying **both** the prior and the likelihood; question (2) therefore has a **cousin**:

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

These questions implicitly acknowledge the strong possibility of **model uncertainty**, in specifying **both** the prior and the likelihood; question (2) therefore has a **cousin**:

(2′) How do you **test** the extent to which the **likelihood** drives the results ***[in settings in which a unique context-specific likelihood doesn't exist]***?

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

These questions implicitly acknowledge the strong possibility of **model uncertainty**, in specifying **both** the prior and the likelihood; question (2) therefore has a **cousin**:

(2′) How do you **test** the extent to which the **likelihood** drives the results ***[in settings in which a unique context-specific likelihood doesn't exist]***?

In a part of the statistical landscape in which we have **few Theorems** on which to rely, we must content ourselves (at present) with **Principles** and **Axioms**.

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

These questions implicitly acknowledge the strong possibility of **model uncertainty**, in specifying **both** the prior and the likelihood; question (2) therefore has a **cousin**:

(2′) How do you **test** the extent to which the **likelihood** drives the results ***[in settings in which a unique context-specific likelihood doesn't exist]***?

In a part of the statistical landscape in which we have **few Theorems** on which to rely, we must content ourselves (at present) with **Principles** and **Axioms**.

I'll now address the questions above by offering **two Principles and an Axiom**,

Some Model Uncertainty Questions

I was asked to **address** the following (highly relevant) **questions** here:

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

These questions implicitly acknowledge the strong possibility of **model uncertainty**, in specifying **both** the prior and the likelihood; question (2) therefore has a **cousin**:

(2′) How do you **test** the extent to which the **likelihood** drives the results ***[in settings in which a unique context-specific likelihood doesn't exist]***?

In a part of the statistical landscape in which we have **few Theorems** on which to rely, we must content ourselves (at present) with **Principles** and **Axioms**.

I'll now address the questions above by offering **two Principles and an Axiom**, illustrated with an **analysis** of the **aspirin data**.

An Axiom and a Principle

I **begin** with the

An Axiom and a Principle

I **begin** with the

Good-Faith Axiom: You want to **help positively advance** the **course of science**, and **repeatedly getting the wrong answer** runs **counter** to this desire.

An Axiom and a Principle

I **begin** with the

Good-Faith Axiom: You want to **help positively advance** the **course of science**, and **repeatedly getting the wrong answer** runs **counter** to this desire.

Under this Axiom, it's a **basic scientific imperative** for **all statistical methods** (Bayesian or otherwise) to be examined for **calibration**: how often they get the right answer.

An Axiom and a Principle

I **begin** with the

Good-Faith Axiom: You want to **help positively advance** the **course of science**, and **repeatedly getting the wrong answer** runs **counter** to this desire.

Under this Axiom, it's a **basic scientific imperative** for **all statistical methods** (Bayesian or otherwise) to be examined for **calibration**: how often they get the right answer.

This **leads** to the

An Axiom and a Principle

I **begin** with the

Good-Faith Axiom: You want to **help positively advance** the **course of science**, and **repeatedly getting the wrong answer** runs **counter** to this desire.

Under this Axiom, it's a **basic scientific imperative** for **all statistical methods** (Bayesian or otherwise) to be examined for **calibration**: how often they get the right answer.

This **leads** to the

Calibration Principle: In **model specification**, it **helps** to **know something** about **how often** {the **methods** You're **using** to choose one **model** over **another**} **get** the **right answer**, and **this** can be **ascertained** by

An Axiom and a Principle

I **begin** with the

Good-Faith Axiom: You want to **help positively advance** the **course of science**, and **repeatedly getting the wrong answer** runs **counter** to this desire.

Under this Axiom, it's a **basic scientific imperative** for **all statistical methods** (Bayesian or otherwise) to be examined for **calibration**: how often they get the right answer.

This **leads** to the

Calibration Principle: In **model specification**, it **helps** to **know something** about **how often** {the **methods** You're **using** to choose one **model** over **another**} **get** the **right answer**, and **this** can be **ascertained** by

(a) **creating simulation environments** (structurally **similar** to the **setup** of the **scientific problem** You're **currently solving**) in which **You know what the right answer is**, and

An Axiom and a Principle

I **begin** with the

Good-Faith Axiom: You want to **help positively advance** the **course of science**, and **repeatedly getting the wrong answer** runs **counter** to this desire.

Under this Axiom, it's a **basic scientific imperative** for **all statistical methods** (Bayesian or otherwise) to be examined for **calibration**: how often they get the right answer.

This **leads** to the

Calibration Principle: In **model specification**, it **helps** to **know something** about **how often** {the **methods** You're **using** to choose one **model** over **another**} **get** the **right answer**, and **this** can be **ascertained** by

(a) **creating simulation environments** (structurally **similar** to the **setup** of the **scientific problem** You're **currently solving**) in which **You know what the right answer is**, and

(b) seeing **how often Your methods recover** known **truth**.

Modeling the Aspirin Data

Study (i)	Aspirin		Placebo			
	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

Modeling the Aspirin Data

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

In **modeling the aspirin data**, let's begin with a **model** for $\boldsymbol{D} = (y_1, \ldots, y_k)$ with $k = 6$:

Modeling the Aspirin Data

Study (i)	Aspirin		Placebo			
	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

In **modeling the aspirin data**, let's begin with a **model** for $\boldsymbol{D} = (y_1, \ldots, y_k)$ with $k = 6$:

$$(y_i \,|\, \Delta_i\, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i)\,. \tag{5}$$

Modeling the Aspirin Data

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

In **modeling the aspirin data**, let's begin with a **model** for $\boldsymbol{D} = (y_1, \ldots, y_k)$ with $k = 6$:

$$(y_i \mid \Delta_i \, \mathcal{B}) \overset{\text{I}}{\sim} N(\Delta_i, V_i) \,. \tag{5}$$

This is the **likelihood** part of the model:

Modeling the Aspirin Data

Study (i)	Aspirin n_i^T	$\hat{p}_i^T$	Placebo n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

In **modeling the aspirin data**, let's begin with a **model** for $\boldsymbol{D} = (y_1, \ldots, y_k)$ with $k = 6$:

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \qquad (5)$$

This is the **likelihood** part of the model: equation (5) says that **Your uncertainty** about y_i (before it's observed) can be captured with a **Normal distribution** centered at an (unknown) mean of Δ_i and a variance of V_i,

Modeling the Aspirin Data

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

In **modeling the aspirin data**, let's begin with a **model** for $\boldsymbol{D} = (y_1, \ldots, y_k)$ with $k = 6$:

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \tag{5}$$

This is the **likelihood** part of the model: equation (5) says that **Your uncertainty** about y_i (before it's observed) can be captured with a **Normal distribution** centered at an (unknown) mean of Δ_i and a variance of V_i, and the y_i are **conditionally independent** given the Δ_i.

Modeling the Aspirin Data

	Aspirin		Placebo			
Study (i)	n_i^T	$\hat{p}_i^T$	n_i^C	$\hat{p}_i^C$	$y_i = (\hat{p}_i^C - \hat{p}_i^T)$	$\widehat{SE}(y_i) \triangleq \sqrt{V_i}$
UK–1	615	7.97	624	10.74	2.77	1.65
CDPA	758	5.80	771	8.30	2.50	1.31
GAMS	317	8.52	309	10.36	1.84	2.34
UK–2	832	12.26	850	14.82	2.56	1.67
PARIS	810	10.49	406	12.81	2.31	1.98
AMIS	2267	10.85	2257	9.10	**−1.15**	0.90

In **modeling the aspirin data**, let's begin with a **model** for $\boldsymbol{D} = (y_1, \ldots, y_k)$ with $k = 6$:

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \,. \qquad (5)$$

This is the **likelihood** part of the model: equation (5) says that **Your uncertainty** about y_i (before it's observed) can be captured with a **Normal distribution** centered at an (unknown) mean of Δ_i and a variance of V_i, and the y_i are **conditionally independent** given the Δ_i.

Since the **sample sizes** in the experiments are **large**, You can take the V_i to be **known** and equal to the **squared standard errors** in the data table.

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \mathcal{B}) \overset{\mathrm{I}}{\sim} N(\Delta_i, V_i)\,. \qquad (6)$$

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i)\,. \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \overset{\mathrm{I}}{\sim} N(\Delta_i, V_i) \,. \tag{6}$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation:

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \tag{6}$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients,

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i —

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \overset{I}{\sim} N(\Delta_i, V_i) \, . \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i:

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \overset{I}{\sim} N(\Delta_i, V_i) \,. \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i: if there were any **study-level covariates**, we could use them to **explain part of the heterogeneity** (e.g., with a **regression** of Δ_i on the covariates),

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \overset{\mathrm{I}}{\sim} N(\Delta_i, V_i) \, . \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i: if there were any **study-level covariates**, we could use them to **explain part of the heterogeneity** (e.g., with a **regression** of Δ_i on the covariates), but in this meta-analysis **there aren't any** such predictors,

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{I}{\sim} N(\Delta_i, V_i) \, . \tag{6}$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i: if there were any **study-level covariates**, we could use them to **explain part of the heterogeneity** (e.g., with a **regression** of Δ_i on the covariates), but in this meta-analysis **there aren't any** such predictors, so the best we can do is to fit a **regression with just an intercept term**:

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i: if there were any **study-level covariates**, we could use them to **explain part of the heterogeneity** (e.g., with a **regression** of Δ_i on the covariates), but in this meta-analysis **there aren't any** such predictors, so the best we can do is to fit a **regression with just an intercept term**:

$$(\Delta_i \mid \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \, . \qquad (7)$$

Modeling the Aspirin Data (continued)

$$(y_i \,|\, \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \tag{6}$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i: if there were any **study-level covariates**, we could use them to **explain part of the heterogeneity** (e.g., with a **regression** of Δ_i on the covariates), but in this meta-analysis **there aren't any** such predictors, so the best we can do is to fit a **regression with just an intercept term**:

$$(\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \, . \tag{7}$$

θ represents the **mean difference (placebo – aspirin)** in mortality in the **population** of **future patients** similar to those in this meta-analysis:

Modeling the Aspirin Data (continued)

$$(y_i \mid \Delta_i \, \mathcal{B}) \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \, . \qquad (6)$$

We need the Δ_i values to **vary** from one experiment to another, to describe the evident **heterogeneity** in patient cohorts.

The **Normality assumption** in equation (6) is **uniquely specified by problem context** $\mathbb{C}$ in this situation: each y_i is a difference of two independent **means**, each of which is based on a **large number** of patients, so the **Central Limit Theorem** gives us Normality of the y_i — this is an example of **unique likelihood specification**.

Now we need a **model** for the Δ_i: if there were any **study-level covariates**, we could use them to **explain part of the heterogeneity** (e.g., with a **regression** of Δ_i on the covariates), but in this meta-analysis **there aren't any** such predictors, so the best we can do is to fit a **regression with just an intercept term**:

$$(\Delta_i \mid \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \, . \qquad (7)$$

θ represents the **mean difference (placebo – aspirin)** in mortality in the **population** of **future patients** similar to those in this meta-analysis: it's the **key unknown** in this problem $\mathbb{P}$.

Modeling the Aspirin Data (continued)

$$(\Delta_i \mid \theta\, \sigma\, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \ . \tag{8}$$

Modeling the Aspirin Data (continued)

$$(\Delta_i \mid \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\big(\theta, \sigma^2\big) \ . \tag{8}$$

The unknown quantity σ is **also important** in this model:

Modeling the Aspirin Data (continued)

$$(\Delta_i \,|\, \theta\, \sigma\, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\big(\theta, \sigma^2\big) \ . \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

Modeling the Aspirin Data (continued)

$$(\Delta_i \,|\, \theta\, \sigma\, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\big(\theta, \sigma^2\big)\ . \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**;

Modeling the Aspirin Data (continued)

$$(\Delta_i \mid \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\big(\theta, \sigma^2\big) \; . \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

Modeling the Aspirin Data (continued)

$$(\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\big(\theta, \sigma^2\big) \ . \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

We now need to **complete the model** by placing a **prior distribution** on (θ, σ):

Modeling the Aspirin Data (continued)

$$(\Delta_i \mid \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \, . \qquad (8)$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

We now need to **complete the model** by placing a **prior distribution** on (θ, σ): at the time of this meta-analysis, **little was known** about aspirin's effect on mortality for heart-attack patients, so we should use a **low-information-content prior** for θ,

Modeling the Aspirin Data (continued)

$$(\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\big(\theta, \sigma^2\big) \ . \qquad (8)$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

We now need to **complete the model** by placing a **prior distribution** on (θ, σ): at the time of this meta-analysis, **little was known** about aspirin's effect on mortality for heart-attack patients, so we should use a **low-information-content prior** for θ, and the **same type of prior** should also be used for σ for a **similar reason**.

Modeling the Aspirin Data (continued)

$$(\Delta_i \,|\, \theta\,\sigma\,\mathcal{B}) \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \;. \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

We now need to **complete the model** by placing a **prior distribution** on (θ, σ): at the time of this meta-analysis, **little was known** about aspirin's effect on mortality for heart-attack patients, so we should use a **low-information-content prior** for θ, and the **same type of prior** should also be used for σ for a **similar reason**.

The **most general prior** for (θ, σ) would look like

$$p(\theta\,\sigma \,|\, \mathcal{B}) = p(\sigma \,|\, \mathcal{B}) \cdot p(\theta \,|\, \sigma\,\mathcal{B})\,, \tag{9}$$

$$(\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N(\theta, \sigma^2) \, . \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

We now need to **complete the model** by placing a **prior distribution** on (θ, σ): at the time of this meta-analysis, **little was known** about aspirin's effect on mortality for heart-attack patients, so we should use a **low-information-content prior** for θ, and the **same type of prior** should also be used for σ for a **similar reason**.

The **most general prior** for (θ, σ) would look like

$$p(\theta \, \sigma \,|\, \mathcal{B}) = p(\sigma \,|\, \mathcal{B}) \cdot p(\theta \,|\, \sigma \, \mathcal{B}) \, , \tag{9}$$

but it turns out that **without essential loss of generality** we can take θ and σ to be **independent** in the prior:

Modeling the Aspirin Data (continued)

$$(\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N\!\left(\theta, \sigma^2\right) \, . \tag{8}$$

The unknown quantity σ is **also important** in this model: it represents the amount of **between-study heterogeneity** in the Δ_i.

In **contrast** with the **likelihood**, the **Normal distribution** in equation (8) is **NOT uniquely specified by problem context**; it's a **conventional assumption** that should undergo **model criticism**.

We now need to **complete the model** by placing a **prior distribution** on (θ, σ): at the time of this meta-analysis, **little was known** about aspirin's effect on mortality for heart-attack patients, so we should use a **low-information-content prior** for θ, and the **same type of prior** should also be used for σ for a **similar reason**.

The **most general prior** for (θ, σ) would look like

$$p(\theta \, \sigma \,|\, \mathcal{B}) = p(\sigma \,|\, \mathcal{B}) \cdot p(\theta \,|\, \sigma \, \mathcal{B}) \, , \tag{9}$$

but it turns out that **without essential loss of generality** we can take θ and σ to be **independent** in the prior:

$$p(\theta \, \sigma \,|\, \mathcal{B}) = p(\sigma \,|\, \mathcal{B}) \cdot p(\theta \,|\, \mathcal{B}) \, . \tag{10}$$

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**;

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

$$(\theta \,|\, \mathcal{B}) \sim N\!\left(0, \sigma^2_{huge}\right) \ . \tag{11}$$

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

$$(\theta \mid \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) \; . \tag{11}$$

It turns out that **considerably more care** is required in models of this type in **specifying the prior** for σ;

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

$$(\theta \mid \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) \, . \qquad (11)$$

It turns out that **considerably more care** is required in models of this type in **specifying the prior** for σ; Gelman (2006) has shown that a **Uniform distribution** from 0 to C (with C chosen large enough to **avoid unintentional truncation** of the likelihood distribution for σ) produces **well-calibrated inferences**:

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

$$(\theta \,|\, \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) \,. \tag{11}$$

It turns out that **considerably more care** is required in models of this type in **specifying the prior** for σ; Gelman (2006) has shown that a **Uniform distribution** from 0 to C (with C chosen large enough to **avoid unintentional truncation** of the likelihood distribution for σ) produces **well-calibrated inferences**:

$$(\sigma \,|\, \mathcal{B}) \sim U(0, C) \,. \tag{12}$$

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

$$(\theta \mid \mathcal{B}) \sim N\left(0, \sigma_{huge}^2\right) \ . \tag{11}$$

It turns out that **considerably more care** is required in models of this type in **specifying the prior** for σ; Gelman (2006) has shown that a **Uniform distribution** from 0 to C (with C chosen large enough to **avoid unintentional truncation** of the likelihood distribution for σ) produces **well-calibrated inferences**:

$$(\sigma \mid \mathcal{B}) \sim U(0, C) \, . \tag{12}$$

So my **initial model** for the aspirin data looks like this:

Modeling the Aspirin Data (continued)

The idea of a **distribution** encoding a **low level of information** corresponds to that distribution being **highly spread out (diffuse, flat)**; to make this idea operational for θ, **long experience** with models of this type has shown that a **Normal distribution centered at 0** with a **huge standard deviation** σ_{huge} produces **well-calibrated inferences**:

$$(\theta \,|\, \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) \,. \tag{11}$$

It turns out that **considerably more care** is required in models of this type in **specifying the prior** for σ; Gelman (2006) has shown that a **Uniform distribution** from 0 to C (with C chosen large enough to **avoid unintentional truncation** of the likelihood distribution for σ) produces **well-calibrated inferences**:

$$(\sigma \,|\, \mathcal{B}) \sim U(0, C) \,. \tag{12}$$

So my **initial model** for the aspirin data looks like this:

$$\begin{aligned} (\theta \,|\, \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) & \qquad (\sigma \,|\, \mathcal{B}) \sim U(0, C) \\ (\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) & \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \\ (y_i \,|\, \Delta_i \, \mathcal{B}) & \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \,. \end{aligned} \tag{13}$$

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in `R`) in the **aspirin meta-analysis** of the following form:

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in `R`) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**:

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in R) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**: it turns out that, on the scale of percentage points of mortality, $\left(\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}\right) =$ **(1.45, 1.24)**.

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in `R`) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**: it turns out that, on the scale of percentage points of mortality, $\left(\hat{\theta}_{MLE}, \hat{\sigma}_{MLE} \right) =$ **(1.45, 1.24)**.

(2) Specify a (3×3) **sensitivity-analysis grid** by halving and doubling the (θ, σ) values in step (1): $\{ \theta_{DG} \in (0.725, 1.45, 2.90) \} \times \{ \sigma_{DG} \in (0.61, 1.24, 2.48) \}$.

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in R) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**: it turns out that, on the scale of percentage points of mortality, $\left(\hat{\theta}_{MLE}, \hat{\sigma}_{MLE} \right) =$ **(1.45, 1.24)**.

(2) Specify a (3×3) **sensitivity-analysis grid** by halving and doubling the (θ, σ) values in step (1):
$\{\theta_{DG} \in (0.725, 1.45, 2.90)\} \times \{\sigma_{DG} \in (0.61, 1.24, 2.48)\}$.

For each $(\theta_{DG}, \sigma_{DG})$ **combination** in this grid,

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in R) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**: it turns out that, on the scale of percentage points of mortality, $\left(\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}\right) =$ **(1.45, 1.24)**.

(2) Specify a (3×3) **sensitivity-analysis grid** by halving and doubling the (θ, σ) values in step (1):
$\{\theta_{DG} \in (0.725, 1.45, 2.90)\} \times \{\sigma_{DG} \in (0.61, 1.24, 2.48)\}$.

For each $(\theta_{DG}, \sigma_{DG})$ **combination** in this grid,

(3) Generate $\boldsymbol{N} =$ **1,000 (say) data sets** from the second and third lines of model (12);

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in `R`) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**: it turns out that, on the scale of percentage points of mortality, $\left(\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}\right) =$ **(1.45, 1.24)**.

(2) Specify a (3×3) **sensitivity-analysis grid** by halving and doubling the (θ, σ) values in step (1):
$\{\theta_{DG} \in (0.725, 1.45, 2.90)\} \times \{\sigma_{DG} \in (0.61, 1.24, 2.48)\}$.

For each $(\theta_{DG}, \sigma_{DG})$ **combination** in this grid,

(3) Generate $\boldsymbol{N} =$ **1,000 (say) data sets** from the second and third lines of model (12);

(4) Use `rjags` to fit variations on model (12) to each simulated data set in step (3), with **several different priors** whose calibration You're investigating,

An Example of the Calibration Principle in Action

If Gelman (2006) were **not available**, guided by the **Calibration Principle** You could conduct a **simulation study** (e.g., in `R`) in the **aspirin meta-analysis** of the following form:

(1) To get a **rough idea** of θ and σ with this data set, calculate their **maximum-likelihood estimates**: it turns out that, on the scale of percentage points of mortality, $\left(\hat{\theta}_{MLE}, \hat{\sigma}_{MLE}\right) =$ **(1.45, 1.24)**.

(2) Specify a (3×3) **sensitivity-analysis grid** by halving and doubling the (θ, σ) values in step (1):
$\{\theta_{DG} \in (0.725, 1.45, 2.90)\} \times \{\sigma_{DG} \in (0.61, 1.24, 2.48)\}$.

For each $(\theta_{DG}, \sigma_{DG})$ **combination** in this grid,

(3) Generate $\boldsymbol{N} =$ **1,000 (say) data sets** from the second and third lines of model (12);

(4) Use `rjags` to fit variations on model (12) to each simulated data set in step (3), with **several different priors** whose calibration You're investigating, and keep track of **how often the posterior 95% intervals** for θ and σ include their **simulation-known-truth** values from step (2).

Calibration Checking in Prior Specification

Actual coverage at nominal 95%: $(\sigma \mid \mathcal{B}) \sim \boldsymbol{U(0,10)}$

	Intervals for μ			Intervals for σ		
	σ_{DG}			σ_{DG}		
μ_{DG}	0.61	1.24	2.48	0.61	1.24	2.48
0.725	0.994	0.988	0.975	0.976	0.963	0.957
1.45	0.996	0.990	0.973	0.971	0.976	0.958
2.90	0.994	0.984	0.970	0.978	0.965	0.963

All actual coverages at or above nominal with the $\boldsymbol{U(0,10)}$ prior for $\boldsymbol{\sigma}$

Calibration Checking in Prior Specification

Actual coverage at nominal 95%: $(\sigma \mid \mathcal{B}) \sim \boldsymbol{U(0, 10)}$

	Intervals for μ			Intervals for σ		
	σ_{DG}			σ_{DG}		
μ_{DG}	0.61	1.24	2.48	0.61	1.24	2.48
0.725	0.994	0.988	0.975	0.976	0.963	0.957
1.45	0.996	0.990	0.973	0.971	0.976	0.958
2.90	0.994	0.984	0.970	0.978	0.965	0.963

All actual coverages at or above nominal with the $\boldsymbol{U(0, 10)}$ prior for $\boldsymbol{\sigma}$

Actual coverage at nominal 95%: $(\sigma \mid \mathcal{B}) \sim$ **Half-Cauchy**

	Intervals for μ			Intervals for σ		
	σ_{DG}			σ_{DG}		
μ_{DG}	0.61	1.24	2.48	0.61	1.24	2.48
0.725	0.994	0.985	0.970	0.979	0.967	0.963
1.45	0.994	0.989	0.967	0.970	0.980	0.964
2.90	0.994	0.984	0.967	0.982	0.970	0.966

All actual coverages at or above nominal with the Half-Cauchy prior for $\boldsymbol{\sigma}$

Calibration Checking in Prior Specification

Actual coverage at nominal 95%: $(\sigma \mid \mathcal{B}) \sim \boldsymbol{U}(\mathbf{0}, \mathbf{10})$

	Intervals for μ			Intervals for σ		
	σ_{DG}			σ_{DG}		
μ_{DG}	0.61	1.24	2.48	0.61	1.24	2.48
0.725	0.994	0.988	0.975	0.976	0.963	0.957
1.45	0.996	0.990	0.973	0.971	0.976	0.958
2.90	0.994	0.984	0.970	0.978	0.965	0.963

All actual coverages at or above nominal with the $\boldsymbol{U}(\mathbf{0}, \mathbf{10})$ prior for $\boldsymbol{\sigma}$

Actual coverage at nominal 95%: $(\sigma \mid \mathcal{B}) \sim$ **Half-Cauchy**

	Intervals for μ			Intervals for σ		
	σ_{DG}			σ_{DG}		
μ_{DG}	0.61	1.24	2.48	0.61	1.24	2.48
0.725	0.994	0.985	0.970	0.979	0.967	0.963
1.45	0.994	0.989	0.967	0.970	0.980	0.964
2.90	0.994	0.984	0.967	0.982	0.970	0.966

All actual coverages at or above nominal with the Half-Cauchy prior for $\boldsymbol{\sigma}$

Actual coverage at nominal 95%: $(\sigma^2 \mid \mathcal{B}) \sim \Gamma^{-1}\left(10^{-6}, 10^{-6}\right)$

	Intervals for μ			Intervals for σ		
	σ_{DG}			σ_{DG}		
μ_{DG}	0.61	1.24	2.48	0.61	1.24	2.48
0.725	0.947	**0.922**	**0.902**	0.998	0.950	**0.825**
1.45	0.943	**0.920**	**0.890**	1.000	0.966	**0.822**
2.90	0.956	**0.909**	**0.884**	0.999	0.953	**0.808**

Poor Coverage with the $\boldsymbol{\Gamma^{-1}(\epsilon, \epsilon)}$ prior for $\boldsymbol{\sigma^2}$ ($\boldsymbol{\epsilon = 10^{-6}}$)

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information);

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors;

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**;

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**; if not, go to (2)).

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**; if not, go to (2)).

(2) **(with simulated data)** Carry out a **simulation study**, similar to the one summarized above with the aspirin meta-analysis, in which you know the right answer, and use a **well-calibrated prior**.

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**; if not, go to (2)).

(2) **(with simulated data)** Carry out a **simulation study**, similar to the one summarized above with the aspirin meta-analysis, in which you know the right answer, and use a **well-calibrated prior**.

An example of (1) arises in the **aspirin meta-analysis**:

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**; if not, go to (2)).

(2) **(with simulated data)** Carry out a **simulation study**, similar to the one summarized above with the aspirin meta-analysis, in which you know the right answer, and use a **well-calibrated prior**.

An example of (1) arises in the **aspirin meta-analysis**: because the AMIS result was an **apparent outlier**,

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**; if not, go to (2)).

(2) **(with simulated data)** Carry out a **simulation study**, similar to the one summarized above with the aspirin meta-analysis, in which you know the right answer, and use a **well-calibrated prior**.

An example of (1) arises in the **aspirin meta-analysis**: because the AMIS result was an **apparent outlier**, perhaps the second line of the model, $(\Delta_i \mid \theta \, \sigma \, \mathcal{B}) \stackrel{\text{IID}}{\sim} N(\theta, \sigma^2)$,

Some Model Uncertainty Questions (continued)

(2) How do you **test** the extent to which the **prior** drives the results ***[in settings in which a unique context-specific prior doesn't exist]***?

In this situation You can perform **two types of sensitivity analyses**:

(1) **(with Your data)** Formulate a **variety of priors** that are responsive to context (e.g., a **low-information-prior** may be called for, but there's **more than one way** to quantify low information); **update Your likelihood** to posterior distributions with each of these priors; and **see if all roads lead to Rome** (if so, **good**; if not, go to (2)).

(2) **(with simulated data)** Carry out a **simulation study**, similar to the one summarized above with the aspirin meta-analysis, in which you know the right answer, and use a **well-calibrated prior**.

An example of (1) arises in the **aspirin meta-analysis**: because the AMIS result was an **apparent outlier**, perhaps the second line of the model, $(\Delta_i \,|\, \theta\, \sigma\, \mathcal{B}) \stackrel{\text{IID}}{\sim} N(\theta, \sigma^2)$, should be replaced by $(\Delta_i \,|\, \theta\, \sigma\, \nu\, \mathcal{B}) \stackrel{\text{IID}}{\sim} t_\nu(\theta, \sigma^2)$ with **unknown degrees of freedom** ν.

Prior Sensitivity Analysis

untitled2

```
{

  theta ~ dnorm( 0.0, 1.0E-4 )
  sigma ~ dunif( 0.0, 10.0 )
  nu ~ dunif( 2.0, 50.0 )

  for ( i in 1:k ) {

    Delta[ i ] ~ dt( theta, tau, nu )
    y[ i ] ~ dnorm( Delta[ i ],
      tau.y[ i ] )

  }

  tau <- 1.0 / ( sigma * sigma )

}
```

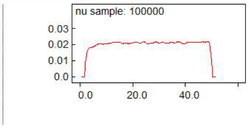
untitled3

```
list( k = 6, y = c( 2.77, 2.50, 1.84, 2.56,
  2.32, -1.15 ), tau.y = c( 0.3673, 0.5827,
  0.1826, 0.3586, 0.2551, 1.235 ) )
```

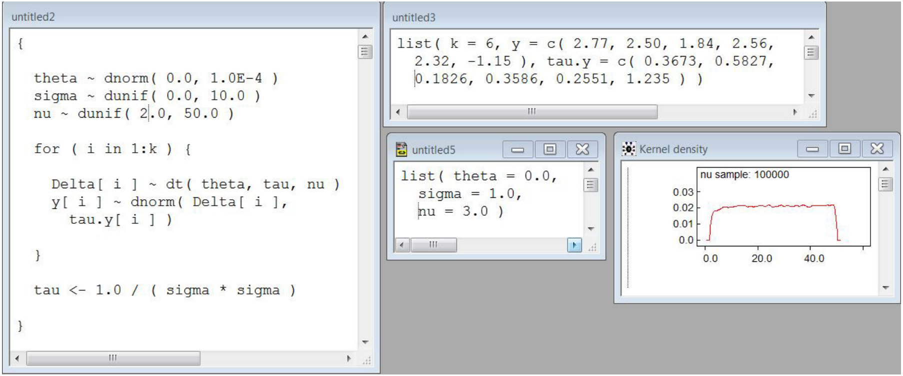
untitled5

```
list( theta = 0.0,
  sigma = 1.0,
  nu = 3.0 )
```

Kernel density

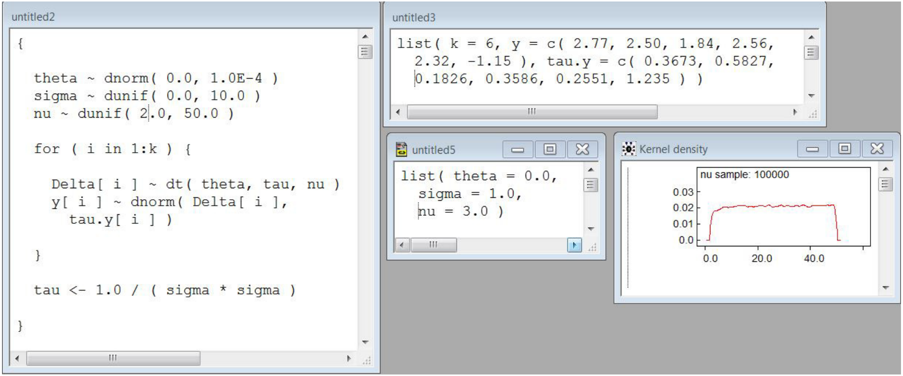


Prior Sensitivity Analysis



With **only $k = 6$ studies** in the meta-analysis, there's **no information in the data about ν**: its posterior and prior distributions **coincide**.

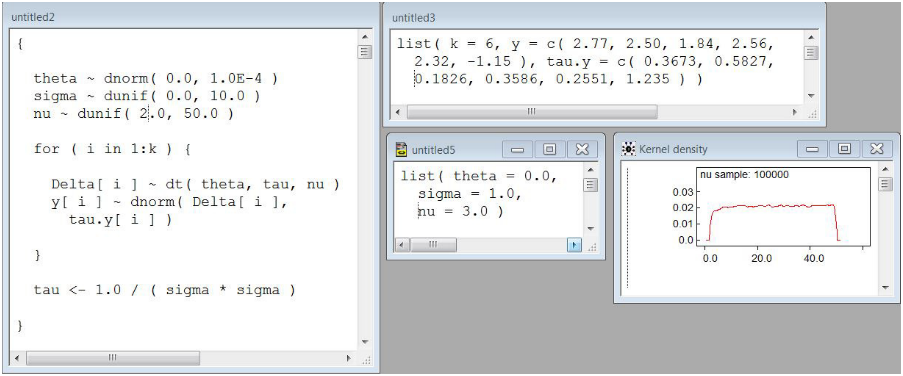
Prior Sensitivity Analysis



With **only $k = 6$ studies** in the meta-analysis, there's **no information in the data about ν**: its posterior and prior distributions **coincide**.

In effect, the data set **can't tell the difference** between (big ν [Normal], big σ)

Prior Sensitivity Analysis



With **only $k = 6$ studies** in the meta-analysis, there's **no information in the data about ν**: its posterior and prior distributions **coincide**.

In effect, the data set **can't tell the difference** between (big ν [Normal], big σ) and (small ν [t], small σ).

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**,

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**);

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**;

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

In the **Bayesian paradigm**,

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

In the **Bayesian paradigm**, with $\boldsymbol{D} = \boldsymbol{y} = (y_1, \ldots, y_n)$ for simplicity,

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

In the **Bayesian paradigm**, with $\boldsymbol{D} = \boldsymbol{y} = (y_1, \dots, y_n)$ for simplicity, making the **Prediction Principle** operational requires **comparing** a data value y^* with $p(y^* \mid \boldsymbol{y} \, \mathbb{M} \, \mathcal{B})$,

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

In the **Bayesian paradigm**, with $\boldsymbol{D} = \boldsymbol{y} = (y_1, \ldots, y_n)$ for simplicity, making the **Prediction Principle** operational requires **comparing** a data value y^* with $p(y^* \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B})$, the **posterior predictive distribution** for y^* under the model $\mathbb{M}$ You're **currently contemplating**;

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

In the **Bayesian paradigm**, with $\boldsymbol{D} = \boldsymbol{y} = (y_1, \dots, y_n)$ for simplicity, making the **Prediction Principle** operational requires **comparing** a data value y^* with $p(y^* \,|\, \boldsymbol{y} \, \mathbb{M} \, \mathcal{B})$, the **posterior predictive distribution** for y^* under the model $\mathbb{M}$ You're **currently contemplating**; a Theorem due to Good (1951) shows that the **best way** to do this is with **log scores**, e.g.,

Some Model Uncertainty Questions (continued)

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

In my view these questions are **best answered** via the

Prediction Principle: **Good models** make **good predictions** of **observable quantities (data)**, especially of **data values not used** in fitting the model (this is **out-of-sample predictive validation**); **bad models** make **bad predictions**; that's one important way **You know if You have a good model**.

In the **Bayesian paradigm**, with $\boldsymbol{D} = \boldsymbol{y} = (y_1, \ldots, y_n)$ for simplicity, making the **Prediction Principle** operational requires **comparing** a data value y^* with $p(y^* \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B})$, the **posterior predictive distribution** for y^* under the model $\mathbb{M}$ You're **currently contemplating**; a Theorem due to Good (1951) shows that the **best way** to do this is with **log scores**, e.g.,

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\, \mathbb{M}\, \mathcal{B}) \; , \tag{14}$$

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\, \mathbb{M}\, \mathcal{B}) \; , \tag{15}$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B}) \; , \tag{16}$$

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\, \mathbb{M}\, \mathcal{B}) \ , \tag{15}$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B}) \ , \tag{16}$$

is a **good approximation** to (15) that's **fast and easy to compute**.

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\,\mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\,\mathbb{M}\,\mathcal{B}) \; , \tag{15}$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\,\mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}\,\mathbb{M}\,\mathcal{B}) \; , \tag{16}$$

is a **good approximation** to (15) that's **fast and easy to compute**.

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \,|\, \boldsymbol{y}\,\mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \,|\, \boldsymbol{y}_{-i}\,\mathbb{M}\,\mathcal{B}) \ , \qquad (15)$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \,|\, \boldsymbol{y}\,\mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \,|\, \boldsymbol{y}\,\mathbb{M}\,\mathcal{B}) \ , \qquad (16)$$

is a **good approximation** to (15) that's **fast and easy to compute**.

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Full-sample log scores can be used to answer a **related question**,

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\, \mathbb{M}\, \mathcal{B}) \ , \qquad (15)$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B}) \ , \qquad (16)$$

is a **good approximation** to (15) that's **fast and easy to compute**.

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Full-sample log scores can be used to answer a **related question**,

(1′) **Could the data set D have arisen** from model $\mathbb{M}$?

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\, \mathbb{M}\, \mathcal{B}) \ , \qquad (15)$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B}) \ , \qquad (16)$$

is a **good approximation** to (15) that's **fast and easy to compute**.

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Full-sample log scores can be used to answer a **related question**,

($1'$) **Could the data set D have arisen** from model $\mathbb{M}$?

by comparing the **realized value** of $LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B})$ with a **distribution of full-sample log scores** obtained by simulating data sets from $\mathbb{M}$ (Gelman, Meng and Stern (1996)),

Out-of-Sample Predictive Validation

$$LS_{CV}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}_{-i}\, \mathbb{M}\, \mathcal{B}) \, , \qquad (15)$$

in which (using a **Cross Validation** (CV) idea) $\boldsymbol{y}_{-i}$ is $\boldsymbol{y}$ with observation i set aside; Draper and Krnjajić (2017) have shown that a **Full Sample** (FS) version of this,

$$LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B}) = \frac{1}{n} \sum_{i=1}^{n} \log p(y_i \mid \boldsymbol{y}\, \mathbb{M}\, \mathcal{B}) \, , \qquad (16)$$

is a **good approximation** to (15) that's **fast and easy to compute**.

(1) How do you evaluate **how well the model fits the data**? What are **formal model checking procedures**?

Full-sample log scores can be used to answer a **related question**,

(1′) **Could the data set D have arisen** from model $\mathbb{M}$?

by comparing the **realized value** of $LS_{FS}(\mathbb{M} \mid \boldsymbol{y}\, \mathcal{B})$ with a **distribution of full-sample log scores** obtained by simulating data sets from $\mathbb{M}$ (Gelman, Meng and Stern (1996)), but this must be **done carefully** to be **well-calibrated** (Draper and Krnjajić (2017)).

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history;**

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)**

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \,|\, \boldsymbol{D}\, \mathcal{B}) > LS_{FS}(\mathbb{M}_2 \,|\, \boldsymbol{D}\, \mathcal{B})$.

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \,|\, \boldsymbol{D}\, \mathcal{B}) > LS_{FS}(\mathbb{M}_2 \,|\, \boldsymbol{D}\, \mathcal{B})$.
    - **(Bayes factors, including *BIC*)**

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \mid \boldsymbol{D}\, \mathcal{B}) > LS_{FS}(\mathbb{M}_2 \mid \boldsymbol{D}\, \mathcal{B})$.
  - **(Bayes factors, including *BIC*)** This approach says —

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \,|\, \boldsymbol{D}\, \mathcal{B}) > LS_{FS}(\mathbb{M}_2 \,|\, \boldsymbol{D}\, \mathcal{B})$.

  - **(Bayes factors, including *BIC*)** This approach says — **pick the model with the highest posterior probability**:

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B}) > LS_{FS}(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})$.
  - **(Bayes factors, including *BIC*)** This approach says — **pick the model with the highest posterior probability**:

$$\left[\frac{p(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B})}{p(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})}\right] = \left[\frac{p(\mathbb{M}_1 \mid \mathcal{B})}{p(\mathbb{M}_2 \mid \mathcal{B})}\right] \cdot \left[\frac{p(\boldsymbol{D} \mid \mathbb{M}_1\,\mathcal{B})}{p(\boldsymbol{D} \mid \mathbb{M}_2\,\mathcal{B})}\right]$$
$$\left(\begin{array}{c}\textbf{posterior}\\\textbf{odds}\end{array}\right) = \left(\begin{array}{c}\textbf{prior}\\\textbf{odds}\end{array}\right) \cdot \left(\begin{array}{c}\textbf{Bayes}\\\textbf{factor}\end{array}\right) \quad (17)$$

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B}) > LS_{FS}(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})$.

  - **(Bayes factors, including *BIC*)** This approach says — **pick the model with the highest posterior probability**:

$$\left[\frac{p(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B})}{p(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})}\right] = \left[\frac{p(\mathbb{M}_1 \mid \mathcal{B})}{p(\mathbb{M}_2 \mid \mathcal{B})}\right] \cdot \left[\frac{p(\boldsymbol{D} \mid \mathbb{M}_1\,\mathcal{B})}{p(\boldsymbol{D} \mid \mathbb{M}_2\,\mathcal{B})}\right]$$
$$\left(\begin{array}{c}\textbf{posterior}\\ \textbf{odds}\end{array}\right) = \left(\begin{array}{c}\textbf{prior}\\ \textbf{odds}\end{array}\right) \cdot \left(\begin{array}{c}\textbf{Bayes}\\ \textbf{factor}\end{array}\right) \tag{17}$$

People who like this approach tend to **ignore the prior odds (!)** and use a different rule:

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B}) > LS_{FS}(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})$.

  - **(Bayes factors, including *BIC*)** This approach says — **pick the model with the highest posterior probability**:

$$\left[\frac{p(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B})}{p(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})}\right] = \left[\frac{p(\mathbb{M}_1 \mid \mathcal{B})}{p(\mathbb{M}_2 \mid \mathcal{B})}\right] \cdot \left[\frac{p(\boldsymbol{D} \mid \mathbb{M}_1\,\mathcal{B})}{p(\boldsymbol{D} \mid \mathbb{M}_2\,\mathcal{B})}\right]$$
$$\left(\begin{array}{c}\textbf{posterior}\\ \textbf{odds}\end{array}\right) = \left(\begin{array}{c}\textbf{prior}\\ \textbf{odds}\end{array}\right) \cdot \left(\begin{array}{c}\textbf{Bayes}\\ \textbf{factor}\end{array}\right) \qquad (17)$$

People who like this approach tend to **ignore the prior odds (!)** and use a different rule: **pick the model with the highest Bayes factor**.

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(**3**) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B}) > LS_{FS}(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})$.

  - **(Bayes factors, including *BIC*)** This approach says — **pick the model with the highest posterior probability**:

$$\left[\frac{p(\mathbb{M}_1 \mid \boldsymbol{D}\,\mathcal{B})}{p(\mathbb{M}_2 \mid \boldsymbol{D}\,\mathcal{B})}\right] = \left[\frac{p(\mathbb{M}_1 \mid \mathcal{B})}{p(\mathbb{M}_2 \mid \mathcal{B})}\right] \cdot \left[\frac{p(\boldsymbol{D} \mid \mathbb{M}_1\,\mathcal{B})}{p(\boldsymbol{D} \mid \mathbb{M}_2\,\mathcal{B})}\right]$$
$$\left(\begin{array}{c}\textbf{posterior}\\ \textbf{odds}\end{array}\right) = \left(\begin{array}{c}\textbf{prior}\\ \textbf{odds}\end{array}\right) \cdot \left(\begin{array}{c}\textbf{Bayes}\\ \textbf{factor}\end{array}\right) \qquad (17)$$

People who like this approach tend to **ignore the prior odds (!)** and use a different rule: **pick the model with the highest Bayes factor**.

This method runs into **additional difficulty** when context implies that You should use a **low-information prior** in either or both of $\mathbb{M}_1$ and $\mathbb{M}_2$,

Bayesian Model Comparison: Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?

There's one last **important model-uncertainty question**:

(3) How do You know when model $\mathbb{M}_1$ is better than model $\mathbb{M}_2$?

This question has a **rich research history; three main methods** have emerged for answering it:

- **(log scores)** $\mathbb{M}_1$ is better than $\mathbb{M}_2$ if $LS_{FS}(\mathbb{M}_1 \mid \boldsymbol{D}\, \mathcal{B}) > LS_{FS}(\mathbb{M}_2 \mid \boldsymbol{D}\, \mathcal{B})$.

  - **(Bayes factors, including *BIC*)** This approach says — **pick the model with the highest posterior probability**:

$$\left[\frac{p(\mathbb{M}_1 \mid \boldsymbol{D}\, \mathcal{B})}{p(\mathbb{M}_2 \mid \boldsymbol{D}\, \mathcal{B})}\right] = \left[\frac{p(\mathbb{M}_1 \mid \mathcal{B})}{p(\mathbb{M}_2 \mid \mathcal{B})}\right] \cdot \left[\frac{p(\boldsymbol{D} \mid \mathbb{M}_1\, \mathcal{B})}{p(\boldsymbol{D} \mid \mathbb{M}_2\, \mathcal{B})}\right]$$
$$\left(\begin{array}{c}\textbf{posterior}\\ \textbf{odds}\end{array}\right) = \left(\begin{array}{c}\textbf{prior}\\ \textbf{odds}\end{array}\right) \cdot \left(\begin{array}{c}\textbf{Bayes}\\ \textbf{factor}\end{array}\right) \quad (17)$$

People who like this approach tend to **ignore the prior odds (!)** and use a different rule: **pick the model with the highest Bayes factor**.

This method runs into **additional difficulty** when context implies that You should use a **low-information prior** in either or both of $\mathbb{M}_1$ and $\mathbb{M}_2$, a setting that **arises frequently**.

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion *BIC*** (Schwarz, 1978):

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion** ***BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \mid \boldsymbol{D}\,\mathcal{B}) = -2\,\log\left[\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B})\right] + k_j\,\log(n)\,, \qquad (18)$$

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion** ***BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \mid \boldsymbol{D}\, \mathcal{B}) = -2 \log \left[\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B}) \right] + k_j \log(n) \, , \qquad (18)$$

in which $\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B})$ is the maximum **likelihood function** under model $\mathbb{M}_j$, k_j is the **number of parameters** in $\boldsymbol{\theta}_j$, and n is the **sample size**.

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion *BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \mid \boldsymbol{D}\,\mathcal{B}) = -2 \log \left[\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B}) \right] + k_j \log(n) \,, \qquad (18)$$

in which $\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B})$ is the maximum **likelihood function** under model $\mathbb{M}_j$, k_j is the **number of parameters** in $\boldsymbol{\theta}_j$, and n is the **sample size**.

Unfortunately the concept of **"number of parameters** in $\boldsymbol{\theta}_j$" is **hard to define** when the model is **hierarchical with random effects**, as in the **aspirin meta-analysis**:

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion *BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \mid \boldsymbol{D}\, \mathcal{B}) = -2 \log \left[\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B}) \right] + k_j \log(n) \, , \qquad (18)$$

in which $\ell(\hat{\boldsymbol{\theta}}_j \mid \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B})$ is the maximum **likelihood function** under model $\mathbb{M}_j$, k_j is the **number of parameters** in $\boldsymbol{\theta}_j$, and n is the **sample size**.

Unfortunately the concept of **"number of parameters** in $\boldsymbol{\theta}_j$" is **hard to define** when the model is **hierarchical with random effects**, as in the **aspirin meta-analysis**:

$$\begin{aligned} (\theta \mid \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) \qquad & (\sigma \mid \mathcal{B}) \sim U(0, C) \\ (\Delta_i \mid \theta\, \sigma\, \mathcal{B}) \quad & \stackrel{\text{IID}}{\sim} \quad N\left(\theta, \sigma^2\right) \\ (y_i \mid \Delta_i\, \mathcal{B}) \quad & \stackrel{\text{I}}{\sim} \quad N(\Delta_i, V_i) \, . \end{aligned} \qquad (19)$$

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion** ***BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \,|\, \boldsymbol{D}\, \mathcal{B}) = -2 \, \log \left[\ell(\hat{\boldsymbol{\theta}}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B}) \right] + k_j \, \log(n) \, , \qquad (18)$$

in which $\ell(\hat{\boldsymbol{\theta}}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B})$ is the maximum **likelihood function** under model $\mathbb{M}_j$, k_j is the **number of parameters** in $\boldsymbol{\theta}_j$, and n is the **sample size**.

Unfortunately the concept of **"number of parameters** in $\boldsymbol{\theta}_j$" is **hard to define** when the model is **hierarchical with random effects**, as in the **aspirin meta-analysis**:

$$\begin{array}{rcl} (\theta \,|\, \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) & & (\sigma \,|\, \mathcal{B}) \sim U(0, C) \\ (\Delta_i \,|\, \theta \, \sigma \, \mathcal{B}) & \stackrel{\text{IID}}{\sim} & N\left(\theta, \sigma^2\right) \\ (y_i \,|\, \Delta_i \, \mathcal{B}) & \stackrel{\text{I}}{\sim} & N(\Delta_i, V_i) \, . \end{array} \qquad (19)$$

This model is **hierarchical with 3 levels**,

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion** ***BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \,|\, \boldsymbol{D}\, \mathcal{B}) = -2 \, \log \left[\ell(\hat{\boldsymbol{\theta}}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B}) \right] + k_j \, \log(n) \,, \qquad (18)$$

in which $\ell(\hat{\boldsymbol{\theta}}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B})$ is the maximum **likelihood function** under model $\mathbb{M}_j$, k_j is the **number of parameters** in $\boldsymbol{\theta}_j$, and n is the **sample size**.

Unfortunately the concept of **"number of parameters** in $\boldsymbol{\theta}_j$" is **hard to define** when the model is **hierarchical with random effects**, as in the **aspirin meta-analysis**:

$$\begin{aligned} (\theta \,|\, \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) & \qquad (\sigma \,|\, \mathcal{B}) \sim U(0, C) \\ (\Delta_i \,|\, \theta\, \sigma\, \mathcal{B}) & \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \\ (y_i \,|\, \Delta_i\, \mathcal{B}) & \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \,. \qquad (19) \end{aligned}$$

This model is **hierarchical with 3 levels**, and the Δ_i are referred to as **random effects** because we're modeling them as if they were **random draws** from the $N\left(\theta, \sigma^2\right)$ distribution;

BIC, and When It Doesn't Work

The only **special case of Bayes factors** that I use in my applied work is the **Bayesian Information Criterion** ***BIC*** (Schwarz, 1978):

$$BIC(\mathbb{M}_j \,|\, \boldsymbol{D}\, \mathcal{B}) = -2 \log \left[\ell(\hat{\boldsymbol{\theta}}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B}) \right] + k_j \log(n) \,, \qquad (18)$$

in which $\ell(\hat{\boldsymbol{\theta}}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B})$ is the maximum **likelihood function** under model $\mathbb{M}_j$, k_j is the **number of parameters** in $\boldsymbol{\theta}_j$, and n is the **sample size**.

Unfortunately the concept of **"number of parameters** in $\boldsymbol{\theta}_j$" is **hard to define** when the model is **hierarchical with random effects**, as in the **aspirin meta-analysis**:

$$\begin{aligned} (\theta \,|\, \mathcal{B}) \sim N\left(0, \sigma^2_{huge}\right) & \qquad (\sigma \,|\, \mathcal{B}) \sim U(0, C) \\ (\Delta_i \,|\, \theta\, \sigma\, \mathcal{B}) & \stackrel{\text{IID}}{\sim} N\left(\theta, \sigma^2\right) \\ (y_i \,|\, \Delta_i\, \mathcal{B}) & \stackrel{\text{I}}{\sim} N(\Delta_i, V_i) \,. \end{aligned} \qquad (19)$$

This model is **hierarchical with 3 levels**, and the Δ_i are referred to as **random effects** because we're modeling them as if they were **random draws** from the $N\left(\theta, \sigma^2\right)$ distribution; ***BIC* doesn't work here**.

DIC

- **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**;

DIC

- **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

DIC

- **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

• **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \mid \boldsymbol{D}\,\mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \,, \qquad (20)$$

DIC

- **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \mid \boldsymbol{D}\, \mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \, , \tag{20}$$

in which $\bar{D}_{\mathbb{M}_j}$ is the posterior mean of the **deviance** $-2 \log \left[\ell(\boldsymbol{\theta}_j \mid \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B}) \right]$ for model $\mathbb{M}_j$

DIC

- **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \,|\, \boldsymbol{D}\,\mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \, , \tag{20}$$

in which $\bar{D}_{\mathbb{M}_j}$ is the posterior mean of the **deviance** $-2 \log [\ell(\boldsymbol{\theta}_j \,|\, \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B})]$ for model $\mathbb{M}_j$ and $(p_D)_{\mathbb{M}_j}$ is the **complexity** of $\mathbb{M}_j$, estimated via

DIC

• **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \mid \boldsymbol{D}\,\mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \;, \tag{20}$$

in which $\bar{D}_{\mathbb{M}_j}$ is the posterior mean of the **deviance** $-2\log\left[\ell(\boldsymbol{\theta}_j \mid \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B})\right]$ for model $\mathbb{M}_j$ and $(p_D)_{\mathbb{M}_j}$ is the **complexity** of $\mathbb{M}_j$, estimated via

$$(p_D)_{\mathbb{M}_j} = \left(\begin{array}{c}\textbf{posterior mean}\\ \textbf{of deviance}\end{array}\right) - \left(\begin{array}{c}\textbf{deviance evaluated at}\\ \textbf{posterior mean of } \boldsymbol{\theta}_j\end{array}\right) . \tag{21}$$

DIC

- **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \,|\, \boldsymbol{D}\, \mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \;, \tag{20}$$

in which $\bar{D}_{\mathbb{M}_j}$ is the posterior mean of the **deviance** $-2 \log [\ell(\boldsymbol{\theta}_j \,|\, \boldsymbol{D}\, \mathbb{M}_j\, \mathcal{B})]$ for model $\mathbb{M}_j$ and $(p_D)_{\mathbb{M}_j}$ is the **complexity** of $\mathbb{M}_j$, estimated via

$$(p_D)_{\mathbb{M}_j} = \left(\begin{array}{c} \textbf{posterior mean} \\ \textbf{of deviance} \end{array} \right) - \left(\begin{array}{c} \textbf{deviance evaluated at} \\ \textbf{posterior mean of } \boldsymbol{\theta}_j \end{array} \right) . \tag{21}$$

As an **example**, consider what the **aspirin meta-analysis model** would look like if there were **no between-study heterogeneity**:

DIC

• **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \mid \boldsymbol{D}\,\mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \,, \tag{20}$$

in which $\bar{D}_{\mathbb{M}_j}$ is the posterior mean of the **deviance** $-2 \log [\ell(\boldsymbol{\theta}_j \mid \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B})]$ for model $\mathbb{M}_j$ and $(p_D)_{\mathbb{M}_j}$ is the **complexity** of $\mathbb{M}_j$, estimated via

$$(p_D)_{\mathbb{M}_j} = \left(\begin{array}{c} \textbf{posterior mean} \\ \textbf{of deviance} \end{array} \right) - \left(\begin{array}{c} \textbf{deviance evaluated at} \\ \textbf{posterior mean of } \boldsymbol{\theta}_j \end{array} \right) . \tag{21}$$

As an **example**, consider what the **aspirin meta-analysis model** would look like if there were **no between-study heterogeneity**:

$$\begin{aligned} (\theta \mid \mathcal{B}) &\sim N\left(0, \sigma^2_{huge}\right) \\ (y_i \mid \theta\,\mathcal{B}) &\overset{\text{I}}{\sim} N(\theta, V_i) \,. \end{aligned} \tag{22}$$

DIC

• **(information-theoretic criteria, including *DIC*)** The idea here is to **explicitly trade off model fit against model complexity**; confusingly, ***BIC*** is based on the same idea but belongs to the world of Bayes factors.

The **most widely used method** in this category is the **Deviance Information Criterion *DIC***:

$$DIC(\mathbb{M}_j \,|\, \boldsymbol{D}\,\mathcal{B}) = \bar{D}_{\mathbb{M}_j} + (p_D)_{\mathbb{M}_j} \;, \tag{20}$$

in which $\bar{D}_{\mathbb{M}_j}$ is the posterior mean of the **deviance** $-2 \log [\ell(\boldsymbol{\theta}_j \,|\, \boldsymbol{D}\,\mathbb{M}_j\,\mathcal{B})]$ for model $\mathbb{M}_j$ and $(p_D)_{\mathbb{M}_j}$ is the **complexity** of $\mathbb{M}_j$, estimated via

$$(p_D)_{\mathbb{M}_j} = \left(\begin{array}{c} \textbf{posterior mean} \\ \textbf{of deviance} \end{array} \right) - \left(\begin{array}{c} \textbf{deviance evaluated at} \\ \textbf{posterior mean of } \boldsymbol{\theta}_j \end{array} \right) . \tag{21}$$

As an **example**, consider what the **aspirin meta-analysis model** would look like if there were **no between-study heterogeneity**:

$$\begin{aligned} (\theta \,|\, \mathcal{B}) & \sim N(0, \sigma^2_{huge}) \\ (y_i \,|\, \theta\,\mathcal{B}) & \stackrel{\text{I}}{\sim} N(\theta, V_i) \;. \end{aligned} \tag{22}$$

This is referred to as a **fixed-effects meta-analysis model**.

Fixed-Effects Versus Random-Effects Meta-Analysis

untitled1

```
{

  theta ~ dnorm( 0.0, 1.0E-4 )

  for ( i in 1:k ) {

    y[ i ] ~ dnorm( theta,
      tau.y[ i ] )

  }
```

untitled2

```
list( k = 6, y = c( 2.77, 2.50,
  1.84, 2.56, 2.32, -1.15 ),
  tau.y = c( 0.3673, 0.5827,
0.1826, 0.3586, 0.2551, 1.235 ) )
```

DIC

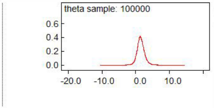
Dbar = post.mean of -2logL; Dhat = -2LogL
at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
y	27.061	26.064	0.997	28.058
total	27.061	26.064	0.997	28.058

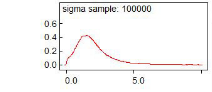
untitled4

```
{

  theta ~ dnorm( 0.0, 1.0E-4 )
  sigma ~ dunif( 0.0, 10.0 )

  for ( i in 1:k ) {

    Delta[ i ] ~ dnorm( theta, tau )
    y[ i ] ~ dnorm( Delta[ i ],
      tau.y[ i ] )

  }

  tau <- 1.0 / ( sigma * sigma )

  positive.effect <- step( theta )
```

Kernel density



Kernel density



DIC

Dbar = post.mean of -2logL; Dhat = -2LogL
at post.mean of stochastic nodes

	Dbar	Dhat	pD	DIC
y	21.671	17.552	4.119	25.790
total	21.671	17.552	4.119	25.790

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%
theta	1.515	1.181	0.006175	-0.6613	1.458	4.001

Node statistics

node	mean
positive.effect	0.9278

Conclusions

- On the basis of this **meta-analysis**,

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks**

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**),

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality)

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**;

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.
- Key issues in **Bayesian model specification** include:

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.
- Key issues in **Bayesian model specification** include:
  - measuring **sensitivity** of results to **alternative prior and likelihood specifications**, when context does not imply **unique choices**;

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.
- Key issues in **Bayesian model specification** include:
  - measuring **sensitivity** of results to **alternative prior and likelihood specifications**, when context does not imply **unique choices**;
  - using **out-of-sample predictive validation** to **find good models**; and

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.
- Key issues in **Bayesian model specification** include:
  - measuring **sensitivity** of results to **alternative prior and likelihood specifications**, when context does not imply **unique choices**;
  - using **out-of-sample predictive validation** to **find good models**; and
  - answering the questions **Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?**

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.
- Key issues in **Bayesian model specification** include:
  - measuring **sensitivity** of results to **alternative prior and likelihood specifications**, when context does not imply **unique choices**;
  - using **out-of-sample predictive validation** to **find good models**; and
  - answering the questions **Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?** and **Could the data have arisen from $\mathbb{M}_1$?**

Conclusions

- On the basis of this **meta-analysis**, **low-dose aspirin** is estimated to **lower one-year all-cause mortality** for people with **one or more previous heart attacks** by about **1.5 percentage points** (relative to a baseline of about **10.5%**), and the **posterior probability** that **low-dose aspirin is beneficial** for such people (in one-year all-cause mortality) is about **93%**.
- Model uncertainty is the **fundamental problem of applied statistics**.
- **All statisticians** should try to be **well-calibrated**; **simulation-based methods** can help You **keep score**.
- Key issues in **Bayesian model specification** include:
  - measuring **sensitivity** of results to **alternative prior and likelihood specifications**, when context does not imply **unique choices**;
  - using **out-of-sample predictive validation** to **find good models**; and
  - answering the questions **Is $\mathbb{M}_1$ better than $\mathbb{M}_2$?** and **Could the data have arisen from $\mathbb{M}_1$?** in a **well-calibrated** way.