



# An Introduction to Bayesian Methods for the Social Sciences

David Kaplan

Department of Educational Psychology



7th *Mplus* Users' Meeting  
12 January 2016  
University of Utrecht

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

- The material for this workshop is drawn from Kaplan, D (2015), *Bayesian Statistics for the Social Sciences*. New York, Guilford Press.

©David Kaplan, 2013



Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

The screenshot shows a web browser displaying the Stats Camp website. The page title is "Euro Camp: Bayesian Analysis". The date is "April 25th - 29th 2016" at the "Mittland Hotel Utrecht BV". An overview states: "The five-day institute sessions will be Monday - Friday from 9:00 a.m. to 5:00 p.m. each day." A paragraph explains the camp's focus on Bayesian statistics and its popularity in applied research. There are three expandable sections: "More", "Venue", and "Transportation". A table lists registration fees for different groups, and a green "Register Now" button is at the bottom.

	Per Course	Qty*
Euro Camp Registration: Faculty (Early)	\$1,795.00 (USD)	0
Euro Camp Registration: Faculty (Regular)	\$2,295.00 (USD)	Rates Increase February 10, 2016
Euro Camp Registration: Faculty (Late)	\$2,795.00 (USD)	Rates Increase March 25, 2016
Euro Camp Registration: Student/Post Doc (Early)	\$1,095.00 (USD)	0
Euro Camp Registration: Student/Post Doc (Regular)	\$1,595.00 (USD)	Rates Increase February 10, 2016
Euro Camp Registration: Student/Post Doc (Late)	\$2,095.00 (USD)	Rates Increase March 25, 2016

Figure 1: <http://www.statscamp.org/eurocamp-bayesian-analysis>



---

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

*[I]t is clear that it is not possible to think about learning from experience and acting on it without coming to terms with Bayes' theorem. - Jerome Cornfield*



- Bayesian statistics has long been overlooked in the quantitative methods training of social scientists.
- Typically, the only introduction that a student might have to Bayesian ideas is a brief overview of Bayes' theorem while studying probability in an introductory statistics class.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

- Bayesian statistics has long been overlooked in the quantitative methods training of social scientists.
- Typically, the only introduction that a student might have to Bayesian ideas is a brief overview of Bayes' theorem while studying probability in an introductory statistics class.
  - ① Until recently, it was not feasible to conduct statistical modeling from a Bayesian perspective owing to its complexity and lack of availability.



- Bayesian statistics has long been overlooked in the quantitative methods training of social scientists.
- Typically, the only introduction that a student might have to Bayesian ideas is a brief overview of Bayes' theorem while studying probability in an introductory statistics class.
  - 1 Until recently, it was not feasible to conduct statistical modeling from a Bayesian perspective owing to its complexity and lack of availability.
  - 2 Bayesian statistics represents a powerful alternative to frequentist (classical) statistics, and is therefore, controversial.





- Bayesian statistics has long been overlooked in the quantitative methods training of social scientists.
- Typically, the only introduction that a student might have to Bayesian ideas is a brief overview of Bayes' theorem while studying probability in an introductory statistics class.
  - 1 Until recently, it was not feasible to conduct statistical modeling from a Bayesian perspective owing to its complexity and lack of availability.
  - 2 Bayesian statistics represents a powerful alternative to frequentist (classical) statistics, and is therefore, controversial.
- Bayesian statistical methods are now more popular owing to the development of powerful statistical software tools that make the estimation of complex models feasible from a Bayesian perspective.



# Outline

- Major differences between the Bayesian and frequentist paradigms of statistics
- Bayes' theorem
- Bayesian hypothesis testing
- Bayesian model building and evaluation
- An example of Bayesian linear regression
- Wrap-up: Additional philosophical issues

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



# Paradigm Differences

- For frequentists, the basic idea is that probability is represented by the model of **long run frequency**.
- Frequentist probability underlies the Fisher and Neyman-Pearson schools of statistics – the conventional methods of statistics we most often use.
- The frequentist formulation rests on the idea of equally probable and stochastically independent events
- The physical representation is the coin toss, which relates to the idea of a very large (actually infinite) number of repeated experiments.

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



- The entire structure of Neyman - Pearson hypothesis testing and Fisherian statistics (together referred to as the **frequentist school**) is based on frequentist probability.
- Our conclusions regarding null and alternative hypotheses presuppose the idea that we could conduct the same experiment an infinite number of times.
- Our interpretation of confidence intervals also assumes a fixed parameter and CIs that vary over an infinitely large number of identical experiments.



- But there is another view of probability as **subjective belief**.
- The physical model in this case is that of the “bet”.
- Consider the situation of betting on who will win the the World Cup (or the World Series).
- Here, probability is not based on an infinite number of repeatable and stochastically independent events, but rather on how much knowledge you have and how much you are willing to bet.
- Subjective probability allows one to address questions such as “what is the probability that my team will win the World Cup?” Relative frequency supplies information, but it is not the same as probability and can be quite different.
- This notion of subjective probability underlies Bayesian statistics.



# Bayes' Theorem

## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Consider the joint probability of two events,  $Y$  and  $X$ , for example observing lung cancer and smoking jointly.
- The joint probability can be written as



# Bayes' Theorem

## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Consider the joint probability of two events,  $Y$  and  $X$ , for example observing lung cancer and smoking jointly.
- The joint probability can be written as

$$p(\text{cancer}, \text{smoking}) = p(\text{cancer} | \text{smoking})p(\text{smoking}) \quad (1)$$



# Bayes' Theorem

## Bayes' Theorem

The  
Assumption of  
Exchangeability  
The Prior  
Distribution  
Conjugate  
Priors

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

- Consider the joint probability of two events,  $Y$  and  $X$ , for example observing lung cancer and smoking jointly.
- The joint probability can be written as

$$p(\text{cancer}, \text{smoking}) = p(\text{cancer} | \text{smoking})p(\text{smoking}) \quad (1)$$

- Similarly,





# Bayes' Theorem

## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Consider the joint probability of two events,  $Y$  and  $X$ , for example observing lung cancer and smoking jointly.
- The joint probability can be written as

$$p(\text{cancer}, \text{smoking}) = p(\text{cancer} | \text{smoking})p(\text{smoking}) \quad (1)$$

- Similarly,

$$p(\text{smoking}, \text{cancer}) = p(\text{smoking} | \text{cancer})p(\text{cancer}) \quad (2)$$



- Because these are symmetric, we can set them equal to each other to obtain the following

#### Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Because these are symmetric, we can set them equal to each other to obtain the following

$$p(\text{cancer}|\text{smoking})p(\text{smoking}) = p(\text{smoking}|\text{cancer})p(\text{cancer}) \quad (3)$$



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Because these are symmetric, we can set them equal to each other to obtain the following

$$p(\text{cancer}|\text{smoking})p(\text{smoking}) = p(\text{smoking}|\text{cancer})p(\text{cancer}) \quad (3)$$

$$p(\text{cancer}|\text{smoking}) = \frac{p(\text{smoking}|\text{cancer})p(\text{cancer})}{p(\text{smoking})} \quad (4)$$



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Because these are symmetric, we can set them equal to each other to obtain the following

$$p(\text{cancer}|\text{smoking})p(\text{smoking}) = p(\text{smoking}|\text{cancer})p(\text{cancer}) \quad (3)$$

$$p(\text{cancer}|\text{smoking}) = \frac{p(\text{smoking}|\text{cancer})p(\text{cancer})}{p(\text{smoking})} \quad (4)$$

- The inverse probability theorem (Bayes' theorem) states



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Because these are symmetric, we can set them equal to each other to obtain the following

$$p(\text{cancer}|\text{smoking})p(\text{smoking}) = p(\text{smoking}|\text{cancer})p(\text{cancer}) \quad (3)$$

$$p(\text{cancer}|\text{smoking}) = \frac{p(\text{smoking}|\text{cancer})p(\text{cancer})}{p(\text{smoking})} \quad (4)$$

- The inverse probability theorem (Bayes' theorem) states

$$p(\text{smoking}|\text{cancer}) = \frac{p(\text{cancer}|\text{smoking})p(\text{smoking})}{p(\text{cancer})} \quad (5)$$



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Why do we care?
- Because this is how you could go from the probability of having cancer given that the patient smokes, to the probability that the patient smokes given that he/she has cancer.
- We simply need the marginal probability of smoking and the marginal probability of cancer.
- We will refer to these marginal probabilities as *prior probabilities*.



# Statistical Elements of Bayes' Theorem

## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- What is the role of Bayes' theorem for statistical inference?
- Denote by  $Y$  a random variable that takes on a realized value  $y$ .
- For example, a person's socio-economic status could be considered a random variable taking on a very large set of possible values.
- Once the person identifies his/her socioeconomic status, the random variable  $Y$  is now realized as  $y$ .
- Because  $Y$  is unobserved and random, we need to specify a probability model to explain how we obtained the actual data values  $y$ .





## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Next, denote by  $\theta$  a parameter that we believe characterizes the probability model of interest.
- The parameter  $\theta$  can be a scalar, such as the mean or the variance of a distribution, or it can be vector-valued, such as a set of regression coefficients in regression analysis or factor loadings in factor analysis.
- We are concerned with determining the probability of observing  $y$  given the unknown parameters  $\theta$ , which we write as  $p(y|\theta)$ .
- In statistical inference, the goal is to obtain estimates of the unknown parameters given the data.
- This is expressed as the likelihood of the parameters given the data, often denoted as  $L(\theta|y)$ .



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- The key difference between Bayesian statistical inference and frequentist statistical inference concerns the nature of the unknown parameters  $\theta$ .
- In the frequentist tradition, the assumption is that  $\theta$  is unknown, but that estimation of this parameter does not incorporate uncertainty.
- In Bayesian statistical inference,  $\theta$  is considered unknown and random, possessing a probability distribution that reflects our uncertainty about the true value of  $\theta$ .
- Because both the observed data  $y$  and the parameters  $\theta$  are assumed random, we can model the joint probability of the parameters and the data as a function of the conditional distribution of the data given the parameters, and the prior distribution of the parameters.



- More formally,

### Bayes' Theorem

The  
Assumption of  
Exchangeability  
The Prior  
Distribution  
Conjugate  
Priors

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



- More formally,

$$p(\theta, y) = p(y|\theta)p(\theta). \quad (6)$$

## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



- More formally,

$$p(\theta, y) = p(y|\theta)p(\theta). \quad (6)$$

where  $p(\theta, y)$  is the joint distribution of the parameters and the data. Following Bayes' theorem described earlier, we obtain

## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- More formally,

$$p(\theta, y) = p(y|\theta)p(\theta). \quad (6)$$

where  $p(\theta, y)$  is the joint distribution of the parameters and the data. Following Bayes' theorem described earlier, we obtain

## Bayes' Theorem

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \quad (7)$$



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- More formally,

$$p(\theta, y) = p(y|\theta)p(\theta). \quad (6)$$

where  $p(\theta, y)$  is the joint distribution of the parameters and the data. Following Bayes' theorem described earlier, we obtain

## Bayes' Theorem

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \quad (7)$$

where  $p(\theta|y)$  is referred to as the *posterior distribution* of the parameters  $\theta$  given the observed data  $y$ .



## Bayes' Theorem

The  
Assumption of  
Exchangeability  
The Prior  
Distribution  
Conjugate  
Priors

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

- From equation (7) the posterior distribution of  $\theta$  given  $y$  is equal to the data distribution  $p(y|\theta)$  times the prior distribution of the parameters  $p(\theta)$  normalized by  $p(y)$  so that the posterior distribution sums (or integrates) to one.
- For discrete variables





## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- From equation (7) the posterior distribution of  $\theta$  given  $y$  is equal to the data distribution  $p(y|\theta)$  times the prior distribution of the parameters  $p(\theta)$  normalized by  $p(y)$  so that the posterior distribution sums (or integrates) to one.
- For discrete variables

$$p(y) = \sum_{\theta} p(y|\theta)p(\theta), \quad (8)$$



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- From equation (7) the posterior distribution of  $\theta$  given  $y$  is equal to the data distribution  $p(y|\theta)$  times the prior distribution of the parameters  $p(\theta)$  normalized by  $p(y)$  so that the posterior distribution sums (or integrates) to one.
- For discrete variables

$$p(y) = \sum_{\theta} p(y|\theta)p(\theta), \quad (8)$$

- For continuous variables



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- From equation (7) the posterior distribution of  $\theta$  given  $y$  is equal to the data distribution  $p(y|\theta)$  times the prior distribution of the parameters  $p(\theta)$  normalized by  $p(y)$  so that the posterior distribution sums (or integrates) to one.

- For discrete variables

$$p(y) = \sum_{\theta} p(y|\theta)p(\theta), \quad (8)$$

- For continuous variables

$$p(y) = \int_{\theta} p(y|\theta)p(\theta)d\theta, \quad (9)$$



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Notice that  $p(y)$  does not involve model parameters, so we can omit the term and obtain the *unnormalized posterior distribution*



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Notice that  $p(y)$  does not involve model parameters, so we can omit the term and obtain the *unnormalized posterior distribution*

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (10)$$



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Notice that  $p(y)$  does not involve model parameters, so we can omit the term and obtain the *unnormalized posterior distribution*

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (10)$$

- This can also be expressed in terms of the likelihood



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Notice that  $p(y)$  does not involve model parameters, so we can omit the term and obtain the *unnormalized posterior distribution*

$$p(\theta|y) \propto p(y|\theta)p(\theta). \quad (10)$$

- This can also be expressed in terms of the likelihood

$$p(\theta|y) \propto L(\theta|y)p(\theta). \quad (11)$$



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Equations (10) and (11) represents the core of Bayesian statistical inference and is what separates Bayesian statistics from frequentist statistics.
- Equation (11) states that our uncertainty regarding the parameters of our model, as expressed by the prior density  $p(\theta)$ , is *weighted* by the actual data  $p(y|\theta)$  (or equivalently,  $L(\theta|y)$ ), yielding an updated estimate of our uncertainty, as expressed in the posterior density  $p(\theta|y)$ .





# Exchangeability

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution  
Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- It is common in statistics to invoke the assumption that the data  $y_1, y_2, \dots, y_n$  are independently and identically distributed – often referred to as the *i.i.d* assumption.
- Bayesians invoke the deeper notion of *exchangeability* to produce likelihoods and address the issue of independence.
- Exchangeability arises from de Finetti's Representation Theorem and implies that the subscripts of a vector of data, e.g.  $y_1, y_2, \dots, y_n$  do not carry information that is relevant to describing the probability distribution of the data.
- In other words, the joint distribution of the data,  $p(y_1, y_2, \dots, y_n)$  is invariant to permutations of the subscripts.



- Consider the response that student  $i$  ( $i = 1, 2, \dots, 10$ ) would make to a PISA question “I like what I read in school”, where

$$y_i = \begin{cases} 1, & \text{if student } i \text{ agrees} \\ 0, & \text{if student } i \text{ disagrees} \end{cases} \quad (12)$$

Next, consider three patterns of responses by 10 randomly selected students

$$p(1, 0, 1, 1, 0, 1, 0, 1, 0, 0) \quad (13a)$$

$$p(1, 1, 0, 0, 1, 1, 1, 0, 0, 0) \quad (13b)$$

$$p(1, 0, 0, 0, 0, 0, 1, 1, 1, 1) \quad (13c)$$

- Note that there are actually  $2^{10}$  possible response patterns.



- If our task were to assign probabilities to all possible outcomes, this could become prohibitively difficult.
- However, suppose we now assume that student responses are independent of one another.
- Exchangeability implies that only the proportion of agreements matter, not the location of those agreements in the vector.
- Given that the sequences are the same length  $n$ , we can exchange the response of student  $i$  for student  $j$ , the without changing our belief about the probability model that generated that sequence.



- Exchangeability means that we believe that there is a parameter  $\theta$  that generates the observed data via a stochastic model and that we can describe that parameter without reference to the particular data at hand.
- The fact that we can describe  $\theta$  without reference to a particular set of data is, in fact, what is implied by the idea of a prior distribution.



# The Prior Distribution

- Why do we specify a prior distribution on the parameters?

Bayes'  
Theorem

The  
Assumption of  
Exchangeability

The Prior  
Distribution

Conjugate  
Priors

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



# The Prior Distribution

- Why do we specify a prior distribution on the parameters?
  - Our view is that progress in science generally comes about by learning from previous research findings and incorporating information from these findings into our present studies.

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



# The Prior Distribution

- Why do we specify a prior distribution on the parameters?
  - Our view is that progress in science generally comes about by learning from previous research findings and incorporating information from these findings into our present studies.
  - Information from previous research is almost always incorporated into our choice of designs, variables to be measured, or conceptual diagrams to be drawn.

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



# The Prior Distribution

- Why do we specify a prior distribution on the parameters?
  - Our view is that progress in science generally comes about by learning from previous research findings and incorporating information from these findings into our present studies.
  - Information from previous research is almost always incorporated into our choice of designs, variables to be measured, or conceptual diagrams to be drawn.
  - Bayesian statistical inference requires that our prior beliefs be made explicit, but then moderates our prior beliefs by the actual data in hand.

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up





# The Prior Distribution

- Why do we specify a prior distribution on the parameters?
  - Our view is that progress in science generally comes about by learning from previous research findings and incorporating information from these findings into our present studies.
  - Information from previous research is almost always incorporated into our choice of designs, variables to be measured, or conceptual diagrams to be drawn.
  - Bayesian statistical inference requires that our prior beliefs be made explicit, but then moderates our prior beliefs by the actual data in hand.
  - Moderation of our prior beliefs by the data in hand is the key meaning behind equations (10) and (11).

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



- But how do we choose a prior?

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- But how do we choose a prior?
  - The choice of a prior is based on how much information we believe we have prior to the data collection and how accurate we believe that information to be.



Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- But how do we choose a prior?
  - The choice of a prior is based on how much information we believe we have prior to the data collection and how accurate we believe that information to be.
  - Leamer (1983), who orders priors on the basis of degree of confidence.



Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- But how do we choose a prior?

- The choice of a prior is based on how much information we believe we have prior to the data collection and how accurate we believe that information to be.
- Leamer (1983), who orders priors on the basis of degree of confidence.
- Leamer's hierarchy of confidence is as follow: truths (e.g. axioms) > facts (data) > opinions (e.g. expert judgement) > conventions (e.g. pre-set alpha levels).



- But how do we choose a prior?

- The choice of a prior is based on how much information we believe we have prior to the data collection and how accurate we believe that information to be.
- Leamer (1983), who orders priors on the basis of degree of confidence.
- Leamer's hierarchy of confidence is as follow: truths (e.g. axioms) > facts (data) > opinions (e.g. expert judgement) > conventions (e.g. pre-set alpha levels).
- The strength of Bayesian inference lies precisely in its ability to incorporate existing knowledge into statistical specifications.



# Non-informative priors

- In some cases we may not be in possession of enough prior information to aid in drawing posterior inferences.
- From a Bayesian perspective, this lack of information is still important to consider and incorporate into our statistical specifications.
- In other words, it is equally important to quantify our ignorance as it is to quantify our cumulative understanding of a problem at hand.
- The standard approach to quantifying our ignorance is to incorporate a non-informative prior into our specification.
- Non-informative priors are also referred to as *vague* or *diffuse* priors.

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



- Perhaps the most common non-informative prior distribution is the uniform distribution  $U(\alpha, \beta)$  over some sensible range of values from  $\alpha$  to  $\beta$ .
- The uniform distribution indicates that we believe that the value of our parameter of interest lies in range  $\beta - \alpha$  and that all values have equal probability.
- Care must be taken in the choice of the range of values over the uniform distribution. The  $U[-\infty, \infty]$  is an *improper* prior distribution because it does not integrate to 1.0 as required of probability distributions.





# Informative-Conjugate Priors

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- It may be the case that some information can be brought to bear on a problem and be systematically incorporated into the prior distribution.
- Such “subjective” priors are called *informative*.
- One type of informative prior is based on the notion of a conjugate distribution.
- A conjugate prior distribution is one that, when combined with the likelihood function yields a posterior that is in the same distributional family as the prior distribution.



# Conjugate Priors for Some Common Distributions

Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

Data Distribution	Conjugate Prior
The normal distribution	The normal distribution or Uniform Distribution
The Poisson distribution	The gamma distribution
The binomial distribution	The Beta Distribution
The multinomial distribution	The Dirichlet Distribution



## Bayes' Theorem

The Assumption of Exchangeability  
The Prior Distribution

### Conjugate Priors

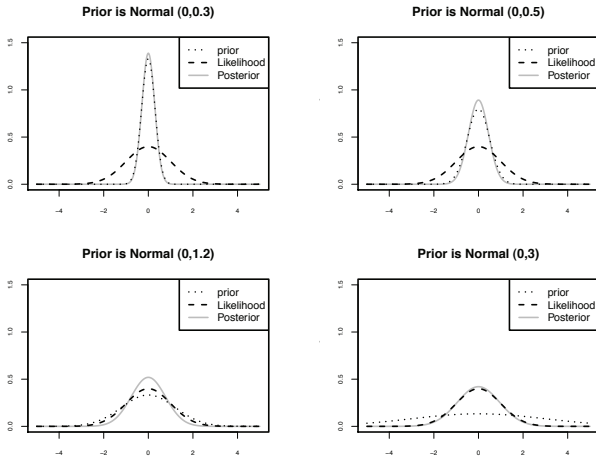
Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up



**Figure 2:** Normal distribution, mean unknown/variance known with varying conjugate priors



## Bayes' Theorem

The Assumption of Exchangeability

The Prior Distribution

Conjugate Priors

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

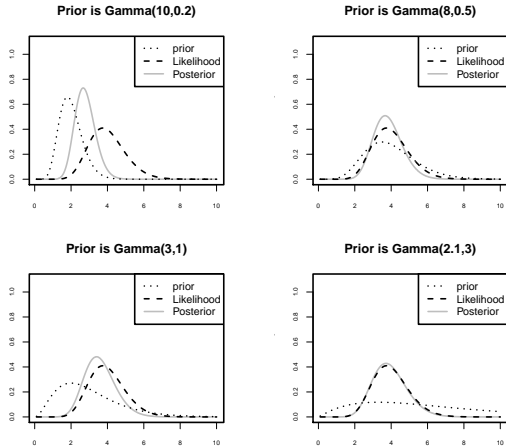


Figure 3: Poisson distribution with varying gamma-density priors



# Bayesian Hypothesis Testing

- A critically important component of applied statistics is hypothesis testing.
- The approach most widely used in the social and behavioral sciences is the Neyman-Pearson approach.
- An interesting aspect of the Neyman-Pearson approach to hypothesis testing is that students (as well as many seasoned researchers) appear to have a very difficult time grasping its principles.
- Gigerenzer (2004) argued that much of the difficulty in grasping frequentist hypothesis testing lies in the conflation of Fisherian hypothesis testing and the Neyman-Pearson approach to hypothesis testing.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Point  
Summaries

Interval  
Summaries

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



- Fisher's early approach to hypothesis testing required specifying only the null hypothesis.
- A conventional significance level is chosen (usually the 5% level).
- Once the test is conducted, the result is either significant ( $p < .05$ ) or it is not ( $p > .05$ ).
- If the resulting test is significant, then the null hypothesis is rejected. However, if the resulting test is not significant, then no conclusion can be drawn.
- Fisher's approach was based on looking at how data could inform evidence for a hypothesis.



- The Neyman and Pearson approach requires that two hypotheses be specified – the null and alternative hypothesis – and is designed to inform specific sets of actions. It's about making a choice, not about evidence against a hypothesis.
- By specifying two hypotheses, one can compute a desired tradeoff between two types of errors: Type I errors ( $\alpha$ ) and Type II errors ( $\beta$ )
- The goal is not to assess the evidence against a hypothesis (or model) taken as true. Rather, it is whether the data provide support for taking one of two competing sets of actions.
- In fact, prior belief as well as interest in “the truth” of the hypothesis is irrelevant – only a decision that has a low probability of being the wrong decision is relevant.



- The conflation of Fisherian and Neyman-Pearson hypothesis testing lies in the use and interpretation of the *p-value*.
- In Fisher's paradigm, the *p-value* is a matter of convention with the resulting outcome being based on the data.
- In the Neyman-Pearson paradigm,  $\alpha$  and  $\beta$  are determined prior to the experiment being conducted and refer to a consideration of the cost of making one or the other error.
- Indeed, in the Neyman-Pearson approach, the problem is one of finding a balance between  $\alpha$ , power, and sample size.
- The Neyman-Pearson approach is best suited for experimental planning. Sometimes, it is used this way, followed by the Fisherian approach for judging evidence. But, these two ideas may be incompatible (Royall, 1997).





- However, this balance is virtually always ignored and  $\alpha = 0.05$  is used.
- The point is that the  $p$ -value and  $\alpha$  are not the same thing.
- This is made worse by the fact that statistical software packages often report a number of  $p$ -values that a researcher can choose from after having conducted the analysis (e.g., .001, .01, .05).
- This can lead a researcher to set  $\alpha$  ahead of time, but then communicate a different level of “significance” after running the test.
- The conventional practice is even worse than described, as evidenced by nonsensical phrases such as results



- However, this balance is virtually always ignored and  $\alpha = 0.05$  is used.
- The point is that the  $p$ -value and  $\alpha$  are not the same thing.
- This is made worse by the fact that statistical software packages often report a number of  $p$ -values that a researcher can choose from after having conducted the analysis (e.g., .001, .01, .05).
- This can lead a researcher to set  $\alpha$  ahead of time, but then communicate a different level of “significance” after running the test.
- The conventional practice is even worse than described, as evidenced by nonsensical phrases such as results

“trending toward significance”.



- The problem with NHST, therefore, seems to lie with the interpretation of the  $p$ -value.
- To be precise, the  $p$ -value rests on a form of argumentation referred to as *reductio ad absurdum*.
- This form of argumentation attempts to establish an assertion by deriving an absurdity from its denial, thus arguing that the assertion must be accepted because its rejection would be indefensible.
- The  $p$ -value as reduction ad absurdum is thus properly interpreted as the probability of observing data at least as extreme as the data that was actually observed, computed under the assumption that the null hypothesis is true.



- The  $p$ -value is based on data that were never observed.
- The  $p$ -value is obtained by referencing the value of the observed test statistic (e.g. the  $t$ -test) based on the study at hand to hypothetically replicated data generated under the assumption that the null hypothesis is true.
- The null hypothesis is never true in reality, and moreover, it is typically not the research question of interest.
- Thus, after Cohen (1994), researchers are typically testing a “nil” hypothesis that is hopefully rejected.
- However, it is important to emphasize that there is nothing within the NHST paradigm that requires testing a null hypothesis of no effect. In fact, any theoretically justifiable hypothesis can serve as the null hypothesis.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Point  
Summaries  
Interval  
Summaries

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

- To summarize, the  $p$ -value does not quantify evidence for a hypothesis.
- Rather, it provides a measure of the probability of an outcome not actually observed, computed under the assumption of a null hypothesis that will likely not to be true in any population and, moreover, is typically not of interest.
- To quote Jeffreys (1961, pg. 385)



- To summarize, the  $p$ -value does not quantify evidence for a hypothesis.
- Rather, it provides a measure of the probability of an outcome not actually observed, computed under the assumption of a null hypothesis that will likely not to be true in any population and, moreover, is typically not of interest.
- To quote Jeffreys (1961, pg. 385)

“This seems a remarkable procedure.”



- Misunderstanding the Fisher or Neyman-Pearson framework to hypothesis testing and/or poor methodological practice is not a criticism of the approach per se.
- However, from the frequentist point of view, a criticism often leveled at the Bayesian approach to statistical inference is that it is “subjective”, while the frequentist approach is “objective”.
- The objection to “subjectivism” is perplexing insofar as frequentist hypothesis testing also rests on assumptions that do not involve data.
- The simplest and most ubiquitous example is the choice of variables in a regression equation or asymptotic distributions of test statistics.



# Point Summaries of the Posterior Distribution

Bayes' Theorem

Bayesian Hypothesis Testing

Point Summaries

Interval Summaries

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

- Hypothesis testing begins first by obtaining summaries of relevant distributions.
- The difference between Bayesian and frequentist statistics is that with Bayesian statistics we wish to obtain summaries of the posterior distribution.
- The expressions for the mean and variance of the posterior distribution come from expressions for the mean and variance of conditional distributions generally.





- For the continuous case, the mean of the posterior distribution of  $\theta$  given the data  $y$  is referred to as the *expected a posteriori* or EAP estimate and can be written as

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Point  
Summaries

Interval  
Summaries

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



- For the continuous case, the mean of the posterior distribution of  $\theta$  given the data  $y$  is referred to as the *expected a posteriori* or EAP estimate and can be written as

$$E(\theta|y) = \int_{-\infty}^{+\infty} \theta p(\theta|y) d\theta. \quad (14)$$

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Point  
Summaries

Interval  
Summaries

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up



- For the continuous case, the mean of the posterior distribution of  $\theta$  given the data  $y$  is referred to as the *expected a posteriori* or EAP estimate and can be written as

$$E(\theta|y) = \int_{-\infty}^{+\infty} \theta p(\theta|y) d\theta. \quad (14)$$

Similarly, the variance of posterior distribution of  $\theta$  given  $y$  can be obtained as



- For the continuous case, the mean of the posterior distribution of  $\theta$  given the data  $y$  is referred to as the *expected a posteriori* or EAP estimate and can be written as

$$E(\theta|y) = \int_{-\infty}^{+\infty} \theta p(\theta|y) d\theta. \quad (14)$$

Similarly, the variance of posterior distribution of  $\theta$  given  $y$  can be obtained as

$$\begin{aligned} \text{var}(\theta|y) &= E[(\theta - E[(\theta|y)])^2|y], \\ &= \int_{-\infty}^{+\infty} (\theta - E[\theta|y])^2 p(\theta|y) d\theta. \end{aligned} \quad (15)$$



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Point  
Summaries

Interval  
Summaries

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

- Another common summary measure would be the mode of the posterior distribution – referred to as the *maximum a posteriori* (MAP) estimate.
- The MAP begins with the idea of maximum likelihood estimation. The MAP can be written as



- Another common summary measure would be the mode of the posterior distribution – referred to as the *maximum a posteriori* (MAP) estimate.
- The MAP begins with the idea of maximum likelihood estimation. The MAP can be written as

$$\hat{\theta}_{MAP} = \arg \max_{\theta} L(\theta|y)p(\theta). \quad (16)$$

where  $\arg \max_{\theta}$  stands for “maximizing the value of the argument”.



## ● Posterior Probability Intervals

- In addition to point summary measures, it may also be desirable to provide interval summaries of the posterior distribution.
- Recall that the frequentist confidence interval requires two
  - imagine an infinite number of repeated samples from the population characterized by  $\mu$ .
- For any given sample, we can obtain the sample mean  $\bar{x}$  and then form a  $100(1 - \alpha)\%$  confidence interval.
- The correct frequentist interpretation is that  $100(1 - \alpha)\%$  of the confidence intervals formed this way capture the true parameter  $\mu$  under the null hypothesis. Notice that the probability that the parameter is in the interval is either zero or one.



## ● Posterior Probability Intervals (cont'd)

- In contrast, the Bayesian framework assumes that a parameter has a probability distribution.
- Sampling from the posterior distribution of the model parameters, we can obtain its quantiles. From the quantiles, we can directly obtain the probability that a parameter lies within a particular interval.
- So, a 95% posterior probability interval would mean that the probability that the parameter lies in the interval is 0.95.
- The posterior probability interval can, of course, include zero. Zero is a credible value for the parameter.
- Notice that this is entirely different from the frequentist interpretation, and arguably aligns with common sense.





# Bayesian Model Building and Evaluation

- The frequentist and Bayesian goals of model building are the same.

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

Posterior Predictive Checking

Bayes Factors

Bayesian Model Averaging

MCMC

An Example

Wrap-up



# Bayesian Model Building and Evaluation

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The frequentist and Bayesian goals of model building are the same.
- ① Model specification based on prior knowledge



# Bayesian Model Building and Evaluation

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

**Bayesian  
Model  
Building and  
Evaluation**

Posterior  
Predictive  
Checking

Bayes Factors  
Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The frequentist and Bayesian goals of model building are the same.
  - 1 Model specification based on prior knowledge
  - 2 Model estimation and fitting to data obtained from a relevant population



# Bayesian Model Building and Evaluation

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The frequentist and Bayesian goals of model building are the same.
  - 1 Model specification based on prior knowledge
  - 2 Model estimation and fitting to data obtained from a relevant population
  - 3 Model evaluation and modification



# Bayesian Model Building and Evaluation

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

**Bayesian  
Model  
Building and  
Evaluation**

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The frequentist and Bayesian goals of model building are the same.
  - 1 Model specification based on prior knowledge
  - 2 Model estimation and fitting to data obtained from a relevant population
  - 3 Model evaluation and modification
  - 4 Model choice



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- Despite these similarities there are important differences.
- A major difference between the Bayesian and frequentist goals of model building lie in the model specification stage.
- Because the Bayesian perspective explicitly incorporates uncertainty regarding model parameters using priors, the first phase of modeling building requires the specification of a full probability model for the data and the parameters of the model.
- Model fit implies that the full probability model fits the data. Lack of model fit may be due to incorrect specification of likelihood, the prior distribution, or both.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors  
Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- Another difference between the Bayesian and frequentist goals of model building relates to the justification for choosing a particular model among a set of competing models.
- Model building and model choice in the frequentist domain is based primarily on choosing the model that best fits the data.
- This has certainly been the key motivation for model building, respecification, and model choice in the context of structural equation modeling (Kaplan 2009).
- In the Bayesian domain, the choice among a set of competing models is based on which model provides the best posterior predictions.
- That is, the choice among a set of competing models should be based on which model will best predict what actually happened.



# Posterior Predictive Checking

- A very natural way of evaluating the quality of a model is to examine how well the model fits the actual data.
- In the context of Bayesian statistics, the approach to examining how well a model predicts the data is based on the notion of *posterior predictive checks*, and the accompanying *posterior predictive p-value*.
- The general idea behind posterior predictive checking is that there should be little, if any, discrepancy between data generated by the model, and the actual data itself.
- Posterior predictive checking is a method for assessing the specification quality of the model. Any deviation between the data generated from the model and the actual data implies model misspecification.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors  
Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up





- In the Bayesian context, the approach to examining model fit and specification utilizes the posterior predictive distribution of replicated data.
- Let  $y^{rep}$  be data replicated from our current model.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors  
Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



- In the Bayesian context, the approach to examining model fit and specification utilizes the posterior predictive distribution of replicated data.
- Let  $y^{rep}$  be data replicated from our current model.

## Posterior Predictive Distribution

$$\begin{aligned} p(y^{rep}|y) &= \int p(y^{rep}|\theta)p(\theta|y)d\theta \\ &= \int p(y^{rep}|\theta)p(y|\theta)p(\theta)d\theta \end{aligned} \tag{17}$$

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors  
Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



- In the Bayesian context, the approach to examining model fit and specification utilizes the posterior predictive distribution of replicated data.
- Let  $y^{rep}$  be data replicated from our current model.

## Posterior Predictive Distribution

$$\begin{aligned} p(y^{rep}|y) &= \int p(y^{rep}|\theta)p(\theta|y)d\theta \\ &= \int p(y^{rep}|\theta)p(y|\theta)p(\theta)d\theta \end{aligned} \quad (17)$$

- Equation (17) states that the distribution of future observations given the present data,  $p(y^{rep}|y)$ , is equal to the probability distribution of the future observations given the parameters,  $p(y^{rep}|\theta)$ , weighted by the posterior distribution of the model parameters.



- To assess model fit, posterior predictive checking implies that the replicated data should match the observed data quite closely if we are to conclude that the model fits the data.
- One approach to model fit in the context of posterior predictive checking is based on Bayesian  $p$ -values.
- Denote by  $T(y)$  a test statistic (e.g.  $\chi^2$ ) based on the data, and let  $T(y^{rep})$  be the same test statistics for the replicated data (based on MCMC). Then, the Bayesian  $p$ -value is defined to be



- To assess model fit, posterior predictive checking implies that the replicated data should match the observed data quite closely if we are to conclude that the model fits the data.
- One approach to model fit in the context of posterior predictive checking is based on Bayesian  $p$ -values.
- Denote by  $T(y)$  a test statistic (e.g.  $\chi^2$ ) based on the data, and let  $T(y^{rep})$  be the same test statistics for the replicated data (based on MCMC). Then, the Bayesian  $p$ -value is defined to be

$$p\text{-value} = pr[T(y^{rep}, \theta) \geq T(y, \theta) | y]. \quad (18)$$



- To assess model fit, posterior predictive checking implies that the replicated data should match the observed data quite closely if we are to conclude that the model fits the data.
- One approach to model fit in the context of posterior predictive checking is based on Bayesian  $p$ -values.
- Denote by  $T(y)$  a test statistic (e.g.  $\chi^2$ ) based on the data, and let  $T(y^{rep})$  be the same test statistics for the replicated data (based on MCMC). Then, the Bayesian  $p$ -value is defined to be

$$p\text{-value} = pr[T(y^{rep}, \theta) \geq T(y, \theta) | y]. \quad (18)$$

- The  $p$ -value is the proportion of replicated test values that equal or exceed the observed test value. High (or low if signs are reversed) values indicate poor model fit.



# Bayes Factors

- A very simple and intuitive approach to model building and model selection uses so-called *Bayes factors* (Kass & Raftery, 1995)
- The Bayes factor provides a way to quantify the odds that the data favor one hypothesis over another. A key benefit of Bayes factors is that models do not have to be nested.
- Consider two competing models, denoted as  $M_1$  and  $M_2$ , that could be nested within a larger space of alternative models. Let  $\theta_1$  and  $\theta_2$  be the two parameter vectors associated with these two models.
- These could be two regression models with a different number of variables, or two structural equation models specifying very different directions of mediating effects.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



- The goal is to develop a quantity that expresses the extent to which the data support  $M_1$  over  $M_2$ . One quantity could be the posterior odds of  $M_1$  over  $M_2$ , expressed as

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up





- The goal is to develop a quantity that expresses the extent to which the data support  $M_1$  over  $M_2$ . One quantity could be the posterior odds of  $M_1$  over  $M_2$ , expressed as

## Bayes Factors

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \times \left[ \frac{p(M_1)}{p(M_2)} \right]. \quad (19)$$

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



- The goal is to develop a quantity that expresses the extent to which the data support  $M_1$  over  $M_2$ . One quantity could be the posterior odds of  $M_1$  over  $M_2$ , expressed as

## Bayes Factors

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \times \left[ \frac{p(M_1)}{p(M_2)} \right]. \quad (19)$$

- Notice that the first term on the right hand side of equation (19) is the ratio of two integrated likelihoods.
- This ratio is referred to as the *Bayes factor* for  $M_1$  over  $M_2$ , denoted here as  $B_{12}$ .
- Our prior opinion regarding the odds of  $M_1$  over  $M_2$ , given by  $p(M_1)/p(M_2)$  is weighted by our consideration of the data, given by  $p(y|M_1)/p(y|M_2)$ .



- This weighting gives rise to our updated view of evidence provided by the data for either hypothesis, denoted as  $p(M_1|y)/p(M_2|y)$ .
- An inspection of equation (19) also suggests that the Bayes factor is the ratio of the posterior odds to the prior odds.
- In practice, there may be no prior preference for one model over the other. In this case, the prior odds are neutral and  $p(M_1) = p(M_2) = 1/2$ .
- When the prior odds ratio equals 1, then the posterior odds is equal to the Bayes factor.



# Bayesian Information Criterion

- A popular measure for model selection used in both frequentist and Bayesian applications is based on an approximation of the Bayes factor and is referred to as the *Bayesian information criterion* (BIC), also referred to as the Schwarz criterion.
- Consider two models,  $M_1$  and  $M_2$  with  $M_2$  nested in  $M_1$ . Under conditions where there is little prior information, the BIC can be written as

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Bayesian Information Criterion

- A popular measure for model selection used in both frequentist and Bayesian applications is based on an approximation of the Bayes factor and is referred to as the *Bayesian information criterion* (BIC), also referred to as the Schwarz criterion.
- Consider two models,  $M_1$  and  $M_2$  with  $M_2$  nested in  $M_1$ . Under conditions where there is little prior information, the BIC can be written as

$$BIC = -2 \log(\hat{\theta}|y) + p \log(n) \quad (20)$$

where  $-2 \log \hat{\theta}|y$  describes model fit while  $p \log(n)$  is a penalty for model complexity, where  $p$  represents the number of variables in the model and  $n$  is the sample size.

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

Posterior Predictive Checking

Bayes Factors

Bayesian Model Averaging

MCMC

An Example

Wrap-up



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The BIC is often used for model comparisons. The difference between two BIC measures comparing, say  $M_1$  to  $M_2$  can be written as



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The BIC is often used for model comparisons. The difference between two BIC measures comparing, say  $M_1$  to  $M_2$  can be written as

$$\begin{aligned}\Delta(BIC_{12}) &= BIC_{(M_1)} - BIC_{(M_2)} \\ &= \log(\hat{\theta}_1|y) - \log(\hat{\theta}_2|y) - \frac{1}{2}(p_1 - p_2) \log(n)\end{aligned}\tag{21}$$



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- The BIC is often used for model comparisons. The difference between two BIC measures comparing, say  $M_1$  to  $M_2$  can be written as

$$\begin{aligned}\Delta(BIC_{12}) &= BIC_{(M_1)} - BIC_{(M_2)} \\ &= \log(\hat{\theta}_1|y) - \log(\hat{\theta}_2|y) - \frac{1}{2}(p_1 - p_2) \log(n)\end{aligned}\tag{21}$$





Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- Rules of thumb have been developed to assess the quality of the evidence favoring one hypothesis over another using Bayes factors and the comparison of BIC values from two competing models. Using  $M_1$  as the reference model,



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- Rules of thumb have been developed to assess the quality of the evidence favoring one hypothesis over another using Bayes factors and the comparison of BIC values from two competing models. Using  $M_1$  as the reference model,

BIC Difference	Bayes Factor	Evidence against $M_2$
0 to 2	1 to 3	Weak
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very strong



# Deviance Information Criterion

- The BIC (ironically) is not fundamentally Bayesian
- An explicitly Bayesian approach to model comparison has been developed based on the notion of *Bayesian deviance*.
- Define *Bayesian deviance* as

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Deviance Information Criterion

- The BIC (ironically) is not fundamentally Bayesian
- An explicitly Bayesian approach to model comparison has been developed based on the notion of *Bayesian deviance*.
- Define *Bayesian deviance* as

## Bayes Deviance

$$D(\theta) = -2 \log[p(y|\theta)] + 2 \log[h(y)], \quad (22)$$

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Deviance Information Criterion

- The BIC (ironically) is not fundamentally Bayesian
- An explicitly Bayesian approach to model comparison has been developed based on the notion of *Bayesian deviance*.
- Define *Bayesian deviance* as

## Bayes Deviance

$$D(\theta) = -2 \log[p(y|\theta)] + 2 \log[h(y)], \quad (22)$$

- To make this Bayesian this, we obtain a posterior mean over  $\theta$  by defining

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Deviance Information Criterion

- The BIC (ironically) is not fundamentally Bayesian
- An explicitly Bayesian approach to model comparison has been developed based on the notion of *Bayesian deviance*.
- Define *Bayesian deviance* as

## Bayes Deviance

$$D(\theta) = -2 \log[p(y|\theta)] + 2 \log[h(y)], \quad (22)$$

- To make this Bayesian this, we obtain a posterior mean over  $\theta$  by defining

$$\overline{D(\theta)} = E_{\theta}[-2 \log[p(y|\theta)|y] + 2 \log[h(y)]]. \quad (23)$$

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



- Let  $D(\bar{\theta})$  be a posterior estimate of  $\theta$ . We can define the *effective dimension* of the model as  $q_D = \overline{D(\theta)} - D(\bar{\theta})$ .
- We then add the model fit term  $\overline{D(\theta)}$  to obtain the *deviance information criterion* (DIC) - namely,

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



- Let  $D(\bar{\theta})$  be a posterior estimate of  $\theta$ . We can define the *effective dimension* of the model as  $q_D = \overline{D(\theta)} - D(\bar{\theta})$ .
- We then add the model fit term  $\overline{D(\theta)}$  to obtain the *deviance information criterion* (DIC) - namely,

## Deviance Information Criterion

$$DIC = \overline{D(\theta)} + q_D = 2\overline{D(\theta)} - D(\bar{\theta}). \quad (24)$$





- Let  $D(\bar{\theta})$  be a posterior estimate of  $\theta$ . We can define the *effective dimension* of the model as  $q_D = \overline{D(\theta)} - D(\bar{\theta})$ .
- We then add the model fit term  $\overline{D(\theta)}$  to obtain the *deviance information criterion* (DIC) - namely,

## Deviance Information Criterion

$$DIC = \overline{D(\theta)} + q_D = 2\overline{D(\theta)} - D(\bar{\theta}). \quad (24)$$

- The DIC can be obtained by calculating equation (23) over MCMC samples.
- Models with the lowest DIC values are preferred.



# Bayesian Model Averaging

- The selection of a particular model from a universe of possible models can also be characterized as a problem of uncertainty. This problem was succinctly stated by Hoeting, Raftery & Madigan (1999) who write

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Bayesian Model Averaging

- The selection of a particular model from a universe of possible models can also be characterized as a problem of uncertainty. This problem was succinctly stated by Hoeting, Raftery & Madigan (1999) who write

*"Standard statistical practice ignores model uncertainty. Data analysts typically select a model from some class of models and then proceed as if the selected model had generated the data. This approach ignores the uncertainty in model selection, leading to over-confident inferences and decisions that are more risky than one thinks they are."(pg. 382)*

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Bayesian Model Averaging

- The selection of a particular model from a universe of possible models can also be characterized as a problem of uncertainty. This problem was succinctly stated by Hoeting, Raftery & Madigan (1999) who write

*"Standard statistical practice ignores model uncertainty. Data analysts typically select a model from some class of models and then proceed as if the selected model had generated the data. This approach ignores the uncertainty in model selection, leading to over-confident inferences and decisions that are more risky than one thinks they are."*(pg. 382)

- An approach to addressing the problem is the method of *Bayesian model averaging* (BMA). We will show this in the regression example.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Bayesian Model Averaging

- Consider a quantity of interest such as a future observation denoted as  $\Upsilon$ .
- Next, consider a set of competing models  $M_k$ ,  $k = 1, 2, \dots, K$  that are not necessarily nested.
- The posterior distribution of  $\Upsilon$  given data  $y$  can be written as

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Bayesian Model Averaging

- Consider a quantity of interest such as a future observation denoted as  $\Upsilon$ .
- Next, consider a set of competing models  $M_k$ ,  $k = 1, 2, \dots, K$  that are not necessarily nested.
- The posterior distribution of  $\Upsilon$  given data  $y$  can be written as

$$p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y). \quad (25)$$

where  $p(M_k|y)$  is the posterior probability of model  $M_k$  written as

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



# Bayesian Model Averaging

- Consider a quantity of interest such as a future observation denoted as  $\Upsilon$ .
- Next, consider a set of competing models  $M_k$ ,  $k = 1, 2, \dots, K$  that are not necessarily nested.
- The posterior distribution of  $\Upsilon$  given data  $y$  can be written as

$$p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y). \quad (25)$$

where  $p(M_k|y)$  is the posterior probability of model  $M_k$  written as

$$p(M_k|y) = \frac{p(y|M_k)p(M_k)}{\sum_{l=1}^K p(y|M_l)p(M_l)}, \quad l \neq k. \quad (26)$$



- The interesting feature of equation (26) is that  $p(M_k|y)$  will likely be different for different models.
- The term  $p(y|M_k)$  can be expressed as an integrated likelihood

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up





- The interesting feature of equation (26) is that  $p(M_k|y)$  will likely be different for different models.
- The term  $p(y|M_k)$  can be expressed as an integrated likelihood

$$p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k, \quad (27)$$

where  $p(\theta_k|M_k)$  is the prior distribution of  $\theta_k$  under model  $M_k$  (Raftery et al., 1997).



- The interesting feature of equation (26) is that  $p(M_k|y)$  will likely be different for different models.
- The term  $p(y|M_k)$  can be expressed as an integrated likelihood

$$p(y|M_k) = \int p(y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k, \quad (27)$$

where  $p(\theta_k|M_k)$  is the prior distribution of  $\theta_k$  under model  $M_k$  (Raftery et al., 1997).

- BMA provides an approach for combining models specified by researchers.
- Madigan and Raftery (1994) show that BMA provides better predictive performance than that of a single model based on a log-score rule.



- BMA is difficult to implement.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- BMA is difficult to implement.

- 1 The number of terms in  $p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y)$  can be quite large and the corresponding integrals are hard to compute.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- BMA is difficult to implement.

- 1 The number of terms in  $p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y)$  can be quite large and the corresponding integrals are hard to compute.
- 2 Eliciting  $p(M_k)$  may not be straightforward. The uniform prior  $1/M$  is often used.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- BMA is difficult to implement.

- 1 The number of terms in  $p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y)$  can be quite large and the corresponding integrals are hard to compute.
- 2 Eliciting  $p(M_k)$  may not be straightforward. The uniform prior  $1/M$  is often used.
- 3 Choosing the class of models to average over is also challenging.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

Posterior  
Predictive  
Checking

Bayes Factors

Bayesian  
Model  
Averaging

MCMC

An Example

Wrap-up

- BMA is difficult to implement.

- 1 The number of terms in  $p(\Upsilon|y) = \sum_{k=1}^K p(\Upsilon|M_k)p(M_k|y)$  can be quite large and the corresponding integrals are hard to compute.
- 2 Eliciting  $p(M_k)$  may not be straightforward. The uniform prior  $1/M$  is often used.
- 3 Choosing the class of models to average over is also challenging.

- The problem of reducing the overall number of models that one could incorporate in the summation has led to solutions based on *Occam's window* and *Markov Chain Monte Carlo Model Composition* –  $M^3$  (Madigan and Raftery, 1994).



- The increased popularity of Bayesian methods in the social and behavioral sciences has been the (re)-discovery of numerical algorithms for estimating the posterior distribution of the model parameters given the data.
- It was virtually impossible to analytically derive summary measures of the posterior distribution, particularly for complex models with many parameters.
- Instead of analytically solving for estimates of a complex posterior distribution, we can instead draw samples from  $p(\theta|y)$  and summarize the distribution formed by those samples. This is referred to as *Monte Carlo integration*.
- The two most popular methods of MCMC are the Gibbs sampler and the Metropolis-Hastings algorithm.

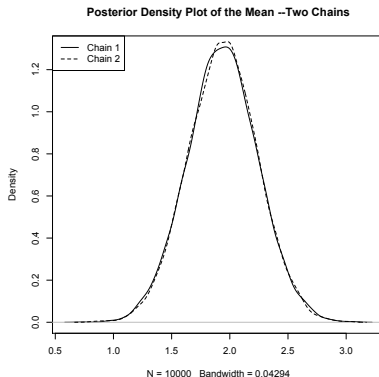




- A decision must be made regarding the number of Markov chains to be generated, as well as the number of iterations of the sampler.
- Each chain samples from another location of the posterior distribution based on starting values.
- With multiple chains, fewer iterations are required, particularly if there is evidence for the chains converging to the same posterior mean for each parameter.
- Once the chain has stabilized, the *burn-in* samples are discarded.
- Summaries of the posterior distribution as well as convergence diagnostics are calculated on the post-burn-in iterations.



- The goal of this simple example is to obtain the posterior distribution of the mean of the normal distribution with known variance where the prior distributions are conjugate. That is, the prior distribution for the mean is  $N(0, 1)$  and the prior distribution for the precision parameter is  $inv\text{-}gamma(a, b)$





Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

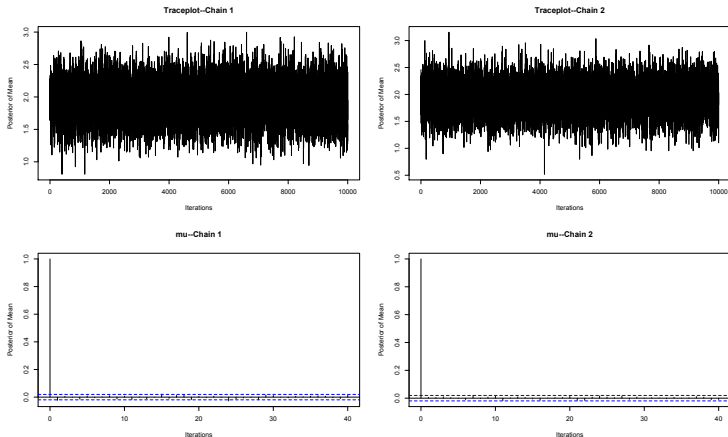


Figure 5: Trace and ACF plots for two chains



Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

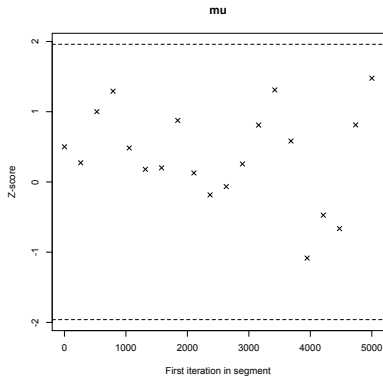
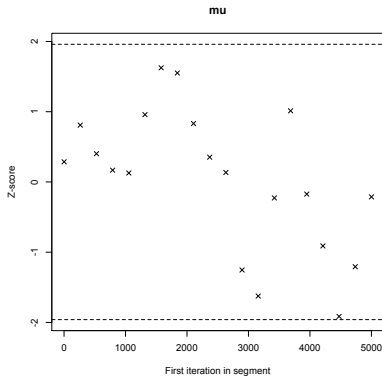


Figure 6: Geweke plots



Bayes'  
Theorem

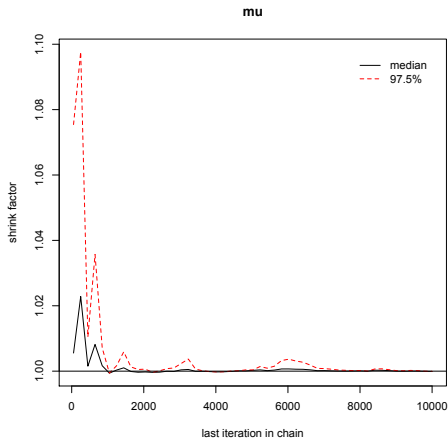
Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

**MCMC**

An Example

Wrap-up



**Figure 7:** Gelman-Rubin-Brooks plot



# Bayesian Linear Regression

- We will start with the basic multiple regression model.
- Let  $\mathbf{y}$  be an  $n$ -dimensional vector  $(y_1, y_2, \dots, y_n)'$  ( $i = 1, 2, \dots, n$ ) of reading scores from  $n$  students on the PISA 2009 reading assessment, and let  $\mathbf{X}$  be an  $n \times k$  matrix containing  $k$  background and attitude measures. Then, the normal linear regression model can be written as

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Bayesian Linear Regression

Wrap-up



# Bayesian Linear Regression

- We will start with the basic multiple regression model.
- Let  $\mathbf{y}$  be an  $n$ -dimensional vector  $(y_1, y_2, \dots, y_n)'$  ( $i = 1, 2, \dots, n$ ) of reading scores from  $n$  students on the PISA 2009 reading assessment, and let  $\mathbf{X}$  be an  $n \times k$  matrix containing  $k$  background and attitude measures. Then, the normal linear regression model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (28)$$

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Bayesian Linear Regression

Wrap-up



# Bayesian Linear Regression

- We will start with the basic multiple regression model.
- Let  $\mathbf{y}$  be an  $n$ -dimensional vector  $(y_1, y_2, \dots, y_n)'$  ( $i = 1, 2, \dots, n$ ) of reading scores from  $n$  students on the PISA 2009 reading assessment, and let  $\mathbf{X}$  be an  $n \times k$  matrix containing  $k$  background and attitude measures. Then, the normal linear regression model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (28)$$

- All the usual regression assumptions apply
- We assume that student level PISA reading scores are generated from a normal distribution.





Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Bayesian  
Linear  
Regression

Wrap-up

- The sample comes from approximately PISA (2009)-eligible students in the United States ( $N \sim 5000$ ).
- Background variables: gender (gender), immigrant status (native), language (1 = test language is the same as language at home, 0 otherwise) and economic, social, and cultural status of the student (escs).
- Student reading attitudes: enjoyment of reading reading (joyread), diversity in reading (divread), memorization strategies (memor), elaboration strategies (elab), control strategies (cstrat).
- The outcome was the first plausible value of the PISA 2009 reading assessment.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Bayesian  
Linear  
Regression

Wrap-up

- Recall that the components of Bayes' theorem require priors on all model parameters.
- We will write the probability distribution for the regression model as



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Bayesian  
Linear  
Regression

Wrap-up

- Recall that the components of Bayes' theorem require priors on all model parameters.
- We will write the probability distribution for the regression model as

$$p(\mathbf{X}, \mathbf{y} | \boldsymbol{\beta}, \sigma^2) \quad (29)$$



- Recall that the components of Bayes' theorem require priors on all model parameters.
- We will write the probability distribution for the regression model as

$$p(\mathbf{X}, \mathbf{y} | \boldsymbol{\beta}, \sigma^2) \quad (29)$$

- Conventional statistics stops here and estimates the model parameters with either maximum likelihood estimation or ordinary least square.
- But for Bayesian regression we need to specify the priors for all model parameters.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Bayesian  
Linear  
Regression

Wrap-up

- First consider non-informative priors.
- In the context of the normal linear regression model, the uniform distribution is typically used as a non-informative prior.
- That is, we assign an improper uniform prior to the regression coefficient  $\beta$  that allows  $\beta$  to take on values over the support  $[-\infty, \infty]$ .
- This can be written as  $p(\beta) \propto c$ , where  $c$  is a constant.
- Recall that the uniform prior can be problematic and other “objective” priors are possible.



- Next, we assign a uniform prior to  $\log(\sigma^2)$  because this transformation also allows values over the support  $[0, \infty]$ .
- From here, the joint posterior distribution of the model parameters is obtained by multiplying the prior distributions of  $\beta$  and  $\sigma^2$  by the likelihood give in equation (29).
- Assuming that  $\beta$  and  $\sigma^2$  are independent, we obtain

$$p(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto L(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) p(\beta) p(\sigma^2). \quad (30)$$

- In virtually all packages, non-informative or weakly informative priors are the default.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Bayesian  
Linear  
Regression

Wrap-up

- What about informative priors?
- A sensible conjugate prior distribution for the vector of regression coefficients  $\beta$  of the linear regression model is the normal prior.
- The conjugate prior for the variance of the disturbance term  $\sigma^2$  is the inverse-Gamma distribution, with shape and scale hyperparameters  $a$  and  $b$ , respectively.
- From here, we can obtain the joint posterior distribution of all model parameters using conjugate priors based on expert opinion or prior research.



Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Bayesian Linear Regression

Wrap-up

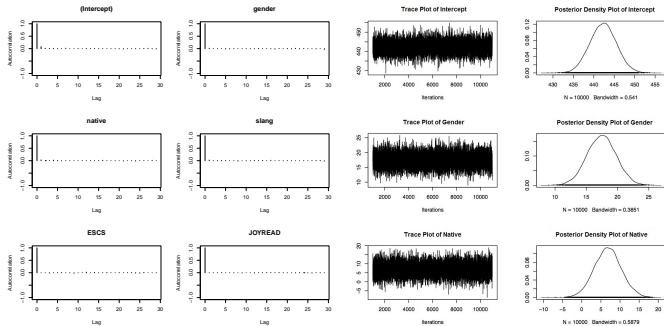


Figure 8: Diagnostic Plots for Regression Example: Selected Parameters





**Table 1:** Bayesian Linear Regression Estimates: Non-informative Prior Case

Parameter	EAP	SD	95% Posterior Probability Interval
<i>Full Model</i>			
INTERCEPT	442.28	3.22	435.99, 448.66
READING on GENDER	17.46	2.29	12.94, 21.91
READING on NATIVE	6.81	3.54	-0.14, 13.9
READING on SLANG	38.45	4.04	30.30, 46.36
READING on ESCS	26.24	1.32	23.69, 28.80
READING on JOYREAD	27.47	1.28	24.97, 29.93
READING on DIVREAD	-5.06	1.21	-7.41, -2.66
READING on MEMOR	-19.03	1.33	-21.65, -16.47
READING on ELAB	-13.74	1.26	-16.26, -11.32
READING on CSTRAT	26.92	1.45	24.07, 29.77

*Note.* EAP = Expected A Posteriori. SD = Standard Deviation.



Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Bayesian Linear Regression

Wrap-up

**Table 2:** Bayesian Regression Estimates: Informative Priors based on PISA 2000

Parameter	EAP	SD	95% PPI
<i>Full Model</i>			
INTERCEPT	487.94	2.82	482.52, 493.59
READING on GENDER	9.46	1.86	5.81, 13.07
READING on NATIVE	-6.87	3.28	-13.30, -0.26
READING on SLANG	9.53	3.56	2.44, 16.48
READING on ESCS	31.19	1.04	29.20, 33.23
READING on JOYREAD	26.50	1.00	24.53, 28.43
READING on DIVREAD	-2.52	0.97	-4.40, -0.59
READING on MEMOR	-18.77	1.09	-20.91, -16.64
READING on ELAB	-13.62	1.06	-15.76, -11.60
READING on CSTRAT	26.06	1.17	23.77, 28.36

*Note.* EAP = Expected A Posteriori. SD = Standard Deviation.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Bayesian  
Linear  
Regression

Wrap-up

**Table 3:** Natural Log Bayes Factors: Informative Priors Case

Sub-Model	BGModel	ATTModel	LSModel	FullModel
BGModel	0.0	27.8	85.6	-539
ATTModel	-27.8	0.0	57.8	-567
LSModel	-85.6	-57.8	0.0	-625
FullModel	539.2	567.0	624.8	0

*Note.* BGModel = Background variables model; ATTModel = Attitude variables model; LSModel = Learning Strategies model; FullModel = Model with all variables.



# Bayesian Model Averaging

**Table 4:** Bayesian model averaging results for full multiple regression model

Predictor	Post Prob	Avg coef	SD	Model 1	Model 2	Model 3	Model 4
<i>Full Model</i>							
INTERCEPT	1.00	493.63	2.11	494.86	491.67	492.77	496.19
GENDER	0.42	2.72	3.54	.	6.46	6.84	.
NATIVE	0.00	0.00	0.00	.	.	.	.
SLANG	0.00	0.00	0.00	.	.	.	.
ESCS	1.00	30.19	1.24	30.10	30.36	30.18	29.90
JOYREAD	1.00	29.40	1.40	29.97	28.93	27.31	28.35
DIVREAD	0.92	-4.01	1.68	-4.44	-4.28	.	.
MEMOR	1.00	-18.61	1.31	-18.47	-18.76	-18.99	-18.70
ELAB	1.00	-15.24	1.26	-15.37	-14.95	-15.43	-15.90
CSTRAT	1.00	27.53	1.46	27.62	27.43	27.27	27.45
$R^2$				0.340	0.341	0.339	0.338
BIC				-1993.72	-1992.98	-1988.72	-1988.51
PMP				0.54	0.37	0.05	0.04



Bayes' Theorem

Bayesian Hypothesis Testing

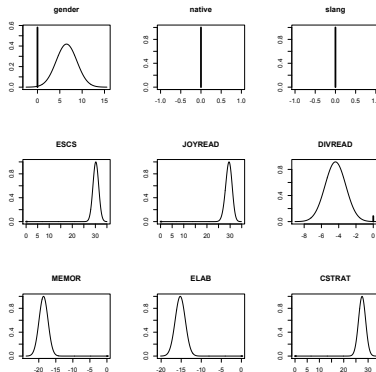
Bayesian Model Building and Evaluation

MCMC

An Example

Bayesian Linear Regression

Wrap-up



**Figure 9:** BMA posterior distributions for model parameters. Note that the narrow spike corresponds to  $p(\beta_k = 0|y)$ . Thus, in the case of gender, note that  $p(\beta_{gender} > 0|y) = .42$  and therefore  $p(\beta_{gender} = 0|y) = 0.58$ . corresponding to the spike found in the BMA posterior distribution plot for GENDER.



Table 5: Comparison of the predictive performance

Regression model	Percent of predictive coverage	Log score of predictive coverage
Bayesian model averaging	95.69%	-0.04
Best model from Bayesian model averaging	90.50%	-0.10
Bayesian model	90.50%	-0.10
Frequentist model	90.45%	-0.10

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage  
Final Thoughts



# Wrap-up: Some Philosophical Issues

- Bayesian statistics represents a powerful alternative to frequentist (classical) statistics, and is therefore, controversial.
- The controversy lies in differing perspectives regarding the nature of probability, and the implications for statistical practice that arise from those perspectives.
- The frequentist framework views probability as synonymous with long-run frequency, and that the infinitely repeating coin-toss represents the canonical example of the frequentist view. frequency.
- In contrast, the Bayesian viewpoint regarding probability was, perhaps, most succinctly expressed by de Finetti

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage  
Final Thoughts



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

**Wrap-up**

The Bayesian  
Advantage  
Final Thoughts

*Probability does not exist.*

- Bruno de Finetti





Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage  
Final Thoughts

- That is, probability does not have an objective status, but rather represents the quantification of our experience of uncertainty.
- For de Finetti, probability is only to be considered in relation to our subjective experience of uncertainty, and, for de Finetti, uncertainty is all that matters.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage  
Final Thoughts

- That is, probability does not have an objective status, but rather represents the quantification of our experience of uncertainty.
- For de Finetti, probability is only to be considered in relation to our subjective experience of uncertainty, and, for de Finetti, uncertainty is all that matters.

*"The only relevant thing is uncertainty – the extent of our known knowledge and ignorance. The actual fact that events considered are, in some sense, determined, or known by other people, and so on, is of no consequence."*(pg. xi).



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage  
Final Thoughts

- That is, probability does not have an objective status, but rather represents the quantification of our experience of uncertainty.
- For de Finetti, probability is only to be considered in relation to our subjective experience of uncertainty, and, for de Finetti, uncertainty is all that matters.

*"The only relevant thing is uncertainty – the extent of our known knowledge and ignorance. The actual fact that events considered are, in some sense, determined, or known by other people, and so on, is of no consequence."*(pg. xi).

- The only requirement then is that our beliefs be coherent, consistent, and have a reasonable relationship to any observable data that might be collected.



# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are

- 1 Coherence

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are
  - 1 Coherence
  - 2 Handling non-nested models

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are
  - 1 Coherence
  - 2 Handling non-nested models
  - 3 Flexibility in handling complex data structures

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are
  - 1 Coherence
  - 2 Handling non-nested models
  - 3 Flexibility in handling complex data structures
  - 4 Inferences based on data actually observed

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts





# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are
  - 1 Coherence
  - 2 Handling non-nested models
  - 3 Flexibility in handling complex data structures
  - 4 Inferences based on data actually observed
  - 5 Quantifying evidence

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Summarizing the Bayesian Advantage

- The major advantages of Bayesian statistical inference over frequentist statistical inference are
  - 1 Coherence
  - 2 Handling non-nested models
  - 3 Flexibility in handling complex data structures
  - 4 Inferences based on data actually observed
  - 5 Quantifying evidence
  - 6 Incorporating prior knowledge

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

The Bayesian Advantage

Final Thoughts



# Subjective v. Objective Bayes

- Subjective Bayesian practice attempts to bring prior knowledge directly into an analysis. This prior knowledge represents the analysts (or others) degree-of-uncertainty.
- An analyst's degree-of-uncertainty is encoded directly into the prior distribution, and specifically in the degree of precision around the parameter of interest.
- The advantages according to Press (2003) include

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Subjective v. Objective Bayes

- Subjective Bayesian practice attempts to bring prior knowledge directly into an analysis. This prior knowledge represents the analysts (or others) degree-of-uncertainty.
- An analyst's degree-of-uncertainty is encoded directly into the prior distribution, and specifically in the degree of precision around the parameter of interest.
- The advantages according to Press (2003) include
  - ① Subjective priors are proper

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



# Subjective v. Objective Bayes

- Subjective Bayesian practice attempts to bring prior knowledge directly into an analysis. This prior knowledge represents the analysts (or others) degree-of-uncertainty.
- An analyst's degree-of-uncertainty is encoded directly into the prior distribution, and specifically in the degree of precision around the parameter of interest.
- The advantages according to Press (2003) include
  - 1 Subjective priors are proper
  - 2 Priors can be based on factual prior knowledge

Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

The Bayesian Advantage

Final Thoughts



# Subjective v. Objective Bayes

- Subjective Bayesian practice attempts to bring prior knowledge directly into an analysis. This prior knowledge represents the analysts (or others) degree-of-uncertainty.
- An analyst's degree-of-uncertainty is encoded directly into the prior distribution, and specifically in the degree of precision around the parameter of interest.
- The advantages according to Press (2003) include
  - 1 Subjective priors are proper
  - 2 Priors can be based on factual prior knowledge
  - 3 Small sample sizes can be handled.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- The disadvantages to the use of subjective priors according to Press (2003) are



Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

The Bayesian Advantage

Final Thoughts

- The disadvantages to the use of subjective priors according to Press (2003) are
  - 1 It is not always easy to encode prior knowledge into probability distributions.





Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- The disadvantages to the use of subjective priors according to Press (2003) are
  - 1 It is not always easy to encode prior knowledge into probability distributions.
  - 2 Subjective priors are not always appropriate in public policy or clinical situations.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- The disadvantages to the use of subjective priors according to Press (2003) are
  - 1 It is not always easy to encode prior knowledge into probability distributions.
  - 2 Subjective priors are not always appropriate in public policy or clinical situations.
  - 3 Prior distributions may be analytically intractable unless they are conjugate priors.



- Within objective Bayesian statistics, there is disagreement about the use of the term “objective”, and the related term “non-informative”.
- Specifically, there are a large class of so-called *reference priors* (Kass and Wasserman, 1996). For example:
  - Uniform prior
  - Vague (diffused) prior
  - Jeffreys prior
  - Zellner's  $g$ -prior
  - Maximum entropy prior



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- An important viewpoint regarding the notion of objectivity in the Bayesian context comes from Jaynes (1968).
- For Jaynes, the “personalistic” school of probability is to be reserved for



Bayes' Theorem

Bayesian Hypothesis Testing

Bayesian Model Building and Evaluation

MCMC

An Example

Wrap-up

The Bayesian Advantage

Final Thoughts

- An important viewpoint regarding the notion of objectivity in the Bayesian context comes from Jaynes (1968).
- For Jaynes, the “personalistic” school of probability is to be reserved for

*“...the field of psychology and has no place in applied statistics. Or, to state this more constructively, objectivity requires that a statistical analysis should make use, not of anybody's personal opinions, but rather the specific factual data on which those opinions are based.”(pg. 228)*



- In terms of advantages Press (2003) notes that

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- In terms of advantages Press (2003) notes that
  - 1 Objective priors can be used as benchmarks against which choices of other priors can be compared



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- In terms of advantages Press (2003) notes that
  - 1 Objective priors can be used as benchmarks against which choices of other priors can be compared
  - 2 Objective priors reflect the view that little information is available about the process that generated the data





Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- In terms of advantages Press (2003) notes that
  - 1 Objective priors can be used as benchmarks against which choices of other priors can be compared
  - 2 Objective priors reflect the view that little information is available about the process that generated the data
  - 3 An objective prior provides results equivalent to those based on a frequentist analysis



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- In terms of advantages Press (2003) notes that
  - 1 Objective priors can be used as benchmarks against which choices of other priors can be compared
  - 2 Objective priors reflect the view that little information is available about the process that generated the data
  - 3 An objective prior provides results equivalent to those based on a frequentist analysis
  - 4 Objective priors are sensible public policy priors.



- In terms of disadvantages of objective priors, Press (2003) notes that

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- In terms of disadvantages of objective priors, Press (2003) notes that
  - ① Objective priors can lead to improper results when the domain of the parameters lie on the real number line.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- In terms of disadvantages of objective priors, Press (2003) notes that
  - 1 Objective priors can lead to improper results when the domain of the parameters lie on the real number line.
  - 2 Parameters with objective priors are often independent of one another, whereas in most multi-parameter statistical models, parameters are correlated.



- In terms of disadvantages of objective priors, Press (2003) notes that
  - 1 Objective priors can lead to improper results when the domain of the parameters lie on the real number line.
  - 2 Parameters with objective priors are often independent of one another, whereas in most multi-parameter statistical models, parameters are correlated.
  - 3 Expressing complete ignorance about a parameter via an objective prior leads to incorrect inferences about functions of the parameter.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- Kadane (2011) states, among other things:



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- Kadane (2011) states, among other things:

*“The purpose of an algorithmic prior is to escape from the responsibility to give an opinion and justify it. At the same time, it cuts off a useful discussion about what is reasonable to believe about the parameters. Without such a discussion, appreciation of the posterior distribution of the parameters is likely to be less full, and important scientific information may be neglected.”(pg. 445)*





# Final Thoughts: Evidenced-based Bayes

- The subjectivist school allows for personal opinion to be elicited and incorporated into a Bayesian analysis.
- The subjectivist school places no restriction on the source, reliability, or validity of the elicited opinion.
- The objectivist school views personal opinion as the realm of psychology with no place in a statistical analysis.
- The objectivist school would require formal rules for choosing reference priors.

Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts



- “Subjectivism” within the Bayesian framework runs the gamut from the elicitation of personal beliefs to making use of the best available historical data available to inform priors.
- I agree with Jaynes (1968) – namely that the requirements of science demand reference to “specific, factual data on which those opinions are based” (pg. 228).
- This view is also consistent with Leamer’s hierarchy of confidence on which priors should be ordered.
- We may refer to this view as an *evidence-based* form of subjective Bayes which acknowledges (1) the subjectivity that lies in the choice of historical data; (2) the encoding of historical data into hyperparameters of the prior distribution; and (3) the choice among competing models to be used to analyze the data.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- What if factual historical data are not available?
- Berger (2006) states that reference priors should be used “in scenarios in which a subjective analysis is not tenable”, although such scenarios are probably rare.
- The goal, nevertheless, is to shift the practice of Bayesian statistics away from the elicitation of personal opinion (expert or otherwise) which could, in principle, bias results toward a specific outcome, and instead move Bayesian practice toward the warranted use prior objective empirical data for the specification of priors.
- The specification of any prior should be explicitly warranted against observable, empirical data and available for critique by the relevant scholarly community.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up

The Bayesian  
Advantage

Final Thoughts

- To conclude, the Bayesian school of statistical inference is, arguably, superior to the frequentist school as a means of creating and updating new knowledge in the social sciences.
- An evidence-based focus that ties the specification of priors to objective empirical data provides stronger warrants for conclusions drawn from a Bayesian analysis.
- In addition predictive criteria should always be used as a means of testing and choosing among Bayesian models.
- As always, the full benefit of the Bayesian approach to research in the social sciences will be realized when it is more widely adopted and yields reliable predictions that advance knowledge.



Bayes'  
Theorem

Bayesian  
Hypothesis  
Testing

Bayesian  
Model  
Building and  
Evaluation

MCMC

An Example

Wrap-up  
The Bayesian  
Advantage

Final Thoughts

Dank u wel