



What's All The Buzz About Bayes?: Bayesian Inference for the Social and Behavioral Sciences

David Kaplan

Department of Educational Psychology



April 13th, 2015
University of Nebraska-Lincoln

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

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



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

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 Statistical Inference

- 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.
- Recently, there has been a renaissance in the development and application of Bayesian statistical methods, owing mostly to developments of powerful statistical software tools that render the specification and estimation of complex models feasible from a Bayesian perspective.



Outline of Workshop

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Morning:

- Major differences between the Bayesian and frequentist paradigms of statistics
- Bayes' theorem
- Bayesian hypothesis testing

Afternoon:

- Bayesian model building and evaluation
- An example of Bayesian linear regression
- Wrap-up: Additional philosophical issues



Paradigm Differences

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

- 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.



- 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 the 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 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 Series?” 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

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
9/80

- 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

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

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

- The inverse probability theorem (Bayes' theorem) states

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



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

11 / 80

- 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 (what we will call *prior probabilities*).



Statistical Elements of Bayes' Theorem

- 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.
- This is the random variable Y . Once the person identifies his/her socioeconomic status, the random variable Y becomes 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 .

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
12 / 80



- Next, denote by θ a parameter that we believe characterizes the probability model of interest.
- The parameter θ 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 θ , 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 $L(\theta|y)$.



- The key difference between Bayesian statistical inference and frequentist statistical inference concerns the nature of the unknown parameters θ .
- In the frequentist tradition, the assumption is that θ is unknown, but that estimation of this parameter does not incorporate uncertainty.
- In Bayesian statistical inference, θ is considered unknown and random, possessing a probability distribution that reflects our uncertainty about the true value of θ .
- Because both the observed data y and the parameters θ 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,

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

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 θ given the observed data y .



- From equation (7) the posterior distribution of θ 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

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

- and for continuous variables

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



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

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

- When expressed in terms of the unknown parameters θ for fixed values of y , the term $p(y|\theta)$ is the *likelihood* $L(\theta|y)$, which we defined earlier. Thus, equation (10) can be re-written as

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



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
18 / 80

- 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 $p(\theta|y)$.



- The likelihood $L(\theta|y)$ is extremely important in Bayesian inference.
- It is the device that we use to summarize the data.
- The likelihood principle states that all of the information in a sample is contained in the likelihood function, and that this function is indexed by model parameters θ .
- Notice that there is no appeal to repeated sampling. All that matters is the sample in our hand.



Exchangeability

- 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 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, $f(y_1, y_2, \dots, y_n)$ is invariant to permutations of the subscripts.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution
Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
20 / 80



- Exchangeability means that we believe that there is a parameter θ 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 θ without reference to a particular set of data is, in fact, what is implied by the idea of a prior distribution.
- The assumption of exchangeability is weaker than the assumption of independence.
- Independence implies that $p(A|B) = p(A)$. If these two events are independent then they are exchangeable – however, exchangeability does not imply independence.



The Prior Distribution

- Why do we specify a prior distribution on the parameters?
- The key philosophical reason concerns our view that progress in science generally comes about by learning from previous research findings and incorporating information from these findings into our present studies.
- The information gleaned 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 simply 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).

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
22 / 80



- 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.
 - This issue has also been discussed by 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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
24 / 80



- Perhaps the most sensible non-informative prior distribution to use in this case is the uniform distribution $U(\alpha, \beta)$ over some sensible range of values from α to β .
- In this case, the uniform distribution essential 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. For example, a $U[-\infty, \infty]$ is an *improper* prior distribution insofar as it does not integrate to 1.0 as required of probability distributions.



Informative-Conjugate Priors

- 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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
26 / 80



Conjugate Priors for Some Common Distributions

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

Conjugate
Priors

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
27 / 80

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

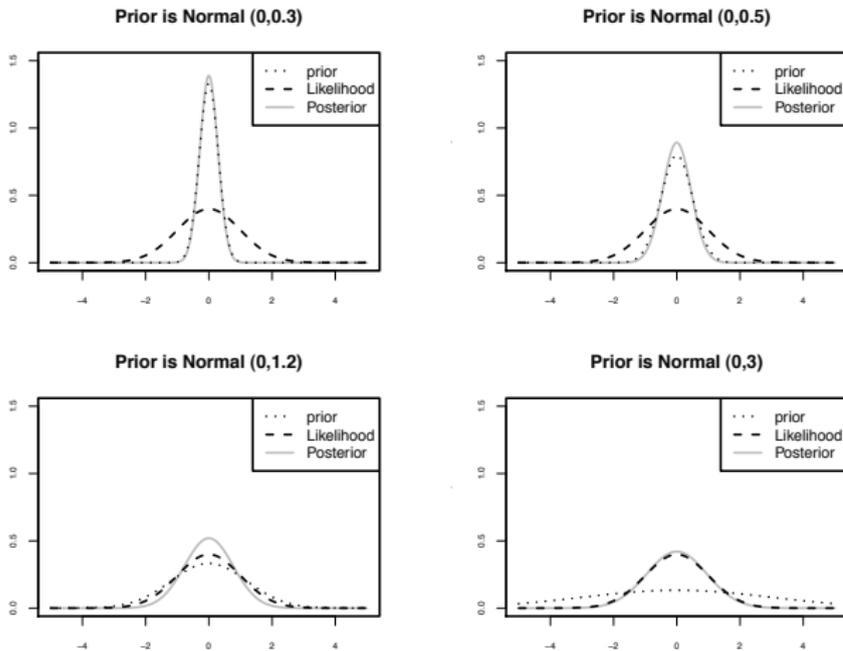


Figure 1: Normal distribution, mean unknown/variance known with varying conjugate priors



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Statistical
Elements of
Bayes'
Theorem

The
Assumption of
Exchangeability

The Prior
Distribution

**Conjugate
Priors**

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
29 / 80

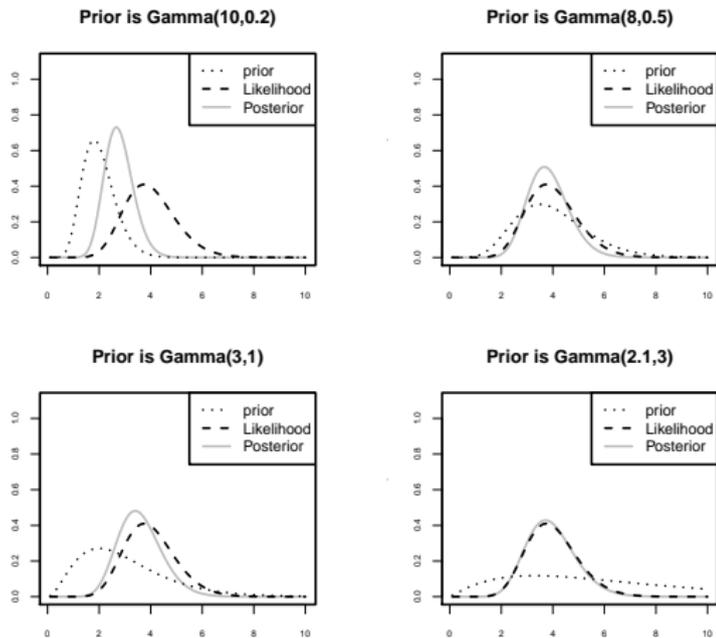


Figure 2: Poisson distribution with varying gamma-density priors



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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries
Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries
Interval
Summaries

Bayesian
Model
Building and
Evaluation

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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries
Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

- 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 (α) and Type II errors (β)
- 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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries
Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

- 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, α and β 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 α , 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).



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries
Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

- However, this balance is virtually always ignored and $\alpha = 0.05$ is used.
- The point is that the p -value and α are not the same thing.
- The confusion between these two concepts 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 α ahead of time, as per the Neyman-Pearson school, 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”.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries
Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

- 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 somewhat 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.



Point Summaries of the Posterior Distribution

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries

Interval
Summaries

Bayesian
Model
Building and
Evaluation

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 θ given the data y is referred to as the *expected a posteriori* or EAP estimate and can be written as

$$E(\theta|y) = \int_{-\infty}^{+\infty} \theta p(\theta|y) d\theta. \quad (12)$$

Similarly, the variance of posterior distribution of θ given y can be obtained as

$$\begin{aligned} \text{var}(\theta|y) &= E[(\theta - E[(\theta|y)])^2|y], \\ &= \int_{-\infty}^{+\infty} (\theta - E[\theta|y])^2 p(\theta|y) d\theta. \end{aligned} \quad (13)$$



- 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

$$\hat{\theta}_{MAP} = \arg \max_{\theta} L(\theta|y)p(\theta). \quad (14)$$

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



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries

Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

● 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 μ .
- 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 μ under the null hypothesis. Notice that the probability that the parameter is in the interval is either zero or one.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Point
Summaries

Interval
Summaries

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

- 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.
- 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.
- First, a researcher will specify an initial model relying on a lesser or greater degree of prior theoretical knowledge.
- Second, these models will be fit to data obtained from a sample from some relevant population.
- Third, an evaluation of the quality of the models will be undertaken, examining where each model might deviate from the data, as well as assessing any possible model violations. At this point, model respecification may come into play.
- Finally, depending on the goals of the research, the “best model” will be chosen for some purpose.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking

Bayes Factors

Bayesian
Model
Averaging

An Example

Wrap-up



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking
Bayes Factors

Bayesian
Model
Averaging

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.
- For a Bayesian, the first phase of modeling building will require the specification of a full probability model for the data and the parameters of the model, where the latter requires the specification of the prior distribution.
- The notion of model fit, therefore, implies that the full probability model fits the data. Lack of model fit may well be due to incorrect specification of likelihood, the prior distribution, or both.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking
Bayes Factors

Bayesian
Model
Averaging

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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking

Bayes Factors

Bayesian
Model
Averaging

An Example

Wrap-up



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking

Bayes Factors

Bayesian
Model
Averaging

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

$$\begin{aligned} p(y^{rep}|y) &= \int p(y^{rep}|\theta)p(\theta|y)d\theta \\ &= \int p(y^{rep}|\theta)p(y|\theta)p(\theta)d\theta \end{aligned} \quad (15)$$

- Equation (15) 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. χ^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

$$p\text{-value} = pr[T(y^{rep}, \theta) \geq T(y, \theta) | y]. \quad (16)$$

- 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 θ_1 and θ_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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking

Bayes Factors

Bayesian
Model
Averaging

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

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \times \left[\frac{p(M_1)}{p(M_2)} \right]. \quad (17)$$

- Notice that the first term on the right hand side of equation (17) 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 $p(y|M_1)/p(y|M_2)$.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking

Bayes Factors

Bayesian
Model
Averaging

An Example

Wrap-up

- 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 (17) 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 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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

Posterior
Predictive
Checking
Bayes Factors

Bayesian
Model
Averaging

An Example

Wrap-up



Bayesian Linear Regression

- We will start with the basic multiple regression model.
- To begin, 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 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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (18)$$

where $\boldsymbol{\beta}$ is an $k \times 1$ vector of regression coefficients and where the first column of $\boldsymbol{\beta}$ contains an n -dimensional unit vector to capture the intercept term. We assume that student level PISA reading scores are generated from a normal distribution.

- We also assume that the n -dimensional vector \mathbf{u} are disturbance terms assumed to be independently, identically, and normally distributed – specifically.



- Recall that the components of Bayes' theorem require the likelihood and the priors on all model parameters.

- We will write the likelihood for the regression model as

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) \quad (19)$$

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



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

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 β that allows β to take on values over the support $[-\infty, \infty]$.
- This can be written as $p(\beta) \propto c$, where c is a constant.
- Note that there is no such thing as an “informationless” prior. This uniform prior says a lot!!



- 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 β and σ^2 by the likelihood given in equation (19).
- Assuming that β and σ^2 are independent, we obtain

$$p(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) \propto L(\beta, \sigma^2 | \mathbf{y}, \mathbf{X}) p(\beta) p(\sigma^2). \quad (20)$$

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



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up

- What about informative priors?
- A sensible conjugate prior distribution for the vector of regression coefficients β of the linear regression model is the normal prior.
- The conjugate prior for the variance of the disturbance term σ^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.



Markov Chain Monte Carlo Sampling

- 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.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example
Bayesian
Linear
Regression

MCMC

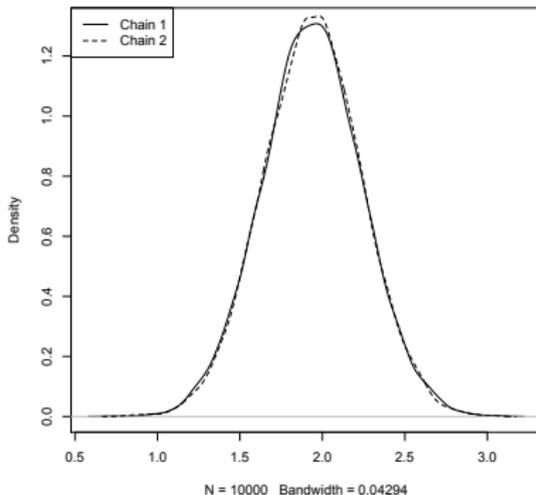
Wrap-up

- 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 $inv\text{-gamma}(a, b)$

Posterior Density Plot of the Mean --Two Chains





Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up

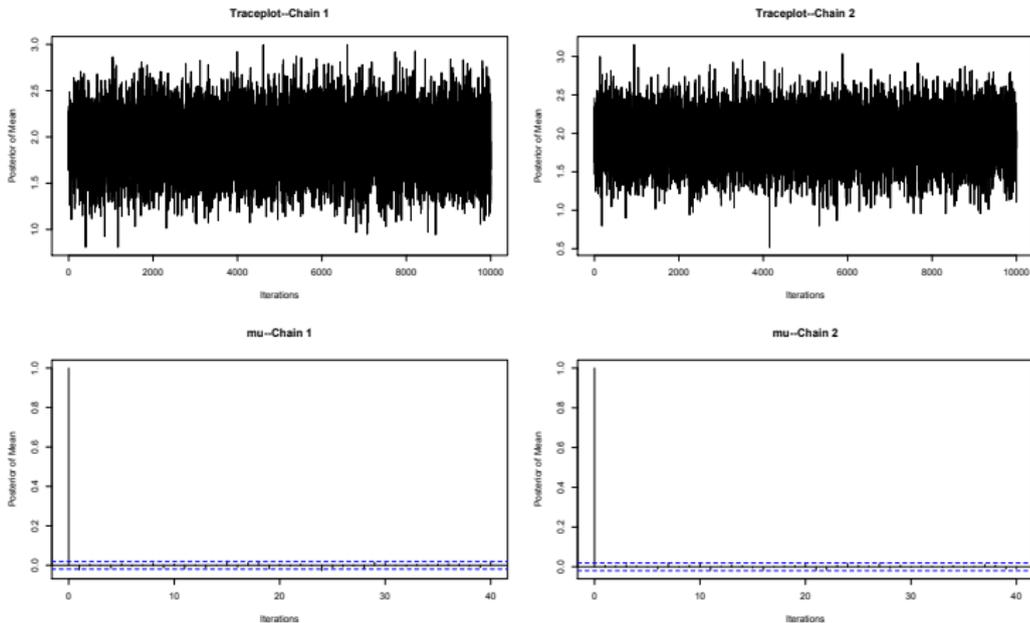


Figure 4: Trace and ACF plots for two chains



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up

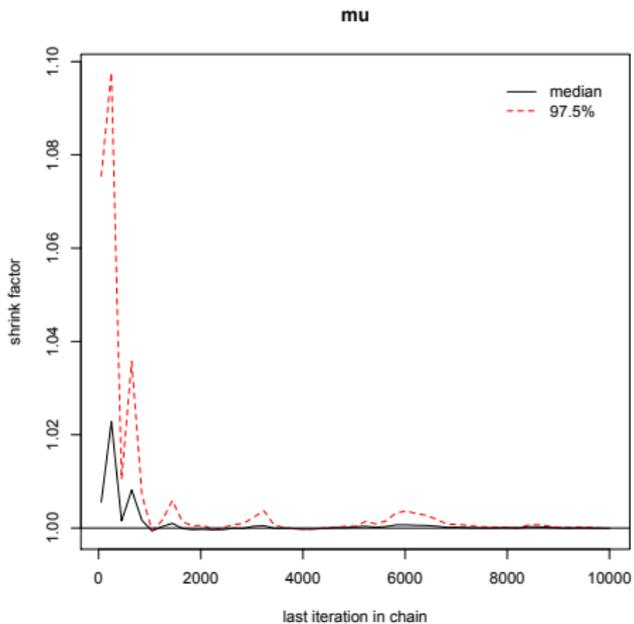


Figure 5: Gelman-Rubin-Brooks plot



- 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 (joyread), diversity in reading (divread), memorization strategies (memor), elaboration strategies (elab), control strategies (cstrat).
- Outcome was the first plausible value of the PISA 2009 reading assessment.
- We use *MCMCpack* in R.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up

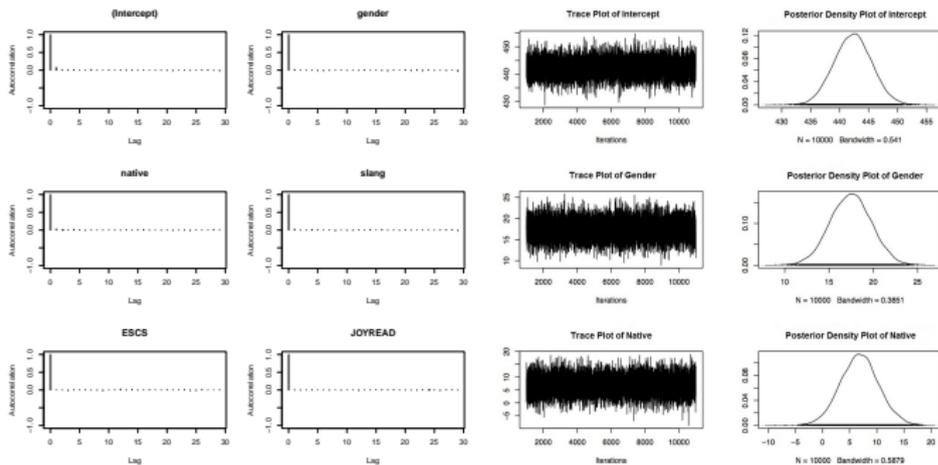


Figure 6: Diagnostic Plots for Regression Example: Selected Parameters

**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.



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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

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

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Bayesian
Linear
Regression

MCMC

Wrap-up



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

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Final Thoughts



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Final Thoughts

Probability does not exist.

- Bruno de Finetti



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

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.



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 specification of the prior distribution, and in particular on the degree of precision around the parameter of interest.
- The advantages include
 - 1 Subjective priors are proper
 - 2 Priors can be based on factual prior knowledge
 - 3 Small sample sizes can be handled.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Final Thoughts



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Final Thoughts

- 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).
- 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)



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Final Thoughts

- In terms of advantages of objective priors 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
 - 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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

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: A Call for Evidenced-based Subjective Bayes

- The subjectivist school, advocated by de Finetti, Savage, and others, allows for personal opinion to be elicited and incorporated into a Bayesian analysis. In the extreme, the subjectivist school would place no restriction on the source, reliability, or validity of the elicited opinion.
- The objectivist school advocated by Jeffreys, Jaynes, Berger, Bernardo, and others, views personal opinion as the realm of psychology with no place in a statistical analysis. In their extreme form, the objectivist school would require formal rules for choosing reference priors.
- The difficulty with these positions lies with the everyday usage of terms such as “subjective” and “belief”.
- Without careful definitions of these terms, their everyday usage might be misunderstood among those who might otherwise consider adopting the Bayesian perspective.

Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

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 argue along the lines of 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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

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.



- 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.



Bayesian
Statistical
Inference

Paradigm
Differences

Bayes'
Theorem

Bayesian
Hypothesis
Testing

Bayesian
Model
Building and
Evaluation

An Example

Wrap-up

Final Thoughts

THANK YOU