Prelude 1

Problems in Sensometrics

Problem 1: Missing Data

- Often, participants do not answer all questions. What do we do? Impute the values? Throw away the data? Or…?

Problem 2: Outlying Data

- Some participants occasionally give random answers, that is, answers that come from some contaminant process. What do we do? How to identify possible outliers? How can we quantify the certainty that an answer is an outlier? And if we are, say, 30% sure that a response is an outlier, how should this affect our inference?

Problem 3: Support for H0

- Sometimes we want to quantify the evidence *in favor of* a null effect. How to do this? Look at power? But power against what specific alternative?

Problem 4: Utilities

- Suppose we find that a particular action **A** will significantly increase the sales of Mars bars (p = .001).
- Should the Mars company take action **A**?
- What if action **A** is "slash the prices in half"?
- What if this quadruples the sales? – but of course you are never sure about how action **A** will affect sales…

Problem 5: Sparsity

- With many predictors, we want sparse solutions, particularly when $p >> n$.
- How to accomplish this? Lasso? Ridge regression? Latent factor models?

Problem 6: Model Uncertainty

- Suppose we have two competing models, M1 and M2, and the goal is prediction.
- When none of the models clearly dominates the other, how should we combine the different predictions that they make into a single overall prediction?

Prelude 2

Problems in Classical, “Frequentist” Statistics

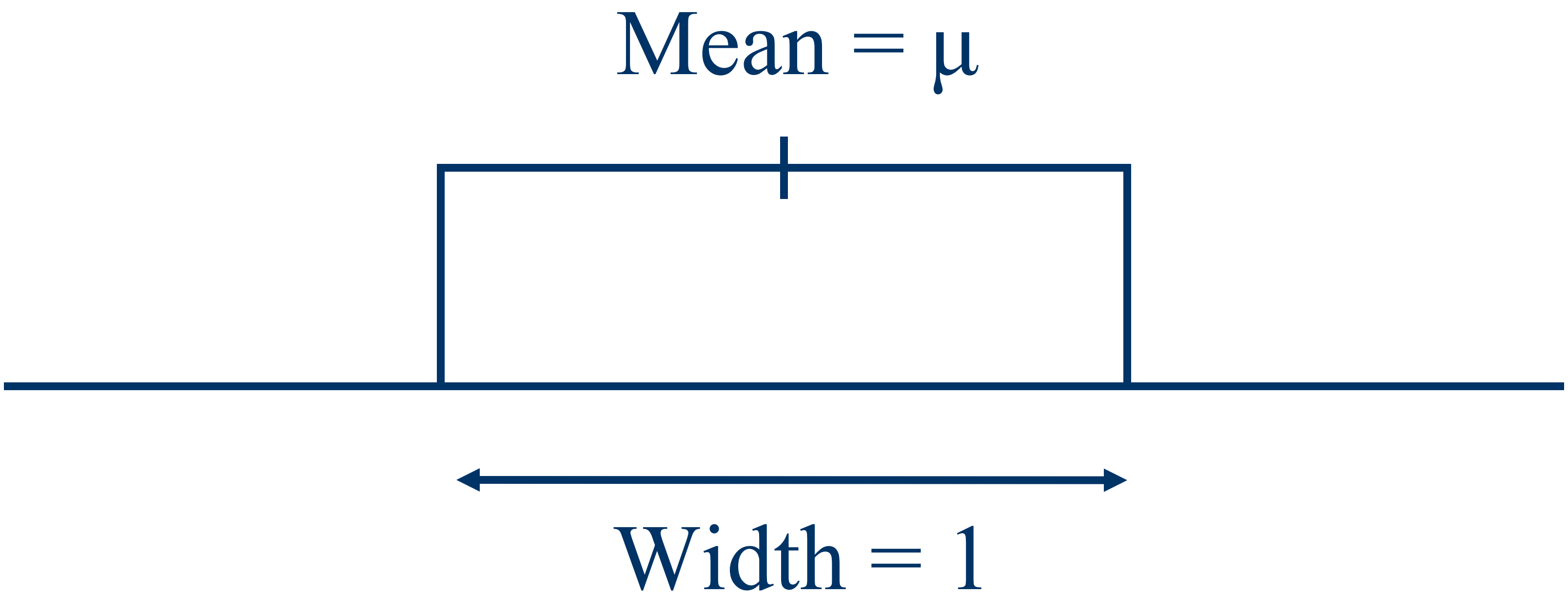
Problem 1: Conceptual Confusion

- What is a p-value?
- What is a confidence interval?

Frequentist Inference

- Procedures are used because they do well in the long run, that is, in many situations.
- Parameters are assumed to be fixed, and do not have a probability distribution.
- Inference is pre-experimental or unconditional.

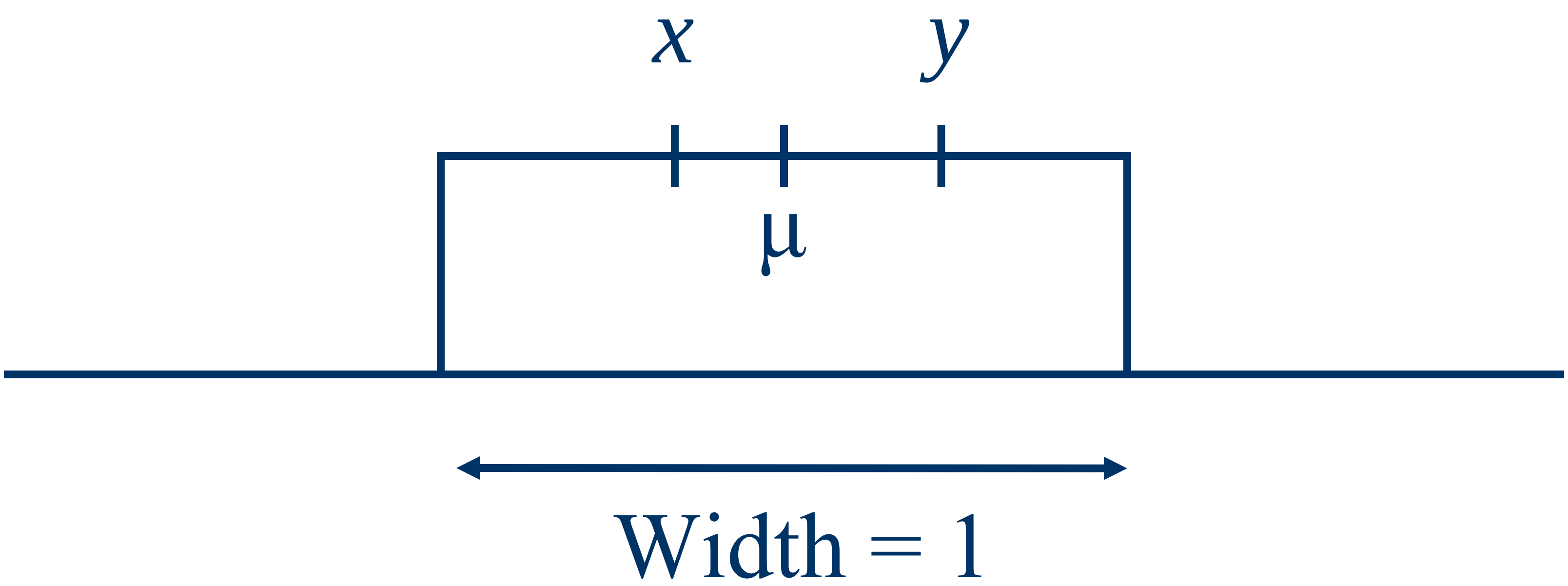
Frequentist Confidence Intervals



Frequentist Confidence Intervals

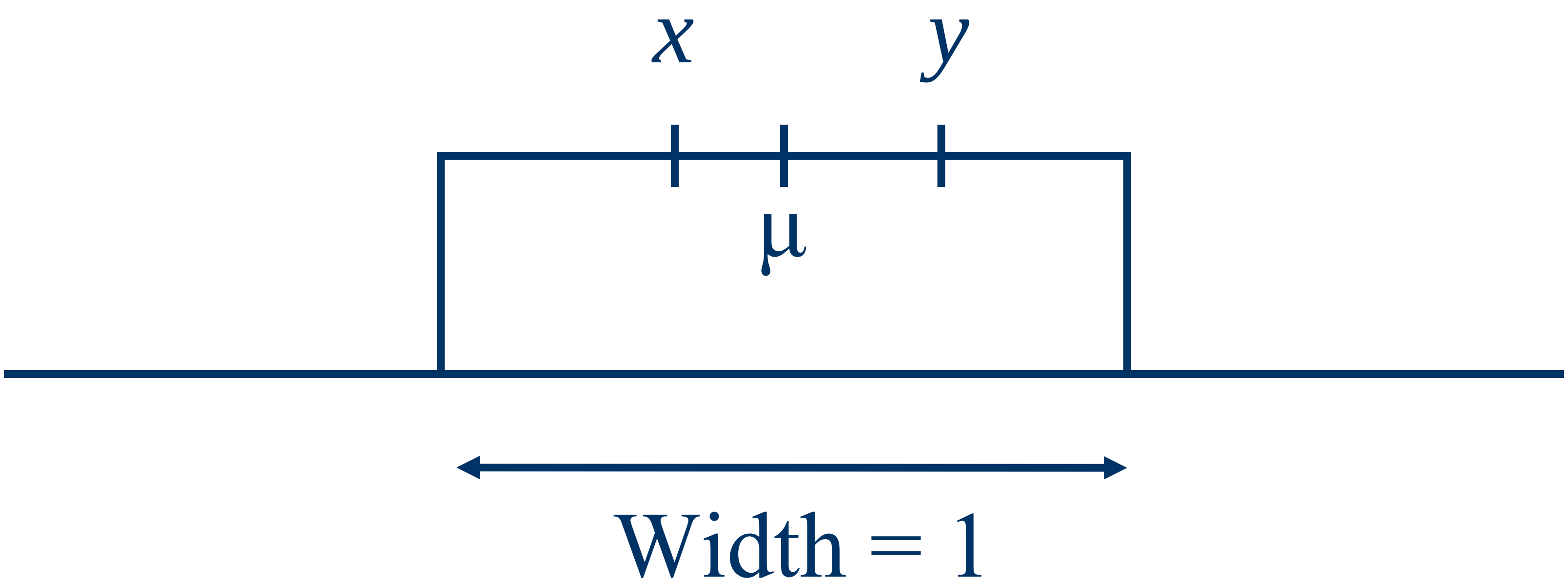
Draw a random number *x*.

Draw another random number *y*. What is the probability that it will lie to the other side of μ?



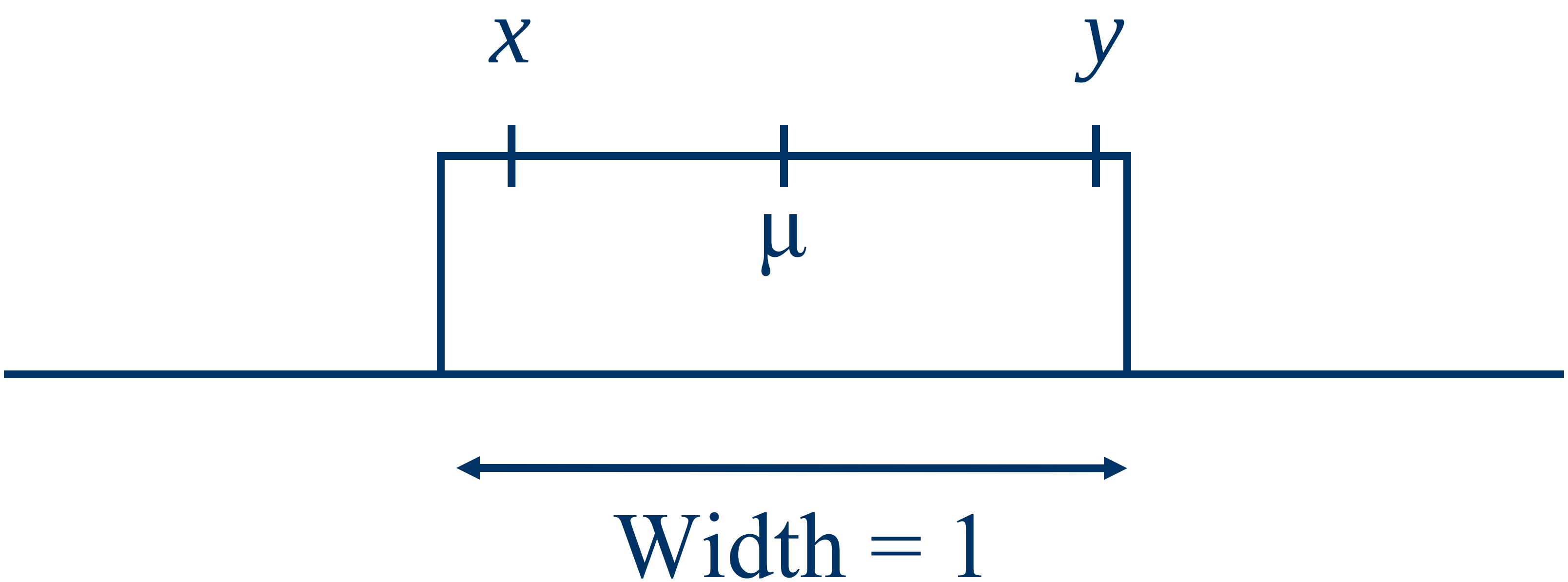
Frequentist Confidence Intervals

When we repeated this procedure many times , the mean μ will lie in the interval in 50% of the cases. Hence, the interval (x, y) with $y > x$ is a 50% confidence interval for μ.



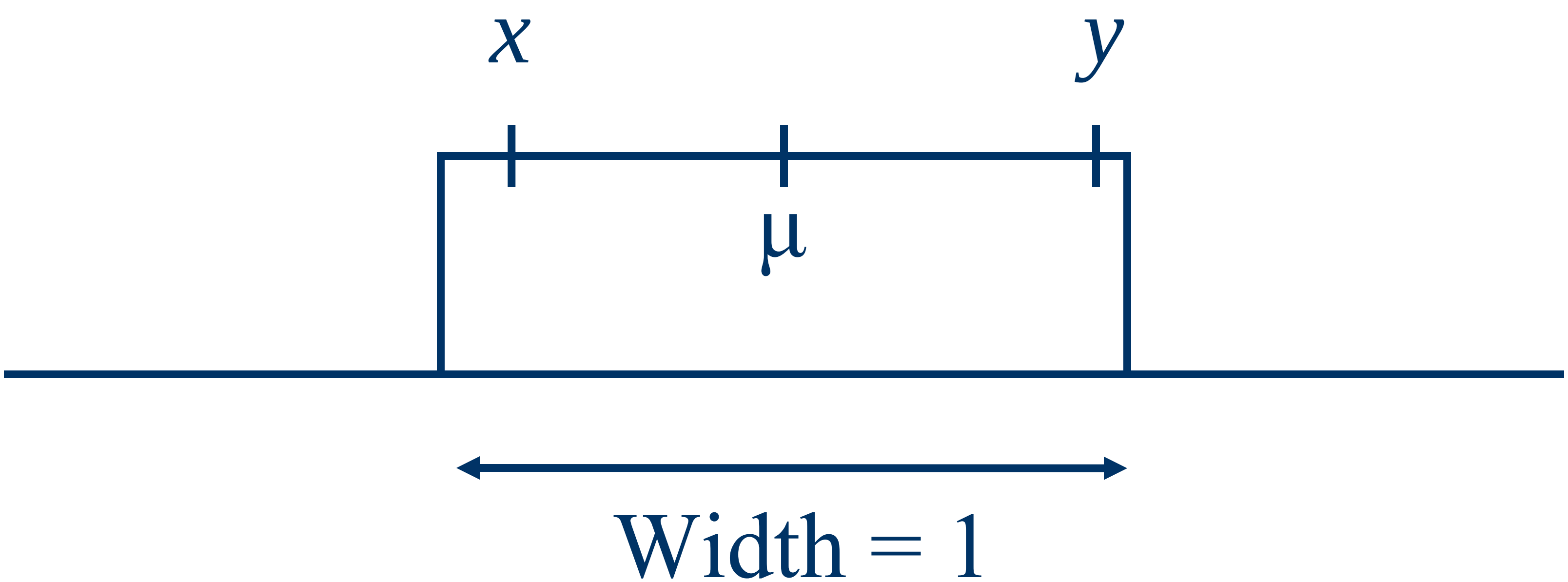
Frequentist Confidence Intervals

But now you observe the following data:



Frequentist Confidence Intervals

Because the width of the distribution is 1, *I am 100% confident that the mean lies in the 50% confidence interval*!



Why?

- Frequentist procedures have good pre-experimental properties and are designed to work well for most data.
- For particular data, however, these procedures may be horrible.

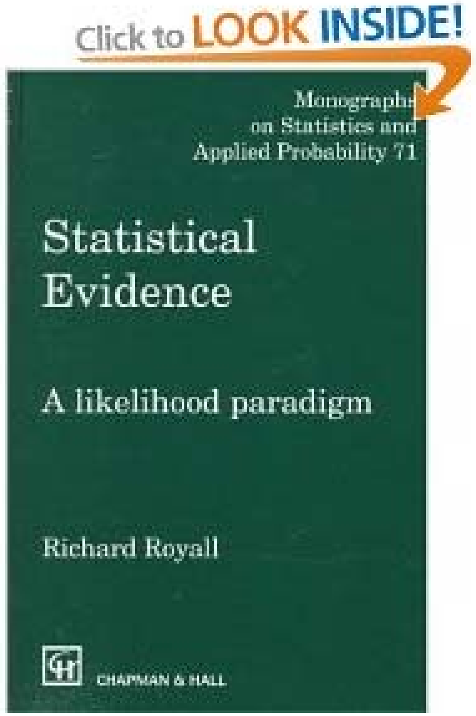
Problem 2: Incorrect Conclusions

- Concluding from a nonsignificant p-value that H0 is true;
- Concluding from a significant p-value that H0 is false!
- Why? Researchers intrinsically want to be able to attach probabilities to parameters and hypotheses.

Problem 3: Violation of the Likelihood Principle

- Conclusions can depend on the (possible unknown) intention with which the researcher collected the data.
- Is this reasonable?

Problem 3: Violation of the Likelihood Principle



& "The Likelihood Principle" by Berger and Wolpert (1988).

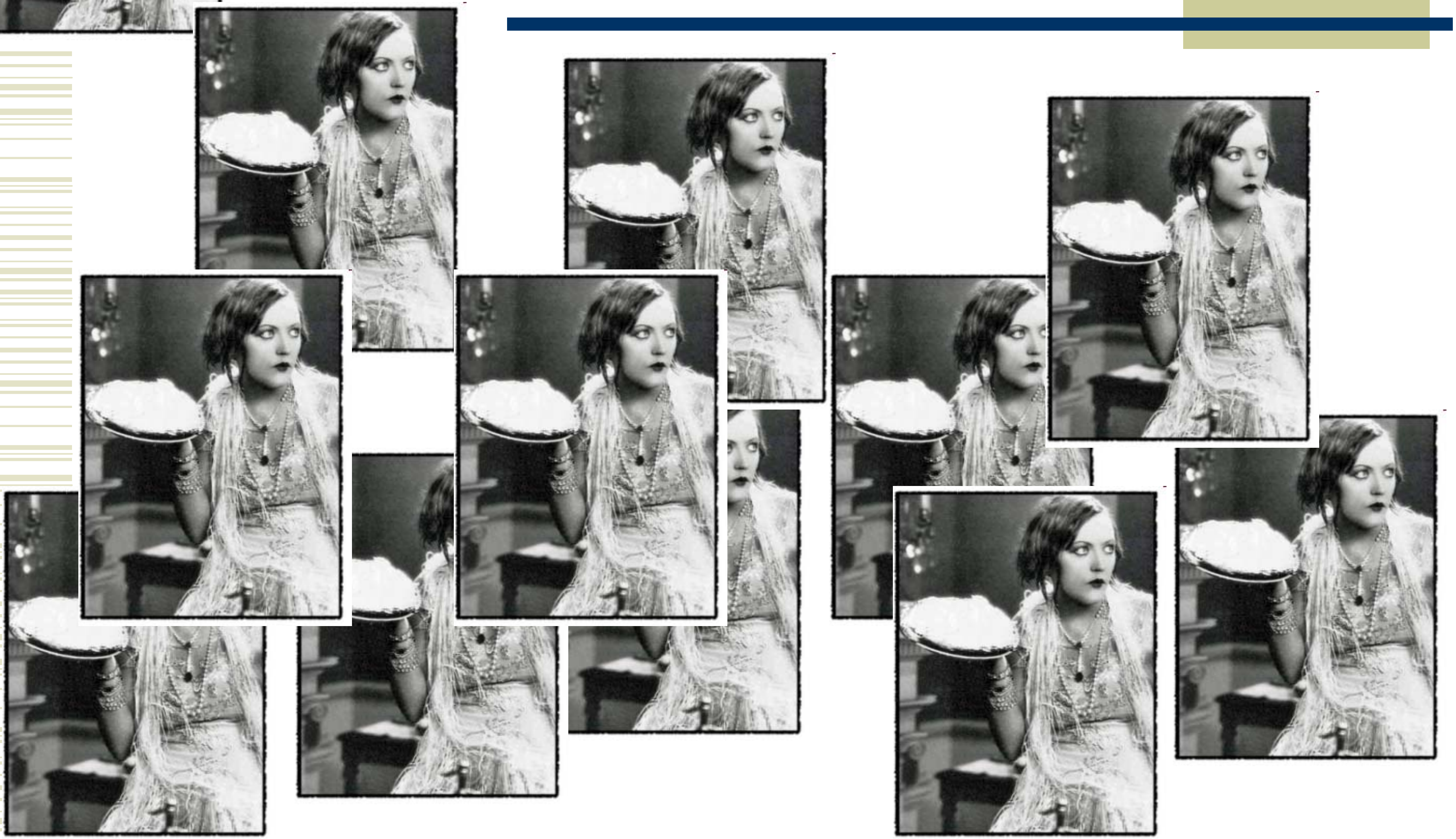
Second hand on Amazon:
£176

New on
http://imstat.org/en/index.html:
$25

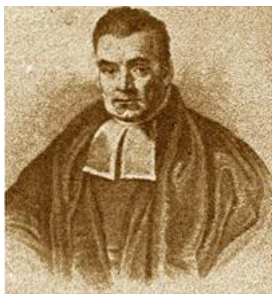
Problem 4: p-values and Evidence

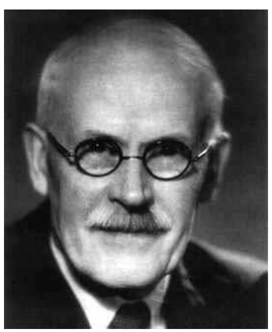
- P-values overestimate the evidence against H0 [see the many works by Jim Berger].
- This tendency gets worse with *n*.
- Some prestigious psychological journals publish mainly surprising findings. But for such findings, can we reject H0 at $p = .05$?

Conclusion: Problems!?

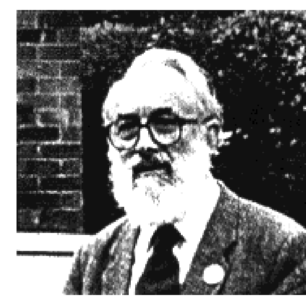


Bayesian Statistics





Eric-Jan Wagenmakers



UNIVERSITEIT VAN AMSTERDAM

Outline

- Bayesian Basics
- Bayesian Revolution and MCMC
- WinBUGS
- Bayesian Sparsity

What is Bayesian Statistics?

"Common sense expressed in numbers"

What is Bayesian Statistics?

"What you think that classical statistics is"

What is Bayesian Statistics?

"The only good statistics"

[For more background see

Lindley, D. V. (2000). The philosophy of statistics. The Statistician, 49, 293-337.]

Bayesian Inference in a Nutshell

- In Bayesian inference, uncertainty or degree of belief is quantified by probability.
- Prior beliefs are updated by means of the data to yield posterior beliefs.

Bayes' Rule

Likelihood

Prior Distribution

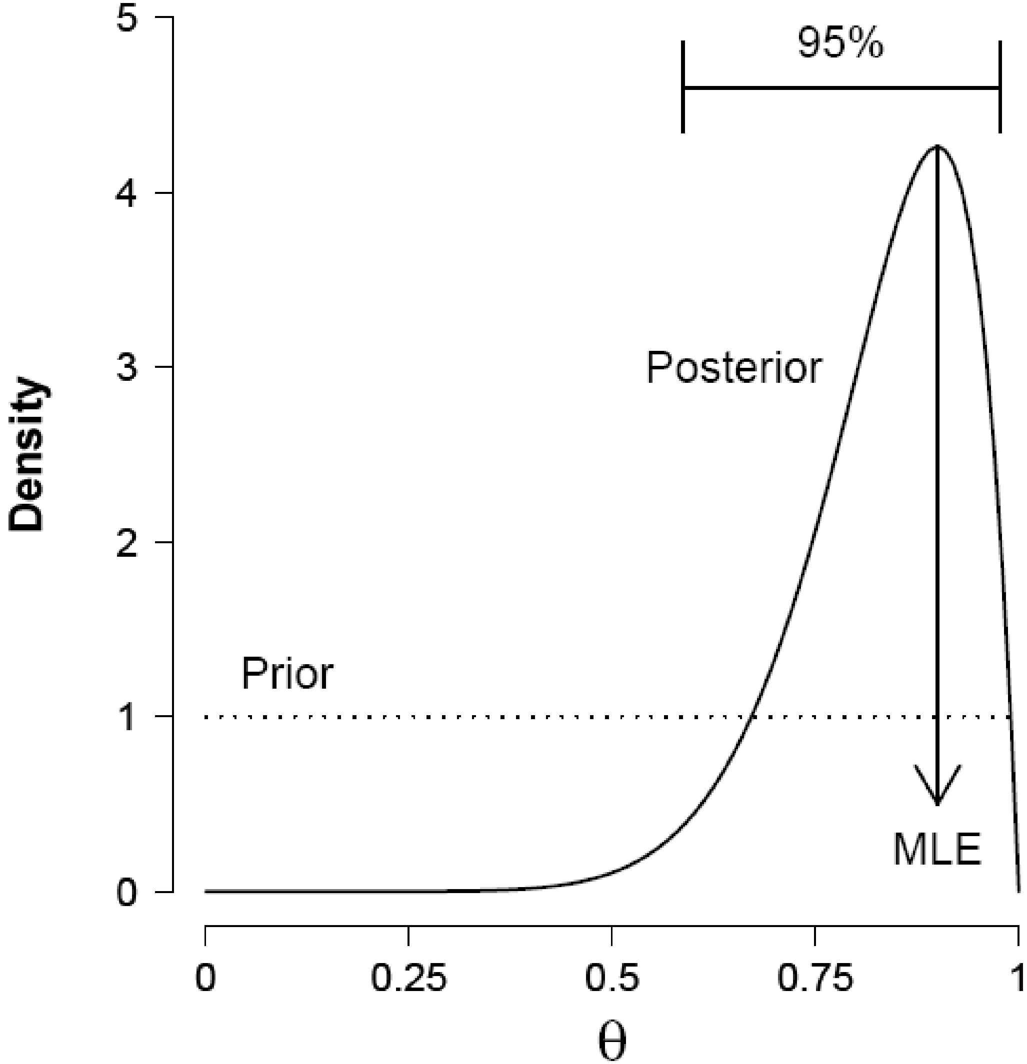
$$P(\theta \mid D, M_1) = \frac{P(D \mid \theta, M_1)\, P(\theta \mid M_1)}{P(D \mid M_1)}$$

Posterior Distribution

Marginal Probability of the Data

Bayesian Parameter Estimation: Example

- Suppose that 9 out of 10 respondents prefer Mars over Snickers.
- What have we learned about the probability θ that people prefer Mars over Snickers?



Mode = 0.9

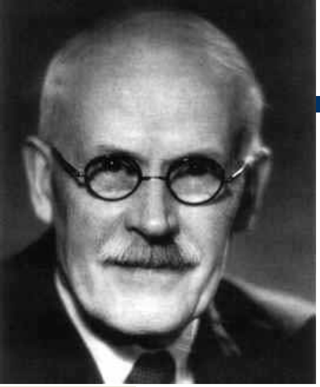
95% confidence interval: (0.59, 0.98)

How to Deal with Utilities

- Choose the action A that maximizes your expected utility,

$$\int u(A,\theta)\,P(\theta \mid A)\,d\theta$$

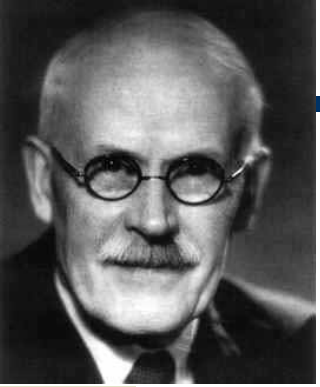
where u(A,θ) is the utility of taking action A when the state of the world is given by a particular value θ. Note how all uncertainty about θ is taken into account.



Bayesian Model Selection

- Suppose we have two models, M_1 and M_2.
- After seeing the data, which one is preferable?

- The one that has the highest posterior probability!
- Compare $P(M_1 \mid D)$ to $P(M_2 \mid D)$.



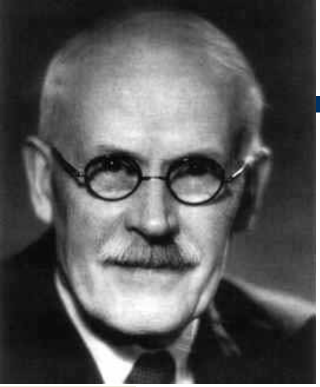
Bayesian Model Selection

$$\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{P(D \mid M_1)}{P(D \mid M_2)} \times \frac{P(M_1)}{P(M_2)}$$

Posterior odds

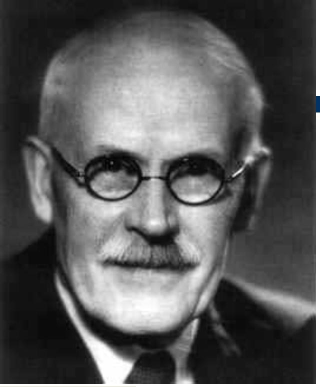
Bayes factor

Prior odds



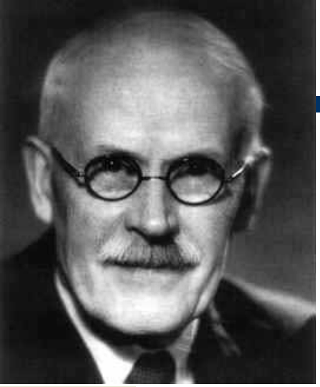
The Bayes Factor

- Is the ratio of the marginal probabilities of the data.
- Is the change from prior to posterior odds brought about by the data.
- Quantifies the evidence for one model versus the other provided by the data.



Interpretation of the Bayes Factor

BF	Evidence
1-3	Anecdotal
3-10	Substantial
10-30	Strong
30-100	Very strong
>100	Decisive

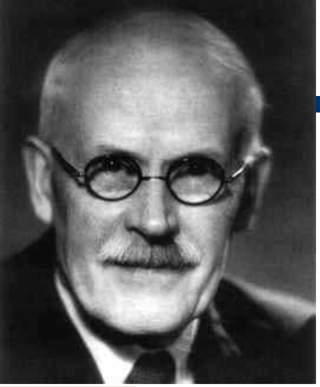


Marginal Likelihood

$$P(D \mid M_i) = \int P(D \mid \theta, M_i) P(\theta \mid M_i) d\theta$$

The marginal likelihood is the probability of the data with the model parameters integrated out.

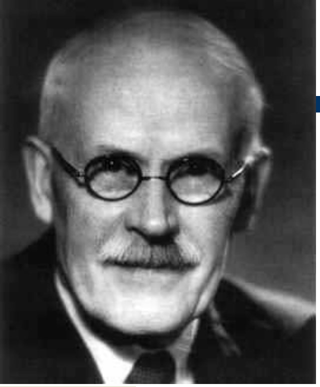
The marginal likelihood is a weighted average of the likelihood, where the weights are given by the prior.



Marginal Likelihood

$$P(D \mid M_i) = \int P(D \mid \theta, M_i) P(\theta \mid M_i) \, d\theta$$

A model with vague priors (i.e., priors that make everything possible by being very spread out) will generally have a low marginal likelihood. Models are punished for making false predictions.



Marginal Likelihood

$$P(D \mid M_i) = \int P(D \mid \theta, M_i) P(\theta \mid M_i) d\theta$$

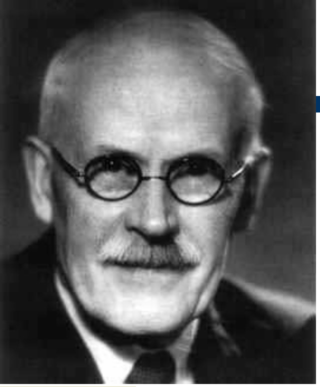
This means that marginal likelihoods prefer simple models over complex models; an automatic Occam's razor!

Marginal Likelihood

$$P(D \mid M_i) = \int P(D \mid \theta, M_i) P(\theta \mid M_i) d\theta$$

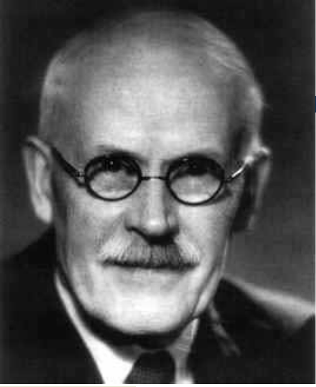
Note that we can use the Bayes factor to compare the support for any two models, nested or not.

In addition, the null hypothesis does not enjoy a special status.



Example Revisited

- Twelve participants choose between a Mars bar and a Snickers bar. In the order that participants choose the bars, the data (D) take the form $D = (M, M, M, S, M, M, M, M, S, M, M, S)$
- Desired is model selection concerning the preference rate θ: What about the hypothesis that the group does not show a preference (i.e., $\theta = 0.5$) for Mars vs. Snickers bars?

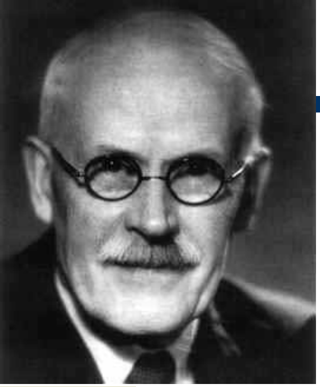


Example Revisited

- We used the binomial model, in which P(D|θ) is given by

$$P(D \mid \theta) = \binom{n}{s} \theta^{s} (1-\theta)^{n-s}$$

where n =12 is the number of trials, and s=3 is the number times that Snickers was selected.

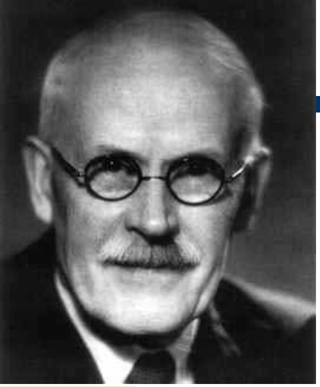


Example Revisited

$$BF_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} = \frac{P\left(D \mid M_1 : \theta = \frac{1}{2}\right)}{P\left(D \mid M_2 : \theta \neq \frac{1}{2}\right)}$$

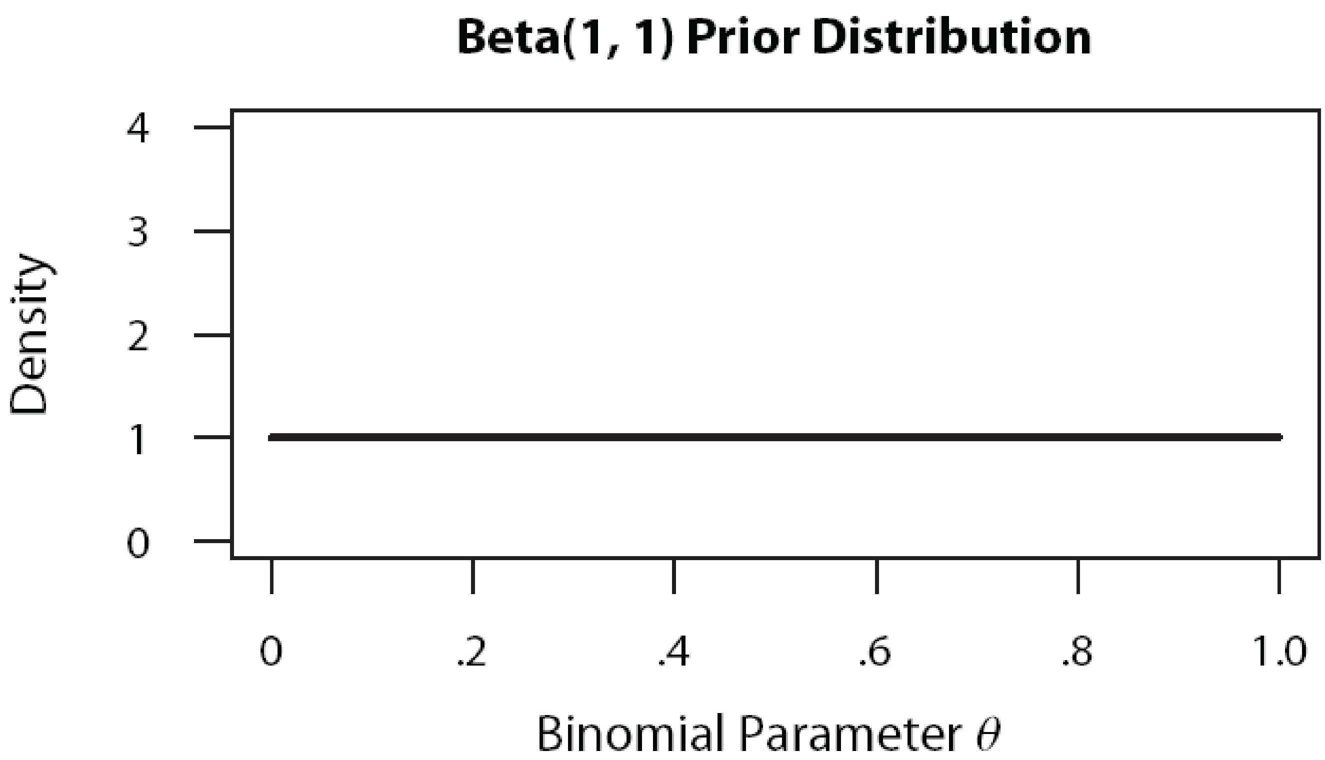
$$= \frac{\binom{12}{3}\left(\frac{1}{2}\right)^{12}}{\int_0^1 P(D \mid \theta) P(\theta) \, d\theta}$$

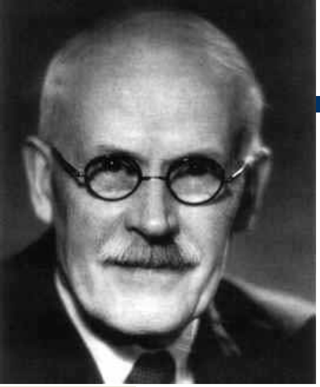
Prior distribution for θ under M_2



Example Revisited

Suppose all values of θ are equally likely a priori:



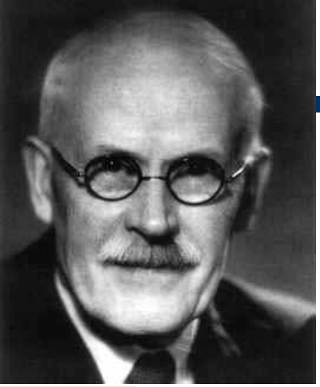


Example Revisited

Then $\int_0^1 P(D \mid \theta) P(\theta)\, d\theta = \dfrac{1}{n+1}$

and $BF_{12} = (n+1)\binom{12}{3}\left(\dfrac{1}{2}\right)^{12} \approx 0.70$

This means that the data are $1/0.70 \approx 1.4$ more likely under M_2 than under M_1.



Example Revisited

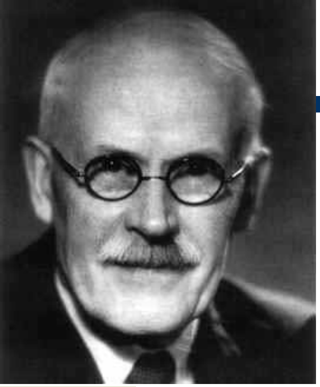
- When we assume that the models are equally likely a priori, we can go from the Bayes factor to the posterior probability:

$$P(M_1 \mid D) \approx \frac{0.7}{1+0.7} \approx .41$$

Example Revisited

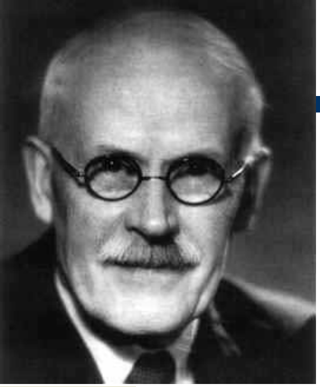
- Suppose M1 predicts X1 and M2 predicts X2.
- Then, a "model-averaged prediction" that takes into account the uncertainty about the models is computed as follows:

$$X_{ave} = P\left(M_1 \mid D\right) \cdot X_1 + P\left(M_2 \mid D\right) \cdot X_2$$



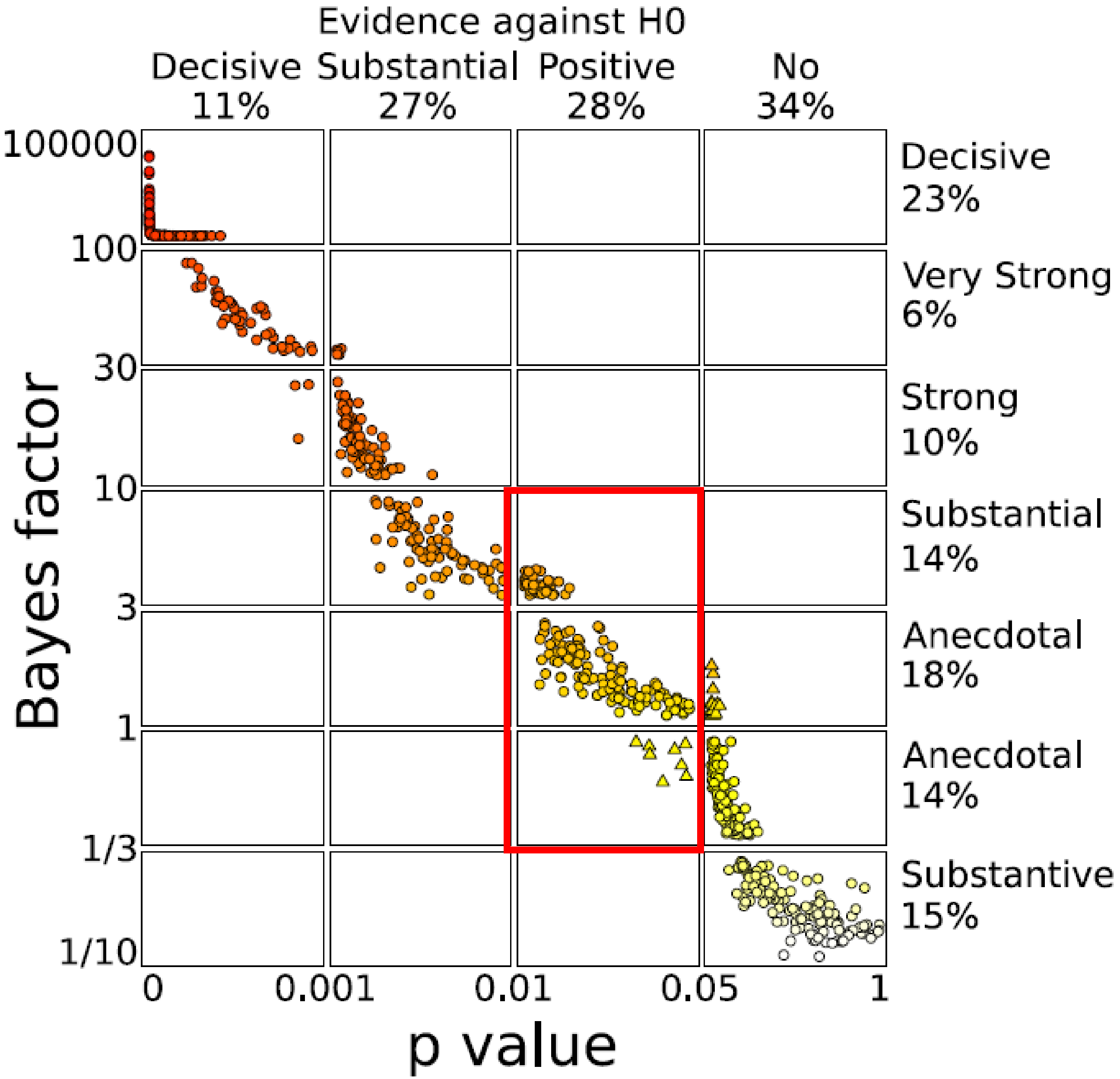
Empirical Comparison

- Does a Bayesian t-test (Rouder et al., 2009, PBR) yield the same conclusions than the traditional t-test?
- “We” collected all t-tests reported in the 2007 issues of Psychonomic Bulletin & Review and JEP:LMC.

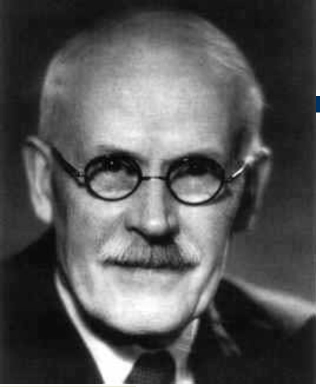


Empirical Comparison

- In 252 articles, spanning 2394 pages, “we” found 855 t-tests.
- This translates to an average of one t-test for every 2.8 pages, or about 3.4 t-tests per article.



Evidence against H0
Decisive 11%
Substantial 27%
Positive 28%
No 34%
Decisive 23%
Very Strong 6%
Strong 10%
Substantial 14%
Anecdotal 18%
Anecdotal 14%
Substantive 15%
100000
100
30
10
3
1
1/3
1/10
Bayes factor
0
0.001
0.01
0.05
1
p value



Empirical Comparison

- Bayes factors and p-values agree on the direction of the effect: p-values < .05 yield evidence against H0, and p-values > .05 yield evidence against H1.
- Bayes factors and p-values often disagree on the strength of the effect: p-values in the .01-.05 interval are likely to yield evidence that is only “anecdotal”.

Outline

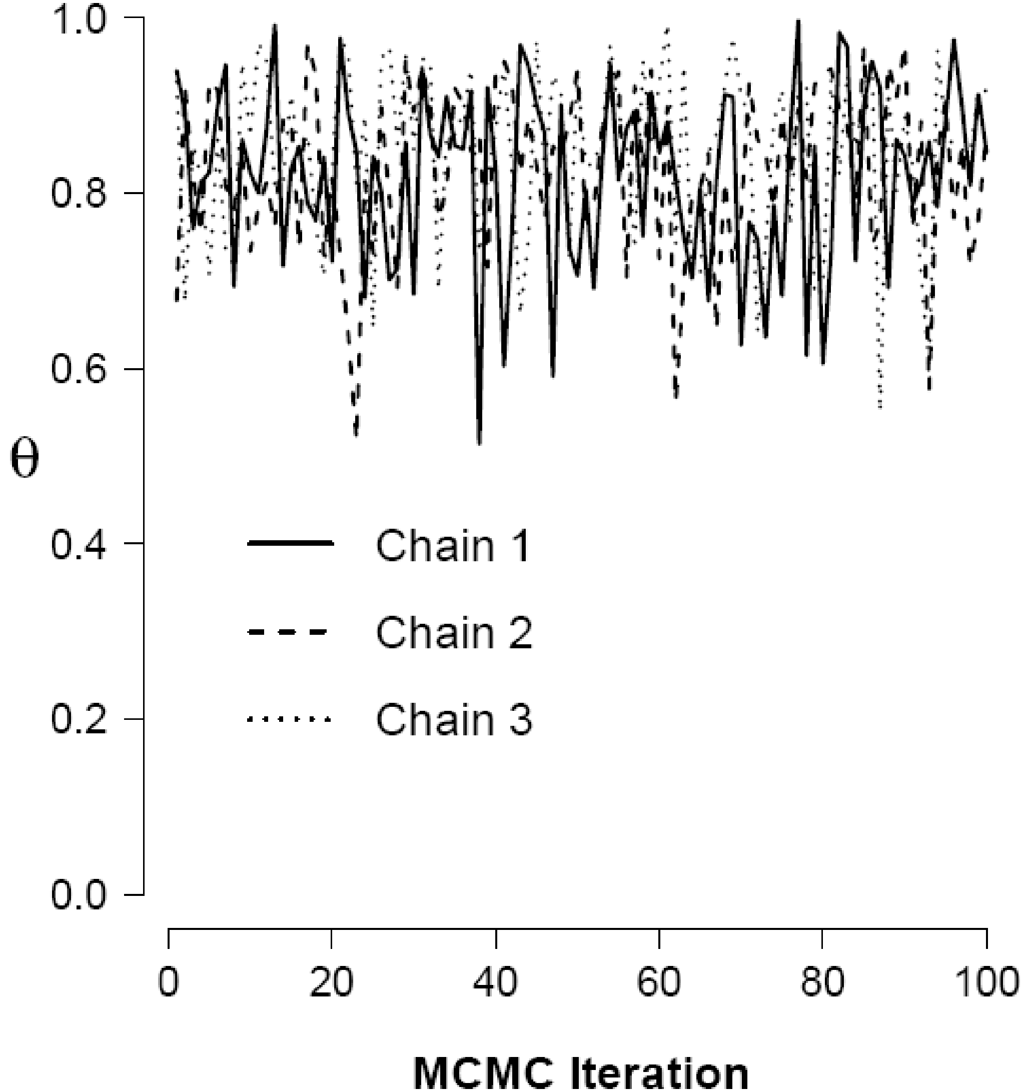
- Bayesian Basics
- Bayesian Revolution and MCMC
- WinBUGS
- Bayesian Sparsity

The Bayesian Revolution

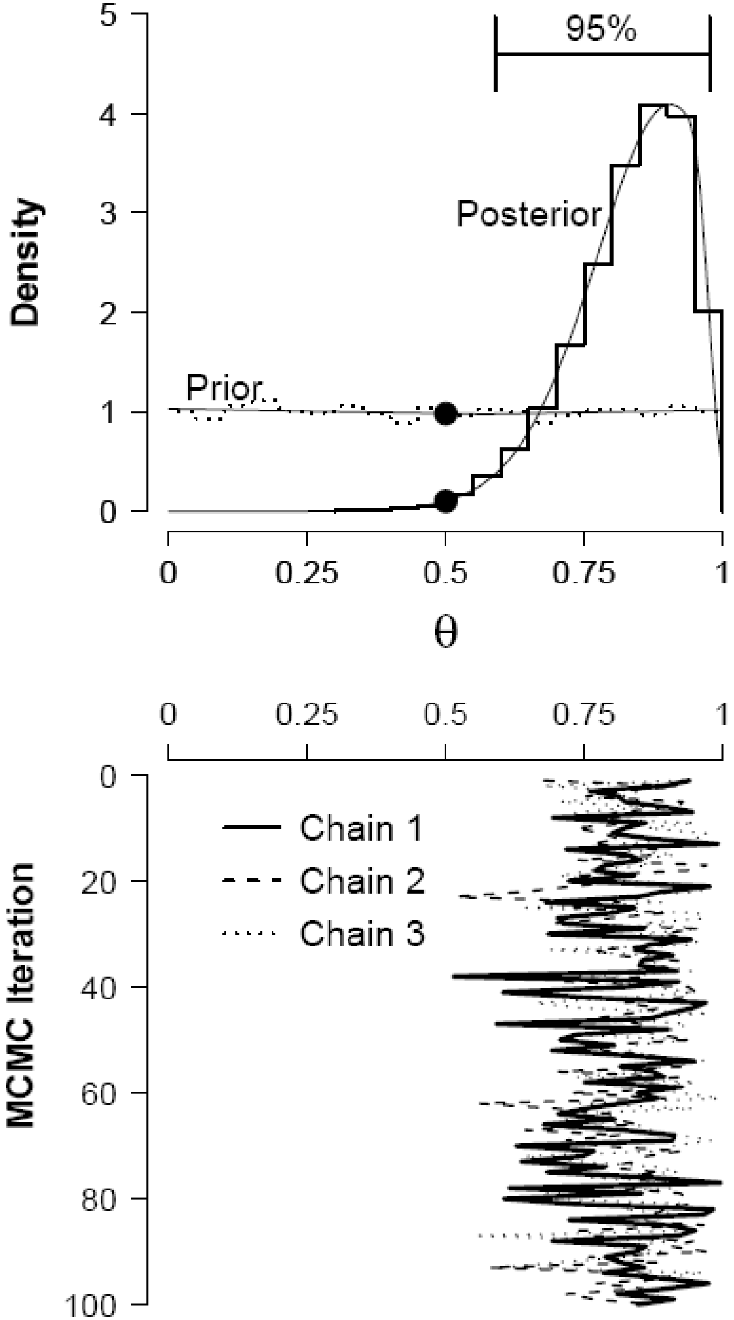
- Until about 1990, Bayesian statistics could only be applied to a select subset of very simple models.
- Only recently, Bayesian statistics has undergone a transformation; With current numerical techniques, Bayesian models are "limited only by the user's imagination."

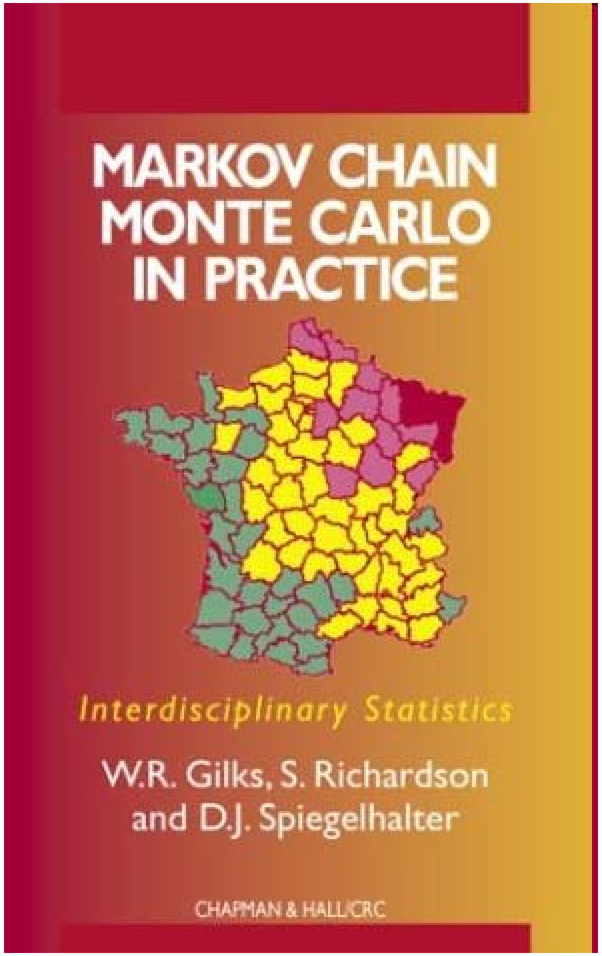
Markov Chain Monte Carlo

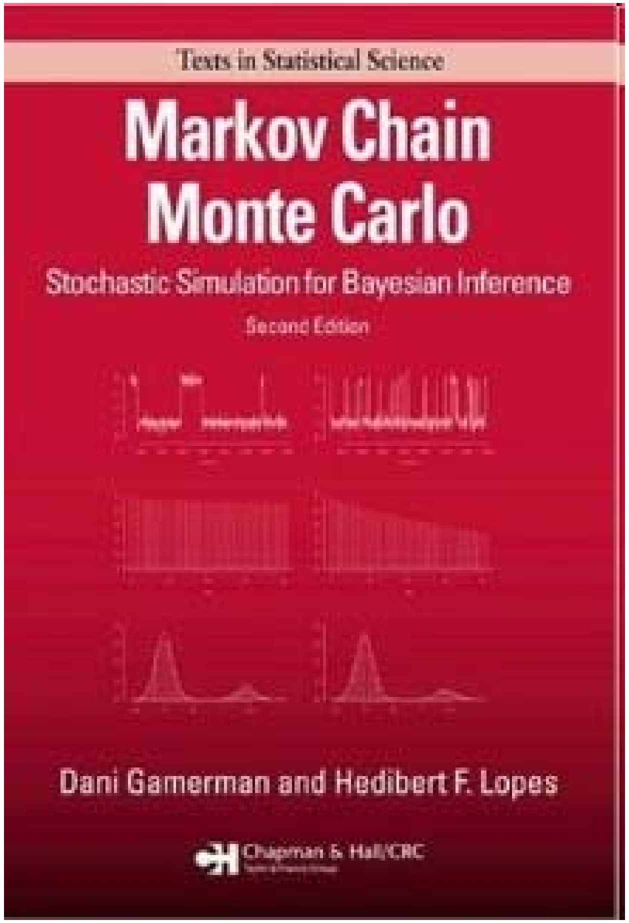
- Instead of calculating the posterior analytically, numerical techniques such as MCMC approximate the posterior by drawing samples from it.
- Consider again our earlier example…



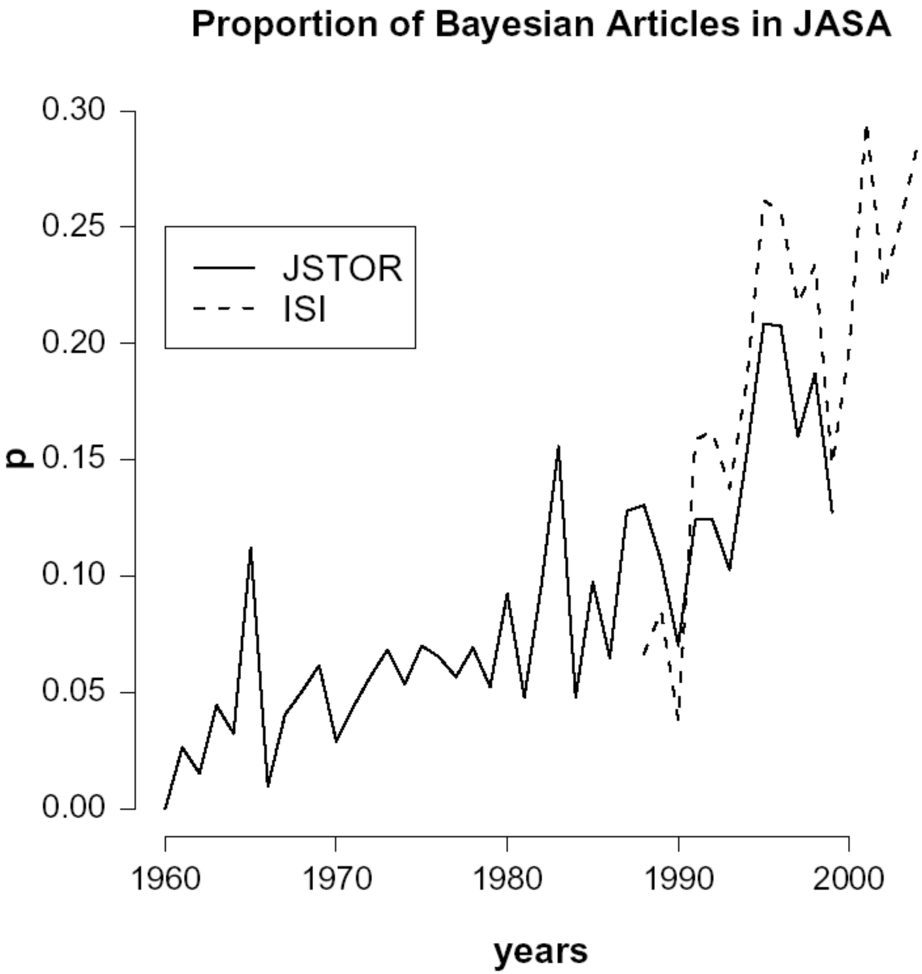
1.0
0.8
0.6
0.4
0.2
0.0
θ
Chain 1
Chain 2
Chain 3
0
20
40
60
80
100
MCMC Iteration


95%
Posterior
Prior
Density
0
1
2
3
4
5
0
0.25
0.5
0.75
1
θ
Chain 1
Chain 2
Chain 3
MCMC Iteration
0
20
40
60
80
100

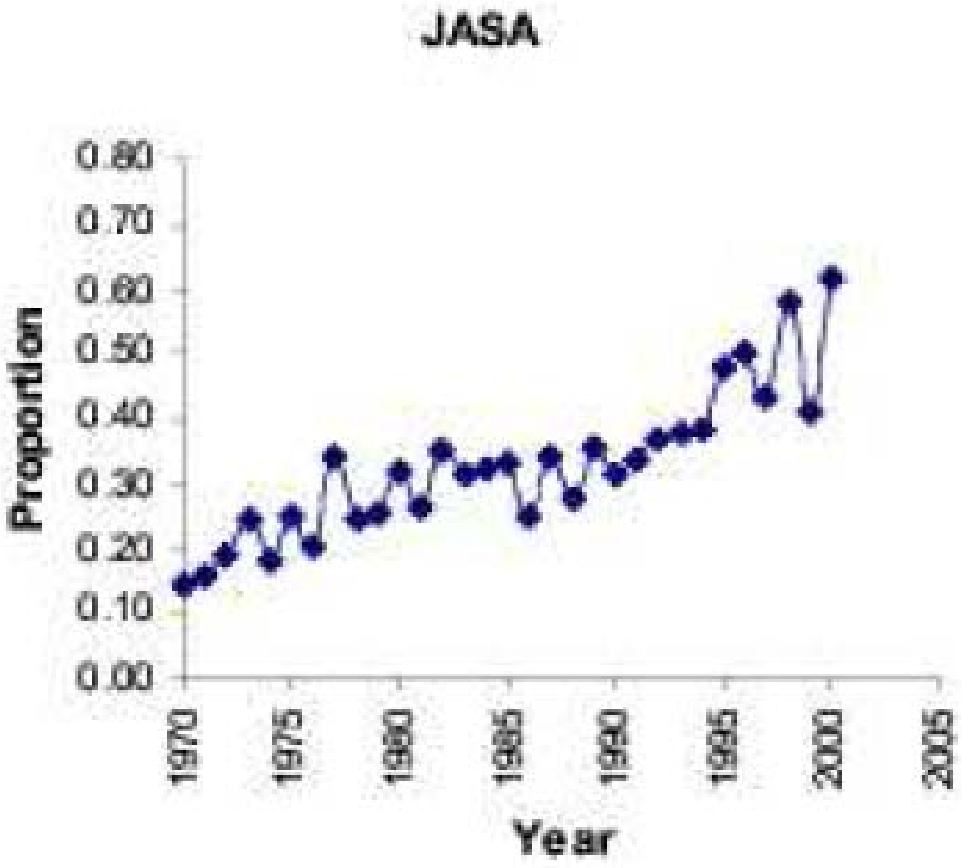

MARKOV CHAIN
MONTE CARLO
IN PRACTICE
Interdisciplinary Statistics
W.R. Gilks, S. Richardson
and D.J. Spiegelhalter
CHAPMAN & HALL/CRC

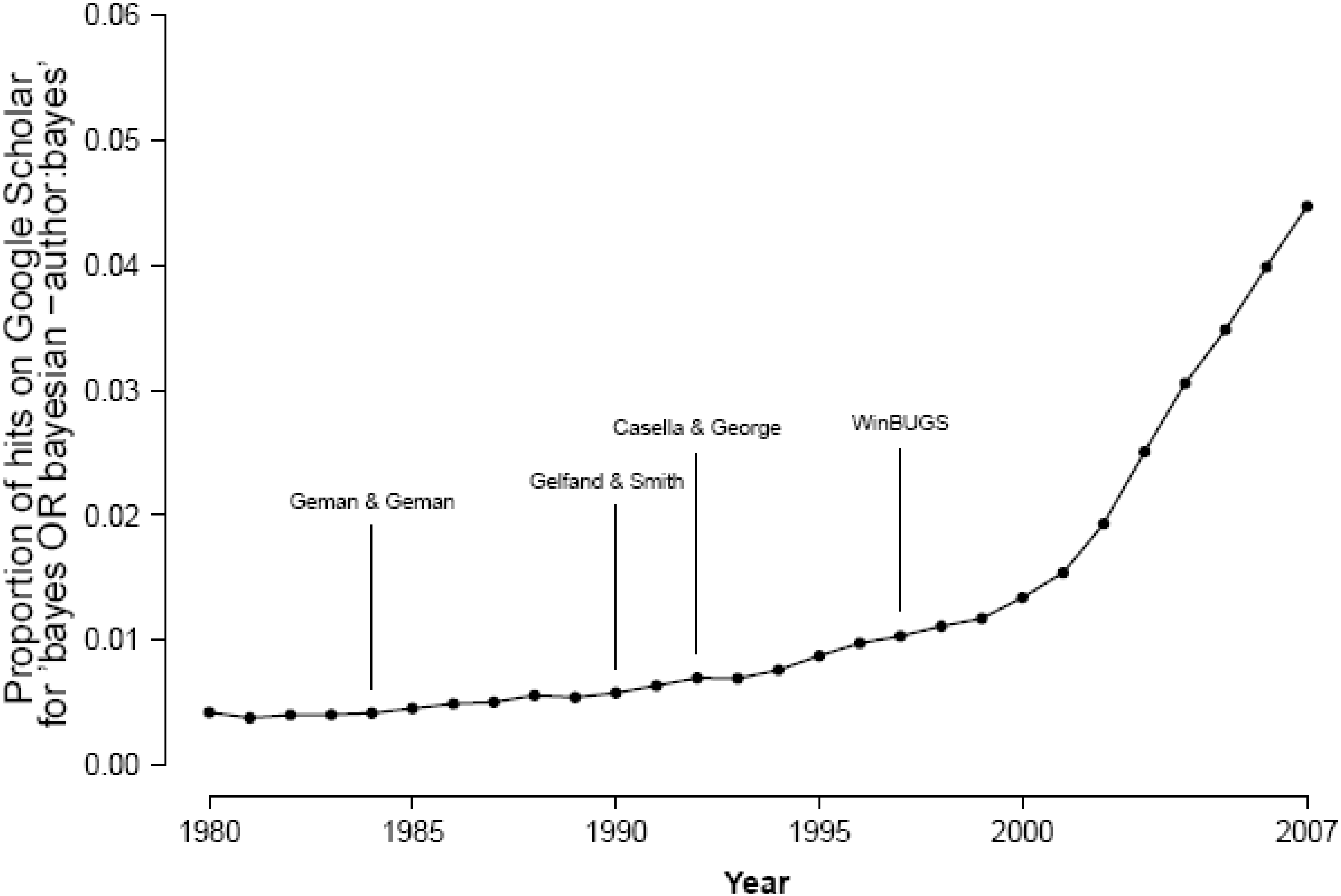

Texts in Statistical Science
Markov Chain
Monte Carlo
Stochastic Simulation for Bayesian Inference
Second Edition
Dani Gamerman and Hedibert F. Lopes
Chapman & Hall/CRC

The Bayesian Revolution in Statistics

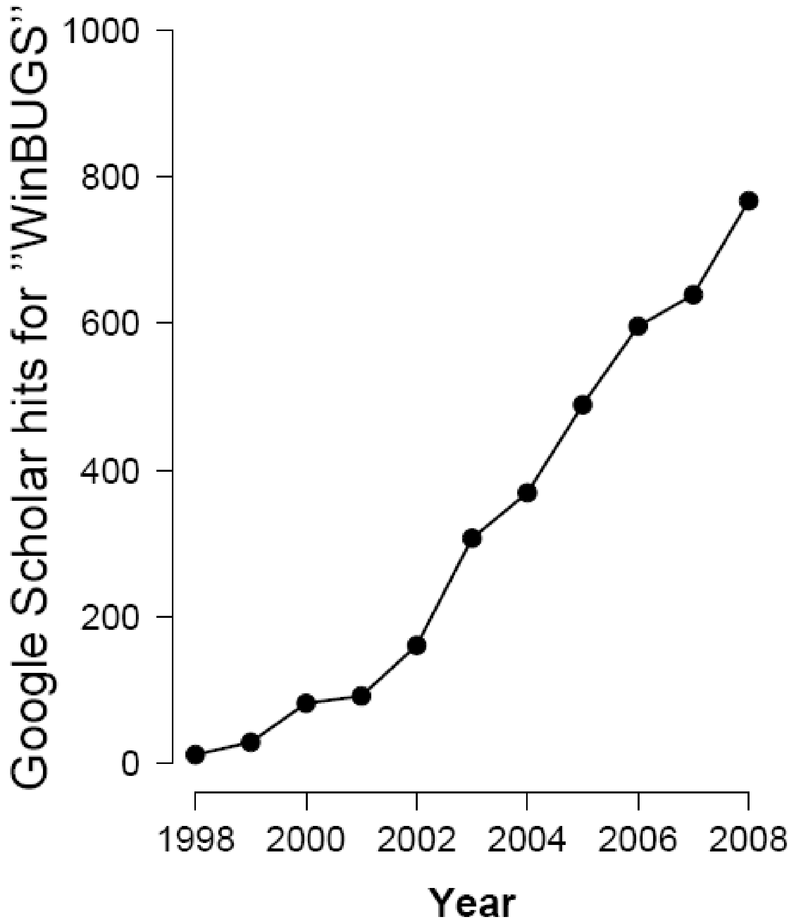


The Bayesian Revolution in Statistics (Poirier, 2006)

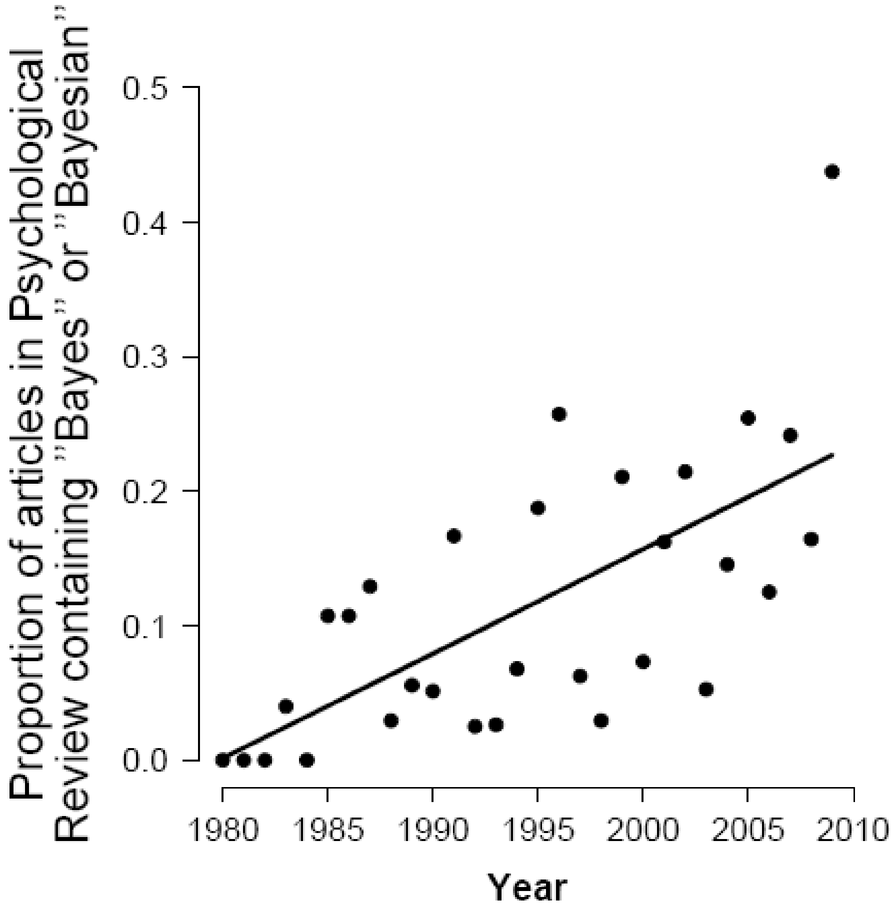



0.06
0.05
0.04
0.03
0.02
0.01
0.00
Proportion of hits on Google Scholar
for 'bayes OR bayesian -author:bayes'
Geman & Geman
Gelfand & Smith
Casella & George
WinBUGS
1980
1985
1990
1995
2000
2007
Year

The Bayesian Revolution in Statistics



The Bayesian Revolution in Psychology?



How About Sensometrics?

?

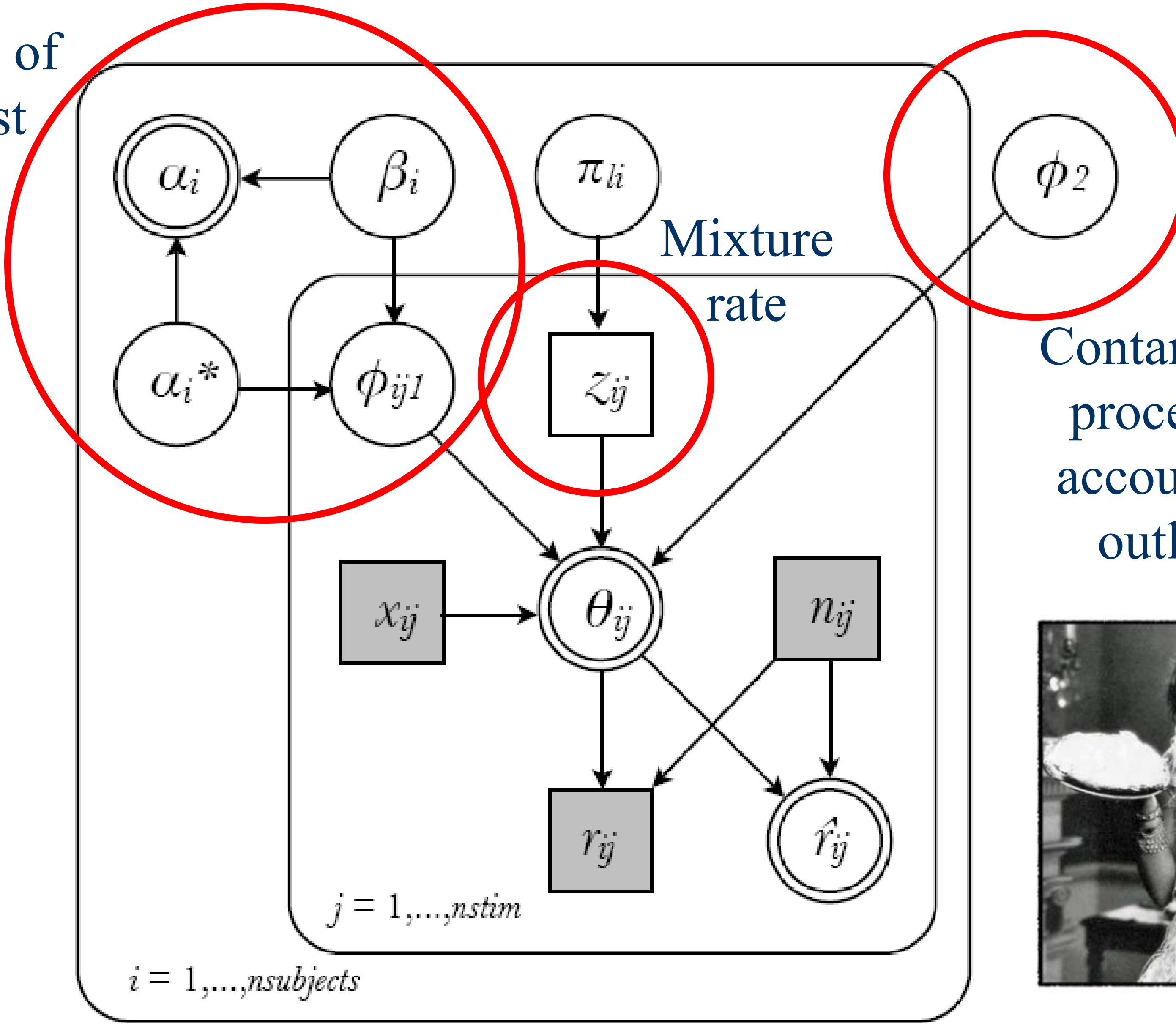
Outline

- Bayesian Basics
- Bayesian Revolution and MCMC
- WinBUGS
- Bayesian Sparsity

WinBUGS

- Allows researchers to specify complex models using simple building blocks.
- Model structure is that of a Directed Acyclical Graph.
- MCMC sampling routines are applied automatically, without users have to hand-code these themselves.

Process of interest

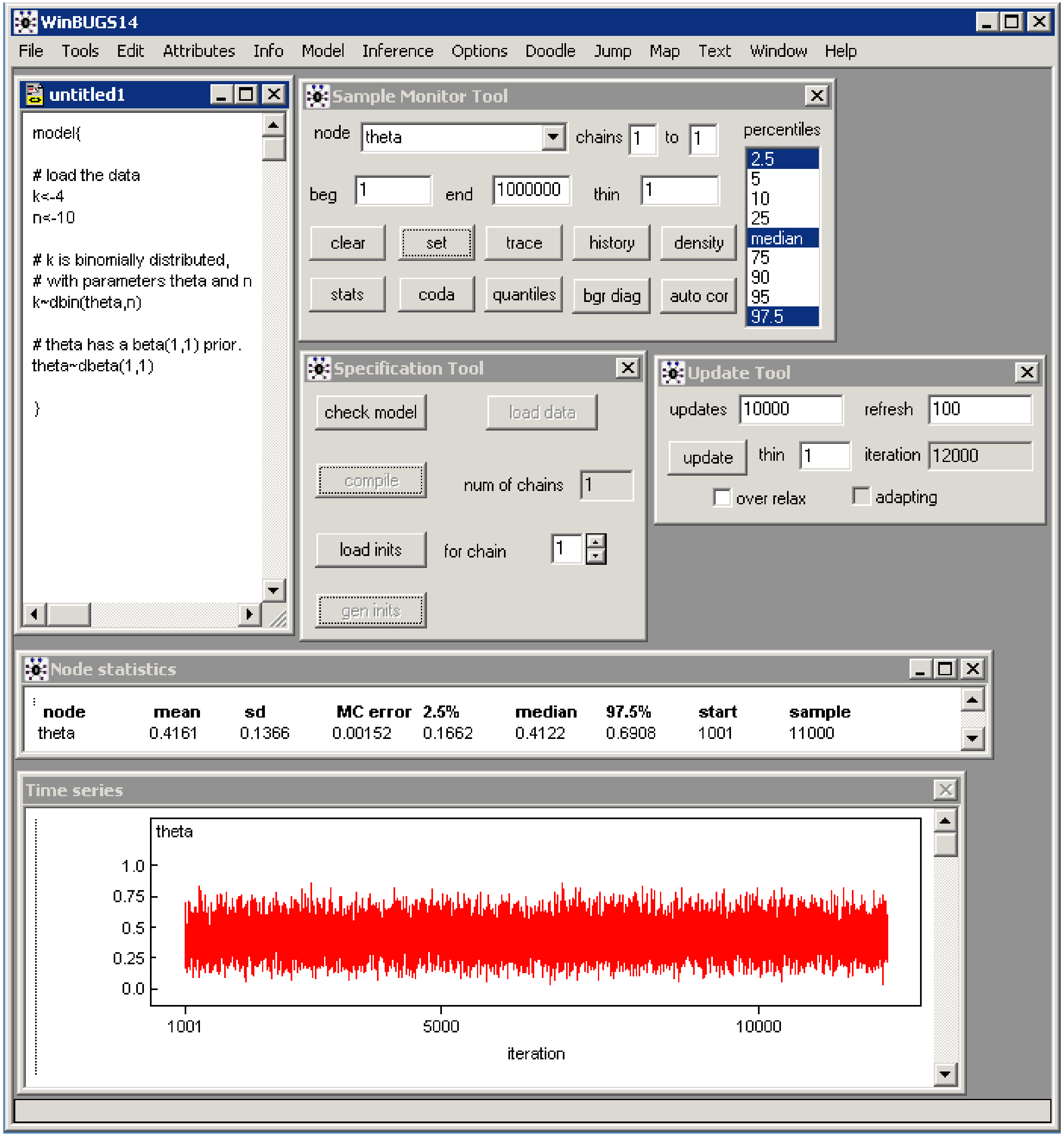


Contaminant process to account for outliers

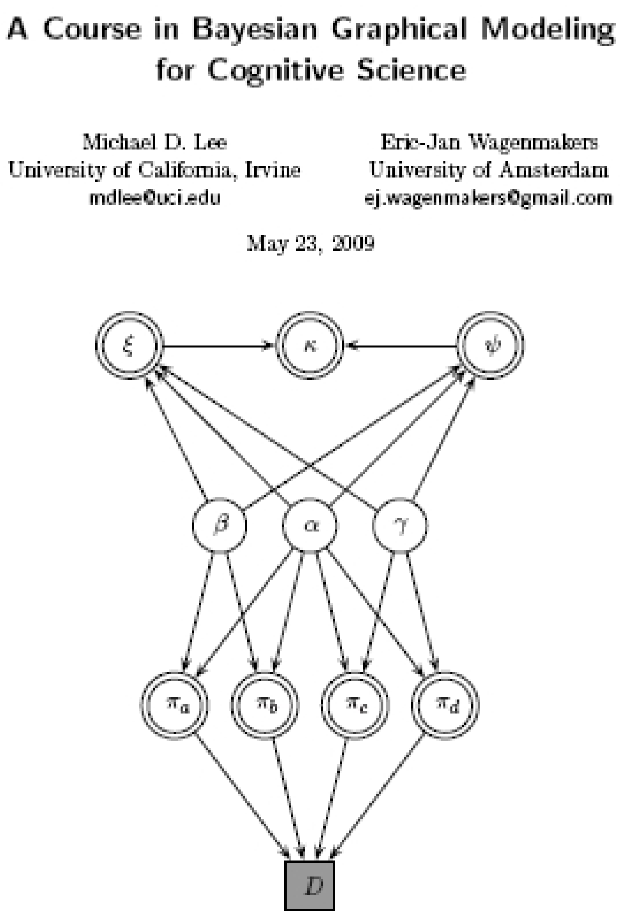
WinBUGS

- Missing dependent variables are encoded as “NA”.
- Subsequently, MCMC sampling generates a series of best guesses for the missing data, in the end generating

$$P\left(D_{mis} \mid D_{obs}, \theta\right)$$


WinBUGS14
File Tools Edit Attributes Info Model Inference Options Doodle Jump Map Text Window Help
untitled1
model{
# load the data
k<-4
n<-10
# k is binomially distributed,
# with parameters theta and n
k~dbin(theta,n)
theta has a beta(1,1) prior.
theta~dbeta(1,1)
}
Sample Monitor Tool
node theta chains 1 to 1 percentiles
beg 1 end 1000000 thin 1
clear set trace history density
stats coda quantiles bgr diag auto cor
2.5 5 10 25 median 75 90 95 97.5
Specification Tool
check model load data
compile num of chains 1
load inits for chain 1
gen inits
Update Tool
updates 10000 refresh 100
update thin 1 iteration 12000
over relax adapting
Node statistics
node mean sd MC error 2.5% median 97.5% start sample
theta 0.4161 0.1366 0.00152 0.1662 0.4122 0.6908 1001 11000
Time series
theta
1.0 0.75 0.5 0.25 0.0
1001 5000 10000
iteration

WinBUGS


A Course in Bayesian Graphical Modeling for Cognitive Science

Michael D. Lee
University of California, Irvine
mdlee@uci.edu

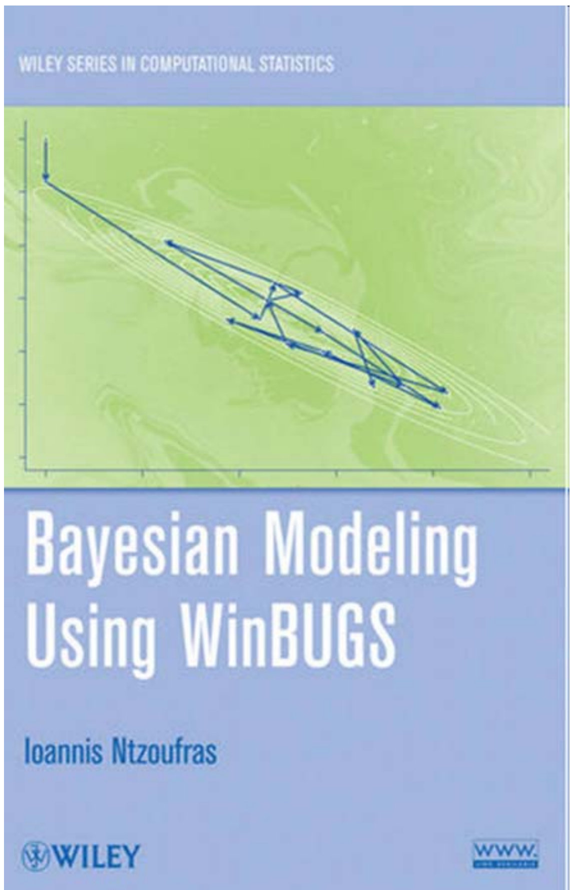
Eric-Jan Wagenmakers
University of Amsterdam
ej.wagenmakers@gmail.com

May 23, 2009

Book (not 100% completed) and code available from

http://www.ejwagenmakers.com/BayesCourse/BayesBook.html

WinBUGS



A review can be found at

http://www.ruudwetzels.com/articles/BMW_review.pdf

Outline

- Bayesian Basics
- Bayesian Revolution and MCMC
- WinBUGS
- Bayesian Sparsity

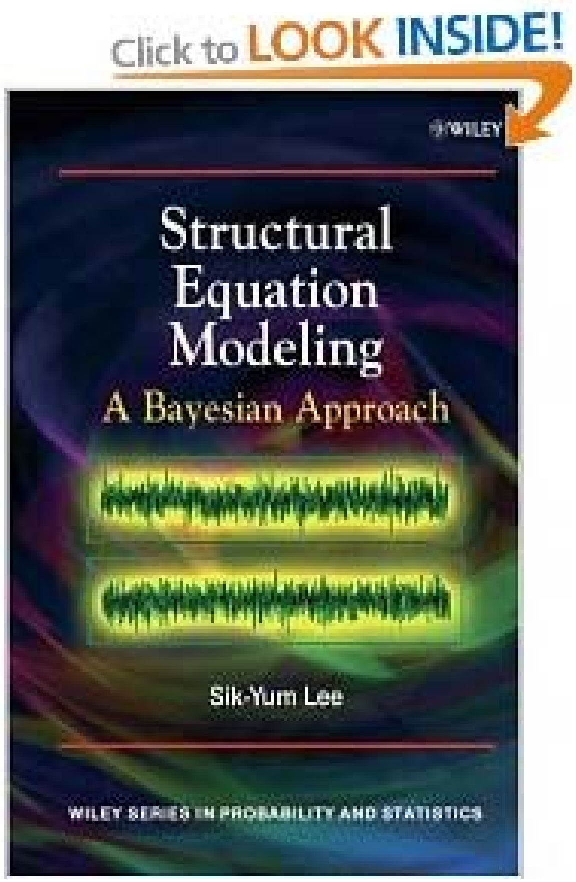
Sparsity

- In exploratory analyses, or with $p >> n$, we prefer parsimonious models.
- How can this be accomplished in the Bayesian paradigm?
- Consider, for concreteness, a linear regression with p regression coefficients.

Sparsity

- Option 1: Bayesian model selection automatically prefers sparse models. See the recent BAS package in R.
- Option 2: Bayesian model selection, but add a penalty that increases with model size.
- Option 3: Bayesian lasso and Bayesian elastic nets.
- Option 4: Bayesian latent factor models.

Bayesian SEM!



Treat latent factors as missing data.

In WinBUGS we can then work directly with the raw data (not the covariances).

Seems to be much simpler than current approaches.

Too few data points simply results in wide posterior distributions.

Concluding Comments

- "Bayesian inference is right, and everything else is wrong".
- "It is better to wrestle with the practicalities of a method that is fundamentally sound than it is to interpret the outcomes of a method that is fundamentally unsound."
- It can be Bayesian not to be Bayesian! Is this why many people still say that they are not Bayesian?

Thanks for Your Attention!