

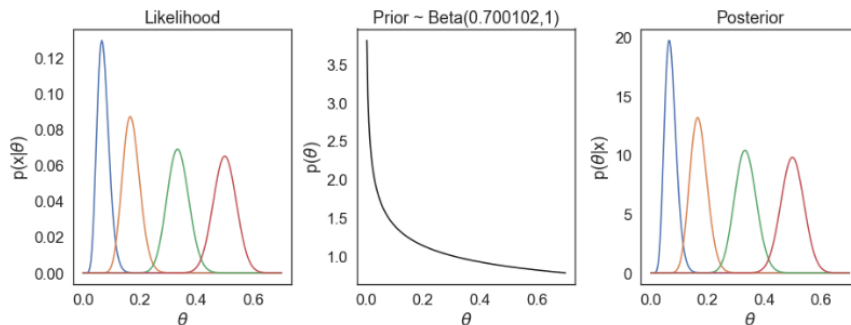
# Mathematical Statistics II

STA2212H S LEC9101

Week 8

March 10 2021

Start recording!





**Boyang Zhao**

Senior data scientist at ING Bank  
Computational biology, genomics,  
systems biology, finance, banking  
@MIT PhD

📍 Netherlands

🐦 Twitter

🌐 LinkedIn

🏠 GitHub

### Vaccine efficacy

The way we measure vaccine efficacy is defined as follows,

$$\text{vaccine efficacy (VE)} = 1 - R$$

The R can be ratio of risks (RR, risk ratio); rates (IRR, incidence rate ratio); or hazards (HR, hazard ratio). Because of the ratios, we see that vaccine efficacy is a relative measure - in how much relative reduction in infection or disease in the vaccinated group compared to the unvaccinated group. A VE of 90% means there are 90% fewer cases in the vaccinated group compared to the placebo group. We will use the subscripts  $v$  and  $p$  to denote vaccinated and placebo groups, respectively; but obviously the discussion is applicable for comparing between any two treatment arms - not necessarily have to be a placebo.

#### With RR

$$\text{VE} = 1 - \text{RR} = 1 - \frac{c_v/N_v}{c_p/N_p}$$

where  $N_v$  and  $N_p$  are the total number of participants in the vaccinated and placebo group, respectively.

#### With IRR

$$\text{VE} = 1 - \text{IRR} = 1 - \frac{c_v/T_v}{c_p/T_p}$$

where  $T_v$  and  $T_p$  are the time-person years for the vaccinated and placebo group, respectively.

#### With HR

$$\text{VE} = 1 - \text{HR} = 1 - \frac{\lambda_v}{\lambda_p}$$

where  $\lambda_v$  and  $\lambda_p$  are the hazard rates for the vaccinated and placebo group, respectively. This measures the relative reduction in the hazard of infection. The hazard ratio can be

$$1 - \frac{8}{162} = \frac{n_v}{n_c}$$

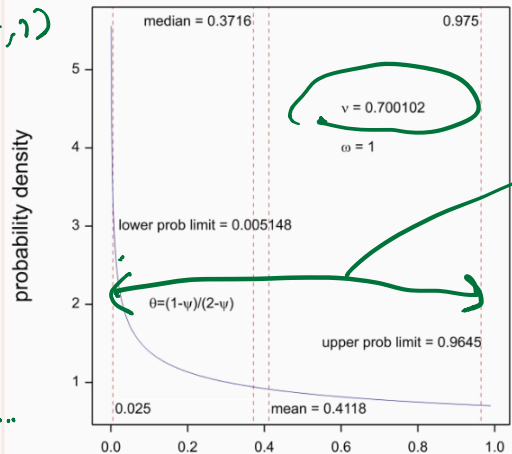
[link](#)

# More on vaccines

Stephen Senn's blog post

[link](#)

Beta prior distribution for Pfizer/BioNTech vaccine trial



$$(0,1) \ni \theta = \pi_v / (\pi_v + \pi_c)$$

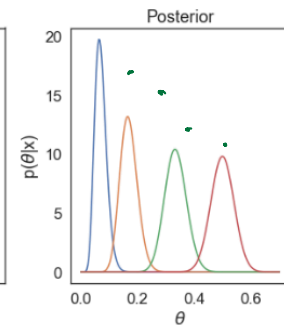
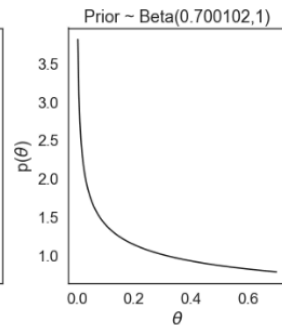
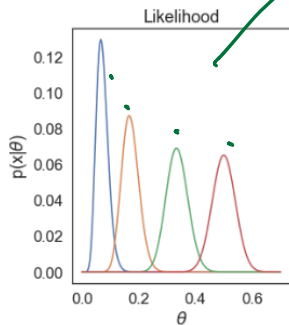
Figure 4 Prior distribution for  $\theta$

95% prior

fake data

$$n_v = 10, 25, 50, 75$$

$$n_v + n_c = 150$$



$$\theta = \frac{\pi_v}{\pi_v + \pi_c} \Rightarrow \pi(VE)$$

$$\theta \in (0,1) \\ VE \in (-\infty, 1)$$

$$\theta = (1 - VE) / (2 - VE)$$



## US election polls: a quick postmortem

How did the 2020 US presidential election polls really do?  
**Ole J. Forsberg** gives his assessment

The American Association for Public Opinion Research (AAPOR) is expected to produce a report early this year that explores the strengths and weaknesses of the polls in the 2020 US election cycle. The polls were criticised in some quarters immediately after the election, when it became clear that Donald Trump had done better than expected and that Joseph R. Biden Jr's margin of victory in the popular vote was not as large as anticipated.<sup>1</sup>

In preparation for this report, I wanted to provide some insight into the polls and some suggestions of my own for moving forward. Specifically, I hope to convince polling houses to use some type of model averaging – or even Bayesian methods – to

closing weeks of the campaign.

The first source of error, faulty weighting, is extremely important for polling houses to take seriously. While the number of US polling houses taking education level into consideration increased in 2020, the education characteristics of the voting population remain uncertain.

“Shy voters” – the second source of error – may be more myth than reality ([53eig.ht/3oNEb6R](https://53eig.ht/3oNEb6R)). But whether shy or not, there are some voters who either choose not to respond to polls, or who choose not to answer honestly when surveyed. Pollsters need to address this, either by asking additional questions to model respondent preference for those who choose not to say how they will vote, or by finding new ways to encourage the public to

interpretation, not of polling.

The mistake happens in how we interpret a poll result such as “48% Biden, 44% Trump”. Do we focus on the two-party vote and claim that Biden is ahead, or do we acknowledge that there is a sizeable portion of voters – 8% – who may only decide how to vote once in the polling booth? Clearly, the latter interpretation is more appropriate, but it makes for a less straightforward story, so these undecided voters tend to be overlooked in media reports.

### Missing data

The majority of polls in the 2020 election cycle contained just three response options for those asked about their intended vote: “Biden”, “Trump”, and “undecided”. The implied fourth

mentioned earlier constitute a huge amount of missing data about voting intention. Ignoring these missing data leads to false precision in the polls’ assessment of the state of the election.

While some undecided voters ultimately will not vote, many will eventually decide between the two candidates. This increases the uncertainty in polling estimates beyond what is reported in terms of confidence intervals and margins of error. As a result, when those late-deciding voters finally vote, polls may look very wrong.

To illustrate this point, compare the polls in the final two weeks of the 2020 election to the final election result (Table 1). In this sample of 174 polls, the actual Biden vote was within the polls’ margins of error 85% of the time, while the actual Trump vote was within the polls’ margins of error only 43% of the time. For the 57% of confidence intervals that missed Trump’s actual vote, they were always too low, never too high – meaning that the polls consistently underestimated Trump’s final vote. The 15% of confidence intervals



**Table 1:** Results from comparing candidate support levels in polls from the last two weeks of the US presidential election with the actual outcome of the election (vote share). Polls are a mix of state-level and national polls from a variety of polling houses, using a variety of methods.

Source	n	Confidence interval hits		Average miss (standard error)	
		Biden	Trump	Biden	Trump
All polls	174	85% (79% to 90%)	43% (35% to 50%)	-0.09	+2.41
Online only	23	96% (78% to 99%)	30% (13% to 53%)	-0.79	+2.21
Online + telephone	26	92% (75% to 99%)	54% (33% to 73%)	-0.78	+2.24
Telephone only	125	82% (74% to 88%)	42% (34% to 52%)	-0.18	+2.48
University	60	92% (82% to 97%)	27% (16% to 40%)	-0.10	+2.99
Non-university	114	82% (73% to 88%)	51% (41% to 60%)	-0.09	+2.10
Partisan	52	79% (65% to 89%)	75% (61% to 86%)	+0.62	+1.33
Non-partisan	122	88% (81% to 93%)	29% (21% to 38%)	-0.40	+2.87

“Personally, I favour the Bayesian solution because it provides a solid statistical structure for estimation and communication of results.”

$$\sqrt{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{n}} \times 2 = \sqrt{\frac{1}{n}} \quad \frac{1}{\sqrt{n}} \quad \tilde{p} = \frac{1}{2} \quad \pm \sqrt{\tilde{p}(1-\tilde{p})/n} \times 2$$

[link](#)

# Latest issue of Applied Statistics (JRSS C)



**Journal of the Royal Statistical Society: Series C (Applied Statistics)**  
Volume 70, Issue 2  
Pages: 249-506  
March 2021

---

## ISSUE INFORMATION

[Free Access](#)

### Issue Information

Pages: 249-250 | First Published: 08 March 2021

---

## ORIGINAL ARTICLES

### **Finite mixtures of semiparametric Bayesian survival kernel machine regressions: Application to breast cancer gene pathway subgroup analysis**

Lin Zhang, Inyoung Kim  
Pages: 251-269 | First Published: 01 December 2020

---

### **Quantile-frequency analysis and spectral measures for diagnostic checks of time series with nonlinear dynamics**

Ta-Hsin Li  
Pages: 270-290 | First Published: 22 November 2020

# Latest issue of Applied Statistics (JRSS C)



Journal of the Royal Statistical Society: Series C (Applied Statistics)

Volume 70, Issue 2

Pages: 249-506

March 2021

## ISSUE INFORMATION

[Free Access](#)

### Issue Information

Pages: 249-250 | First Published: 08 March 2021

## ORIGINAL ARTICLES

### Finite mixtures of semiparametric Bayesian survival kernel Application to breast cancer gene pathway subgroup analysis

Lin Zhang, Inyoung Kim

Pages: 251-269 | First Published: 01 December 2020

### Quantile-frequency analysis and spectral measures for data with nonlinear dynamics

Ta-Hsin Li

Pages: 270-290 | First Published: 22 November 2020

[Open Access](#)

### Future proofing a building design using history matching inspired level-set techniques

Evan Baker, Peter Challenor, Matt Eames

Pages: 335-350 | First Published: 19 December 2020

### Recurrent events modelling of haemophilia bleeding events

Andrew C. Titman, Martin J. Wolfsegger, Thomas F. Jaki

Pages: 351-371 | First Published: 07 January 2021

### Multiscale null hypothesis testing for network-valued data: Analysis of brain networks of patients with autism

Ilenia Lovato, Alessia Pini, Aymeric Stamm, Maxime Taquet, Simone Vantini

Pages: 372-397 | First Published: 22 January 2021

[Open Access](#)

### Bayesian semi-parametric G-computation for causal inference in a cohort study with MNAR dropout and death

Maria Josefsson, Michael J. Daniels

Modeling multilevel spatial behaviour in binary-mark muscle fibre configurations .....	TILMAN M. DAVIES, MATTHEW R. SCHOFIELD, JON CORNWALL AND PHILIP W. SHEARD	1329
Identifying and estimating principal causal effects in a multi-site trial of Early College High Schools .....	LO-HUA YUAN, AVI FELLER AND LUKE W. MIRATRIX	1348
Imputation and post-selection inference in models with missing data: An application to colorectal cancer survival guidelines	LIN LIU, YUQI QIU, LOKI NATARAJAN AND KAREN MESSER	1370
A hidden Markov model approach to characterizing the photo-switching behavior of fluorophores .....	LEKHA PATEL, NILS GUSTAFSSON, YU LIN, RAIMUND OBER, RICARDO HENRIQUES AND EDWARD COHEN	1397
Identifying multiple changes for a functional data sequence with application to freeway traffic segmentation .....	JENG-MIN CHIOU, YU-TING CHEN AND TAILIEN HSING	1430
The classification permutation test: A flexible approach to testing for covariate imbalance in observational studies .....	JOHANN GAGNON-BARTSCH AND YOTAM SHEM-TOV	1464
Spatio-temporal short-term wind forecast: A calibrated regime-switching method	AHMED AZIZ EZZAT, MIKYOUNG JUN AND YU DING	1484
Network modelling of topological domains using Hi-C data .....	Y. X. RACHEL WANG, PURNAMITA SARKAR, OANA URSU, ANSHUL KUNDJE AND PETER J. BICKEL	1511
Fast dynamic nonparametric distribution tracking in electron microscopic data	YANJUN QIAN, JIANHUA Z. HUANG, CHIWOO PARK AND YU DING	1537
Distributional regression forests for probabilistic precipitation forecasting in complex terrain .....	LISA SCHLOSSER, TORSTEN HOTHORN, RETO STAUFER AND ACHIM ZEILEIS	1564
Modeling seasonality and serial dependence of electricity price curves with warping functional autoregressive dynamics .....	YING CHEN, J. S. MARRONE AND JIEJIE ZHANG	1590
RCNorm: An integrated system of random-coefficient hierarchical regression models for normalizing NanoString nCounter data .....	GAOXIANG JIA, XINLEI WANG, QIWEI LI, WEI LU, XIMING TANG, IGNACIO WISTUBA AND XIAN JIE	1617
Network classification with applications to brain connectomics	JESÚS D. ARROYO RELIÓ, DANIEL KESSLER, ELIZAVETA LEVINA AND STEPHAN F. TAYLOR	1648
Sequential decision model for inference and prediction on nonuniform hypergraphs with application to knot matching from computational forestry .....	SEONG-HWAN JUNG, SAMUEL W. K. WONG, JAMES V. ZIDEK AND ALEXANDRE BOULCHARD-CÔTÉ	1678
A Bayesian mark interaction model for analysis of tumor pathology images	QIWEI LI, XINLEI WANG, FANFANG LIANG AND GUANGHUA XIAO	1708
A hierarchical Bayesian model for single-cell clustering using RNA-sequencing data	YIYI LU, JOSHUA L. WARREN AND HONGYU ZHAO	1733

Gaussian CAR  
scaly inf. +  
vogue  
chick

BIC for A

3



$\frac{9}{24}$  B.  $\frac{14}{24}$  collab.

de bac

AN OFFICIAL JOURNAL OF THE  
INSTITUTE OF MATHEMATICAL STATISTICS

Incorporating conditional dependence in latent class models for probabilistic record linkage: Does it matter?.....	HUIPING XU, XIAOCHUN LI, CHANGYU SHEN, XIU L. HUI and SHAUN GRANNIS	1753
Bayesian modeling of the structural connectome for studying Alzheimer's disease .....	ARKAPRABA ROY, SUBHASHIS GHOSAL, JEFFREY PRESCOTT and KINGSHUK ROY CHOUDHURY	1791

B  
Sims  
con. j. priors  
(no chad)

B

2

2

9

Wavelet spectral testing: Application to nonstationary circadian rhythms  
JESSICA K. HARGREAVES, MARINA I. KNIGHT, JON W. PITCHFORD, RACHAEL J. OAKENFULL, SANGEETA CHAWLA, JACK MUNNS AND SETH J. DAVIS 1817

Oblique random survival forests..... BYRON C. JAEGER, D. LEANN LONG, DUSTIN M. LONG, MARIO SIMS, JEFF M. SZYCHOWSKI, YUAN-I MIN, LESLIE A. MCCLURE, GEORGE HOWARD AND NOAH SIMON 1847

Approximate inference for constructing astronomical catalogs from images  
JEFFREY REGIER, ANDREW C. MILLER, DAVID SCHLEGEL, RYAN P. ADAMS, JON D. MCALUFFE AND PRABHAT 1884

Bayesian methods for multiple mediators: Relating principal stratification and causal mediation in the analysis of power plant emission controls  
CHANYMIN KIM, MICHAEL J. DANIELS, JOSEPH W. HOGAN, CHRISTINE CHOIRAT AND CORWIN M. ZIGLER 1927

Radio-iBAG: Radiomics-based integrative Bayesian analysis of multipplatform genomic data..... YUUYI K. K. RAO AND VEERABHADRA BALADANDAYUTHAPANI 1957

A semiparametric modeling approach using Bayesian Adaptive Regression Trees with an application to evaluate heterogeneous treatment effects  
BRET ZELDOV, VINCENT LO RE III AND JASON ROY 1989

BRET ZELDOW, VINCENT LO RE III AND JASON ROY 198

compared

- approximate posterior normality

if  $n$  not large, need better

$$\tilde{\theta}_B \pm z_{\alpha/2} \tilde{se}_B \quad \text{if } n \text{ is large - } \hat{\theta} \pm z_{\alpha/2} \hat{se}$$

- choosing a prior: subjective, conjugate, flat, convenience

- matching priors; Jeffreys' prior

$$O\left(\frac{1}{n}\right)$$

(not m.)

$$\pi_J(\theta) \propto |i(\theta)|^{1/2}$$

$$i(\theta) = E\left\{-\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right\}$$

- multiple parameters, marginal posterior

$$\pi(p_1, p_2 | x^n) \rightarrow \pi(p_1, p_2 | x^n)$$

$$\pi(\underline{\theta} | \underline{x}^n) : \pi_m(\psi | \underline{x}^n) = \int \pi(\underline{\theta} | \underline{x}^n) d\theta \quad \text{if } \psi(\underline{\theta}) = \psi$$

- Bayesian and frequentist philosophy

AoS 11.1 (\*)

AoS §11.1

- empirical and epistemic probability

freq.

uncertainty of knowledge

1. Friday: Jeffreys-Lindley paradox (HW 6 (c)); DF re  $\chi^2$ ; Pf. of B-H (?)

2. Bayesian inference overview ✓

→ 3. Two weird examples

AoS 11.8, 11.10

→ 4. Empirical Bayes }  
5. Hierarchical Bayes }

- Mar 15 5.15 – 6.15 pm EDT  
[Data Science Speaker Series](#)  
 Jesse Cresswell  
 Machine Learning Scientist, Layer 6 AI at TD  
 “Evaluating Model Performance on  
 Highly Imbalanced Datasets”



- all information about  $\theta$  contained in posterior density  $\pi(\theta \mid x^n) = f(x^n \mid \theta)\pi(\theta)/f_{X^n}(x^n)$

- all information about  $\theta$  contained in posterior density  $\pi(\theta \mid x^n) = f(x^n \mid \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior



# Overview

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$

# Overview

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$
- for choosing between models, **do need** marginal distribution of  $X^n$ , as in HW 6 (c)

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$
- for choosing between models, **do need** marginal distribution of  $X^n$ , as in HW 6 (c)
- Bayesian predictions of future values: posterior predictive

$$\pi(x_{new} | x^n) = \int f(x_{new} | x^n, \theta)\pi(\theta | x^n)d\theta,$$

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$
- for choosing between models, **do need** marginal distribution of  $X^n$ , as in HW 6 (c)
- Bayesian predictions of future values: posterior predictive

$$\pi(x_{new} | x^n) = \int f(x_{new} | x^n, \theta)\pi(\theta | x^n)d\theta,$$

- probability statements refer to uncertainty of knowledge

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$
- for choosing between models, **do need** marginal distribution of  $X^n$ , as in HW 6 (c)
- Bayesian predictions of future values: posterior predictive

$$\pi(x_{new} | x^n) = \int f(x_{new} | x^n, \theta)\pi(\theta | x^n)d\theta,$$

- probability statements refer to uncertainty of knowledge
- choosing priors can be difficult, and can have large impact in high-dimensional settings

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta)\pi(\theta)/f_{X^n}(x^n)$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$
- for choosing between models, **do need** marginal distribution of  $X^n$ , as in HW 6 (c)
- Bayesian predictions of future values: posterior predictive

$$\pi(x_{new} | x^n) = \int f(x_{new} | x^n, \theta)\pi(\theta | x^n)d\theta,$$

- probability statements refer to uncertainty of knowledge
- choosing priors can be difficult, and can have large impact in high-dimensional settings
- most applications of Bayesian inference involve **sampling** from the posterior density

# Overview

$$\int \text{num } d\theta$$

$$\pi(\theta) = f(\theta)$$

$$\pi(\theta | x) \propto L(\theta) \pi(\theta)$$

- all information about  $\theta$  contained in posterior density  $\pi(\theta | x^n) = f(x^n | \theta) \pi(\theta) / \underline{\underline{f_{X^n}(x^n)}}$
- inference about  $\psi(\theta)$  based on marginal posterior
- for comparing two (or more) points  $\theta$  in  $\pi(\theta | x^n)$ , don't need marginal distribution of  $X^n$
- for choosing between models, **do need** marginal distribution of  $X^n$ , as in HW 6 (c)
- Bayesian predictions of future values:

posterior predictive

$$p(\mu | X) = \int \pi(\mu | \underline{x}^n) \cdot d\mu$$

$$\pi(x_{\text{new}} | x^n) = \int \underbrace{f(x_{\text{new}} | x^n, \theta)}_{\approx} \underbrace{\pi(\theta | x^n)}_{\approx} d\theta,$$

$$\text{freq.: } f(x_{\text{new}} | x^n; \hat{\theta})$$

- probability statements refer to uncertainty of knowledge
- choosing priors can be difficult, and can have large impact in high-dimensional settings
- most applications of Bayesian inference involve **sampling** from the posterior density
- or approximating the posterior density

normal, Laplace

- excellent overview of Bayesian computational methods (in SM book)

- Laplace approximation of integrals

analytic approx<sup>n</sup> using  
Normal + higher order  
terms

- importance sampling

- Markov chain Monte Carlo sampling

- Gibbs sampling

- Metropolis-Hastings algorithm

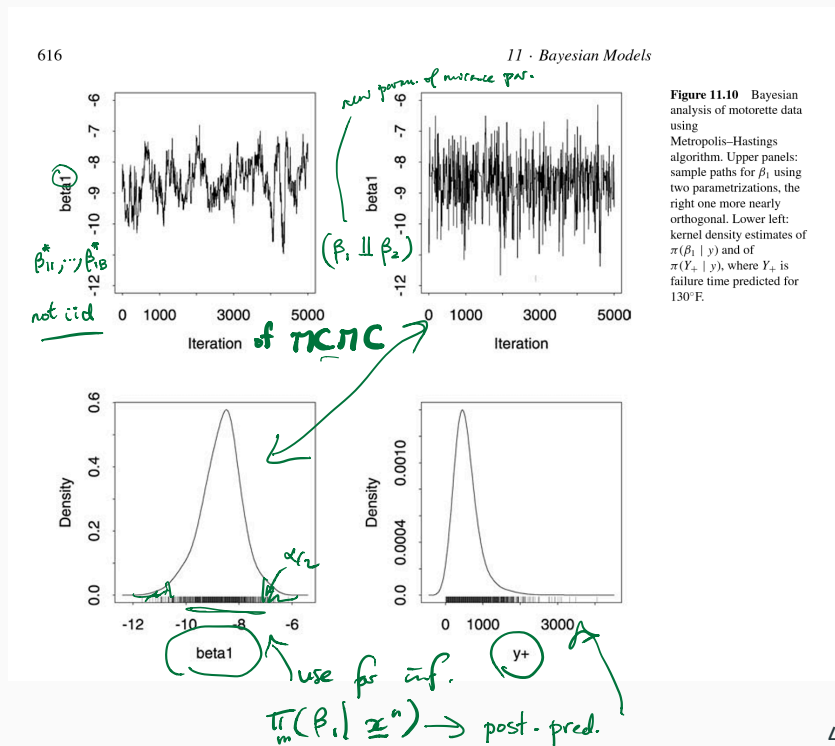
(improvement on)

$$\theta_1^*, \theta_2^*, \dots, \theta_B^* \sim \pi(\theta | \underline{x}^n)$$

↓ kernel  
hist, or cdf, or  $\hat{\pi}(\theta | \underline{x})$



- excellent overview of Bayesian computational methods
- Laplace approximation of integrals
- importance sampling
- Markov chain Monte Carlo sampling
  - Gibbs sampling
  - Metropolis-Hastings algorithm



## Example 11.9 AoS

$(X_1, R_1, Y_1), \dots, (X_n, R_n, Y_n)$  i.i.d.;

parameter  $\theta = (\theta_1, \dots, \theta_B)$ ,  $B$  very large

parameter  $\theta = (\theta_1, \dots, \theta_B)$   $B$  v. large  $\gg n$   $0 \leq \theta_j \leq 1, \forall j$   
 known v.  $\xi = (\xi_1, \dots, \xi_B)$   $0 \leq \delta < \xi_j < 1 - \delta \leq 1$

Sampling:  $X_i \sim U\{1, \dots, B\}$  rand. integer  
 $R_i \sim \text{Ber}(\xi_{X_i}) \left\{ \begin{array}{l} 1 \text{ w.p. } \xi_{X_i} \\ 0 \text{ w.p. } 1 - \xi_{X_i} \end{array} \right.$

if  $R_i = 1$ , obs.  $Y_i \sim \text{Ber}(\theta_{X_i}) \leftarrow$

if  $R_i = 0$  don't obs  $Y_i$

missing data

$\rightarrow$  par. of interest =  $P_n(Y_i = 1) = \psi = \frac{\sum_{j=1}^B \theta_j}{B}$  }  $\theta$

## Example 11.9 AoS

$$L\{\theta | (\underline{x}, \underline{y}, \underline{z})\} = \prod_{i=1}^n f(x_i) \underbrace{f(z_i | x_i)}_{\substack{\uparrow \\ \text{Bernoulli}}} f(y_i | x_i, z_i) \\ = \prod_{i=1}^n \frac{1}{B} \cdot \sum_{x_i}^{n_i} (1 - \sum_{x_i})^{1-n_i} \underbrace{\theta_{x_i}^{y_i z_i}}_{\substack{\uparrow \\ \text{Bernoulli}}} \underbrace{(1 - \theta_{x_i})^{(1-y_i)z_i}}_{\substack{\uparrow \\ \text{Bernoulli}}}$$

$$\propto \prod \theta_{x_i}^{y_i z_i} (1 - \theta_{x_i})^{(1-y_i)z_i}$$

$$l(\theta | (\underline{x}, \underline{y}, \underline{z})) = \sum_{i=1}^n \{ z_i y_i \log \theta_{x_i} + (1 - y_i) z_i \log (1 - \theta_{x_i}) \}$$

$$\text{for most } \theta_j \quad \log l(\theta | (\underline{x}, \underline{y})) = \sum_{j=1}^B n_j \log \theta_j + \sum_{j=1}^B m_j \log (1 - \theta_j)$$

$$m_j = \#\{y_i = 0, z_i = 1, x_i = j\} \quad n_j = \#\{y_i = 1, z_i = 1, x_i = j\}$$

To use this for B. inf.

$$\pi(\underline{\theta} | \underline{x}, \underline{y}, \underline{z}) \propto \underbrace{L(\underline{\theta} | (\underline{x}, \underline{y}, \underline{z}))}_{\text{many are 1}} \pi(\underline{\theta})$$

on  $\mathbb{R}^B$

$$\prod_{j \in \text{small set}} L(\theta_j) \pi(\theta_j) \propto \prod_{\text{most } j \in \{1, \dots, B\}} 1 \cdot \pi(\theta_j)$$

$$\psi = \frac{1}{B} \sum_{j=1}^B \theta_j$$

$$P(\psi | (\underline{x}, \underline{y}, \underline{z})) = \sum_{\theta_j} P(\theta_j | \underline{x}, \underline{y}, \underline{z})$$

$$\approx \pi(\psi)$$

only

Bayesian inf.  
fails

$$\hat{\psi} = \frac{1}{n} \sum_{i=1}^n \frac{R_i Y_i}{\sum x_i}$$

weighting by  $p_i(R_i=1)$

$$E(\hat{\psi} | \psi) = \psi \quad \text{in this weird model}$$

perfectly good estimate of  $\psi$

$$\text{var}(\hat{\psi} | \psi) \leq \frac{1}{n \delta^2}$$

"if the likelihood is weird, then Bayesian methods fail"  
if your model parametrization could be in trouble

model  $\left\{ x_i \mid \theta_i \sim N(\underline{\theta}_i, \underline{\sigma}_i^2) \right\} \quad i=1, \dots, n \text{ ind't}$

prior  $\theta_i \mid \mu \sim N(\mu, \sigma^2) \quad \text{iid} \quad i=1, \dots, n$

hyperprior  $\underline{\mu} \sim N(\mu_0, \tau^2) \quad \mu_0, \tau^2 \text{ are } \underline{\text{known}}$

$$\begin{aligned} \pi(\underline{\theta}, \underline{\mu} \mid \underline{x}) &\propto f(\underline{x} \mid \underline{\theta}) f(\underline{\theta} \mid \underline{\mu}) \pi(\underline{\mu}) \\ &= \prod_{i=1}^n \downarrow \prod_{i=1}^n \downarrow \prod_{i=1}^n \downarrow \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{1}{2\tau^2}(\mu - \mu_0)^2} \\ &\downarrow \\ &\text{Multivariate } n \text{ on } \mathbb{R}^{n+1} \end{aligned}$$

$$\tilde{\mu}_{\cdot B} = E(\mu | \underline{x}) = \frac{\mu_0 / \tau^2 + \sum x_i / (\sigma^2 + v_i)}{1/\tau^2 + \sum 1/(\sigma^2 + v_i)}$$

Bayes est. of  $\mu$

$$\text{var}(\mu | \underline{x}) = \frac{1}{1/\tau^2 + \sum 1/(\sigma^2 + v_i)}$$

hyper-prior  $\mu$

"pop" mean

$$\tilde{\theta}_{\cdot B} = E(\theta_i | \underline{x}) = E(\mu | \underline{x}) + \frac{\sigma^2}{\sigma^2 + v_i} (x_i - E(\mu | \underline{x}))$$

$$\text{Bayes est. of } \theta_i = x_i \underbrace{\frac{\sigma^2}{\sigma^2 + v_i}}_{\uparrow} + E(\mu | \underline{x}) \underbrace{\left(1 - \frac{\sigma^2}{\sigma^2 + v_i}\right)}_{\downarrow}$$

$$v_i = \text{var}(x_i | \theta_i) \quad (\sigma^2 = \text{var}(\theta_i | \mu))$$

Fig 11.11

$$\tilde{\theta}_{iB} = w_i x_i + (1 - w_i) \tilde{\mu}_B$$

$$\tilde{\mu}_B = \bar{w} \mu_0 + (1 - \bar{w}) \quad \text{see previous slide}$$

$$\text{var}(\tilde{\theta}_{iB}) = ? = \text{var}(\tilde{\theta}_{iB} | \tilde{x})$$

— not available in closed form —







