Mathematical Statistics II

STA2212H S LEC9101

Week 7

March 3 2021

Start recording!

Behind the numbers: what does it mean if a Covid vaccine has '90% efficacy'?

David Spiegelhalter and Anthony Masters

Confusion surrounds the vaccines' effectiveness. The leading Cambridge professor clarifies the data behind the trials

People rest in Salisbury Cathedral, England, after receiving the Pfizer/BioNTech vaccine. Photograph: Neil Hall/EPA

Vaccine efficacy

Guardian, Jan 24

Link to Guardian

Pfizer-BioNTech vaccine trial:

vaccine: 22000 subjects, 8 cases

placebo: 22000 subjects, 162 cases

$8/162 = 5\% \Longrightarrow 95\%$ efficacy *

data released November 18 2020 link

published December 31 2020 in NEJM link

Behind the numbers: what does it mean if a Covid vaccine has '90% efficacy'?

David Spiegelhalter and Anthony Masters

Confusion surrounds the vaccines' effectiveness. The leading Cambridge professor clarifies the data behind the trials

People rest in Salisbury Cathedral, England, after receiving the Pfizer/BioNTech vaccine. Photograph: Neil Hall/EPA

* "there is some uncertainty around these est."

Pfizer publication

NEJM

Editor's Note: This article was published on December 10, 2020, at NEJM.org.

ORIGINAL ARTICLE

Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine

Fernando P. Polack, M.D., Stephen J. Thomas, M.D., Nicholas Kitchin, M.D., Judith Absalon, M.D., Alejandra Gurtman, M.D., Stephen Lockhart, D.M., John L. Perez, M.D., Gonzalo Pérez Marc, M.D., Edson D. Moreira, M.D., Cristiano Zerbini, M.D., Ruth Bailey, B.Sc., Kena A. Swanson, Ph.D., et al., for the C4591001 Clinical Trial Group*

Article Figures/Media Metrics

13 References 263 Citing Articles Letters

December 31, 2020
N Engl J Med 2020; 383:2603-2615
DOI: 10.1056/NEJMoa2034577
Chinese Translation 中文翻译

Results: A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6).

? Bayesian proc. ?

Pfizer publication

NEJM

Table 2. Vaccine Efficacy against Covid-19 at Least 7 days after the Second Dose.*

Efficacy End Point	BNT162b2		Placebo		Vaccine Efficacy, % (95% Credible Interval)‡	Posterior Probability (Vaccine Efficacy >30%)§
	No. of Cases	Surveillance Time (n)†	No. of Cases	Surveillance Time (n)†		
	(N=18,198)		(N=18,325)			
Covid-19 occurrence at least 7 days after the second dose in participants without evidence of infection	8	2.214 (17,411)	162	2.222 (17,511)	95.0 (90.3–97.6)	>0.9999
	(N=19,965)		(N=20,172)			
Covid-19 occurrence at least 7 days after the second dose in participants with and those without evidence of infection	9	2.332 (18,559)	169	2.345 (18,708)	94.6 (89.9–97.3)	>0.9999

* The total population without baseline infection was 36,523; total population including those with and those without prior evidence of infection was 40,137.

† The surveillance time is the total time in 1000 person-years for the given end point across all participants within each group at risk for the end point. The time period for Covid-19 case accrual is from 7 days after the second dose to the end of the surveillance period.

‡ The credible interval for vaccine efficacy was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

§ Posterior probability was calculated with the use of a beta-binomial model with prior beta (0.700102, 1) adjusted for the surveillance time.

$$\left\{\begin{array}{llll} 18{,}000 & \text{vaccine group} & 8 & \text{cases} \\ 18{,}000 & \text{placebo group} & 162 & \text{cases} \end{array}\right\}$$

? $\frac{8}{162} = 0.05$?

$X_1 \sim \text{Poisson}(\lambda\psi)$ vacc.

$X_2 \sim \text{Poisson}(\lambda)$ plac.

$$\Pr(\underset{\text{vaccine}}{X_1} = x_1 \mid X_1 + X_2 = s) = \binom{s}{x_1}\left(\frac{\psi}{1+\psi}\right)^{x_1}\left(\frac{1}{1+\psi}\right)^{s-x_1}$$

$$\widehat{\left(\frac{\psi}{1+\psi}\right)} = \frac{x_1}{s} = \frac{8}{170}$$

$x_1 = 0, 1, \ldots, s = x_1 + x_2 = 170$

$s = 8 + 162$

$\Rightarrow \hat{\psi} = 8/162$

1. Bayesian: Beta(a,b) prior on bin. prob $a = 0.7$, $b = 1$

posterior $\text{Be}(a + \underset{8}{x}, \underset{170}{s} + b + 1)$

$\pi(p \mid x_1, s) = \text{Beta}(8, 172)$ ⟵ for credible interval

2. Frequentist sol^{n} – "Clopper–Pearson" intervals using exact Bin.

$$\left.\begin{array}{l} \Pr\{\text{Bin}(170,p) \geq x^0\} = \alpha/2 \\ \Pr\{\text{Bin}(170,p) \leq x^0\} = \alpha/2 \end{array}\right] \text{find } p \text{ to satisfy this}$$

$x^0 = 8$

Main paper refers to Bayesian credible intervals

Supplement " " frequentist " "

effectively same $\sim (90\%, 97\%)$

- adjustment for person-years of exposure (no details)

Recap

STA 2212S: Mathematical Statistics II
Syllabus Spring 2021

Updated Mar 3

Week	Date	Methods	References
1	Jan 13/15	Review of parametric inference	AoS Ch 9
2	Jan 20/22	~~Significance testing~~Hypothesis testing	AoS Ch 10.1,2,~~6~~,7; SM Ch 7.3.2, 4
3	Jan 27/29	Significance testing	AoS Ch 10.2, 6; SM Ch 7.3.1, Ch 4
4	Feb 3/5	Goodness of fit testing, Intro to multiple testing	AoS Ch 10.3,4,5,8; SM p.327-8 (hard)
5	Feb 10/12	Multiple testing and FDR	AoS Ch 10.7, EH Ch 15.1,2
	Feb 17/19	Break	
6	Feb 24/26	Bayesian Inference	AoS Ch 11.1-4; SM Ch 11.1,2; EH Ch 3, 13
7	Mar 3/5	Bayesian Inference	AoS Ch 11.5-9; SM Ch 11.4
8	Mar 10/12	Empirical Bayes	EH Ch 6, SM Ch 11.5
9	Mar 17/19	Statistical decision theory	AoS Ch 12, SM Ch 11.5.2
10	Mar 24/26	Multivariate Models	AoS Ch 14; SM Ch 6.3
11	Mar 31	Causal Inference and Graphical Models	AoS Ch 16, 17
12	Apr 7	Recap	

Recap 2

- Bayes theorem – conditional probability
- twin boys, diagnostic tests
- ingredients for inference: prior, posterior
- Bernoulli, normal correlation coefficient
- point estimates, equi-tailed posterior intervals, HPD intervals
- transformation of parameters
- approximate normality of posterior

$$(\theta \mid x^n) \;\dot\sim\; N(\tilde\theta, \tilde{j}^{-1}(\tilde\theta))$$

$$\tilde{j} = -\tilde\ell''(\tilde\theta)$$

$\hat\rho$ = $\pi(\rho \mid r)$

discrete

in mv normal

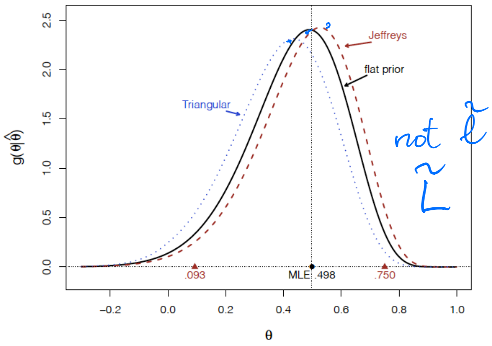


Figure 3.2 Student scores data; posterior density of correlation θ for three possible priors.

not $\tilde\ell$

$\tilde{L}$

$$\tilde\ell(\theta) = \ell(\theta) + \log \pi(\theta)$$

Today

Start Recording

1. Bayesian inference Part II

   Multi-par Bayes; noninformative priors; subjective priors; philosophy; Laplace approx; ; AoS 11.6-9; SM 11.1.3;11.2;11.311.4

2. ~~Friday – I owe you: proof of BH; careful χ^2; careful LRT; notes from Feb 26~~ → Mar 12

   OH ✓ 11-12

- Mar 5 12.00 pm EST Olufunmilayo I. Olopade "What African Genomes Tell Us About the Origins of Breast Cancer" Stage ISSS

STAGE ISSS: Olufunmilayo I. Olopade

- Mar 5 9.00 am EST ME

Gronsbell

Prof Alexander

Sun
Wang
Strug

} genetics

Park — psychology (neuro?)

Kong theory + genetics

DSS Statistics Seminar
March 5, 2021, 15:00
https://uniroma1.zoom.us/j/86881977368?pwd=SWRFcVFjMDZTa0lXZk05TE1zNm5adz09
Passcode: 432940

Replicability and Reproducibility: the interplay between statistical science and data science

N. Reid
University of Toronto

The current pandemic has brought into sharp relief the essential role of data in nearly all aspects of science, government and public health. But data is useless without explanation and interpretation, and statistical science has a long history and rich traditions for providing explanation and interpretation. In this talk I describe how data science and statistical science together can provide a robust framework for extracting insights from data reliably, and thus contribute to both replicability and reproducibility. This is illustrated with a selection of examples from recent news articles, along with some discussion on the role of the theory of inference in this framework.

Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(x^n \mid \theta)$, $\quad \theta \sim \pi(\theta)$, $\quad \pi(\theta \mid x^n) = \dfrac{f(x^n \mid \theta)\pi(\theta)}{f(x^n)}$ $\quad x^n = (x_1, \ldots, x_n)$

$f(\underset{\sim}{x} \mid \theta)$

- $\pi(\theta \mid x^n) \approx N\{\hat{\theta}, j^{-1}(\hat{\theta})\}$; $\quad \pi(\theta \mid x^n) \approx N\{\tilde{\theta}, \tilde{j}(\tilde{\theta})\}$ $\quad \tilde{\theta}$ = post. mode

mle

- careful statement $\quad$ Berger, 1985; Ch.4
  - For any $a, b \in \mathbb{R}, a < b$ $\quad \theta \in \mathbb{R}$
  - let $a_n = \hat{\theta}_n + aj^{-1/2}(\hat{\theta}_n)$, $b_n = \hat{\theta}_n + bj^{-1/2}(\hat{\theta}_n)$ $\quad a_n = a_n(x^n)$
  - $\hat{\theta}_n$ is the solution of $\ell'(\theta; x^n) = 0$, assumed unique, and $j(\theta) = -\ell''(\theta; x^n)$

Then

$$\boxed{\int_{a_n}^{b_n} \pi(\theta \mid x^n)\, d\theta \xrightarrow[n\to\infty]{} \Phi(b) - \Phi(a),} \quad n \to \infty.$$

need $\pi(\theta) > 0$, $\pi'(\theta)$ continuous

$\theta \,\dot{\sim}\, N\{\hat{\theta}_n, j^{-1}(\hat{\theta}_n)\}$

$\dot{\sim}\, N\{\tilde{\theta}_n, \tilde{j}^{-1}(\tilde{\theta}_n)\}$

Approximate normality of posterior

- $X_1, \ldots, X_n \sim f(x^n \mid \theta), \qquad \theta \sim \pi(\theta), \qquad \pi(\theta \mid x^n) = \dfrac{f(x^n \mid \theta)}{f(x^n)}$

$x^n = (x_1, \ldots, x_n)$

- $\pi(\theta \mid x^n) \approx N\{\hat\theta, j^{-1}(\hat\theta)\}; \qquad \pi(\theta \mid x^n) \approx N\{\tilde\theta, \tilde\jmath(\tilde\theta)\}$

θ random x fixed

to the order of these approx ($\sqrt{n}$)

$\tilde\theta \doteq \hat\theta$

$\tilde\theta = \hat\theta + O(\frac{1}{n})$

- approximate posterior probability intervals

$\theta \in \tilde\theta \pm z_{\alpha/2} \cdot \tilde{se}$

$\hat\theta \pm z_{\alpha/2} \, \hat{se}$

- exact posterior probability intervals (θ_L, θ_U)

$$\int_{\theta_L}^{\theta_U} \pi(\theta \mid x^n)\, d\theta = 1-\alpha$$

$\tilde\theta \approx \hat\theta$

same as Wald intervals (check)

- subjective "I" think $\theta \approx$ not > 100, not < 0, prob bet 50 & 80"
  need a problem.
  personal decision making.
- conjugate ← make calculations easy
- "flat" "ignorance" Bayes : Laplace 1763 "insufficient reason"
  $\pi(p) = 1, \; 0 \le p \le 1$
  theory of Bayesian inference based on subjective priors has compelling philosophical / foundational support
- "matching" ← give B ≐ F → (noninformative priors)
- convenience
- empirical

? ∃ ?
yes if $\theta \in \mathbb{R}$; no if $\underline{\theta} \in \mathbb{R}^k$

(penalized complexity)

Choosing a prior

✓ • subjective

✓ • conjugate

✓ • "flat"

✓ • "matching"

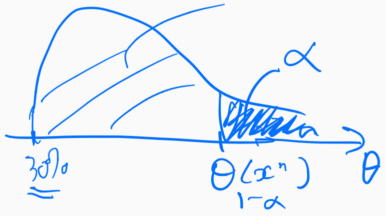
• convenience

data twins

• empirical

$$\Pr\{\theta \geq \underbrace{\theta_{1-\alpha}(x^n)} \mid x^n\} = \alpha$$

$\pi(\theta \mid x^n)$



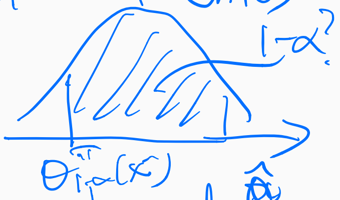
$$\Pr\{\theta^{\pi}_{1-\alpha}(X^n) \leq \theta_0 \mid \theta_0\} \stackrel{?}{=} 1-\alpha$$

lower confidence bound

$X^n \sim f(x^n \mid \theta_0)$

is also a freq. bound $\hat{\theta}$

$f(\hat{\theta}_n \mid \theta)$



Can we find prior to matching post / conf bounds

yes:

Noninformative prior: Jeffreys'

matching: if $X_i \sim f(x_i - \theta)$ θ loc^n par.

$\pi(\theta) \propto 1$ is matching

if $X_i \sim \frac{1}{\tau} f\left(\frac{x_i}{\tau}\right)$ τ scale par

$\pi(\tau) \propto 1/\tau$ matching

extend a little bit to model with a prop str.

$\exists\ \pi(\theta)$ s.t.

~~mat~~ post prob

$\doteq$ conf prob $+ \underline{O\left(\frac{1}{n}\right)}$

yes $\underline{\underline{\pi(\theta)}} \propto \{i(\theta)\}^{1/2}$

Matching prior: scalar parameter

Jeffreys' prior

$X_1 \sim f(x|\theta)$

$$i(\theta) = E\left\{\frac{\partial \log f(x;\theta)}{\partial\theta}\right\}^2$$

$$= -E\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]$$

exp'd Fisher info. in 1 obs $\hat{=}$

If $\theta \in \mathbb{R}^k$, suggested $\pi(\underline{\theta}) \propto |i(\underline{\theta})|^{1/2}$

k×k matrix

Matching prior: scalar parameter

$\pi(\theta) \propto \{i(\theta)\}^{1/2}$ is invariant to 1-1 transf. of θ

parameter of interest $= \tau = g(\theta)$

$$\{i(\tau)\}^{1/2} = \{i(\theta)\}^{1/2}$$

$$\{i(\tau)\}^{1/2} = E\left\{-\frac{\partial^2 \ell(\tau; X)}{\partial \tau^2}\right\} \qquad \{i(\theta)\}^{1/2} \leftarrow E\left\{-\frac{\partial^2 \ell(\theta; X)}{\partial \theta^2}\right\}$$

$\Leftarrow (g')^2$ from chain rule

$X \sim Bin(n, p)$ $\qquad \binom{n}{x} p^{x}(1-p)^{n-x}$

$\ell(p) = x\log p + (n-x)\log(1-p)$

$\ell'(p) = \frac{x}{p} - \left(\frac{n-x}{1-p}\right)$ $\qquad (1-p)^{-1} \quad -(1-p)^{-2}$

$\ell''(p) = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}$

$E\{\ell''(p)\} = \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2}$

$= n\cdot\left(\frac{1}{p(1-p)}\right) \qquad = n\,i(p)$

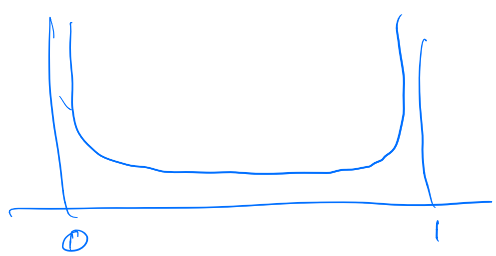
$\pi_J(p) \propto \{i^{1/2}(p)\} = p^{-1/2}(1-p)^{-1/2} \Longleftarrow \|$

$Be(\frac{1}{2},\frac{1}{2}) \qquad 0 \le p \le 1$

$\psi = \log\left(\frac{p}{1-p}\right)$ $\qquad$ H&S "almost unif."

ntbc $\pi(\psi)$ same

$\pi_J(p)$



$\mathcal{N}(\mu, \sigma^2)$ $\pi(\mu) \propto 1$ $\pi(\sigma^2) \propto \frac{1}{\sigma}$

582 *11 · Bayesian Models*

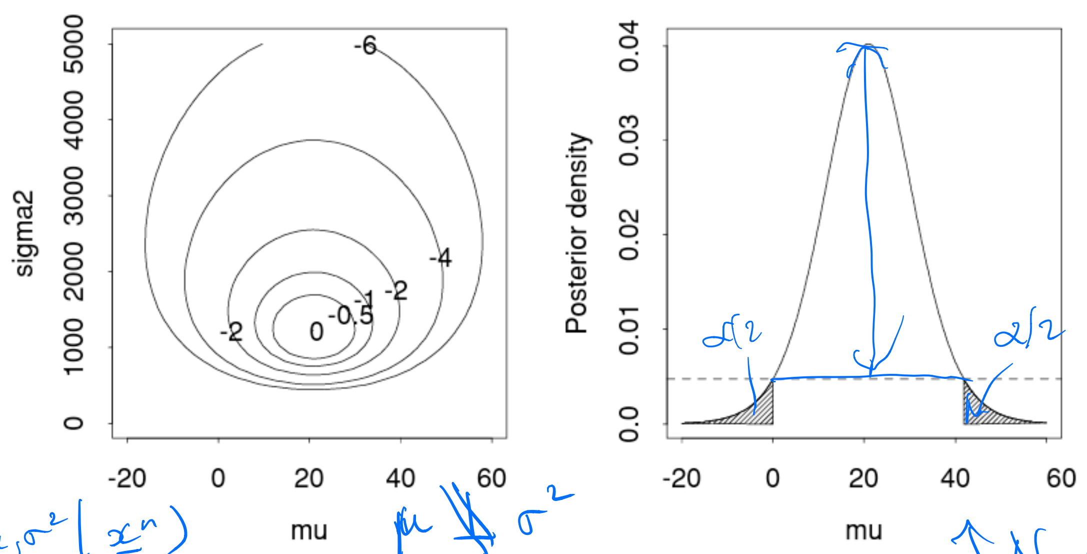


Figure 11.2 Posterior densities of (μ, σ^2) of normal model for maize data. Left: contours of the normalized log joint posterior density. Right: marginal posterior density for μ, showing 95% HPD credible set, which is the set of values of μ whose values of the posterior density $\pi(\mu \mid y)$ lie above the dashed line. The shaded region has area 0.05.

$\pi(\mu, \sigma^2 \mid x^n)$ $\mu \not\perp \sigma^2$ in posterior

$\uparrow \mathcal{N}$ posterior

$X^n \sim f(x^n \mid \underset{\sim}{\theta}) \quad \underset{\sim}{\theta} \in \mathbb{R}^k \quad \theta = (\psi, \underset{\sim}{\lambda})$

scalar par. of interest $k-1$ vector

example $\quad \underset{n\times 1}{X} = Z\beta + \varepsilon$

$\varepsilon \sim N(0, \sigma^2 I_n)$

$\psi = \beta_5$ or $\psi = \sigma^2$, or ... β_3 / β_2

[extension]

$\pi(\underset{\sim}{\theta} \mid \underset{\sim}{x}) \propto \pi(\underset{\sim}{\theta}) f(\underset{\sim}{x}^n \mid \underset{\sim}{\theta})$ density on $\mathbb{R}^k$

$$\pi_m(\psi \mid \underline{x}) \propto \int \pi(\psi, \lambda_1, \ldots, \lambda_{k-1} \mid \underline{x}^n)\, d\lambda_1 \cdots d\lambda_{k-1}$$

$$\underbrace{\pi_n(\psi \mid \underline{x})} = \frac{\int \pi(\underline{\theta})\, L(\underline{\theta} \mid x^n)\, d\lambda_1 \cdots d\lambda_{k-1}}{\int \pi(\underline{\theta})\, L(\underline{\theta} \mid x^n)\, d\psi\, d\lambda_1 \cdots d\lambda_{k-1}}$$

need to integrate over $\mathbb{R}^{k-1}$ top

$\mathbb{R}^{k}$ bottom

$\boxed{\text{MCMC sampling permits calc}^{\underline{n}}}$

1) choose prior 2) compute posterior

Multiparameter models

$\theta = (\psi, \underset{\sim}{\lambda})$

1. $\pi_m(\psi \mid \underset{\sim}{x}) = \int \pi(\underset{\sim}{\theta} \mid \underset{\sim}{x})\, d\lambda_1, \cdots d\lambda_{k-1}$

2. $\pi(\mu_1, \cdots, \mu_k \mid \underset{\sim}{x}^n)$ $\qquad = (\sum \mu_i^2) = \psi$

$$\int_{\{\underset{\sim}{\mu} \in \mathbb{R}^n : \sum \mu_i^2 = \psi\}} \pi(\underset{\sim}{\mu} \mid \underset{\sim}{x}^n)\, d\underset{\sim}{\mu} = \pi_m(\psi \mid \underset{\sim}{x}^n)$$

3. In practice, either simulate (usually MCMC) or integrate

4. Matching? can we find $\pi(\underset{\sim}{\theta})$ s.t.

post. prob bound for $\psi \overset{?}{\doteq}$ lower conf. bd. for ψ

Tibshirani priors

$\psi = \theta_1 \quad \underset{\sim}{\lambda} = \theta_2 \cdots \theta_k$

$$\pi(\underset{\sim}{\theta}) \propto i_{\psi\psi}^{1/2}(\underset{\sim}{\theta}) \cdot g(\underset{\sim}{\lambda})$$

is matching to $O(n^{-1})$ for $\boxed{\psi}$

$$i(\underset{\sim}{\theta}) = E_\theta\left[-\frac{\partial^2 \ell(\theta)}{\partial \underset{\sim}{\theta}\, \partial \underset{\sim}{\theta}^T}\right]_{k \times k}$$

$\left[\{i^{\psi\psi}(\theta)\}^{-1/2}\right] = \begin{bmatrix} i_{\psi\psi}(\underset{\sim}{\theta}) & \cancel{i_{\psi\lambda}} \\ & i_{\lambda\lambda} \end{bmatrix}$

no single $\pi(\underset{\sim}{\theta})$ that can match each comp. in vector

Also, a lot of convenience priors, and 'matching' priors, and Jeffreys' prior are improper $\int \pi(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} = \infty$

$\pi(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$ improper prior

if $\int \pi(\underset{\sim}{\theta} \mid \underset{\sim}{x}^n)\, d\underset{\sim}{\theta} = 1$ we're okay

$= \infty$ we are not okay

? MCMC doesn't seem to be converging ?

- In multi parameter problems, "flat" priors for $\underset{\sim}{\theta}$ can be quite informative for functions of $\underset{\sim}{\theta}$.

$\pi(\mu_i) \propto 1$ induced prior for $(\sum \mu_i^2)$ is very informative

Bayesian philosophy

- empirical probability "physical prop. of a randomization event"

F 1 – 3 in AoS

to justify (matching) (any th!)

- epistemic probability "uncertainty of knowledge"

B 1– 3 in AoS

→ AoS.pdf 1

? $\pi(\theta) \propto 1$? Prior Pr$(\theta > 10^{10})$ = Prior Pr $(\theta > 0)$

$\pi(\theta|x^n) \propto \pi(\theta) f(x^n|\theta)$

changes our prior

prob on Θ

fixed

Bayesian philosophy

- Example 11.8

- Example 11.10

"objective priors are the perpetual motion machine of Bayesian inference"

⟶ AoS.pdf 2