Distributions for Parameters

Nancy Reid
University of Toronto

Monash University

Feb 21 2020

Reproducibility and Statistical theory

Those pesky *p*-values

medium.com

David Spiegelhalter
@d_spiegel

This paper motivates the call for the end of significance. A 25% mortality reduction, but because P=0.06 (two-sided), they declare it 'did not reduce' mortality. Appalling.
jamanetwork.com/journals/jama/…

Research

JAMA | Original Investigation | CARING FOR THE CRITICALLY ILL PATIENT

Effect of a Resuscitation Strategy Targeting Peripheral Perfusion Status vs Serum Lactate Levels on 28-Day Mortality Among Patients With Septic Shock

The ANDROMEDA-SHOCK Randomized Clinical Trial

Glenn Hernández, MD, PhD; Gustavo A. Ospina-Tascón, MD, PhD; Lucas Petri Damiani, MSc; Elisa Estenssoro, MD;
Arnaldo Dubin, MD, PhD; Javier Hurtado, MD; Gilberto Friedman, MD, PhD; Ricardo Castro, MD, MPH;
Leyla Alegría, RN, MSc; Jean-Louis Teboul, MD, PhD; Maurizio Cecconi, MD, FFICM; Giorgio Ferri, MD;
Manuel Jibaja, MD; Ronald Pairumani, MD; Paula Fernández, MD; Diego Barahona, MD;
Vladimir Granda-Luna, MD, PhD; Alexandre Biasi Cavalcanti, MD, PhD; Jan Bakker, MD, PhD; for the
ANDROMEDA-SHOCK Investigators and the Latin America Intensive Care Network (LIVEN)

Andromeda Trial

- comparing two treatments for septic shock
- randomized clinical trial

- estimated hazard ratio 0.75 [0.55, 1.02] — after adjusting for confounders
- 2-sided p-value 0.06 — 34.9% vs 43.4% unadjusted

- Discussion: " a peripheral perfusion-targeted resuscitation strategy did not result in a significantly lower 28-day mortality when compared with a lactate level-targeted strategy"
- Abstract: "Among patients with septic shock, a resuscitation strategy targeting normalization of capillary refill time, compared with a strategy targeting serum lactate levels, did not reduce all-cause 28-day mortality."

A recent timeline

- 2014: *Basic and Applied Social Psychology* published an editorial banning *p*-values

  actually “null hypothesis significance testing”

- “prior to publication, authors will need to remove all vestiges of the NHSTP … *p*-values, … , statements about ‘significant differences’ or lack thereof, and so on”

  “confidence intervals are also banned”

- 2014: *Nature* published a News Feature by R. Nuzzo: “*p*-values, the gold standard of statistical validity, are not as reliable as many scientists assume”

- 2016: American Statistical Association released a public statement on statistical significance and *p*-values

ASA News

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

732 North Washington Street, Alexandria, VA 22314 • (703) 684-1221 • Toll Free: (888) 231-3473 • *www.amstat.org* • *www.twitter.com/AmstatNews*

AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND *P*-VALUES

Provides Principles to Improve the Conduct and Interpretation of Quantitative Science

March 7, 2016

- 2017: Another *Nature* article $p < 0.005$
- Articles solicited for special issue of *American Statistician*

comment

Redefine statistical significance

We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, David Cesarini, Christopher D. Chambers, Merlise Clyde, Thomas D. Cook, Paul De Boeck, Zoltan Dienes, Anna Dreber, Kenny Easwaran, Charles Efferson, Ernst Fehr, Fiona Fidler, Andy P. Field, Malcolm Forster, Edward I. George, Richard Gonzalez, Steven Goodman, Edwin Green, Donald P. Green, Anthony Greenwald, Jarrod D. Hadfield, Larry V. Hedges, Leonhard Held, Teck Hua Ho, Herbert Hoijtink, Daniel J. Hruschka, Kosuke Imai, Guido Imbens, John P. A. Ioannidis, Minjeong Jeon, James Holland Jones, Michael Kirchler, David Laibson, John List, Roderick Little, Arthur Lupia, Edouard Machery, Scott E. Maxwell, Michael McCarthy, Don Moore, Stephen L. Morgan, Marcus Munafó, Shinichi Nakagawa, Brendan Nyhan, Timothy H. Parker, Luis Pericchi, Marco Perugini, Jeff Rouder, Judith Rousseau, Victoria Savalei, Felix D. Schönbrodt, Thomas Sellke, Betsy Sinclair, Dustin Tingley, Trisha Van Zandt, Simine Vazire, Duncan J. Watts, Christopher Winship, Robert L. Wolpert, Yu Xie, Cristobal Young, Jonathan Zinman and Valen E. Johnson

- 2019: *American Statistician* publishes special issue 43 articles; 400 pages
- Editorial introduction advises "abandon 'statistical significance' "

- *Nature* publishes a letter agreeing with this
- "we are not advocating a ban on P values, confidence intervals or other statistical measures – only that we should not treat them categorically
- "This includes dichotomization as statistically significant or not, as well as categorization based on other statistical measures such as Bayes factors."

COMMENT

EVOLUTION Cooperation and conflict from ants and chimps to us p.308

HISTORY To fight denial, study Galileo and Arendt p.309

CHEMISTRY Three more unsung women — of astatine discovery p.311

PUBLISHING As well as ORCID ID and English, list authors in their own script p.311

ILLUSTRATION BY DAVID PARKINS

Retire statistical significance

Valentin Amrhein, **Sander Greenland**, **Blake McShane** and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

When was the last time you heard a seminar speaker claim there was 'no difference' between two groups because the difference was 'statistically non-significant'?

If your experience matches ours, there's a good chance that this happened at the last talk you attended. We hope that at least someone in the audience was perplexed if, as frequently happens, a plot or table showed that there actually was a difference.

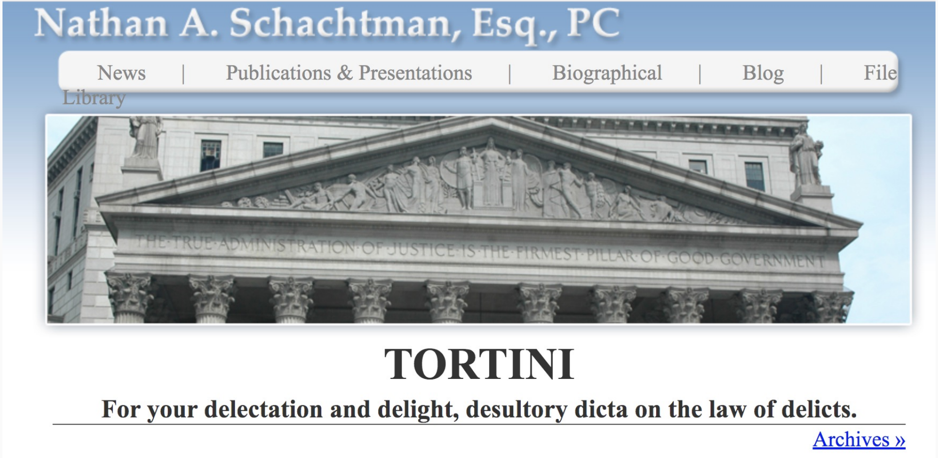
How do statistics so often lead scientists to deny differences that those not educated in statistics can plainly see? For several generations, researchers have been warned that a statistically non-significant result does not 'prove' the null hypothesis (the hypothesis that there is no difference between groups or no effect of a treatment on some measured outcome)[1]. Nor do statistically significant results 'prove' some other hypothesis. Such misconceptions have famously warped the literature with overstated claims and, less famously, led to claims of conflicts between studies where none exists.

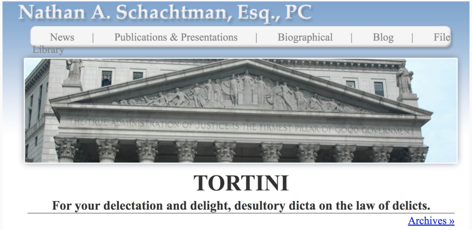
We have some proposals to keep scientists from falling prey to these misconceptions.

PERVASIVE PROBLEM

Let's be clear about what must stop: we should never conclude there is 'no difference' or 'no association' just because a *P* value is larger than a threshold such as 0.05

21 MARCH 2019 | VOL 567 | NATURE | 305


Nathan A. Schachtman, Esq., PC
News | Publications & Presentations | Biographical | Blog | File Library
THE·TRUE·ADMINISTRATION·OF·JUSTICE·IS·THE·FIRMEST·PILLAR·OF·GOOD·GOVERNMENT
TORTINI
For your delectation and delight, desultory dicta on the law of delicts.
Archives »



"Lawyers and judges pay close attention to standards, guidances, and consenus statements from respected and recognized professional organizations."

"Despite the fairly clear and careful statement of principles, legal actors did not take long to misrepresent the ASA principles." 2016

"distorted into strident assertions that statistical significance was unnecessary for scientific conclusions."

HARVARD DATA SCIENCE REVIEW

P-Values on Trial: Selective Reporting of (Best Practice Guides Against) Selective Reporting

by Deborah Mayo

outlines a 2018 Supreme Court case appealing a conviction for wire fraud, based on misleading investors

Harkonen v. United States 13-180

the fraud centered on *p*-hacking the results of a Phase III trial of a drug

marketed by Harkonen

in the appeal "his defenders argued that the ASA guide provides compelling new evidence that the scientific theory upon which petitioner's conviction was based [that of statistical significance testing] is demonstrably false"

What to do?

- report actual p-value, not "*", $p < 0.05$, etc. — to sensible number of decimal points
- supplement p-value with sample size, estimated power, etc.
- clarify 'exploratory' and 'confirmatory' p-values — Spiegelhalter 2017

- report effect sizes and estimated standard errors
- report confidence intervals

- pre-register trials, specifying primary and secondary outcomes
- pre-specify data analysis — NEJM

- provide a p-value function — significance function
- or some analogous distribution — Bayes posterior

Distributions for parameters

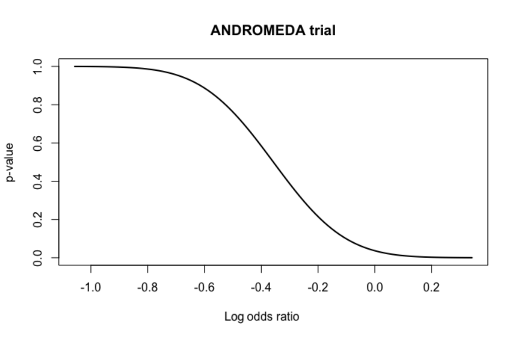
A p-value function

ANDROMEDA trial

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided p-value = 0.07

likelihood ratio test
no adjustment for covariates



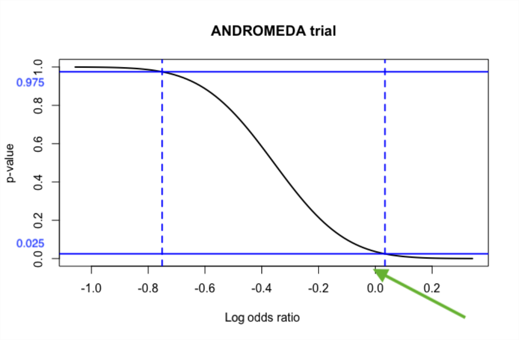
90% confidence interval: $[-0.688, -0.030]$
95% confidence interval: $[-0.751, 0.034]$
99% confidence interval: $[-0.825, 0.107]$

ANDROMEDA trial

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

2-sided *p*-value = 0.07

likelihood ratio test
no adjustment for covariates



90% confidence interval: [−0.688, −0.030]
95% confidence interval: [−0.751, 0.034]
99% confidence interval: [−0.825, 0.107]

Confidence distribution function

SOME PROBLEMS CONNECTED WITH STATISTICAL INFERENCE

By D. R. Cox

Birkbeck College, University of London[1]

1. Introduction. This paper is based on an invited address given to a joint meeting of the Institute of Mathematical Statistics and the Biometric Society at Princeton, N. J., 20th April, 1956. It consists of some general comments, few of them new, about statistical inference.

Since the address was given publications by Fisher [11], [12], [13], have produced a spirited discussion [7], [21], [24], [31] on the general nature of statistical methods. I have not attempted to revise the paper so as to comment point by point on the specific issues raised in this controversy, although I have, of course, checked that the literature of the controversy does not lead me to change the opinions expressed in the final form of the paper. Parts of the paper are controversial; these are not put forward in any dogmatic spirit.

2. Inferences and decisions. A statistical inference will be defined for the

- "… the method of confidence intervals, as usually formulated, gives only one interval at some preselected level of probability"
- "… in … simple cases … there seems no reason why we should not work with confidence distributions for the unknown parameter
- "These can either be defined directly, or … introduced in terms of the set of all confidence intervals"

Confidence Distribution

Efron 1993

The idea of obtaining Bayesian results from confidence intervals goes back at least to Fisher's work on fiducial inference in the 1930's. Suppose that a data set x is observed from a parametric family of densities $g_\mu(x)$, depending on an unknown parameter vector μ, and that inferences are desired for $\theta = t(\mu)$, a real-valued function of μ. Let $\theta_x(\alpha)$ be the upper endpoint of an exact or approximate one-sided level-α confidence interval for θ. The standard intervals for example have

$$\boxed{\theta_x(\alpha) = \hat\theta + \hat\sigma z^{(\alpha)},} \qquad (1{\cdot}1)$$

where $\hat\theta$ is the maximum likelihood estimate of θ, $\hat\sigma$ is the Fisher information estimate of standard error for $\hat\theta$, and $z^{(\alpha)}$ is the α-quantile of a standard normal distribution, $z^{(\alpha)} = \Phi^{-1}(\alpha)$. We write the inverse function of $\theta_x(\alpha)$ as $\alpha_x(\theta)$, meaning the value of α

corresponding to upper endpoint θ for the confidence interval, and assume that $\alpha_x(\theta)$ is smoothly increasing in θ. For the standard intervals, $\alpha_x(\theta) = \Phi((\theta - \hat\theta)/\hat\sigma)$, where Φ is the standard normal cumulative distribution function.

The confidence distribution for θ is defined to be the distribution having density

$$\boxed{\pi_x^\dagger(\theta) = d\alpha_x(\theta)/d\theta.} \qquad (1{\cdot}2)$$

We shall call (1·2) the confidence density. This distribution assigns probability 0·05 to θ lying between the upper endpoints of the 0·90 and 0·95 confidence intervals, etc. Of

$\theta_y(\alpha)$ upper endpt of interval

$\alpha_y(\theta)$ inverse function

$\pi_y(\theta) = d\alpha_y(\theta)/d\theta$
confidence density

- "assigns probability 0.05 to θ between upper endpoints of 0.90 and 0.95 confidence intervals, ..."
- "Of course this is logically incorrect, but it has powerful intuitive appeal"
- "... no nuisance parameters [this] is exactly Fisher's fiducial distribution"

Seidenfeld 1992; Zabell 1992

528 *Dr Fisher, Inverse probability*

Inverse Probability. By R. A. FISHER, Sc.D., F.R.S., Gonville and Caius College; Statistical Dept., Rothamsted Experimental Station.

[*Received* 23 July, *read* 28 July 1930.]

$$\mathrm{d}f = -\frac{\partial}{\partial\theta}F(Y,\theta)\mathrm{d}\theta$$

fiducial probability density for θ, given statistic Y

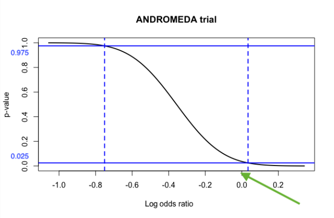
"It is not to be lightly supposed that men of the mental calibre of Laplace and Gauss … could fall into error on a question of prime theoretical importance, without an uncommonly good reason"

Distributions for parameters

- significance function $p(\theta) = \Pr(y \geq y^{o} \mid \theta)$
- confidence distribution $\alpha_y(\theta) = \theta_y^{-1}(\alpha)$
- fiducial probability $df = -(\partial F/\partial\theta)(Y;\theta)d\theta$
- structural probability a re-formulation of fiducial probability for transformation models

Fraser 1966
- belief functions

Dempster ’66; Schafer ’76



In spite of the naming, these are not ‘real’ probability distributions

Don’t obey the rules of probability calculus

Isn't it obvious?

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

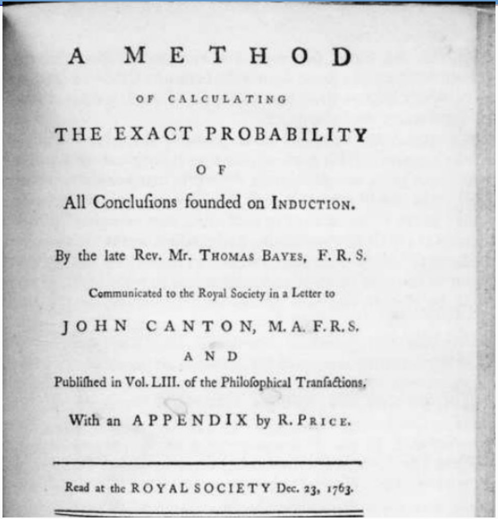
I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

$$\pi(\theta \mid y^o) = f(y^o; \theta)\pi(\theta)/m(y^o)$$

a 'real' probability distribution for θ

y^o is fixed

probability comes from $\pi(\theta)$


A METHOD

OF CALCULATING

THE EXACT PROBABILITY

OF

All Conclufions founded on INDUCTION.

By the late Rev. Mr. THOMAS BAYES, F. R. S.

Communicated to the Royal Society in a Letter to

JOHN CANTON, M. A. F. R. S.

AND

Publifhed in Vol. LIII. of the Philofophical Tranfactions.

With an APPENDIX by R. PRICE.

Read at the ROYAL SOCIETY Dec. 23, 1763.

$$\pi(\theta \mid y^o) = f(y^o; \theta)\pi(\theta)/m(y^o)$$

a 'real' probability distribution for θ

y^o is fixed

probability comes from $\pi(\theta)$

$$\Pr(\Theta \in A \mid y^o) = \int_A \pi(\theta \mid y^o) d\theta$$

Stigler 2013

Why do we want distributions on parameters?

- inference is intuitive
- combines easily with decision theory
- de-emphasizes point estimation and arbitrary cut-offs

- “it’s tempting to conclude that μ is more likely to be near the middle of this interval, and if outside, not very far outside”

  Cox 2006

- “assigns probability 0.05 to θ lying between the upper endpoints of the 0.90 and 0.95 confidence intervals, etc.”

  Efron 1993

- all inference statements become probability statements about unknowns   hmm...

Nature of Probability

Cox 2006; R & Cox 2015

- probability to describe physical haphazard variability — aleatory/empirical
  - probabilities represent features of the "real" world in somewhat idealized form
  - subject to empirical test and improvement
  - conclusions of statistical analysis expressed in terms of interpretable parameters
  - enhanced understanding of the data generating process
- probability to describe the uncertainty of knowledge — epistemic
  - measures rational, supposedly impersonal, degree of belief, given relevant information — Jeffreys
  - measures a particular person's degree of belief, subject typically to some constraints of self-consistency — Ramsey, de Finetti, Savage
- "In short, the [Bayesian] paradigm does not produce probabilities from no probabilities"

Fraser 2011

The First Workshop on BFF Inference and Statistical Foundations (BFF 2014)

November 10 – November 14, 2014

7th Bayes, Fiducial and Frequentis Statistics Conference

Methodological, Computational, and Ethical Principles for Data Science

May 6 - 8, 2020, The Fields Institute

Location: Fields Institute, Room 230

… BFF 1 – 6

- posterior distribution 1763
- fiducial probability 1930
- confidence distribution 1958
- structural probability 1964
- significance function
- belief functions 1967

- objective Bayes
- generalized fiducial inference Hannig; Taraldsen
- confidence distributions and confidence curves Hjort, Schweder, Xie
- approximate significance functions Brazzale et al 2007; Fraser & R 1993
- inferential models Martin & Liu

high-dimensional inference and model selection

What has changed?

computation

model complexity

model dimension

data

science

… what has changed

From geeky to cool: Statistics is Berkeley’s fastest-growing major

... BFF 1 – 6

- posterior distribution
- fiducial probability
- confidence distribution
- structural probability
- significance function
- belief functions

- objective Bayes
- generalized fiducial inference Hannig; Taraldsen
- confidence distributions and confidence curves Hjort, Schweder, Xie
- approximate significance functions Brazzale et al 2007; Fraser & R 1993
- inferential models Martin & Liu

high-dimensional inference and model selection

Objective Bayes

Objective Bayes

- there are many proposals for priors meant to be non-informative
- examples include reference, default, matching, vague, … priors
- a popular choice is Jeffreys' prior $\pi(\theta) \propto |i(\theta)|^{1/2}$ expected Fisher information

- what interpretation do we put on the posterior distribution? empirical? epistemic?

- we may avoid the need for a different version of probability by appeal to a notion of calibration Cox 2006, R & Cox 2015
- as with other measuring devices,
  within this scheme of repetition, probability is defined as a hypothetical frequency
- it is unacceptable if a procedure yielding high-probability regions in some non-frequency sense are poorly calibrated

- such procedures, used repeatedly, give misleading conclusions

Bayesian Analysis, V1(3) 2006

... objective Bayes

- there are many proposals for priors meant to be non-informative
- examples include reference, default, matching, vague, ... priors

not flat

- some versions may not be correctly calibrated

e.g. Jeffreys'

- requires checking in each example
- calibrated versions must be targetted on the parameter of interest

Fraser 2011

- only in very special cases can calibration be achieved for more than one parameter in the model, from the same prior
- the simplicity of a fully Bayesian approach to inference is lost

Gelman 2008; PPM LW

... objective Bayes

- the simplicity of a fully Bayesian approach to inference is lost

Gelman 2008

- for example

$$\pi(\psi \mid y) = \int_{\psi(\theta)=\psi} \pi(\theta \mid y) d\theta, \quad \text{for any } \psi : \Theta \searrow \Psi$$

lower dimension

- the prior can have unexpected influence on the posterior
- even if they are seemingly noninformative

objective Bayes fails

- Stein's example:

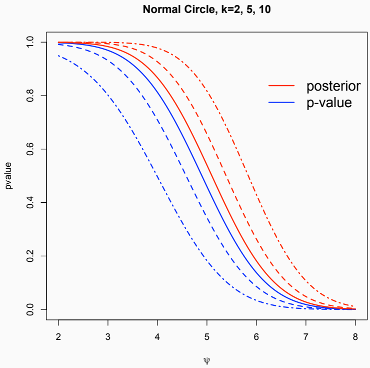
$$\begin{aligned} y_i &\sim N(\theta_i, 1/n), \quad i = 1, \ldots, k \\ \pi(\theta_i) &\propto 1 \\ \pi(\theta \mid y) &\propto N(y, I_k/n) \end{aligned}$$

Example

Stein, 1959

- $y_i \sim N(\theta_i, 1/n), \quad i = 1, \ldots, k; \quad \pi(\theta_i) \propto 1$
- posterior distribution of $a^{\mathrm{T}}\mu$ is well-calibrated
- marginal posterior distribution of $\psi = ||\mu||$ is not
- discrepancy is a function of $\dfrac{k-1}{\psi\sqrt{n}}$
- $p(\psi) = \Pr\{\chi^2_k(n\psi^2) \geq n||y||^2\}$

  $s(\psi) = \Pr\{\chi^2_k(n||y||^2) \geq n\psi^2\}$



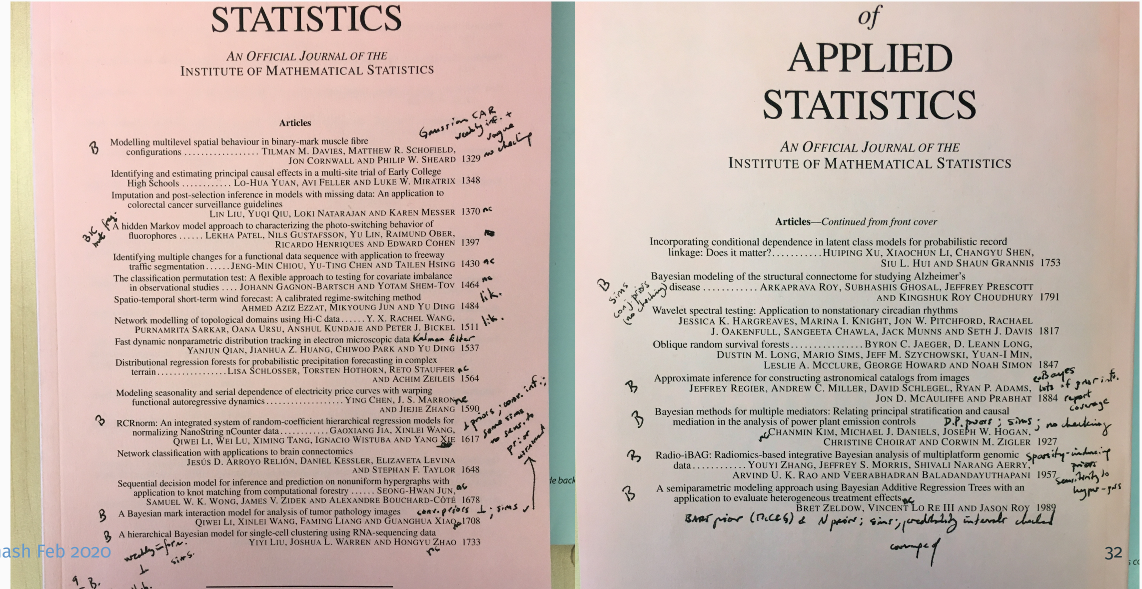
- calibrated posterior distributions must be targetted on the parameter of interest
- matching priors set out this requirement explicitly

  defined by calibration of posterior quantiles
- reference priors are also targetted

  although with a different goal than calibration
- vague priors, hierarchical priors, weakly informative priors, ... are not (usually) targetted on a particular parameter of interest
-
- “in short
- AoAS September 2018: 9/24 articles used Bayesian methods
- one checked the coverage of posterior intervals
- one used simulations to evaluate point estimates

Haphazard sample

AoAS Sept 2019



STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles

of

APPLIED STATISTICS

AN OFFICIAL JOURNAL OF THE
INSTITUTE OF MATHEMATICAL STATISTICS

Articles—*Continued from front cover*

4. Bayesian approach.

4.1. *Full probability model.* Based on the system of equations (2)–(5), the parameters of our data model include μ_a, σ_a^2, μ_b, σ_b^2, σ_d^2, σ_e^2, σ_{d-}^2, σ_{e-}^2, $\{\phi_i\}_{i=1}^{I-1}$, $\{\lambda_h^*\}_{h=1}^{H}$, $\sigma_{\kappa*}^2$, $\{\lambda_r\}_{r=1}^{R}$, σ_κ^2 and c, among which μ_a, μ_b, c, ϕ_i's, λ_r's and λ_h^*'s are location parameters, and all others are variance parameters.

“ we assume all these parameters are *a priori* independent”

For any other location parameter (say θ), we use a noninformative uniform distribution, $\theta \sim \text{Uniform}(L_\theta, U_\theta)$, which should provide a sufficiently wide coverage for all plausible values of θ suggested by data. For the added small value

joint posterior distribution.

$\omega_Q = 1$ and $\theta_{QQ} = 1$. For the other parameters in $\boldsymbol{\omega}$ and $\boldsymbol{\Theta}$, we consider normal priors and set $\omega_q \sim \mathrm{N}(\mu_\omega, \sigma_\omega^2)$, $q = 1, \ldots, Q-1$ and $\theta_{qq'} \sim \mathrm{N}(\mu_\theta, \sigma_\theta^2)$, $q = 1, \ldots, Q-1$, $q' = q, \ldots, Q$. We suggest users choose the standard normal distribution; that is, $\mu_\omega = \mu_\theta = 0$ and $\sigma_\omega = \sigma_\theta = 1$. For the decay parameter λ, we specify a gamma prior $\lambda \sim \mathrm{Ga}(a_\lambda, b_\lambda)$. One standard way of setting a weakly informative gamma prior is to choose small values for the two parameters, such as $a_\lambda = b_\lambda = 0.001$. The $\mathrm{Ga}(a, b)$ prior is an attempt at noninformativeness within the conditionally conjugate family, with both a and b set to a low value, such as 0.1, 0.01 and 0.001 (Gelman et al. (2014)).

Distributions for parameters

- what about fiducial, confidence, significance, inferential models, etc.?
- do they provide a way around the problems with objective Bayes?
- no
- confidence approach is to pre-specify, or identify, from the model, a quantity that measures the parameter of interest
  focus parameter, Hjort & Schweder, 2016
- significance function approach is to use higher-order asymptotic theory to tell you what that quantity should be
  F & R, 2018

… distributions for parameters

- posterior distribution
- fiducial probability
- confidence distribution
- structural probability
- significance function
- belief functions

- objective Bayes
- generalized fiducial inference
  Hannig; Taraldsen
- confidence distributions and confidence curves — Hjort, Schweder, Xie
- approximate significance functions
  Brazzale et al 2007; Fraser & R 1993
- inferential models — Martin & Liu

high-dimensional inference and model selection

High-dimensional inference and model selection

- $Y = X\beta + \epsilon, \quad \epsilon \sim N_n(0, I), \quad \beta \in \mathbb{R}^p, \quad p >> n$
- assumption of sparsity – many components of β are 0

Lasso

- prior specification first on dimension s, then on subset $S \subset \{1, \dots, p\}$ with $|S| = s$, finally on $\beta_S = \{\beta_i, i \in S\}$

$$\pi(S, \beta) \propto \pi_p(|S|) \frac{1}{\binom{p}{|S|}} g_S(\beta_S) \delta_0(\beta_{S^c})$$

- example

$$\pi_p(|S|) \propto (cp^a)^{-s}, \quad \beta_j \sim \text{i.i.d. Laplace}$$

Castillo et al.

- example

$$\pi_p = \text{Bin}(p, r), \quad \beta_j \sim \text{i.i.d. } (1 - r)\delta_0 + r \text{ Laplace}, \quad r \sim \text{Beta}(1, p^u)$$

spike and slab

- under conditions on design matrix X $\inf\{\frac{||X\beta||_2}{||X||||\beta||_2} : |S_\beta| \leq s\} > 0$
- and on the scale parameter in the Laplace prior $\frac{||X||}{p} \leq \lambda \leq 2||X||(\log p)^{1/2}$
- obtain various consistency results on posterior estimates of $|S|$, S and β
- in particular, Bayesian credible sets for β are well-calibrated
- special case $n = p, X = I$: “sequence model”:

$$Y_i \sim N(\beta_i, 1), \quad i = 1, \ldots, n$$

Stein’s example

- $y_i \sim N(\theta_i, 1), \quad i = 1, \ldots, n; \quad \theta$ sparse

nonparametric Bayes

- prior specification first on dimension s, then on subset $S \subset \{1, \ldots, p\}$ with $|S| = s$, finally on $\theta_S = \{\theta_i, i \in S\}$

$$\pi(S, \theta) \propto \pi_p(|S|) \frac{1}{\binom{p}{|S|}} g_S(\theta_S) \delta_0(\theta_{S^c})$$

- example

$$\pi_p(|S|) \propto (cp^a)^{-s}, \quad \theta_j \sim \text{i.i.d. } \cancel{\text{Laplace}} \rightarrow \text{Normal}$$

$N(Y_S, \sigma^2 \tau^{-1} I_{|S|})$

- tempered likelihood

generalized Bayes

$$\pi(\theta \mid y) \propto \{L_n(\theta; y)\}^\alpha \pi(\theta)$$

- posterior coverage for linear functions of θ

e.g. components

...sequence model

- $Y_i = \theta_i + \epsilon_i, \quad \epsilon_i \sim N(0,1)$
- "horseshoe" prior $\theta_i \sim N(0, \nu_i^2\tau^2), \quad \nu_i \sim C^+(0,1)$

  τ hyperparameter

  Carvalho et al. 2010
- frequentist coverage of Bayesian credible sets for θ based on horseshoe posterior
- identify three regions, or three "types" of parameters: small, medium and large
- posterior calibrated for small and large θ but not intermediate values

$$\begin{aligned}
&\text{small:} && |\theta_i| < k_S\tau \\
&\text{medium:} && f_\tau\tau \le |\theta_i| \le k_M\sqrt{2\log(1/\tau)} \\
&\text{large:} && k_L\sqrt{2\log(1/\tau)} \le |\theta_i|
\end{aligned}$$

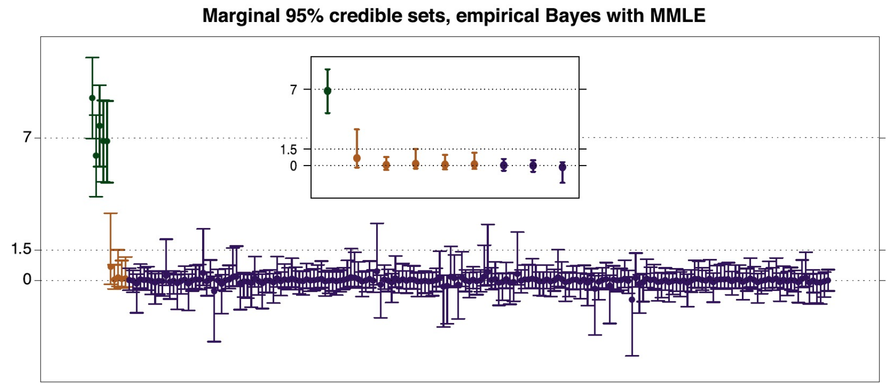


Figure 1: 95% marginal credible intervals based on the MMLE empirical Bayes method, constructed using the 2.5% and 97.5% quantiles, for a single observation Y^n of length $n = 200$ with $p_n = 10$ nonzero parameters, the first 5 (from the left) being 7 (green), the next 5 equal to 1.5 (orange); the remaining 190 parameters are coded (blue). The inserted plot zooms in on credible intervals 5 to 13, thus showing one large mean and

Horseshoe prior

- also called global-local shrinkage prior
- proposed here as a default (objective) prior
- for regular (low-dimensional) models
- "Global-local shrinkage priors can separate a low-dimensional signal from high-dimensional noise even for nonlinear functions." abstract
- Example $y_i = \theta_i + \epsilon_i, \quad \psi = ||\theta||$

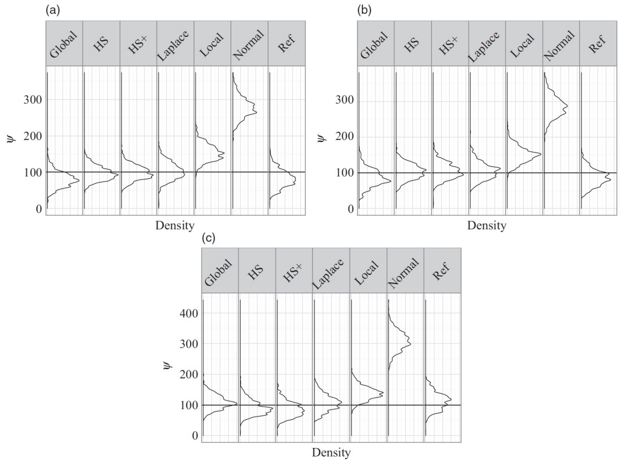


Fig. 1. Posterior densities of $\psi = \sum_{i=1}^{100} \theta_i^2$ for the horseshoe (HS), horseshoe+ (HS+), Laplace, normal, pure-local, pure-global and reference priors. In each panel, q_p is the number of nonzero means and A is the magnitude of the nonzero means: (a) $A = 10$ and $q_p = 1$ (sparse); (b) $A = 5$ and $q_p = 4$ (sparse); (c) $A = 1$ and $q_p = 100$ (dense). The horizontal line is at the true value $\psi = 100$.

- Miller and Dunson propose a *c*-posterior

$$\pi^c(\theta \mid y) \propto \{L(\theta; y)\}^{a/(a+n)} \pi(\theta)$$

a a hyper-parameter

- based on modelling misspecification via a distance between observations $y_1, \ldots, y_n$ and 'ideal' observations $Y_1, \ldots, Y_n$

robust Bayes

- Grünewald and van Ommen refer to 'generalized posterior'

$$\pi(\theta \mid y) \propto \{L(\theta; y)\}^{\eta_n} \pi(\theta), \eta < 1$$

- with a complicated method for estimating η_n from the data
- this can also be interpreted as 'ordinary Bayes' with a data-dependent prior

Walker and Hjort 2002

Too much notation

- sparse regression or sparse normal means
- prior on sparsity $\times$ prior on means **or** spike and slab; Castillo et al
- normal prior on means; Cauchy hyperprior on variance **or** horseshoe; Carvalho et al
- empirical prior on means; tempered likelihood function Martin & Ning
- consistency and asymptotic normality theorems re posterior for the means
- empirically good behaviour of horseshoe for non-linear functions of means Bhadra et al

… distributions for parameters

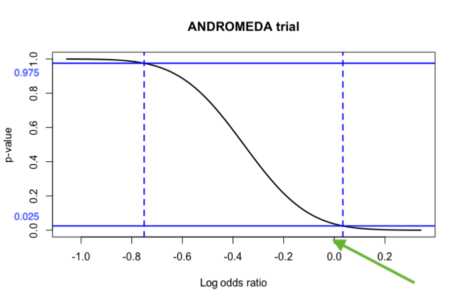
- posterior distribution
- fiducial probability
- confidence distribution
- structural probability
- significance function
- belief functions

- objective Bayes
- generalized fiducial inference
  Hannig; Taraldsen
- confidence distributions and confidence curves Hjort, Schweder, Xie
- approximate significance functions
  Brazzale et al 2007; Fraser & R 2013
- inferential models Martin & Liu

high-dimensional inference and model selection

Approximate significance functions

Approximate significance functions



	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

$p(\psi)$ based on saddlepoint approximation to conditional density of s_1, given s_2

full exponential family

Summary

Statistical theory, reproducibility and data science

- dichotomizing conclusions based on *p*-values is not a good idea
- statistical science is more nuanced than that
- science rarely advances on the basis of a single study

- posterior distributions need to be treated with care
- they can depend heavily on the prior, even when it seems uninformative

- several current versions of fiducial inference: confidence, significance, generalized fiducial
- "making a Bayesian omelette without cracking the Bayesian eggs"

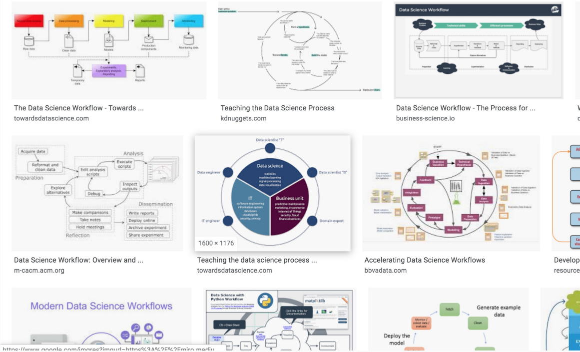
- all these methods require a reduction of data and parameter space to a scalar dimension
- how do we do this ??
- foundations

Role of Foundations

- avoid apparent discoveries based on spurious patterns
- shed light on the structure of the problem
- obtain calibrated inferences about interpretable parameters
- provide a realistic assessment of precision
- understand when/why methods work/fail

Needs in applications

- something that works
- gives ‘sensible’ answers
- not too sensitive to model assumptions
- computable in reasonable time
- provides interpretable parameters



Thank you!

References

Bayesian Analysis (2006). Volume 1 Issue 3.

Brazzale, A.R., Davison, A.C. and Reid, N. (2007). Applied Asymptotics. CUP

Berger, J. and Bernardo, J. (1992). in Bayesian Statistics 4. OUP

Bhadra, A., Datta, J., Polson, N. and Willard, B. (2016). Biometrika

Carvalho, C.M., Polson, N.G. and Scott, J.G. (2010). Biometrika

Castillo, I., Schmidt-Hieber, J. and van der Vaart, A. (2015). Ann. Statist.

Cox, D.R. (1958). Ann. Math. Statist.

Cox, D.R. (2006). Principles of Statistical Inference. CUP.

Datta, G. and Mukerjee, R. (2004). Probability Matching Priors: Higher-order Inference. Springer.

Dempster, A. (1966). Ann. Math. Statist.

Efron, B. (1993). Biometrika

Fisher, R.A. (1930). Proc. Cam. Phil. Soc.

Fraser, D.A.S. (1966). Biometrika

Fraser, D.A.S. (1991). J. Amer. Statist. Assoc.

Fraser, D.A.S. (2011). Statist. Sci.

Fraser, D.A.S. and Reid, N. (1993). Statist. Sinica

Fraser, D.A.S. and Reid, N. (2018). Ann. Appl. Statist.

Gelman, A. (2008). Ann. Appl. Statist.

Grünewald, P. and van Ommen, T. (2017) Bayesian Anal.

Hernández et al. (2019). J. Amer. Medic. Assoc. **321**, 654–664.

Hannig, J., Iyer, H., Lai, R.C.S. and Lee, T.C.M. (2016). J. Amer. Statist. Assoc.

Hjort, N. and Schweder, T. (2016). Confidence, Likelihood, Probability: Inference with Confidence Distributions

Martin, R. and Liu, C. (2015). Inferential Models

Martin, R. and Ning, B. (2018). `http://www.researchers.one/article/2018-12-6`

Mayo, D. (2020). Harvard Data Science Review **2** `https://hdsr.mitpress.mit.edu/pub/bd5k4gzf`. Published Feb 2.

Reid, N. and Cox, D.R. (2015). Intern. Statist. Rev.

Schachtman, N.A. (2019). `http://schachtmanlaw.com/american-statistical-association-consensus-versus-personal-opinion/` Posted Dec 13.

Seidenfeld, T. (1992). Statist. Sci.

Shafer, G. (1976). A Mathematical Theory of Evidence.

Spiegelhalter, D. (2019). Medium Jan. 19 `https://medium.com/wintoncentre/andromeda-and-appalling-science-a-response-to-hardwicke-and-ioannidis-a79458efdba1`

Stein, C. (1959). Ann. Math. Statist.

Stigler, S. (2013). Statist. Sci.

Taraldsen, G. and Lindqvist, B.H. (2018). JSPI

van der Pas, S., Szabó, B. and van der Vaart, A. (2017). Bayes. Anal.

Walker, S. and Hjort, N. (2002) JRSS B

Xie, M. and Singh, K. (2013) Int. Statist. Rev.

Zabell, X. (1992). Statistical Science