Distributions for Parameters

Nancy Reid
University of Toronto

WIML Unconference

July 13 2020

Those pesky p-values

Example 1: Andromeda trial

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)

- 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."

Example 2: Imperial College COVID Report 17

April 2020

- Clinical characteristics and predictors of outcomes of hospitalized patients with COVID-19 in a London NHS trust: a retrospective cohort study

- "all patients hospitalised with laboratory-confirmed SARS-CoV-2 infection at Imperial College Healthcare NHS Trust between February 25 and April 5, 2020

outcomes were recorded as of April 19, 2020

- "logistic regression models, survival analyses and cumulative competing risk analyses ... to evaluate factors associated with COVID-19 hospital mortality

- "ethnic minority groups were over-represented in our cohort
- "compared to whites, people of black ethnicity may be at increased odds of mortality

- "Further research is urgently needed to investigate these associations "

several other conclusions around co-morbidities and various blood-test indicators

- "the crude OR of death of black compared to white patients was not significant"
  odds ratio: 1·14 95%CI 0·69-1·88 p=0.62
- "adjusting for age and comorbidity showed a trend towards significance"
  adj. odds ratio: 1·72 95%CI 0·98-3·02 p=0.06
- "further accounting for admission severity ... was significant"
  adj. odds ratio: 1·83 95%CI 1·02-3·30 p=0.04
- "when adjusting for age, Elixhauser comorbidity score and severity of disease on admission, we observed higher odds of death for those from black ethnic background compared with whites
- "this finding merits further investigation **given its borderline statistical significance**"

my emphasis

Table 3: Clinical characteristics by ethnicity and logistic regression of odds of death

*Adjusted logistic regression for age, Elixhauser and EWS differences of ethnic groups

**Additional logistic regression models adjusted for individual comorbidities that had statistically significant variation across ethnic groups,but were inferior predictors compared to the selected model with Elixhauser score. CKD, cirrhotic liver disease and HIV/AIDS were not used in logistic regression models, as they had low n values for ethnic groups

Abbreviations: COPD, chronic obstructive pulmonary disease; DVT/PE, deep vein thrombosis / pulmonary embolism; HDU/ITU, high dependency unit / intensive treatment unit; HIV/AIDS, human immunodeficiency virus / acquired immunodeficiency syndrome; IVS, invasive ventilation support; SD, standard deviation

	White (n = 196)	Black (n = 116)	Asian (n = 78)	Other (n = 15)	NA (n = 115)	p-value
Male, n (%)	119 (61%)	66 (57%)	51 (65%)	10 (67%)	76 (66%)	0·60
Mean age (SD)	69-94 (15-6)	62-8 (19-53)	66-05 (15-45)	55-07 (16-01)	59-67 (17-91)	<0·01
Mean days to admission (SD)	6-78 (5-76)	7-3 (6-12)	5-87 (4-84)	7-47 (6-8)	8-34 (5-09)	0·28
Mean Elixhauser score (SD)	4-89 (6-81)	3-63 (5-84)	4-96 (6-66)	2-47 (4-53)	2-38 (4-72)	<0·01
Mean EWS score (SD)	5-01 (3-26)	4-98 (3-11)	4-53 (3-29)	5-47 (2-75)	5-53 (2-76)	0·15
Outcomes						
FiO2 >=60%, n (%)	88 (45%)	54 (47%)	38 (49%)	5 (33%)	67 (58%)	0·14
Admitted HDU/ITU, n (%)	22 (11%)	18 (16%)	12 (15%)	2 (13%)	26 (23%)	0·12
Received IVS, n (%)	21 (11%)	16 (14%)	11 (14%)	2 (13%)	23 (20%)	0·27
Died in hospital, n (%)	54 (28%)	35 (30%)	23 (29%)	3 (20%)	29 (25%)	0·86
Discharged alive, n (%)	111 (57%)	67 (58%)	43 (55%)	12 (80%)	69 (60%)	0·46
Pending outcome, n (%)	31 (16%)	14 (12%)	12 (15%)	0 (0%)	17 (15%)	0·49
Comorbidities						
Ischaemic heart disease, n (%)	18 (9%)	9 (8%)	10 (13%)	0 (0%)	6 (5%)	0·27
Chronic heart failure, n (%)	10 (5%)	6 (5%)	0 (0%)	0 (0%)	5 (4%)	0·30
Hypertension, n (%)	68 (35%)	45 (39%)	27 (35%)	5 (33%)	42 (37%)	0·96
Hyperlipidaemia, n (%)	32 (16%)	13 (11%)	18 (23%)	1 (7%)	18 (16%)	0·20
Diabetes**, n (%)	36 (18%)	39 (34%)	31 (40%)	3 (20%)	29 (25%)	<0·01
Chronic kidney disease**, n (%)	31 (16%)	12 (10%)	19 (24%)	0 (0%)	8 (7%)	<0·01
Peripheral vascular disease, n (%)	7 (4%)	2 (2%)	0 (0%)	0 (0%)	2 (2%)	0·38
Stroke, n (%)	17 (9%)	9 (8%)	6 (8%)	0 (0%)	2 (2%)	0·12
Atrial fibrillation**, n (%)	33 (17%)	10 (9%)	9 (12%)	2 (13%)	7 (6%)	<0·05
DVT/PE history, n (%)	4 (2%)	2 (2%)	0 (0%)	0 (0%)	1 (1%)	0·68
Hemiplegia, n (%)	3 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0·29
Dementia**, n (%)	23 (12%)	5 (4%)	4 (5%)	0 (0%)	3 (3%)	<0·01
Asthma, n (%)	9 (5%)	11 (9%)	5 (6%)	3 (20%)	9 (8%)	0·15
COPD**, n (%)	15 (8%)	2 (2%)	2 (3%)	0 (0%)	1 (1%)	<0·01
Connective tissue disease, n (%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	1 (1%)	0·93
Peptic ulcer, n (%)	5 (3%)	2 (2%)	1 (1%)	0 (0%)	2 (2%)	0·92
Liver (non-cirrhotic), n (%)	12 (6%)	9 (8%)	7 (9%)	3 (20%)	5 (4%)	0·20
Liver (cirrhotic)**, n (%)	2 (1%)	2 (2%)	5 (6%)	0 (0%)	1 (1%)	<0·05

		White (n = 196)	Black (n = 116)	Asian (n = 78)	Other (n = 15)	NA (n = 115)	p-value
Logistic regression of odds of death by ethnicity							
Unadjusted OR (95%CI)	Intercept	1·14 (0·69, 1·88)	1·10 (0·62, 1·96)	0·66 (0·18, 2·42)	0·89 (0·52, 1·50)	---	
Adjusted OR (95%CI) *	Intercept	1·86 (1·03, 3·35)	1·74 (0·90, 3·36)	1·72 (0·42, 7·01)	1·73 (0·94, 3·18)	---	

… Example 2

Perez-Guzman et al. 2020

Table 3: Clinical characteristics by ethnicity and logistic regression of odds of death

*Adjusted logistic regression for age, Elixhauser and EWS differences of ethnic groups

**Additional logistic regression models adjusted for individual comorbidities that had statistically significant variation across ethnic groups,but were inferior predictors compared to the selected model with Elixhauser score. CKD, cirrhotic liver disease and HIV/AIDS were not used in logistic regression models, as they had low n values for ethnic groups

Abbreviations: COPD, chronic obstructive pulmonary disease; DVT/PE, deep vein thrombosis / pulmonary embolism; HDU/ITU, high dependency unit / intensive treatment unit; HIV/AIDS, human immunodeficiency virus / acquired immunodeficiency syndrome; IVS, invasive ventilation support; SD, standard deviation

	White (n = 196)	Black (n = 116)	Asian (n = 78)	Other (n = 15)	NA (n = 115)	p-value
Male, n (%)	119 (61%)	66 (57%)	51 (65%)	10 (67%)	76 (66%)	0·60
Mean age (SD)	69-94 (15-6)	62-8 (19-53)	66-05 (15-45)	55-07 (16-01)	59-67 (17-91)	<0·01
Mean days to admission (SD)	6-78 (5-76)	7-3 (6-12)	5-87 (4-84)	7-47 (6-8)	8-34 (5-09)	0·28
Mean Elixhauser score (SD)	4-89 (6-81)	3-63 (5-84)	4-96 (6-66)	2-47 (4-53)	2-38 (4-72)	<0·01
Mean EWS score (SD)	5-01 (3-26)	4-98 (3-11)	4-53 (3-29)	5-47 (2-75)	5-53 (2-76)	0·15
Outcomes						
FiO2 >=60%, n (%)	88 (45%)	54 (47%)	38 (49%)	5 (33%)	67 (58%)	0·14
Admitted HDU/ITU, n (%)	22 (11%)	18 (16%)	12 (15%)	2 (13%)	26 (23%)	0·12
Received IVS, n (%)	21 (11%)	16 (14%)	11 (14%)	2 (13%)	23 (20%)	0·27
Died in hospital, n (%)	54 (28%)	35 (30%)	23 (29%)	3 (20%)	29 (25%)	0·86
Discharged alive, n (%)	111 (57%)	67 (58%)	43 (55%)	12 (80%)	69 (60%)	0·46
Pending outcome, n (%)	31 (16%)	14 (12%)	12 (15%)	0 (0%)	17 (15%)	0·49
Comorbidities						
Ischaemic heart disease, n (%)	18 (9%)	9 (8%)	10 (13%)	0 (0%)	6 (5%)	0·27
Chronic heart failure, n (%)	10 (5%)	6 (5%)	0 (0%)	0 (0%)	5 (4%)	0·30
Hypertension, n (%)	68 (35%)	45 (39%)	27 (35%)	5 (33%)	42 (37%)	0·96
Hyperlipidaemia, n (%)	32 (16%)	13 (11%)	18 (23%)	1 (7%)	18 (16%)	0·20
Diabetes**, n (%)	36 (18%)	39 (34%)	31 (40%)	3 (20%)	29 (25%)	<0·01
Chronic kidney disease**, n (%)	31 (16%)	12 (10%)	19 (24%)	0 (0%)	8 (7%)	<0·01
Peripheral vascular disease, n (%)	7 (4%)	2 (2%)	0 (0%)	0 (0%)	2 (2%)	0·38
Stroke, n (%)	17 (9%)	9 (8%)	6 (8%)	0 (0%)	2 (2%)	0·12
Atrial fibrillation**, n (%)	33 (17%)	10 (9%)	9 (12%)	2 (13%)	7 (6%)	<0·05
DVT/PE history, n (%)	4 (2%)	2 (2%)	0 (0%)	0 (0%)	1 (1%)	0·68
Hemiplegia, n (%)	3 (2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0·29
Dementia**, n (%)	23 (12%)	5 (4%)	4 (5%)	0 (0%)	3 (3%)	<0·01
Asthma, n (%)	9 (5%)	11 (9%)	5 (6%)	3 (20%)	9 (8%)	0·15
COPD**, n (%)	15 (8%)	2 (2%)	2 (3%)	0 (0%)	1 (1%)	<0·01
Connective tissue disease, n (%)	1 (1%)	1 (1%)	0 (0%)	0 (0%)	1 (1%)	0·93
Peptic ulcer, n (%)	5 (3%)	2 (2%)	1 (1%)	0 (0%)	2 (2%)	0·92
Liver (non-cirrhotic), n (%)	12 (6%)	9 (8%)	7 (9%)	3 (20%)	5 (4%)	0·20
Liver (cirrhotic)**, n (%)	2 (1%)	2 (2%)	5 (6%)	0 (0%)	1 (1%)	<0·05

29 April 2020 Imperial College COVID-19 response team

		White (n = 196)	Black (n = 116)	Asian (n = 78)	Other (n = 15)	NA (n = 115)	p-value
Logistic regression of odds of death by ethnicity							
Unadjusted OR (95%CI)	Intercept		1·14 (0·69, 1·88)	1·10 (0·62, 1·96)	0·66 (0·18, 2·42)	0·89 (0·52, 1·50)	---
Adjusted OR (95%CI) *	Intercept		1·86 (1·03, 3·35)	1·74 (0·90, 3·36)	1·72 (0·42, 7·01)	1·73 (0·94, 3·18)	---

adj. odds ratio: 1·83 p=0.04

odds ratio: 1.14 p=0.62

Examples 1 and 2

unadjusted for confounders:

Example 1: ANDROMEDA trial

	Died	Lived	
New	74 34%	138	212
Old	92 43%	120	212
Total	166	258	424

randomized controlled trial
sepsis very difficult to treat
$p = 0.06$

Example 2: IC Report 17

	Died*	did not**	
White	54 28%	142	196
Black	35 30%	81	116
Total	89	223	312

*in hospital; **unresolved or discharged alive

observational study
large number of potential confounders
$p = 0.67$ (0.04)

" From now on, BASP is banning the NHSTP" [Null Hypothesis Significance Testing]

p-values, confidence intervals, ...

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

nature > nature human behaviour > comment > article

nature human behaviour

Comment | Published: 01 September 2017

Redefine statistical significance

Daniel J. Benjamin, James O. Berger, [...] Valen E. Johnson

Nature Human Behaviour **2**, 6–10(2018) | Cite this article

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

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

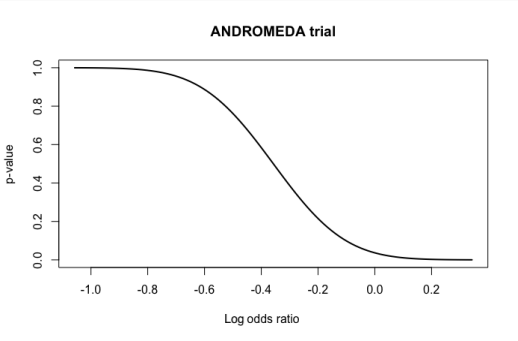
Fraser 1991

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



A *p*-value function

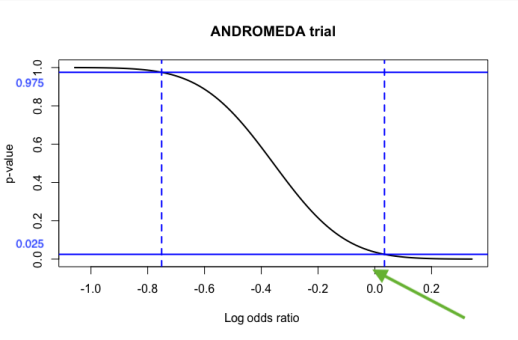
Fraser 1991

ANDROMEDA trial

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

2-sided *p*-value = 0.07

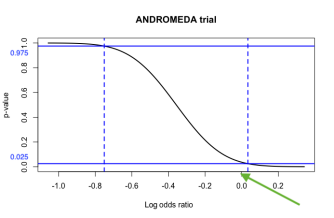
likelihood ratio test
no adjustment for covariates



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

Distributions for parameters

- significance function $p(\theta) = \Pr(y \geq y^{o} \mid \theta)$
- confidence distribution — set of all confidence intervals
- fiducial probability — inversion of data-generating equation



- structural probability — a re-formulation of fiducial probability for transformation models

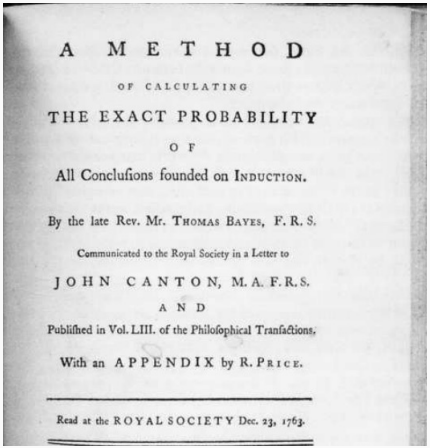
  Fraser 1966
- belief functions — upper and lower probabilities

  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?



A METHOD

OF CALCULATING

THE EXACT PROBABILITY

OF

All Conclusions 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

Published in Vol. LIII. of the Philosophical Transactions,

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 ~~are~~ **seem to be** probability statements about unknowns

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

November 10 – November 14, 2014

7th Bayes, Fiducial and Frequentist Statistics Conference

Methodological, Computational, and Ethical Principles for Data Science

May 6 - 8, 2020, The Fields Institute

Location: Fields Institute, Room 230

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

- 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; PMM 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:

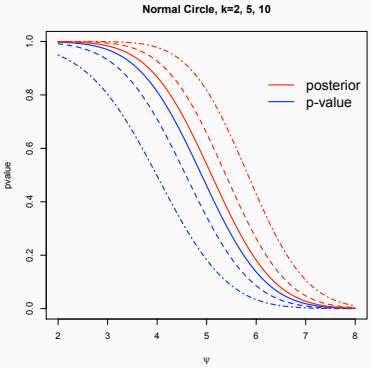
$$\begin{aligned} y_i &\sim N(\theta_i, 1/n), \quad i = 1, \ldots, k \\ \pi(\theta_i) &\propto 1 \\ \pi(\theta \mid y) &\propto N(y, I_k/n) \end{aligned}$$

Example 3

Stein, 1959

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

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



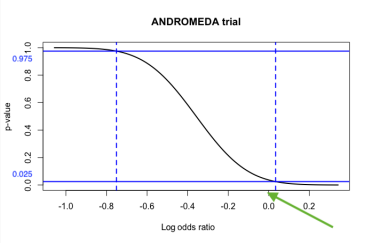
True prob*	0.001	0.01	0.025	0.05
Bayes prob*	10^{-6}	10^{-4}	0.001	0.003

* left tail

Targetting on parameter

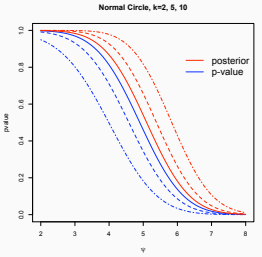
Example 1: 2×2 table

Based on conditional distribution of odds-ratio, given marginal totals



Example 3: $y_i \sim N(\theta_i, 1/n)$

Based on marginal distribution of Σy_j^2



High-dimensional inference and model selection

- $Y_i \sim N(\beta_i, 1), \quad i = 1, \ldots, n$ $y_i \sim N(\theta_i, 1/n), \quad i = 1, \ldots, k$
- prototype for nonparametric modelling
- sparsity assumption needed on $\beta = (\beta_1, \ldots, \beta_n)$
- special case of high-dimensional regression

$$y = X\beta + \epsilon, \quad \epsilon \sim N_n(0, I), \quad \beta \in \mathbb{R}^p, \quad p >> n$$

- can we find an objective Bayes prior?
- or at least a prior leading to frequentist coverage of posterior credible intervals

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

Lasso

- prior

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

- prior specification first on dimension:

$$\pi_p(|S|) \propto (cp^a)^{-s}$$

- then on subset $S \subset \{1, \ldots, p\}$, given dimension: $\frac{1}{\binom{p}{|S|}}$
- finally on $\beta_S = \{\beta_j, j \in S\}$, given S:

$$\beta_j \sim \text{i.i.d. Laplace}(\lambda), \quad j \in S$$

- 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
- Lasso posterior not useful for inference

  Lasso point estimate is posterior mode (from Laplace prior)

- $y_i \sim N(\beta_i, 1), \quad i = 1, \ldots, n; \quad \beta$ sparse — nonparametric Bayes

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

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

- as above

$$\pi_p(|S|) \propto (cp^a)^{-s}, \quad \text{but } \beta_S \sim \sout{\text{Laplace}} \to N(Y_S, \sigma^2 \tau^{-1} I_{|S|})$$

- and tempered likelihood — generalized Bayes

$$\pi(\beta, S \mid y) \propto \{L_n(\beta_S; y)\}^{\alpha} \pi(\beta, S)$$

- posterior coverage for linear functions of β

Various Bayesian approaches

incomplete

- van der Pas et al 2018: horseshoe prior $\theta_i \sim N(0, \nu_i^2\tau^2), \quad , \nu_i \sim C^+(0,1)$

  τ hyperparameter

- Bhadra et al 2016: horseshoe plus with hyperparameters e.g. $\nu_i \sim C^+(0, \tau\eta_i)$

  nonlinear marginals

- Miller & Dunson 2017: c-posterior based on tempered likelihood — a a hyper-parameter

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

- Grünewald and van Ommen 2018: 'generalized posterior' — large ML literature

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

- Giordano et al. 2018: 'local robustness' and mean-field variational Bayes

  a link to frequentist properties of Bayes estimates; Efron 2015

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

- calibrated inference in high-dimensional models still a w.i.p.

- considerable scope for increased interactions between ML and Stat

Thank you!