# An example of application of PC priors to a Bayes factor computation

# Hommages

- Gustave Malécot (1911-1998)
  - ENS élève de Borel, Cartan, Darmois & Fréchet
  - Thèse sur les travaux de Fisher
  - Maitre de conférences, Université de Lyon
  - Probabiliste des processus stochastiques de l'hérédité
  - Concept probabiliste de la  parenté (IBD
  - Travaux diffusés par Gillois, Jacquard, Lamotte, Lalouel, Langaney
  - USA: Kempthorne, Nagylaki,
  - *Les mathématiques de l'hérédité*
  - *La subjectivité de la connaissance scientifique, 1947*
- Guy Lefort (1921-1979)
  - Cf Applibugs, 4 juin, 2009
  - *Introduction à la théorie de la décision et à la statistique bayésienne, 1975,INA-PG*

# Contents


- Introduction
- Basics on PC priors
- Context of application
- Benjamin & Berger's recommendation(s)
- Variants of BF bounds
  - Uniform prior
  - PC priors
- Discussion

# Why be concerned with PC priors

- Penalising Complexity (PC) priors as another quest for default (weakly informative) priors
- Paper by Simpson, Rue, Martins, Riebler & Sorbye (2014) commented in https://xianblog.wordpress.com/ (April 1st, 18 /2014)
- Recurrent issues with priors for variance components (Applibugs 27-11-2008)
  - ✓ Overfitting with Gamma (a,b) on precision components
  - ✓ PC: Exponential on standard deviation
- Reminiscence of same (old) ideas prior to PC priors
  - ➢ MINQUE-0 (Rao, 1971)
- Close links with hypothesis testing (Applibugs, 19-12-2013)

# Basic principles of PC priors

1)Principle of parsimony

Occam's (1287-1347) razor: "*Pluralitas non est ponenda sine necessitate*"

a model viewed as an extension of a "base model" characterized by a value of a flexibility parameter $\xi$

Ex $\mathbf{u} \mid \xi \sim \mathcal{N}\left(0, \xi \mathbf{I}_p\right)$ vs $\pi\left(\mathbf{u} \mid \xi \rightarrow 0\right)$ for the base which mimics H1 vs H0

2)Complexity of $f_1$ measured via a divergence from the base $f_0$

$$\boxed{\mathrm{D}\left(f_1 \| f_0\right) = \sqrt{2KL\left(f_1 \| f_0\right)}} \text{ where } \boxed{KL\left(f_1 \| f_0\right) = \int f_1(t) \ln \frac{f_1(t)}{f_0(t)} dt}$$

Measure of discrepancy between $f_0$ and $f_1$, asymmetric, not a metric

# Basic principles of PC priors/divergence

3) Penalising the deviation from the base via a constant decay rate

$$\pi_d(d+\delta)/\pi_d(d) = r^{\delta}$$

$(0 < r < 1)$ implies an exponential form of the prior on $d$

$$\pi_d(d) = \lambda \exp(-\lambda d)$$

$$\pi(\xi) = \lambda \exp\left[-\lambda d(\xi)\right]\left|d'(\xi)\right|$$

$\lambda > 0$ parameter monitoring the change of $\pi(\xi)$ $\lambda = -\log r$

Invariance by construction

# Basic principles of PC priors/choice of lambda

## 4) Choice of $\lambda$

$$\Pr\left(q\left(\xi\right)\geq U\right)=\alpha$$

$$Ex:\mathbf{u}\,|\,\xi\sim\mathcal{N}\left(0,\xi\mathbf{I}_p\right)\xi\text{ variance}$$

$$\text{Let }q\left(\xi\right)=\sqrt{\xi}=\sigma\Rightarrow\sigma\sim Exp\left(\lambda\right)$$

$$\Pr\left(\sigma\geq\sigma_b\right)=e^{-\lambda\sigma_b}=\alpha$$

$$\Rightarrow\boxed{\lambda=-\ln\left(\alpha\right)/\sigma_b}$$

# Distribution on divergence: expo (solid); G(1,a) (dotted)

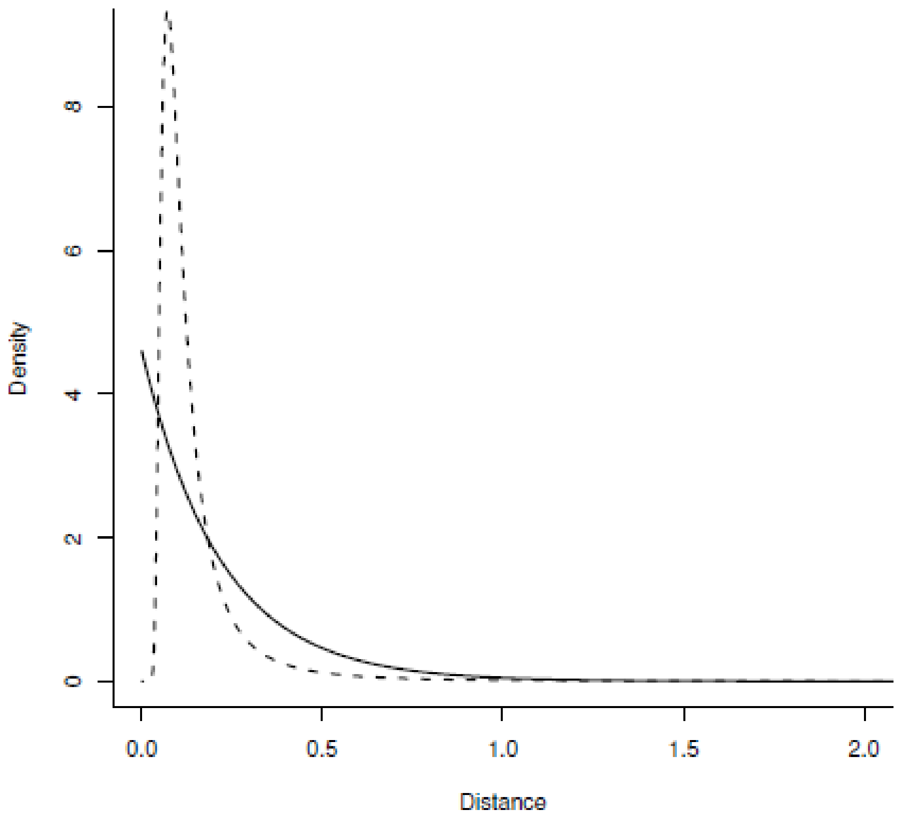
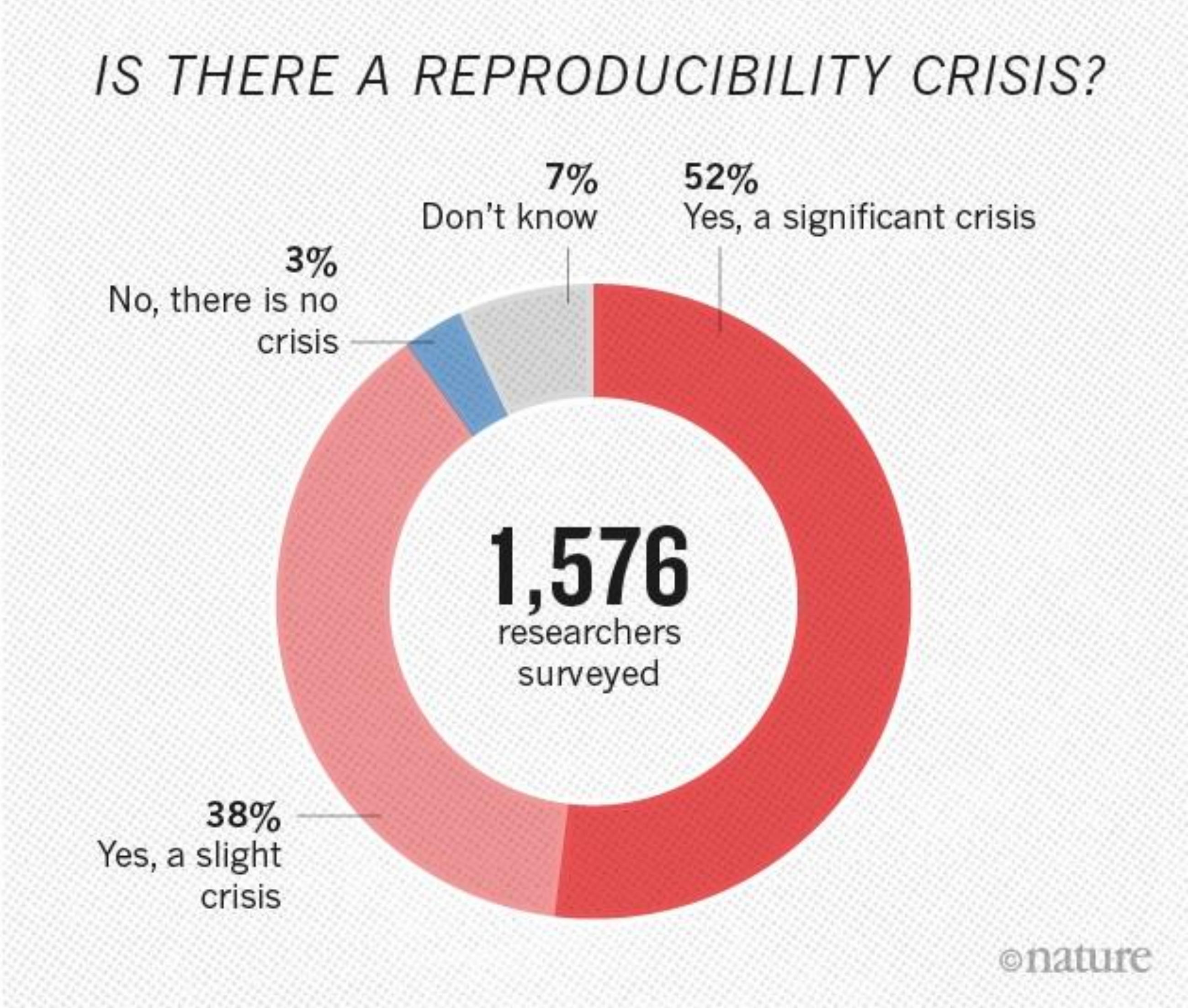
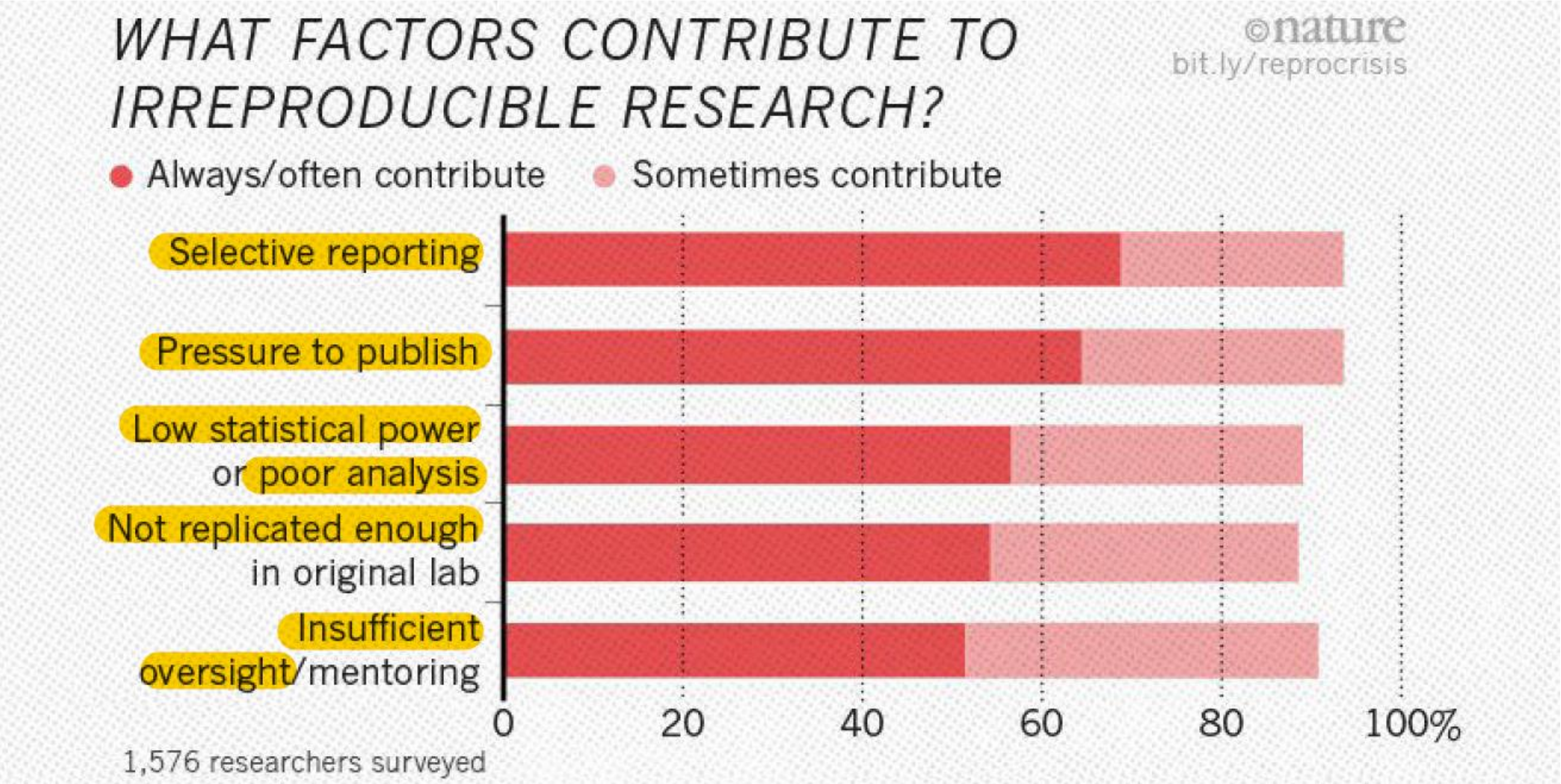



Figure 4: Panel (a) displays the new prior (solid) with parameters $(U = 0.3/0.31, \alpha = 0.01)$, and the Gamma(shape $= 1$, rate $= a$) prior (dashed). The value of $a$ is computed so that the marginal variance for the random effect are equal for the two priors, which leads to $a = 0.0076$. Panel (b) shows the same two priors on the distance scale demonstrating that the density for the Gamma-prior is zero at distance zero.

# Reproductibility crisis
# (Baker, Nature, 2016,533)



# Reproductibility crisis: factors involved



# Erroneous intuition about probability of replicating results (Tversky & Kahneman, 1971)

"Suppose you have run an experiment on 20 subjects, and have obtained a significant result which confirms your theory ($z = 2.23$, $p < .05$, two-tailed). You now have cause to run an additional group of 10 subjects. What do you think the probability is that the results will be significant, by a one-tailed test, separately for this group?"

# Erroneous intuition about probability of replicating results (Tversky & Kahneman, 1971)

- Questionnaire distributed at meetings of the Mathematical Psychology Group and of the American Psychological Association
- 9 of 84 (**10.7%**) gave answers between 0.40-0.60
- Correct answer: 0.473 (0.478 Bayesian flavour)
- « Believer of small numbers pratice science as follows:
  - **Undue confidence in early trends** and stability of observed patterns
  - **Unreasonably high expectations** about replicability of significant results»
- « Acquaintance with formal logic and with probability theory does not extinguish erroneous intuitions »

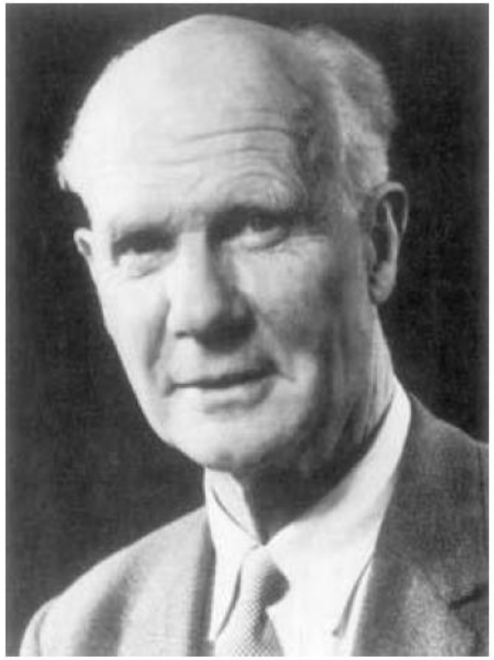
# Significance testing & P-value

Sir Ronald Fisher (1890-1962)

# Hypothesis testing/NP

**Jerzy Neyman (1894-1981)**

**Egon Pearson (1895-1980)**



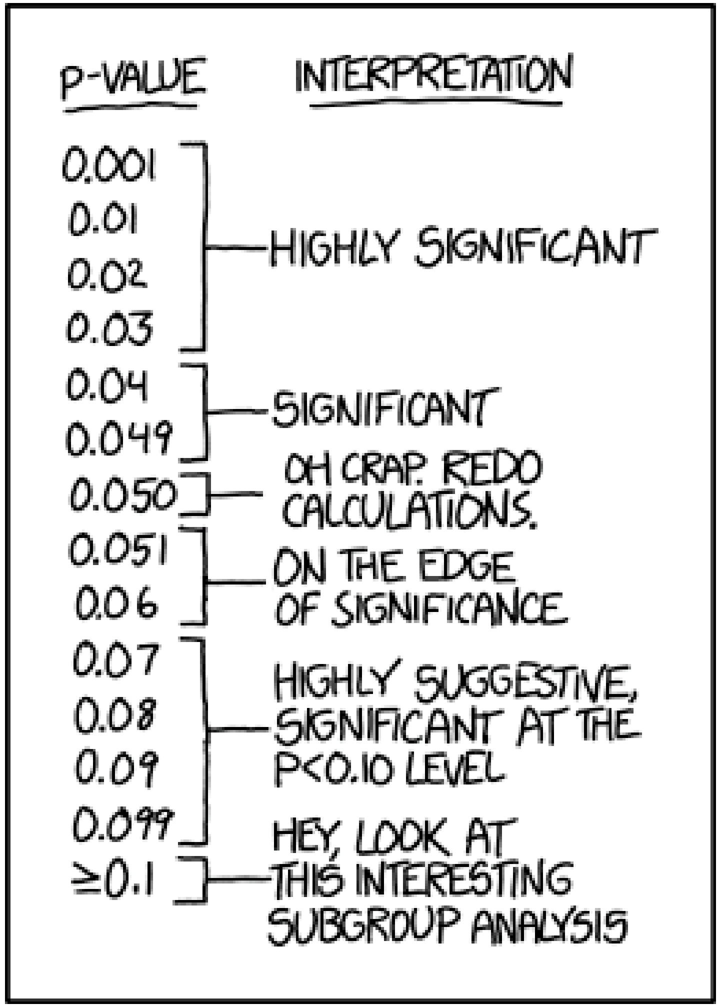
# P-value vs Hypothesis testing
# (Biau et al, 2010)

| Fisher's p value | Hypothesis testing |
| --- | --- |
| Ronald Fisher | Jerzy Neyman and Egon Pearson |
| Significance test | Hypothesis test |
| p Value | $\alpha$ |
| The p value is a measure of the evidence against the null hypothesis | $\alpha$ and $\beta$ levels provide rules to limit the proportion of errors |
| Computed a posteriori from the data observed | Determined a priori at some specified level |
| Applies to any single experiment | Applies in the long run through the repetition of experiments |
| Subjective decision | Objective behavior |
| Evidential, ie, based on the evidence observed | Nonevidential, ie, based on a rule of behavior |

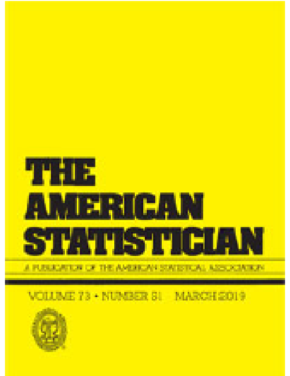
# NHST as a synthesis

- Most statisticians do not make the difference between Fisher P-values/Significance testing and Neyman-Pearson hypothesis testing « Null Hypothesis Significance Testing » (Lecoutre & Poitevineau, 2014. The Significant Test Controversy Revisited)
- « It is an incoherent mismah of some of Fisher's ideas on one hand and some of the ideas of Neyman and ES Pearson on the other » Gigerenzer, 1993
- « I don't care about the people, Neyman, Fisher, and Pearson. I care about what researchers do. They do something called NHST, and it's a disaster", Gelman, 2019
- Ex: ICH E9 Guidelines for Clinical Trials
- Synthesis responsible for many bad uses of p-values (Greenland, Senn et al, 2016)

# NHST as a synthesis



# Propositions




Taylor & Francis
Taylor & Francis Group

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

## Moving to a World Beyond "$p < 0.05$"

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

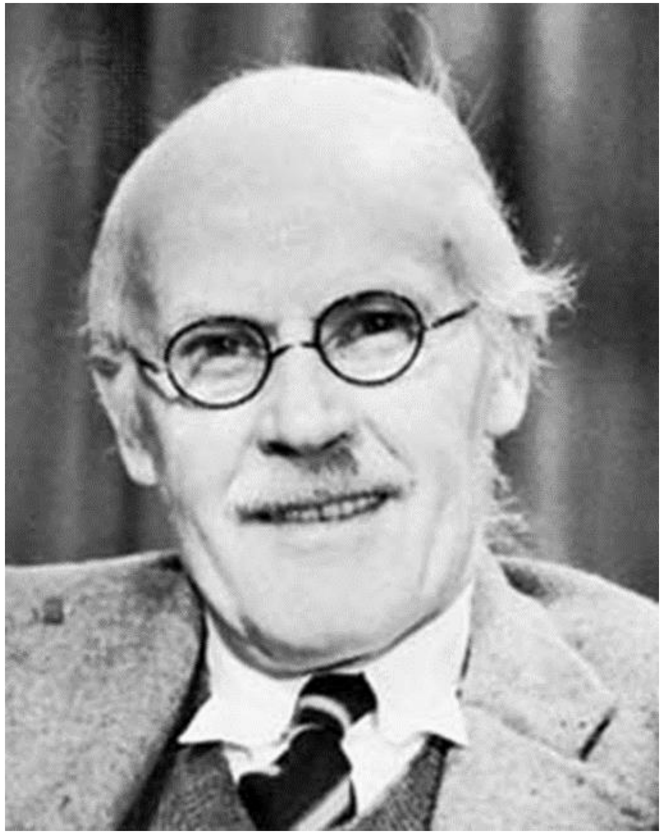
**To cite this article:** Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond "$p$<0.05", The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

**To link to this article:** https://doi.org/10.1080/00031305.2019.1583913

# Three recommendations for improving the use of p-values

- Benjamin & Berger (2019) *The American Statistician*, 73, 186-191
- R 0.1-Refer to discoveries,with a p-value between 0.05 and 0.005 as « suggestive » rather than « significant »
- R 0.2-When reporting a p-value, p, in a test of the null hypothesis H0 vs an alternative H1, also report that the data-based odds of H1 to H0 being true are at most 1/[-eplogp]
- R 0.3 –Determine and report your prior odds of H1 to H0 and derive and report the final (posterior) odds of H1 to H0

# Sir Harold Jeffreys (1891-1989)



# Posterior odds and BF

Bayes rule $\underbrace{\frac{\Pr(H_1 \mid y)}{\Pr(H_0 \mid y)}}_{\text{"Posterior odds"}} = \underbrace{\frac{\Pr(H_1)}{\Pr(H_0)}}_{\text{"Prior odds"}} \text{ x } \underbrace{\frac{m(y)}{f(y \mid \theta_0)}}_{\text{"Bayes Factor" } B_{10}}$

$H_0 : \theta = \theta_0 \qquad H_1 : \theta \sim \pi(.)$

where $m(y) = \int f(y \mid \theta)\pi(\theta)\,d\theta$

("$K^{-1}$" de Jeffreys,1939), proposed independently by Turing at Blechtley Park (Good,1979)

"Weight of evidence" defined as $10\log_{10} B_{10}$ (deciban)

# Posterior probability & BF

$$\rho_{10} = \Pr(\mathrm{H}_1) / \Pr(\mathrm{H}_0)$$

$$\Pr(\mathrm{H}_1 \mid y) = \frac{B_{10}\rho_{10}}{1 + B_{10}\rho_{10}} \quad \Pr(\mathrm{H}_0 \mid y) = \frac{1}{1 + B_{10}\rho_{10}}$$

$$\boxed{\Pr(\mathrm{H}_1 \mid y, \rho_{10} < 1) < \Pr(\mathrm{H}_1 \mid y, \rho_{10} = 1)}$$

# BF Calibration

BF Calibration according to Jeffreys

| $K^{-1} = B_{10}$ | $\Pr(H_1 \mid \mathbf{y})^*$ | Deciban (dB) | Deviance (ΔD) | Strength of evidence against the null |
|---|---|---|---|---|
| $<1$ | $>1/2$ | $<0$ | | 0-Null hypothesis supported |
| 1.0 à 3.16 | 0.50 à 0.76 | 0 à 5 | 0 à 2.3 | 1-Not worth than a bare mention |
| 3.16 à 10 | 0.76 à 0.91 | 5 à 10 | 2.3 à 4.6 | 2-Substantial |
| 10.0 à 31.62 | 0.91 à 0.97 | 10 à 15 | 4.6 à 6.9 | 3-Strong |
| 31.62 à 100 | 0.97 à 0.99 | 15 à 20 | 6.9 à 9.2 | 4-Very strong |
| $>100$ | $>0.99$ | $>20$ | $>9.2$ | 5-Decisive |

$K = B_{01} = f(y \mid H_0) / f(y \mid H_1)$: « grade of decisiveness of evidence », Jeffreys (1961) Appendix B

$\Pr(H_1 \mid \mathbf{y})^*$ setting $\Pr(H_0) = \Pr(H_1) = 1/2$

Deciban pertaining to $K^{-1}: dB = 10\log_{10} B_{10}$ (Turing A, Good IJ, 1940)

Deviance defined as : $\Delta D_{01} = D_0 - D_1 = -2\log \mathrm{L}_0 / \mathrm{L}_1$ (reduced vs complete models)

# Frequentist justification of BF

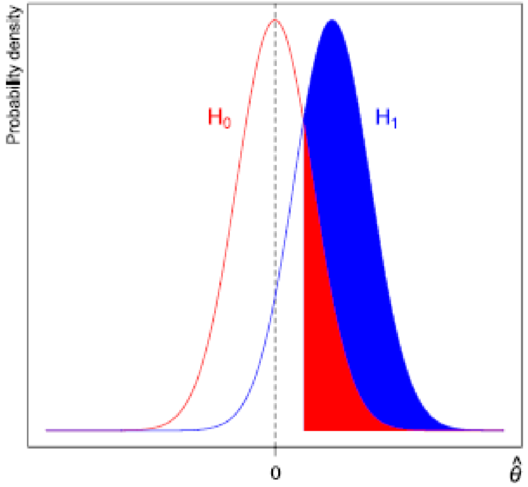
Bayarri, Benjamin, Berger, Sellke (2015) . Let $BF_{10}(y)=\frac{m(y)}{f(y|\theta_0)}$

Under Neyman-Pearson $\alpha=\Pr(R_0|H_0)$; $1\text{-}\beta(\theta)=\Pr(R_0|H_1:\theta)$

$R_0$ = rejection of $H_0$

$$\frac{\Pr(H_1|R_0)}{\Pr(H_0|R_0)}=\frac{\pi_1}{\pi_0}\frac{1\text{-}\bar{\beta}}{\alpha}=\frac{\text{Pr(Correct Rejections)}}{\text{Pr(False Rejections)}}=\frac{\text{Pr(True Positive)}}{\text{Pr(False Positive)}}$$

$1-\bar{\beta}=\int(1-\beta(\theta))\pi(\theta)d\theta$ "expected power"




Then $E\left[BF_{10}(y)|H_0,R_0\right]=\frac{1\text{-}\bar{\beta}}{\alpha}$

Under $H_0$, the frequentist expectation of the post-experimental rejection ratio equals the pre-experimental rejection ratio

$\pi_1/\pi_0$ could be very low ($10^{-5}$ : $GWAS, \alpha=5x10^{-8}$)Report: $10^{-5}BF_{10}(y)$

# Upper bound of BF10

Sellke, Bayarri & Berger (2001) proposed

$$BF_{10}(p) \leq -\frac{1}{ep \log p}$$

for $p < 1/e$ , else $B_{10}(p) = 1$

see Vovk (1993) Held & Ott (2018)

# Distribution of P-values

Example: unilateral test under normality

$\mathrm{H}_0 : \mu = \mu_0$ vs $\mathrm{H}_1 : \mu \geq \mu_0$

$X_i \sim_{id} \mathcal{N}\left(\mu, \sigma^2\right)\ i = 1,..,n\ \ \sigma$ known

$p_{obs} = \Pr\left(T_{H_0} \geq t_{obs}\right)\ \ t_{obs} = \sqrt{n}\left(\bar{x} - \mu_0\right)/\sigma$

$T_{H_0} \sim \mathcal{N}\left(0,1\right)$: Distribution of statistical test under $\mathrm{H}_0$

If now $\bar{x}$ sampled from $\bar{X} \Rightarrow$ P random variable

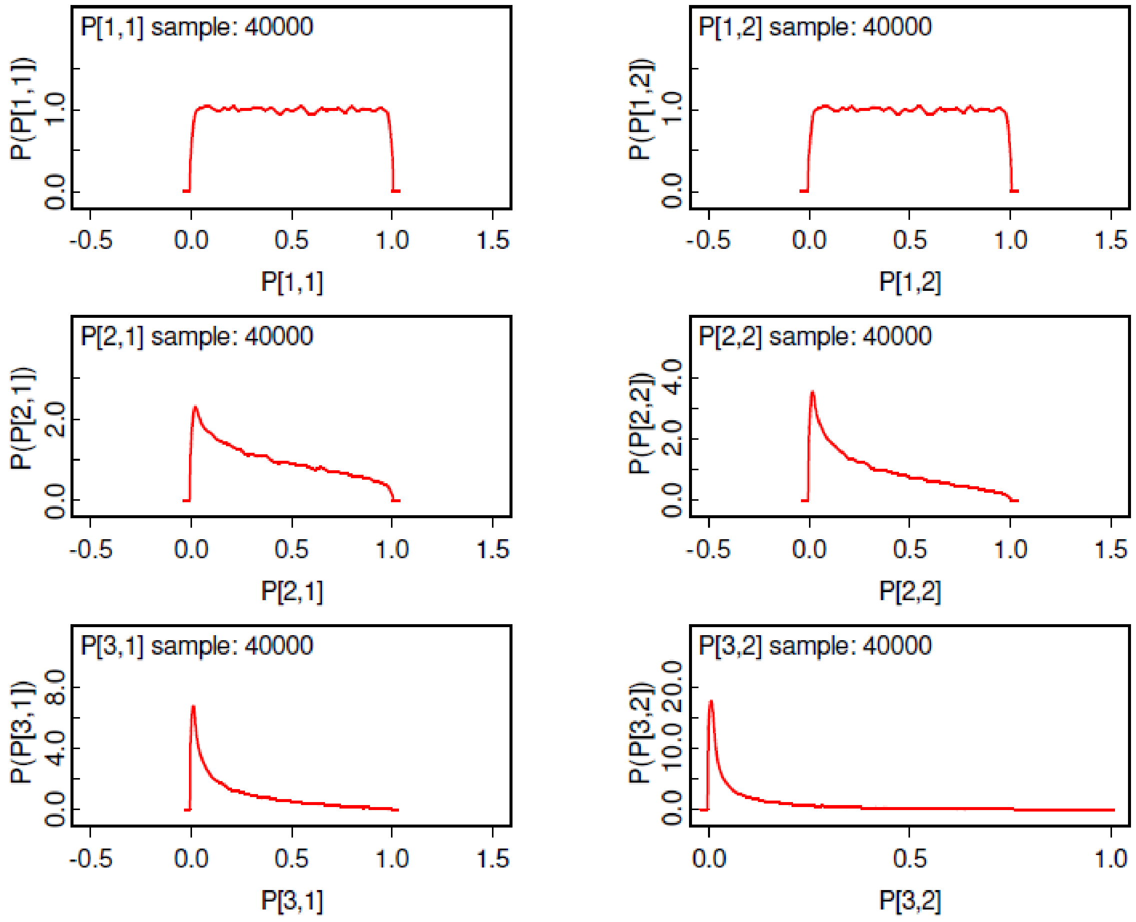
$\bar{X} = \mu_0 + \Delta\mu + Z\sigma/\sqrt{n}$ where $Z \sim \mathcal{N}\left(0,1\right)$

$$P = \Pr\left(T_{H_0} \geq \underbrace{\sqrt{n}\Delta\mu/\sigma}_{\delta} + Z\right) = 1 - \Phi(\delta + Z) = \Phi(-\delta - Z)$$

$$\boxed{\text{P-val} \mid \mathrm{H}_1 = \Phi\left[\mathcal{N}\left(-\delta,1\right)\right]}$$

$$\delta = 0 \Rightarrow \boxed{\text{P-val} \mid \mathrm{H}_0 = \Phi\left[\mathcal{N}\left(0,1\right)\right] \sim \mathcal{U}\left(0,1\right)}$$

# Distribution of p-values/suite




P[i,j]: P-value for i=1 (d=0) , 2 (d=0.10), 3 (d=0.5) and j=1 (N=20) and 2 (N=50)

# Deriving BF10 from p-values

$$H_0 : p \sim U(0,1)$$

Choice of a generic function

$$H_1 : f\left(p\right) \sim \text{"Standard power function distribution"}$$

$$H_1 : p \sim Beta(\xi,1),\ 0<\xi \leq 1$$

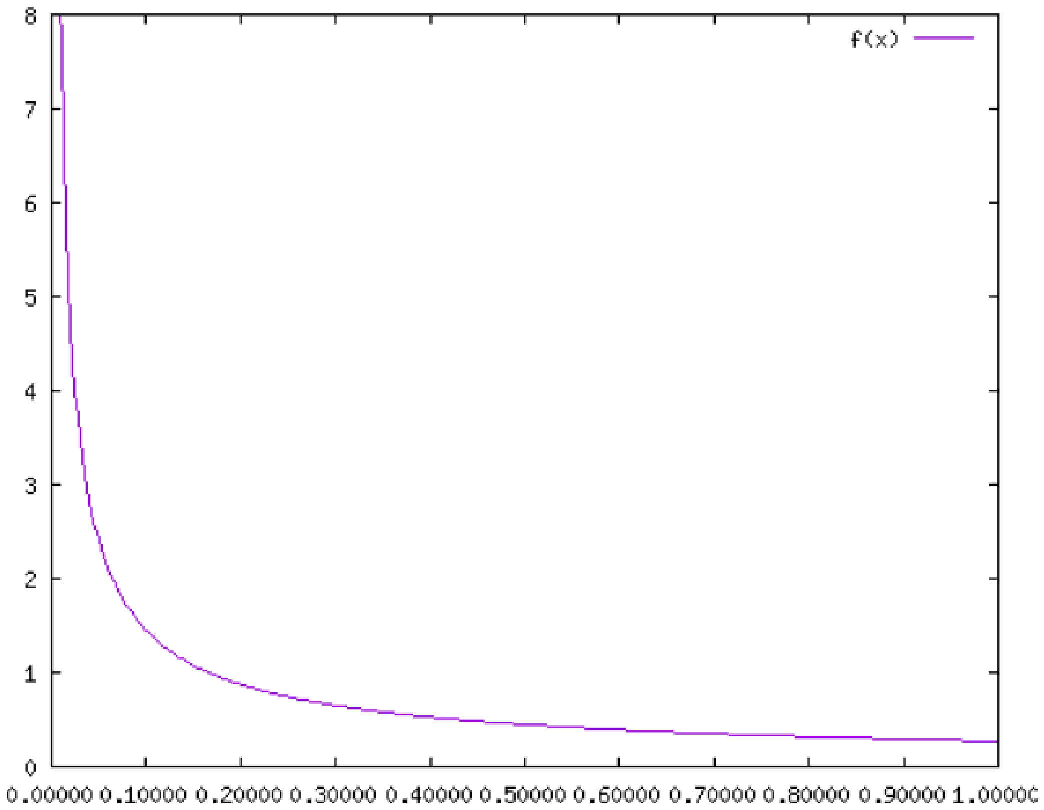
$$f\left(p \mid \xi\right) = \xi p^{\xi-1} \text{ (decreasing function)}$$

# Distribution of p-values under H1

You have entered : $f(x) = 0.27(x^{-0.73})$.

f(0.005) = 12.91563195995249156 . (click on a value to check its meaning in another window)

$$\int_{0.0}^{1.00} f(x)\mathrm{dx} = 0.99999999999999976313$$



# Upper bound of BF10/proof

$$BF_{10}(p) = \frac{\int_0^1 f\left(p \mid \xi\right)\pi\left(\xi\right)d\xi}{f\left(p \mid \xi = 1\right)}$$

$$0 \le \int_0^1 f\left(p \mid \xi\right)\pi\left(\xi\right)d\xi \le \left[\mathrm{Max}_{\xi}\, \mathrm{f}\left(p \mid \xi\right)\right]\underbrace{\int_0^1 \pi\left(\xi\right)d\xi}_{1}$$

# Upper bound of BF10/proof

Ici $f\left(p \mid \xi\right) = \xi p^{\xi-1}, 0 < \xi \leq 1 \Rightarrow \xi_{ML} = \min(-1/\log p, 1)$

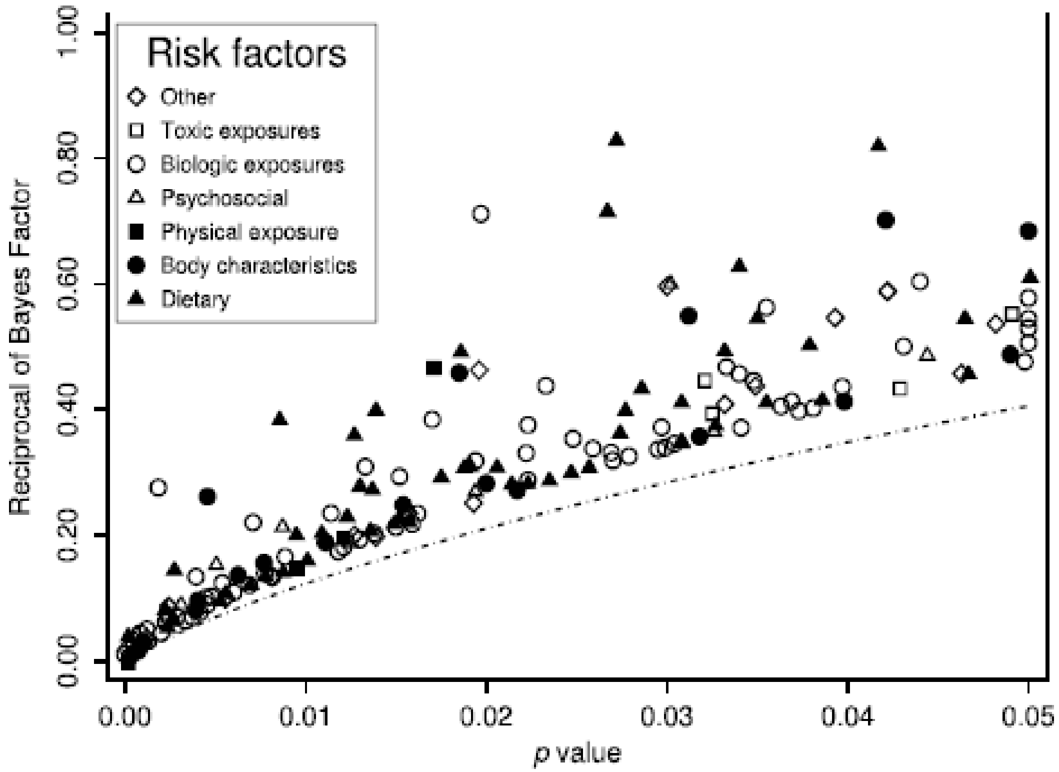
$$\boxed{BF_{10}\left(p\right) \leq BFB(p) = -\frac{1}{ep\log p}} \text{ for } p < 1/e \text{ , 1 otherwise}$$

$p = 0.05 \ \Rightarrow BFB_{10} = 2.46 \ \ (\Pr(\mathrm{H1} \mid \mathrm{p}) \leq 0.71$

BFB viewed as GLR (W. Edwards et al,1963)

# Checking BFB10(p) on real data

Bayarri et al, 2016, J Math Psycho : 272 studies (Ioannidis, 2008)



# BFB10(p)/Bias towards H1

THE AMERICAN STATISTICIAN
2020, VOL. 74, NO. 1, 101–102: Letter to the Editor
https://doi.org/10.1080/00031305.2019.1668850

Taylor & Francis
Taylor & Francis Group

Check for updates

LETTER TO THE EDITOR

**Benjamin, D. J., and Berger, J. O. (2019), "Three Recommendations for Improving the Use of *p*-Values", *The American Statistician*, 73, 186–191: Comment by Foulley**

- BFB10 favors H1
- Quantify amount of bias towards H1
- Test priors on ξ (Foulley, 2020, TAS,74:1, 101-102,)
  - Uniform
  - PC prior (Simpson & al, 2015)

# Upper bound of BF10: uniform prior

Jeffreys' prior $\pi_J(\xi) \propto \xi^{-1}$ improper

$\xi \in ]0,1[$ uniform

Analytical expression of $\int_0^1 f(p \mid \xi)\, d\xi$

$$\boxed{BFU_{10} = \frac{1}{\log p}\left(\frac{1-p}{p\log p} + 1\right)}$$

$p = 0.05 \Rightarrow BFU_{10} = 1.78$ $(\Pr(\mathrm{H1} \mid \mathrm{p}) \le 0.64$

# BFB vs BFU (p fixed)

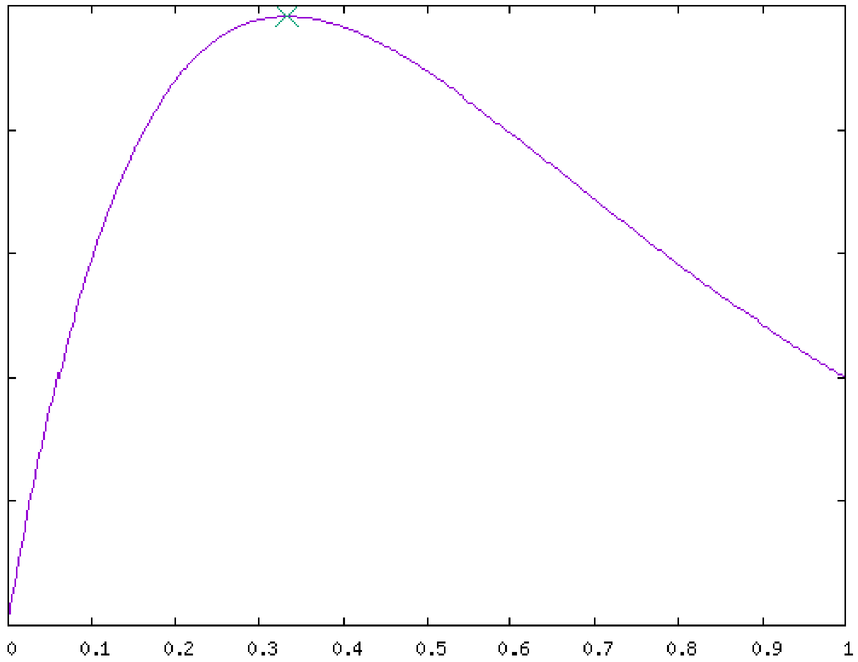
$$BFU_{10}(p)=\int_0^1 f\left(p\,|\,\xi\right)\pi\left(\xi\right)d\xi \quad \text{with} \quad f\left(p\,|\,\xi\right)=\xi p^{\xi-1} \quad Ex:\ p=0.05$$

1

You have entered : $f(x)=x\times 0.05^{x-1}$.

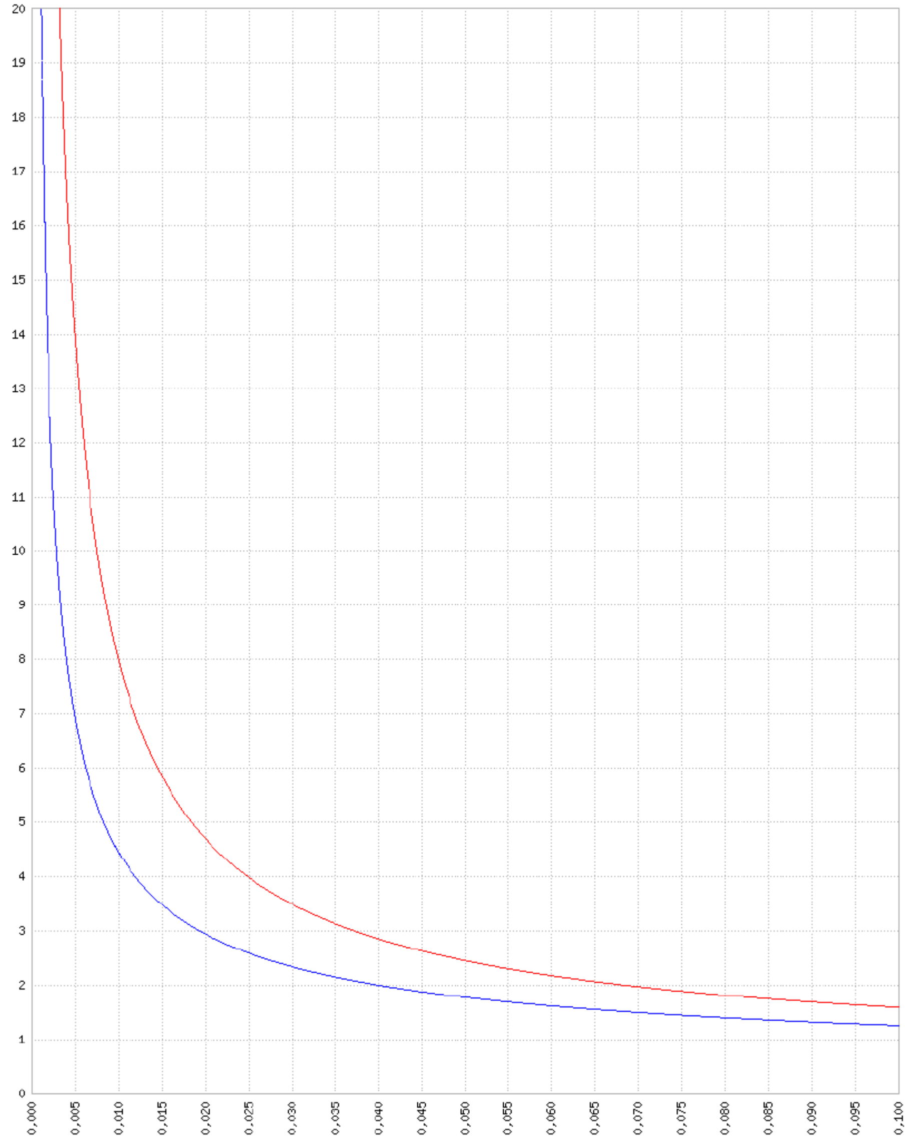
$$\int_{0.001}^{0.999} f(x)\mathrm{dx}=1.78231120307$$

Click on a root or an extremum of the curve, to compute its value up to 12 digits.

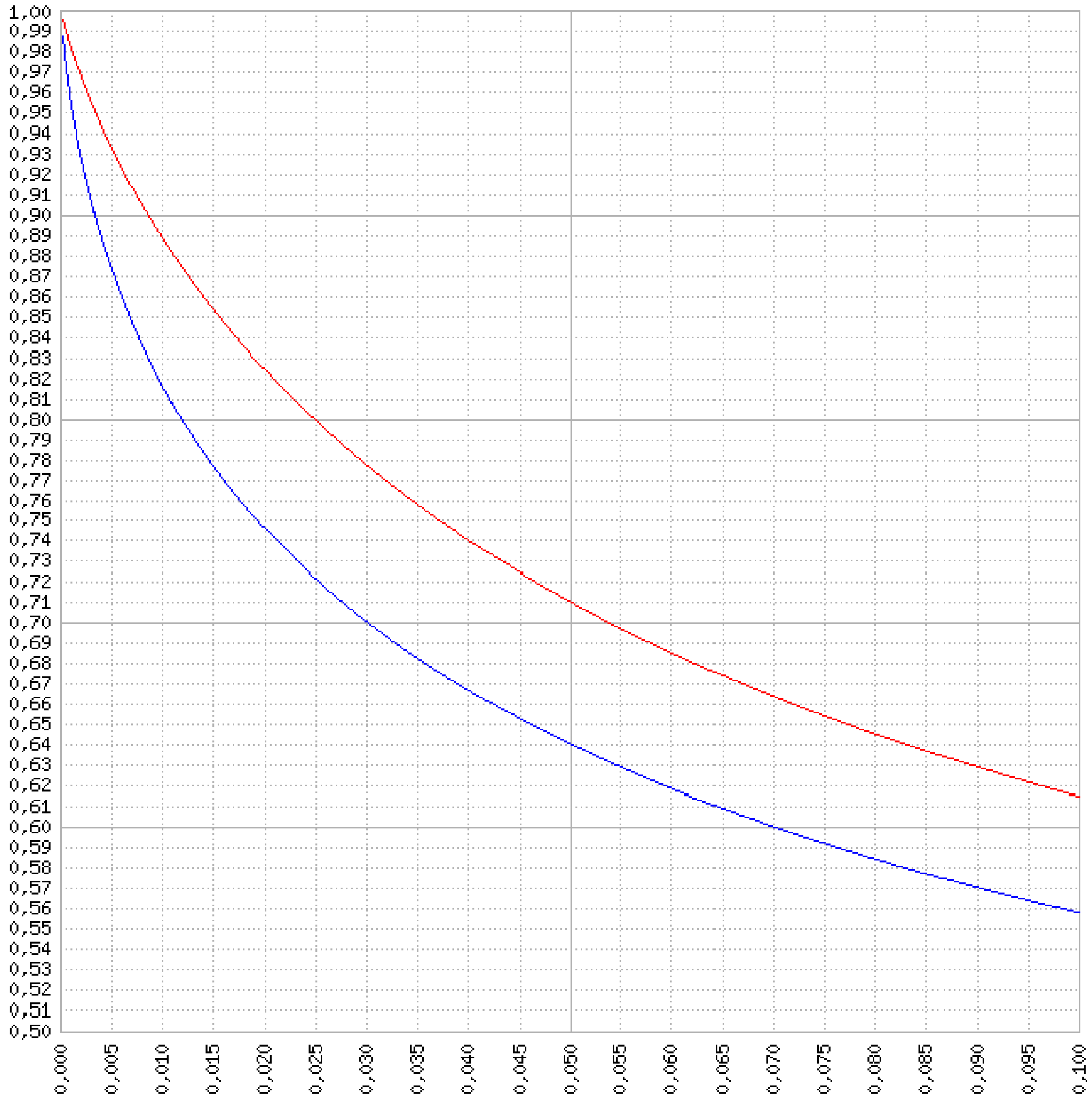



Maximum local : f(0.333808200695) = 2.4560234866. (click on a value to check its meaning

# BFB (red) vs BFU (blue)



# Pr(H1/p) bsup (red) vs Pr(H1/p) unif (blue)



# PC priors/Cont.

Looking for priors as decreasing functions of $\xi = 1\ (H_0)$ to $\xi \to 0\ (H_1)$

$\pi(\xi) = \lambda \exp\left[-\lambda D(\xi)\right]\left|D'(\xi)\right|$ où $D=\sqrt{2KL}$

$KL(\xi) = \ln \xi + \xi^{-1} - 1 \quad D'(\xi) = \xi^{-2}(\xi - 1)(2KL)^{-1/2}$

$\pi(\xi), f$ monotically increasing function for $\lambda \geq 4/3\ (r = 0.264)$

$\lambda = 4/3$ least unfavorable to $H_1$ among those penalising $H_1$

Numerical computation de $\int_0^1 f(p|\xi)\pi(\xi)d\xi$ via

*https://wims.auto.u-psud.fr/wims/*

# PC priors/Cont.

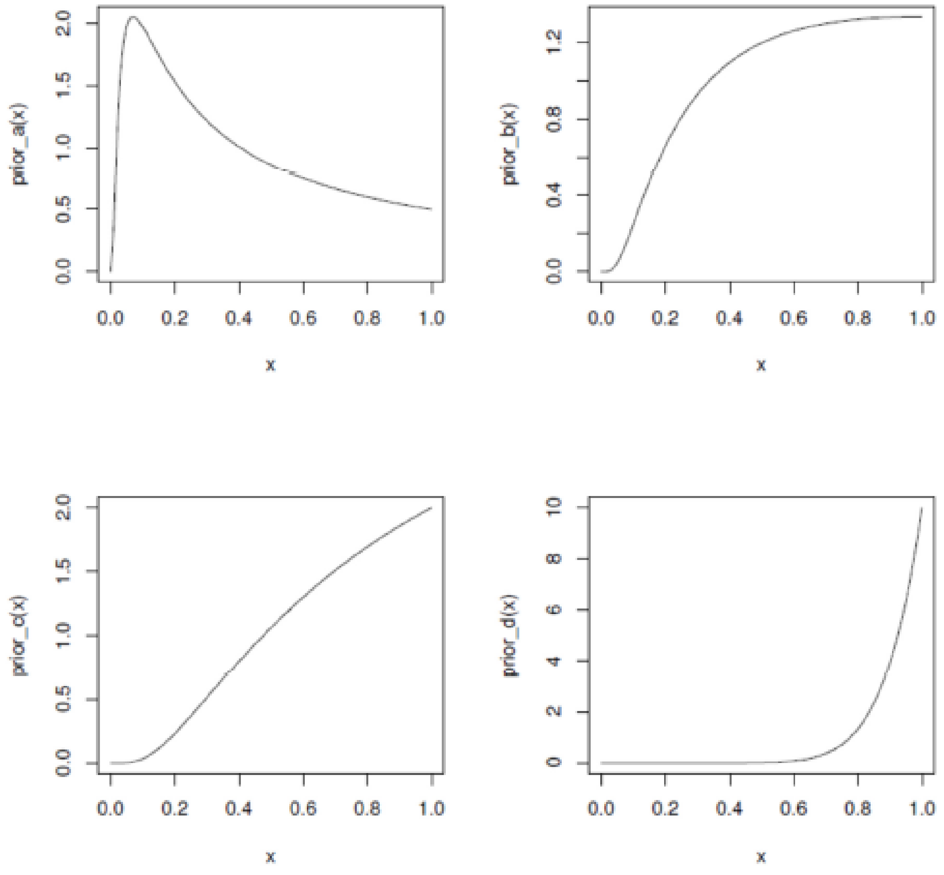



Figure 1. Examples of PC prior pdf's for the following $\lambda$ values: a=0.5, b=4/3; c=2 and d=10

# BF10 under PC priors

| p | 0.10 | 0.05 | 0.01 | 0.005 | 0.001 |
|---|---|---|---|---|---|
| **BFB**<br>**Pr(H1\|p)** | **1.60**<br>**0.62** | **2.46**<br>**0.71** | **7.99**<br>**0.89** | **13.9**<br>**0.93** | **53.3**<br>**0.981** |
| BFU<br>Pr(H1\|p) | 1.26<br>0.56 | 1.78<br>0.64 | 4.45<br>0.82 | 6.90<br>0.87 | 20.8<br>0.954 |
| BFP<br>Pr(H1\|p) | 1.36<br>0.57 | 1.84<br>0.65 | 4.04<br>0.80 | 5.89<br>0.85 | 15.3<br>0.939 |

Table 1. Bayes Factors and corresponding probabilities of the alternative hypothesis Pr(H1|p) under different prior distributions
BFB: (upper) bound based on generalized likelihood ratio by Benjamin & Berger (2019)
BFU: uniform prior
BFP: penalizing complexity prior with $\lambda$=4/3

# Checking the rule in standard testing situations

- Rule BFB(p) works well for
  - One & Two-sided z-tests
  - Two-sided t-tests with df≥10
- Rule plausible for Chi-squared tests

# Discussion

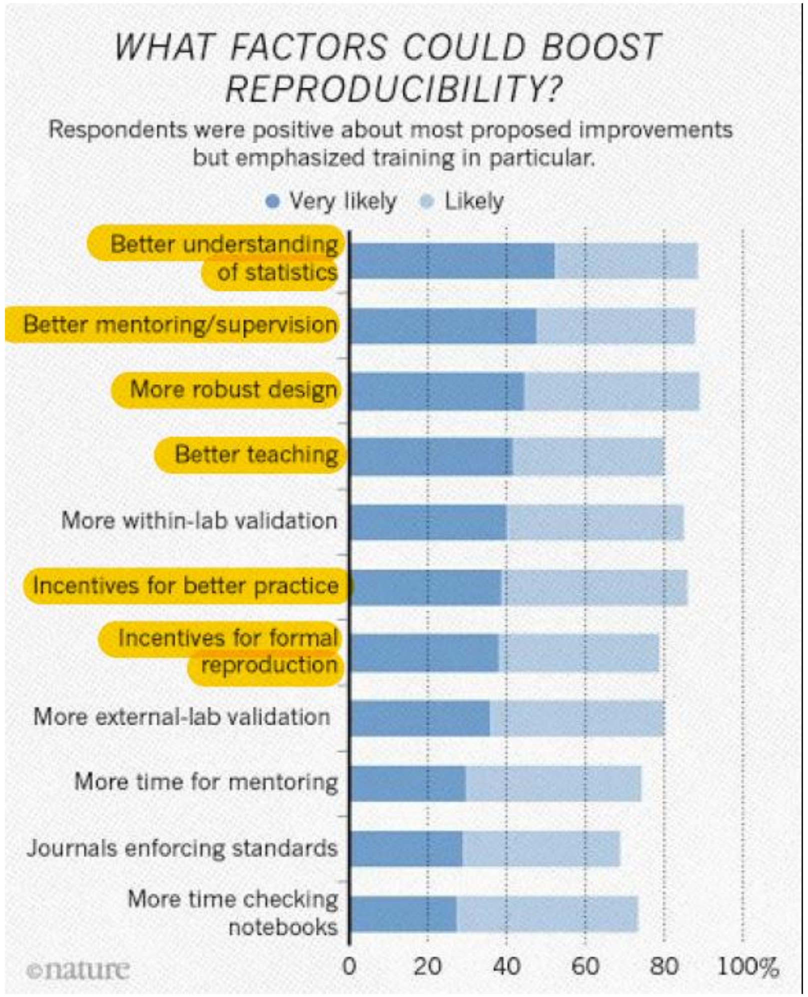
- Issues with BFB10 being **BF10 upper bounds**
  - OK if bounds small then do not reject H0; uncertainty otherwise
  - Bias towards H1 for BFB10, but BFB10 remains less optimistic wrt H1 than supposed from the p-value: good warning against H1 (adjustable according to choice of other priors)
- A fixed P-value does mean the same thing at different sample sizes (decreasing evidence with n incresing): not taken into account here . But from Good (1992)we can suggest $p^{*}=\min\left(p\sqrt{N/N_0},0.5\right)$

  or alternatively adjust directly lambda $\lambda^{*}=(4/3)(N/N_0)^{\alpha},0<\alpha<1$
- BFB can be implemented without a prior elicitation & is a good entry to a complete Bayesian analysis (Goodman, 1999)
- Choice of prior under H1 very influential on BF

# Discussion : other criteria

- Killeen replication and Lecoutre predictive probabilities (Lecoutre et al,2010)
- Posterior predictive p value (Gelman et al, 2004)
- Mixture models (Kamary et al, 2018)
- Analysis of credibility (AnCred, Matthews,2018)
- Severe testing (Mayo & Spanos, 2006)

# Conclusion


WHAT FACTORS COULD BOOST REPRODUCIBILITY?
Respondents were positive about most proposed improvements but emphasized training in particular.
Very likely
Likely
Better understanding of statistics
Better mentoring/supervision
More robust design
Better teaching
More within-lab validation
Incentives for better practice
Incentives for formal reproduction
More external-lab validation
More time for mentoring
Journals enforcing standards
More time checking notebooks
0
20
40
60
80
100%
©nature

# Appendix A: Kahneman & Tversky

$$1)\ \bar{X}_2 = \mu_0 + \Delta\mu + Z\sigma / \sqrt{N_2} \ \textit{where}\ Z \sim \mathcal{N}(0,1)$$

After standardisation

$$P = \underbrace{\frac{\bar{X}_2 - \mu_0}{\sigma}\sqrt{N_2}}_{Z_2} = \underbrace{\boxed{\frac{\Delta\mu\sqrt{N_1}}{\sigma}}}_{2.23\ if\ \Delta\mu=\Delta\hat{\mu}} \frac{\sqrt{N_2}}{\sqrt{N_1}} + Z$$

$$P = \Pr(Z_2 \geq 1.645) = \Pr\left[\mathcal{N}(0,1)\right] > 1.645 - 2.23/\sqrt{2} = 0.0682$$

$$\boxed{P = 0.4728}$$

$$2)\ \textit{In fact}\ \Delta = \frac{\Delta\mu\sqrt{N_1}}{\sigma} \sim \mathcal{N}\left(\frac{\Delta\hat{\mu}\sqrt{N_1}}{\sigma}, \frac{N_2}{N_1}\right)$$

$$P = \Pr\left(\Delta + \mathcal{N}(0,1) > 1.645\right) = 1 - \Phi\left(\frac{1.645 - \hat{\Delta}}{\sqrt{1 + Var\Delta}}\right)$$

$$P = 1 - \Phi\left(\frac{0.0682}{\sqrt{1+1/2}}\right) = 1 - \Phi(0.0556)$$

$$\boxed{P = 0.4778}$$

# Appendix B: Variance components

Consider a one way ANOVA model $y_{ij} = \mu + a_i + e_{ij}$ , $i = 1,\ldots,I, j = 1,\ldots,n_i$ where $a_i \sim_{iid} \left(0, \sigma_a^2\right)$, $e_{ijk} \sim_{iid} \left(0, \sigma_e^2\right)$ and $a_i \perp e_{ij}$ .

1) Specify the analytical expressions of $SSA = R(\mu, a) - R(\mu)$ and $SSE = \mathbf{y}'\mathbf{y} - R(\mu, a)$ as functions of $y_{ij}$, $y_{i0} = \sum_{j=1}^{n_i} y_{ij}$, $y_{00} = \sum_{i=1}^{I} y_{i0}$, and $N = \sum_{ij} n_{ij}$ .

2) Give the formulae for the expectations of $SSA$ and $SSE$ under the true random model and for the moment estimators of $\sigma_e^2$ and $\sigma_a^2$ based on the mean squares $MSA = SSA/(I-1)$ and $MSE = SSE/(N-I)$.

3) Now we consider the following quadratic forms of MINQUE-0, i.e., $SS_0 = \mathbf{y}'\mathbf{M}\mathbf{y}$ and $SS_1 = \mathbf{y}'\mathbf{M}\mathbf{Z}\mathbf{Z}'\mathbf{M}\mathbf{y}$ where $\mathbf{M} = \mathbf{I}_N - \mathbf{J}_N / N$, $\mathbf{I}_N$ being the identity matrix of order $N$ and $\mathbf{J}_N = \mathbf{1}_N \mathbf{1}_N'$, the $(N \text{x} N)$ matrix made of ones.

a) Specify the analytical expressions of $SS_0$ and $SS_1$ as functions of $N$, $\sum_{ij} y_{ij}^2$ and $y_{00} = \sum_{i=1}^{I} y_{i0}$ on the one hand and of $n_i$, $y_{i.} = y_{i0} / n_i$, and $y_{..} = \left(\sum_{ij} y_{ij}\right)/N$ on the other hand.

# Appendix B: Variance components (cont.)

b) How would you interpret $SS_0$ and how does $SS_1$ differ from $SSA$ ?

c) Derive the analytical expression of the coefficients of $\sigma_e^2$ and $\sigma_a^2$ in $E(SS_0)$ and $E(SS_1)$.

4) Compute $\hat{\sigma}_e^2$ and $\hat{\sigma}_a^2$ for ANOVA and MINQUE-0 from the following data set

| A-levels | $n$ | $\sum y$ | $\sum y^2$ |
|---|---|---|---|
| 1 | 7 | 40 | 274 |
| 2 | 11 | 95 | 909 |
| 3 | 6 | 48 | 416 |
| All | 24 | 183 | 1599 |

# App. B: Variance components (cont.)

Estimation of $\mathbf{p}'\boldsymbol{\sigma}^2$ from $\mathbf{y}'\mathbf{A}\mathbf{y}$

$\mathbf{p}' = (p_0, p_1, ..., p_K)$ $\boldsymbol{\sigma}^{2'} = (\sigma_0^2, \sigma_1^2, ..., \sigma_K^2)$

$E(\mathbf{y}'\mathbf{A}\mathbf{y}) = \boldsymbol{\beta}'\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} + \sum_k \mathrm{tr}(\mathbf{A}\mathbf{Z}_k\mathbf{Z}_k')\sigma_k^2$

1) $\mathbf{y}'\mathbf{A}\mathbf{y}$ translation invariant $\Rightarrow \mathbf{A}\mathbf{X} = \mathbf{0}$

2) $\mathbf{y}'\mathbf{A}\mathbf{y}$ unbiased $\Rightarrow \mathrm{tr}(\mathbf{A}\mathbf{Z}_k\mathbf{Z}_k') = p_k$

3) $\mathbf{y}'\mathbf{A}\mathbf{y}$ locally of minimum (norm) variance

$\mathrm{var}(\mathbf{y}'\mathbf{A}\mathbf{y}) = 2\mathrm{tr}(\mathbf{A}\mathbf{V}\mathbf{A}\mathbf{V})$ under normality

Pb reduces to minimise $\mathrm{tr}(\mathbf{A}\mathbf{V}^*\mathbf{A}\mathbf{V}^*)$

where $\mathbf{V}^* = \sum_k \mathbf{Z}_k\mathbf{Z}_k'\alpha_k$ ; $\alpha_k$ = "prior" value

Minimisation of 3) under constraints 1) et 2)

# Appendix B: Variance components/MIVQUE (cont.)

Equations same as in REML replacing **V** by $\mathbf{V}^*$

$$\boxed{\mathbf{F}^*\hat{\boldsymbol{\sigma}}^2 = \mathbf{g}^*}$$

$$\mathbf{F}^* = \left\{ f_{kl}^* = \mathrm{tr}\left( \mathbf{P}^*\mathbf{Z}_k\mathbf{Z}_k'\mathbf{P}^*\mathbf{Z}_l\mathbf{Z}_l' \right)\right\}$$

$$\mathbf{g}^* = \left\{ g_k^* = \mathbf{y}'\mathbf{P}^*\mathbf{Z}_k\mathbf{Z}_k'\mathbf{P}^*\mathbf{y} \right\}$$

$$\mathbf{P}^* = \mathbf{V}^{*-1}\left( \mathbf{I}_N - \mathbf{Q}^* \right) \text{where } \mathbf{Q} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}$$

$$\text{MINQUE} - 0: \ \mathbf{V}^* = \mathbf{I}_N\sigma_0^{2*}$$

$$\boxed{\mathbf{P}^* \text{ replaced by } \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'}$$

I-MINQUE = REML

# References


- Assaf AG, Tsionas M (2018) Bayes Factors vs P-values. *Tourism Management*, 67, 17-31
- **Benjamin DJ, Berger JO, (2020) Three recommandations for improving the use of p-values. The American Statistician, 73, 186-191**
- Benjamin DJ, Berger JO, Johannesson M et al (2017) Redefine Statistical Significance. *Nature Human Behaviour*
- Berger JO (1985) Statistical theory and Bayesian analysis. Springer
- Berger JO (2003) Could Fisher, Jeffreys and Neyman have agreed on testing. (with discussion), *Statistical Science*, 18, 1-32
- Berger JO, Sellke T (1987) Testing a point null hypothesis: the irreconcilability of P values and evidence. *JASA*, 82, 112-122
- Berger JO, Delampady M (1987) Testing precise hypotheses. *Statistical Science*, 2, 317-352
- Berger JO, Perrichi LR (1996) The intrinsic Bayes factor for model selection and prediction. *JASA*, 91, 109-122
- Berger JO, Boukai B, Wang Y (1997) Unified frequentist and Bayesian testing of a precise hypothesis (with discussion). *Statistical Science*, 3, 133-160
- Bernardo JM (1999) Nested hypothesis testing: the Bayesian reference criterion. *Bayesian Statistics*, 6, 101-130
- Biau DJ, Jolles BM Porcher R (2010) P value and the Theory of Hypothesis Testing. *Clin Orthop Relat Res* , 468,885-892
- Cumming G (2005) Understanding the average probability of replication: Comment on Killeen (2005). *Psychological Science*. 16, 1002–1004
- Cumming G, Fidler F (2009) Confidence intervals. *Journal of Psychology*, 217, 15-26.
- Cumming G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge, New York

# References/Cont.


- Edwards W, Lindman H, Savage LJ (1963) Bayesian statistical inference for psychological research. *Psychological Review*, 70, 193-242
- Foulley JL (2013) Le paradoxe de Jeffreys-Lindley: pierre dans le jardin des fréquentistes ou épine dans le pied des bayésiens. Applibugs, 19/12/2013, https://www.researchgate.net/publication/259575305_applibugs13_12_19JLF
- Foulley JL (2020) Comment on « Three Recommandations for Improving the Use of p-values ». *The American Statistician*,74:1,101-102 https://doi.org/10.1080/00031305.2019.1668850

  https://www.tandfonline.com/eprint/KAJWATZ25ZPZRFT2P3TM/full?target=10.1080/00031305.2019.1668850
- Gelman A, Shalizi CR (2013) Philosophy and the practice of Bayesian statistics. *British J of Mathematical and Statistical Psychology*, 66, 8-38
- Good IJ (1992) The Bayes/Non-Bayes Compromise: A Brief Review. JASA, 87, 597-606
- Goodman SN (1999) Toward evidence based statistics. 1: the P value fallacy. *Annals of Internal Medicine*, 130, 996-1004
- Goodman SN (1999) Toward evidence based statistics. 2: the Bayes Factor. *Annals of Internal Medicine*, 130, 1005-1013
- Greenland S, Senn SJ, Rothman K, Carlin JB, Poole C, Goodman SN, Altman DG (2016) Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, 31, 337-350
- Held, L., and Ott, M. (2018), On p-values and Bayes Factors. *Annual Review of Statistics and Its Application*, 5, 393-519.
- Hubbard R, Bayrarri MJ (2003) Confusion over measures of evidence (p's) versus errors (α's) in classical testing. *The American Statistician*, 57, 171-178
- Hung HM, O'Neill RT, Bauer P, Kohne K (1997) The behavior of the P-value when the alternative hypothesis is true. *Biometrics*, 57, 11-22
- Ioannidis JPA (2005) Why most published research findings are false. *PloS Medicine*, 2, (8), 696-701
- Jeffreys H (1961) Theory of probability (3rd edition) Oxford-Clarendon Press
- Johnson VE (2013) Revised standards for statistical evidence. PNAS, 110 (48);19313-1931
- Kamary K, Mengersen K, Robert CP, Rousseau J (2014) Testing hypotheses via a mixture model; arXiv
- Kass RE, Raftery AE (1995) Bayes factors. *JASA*, 90, 773-795
- Killeen PR (2005) An alternative to null hypothesis tests. *Psychological Science*, 16, 345-353

# References/Cont.


- **Lecoutre B, Poitevineau J (2014) The significance test controversy revisited: the fiducial Bayesian alternative. Springer Briefs in Statistics**
- Lecoutre B, Lecoutre M-P, Poitevineau J (2010) Killeen's probability of replication and predictive probabilities: How to compute, use and interpret them. *Psychological Methods*, 15, 158-171.
- Lindley DV (1957) A statistical paradox. *Biometrika*, 44, 187-192
- Mayo DG, Spanos A (2006) Severe testing as a basic concept in a Neyman-Pearson philosophy of induction. *British J of Philosophy of Science*, 57, 323-357
- Rougier, J. (2019), p-values, Bayes Factors, and Sufficiency, *The American Statistician*, 73: sup1, 148-151.
- Royall RM (1986) The effect os sample size on the meaning of significance tests. *The American Statistician*, 40, 313-315.
- Sellke, TM. (2012) On the interpretation of p-values. Technical report #17-01, Department of Statistics, Purdue University.
- **Sellke, T., Bayarri, M.J., and Berger, J.O. (2001), Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55, 62–71.**
- Senn S (2001) Two cheers for P-values. *J of Epidemiology & Statistics*, 6, 193-204
- Spanos A (2013) Who should be afraid of the Jeffreys-Lindley paradox? *Philosophy of Science*, 80, 73-93
- Simpson, D.P., Rue, H., Martins, T. G., Riebler, A., Martins. T. G., and Sørbye, S.H. (2015), Penalising model component complexity: A principled, practical approach to constructing priors. arXiv:1403.4630v4
- Spiegelhalter DJ, Abrams KR, Myles JP (2004) Bayesian approaches to clinical trials and health-care evaluation. J Wiley &Sons.
- Tversky A., Kahneman, D (1971) Belief in the law of small numbers. *Psychological Bulletin*, 6, 105-110.
- Vovk V . (1993) Mellin transforms and asymptotics: harmonic sums. *Journal of the Royal Statistical Society*, B, 55, 317-351
- Wagenmakers EJ (2007) A practical solution to the pervasive problem of p values. *Psychonomic Bulletin & Review*, 14, 779-804