

# The Statistical Tool Box: Are the tools calibrated?

D A S Fraser  
Statistical Sciences  
University of Toronto

Imperial College  
2019 March 22

12 Slides:

[www.utstat.utoronto.ca/dfraser/documents/Imperial2019](http://www.utstat.utoronto.ca/dfraser/documents/Imperial2019)

- 1 Brief history
- 2 Exponential models
- 3 Interest parameters
- 4 Scalar interest
- 5 General models
- 6 Mysterious Bayes
- 7 Location Normal - linear
- 8 Location Normal - curved
- 9 What can Bayes do?
- 10 Gamma model
- 11 Gamma data
- 12 References

## Brief history

Model - by prob. density fn  $f(y; \theta)$  Data  $y^o$  get  $f^o(\theta) = f(y^o; \theta) = L(\theta)$   
- by data-gen. fn  $y = y(z; \theta)$  get  $y^o = y(z; \theta)$  Invert

1763 Bayes Use  $\theta \sim \pi(\theta) L(\theta)$ , "some"  $\pi(\theta)$

1812 Laplace: Endorsed (didn't believe!) ; sought "Principle"; Non-Info.

1930 Fisher Use  $y^o = y(z; \theta)$  fiducial ...!

1937 Neyman: Used "Fisher" but fixed sets for  $z$ : conf (new name)

From tradition: Standardized departure

$$z = (\text{obs} - \text{exp}) / \text{SD} \quad \text{cf } \phi(z), \Phi(z) \quad \text{Exploratory} \quad \begin{matrix} \text{-bias} \\ \text{-variance?} \end{matrix}$$

Basis and essence of all generalizations

but needs more in model structure!

Generalize: Exponential models

$$f(y; \theta) = \exp\{\varphi(s - h(\varphi))\} h(s)$$

$$\varphi = \varphi(\theta)$$

Exponential tilt

$$s = s(y)$$

Continuous, Discrete

Advantage: SP

$$g(\theta; \varphi) = c e^{-n^2/2} |\hat{J}_{\varphi\varphi}|^{-1/2}$$

$$-n^2/2 = \ell(\varphi) - \ell(\hat{\varphi})$$

log lik stat

$$\hat{J}_{\varphi\varphi} = -\ell_{\varphi\varphi}(\hat{\varphi})$$

obs info

Contains full 3rd-order accuracy

Reid 1988

Uses familiar basic stat. quantities

Power

Interest in:  $\psi(\theta)$  (dim d) from  $\theta$  (dim p) Canonical wlog

In general: nuisance  $\lambda \dots$  "affects conditioned variable"

use Laplace integration

$$g(s, \psi) = C e^{-R^2/2} \left[ \int \int_{\psi \psi} \left| \frac{\partial \tilde{J}(\lambda \lambda)}{\partial \lambda \lambda} \right|^{1/2} \right]$$

$-R^2/2 = \ell(\hat{\psi}_+) - \ell(\hat{\psi})$

$\tilde{J}_{\lambda \lambda} = \text{inference } \lambda, \text{ rescaled}$

"Continuity" gives uniqueness . F2014

Have: full resolution: Marginal variable; density!

Will use!  $e^{\text{logLik}} [\dots]$  Reciprocal of [] later

Scalar Interest  $\psi$  ... widely of interest  
Here:  $\psi$  canonical

Want distr fin from  $c e^{-n^2/2} \left[ |\hat{J}_{\psi\psi}|^{-1/2} |\tilde{J}_{(\lambda\lambda)}|^{1/2} \right]$  ← density

Integrate by parts  $G(s,; \psi) = \Phi(n - n' \log \frac{n}{g})$   
Work to 3rd

$$n = n_\psi \text{ from } \ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)$$
$$g = (\hat{\psi} - \psi) [ ]^{-1}$$

Uses Barndorff-Nielsen (1991)

Widely: all needed accuracy for applications

get  $p(\psi)$  .. p-value function

FRW (1999)  
F (2019)

- never been beaten
- available in tool box
- well calibrated
- unique 3rd order

FRW 1999  
F 2019

And for general models: regularity

There is Exponential model, tangent at data, gives 3rd accuracy.

Just get direction of  $\theta$ -effect on sample space

Use data generating version of model  $y = y(z; \theta)$

$$\frac{\partial y}{\partial \theta} \Big|_{\hat{z}^0} = V \quad \frac{\partial \ell(\theta; y)}{\partial V} \Big|_{y^0} = \varphi(\theta) \quad \text{FR1995 FRW 1999}$$

Then differentiate  $\ell(\theta; y)$  in directions  $V$  gives canonical  $\varphi(\theta)$

Treat as "Exponential" ... full 3rd!

Ex:  $\ell(\theta; y) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum (y_i - \mu)^2 \Rightarrow \varphi = \left( \frac{\mu}{\sigma^2}, \frac{1}{\sigma^2} \right)$  or equiv

Ex: Box & Cox  $\hat{y} = X\beta + \sigma z \Rightarrow \varphi \dots$

Get p-value for  $\lambda$  or any other parameter

Ex's wealth in literature, in toolbox | Forget sufficiency  
High simulation accuracy! It's calibrated! | Doesn't work!

## Mysterious Bayes

Have observed Model:  $f(y^o; \theta) = L(\theta)$  .... no quarrel ... OK!

Append a  $\pi(\theta)$ ?  $\pi(\theta)L(\theta)$  Many major math. said "arbitrary"!

Laplace found it often "worked" but didn't know why?

Went looking for a principle: Non-informative; not convinced!

Try "Data-generating" model:  $y^o = y(z; \theta)$  Invert Fisher confidence  
Neyman added "Fixed set for  $z$  (Changed name; who gets credit?)  
Currently being undone:

BFF! Confidence sets, distributions: try anything: libertarian!

Anything goes

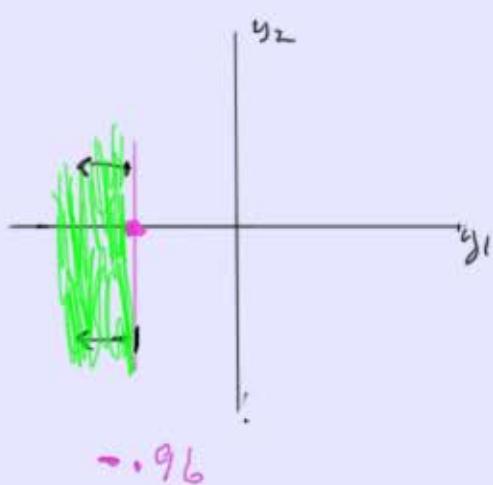
Tool box in trouble

Cautionary tale but there is

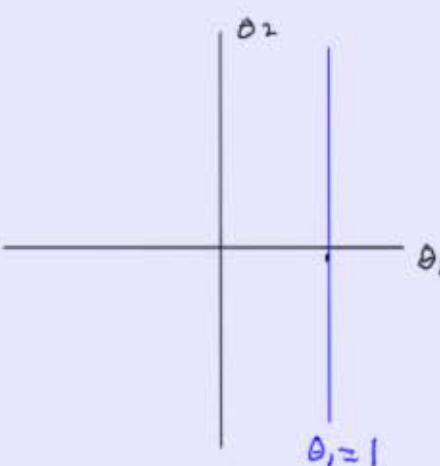
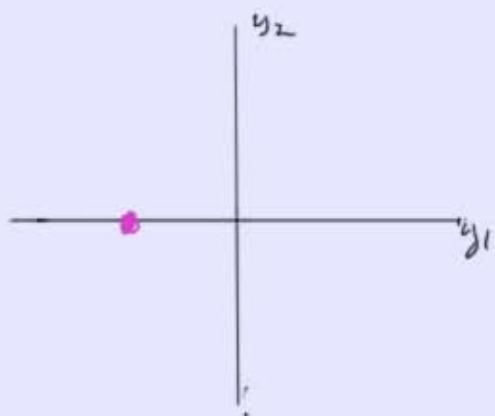
Calibration!  
(Confidence!)

Berger 2006  
Goldstein 2006

Example: Location Normal, generic 1st order. Linear parameter



$$\rho = 2.5\%$$



Test  $\theta_1 = 1$  linear  
Data  $y_1 = -0.96$   
p-value  $\rho = 2.5\%$



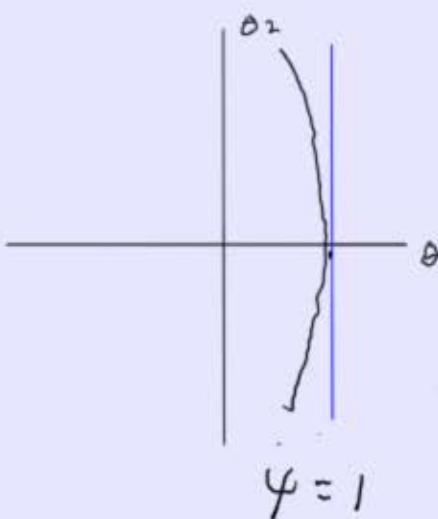
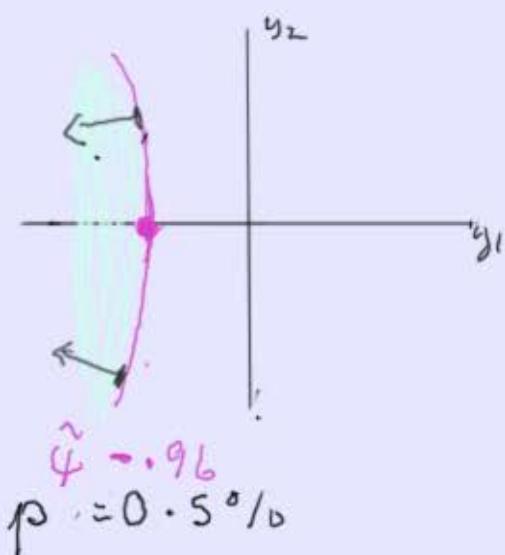
Try Bayes

Distribution for  $\theta$  centered at Data

Bayes survivor  $\Delta = 2.5\%$

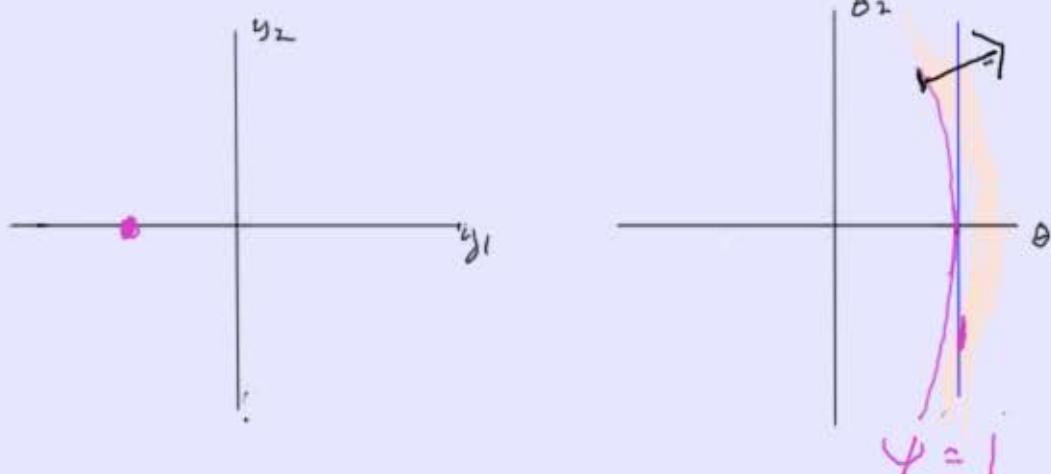
Equal  $2.5\% = 2.5\%$   
 $\beta$  linear works!

Example: Location Normal / Generic 1st order Curved



Test  $\psi = 1$   
Data  $\hat{\psi} = -0.96$   
p-value  $p = .5\%$

Try Bayes

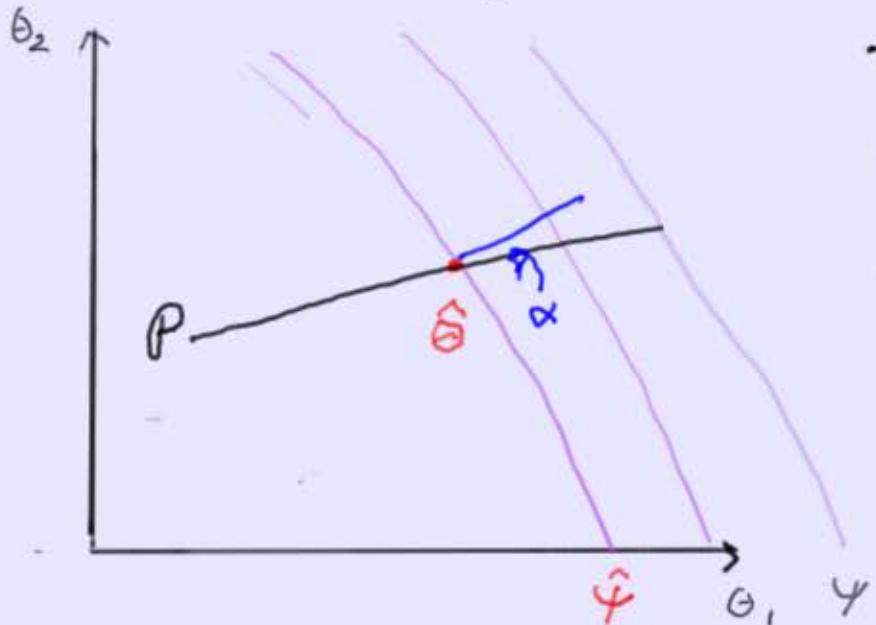


Distribution for  $\theta$  centered at Data  
Bayes survivor  $\Delta = 4.7\%$

$.5\% \neq 4.7\%$

Bayes can't handle curvature! Dawid SZ 1973

What can Bayes do? Try scalar  $\psi$  with vector  $\theta$



$$\text{Test } \psi(\theta) = \psi$$

Data  $\hat{\theta}$

$$\text{Try Jeffreys: } \pi(\theta) \propto |\mathcal{J}_{\theta\theta}(\theta)|^{1/2} d\theta$$

Known to be bad!

So?

1) Calculate profile contour  $P = \{\psi = \hat{\psi}_P\}$

2) Use p-dim Jeffreys on scalar profile, on 1-dim

3) Add a Jacobian  $\cos \alpha$

4) get 2nd order Bayes  $|\mathcal{J}_{\theta\theta}(\theta)|^{1/2} \cos \alpha d\theta$

Has NOT been available otherwise

Bedard F 2019

Ex: Gamma model; Positive arcsis data

$$\frac{\beta^\alpha}{\Gamma(\alpha)} y^\alpha e^{-\beta y} \cdot \frac{dy}{y} = \exp\left\{\alpha \ln y - \beta y - \alpha \ln \beta - \ln \Gamma(\alpha)\right\} \cdot d \ln y$$

$$\mu = \frac{\alpha}{\beta} \quad \sigma^2 = \frac{\alpha}{\beta^2} = \underline{\psi} = \text{Interest curved} \quad \text{Say } \lambda = \beta$$

Data: .20 .45 .78 1.28 2.28

Brazzaile et al 2007

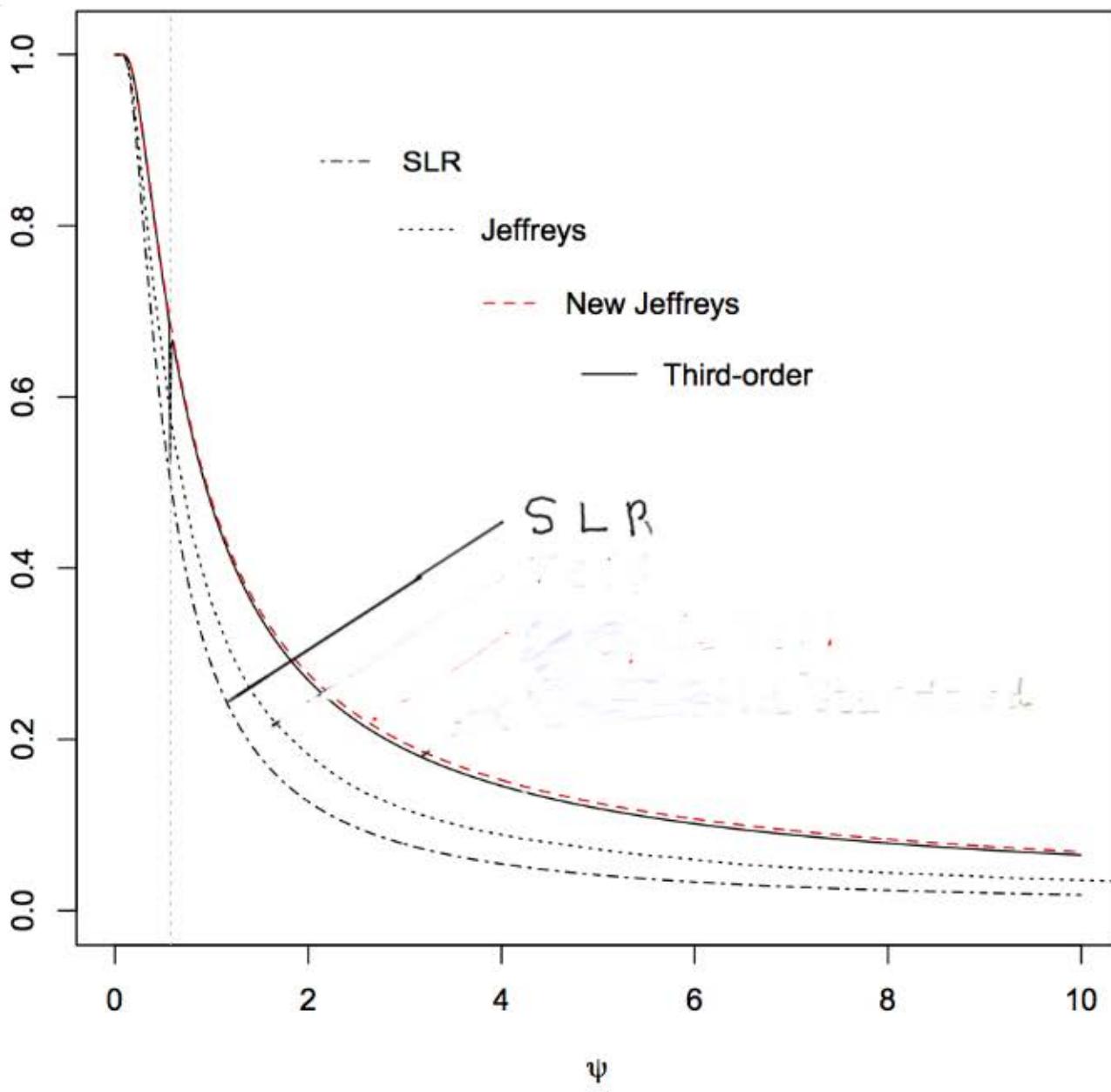
Get:  $\ell(y, \lambda)$  mle's info's

$$\text{Jeffreys (on profile) } \dots \quad \pi_J(y, \hat{\lambda}_y) = \frac{n}{\hat{\beta}_y} \left\{ \hat{\alpha}_y D(\hat{\alpha}_y)^{-1} \right\}^{1/2}$$

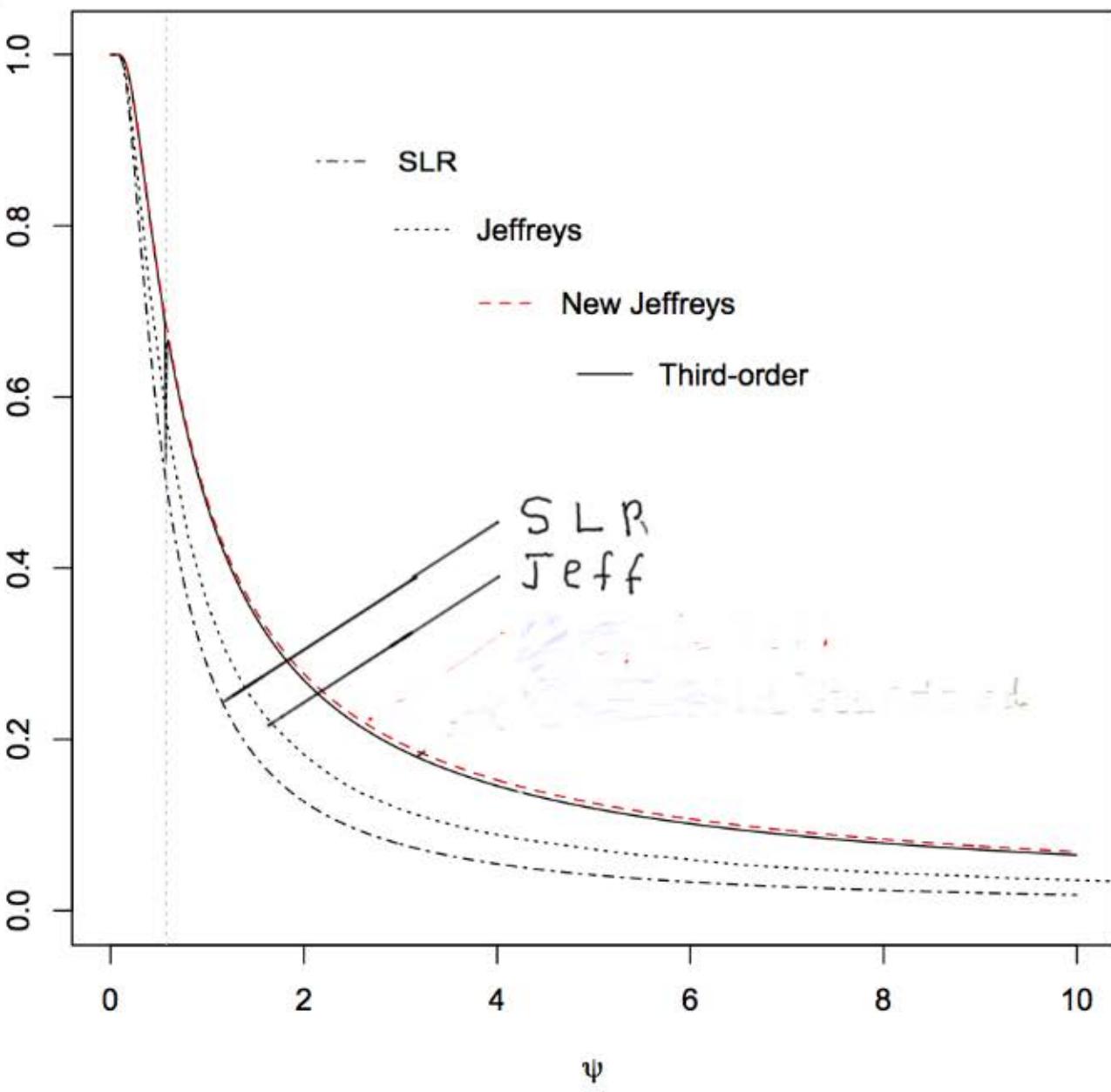
Compute, MCMC

Bédard F (2019)

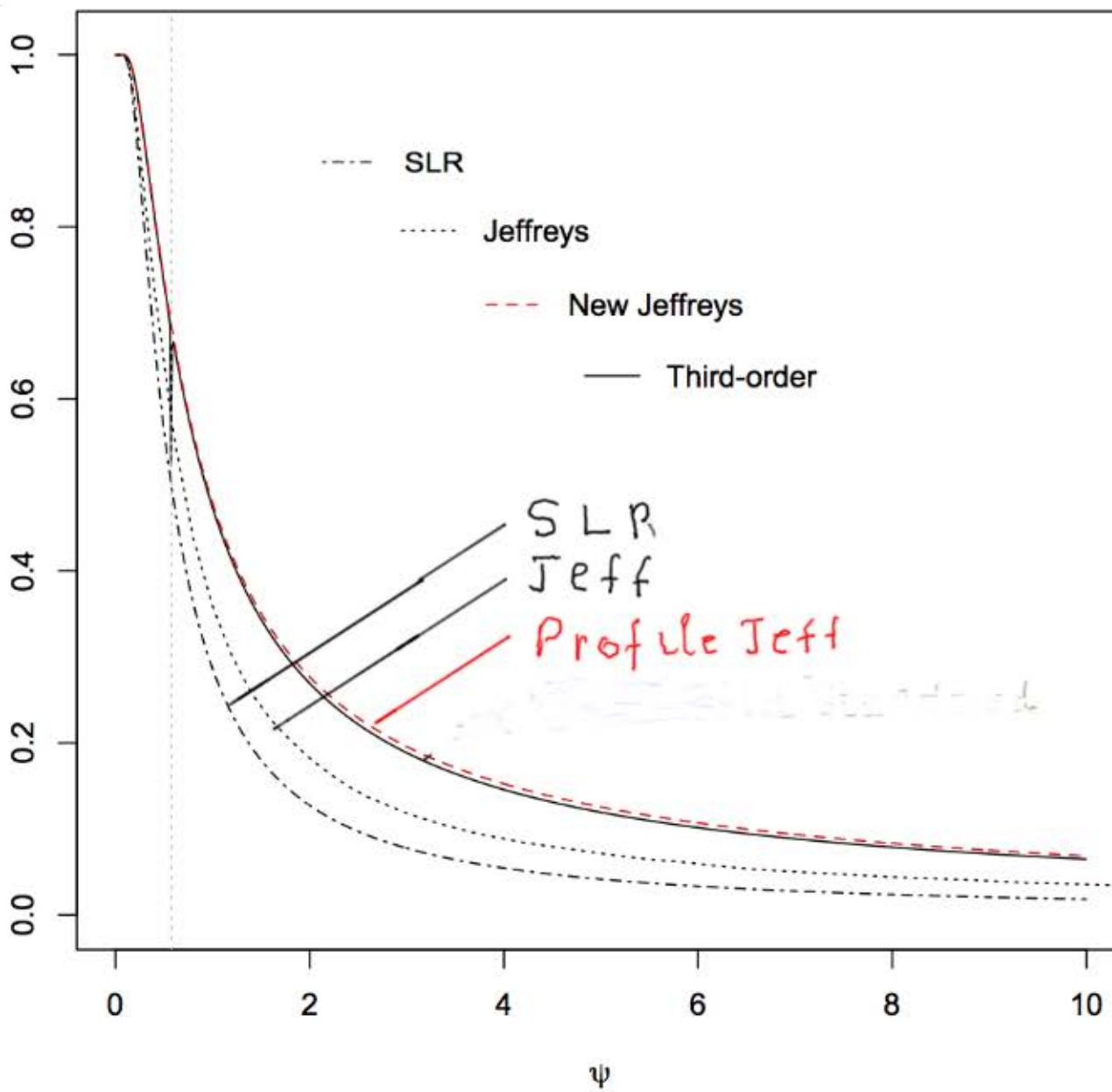
$p(\psi)$



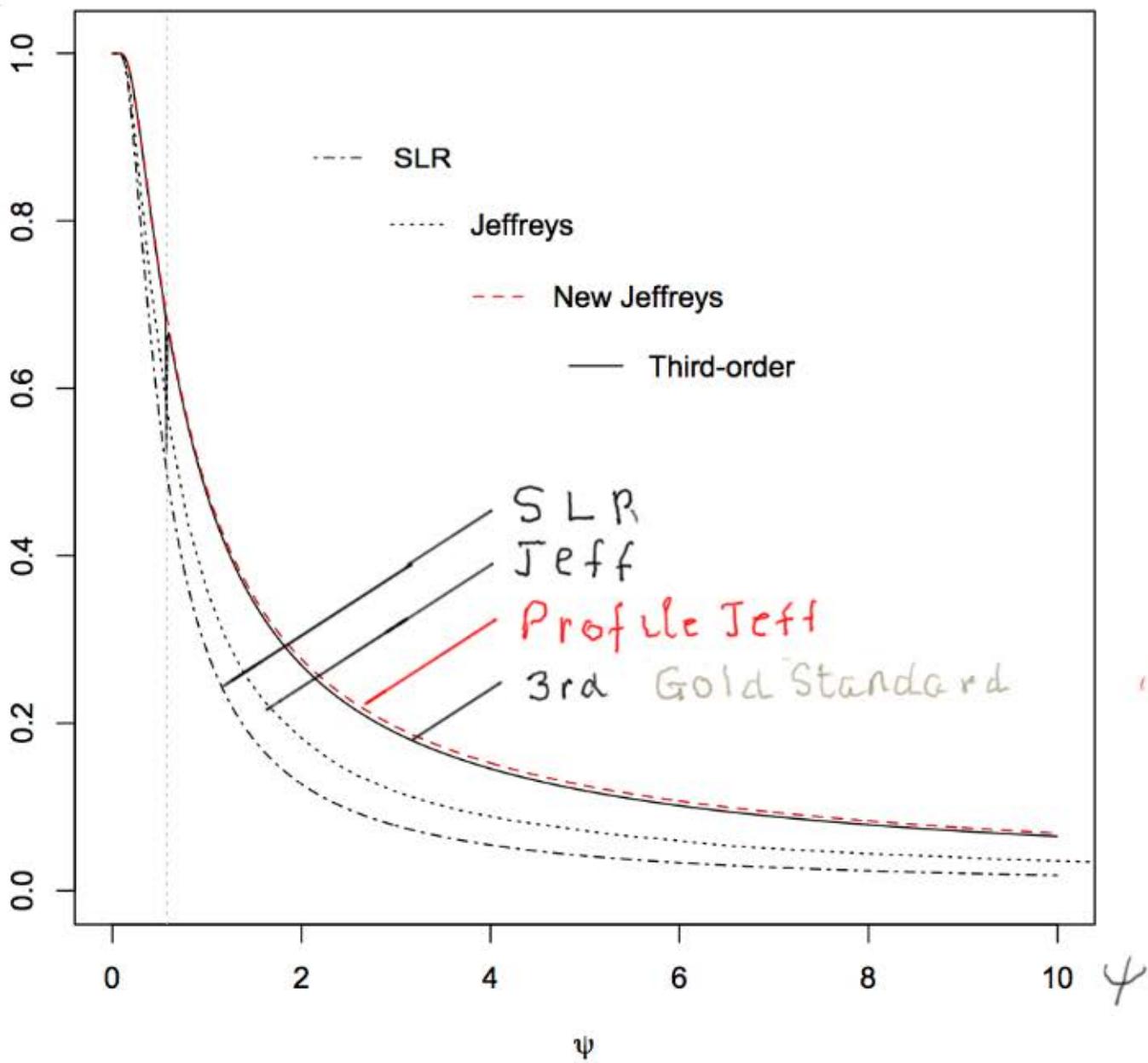
$p(\psi)$



$p(\psi)$



$p(\psi)$



Thank you

## REFERENCES

- Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio. *Biometrika* 78, 557–563.
- Dawid, A. P., M. Stone, and J. V. Zidek (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. B* 35, 189–233.
- Fraser, D. and M. Bedard (2019). Can bayes be calibrated?
- Fraser, D. A. S. (201). The *p*-value function and statistical inference. *The American Statistician* in press, 153–165.
- Fraser, D. A. S. (2011). Is Bayes posterior just quick and dirty confidence? (with discussion). *Statistical Science* 26, 299–316.
- Fraser, D. A. S. (2014). Why does statistics have two theories? In X. Lin, C. Genest, D. L. Banks, G. Molenberghs, D. W. Scott, and J.-L. Wang (Eds.), *Past, Present and Future of Statistical Science*, pp. 237–252. Florida: CRC Press.
- Fraser, D. A. S. and N. Reid (1995). Ancillaries and third order significance. *Utilitas Mathematica* 47, 33 – 53.
- Fraser, D. A. S., N. Reid, and J. Wu (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika* 86, 249–264.
- Reid, N. (1988). Saddlepoint methods and statistical inference. *Statistical Science* 3, 213–238.