

Parameter curvature and the Bayesian frequentist divergence

D A S Fraser
Statistics
University of Toronto

www.toronto.edu/dfrazer/documents/warwphi8.pdf

Overview

- 1 Background:
Output conflicts: Time line
Input conflicts: Poisson example
Issues?

- 2 Scalar case:
 $f(y - \theta)$ Linear
 $f(y; \theta)$ Curved
The Neyman diagram

- 3 Curved models
 $N(\theta; I)$ $\theta > 0$... 1st order
 $N(\theta; I)$ $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$... 2nd order
 $N(\theta; I + \sigma\theta^2/2n)$... 3rd order

- 4 The B Paradigm
The claim that posterior has status as probability

- 5 The Theorem
The claim that a prior is needed for optimality

- 6 Message / Summary

1 Background:

i) Two outputs... time line

1763 Bayes $f(y|\theta)$

$\Rightarrow \pi_B(\theta | y^o)$

(Linearity
invariance

1774 Laplace endorsed

authority

Inverse prob.

1922 Fisher $L(\theta)$

+ 150 years

1930 Fisher Confidence

\Rightarrow

$\pi_F(\theta | y^o)$

(pivotal inversion
"Bayes wrong")

1958 Lindley

" π_F wrong"

Here: Bayes 1763: Invariance, non informative, default,

NOT: Subjective probs ... Rama

Not: Real probs (like there is a random source for θ)

... use it if appropriate -- not Bayes

1 Background:

i) Two outputs

1763 Bayes $f(y|\theta)$

$\Rightarrow \underline{\pi_B(\theta | y)}$

(translation
invariance)

1774 Laplace endorsed

authority

1922 Fisher $L(\theta)$

150 years later

1930 Fisher Confidence

\Rightarrow

$\underline{\pi_F(\theta | y)}$

(pivotal inversion
"Bayes wrong")

1958 Lindley

" π_F wrong"

1946 Jeffreys

1963 Welch Peers

1964 Box Cox

1973 Dawid Stone Zidek

1979 Bernardo

2000 Wasserman

1 Background:

i) Two outputs

1763 Bayes $f(y|\theta)$

$\Rightarrow \pi_B(\theta | y^\circ)$

(translation
invariance)

1774 Laplace endorsed

authority

1922 Fisher $L(\theta)$

150 years later

1930 Fisher Confidence

\Rightarrow

$\pi_F(\theta | y^\circ)$

(pivotal inversion
"Bayes wrong")

1958 Lindley

" π_F wrong"

Two versions of $\pi(\theta | y^\circ)$

Both can't be right

coexist

Q's

2) Two different inputs • Likelihood or No likelihood

Ex: Poisson(θ)

Known $\theta \geq \theta_0$ θ_0 = Background
Want To detect $\theta > \theta_0$ Nobel, LHC

Physicist: Mendelkern in Stats journal Kendall & Stuart

Discussion: 5 statisticians Central confidence intervals

Mendelkern (2002) Statistical Science w Discussion

2) Two different inputs • Likelihood or No likelihood

Ex: Poisson(θ)

Known $\theta \geq \theta_0$ θ_0 = Background
 Want To detect $\theta > \theta_0$ Nobel, LHC

Physicist: Mendelkern in Stats journal / Kendall & Stuart

Discussion: 5 statisticians Central confidence intervals

- || almost nothing on Likelihood
- || " " " " p-value
- || little on Bayes

Mendelkern (2002) Statistical Science w Discussion

2) Two different inputs • Likelihood or No likelihood

Ex: Poisson(θ)

Known $\theta \geq \theta_0$ θ_0 = Background
Want To detect $\theta > \theta_0$ Nobel, LHC

Physicist: Mendelkern in Stats journal / Kendall & Stuart

Discussion: 5 statisticians Central confidence intervals

almost nothing on L
 " " " p-value
 little on Bayes

Rebuttal:

Need $L(\theta)$, $p(\theta)$

FRW (2004) but in Physics literature

Contradictions No common view

Mendelkern (2002) Statistical Science w Discussion
F & Reid Wong (2004) Physics Rev D

3) Issues

out a) $\pi^B(\theta y^\circ)$ $\pi^F(\theta y^\circ)$	- Not equal - Contradictory	Two outputs Two inputs
in b) Bayesian ... Use <u>only</u> $L^\circ(\theta)$ frequentist... Ignores $L^\circ(\theta)$		

flow come?

3) Issues

a) $\pi^B(\theta | y^\circ)$
 $\pi^F(\theta | y^\circ)$

Not equal
 Contradictory

b) Bayesian ... Use only $L^\circ(\theta)$
 frequentist... Ignores $L^\circ(\theta)$

c) Authorities Bayes

Laplace

Jeffreys

Disney

No continental drift

No "Big Bang"

Fortey, The Earth 2004

Amer. Scientist 2007

Back to statistics a) π^B vs. π^F

b) Only L vs. No L

c) Serious?

2 Scalar variable, parameter

a) Location / linear $f(y-\theta)$... original Bayes 1763

Bayes: Invariant prior $\pi(\theta) = C$

Posterior upper/survivor prob $S(\theta; y^*) = \int_{\theta}^{\infty} C f(y^* - \theta) d\theta$

2 Scalar variable, parameter

a) Location / linear $f(y-\theta)$... original Bayes 1763

Bayes: Invariant prior $\pi(\theta) = C$

$$\text{Posterior upper/survivor prob } S(\theta; y^*) = \int_{\theta}^{\infty} C f(y^* - \theta) d\theta$$

freq. %age position of data ... tell it straight!

$$\text{Probability left of data } p(\theta; y^*) = \int_{-\infty}^{y^*} f(y - \theta) dy$$

Equal: Same calculus integral: $S(\theta; y) = p(\theta; y)$

2 Scalar variable, parameter

a) Location / linear $f(y-\theta)$... original Bayes 1763

Bayes: Invariant prior $\pi(\theta) = C$

$$\text{Posterior upper/survivor prob } \alpha(\theta; y^*) = \int_{\theta}^{\infty} C f(y^* - \theta) d\theta$$

freq. % age position of data

Probability left of data

$$p(\theta; y^*) = \int_{-\infty}^{y^*} f(y - \theta) dy$$

Equal: Same calculus integral: $\alpha(\theta; y) = p(\theta; y)$

But $p(\theta; y)$ is just p-value

is Uniform (0, 1)

gives confidence intervals etc ...

2 Scalar variable, parameter

a) Location / linear $f(y-\theta)$... original Bayes 1763

Bayes: Invariant prior $\pi(\theta) = C$

Posterior upper/survivor prob $S(\theta; y^*) = \int_{\theta}^{\infty} C f(y^* - \theta) d\theta$

freq.: % age position of data

Probability left of data

$$p(\theta; y^*) = \int_{-\infty}^{y^*} f(y - \theta) dy$$

Equal: Same calculus interval: $S(\theta; y) = p(\theta; y)$

But $p(\theta; y)$ is just p-value ... is just $S(\theta; y)$
 is Uniform(0, 1)
 gives confidence intervals ...

If Linear/Location \Rightarrow Bayes is confidence

Maybe: Bayes is only confidence
at most !

b) Curved case $f(y; \theta)$

$$\begin{aligned} \text{freq} \quad p(\theta) &= \int_{-\infty}^{y^*} f(y; \theta) dy = F(y^*; \theta) \\ &= \int_{\theta}^{\infty} F_y(y^*; \theta) d\theta \end{aligned}$$

$$\begin{aligned} \text{Bayes} \quad s(\theta) &= \int_{\theta}^{\infty} c \cdot f(y^*; \theta) d\theta \\ &= \int_{\theta}^{\infty} \pi(\theta) F_y(y^*; \theta) d\theta \end{aligned}$$

Still equal?

... percentage position of data

$\left. \begin{array}{l} \text{Just derive, re } \theta \\ \text{Equality by calculus} \end{array} \right\} \text{Seen before?}$

b) Curved case $f(y; \theta)$

Usually not equal

freq $P(\theta) = \int_{-\infty}^{y^*} f(y; \theta) = F(y^*; \theta)$

 $= \int_{\theta}^{\infty} F_y(y^*; \theta) d\theta$

... % age position of data

Bayes $\pi(\theta) = \int_{\theta}^{\infty} c \cdot f(y^*; \theta) d\theta$

 $= \int_{\theta}^{\infty} \pi(\theta) F_y(y^*; \theta) d\theta$

If agree $\Rightarrow \pi(\theta) = \frac{-F_y(y^*; \theta)}{F'_y(y^*; \theta)} = \frac{d y}{d \theta} \Big|_{y^*}$

Quantile function

$y = y(u; \theta)$

(is inverse of $u = F(y; \theta)$)

= Data-dependent prior

Lindley's calculations reversed
Bayes has confidence property ONLY
if location model.

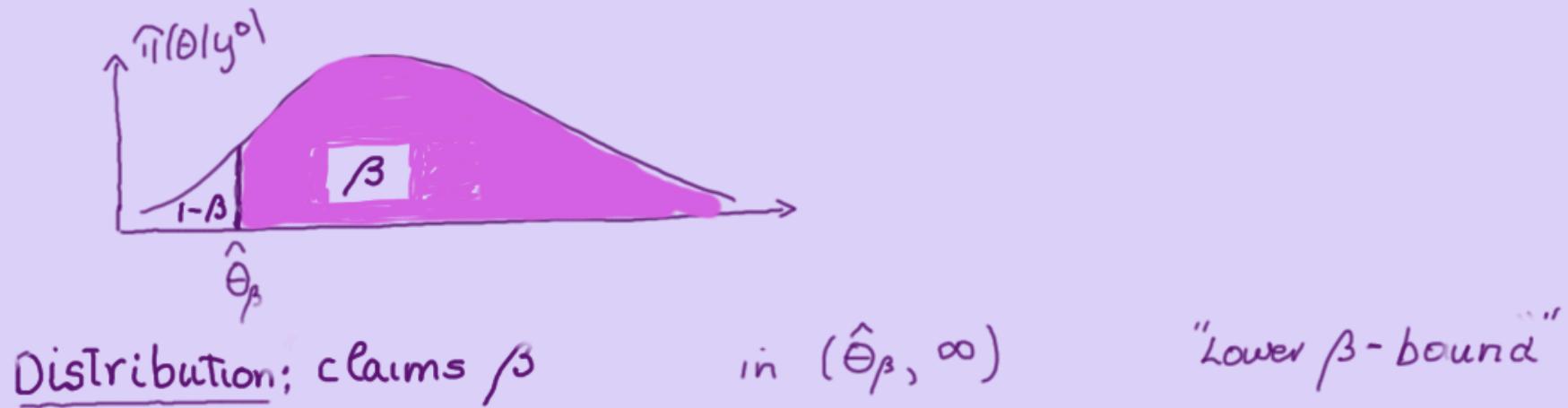
Box Cox 1964

Wasserman (2000)

F. Reid Marras Yi (2008)

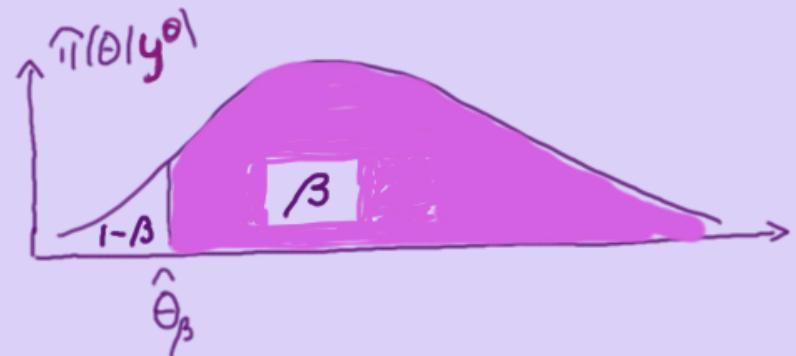
c) $\hat{\pi}(\theta | y^0)$: How To assess a "Distribution" for θ ? Try Neyman

(i) Try a quantile: say 50%ile; 90%ile; β -%ile



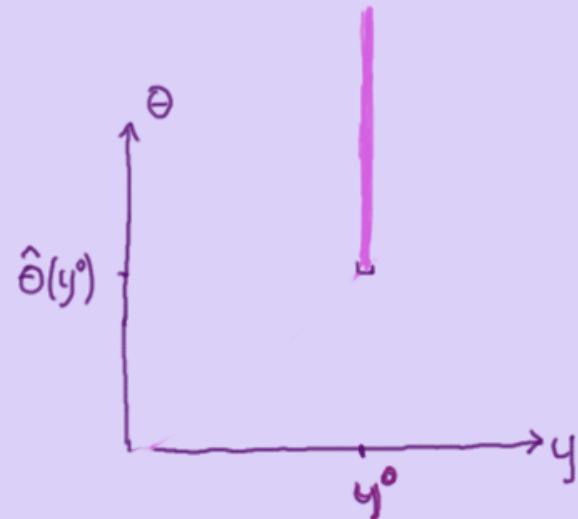
c) $\pi(\theta | y^{\circ})$ How To assess "Distribution" Try Neyman

(i) Try a quantile: say 50%ile; 90%ile; β -%ile



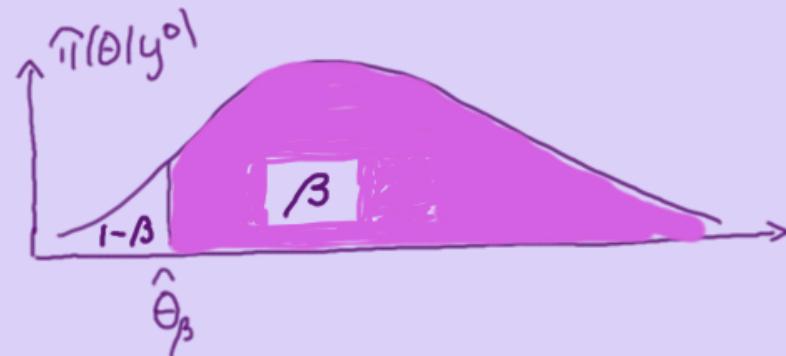
"Distribution" claims β that θ in $(\hat{\theta}_{\beta}, \infty)$ "Lower β bound"

(ii) Neyman diagram



c) $\pi(\theta | y^\circ)$ Assess "Distribution": Neyman diagram

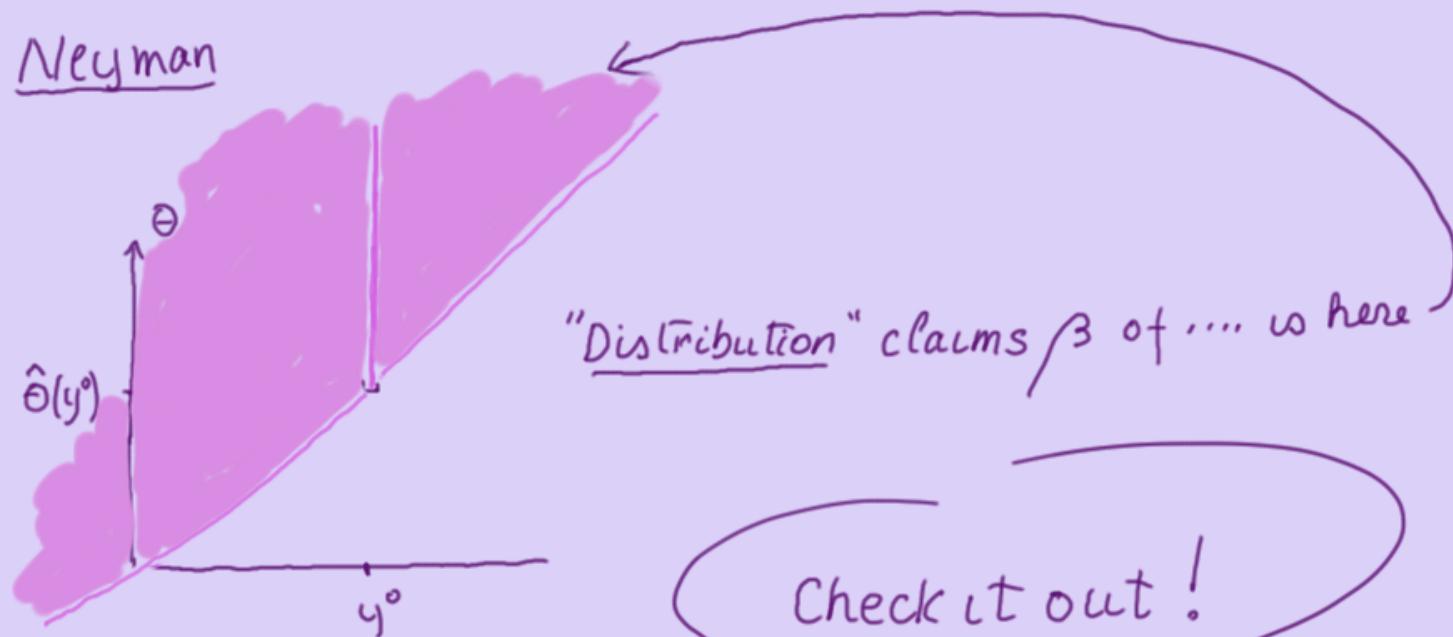
(i) Try a quantile: say 50%ile; 90%ile; β -%ile



"Distribution" claims β that θ in $(\hat{\theta}_\beta, \infty)$

"Lower β bound"

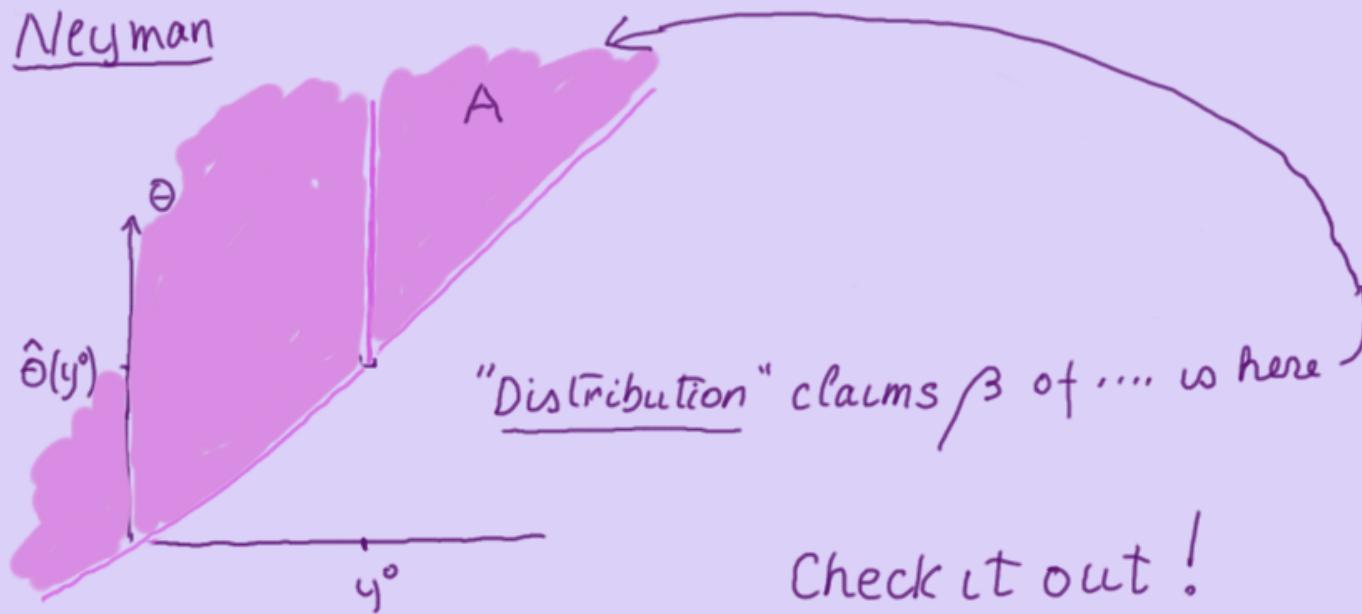
(ii) Neyman



"Distribution" claims β of is here

c) $\pi(\theta | y^*)$ Assess "Distribution": Neyman diagram

(ii) Neyman



(iii) Do Reality check

$$\text{Reality}(\theta) = \text{Prob}(A) = \text{Prob}\{(\hat{\theta}_\beta, \infty) \text{ includes True } \theta ; \theta\}$$

...for any particular θ ...

...for any blend $\pi(\theta)$

3 Simple examples: Curvature i.e. not linear

Example 1 $N(\theta; 1)$ Know $\theta \geq 0$ like Poisson example

$O(1)$

$$\begin{aligned}L(\theta) &= \phi(y^* - \theta) \\&= 0\end{aligned}\quad \theta \geq 0 \quad 0/\omega$$



Example 1 $N(\theta; 1)$

Know $\theta \geq 0$

.... like Poisson example

$$L(\theta) = \phi(y^* - \theta) \quad \theta \geq 0$$

$$= 0$$



Bayes: $\pi(\theta; y^*) = \frac{\phi(y^* - \theta)}{\Phi(y^*)}$

$$\pi(\theta) = \frac{\Phi(y^* - \theta)}{\Phi(y^*)}$$

β -quantile $\hat{\theta}_\beta = y^* - z_{\beta} \Phi(y^*)$

$$\text{Reality}(\theta) = \Pr\{z < z_{\beta} \Phi(\theta + z)\} \quad z \sim N(0, 1)$$

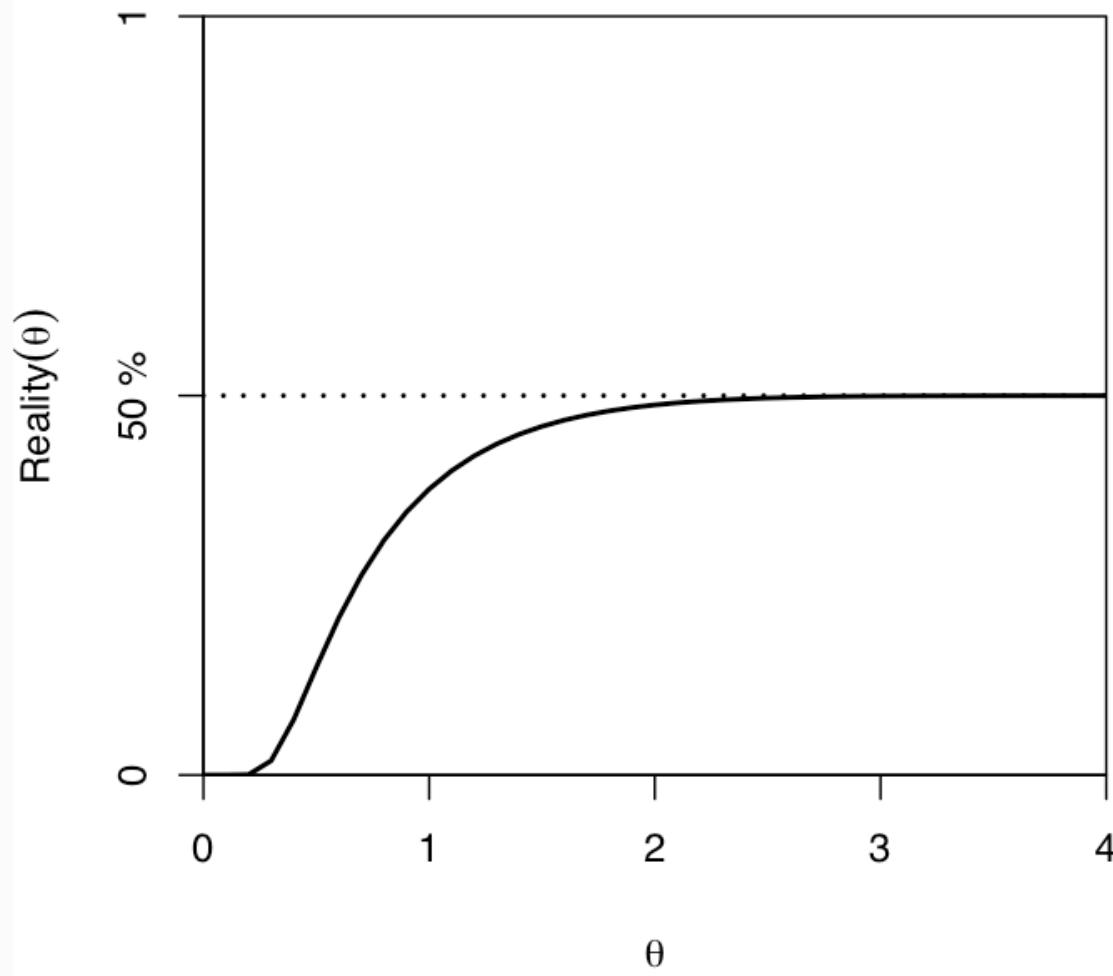
$O(1)$

= Plot ...

Woodroofe (2000)

Ye Sun

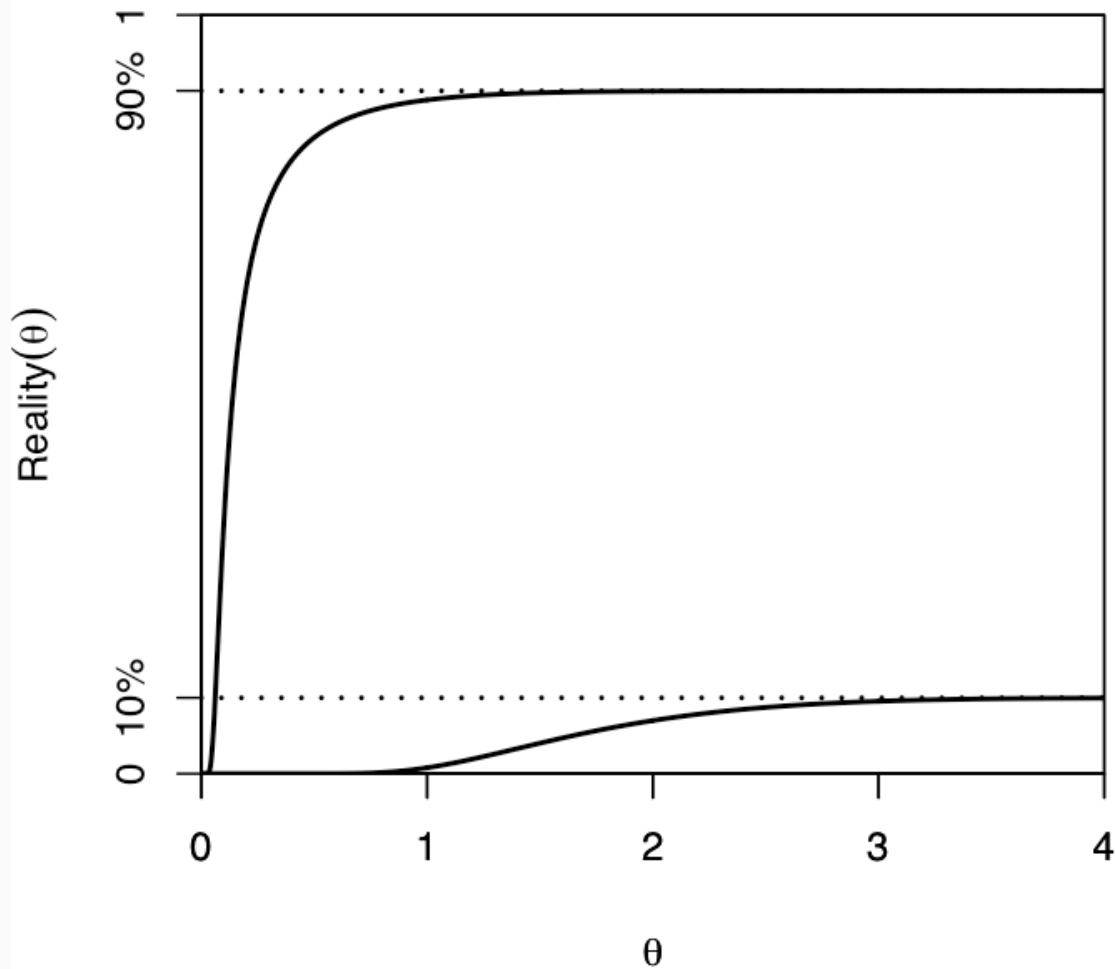
$$\beta = 50\%$$



"Reality" is strictly less than "claimed" Neyman!
No way that posterior can be probability

$$\beta = 90\%$$

$$\beta = 10\%$$



"Reality" is strictly less than "claimed"

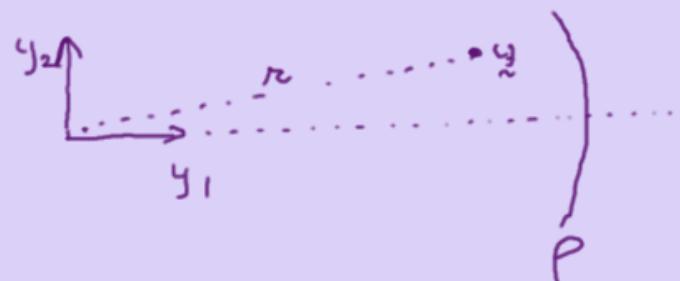
No way that posterior can be / relate to probability

Example 2 $y \sim N(\theta; I)$ $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$... curved $O(n^{1/2})$

f: $p(\rho) = \Pr\{\chi^2(\rho^2) \leq n^2\}$ Non Central Chi² with 2 df
 $n^2 = y_1^2 + y_2^2$

B: $s(\rho) = \Pr\{\rho^2 \leq \chi^2(n^2)\}$

Differ! Bug!



Example 2 $y \sim N(\theta; I)$ $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$... curved $O(n^{1/2})$

f: $p(\rho) = \Pr\{\chi^2(\rho^2) \leq n^2\}$ Non Central Chi² with 2 df
 $n^2 = y_1^2 + y_2^2$

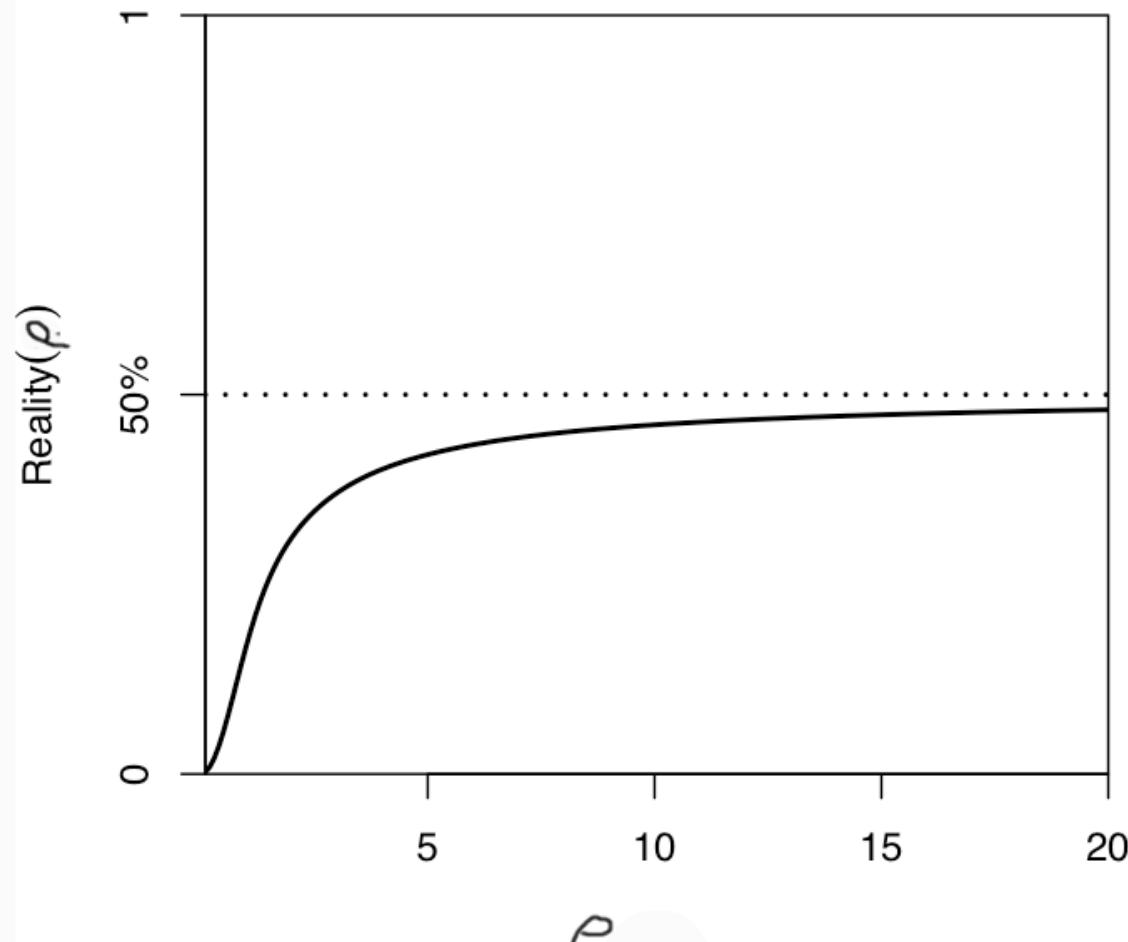
B: $s(\rho) = \Pr\{\rho^2 \leq \chi^2(n^2)\}$

Differ! Bug

Reali_Y(ρ) = Prob $\left[\chi_{1-\beta}^2 \left\{ \chi^2(\rho^2) \right\} \leq \rho^2\right]$ --- for the B posterior

= Plot ...

$$\beta = 50\%$$

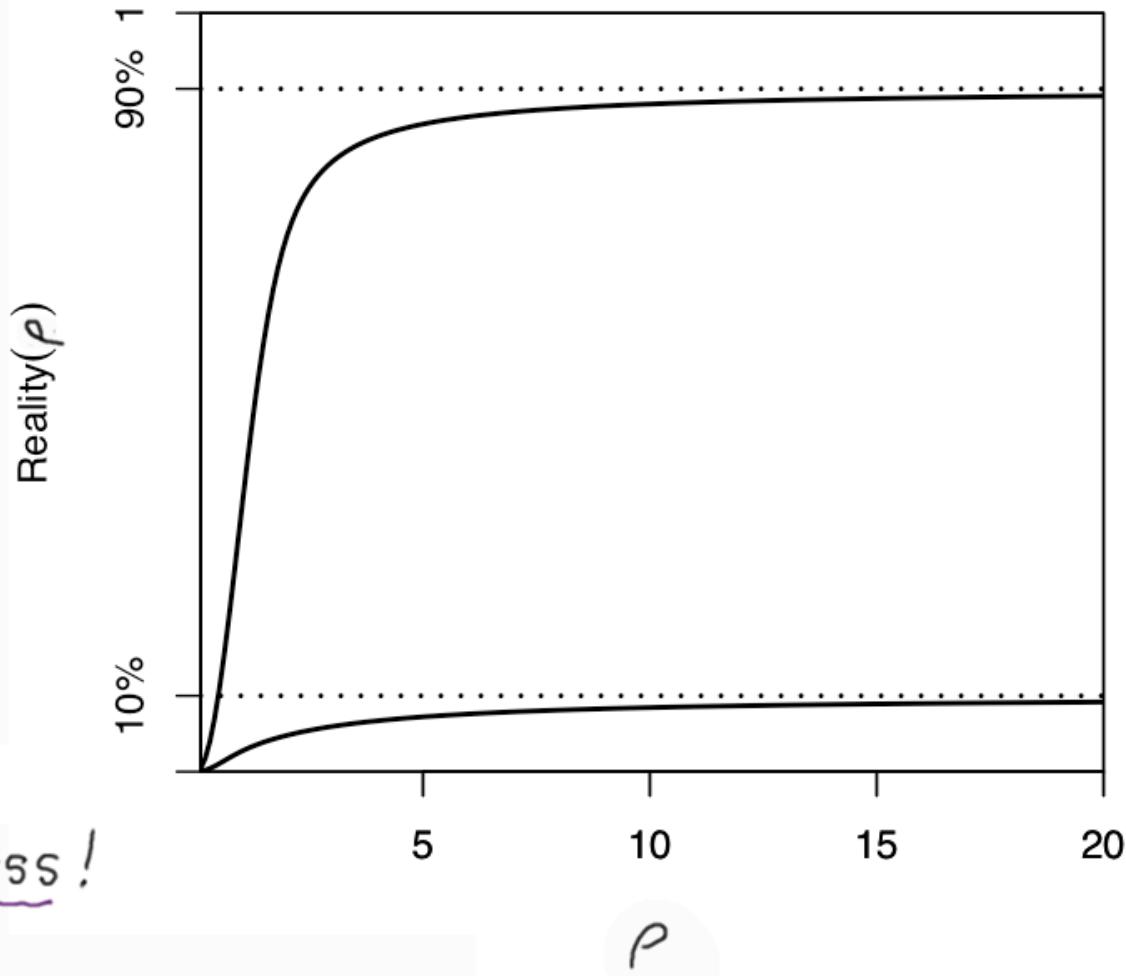


Strictly less

No way that posterior can be / relate to probability

$\beta = 90\%$

$\beta = 10\%$



Again...

Strictly less!

ρ

Example 3 $y \sim N(\theta; 1 + \gamma\theta^2/2n)$

| Variance has $O(n^4)$
| dependence on θ

$$\hat{\theta}_\beta^f = y - \bar{z}_\beta - \gamma z_\beta (y - \bar{z}_\beta)^2 / 4n \quad \dots \text{easy}$$

Example 3 $y \sim N(\theta; 1 + \sigma^2/2n)$

| Variance has
| dependence on θ $O(n^{-1})$

$$\hat{\theta}_\beta^f = y - \bar{z}_\beta - \gamma z_\beta (y - \bar{z}_\beta)^2 / 4n$$

- Bayes $\pi(\theta) = \exp\left\{a\theta/n^{1/2} + c\theta^2/2n\right\}$

Near $\theta=0$ choose $a=0, c=0$ e. flat

Example 3 $y \sim N(\theta; 1 + \sigma^2/2n)$

| Variance has
| dependence on θ

$$\hat{\theta}_\beta^f = y - z_\beta - \gamma z_\beta (y - z_\beta)^2 / 4n$$

$$\pi(\theta) = \exp\{a\theta/n^{1/2} + c\theta^2/2n\}$$

Near $\theta=0$ choose $a=0, c=0$ e. flat

Diff in quantiles $\hat{\theta}^B - \hat{\theta}^f = \left\{ \frac{\sigma}{2n} + \frac{c}{2n} \right\} (y - z_\beta) + \frac{a}{n^{1/2}} + \frac{c}{2n} y$ asymptotics
Likelihood

$$\text{Reality}(\theta) = \beta - \frac{\sigma}{2n} \theta \phi(z_\beta) \quad \text{with flat } a=c=0$$

Example 3 $y \sim N(\theta; 1 + \sigma^2/2n)$

| Variance has
dependence on θ

$$\hat{\theta}_\beta^f = y - z_\beta - \gamma z_\beta (y - z_\beta)^2 / 4n$$

$$\pi(\theta) = \exp\{a\theta/n^{1/2} + c\theta^2/2n\}$$

Near $\theta=0$ choose $a=0, c=0$ e. flat

$$\hat{\theta}^B - \hat{\theta}^f = \left\{ \frac{\sigma}{2n} + \frac{c}{2n} \right\} (y - z_\beta) + \frac{a}{n^{1/2}} + \frac{c}{2n} y$$

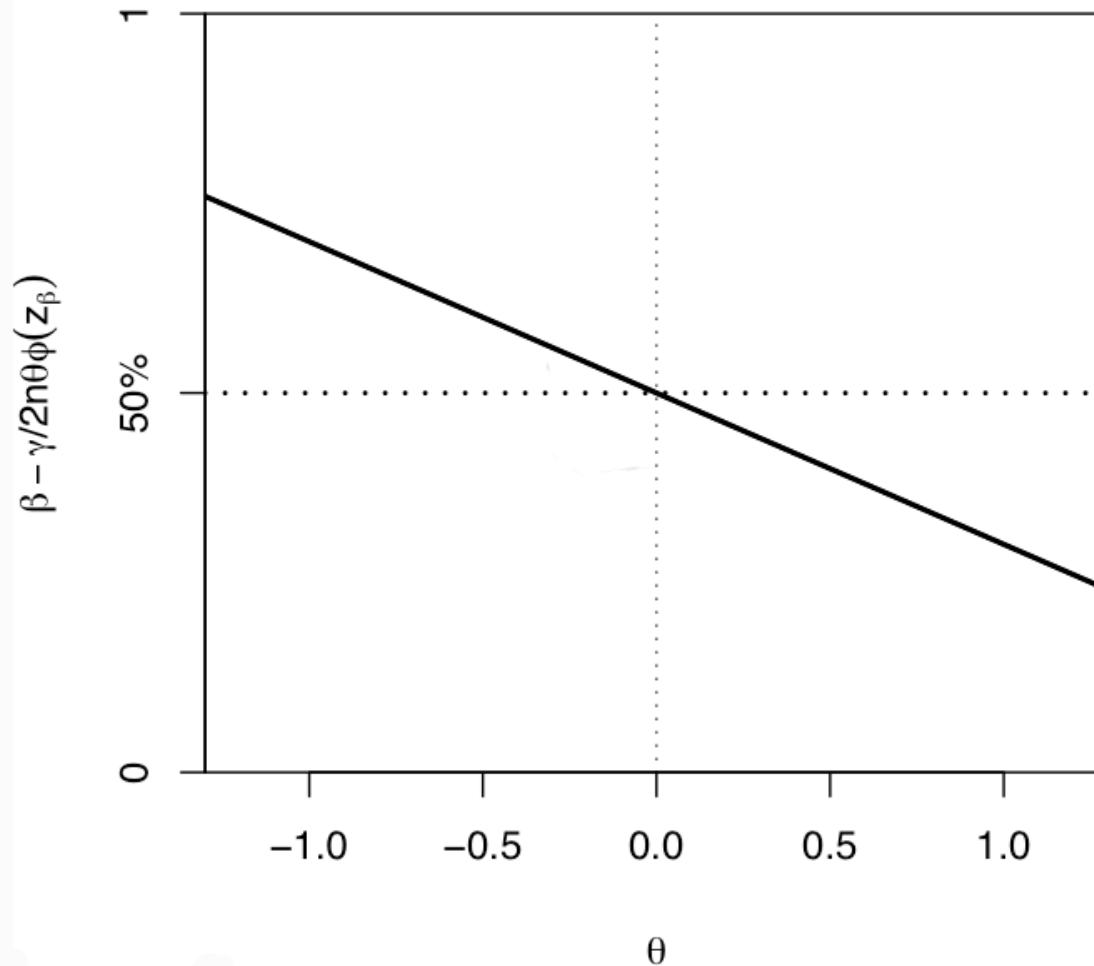
$$\text{Reality}(\theta) = \beta - \frac{\sigma}{2n} \theta \phi(z_\beta) \quad \text{with flat}$$

No prior can give "Reality" = β
anomaly directly tied to curvature

| If not linear
Bayes can't do it!

Ye Sun

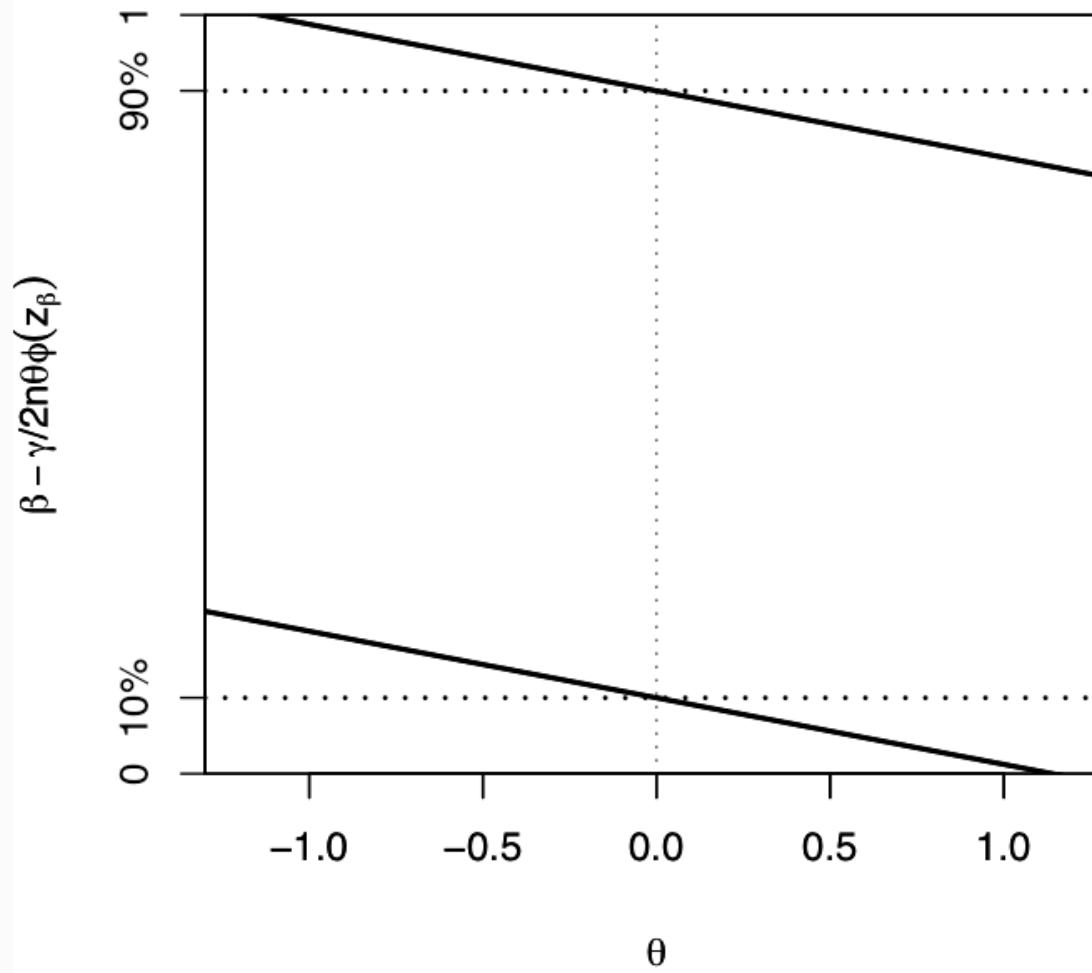
$$\beta = 50\%$$



Above claimed, on either side of centre of curvature
Below

No way that posterior can be / relate to probability

$$\beta = 90\%$$
$$\beta = 10\%$$



Above "claimed", on either side of centre of curvature
Below

4 "The paradigm"

what does it tell us?



If θ values come in a probability pattern $\pi(\theta)$

and

if resulting y values close to \hat{y} are examined

4 "The paradigm"

what does it tell us?



If θ values come in a probability pattern $\pi(\theta)$

and

if resulting y values close to y° are examined

then associated θ values have pattern $\pi(\theta; y^{\circ})$

Conditional probability!

4 "The paradigm"

what does it tell us?



If θ values come in a probability pattern $\pi(\theta)$

and

if resulting y values close to y° are examined

then associated θ values have pattern $\pi(\theta; y^{\circ})$

Conditional probability!

It does not tell us

If there isn't a $\pi(\theta)$

then we will still get a $\pi(\theta; y^{\circ})$

4 "The paradigm"

what does it tell us?



If θ values come in a probability pattern $\pi(\theta)$

and

if resulting y values close to y° are examined

then associated θ values have pattern $\pi(\theta; y^{\circ})$

Conditional probability!

It does not tell us

If there isn't a $\pi(\theta)$

then we will still get a $\pi(\theta; y^{\circ})$

Like

"Probabilities from No probabilities."

Magic!

Mysticism!

Fraud!

5 "The Theorem"

"An optimum procedure is obtained from a prior"

Relevance here?

5 "The Theorem"

"An optimum procedure is obtained from a prior"

Relevance here?

1) Lindley (1958) said confidence was wrong:

Confidence is only a posterior if model is location fly- θ)

5 "The Theorem"

"An optimum procedure is obtained from a prior"

Relevance here?

1) Lindley (1958) said confidence was wrong:

Confidence is only a posterior if model is location fly- θ)

2) but what does Neyman diagram tell us?

Posterior is only confidence if model is location!
% w gamble

5 "The Theorem"

"An optimum procedure is obtained from a prior"

Relevance here?

1) Lindley (1958) said confidence was wrong:

Confidence is only a posterior if model is location $f(y|\theta)$

2) but what does Neyman diagram tell us?

Posterior is only confidence if model is location!
% w gamble

3) Bayes posterior.... just a linear approximation to confidence

6. The Message

1. Bayes posterior probabilities

- around for a long time.
- Use only Likelihood

6. The Message

1. Bayes posterior probabilities

- around for a long time.

- Use only Likelihood

2 If model Linear $f(y|\theta) \quad f(Ay - B\theta)$

Ex 1

Bayes posterior can be confidence ... but easier

6. The Message

1. Bayes posterior probabilities

- around for a long time.

- Use only Likelihood

2 If model Linear $f(y-\theta)$ $f(Ay-B\theta)$

Ex 1

Bayes posterior can be confidence ... but easier

3 If parameter is curved,

Ex 2

- then can have no resemblance to prob. or confidence

Stainforth et al (2007) Phil Trans. Roy. Soc A 365 - 61

- Economist (Aug 18, 2007)

6. The Message

1. Bayes posterior probabilities

- around for a long time.

- Use only Likelihood

2 If model Linear $f(y-\theta)$ $f(Ay-B\theta)$

Ex 1

Bayes posterior can be confidence ... but easier

3 If parameter is curved,

Ex 2

- then can have no resemblance to prob. or confidence

Stainforth et al (2007) Phil Trans. Roy. Soc A 365 - 61

- Economist (Aug 18, 2007)

4 If model is curved,

Ex 3

Bayes cannot achieve confidence

can be
consistently wrong

6. The Message

1. Bayes posterior probabilities

- around for a long time.

- use only Likelihood

2 If model Linear $f(y - \theta)$ $f(Ay - B\theta)$

.... Ex 1

Bayes posterior can be confidence ... but easier

3 If parameter is curved,

.... Ex 2

- then can have no resemblance to prob. or confidence

Stainforth et al (2007) Phil Trans. Roy. Soc A 362: 2145-61

- Economist (Aug 18, 2007)

4 If model is curved,

.... Ex 3

Bayes cannot achieve confidence

An old story writ large

- Linear approximation can be helpful

- but can also be very wrong

- Bayes provides a linear approximation to confidence