

The Bayes myth - Probabilities or just approximate confidence?

D A S Fraser
Collaborators
University of Toronto

February 7 2008
University of Western Ontario

www.toronto.edu/dfrazer/documents/uwo08.pdf

Why ?

Outline :

Background: Timeline
 Poisson example
 Issues

Scalar case: $f(y - \theta)$ (linear)
 $f(y; \theta)$ (curved)
 The Neyman diagram

Curved ... : $N(\theta; I)$ $\theta > 0$
 $N(\theta; I)$ $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$
 $N(\theta; I + \sigma\theta^2/2n)$

"The Paradigm"

"The Theorem"

Message

I Background:

i) Time Line:

1763 Bayes $f(y|\theta)$

... $p^\theta(\theta|y^\circ)$

1774 Laplace endorsed

... authority

1922 Fisher $L(\theta)$

... 150 years later

1930 Fisher Confidence

... $\pi^F(\theta|y^\circ)$ & π^β wrong

1958 Lindley

... π^F wrong & π^β right

/ Background:

i) Time Line:

1763 Bayes $f(y|\theta)$

... $\pi^\theta(\theta|y^*)$

1774 Laplace endorsed

... authority

1922 Fisher $L(\theta)$

... after 150 years

1930 Fisher Confidence

.. $\pi^F(\theta|y^*)$ & π^θ wrong

1958 Lindley

.. π^F wrong & π^θ right

1946 Jeffreys

1963 Welch Peers

1964 Box Cox

1973 Dawid Stone Zidek

1979 Bernardo

2000 Wasserman

/ Background:

i) Time Line:

1763	Bayes	$f(y \theta)$... $\pi^\theta(\theta y^\circ)$
1774	Laplace	endorsed	... authority
1922	Fisher	$L(\theta)$... after 150 years
1930	Fisher	Confidence	.. $\pi^F(\theta y^\circ)$ & π^θ wrong
1958	Lindley		.. π^F wrong

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

Both can't be right

coexist

Q's

2) 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) 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

Mendelkern (2002) Statistical Science w Discussion

2) 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 Woog (2004) Physics Rev D

3) Issues

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

Not equal
Contradictory

Exploratory
"Not the claim!"

- b) Bayesian ... Use only $L^\circ(\theta)$
frequentist... Ignores $L^\circ(\theta)$
- 2 flavours !

3) Issues

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

Not equal
 Contradictory

Exploratory
 Not the claim!

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

Serious?

2 Scalar case

a) Simple linear $f(y-\theta)$ "you can't go wrong"

freq, "Where data is re θ :"

$$p(\theta) = \int_{-\infty}^y f(y-\theta) dy \quad \text{"% age position of data"}$$

at 10% point in

Stats class
On 40%

"No grad school"
?

2 Scalar case

a) Simple linear $f(y-\theta)$ You can't go wrong

freq: "Where data is re θ :"

$$p(\theta) = \int_{-\infty}^{y^*} f(y-\theta) dy \quad \% \text{ age position of data}$$

at 10% point in Stats class No grad school
 Stats problem ?

Bayes: "Where θ is re data:"

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

Equal: Same calculus interval: equal!

If linear! Bayes....OK

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

You can screw up BIG!

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

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

... % age position of data
Just derive θ
Equality by calculus

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

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

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

You can screw up BIG!

freq $p(\theta) = \int_{-\infty}^{y^*} f(y; \theta) d\theta = F(y^*; \theta)$... % age position of data

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

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

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



If agree \Rightarrow

$$\pi(\theta) = \frac{-F_y(y^*; \theta)}{F_y(y^*; \theta)} = \frac{d}{d\theta} \underbrace{y}_{|y^*}$$

Quantile function

$$y = y(u; \theta)$$

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

= Data dependent prior

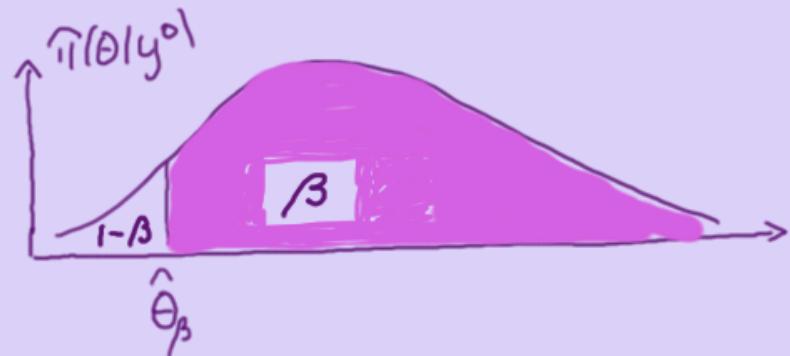
Box Cox 1964

Wasserman (2000)

F. Reid Marras Yi (2008)

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

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

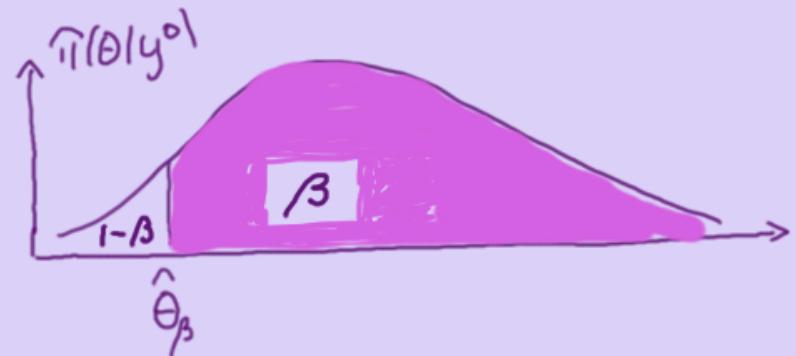


"Distribution" claims β that θ in $(\hat{\theta}_\beta, \infty)$

"Lower β bound"

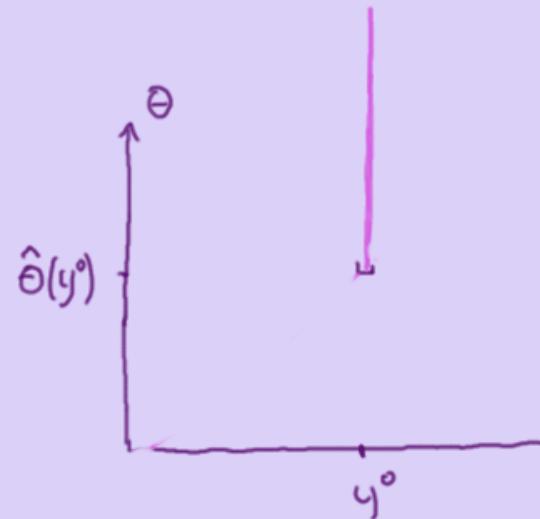
c) $\pi(\theta | y^{\circ})$ How To assess "Distribution" Try Neyman

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



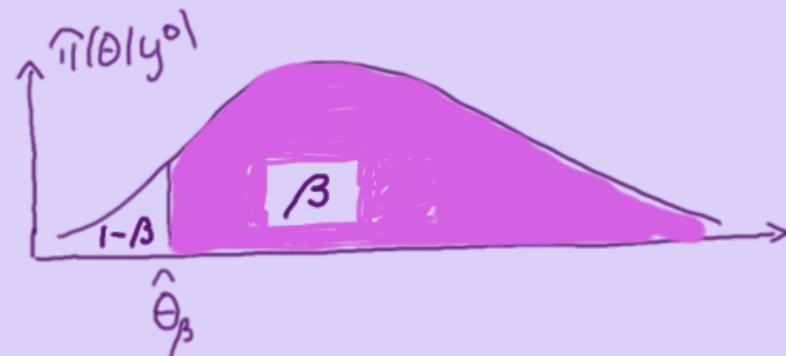
"Distribution" claims β that θ in $(\hat{\theta}_{\beta}, \infty)$ "Lower β bound"

(ii) Neyman diagram



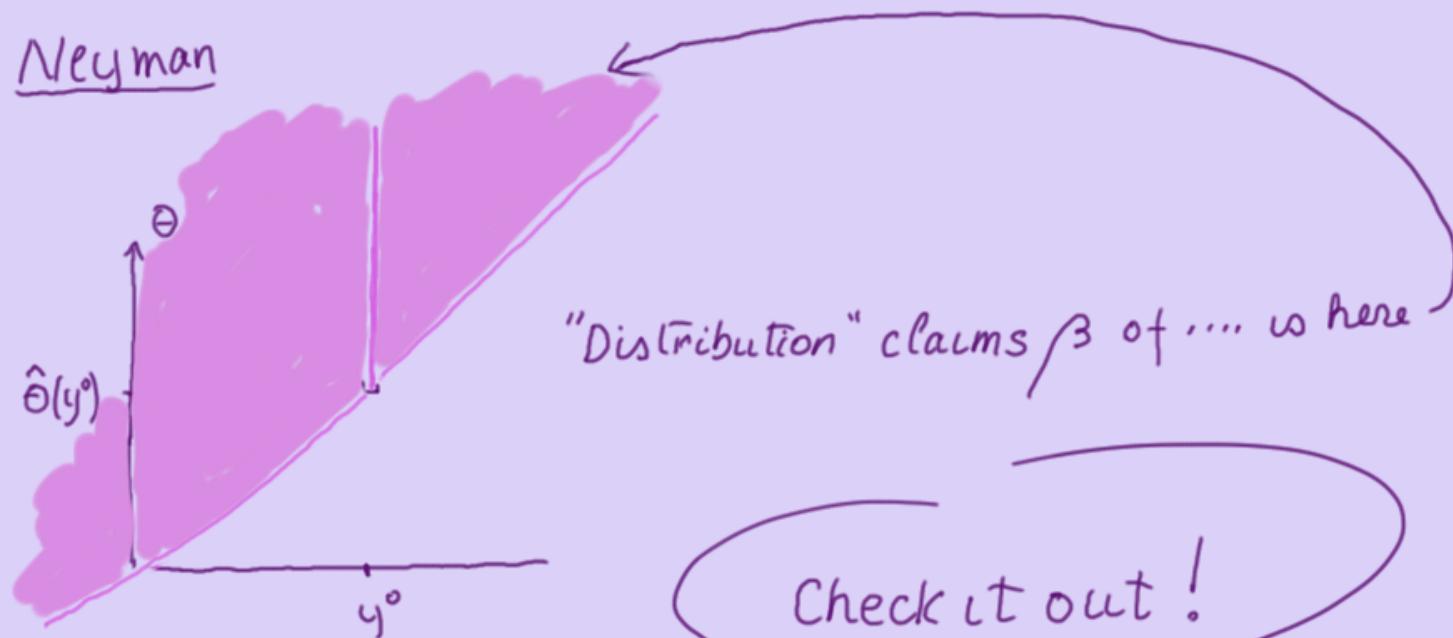
c) $\pi(\theta | y^\circ)$ Assess "Distribution": Neyman diagram

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



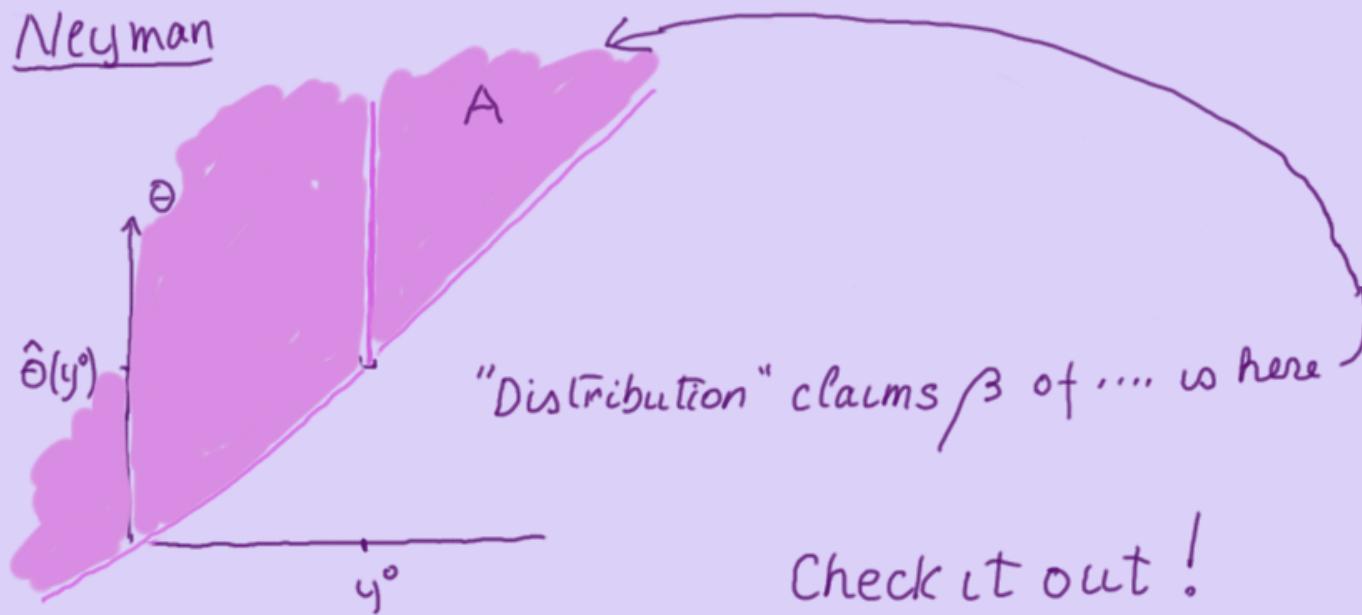
"Distribution" claims β that θ in $(\hat{\theta}_\beta, \infty)$ "Lower β bound"

(ii) Neyman



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

(ii) Neyman



(iii) Do Reality check

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

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

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



Example 1 $N(\theta; 1)$

Know $\theta \geq 0$

.... like Poisson example

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

$\theta \geq 0$
0/ω



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

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

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

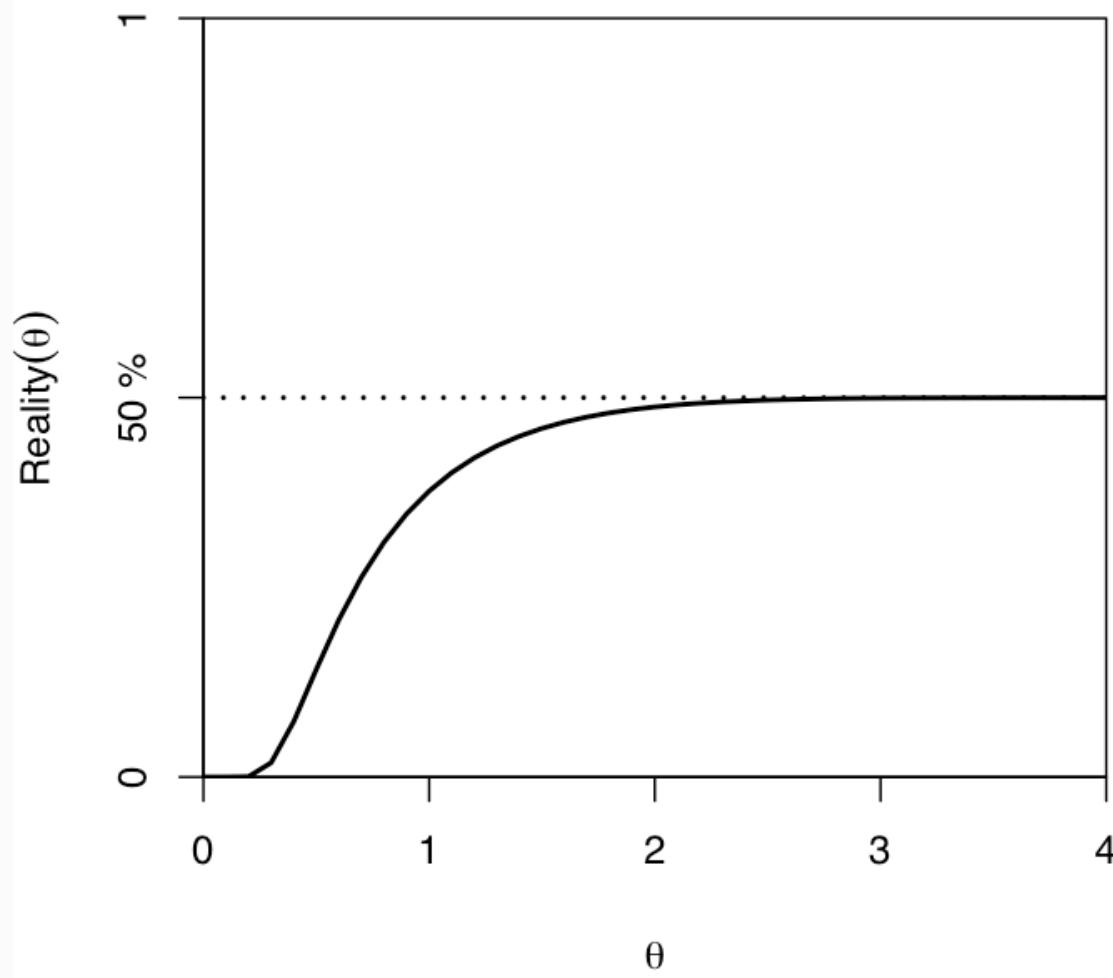
$$\text{Reality}(\theta) = \Pr\{z < z_{\beta} \bar{\Phi}(\theta + z)\} \quad z \sim N(0, 1)$$

= Plot ...

Woodroofe (2000)

Ye Sun

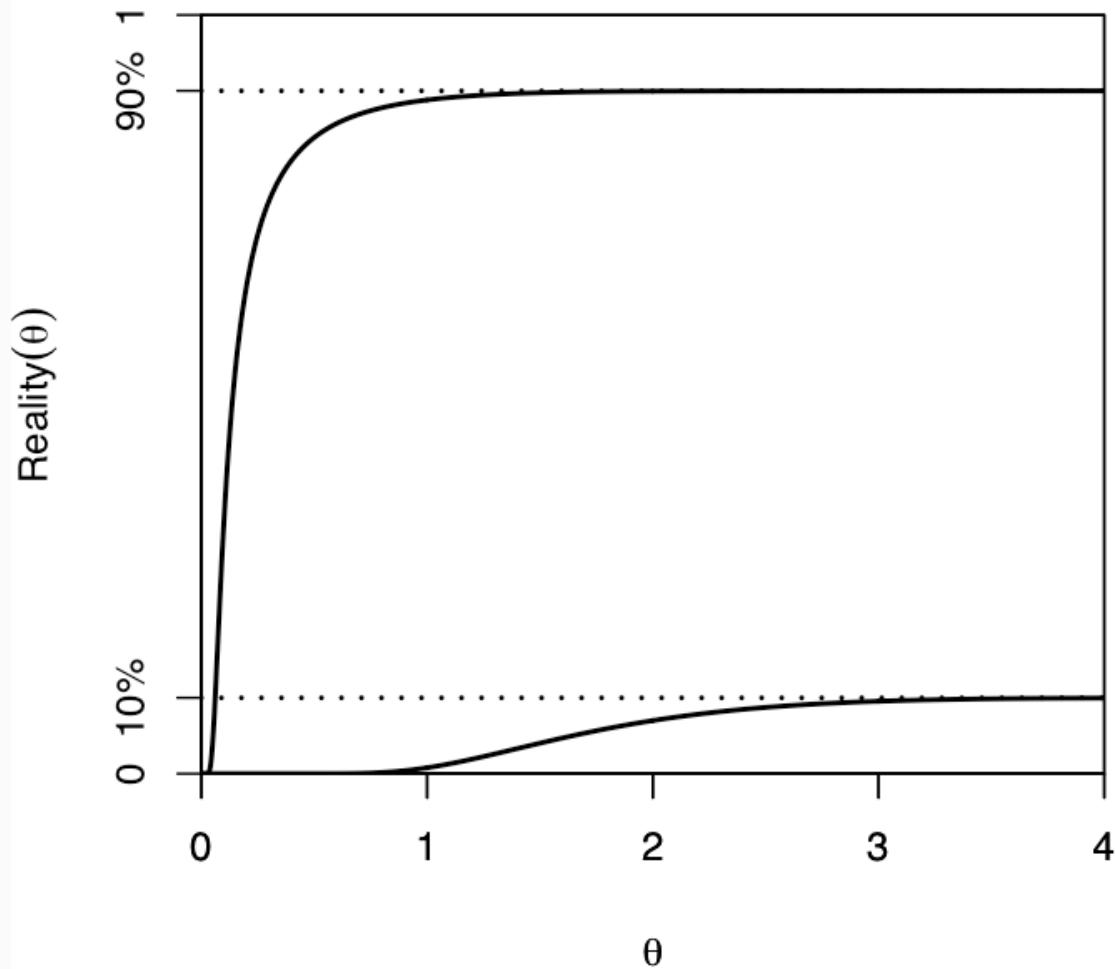
$$\beta = 50\%$$



"Reality" is strictly less than "claimed" Neyman!

$$\beta = 90\%$$

$$\beta = 10\%$$



"Reality" is strictly less than "claimed"

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

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

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

Differ! Bug!



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

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

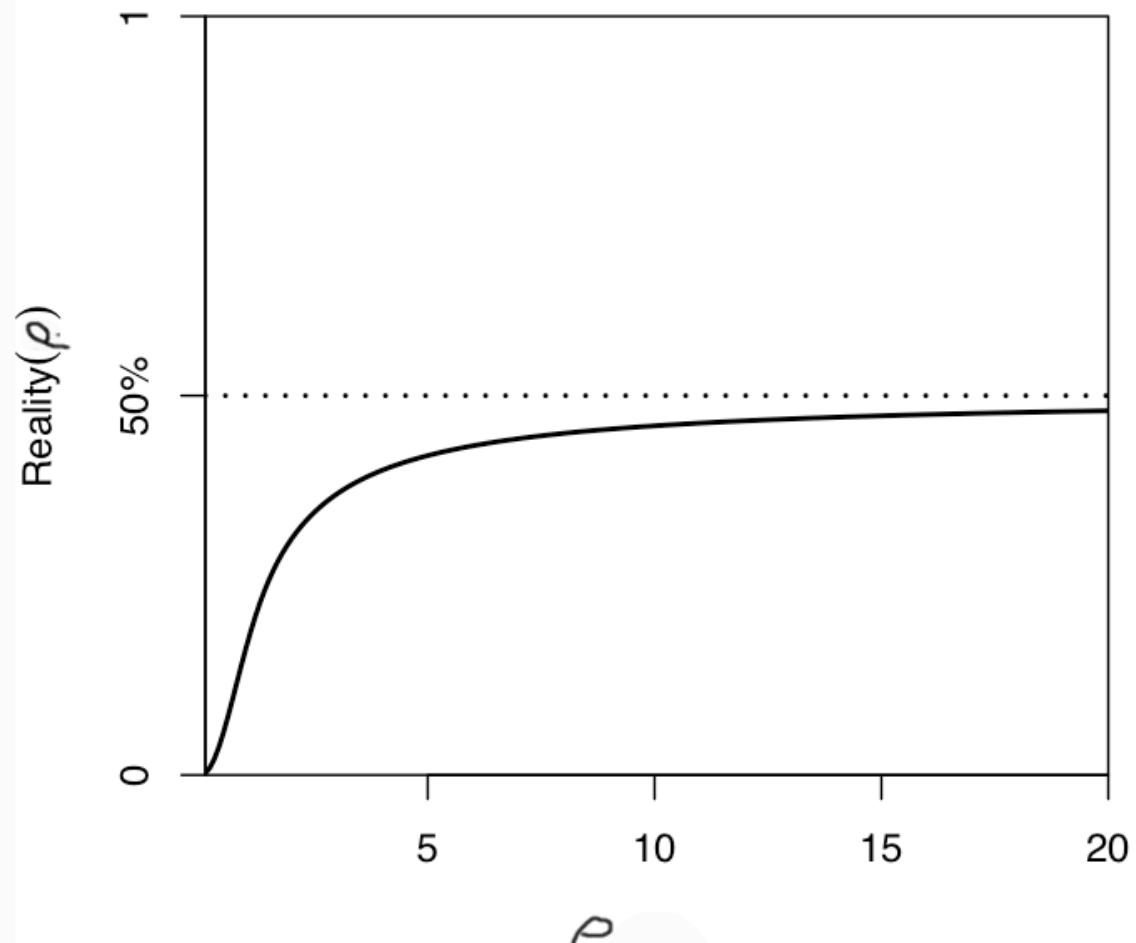
B: $s(\rho) = \Pr\{\rho^2 \leq \chi^2(n^2)\}$

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 ...

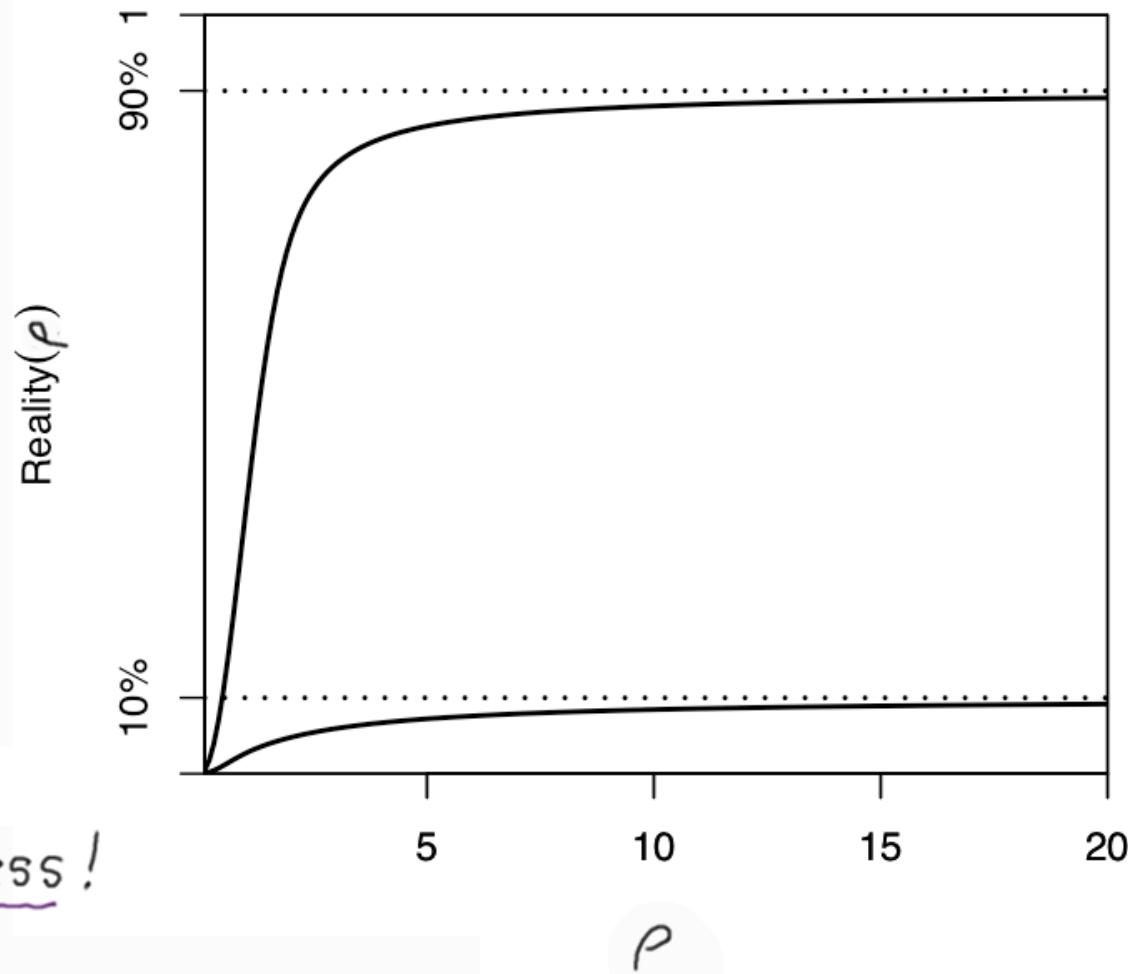
$$\beta = 50\%$$



strictly less

$$\beta = 90\%$$

$$\beta = 10\%$$



Stainforth et al (2007) Phil Trans. Roy. Soc A 366-61
ECONOMIST (Aug 18, 2007) Reports on Stainforth (2007)

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

| Variance has
| dependence on θ

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

...easy

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

| Variance has
| dependence on θ

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

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

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

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

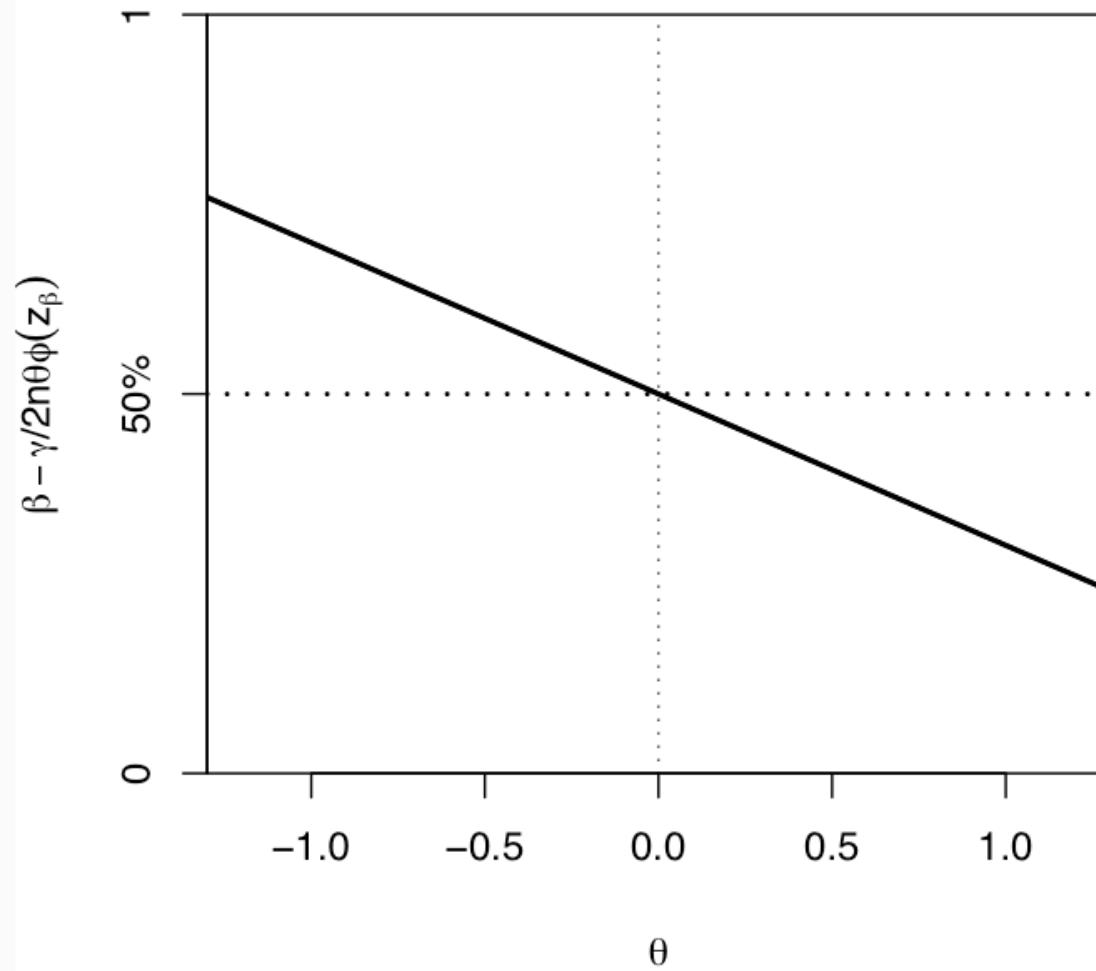
$$\text{Reality}(\theta) = \beta - \frac{\sigma}{2n} \theta \phi(z_\beta) \quad \text{with flat}$$

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

| If not linear
Bayes can't do it!

Ye Sun

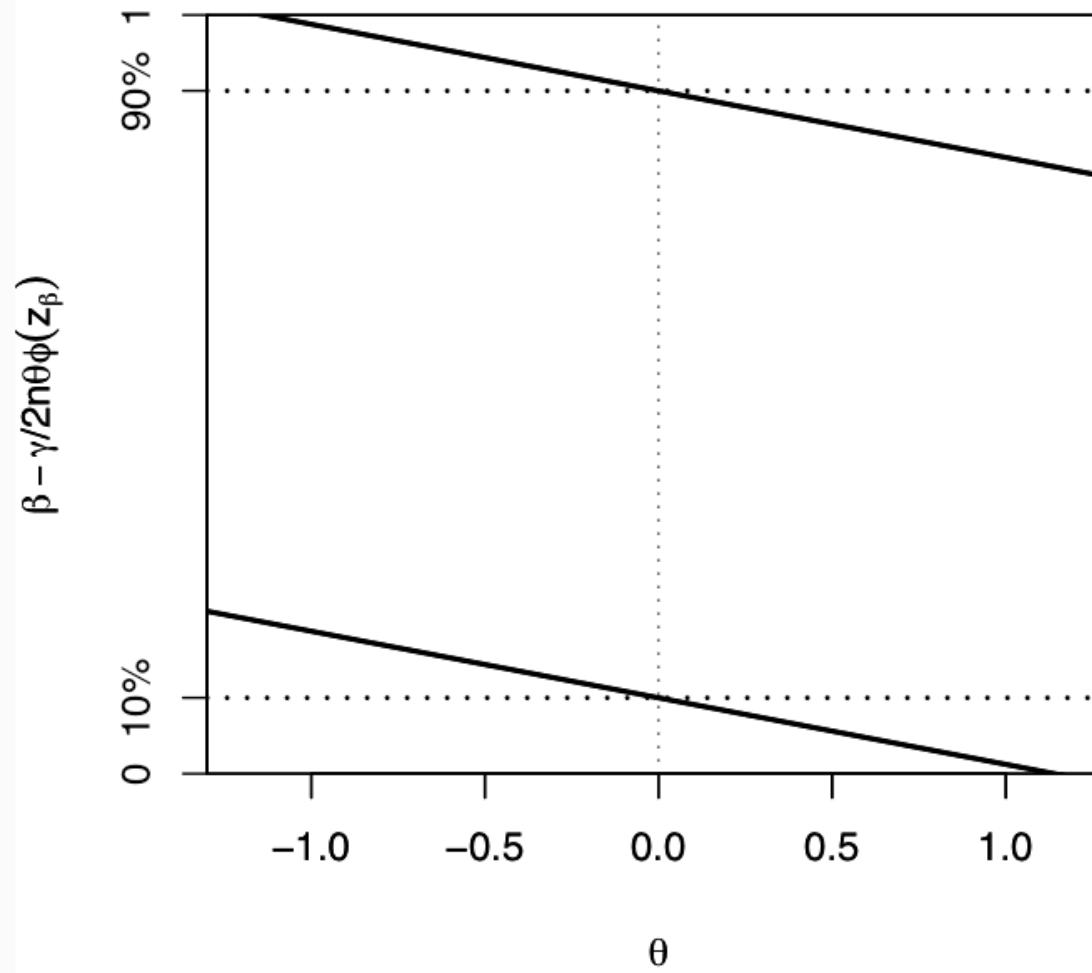
$$\beta = 50\%$$



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

$$\beta = 90\%$$

$$\beta = 10\%$$



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 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^o are examined

then associated θ values have pattern $\pi(\theta; y^o)$

Conditional probability!

Here: "mathematical prior",

"what if"

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!

Here: mathematical prior,

θ -values don't come in a prob. pattern $\pi(\theta)$...

then the paradigm says nothing.

You don't get Probabilities from No probabilities.
... misrepresentation ... Mysticism BIG

5 "The Theorem"

An optimum procedure ... obtained from a prior
Relevance here?

i) Lindley (1958) said confidence was 'wrong':

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

5 "The Theorem"

"An optimum procedure ... 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 the Neyman diagram tell us?

Posterior is only "relative frequency" if model is location!

--- it is then confidence

5 "The Theorem"

"An optimum procedure ... 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 the Neyman diagram tell us?

Posterior is only "relative frequency" if model is location!
--- it is then confidence

i.e. Bayes doesn't get benefit of confidence argument
unless linearly

Theorem: True, but irrelevant

Lindley: True, but implications were the wrong way around

6. The Message

If linearity - easy approximations work
- Bayes give confidence statements

Likelihood is powerful ... $O(n^{1/2})$
frequentists don't use it ! and Should !
Bayesian do use it (+63+) and maybe over-play it

Integrate Likelihood
Get first order confidence (If linear exact
%w approximate go for it!
Misrepresentation to claim more

Higher order accuracy available
Needs more than likelihood
" data-dependent priors

If linear, then things "easy" ... math, applied math, stat (but tell I straight)
curved, need { more than Bayes !
 more than likelihood !

Thank you!