Bayesian posterior probability is just Confidence and inconveniently needs linearity

D A S Fraser
Statistics
University of Toronto

April 1 2008
McMaster University

www.toronto.edu/dfraser/documents/mcm08.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
$N(\underset{\sim}{\theta}; I)$ $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ ... 2nd order
$N(\theta; I + \gamma\theta^2/2n)$ ... 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:

1) Two outputs ... time line

1763	Bayes	$f(y-\theta)$	$\Rightarrow$ $\pi_B(\theta \mid y^o)$	(Linearity invariance
1774	Laplace	endorsed	authority	Inverse prob.
1922	Fisher	$L(\theta)$	+ 150 years	
1930	Fisher	Confidence	$\Rightarrow$ $\pi_F(\theta \mid y^o)$	("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:

1) Two outputs

1763 Bayes $f(y-\theta)$ $\Rightarrow$ $\pi_B(\theta \mid y^\circ)$ (translation invariance

1774 Laplace endorsed authority

1922 Fisher $L(\theta)$ 150 years later

1930 Fisher Confidence $\Rightarrow$ $\pi_F(\theta \mid y^\circ)$ ("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:

1) Two outputs

1763 Bayes $f(y-\theta)$ $\Rightarrow$ $\pi_B(\theta \mid y^o)$ (translation invariance

1774 Laplace endorsed authority

1922 Fisher $L(\theta)$ 150 years later

1930 Fisher Confidence $\Rightarrow$ $\pi_F(\theta \mid y^o)$ ("pivotal inversion "Bayes wrong"

1958 Lindley "π_F wrong"

Two versions of $\pi(\theta \mid y^o)$

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) $\pi^B(\theta|y^o)$
$\pi^F(\theta|y^o)$
| - Not equal
- Contradictory

Two outputs

in

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

Two inputs

} How come?

3) Issues

a) $\pi^B(\theta|y^o)$ Not equal
$\pi^F(\theta|y^o)$ Contradictory

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

c) Authorities Bayes
Laplace
Jeffreys No continental drift Fortey The Earth 2004
Disney No "Big Bang" 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 $S(\theta; y^\circ) = \int_\theta^\infty c\, f(y^\circ - \theta)\, d\theta$

2 Scalar variable, parameter

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

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

Posterior upper/survivor prob $\quad s(\theta; y^o) = \int_\theta^\infty c\, f(y^o - \theta)\, d\theta$

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

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

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

2 Scalar variable, parameter

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

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

Posterior upper/survivor prob $\quad s(\theta; y^o) = \int_\theta^\infty c\, f(y^o - \theta)\, d\theta$

freq. % age position of data

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

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

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 $s(\theta; y^\circ) = \int_\theta^\infty c\, f(y^\circ - \theta)\, d\theta$

freq. % age position of data

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

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 $f(y;\theta)$

Still equal?

freq $$p(\theta) = \int_{-\infty}^{y^\circ} f(y;\theta)\,dy = F(y^\circ;\theta)$$
$$= \int_{\theta}^{\infty} -F_{;\theta}(y^\circ;\theta)\,d\theta$$

... % age position of data

Just deriv. re θ — Seen before?
Equality by calculus

Bayes $$s(\theta) = \int_{\theta}^{\infty} c \cdot f(y^\circ;\theta)\,d\theta$$
$$= \int_{\theta}^{\infty} \pi(\theta)\, F_y(y^\circ;\theta)\,d\theta$$

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

Usually not equal

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

$= \int_{\theta}^{\infty} -F_{;\theta}(y^\circ;\theta)\, d\theta$

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

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

If agree => $\pi(\theta) = \frac{-F_{;\theta}(y^\circ;\theta)}{F_y(y^\circ;\theta)} = \frac{dy}{d\theta}\Big|_{y^\circ}$

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

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

= Data-dependent prior

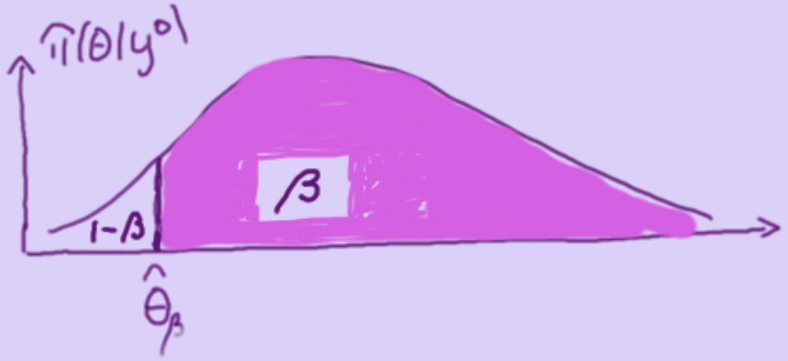
Lindley's calculations reversed

Bayes has confidence property ONLY if location model.

Box Cox 1964
Wasserman (2000)
F Reid Marras Yi (2008)

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

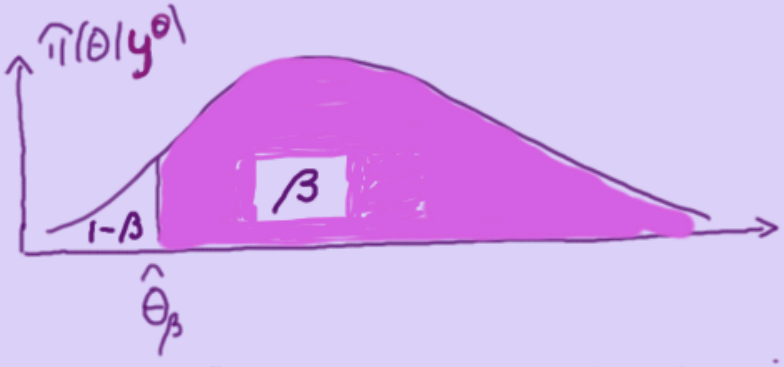
(i) Try a quantile: say 50%ile; 90%ile; β-%ile



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

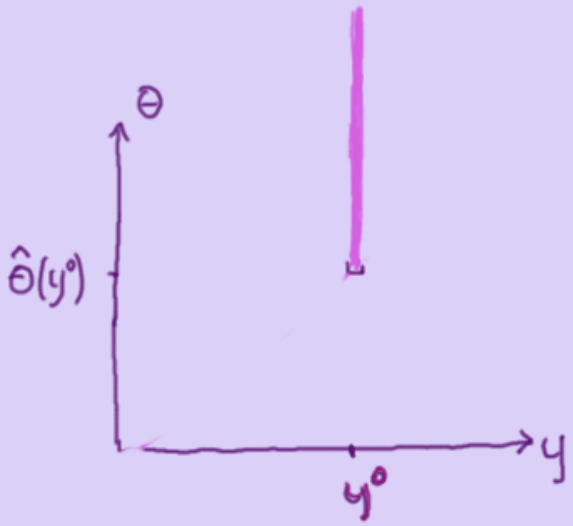
c) $\pi(\theta|y^o)$ How to assess "Distribution" .... Try Neyman

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



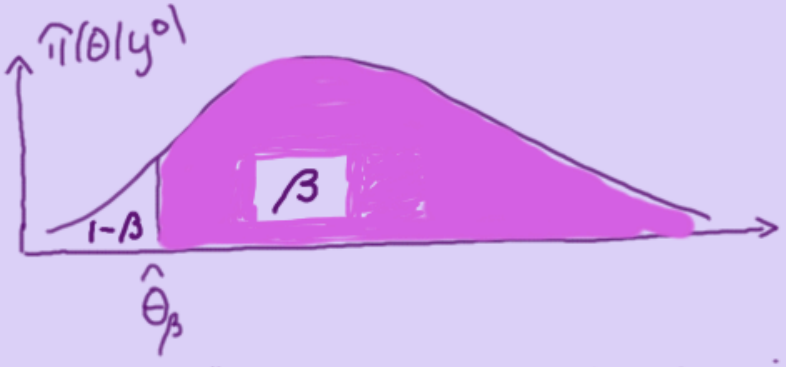
"Distribution" claims β that θ in $(\hat{\theta}_\beta, \infty)$ "Lower β bound"

(ii) Neyman diagram



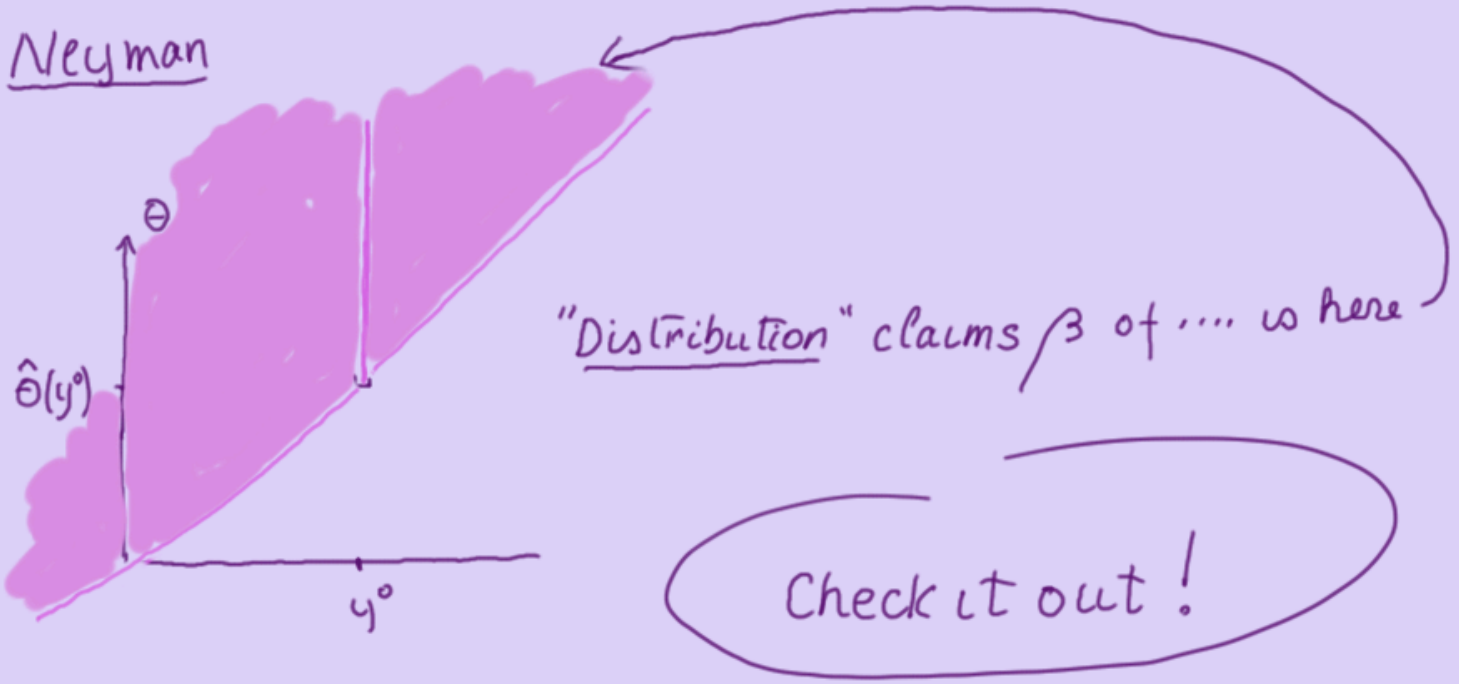
c) $\pi(\theta|y^o)$ Assess "Distribution": Neyman diagram

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



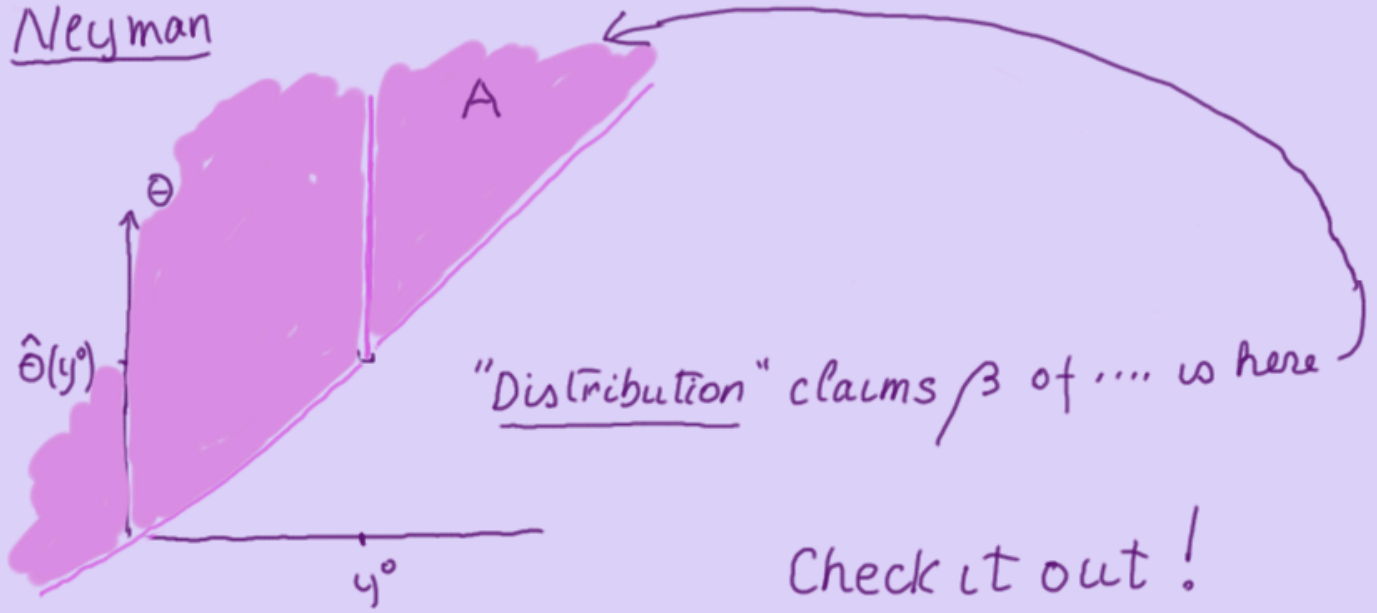
"Distribution" claims β that θ in $(\hat{\theta}_\beta, \infty)$ "Lower β bound"

(ii) Neyman



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

(ii) Neyman



Check it out!

(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 $O(1)$

$$L(\theta) = \phi(y^\circ - \theta) \quad \theta \geq 0$$
$$= 0 \quad o/w$$

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

$$L(\theta) = \phi(y^\circ - \theta) \quad \theta \geq 0$$
$$= 0 \quad o/w$$

Bayes: $\pi(\theta; y^\circ) = \dfrac{\phi(y^\circ - \theta)}{\bar{\Phi}(y^\circ)}$

$$s(\theta) = \frac{\bar{\Phi}(y^\circ - \theta)}{\bar{\Phi}(y^\circ)}$$

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

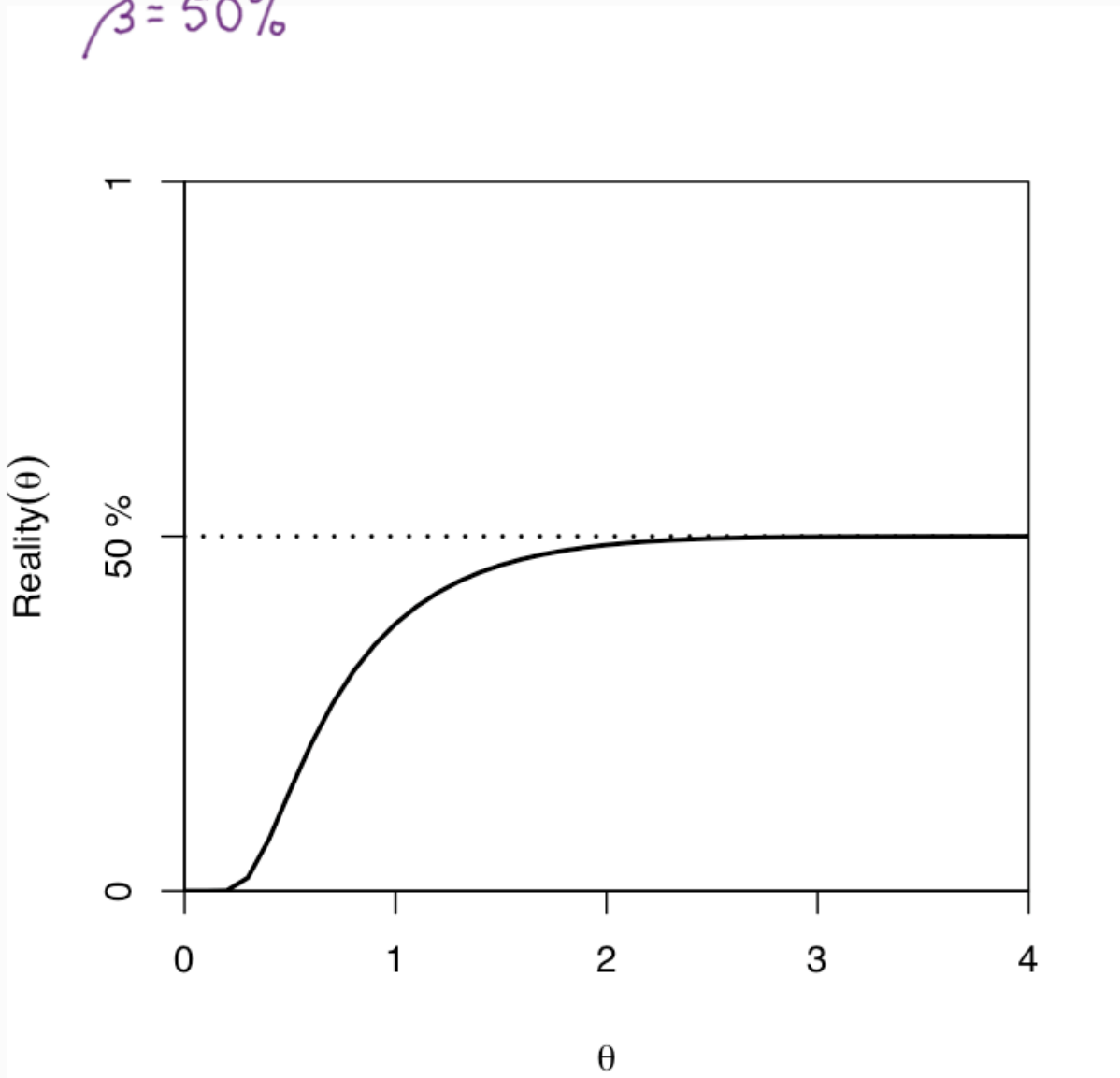
Reality $(\theta) = Pr\{ z < z_\beta \bar{\Phi}(\theta + z) \}$ $z \sim N(0,1)$ $O(1)$

$$= \text{Plot...}$$

Woodroofe (2000)

Ye Sun

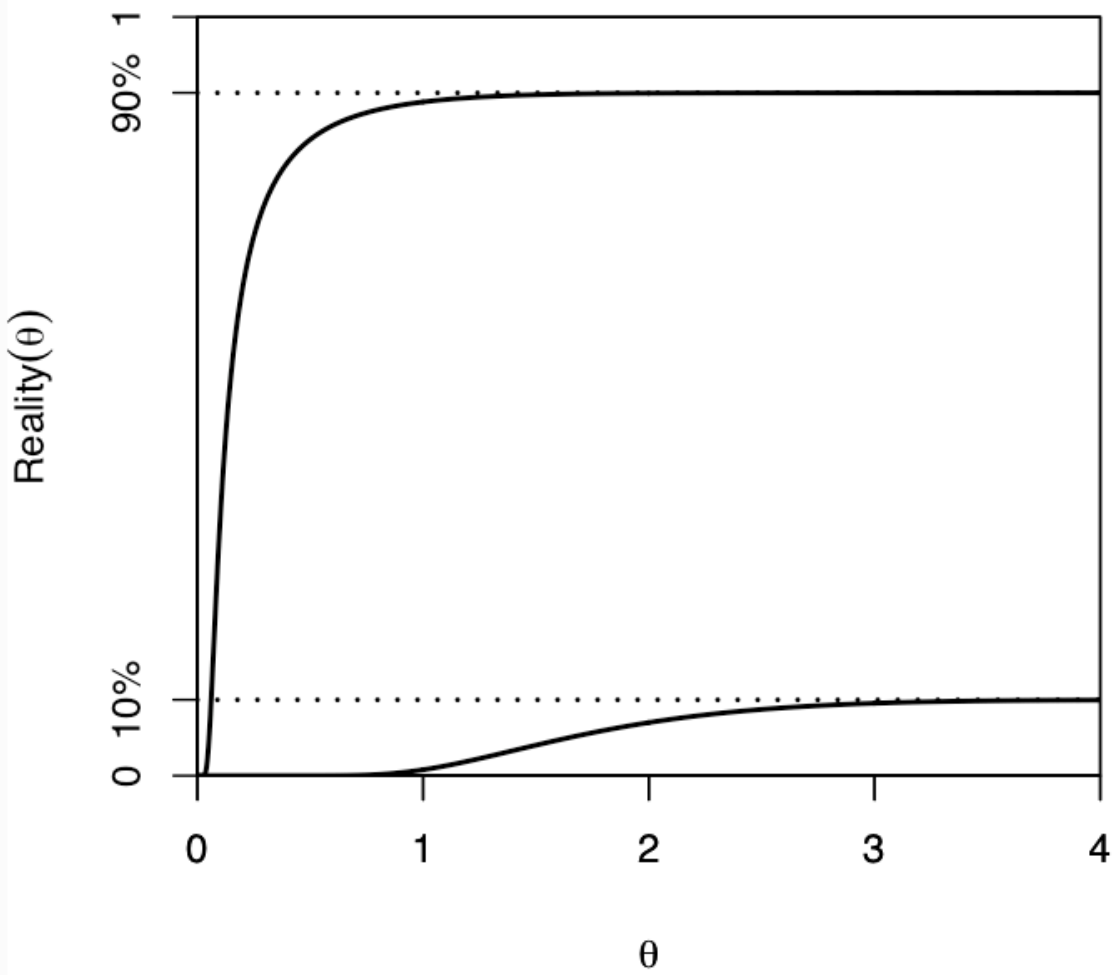
$\beta = 50\%$



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

f: $p(\rho) = P_r\{\chi^2(\rho^2) \leq r^2\}$

Non Central Chi^2 with 2 df
$r^2 = y_1^2 + y_2^2$

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

Differ! Big!

F, Reid (2001) Ye Sun (2008)

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

f: $p(\rho) = Pr\{\chi^2(\rho^2) \leq r^2\}$

Non Central Chi^2 with 2 df
$r^2 = y_1^2 + y_2^2$

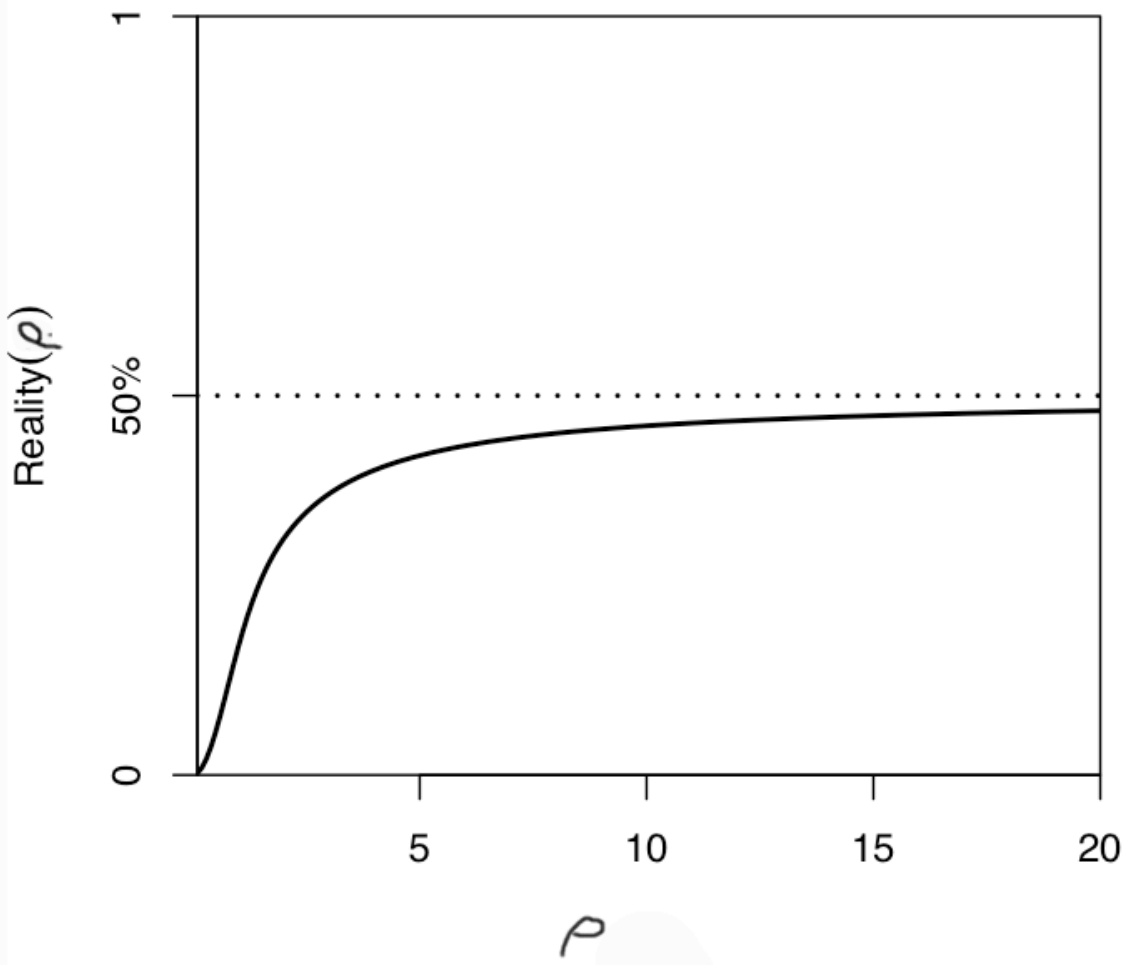
B: $s(\rho) = Pr\{\rho^2 \leq \chi^2(r^2)\}$

Differ! Big

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

= Plot ...

$\beta = 50\%$

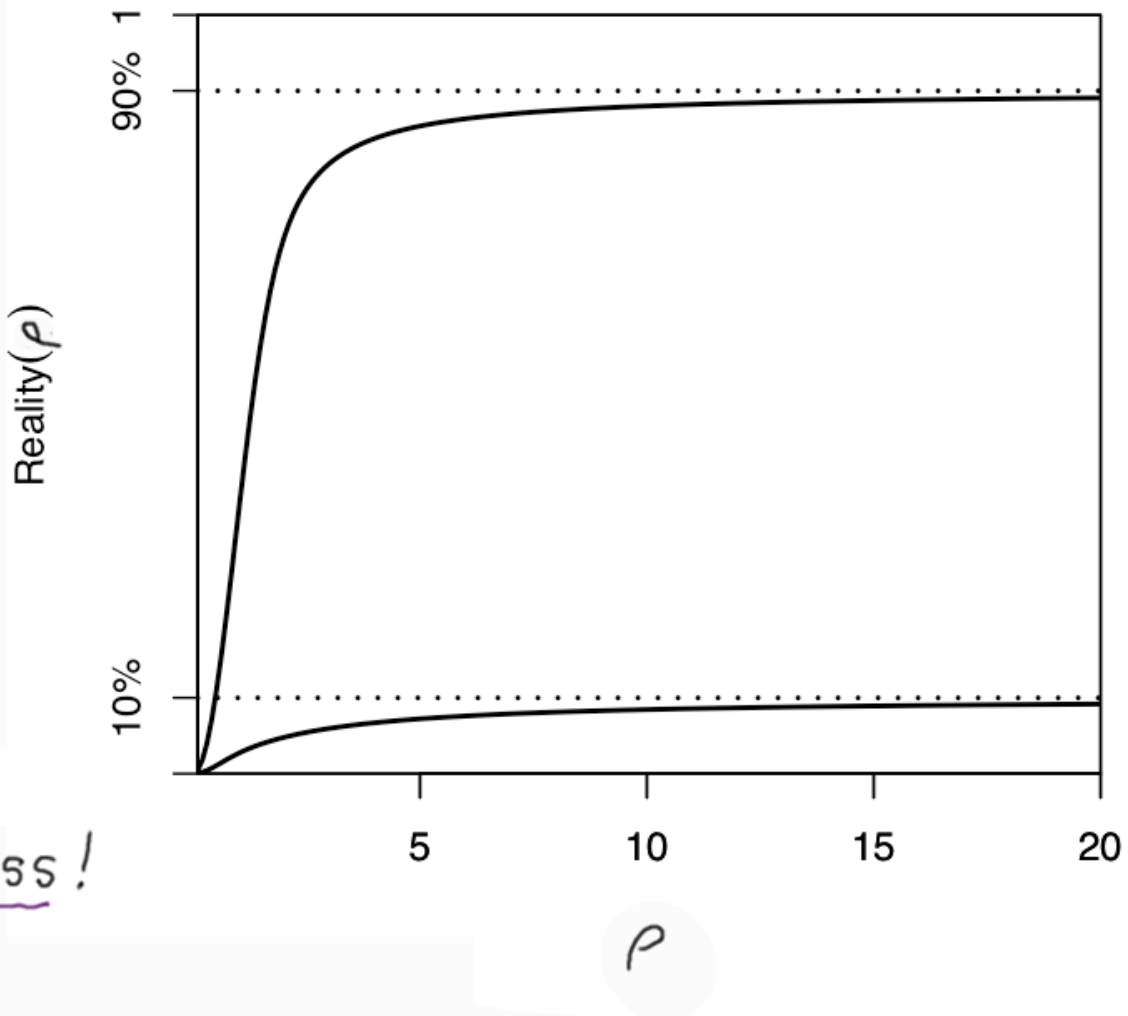


Strictly less

No way that posterior can be / relate to probability

$\beta = 90\%$

$\beta = 10\%$



Again...

Strictly less!

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

Variance has dependence on θ $O(\tilde{n}^1)$

$\hat{\theta}^f_\beta = y - z_\beta - \gamma z_\beta (y - z_\beta)^2/4n$ ...easy

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

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

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

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

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

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

Variance has dependence on θ

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

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

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

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

asymptotics Likelihood

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

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

Variance has dependence on θ

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

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

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

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

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

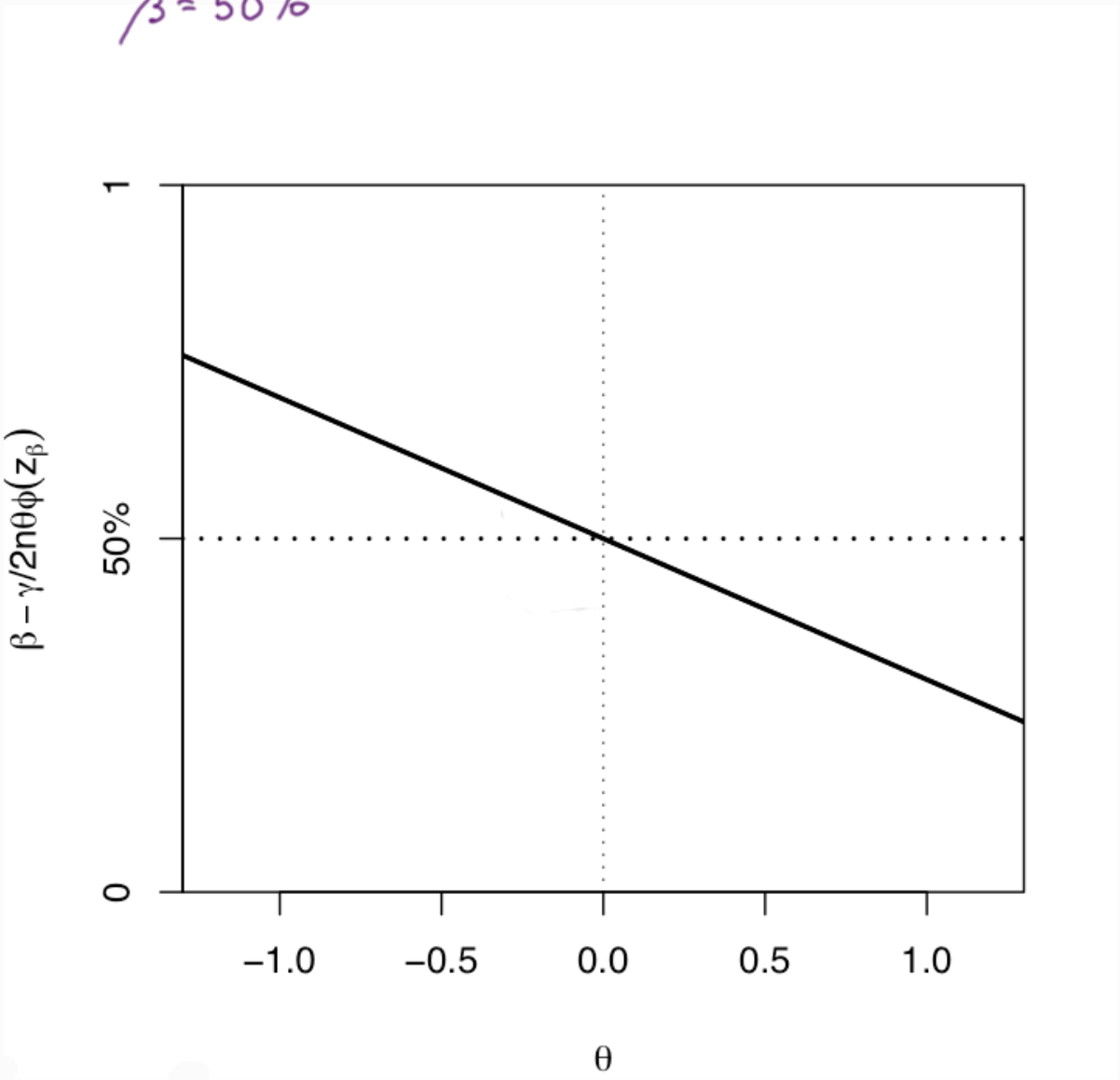
No prior can give "Reality" = β

anomaly directly tied to curvature

If not linear Bayes can't do it!

Ye Sun

$\beta = 50\%$

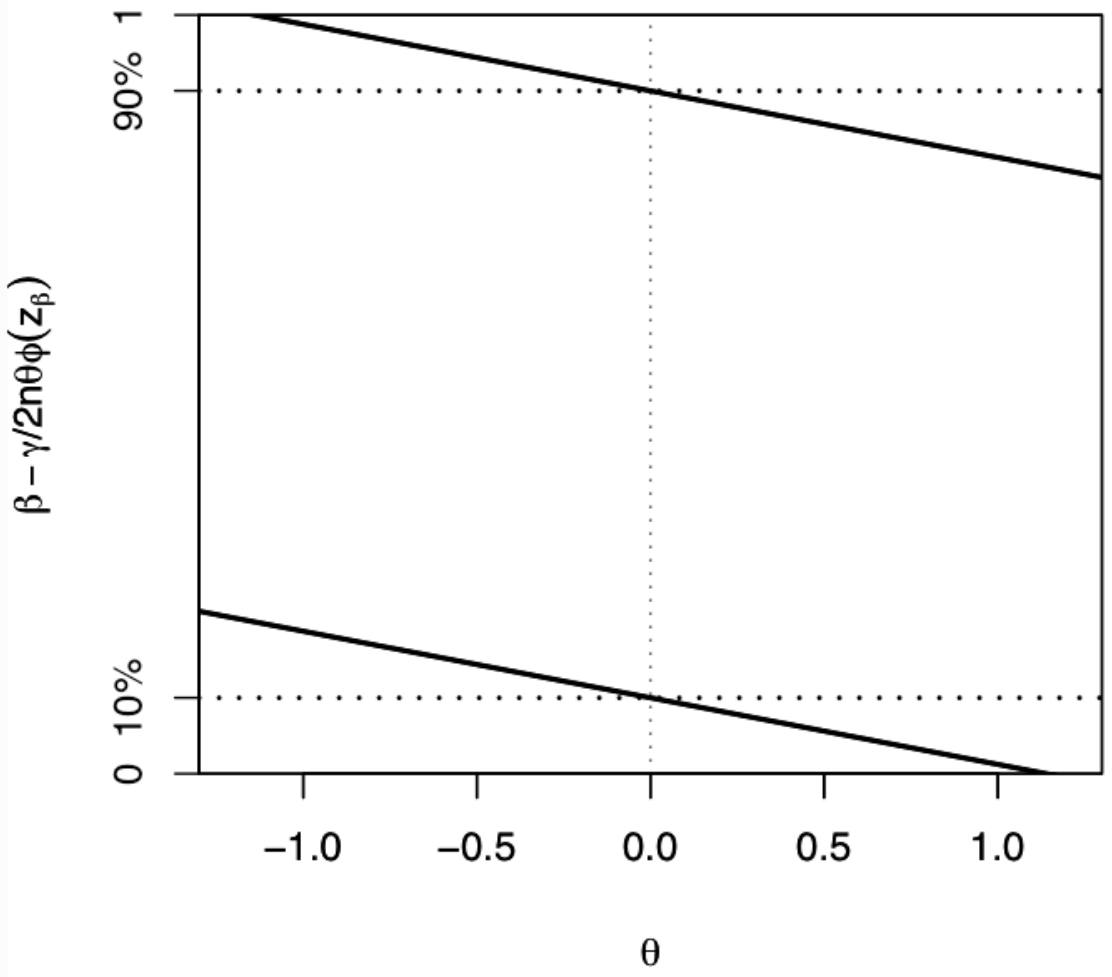


Above / Below claimed, on either side of centre of curvature

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 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 $f(y-\theta)$

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!

$\forall w$ 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 $f(y-\theta)$ $f(Ay-B\theta)$ 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

Ex 2

3 If parameter is curved,

- then can have no resembance to prob. or confidence

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

- 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

Ex 2

3 If parameter is curved,

- then can have no resembance to prob. or confidence

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

- Economist (Aug 18, 2007)

Ex 3

4 If model is curved,

Bayes can not 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 resembance to prob. or confidence

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

- Economist (Aug 18, 2007)

4 If model is curved, ... Ex 3

Bayes can not 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