Higher order Likelihood and the Curse of Curvature

D A S Fraser
Statistics
Univ Toronto

JSM 2010
Vancouver
Aug 5 2010

http://www.utstat.toronto.edu/dfraser/documents/jsm10.pdf

f | p-value: pragmatic
  | p-value: model-based

Bayes uses ONLY likelihood

Parameter without curvature

Parameter with curvature

Default priors

Intrinsically linear parameters

Is a prior 'flat'

Prior (tilted) shifts confidence

Effect on scalar interest parameter

Prior (tilted) translates p-value function

Overview

Linear	Reproducible
Curvature	Need Calibration

p-value: pragmatic

p-value: model-based

B | Bayes uses ONLY likelihood
| Parameter without curvature
| with curvature

p-value: pragmatic

p-value: model-based

Bayes uses ONLY likelihood

Parameter without curvature

Parameter with curvature

π | Default priors

- linear parameters

Is a prior 'flat'

Prior (tilted) shifts confidence

Effect on scalar interest parameter

Prior (tilted) translates p-value function

Overview

Linear	Reproducible
Curvature	Need Calibration

p-value: pragmatic

p-value: model-based

Bayes uses ONLY likelihood

Parameter without curvature

Parameter with curvature

Default priors

Intrinsically linear parameters

Is a prior 'flat'

Prior (tilted) shifts confidence

flat Tilt — Effect on scalar interest parameter

Prior (tilted) translates p-value function

p-value: pragmatic

p-value: model-based

Bayes uses ONLY likelihood

Parameter without curvature

Parameter with curvature

Default priors

Intrinsically linear parameters

Is a prior 'flat'

Prior (tilted) shifts confidence

Effect on scalar interest parameter

Prior (tilted) translates p-value function

Overview

Linear: Reproducible

Curvature: Need Calibration

p-value: pragmatic

Have data $\bar{y}$ Is true mean θ ?

Departure

Statistical units: $\frac{\bar{y}-\theta}{s_y/\sqrt{n}}$

p-value = $p(\theta) = \Phi\left(\frac{\bar{y}-\theta}{s_y/\sqrt{n}}\right)$ = where data is re theta

↑
- N 0 1 df CLT
- Bootstrap

Invert: $p(\theta) = \beta$

get: β-level confidence bound $\hat{\theta}_\beta$ ... (lower bound)

Use $\beta = 97.5\%$, $\beta = 2.5\%$ get ... 95% CI ... $(\hat{\theta}_{97.5\%}, \hat{\theta}_{2.5\%})$

General: e.g. Regression

$p(\beta_h) = \Phi\left(\frac{\hat{\beta}_h - \beta_h}{s_E \, c_{hh}^{1/2}}\right)$ $C = (X'X)^{-1} = (c_{ij})$

← CLT / BS

p-value: pragmatic

Have data $\bar{y}$ Is true mean θ ?

Departure

Standardized units = $\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}$

Statistical units = p-value = $p(\theta)$

Nor1 df CLT / Bootstrap $= \Phi\left(\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}\right)$ = "where data is re theta"

Invert: $p(\theta) = \beta$

get: β-level confidence bound $\hat{\theta}_\beta$ ... (lower bound)

Use $\beta = 97.5\%$, $\beta = 2.5\%$ get ... 95% CI ... $(\hat{\theta}_{97.5}, \hat{\theta}_{2.5})$

General: e.g. Regression

$p(\beta_h) = \Phi\left(\dfrac{\hat{\beta}_h - \beta_h}{s_E\, c_{hh}^{1/2}}\right)$ (CLT / BS)

$C = (X'X)^{-1} = (c_{ij})$

p-value: pragmatic

Have data $\bar{y}$ Is true mean θ ?

Departure

Standardized units = $\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}$

Statistical units = p-value = $p(\theta)$

No 1 df CLT / Bootstrap $= \Phi\left(\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}\right)$ = "Where data is re theta"

Invert: $p(\theta) = \beta$

get: β-level confidence bound $\hat{\theta}_\beta$ ... (lower bound)

Use $\beta = 97.5\%$, $\beta = 2.5\%$ get ... 95% CI ... $(\hat{\theta}_{97.5\%}, \hat{\theta}_{2.5\%})$

General: e.g. Regression

$p(\beta_r) = \Phi\left(\dfrac{\hat{\beta}_r - \beta_r}{s_E\, c_{rr}^{1/2}}\right)$ ← CLT, BS

$C = (X'X)^{-1} = (c_{ij})$

p-value: pragmatic

Have data $\bar{y}$ Is true mean θ ?

Departure

Standardized units = $\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}$

Statistical units = p-value = $p(\theta)$

Nol df CLT / Bootstrap → $= \Phi\left(\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}\right)$ = "Where data is re theta"

Invert $p(\theta) = \beta$ to get quantile...

Get: β-level confidence bound $\hat{\theta}_\beta$ ... (lower bound)

Use $\beta = 97.5\%$, $\beta = 2.5\%$ $\Rightarrow$ 95% CI ... $(\hat{\theta}_{97\frac{1}{2}}, \hat{\theta}_{2\frac{1}{2}\%})$

General: e.g. Regression

$p(\beta_h) = \Phi\left(\dfrac{\hat{\beta}_h - \beta_h}{s_E\, c_{hh}^{1/2}}\right)$ $C = (X'X)^{-1} = (c_{ij})$

CLT / BS

p-value: pragmatic

Have data $\bar{y}$ Is true mean θ ?

Departure

Standardized units = $\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}$

Statistical units = p-value = $p(\theta)$

No 1 df CLT / Bootstrap $= \Phi\left(\dfrac{\bar{y}-\theta}{s_y/\sqrt{n}}\right)$ = "Where data is re theta"

Invert: $p(\theta) = \beta$

Get: β-level confidence bound $\hat{\theta}_\beta$ ... (lower bound)

Use $\beta = 97.5\%$, $\beta = 2.5\%$ get .... 95% CI .... $(\hat{\theta}_{97\frac{1}{2}}, \hat{\theta}_{2\frac{1}{2}\%})$

General: e.g. Regression

$$p(\beta_r) = \Phi\left(\frac{\hat{\beta}_r - \beta_r}{s_E\, c_{rr}^{1/2}}\right) \qquad C = (X'X)^{-1} = (c_{ij})$$

CLT / BS

p-value: model based

$\ell(\theta) = \log L(\theta; y^{\circ})$

1) Likelihood ratio (SLR)

p-value: model based

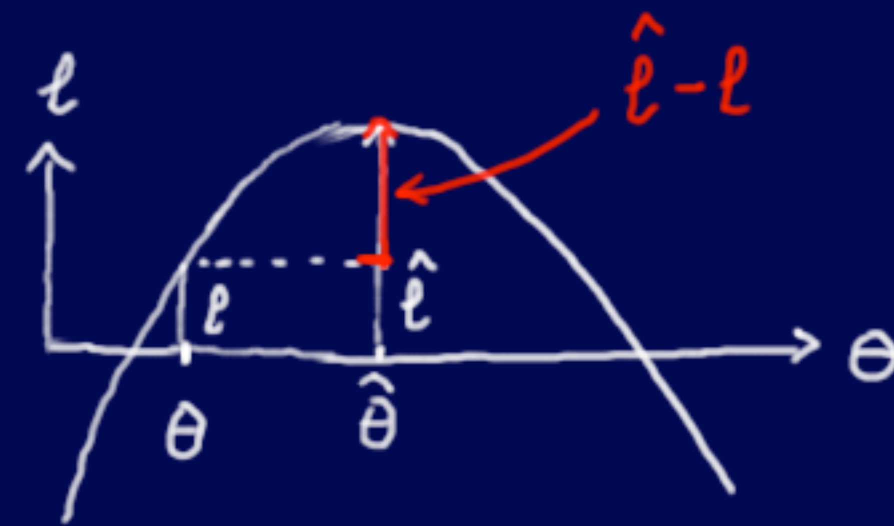
$\ell(\theta) = \log L(\theta; y^\circ)$

1) Likelihood ratio (SLR)

Regularity, Continuity

$r = \pm\sqrt{2(\hat{\ell} - \ell)} = r(\theta, y)$

$\uparrow$ sign$(\hat{\theta} - \theta)$



$p(\theta) = \Phi(r)$ ...... 1st order

p-value: model based

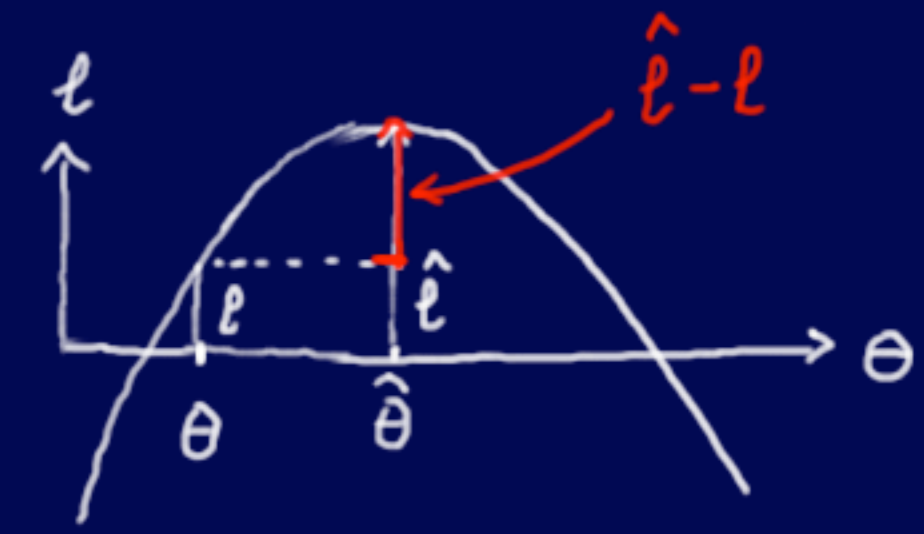
$\ell(\theta) = \log L(\theta; y^0)$

1) Likelihood ratio (SLR)

Regularity, Continuity

$r = \pm\sqrt{2(\hat{\ell} - \ell)} = r(\theta, y)$

↑ $\text{Sign}(\hat{\theta} - \theta)$



$p(\theta) = \Phi(r)$ ...... 1st order

2) Higher order: Need more than Likelihood!

Must have:
Exponential-type
reparameterization

mle departure in φ — Can add nuisance info

$q = \hat{j}_{\varphi\varphi}^{1/2}(\hat{\varphi} - \varphi) = q(\theta, y)$

$p(\theta) = \Phi\left(r - r^{-1}\log\frac{r}{q}\right)$ ...... 3rd order

Barndorff-Nielsen 1986

F & Reid 1993

Brazzale Davison Reid 2007

Need
- Likelihood
- Reparameterization (Data based)

$2 \cdot 10^6$ MCMC

p-value: model based

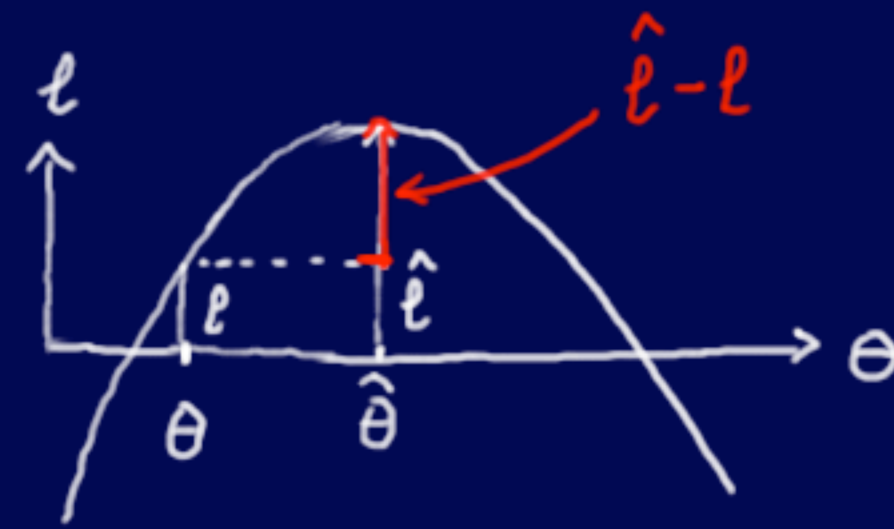
$\ell(\theta) = \log L(\theta; y^\circ)$

1) Likelihood ratio (SLR)

Regularity, Continuity

$r = \pm\sqrt{2(\hat{\ell} - \ell)} = r(\theta, y)$

↑ $\text{sign}(\hat{\theta} - \theta)$



$p(\theta) = \Phi(r)$ …… 1st order

2) Higher order: Need more than Likelihood!

Must have: Exponential-type reparameterization

mle departure in φ — Can 'add' nuisance info

$q = \hat{j}_{\varphi\varphi}^{1/2} (\hat{\varphi} - \varphi) = q(\theta, y)$

$p(\theta) = \Phi\left(r - r^{-1}\log\frac{r}{q}\right)$ …… 3rd order

Barndorff-Nielsen 1986

F & Reid 1993

Brazzale Davison Reid 2007

Need
- Likelihood
- Reparameterization (Data based)

$2 \cdot 10^6$ MCMC

p-value : model based

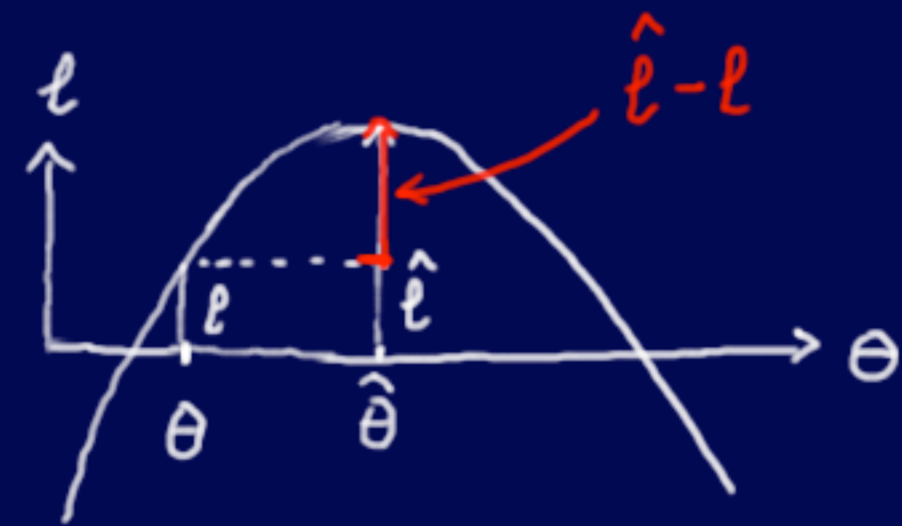
$\ell(\theta) = \log L(\theta; y^\circ)$

1) Likelihood ratio (SLR)

Regularity, Continuity

$r = \pm\sqrt{2(\hat{\ell} - \ell)} = r(\theta, y)$

↑ $\mathrm{Sign}(\hat{\theta} - \theta)$



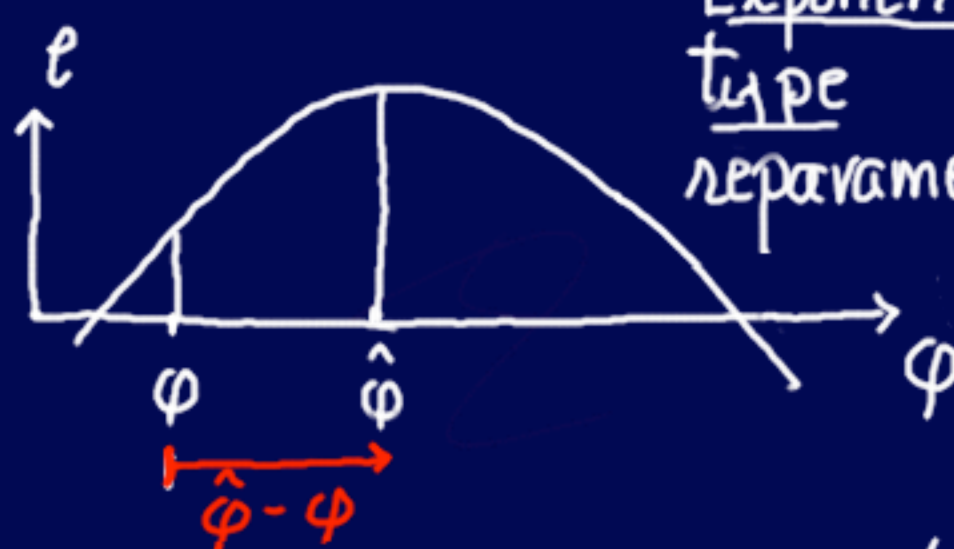
$p(\theta) = \Phi(r)$ ...... 1st order

2) Higher order : Need more than Likelihood ! ..... Daniels 1949

mle departure in φ — Can 'add' nuisance info

$q = \hat{j}_{\varphi\varphi}^{1/2}(\hat{\varphi} - \varphi) = q(\theta, y)$

$p(\theta) = \Phi\left(r - r^{-1}\log\frac{r}{q}\right)$ ...... 3rd order



Must have: Exponential-type reparameterization

Barndorff-Nielsen 1986

F & Reid 1993

Brazzale Davison Reid 2007

Need
- Likelihood
- Reparameterization (Data based)

$2 \cdot 10^6$ MCMC

Bayes uses ONLY likelihood

$\pi(\theta|y^0) = \pi(\theta) L(\theta; y^0)$ ...... Can it do as well?

Types of prior!

1 Mathematical Original Bayes 1763; Jeffreys; Bernardo; default*

- Flat in 'some' metric ...

2 Subjective/opinion ...

- Nothing of substance says: prior has to be used to analyse! ⇒ Can record
  - opinion
  - frequentist

  separately

3 Real prior

- There is/was a real frequency source $\pi(\theta)$ ⇒ Don't have to combine
  - Can record 'f' and prior separately

All goes back to: Mathematical

Here consider: Mathematical

* F Reid Marras Yi 2010 JRSSB

Bayes uses ONLY likelihood

$\pi(\theta|y^0) = \pi(\theta) L(\theta; y^0)$ …… Can it do as well?

<u>Types of prior</u>!

1 <u>Mathematical</u> Original Bayes 1763; Jeffreys; Bernardo; default*

- Flat in 'some' metric…

* F Reid Marras Yi 2010 JRSS B

Bayes uses ONLY likelihood

$\pi(\theta \mid y^0) = \pi(\theta) L(\theta; y^0)$ ...... Can it do as well?

<u>Types of prior</u>!

1 <u>Mathematical</u> Original Bayes 1763; Jeffreys; Bernardo; default*

- Flat in 'some' metric...

2 <u>Subjective</u> /opinion...

- Nothing of substance says: prior <u>has</u> to be used to analyze!

⇒ | <u>Can</u> record <u>separately</u>
- <u>opinion</u>
- <u>frequentist</u>

---

3 <u>Real</u> prior

- There <u>is</u>/<u>was</u> a real frequency source $\pi(\theta)$

⇒ | <u>Don't</u> <u>have</u> to combine
- <u>Can</u> record "f" and prior <u>separately</u>

All goes back to: <u>Mathematical</u>

Here consider: <u>Mathematical</u>

* F Reid Marras Yi 2010 JRSSB

Bayes uses ONLY likelihood

$\pi(\theta|y^0) = \pi(\theta) L(\theta; y^0)$ ...... Can it do as well?

Types of prior!

1 Mathematical Original Bayes 1763; Jeffreys; Bernardo; default*

- Flat in 'some' metric...

2 Subjective/opinion...

- Nothing of substance says: prior has to be used to analyze!

⇒ Can record
- opinion
- frequentist

separately

3 Real prior

- There is/was a real frequency source $\pi(\theta)$

⇒ Don't have to combine

Can record separately
- "real"
- frequentist

All goes back to: Mathematical

Here consider: Mathematical

* F Reid Marras Yi 2010 JRSSB

Bayes uses ONLY likelihood

$\pi(\theta|y^0) = \pi(\theta)L(\theta;y^0)$ ...... Can it do as well?

<u>Types of prior</u>!

1 <u>Mathematical</u> Original Bayes 1763; Jeffreys; Bernardo; default*

- Flat in 'some' metric...

2 <u>Subjective</u> /opinion...

- Nothing of substance says: prior <u>has</u> to be used to analyze! ⇒ | Can record - <u>opinion</u> - <u>frequentist</u> | <u>separately</u>

3 <u>Real</u> prior

- There <u>is</u> /<u>was</u> a real frequency source $\pi(\theta)$ ⇒ | <u>Don't</u> <u>have</u> to combine - <u>Can</u> record "f" and prior <u>separately</u>

All goes back to: <u>Mathematical</u>

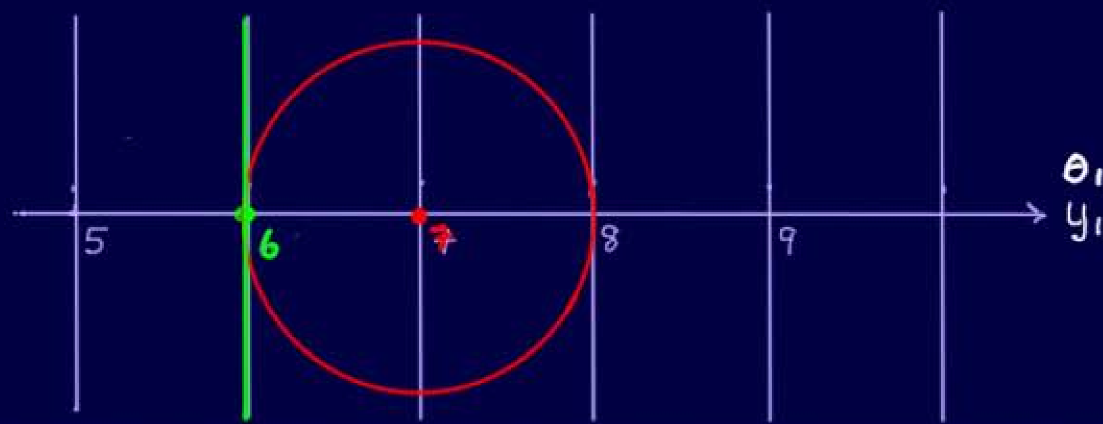
Here consider: <u>Mathematical</u>

* F Reid Marras Yi 2010 JRSS B

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}6\\ 0\end{pmatrix}$

Parameter: <u>Linear</u> θ_1 Assess $\theta_1 = 7$ Departure $y_1 - \theta_1$

freq



Normal at True $\begin{pmatrix}7\\ 0\end{pmatrix}$

One-σ contour

Bayes

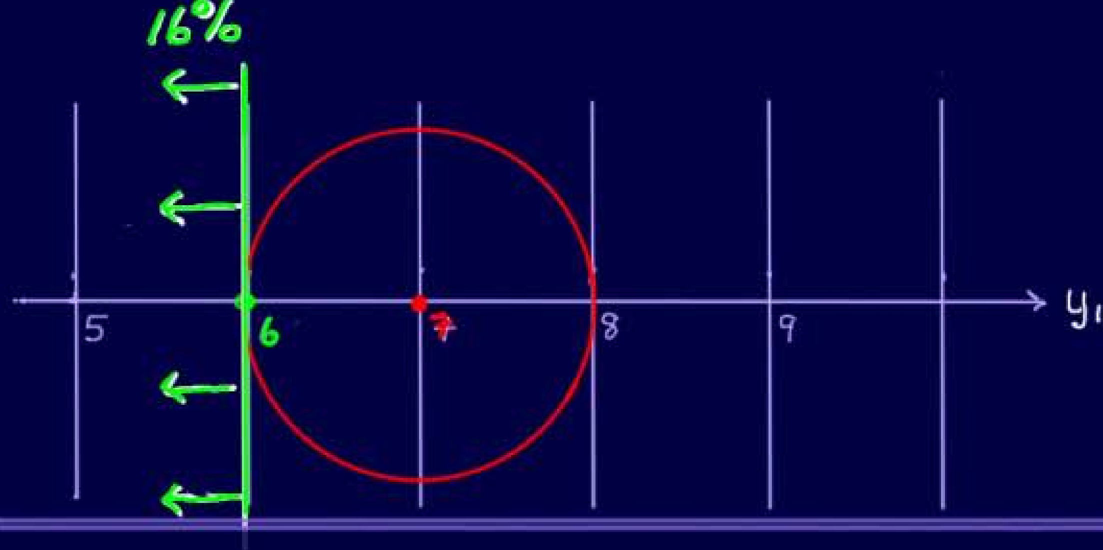
Normal at data $\begin{pmatrix}6\\ 0\end{pmatrix}$

One-σ contour

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}6\\0\end{pmatrix}$

Parameter: <u>No curvature</u> θ_1 Assess $\theta_1 = 7$ Departure $y_1 - \theta_1$

freq



Normal at true $\begin{pmatrix}7\\0\end{pmatrix}$

One-σ contour

Bayes

Normal at data $\begin{pmatrix}6\\0\end{pmatrix}$

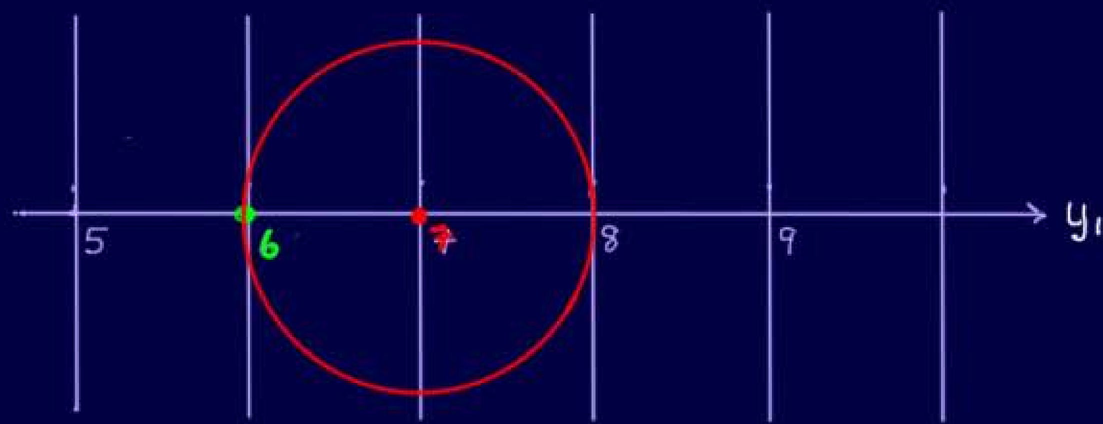
One-σ contour

f Prob. left of data $= p(\ ) = \Phi(-1) = 16\%$

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $=\begin{pmatrix}6\\0\end{pmatrix}$

Parameter: <u>No curvature</u> θ_1 Assess $\theta_1 = 7$ Departure $y_1 - \theta_1$

freq



Normal at True $\begin{pmatrix}7\\0\end{pmatrix}$
One-σ contour

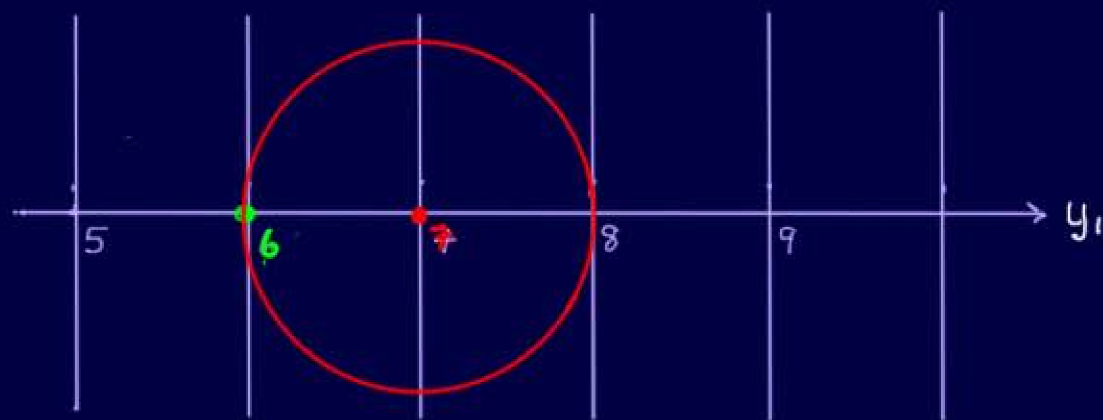
Bayes

Normal at data $\begin{pmatrix}6\\0\end{pmatrix}$
One-σ contour

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}6\\0\end{pmatrix}$

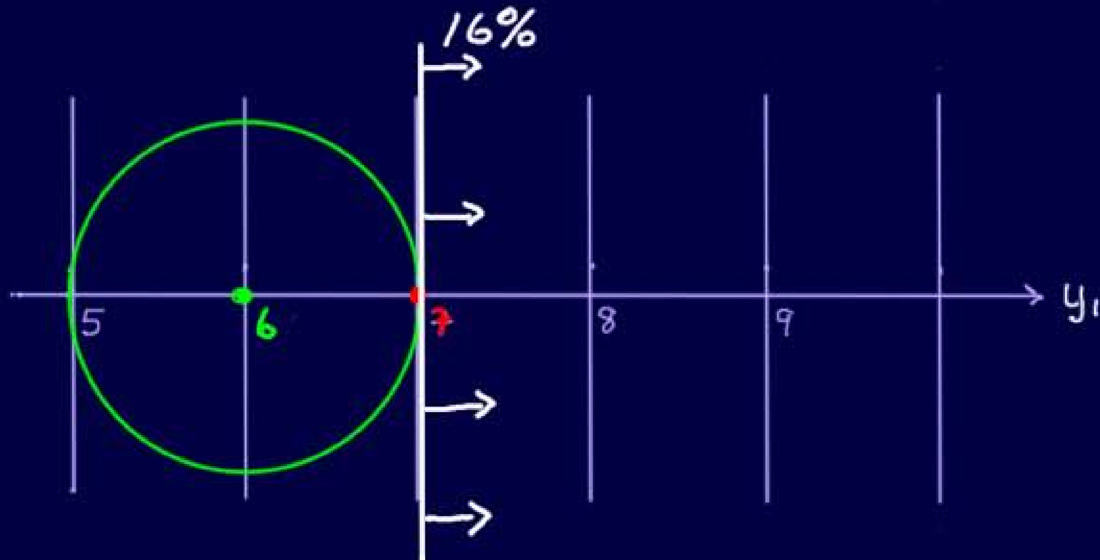
Parameter: <u>No curvature</u> θ_1 Assess $\theta_1 = 7$ Departure $y_1 - \theta_1$

freq



Normal at True $\begin{pmatrix}7\\0\end{pmatrix}$
One-σ contour

Bayes



Normal at data $\begin{pmatrix}6\\0\end{pmatrix}$
One-σ contour

f Prob. left of data = p( ) = $\Phi(-1)$ = 16%

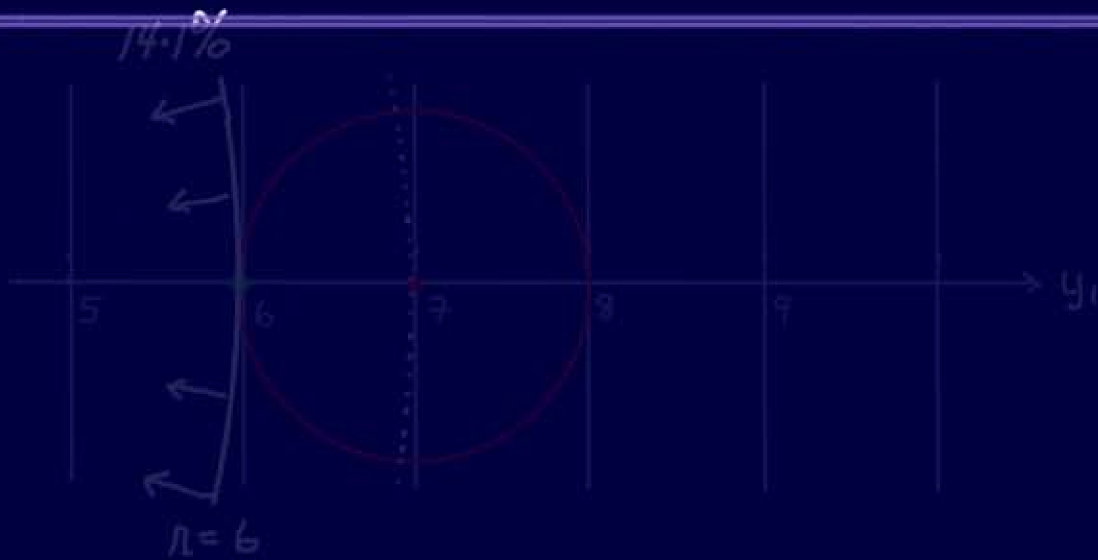
B Posterior right of parameter = Post. "survivor" = $s(7)$ = 16%

Equal!
Works!

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane   Data $= \begin{pmatrix}6\\0\end{pmatrix}$

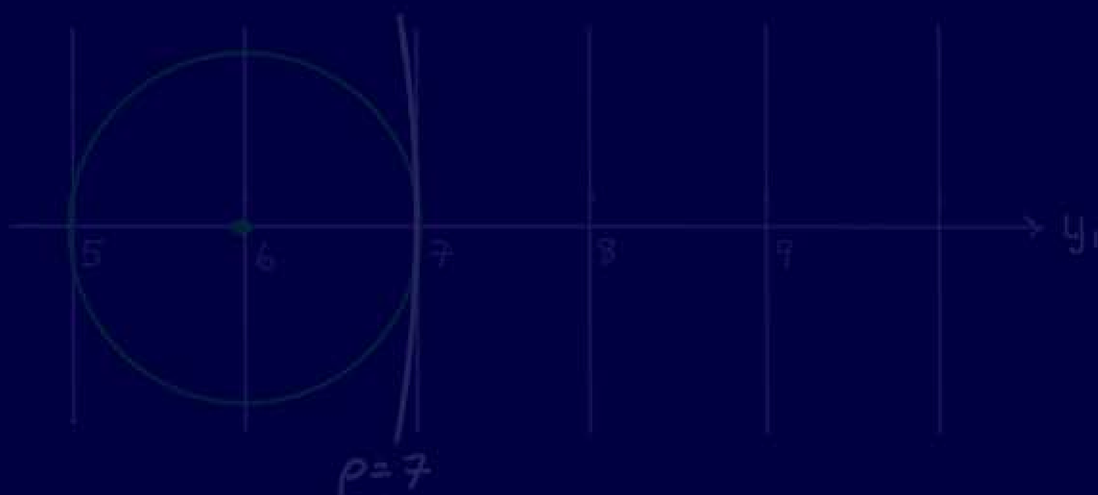
Parameter: <u>Curved</u> $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ Assess: $\rho = 7$

freq



Normal at true $\begin{pmatrix}7\\0\end{pmatrix}$
One-σ contour

Bayes

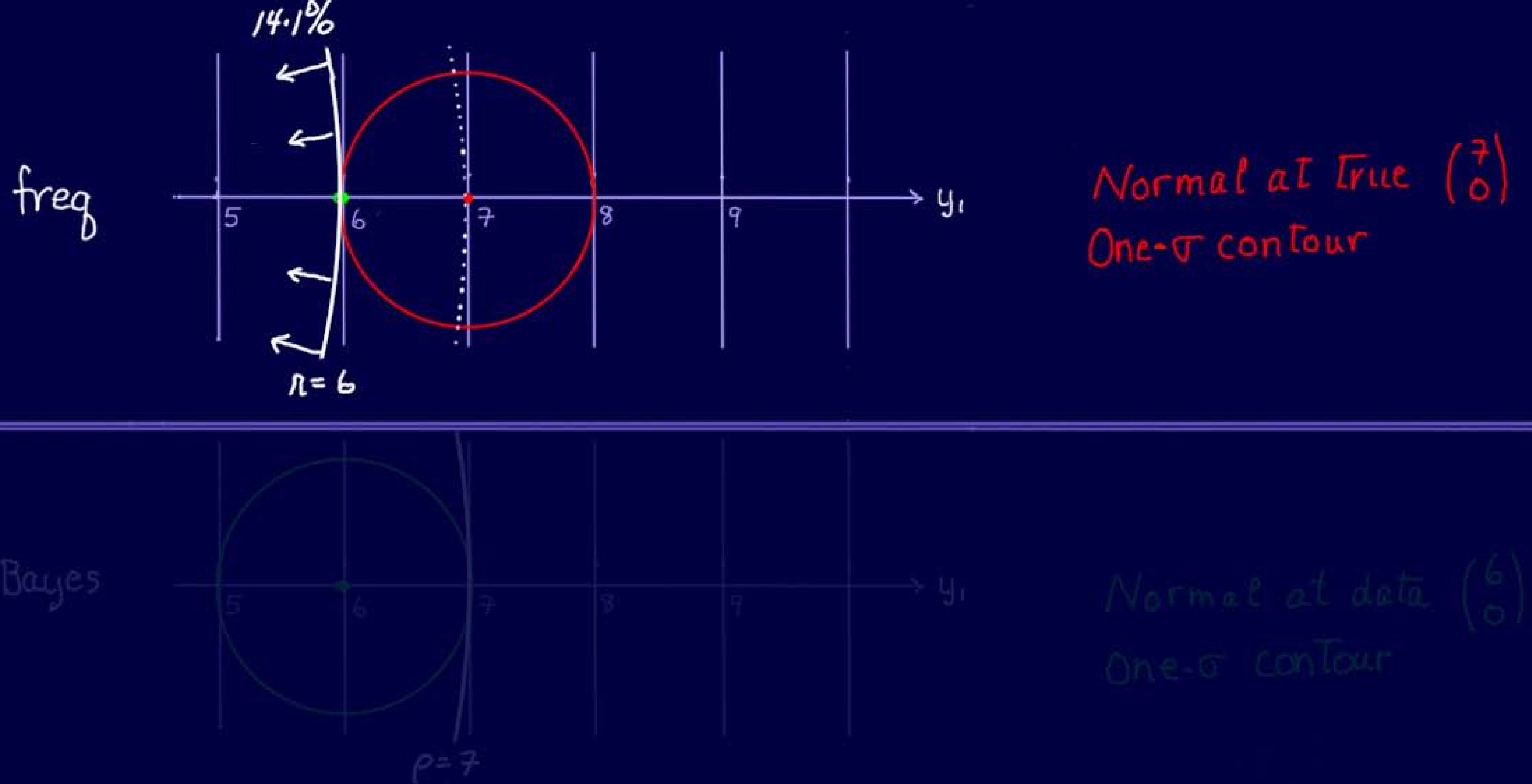


Normal at data $\begin{pmatrix}6\\0\end{pmatrix}$
One-σ contour

f Prob. left of data $= p(7) =$ 14.1% $NC\chi^2$ 2df 7^2

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}6\\ 0\end{pmatrix}$

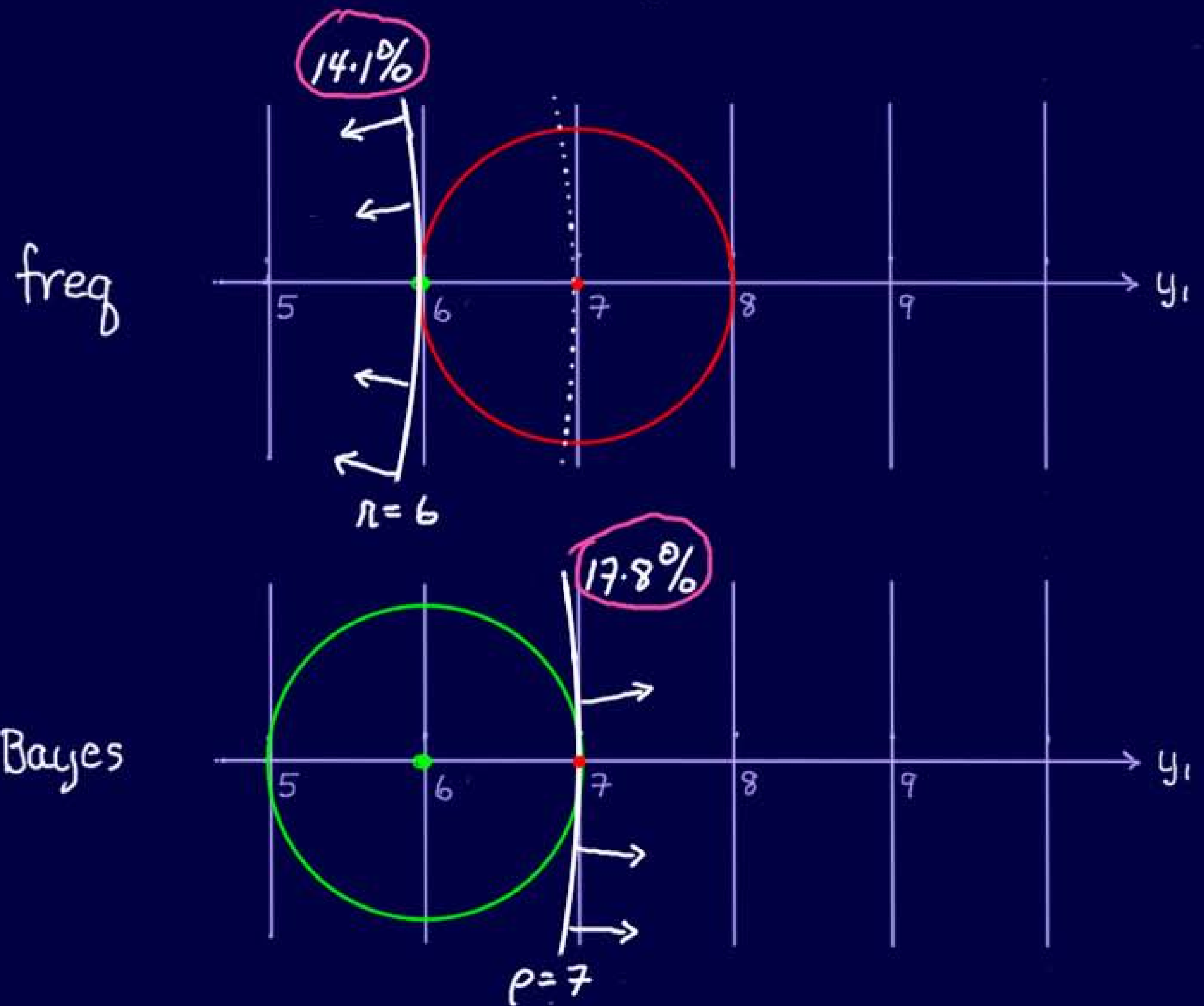
Parameter: Curved $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ Assess: $\rho = 7$



f Prob. left of data $= p(7) =$ 14.1% $NC\chi^2$ 2df 7^2

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}6\\0\end{pmatrix}$

Parameter: Curved $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ Assess: $\rho = 7$



Normal at True $\begin{pmatrix}7\\0\end{pmatrix}$
One-σ contour

Normal at data $\begin{pmatrix}6\\0\end{pmatrix}$
One-σ contour

f Prob. left of data $= p(7) = 14.1\%$ $NC\chi^2$ 2df 7^2

B Post. right of $(\rho = 7) = s(7) = 17.8\%$ $NC\chi^2$ 2df 6^2

F & Reid 2003 JSPI Curvature $\Rightarrow$ Bayes & freq. move oppositely!

Default priors

$f(y;\theta)$ $\quad y^0$ $\quad$ Regularity

$d\hat{\theta} = M(\theta)\, d\theta$ .... follows from continuity ... differentiate

get: Default prior $= \pi(\theta)d\theta = |M(\theta)|\, d\theta$ ... right invariant if available

... otherwise ... like right invariant

Ex: $N(X\beta, \sigma^2)$

$$M(\theta) = \begin{pmatrix} I & (\hat{\beta}^0-\beta)/2\sigma^2 \\ 0 & \hat{\sigma}^2/\sigma^2 \end{pmatrix} \qquad |M(\theta)|\, d\theta = \frac{\hat{\sigma}^2}{\sigma^2}\, d\beta\, d\sigma^2 = \text{right invar.}$$

get: Intrinsic linear parameter $d\varphi = M(\theta)d\theta$ near $\hat{\theta}^0$ $\quad$ 2nd order, local

Linear/location ... $\quad$ Frobenius (radial)

Ex: Linear in β

Linear in $\log \sigma$ ... radially from $\hat{\theta}^0$

F Reid Marras Yi 2010 JRSSB

Fraser F Fraser 2010 JSR Efron Vol.

Default priors

$f(y;\theta)$ $\quad y^0$ $\quad$ Regularity

$d\hat{\theta} = M(\theta)\, d\theta$ .... follows from continuity ... differentiate, 2nd

at data $\quad p \times p \quad p \times 1$

Default priors

$f(y;\theta)$ y^0 Regularity

$d\hat{\theta} = M(\theta)\, d\theta$ .... follows from continuity ... differentiate

at data $\quad p\times p \quad p\times 1$

get:

Default prior $= \pi(\theta)d\theta = |M(\theta)|\, d\theta$ ... right invariant if available
... otherwise .. like right invariant

Ex: $N \; X\beta \; \sigma^2$

$$M(\theta) = \begin{pmatrix} I & (\hat\beta^0-\beta)/2\sigma^2 \\ 0 & \hat\sigma^2/\sigma^2 \end{pmatrix} \qquad |M(\theta)|\,d\theta = \frac{\hat\sigma^2}{\sigma^2}\, d\beta\, d\sigma^2 = \text{right invar.}$$

get:

Intrinsic linear parameter $d\beta = M(\theta)d\theta$ near $\hat\theta^0$ — 2nd order, local
Linear/location ... — Frobenius (radial)

Ex: Linear in β
Linear in $\log\sigma$ ... radially from $\hat\theta^0$

F Reid Marras Yi 2010 JRSSB

Fraser F Fraser 2010 JSR Efron Vol.

Default priors

$f(y;\theta)$ $\quad y^0$ $\quad$ Regularity

$d\hat{\theta} = M(\theta)\, d\theta$ .... follows from continuity ... differentiate

at data $\quad p\times p \quad p\times 1$

get:

Default prior $= \pi(\theta)d\theta = |M(\theta)|\, d\theta$ ... right invariant if available
... otherwise .. like right invariant

Ex: N $X\beta$ σ^2

$$M(\theta) = \begin{pmatrix} I & (\hat{\beta}^0-\beta)/2\sigma^2 \\ 0 & \hat{\sigma}^2/\sigma^2 \end{pmatrix} \qquad |M(\theta)|\,d\theta = \frac{\hat{\sigma}^2}{\sigma^2}\, d\beta\, d\sigma^2 = \text{right invar.}$$

Get:

Intrinsic linear parameter $d\beta = M(\theta)d\theta$ near $\hat{\theta}^0$ $\quad$ 2nd order; local
Linear/location ... $\quad$ Frobenius (radial)

Ex: Linear in β
Linear in $\log\sigma$ ... radially from $\hat{\theta}^0$

F Reid Marras Yi 2010 JRSSB

Fraser F Fraser 2010 JSR Efron Vol.

Default priors

$f(y;\theta)$ y^0 Regularity

$d\hat{\theta} = M(\theta)\, d\theta$ .... follows from continuity ... differentiate

at data $\quad p\times p \quad p\times 1$

get:
Default prior $= \pi(\theta)d\theta = |M(\theta)|\, d\theta$ ... right invariant if available
... otherwise .. like right invariant

Ex: N $X\beta$ σ^2

$$M(\theta) = \begin{pmatrix} I & (\hat{\beta}^0-\beta)/2\sigma^2 \\ 0 & \hat{\sigma}^2/\sigma^2 \end{pmatrix} \qquad |M(\theta)|\, d\theta = \frac{\hat{\sigma}^2}{\sigma^2}\, d\beta\, d\sigma^2 = \text{right invar.}$$

Get:
Intrinsic linear parameter $d\beta = M(\theta)d\theta$ near $\hat{\theta}^0$ 2nd order; local
Linear/location ... Frobenius (radial)

Ex: Linear in β
Linear in $\log\sigma$ ... radially from $\hat{\theta}^0$

F Reid Marras Yi 2010 JRSS B

Fraser F Fraser 2010 JSR Efron Vol.

Is a prior intrinsically flat ?

Model: regular; continuous, ...

Check prior at $\hat{\theta}^0$

Calculate: $a^0 = \frac{d}{d\theta} \log \pi(\theta) \Big|_{\hat{\theta}^0}$ = log slope of prior at mle p×1

If $a^0 \neq 0$ ... then Bayes has bias $\pi(\theta) \approx e^{a^0 \theta}$ nr. mle

= 0 ... " " maybe OK ..... for linear parameters

Example: Suppose location model

Likelihood $L(\theta) = e^{-\theta^2/2}$

Confidence distn $L(\theta)d\theta = e^{-\theta^2/2} d\theta$

Prior (say) $\pi(\theta) = e^{a\theta}$

Posterior $\pi(\theta; y^0) = e^{-(\theta - a)^2/2}$

Prior shifts likelihood, Confidence reproducibility lost !

Is a prior intrinsically flat ?

Model: regular; continuous, ...

Check prior at $\hat{\theta}^0$

Calculate: $a = \frac{d}{d\theta} \log \pi(\theta) \Big|_{\hat{\theta}^0}$ = log slope of prior at mle $p \times 1$

Is a prior intrinsically flat ?

Model: regular; continuous, ...

Check prior at $\hat{\theta}^0$

Calculate: $a = \frac{d}{d\theta} \log \pi(\theta) \Big|_{\hat{\theta}^0}$ = log slope of prior at mle $p \times 1$

If $a^0 \neq 0$ ... then Bayes has bias $\qquad \pi(\theta) \approx e^{a'\theta}$ nr. mle

$= 0$ ... " " maybe OK ..... for linear parameters

Example: Suppose location model

Likelihood $\qquad L(\theta) = e^{-\theta^2/2}$

Confidence distn $\qquad L(\theta)d\theta = e^{-\theta^2/2} d\theta$

Prior (say) $\qquad \pi(\theta) = e^{a\theta}$

Posterior $\qquad \pi(\theta, y^0) = e^{-(\theta-a)^2/2}$

Prior shifts likelihood, Confidence reproducibility lost !

Is a prior intrinsically flat?

Model: regular; continuous, ...

Check prior at $\hat{\theta}^0$

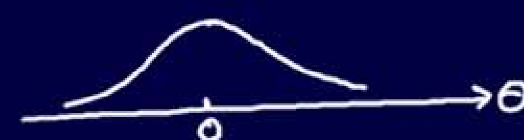
Calculate: $a = \frac{d}{d\theta} \log \pi(\theta) \Big|_{\hat{\theta}^0}$ = log slope of prior at mle $p \times 1$

If $a^0 \neq 0$ ... then Bayes has bias $\quad \pi(\theta) \approx e^{a'\theta}$ nr. mle

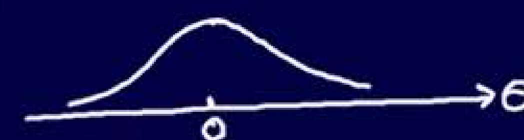
$\quad = 0$ ... " " maybe OK ..... for linear parameters

Example: Suppose location model

Likelihood $\quad L(\theta) = e^{-\theta^2/2}$

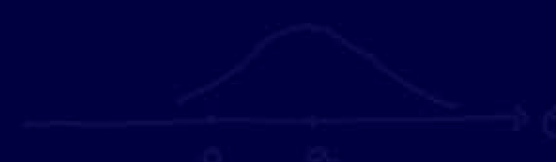


Confidence distn $\quad L(\theta)\,d\theta = e^{-\theta^2/2}\,d\theta$



Prior (say) $\quad \pi(\theta) = e^{a\theta}$

Posterior $\quad \pi(\theta, y^0) = e^{-(\theta-a)^2/2}$



Prior shifts likelihood, Confidence reproducibility lost!

Is a prior intrinsically flat ?

Model: regular; continuous, ...

Check prior at $\hat{\theta}^0$

Calculate: $a = \frac{d}{d\theta} \log \pi(\theta) \Big|_{\hat{\theta}^0}$ = log slope of prior at mle $p \times 1$

If $a^0 \neq 0$ ... then Bayes has bias $\quad \pi(\theta) \approx e^{a'\theta}$ nr. mle

$= 0$ ... " " maybe OK ..... for linear parameters

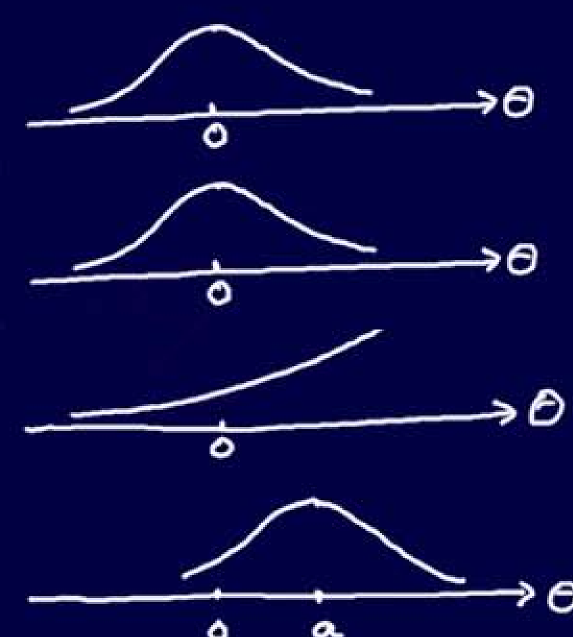
Example: Suppose location model

Likelihood $\quad L(\theta) = e^{-\theta^2/2}$

Confidence distn $\quad L(\theta)d\theta = e^{-\theta^2/2} d\theta$

Prior (say) $\quad \pi(\theta) = e^{a\theta}$

Posterior $\quad \pi(\theta; y^0) = e^{-(\theta-a)^2/2}$



Prior shifts likelihood, Confidence reproducibility lost !

Prior shifts Confidence posterior ... Vector case

<u>General</u> $f(y;\theta)$ regular

<u>Center</u> θ at $\hat{\theta}^{\circ} = 0$

<u>Scale</u> $\hat{j} = I$

* <u>Linearized near</u> $\hat{\theta}^{\circ}$. See #35

General prior: $\pi(\theta) = \exp\{a_\theta'\theta/n^{1/2} + \theta' a_{\theta\theta}\theta/2n\}$ — 2nd order: Ignore quadratic

Likelihood $L(\theta) = e^{-\theta^2/2} \Rightarrow N^{+}(\hat{\theta}^{\circ}; I)$ $\left\{\begin{array}{l} N\ O\ I \\ \text{with cubic } n^{-1/2} \end{array}\right.$

Confidence dist'n $L(\theta)\,d\theta = e^{-\theta^2/2}\,d\theta \Rightarrow N^{+}(\hat{\theta}^{\circ}; I)$

Prior (say) $\pi(\theta) = e^{a\theta} \Rightarrow \exp\{a'\theta\}$

Posterior $\pi(\theta, y^{\circ}) = e^{-|\theta - a|^2/2} \Rightarrow N'(\hat{\theta}^{\circ} + a; I)$ — 2nd

Posterior with <u>flat</u> prior <u>gives</u> confidence (for linear parameters)

" " <u>tilt</u> $\exp\{a'\theta\}$ <u>shifts</u> 'confidence' dis'tn

$\Rightarrow$
- <u>NOT</u> confidence !
- <u>NOT</u> reproducible !
- <u>NOT</u> "probability" !

Prior shifts Confidence posterior ... Vector case

General $f(y;\theta)$ regular — **Center** θ at $\hat\theta^0 = 0$; **Scale** $\hat{j} = I$

* **Linearized near** $\hat\theta^0$. See #35

general prior: $\pi(\theta) = \exp\{a_\theta'\theta/n^{1/2} + \theta' a_{\theta\theta}\theta/2n\}$ — 2nd order: Ignore quadratic

---

Likelihood $L(\theta) = e^{-\theta^2/2} \Rightarrow N^+(\hat\theta^0; I)$ {N 0 1 with cubic $n^{-1/2}$}

Confidence distn $L(\theta)d\theta = e^{-\theta^2/2}d\theta \Rightarrow N^+(\hat\theta^0; I)$

Prior (say) $\pi(\theta) = e^{a\theta} \Rightarrow \exp\{a'\theta\}$

Posterior $\pi(\theta, y^0) = e^{-|\theta-a|^2/2} \Rightarrow N(\hat\theta^0 + a; I)$ 2nd

Posterior with flat prior gives confidence (for linear parameters)

" " tilt $\exp[a'\theta]$ shifts 'confidence' distn

$\Rightarrow$ NOT confidence !

NOT reproducible !

NOT "probability" !

Prior shifts Confidence posterior ... Vector case

General $f(y;\theta)$ regular

Center θ at $\hat{\theta}^{o} = 0$

Scale $\hat{j} = I$

* Linearized near $\hat{\theta}^{o}$. See #35

general prior: $\pi(\theta) = \exp\{a_\theta'\theta/n^{1/2} + \theta' a_{\theta\theta}\theta/2n\}$ 2nd order Ignore quadratic

Likelihood $L(\theta) = e^{-\theta^2/2} \Rightarrow N^+(\hat{\theta}^{o}; I)$ $\left\{\begin{array}{l} N\ 0\ I \\ \text{with cubic } n^{-1/2} \end{array}\right.$

Confidence distn $L(\theta)d\theta = e^{-\theta^2/2}d\theta \Rightarrow N^+(\hat{\theta}^{o}; I)$

Prior shifts Confidence posterior ... Vector case

General $f(y;\theta)$ regular — Center θ at $\hat\theta^0 = 0$; Scale $\hat{j} = I$

* Linearized near $\hat\theta^0$. See #35

general prior: $\pi(\theta) = \exp\{a_\theta'\theta/n^{1/2} + \theta' a_{\theta\theta}\theta/2n\}$ — 2nd order: Ignore quadratic

Likelihood	$L(\theta) = e^{-\theta^2/2}$	$\Rightarrow$	$N^+(\hat\theta^0; I)$	$\{$ N 0 I with cubic $n^{-1/2}$
Confidence distn	$L(\theta)d\theta = e^{-\theta^2/2}d\theta$	$\Rightarrow$	$N^+(\hat\theta^0; I)$	
Prior (say)	$\pi(\theta) = e^{a\theta}$	$\Rightarrow$	$\exp\{a'\theta\}$	
Posterior	$\pi(\theta;y^0) = e^{-\lvert\theta-a\rvert^2/2}$	$\Rightarrow$	$N'(\hat\theta^0 + a; I)$	2nd

Posterior with flat prior gives confidence (for linear parameters)

" " tilt $\exp[a'\theta]$ shifts 'confidence' distn

$\Rightarrow$ NOT confidence ! NOT reproducible ! NOT "probability" !

Prior shifts Confidence posterior

General $f(y;\theta)$ regular — Center θ at $\hat{\theta}^0 = 0$

Scale $\hat{j} = I$

Linearized near $\hat{\theta}^0$. See #35

general prior: $\pi(\theta) = \exp\{a_0'\theta/n^{1/2} + \theta' a_{\theta\theta}\theta/2n\}$ — 2nd order, Ignore quadratic

Likelihood $L(\theta) = e^{-\theta^2/2} \Rightarrow N^+(\hat{\theta}^0; I)$ — $\{ N\ 0\ I$ with cubic $n^{-1/2}$

Confidence distn $L(\theta)d\theta = e^{-\theta^2/2}d\theta \Rightarrow N^+(\hat{\theta}^0; I)$

Prior (say) $\pi(\theta) = e^{a\theta} \Rightarrow \exp\{a'\theta\}$

Posterior $\pi(\theta; y^0) = e^{-|\theta-a|^2/2} \Rightarrow N'(\hat{\theta}^0 + a; I)$ — 2nd

Posterior with flat prior gives confidence (for linear parameters)

" " : tilt $\exp\{a'\theta\}$ shifts 'confidence' distn

$\Rightarrow$
- NOT confidence !
- NOT reproducible !
- NOT "probability" !

(a) Now consider <u>scalar</u> interest $\psi(\theta)$

Assume: Sta<u>ndar</u>dized & Li<u>neari</u>zed .... convenience

Marginal Likelihood/Conf $\quad N^{+}(\hat{\psi}^{0}, 1)$ $\quad$ with cubic is $^{-1/2}$

Posterior $\quad N^{+}(\hat{\psi}^{0} + \delta, 1)$ $\quad \delta = \psi' \cdot a_0$ Inner product

$\uparrow \psi'(\hat{\theta}^{0})$

Prior (with tilt) at mle $\Rightarrow$ "<u>Biased</u>" Confidence

Not reproducible!

(b) Again consider <u>scalar</u> interest $\psi(\theta)$ ... assume <u>linear</u>

Assume: <u>Just</u>: info - Standardized

<u>Know</u>: linearization exists ... !

For Bayes p-value $p_B(\psi)$ at 'centre' $\hat{\psi}^{0}$

coming from a prior with gradient a_0

$$p(\hat{\psi}^{0}) = \Phi^{+}(\hat{\psi} - \psi' a_0)$$

2nd order

... <u>Tilted prior translates</u> p-<u>value function</u>.

(a) Now consider scalar interest $\psi(\theta)$

Assume: Standardized & Linearized .... convenience

Marginal Likelihood/Conf $\quad N^+(\hat{\psi}^0, 1)$ $\quad$ with cubic $n^{-1/2}$

Posterior $\quad N^+(\hat{\psi}^0 + \delta, 1)$ $\quad \delta = \psi' \cdot a_0$ Inner product

$\uparrow \psi'(\hat{\theta}^0)$

Prior (with tilt) at mle $\Rightarrow$ "Biased" Confidence

Not reproducible !

(b) Again consider scalar interest $\psi(\theta)$ ... assume linear

Assume: Just: info - Standardized

Know: linearization exists ... !

For Bayes p-value $p_B(\psi)$ at 'centre' $\hat{\psi}^0$

coming from a prior with gradient a_0

$p(\hat{\psi}^0) = \Phi^+(\hat{\psi} - \psi' a_0)$ $\quad$ 2nd order

... Tilted prior translates p-value function.

(a) Now consider <u>scalar</u> interest $\psi(\theta)$

Assume: Standardized & Linearized .... convenience

Marginal Likelihood/Conf $\quad N^+(\hat{\psi}^0, 1)$ $\quad$ with cubic $n^{-1/2}$

Posterior $\quad N^+(\hat{\psi}^0 + \delta, 1)$ $\quad \rightarrow \quad \delta = \psi' \cdot a_0$ Inner product

$\quad \uparrow \psi'(\hat{\theta}^0)$

Prior (with tilt) at mle $\Rightarrow$ "Biased" Confidence

Not reproducible !

(b) Again consider <u>scalar</u> interest $\psi(\theta)$ ... assume <u>Linear</u>

Assume: <u>Just</u> info - Standardized

<u>Know</u>: linearization exists ... !

For Bayes p-value $p_B(\psi)$ at 'centre' $\hat{\psi}^0$

coming from a prior with gradient a_0

$$p(\hat{\psi}^0) = \Phi^+(\hat{\psi} - \psi' a_0) \qquad \text{2nd order}$$

... <u>Tilted prior translates</u> <u>p-value function</u>.

(a) Now consider scalar interest $\psi(\theta)$

Assume: Standardized & Linearized .... convenience

Marginal Likelihood/Conf $N^+(\hat\psi^0, 1)$ with cubic $n^{-1/2}$

Posterior $N^+(\hat\psi^0 + \delta, 1)$ → $\delta = \psi' \cdot a_0$ Inner product

$\psi' \leftarrow \psi'(\hat\theta^0)$

Prior (with tilt) at mle ⇒ "Biased" Confidence

Not reproducible !

(b) Again consider scalar interest $\psi(\theta)$ ... assume linear

Assume: Just: info - Standardized

Know: linearization exists ... !

For Bayes p-value $p_B(\psi)$ at centre $\hat\psi^0$

coming from a prior with gradient a_0

$p(\hat\psi^0) = \Phi^+(\hat\psi - \psi' a_0)$ 2nd order

... Tilted prior translates p-value function.

(a) Now consider <u>scalar</u> interest $\psi(\theta)$

Assume: Standardized & Linearized .... convenience

Marginal Likelihood/Conf $N^{+}(\hat{\psi}^{0}, 1)$ with cubic $n^{-1/2}$

Posterior $N^{+}(\hat{\psi}^{0} + \delta, 1)$ → $\delta = \psi' \cdot a_0$ Inner product

$\psi'(\hat{\theta}^{0})$

Prior (with tilt) at mle ⇒ "Biased" Confidence

Not reproducible !

(b) Again consider <u>scalar</u> interest $\psi(\theta)$ ... assume <u>linear</u>

Assume: Just: info - Standardized

Know: linearization exists ... !

For Bayes p - Value $p_B(\psi)$ at 'centre' $\hat{\psi}^{0}$

coming from a prior with gradient a_0

$$p(\hat{\psi}^{0}) = \Phi^{+}(\hat{\psi} - \psi' a_0)$$

2nd order

... Tilted prior "translates" p-value function.

Overview

1 Interest parameter curvature

Ex: $N\begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}5 \\ 0\end{pmatrix}$

Parameter: <u>Curved</u> $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ Assess: ρ

Bayes − frequentist $= s(\rho) - p(\rho) =$ Bias

Max Bias near mle $\hat{\rho}^{\circ}$

Here: near 8 %-age points

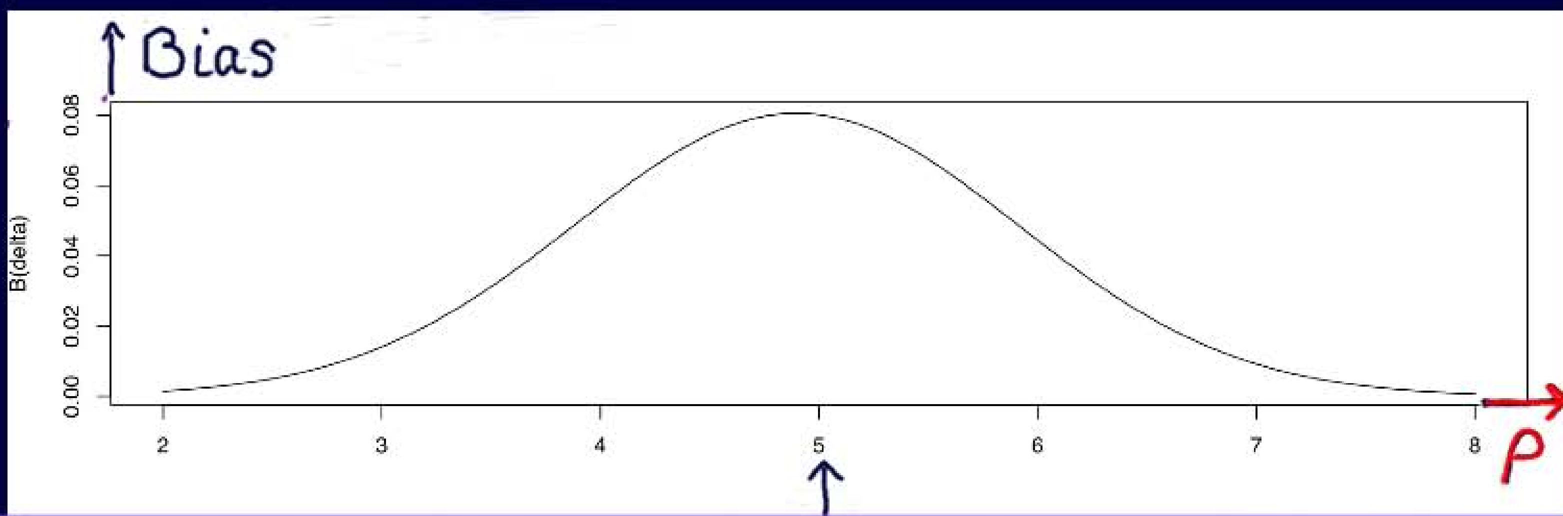
Overview

1 Interest parameter curvature

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}5\\ 0\end{pmatrix}$

Parameter: Curved $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ Assess: ρ

Bayes − frequentist $= s(\rho) - p(\rho) =$ Bias



Max Bias near mle $\hat{\rho}^0$

Here: near 8 %-age points

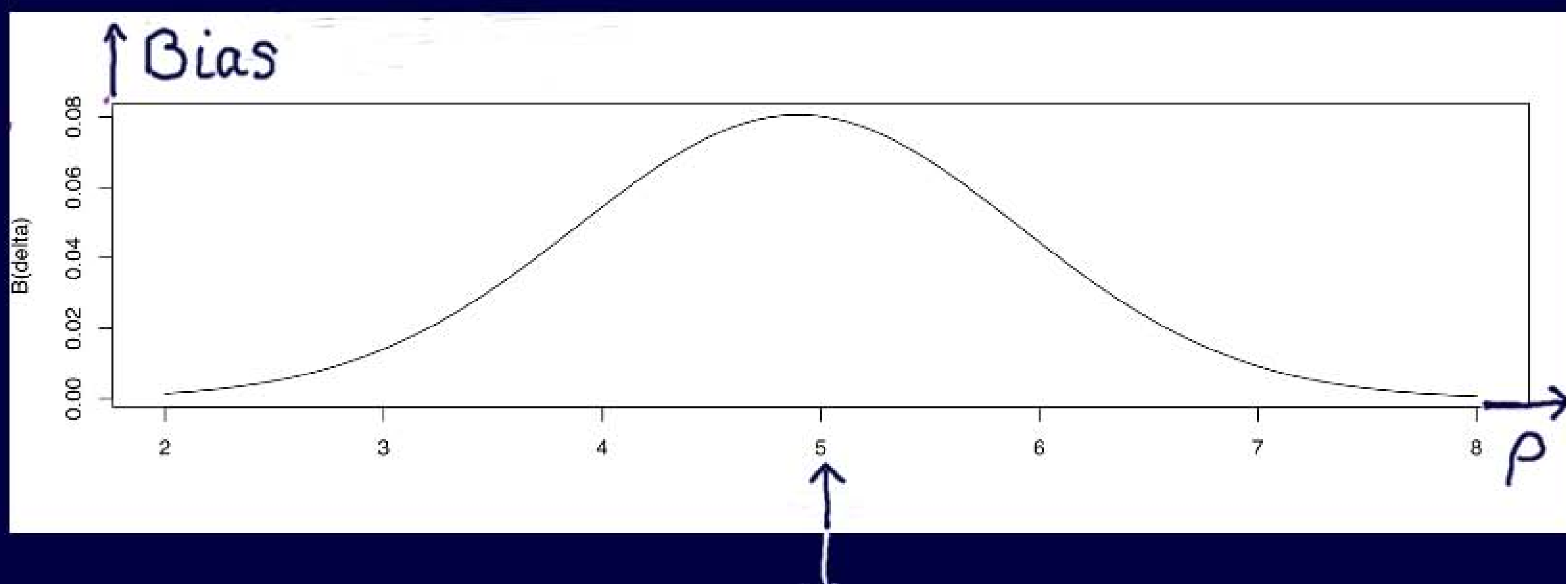
Overview

1 Interest parameter curvature

Ex: $N\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}; I$ Normal on plane Data $= \begin{pmatrix}5\\0\end{pmatrix}$

Parameter: Curved $\rho = (\theta_1^2 + \theta_2^2)^{1/2}$ Assess: ρ

Bayes – frequentist $= s(\rho) - p(\rho) =$ Bias



Max Bias near mle $\hat{\rho}^0$

Here: 8 %-age points

2. Model continuity gives:

$$d\hat{\theta} = M(\theta)\, d\theta$$

(at data; $p \times p$; $p \times 1$)

$d\hat{\theta}$ change at data

$d\theta$ change at θ in mod. deviations

---

and gives

Default prior $= |M(\theta)|\, d\theta$

2nd order (gen. left inv.)

All linear parameters

and gives

Linear parameter: $d\beta = M(\theta)\, d\theta$

$$\beta - \hat{\beta}^\circ = M(\hat{\theta}^\circ)\theta + \frac{1}{2}\theta' M_\theta(\hat{\theta}^\circ)\theta + \ldots$$

Radial Frobenius M_θ $p \times p \times p$

3. Prior shifts Confidence posterior

$$\pi(\theta, y^\circ) = e^{-(\theta - a_\theta)^2/2} \Rightarrow N^+(\hat{\theta}^\circ + a_\theta;\, I)$$

4 Prior shifts p-value for scalar interest

$$p(\hat{\psi}^\circ) = \Phi^+(\hat{\psi} - \psi' a_\theta)$$

Linearized coord.

2. Model <u>continuity</u> gives:

$$d\hat{\theta} = M(\theta)\, d\theta$$

at data $\quad p \times p \quad p \times 1$

$d\hat{\theta}$ change at data

$d\theta$ change at θ in mod. deviations

and gives

<u>Default prior</u> $= |M(\theta)|\, d\theta$

2nd order (gen. left. inv.)

All linear parameters

and gives

<u>Linear parameter</u>: $d\beta = M(\theta)\, d\theta$

$$\beta - \hat{\beta}^0 = M(\hat{\theta}^0)\theta + \frac{1}{2}\theta' M_\theta(\hat{\theta}^0)\theta +$$

Radial Frobenius

$M_\theta \quad p \times p \times p$

---

3. <u>Prior</u> shifts <u>Confidence</u> posterior

$$\pi(\theta, y^0) = e^{-(\theta - a_\theta)^2/2} \Rightarrow N^+(\hat{\theta}^0 + a_\theta; I)$$

4 Prior shifts p-<u>value</u> for scalar interest

$$p(\hat{\psi}^0) = \Phi^+(\hat{\psi} - \psi' a_0)$$

Linearized coord.

2. Model <u>continuity</u> gives:

$$d\hat{\theta} = M(\theta)\, d\theta$$

at data $\quad p\times p \quad p\times 1$

$d\hat{\theta}$ change at data

$d\theta$ change at θ in mod. deviations

and gives

<u>Default prior</u> $= |M(\theta)|\, d\theta$

2nd order (gen. left. inv.)

All linear parameters

and gives

<u>Linear parameter</u>: $d\beta = M(\theta)\, d\theta$

$$\beta - \hat{\beta}^{\circ} = M(\hat{\theta}^{\circ})\theta + \frac{1}{2}\theta' M_{\theta}(\hat{\theta}^{\circ})\theta +$$

Radial Frobenius

M_{θ} $p\times p\times p$

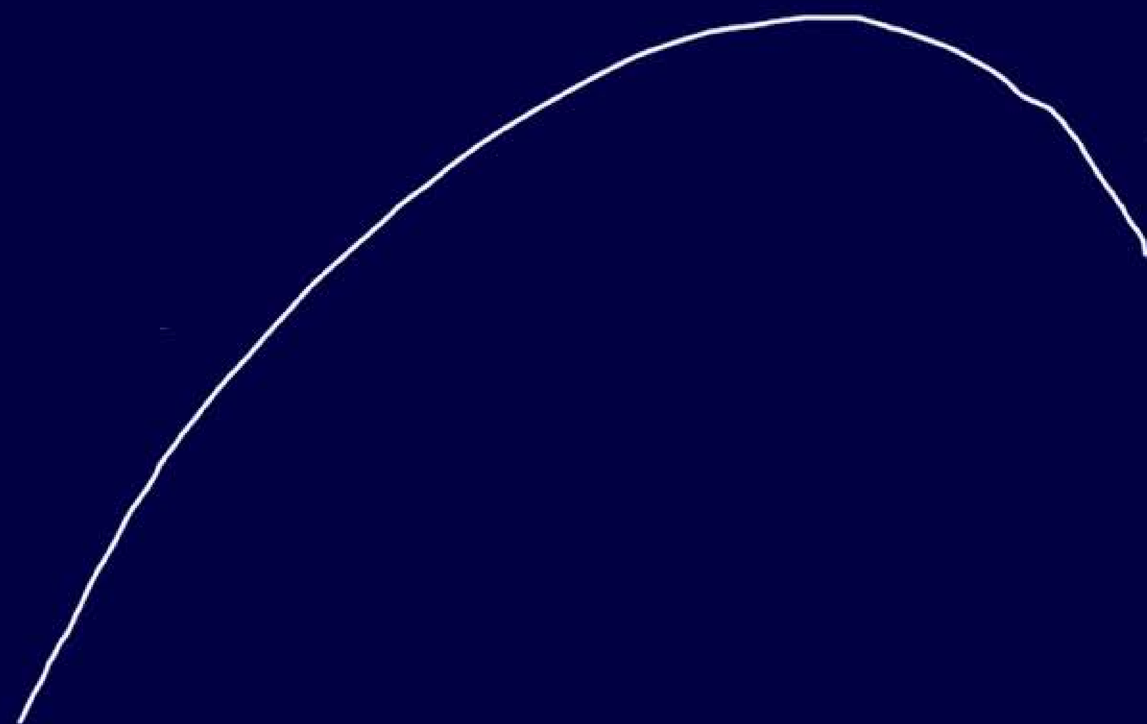
3. <u>Prior</u> shifts <u>Confidence</u> posterior

$$\pi(\theta; y^{\circ}) = e^{-|\theta - a|_{\theta}^{2}/2} \Rightarrow N^{+}(\hat{\theta}^{\circ} + a_{\theta}; I)$$

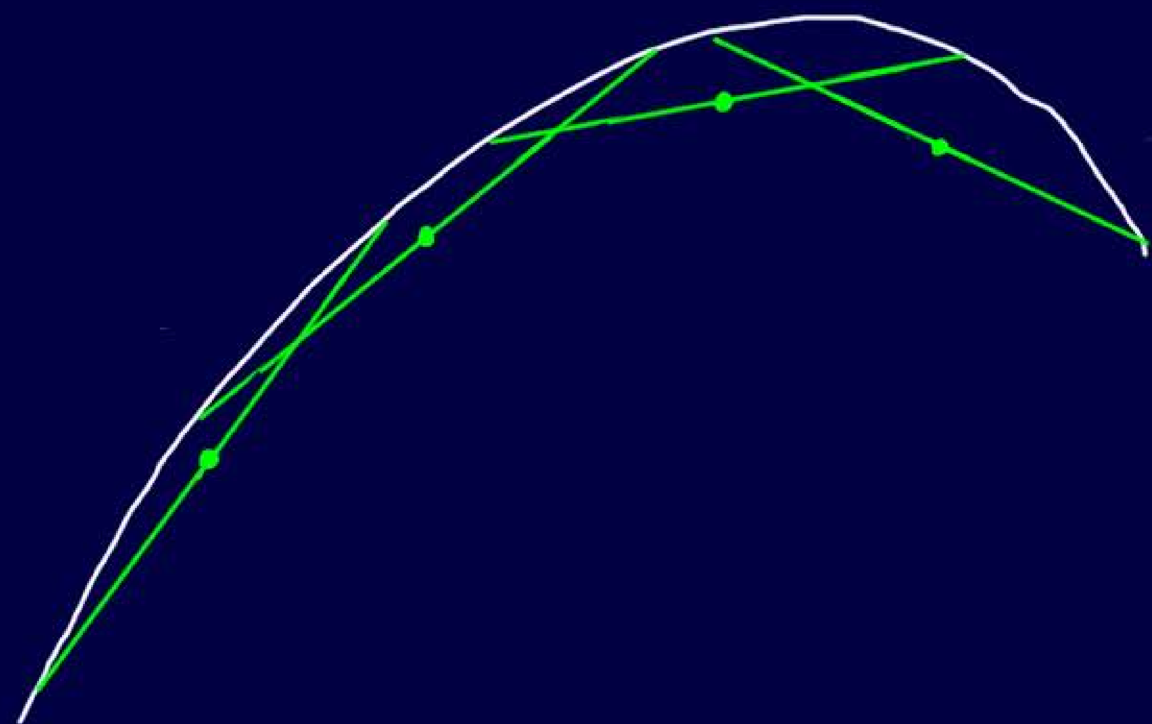
Prior shifts p-<u>value</u> for scalar interest

$$p(\hat{\psi}^{\circ}) = \Phi^{+}(\hat{\psi} - \psi' a_{\theta})$$

Linearized coord.

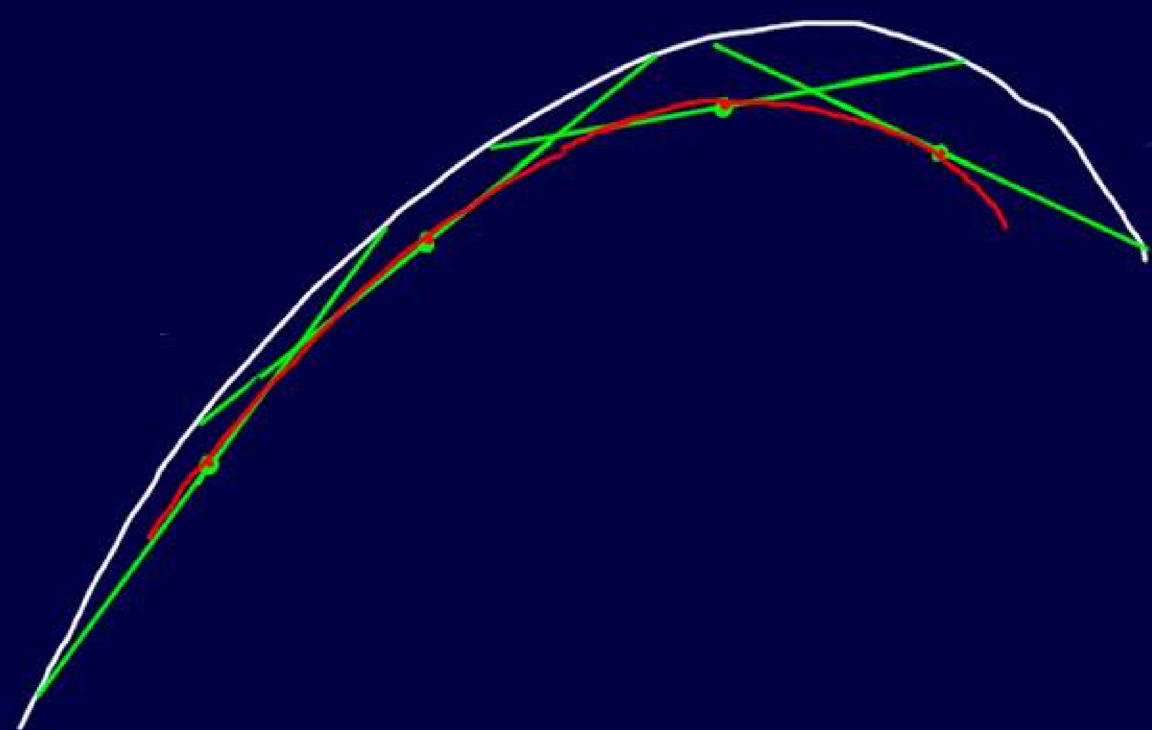


a curve



a curve

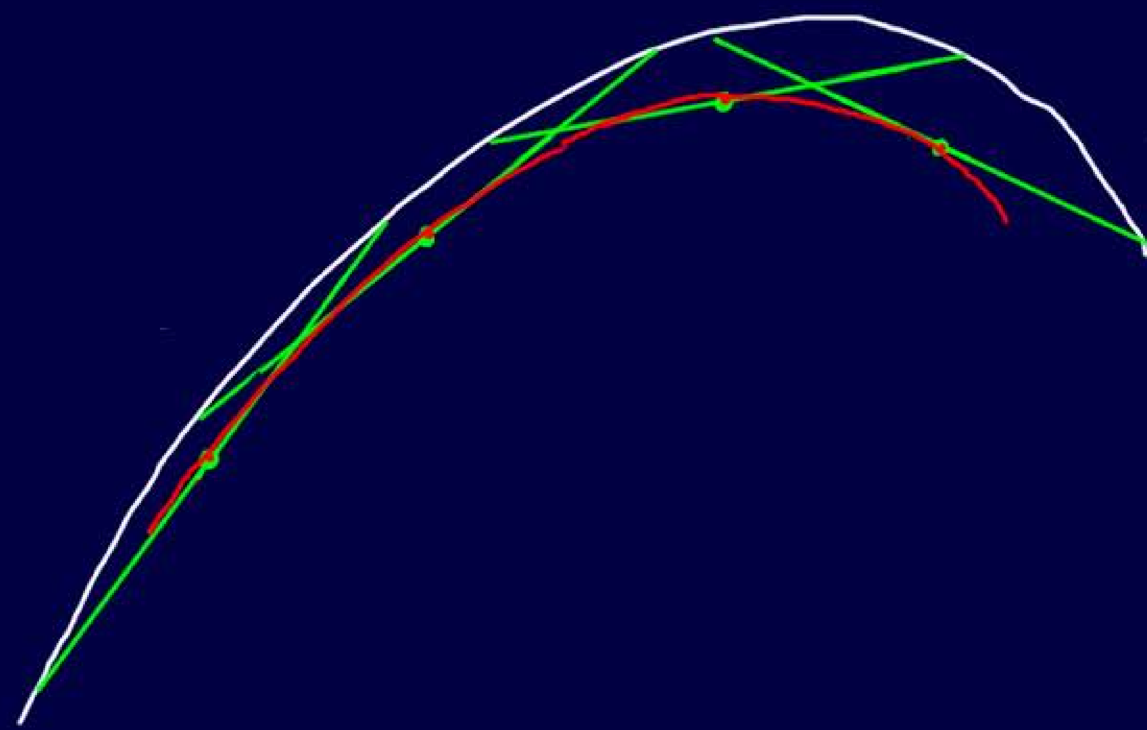
Average (of points on curve)



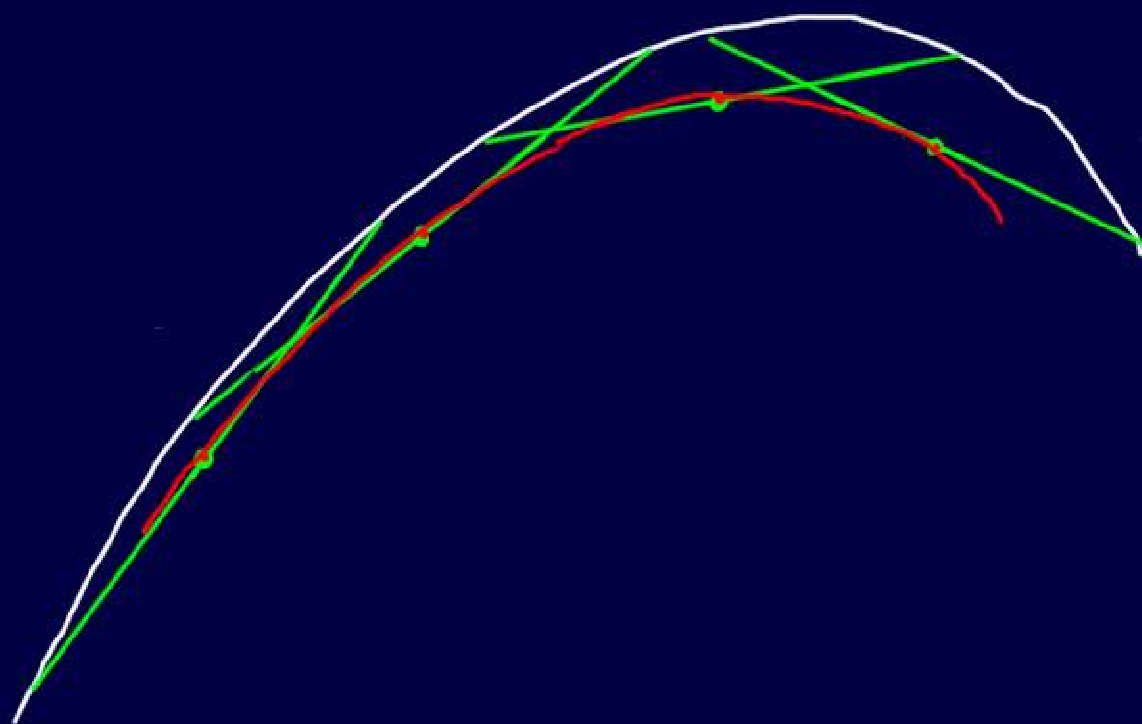
a curve

average (of points on curve)

a new curve



a model with curvature



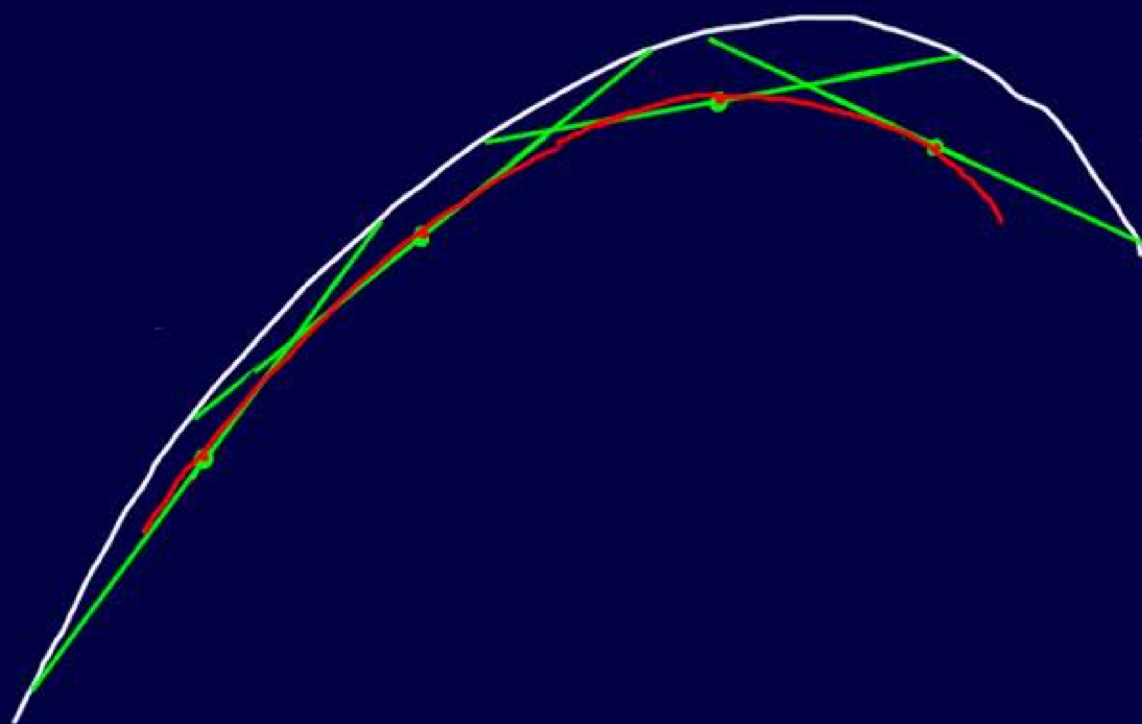
a curve

average (of points on curve)

a new curve

a model with curvature

Multilevel modelling (Bayes mixture)



a curve

Average (of points on curve)

a new curve

a model with curvature

Multilevel modelling (Bayes mixture)

Composite model (different)

Stainforth et al 2007 Phil Trans Royal Soc A

- Economist Aug 18 2007

Linearity $\Rightarrow$ Bayes reproducible

Curvature $\Rightarrow$ Bayes <u>biased</u>