Inference distributions for a parameter: are they calibrated?
(Do they mean what they say?)

D A S Fraser
Statistical Sciences U of Toronto

59th World Statistics Congress 2013
IPS028
August 27 2013

www.utstat.toronto.edu/dfraser/documents/HK-III.pdf

Topics: Distributions for a parameter

1) History
2) Examples: scalar θ
3) Difficulties: - Selection
4) - Curvature
5) "All relevant information re interest ψ"
6) Exact distribution for θ
7) Welch Peers and scalar Jeffreys
8) Vector parameter: second-order directional prior
9) Summary

1) Distribution for a parameter : History

- Arose from Bayes 1763 ... clearly

1) Distribution for a parameter: History

- Arose from Bayes 1763... clearly
  - Binomial data, parameter p
  - Suppose p had a random source

1) <u>Distribution</u> for a parameter: <u>History</u>

- Arose from Bayes 1763... clearly
  - Binomial data, parameter p
  - <u>Suppose</u> p had a <u>random</u> source
    - <u>no</u> objective reference
    - just convenience for <u>conditional</u> pro
    - expediency

1) <u>Distribution</u> for a parameter : <u>History</u>

- Arose from Bayes 1763 ... clearly
  - Binomial data, parameter p
  - <u>Suppose</u> p had a <u>random</u> source
    - <u>no</u> objective reference
    - just convenience for <u>conditional</u> pro
    - expediency

- Laplace (1812) Anticipated <u>confidence</u> via Non-informative priors
  - Saw that it <u>worked</u>, but not <u>why</u> it worked!

1) Distribution for a parameter: History

- Arose from Bayes 1763... clearly
  - Binomial data, parameter p
  - Suppose p had a random source
    - no objective reference
    - just convenience for conditional pro
    - expediency
- Laplace (1812) Anticipated confidence via Non-informative priors
  - Saw that it worked, but not why it worked!
- Fisher (1930) invented confidence via pivotal inversion $u = F(y^0; \theta)$

1) Distribution for a parameter: History

- Arose from Bayes 1763... clearly
  - Binomial data, parameter p
  - Suppose p had a random source
    - no objective reference
    - just convenience for conditional pro
    - expediency
- Laplace (1812) Anticipated confidence via Non-informative priors
  - Saw that it worked, but not why it worked!
- Fisher (1930) invented confidence via pivotal inversion $u = F(y^0; \theta)$
  - Neyman (1937) clarified confidence: use pivotal set

1) Distribution for a parameter : History

- Arose from Bayes 1763... clearly
  - Binomial data, parameter p
  - Suppose p had a random source
    - no objective reference
    - just convenience for conditional pro
    - expediency
- Laplace (1812) Anticipated confidence via Non-informative priors
  - Saw that it worked, but not why it worked!
- Fisher (1930) invented confidence via pivotal inversion $u = F(y^0; \theta)$
  - Neyman (1937) clarified confidence : use pivotal set
- Fisher lecture 2013 JSM Montreal : Overview on Fisher

1) Distribution for a parameter : History

- Arose from Bayes 1763 ... clearly
  - Binomial data, parameter p
  - Suppose p had a random source
    - no objective reference
    - just convenience for conditional pro
    - expediency
- Laplace (1812) Anticipated confidence via Non-informative priors
  - Saw that it worked, but not why it worked!
- Fisher (1930) invented confidence via pivotal inversion $u = F(y^0; \theta)$
  - Neyman (1937) clarified confidence : use pivotal set
- Fisher lecture 2013 JSM Montreal : Overview on Fisher
  - Peter Bickel ... on fiducial :
    "... like to say as little as possible" ... innuendo
    - no love lost:
    - Berkeley - Fisher

1) Distribution for a parameter: History

- Arose from Bayes 1763 ... clearly
  - Binomial data, parameter p
  - Suppose p had a random source { - no objective reference; - just convenience for conditional pro; - expediency }
- Laplace (1812) Anticipated confidence via Non-informative priors
  - Saw that it worked, but not why it worked!
- Fisher (1930) invented confidence via pivotal inversion $u = F(y^0; \theta)$
  - Neyman (1937) clarified confidence: use pivotal set
- Fisher lecture 2013 JSM Montreal: Overview on Fisher
  - Peter Bickel ... on fiducial:
    "... like to say as little as possible" ... innuendo { no love lost: Berkeley–Fisher }
- Fisher: Should get full credit for production of "confidence"!

1) Distribution for a parameter: History

- Arose from Bayes 1763... clearly
  - Binomial data, parameter p
  - Suppose p had a random source { - no objective reference; - just convenience for conditional pro; - expediency }
- Laplace (1812) Anticipated confidence via Non-informative priors
  - Saw that it worked, but not why it worked!
- Fisher (1930) invented confidence via pivotal inversion $u = F(y^0; \theta)$
  - Neyman (1937) clarified confidence: use pivotal set
- Fisher lecture 2013 JSM Montreal: Overview on Fisher
  - Peter Bickel ... on fiducial:
    "... like to say as little as possible" ... innuendo { no love lost: Berkeley - Fisher }
- Fisher: Should get full credit for production of "confidence"
- History of statistics: "Trash the other guy!"

2) Simple scalar examples:

$y \sim N(\theta; \sigma_0^2/n)$ $\quad$ $y \sim EV(\theta; 1)$ $\quad$ $\boxed{y \sim f(y-\theta)}$

Continuity has implications ...

2) Simple scalar examples:

$y \sim N(\theta; \sigma_0^2/n)$ $\qquad$ $y \sim EV(\theta; 1)$ $\qquad$ $\boxed{y \sim f(y-\theta)}$

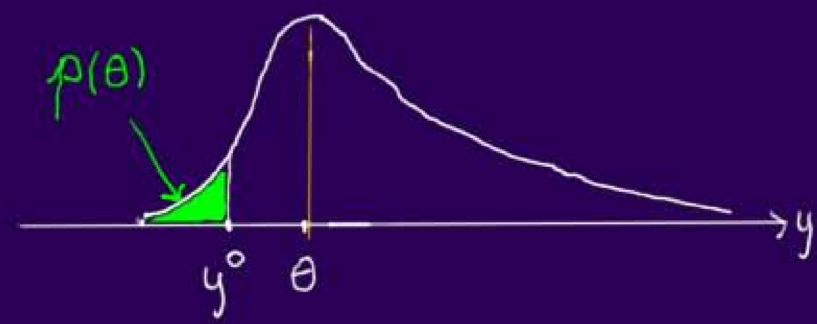
Continuity has implications ...

Consider: model & Data

$f(y-\theta)$ $\quad y^0$

p-value $= p(\theta) = F(y^0; \theta)$

= where "y^0 is re θ"



2) <u>Simple scalar</u> <u>examples</u>:

$y \sim N(\theta; \sigma_0^2/n)$ $\qquad$ $y \sim EV(\theta; 1)$ $\qquad$ $\boxed{y \sim f(y-\theta)}$

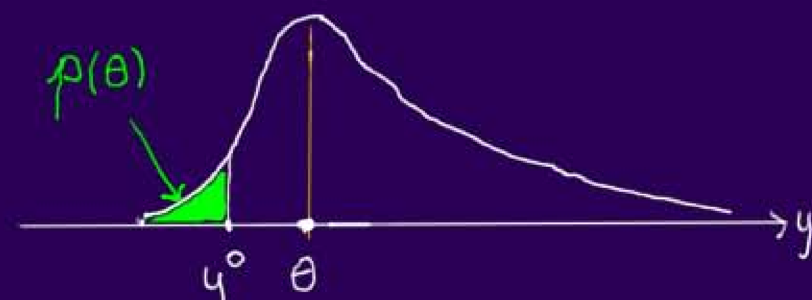
Continuity has implications ...

Consider: model & Data

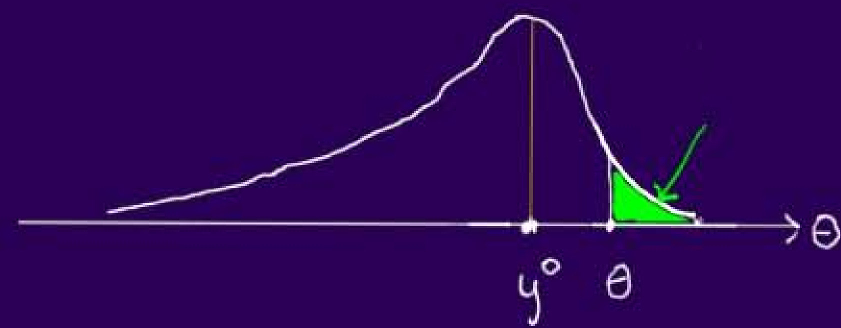
$f(y-\theta)$ $\quad$ y^0

p-value $= p(\theta) = F(y^0; \theta)$

= where "<u>y^0 is re θ</u>"



Posterior: (pivot or Bayes)



<u>Flip</u> dist'n

Put it at data <u>y^0</u>

2) Simple scalar examples:

$y \sim N(\theta; \sigma_0^2/n)$ $\qquad$ $y \sim EV(\theta; 1)$ $\qquad$ $\boxed{y \sim f(y-\theta)}$

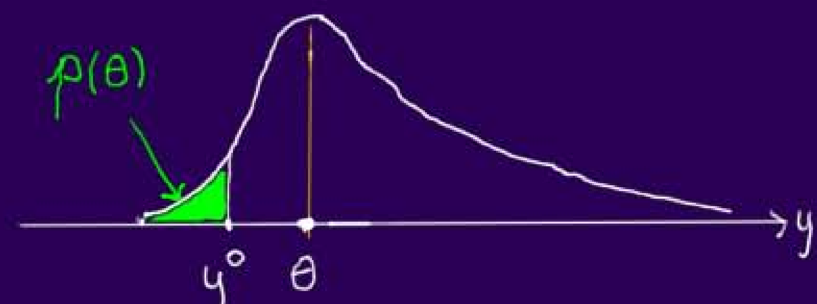
Continuity has implications ...

Consider: model & Data

$f(y-\theta)$ $\quad y^0$

p-value $= p(\theta) = F(y^0; \theta)$

= where "y^0 is re θ"

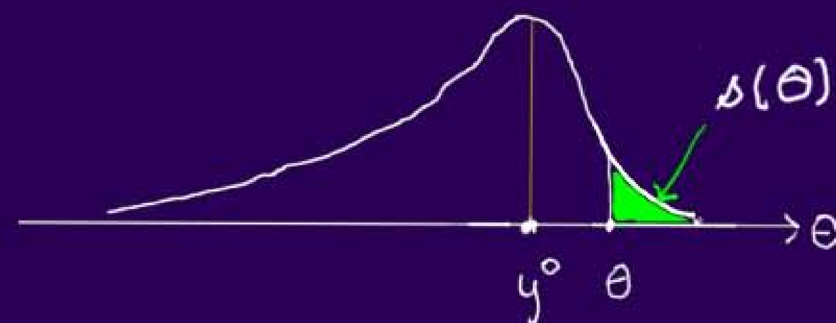


Posterior: pivot or Bayes

s-value $= s(\theta)$

= "parameter survivor value"

= where "θ is re y^0"



Flip dist'n

Put it at data y^0

2) Simple scalar examples:

$y \sim N(\theta; \sigma_0^2/n)$ $\quad$ $y \sim EV(\theta; 1)$ $\quad$ $\boxed{y \sim f(y-\theta)}$

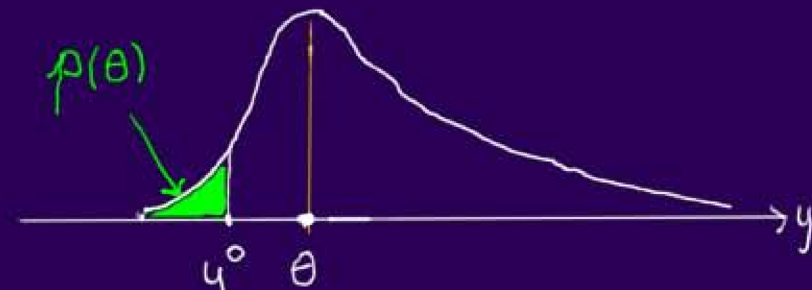
Continuity has implications ...

Consider: model & Data

$f(y-\theta)$ $\quad$ y^0

p-value $= p(\theta) = F(y^0; \theta)$

= where "y^0 is re θ"

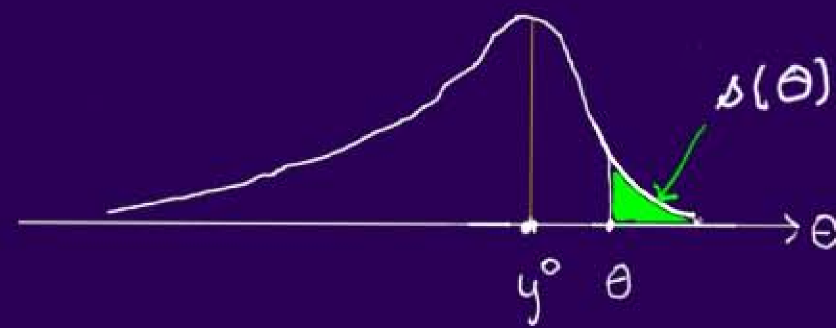


Posterior: pivot or Bayes

s-value $= s(\theta)$

= "parameter survivor value"

= where "θ is re y^0"



Flip dist'n

Put it at data y^0

$p(\theta) = s(\theta) \sim U(0,1)$

LH df for y = RH df for θ

Always so simple?

NB

Laplace anticipated?

3. Difficulties / Risks

Simple $y \sim N(\theta; \sigma_0^2/n)$ $\quad\quad \theta \sim N(y^0; \sigma_0^2/n)$

Any problems? Is it <u>really</u> probability: "$\theta \sim N(y^0; \sigma_0^2/n)$" ?

3. Difficulties / Risks

Simple $y \sim N(\theta; \sigma_0^2/n)$ $\quad\quad$ $\theta \sim N(y^o; \sigma_0^2/n)$

Any problems? Is it <u>really</u> probability: "$\theta \sim N(y^o; \sigma_0^2/n)$" ?

Suppose an aversion to small (alg.) values

$$PI_{50\%} = (y^o, \infty) \quad \text{if } y^o > 0$$
$$= (-\infty, y^o) \quad \text{if } y^o < 0$$

3. Difficulties / Risks

Simple $y \sim N(\theta; \sigma_0^2/n)$ $\qquad \theta \sim N(y^o; \sigma_0^2/n)$

Any problems? Is it <u>really</u> probability: "$\theta \sim N(y^o; \sigma_0^2/n)$" ?

Suppose an aversion to small (alg.) values

$$PI_{50\%} = (y^o, \infty) \qquad \text{if } y^o > 0$$
$$= (-\infty, y^o) \qquad \text{if } y^o < 0$$

Neyman diagram $\Rightarrow$ If $\theta = 0$, PI <u>never</u> catches θ ! $\quad$ 0% cov

3. Difficulties / Risks

Simple $y \sim N(\theta; \sigma_0^2/n)$ $\qquad \theta \sim N(y^o; \sigma_0^2/n)$

Any problems? Is it <u>really</u> probability: "$\theta \sim N(y^o; \sigma_0^2/n)$" ?

Suppose an aversion to small (alg.) values

$$PI_{50\%} = (y^o, \infty) \quad \text{if } y^o > 0$$
$$= (-\infty, y^o) \quad \text{if } y^o < 0$$

Neyman diagram $\Rightarrow$ If $\theta = 0$, PI <u>never</u> catches θ ! $\quad$ 0% cov

aversion to large (alg.) values

$$PI_{50\%} = (-\infty, y^o) \quad \text{if } y^o > 0$$
$$= (y^o, \infty) \quad \text{if } y^o < 0$$

3. Difficulties / Risks

Simple $y \sim N(\theta; \sigma_0^2/n)$ $\quad\quad$ $\theta \sim N(y^o; \sigma_0^2/n)$

Any problems? Is it <u>really</u> probability: "$\theta \sim N(y^o; \sigma_0^2/n)$" ?

Suppose an aversion to small (alg.) values

$$PI_{50\%} = (y^o, \infty) \quad \text{if } y^o > 0$$
$$= (-\infty, y_o) \quad \text{if } y^o < 0$$

Neyman diagram $\Rightarrow$ If $\theta = 0$, PI <u>never</u> catches θ ! $\quad$ 0% cov

aversion to large (alg.) values

$$PI_{50\%} = (-\infty, y^o) \quad \text{if } y^o > 0$$
$$= (y^o, \infty) \quad \text{if } y^o < 0$$

Neyman diagram $\Rightarrow$ If $\theta = 0$, PI <u>always</u> catches θ ! $\quad$ 100% cov

3. Difficulties / Risks

Simple $y \sim N(\theta; \sigma_0^2/n)$ $\quad \theta \sim N(y^o; \sigma_0^2/n)$

Any problems? Is it really probability: "$\theta \sim N(y^o; \sigma_0^2/n)$" ?

Suppose an aversion to small (alg.) values $\quad PI_{50\%} = (y^o, \infty)$ if $y^o > 0$

$= (-\infty, y^o)$ if $y^o < 0$

Neyman diagram $\Rightarrow$ If $\theta = 0$, PI never catches θ ! 0% cov

aversion to large (alg.) values $\quad PI_{50\%} = (-\infty, y^o)$ if $y^o > 0$

$= (y^o, \infty)$ if $y^o < 0$

Neyman diagram $\Rightarrow$ If $\theta = 0$, PI always catches θ ! 100% cov

Can't really be probability ! $\leftarrow$ NB

Neyman 1937 Need "pivot set" Why: fiducial $\rightarrow$ Confidence

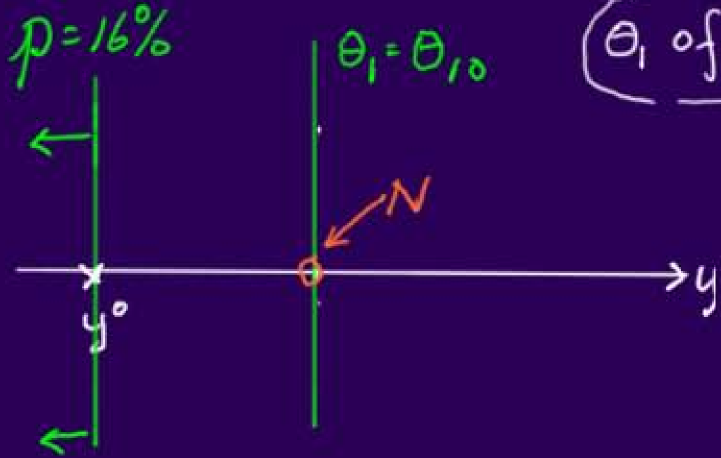
Set-selection bias

4

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; \underline{I}\right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{ \begin{matrix} y_1^0 \\ y_2^0 \end{matrix} ; \underline{I} \right\}$$

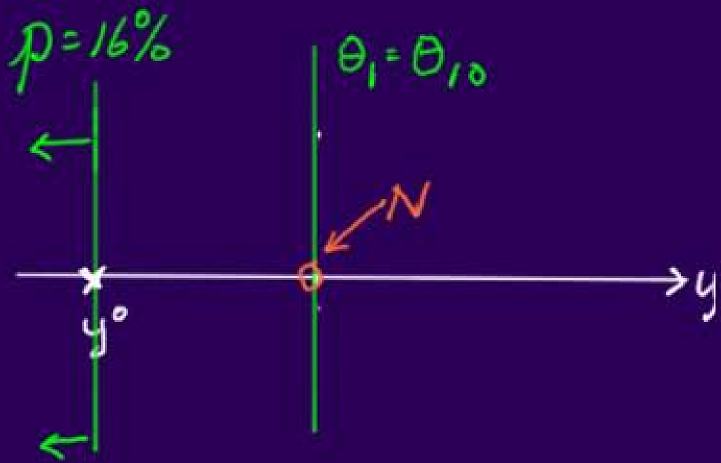
4 $\quad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; \underline{I} \right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{ \begin{matrix} y_1^0 \\ y_2^0 \end{matrix} ; \underline{I} \right\}$

θ_1 of interest



4

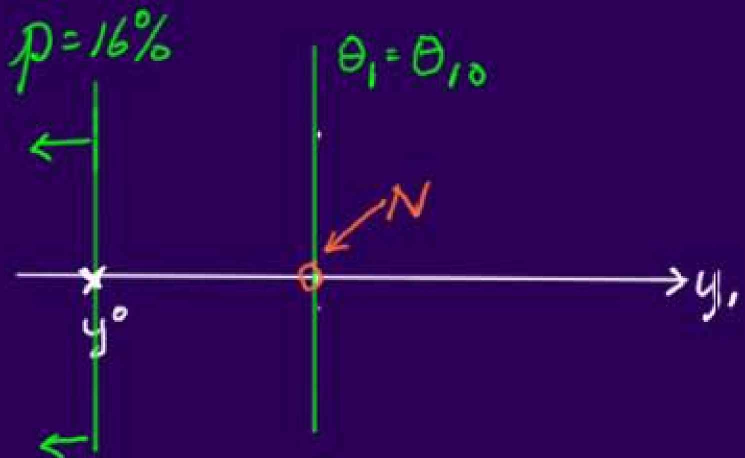
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; \underline{I} \right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{ \begin{matrix} y_1^0 \\ y_2^0 \end{matrix} ; \underline{I} \right\}$$



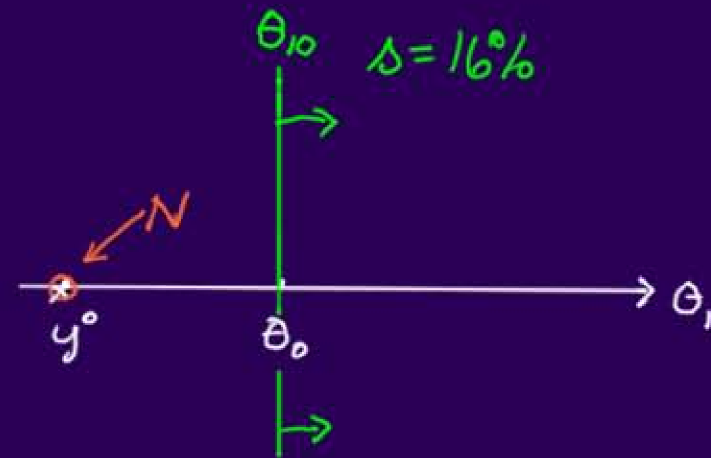
Posterior:

At y^0; flip

4 $\quad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; I \right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{ \begin{matrix} y_1^0 \\ y_2^0 \end{matrix} ; I \right\}$

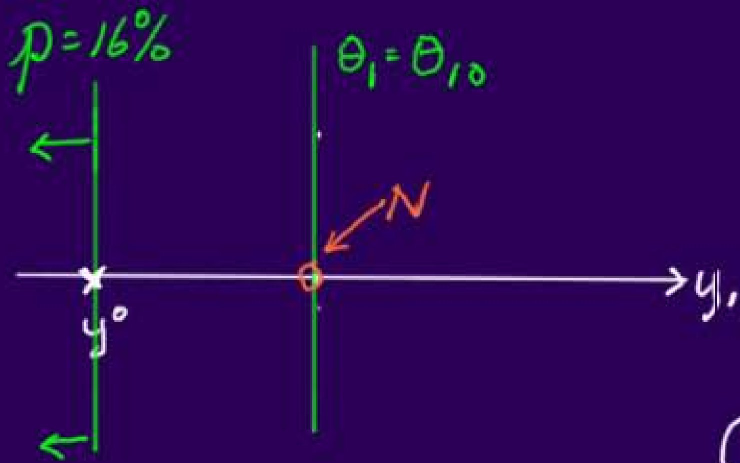


Posterior:
At y^0; flip



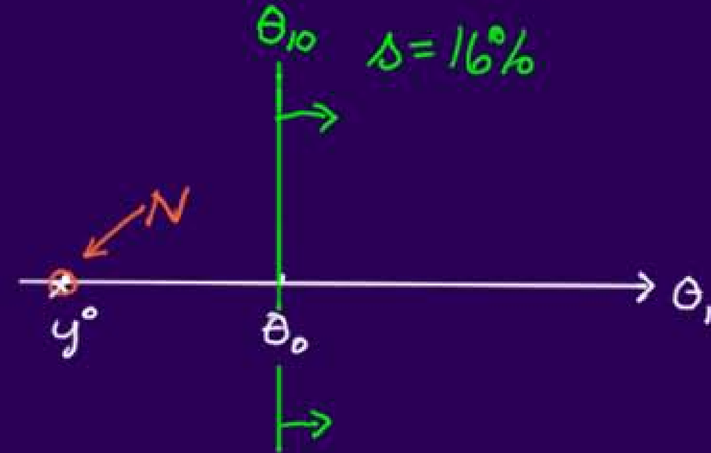
4

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{ \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; \underline{I} \right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{ \begin{matrix} y_1^0 \\ y_2^0 \end{matrix} ; \underline{I} \right\}$$



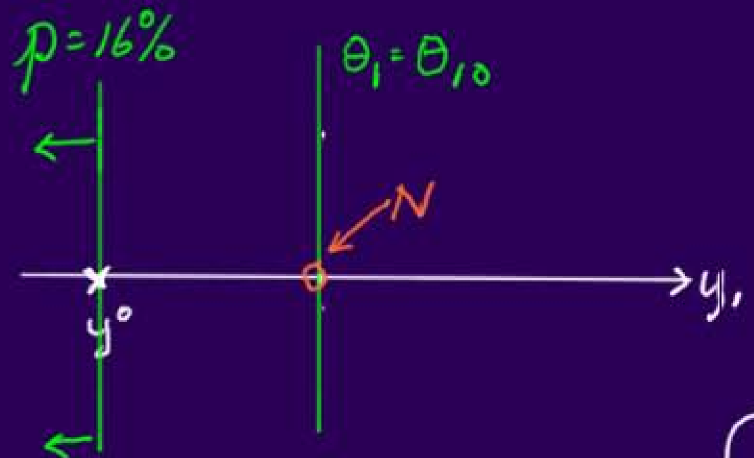
Posterior:
At y^0; flip

Parameter linear
→ pvalue = svalue

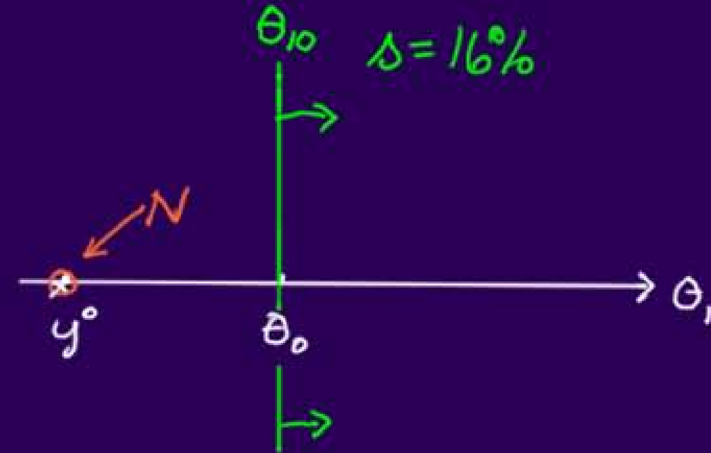


4

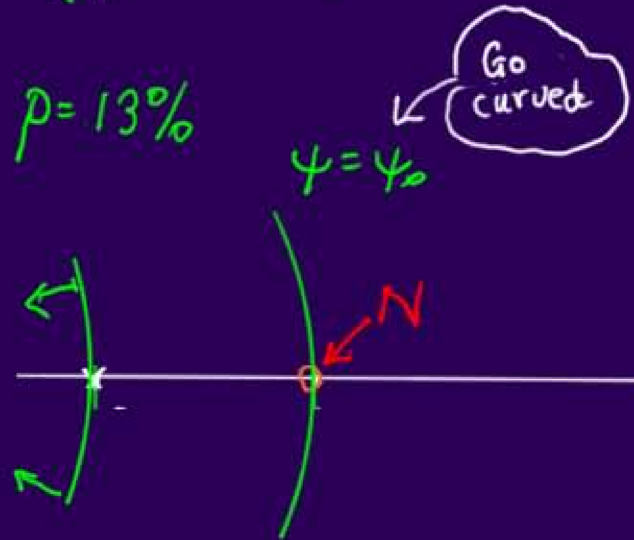
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} ; I\right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{\begin{matrix} y_1^0 \\ y_2^0 \end{matrix} ; I\right\}$$

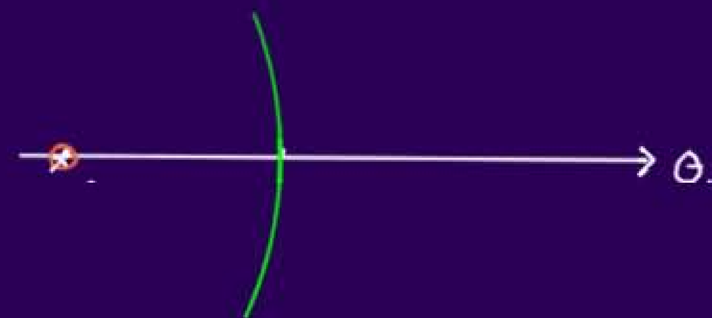


Posterior:
At y°; flip

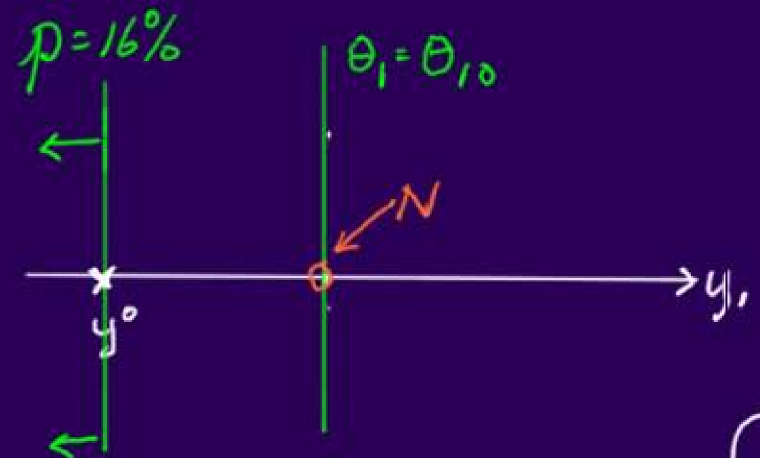


Parameter linear
pvalue = svalue

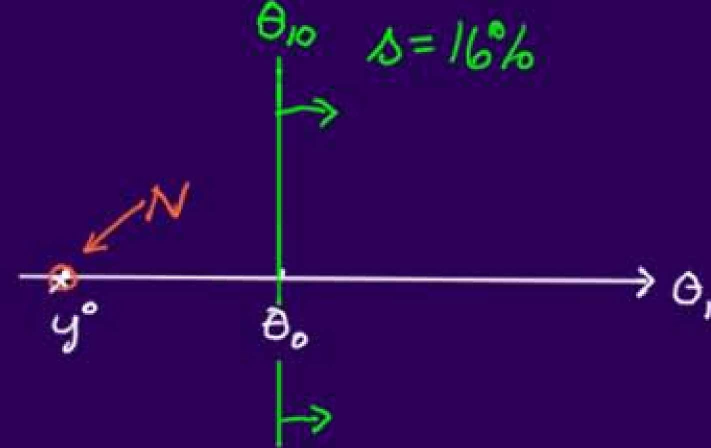




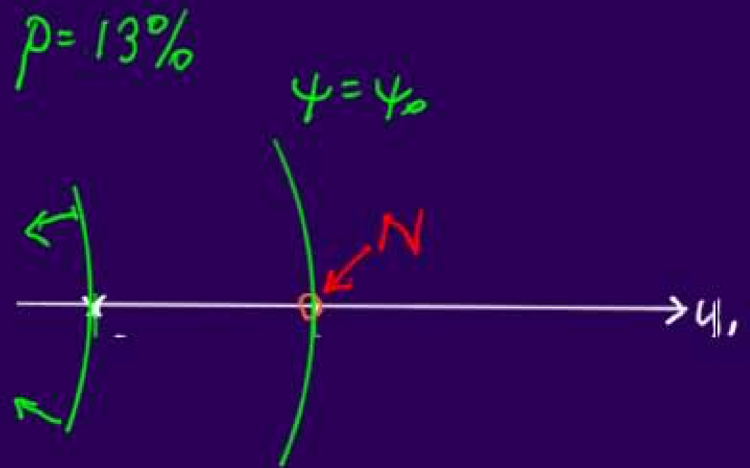
4 $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}; \underline{I}\right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{\begin{matrix} y_1^0 \\ y_2^0 \end{matrix}; \underline{I}\right\}$

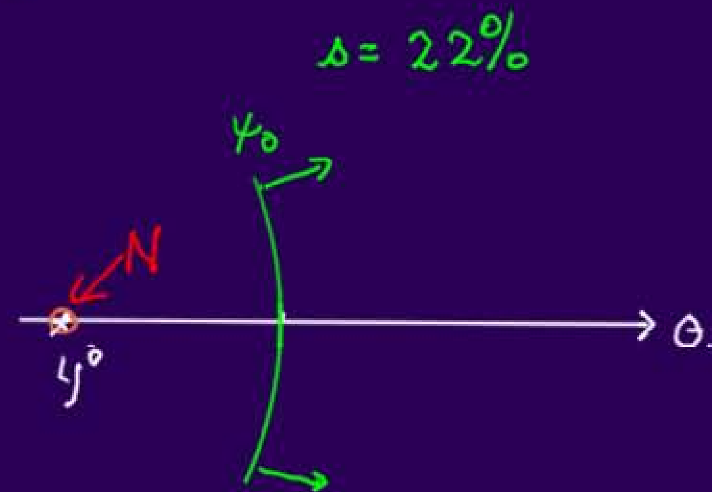


Posterior:
At y^0; flip

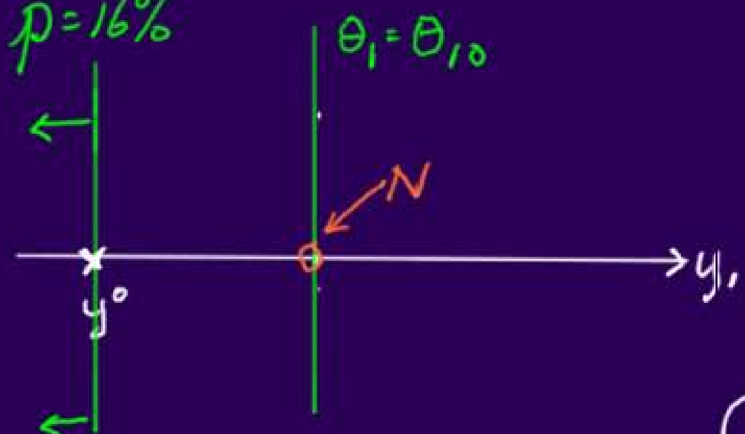


Parameter linear
→ pvalue = svalue

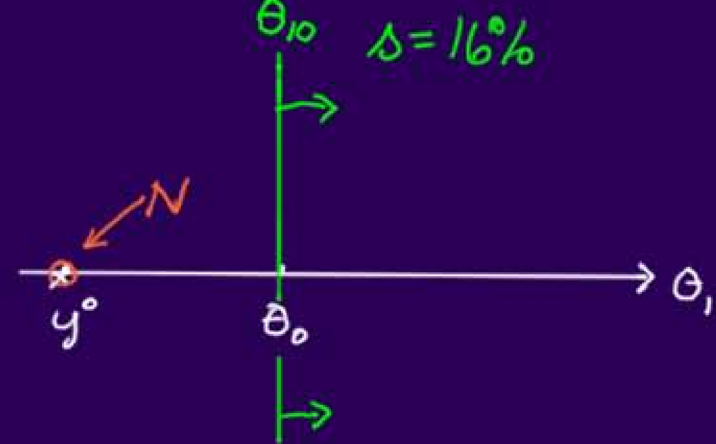




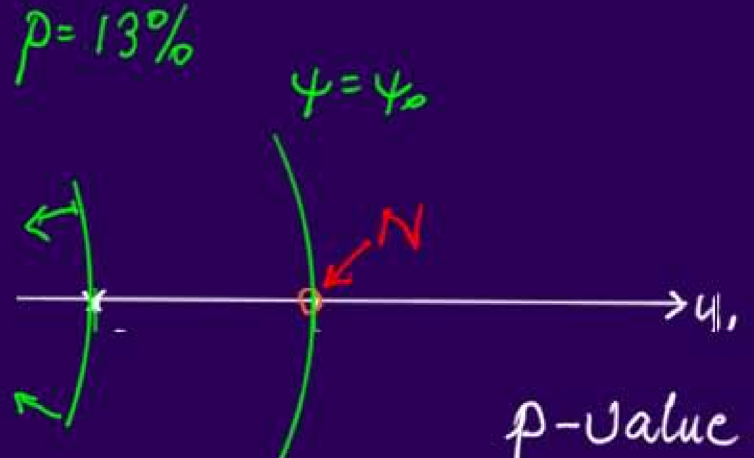
4 $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left\{\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}; I\right\} \Rightarrow \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim N\left\{\begin{matrix} y_1^0 \\ y_2^0 \end{matrix}; I\right\}$

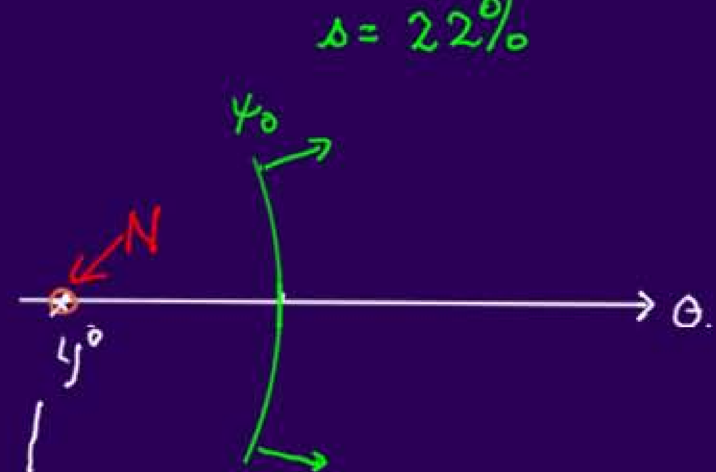


Posterior:
At y°; flip



Parameter linear → pvalue = svalue





p-value decreases !
s-value increases

p-value & s-value change in opposite directions, but --- p-value has repetition validity !

NB

5 <u>All relevant information</u>

Ex: $y_1, \dots, y_n$ $N(\theta; \sigma_0^2)$ Base confidence on y_1 ?

<u>or</u> Base " on $y_1, \dots, y_n$? ✓

"Higher order" can achieve widely, sufficiency unavailable !

5 All relevant information

Ex: $y_1, \dots, y_n$ $N(\theta; \sigma_0^2)$ Base confidence on y_1 ?

or Base " on $y_1, \dots, y_n$?

"Higher order" can achieve widely, when sufficiency unavailable

Now: $y \sim$ Exp't'l model, can par $\varphi(\theta)$, dim $= p$; Interest $\psi(\varphi)$, dim $= d$

5 <u>All relevant information</u>

Ex: $y_1, \dots, y_n$ $N(\theta; \sigma_0^2)$ Base confidence on y_1 ?

or Base " on $y_1, \dots, y_n$?

"Higher order" can achieve widely, when sufficiency unavailable

$y \sim$ Exp't'l model; can par $\varphi(\theta)$, dim $= p$; Interest $\psi(\varphi)$, dim $= d$

<u>Relevant</u> dist'n: SP $\dfrac{e^{k/n}}{(2\pi)^{d/2}} \exp\left\{-\dfrac{r^2(\psi; s)}{2}\right\} |\hat{j}_{\varphi\varphi}(s)|^{-1/2} \, |\hat{j}_{(\lambda\lambda)}(s)|^{1/2} \cdot ds$ | Nuis $\hat{\lambda}_\psi = \hat{\lambda}^0_\psi$, $s^0 = 0$, 3rd order

New NB

5 All relevant information

Ex: $y_1, \dots, y_n$ $N(\theta; \sigma_0^2)$ Base confidence on y_1 ?

or Base " on $y_1, \dots, y_n$?

"Higher order" can achieve widely, when sufficiency unavailable

$y \sim$ Exp't'l model; can par $\varphi(\theta)$, dim $= p$; Interest $\psi(\varphi)$, dim $= d$

Relevant dist'n: SP

$$\frac{e^{k/n}}{(2\pi)^{d/2}} \exp\left\{-\frac{r^2(\psi; s)}{2}\right\} |\hat{j}_{\varphi\varphi}(s)|^{-1/2} |\hat{j}_{(\lambda\lambda)}(s)|^{1/2} \cdot ds$$

Nuis $\hat{\lambda}_\psi = \hat{\lambda}_\psi^0$
$s^0 = 0$
3rd order

As if N

log-lik ratio re $\psi(\varphi) = \psi$

Full info determinant

Nuis. info det in φ scaling

5 All relevant information

Ex: $y_1, \dots, y_n$ $N(\theta; \sigma_0^2)$ Base confidence on y_1 ?

or Base " on $y_1, \dots, y_n$?

"Higher order" can achieve widely, when sufficiency unavailable

$y \sim$ Exp't'l model; can par $\varphi(\theta)$, dim $= p$; Interest $\psi(\varphi)$, dim $= d$

Relevant dist'n: SP $\dfrac{e^{k/n}}{(2\pi)^{d/2}} \exp\left\{-\dfrac{r^2(\psi; s)}{2}\right\} |\hat{j}_{\varphi\varphi}(s)|^{-1/2} |\hat{j}_{(\lambda\lambda)}(s)|^{1/2} \cdot ds$

Nuis $\hat{\lambda}_\psi = \hat{\lambda}_\psi^0$
$s^0 = 0$
3rd order

As if N

log-lik ratio re $\psi(\varphi) = \psi$

Full info determinant

Nuis. info det in φ scaling

For any scalar $\psi(\varphi)$, the p-value function available from r^* to 3rd

uniquely

$$p(\psi) = \Phi(r - r^{-1} \log \frac{r}{q})$$

r = SLR

q = mle departure standardized

5 All relevant information

Ex: $y_1, \dots, y_n$ $N(\theta; \sigma_0^2)$ Base confidence on y_1 ?

or Base " on $y_1, \dots, y_n$?

"Higher order" can achieve widely, when sufficiency unavailable

$y \sim$ Exp't'l model; can par $\varphi(\theta)$, dim $= p$; Interest $\psi(\varphi)$, dim $= d$

Relevant dist'n: SP

$$\frac{e^{k/n}}{(2\pi)^{d/2}} \exp\left\{-\frac{r^2(\psi; s)}{2}\right\} \left|\hat{j}_{\varphi\varphi}(s)\right|^{-1/2} \left|\hat{j}_{(\lambda\lambda)}(s)\right|^{1/2} \cdot ds$$

Nuis $\hat{\lambda}_\psi = \hat{\lambda}^0_\psi$

$s^0 = 0$

3rd order

As if N

log-lik ratio re $\psi(\varphi) = \psi$

Full info determinant

Nuis. info det in φ scaling

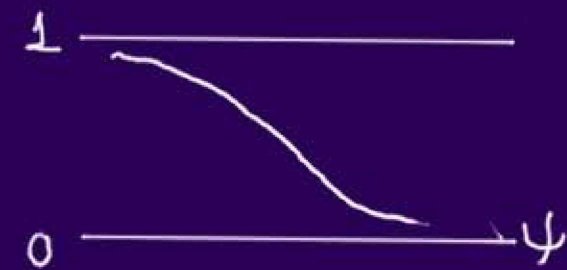
For any scalar $\psi(\varphi)$, the p-value function available from r^* to 3rd

$$p(\psi) = \Phi\left(r - r^{-1} \log \frac{r}{q}\right)$$

r = SLR

q = mle departure standardized

Right Tail df for confidence; uniqueness!



6a <u>Exact distribution for θ</u>: scalar case (via pivot)

$y \sim F(y;\theta)$ Continuous

6a **Exact distribution for θ : scalar case**

$y \sim F(y;\theta)$ Continuous

$u = F(y;\theta) \sim U(0,1)$

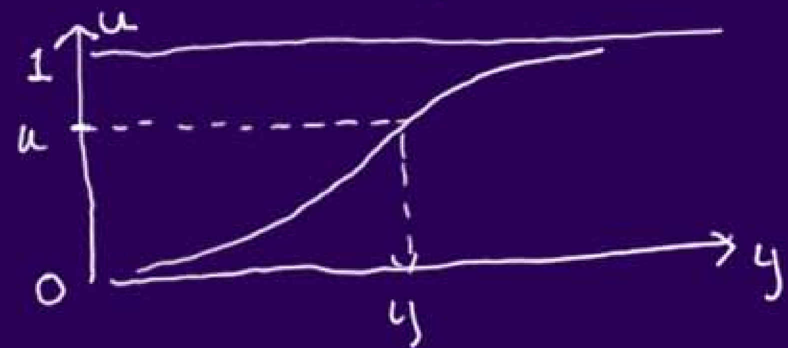
Invert $U(0,1)$ for $u \to y$ then y has df $F(y;\theta)$

6a <u>Exact distribution for θ</u>: scalar case

$y \sim F(y;\theta)$ Continuous

$u = F(y;\theta) \sim U(0,1)$

Invert $U(0,1)$ for $u \to y$ then y has df $F(y;\theta)$

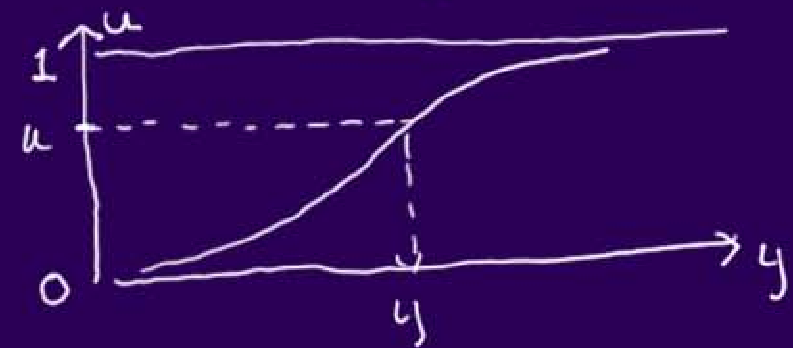


6a **Exact distribution for θ**: scalar case

$y \sim F(y;\theta)$ Continuous

$u = F(y;\theta) \sim U(0,1)$

Invert $U(0,1)$ for $u \to y$ then y has df $F(y;\theta)$



Invert $U(0,1)$ for $u \to \theta$ then θ has RH df $F(y^0;\theta)$

- Definitive dist'n for θ
- Fisher 1930

Vindicates Fisher!

- All seems obvious now

6b <u>Exact distribution for</u> Θ : scalar case; try ⟨via <u>Bayes</u>⟩

Can view a prior – as <u>means</u> to get <u>all</u> info from <u>Likelihood</u>

6b Exact distribution for θ : scalar case ; try with Bayes

Can view a prior - as means to get all info from Likelihood

Start with scalar $f(y;\theta)$ with y^o :

$$p(\theta) = F(y^o;\theta) = \int_{-\infty}^{y^o} F_y(y;\theta)\,dy = -\int_{\theta}^{\infty} F_\theta(y^o;\theta)\,d\theta \quad (\text{just calculus})$$

6b Exact distribution for θ : scalar case ; try with Bayes

Can view a prior – as means to get all info from Likelihood

Start with scalar $f(y;\theta)$ with y^0 :

$$p(\theta) = F(y^0;\theta) = \int_{-\infty}^{y^0} F_y(y;\theta)\,dy = -\int_{\theta}^{\infty} F_\theta(y^0;\theta)\,d\theta$$

$$s(\theta) = \int_{\theta}^{\infty} \pi(\theta)\, F_y(y^0;\theta)\,d\theta$$

6b Exact distribution for θ : scalar case ; try with Bayes

Can view a prior – as means to get all info from Likelihood

Start with scalar $f(y;\theta)$ with y° :

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

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

$s(\theta)$ has repetition validity $\Leftrightarrow$ $\pi(\theta) = -\dfrac{F_\theta(y^\circ;\theta)}{F_y(y^\circ;\theta)} = \left.\dfrac{dy}{d\theta}\right|_{y^\circ}$

6b Exact distribution for θ: scalar case; try with Bayes

Can view a prior – as means to get all info from Likelihood

Start with scalar $f(y;\theta)$ with y^0:

$$p(\theta) = F(y^0;\theta) = \int_{-\infty}^{y^0} F_y(y;\theta)\,dy = -\int_{\theta}^{\infty} F_\theta(y^0;\theta)\,d\theta$$

$$s(\theta) = \int_{\theta}^{\infty} \pi(\theta)\, F_y(y^0;\theta)\,d\theta$$

$s(\theta)$ has repetition validity $\Leftrightarrow$ $\pi(\theta) = -\dfrac{F_\theta(y^0;\theta)}{F_y(y^0;\theta)} = \left.\dfrac{dy}{d\theta}\right|_{y^0}$

where $y = y(\theta;u)$ is quantile function depends on θ & p-value u

6b Exact distribution for θ: scalar case; try with Bayes

Can view a prior – as means to get all info from Likelihood

Start with scalar $f(y;\theta)$ with y^0:

$$p(\theta) = F(y^0;\theta) = \int_{-\infty}^{y^0} F_y(y;\theta)\,dy = -\int_{\theta}^{\infty} F_\theta(y^0;\theta)\,d\theta$$

$$s(\theta) = \int_{\theta}^{\infty} \pi(\theta)\, F_y(y^0;\theta)\, d\theta$$

$s(\theta)$ has repetition validity $\Leftrightarrow$ $\pi(\theta) = -\dfrac{F_\theta(y^0;\theta)}{F_y(y^0;\theta)} = \left.\dfrac{dy}{d\theta}\right|_{y^0}$

where $y = y(\theta;u)$ is quantile function depends on θ & p-value u

$\pi(\theta) = \left.\dfrac{dy}{d\theta}\right|_{y^0}$
- each θ change gives same effect at y^0
- Makes it "location like"
- But: data dependent

7 Welch-Peers (1963):

Exponential model (scalar)

$$f(y;\theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\} h(y) \qquad \varphi \overset{1-1}{\longleftrightarrow} \theta$$

7 Welch-Peers (1963):

Exponential model (scalar)

$$f(y;\theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\}h(y) \qquad \varphi \overset{1-1}{\longleftrightarrow} \theta$$

Showed: Jeffreys prior gives 2nd order Conf. Intervals ! $\left\{\begin{array}{l}\text{Profound!}\\ \text{Neglected!}\end{array}\right.$

$$\pi(\varphi) = j_{\varphi\varphi}^{1/2}(\varphi)$$

7 Welch-Peers (1963):

Exponential model (scalar)

$$f(y;\theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\}h(y) \qquad \varphi \overset{1-1}{\longleftrightarrow} \theta$$

Showed: Jeffreys prior gives 2nd order Conf. Intervals ! $\left\{\begin{array}{l}\text{Profound!}\\ \text{Neglected!}\end{array}\right.$

$$\pi(\varphi)\, d\varphi = j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi$$

Another way to view ...

"Constant info par." $\beta = \int^{\varphi} j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi \qquad j_{\beta\beta}(\varphi) = \text{const} = 1$, 2nd order

7 Welch-Peers (1963):

Exponential model (scalar)

$$f(y;\theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\} h(y) \qquad \varphi \overset{1-1}{\longleftrightarrow} \theta$$

Showed: Jeffreys prior gives 2nd order Conf. Intervals ! $\begin{cases} \text{Profound!} \\ \text{Neglected!} \end{cases}$

$$\pi(\varphi)\, d\varphi = j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi$$

Another way to view...

"Constant info par." $\beta = \int^{\varphi} j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi$ $\qquad j_{\beta\beta}(\varphi) = \text{const} = 1$, 2nd order

$\Rightarrow \hat{\beta} - \beta$ has fixed dist'n, 2nd order $\quad$ i.e. $h(\hat{\beta} - \beta)$

7 Welch-Peers (1963):

Exponential model (scalar)

$$f(y;\theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\} h(y) \qquad \varphi \overset{1-1}{\longleftrightarrow} \theta$$

Showed: Jeffreys prior gives 2nd order Conf. Intervals ! $\left\{\begin{array}{l}\text{Profound!}\\ \text{Neglected!}\end{array}\right.$

$$\pi(\varphi)\, d\varphi = j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi$$

Another way to view...

"Constant info par." $\beta = \int^{\varphi} j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi \qquad j_{\beta\beta}(\varphi) = \text{const} = 1$, 2nd order

Then: $\hat{\beta} - \beta$ has fixed dist'n, 2nd order i.e. $h(\hat{\beta} - \beta)$

$\Rightarrow$ Survivor interval for $\beta \equiv$ Conf. interval for β ... 2nd order W-P!

7 Welch-Peers (1963):

Exponential model (scalar)

$f(y;\theta) = \exp\{s(y)\varphi(\theta) + k(\theta)\} h(y)$ $\qquad \varphi \overset{1-1}{\longleftrightarrow} \theta$

Showed: Jeffreys prior gives 2nd order Conf. Intervals ! $\left\{ \begin{array}{l} \text{Profound!} \\ \text{Neglected!} \end{array} \right.$

$$\pi(\varphi)\, d\varphi = j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi$$

Another way to view ...

"Constant info par." $\beta = \int^{\varphi} j_{\varphi\varphi}^{1/2}(\varphi)\, d\varphi$ $\qquad j_{\beta\beta}(\varphi) = \text{const} = 1$, 2nd order

Then: $\hat{\beta} - \beta$ has fixed dist'n, 2nd order i.e. $h(\hat{\beta} - \beta)$

$\Rightarrow$ Survivor interval for β $\equiv$ Conf. interval for β ... 2nd order W-P!

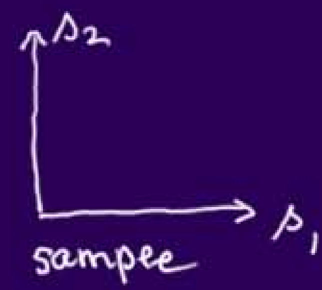
Vector parameter? Jeffreys had shown his prior not OK! !

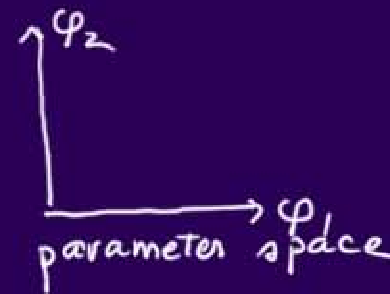
i.e. Location scale: $\dfrac{d\mu\, d\sigma}{\sigma^2}$ not OK ... but $\dfrac{d\mu\, d\sigma}{\sigma}$ OK! ile

8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

Exp Model: can par $\varphi(\theta)$, can. var $s(y)$



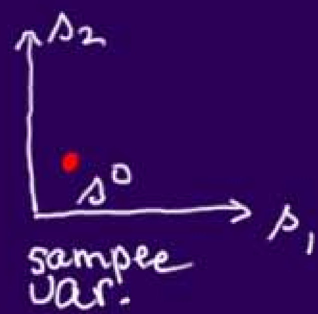


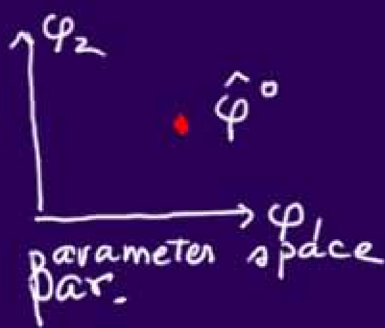
8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

Exp Model: can par $\varphi(\theta)$, can. var $s(y)$

Obs data s° Plot $\Rightarrow$

"Obs" mle $\hat{\varphi}^\circ$





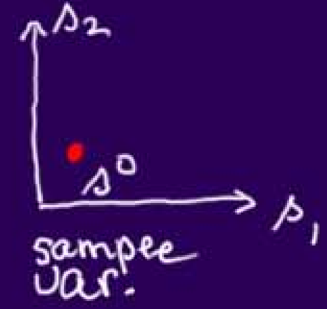
8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

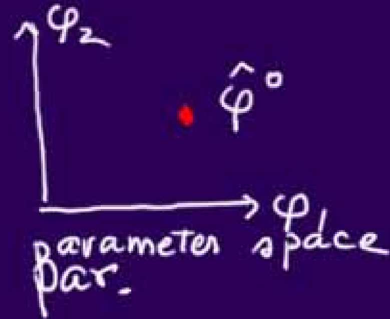
Exp Model: can par $\varphi(\theta)$, can. var $s(y)$

Obs data s°

Plot $\Rightarrow$

"Obs" mle $\hat{\varphi}^\circ$





1) For convenience: standardige re obs. info.

$$\begin{pmatrix}\bar{\varphi}_1\\ \bar{\varphi}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}\varphi_1-\hat{\varphi}_1^\circ\\ \varphi_2-\hat{\varphi}_2^\circ\end{pmatrix}$$

$$\begin{pmatrix}\bar{s}_1\\ \bar{s}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}s_1-s_0^\circ\\ s_2-s_2^\circ\end{pmatrix}$$

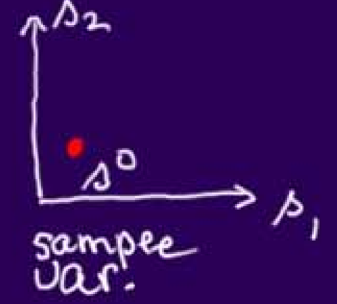
8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

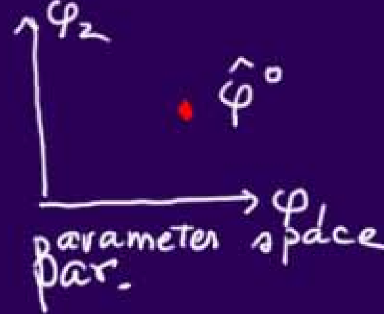
Exp Model: can par $\varphi(\theta)$, can. var $s(y)$

Obs data s^0

"Obs" mle $\hat{\varphi}^0$

Plot $\Rightarrow$





1) For convenience: standardize re obs. info.

$$\begin{pmatrix} \bar{\varphi}_1 \\ \bar{\varphi}_2 \end{pmatrix} = j^{1/2} \begin{pmatrix} \varphi_1 - \hat{\varphi}_1^0 \\ \varphi_2 - \hat{\varphi}_2^0 \end{pmatrix}$$

Right root info $j^{1/2}$

$j^0 = (j^{1/2})' j^{1/2}$ Beware of R!

$$\begin{pmatrix} \bar{s}_1 \\ \bar{s}_2 \end{pmatrix} = j^{1/2} \begin{pmatrix} s_1 - s_1^0 \\ s_2 - s_2^0 \end{pmatrix}$$

where

Obs info $= j^0 = \begin{pmatrix} j^0_{11} & j^0_{12} \\ j^0_{21} & j^0_{22} \end{pmatrix}$

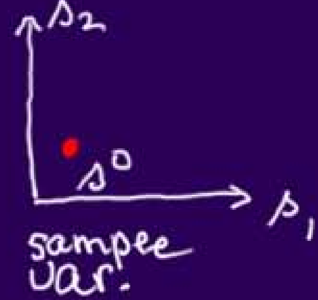
8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

Exp Model: can par $\varphi(\theta)$, can. var $s(y)$

Obs data s^o

"Obs" mle $\hat{\varphi}^o$

Plot $\Rightarrow$



1) For convenience: standardize re obs. info.

$$\begin{pmatrix}\bar{\varphi}_1\\ \bar{\varphi}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}\varphi_1-\hat{\varphi}_1^o\\ \varphi_2-\hat{\varphi}_2^o\end{pmatrix}$$

Right root info $j^{1/2}$ $\qquad j^o = (j^{1/2})' j^{1/2}$ Beware of R!

$$\begin{pmatrix}\bar{s}_1\\ \bar{s}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}s_1-s_1^o\\ s_2-s_2^o\end{pmatrix}$$

where

$$\text{Obs info} = j^o = \begin{pmatrix} j_{11}^o & j_{12}^o\\ j_{21}^o & j_{22}^o\end{pmatrix}$$

2) Do a scalar Jeffreys on each radius from data $\bar{s}^o$

$$\pi(\theta) = j_{\varphi\varphi}^{1/2}(\bar{\varphi})$$

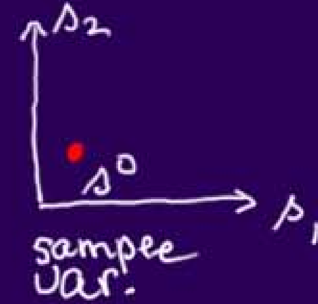
$\bar{s} = \bar{s}^o + \rho \underset{\sim}{e}$ $\qquad \underset{\sim}{e}$ unit vector $\quad$ polar coordinates

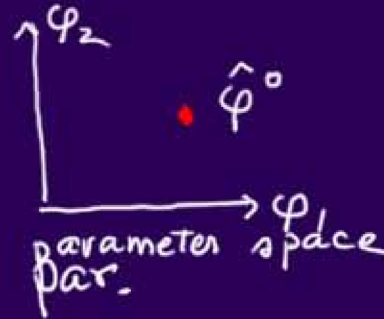
8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

Exp Model: can par. $\varphi(\theta)$, can. var $s(y)$

Obs data s^0 Plot $\Rightarrow$

"Obs" mle $\hat{\varphi}^0$





1) For convenience: standardize re obs. info.

$$\begin{pmatrix}\bar{\varphi}_1\\ \bar{\varphi}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}\varphi_1-\hat{\varphi}_1^0\\ \varphi_2-\hat{\varphi}_2^0\end{pmatrix}$$

Right root info $j^{1/2}$ $\quad j^0 = (j^{1/2})' j^{1/2}$ Beware of R!

$$\begin{pmatrix}\bar{s}_1\\ \bar{s}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}s_1-s_1^0\\ s_2-s_2^0\end{pmatrix}$$

where Obs info $= j^0 = \begin{pmatrix} j^0_{11} & j^0_{12}\\ j^0_{21} & j^0_{22}\end{pmatrix}$

2) Do a scalar Jeffreys on each radius from data $\bar{s}^0$

$$\pi(\theta) = k_e^{1/2}(\bar{\varphi})$$

= radial scalar Jeffreys

$$= \{e' j_{\varphi\varphi} e\}^{1/2}$$

$\bar{s} = \bar{s}^0 + \rho \underset{\sim}{e}$ $\quad \underset{\sim}{e}$ unit vector polar coordinates

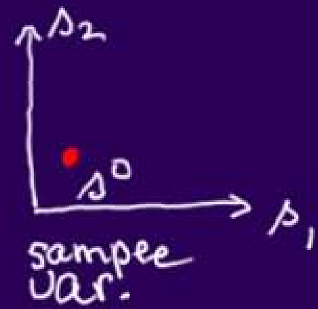
8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

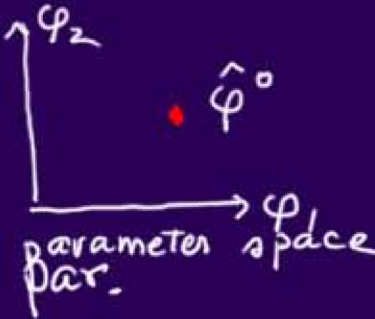
Exp Model: can par $\varphi(\theta)$, can. var $s(y)$

Obs data s^0

"Obs" mle $\hat{\varphi}^0$

Plot =>





1) For convenience: standardize re obs. info.

$$\begin{pmatrix} \bar{\varphi}_1 \\ \bar{\varphi}_2 \end{pmatrix} = j^{1/2} \begin{pmatrix} \varphi_1 - \hat{\varphi}_1^0 \\ \varphi_2 - \hat{\varphi}_2^0 \end{pmatrix}$$

Right root info $j^{1/2}$ $\quad j^0 = (j^{1/2})' j^{1/2}$ Beware of R!

$$\begin{pmatrix} \bar{s}_1 \\ \bar{s}_2 \end{pmatrix} = j^{1/2} \begin{pmatrix} s_1 - s_0^0 \\ s_2 - s_2^0 \end{pmatrix}$$

where

Obs info $= j^0 = \begin{pmatrix} j_{11}^0 & j_{12}^0 \\ j_{21}^0 & j_{22}^0 \end{pmatrix}$

2) Do a scalar Jeffreys on each radius from data $\bar{s}^0$

$$\pi(\theta) = j_{\rho\rho}^{1/2}(\bar{\varphi})$$

$\bar{s} = \bar{s}^0 + \rho \underset{\sim}{e}$ $\quad \underset{\sim}{e}$ unit vector $\quad$ polar coordinates

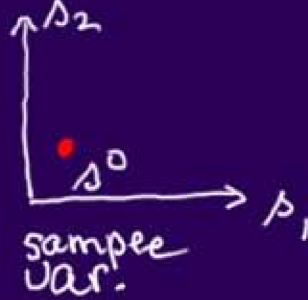
= radial scalar Jeffreys

$$= \{ e' j_{\varphi\varphi} \underset{\sim}{e} \}^{1/2}$$

Reproduces good priors!
Full 2nd order!
Simple to calculate!

8 Vector parameter: Is there a 2nd order prior? Welch-Peers? $O(n^{-1})$

Exp Model: can par $\varphi(\theta)$, can. var $s(y)$



Obs data s^0 Plot =>

"Obs" mle $\hat{\varphi}^0$

1) For convenience: standardize re obs. info.

$$\begin{pmatrix}\bar{\varphi}_1\\ \bar{\varphi}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}\varphi_1-\hat{\varphi}_1^0\\ \varphi_2-\hat{\varphi}_2^0\end{pmatrix}$$

$$\begin{pmatrix}\bar{s}_1\\ \bar{s}_2\end{pmatrix} = j^{1/2}\begin{pmatrix}s_1-s_0^0\\ s_2-s_2^0\end{pmatrix}$$

Right root info $j^{1/2}$ $\quad j^0 = (j^{1/2})' j^{1/2}$ Beware of R!

where Obs info $= j^0 = \begin{pmatrix} j_{11}^0 & j_{12}^0 \\ j_{21}^0 & j_{22}^0 \end{pmatrix}$

2) Do a scalar Jeffreys on each radius from data $\bar{s}^0$

$$\pi(\theta) = k_{ee}^{1/2}(\bar{\varphi})$$

= radial scalar Jeffreys

$$= \{e' j_{\varphi\varphi}\, e\}^{1/2}$$

$\bar{s} = \bar{s}^0 + \rho e$ $\quad e$ unit vector polar coordinates

Reproduces good priors!
Full 2nd order!
Simpee to calculate!

Aggrees with frequentist inference!

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7
then term "Prior" misleading → Just modeling theory Silver 2013

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7

then term "Prior" misleading → Just modeling theory Silver 2013

Mathematical priors: Efron: "Laplace ... a prime violator" 2013

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7

then term "Prior" misleading → Just modeling theory Silver 2013

Mathematical priors: Efron: "Laplace... a prime violator" 2013

but can give approximate confidence: Welch Peers & new prior!

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7

then term "Prior" misleading → Just modeling theory Silver 2013

Mathematical priors: Efron: "Laplace ... a prime violator" 2013

but can give approximate confidence: Welch Peers & new prior!

c) Confidence dist'n: powerful expression, but risks, should not be ignored

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7

then term "Prior" misleading → Just modeling theory Silver 2013

Mathematical priors: Efron: "Laplace ... a prime violator" 2013

but can give approximate confidence: Welch Peers & new prior!

c) Confidence dist'n: powerful expression, but risks, should not be ignored

d) Use of conditional probability (Often labeled "Bayes")

i) Genuine "prior": Just modeling not Bayes

9 Summary conclusions

a) Confidence dist'ns should use all relevant information $O(n^{-3/2})$ $O(n^{-1})$

b) Use of priors: Based on "genuine prior information" Efron 2013 Science June 7
then term "Prior" misleading → Just modeling theory Silver 2013
Mathematical priors: Efron: "Laplace... a prime violator" 2013
but can give approximate confidence: Welch Peers & new prior!

c) Confidence dist'n: powerful expression, but risks, should not be ignored

d) Use of conditional probability (Often labeled "Bayes")

1) Genuine "prior": Just modeling

2) Mathematical prior can give confidence

1) History

2) Examples: scalar θ

3) Difficulties: Selection

4) " Curvature

5) All relevant information re interest ψ

6) Exact distribution for θ

7) Welch Peers and scalar Jeffreys

8) Vector parameter and second-order prior

9) Summary

Calibration? - First order: for sure!

- Second $O(n^{-1})$ No!

- Just linear parameters

Thank you!