Objective and other priors

D A S Fraser
N Reid
Dept Statistics
Univ Toronto

OBayes 5
Branson MO
June 8 2005

$$\boxed{\pi(\theta)} \xrightarrow{\ ?\ } \boxed{f(y;\theta)} \longrightarrow y$$

Have a model $f(y;\theta)$

Add a prior ?

1) Types of prior
2) Examples w/ messages
3) Stuff from likelihood
4) Enigmatic examples
5) p-values

OBayes 5
Branson MO
June 8 2005

$$\boxed{\pi(\theta)} \longrightarrow ? \longrightarrow \boxed{f(y;\theta)} \longrightarrow y$$

Have a model $f(y;\theta)$

Add a prior ?

1) Types of prior
2) Examples w/ messages
3) Stuff from likelihood
4) Enigmatic examples
5) p-values

Propose:

1) Use: flat (re model) prior for interest
2) Add: if wanted, subjective for interest
3) Combine: if needed, for decisions

OBayes 5
Branson MO
June 8 2005

Types of prior:

① <u>Objective</u>

There is an <u>identified random source</u>

<u>Model</u> (parameters): sampling; theory (e.g. genetics); experiment; related investigation

Types of prior:

① <u>Objective</u>

There is an <u>identified random source</u>

<u>Model</u> (parameters): sampling; theory (e.g. genetics); experiment; related investigation

② <u>Flat</u> (re model)

Invariance; original Thomas B; Laplace; Insufficient r. ... (some versions)

Types of prior:

① Objective

There is an identified random source

Model (parameters): sampling; theory (e.g. genetics); experiment; related investigation

② Flat (re model)

Invariance; original Thomas B; Laplace; Insufficient r. ... (some versions)

③ Subjective

Individual, group

Personal assessment, impression, one basis for action

Types of prior:	Names: f	B
① Objective There is an identified random source Model (parameters): sampling; theory (e.g. genetics); experiment; related investigation	objective	— ⊛
② Flat (re model) Invariance; original Thomas B; Laplace; Insufficient r. ... (some versions)	flat	objective*
③ Subjective Individual, group Personal assessment, impression, one basis for action	subjective	subjective

---

⊛ No apparent term

* 2nd order objective

- Describes a model
- Model descibes the process

Types of prior:

Names: f

① Objective

objective

There is an identified random source

Model (parameters): sampling; theory (e.g. genetics); experiment; related investigation

② Flat (re model)

flat

Invariance; original Thomas B; Laplace; Insufficient r. ... (some versions)

③ Subjective

subjective

Individual, group

Personal assessment, impression, one basis for action

↑

Use here...

An issue!

Analysis ? $\boxed{\pi(\theta)} \longrightarrow \boxed{f(y;\theta)} \longrightarrow y$

① Objective

Probability model, no parameter
Concealed internal variable
Standard probability theory
... but Forbidden Knowledge ?? $\left\{\begin{array}{l}\text{Science}\\ \text{11 Feb 2005}\end{array}\right.$

② Flat (re model)

③ Subjective

$\boxed{\pi(\theta)} \longrightarrow \boxed{f(y;\theta)} \longrightarrow y$

What analysis ?

1. Objective
   - Probability model, no parameter
   - Concealed internal variable
   - Standard probability theory
   - ... but Forbidden Knowledge ?? {Science 11 Feb 2005

2. Flat (re model)
   - Bayes paradigm to extract model-data info ?
   - Are there other ways ?
   - Combined ways ?

3. Subjective

$\boxed{\pi(\theta)} \rightarrow \boxed{f(y;\theta)} \rightarrow y$ What analysis ?

① **Objective**

- Probability model, no parameter
- Concealed internal variable
- Standard probability theory
- ... but Forbidden Knowledge ?? { Science 11 Feb 2005

② **Flat** (re model)

- Bayes paradigm to extract model-data info ?
- A basis for decisions ?
- Do posterior probs. describe ... ?

③ **Subjective**

- For individual decisions ...
- ... but "Forbidden Knowledge (ibid Science)
- ... but profiling, prejudice,

Summary

$\boxed{\pi(\theta)} \rightarrow \boxed{f(y;\theta)} \rightarrow y$

① If prior is objective (truly)
- Just probability analysis
- Recall "Forbidden knowledge"

② Choosing Flat prior
- a convenient route to model-data info!
- But need relevant model
  - Data can condition / restrict model

OBayes 1999
Severini, Fraser, Reid
U Valencia
and noted

③ If subjective is available
- Record in parallel
  - Others may want to use own prior
  - For individual decisions !
  - "Not for full model analysis !

Simple Examples

① Location model, scalar $f(y-\theta)$

frequentist

$\longleftrightarrow$ translation / location

(a) percentile position of data

$$p(\theta) = F(y^o - \theta) = \int_{-\infty}^{y^o} f(y-\theta)\,dy$$

"as it is!"

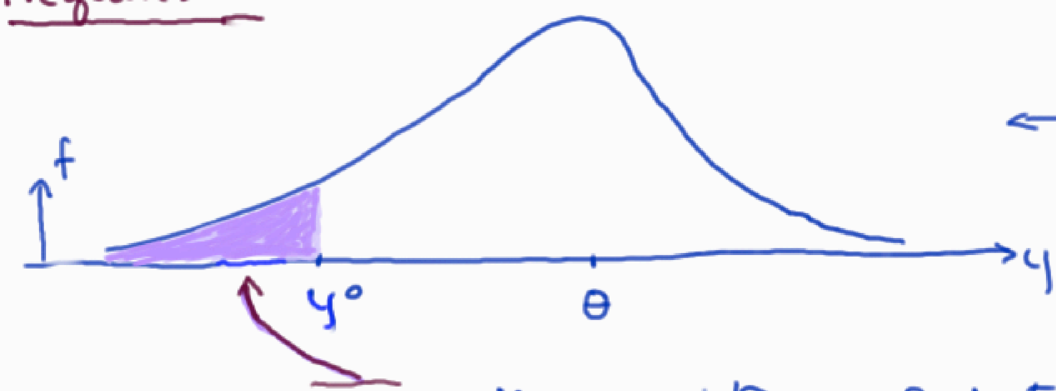
(b) Likelihood

$$L(\theta) = f(y^o - \theta)$$

Simple Examples

1 Location model, scalar $f(y-\theta)$

frequentist



$\leftrightarrow$ translation / location

(a) percentile position of data

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

"as it is!"

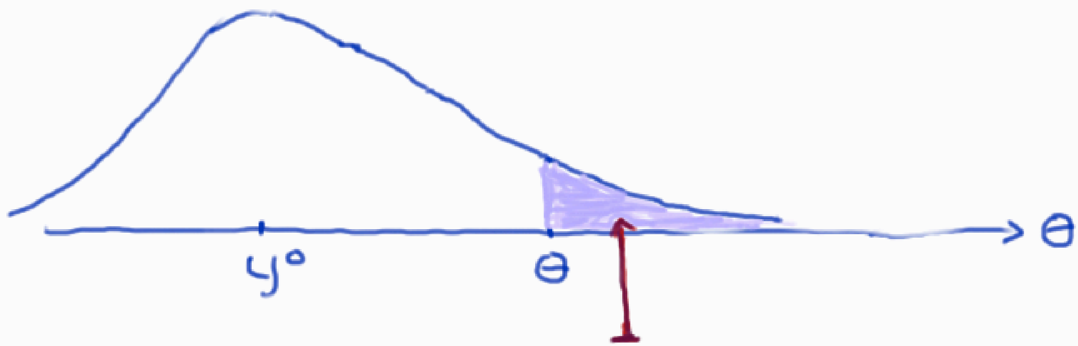
df F

(b) Likelihood

$$L(\theta) = f(y^\circ - \theta)$$

pdf f at data

Bayes: flat (model) prior $\quad d\pi(\theta) = d\theta$



$$s(\theta) = \int_\theta^\infty c\,L(t)\,dt$$

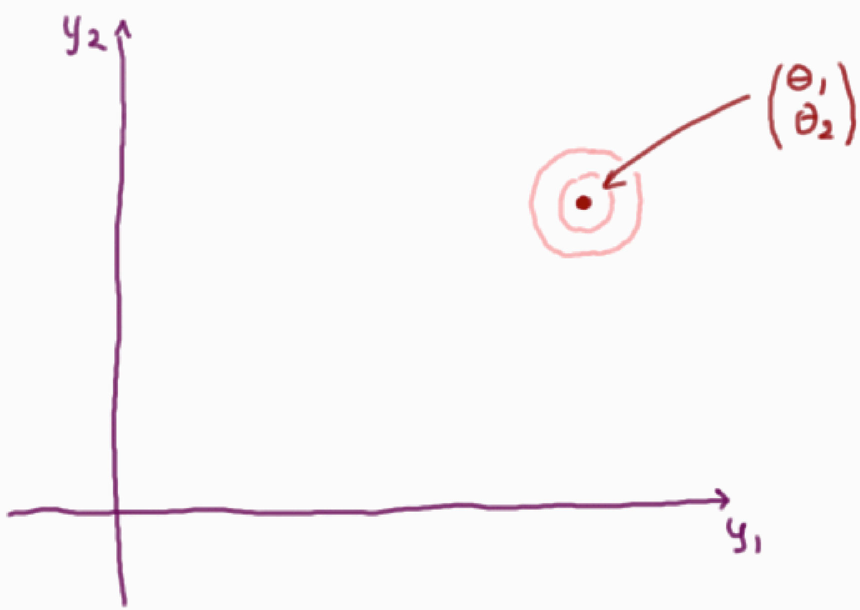
$\Rightarrow$ freq & Bayes say same: $\quad p(\theta) \equiv s(\theta) \equiv$ same integral

Message 1: Scalar location $\Rightarrow$ freq $\equiv$ Bayes $\equiv$ same "info"

Can life be so simple?
Conflict free?

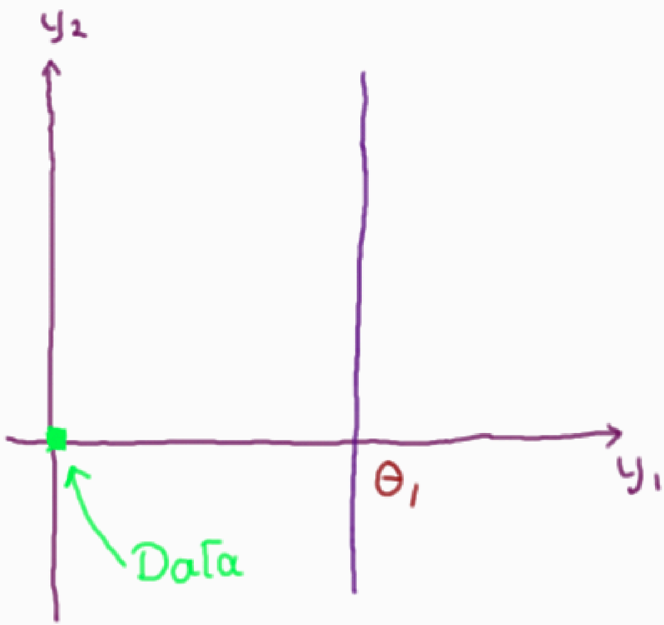
Example 2: Location on R^2 Flat = $d\underset{\sim}{\theta}$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \qquad z_i \sim N(0,1)$$



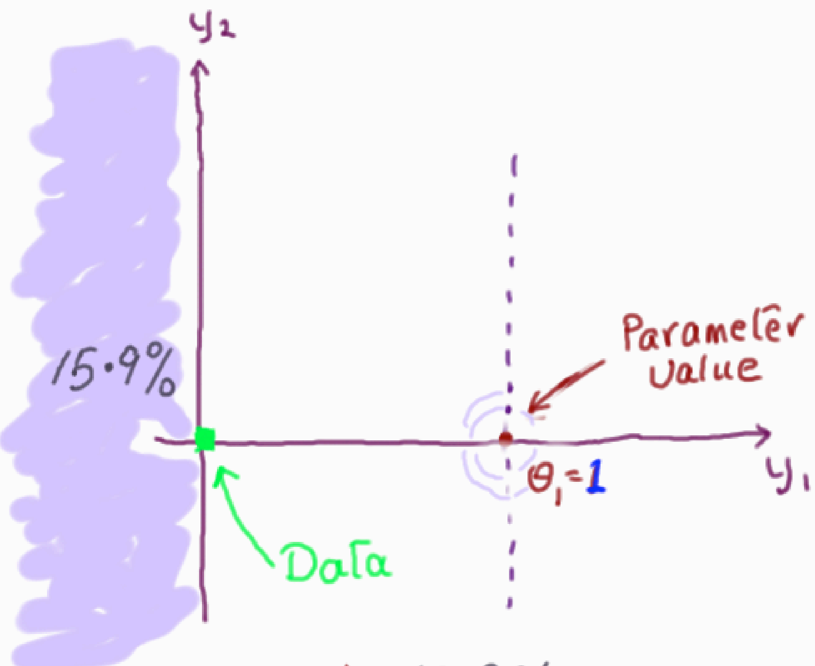
Data : $\begin{pmatrix} y_1^{\circ} \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ... say

Interest: $\psi(\theta) = \theta_1$ which is linear



Interest: θ_1 ..... linear Assess: $\theta_1 = 1$

frequentist: $y_1 = 1 + z$



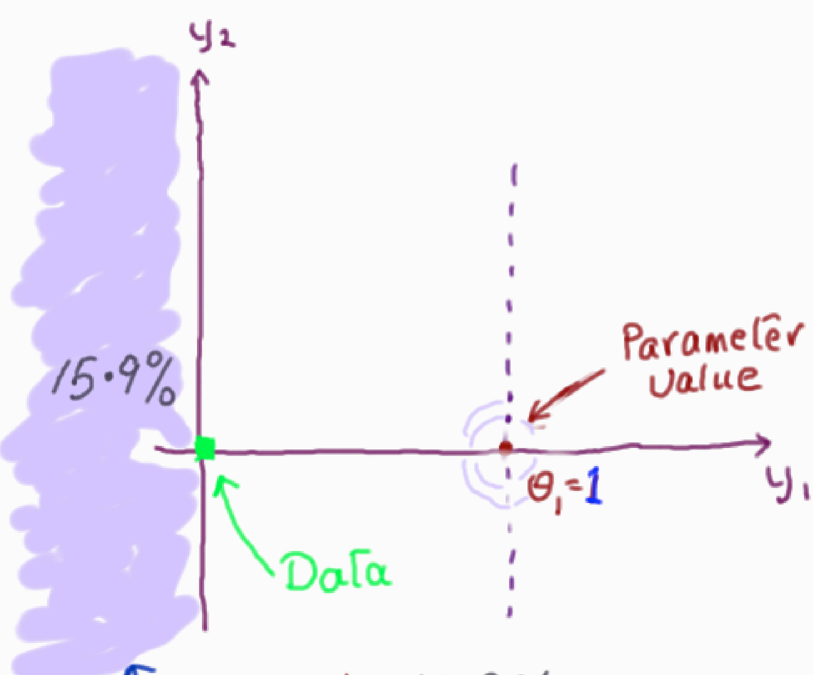
$p(1) = 15.9\%$

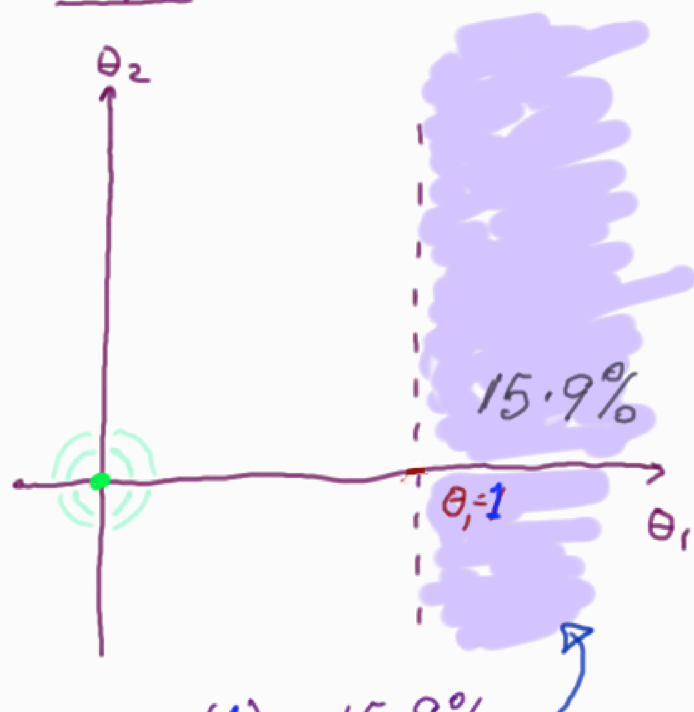
Interest: θ_1 linear

Assess: $\theta_1 = 1$

frequentist: $y_1 = 1 + z$

Bayes: $\theta_1 = 0 + z$





$p(1) = 15.9\%$
$= P(y_1 \le 0)$

$s(1) = 15.9\%$
$= P(\theta_1 \ge 1)$

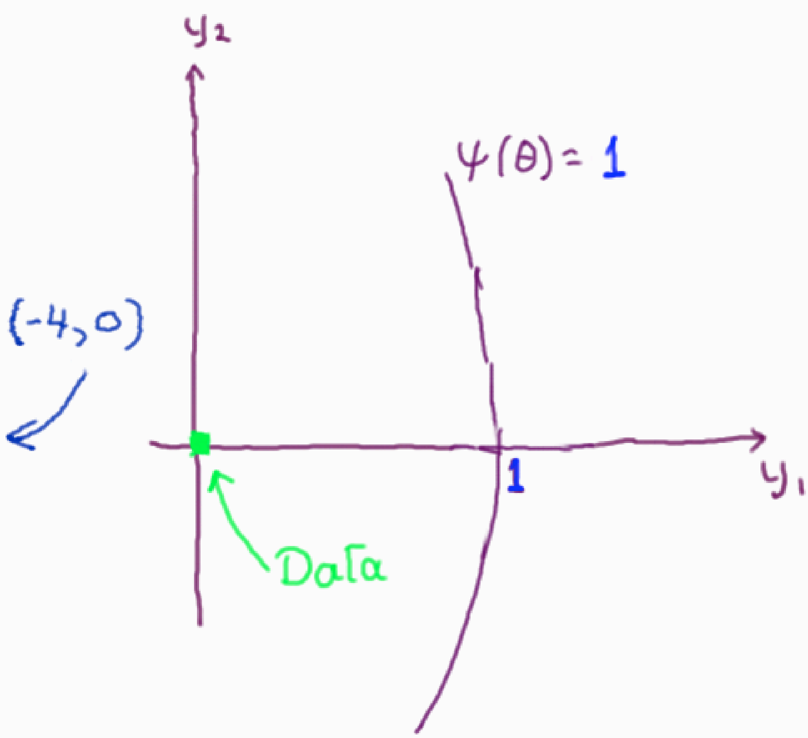
* $p(1) = s(1) =$ same "integral"!

- But: the interest parameter is linear

Now take a...

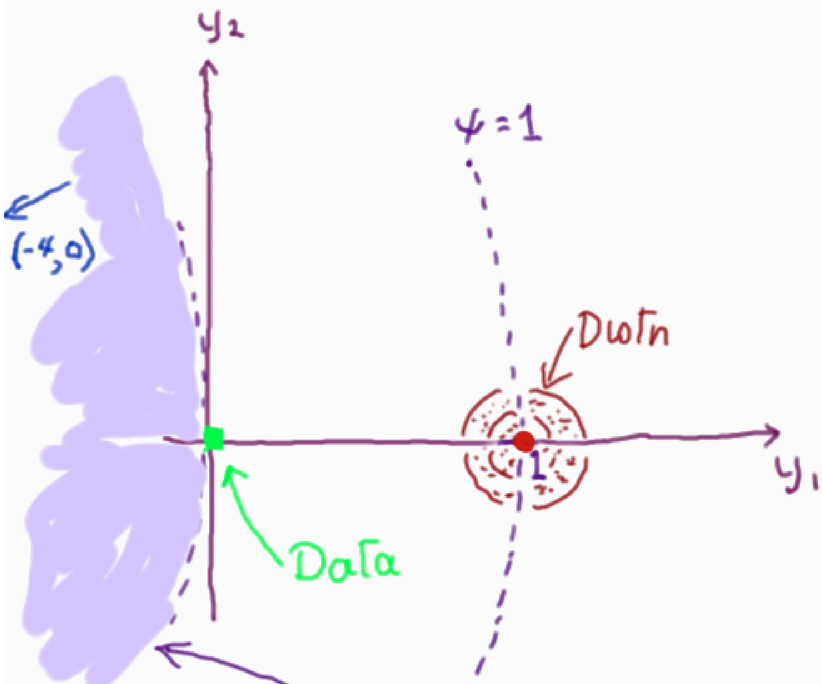
Curved interest : $\psi(\theta) = \{(4+\theta_1)^2 + \theta_2^2\}^{1/2} - 4$; assess $\psi = 1$

$= \text{distance from } (-4, 0) - 4$



Curved interest : $\psi = \{(4+\theta_1)^2 + \theta_2^2\}^{1/2} - 4$; assess $\psi = 1$

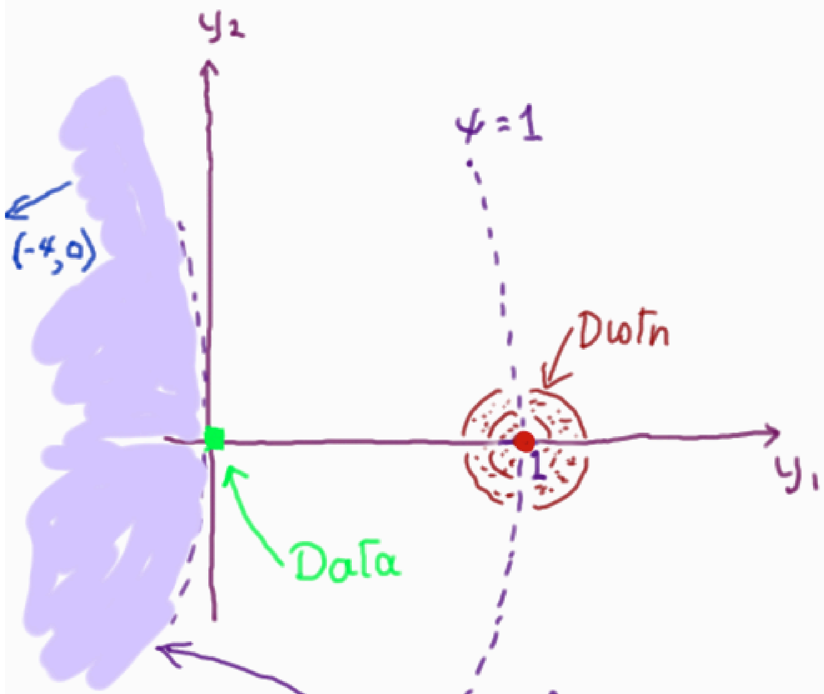
f: Distn at $\begin{pmatrix}1\\0\end{pmatrix}$ B: Post at $\begin{pmatrix}0\\0\end{pmatrix}$

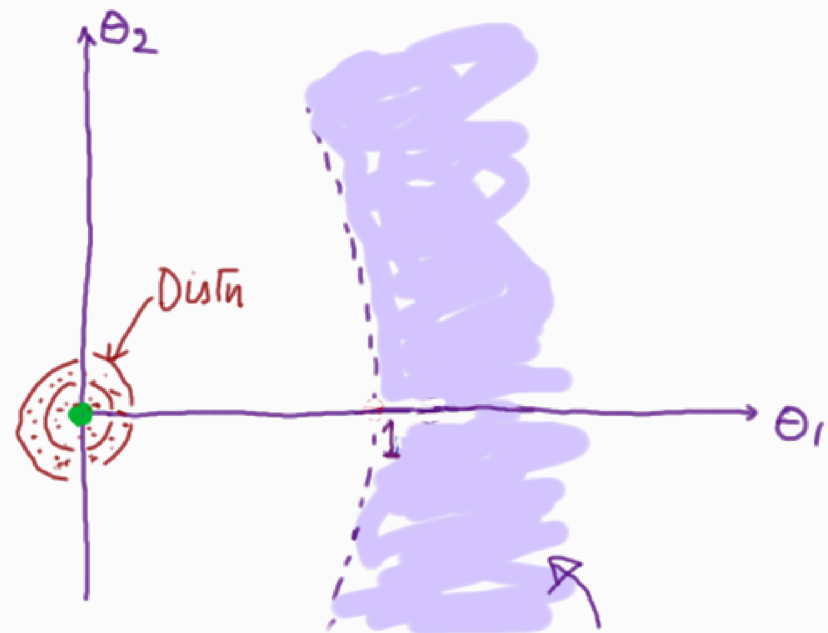


freq $p = P(\hat{\psi} < 0; \psi = 1)$
$= 13.5\%$

Curved interest : $\psi = \{(4+\theta_1)^2 + \theta_2^2\}^{1/2} - 4$; assess $\psi = 1$

f: Distn at $\binom{1}{0}$ B: Post at $\binom{0}{0}$





freq $p = P(\hat{\psi} < 0 ; \psi = 1)$

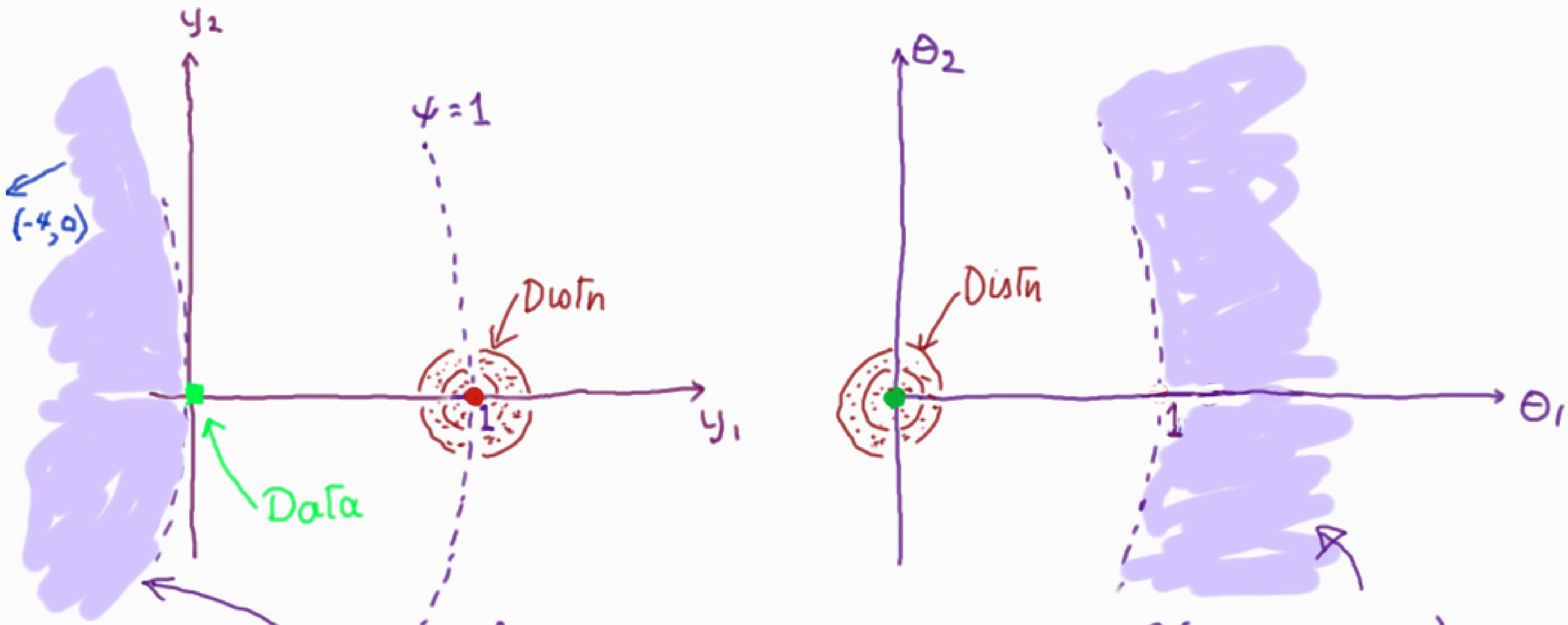
$= 13.5\%$

B $s = P(\psi > 1 ; y_1 = 0)$

$= 18.7\%$

Curved interest : $\psi = \{(4+\theta_1)^2 + \theta_2^2\}^{1/2} - 4$; assess $\psi = 1$

f: Distn at $\binom{1}{0}$ B: Post at $\binom{0}{0}$



freq $p = P(\hat{\psi} < 0 ; \psi = 1)$ $= 13.5\%$

B $s = P(\psi > 1 ; y_1 = 0)$ $= 18.7\%$

Different 40%

Clarifying example; old problem

Marginalization ; Dawid Stone Zidek 1973

Vector flat prior can't handle curved parameters

Curved interest: $\psi = \{(4+\theta_1)^2 + \theta_2^2\}^{1/2} - 4$; assess $\psi = 1$

- There is a variable r that measures ψ
- And it has a flat prior
- Use it!

So (2) Use identified model for parameter of interest & corresponding flat prior

available from model structure
not " " behavior

Example 3 $\quad f_0(y-\theta) \quad f_1(y-\log\theta) \quad f_i(y;\theta) \quad i = \text{Bern}(1/2)$

$i=0$ Jeffreys $d\theta$

$=1$ Jeffreys $d\log\theta$

$I(\theta) =$ average

general: $\left| j_{(\theta\theta)}(\theta; y) \right|^{1/2} d\theta \qquad p(\theta) = p(y)$

$$\left| \frac{\ell_\theta(\theta; y)}{\varphi(\theta)} \right| d\theta \qquad \text{"} \qquad \text{"}$$

Fraser & Reid OBayes U Valencia Jun 13 1999
Severini " " "
Conditioning gives ... indicated/appropriate model

Use model determined by data
available information

(3) Use model indicated by observed data

Flat (model) prior available in Exs 1, 2, 3

Obtain: $p(\theta) = s(\theta)$ provided you model interest parameter

Can everything be so simple? Yes!

but you need the reparameterization $\varphi(\theta)$

Finding flat prior ... from Likelihood information...

Prior $\longrightarrow$ Likelihood $\longrightarrow$ Posterior

① - How likelihood processes information

or

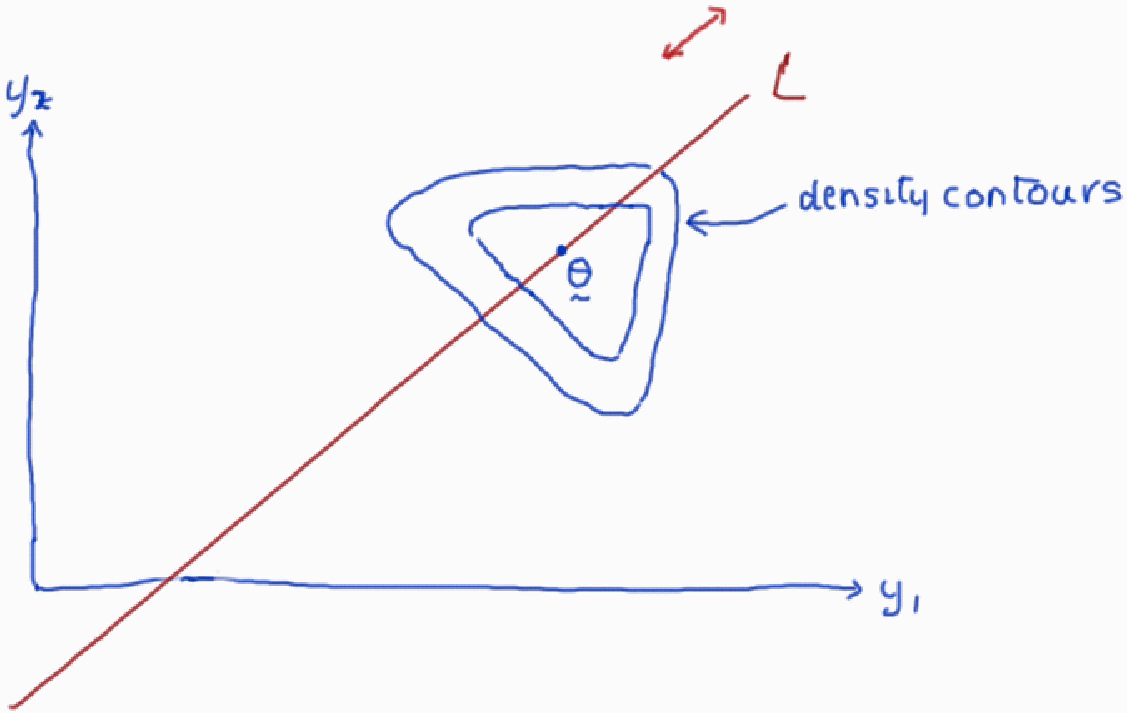
② - What model information is available ... via $\varphi(\theta)$

An example:

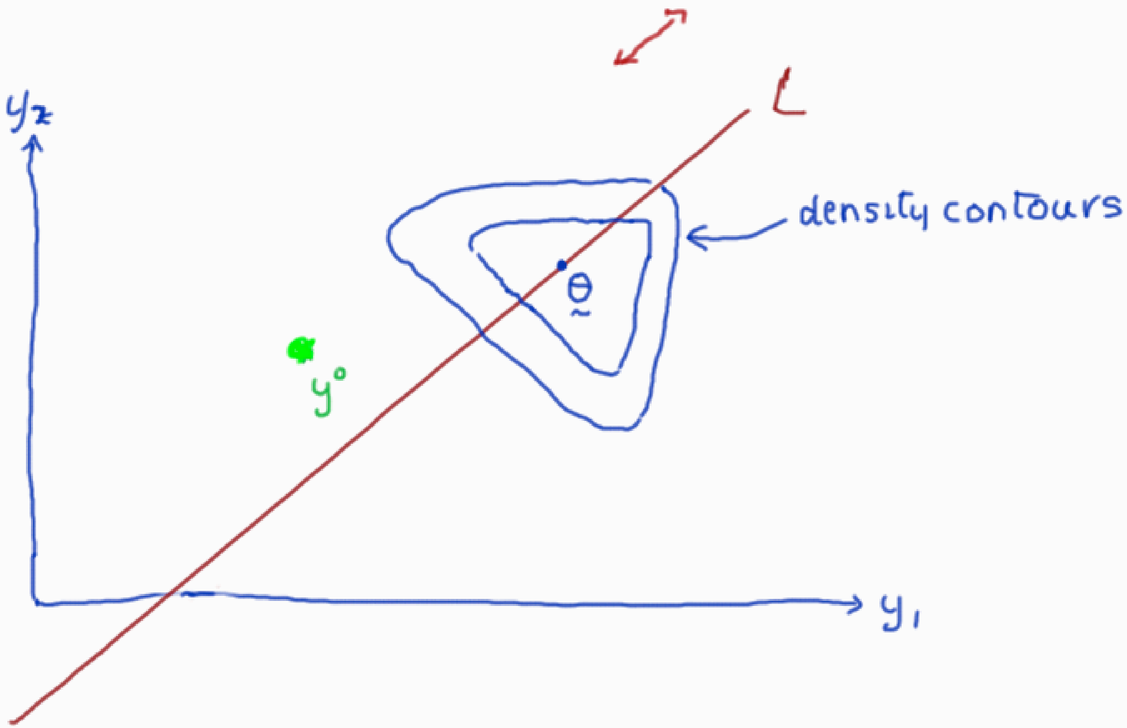
Example 4 Location model on R^2

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \qquad \underset{\sim}{z} \sim f(\underset{\sim}{z})$$

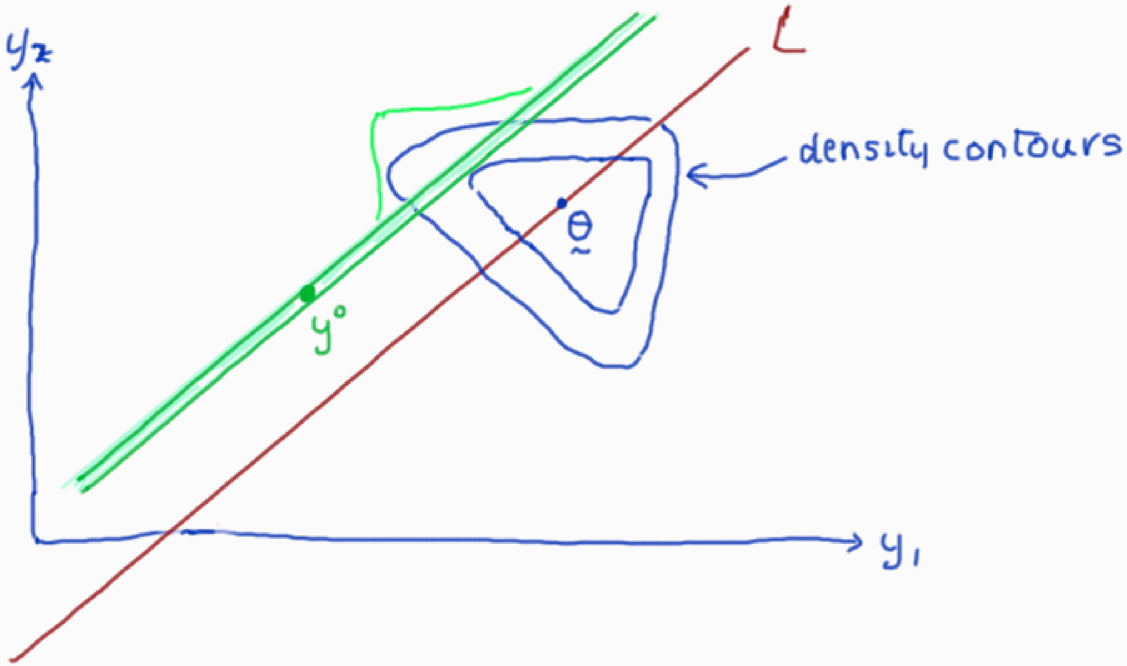
Location on plane ; $\underset{\sim}{\theta}$ on line L



Location on plane ; $\underset{\sim}{\theta}$ on line L ; data y^o



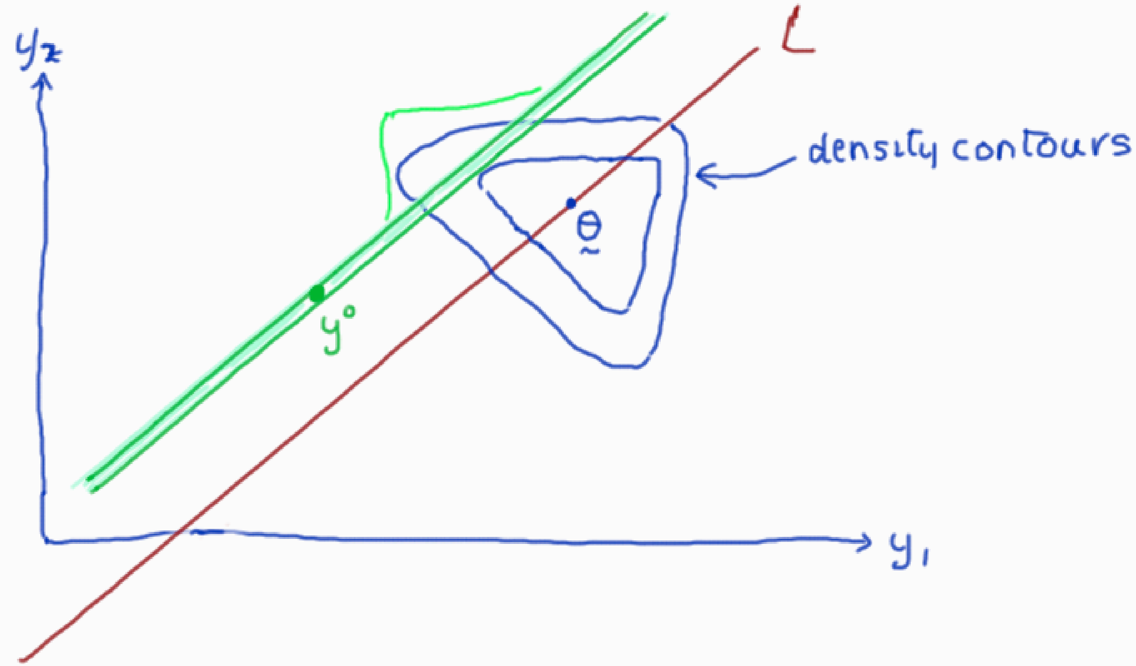
Location on plane ; $\underset{\sim}{\theta}$ on line L ; data y^{o}



Can identify what part of model is relevant

Condition: parallel to line L ; density indicated

Location on plane ; $\underset{\sim}{\theta}$ on line L ; data y°



Can identify what part of model is relevant
Condition parallel to line L ; density indicated

So: Use the identified model & related flat prior
but more generally ?

Flat (model) prior available for full θ, for interest $\psi(\theta)$ from Likelihood analysis

Moderate regularity, wide generality.

Model: $f(y;\theta)$
Data: y^0
How y measures θ $\qquad$ piv $= \underset{\sim}{z} = p(y;\theta)$

$\downarrow$

Lik: $L(\theta) = \log f(y;\theta)$
Repar: $\varphi(\theta) = \nabla_V \ell(\theta; y^0)$ ... { Provides all recalibration information .. for 2nd/3rd order flat priors

$\downarrow$

Lik: $L(\theta)$
Flat: $d\beta(\theta)$

Model
Data $\Longrightarrow$ $L(\theta)$, $\beta(\theta)$, $\beta(\psi)$

(well defined calculations)

F Reid Wu Bka 1999
Fraser Bka 2003

More Examples

(5) <u>An old enigmatic example</u> Behrens-Fisher (1929) (1934) ; Ghosh Kim CJS (2001)

$y_{11}, \dots, y_{1n}$ N μ_1 σ_1^2 Interest $\psi = \mu_1 - \mu_2$

$y_{21}, \dots, y_{2n}$ N μ_2 σ_2^2

GK: $\pi = \sigma_1^{-3} \sigma_2^{-3} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$ J: $\pi = \sigma_1^{-1} \sigma_2^{-1}$ (cf fisher) Jeffreys Indep

Some examples:

5 An old enigmatic example   Behrens-Fisher (1929) (1934); Ghosh Kim CJS (2001)

$y_{11}, \dots, y_{1n}$ N μ_1, σ_1^2

$y_{21}, \dots, y_{2n}$ N μ_2 σ_2^2

Interest $\psi = \mu_1 - \mu_2$

GK: $\pi = \sigma_1^{-3} \sigma_2^{-3} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$ vs Jeffreys(right): $\sigma_1^{-1} \sigma_2^{-1}$

Some simulation numbers: Case

n	m	σ_1^2	σ_2^2	ψ
2	2	2	1	2

Nominal 5% 95% points by various methods including flat prior

Nominal	5%	95%
Jeffreys	.7%	99.1%
Kim Ghosh	1.7%	97.9%
Lik ratio	13.2%	86.9%
$p(\psi), s(\psi)$	4.23%	95.8%
Sim 95% limits	(4.86, 5.14)	(94.9, 95.14)

$N = 100{,}000$

(6) Not-so-old example

Power transformed regression   Box Cox (1964, 1982)

$$y_i = (\alpha + \beta x_i + \sigma z_i)^{1/\lambda} \qquad z_i \sim N(0,1) \text{ say}$$

Interest: How y depends on x

Maybe β ?

Chen Lockhart Stephens 2002   CJS

Maybe β/σ ?

6 Not-so-old example

Fraser, Wong 2004

Power transformed regression Box Cox (1964, 1982)

$$y_i = (\alpha + \beta x_i + \sigma z_i)^{1/\lambda}$$

$z_i \sim N(0,1)$ say   or Student or ...

Interest: How y depends on x

Maybe β ?

Chen Lockhart Stephens 2002   CJS

Yang 2002

Maybe β/σ ?

$$\psi = \beta\lambda^{-1}(\alpha + \beta x_0)^{\frac{1}{\lambda}-1} \longleftarrow \frac{d}{dx}\tilde{E}(y\,|\,x)\Big|_{x_0}$$

Increase in miles for increase in gasoline

"True derivative"

Prior on $\alpha, \beta, \sigma, \lambda$?

Flat (model) prior on particular interest ψ

Usual (L based) p-values .... conditional

but assessment ------ marginal

Try conditional assessment...

a simple example: —

Example 7 Non Normal regression

$y_i = \alpha + \beta x_i + \sigma z_i \qquad z_i \sim \text{Student}(7)$

$n = 7$

Data							
x	-3	-2	-1	0	1	2	3
y	-2.68	-4.02	-2.91	$.22$	$.38$	$-.28$	$.03$

from $\boxed{\alpha = 0 \quad \beta = 1 \quad \sigma = 1}$

Bédard, F, Wong

7 Non Normal regression

$y_i = \alpha + \beta x_i + \sigma z_i$ $z_i \sim$ Student(7)

$n=7$

Data							
x	-3	-2	-1	0	1	2	3
y	-2.68	-4.02	-2.91	$.22$	$.38$	$-.28$	$.03$

from $\boxed{\alpha = 0 \quad \beta = 1 \quad \sigma = 1}$ ← "True"

Test t-statistic $t^o = -.122178$ $\pm\left(\frac{SST}{SSE}\right)^{1/2}$

f 3rd $p^o = .105255 \approx 10.5\%$

(a) What is true p-value ? (b) true s-value ?

MCMC and interesting things !

Bédard, F, Wong

$y_i = \alpha + \beta x_i + \sigma z_i$ $\quad z_i \sim$ Student(7) $\quad n=7$

(a) what is "true" p-value for data?

Distn of $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}) = (a, b, c)$ $\quad$ LS & residual length

$$\boxed{f(a,b,s) = c\, s^4 \prod_1^7 \left\{1 + [a + bx_i + s d_i]^2\right\}^{-4}}$$

MCMC | Proposal $N(0, .35)$

Dump 50, 950 ; repeat 5000 $\quad$ 5M

Record whether "$t \le .105255$"

	By 3rd	By MCMC	
p-value $p(1)$	10.52%	10.76%	Sim SD = .2%

$y_i = \alpha + \beta x_i + \sigma z_i \qquad z_i \sim \text{Student}(7) \qquad n = 7$

(b) What is "true" s-value for data? $\qquad E(\beta)$? $SD(\beta)$?

$$\pi(\theta) L(\theta) = \prod_1^7 \left\{ 1 + \frac{1}{7\sigma^2}[y_i - \alpha - \beta x_i]^2 \right\}^{-4} \frac{1}{\sigma^7} \cdot \frac{1}{\sigma}$$

"True $\beta = 1$"

	By 3rd order	By AMCMC 5M
p-value $p(1)$	10.52%	10.76
s-value $s(1)$	10.53%	10.82%
$E(\beta)$	.676	.672
$SD(\beta)$	.285	.294

Sun Ye, Wong YorkU $\qquad$ Bédard UofToronto

Example 8 SUR

Seemingly Unrelated Regression   Zellner 1962, 1963

$y_1 = X_1 \beta_1 + \sigma_1 z_1$ say N errors

Independent for each regression

$y_2 = X_2 \beta_2 + \sigma_2 z_2$ Correlated at each Time point i

assess $\delta = \beta_{2r} - \beta_{1r}$ ..... ex GE vs Westinghouse

Flat (model) prior for δ ; eliminate remaining (Laplace or other)

Fraser Rekkas Wong   J. Econometrics   2005

Example 8 SUR Seemingly Unrelated Regression Zellner 1962, 1963

$y_1 = X_1\beta_1 + \sigma_1 z_1$ say N errors
Independent for each regression
$y_2 = X_2\beta_2 + \sigma_2 z_2$ Correlated at each time point i

assess $\delta = \beta_{2r} - \beta_{1r}$ ..... ex GE vs Westinghouse

Flat (model) prior for δ; eliminate remaining, Laplace or other

Simulations: Two commodity demand model

$$\log q_1 = \alpha_1 + \beta_1 \log p_1 + e_1$$
$$\log q_2 = \alpha_2 + \beta_2 \log p_2 + e_2$$

$n = 15$

Int $= \delta = \beta_2 - \beta_1$

q = quantity
p = price

Ex: 90% central intervals $N = 5{,}000$

Fraser Rekkas Wong J. Econometrics 2005

<u>Simulations</u>: Two commodity demand model

$$\log q_1 = \alpha_1 + \beta_1 \log p_1 + e_1 \qquad n = 15$$
$$\log q_2 = \alpha_2 + \beta_2 \log p_2 + e_2 \qquad \underline{\text{Int}} = \delta = \beta_2 - \beta_1$$

| q = quantity
| p = price

Ex: 90% central intervals $N = 5{,}000$

Percent outside	Lower Limit	Upper Limit
1) Bootstrap	6.81 %	6.47 %
2) Likelihood ratio	7.00 %	6.83 %
3) Bartlett	5.89 %	5.68 %
4) Lik prior/3rd	4.69 %	5.27 %
Nominal	5 %	5 %

Fraser Rekkas Wong J. Econometrics 2005

General Inference Context

Interest in	$\psi(\theta)$	say	μ	β
Nuisance	λ		σ^2	α, σ^2

Analyze

① measure of departure $t(y)$

say $t = \bar{y} - \mu$

② Modify $t(y)$ to say $\tilde{t}(y)$ --- to eliminate dependence on λ

③ Calibrate $\tilde{t}(y)$ to get p-value $F\{\tilde{t}(y^0); \psi\} = p^0$ free of λ?

Hope: for simple $N\ \mu\ \sigma^2$ case

① Take $\bar{y} - \mu \longmapsto \tilde{t}(y) = \frac{\bar{y} - \mu}{s_y/\sqrt{n}}$

② Would give $p^0(\mu) = H_{n-1}\{\tilde{t}(y^0)\}$

$\uparrow$ Student $(n-1)$ df

General Inference Context

Interest in $\psi(\theta)$
Nuisance λ
What can you do? .... when $t(y)$ is given

Bayarri Berger p-values for composite null models JASA 2000
Robins van der Vaart Ventura asymptotic distn of p-values in composite null models JASA 2000

Many Bayesian ways to eliminate λ, calibrate!
cf "Plug in" $P\{t(y) \le t(y^0) \,;\, (\psi, \hat{\lambda}_\psi)\}$

General options:

1) Bootstrap $t(y)$

2) Version of p_{cpred} $\begin{cases} p \text{ value from conditional given mle} \\ \text{average such using Bayes/Lik from mle distn} \end{cases}$

3) Use a frequentist ancillary to calculate $p\{t(y) \le t(y^0)\}$

1) Bootstrap $t(y)$

2) Version of p_{cpred} { p value from conditional given mle
average such using Bayes/Lik from mle distn

3) Use a frequentist cancillary to calculate $p\{t(y) \le t(y^0)\}$

.....

Rousseau Fraser 2005

Under moderate regularity

① Four BS steps
② Bayesian p_{cpred} calculate
③ frequentist ancillary calculation

give same result $O(n^{-3/2})$

1) Bootstrap $t(y)$

2) Version of p_{cpred} { p value from conditional given mle; average such using Bayes/Lik from mle distn }

3) Use a frequentist cancillary to calculate $p\{t(y) \leq t(y^0)\}$

.....

Rousseau Fraser 2005

Under moderate regularity

① Four BS steps
② Bayesian p_{cpred} calculate
③ frequentist ancillary calculation

give same result $O(n^{-3/2})$

But....

Statistic $t(y)$

Moderate regularity, continuous parameter

BS B f all lead to same result!

But: How to get the $t(y)$? ... available <u>lik</u> theory

Statistic $t(y)$

Moderate regularity, continuous parameter

BS B f all lead to same result!

But: How to get the $t(y)$? ... available <u>lik</u> theory

Need: $\ell(\theta) = \log f(y^0; \theta)$

$\varphi(\theta) = \frac{d}{dy} \ell(\theta; y)\big|_{y^0}$ in directions V from piv

Get: 'Right' $t(y)$ for your chosen interest

Priors?

1) Use Hessian $j_{\varphi\varphi}(\theta; y_0) = -\frac{\partial}{\partial \varphi} \frac{\partial}{\partial \varphi'} \ell(\theta)$

2) For interest $\psi(\theta)$

$$\left[j^{\{\psi\psi\}}(\psi, \hat{\lambda}_\psi) \right]^{-\frac{1}{2}} d\psi \quad \left[j_{\{\lambda\lambda\}}(\psi, \lambda) \right]^{1/2} d\lambda \quad \Bigg\} \begin{array}{l} \{\,\}\ \text{infós} \\ \text{via } \varphi \\ \text{"Obs" Jeffreys...} \end{array}$$