Some thoughts on model-based priors

D Fraser
Statistics Univ Toronto

Thursday Nov 17 2005
Univ of Waterloo

www.utstat.toronto.edu/dfraser

Some thoughts on model-based priors

Model-based prior? Sort of a contradiction

A | affiliation of statisticians...

B | Prior

C | Model

D | Inference

E | B-f interplay

F | B-f exclusiveness

G | Wrap up

A Large affiliation of statisticians...

Persuasion.... "<u>model-based priors</u>"

<u>Name</u>: invariant, default, flat, non-subjective, objective ?

Huge 'industry', committed, power, ...

Believe in 'Cause'; spin; turf

Dare frequentists ignore ? ... Not an idle question !

A Large affiliation of statisticians...

Persuasion.... <u>model-based</u> priors

Name: invariant, default, flat, non-subjective, objective ?

Huge 'industry', committed, power, ...

Believe in 'cause', spin, turf

Dare frequentists ignore ? ... Not an idle question !

"They" think frequentists are <u>asleep</u>: "lost their way!"

Belief
Mystique
Pontification
Statistical version of ID ... this isnt Kansas !

A Large affiliation of statisticians...

Persuasion.... model-based priors

Name: invariant, default, flat, non-subjective, objective ?

Huge 'industry', committed, power, ...

Believe in 'cause', spin, turf

Dare frequentists ignore? ... Not an idle question!

"They" think frequentists are asleep ... lost their way!

Belief
Mystique
Pontification
Statistical version of ID ... this isnt Kansas!

$$\left| \begin{array}{ll} f(y;\theta) & \pi(\theta) \\ y^0 & \end{array} \right\}$$ Isn't this just "frequentist"?

Welch & Peers 1963
Re-designing "frequentist"

B Priors

Classify:

1) Objective
2) Subjective
3) Model-based

B Priors

Classify:

1) Objective → There really is a $\pi(\theta)$ distribution

Fisher 1956 SMSI .... examples

... and many before

Just prob. modeling & conditional prob. !

Hadn't been addressed as such !

Brad Efron Nov 9 at U of T

"50 years of empirical Bayes"

All examples: an objective distribution!

Bayesian ? How so ?

Why "new" ?

B Priors

Classify:

1) Objective

2) <u>Subjective</u> Jimmy Savage 1954 ... but

3) Model-based

B Priors

Classify:

1) Objective

2) Subjective

3) Model-based

Bayes 1763
Laplace 1812
Welch & Peers 1963
Neo-Bayesian Bernardo 1979 DSZ 1973

Name?? Evolutionary....

Invariant
Default
Non-subjective
Objective Call "black white"
Obfuscation
Pre-empt....

Just use $L^{\circ}(\theta)$ and weight it F AMS 1972

but then... frequentists neglected $L^{\circ}(\theta)$
Stanford 1962

B Priors

Classify:

1) Objective

2) Subjective

3) Model-based Neo-Bayesian Just use $L^0(\theta)$ and weight it

Method:

$y ; \theta$ $\quad f(y;\theta)$

y^0

θ $\quad \pi(\theta)$

frenquentist ?

B Priors

Classify:

1) Objective
2) Subjective
3) <u>Model-based</u> <u>Neo-Bayesian</u> Just use $L^{\circ}(\theta)$ and weight it

<u>Method:</u>

$y;\theta$ $\quad f(y;\theta)$

y°

θ $\quad \pi(\theta)$

frenquentist?

<u>Credo:</u>

Prior of convenience....

Use algorithm

and how the algorithm processes information

- Information processing

Eschew ..... <u>any</u> <u>other</u> <u>model</u> <u>structure</u> ! More than "Ety..." shy !

Belief : <u>Suppress</u> things you <u>know</u> Neo-science ID

Laugh at "f" afraid to use $L^{\circ}(\theta)$

C <u>Models</u>

Assume: $f(y;\theta)$ .... in <u>all</u> Bayesian analysis

in <u>some</u> frequentist "

scalar fn on $\{y\} \times \{\theta\}$

How much model info is <u>used</u> / <u>allowed</u> ? .... in Bayesian context ?

Plot $f(\cdot;\cdot)$ over $S \times \Omega$ $\rightarrow$

C Models

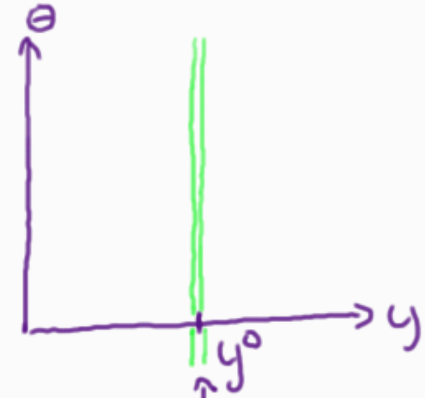
Assume: $f(y;\theta)$ .... in all Bayesian analysis

in some frequentist "

scalar fn on $\{y\} \times \{\theta\}$

How much model info is used / allowed ..... Bayesian context?

Plot $f(\cdot;\cdot)$ over $S \times \Omega$ →



Bayesian uses →

L° .... y°-section of model

plus info-processing .... IP Zellner

ie How model processes $\pi(\theta) \to \pi(\theta|y)$

$$\pi(\cdot) \longrightarrow \boxed{} \longrightarrow \pi(\cdot|y)$$

K-L distance

Only model structure to be used ???

D Inference

$f(y;\theta) + y^o \Rightarrow$?

Case: Scalar y & θ ... Where Things should start!

Obvious: $f^o = f(y^o;\theta)$ $\quad L^o(\theta)$ $\quad$ Prob at

$F^o = F(y^o;\theta)$ $\quad p^o(\theta)$ $\quad$ Prob left

Is there anything more?

D Inference

$f(y;\theta) + y^0 \Rightarrow ?$

Case: Scalar y & θ ... Where Things should start!

Obvious: $f^0 = f(y^0;\theta)$ $\quad L^0(\theta)$ $\quad$ Prob at

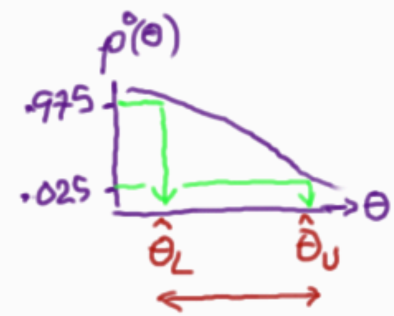
$F^0 = F(y^0;\theta)$ $\quad p^0(\theta)$ $\quad$ Prob left

Is there anything more ?

$u = F(y;\theta)$ is a pivot

Invert; get all confidence bounds, conf. int's

$95\%\ CI = \{p^{-1}(.975),\ p^{-1}(.025)\}$



D Inference

$f(y;\theta) + y^o \Rightarrow$?

Case: Scalar y & θ ... Where Things should start!

Obvious: $f^o = f(y^o;\theta)$ $\quad L^o(\theta)$ $\quad$ Prob at

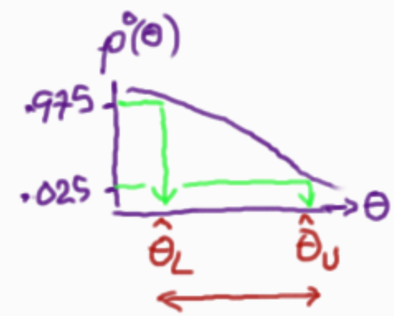
$F^o = F(y^o;\theta)$ $\quad p^o(\theta)$ $\quad$ Prob left

Is there anything more?

$u = F(y;\theta)$ is a pivot

Invert; get all confidence bounds, conf. int's

$$95\%\ CI = \{p^{-1}(.975),\ p^{-1}(.025)\}$$



you better use it or you'll lose it

There is an end run in play !! $\rightarrow L^o(\cdot)$

...think bigger

D Inference

$f(y;\theta) \;+\; y^\circ \;\Rightarrow\; ?$

Case: Scalar y & θ ... Where Things should start!

obvious: $f^\circ = f(y^\circ;\theta)$ $\quad L^\circ(\theta)$ $\quad$ Prob at

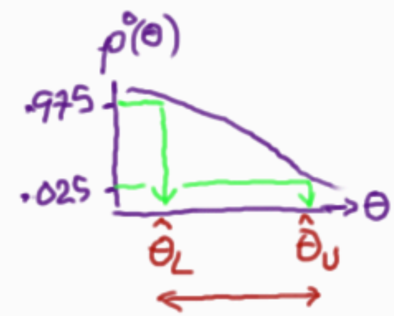
$\qquad\qquad F^\circ = F(y^\circ;\theta)$ $\quad p^\circ(\theta)$ $\quad$ Prob left

Is there anything more?

$u = F(y;\theta)$ is a pivot

Invert; get all confidence bounds, conf. int's

$$95\%\ CI = \{p^{-1}(.975),\ p^{-1}(.025)\}$$

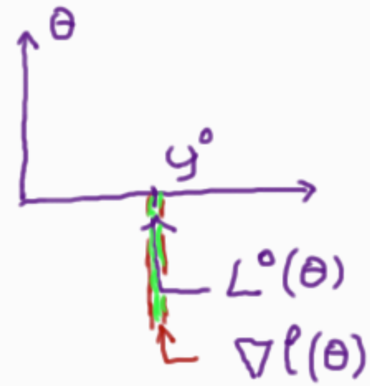


you better use it or you'll lose it

There is an end run in play!! $\rightarrow L^\circ(\cdot)$

Can do in general;

Min. assumptions!



$\ell(\theta)$

$\varphi(\theta)$

you can answer questions
the Bayesians can't even get near!

E B-f interplay

1958 Lindley

Case: y and θ scalar

When is Confidence distribution = Posterior distribution ?

$$p_\theta(\theta) = L(\theta)\,\pi(\theta)\,/\,m(y)$$

$$-F_{;\theta}(y;\theta) = F_y(y;\theta)\,\pi(\theta)\,/\,m(y)$$

Fear
Stigma
real

elim.
nuis.

E B-f interplay

1958 Lindley

Case: y and θ scalar

When is Confidence distribution = Posterior distribution ?

$$p_\theta(\theta) = L(\theta)\,\pi(\theta)\,/\,m(y)$$

$$-F_{;\theta}(y;\theta) = F_y(y;\theta)\,\pi(\theta)\,/\,m(y)$$

$$-\frac{F_{;\theta}}{F_y} = \frac{dy}{d\theta} = \frac{\pi(\theta)}{m(y)}$$

Diff thro' pivot

$$m(y)\,dy = \pi(\theta)\,d\theta$$

$$\tilde{y} = \tilde{\theta} + z \qquad \text{... ie Location model}$$

E B-f interplay

1958 Lindley

Case: y and θ scalar

When is Confidence distribution = Posterior distribution ?

$$p_\theta(\theta) = L(\theta)\,\pi(\theta)\,/\,m(y)$$

$$-F_{;\theta}(y;\theta) = F_y(y;\theta)\,\pi(\theta)\,/\,m(y)$$

$$-\frac{F_{;\theta}}{F_y} = \frac{dy}{d\theta} = \frac{\pi(\theta)}{m(y)}$$

Diff thro' pivot

$$m(y)\,dy = \pi(\theta)\,d\theta$$

$$\tilde{y} = \tilde{\theta} + z \qquad \text{... ie. Location model}$$

Lindley: Carping at upstart Fisher ('30) .... "Conf. is wrong"

Real message: Global Bayes is wrong

E B-f interplay

1958 Lindley

Case: y and θ scalar

When is Confidence distribution = Posterior distribution ?

$$p_\theta(\theta) = L(\theta)\,\pi(\theta)\,/\,m(y)$$

$$-F_{;\theta}(y;\theta) = F_y(y;\theta)\,\pi(\theta)\,/\,m(y)$$

$$-\frac{F_{;\theta}}{F_y} = \frac{dy}{d\theta} = \frac{\pi(\theta)}{m(y)}$$

Diff thro' pivot

$$m(y)\,dy = \pi(\theta)\,d\theta$$

$$\tilde{y} = \tilde{\theta} + z \qquad \ldots \text{ ie Location model}$$

Just examine near data point!

Answer: Data-dependent priors

- Box and Cox
- David C here at Waterloo
- Wasserman ... and many more

E B-f interplay

1958 Lindley

Case: y and θ scalar

When is Confidence distribution = Posterior distribution ?

$$p_\theta(\theta) = L(\theta)\,\pi(\theta)\,/\,m(y)$$

$$-F_{;\theta}(y;\theta) = F_y(y;\theta)\,\pi(\theta)\,/\,m(y)$$

$$-\frac{F_{;\theta}}{F_y} = \frac{dy}{d\theta} = \frac{\pi(\theta)}{m(y)}$$

Diff thro' pivot

$$m(y)\,dy = \pi(\theta)\,d\theta$$

$$\tilde{y} = \tilde{\theta} + z \qquad \ldots \text{ ie Location model}$$

Answer: Data-dependent priors

- Box and Cox
- David C here at Waterloo
- Wasserman ... and many more

Locally (near data)

- Model is location
- f & B agree ! ... local flat prior

Can't use just $\ell(\theta)$

Also need $\varphi(\theta) = \nabla \ell(\theta; y^0)$

$\{\ell(\theta), \varphi(\theta)\}$

F B-f exclusiveness

Bayesian : Always use $\mathcal{L}^o(\theta) = f^o(\theta)$ - - - - - - - - - - and only $\mathcal{L}^o(\theta)$

and conditional probability - - - - - - - - - - plus MCMC (Nuclear power)

frequentist : Ignores $\mathcal{L}^o(\theta)$

- global probabilities - - - easy !
- novel - - - "theory" bankrupt

F B-f exclusiveness

Bayesian : Always use $\mathcal{L}^{o}(\theta) = f^{o}(\theta)$
and conditional probability - - - - - - - - - plus MCMC (Nuclear power)

frequentist : Ignores $\mathcal{L}^{o}(\theta)$
- global probabilities
- novel

Two solitudes !
... in fundamental principles
... just playing 'their own game' fair enough

F B-f exclusiveness

Bayesian : Always use $L^{o}(\theta) = f^{o}(\theta)$
and conditional probability - - - - - - - - - plus MCMC
(Nuclear power)

frequentist : Ignores $L^{o}(\theta)$
- global probabilities
- novel

Two solitudes !

... in fundamental principles
... just playing 'their own game' fair enough

but Is what you know from an investigation a game ?

F B-f exclusiveness

Bayesian : Always use $\mathcal{L}^{o}(\theta) = f^{o}(\theta)$
and conditional probability -------- plus MCMC (Nuclear power)

frequentist : Ignores $\mathcal{L}^{o}(\theta)$
- global probabilities
- novel

Two solitudes ! ... in fundamental principles
... just playing 'their own game' fair enough

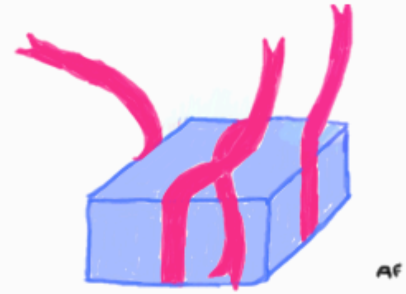
but... Is what you know from an investigation "a game" ?

- I use ONLY "likelihood" .... fully condition
- I use "global" probabilities".... never condition

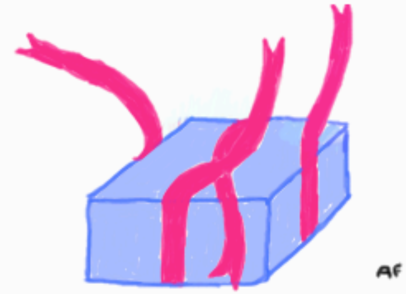
Sand box "TIFF"

... Let's play ... !

G Wrap Up:



G Wrap Up: What are options?



On,... High Principle

- Use only the observed likelihood $L^{\circ}(\theta)$ ---- Bayesian

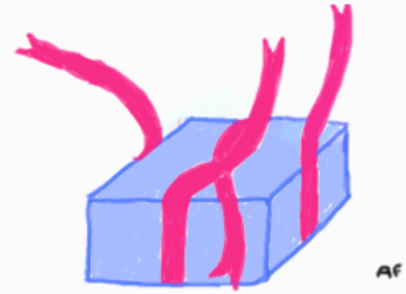
or

- Use only "novel" & global probability assessments ---- frequentist

or

- Use likelihood $L^{\circ}(\theta)$ and recalibrate parameter $\varphi(\theta)$ and more!

G Wrap Up: What are options?



go for it … on High Principle

- use only the observed likelihood $L^\circ(\theta)$ ---- Bayesian

or

- use only novel & global probability assessment ---- frequentist

or

- use likelihood $L^\circ(\theta)$ and calibrate parameter $\varphi(\theta)$ ---

Why exclusiveness? … Dumbing down? Turf?

Structure on parameter … $\varphi(\theta)$ | … go from 1st to 3rd & Deconstruct the model

An example: Q since 1929!

<u>Example</u> Behrens-Fisher 1929 1934 Ghosh Kim 2001

Example Behrens-Fisher 1929 1934 Ghosh Kim 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): $\sigma_1^{-1} \sigma_2^{-1}$ familiar?

Example Behrens-Fisher 1929 1934 Ghosh Kim 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: $\pi = \sigma_1^{-3} \sigma_2^{-3} \left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} \right)$ 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 (data dependent!)

Example Behrens–Fisher 1929 1934 Ghosh Kim 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%

Example Behrens–Fisher 1929 1934 Ghosh Kim 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)

flat prior, data dep., $O(n^{-3/2})$, $B \equiv f$

N=100,000

Recall

1) Priors: Objective, Subjective, Model-based

Recall

1) Priors: Objective, Subjective, <u>Model-based</u>

2) Implications: Use model-based <u>locally</u>: B & f agree to $O(n^{-3/2})$

Recall

1) Priors: Objective, Subjective, Model-based

2) Implications: Use model-based locally: B & f agree to $O(n^{-3/2})$

3) Scalar case $L = f^{\circ}$ $p = F^{\circ}$ Is there more? less?

Recall

1) Priors: Objective, Subjective, Model-based
2) Implications: use model-based locally: B & f agree to $O(n^{-3/2})$
3) Scalar case $L = f^{\circ}$ $p = F^{\circ}$ Is there more? less?
4) Lindley; example ~~confidence is wrong~~ → Bayes Too simplistic, wrong!

Recall

1) Priors: Objective, Subjective, Model-based
2) Implications: Use model-based locally: B & f agree to $O(n^{-3/2})$
3) Scalar case $L = f^{\circ}$ $p = F^{\circ}$ Is there more? less?
4) Lindley; example ~~confidence is wrong~~ → Bayes Too simplistic, wrong!
5) Exclusiveness? I only use L°
I only use 'novel', global prob's
I fully condition
I dont condition
!

Recall

1) Priors: Objective, Subjective, <u>Model-based</u>

2) Implications: Use model-based <u>locally</u>: B & f agree to $O(n^{-3/2})$

3) Scalar case $L = f^{\circ}$ $p = F^{\circ}$ Is there more? less?

4) Lindley; example ~~confidence is wrong~~ → Bayes Too simplistic, wrong!

5) Exclusiveness? I <u>only</u> use L°
I <u>only</u> use 'novel', global prob's
I fully condition
I dont condition
!

6) There's a <u>lot</u> more than L°

Recall

1) Priors: Objective, Subjective, Model-based
2) Implications: use model-based locally: B & f agree to $O(n^{-3/2})$
3) Scalar case $L = f^\circ$ $p = F^\circ$ Is there more? less?
4) Lindley; example ~~confidence is wrong~~ → Bayes too simplistic, wrong!
5) Exclusiveness? I only use L°
   I only use 'novel', global prob's
   I fully condition
   I dont condition
   } !
6) There's a lot more than L°
7) ..... like p-values

Recall

1) Priors: Objective, Subjective, Model-based
2) Implications: use model-based locally: B & f agree to $O(n^{-3/2})$
3) Scalar case $L = f^o$ $p = F^o$ Is there more? less?
4) Lindley; example ~~confidence is wrong~~ → Bayes Too simplistic, wrong!
5) Exclusiveness? I only use L^o
   I only use 'novel', global prob's
   I fully condition
   I dont condition } !
6) There's a lot more than L^o
7) ..... like p-values
8) Model-context: at minimum, calculate $\ell(\theta)$, $\varphi(\theta)$.
   - Get 3rd order $p(\psi)$ for any nice $\psi(\theta)$ scalar
   - Get " " $L(\psi)$ " " " " scalar or vector
   - Get 2nd " $p(\psi)$ " " " " vector
   - Tell the neo-B's how they can do it right

H9

Recall

1) Priors: Objective, Subjective, Model-based

2) Implications: Use model-based locally: B & f agree to $O(n^{-3/2})$

3) Scalar case $L = f^{\circ}$ $p = F^{\circ}$ Is there more? less?

4) Lindley; example ~~confidence is wrong~~ → Bayes Too simplistic, wrong!

5) Exclusiveness? I only use L°
I only use 'novel', global prob's
I fully condition
I dont condition
!

6) There's a lot more than L°

7) ..... like p- values

8) Model-context: at minimum, calculate $\ell(\theta)$, $\varphi(\theta)$

Get 3rd order $p(\psi)$ for any nice $\psi(\theta)$ scalar
Get " " $L(\psi)$ " " " " scalar or vector
Get 2nd " $p(\psi)$ " " " " vector
Tell the neo-B's how they can do it right

Thank you ....