Some thoughts on the
Bayesian-frequentist divergence

D A S Fraser
Statistics
University of Toronto

XL I Congreso Nacional de la Sociedad Matemática Mexicana
22 Octubre 2008
Valle de Bravo

Thanks and Appreciation
to the organizing committee...
You can't be around statistics for long without noticing that there are various "flavors":

frequentist:
Bayesian:
You can't be around statistics for long without noticing that there are various "flavors":

frequentist: what statistic to use...
Bayesian: what prior to use...

Some thoughts on the Bayesian-frequentist divergence.
You can't be around statistics for long without noticing that there are various "flavors":

frequentist: what statistic to use...
Bayesian: what prior to use...

Some thoughts on the Bayesian-frequentist divergence.

Why two cults?

Other cults? Just calculating prob's for ŷ or θ ...
BS, Sim's, MCMC, SP, HOL, ...
Two cults!
Can give different results
Both can't be right!
Two cults!
Can give different results
Both can’t be right!

Particle Collider delayed; \( B-f \) differences;
Many $’s, Mega

CERN LHC

\[ y \sim \text{Poisson}(\theta) \quad \theta > 0. \]
Is \( \theta > 0 \).
Nobel
Overview

Preamble:

Four significant dates: (1) 1763
" " " (2) 1922
" " " (3) 1930
" " " (4) 1958

Bayes & confidence intervals

Nonlinearity corrupts Bayes:

Summary
Four significant dates:

1763 Bayes Posterior

1922 Fisher Likelihood

1930 Fisher Confidence

1958 Lindley Dispute
(1) \( 1763 \) Bayes examined \( f(y - \theta) \) location model (now called)
1763 Bayes examined \( f(y-\theta) \)

he noted:

- location model (now called)
- Translation invariance (now called)
(1) 1763 Bayes examined
\[ f(y - \theta) \]

Noted:

Proposed: weight function \( \pi(\theta) \) with \( \pi(\theta) = C \) by invariance of location model and translation invariance (now called)
(1) 1763 Bayes examined

\[ f(y - \theta) \]

location model (now called)

Translational invariance (now called)

Noted:

Proposed: Weight function \( \Pi(\theta) \) with \( \Pi(\theta) = C \) by invariance

Proposed: "Mathematical" \( \Pi(\theta) \) be treated as physical/real source (for \( \theta \))
1763 Bayes examined

\[ f(y - \Theta) \]

\textbf{Noted:}

\textbf{Location model} (now called)

\textbf{Translation invariance} (now called)

Proposed: Weight function \( \pi(\Theta) \) with \( \pi(\Theta) = C \) by invariance

Proposed: Mathematical \( \pi(\Theta) \) be treated as physical/real source

\textbf{If} \( \pi(\Theta) \text{ is source of } \Theta \), then

\[ (\Theta, y) \sim \pi(\Theta) f(y - \Theta) \]

\[ \Theta | y^0 \sim C \pi(\Theta) f(y^0 - \Theta) \ldots \text{Posterior } \pi(\Theta | y^0) \]
(1) 1763 Bayes examined

\[ f(y-\theta) \]

Location model (now called)

Noted:

Translation invariance (now called)

Proposed: Weight function \( \pi(\theta) \) with \( \pi(\theta) = c \) by invariance proposed: Mathematical \( \pi(\theta) \) be treated as physical/real source

If \( \pi(\theta) \) is source of \( \theta \), then

\[ (\theta, y) \sim \pi(\theta) f(y-\theta) \]

\[ \theta | y^o \sim c \pi(\theta) f(y^o - \theta) \]

But \( \pi(\theta) \) is just a mathematical construct

Conditional probability derivation does NOT apply.
But it sort of worked...?

\[ \theta | y^o \sim \mathcal{N}(\theta) f(y^o - \theta) \]

Why? \hspace{1cm} (Needs linearity)
1922 Fisher. Likelihood: $L(\theta) = f(y^o; \theta)$.

Have $f(y; \theta)$; Have $y^o$; Have model section $f(y^o; \theta)$!
(2) 1922 Fisher Likelihood: $L(\theta) = f^0(\theta)$

Have $f(y; \theta)$; Have $y^0$; Have model section $f(y^0; \theta)$!

But: Bayes, Laplace, ... had been using it for 150+ years "Operational" in a full sense
(2) 1922 Fisher. Likelihood: \( L(\theta) = f^*(\theta) \)

Have \( f(y; \theta) \); Have \( y^* \); Have model section \( f(y^*; \theta) \)!

But: Bayes, Laplace, ... had been using it for 150+ years "Operational"

Just integrate with a strategic \( \pi(\theta) \)!

Naming Likelihood ??
(3) 1930 Fisher Confidence: \( F(y^0; \theta) \) is right-tail disln function for \( \theta \)
(3) 1930 Fisher Confidence: $F(y^o; \Theta)$ = right-tail disfn function for $\Theta$

Has good probability behavior! Neyman 1937 "classical"
(3) 1930 Fisher Confidence: $F(y^0; \theta) =$ right-tail distribution function for $\theta$

Has good probability behavior! Neyman 1937 "classical"

$$F(y^0; \theta) = \beta \quad \Rightarrow \quad \hat{\theta}_\beta = \beta \text{-level confidence quantile}$$

\[\text{"}\hat{\theta}_\beta(y) < \theta\text{" is true with probability } \beta\] Neyman said so!
(3) 1930 Fisher confidence: $F(y^0; \Theta) =$ right-tail disfn function for $\Theta$

Has good probability behavior! Neyman 1937 "classical"

$F(y^0; \Theta) = \beta \quad \Rightarrow \quad \hat{\Theta}_\beta = \beta$-level confidence quantile

"$\hat{\Theta}_\beta(y) < \Theta$" is true with probability $\beta$

In Bayes' special case: $f(y-\Theta)$

$\int_{\Theta}^\infty f(y^0 - \Theta) \, d\Theta = \beta \quad \Rightarrow \quad \hat{\Theta}_\beta = \beta$-level posterior quantile

= $\hat{\Theta}_\beta$ same as confidence

In this case (location) Bayes does have (probability behavior)

Bayes got confidence first! ~170 years
For example:
\[ N(\mu, 1) \]
\[ N(y^0, 1) \]
\[ \rho(\theta) = \Delta(\theta) \]
Freq \( \Rightarrow \) Bayes
(4) 1958 Lindley:

Bayes = Confidence ... only for Location models

"F Treaded on Bayesian territory"
1958 Lindley:

\[ \text{Bayes} = \text{Confidence} \ldots \text{only for Location model} \]

**In Bayes case:** \( f(y-\theta) \)

"Bayes quantile = Confidence quantile" (probability properly)

**If not in Bayes case** \( f(y-\theta) \)

Bayes is NOT confidence

Bayes does NOT have (probability properly)
(4) 1958 Lindley:

Bayes = Confidence... only for Location model

In Bayes case: \( f(y - \theta) \)

Bayes quantile = Confidence quantile (probability properly)

If not in Bayes case \( f(y - \theta) \)

Bayes is NOT confidence

Bayes does NOT have (probability properly)

Lindley implied... confidence is wrong!

Fisher is wrong!
Scalar case: Bayes & Confidence dist'ns

(i) Bayes
\[ \Pi(\theta; y^*) = c \pi(\theta) f(y^*; \theta) \]

Upper tail
\[ \delta(\theta) = \int_\theta^\infty c \pi(\alpha) F_y(y^*; \alpha) \, d\alpha \]

\[ F_y = \frac{d}{dy} F(y) = f(y) \]
Scalar case: Bayes and Confidence dist'ns

(i) Bayes
\[ \Pi(\theta; y^o) = c \frac{\Pi(\theta)}{f(y^o; \theta)} \]
\[ \text{upper-tail}_{\text{Bayes}} = \tilde{\delta}(\theta) = \int_{\theta}^{\infty} c \Pi(\alpha) F_{\theta}(y^o; \alpha) d\alpha \]
\[ F_y = \frac{\partial}{\partial y} F(y) = f(y) \]

(ii) Fisher
\[ \Pi^*(\theta; y^o) = \text{Differentiate} \ F(y^o; \theta) \text{ i.e.} \frac{\partial}{\partial \theta} F(y^o; \theta) \]
\[ \text{upper-tail}_{\text{confidence}} = \delta(\theta) = \int_{\theta}^{\infty} -F_{\theta}(y^o; \alpha) d\alpha \]
\[ F_{\theta} = \frac{\partial}{\partial \theta} F(y; \theta) \]
Scalar case: Bayes and confidence dist’ns

(i) Bayes
\[ \Pi(\theta; y^0) = c \frac{\pi(\theta)}{f(y^0; \theta)} \]
\[ \text{upper tail} = \delta(\theta) = \int_\theta^\infty c \pi(\alpha) \frac{F_{y^0}(y^0; \alpha)}{f(y^0)} \, d\alpha \]
\[ F_{y^0} = \frac{d}{dy} F(y) = f(y) \]

(ii) Fisher
\[ \widetilde{\Pi}(\theta; y^0) = -\frac{d}{d\theta} F(y^0; \theta) \]
\[ \text{upper tail} = \tilde{\delta}(\theta) = \int_\theta^\infty -F_{y^0}(y^0; \alpha) \, d\alpha \]
\[ F_{\theta y^0} = \frac{d}{d\theta} F(y; \theta) \]

(iii) Bayes posterior \equiv confidence iff integrands equal!

Solve \[ \Pi(\theta) = -\frac{F_{\theta y^0}(y^0; \theta)}{F_{y^0}(y^0; \theta)} = \frac{dy}{d\theta}|_{y^0} \]
\[ y = y(u; \theta) \]
"Quantifies" Lindley

quantile function

Solution of \( u = F(y; \theta) \)
Scalar case: Bayes and Confidence dist'ns

(i) Bayes
\[ \Pi(\theta; y^o) = c \Pi(\theta) f(y^o; \theta) \]
\[ \text{upper tail} = \delta(\theta) = \int_\Theta^\infty c \Pi(\alpha) F_{y^o}(y^o; \alpha) \, d\alpha \]
\[ F_{y^o} = \frac{\partial}{\partial y} F(y) = f(y) \]

(ii) Fisher
\[ \tilde{\Pi}(\theta; y^o) = -\frac{\partial}{\partial \theta} F(y^o; \theta) \]
\[ \text{upper tail} = \tilde{\delta}(\theta) = \int_\Theta^\infty -F_{y^o}(y^o; \alpha) \, d\alpha \]
\[ F_{\tilde{\delta}} = \frac{\partial}{\partial \delta} F(y; \theta) \]

(iii) $\beta$-posterior = confidence iff integrands equal!

Solve
\[ \Pi(\theta) = -\frac{F_{\tilde{\delta}}(y^o; \theta)}{F_{y^o}(y^o; \theta)} = \frac{dy}{d\theta} \bigg|_{y^o} \]
\[ y = y(u; \theta) = \text{quantile function} \]
\[ y = \text{Solution of } u = F(y; \theta) \]

Quantifies Lindley

Data dependent priors
Box & Cox 1964
Wasserman 2000
F Reid Marras Yi 2008 "default priors"
4 Can things go wrong? Three examples!

Ex 1 Bounded parameter range

\[ \text{Ex: } y \sim \text{Normal}(\theta; 1) \quad \theta_0 \leq \theta \]

- Frequentist: \( p(\theta) = \Phi(y^0 - \theta) \)
- \( \hat{\theta}_\beta(\theta) = y^0 - z_\beta \)

Use \( \theta_0 = 0 \)
Like Poisson(\theta) example
\[ y_1 \sim \text{Normal}(\theta; 1) \]

**frequentist:**

\[ p(\theta) = \Phi(y^0 - \theta) \]

\[ \hat{\theta}_\beta (\theta) = y^0 - z_\beta \]

**Bayes:**

\[ \lambda(\theta) = \Phi(y^0 - \theta) / \Phi(y^0) \quad \theta > 0 \]

\[ \tilde{\theta}_\beta (\theta) = y^0 - z_\beta \Phi(y^0) \]
$y_1 \sim \text{Normal} (\theta; 1)$

**Frequentist:**

$p(\theta) = \Phi(y^0 - \theta)$

$\hat{\theta}_p (\theta) = y^0 - z_{\beta}$

**Bayes:**

$
\Delta (\theta) = \Phi(y^0 - \theta)/\Phi(y^0) \quad \theta > 0
$

$\tilde{\theta}_p (\theta) = y^0 - z_{\beta} \Phi(y^0)$

Can calculate:

$Pr["\tilde{\theta} < \theta"] = \beta$ ?

Actual falls short of claimed 50%
Normal on the plane

\[
\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \text{Normal}\left( \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)
\]

Interest in \( \rho = (\theta_1^2 + \theta_2^2)^{1/2} \)

Write \( \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} r \cos \alpha \\ r \sin \alpha \end{pmatrix} \)

\( r \) "measures" radial distance \( r \)

Normal on the plane
Ex2 Normal on the plane \[ O(n^{-1/2}) \]

\[
\begin{pmatrix}
 y_1 \\
 y_2
\end{pmatrix}
\sim \text{Normal}\left( \begin{pmatrix}
 \theta_1 \\
 \theta_2
\end{pmatrix}; \begin{pmatrix}
 1 & 0 \\
 0 & 1
\end{pmatrix} \right)
\]

Interest in \( p = (\theta_1^2 + \theta_2^2)^{1/2} \)

Write \( \begin{pmatrix}
 y_1 \\
 y_2
\end{pmatrix} = \begin{pmatrix}
 r \cos \alpha \\
 r \sin \alpha
\end{pmatrix} \) \( r \) "measured" radial distance \( r \)

freq p-value \( p(p) = \Pr \{ \chi^2_2(p^2) < r^2 \} \)

Bayes s-value \( \tilde{s}(p) = \Pr \{ \chi^2_2(r^2) > p^2 \} \)

Can differ "BIG"!

Bayes Excess = \( \tilde{s}(p) - p(p) \)

Non Central \( \chi^2 \) with 2 df

Dawid Stone Židek (1973)
Short fall

Actual when claimed 90%
(c) Normal \{ \theta; \sigma^2(\theta) \}

\[ \sigma^2(\theta) = 1 + \gamma \theta^2/2n \]

Confidence:
\[ \tilde{\theta}_\beta(y) = y - z_\beta - \gamma z_\beta \left( y - z_\beta \right)^2 / 4n \]

Bayes:
\[ \text{prior} = \exp \left\{ a \theta/n + b \theta^2 / 2n \right\} \]

Difference:
\[ \tilde{\theta}(y) - \hat{\theta}(y) = \left\{ \frac{x}{2n} + \frac{c}{2n} \right\} (y - z_\beta) + \frac{a}{n^{1/2}} + \frac{c}{2n} y \]

Bayes prior cannot match confidence \( \text{i.e. No } a, c \text{ that works!} \)

Actual (\theta) = Probability \{ \tilde{\theta}_\beta(y) < \text{true } \theta \} \]

\[ = \beta - \frac{x}{2n} \theta \phi(z_\beta) \ldots \text{using best choice of prior} \]
When $\beta = 50\%$

- Overslanted
- Short fall
SUMMARY

With scalar $\theta$, continuous case

(i) Use a $\beta$-level conf. bound $\hat{\theta}_\beta$:

then $\hat{\theta}_\beta < \text{the True } \theta \ldots$ a prop' $\beta$ of the time

$\hat{\theta}_\beta > \ldots \ldots \ldots \ldots 1-\beta \ldots \ldots \ldots$

over investigations current or coming!
SUMMARY

With scalar $\Theta$, continuous case

(i) Use a $\beta$-level confidence bound $\hat{\Theta}_\beta$:

then $\hat{\Theta}_\beta < \text{the True } \Theta$ ... a prop'n $\beta$ of the time

$\hat{\Theta}_\beta > " " \ldots " " 1-\beta " "$

over investigations current or coming! ... whatever

(ii) Use a $\beta$-level Bayes bound $\tilde{\Theta}_\beta$

(and if different from confidence)

Then sometimes larger & prop'n with $\tilde{\Theta}_\beta < \text{True } \Theta$ is decremented

" " smaller & " " is increased

For these "To average out" and give claimed prop'n $\beta$

the real prior has to precisely balance out preceding discrepancies
SUMMARY

With scalar θ, continuous case

(i) Use a β-level conf. bound \( \hat{\Theta}_β \):

then \( \hat{\Theta}_β < \text{the True } \theta \) ... a prop'n β of the time \( \hat{\Theta}_β > " " ... " " 1-β " " \)
over investigations current or coming! ... whatever

(ii) Use a β-level Bayes bound \( \tilde{\Theta}_β \)
(and if different from confidence)

Then sometimes larger & prop'n with \( \tilde{\Theta}_β < \text{True } \theta \) is decremented
"smaller & " " " " is increased
For these "to average out" and we claimed prop'n β
the real prior has to precisely balance out preceding discrepancies

(iii) A prior in general cannot produce confidence
- More is needed than \( L(\theta) \)
- Data dependent priors
Thank you :)