Bayes linearity confidence
D A S Fraser
Statistics U of Toronto

November 24 2008
Columbia University

http://fisher.utstat.toronto.edu/documents/col08.pdf
Background

Bayes' proposal

Extended Bayes

\underline{Compare B & f ( Scalar )}

\underline{Compare B & f ( by quantiles )}

Two generic examples

Likelihood asymptotics example

Summary
Background
Bayes' proposal
Extended Bayes

\underline{Compare B & f (Scalar )}
\underline{Compare B & f (by quantiles )}

Two generic examples
Likelihood asymptotics example
Summary

Simple real problem:
\( f(y; \Theta) \) is stochastically increasing, \( \not\equiv f(y-\Theta) \)
Want: Power bound \( \hat{\Theta}(y) \) with a 90% "meaning"

Frequentist: Easy; Solve \( F(y; \Theta) = 90\% \)
Bayesian: Can't do it?
1a) Bayesian-frequentist “differences” cost many millions of Dollars


High energy physics  Colliders  CERN  LHC
Bayes-frequentist "differences" cost many millions of Dollars


High energy physics Colliders CERN LHC

Core problem:

\[ y \sim \text{Poisson}(\Theta) \quad \Theta > \Theta_0 \quad \text{Background radiation } \Theta_0 \]

Detected \( \Theta > \Theta_0 \) New particle Nobel

B-f differences caused delays ... many millions ... several years

Bounded parameters: $\text{Poisson}(\theta), \theta > \theta_0 \quad \text{and} \quad N(\theta, \sigma^2), \theta > \theta_0$

Bounded parameters:  Poisson(θ), θ ≥ θ₀  N(θ, σ₀²), θ ≥ θ₀

Central β-confidence intervals

... but problem is one-sided

Almost nothing (paper or discussion)

... p-value, Bayes, likelihood

Bounded parameters: $\text{Poisson}(\theta), \theta \geq \theta_0$ $N(\theta, \sigma_0^2), \theta \geq \theta_0$

Central $\beta$-confidence intervals

... but problem is one-sided

Almost nothing (paper or discussion)

... $p$-value, Bayes, likelihood

Coherent view? ... (author, discussants)

... scattered

HOL Higher Order Likelihood

Bayes (1763) proposed:

Model: \( f(y - \theta) \) ... now called: location model; data \( y^0 \)
Bayes (1763) proposed:

Model: \( f(y; \theta) \) ... now called: location model; data \( y^0 \)

Mathematical prior: \( \Pi(\theta) = 1 \) ... invariance

Use with: \[
\Pi(\theta | y = c \Pi(\theta) f(y; \theta))
\]
2a) Bayes (1763) proposed:

Model: \( f(y - \theta) \) ... now called: location model; data \( y^0 \)

Mathematical prior: \( \Pi(\theta) = 1 \) ... invariance

Use with: \[
\Pi(\theta | y) = c \Pi(\theta) f(y; \theta)
\]

Mathematical posterior \( \Pi(\theta | y^0) = f(y^0 - \theta) \)

Posterior survivor/right tail df \( = \int_\theta^\infty f(y^0 - \theta) \, d\theta = \Delta(\theta) \)
2b) Location case $f(y - \mu)$

$p$-value = $F^*(\mu)$

$y \sim EV(\mu)$
2b) Location case $f(y-\mu)$

Frequentist

$\rho$-value = $F^*(\mu)$

Bayesian

$\ell(\mu)$: density of $\mu$

posterior survivor

$\mu \sim -EV(y^*)$
2b) Location case \( f(y - \mu) \)

**Frequentist**

\[ \text{p-value} = F^*(\mu) \]

\[ y \sim EV(\mu) \]

**Bayesian**

\[ \mu \sim ^{-\text{EV}}(y^0) \]

\[ \text{Posterior survivor} \]

\[ f: \text{Pivot: } z = y - \mu \sim f(z) \]

\[ \text{p-value} = p(\mu) = \int_{y^0}^{\infty} f(y - \mu) \, dy \]

\[ \mu \int_{\mu}^{\infty} f(y - \mu) \, d\mu = p(\mu) = \text{p-value} \]
2b) Location case $f(y - \mu)$

Frequentist

$p$-value $= F^0(\mu)$

Bayesian

Posterior $\mu$ survivor

$f$: **Pivot:** $z = y - \mu \sim f(z)$

$p$-value $= P(\mu) = \int_{-\infty}^{y^0} f(y - \mu) dy = \int_{-\infty}^{y^0 - \mu} f(z) dz = \int_{\mu}^{\infty} f(y^0 - \mu) d\mu = P(\mu) = S$-value

Just confidence! Does have meaning: Neyman 1937
3. **Extended Bayes**

\[ P(\theta | y^0) = cP(\theta) f(y^0; \theta) \] \hspace{1cm} \text{Laplace, ISBA}

It sort-of worked! \hspace{1cm} \text{but why?}
3. Extended Bayes

\[ \pi(\theta | y^0) = c \pi(\theta) f(y^0; \theta) \]  
Laplace, ISBA

It sort-of worked! but why?

(i) Uses Likelihood

\[ \pi(\theta | y^0) = c \pi(\theta) L^0(\theta) \]  
L^0(\theta) "+" Calibration \pi(\theta)

L^0(\theta)… fundamental frequentist concept!

Who gets credit? But "frequentists" don't use it ---!
3. Extended Bayes

\[ \Pi(\theta | y^o) = c \pi(\theta) f(y^o; \theta) \quad \text{Laplace, ISBA} \]

It sort-of worked! but why?

(i) Uses Likelihood

\[ \Pi(\theta | y^o) = c \pi(\theta) L^o(\Theta) \quad L^o(\Theta) \text{ "+" Calibration } \pi(\theta) \]

\[ L^o(\Theta) \ldots \text{ fundamental frequentist concept!} \]

Who gets credit? But frequentists don't use it ... !

(ii) Special Bayes case: \( f(y - \Theta) \) ?

Posterior is just confidence ... !

But: is it more?
3. Extended Bayes

\[ \Pi(\theta | y^0) = c \cdot \Pi(\theta) f(y^0; \theta) \]  
Laplace, ISBA

It sort-of worked! but why?

(i) Uses Likelihood

\[ \Pi(\theta | y^0) = c \cdot \Pi(\theta) L^0(\theta) \]  
\[ L^0(\theta) \text{ is fundamental frequentist concept!} \]

Who gets credit? But frequentists don’t use it...!

(ii) Special Bayes case: \( f(y - \theta) \)?

Posterior is just confidence...!

But is it more?

(iii) \( \Pi(\theta | y^0) \) is "mathematical" posterior! meaning?

Does it satisfy probability rules?

Beyond "belief"?
Bayes summary

- Introduced use of $L^0(\theta)$
- Added a calibration $\Pi(\theta)$
- If linear ($y \leftrightarrow \theta$), gives confidence
- Likelihood is 1st order $O(n^{1/2})$ accurate
  is neglected by frequentists!

Fisher 1930 Neyman 1937
Bayes summary

- Introduced use of $L^o(\theta)$
- Added a calibration $\pi(\theta)$
- If linear ($y \leftrightarrow \theta$), gives confidence
- Likelihood is 1st order $O(n^{1/2})$ accurate
  is neglected by frequentists!

Disclaimer:

Likelihood is a powerful ingredient
- Use it, weight it, play with it
Can mislead

$L$ is not whole story
Bayes summary

- Introduced use of $L^0(\theta)$
- Added a calibration $\Pi(\theta)$
- If linear ($y \leftrightarrow \theta$), gives confidence
- Likelihood is 1st order $O(n^{1/2})$ accurate ...

... is neglected by frequentists!

Fisher 1930 Neyman 1937

Disclaimer:

Likelihood is a powerful ingredient
- Use it, weight it, play with it
Can mislead

$L$ is not whole story

Might want higher accuracy!

What if $y \leftrightarrow \theta$ nonlinear?
4 Compare B & f (Scalar case)

(i) Bayes

\[ \pi(\theta) f(y^0; \theta) \]

\[ \text{Upper Tail } \rightarrow s(\theta) = \int_0^{\infty} c \pi(\theta) F_y(y^0; \theta) d\theta \]

\[ F_y = \frac{\partial}{\partial y} F(y; \theta) \]

\[ = \text{pdf} \]
\( (i) \) Bayes
\[ \pi(\theta) f(y^0; \theta) \]
\[ \text{Upper Tail} = s(\theta) = \int_\theta^\infty c \pi(\theta) F_y(y^0; \theta) \, d\theta \]

\( (ii) \) (Fisher-Neyman)
\[ \pi(\theta; y^0) = -\frac{\partial}{\partial \theta} F(y^0; \theta) \]
\[ \text{Upper Tail} = \tilde{s}(\theta) = \int_\theta^\infty -F_y(y^0; \theta) \, d\theta. \]

\[ F_y = \frac{\partial}{\partial y} F(y; \theta) \]

\[ u = F(y; \theta) \text{ is pivotal} \]
\[ F_{\theta} = \frac{\partial}{\partial \theta} F(y; \theta) \]
(i) Bayes

\[ \pi(\theta | y^0, \theta) \]

Upper tail = \( S(\theta) = \int_\theta^\infty c \pi(\theta) F_y(y^0; \theta) \, d\theta \)

(ii) [Fisher-Neyman]

\[ \pi(\theta; y^0) = -\frac{\partial}{\partial \theta} F(y^0; \theta) \]

Upper tail = \( S(\theta) = \int_\theta^\infty -F_{\theta}(y^0; \theta) \, d\theta \)

(iii) \( B = \text{Conf} \) iff Integrands same

\[ \pi(\theta) = \frac{-F_{\theta}(y^0; \theta)}{F_y(y^0; \theta)} = \frac{dy}{d\theta | y^0} \]

Quantile fn \( y = y(u; \theta) \)

Total diff \( u = F(y; \theta) \)

Intrinsic!
(i) Bayes

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

UpperTail = $$\delta(\theta) = \int_\theta^\infty c \pi(\theta) F_y(y^0; \theta) d\theta$$

$$E_{\pi} = \frac{\partial}{\partial \theta} F(y; \theta)$$

(ii) Fisher-Neyman

$$\pi(\theta; y^0) = -\frac{\partial}{\partial \theta} F(y^0; \theta)$$

Upper tail = $$\delta(\theta) = \int_\theta^\infty -F_{\theta}(y^0; \theta) d\theta$$

$$u = F(y^0; \theta)$$ is pivot

$$F_{\theta} = \frac{\partial}{\partial \theta} F(y^0; \theta)$$

(iii) $$B = \text{Conf \ if } \text{Integrands same}$$

$$\pi(\theta) = -\frac{F_{\theta}(y^0; \theta)}{F_y(y^0; \theta)} = \frac{du}{d\theta} |_{y^0} \quad \text{Quantile function} \quad y = y(u; \theta)$$

$$\text{Total diff} \quad u = F(y^0; \theta)$$

Intrinsic!

Data-dependent prior:

- Box & Cox 1964
- Wasserman 2000
- F Reid Marinav Yi 2008 default priors!
4. Compare $B \& f$ (more general)

(i) Use quantile of distribution (posterior or confidence)

Try 50\%-ile, 90\%-ile, $\beta$-\%ile

Claim: "$\beta$" that $\theta$ in $(\hat{\theta}_\beta, \infty)$
4 Compare B & f (more general)

(i) Use quantile of Distribution (posterior or confidence)
Try 50%ile 90%ile \( \beta \)-%ile

Claim: "\( \beta \)" that \( \theta \) in \( (\hat{\theta}_\beta, \infty) \)

(ii) Then Neyman diagram

"Distribution" claims "\( \beta \)" that \((y, \theta)\) in \(A_\beta\) | True \( \theta \)
Resultantly...
(iii) **Calculate**: Proportion of "True statements"

If $\theta$, \[ \text{Propn}(\theta) = \Pr \{ \theta \in (\hat{\theta}(y), \infty) ; \theta \} \]

$\beta$ ?
(iii) **Calculate**: Proportion of True statements

If $\theta$, \[ \text{Propn}(\theta) = \Pr \{ \theta \in (\hat{\theta}_\beta(y), \infty) ; \theta \} \]

If $\tau(\theta)$, \[ \text{Propn}(\tau) = \Pr \{ \theta \in (\hat{\theta}_\beta(y), \infty) ; \theta \sim \tau(\cdot) \} \]

Are these equal to $\beta$, or to what?
5 Examples: non linear

Example 1 \( N(\theta; 1) \) \( \text{Know } \theta \geq 0 \) .... like Poisson example

\[ L(\theta) = \phi(y^0 - \theta) \quad \theta \geq 0 \]
\[ = 0 \quad \theta < 0 \]

[Graph showing the distribution of \( L(\theta) \) with \( y^0 \) as an estimate of \( \theta \).]
Example 1 \( N(\theta; 1) \) know \( \theta \geq 0 \)

\[
L(\theta) = \phi(y^0 - \theta) \quad \theta \geq 0
\]

\[
= 0 \quad \text{otherwise}
\]

Bayes:

\[
\pi(\theta; y^0) = \frac{\phi(y^0 - \theta)}{\Phi(y^0)}
\]

\[
\lambda(\theta) = \frac{\Phi(y^0 - \theta)}{\Phi(y^0)}
\]

\( \beta \)-quantile \( \hat{\theta}_\beta = y^0 - z_\beta \Phi(y^0) \)
Example 1: \( N(\theta; 1) \) \( \text{Known } \theta \geq 0 \)  

\[
L(\theta) = \phi(y^0 - \theta) \quad \theta \geq 0
\]

\[
= 0 \quad o/w
\]

Bayes:  
\[
\pi(\theta; y^0) = \frac{\phi(y^0 - \theta)}{\Phi(y^0)}
\]

\[
\lambda(\theta) = \frac{\Phi(y^0 - \theta)}{\Phi(y^0)}
\]

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

\[
\text{pr}(\theta) = \text{Pr}\{ z < z_\beta \Phi(\theta + z) \} \quad z \sim N(0, 1)
\]

\[
\text{Pr}(\theta) = \text{Plot}... 
\]

Woodroffe (2000)
Ye Sun
"Propn" is strictly less than "claimed" Neyman! No way that posterior can be probability.
\[ \beta = 90\% \]
\[ \beta = 10\% \]

"Propn" is strictly less than "Claimed"
No way that posterior can be probability
Example 2 \[ y \sim N(\theta, I) \]  
\[ \rho = (\theta_1^2 + \theta_2^2)^{1/2} \]

\[ f: \quad \rho(\rho) = P_n \{ \chi^2(\rho^2) \leq n^2 \} \]

Non-central \( \chi^2 \) with 2 df
\[ \chi^2 = y_1^2 + y_2^2 \]

\[ F, \text{Reid (2001)} \quad \text{Ye Sun (2008)} \]
Example 2 \( y \sim N(\bar{\theta}, I) \)  
\( \rho = (\theta_1^2 + \theta_2^2)^{1/2} \) ... curved  
\( o(\bar{n}^{1/2}) \)

\[ f: \quad \rho(\rho) = \Pr\{ \chi^2(\rho^2) \leq n^2 \} \]

\[ B: \quad \delta(\rho) = \Pr\{ \rho^2 \leq \chi^2(n^2) \} \]

Differ! Bug!

Non Central Chi\(^2\) with 2 df
\[ n^2 = y_1^2 + y_2^2 \]

\[ \Theta \sim N(\bar{\theta}, I) \]

Example 2 \( y \sim N(\theta; I) \) \( \rho = \left( \hat{\theta}_1 + \hat{\theta}_2 \right)^{1/2} \) curved

\[
f(\rho) = P_n \{ \chi^2(p^2) \leq \rho^2 \}
\]

\[
B: \quad s(\rho) = P_n \{ \rho^2 \leq \chi^2(n^2) \}
\]

Differ! Bug

\[
\text{Propn}(\rho) = \text{Prob} \left[ \chi^2_{1-\beta} \left\{ \chi^2(p^2) \right\} \leq \rho^2 \right]
\]

--- for the B posterior

\( h^2 = y_1^2 + y_2^2 \)

Non-Central Chi\(^2\) with 2 df

\( \rho = (\hat{\theta}_1 + \hat{\theta}_2)^{1/2} \)
\[ \beta = 50\% \]

Strictly less than \( \beta \)

No way that posterior can be related to probability
\( \beta = 90\% \)

\( \beta = 10\% \)

again...

strictly less!
Ex: Lik. Asymptotics \[ y \sim N(\theta; 1 + \sigma^2/2n) \]

SD depends on mean
Linear ? Quadratic
Ex: Lik. Asymptotics \( y \sim N(\theta; 1 + \sigma^2/2n) \)  SD depends on mean

Linear? Quadratic

f: Confidence quantile... easy!

\[ \hat{\theta}_\beta(y) = y - z_\beta - \sigma z_\beta (y - z_\beta)^2 / 4n \]

\( O(n^{-3/2}) \)
Ex: Lik. Asymptotics \( y \sim N(\theta; 1 + \sigma^2/2n) \) \( \text{SD depends on mean} \) \( \text{Linear? Quadratic} \)

f: **Confidence quantile** ... easy!

\[ \hat{\theta}_\beta(y) = y - z_\beta - \gamma z_\beta (y - z_\beta)^2 / 4n \]

\( \mathcal{O}(n^{-3/2}) \)

B: **Posterior quantile** ... harder

Expand prior:

\[ \pi(\theta) = \exp \left\{ a \theta/n + c \theta^2/2n \right\} \]

\( \theta/n^{1/2} \ldots \text{discards} \)

\[ \hat{\pi}(\theta) = \exp \left\{ a \theta/n + c \theta^2/2n \right\} \]

\( \text{If not linear} \)
Ex: Lik. Asymptotics \( y \sim N(\theta ; 1 + \sigma \theta^2/2n) \)  SD depends on mean
Linear ? Quadratic

\[ \hat{\theta}_\beta(y) = y - z_\beta - \sigma z_\beta (y - z_\beta)^2/4n \quad O(n^{-3/2}) \]

B: Posterior quantile ... harder

Expand prior: \( \pi(\theta) = e^{\theta^2/2n} \)

\[ \pi(\theta) = e^{a \theta/n + c \theta^2/2n} \]

\[ \theta/n^{1/2} \ldots \text{discards...} \]

If not linear

Diff: Can get difference of quantiles: "Bayes-confidence"! Bayes can't do it
and see what is feasible

\[ \hat{\theta}_B - \hat{\theta}_f = \left\{ \frac{x}{2n} + \frac{c}{2n} \right\} (y - z_\beta) + \frac{a}{n^{1/2}} + \frac{c}{2n} y \]
Ex: Lik. Asymptotics \( y \sim N(\theta; 1 + \sigma^2/2n) \) SD depends on mean
Linear? Quadratic

f: Confidence quantile ... easy!

\[
\hat{\Theta}_\beta(y) = y - \frac{z_\beta}{2} - \frac{\sigma z_\beta}{2} (y - z_\beta)^2 / 4n = O(n^{-3/2})
\]

B: Posterior quantile ... harder

Expand prior: \( \pi(\theta) = \exp\{a \theta/n + c \theta^2/2n\} \) \( \theta/n^{1/2} \) ... discredit...

\[
\pi(\theta) = \exp\{a \theta/n + c \theta^2/2n\} \quad \text{If not linear}
\]

Diff: Can get difference of quantiles: "Bayes-confidence"! Bayes can't do it

and see what is feasible

\[
\hat{\Theta}_B - \hat{\Theta}_f = \left\{ \frac{x}{2n} + \frac{c}{2n} \right\} (y - z_\beta) + \frac{a}{n^{1/2}} + \frac{c}{2n} y_j
\]

Sum: No choice prior can give confidence. i.e. via weighted Lik

"Best" choice: flat at center ... i.e. \( a = 0 \) \( c = 0 \)
Ex: Lik. Asymptotics \( y \sim N(\theta ; 1 + \frac{\sigma \theta^2}{2n}) \) SD depends on mean Linear? Quadratic

f: Confidence quantile ... easy!

\[
\hat{\theta}_y(y) = y - z_\beta - \frac{\sigma}{2} z_\beta (y - z_\beta)^2 / 4n \quad O(n^{-3/2})
\]

B: Posterior quantile ... harder

Expand prior: \( \pi(\theta) = \exp\{a \theta / n^{1/2} + c \theta^2 / 2n\} \)

\[\pi(\theta) = \exp\{a \theta / n + c \theta^2 / 2n\}\]

If not linear \( \theta / n^{1/2} \) ... discredits...

Diff: Can get difference of quantiles: "Bayes-confidence"! Bayes can't do it and see what is feasible

\[
\hat{\theta}_B - \hat{\theta}_y = \left\{ \frac{\sqrt{n}}{2n} + \frac{c}{2n} \right\} (y - z_\beta) + \frac{a}{n^{1/2}} + \frac{c}{2n} y
\]

Sum: No choice prior can give confidence. "Best" choice: flat at center ... i.e. \( a = 0 \) \( c = 0 \)

Proportion True: \( \text{Proportion}(\theta) = \beta - \frac{\sqrt{n}}{2n} \theta \phi(z_\beta) \)

= Plot
\[ \beta = 50\% \]

\[ \beta - \gamma/2n_0 \phi(z_0) \]

Above claimed, on either side of centre of curvature, no way that posterior can give confidence.
Above "claimed", on either side of centre of curvature.
Summary

- Likelihood with calibration (i.e. "Bayes")
  - a powerful fruitful way to explore
Summary

- Likelihood with calibration (i.e. Bayes)
  - a powerful fruitful way to explore and

- With linearity, can get confidence
Summary

- Likelihood with calibration (i.e. Bayes)
  - a powerful fruitful way to explore
  and
- With linearity, can give confidence
  but
- Without linearity, can be grossly misleading
  to claim otherwise is ...
Summary

- **Likelihood with calibration** (i.e. Bayes)
  - a powerful fruitful way to explore

and

- With linearity, can give confidence

but

- Without linearity, can be grossly misleading

to claim otherwise is ...

Recent:

Economist Aug 18 (2007) p69
Heinrich, J. (2006) PHYSTAT'05
Summary

- Likelihood with calibration (i.e. Bayes)
  - a powerful fruitful way to explore
and
- With linearity, can give confidence
but
- Without linearity, can be grossly misleading

to claim otherwise is...

Recent:

Economist Aug 18 (2007) p.69

Heinrich, J. (2006) PHYSTAT05

Exploration and maybe approx. confidence