Combining p-values from independent sources

D A S Fraser
A K Md E Saleh

U Toronto
Carleton U

Friday Sept 15 2006

International Conference on Nonparametric Statistics & related Topics

Carleton U Sept 15-17 2006
$$\frac{1}{100}$$  \hspace{1cm}  $$\frac{1}{120}$$  \hspace{1cm}  $$\frac{1}{10^2}$$  \hspace{1cm}  $$\frac{1}{5^2}$$  \hspace{1cm}  $$\frac{1}{5^2}$$  \hspace{1cm}  $$\frac{1}{10^2}$$

$$\left(\frac{1}{10^2} + \frac{1}{5^2}\right)^{-1} \left\{ \frac{1}{10^2} 100 + \frac{1}{5^2} 120 \right\}$$

$$= 116$$  \hspace{1cm}  $$4.47^2$$

- Combining
- Bayes
- Estimation
- Vector case  variance matrix
- Dependence  \( \text{var} \rightarrow \text{var-cov} \)
- Shrinkage
Overview

2. "Weight by reciprocal variance" (information)
4. What are p-values?
5. " Bayesian p-values?"
6. How to get p-values?
7. p-values from likelihood
8. How to combine? Fisher; likelihood
9. Two investigations
10. Reparameterization $\phi(\Theta)$
11. Combining formula
12. Simple $N$ example
13. Just LS calculations
14. Calculations in general
15. Overview
Position of data re theta... or vice versa

i.e. a \( p \)-value, a \( p \)-value function

What are they?

\[
f(y - \theta)
\]

\[
y^o \ \ f(y^o; \theta) \ \\
\ \\
F^o(\theta) \ \ L \ \\
F(\theta) \ \ p
\]

freq:

\[
p-value = p(\theta) = \int_{-\infty}^{y^o} f(y - \theta) \, dy = F(y^o; \theta) = F^o(\theta)
\]

= Statistical position of data \( y \) re \( \theta \) distribution

= % age position of \( y \) re \( \theta \)

one-sided
two-sided
dumb-sided

or... real: "Tell it as it is!"
Where is \( \theta \) re data?

\[ L(\theta; y) \]

Bayes \( p \)-values

\[ s(\theta) = \int \frac{L(\theta; y)}{\theta} d\theta = \int \frac{f(y - \theta)}{\theta} d\theta \]

= Statistical position of \( \theta \) re \( y \)

Get here...

\[ p(\theta) = s(\theta) \quad \text{--- Location model; just a rewrite of integral} \]

Contestation:

frequentist \( p \)-value = Bayesian \( p \)-value (\( s \)-value)

provided....

- you don't mumble-jumble the available info
- i.e. you don't go "proprietary"
How to get p-values

(a) Easy? Just \( F^*(\theta) = F(\hat{\theta}; \theta) \)

But what \( \Delta(y) \), what departure measure?
Say \( s = \hat{\theta} - \theta \)  
Bootstrap  
Maybe MCMC

But: mle departures ... can be unreliable!

(b) Use more ... typical continuity ... coordinate by coordinate
--- related likelihood theory

\[ \rightarrow \text{To 2nd/3rd order} \]

variable is location re interest \( \psi(\theta) \)

\( \psi(\theta) \) is location re data

(c) Get unique p-values... high accuracy!

Bhindra-Fisher --- 1929  
\( n = m = 2 \)  
\( t = \mu_1 - \mu_2 \)  
\( N = 100,000 \)

<table>
<thead>
<tr>
<th>Nominal conf. limits @ 5%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Confidence (sim)</td>
<td></td>
</tr>
<tr>
<td>Lik. ratio</td>
<td>13.2</td>
</tr>
<tr>
<td>Jeffreys</td>
<td>.7</td>
</tr>
<tr>
<td>Kim Ghosh 2nd</td>
<td>1.7</td>
</tr>
<tr>
<td>3rd p &amp; g</td>
<td>4.2</td>
</tr>
</tbody>
</table>

95% confidence limits: (4.8, 5.1)  
(94.9, 95.1)
\(p\)-values: Where from?

Need just: 1) log likelihood \(l(\theta)\)

2) Reparameterization \(\varphi(\theta)\)

\[
l^2/2 = l(\hat{\theta}) - l(\theta)
\]

Get

\[r = \pm \sqrt{2 \{l(\hat{\theta}) - l(\theta)\}}\]

\[q = (\hat{\varphi} - \varphi) \frac{\hat{r}}{\sqrt{\varphi}}\]

Gives

\[p(\theta) = \Phi (r - r' \log \frac{r}{q})\]

Vector case? Smooth \(\varphi(\theta)\) scalar

\(r, q\) by routine calculations

\(y^0 + \text{Model} \rightarrow \{l(\theta), \varphi(\theta)\} \rightarrow \{r(\hat{y}), q(\hat{y})\} \rightarrow p(\hat{y})\)

Based on: Continuity re \(\theta\) of dist. \(\text{fns: } F_1(y_1, \theta), \ldots\)

Behrens–Fisher
How to combine?

Fisher SMRW 1948 p100

\[ p_1 \quad \text{Calculate} \quad -2 \log (1-p_1) \]

\[ p_2 \quad " \quad -2 \log (1-p_2) \]

\[ \text{Add} \quad \frac{-2 \log (1-p_2)}{S} \]

\[ p = H_4(S) \]

- Pragmatic
- Takes no account of sensitivity
- Criticized by "Power" people

- Do better?
- Use: \( l(\theta) \rightarrow \phi(\theta) \rightarrow r(\psi) \rightarrow q(\psi) \rightarrow p(\psi) \)

Routine computation!

Q How to combine independent \( l_i(\theta) \) \( \phi_i(\theta) \) and solve the p-value enigma.
Combine: 2 investigations

\[ l_1(\theta) \quad q_1(\theta) \]
\[ l_2(\theta) \quad q_2(\theta) \]

Want: \( l(\theta) \quad q(\theta) \) for composite investigation

1. \( l(\theta) \) is immediate:
   \[ l(\theta) = l_1(\theta) + l_2(\theta) \]
   \text{---- additivity of likelihood}

2. \( q(\theta) \) ?

Where does info come from?

1. At data point
   Likelihood \( L(\theta) = f(y; \theta) \)
   All the Bayesians use Likelihood and some frequentists do!

2. Near data point
   Gotta be some minimum extra to look at!
   At least look locally
   \( 1st \ derivative \ at \ y^0 \)
   "Sensitivity"
   "Wiggle the data point!"
Reparameterization $\varphi(\theta)$

$$
\varphi(\theta) = \text{gradient of } \ell(\theta; y) \text{ at } y^0 \\
= \sum_{i=1}^{n} \frac{\partial}{\partial y_i} \ell(\theta; y_i) \bigg|_{y_i} \times \frac{dy_i}{d\theta} \bigg|_{(y_i^0, \hat{\theta}^0)} = \sum_{i=1}^{n} \varphi_i(\theta) \times n_i
$$

(a) How likelihood is affected by $i$th coord

(b) How $i$th coordinate is affected by $\theta$ change

\[\uparrow\]

Crucial to modern likelihood theory

Why?

Good theory!

Works 1) Behrens Fisher 1929

Decisive answer Simulations!

2) Box & Cox 1964

Any parameter... !

There is Theory!
a) For $i$th investigation: 

\[ l(\theta) \phi(\theta) \]

Nominal log-model: 
\[ l(\theta) + \Delta \phi(\theta) \]

\[ \left( \phi(\theta) = \frac{\partial \phi}{\partial \lambda} \bigg|_{\lambda=0} \right) \]

Scale re $\hat{\theta}$ at $\hat{\theta}^o$. Use \( \frac{\phi(\theta)-\phi(\hat{\theta}^o)}{\phi_{\theta}(\hat{\theta}^o)} \) for effective $\phi(\theta)$.

\[ \left( \frac{\partial \phi}{\partial \lambda} = \hat{j} \right) \]

b) How does $\theta$ affect $\hat{\theta}$ from $i$th investigation?

Revealed by location reparameterization $\beta(\theta)$

\[ d\beta(\theta) = \frac{l_{\theta}(\theta) - l_{\theta}(\hat{\theta}^o)}{\phi(\theta) - \phi(\hat{\theta}^o)} \ d\theta \]

and use above $\phi_i(\theta)$

\[ n_i = \frac{l_i^{\hat{\theta}}(\theta) - l_i^{\hat{\theta}}(\hat{\theta})}{\phi_i(\theta) - \phi_i(\hat{\theta})} \ 
\]

\[ \phi(\theta) = \sum \phi_i(\theta) \times n_i \]

\[ = \sum \left[ \phi_i(\theta) - \phi_i(\hat{\theta}^o) \right] \times \frac{l_i^{\hat{\theta}}(\theta) - l_i^{\hat{\theta}}(\hat{\theta})}{\phi_i(\theta) - \phi_i(\hat{\theta})} \]
Simple example

\[ y_1 \sim N(a_1 \theta, 1) \]
\[ y_2 \sim N(a_2 \theta, 1) \]

\[ y_i \]
\[ l_i(\theta) = -\frac{1}{2} (y_i - a_i \theta)^2 \]
\[ r_i = y_i - a_i \theta \]
\[ \phi_i(\theta) = a_i \theta - y_i \]
\[ \hat{\phi}_i = 0 \]
\[ \hat{\phi}_i = \frac{1}{2} (\phi - \hat{\phi}) = y_i - a_i \theta \]

\[ (y_1, y_2) \]
\[ l(\theta) = -\frac{1}{2} (y_1 - a_1 \theta)^2 - \frac{1}{2} (y_2 - a_2 \theta)^2 \]

Least squares

\[ \theta(\theta) = \text{linear in } \theta \ldots \text{ equivalent to } \theta \]
\begin{align*}
\text{Estimate} & \quad \text{variance} \\
\frac{y_1}{a_1} & \quad \frac{1}{a_1^2} \\
\frac{y_2}{a_2} & \quad \frac{1}{a_2^2} \\
(a_1^2 + a_2^2)^{-1} \left\{ \frac{a_2}{a_1} \frac{y_1}{a_1} + \frac{a_2}{a_2} \frac{y_2}{a_2} \right\} & \quad (a_1^2 + a_2^2)^{-1} \\
& \quad \frac{a_1 y_1 + a_2 y_2}{a_1^2 + a_2^2} - θ \\
\lambda & = \frac{a_1 y_1 + a_2 y_2}{(a_1^2 + a_2^2)^{1/2}} - (a_1^2 + a_2^2)^{1/2} θ \\
& \quad \text{... just LS estimate}
\end{align*}
\[ p(\theta) = \Phi \left( r - n^{\prime} \log \frac{r}{\delta} \right) \]

but \( n = q \) here

\[ = \Phi \left( r \right) \]

\[ = \Phi \left\{ \frac{a_1 y_1 + a_2 y_2}{(a_1^2 + a_2^2)^{1/2}} - \left( a_1^2 + a_2^2 \right)^{1/2} \Theta \right\} \]

In general: automatic

\[ \ell = \ell_1(\theta) + \ell_2(\theta) \]

\[ q = q_1(\theta) n_1 + q_2 n_2 \]

then \( r(\theta) \) and \( q(\theta) \) -- an algorithm
Overview

1) $p$ values come from likelihood reparameterization
- $p(\theta)$
- $\varphi(\theta)$
- Standard formulae give
  - $S L R \quad r(\theta)$
  - MLE dep $\varphi(\theta)$
  - and then
  - $p$-value $p(\theta)$

2) Independent data
- $l(\theta) = l_1(\theta) + l_2(\theta)$
- $\varphi(\theta) = \varphi_1(\theta) n_1 + \varphi_2(\theta) n_2$
  - $n_1, n_2$ from $l_i$ and $\varphi_i$

3) Calculate new $p(\theta)$

All routine 'calculus' of likelihood methodology