Objective Bayes Factors for Inequality Constrained Hypotheses

Herbert Hoijtink

Department of Methods and Statistics
University Utrecht, The Netherlands

All Models are Wrong, Groningen, March 2011
Two Traditional Hypotheses

- Nothing is Going On
  \[ H_0 : \theta_1 = \ldots = \theta_p \]

- Something is Going On But I don’t Know What
  \[ H_a : \text{not } H_0 \]
An Informative Hypothesis

Table: Mean Confidence Ratings for the Data of Hasel and Kassin (2009)

<table>
<thead>
<tr>
<th>Condition</th>
<th>Phase a</th>
<th>Phase b</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person Identified Confessed</td>
<td>5.95</td>
<td>8.34</td>
<td>41</td>
</tr>
<tr>
<td>All Suspects Denied</td>
<td>5.63</td>
<td>4.47</td>
<td>43</td>
</tr>
<tr>
<td>The Person Identified Denied</td>
<td>4.48</td>
<td>3.63</td>
<td>43</td>
</tr>
<tr>
<td>Another Person Confessed</td>
<td>5.61</td>
<td>2.65</td>
<td>46</td>
</tr>
</tbody>
</table>

Leading to the informative hypothesis:

\[ H_i : \]

\[
\begin{align*}
\theta_{1b} &> \theta_{1a} \\
\theta_{2b} &< \theta_{2a} \\
\theta_{3b} &< \theta_{3a} \\
\theta_{4b} &< \theta_{4a}
\end{align*}
\]

\[ \theta_{1b} > \theta_{2b} > \theta_{3b} > \theta_{4b} \]

and its complement

\[ H_c : \text{not } H_i \]
Another Informative Hypothesis

Table: Item Responses Presented by Bock and Lieberman (1970)

<table>
<thead>
<tr>
<th>Item Responses</th>
<th>Frequency</th>
<th>Item Responses</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000</td>
<td>12</td>
<td>10000</td>
<td>7</td>
</tr>
<tr>
<td>00001</td>
<td>19</td>
<td>10001</td>
<td>39</td>
</tr>
<tr>
<td>00010</td>
<td>1</td>
<td>10010</td>
<td>11</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>01101</td>
<td>23</td>
<td>11101</td>
<td>136</td>
</tr>
<tr>
<td>01110</td>
<td>8</td>
<td>11110</td>
<td>32</td>
</tr>
<tr>
<td>01111</td>
<td>28</td>
<td>11111</td>
<td>308</td>
</tr>
</tbody>
</table>

Table: A Latent Class Model

<table>
<thead>
<tr>
<th>Responses Class 1</th>
<th>Responses Class 2</th>
<th>Responses Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>12345</td>
<td>12345</td>
</tr>
<tr>
<td>00000</td>
<td>10101</td>
<td>11110</td>
</tr>
<tr>
<td>10000</td>
<td>11100</td>
<td>00111</td>
</tr>
<tr>
<td>11000</td>
<td>11001</td>
<td>11010</td>
</tr>
<tr>
<td>01000</td>
<td>00111</td>
<td>11011</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Leading to the informative hypothesis:

\[ H_i : \theta_{1j} < \theta_{2j} < \theta_{3j} \text{ for } j = 1, \ldots, 5, \]

and its complement

\[ H_c : \text{not } H_i \]

Universiteit Utrecht
A General Formulation of Informative Hypotheses

\[ H_i : \mathbf{R}\theta > 0 \]
\[ H_c : \text{not } H_i \]

\( \mathbf{R} \) is a \( K \times P \) matrix containing real numbers and \( \mathbf{0} \) denotes a vector of length \( K \) containing zero's.
Bayesian Evaluation of Informative Hypotheses

The Bayes Factor

\[
BF_{ia} = \frac{\int_{\theta, \phi} f(x | \theta, \phi) h(\theta, \phi | H_i) d\theta, \phi}{\int_{\theta, \phi} f(x | \theta, \phi) h(\theta, \phi | H_a) d\theta, \phi} = \frac{f(x | \theta, \phi) h(\theta, \phi | H_i)}{g(\theta, \phi | x, H_i)} \Big/ \frac{f(x | \theta, \phi) h(\theta, \phi | H_a)}{g(\theta, \phi | x, H_a)},
\]

- \(\theta\) denotes the parameters that are subjected to constraints
- \(\phi\) denotes the parameters that are not subjected to constraints
- \(H_a\) denotes an unconstrained hypothesis
- \(f(.)\) denotes the likelihood
- \(h(.)\) denotes the prior distribution
- \(g(.)\) denotes the posterior distribution
"If nothing was elicited to indicate that the two priors should be different, then it is sensible to specify [the prior of the inequality constrained hypothesis] to be, ..., as close as possible to [the prior of the alternative hypothesis]. In this way the resulting Bayes factor should be least influenced by dissimilarities between the two priors due to differences in the construction processes, and could thus more faithfully represent the strength of the support that the data lend to each [hypothesis]." (Leucari and Consonni, 2003)
\[ h(\theta, \phi \mid H_i) = \frac{h(\theta, \phi \mid H_a) l_{\theta \in H_i}}{\int_{\theta, \phi} h(\theta, \phi \mid H_a) l_{\theta \in H_i} d\theta, \phi} = \frac{1}{c_i} h(\theta, \phi \mid H_a) \]

Note that \( c_i \) is the proportion of the prior distribution of \( H_a \) in agreement with \( H_i \). Note also that in the sequel
\[ h(\theta, \phi \mid H_a) = h(\theta \mid H_a) h(\phi \mid H_a) \]
$H_i: \theta_1 > 0 \text{ and } \theta_2 > 0$

$H_c: \text{ not } H_i$
Bayesian Evaluation of Informative Hypotheses

The Posterior Distribution

\[
g(\theta, \phi \mid x, H_i) = \frac{g(\theta, \phi \mid x, H_a)I_{\theta \in H_i}}{\int_{\theta, \phi} g(\theta, \phi \mid x, H_a)l_{\theta \in H_i} d\theta, \phi} = \frac{1}{f_i} g(\theta, \phi \mid x, H_a)
\]

Note that \( f_i \) is the proportion of the posterior distribution of \( H_a \) in agreement with \( H_i \).
Bayesian Evaluation of Informative Hypotheses
A Simple Formula for the Bayes Factor

Substitution of the formulae for prior and posterior distribution in the formula of the Bayes factor renders:

\[ BF_{ia} = \frac{f_i}{c_i} \]

Since \( H_c \) is the complement of \( H_i \), the proportion of the prior distribution of \( H_a \) in agreement with \( H_c \) is \( 1 - c_i \). Similarly, the proportion of the posterior distribution of \( H_a \) in agreement with \( H_c \) is \( 1 - f_i \). Using this result it follows that

\[ BF_{ic} = \frac{BF_{ia}}{BF_{ca}} = \frac{f_i/c_i}{((1 - f_i)/(1 - c_i))} \]
Consider the hypothesis $\theta_1 > \theta_2 > \theta_3$.

There are in total $3! = 6$ hypotheses with an equivalent structure, e.g., $\theta_3 > \theta_1 > \theta_2$.

Each of these hypotheses is of the same complexity, that is, neither is more or less parsimonious or simple than another.

Conclusion, if the complexity of the total unconstrained parameter space is 1.0, the complexity of each of these hypotheses should be $1/6$. 
Consider the hypothesis $\theta_1 > \{\theta_2, \theta_3\}$.

There are in total 3 hypotheses with an equivalent structure. The other two are $\theta_2 > \{\theta_1, \theta_3\}$ and $\theta_3 > \{\theta_1, \theta_2\}$.

Each of these hypotheses is of the same complexity, that is, neither is more or less parsimonious or simple than another.

Conclusion, if the complexity of the total unconstrained parameter space is 1.0, the complexity of each of these hypotheses should be 1/3.
The prior distribution has to be chosen such that the complexity of an informative hypothesis is adequately reflected in the Bayes factor.

If this can be done such that the prior is objective, that is, determined without using the data or subjective input from the researcher, the ”subjectivity discussion” can be avoided.
Theorem 1 If \( h(\theta_p | H_a) = \mathcal{N}(\mu_0, \sigma_0^2) \) for \( p = 1, \ldots, P \), \( c_i \) is independent of \( \mu_0 \) for \( \sigma_0^2 \to \infty \).

Applied to the hypothesis \( \theta_1 > \theta_2 > \theta_3 \) it can be seen that

- if \( \theta_1, \theta_2 \) and \( \theta_3 \) are sampled from \( \mathcal{N}(\mu_0, \sigma_0^2) \) the probability that \( \theta_1 > \theta_2 > \theta_3 \) is \( 1/6 \), in agreement with the definitions given earlier.
- Similarly, the probability that \( \theta_1 > \{\theta_2, \theta_3\} \) is \( 1/3 \).
- this still holds for \( \sigma_0^2 \to \infty \).
- for \( \sigma_0^2 \to \infty \) it also holds that \( f_i \) is completely determined by the data.
Complexity

Theorem 1 does apply to the example from Hasel and Kassin (2009) and any matrix $R$ that is of full rank, but not to Bock and Lieberman (1970) because a normal prior distribution is not a good choice for parameters that are probabilities. However, the following theorem applies to the informative hypothesis formulated for the latent class example:

- **Theorem 2** If $h(\theta_p \mid H_a) = \text{Beta}(a, b)$ for $p = 1, \ldots, P$, $c_i$ is independent of the choice of $a$ and $b$.

Note that:

- the resulting $c_i$ measures model complexity according to the definitions given earlier.
- for $a = 1$ and $b = 1$ the element $f_i$ is completely determined by the data.
If the prior distribution is specified such that it reflects model complexity along the lines sketched above, the Bayes factor for the comparison of $H_i$ with $H_c$ is objective in the sense that there are default choices for the parameters of the prior distribution such that $c_i$ is a measure of model complexity, and $f_i$ is completely determined by the data.
Computation of $BF_{ic}$

- The software package WinBUGS http://www.mrc-bsu.cam.ac.uk/bugs/ can be used for the computation of $BF_{ic}$ for a wide variety of models and hypotheses. It is rather easy to instruct WinBUGS to render a sample from the prior distribution of $H_a$ and the corresponding posterior distribution of $H_a$. The proportion of the prior distribution in agreement with $H_i$ is an estimate of $c_i$ and the proportion of the posterior distribution in agreement with $H_i$ is an estimate of $f_i$.

- Note that models that are not identified like the latent class model (label switching) need special attention when determining whether or not a parameter vector sampled is in agreement with $H_i$. For the example at hand agreement is obtained if one of $3!$ ways in which three latent classes can be ordered renders a parameter vector in agreement with $H_i$. 
The density for the data of Hasel and Kassin (2009) is:

\[
\begin{bmatrix}
  x_{ija} \\
  x_{ijb}
\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}
  \theta_{ja} \\
  \theta_{jb}
\end{bmatrix}, \begin{bmatrix}
  \phi_{aa} & \phi_{ab} \\
  \phi_{ab} & \phi_{bb}
\end{bmatrix}\right),
\]

- \(x_{ija}\) denotes the response of person \(i\) in Condition \(j\) and Phase \(a\), \(\theta_{ja}\) denotes the means of Condition \(j\) in Phase \(a\), and \(\phi_{aa}\) the residual variance in Phase \(a\).
- \(h(\theta_{jb} \mid H_a) = h(\theta_{ja}) \sim \mathcal{N}(0, \sigma_0^2)\) for \(j = 1, \ldots, 4\) with \(\sigma_0^2 \to \infty\) and \(h(\phi \mid H_a) = \mathcal{W}^{-1}(1, \mathbf{I})\).
- a sample of \(T = 10,000\) from the posterior distribution rendered \(f_i = .950\). Using \(T = 10,000\) the estimate of \(c_i = .008\). The resulting \(\hat{BF}_{ic} = 2375\), that is, strong evidence in favor of \(H_i\).
The density for the data of Bock and Lieberman (1970) is:

\[
f(x \mid \theta, \phi) = \prod_{i=1}^{N} \sum_{p=1}^{P} \prod_{j=1}^{J} \theta_{pj}^{x_{ij}} (1 - \theta_{pj})^{1-x_{ij}} \phi_p
\]

- \(x_{ij} \in \{0, 1\}\) denotes the response of unit \(i = 1, \ldots, N\) to variable \(j\), \(\theta_{pj}\) denotes the probability of the response 1 to item \(j\) in class \(p\), and \(\phi_p\) denotes the proportion of units allocated to class \(p\).
- \(h(\theta_{pj} \mid H_a) \sim \text{Beta}(1, 1)\) for \(p = 1, \ldots, P\) and \(j = 1, \ldots, J\).
- \(h(\phi \mid H_a) \sim \mathcal{D}(1, \ldots, 1)\).
- A sample of \(T = 100,000\) from prior and posterior distribution renders \(c_i = .00085\) and \(f_i = .086\), respectively. The resulting \(\hat{BF}_{ic} = 111\), that is, strong evidence in favor of \(H_i\).
... of $H_0$ ...? In fact to paraphrase Richard Royal from his book Statistical Evidence, A Likelihood Paradigm, 1999, New York: Chapman & Hall/CRC: $H_0$ is wrong anyway, so why bother to collect data to falsify it. Or, as Jacob Cohen summarized this in the title of one of his papers: The Earth is Round, $p < .05$ (American Psychologist, 49, 997-1003).
Further information with respect to informative hypotheses can be found at http://tinyurl.com/informativehypotheses and http://tinyurl.com/hoijtink and h.hoijtink@uu.nl. This presentation was based on two papers that are currently under review: