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)

Condition	Phase *a*	Phase *b*	*N*
Person Identified Confessed	5.95	8.34	41
All Suspects Denied	5.63	4.47	43
The Person Identified Denied	4.48	3.63	43
Another Person Confessed	5.61	2.65	46

Hasel, L.E., Kassin, S.M. (2009). On the presumption of evidentiary independence. Can confessions corrupt eyewitness identifications? *Psychological Science, 20*, 122-126.

Leading to the informative hypothesis:

$$H_i: \quad \begin{array}{c} \theta_{1b} > \theta_{1a} \\ \theta_{2b} < \theta_{2a} \\ \theta_{3b} < \theta_{3a} \\ \theta_{4b} < \theta_{4a} \\ \theta_{1b} > \theta_{2b} > \theta_{3b} > \theta_{4b} \end{array}$$

and its complement

$$H_c : \text{not } H_i$$

Another Informative Hypothesis

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

Item Responses	Frequency	Item Responses	Frequency
00000	12	10000	7
00001	19	10001	39
00010	1	10010	11
	...	...	
01101	23	11101	136
01110	8	11110	32
01111	28	11111	308

Bock, R.D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35*, 179-197.

Table: A Latent Class Model

Responses Class 1	Responses Class 2	Responses Class 3
12345	12345	12345
00000	10101	11110
10000	11100	00111
11000	11001	11010
01000	00111	11011
...	...	...

Leading to the informative hypothesis:

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

and its complement

$$H_c : \text{not } H_i$$

A General Formulation of Informative Hypotheses

$$H_i : \mathbf{R}\theta > \mathbf{0}$$

$$H_c : \text{not } H_i$$

$\mathbf{R}$ 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

$$BF_{ia} = \frac{\int_{\boldsymbol{\theta},\phi} f(\mathbf{x} \mid \boldsymbol{\theta}, \phi) h(\boldsymbol{\theta}, \phi \mid H_i) d\boldsymbol{\theta}, \phi}{\int_{\boldsymbol{\theta},\phi} f(\mathbf{x} \mid \boldsymbol{\theta}, \phi) h(\boldsymbol{\theta}, \phi \mid H_a) d\boldsymbol{\theta}, \phi} =$$

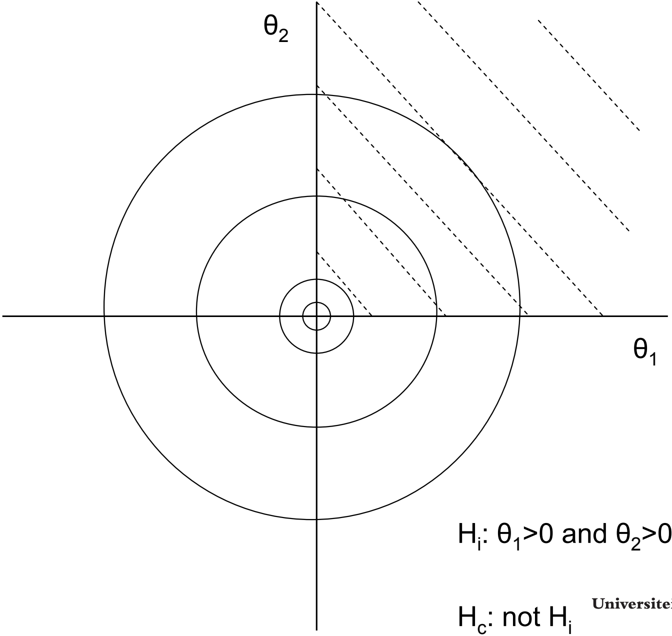
$$\frac{f(\mathbf{x} \mid \boldsymbol{\theta}, \phi) h(\boldsymbol{\theta}, \phi \mid H_i)}{g(\boldsymbol{\theta}, \phi \mid \mathbf{x}, H_i)} / \frac{f(\mathbf{x} \mid \boldsymbol{\theta}, \phi) h(\boldsymbol{\theta}, \phi \mid H_a)}{g(\boldsymbol{\theta}, \phi \mid \mathbf{x}, H_a)},$$

- $\boldsymbol{\theta}$ denotes the parameters that are subjected to constraints
- ϕ 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)

$$h(\boldsymbol{\theta}, \phi \mid H_i) = \frac{h(\boldsymbol{\theta}, \phi \mid H_a) I_{\boldsymbol{\theta} \in H_i}}{\int_{\boldsymbol{\theta}, \phi} h(\boldsymbol{\theta}, \phi \mid H_a) I_{\boldsymbol{\theta} \in H_i} d\boldsymbol{\theta}, \phi} = 1/c_i \; h(\boldsymbol{\theta}, \phi \mid H_a)$$

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 $h(\boldsymbol{\theta}, \phi \mid H_a) = h(\boldsymbol{\theta} \mid H_a) h(\phi \mid H_a)$



H_i: θ_1>0 and θ_2>0

H_c: not H_i

Bayesian Evaluation of Informative Hypotheses

The Posterior Distribution

$$g(\boldsymbol{\theta}, \phi \mid \mathbf{x}, H_i) = \frac{g(\boldsymbol{\theta}, \phi \mid \mathbf{x}, H_a) I_{\boldsymbol{\theta} \in H_i}}{\int_{\boldsymbol{\theta}, \phi} g(\boldsymbol{\theta}, \phi \mid \mathbf{x}, H_a) I_{\boldsymbol{\theta} \in H_i} d\boldsymbol{\theta}, \phi} = 1/f_i \, g(\boldsymbol{\theta}, \phi \mid \mathbf{x}, H_a)$$

Note that f_i is the proportion of the posterior distribution of H_a in agreement with H_i.

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

$$BF_{ia} = f_i / c_i$$

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

$$BF_{ic} = BF_{ia} / BF_{ca} = (f_i / c_i) / ((1 - f_i) / (1 - c_i))$$

Complexity

- 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 $1/6$.

Complexity

- 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 $1/3$.

Complexity

- 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 wihout using the data or subjective input from the researcher, the "subjectivity discussion" can be avoided.

Complexity

- **Theorem 1** If $h(\theta_p \mid H_a) = \mathcal{N}(\mu_0, \sigma_0^2)$ for $p = 1, \ldots, P$, c_i is independent of μ_0 for $\sigma_0^2 \to \infty$.

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

- if θ_1, θ_2 and θ_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.

Objective Bayes Factors

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.

Hasel and Kassin, 2009, continued

The density for the data of Hasel and Kassin (2009) is:

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

- x_{ija} denotes the response of person i in Condition j and Phase a, θ_{ja} denotes the means of Condition j in Phase a, and ϕ_{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 $h(\phi|H_a) = \mathcal{W}^{-1}(1, \mathbf{I})$
- 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 $B\hat{F}_{ic} = 2375$, that is, strong evidence in favor of H_i.

Bock and Lieberman, 1970, continued

The density for the data of Bock and Lieberman (1970) is:

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

- $x_{ij} \in \{0, 1\}$ denotes the response of unit $i = 1, \ldots, N$ to variable j, θ_{pj} denotes the probability of the response 1 to item j in class p, and ϕ_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 $j = 1, \ldots, J$
- $h(\phi|H_a) \sim \mathcal{D}(1, \ldots, 1)$
- A sample of $T = 100,000$ from prior and posterior distribution renders $c_i = .00085$ and $f_i = .086$, respectively. The resulting $B\hat{F}_{ic} = 111$, that is, strong evidence in favor of H_i.

The End

... 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

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:

- Hoijtink, H. (2010). Objective Bayes Factors for Inequality Constrained Hypotheses.
- Schoot, R. van de, and Hoijtink, H. (2010). Evaluation of Inequality Constrained Hypotheses using WinBUGS.