

## 22S:138 Bayes factors

Lecture 19  
Nov. 14, 2008

Kate Cowles  
374 SH, 335-0727  
kcowles@stat.uiowa.edu

## Bayes factors for model comparison and hypothesis testing

- simplest case: null and alternative hypotheses both simple
- equivalently: comparing two models that differ according to point values of one parameter

### Bayes' rule applied to the example (from lectures 1)

You take the blood test and the result is positive (+). This is the data or observation.

MODEL	Prior	Like for +	Product	Posterior
Have disease	.001	.95	.00095	.019
Don't have disease	.999	.05	.04995	.981
			.05090	1

Hypothesis-testing view:

$$H_0 : p = 0.05$$

$$H_A : p = 0.95$$

“simple hypotheses” regarding probability of positive test

Model	description	prior probability
$M_0$	don't have disease	.999
$M_1$	have disease	.001
Equivalently		
Hypothesis	description	prior probability
$H_0$	$p = .05$	.999
$H_1$	$p = .95$	.001

Prior odds in favor of Model 1 vs. Model 0:

$$\frac{Pr(M_1)}{Pr(M_0)} = \frac{.001}{.999} = \frac{1}{999}$$

Bayes factor in favor of Model 1 vs. Model 0

$$BF_{10} = \frac{Pr(\text{data}|M_1)}{Pr(\text{data}|M_0)} = \frac{.95}{.05} = 19$$

where “data” is positive test

## Bayes factor in simple/simple case

- $BF_{10}$  is weight of evidence contained *in the data* in favor of  $M_1$  vs  $M_0$
- usually reported on  $\log_{10}$  scale
- interpretation (Kass and Raftery, JASA, 1995)

$\log_{10}(B_{10})$	$B_{10}$	Evidence against $H_0$ (or $M_0$ )
0 to 1/2	1 to 3.2	Not worth more than bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

## Posterior probabilities and posterior odds

- posterior odds in example

$$\frac{Pr(M_1|data)}{Pr(M_0|data)} = \frac{.019}{.981} = .0194$$

- relationship among BF, posterior odds and prior odds in simple/simple case

$$BF_{10} = \frac{\frac{Pr(M_1|data)}{Pr(M_0|data)}}{\frac{Pr(M_1)}{Pr(M_0)}}$$

BF is ratio of posterior odds to prior odds

Before considering more general case, recall Bayes' rule:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}$$

Denominator is:

$$\begin{aligned} & \int p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta} \\ &= \int p(\boldsymbol{\theta}, \mathbf{y})d\boldsymbol{\theta} \\ &= p(\mathbf{y}) \end{aligned}$$

- the “marginal likelihood” of the data
- depends on
  - data
  - model (*form* of likelihood and prior)

## More general case

To compare two competing models,  $M_1$  and  $M_0$ :

- Compute the marginal likelihood of the data under each model
  - let  $\boldsymbol{\theta}_1$  = parameters under  $M_1$
  - let  $\boldsymbol{\theta}_0$  = parameters under  $M_0$

$$\begin{aligned} p(\mathbf{y}|M_1) &= \int p(\boldsymbol{\theta}_1)p(\mathbf{y}|\boldsymbol{\theta}_1)d\boldsymbol{\theta}_1 \\ p(\mathbf{y}|M_0) &= \int p(\boldsymbol{\theta}_0)p(\mathbf{y}|\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0 \\ BF_{10} &= \frac{p(\mathbf{y}|M_1)}{p(\mathbf{y}|M_0)} \end{aligned}$$

## More general case

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_1$$

Bayesian hypothesis test involves calculating posterior probabilities:

$$P(\Theta_0|\mathbf{y})$$

$$P(\Theta_1|\mathbf{y})$$

## Example continued

$$H_0 : \theta \leq 100$$

$$H_1 : \theta > 100$$

Prior probabilities and prior odds:

$$p(\theta) = N(100, 225)$$

$$\rightarrow Pr(\theta \leq 100) = .5$$

$$Pr(\theta > 100) = .5$$

$$\text{Prior odds} = \frac{0.5}{0.5} = 1$$

Posterior probabilities and odds:

$$Pr(\theta \leq 100|y) = .106$$

$$Pr(\theta > 100|y) = .894$$

## Example

- Child is given intelligence test, with resulting score  $Y$ .
- $Y \sim N(\theta, 100)$ 
  - where  $\theta$  represents child's own true IQ
  - 100 is variance if same child takes repeated IQ tests of the same kind

- in population as a whole, IQ scores are distributed as:

$$\theta \sim N(100, 225)$$

- if child scores  $y = 115$ , then posterior distribution of  $\theta$  is:

$$(\theta|y) \sim N(110.4, 69.2)$$

posterior odds

$$\frac{0.106}{0.894} = 0.119$$

Bayes factor in favor of  $H_0$  vs  $H_1$  is

$$BF_{01} = \frac{0.106}{\frac{0.894}{0.5}} = 0.119$$

Baey's factor in favor of  $H_1$  vs  $H_0$  is

$$BF_{10} = 8.44$$

## Bayesian hypothesis testing and frequentist p-values

- In one-sided testing situations like this, frequentist p-value will sometimes have a Bayesian justification.

- Example:

– normal likelihood, variance known

$$Y \sim N(\theta, \sigma^2)$$

– noninformative prior

$$p(\theta) \propto 1$$

– posterior

$$p(\theta|y) \sim N(y, \sigma^2)$$

- Hypotheses

$$H_0 : \theta \leq \theta_0$$

$$H_1 : \theta > \theta_0$$

- posterior probability of  $H_0$

$$Pr(\theta \leq \theta_0|y) = \Phi\left(\frac{\theta_0 - y}{\sigma}\right)$$

- classical p-value

$$pval = Pr(Y \geq y|\theta = \theta_0)$$

$$= 1 - \Phi\left(\frac{y - \theta_0}{\sigma}\right)$$

- By symmetry of normal distribution,  $Pr(\theta \leq \theta_0|y) = \text{p-value against } H_0$

## Testing a point null hypothesis

- common in frequentist practice

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

- where  $\theta$  could have any value on a continuum
- Bayesian answers may differ radically from frequentist answers
- almost never do we seriously consider that  $\theta = \theta_0$  exactly
- more reasonable:

$$H_0 : \theta \in (\theta_0 - b, \theta_0 + b)$$

for some small  $b$

“region of indifference”

## Consider Bayesian test of point null so as to compare with frequentist

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

- cannot use a continuous prior on  $\theta$   
Why?
- reasonable approach to constructing a prior  
– put positive prior *probability* on  $\theta_0$

$$Pr(\theta = \theta_0) = \pi_0 > 0$$

– give  $\{\theta : \theta \neq \theta_0\}$  the prior

$$(1 - \pi_0)g_1(\theta)$$

where

\*  $g_1$  is a proper density

## Bayesian analysis

- let  $f(\mathbf{y}|\theta)$  denote the sampling density of  $\mathbf{y}$
- then the marginal likelihood is

$$m(\mathbf{y}) = f(\mathbf{y}|\theta_0)\pi_0 + m_1(\mathbf{y})(1 - \pi_0)$$

where

$$m_1(\mathbf{y}) = \int_{\theta \neq \theta_0} f(\mathbf{y}|\theta)g_1(\theta)d\theta$$

- so posterior probability of  $H_0$  is

$$Pr(\theta = \theta_0|\mathbf{y}) = \frac{f(\mathbf{y}|\theta_0)\pi_0}{f(\mathbf{y}|\theta_0)\pi_0 + m_1(\mathbf{y})(1 - \pi_0)}$$

- posterior odds in favor of  $H_0$  vs  $H_1$

$$\frac{\pi_0}{1 - \pi_0} \frac{f(\mathbf{y}|\theta_0)}{m_1(\mathbf{y})}$$

- and the Bayes factor in favor of  $H_0$  vs  $H_1$  is

$$BF_{01} = \frac{f(\mathbf{y}|\theta_0)}{m_1(\mathbf{y})}$$

## Example: child's intelligence

- sampling distribution of data

$$f(y|\theta) = N(\theta, \sigma^2 = 100)$$

- hypotheses to be tested

$$H_0 : \theta = 100$$

$$H_1 : \theta \neq 100$$

- priors

$$Pr(\theta = 100) = \pi_0 = 0.5$$

$$g_1(\theta) = N(\mu_0, \sigma_0^2) = N(100, 100)$$

- note: prior mean  $\mu_0 = \theta_0$  (value from  $H_0$ )
- prior variance  $\sigma_0^2$  equals variance of sampling distribution

## Example continued

- What statistical test would a frequentist use when sampling distribution is assumed to be normal with known variance?
- results for frequentist test with different possible data values  $y$

		frequentist	
y	z	p-value	$Pr(H_0 y)$
116.45	1.645	0.1	0.42
119.60	1.960	0.05	0.35
125.76	2.576	0.01	0.21
132.91	3.291	0.001	0.086

## Similar table for different sample sizes

- Table 4.2, p. 151 from  
Berger, JO (1985) *Statistical Decision Theory and Bayesian Analysis, 2nd ed.*, New York, Springer-Verlag
- applies when  $\sigma^2$  is assumed known,  $\mu_0 = \theta_0$ ,  
 $\pi_0 = 0.5$ ,  $\sigma_0^2 = \sigma^2$