Prior Exposure 1: Bayes for beginners: **Bayes factors**

Thom Baguley and Mark Andrews
Nottingham Trent University

Reprise: An approximate Bayes factor

A difference in BIC between models on the same data can be expressed as a likelihood ratio:

$$e.g., \Delta_{BIC} = BIC_{M0} - BIC_{M1}$$

$$LR_{BIC} = e^{1/2\Delta_{BIC}}$$

This quantity is a well-known approximation to a Bayes factor with a *unit-information prior*:

$$BF \approx LR_{BIC} = e^{1/2\Delta_{BIC}}$$

Here the prior is meant to carry as much information as a single observation (Kass & Wasserman, 1995)

What is a Bayes factor?

Background

Bayes factors were developed (independently) by Harold Jeffries in the 1930s and Alan Turing in the 1940s (Good, 1979)

What is a Bayes factor?

Background

Bayes factors were developed (independently) by Harold Jeffries in the 1930s and Alan Turing in the 1940s (Good, 1979)

... though Jeffries may have been influenced by J. B. S. Haldane (Etz & Wagenmakers, 2015)

What is a Bayes factor?

Background

Bayes factors were developed (independently) by Harold Jeffries in the 1930s and Alan Turing in the 1940s (Good, 1979)

... though Jeffries may have been influenced by J. B. S. Haldane (Etz & Wagenmakers, 2015)

They are a Bayesian tool for *hypothesis testing* (Jeffries, 1961)

What is a Bayes factor?

Background

Bayes factors were developed (independently) by Harold Jeffries in the 1930s and Alan Turing in the 1940s (Good, 1979)

... though Jeffries may have been influenced by J. B. S. Haldane (Etz & Wagenmakers, 2015)

They are a Bayesian tool for *hypothesis testing* (Jeffries, 1961)

... consisting of a ratio of marginal likelihoods for two competing hypotheses

What is a Bayes factor?

Background

Bayes factors were developed (independently) by Harold Jeffries in the 1930s and Alan Turing in the 1940s (Good, 1979)

... though Jeffries may have been influenced by J. B. S. Haldane (Etz & Wagenmakers, 2015)

They are a Bayesian tool for *hypothesis testing* (Jeffries, 1961)

... consisting of a ratio of marginal likelihoods for two competing hypotheses

They are equal to likelihood ratios only for nested models with simple hypotheses and no nuisance parameters (Kass & Raftery, 1995)

What is a Bayes factor?

From likelihood to posterior probability

Recall that the likelihood of θ is proportional to the probability of the data given θ

$$\ell(\theta|\mathcal{D}) \propto P(\mathcal{D}|\theta)$$

What is a Bayes factor?

From likelihood to posterior probability

Recall that the likelihood of θ is proportional to the probability of the data given θ

$$\ell(\theta|\mathcal{D}) \propto P(\mathcal{D}|\theta)$$

We can extend this idea to the posterior probability of θ:

$$\overbrace{P(\theta|\mathcal{D})}^{\text{posterior}} \propto \overbrace{\ell(\theta|\mathcal{D})}^{\text{likelihood}}\overbrace{P(\theta)}^{\text{prior}}$$

What is a Bayes factor?

From likelihood to posterior probability

Recall that the likelihood of θ is proportional to the probability of the data given θ

$$\ell(\theta|\mathcal{D}) \propto P(\mathcal{D}|\theta)$$

We can extend this idea to the posterior probability of θ:

$$\overbrace{P(\theta|\mathcal{D})}^{\text{posterior}} \propto \overbrace{\ell(\theta|\mathcal{D})}^{\text{likelihood}}\overbrace{P(\theta)}^{\text{prior}}$$

... thus the posterior probability is proportional to the likelihood times the prior

What is a Bayes factor?

The odds ratio form of Bayes' theorem

A likelihood ratio expresses the relative odds of two hypotheses given the data at hand

What is a Bayes factor?

The odds ratio form of Bayes' theorem

A likelihood ratio expresses the relative odds of two hypotheses given the data at hand

The odds form of Bayes' theorem introduces combines the likelihood with the prior probability:

$$\frac{P\left(H_0|\mathcal{D}\right)}{P\left(H_1|\mathcal{D}\right)} = \frac{P\left(\mathcal{D}|H_0\right)}{P\left(\mathcal{D}|H_1\right)} \times \frac{P\left(H_0\right)}{P\left(H_1\right)}$$

What is a Bayes factor?

The odds ratio form of Bayes' theorem

A likelihood ratio expresses the relative odds of two hypotheses given the data at hand

The odds form of Bayes' theorem introduces combines the likelihood with the prior probability:

$$\overbrace{\frac{P\left(H_0 \middle| \mathcal{D}\right)}{P\left(H_1 \middle| \mathcal{D}\right)}}^{\text{posterior odds}} = \overbrace{\frac{P\left(\mathcal{D} \middle| H_0\right)}{P\left(\mathcal{D} \middle| H_1\right)}}^{\text{Bayes factor}} \times \overbrace{\frac{P\left(H_0\right)}{P\left(H_1\right)}}^{\text{prior odds}}$$

... so the Bayes factor is the degree to which the data change the prior odds to obtain the posterior odds (It can also be interpreted as the strength of evidence provided by the data in the context of the prior)

What is a Bayes factor?

The odds ratio form of Bayes' theorem

A likelihood ratio expresses the relative odds of two hypotheses given the data at hand

The odds form of Bayes' theorem introduces combines the likelihood with the prior probability:

$$\overbrace{\frac{P\left(H_0|\mathcal{D}\right)}{P\left(H_1|\mathcal{D}\right)}}^{\text{posterior odds}} = \overbrace{\frac{P\left(\mathcal{D}|\,H_0\right)}{P\left(\mathcal{D}|\,H_1\right)}}^{\text{Bayes factor}} \times \overbrace{\frac{P\left(H_0\right)}{P\left(H_1\right)}}^{\text{prior odds}}$$

... so the Bayes factor is the degree to which the data change the prior odds to obtain the posterior odds (It can also be interpreted as the strength of evidence provided by the data in the context of the prior)

Bayes factors in practice

Advantages

- provide a continous measure of strength of evidence

Bayes factors in practice

Advantages

- provide a continous measure of strength of evidence
- implement the *fully automatic Occam's razor*

Bayes factors in practice

Advantages

- provide a continous measure of strength of evidence
- implement the *fully automatic Occam's razor*
- not vulnerable to optional stopping

Bayes factors in practice

Advantages

- provide a continous measure of strength of evidence
- implement the *fully automatic Occam's razor*
- not vulnerable to optional stopping
- offer a 'simple' alternative to p values

Bayes factors in practice

Advantages

- provide a continous measure of strength of evidence
- implement the *fully automatic Occam's razor*
- not vulnerable to optional stopping
- offer a 'simple' alternative to p values
- it is possible to quantify support for the null hypothesis (not just against)

Bayes factors in practice

Advantages

- provide a continous measure of strength of evidence
- implement the *fully automatic Occam's razor*
- not vulnerable to optional stopping
- offer a 'simple' alternative to *p* values
- it is possible to quantify support for the null hypothesis (not just against)
- promising 'default' objective priors have been developed

Bayes factors in practice

Disadvantages

- can be difficult to compute

Bayes factors in practice

Disadvantages

- can be difficult to compute
- more sensitive to prior than full Bayesian analysis

Bayes factors in practice

Disadvantages

- can be difficult to compute
- more sensitive to prior than full Bayesian analysis
- less transparent than a full Bayesian analysis

Bayes factors in practice

Disadvantages

- can be difficult to compute
- more sensitive to prior than full Bayesian analysis
- less transparent than a full Bayesian analysis
- less informative than a full Bayesian analysis

The fully automatic Occam's razor

In any Bayesian analysis the posterior is a weighted average of prior and likelihood

The fully automatic Occam's razor

In any Bayesian analysis the posterior is a weighted average of prior and likelihood

If the prior is thinly spread over a parameter range, the weighted likelihood is also thinly spread

The fully automatic Occam's razor

In any Bayesian analysis the posterior is a weighted average of prior and likelihood

If the prior is thinly spread over a parameter range, the weighted likelihood is also thinly spread

This weakens the evidential support for any particular value

The fully automatic Occam's razor

In any Bayesian analysis the posterior is a weighted average of prior and likelihood

If the prior is thinly spread over a parameter range, the weighted likelihood is also thinly spread

This weakens the evidential support for any particular value

If the prior falls into a narrow range, then the likelihood is focussed around those parameter values

The fully automatic Occam's razor

In any Bayesian analysis the posterior is a weighted average of prior and likelihood

If the prior is thinly spread over a parameter range, the weighted likelihood is also thinly spread

This weakens the evidential support for any particular value

If the prior falls into a narrow range, then the likelihood is focussed around those parameter values

This increases the support those values

The fully automatic Occam's razor

In any Bayesian analysis the posterior is a weighted average of prior and likelihood

If the prior is thinly spread over a parameter range, the weighted likelihood is also thinly spread

This weakens the evidential support for any particular value

If the prior falls into a narrow range, then the likelihood is focussed around those parameter values

This increases the support those values

Unlike AIC or BIC there is no need to add a penalty for complex models or hypotheses

Interpreting a Bayes factor (BF)

Transforming a bayes factor

The BF is a *continuous* measure of evidence for H_0 over H_1

Interpreting a Bayes factor (BF)

Transforming a bayes factor

The BF is a *continuous* measure of evidence for H_0 over H_1

As an odds ratio, the Bayes factor for H_0 over H_1 is therefore:

$$BF_{10} = \frac{P\left(\mathcal{D} \middle| H_1\right)}{P\left(\mathcal{D} \middle| H_0\right)} = \frac{1}{BF_{01}}$$

Interpreting a Bayes factor (BF)

Transforming a bayes factor

The BF is a *continuous* measure of evidence for H_0 over H_1

As an odds ratio, the Bayes factor for H_0 over H_1 is therefore:

$$BF_{10} = \frac{P(\mathcal{D}|\, H_1)}{P(\mathcal{D}|\, H_0)} = \frac{1}{BF_{01}}$$

It can also be transformed onto a posterior probability scale:

Interpreting a Bayes factor (BF)

Transforming a bayes factor

The BF is a *continuous* measure of evidence for H_0 over H_1

As an odds ratio, the Bayes factor for H_0 over H_1 is therefore:

$$BF_{10} = \frac{P(\mathcal{D}|\, H_1)}{P(\mathcal{D}|\, H_0)} = \frac{1}{BF_{01}}$$

It can also be transformed onto a posterior probability scale:

$$Pr(H_0|\mathcal{D}) = \frac{BF_{01}}{1 + BF_{01}}$$

... assuming H_0 and H_1 are equally likely *a priori*

Interpreting a Bayes factor (BF)

Transforming a bayes factor

The BF is a *continuous* measure of evidence for H_0 over H_1

As an odds ratio, the Bayes factor for H_0 over H_1 is therefore:

$$BF_{10} = \frac{P\left(\mathcal{D}\middle|\, H_1\right)}{P\left(\mathcal{D}\middle|\, H_0\right)} = \frac{1}{BF_{01}}$$

It can also be transformed onto a posterior probability scale:

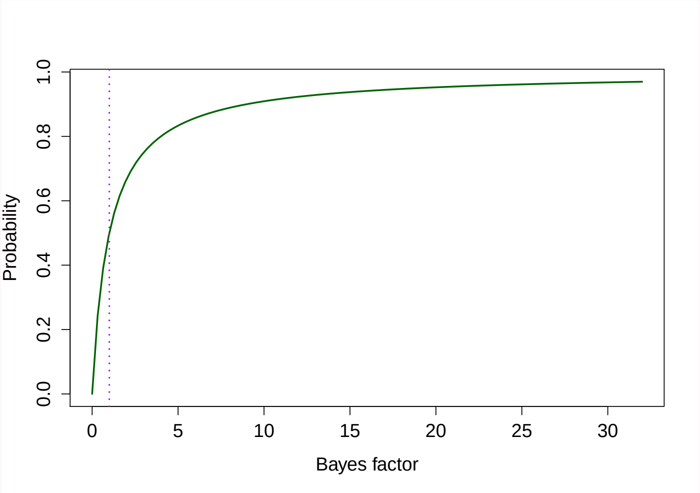
$$Pr\left(H_0\middle|\mathcal{D}\right) = \frac{BF_{01}}{1 + BF_{01}}$$

... assuming H_0 and H_1 are equally likely *a priori*

(It can also be useful to use the logarithm $ln(BF)$ when the Bayes factor is large)

Interpreting a Bayes factor (BF)

Transforming from an odds to a probability scale



Interpreting a Bayes factor (BF)

Jeffries (1961) proposed the following interpretation:

BF_{10}	Strength of evidence for H_1
< 1	Negative evidence
1 - 3	Barely worth mentioning
3 - 10	Substantial
10 - 30	Strong
30 - 100	Very strong
> 100	Decisive

Interpreting a Bayes factor (BF)

Jeffries (1961) proposed the following interpretation:

BF_{10}	Strength of evidence for H_1
< 1	Negative evidence
1 - 3	Barely worth mentioning
3 - 10	Substantial
10 - 30	Strong
30 - 100	Very strong
> 100	Decisive

Alternatively, for likelihood ratios Royall (1997) suggested:

< 8	Weak evidence
8 - 32	Moderate evidence
> 32	Strong evidence

Interpreting a Bayes factor (BF)

Jeffries (1961) proposed the following interpretation:

BF_{10}	Strength of evidence for H_1
< 1	Negative evidence
1 - 3	Barely worth mentioning
3 - 10	Substantial
10 - 30	Strong
30 - 100	Very strong
> 100	Decisive

Alternatively, for likelihood ratios Royall (1997) suggested:

< 8	Weak evidence
8 - 32	Moderate evidence
> 32	Strong evidence

... but remember that it is a continuous measure!

Objective Bayes factors with default priors

One appealing approach to is to use the prior to place a weak constraint on the standardised effect size δ

Objective Bayes factors with default priors

One appealing approach to is to use the prior to place a weak constraint on the standardised effect size δ

This builds in a general expectation of the plausible range of effect sizes a research might find

Objective Bayes factors with default priors

One appealing approach to is to use the prior to place a weak constraint on the standardised effect size δ

This builds in a general expectation of the plausible range of effect sizes a research might find

... it can also be scaled for a particular literature or discipline

Objective Bayes factors with default priors

One appealing approach to is to use the prior to place a weak constraint on the standardised effect size δ

This builds in a general expectation of the plausible range of effect sizes a research might find

... it can also be scaled for a particular literature or discipline

Two main types of prior on δ have been proposed:

- unit-information prior

(e.g., Kass & Raftery, 1995; Wagenmakers, 2007)

Objective Bayes factors with default priors

One appealing approach to is to use the prior to place a weak constraint on the standardised effect size δ

This builds in a general expectation of the plausible range of effect sizes a research might find

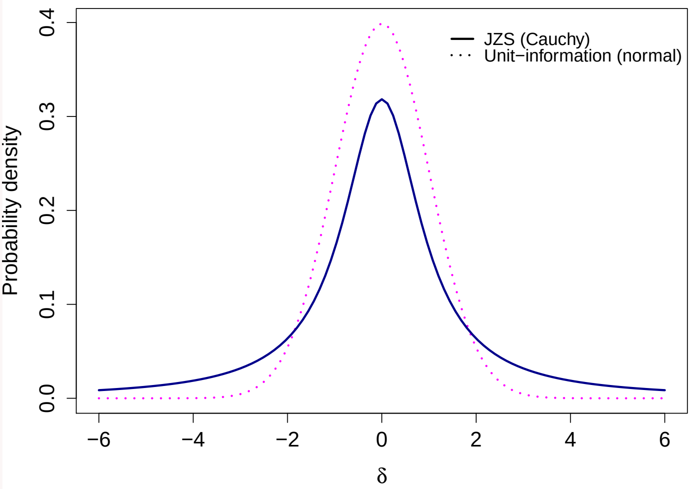
... it can also be scaled for a particular literature or discipline

Two main types of prior on δ have been proposed:

- unit-information prior
  (e.g., Kass & Raftery, 1995; Wagenmakers, 2007)
- Cauchy or JZS prior
  (e.g., Liang et al., 2008; Rouder et al., 2009)

Objective Bayes factors with default priors

Standard Cauchy versus standard normal priors



Bayesian t test with unit-information prior

Gönen et al. (2005) proposed a unit-information prior for a Bayesian t test

Bayesian t test with unit-information prior

Gönen et al. (2005) proposed a unit-information prior for a Bayesian t test

This uses the non-central distribution (easily obtained in packages such as R or SPSS)

Bayesian *t* test with unit-information prior

Gönen et al. (2005) proposed a unit-information prior for a Bayesian *t* test

This uses the non-central distribution (easily obtained in packages such as R or SPSS)

However, when H_0 is that $\delta = 0$ it can be calculated by hand (e.g., for one sample *t*):

$$BF_{01} = \frac{\left(1 + \frac{t^2}{n-1}\right)^{-n/2}}{(1 + nr^2)^{-1/2}\left(1 + \frac{t^2}{(1+nr^2)(n-1)}\right)^{-n/2}}$$

... here r is the scale factor for the variance (typically set at 1)

Bayesian t test with a Jeffreys-Zellner-Siow (JZS) prior

Rouder et al. (2009) argue that the unit-information prior typically provides too much information

Bayesian t test with a Jeffreys-Zellner-Siow (JZS) prior

Rouder et al. (2009) argue that the unit-information prior typically provides too much information

They propose a Cauchy (JZS) prior instead ...

Bayesian *t* test with a Jeffreys-Zellner-Siow (JZS) prior

Rouder et al. (2009) argue that the unit-information prior typically provides too much information

They propose a Cauchy (JZS) prior instead ...

The analytic solution for the one sample case is:

$$BF_{01} = \frac{\left(1 + \frac{t^2}{n-1}\right)^{-n/2}}{\int_0^\infty (1 + ngr^2)^{-1/2} \left(1 + \frac{t^2}{(1+ngr^2)(n-1)}\right)^{-n/2} (2\pi)^{-1/2} g^{-3/2} e^{-1/(2g)} \, dg}$$

Bayesian t test with a Jeffreys-Zellner-Siow (JZS) prior

Rouder et al. (2009) argue that the unit-information prior typically provides too much information

They propose a Cauchy (JZS) prior instead ...
The analytic solution for the one sample case is:

$$BF_{01} = \frac{\left(1+\frac{t^2}{n-1}\right)^{-n/2}}{\int_0^\infty (1+ngr^2)^{-1/2}\left(1+\frac{t^2}{(1+ngr^2)(n-1)}\right)^{-n/2}(2\pi)^{-1/2}g^{-3/2}e^{-1/(2g)}\,dg}$$

... and here r is the scale factor for the variance typically set at $1/2$ for a one sample or $\sqrt{2}/2$ for two samples

Can Bayes factors assess support for H_0?

Rouder et al. (2009) argue that they can assess support for a true null or an approximate one (a negligible effect)

Can Bayes factors assess support for H_0?

Rouder et al. (2009) argue that they can assess support for a true null or an approximate one (a negligible effect)

However, for small effects there is a tipping point at which large samples begin to favour H_1

Can Bayes factors assess support for H_0?

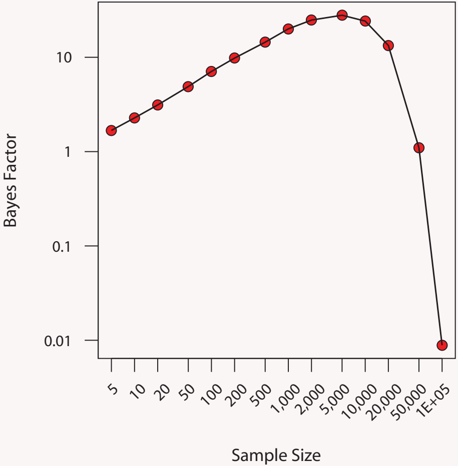
Rouder et al. (2009) argue that they can assess support for a true null or an approximate one (a negligible effect)

However, for small effects there is a tipping point at which large samples begin to favour H_1

This is around $n > 5{,}000$ for the unit-information prior or and $n > 50{,}000$ for the JZS prior

Can Bayes factors assess support for H_0?

Tipping point for the unit-information prior when $\delta = 0.02$



Reprise: Voting intention example

Does a single exposure to an US flag influence voting intention?

Carter et al. (2011) report a priming study in which a single exposure to a US flag changes voting intentions

They report an 'effect size' of over a quarter of a standard deviation:

$$\hat{\mu}_E - \hat{\mu}_C = 0.142, \quad t_{181} = 2.02, \quad p = .04$$

Reprise: Voting intention example

Does a single exposure to an US flag influence voting intention?

Carter et al. (2011) report a priming study in which a single exposure to a US flag changes voting intentions

They report an 'effect size' of over a quarter of a standard deviation:

$$\hat{\mu}_E - \hat{\mu}_C = 0.142, \quad t_{181} = 2.02, \quad p = .04$$

How do our objective Bayes factors work in this case?

Voting intention example

Unit-information prior Bayes factor

Baguley (2012) provides R functions for the unit-information prior Bayes factor:

Voting intention example

Unit-information prior Bayes factor

Baguley (2012) provides R functions for the unit-information prior Bayes factor:

```
> unit.prior.Bf.2s(t=2.02, n1=91.5, n2=91.5)
$`Bayes factor for H0`
[1] 0.9394962

$`Bayes factor for H1`
[1] 1.0644
```

Voting intention example

JZS Bayes factor

The BayesFactor package provides functions for JZS t tests (Morey & Rouder, 2013)

Voting intention example

JZS Bayes factor

The BayesFactor package provides functions for JZS t tests (Morey & Rouder, 2013)

These are designed for raw data but can be persuaded to work from summary statistics.

```
> log.bf <- ttest.tstat(t=2.02, n1=91.5, n2=91.5, rscale=sqrt(2)/2)
> exp(log.bf[['bf']])
[1] 1.058843
```

Here the Bayes factor for H_1 is only slight smaller than for the unit-uniformation prior

Voting intention example

JZS Bayes factor

The BayesFactor package provides functions for JZS t tests (Morey & Rouder, 2013)

These are designed for raw data but can be persuaded to work from summary statistics.

```
> log.bf <- ttest.tstat(t=2.02, n1=91.5, n2=91.5, rscale=sqrt(2)/2)
> exp(log.bf[['bf']])
[1] 1.058843
```

Here the Bayes factor for H_1 is only slight smaller than for the unit-uniformation prior

Baguley (2012) also provides R functions for these tests.

Constraints on the range of the effect

Directional tests

The BayesFactor package can also incorporate constraints

e.g., perhaps we are convinced that a US flag can only make people more right-wing:

Constraints on the range of the effect

Directional tests

The BayesFactor package can also incorporate constraints

e.g., perhaps we are convinced that a US flag can only make people more right-wing:

```
> log.bf <- ttest.tstat(t=2.02, n1=91.5, n2=91.5,
                        nullInterval=c(0, Inf))
> exp(log.bf[['bf']])
[1] 2.062125
```

Now H_1 is a little more strongly supported (but still not convincingly)

Constraints on the range of the effect

Directional tests

The BayesFactor package can also incorporate constraints

e.g., perhaps we are convinced that a US flag can only make people more right-wing:

```
> log.bf <- ttest.tstat(t=2.02, n1=91.5, n2=91.5,
                        nullInterval=c(0, Inf))
> exp(log.bf[['bf']])
[1] 2.062125
```

Now H_1 is a little more strongly supported (but still not convincingly)

... other constraints are possible