

Chapter 13: Bayes factors

Shravan Vasishth (vasishth.github.io)

July 2025

Contents

Textbook	1
Introduction	2
Discrete example of computing Bayes factor	4
Transitioning from discrete to continuous priors	5
The prior sensitivity of the Bayes factor	7
The Savage-Dickey method	8
Proof: Savage-Dickey density ratio method	9
Four ways to obtain approximate Bayes factors	10
Method 1: Computing Bayes factors with brms (bridgesam- pling)	10
Set up data	10
Define priors for full model	10
Some things to pay attention to	12
Method 2: Using bayestestR (bridgesampling)	12
Method 3: Bayes factor using the hypothesis function (brms)	13
Method 4: Using an approximation using the normal-normal conjugate case	14
Summary: Comparison of the BF estimates	16
Method 4: Practical example of using the normal-normal conjugate case analytically	17
There is no free lunch	18
Summary	18

Textbook

Introduction to Bayesian Data Analysis for Cognitive Science

Nicenboim, Schad, Vasishth

- Online version: <https://bruno.nicenboim.me/bayescogsci/>
- Source code: <https://github.com/bnicenboim/bayescogsci>

- Physical book: [here](#)

Introduction

Suppose we have two models M_1 and M_2 , and we want to know which model better accounts for the observed data D . The probability of each model given the data can be quantified via Bayes' rule:

$$P(M_1 | D) = \frac{P(D | M_1)P(M_1)}{\sum_{i=1}^2 P(D | M_i)P(M_i)} \quad (1)$$

$$P(M_2 | D) = \frac{P(D | M_2)P(M_2)}{\sum_{i=1}^2 P(D | M_i)P(M_i)} \quad (2)$$

Here, $P(M_1)$ and $P(M_2)$ are the a priori model probabilities.

If we take the ratio: $\frac{P(M_1|D)}{P(M_2|D)}$, the denominators cancel out, so we'd get:

$$\frac{P(M_1 | D)}{P(M_2 | D)} = \frac{\frac{P(D|M_1)P(M_1)}{\sum_{i=1}^2 P(D|M_i)P(M_i)}}{\frac{P(D|M_2)P(M_2)}{\sum_{i=1}^2 P(D|M_i)P(M_i)}} \quad (3)$$

$$\frac{P(M_1 | D)}{P(M_2 | D)} = \frac{P(D | M_1)P(M_1)}{P(D | M_2)P(M_2)} \quad (4)$$

Rewriting this:

$$\frac{P(M_1 | D)}{P(M_2 | D)} = \frac{P(D|M1) P(M_1)}{P(D|M2) P(M_2)} \quad (5)$$

So, for example, if the two models are a priori equally likely, the BF (the term $\frac{P(D|M1)}{P(D|M2)}$ below) tells you the posterior odds of M1 compared to M2.

$$\frac{P(M_1 | D)}{P(M_2 | D)} = \frac{P(D|M1) P(M_1)}{P(D|M2) P(M_2)} \quad (6)$$

$P(D | M_1)$ and $P(D | M_2)$ are the likelihood of the observed data D given the model M_1 (respectively M_2).

Obviously, you would prefer a model that gives a higher likelihood. For example, and speaking informally, if you have data that were generated from a $\text{Normal}(0,1)$ distribution, then the likelihood of the data with $\mu = 0$ will be higher than the likelihood with some other value like $\mu = 10$.

The higher likelihood is telling us that the underlying model is more likely to have produced the data. So we would prefer the model with the higher likelihood: we would prefer $\text{Normal}(0,1)$ over $\text{Normal}(10,1)$ as the presumed distribution that generated the data.

Assume for simplicity that $\sigma = 1$.

```
## sample 100 iid data points:
x<-rnorm(100)
## compute log likelihood under mu=0
(loglikmu0<-sum(dnorm(x,mean=0,sd=1,log=TRUE)))

## [1] -149.4709

## compute log likelihood under mu=10
(loglikmu10<-sum(dnorm(x,mean=10,sd=1,log=TRUE)))

## [1] -5221.701

## the likelihood ratio is a difference of logliks
## on the log scale:
loglikmu0-loglikmu10

## [1] 5072.23
```

So, one way to compare two models M_1 and M_2 is to compute the Bayes factor:

$$BF_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} \quad (7)$$

The Bayes factor is similar to the frequentist likelihood ratio test (or ANOVA), with the difference that in the Bayes factor, the likelihood is integrated over the parameter space, not maximized (shown below).

How to compute the likelihood? Consider the simple binomial case where we have a subject answer 10 questions, and they get 9 right. That's our data.

Discrete example of computing Bayes factor

Assuming a binomial likelihood function, $\text{Binomial}(n, \theta)$, the models we will compare pairwise (not all possible pairs) are

- M_1 , the parameter has a point value $\theta = 0.5$ with probability 1 (a very sharp prior).
- M_2 , the parameter has a discrete prior where $P(\theta = 0.1) = P(\theta = 0.5) = P(\theta = 0.9) = \frac{1}{3}$.
- M_3 , the parameter has a vague prior $\theta \sim \text{Beta}(1, 1)$. Recall that this $\text{Beta}(1, 1)$ distribution is $\text{Uniform}(0, 1)$.

The likelihood under M_1 is:

$$\binom{n}{k} \theta^k (1 - \theta)^{n-k} = \binom{10}{9} 0.5^{10} \quad (8)$$

We already know how to compute this:

```
(probDataM1<-dbinom(9,p=0.5,size=10))
```

```
## [1] 0.009765625
```

In Model 2, our θ can take on only three possible values: $\theta_1 = 0.1, \theta_2 = 0.5, \theta_3 = 0.9$, and each has probability $1/3$. Then, the marginal likelihood of the data given this prior specification of θ would be:

$$\begin{aligned} P(D \mid M_2) &= P(\theta_1)P(D \mid \theta_1) + P(\theta_2)P(D \mid \theta_2) + P(\theta_3)P(D \mid \theta_3) \\ &= \sum P(D \mid \theta_i, M_2)P(\theta_i \mid M_2) \end{aligned} \quad (9)$$

In our discrete example, this evaluates to:

```
(probDataM2<-(1/3)* (choose(10,9)* (0.1)^9 * (1-0.1)^1) + (1/3)*  
  (choose(10,9)* (0.5)^9 * (1-0.5)^1) +  
  (1/3)* (choose(10,9)* (0.9)^9 * (1-0.9)^1))
```

```
## [1] 0.1323954
```

This may be easier to read in mathematical form:

$$\begin{aligned}
P(D \mid M_2) &= P(\theta_1)P(D \mid \theta_1) + P(\theta_2)P(D \mid \theta_2) + P(\theta_3)P(D \mid \theta_3) \\
&= \frac{1}{3} \left(\binom{10}{9} 0.1^9 (1 - 0.1)^1 \right) + \frac{1}{3} \left(\binom{10}{9} 0.5^9 (1 - 0.5)^1 \right) \\
&\quad + \frac{1}{3} \left(\binom{10}{9} 0.9^9 (1 - 0.9)^1 \right) \quad (10) \\
&= 0.132
\end{aligned}$$

Essentially, we are computing the marginal likelihood $P(D \mid M_2)$ by averaging the likelihood across possible parameter values (here, only three possible values), with the prior probabilities for each parameter value serving as a weight.

The Bayes factor for Model 1 vs Model 2 would then be

```
(BF12 <- probDataM1 / probDataM2)
```

```
## [1] 0.07376107
```

Model 1, which assumes that θ has a point value 0.5, is less likely than the Model 2 with the discrete prior over θ ($\theta_1 = 0.1, \theta_2 = 0.5, \theta_3 = 0.9$, each with probability $1/3$). Model M2 is 14 times more likely than M1.

Transitioning from discrete to continuous priors

The marginal likelihood under M_3 involves solving the following integral:

$$P(D \mid M_3) = \int P(D \mid \theta, M_3) P(\theta \mid M_3) d\theta \quad (11)$$

The integral is simply integrating out (“summing over”) all possible values of the parameter θ .

The integral shown above is summing over the entire continuous space that is the range of possible values of θ :

$$P(D \mid M_3) = \int P(D \mid \theta, M_3) P(\theta \mid M_3) d\theta \quad (12)$$

Let's solve this integral analytically. We need to know only one small detail from integral calculus:

$$\int_a^b x^9 dx = \left[\frac{x^{10}}{10} \right]_a^b = \frac{b^{10}}{10} - \frac{a^{10}}{10} \quad (13)$$

Similarly:

$$\int_a^b x^{10} dx = \left[\frac{x^{11}}{11} \right]_a^b = \frac{b^{11}}{11} - \frac{a^{11}}{11} \quad (14)$$

Having reminded ourselves of how to solve this simple integral, we proceed as follows.

Our prior for θ in a new model M_3 is $Beta(\alpha = 1, \beta = 1)$:

$$\begin{aligned} P(\theta \mid M_3) &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} \theta^{\beta-1} \\ &= \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \theta^{1-1} \theta^{1-1} \\ &= 1 \end{aligned} \quad (15)$$

So, our integral simplifies to:

$$\begin{aligned} P(D \mid M_3) &= \int_0^1 P(D \mid \theta, M_3) d\theta \\ &= \int_0^1 \binom{10}{9} \theta^9 (1 - \theta)^1 d\theta \\ &= \int_0^1 \binom{10}{9} (\theta^9 - \theta^{10}) d\theta \\ &= 10 \left[\frac{\theta^{10}}{10} - \frac{\theta^{11}}{11} \right]_0^1 \\ &= 10 \times \frac{1}{110} = \frac{1}{11} \end{aligned} \quad (16)$$

So, when Model 1 assumes that the θ parameter is 0.5, and Model 3 has a vague prior $Beta(1, 1)$ on the θ parameter, our Bayes factor will be:

$$BF_{13} = \frac{P(D \mid M_1)}{P(D \mid M_3)} = \frac{0.00977}{1/11} = 0.107 \quad (17)$$

Thus, the model with the vague prior (M3) is about 9 times more likely than the model with $\theta = 0.5$:

$$\frac{1}{0.10742} = 9.309 \quad (18)$$

We could conclude that we have some evidence against the guessing model M1 in this case compared to model M2 and M3. @jeffreys1998theory has suggested the following decision criteria using Bayes factors. Here, we are comparing two models, labeled arbitrarily as 1 and 2 (but the label on the BF would be whatever the models' likelihoods are in the numerator and denominator. E.g., BF_{12} or BF_{21}).

- $BF_{12} > 100$: Decisive evidence
- $BF_{12} = 32 - 100$: Very strong
- $BF_{12} = 10 - 32$: Strong
- $BF_{12} = 3 - 10$: Substantial
- $BF_{12} = 2 - 3$: Not worth more than a bare mention

Do not interpret these as absolute divisions!

The prior sensitivity of the Bayes factor

The Bayes factor is sensitive to the choice of prior. It is therefore important to do a sensitivity analysis with different priors.

Imagine a fourth model, M_4 , with a prior on θ such that there are 10 possible values for θ , 0.1, 0.2, 0.3, ..., 1, and the probabilities of each value of θ are 1/10.

```
theta<-seq(0.1,1,by=0.1)
w<-rep(1/10,10)

prob<-rep(NA,length(w))
for(i in 1:length(theta)){
  prob[i]<-(w[i])*choose(10,9)*theta[i]^9*(1-theta[i]^1)
}
## Likelihood for model M4 with
```

```
## new prior on theta:  
(probDataM4<-sum(prob))
```

```
## [1] 0.08287079
```

Now the Bayes factor for M1 compared to M4 is:

```
probDataM1/probDataM4
```

```
## [1] 0.1178416
```

Now, model M4 is about 8.5 times more likely compared to model M1:

```
1/(probDataM1/probDataM4)
```

```
## [1] 8.485969
```

This toy example illustrates the effect of prior specification on the Bayes factor. It is therefore very important to display the Bayes factor under both uninformative and informative priors for the parameter that we are interested in.

One should never use a single ‘default’ prior or report a single Bayes factor.

The Savage-Dickey method

This method has the limitation that it only works for nested models: the null model has to be a proper subset of the full model. But it is relatively fast.

Example: Assume a baseline model M_0 that posits a point value (point null hypothesis) for the parameter of interest: $M_0 : \theta = \theta_0$.

The Savage-Dickey approximation is

$$\frac{P(D|M_0)}{P(D|M_1)} = \frac{P(\theta=\theta_0|D,M_1)}{P(\theta=\theta_0|M_1)}$$

In English, the BF is the ratio of the posterior density at θ_0 to the prior density in M_1 at θ_0 .

Intuitive idea: if the posterior density at θ_0 is lower than in the prior, then θ_0 is less likely given the data.

The proof is pretty surprising.

Proof: Savage-Dickey density ratio method

Assuming that the two models being compared are properly nested (one is a proper subset of the other), and assuming that the parameter of interest is θ , and there are other parameters δ that are common to both models, we will show that:

$$\frac{P(D|M_0)}{P(D|M_1)} = \frac{P(\theta=\theta_0|D,M_1)}{P(\theta=\theta_0|M_1)} = \frac{\text{Posterior density at the hypothesized value under M1}}{\text{Prior density at the hypothesized value under M1}}$$

The term $P(D|M_0)$ can be rewritten to integrate out the δ parameter(s):

$$P(D|M_0) = \int P(D|\delta, M_0)P(\delta|M_0) d\delta \quad (19)$$

Now, if M_0 and M_1 have the same structure when $\theta = \theta_0$, the RHS can be written as follows (we can replace M_0 with M_1):

$$\int P(D|\delta, M_0)P(\delta|M_0) d\delta = \int P(D|\delta, \theta = \theta_0, M_1)P(\delta|\theta = \theta_0, M_1) d\delta$$

But the RHS in the equation above is simply integrating out the δ parameter(s), so the RHS can be rewritten as:

$$\int P(D|\delta, \theta = \theta_0, M_1)P(\delta|\theta = \theta_0, M_1) d\delta = P(D|\theta = \theta_0, M_1)$$

Now recall Bayes' rule:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Setting $A = D$ and $B = (\theta = \theta_0)$, we get:

$$P(D|\theta = \theta_0) = \frac{P(\theta=\theta_0|D)P(D)}{P(\theta=\theta_0)}$$

The above can be conditioned on model M_1 everywhere.

$$P(D|\theta = \theta_0, M_1) = \frac{P(\theta=\theta_0|D, M_1)P(D|M_1)}{P(\theta=\theta_0|M_1)}$$

Going back to the original equation and plugging in this term in the numerator:

$$\frac{P(D|M_0)}{P(D|M_1)} = \frac{\frac{P(\theta=\theta_0|D, M_1)P(D|M_1)}{P(\theta=\theta_0|M_1)}}{P(D|M_1)} = \frac{P(\theta=\theta_0|D, M_1)}{P(\theta=\theta_0|M_1)}$$

What this means is that the Bayes factor is the ratio of the posterior to prior density at θ_0 .

Four ways to obtain approximate Bayes factors

- Method 1: `bayes_factor` (brms, bridgesampling)–time-consuming (and may not work for large data-sets)
- Method 2: `bayestestr` (bridgesampling)–as slow as `bayes_factor`
- Method 3: `hypothesis` (brms, Savage-Dickey ratio)–somewhat faster, but limited to point null hypotheses, plus the null model must be a proper subset of the full model
- Method 4: Using the Normal-Normal conjugate case (with Savage-Dickey): very fast approximation. I only use this when I have no choice

Method 1: Computing Bayes factors with brms (bridgesampling)

`brms` has a function for computing Bayes factors:

- `bayes_factor(m0,m1)`

`bridgesampling` is more flexible (models don't need to be nested), but more time-consuming and may not work well with large data-sets.

Set up data

```
library(bcogsci)
data("df_gg05_rc")
df_gg05_rc$so<-ifelse(df_gg05_rc$condition=="objgap",1/2,-1/2)
```

Define priors for full model

```
priors <- c(set_prior("normal(6, 0.6)", class = "Intercept"),
            set_prior("normal(0, 0.2)", class = "b",
                      coef = "so"),
            set_prior("normal(0, 0.2)", class = "sd"),
            set_prior("normal(0, 0.5)", class = "sigma"),
            set_prior("lkj(2)", class = "cor"))
```

```

brm1 <- brm(RT ~ so +
            (1+so|subj) + (1+so|item), df_gg05_rc,
            family=lognormal(), prior=priors,
            warmup=2000,
            iter=10000,
            chains = 4,
            cores=4,
            save_pars = save_pars(all = TRUE),
            control=list(adapt_delta=0.99,
                        max_treedepth=15))

priorsNULL <- c(set_prior("normal(6, 0.6)",
                        class = "Intercept"),
               set_prior("normal(0, 0.2)", class = "sd"),
               set_prior("normal(0, 0.5)", class = "sigma"),
               set_prior("lkj(2)", class = "cor"))

brm0 <- brm(RT ~ 1 +
            (1+so|subj) + (1+so|item), df_gg05_rc,
            family=lognormal(),
            prior=priorsNULL,
            warmup=2000,
            iter=10000,
            cores=4,
            save_pars = save_pars(all = TRUE),
            control=list(adapt_delta=0.99,
                        max_treedepth=15))

```

The evidence for m0 over m1:

```

bayes_factor(brm0,brm1)

```

```

## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5

```

```
## Estimated Bayes factor in favor of brm0 over brm1: 0.22421
```

Reverse the order of models to get evidence for m1 over m0:

```
bayes_factor(brm1,brm0)
```

```
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 6
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 6
```

```
## Estimated Bayes factor in favor of brm1 over brm0: 4.54188
```

Some things to pay attention to

- Run the command several times to check stability.
- Run a lot more iterations than the default values (10000 or even more).
- For large data-sets, you may have to simplify the model (e.g., varying intercepts only), which is obviously problematic (Barr et al 2013).

Method 2: Using bayestestR (bridgesampling)

For this analysis, I consulted: https://easystats.github.io/bayestestR/articles/bayes_factors.html

```
library(bayestestR)
```

Refit models with even more samples (recommended):

```
gg05m1 <- brm(RT ~ so + (1 + so | subj) +
  (1 + so | item), df_gg05_rc,
family = lognormal(), prior = priors,
warmup = 2000,
iter = 40000,
```

```
cores = 4,
save_pars = save_pars(all = TRUE),
control=list(adapt_delta=0.99, max_treedepth=15)
)
```

```
gg05m0 <- brm(RT ~ 1 + (1 + so | subj) + (1 + so | item),
              df_gg05_rc,
              family = lognormal(), prior = priorsNULL,
              warmup = 2000,
              iter = 40000,
              cores = 4,
              save_pars = save_pars(all = TRUE)
)
```

Bayes factor:

```
comparison <- bayesfactor_models(gg05m1, denominator = gg05m0)
```

```
## Computation of Marginal Likelihood: estimating marginal likelihood,
## please wait...
## Computation of Marginal Likelihood: estimating marginal likelihood,
## please wait...
```

```
comparison
```

```
## Bayes Factors for Model Comparison
```

```
##
##      Model                                BF
## [1] so + (1 + so | subj) + (1 + so | item) 4.44
##
## * Against Denominator: [2] 1 + (1 + so | subj) + (1 + so | item)
## * Bayes Factor Type: marginal likelihoods (bridgesampling)
```

Method 3: Bayes factor using the hypothesis function (brms)

This uses the Savage-Dickey density ratio method.

```
brm1 <- brm(RT ~ so +
            (1+so|subj) + (1+so|item), df_gg05_rc,
            family=lognormal(), prior=priors,
            warmup=2000,
            iter=10000,
            chains = 4,
            cores=4,
            save_pars = save_pars(all = TRUE),
            control=list(adapt_delta=0.99, max_treedepth=15),
            sample_prior="yes")
```

Evidence for OR-SR difference:

```
(h <- hypothesis(brm1, "so = 0"))

## Hypothesis Tests for class b:
##   Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star
## 1   (so) = 0      0.12      0.05      0.02      0.21      0.24      0.19   *
## ---
## 'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.
## '*': For one-sided hypotheses, the posterior probability exceeds 95%;
## for two-sided hypotheses, the value tested against lies outside the 95%-CI.
## Posterior probabilities of point hypotheses assume equal prior probabilities.
(BF10_SavageDickey <- 1 / h$hypothesis$Evid.Ratio)

## [1] 4.251215
```

Method 4: Using an approximation using the normal-normal conjugate case

A well-known result is that if the prior for the mean μ is defined in terms of a normal distribution, and the likelihood is normal, then the posterior for μ is normal. This is the Normal-Normal conjugate case.

- Suppose we have a sample mean \bar{x} and the standard error of the sample mean is SE .
- Suppose also that the prior's mean is μ_{prior} and the prior's standard deviation is σ_{prior} .

The posterior mean μ_{post} :

$$\mu_{post} = \frac{(w_1 \times \mu_{prior} + w_2 \times \bar{x})}{w_1 + w_2}$$

where $w_1 = \frac{1}{\sigma_{prior}^2}$, and $w_2 = \frac{1}{SE^2}$.

Recall the Poisson-Gamma conjugate case! The posterior mean is a compromise between the prior mean and the MLE.

Sample mean and SE (estimated using lmer so that SE includes all variance components of interest):

```
library(lme4)

m_lmer<-lmer(log(RT)~so + (1+so|subj) + (1+so|item),
             df_gg05_rc)

## boundary (singular) fit: see help('isSingular')
summary(m_lmer)$coefficients
```

```
##           Estimate Std. Error   t value
## (Intercept) 5.8830560 0.05202467 113.082051
## so          0.1240335 0.04932203   2.514768
```

```
xbar<-0.12403
se<- 0.04932
mu_prior<-0
prior_variance<- 0.2^2
w1<-1/prior_variance
w2<-1/se^2

mu_posterior<-(w1*mu_prior+w2*xbar)/(w1+w2)
mu_posterior
```

```
## [1] 0.1169199
```

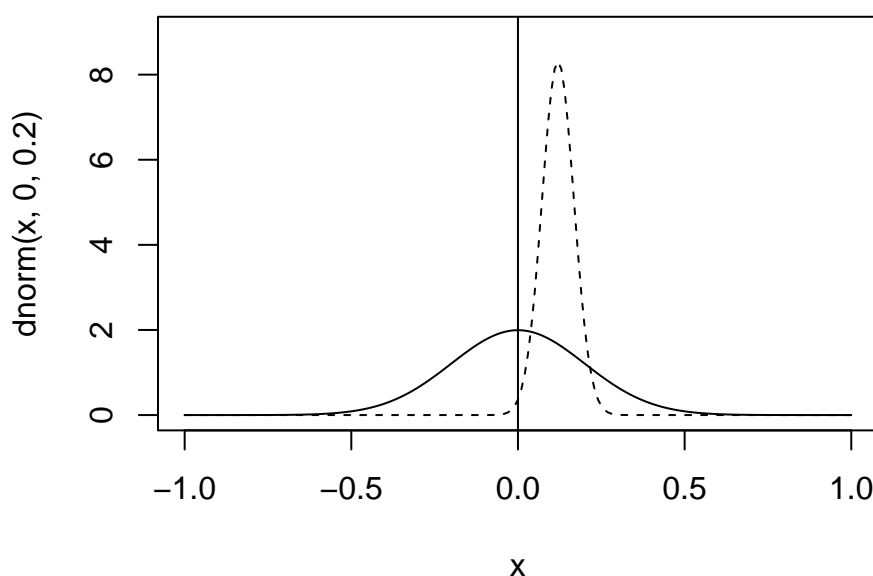
```
var_posterior<-1/(w1+w2)
sqrt(var_posterior)
```

```
## [1] 0.04788549
```

Posterior: Normal(0.12,0.048).

Next, compute the Savage-Dickey approximation by hand:

```
x<-seq(-1,1,by=0.01)
plot(x,dnorm(x,0,0.2),type="l",ylim=c(0,9))
lines(x,dnorm(x,0.12,0.048),lty=2)
abline(v=0)
```



We want to find out the density at 0 under the prior and compare it with the density at 0 under the pos-

terior. If the density under the posterior is lower than under the prior, we have evidence for the effect. The Savage-Dickey density ratio method quantifies the Bayes factor as the ratio of these two densities:

```
## density under posterior at 0:  
dnorm(0,0.12,0.048)
```

```
## [1] 0.3651729
```

```
## density under prior at 0:  
dnorm(0,0,0.2)
```

```
## [1] 1.994711
```

The evidence in favor of the full model is:

```
1/(dnorm(0,0.12,0.048)/dnorm(0,0,0.2))
```

```
## [1] 5.462375
```

With an uninformative prior:

```
1/(dnorm(0,0.12,0.048)/dnorm(0,0,1))
```

```
## [1] 1.092475
```

This approximation is smaller than the bridgesampling estimate in method 1.

Summary: Comparison of the BF estimates

- Method 1 (bayes_factor in brms, bridgesampling): $BF_{12} = 4.44$
- Method 2 (hypothesis in brms, Savage-Dickey): $BF_{12} = 3.8$
- Method 3 (bayestestr, bridgesampling): $BF_{12} = 4.5$
- Method 4 (Normal-Normal conjugate case + Savage-Dickey): $BF_{12} = 5.5$

The BFs all point to the full model being better than the model M_1 , but the numbers differ slightly due to the fact that these are all approximations.

Method 4: Practical example of using the normal-normal conjugate case analytically

Given: effect size -50 and CI [-115, 15] (from the published paper). That is all the information we have. No raw data available.

How to do a Bayes factor analysis?

Here, we are forced to work on the ms scale.

```
xbar<- -50
se<- (15+115)/4

mu_prior<-0
prior_variance<-50^2
w1<-1/prior_variance
w2<-1/se^2

mu_posterior<-(w1*mu_prior+w2*xbar)/(w1+w2)
mu_posterior

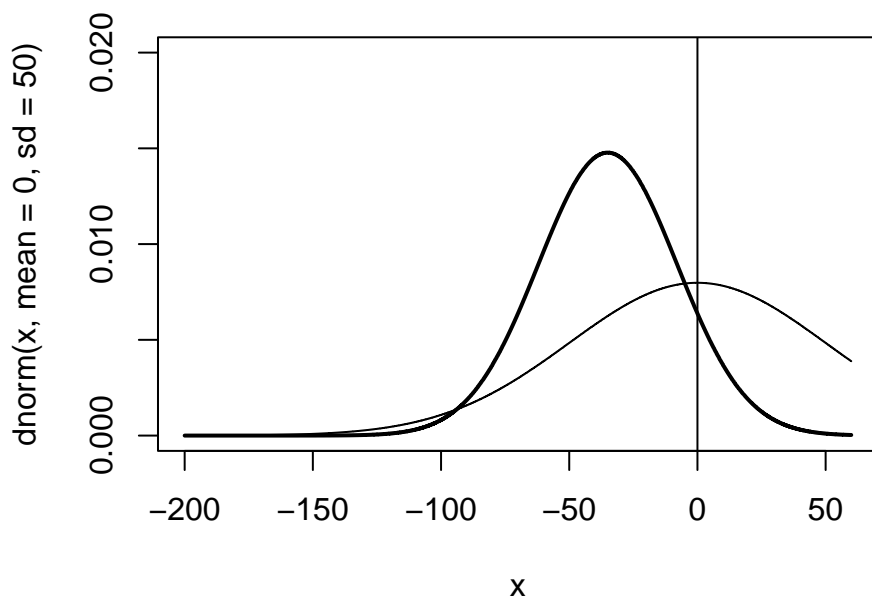
## [1] -35.14938

var_posterior<-1/(w1+w2)
sqrt(var_posterior)

## [1] 27.24942
```

Posterior: Normal(-35,27). Graphically:

```
x<-seq(-200,60,by=0.01)
plot(x,dnorm(x,mean=0,sd=50),type="l",ylim=c(0,0.02))
lines(x,dnorm(x,mean=-35,sd=27),lwd=2)
abline(v=0)
```



```
## density under posterior at 0:
```

```
dnorm(0,-35,27)
```

```
## [1] 0.006377574
```

```
## density under prior at 0:
```

```
dnorm(0,0,50)
```

```
## [1] 0.007978846
```

The evidence in favor of the full model is:

```
1/(dnorm(0,-35,27)/dnorm(0,0,50))
```

```
## [1] 1.251078
```

There is no free lunch

- Just because one gets a large Bayes factor does not entail that we made a discovery claim.
- The same problems exist with BFs as with p-values: inflation of BFs due to Type M error.
- The most convincing thing you can do to establish the robustness of effects is: run the highest power studies you can, and replicate, replicate, replicate.

Summary

- The Bayes factor is sensitive to the prior specification

- Therefore, always report a sensitivity analysis (not usually done)
- For nested comparisons, Savage-Dickey is good enough
- For more general model comparisons, you may need to use bridgesampling