

Corrected ABF

Karl Tayeb, Matthew Stephens

Introduction

The Bayes factor (BF) is a ratio of the marginal likelihoods of two competing hypotheses, usually denoted as H_0 and H_1 . Given a set of observed data, the Bayes factor measures the extent to which the data supports one hypothesis compared to the other, taking into account the prior probabilities of the hypotheses. A BF greater than 1 indicates that the data supports H_1 over H_0 , whereas a BF less than 1 implies the opposite. A BF close to 1 suggests that the data provides no strong evidence in favor of either hypothesis. Bayes factors are particularly valuable because they do not require specifying a precise value for the alternative hypothesis, which makes them more flexible than classical hypothesis tests.

A simple but widely applicable case is where we are interested in testing for the effect of a single covariate. The null hypothesis is that there is no effect. The alternative hypothesis is that there is a non-zero effect. Computing the Bayes factor comparing these two models involves specifying a prior over effect sizes, and marginalizing over possible effects. This can be done exactly when we put a normal prior on the effect, and the covariate is related to the response via a linear model with normal error model (assuming variance is known).

“The more widespread use of the Bayes factor has been hampered by the need for prior distributions to be specified for all of the unknown parameters in the model, and the need to evaluate multidimensional integrals, a complex computational task.”

Asymptotic approximations to the Bayes Factor

Computing the Bayes Factor

Computing the BF involves computing the following integral for each model M

$$I_M = \int Pr(\mathcal{D}|\theta, M)Pr(\theta|M)d\theta$$

I_M can be computed analytically only in special cases e.g. exponential family models with conjugate priors. The normal-linear model with normal prior is a particular important special case.

Laplace approximation

For some non-negative function f we want to approximate the integral

$$I = \int f(x) dx$$

if f is sufficiently peaked around it's maximum, the integral I can be well approximated by a local approximation of f around it's maximum. We can choose that local approximation to be easy to integrate. Specifically we can define $h(x) = \log f(x)$ and take a second order Taylor approximation of h around it's maximum x^* . $h(x) \approx h(x^*) + h''(x^*)(x - x^*)^2$. x^* is a maximum so the first order term vanishes and $h''(x^*) < 0$. Substituting the approximation we get a Gaussian integral approximating the original integral, known as the *Laplace approximation*.

$$\hat{I} = \int \exp\{h(x^*) + h''(x^*)(x - x^*)^2\} dx = \exp h(x^*) \left(-\frac{2\pi}{h''(x^*)} \right)^{1/2}$$

We will also refer to $f(x) \approx \exp\{h(x^*) + h''(x^*)(x - x^*)^2\}$ as a Laplace approximation to f , since substituting f with it's approximation in the integral results in the Laplace approximation for \hat{I} .

Kass and Raftery propose using a Laplace approximation to approximate BFs. This involves approximating the joint log-likelihood with a quadratic approximation about the MAP. Another reasonable approximation that avoids specification of the prior is to approximate the log-likelihood with a quadratic approximation about the maximum likelihood estimate $\hat{\beta}_{MLE}$. This approximation has the advantage of requiring only the outputs of standard statistical software.

Under certain conditions, the Laplace approximation has relative error $O(n^{-1})$, so it can be applied to the numerator and the denominator of the BF to approximate the BF with relative error $O(n^{-1})$. Note in the case we are concerned with, the null hypothesis is *simple*, the denominator of the BF is simply the likelihood of the data under $\beta = 0$. No integration or approximation is needed.

Wakefield's ABF

The approximate Bayes Factor (ABF) is a particularly simple approximation to the BF is introduced by Jon Wakefield. The ABF leverages asymptotic normality of the maximum likelihood estimates to approximate the BF.

The ABF approximates $I_{\beta=0}$ and $I_{\beta \neq 0}$ using the normal approximation for the likelihood term centered around the MLE $\hat{\beta} \sim N(\beta, s^2)$. With the prior $\beta \sim N(0, \sigma^2)$ we get $I_{\beta \neq 0} = N(\hat{\beta}; 0, \sigma^2 + s^2)$ and $I_{\beta=0} = N(\hat{\beta}; 0, s^2)$.

Another way to think of the ABF is to think of it as a particular approximation of the likelihood ratio, integrated over a normal prior on the effect. Start by noting that when the BF is comparing against a simple null hypothesis, the BF can be thought of as marginalizing over the likelihood ratio.

$$BF = \frac{Pr(y|\beta \neq 0)}{Pr(y|\beta = 0)} = \int \frac{Pr(y|\beta)}{Pr(y|\beta = 0)} Pr(\beta) d\beta$$

For large n , $\hat{\beta}$ is approximately sufficient, that is y is approximately conditionally independent of β given $\hat{\beta}$.

$$Pr(y|\beta) = Pr(y, \hat{\beta}|\beta) = Pr(y|\hat{\beta}, \beta) Pr(\hat{\beta}|\beta) \approx Pr(y|\hat{\beta}) N(\hat{\beta}; \beta, s^2)$$

We can then approximate the likelihood ratio as

$$\widehat{LR}_0(\beta) = \frac{N(\hat{\beta}; \beta, s^2)}{N(\hat{\beta}; 0, s^2)}$$

Where the $p(y|\hat{\beta})$ terms cancel. We get the ABF by substituting this approximation and marginalizing over β with a normal prior.

$$ABF = \frac{N(\hat{\beta}|0, s^2 + \sigma^2)}{N(\hat{\beta}|0, s^2)} = \int \widehat{LR}_0(\beta) \times N(\beta|0, \sigma^2) d\beta$$

The ABF uses the normal approximate to approximate the marginal likelihood under both the alternative and null model. This choice is attractive because $\widehat{LR}_0(0) = 1$, and computing the ABF requires only the effect estimate and it's standard error. Wakefield notes that the ABF will result in a consistent *decision* about a single hypothesis, “ABF is consistent under both the null and the alternative—the correct model is chosen with probability 1 as the sample sizes increase.”. However, as we will discuss, the ABF is *not* a consistent estimator of the BF under the alternative hypothesis— the approximation error can grow as evidence for the alternative hypothesis increases.

An issue with Wakefield's approximation

We observe that when the z score is very large the ABF can significantly underestimate the BF. While the normal approximation is good around the MLE, the normal approximation can be quite poor in the tail. If the z -score is very large ($\hat{\beta} \gg s^2$) $N(\hat{\beta}|0, s^2)$ is a poor approximation for $p(\hat{\beta}|\beta = 0)$.

The ABF uses a quadratic approximation of the log likelihood ratio which attains its maximum at the MLE, with a curvature equal to the curvature of the exact log likelihood ratio at the MLE. However, the likelihood ratio approximation is made to equal 1 and $\beta = 0$. To most accurately approximate the BF, we want our likelihood ratio approximation to be good in the regions that contribute most to the integral. If the normal approximation is bad at $\beta = 0$, our approximation of the likelihood ratio can be quite poor. Requiring our approximate LR to agree with the exact LR at $\beta = 0$ forces us to incur a large error in the regions that contribute most to the integral.

An alternative to the likelihood ratio approximation implied by the ABF, we could simply take a Laplace approximation to the exact likelihood ratio. This amounts to using the same approximation of the likelihood ratio used in the ABF and rescaling it (or translating on the log scale) so the the approximate LR and the exact LR agree at $\beta = \hat{\beta}_{MLE}$ rather than at $\beta = 0$.

When the alternative is true, we will get a better approximation of the BF compared to the ABF (is the Laplace approximation the “best” in some sense?). When the null is true, $\hat{\beta} \rightarrow 0$ and the quadratic approximation becomes very good near $\beta = 0$.

One view is that both ABF and our proposed approximate BF both use a Laplace approximation to the numerator of the BF. The ABF *uses the same approximation of the likelihood to approximate the likelihood under the null model*. This has the nice feature that $\widehat{LR}(0) = 1$, but it comes at the expense of \widehat{LR} having potentially large error around $\hat{\beta}_{MLE}$ —the region that contributes most to the integral. In contrast, we propose using the exact likelihood in the denominator. This makes our approximate likelihood ratio most accurate around $\hat{\beta}_{MLE}$, and results in a better estimate of the BF.

Computing the LABF

We can relate the Laplace approximation for the likelihood ratio to the approximation implied by the ABF, \widehat{LR} .

$$LR(a, c) = LR(a, b) \times LR(b, c)$$

$$\begin{aligned} LR_0(\beta) &= LR(\beta, \hat{\beta}_{MLE}) \times LR(\hat{\beta}_{MLE}, 0) \\ &\approx \widehat{LR}(\beta, \hat{\beta}_{MLE}) \times LR(\hat{\beta}_{MLE}, 0) := \tilde{LR}_0(\beta) \end{aligned}$$

For β close to $\hat{\beta}_{MLE}$, $\tilde{LR}(\beta, \hat{\beta}_{MLE})$ is a good approximation, and when $\beta = \hat{\beta}_{MLE}$, the approximate and exact likelihood ratios agree.

We can write our new approximate Bayes Factor, the Laplace approximate Bayes Factor (LABF) in terms of the ABF:

$$\begin{aligned} LABF &= \int \tilde{LR}_0(\beta) N(\beta, 0, \sigma^2) d\beta, \\ &= LR(\hat{\beta}, 0) \widehat{LR}(0, \hat{\beta}) \int \widehat{LR}(\beta, 0) N(\beta|0, \sigma^2) d\beta, \\ &= LR(\hat{\beta}, 0) \times \widehat{LR}(0, \hat{\beta}) \times ABF. \end{aligned}$$

$LR(\hat{\beta}, 0)$ is the likelihood ratio between the MLE and the null model, which is commonly reported by statistical software. $\widehat{LR}(0, \hat{\beta}) \times ABF < 1$ is the asymptotic approximation to the BF comparing the alternative model against the MLE.

Example: Single Effect Regression

Here we demonstrate the impact of using ABF vs LABF in the context of variable selection.

Perhaps it is not surprising that the ABF does not work well in the SER. The denominator of all the BFs should be the same— but the use of ABF implies using a different approximation of the denominator for each BF. We see that this can lead to reordering of the ABFs compared to the exact BFs.

As Matthew notes, for increasingly correlated variables the ratio of ABF_i/ABF_j should give better estimates for BF_{ij} even if ABF_i and ABF_j are off. However, it is not hard to find examples where the ordering of ABFs differ from the ordering of BFs which is

Discussion

We describe a correction to Wakefield's ABF, which really amounts to taking a Laplace approximation of the likelihood ratio, rather than approximating the numerator and denominator of the Bayes factor with the asymptotic approximation. Since we are computing the BF against a simple null hypothesis, this is also equivalent to using a Laplace approximation of the likelihood which is an old idea presented in Kass and Rafteries work popularizing Bayes Factor. The laplace approximation and the asymptotic approximation are closely related, since they are rescaled versions of each other. The main difference is that the asymptotic approximation

to the likelihood ratio is made exact at $\beta = 0$, whereas the Laplace approximation is made exact at $\beta = \hat{\beta}_{MLE}$. Since the likelihood ratio near the MLE contributes most to the integral involved in approximating the BF—making the latter choice results in a better approximation of the BF, particularly when the BF is large.

Compared to the ABF, the LABF requires the likelihood ratio of the MLE and null model. This information wins substantial improvements in the approximate BF. This should not be surprising, when the likelihood surface is very peaked the likelihood ratio is *almost* the Bayes Factor (in the sense that, the integral is dominated by the likelihood ratio near this point). Having accurate approximations of the BF are important when we are making comparisons between many BFs (e.g. when performing variable selection).