Gamma function approximations for computing closed-form Bayes factors

Thomas J. Faulkenberry

Tarleton State University

MAA-TX 2023

Background – common goal in applied settings is to compare means from treatment groups.

- Hypothesis test: compare competing models of data $\boldsymbol{y}$

$$\mathcal{H}_0 : \mu_1 = \mu_2$$
$$\mathcal{H}_1 : \mu_1 \neq \mu_2$$

- How to compare? – Bayes factor (Kass & Raftery, 1995)

$$\text{BF}_{10} = \frac{P(\boldsymbol{y} \mid \mathcal{H}_1)}{P(\boldsymbol{y} \mid \mathcal{H}_0)}$$

Problem – while conceptually simple, Bayes factor is hard to compute. For each model $\mathcal{H}_i$, must compute marginal likelihood:

$$P(\boldsymbol{y} \mid \mathcal{H}_i) = \int P(\boldsymbol{y} \mid \boldsymbol{\theta}, \mathcal{H}_i)\pi(\boldsymbol{\theta}, \mathcal{H}_i)d\boldsymbol{\theta}$$

Difficulties:

- must assign prior distribution to model parameters $\boldsymbol{\theta}$
- resulting marginal likelihood usually involves integral representation

A solution (Wang & Liu, 2016; Faulkenberry, 2020):

With a specific choice of prior, the Bayes factor has an analytic representation:

$$\mathsf{BF}_{10} = \frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}+\frac{1}{2}\right)\sqrt{\pi}}\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu-1}{2}}$$

But – could we get rid of the need for computing the gamma function?

That is, can we approximate

$$C_\nu = \frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2} + \frac{1}{2}\right)} \ ?$$

As it happens, this problem has had some interest since the 1940s!

(see Borwein & Corless (2018) – American Mathematical Monthly)

Today, I'll describe three approaches to approximating C_ν:

1. using Stirling's formula for gamma function
2. using classical asymptotic formula of Wendel (1948)
3. using improved approximation of Frame (1949)

Approach 1: Stirling's Formula

$$\Gamma(x) \sim \sqrt{2\pi} \cdot x^{x-\frac{1}{2}} \cdot e^{-x} \text{ (see Jameson, 2013)}$$

$$\begin{aligned} C_\nu &= \frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)} \\ &= \frac{\sqrt{2\pi}\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}-\frac{1}{2}} \cdot e^{-\frac{\nu}{2}}}{\sqrt{2\pi}\left(\frac{\nu+1}{2}\right)^{\frac{\nu}{2}} \cdot e^{-\frac{\nu+1}{2}}} = \cdots \\ &= \sqrt{2e\frac{\nu^{\nu-1}}{(\nu+1)^\nu}} \end{aligned}$$

Approach 2: Wendel's approximation (1948)

$$\left(1 + \frac{a}{x+a}\right)^{1-a} \leq \frac{\Gamma(x+a)}{x^a \cdot \Gamma(x)} \leq 1$$

For fixed a, letting $x \longrightarrow \infty$ gives

$$\frac{\Gamma(x+a)}{x^a \cdot \Gamma(x)} \longrightarrow 1$$

implying

$$\frac{\Gamma(x)}{\Gamma(x+a)} \longrightarrow x^{-a}$$

We can enlist this asymptotic result in our current context by letting $x = \nu/2$ and $a = 1/2$, giving a very simple approximation:

$$C_\nu = \frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}$$

$$\sim \left(\frac{\nu}{2}\right)^{-\frac{1}{2}}$$

$$= \sqrt{\frac{2}{\nu}}$$

Approach 3: Frame's improved approximation (1949):

$$\frac{\Gamma\left(n+\frac{1+u}{2}\right)}{\Gamma\left(n+\frac{1-u}{2}\right)} \sim \left(n^2+\frac{1-u^2}{12}\right)^{\frac{u}{2}}$$

The main work here is to show that with the following specific choices:

- $u=-\frac{1}{2}$
- $n=\frac{2\nu-1}{4}$

we can transforms LHS directly into C_ν.

So, Frame's approximation gives us

$$C_\nu \sim \left(n^2 + \frac{1-u^2}{12}\right)^{\frac{u}{2}}$$

$$= \left[\left(\frac{2\nu-1}{4}\right)^2 + \frac{1-\left(-\frac{1}{2}\right)^2}{12}\right]^{-\frac{1}{4}}$$

$$= \cdots$$

$$= \left(\frac{8}{2\nu^2 - 2\nu + 1}\right)^{\frac{1}{4}}$$

Three approximations – how good are they?

1. $C_\nu = \sqrt{2e\frac{\nu^{\nu-1}}{(\nu+1)^\nu}}$ (Stirling)

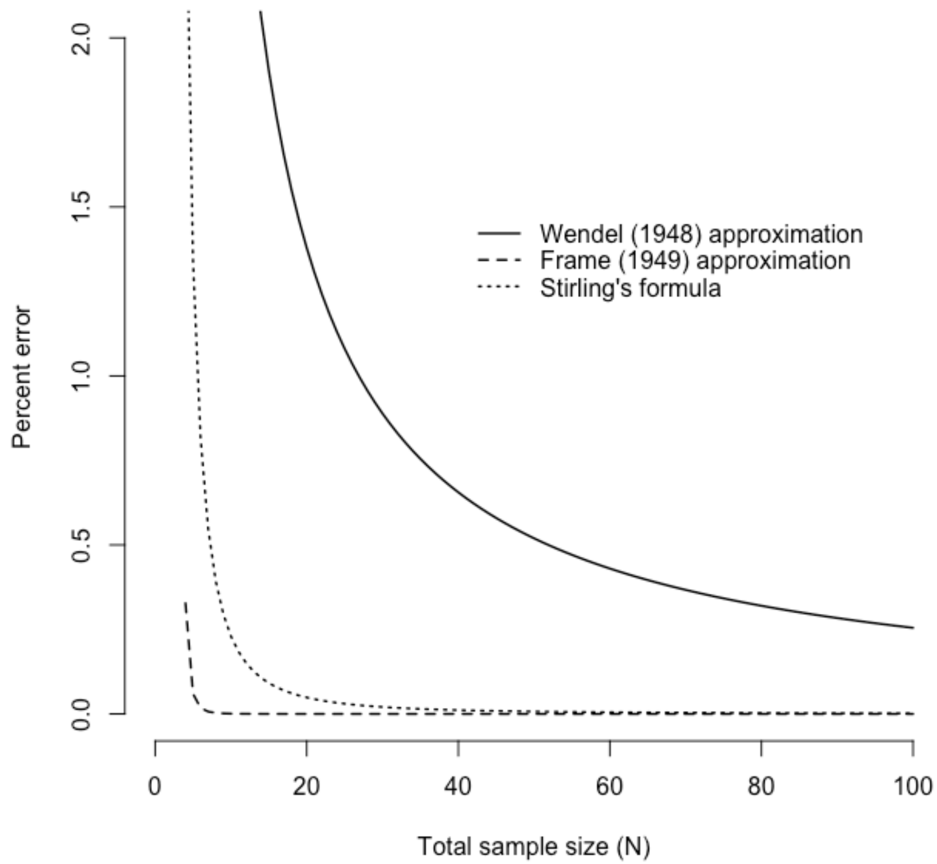
2. $C_\nu = \sqrt{\frac{2}{\nu}}$ (Wendel)

3. $C_\nu = \left(\frac{8}{2\nu^2-2\nu+1}\right)^{\frac{1}{4}}$ (Frame)

Simulation study:

- for $N = 4, \ldots, 100$, generate 1000 random datasets of size N
- compute analytic Bayes factor
- compute closed-form approximations
- plot percent error as a function of N

They're all pretty good!



An example – Borota et al. (2014) observed significantly better test scores for participants who received 200 mg of caffeine compared to those who took a placebo, $t(71) = 2.0$, $p = 0.049$.

Recall:

$$\begin{aligned}\mathsf{BF}_{10} &= \frac{C_\nu}{\sqrt{\pi}}\left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu-1}{2}} \\ &= \frac{C_{71}}{\sqrt{\pi}}\left(1 + \frac{(2.0)^2}{71}\right)^{\frac{71-1}{2}} \\ &= C_{71} \cdot 3.8417\end{aligned}$$

C_{71} – Stirling approximation

$$
\begin{aligned}
C_{71} &\approx \sqrt{2e\frac{\nu^{\nu-1}}{(\nu+1)^{\nu}}} \\
&= \sqrt{2e\frac{71^{71-1}}{(71+1)^{71}}} \\
&= \sqrt{2e\frac{71^{70}}{(72)^{71}}} \\
&= 0.1684
\end{aligned}
$$

This gives $\mathsf{BF}_{10} = 0.1684 \cdot 3.8417 = 0.647$

C_{71} – Wendel approximation

$$
\begin{aligned}
C_{71} &\approx \sqrt{\frac{2}{\nu}} \\
&= \sqrt{\frac{2}{71}} \\
&= 0.1678
\end{aligned}
$$

This gives $\mathsf{BF}_{10} = 0.1678 \cdot 3.8417 = 0.645$

C_{71} – Frame approximation

$$\begin{aligned} C_{71} &\approx \left(\frac{8}{2\nu^2 - 2\nu + 1}\right)^{\frac{1}{4}} \\ &= \left(\frac{8}{2(71)^2 - 2(71) + 1}\right)^{\frac{1}{4}} \\ &= 0.1684 \end{aligned}$$

This gives $\text{BF}_{10} = 0.1684 \cdot 3.8417 = 0.647$

Recall – Borota et al. (2014) observed significantly better test scores for participants who received 200 mg of caffeine compared to those who took a placebo, $t(71) = 2.0$, $p = 0.049$.

Note: $\text{BF}_{01} = \dfrac{1}{\text{BF}_{10}} = \dfrac{1}{0.65} = 1.54$.

This means that Borota's data are 1.54 times more likely under the null hypothesis $\mathcal{H}_0$ than under the alternative hypothesis $\mathcal{H}_1$ (this is an example of something called Lindley's paradox).

Moral – don't trust p-values just below 0.05!

Thank you!

- slides available at `https://tomfaulkenberry.github.io`
- Twitter/Mastodon: `@tomfaulkenberry@mathstodon.xyz`
- Email: `faulkenberry@tarleton.edu`