Partial Bayes Factors v1.3_20240619

$$\frac{p(\mathcal{M}_j \mid \boldsymbol{y})}{p(\mathcal{M}_i \mid \boldsymbol{y})} = \underbrace{\frac{p(\boldsymbol{y} \mid \mathcal{M}_j)}{p(\boldsymbol{y} \mid \mathcal{M}_i)}}_{B_{ji} := m_j(\boldsymbol{y}) / m_i(\boldsymbol{y})} \cdot \frac{p(\mathcal{M}_j)}{p(\mathcal{M}_i)} = \frac{\int_{\boldsymbol{\theta} \in \Theta_j} p(\boldsymbol{y} \mid \boldsymbol{\theta}, \mathcal{M}_j) \cdot \pi(\boldsymbol{\theta} \mid \mathcal{M}_j)\, \mathrm{d}\boldsymbol{\theta}}{\int_{\boldsymbol{\theta} \in \Theta_i} p(\boldsymbol{y} \mid \boldsymbol{\theta}, \mathcal{M}_i) \cdot \pi(\boldsymbol{\theta} \mid \mathcal{M}_i)\, \mathrm{d}\boldsymbol{\theta}} \cdot \frac{p(\mathcal{M}_j)}{p(\mathcal{M}_i)} \equiv \frac{\int p_j(\boldsymbol{y} \mid \boldsymbol{\theta}_j) \cdot \pi_j(\boldsymbol{\theta}_j)\, \mathrm{d}\boldsymbol{\theta}_j}{\int p_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i) \cdot \pi_i(\boldsymbol{\theta}_i)\, \mathrm{d}\boldsymbol{\theta}_i} \cdot \frac{p(\mathcal{M}_j)}{p(\mathcal{M}_i)} \tag{1}$$

Problem

Given a prior $\pi_i(\boldsymbol{\theta}_i) = c_i \cdot \tilde{\pi}_i(\boldsymbol{\theta}_i)$, if the normalizing constant c_i does not exist, then $\tilde{\pi}_i(\boldsymbol{\theta}_i)$ is improper. Provided that the integral in Eq. 1 converges, c_i cancels out in the posterior $\pi_i(\boldsymbol{\theta}_i \mid \boldsymbol{y}) = p_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i) \cdot \tilde{\pi}_i(\boldsymbol{\theta}_i) \big/ \tilde{m}_i(\boldsymbol{y})$ (an improper prior leads to a proper posterior), but does not cancel out in the Bayes factor $B_{ji} = (c_j / c_i) \cdot \tilde{B}_{ji}$. B_{ji} is undetermined since c_j / c_i is arbitrary.

Motivation 1

A model selection method that is automatic and produces actual Bayes factors with respect to proper priors.

Method 1

The intrinsic Bayes factor (IBF)

- Used a *training sample* $\boldsymbol{y}_l$ to convert the noninformative priors $\tilde{\pi}_i(\boldsymbol{\theta}_i)$ to proper posteriors $\pi_i(\boldsymbol{\theta}_i \mid \boldsymbol{y}_l)$,
  assuming $\boldsymbol{y}_l$ is *proper* if its marginal likelihoods $\tilde{m}_i(\boldsymbol{y}_l) \in (0, +\infty), \forall i$,
  and *minimal* if it is proper and no subset is proper;
- Computed the IBF $B_{ji}\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right)$ based on the remaining data partition $\boldsymbol{y}_l^{\complement}$ and the "priors" $\pi_i(\boldsymbol{\theta}_i \mid \boldsymbol{y}_l)$;

$$m_i\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right) := \int p_i\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{\theta}_i, \boldsymbol{y}_l\right) \cdot \pi_i(\boldsymbol{\theta}_i \mid \boldsymbol{y}_l)\, \mathrm{d}\boldsymbol{\theta}_i \iff m_i(\boldsymbol{y}) \;/\; m_i(\boldsymbol{y}_l) \tag{2}$$

$$B_{ji}\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right) := m_j\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right) \;/\; m_i\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right) = \tilde{B}_{ji} \cdot \tilde{B}_{ij}(\boldsymbol{y}_l) \tag{3}$$

- Selected another minimal training sample, iterated l, and averaged the IBFs (arithmetic or geometric).

$$AIBF_{ji} := \frac{1}{L} \sum_{l=1}^{L} B_{ji}\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right) = \tilde{B}_{ji} \cdot \frac{1}{L} \sum_{l=1}^{L} \tilde{B}_{ij}(\boldsymbol{y}_l) \qquad GIBF_{ji} := \left(\prod_{l=1}^{L} B_{ji}\left(\boldsymbol{y}_l^{\complement} \mid \boldsymbol{y}_l\right) \right)^{1/L} = \tilde{B}_{ji} \cdot \left(\prod_{l=1}^{L} \tilde{B}_{ij}(\boldsymbol{y}_l) \right)^{1/L}$$

* Caution the ~ notation and the subscripts $_{ij}$ or $_{ji}$

Motivation 2

$L = \binom{n}{m}$

An alternative to the IBFs that avoids the selection of and the subsequent averaging over training samples.

n and m are the sample and minimal training sample sizes, respectively.

Method 2

The fractional Bayes factor (FBF)

- Cancelled out c_i and kept $\tilde{\pi}_i(\boldsymbol{\theta}_i)$, but used a fraction $b = m / n \in (0, 1)$ of the likelihood to *properize* Eq. 4.
- ◦ Show that the likelihood $p_i(\boldsymbol{y}_l \mid \boldsymbol{\theta}_i) \to p_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)^b$ as $m, n \to +\infty$.

$$m_i(b, \boldsymbol{y}) := \frac{\int p_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i) \cdot \tilde{\pi}_i(\boldsymbol{\theta}_i)\, \mathrm{d}\boldsymbol{\theta}_i}{\int p_i(\boldsymbol{y} \mid \boldsymbol{\theta}_i)^b \cdot \tilde{\pi}_i(\boldsymbol{\theta}_i)\, \mathrm{d}\boldsymbol{\theta}_i} \;\to\; \tilde{m}_i(\boldsymbol{y}) \;/\; \tilde{m}_i(\boldsymbol{y}_l) \tag{4}$$

$$FBF_{ji} := m_j(b, \boldsymbol{y}) \;/\; m_i(b, \boldsymbol{y}) = \tilde{B}_{ji} \cdot \tilde{B}_{ij}(b) \tag{5}$$

Example: A simple linear regression without an intercept, $y_i = \beta x_i + \epsilon_i$, $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ for $i = 1, \cdots, n$.

$\mathcal{H}_0 : \beta = 0$ versus $\mathcal{H}_1 : \beta \neq 0$. Assume a Jeffreys prior $\tilde{\pi}_1(\beta, \sigma_\epsilon) = 1 \,/\, \sigma_\epsilon$.

$FBF_{01} = \dfrac{\Gamma(n/2) \cdot \Gamma((nb-1)/2)}{\Gamma((n-1)/2) \cdot \Gamma(nb/2)} \cdot \left(1 + (n-1)^{-1} F\right)^{-n(1-b)/2}$, where $F = \dfrac{\text{MSR}}{\text{MSE}} = \dfrac{\text{SSR} \;/\; 1}{\text{SSE} \;/\; (n-1)}$. **Questions**: What is m?

DISCUSSIONS

1. The Valencia International Meetings on Bayesian Statistics (every four years from 1979 to 2010).

"*a growing emphasis on computational issues, concerned with making Bayesian methods routinely available to applied practitioners*"

Debate over *subjective* or *objective* Bayesian methods.

The objective Bayesian methods are typically based on noninformative priors and are sometimes called "default" or "automatic" (including but not limited to "intrinsic") Bayesian methods, i.e., not requiring substantial prior input from the user, to avoid the loaded connotation of the label "objective".

"*I (Berger) am more of an objective Bayesian than a subjective Bayesian. We were conducting objective Bayesian analyses in very large model spaces, such as variable selection with 100 variables. In Bayesian analysis, you have to assign prior probabilities to models, and you cannot do subjective assignment of* 2^{100} *models.*"

2. Hierarchical model specification.

Improper priors combined with proper hyperpriors can result in proper marginal priors.

3. Derive $\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n \left(\mathcal{N}(y_i \mid \mu, \sigma^2)\right)^b \cdot \frac{1}{\sigma}\, \mathrm{d}\mu\, \mathrm{d}\sigma$. What is m? Note that $\Gamma(z)$ requires $z > 0$.

4. Choose $\boldsymbol{y}_l$ for IBF and b for FBF.

$\boldsymbol{y}_l$ is the data for which all $\boldsymbol{\theta}_i$ are identifiable; $m = \max\{\dim(\boldsymbol{\theta}_i)\}$. If $\pi_i(\boldsymbol{\theta}_i)$ is actually proper, then $m = 0$.

5. Recall the Pearson Bayes factor for one-way analysis of variance in between-subjects designs.

$$PBF_{10} = \frac{\Gamma\left(\frac{df_\mathrm{B}}{2} + 1 + \gamma\right) \cdot \Gamma\left(\frac{df_\mathrm{W}}{2}\right)}{\Gamma\left(\frac{df_\mathrm{T}}{2}\right) \cdot \Gamma(1+\gamma)} \left(\frac{df_\mathrm{W}}{df_\mathrm{W} + df_\mathrm{B}\, F}\right)^{-\frac{df_\mathrm{W}}{2} + 1 + \gamma},$$

where $df_\mathrm{B} = a - 1$, $df_\mathrm{W} = N - a$, and $df_\mathrm{T} = N - 1$ are the degrees of freedom for between-groups, within-group, and total sources of variation, respectively. $N = \sum_{i=i}^{a} n_i$

6. *Coherence* conditions are (i) $B_{12} = B_{10} \cdot B_{02}$ and (ii) $B_{01} = 1 / B_{10}$ when the priors are proper. *TBC* p. 270

Show that the FBF satisfies (ii) but not (i). Show that neither the AIBF nor the GIBF satisfy (i) and (ii).

7. Read more:

asymptotic and consistency, non-nested models, sensitivity analysis, *etc*.

Berger, J. O. (2023). Frequentist and/or Bayesian adjustment for multiple testing. *29th Distinguished Statistician's Colloquium Sponsored by Pfizer/ASA/UConn*. https://youtu.be/3gtd9-wjbBQ?t=276
Berger, J. O., & Pericchi, L. R. (1993). The intrinsic Bayes factor for model selection and prediction. *Technical Report, 93-43C*, 1–51. https://www.stat.purdue.edu/docs/research/tech-reports/1993/tr93-43c.pdf
cf.
Berger, J. O., & Pericchi, L. R. (1994). The intrinsic Bayes factor for linear models. *Technical Report, 94-10C*, 1–39. https://www.stat.purdue.edu/docs/research/tech-reports/1994/tr94-10c.pdf
Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*, 109–122. DOI: 10.1080/01621459.1996.10476668
Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for linear models. *Bayesian Statistics*, *5*, 25–44. DOI: 10.1093/oso/9780198523567.003.0002
Faulkenberry, T. J. (2021). The Pearson Bayes factor: An analytic formula for computing evidential value from minimal summary statistics. *Biometrical Letters*, *58*, 1–26. DOI: 10.2478/bile-2021-0001
O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 99–118. DOI: 10.1111/j.2517-6161.1995.tb02017.x
O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test*, *6*, 101–118. DOI: 10.1007/BF02564428
Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation* (2nd ed.). New York: Springer.

$$\int_0^\infty \int_{-\infty}^\infty \prod_{i=1}^n \left(\mathcal{N}(y_i \mid \mu,\, \sigma^2)\right)^b \cdot \frac{1}{\sigma}\, \mathrm{d}\mu\, \mathrm{d}\sigma$$

$$= \int_0^\infty \int_{-\infty}^\infty \left(2\pi\sigma^2\right)^{-\frac{nb}{2}} \cdot \exp\left\{ -\frac{b}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right\} \cdot \frac{1}{\sigma}\, \mathrm{d}\mu\, \mathrm{d}\sigma \tag{1}$$

$$= \int_0^\infty (2\pi)^{-\frac{nb}{2}}\, \sigma^{-nb-1} \int_{-\infty}^\infty \exp\left\{ -\frac{b}{2\sigma^2} \sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2) \right\} \mathrm{d}\mu\, \mathrm{d}\sigma \tag{2}$$

$$= \int_0^\infty (2\pi)^{-\frac{nb}{2}}\, \sigma^{-nb-1} \int_{-\infty}^\infty \exp\left\{ -\frac{b}{2\sigma^2} (\boldsymbol{y}^\top \boldsymbol{y} - 2\mu n \bar{y} + n\mu^2) \right\} \mathrm{d}\mu\, \mathrm{d}\sigma \tag{3}$$

$$= \int_0^\infty (2\pi)^{-\frac{nb}{2}}\, \sigma^{-nb-1} \cdot \exp\left\{ -\frac{b}{2\sigma^2} \boldsymbol{y}^\top \boldsymbol{y} \right\} \int_{-\infty}^\infty \exp\left\{ -\frac{nb}{2\sigma^2} \left((\mu - \bar{y})^2 - \bar{y}^2\right) \right\} \mathrm{d}\mu\, \mathrm{d}\sigma \tag{4}$$

$$= \int_0^\infty (2\pi)^{-\frac{nb}{2}}\, \sigma^{-nb-1} \cdot \exp\left\{ -\frac{b}{2\sigma^2} \left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) \right\} \int_{-\infty}^\infty \exp\left\{ -\frac{nb}{2\sigma^2} (\mu - \bar{y})^2 \right\} \mathrm{d}\mu\, \mathrm{d}\sigma \tag{5}$$

$$= \int_0^\infty (2\pi)^{-\frac{nb}{2}}\, \sigma^{-nb-1} \cdot \exp\left\{ -\frac{b}{2\sigma^2} \left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) \right\} \cdot \sqrt{\frac{2\pi\sigma^2}{nb}}\, \mathrm{d}\sigma \tag{6}$$

$$= (nb)^{-\frac{1}{2}}\, (2\pi)^{-\frac{nb-1}{2}} \int_0^\infty \sigma^{-nb} \cdot \exp\left\{ -\frac{b}{2\sigma^2} \left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) \right\} \mathrm{d}\sigma \tag{7}$$

$$= (nb)^{-\frac{1}{2}}\, (2\pi)^{-\frac{nb-1}{2}} \int_\infty^0 u^{\frac{nb}{2}} \cdot \exp\left\{ -\frac{b}{2}\, u \left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) \right\} \mathrm{d}u^{-\frac{1}{2}} \tag{8}$$

$$= (nb)^{-\frac{1}{2}}\, (2\pi)^{-\frac{nb-1}{2}} \left(-\frac{1}{2}\right) \int_\infty^0 u^{\frac{nb-1}{2}-1} \cdot \exp\left\{ -\frac{b}{2}\, u \left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) \right\} \mathrm{d}u \tag{9}$$

$$\begin{aligned}
&= (nb)^{-\frac{1}{2}}\, (2\pi)^{-\frac{nb-1}{2}}\, \frac{1}{2} \cdot \left(\frac{b}{2}\left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right)\right)^{-\frac{nb-1}{2}} \cdot \\
&\quad \int_0^\infty \left(\frac{b}{2}\left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) u\right)^{\frac{nb-1}{2}-1} \cdot \exp\left\{ -\frac{b}{2}\left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) u \right\} \mathrm{d}\left(\frac{b}{2}\left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right) u\right)
\end{aligned} \tag{10}$$

$$= (nb)^{-\frac{1}{2}}\, (2\pi)^{-\frac{nb-1}{2}}\, \frac{1}{2} \left(\frac{b}{2}\left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right)\right)^{-\frac{nb-1}{2}} \cdot \Gamma\left(\frac{nb-1}{2}\right) \tag{11}$$

$$= \frac{1}{2} \cdot n^{-\frac{1}{2}} \cdot b^{-\frac{nb}{2}} \cdot \left(\boldsymbol{\pi} \cdot \left(\boldsymbol{y}^\top \boldsymbol{y} - n\bar{y}^2\right)\right)^{-\frac{nb-1}{2}} \cdot \Gamma\left(\frac{nb-1}{2}\right) \tag{12}$$