Bayesian Statistics III/IV (MATH3341/4031) Michaelmas term, 2019

Handout 14: Jeffreys-Lindley 'Paradox' (?) [a]

Lecturer: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim: To explain, and theorize Jeffreys-Lindley Paradox

References:

- Berger, J. O. (2013; Section 4.3.3). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.
- Robert, C. (2007; Section 5.2(exclude 5.2.6)). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.

[a] Author: Georgios P. Karagiannis.

1 Improper priors situations

Jeffreys-Lindley Paradox describes the overwhelming statistical evidence in favor a hypothesis when the priors of the alternative hypotheses diverge faster. It occurs in Bayesian hypothesis test and model selection with improper priors.

Example 1. [Single-vs-General-alternative]

Let $y = (y_1, ..., y_n)$ observables and consider the Bayesian hypothesis test

$$\mathrm{H}_0 : y_i|\theta_0 \overset{\text{IID}}{\sim} \mathrm{N}(\theta_0, \sigma^2),\ i = 1, ..., n \qquad \text{vs} \qquad \mathrm{H}_1 : \begin{cases} y_i|\mu & \overset{\text{IID}}{\sim} \mathrm{N}(\mu, \sigma^2),\ i = 1, ..., n \\ \mu & \sim \mathrm{N}(\mu_0, \sigma_0^2) \end{cases} \tag{1}$$

with $\pi_j = \mathsf{P}_\Pi(\theta \in \Theta_j)$ for $j = 0, 1$. Let $\theta_0, \sigma^2, \mu_0, \sigma_0^2$ be fixed values. It can be computed (Appendix A) that

$$\mathrm{B}_{01}(y) = \frac{\left(\frac{\sigma^2}{n}\right)^{-\frac{1}{2}}}{\left(\frac{\sigma^2}{n} + \sigma_0^2\right)^{-\frac{1}{2}}} \frac{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\theta_0)^2}{\frac{\sigma^2}{n}}\right)}{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n}+\sigma_0^2}\right)} ; \quad \mathsf{P}_\Pi(\mathrm{H}_0|y) = \left(1 + \frac{1-\pi_0}{\pi_0} \frac{\left(\frac{\sigma^2}{n} + \sigma_0^2\right)^{-\frac{1}{2}}}{\left(\frac{\sigma^2}{n}\right)^{-\frac{1}{2}}} \frac{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n}+\sigma_0^2}\right)}{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\theta_0)^2}{\frac{\sigma^2}{n}}\right)}\right)^{-1}$$

1. The Jeffreys' (and Laplace) prior of μ is $\pi^{(\mathrm{J})}(\mu) \propto 1(\mu \in \mathbb{R})$ (see Handout 7). Find values μ_*, σ_*^2 such that
$$\tilde{\pi}(\mu|\mathrm{H}_1) \to \tilde{\pi}^{(\mathrm{J})}(\mu)\ ; \ \text{as} \quad (\mu_0, \sigma_0^2) \to (\mu_*, \sigma_*^2) \tag{2}$$
where $\tilde{\pi}(\mu|\mathrm{H}_1)$ denotes the kernel of the pdf of the conditional prior $\mu \sim \mathrm{N}(\mu_0, \sigma_0^2)$ of μ under H_1in (1), and $\tilde{\pi}^{(\mathrm{J})}(\mu)$ denotes that of the the Jeffreys prior $\pi^{(\mathrm{J})}(\mu) \propto 1(\mu \in \mathbb{R})$.
2. [Lindley's Paradox] Let $\sigma_0^2 \to \infty$. Investigate, how, and why
    - the conditional prior $\pi(\mu|\mathrm{H}_1)$ given the hypothesis H_1 behaves?
    - the Bayes Factor $\mathrm{B}_{01}(y)$ and the posterior $\mathsf{P}_\Pi(\mathrm{H}_0|y)$ behave?

Solution. The calculation of $\mathrm{B}_{01}(y)$ and of $\mathsf{P}_\Pi(\mathrm{H}_0|y)$ is presented in the Appendix A.

1. The kernel $\tilde{\pi}(\mu|\mathrm{H}_1)$ of the prior $\pi(\mu|\mathrm{H}_1)$ is

$$\pi(\mu|\mathrm{H}_1) = \mathrm{N}(\mu|\mu_0, \sigma_0^2) \propto \exp\left(-\frac{1}{2}\frac{1}{\sigma_0^2}(\mu-\mu_0)^2\right) 1(\mu \in \mathbb{R}) = \tilde{\pi}(\mu|\mathrm{H}_1)$$

   So for $\sigma_*^2 = \infty$ and any $|\mu_*| < \infty$, I get the (2).

2. If $\sigma_0^2 \to \infty$, the prior $\mathrm{N}(\mu_0, \sigma_0^2)$ of H_1 becomes flat (the mass spreads uniformly around $\mathbb{R}$) and improper:

$$\pi(\mu|\mathrm{H}_1) = \mathrm{N}\left(\mu|\mu_0, \sigma_0^2\right) \propto \exp\left(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}\right) 1(\mu \in \mathbb{R}) \xrightarrow{\sigma_0^2 \to \infty} 1(\mu \in \mathbb{R})$$

   In fact $\pi(\mu|\mathrm{H}_1)$ meets Jeffreys' (or Laplace) prior of μ in the limit $\sigma_0^2 \to \infty$.

   Then in the limit the Bayes factor $\mathrm{B}_{01}(y)$ and the posterior $\mathsf{P}_\Pi(\mathrm{H}_0|y)$ approach the values

$$\mathrm{B}_{01}(y) \to \infty, \text{ and } \mathsf{P}_\Pi(\mathrm{H}_0|y) \to 1 \text{ as } \quad \sigma_0^2 \to \infty$$

   We observe that regardless the number of the observables n, when the prior variance σ_0^2 of μ given H_1 becomes huge $\left(\sigma_0^2 \to \infty\right)$, the prior of μ in H_1 becomes more diverge

$$\pi(\mu|\mathrm{H}_1) \xrightarrow{\sigma_0^2 \to \infty} 1(\mu \in \mathbb{R}), \text{ as } \quad \sigma_0^2 \to \infty$$

   spreading the prior mass uniformly in $\mathbb{R}$ while the evidence in favor $\mathrm{H}_0 : \mu = \theta_0$ that μ is equal to single value θ_0 becomes overwhelming.

$$\mathrm{B}_{01}(y) \xrightarrow{\sigma_0^2 \to \infty} \infty \quad \text{and} \quad \mathsf{P}_\Pi(\mathrm{H}_0|y) \xrightarrow{\sigma_0^2 \to \infty} 1$$

   Paradoxical behavior: One would not expect for $\sigma_0^2 \to \infty$ favor $\mathrm{H}_0 : \mu = \theta_0$ because $\sigma_0^2 \to \infty$ gives increases ignorance and give uniformly positive mass to more values. Also, we wouldn't expect increasing the sample size n to have no effect.

   Essentially, the above paradoxical behavior says that: When $\sigma_0^2 \to \infty$, it means that a priori, I know $\mu = \theta_0$ with probability π_0, but I know nothing about μ probability $1 - \pi_0$; however after I get any observation y I am a posteriori certain that $\mu = \theta_0$.

   Observe that the overall posterior is

$$\pi(\mu) = \pi_0 1(\mu \in \{\theta_0\}) + (1-\pi_0)\left[(\frac{1}{2\pi\sigma_0^2})^{\frac{1}{2}} \exp(-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0^2}) 1(\mu \in \mathbb{R})\right]$$

   When $\sigma_0^2 \to \infty$, it involves one conditional component that concentrates the mass $\pi(\mu|\mathrm{H}_0) = 1_{\{\theta_0\}}(\mu)$ and one component with infinite normalizing constant. So we cannot apply the Bayes theorem by using the kernels of the priors and in the sense of canceling normalizing constants as $\pi(\mu|y) \propto f(y|\mu)\tilde{\pi}(\mu)$ where $\pi(\mu) \propto \tilde{\pi}(\mu)$.

Lindley's paradox also appears in composite-vs-composite tests with improper priors, when the one prior diverges faster than the other in the limit. See the following realistic example in Bayesian regression Variable selection problem.

Example 2. [Composite-vs-Composite]

Consider a Normal linear regression model with dependent variable y and a set of regressors $\{\Phi_j\}_{j\in\mathscr{M}}$ where $\mathscr{M}$ is the set of size d that includes the labels of the available regressors; e.g.

$$y_i|\beta, \sigma^2 \sim \mathrm{N}\left(\sum_{j\in\mathscr{M}} \Phi_{i,j}\beta_j, I\sigma^2\right), \quad \text{for } i = 1, ..., n$$

where the regression coefficients $\{\beta_j\}_{j\in\mathscr{M}}$ and the noise variance σ^2 are unknown.

Let $\mathscr{M}_0$ and $\mathscr{M}_1$ denote two sets of regressors (nested or not) with $\dim(\mathscr{M}_j) = d_j$. We are interested in learning whether the linear model with $\mathscr{M}_0$ set of regressors or that with $\mathscr{M}_0$ set of regressors models the data generating processes 'better'. I.e. we may test

$$\mathrm{H}_0 : \begin{cases} y|\beta_{\mathscr{M}_0}, \sigma^2 & \sim \mathrm{N}(\Phi_{\mathscr{M}_0}\beta_{\mathscr{M}_0}, I\sigma^2) \\ \beta_{\mathscr{M}_0}|\sigma^2 & \sim \mathrm{N}(\mu_{\mathscr{M}_0}, V_{\mathscr{M}_0}\sigma^2) \\ \sigma^2 & \sigma^2 \sim \mathrm{IG}(a,k) \end{cases} \qquad \text{v.s.} \qquad \mathrm{H}_1 : \begin{cases} y|\beta_{\mathscr{M}_1}, \sigma^2 & \sim \mathrm{N}(\Phi_{\mathscr{M}_1}\beta_{\mathscr{M}_1}, I\sigma^2) \\ \beta_{\mathscr{M}_1}|\sigma^2 & \sim \mathrm{N}(\mu_{\mathscr{M}_1}, V_{\mathscr{M}_1}\sigma^2) \\ \sigma^2 & \sigma^2 \sim \mathrm{IG}(a,k) \end{cases}$$

The Bayes factor $B_{01}(y)$ and the posterior marginal probability $\mathsf{P}_\Pi(\mathrm{H}_0|y)$ are (See the Appendix B)

$$B_{01}(y) = \sqrt{\frac{|V_1|}{|V_0|}}\sqrt{\frac{|V_0^*|}{|V_1^*|}}\left(\frac{k_0^*}{k_1^*}\right)^{-\frac{n}{2}-a}; \qquad \mathsf{P}_\Pi(\mathrm{H}_0|y) = \left(1 + \frac{1-\pi_0}{\pi_0}B_{01}^{-1}(y)\right)^{-1}$$

where for $j = 0, 1$

$$k_j^* = k + \frac{1}{2}\mu_j^\top V_j^{-1}\mu_j - \frac{1}{2}\left(\mu_j^*\right)^\top \left(V_j^*\right)^{-1}\mu_j^* + \frac{1}{2}y^\top y$$

$$V_j^* = \left(V_j^{-1} + \Phi_j^\top \Phi_j\right)^{-1}; \qquad \mu_j^* = V_j^*\left(V_j^{-1}\mu_j + \Phi_j^\top y\right)$$

Assume that $V_{\mathscr{M}_0} = vI_{d_0}$ and $V_{\mathscr{M}_1} = vI_{d_1}$. Let $v \to \infty$, how $B_{01}(y)$ and $\mathsf{P}_\Pi(\mathrm{H}_0|y)$ behave when (1.) $d_0 < d_1$, (2.) $d_0 > d_1$, and (3.) $d_0 = d_1$

Solution. For $V_0 = vI_{d_0}$ and $V_1 = vI_{d_1}$ it is

$$\lim_{v\to\infty} B_{01}(y) = \lim_{v\to\infty}(v)^{\frac{d_1-d_0}{2}} \times \sqrt{\frac{|\Phi_0^\top\Phi_0|}{|\Phi_1^\top\Phi_1|}}\left(\frac{k - \frac{1}{2}y^\top\Phi_0\left(\Phi_0^\top\Phi_0\right)^{-1}\Phi_0^\top y_0 + y^\top y}{k - \frac{1}{2}y^\top\Phi_1\left(\Phi_1^\top\Phi_1\right)^{-1}\Phi_1^\top y_1 + y^\top y}\right)^{-\frac{n}{2}-a}$$

$$\text{So } B_{01}(y) \xrightarrow{v\to\infty} \begin{cases} +\infty, & d_0 < d_1 \\ 0, & d_0 > d_1 \\ < \infty, & d_0 = d_1 \end{cases} \quad \text{and} \quad \mathsf{P}_\Pi(\mathrm{H}_0|y) \xrightarrow{v\to\infty} \begin{cases} 1, & d_0 < d_1 \\ 0, & d_0 > d_1 \\ \in (0,1), & d_0 = d_1 \end{cases}$$

As $v \to \infty$, both conditional priors $\beta_0|\sigma^2 \sim \mathrm{N}(\mu_0, \sigma^2 Iv)$ and $\beta_1|\sigma^2 \sim \mathrm{N}(\mu_1, \sigma^2 Iv)$ under hypothesis H_0 and H_1 become more and more diverge, while the evidence becomes more overwhelming in favor of the hypothesis that the prior diverges slower. In our example, when $d_0 < d_1$ the conditional prior $\pi(\beta|\sigma^2, \mathrm{H}_0)$ given H_0 diverges slower.

2 Informal but intuitive investigation of the phenomenon

Consider

$$\mathrm{H}_0 : \theta \in \Theta_0, \text{ vs } \mathrm{H}_1 : \theta \in \Theta_1$$

where $\{\Theta_0, \Theta_1\}$ unbounded sets partitioning $\Theta \subseteq \mathbb{R}^d$. Consider overall prior with pdf

$$\pi(\theta) = \pi_0\ \pi_0(\theta) + \pi_1\ \pi_1(\theta)$$

Let conditional priors $\pi_0(\theta) = \frac{1}{C_0}\tilde{\pi}_0(\theta)$ and $\pi_1(\theta) = \frac{1}{C_1}\tilde{\pi}_1(\theta)$ where $\tilde{\pi}_0(\theta)$ and $\tilde{\pi}_1(\theta)$ are pdf kernels.

The posterior probability $\mathsf{P}_\Pi(\mathrm{H}_0|y)$ of the hypothesis H_0 is

$$\mathsf{P}_\Pi(\mu \in \Theta_0|y) = \frac{\pi_0 C_0^{-1}\int_{\Theta_0}\tilde{\pi}_0(\theta)f(y|\theta)\mathrm{d}\theta}{\pi_0 C_0^{-1}\int_{\Theta_0}\tilde{\pi}_0(\theta)f(y|\theta)\mathrm{d}\theta + \pi_1 C_1^{-1}\tilde{\pi}_1(\theta)\int_{\Theta_1}f(y|\theta)\mathrm{d}\theta} = \left[1 + \frac{C_0}{C_1}\frac{\int_{\Theta_1}\tilde{\pi}_1(\theta)f(y|\theta)\mathrm{d}\theta}{\int_{\Theta_0}\tilde{\pi}_0(\theta)f(y|\theta)\mathrm{d}\theta}\right]^{-1}$$

The Bayes factor is

$$\mathrm{B}_{01}(y) = \frac{C_1}{C_0}\frac{\int_{\Theta_0}\tilde{\pi}_0(\theta)f(y|\theta)\mathrm{d}\theta}{\int_{\Theta_1}\tilde{\pi}_1(\theta)f(y|\theta)\mathrm{d}\theta}.$$

When $\tilde{\pi}_0(\theta) \to$ const and $\tilde{\pi}_1(\theta) \to$ const, the normalizing constants of $\pi_0(\theta)$ and $\pi_1(\theta)$ become infinite $C_0 \to \infty$ and $C_1 \to \infty$. Then whether $\mathrm{B}_{01}(y) \to 0$ or $\mathrm{B}_{01}(y) \to \infty$ depends on which prior $\pi_0(\theta)$ or $\pi_1(\theta)$ becomes improper faster than the other, namely

$$\frac{C_1}{C_0} \xrightarrow[\tilde{\pi}_1(\theta)\to\text{const}]{\tilde{\pi}_0(\theta)\to\text{const}} \begin{cases} \infty & \text{, if } C_0 << C_1 \\ 0 & \text{, if } C_1 << C_0 \end{cases}$$

In Example 2,

$$C_j = (2\pi)^{\frac{n}{2}}\,|V_j|^{\frac{1}{2}} \overset{V_j=vI}{=} (2\pi)^{\frac{n}{2}}\,v^{\frac{d_j}{2}};\ \Longrightarrow \qquad \frac{C_1}{C_0} = (v)^{\frac{d_1-d_0}{2}} \xrightarrow{v\to\infty} \begin{cases} \infty & \text{, if } d_0 < d_1 \\ 0 & \text{, if } d_0 > d_1 \end{cases} \text{.as}$$

Summary 3. Improper priors do not really work in hypothesis tests, or model comparison. Be cautious when You use improper priors.

3 Large samples situation

It is not a secrete[1] that in Frequentist hypothesis tests you can reject H_0 if you get a large enough sample. P-values is a misleading description of the evidence against H_0. What is the case in Bayesian hypothesis tests?

Example 4. (Cont. of Example 1) Set $\mu_0 = \theta_0$ and let $z_n = \frac{\bar{y}-\mu_0}{\sigma}\sqrt{n}$. Then, it is

$$\mathrm{B}_{01}(y) \overset{\text{calc.}}{=} \frac{\left(1 + n\sigma_0^2/\sigma^2\right)^{\frac{1}{2}}}{\exp\left(\frac{1}{2}z_n^2\left(1+\sigma^2/(\sigma_0^2 n)\right)^{-1}\right)} \geq \frac{\left(1 + n\sigma_0^2/\sigma^2\right)^{\frac{1}{2}}}{\exp\left(\frac{1}{2}z_n^2\right)},$$

Assume that $|z_n| \geq z^*_{1-\frac{a}{2}}$; compare the result of the Bayesian hypothesis test against the a-sig. level Frequentist test.

Solution. Consider z_n as fixed. For the Bayesian hypothesis test, I get

$$\mathrm{B}_{01}(y) \geq \frac{\left(1 + n\sigma_0^2/\sigma^2\right)^{\frac{1}{2}}}{\exp\left(\frac{1}{2}z_n^2\right)},\ \text{and } \mathrm{P}_\Pi(\mathrm{H}_0|y) \geq \left(1 + \frac{1-\pi_0}{\pi_0}\frac{\exp\left(\frac{1}{2}z_n^2\right)}{\left(1 + n\sigma_0^2/\sigma^2\right)^{\frac{1}{2}}}\right)^{-1}$$

As $n \to \infty$ $|z_n| < \infty$ is bounded due to the CLT and hence

$$\mathrm{B}_{01}(y) \to \infty,\ \text{and } \mathrm{P}_\Pi(\mathrm{H}_0|y) \to 1$$

so H_0 is accepted for sure. In the frequentist hypothesis test, I reject H_0 at a-sig. level, because $|z_n| > z^*_{1-\frac{a}{2}}$.

The Bayesian behavior is the reasonable one: It has to be $\bar{y} \approx \mu_0$ for the observed $z_n = \frac{\bar{y}-\mu_0}{\sigma}\sqrt{n}$ to be a fixed finite number and not an infinite as $n \to \infty$

Note 5. It is not difficult to see that for Single-vs-General-alternative tests $\mathrm{H}_0 : \mu = \theta_0 \quad \text{vs} \quad \mathrm{H}_1 : \mu \neq \theta_0$

$$\mathrm{B}_{01}(y) = \frac{f_0(y)}{f_1(y)} \geq \frac{f(y|\theta_0)}{\sup_{\mu\neq\theta_0} f(y|\mu)} = \frac{f(y|\theta_0)}{f(y|\hat{\mu}_{\text{MLE}})} \equiv \text{Max. Likl. Ratio}$$

implying that the $\mathrm{B}_{01}(y)$ has the maximum likelihood ratio as a lower bound.

Question. *Practice with Exercises 67 and 68 in the Exercise sheet.*

[1] ...maybe it is kept as a secrete from students attending only frequentist courses in any university ...

Appendix

The following calculations are given for completeness. They are part of Exercises 67 and 68 in the Exercise sheet.

A Calculations

Hint-1: It is

$$-\frac{1}{2}\sum_{i=1}^{n}\frac{(x-\mu_i)^2}{\sigma_i^2}=-\frac{1}{2}\frac{(x-\hat{\mu})^2}{\hat{\sigma}^2}+C(\hat{\mu},\hat{\sigma}^2)$$

$$\hat{\sigma}^2=(\sum_{i=1}^{n}\frac{1}{\sigma_i^2})^{-1};\qquad \hat{\mu}=\hat{\sigma}^2(\sum_{i=1}^{n}\frac{\mu_i}{\sigma_i^2});\qquad C(\hat{\mu},\hat{\sigma}^2)=\underbrace{\frac{1}{2}\frac{(\sum_{i=1}^{n}\frac{\mu_i}{\sigma_i^2})^2}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}-\frac{1}{2}\sum_{i=1}^{n}\frac{\mu_i^2}{\sigma_i^2}}_{=\text{independent of } x}$$

Hint-2: It is $\sum_{i=1}^{n}(x_i-\theta)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\theta)^2$

Solution. The overall prior is

$$\pi(\mu)=\pi_0 1_{\{\theta_0\}}(\mu)+(1-\pi_0)\mathrm{N}(\mu|\mu_0,\sigma_0^2)$$

with $\pi_0=1/2$ (although the value does not play any role here), and known μ_0 and σ_0^2.

The Bayes factor is

$$\mathrm{B}_{01}(y)=\frac{f_0(y)}{f_1(y)}$$

So

$$\begin{aligned}
f_0(y)=f(y|\theta_0)&=\prod_{i=1}^{n}\mathrm{N}(y_i|\theta_0,\sigma^2)=(\frac{1}{2\pi\sigma^2})^{\frac{n}{2}}\exp(-\frac{1}{2}\sum_{i=1}^{n}\frac{(y_i-\theta_0)^2}{\sigma^2})\\
&=(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\exp(-\frac{1}{2}\frac{1}{\sigma^2}(\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{=ns^2}+n(\bar{y}-\theta_0)^2))\\
&=(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\exp(-\frac{1}{2}\frac{1}{\sigma^2/n}s^2)\exp(-\frac{1}{2}\frac{1}{\sigma^2/n}(\bar{y}-\theta_0)^2)
\end{aligned}$$

$$\begin{aligned}
f_1(y)=\int_{\mathbb{R}}f(y|\theta)\mathrm{d}\Pi_1(\theta)&=\int_{\mathbb{R}}\prod_{i=1}^{n}\mathrm{N}(y_i|\mu,\sigma^2)\mathrm{N}(\mu|\mu_0,\sigma_0^2)\mathrm{d}\mu\\
&=\int_{\mathbb{R}}(2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\exp(-\frac{1}{2}\frac{1}{\sigma^2}(\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})^2}_{=ns^2}+n(\bar{y}-\mu)^2))\times\\
&\quad\times(2\pi)^{-\frac{1}{2}}(\sigma_0^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}\frac{1}{\sigma_0^2}(\mu-\mu_0)^2)\mathrm{d}\mu\\
&=(2\pi)^{-\frac{n}{2}-\frac{1}{2}}(\sigma^2)^{-\frac{n}{2}}(\sigma_0^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}\frac{1}{\sigma^2/n}s^2)\\
&\quad\times\int_{\mathbb{R}}\exp(\underbrace{-\frac{1}{2}\frac{1}{\sigma^2/n}(\bar{x}-\mu)^2-\frac{1}{2}\frac{1}{\sigma_0^2}(\mu-\mu_0)^2}_{=A(\mu)})\mathrm{d}\mu
\end{aligned}$$

$$A(\mu) = -\frac{1}{2}\frac{1}{\sigma^2/n}(\bar{y}-\mu)^2 - \frac{1}{2}\frac{1}{\sigma_0^2}(\mu-\mu_0)^2 = -\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\hat{\sigma}^2} - \frac{1}{2}(\frac{\bar{y}^2}{\sigma^2/n} + \frac{\mu_0^2}{\sigma_0^2}) + \frac{1}{2}\frac{(\frac{\bar{y}}{\sigma^2/n} + \frac{\mu_0}{\sigma_0^2})^2}{\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2}}$$

$$= -\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\hat{\sigma}^2} - \frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n} + \sigma_0^2}$$

where

$$\hat{\sigma}^2 = (\frac{1}{\sigma^2/n} + \frac{1}{\sigma_0^2})^{-1}$$

and

$$\int_{\mathbb{R}} \exp(A(\mu))\,\mathrm{d}\mu = \exp\left(-\frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n} + \sigma_0^2}\right)\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\frac{(\mu-\hat{\mu})^2}{\hat{\sigma}^2}\right)\mathrm{d}\mu = \exp\left(-\frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n} + \sigma_0^2}\right)(2\pi)^{\frac{1}{2}}(\hat{\sigma}^2)^{\frac{1}{2}}$$

So

$$f_1(y) = (2\pi)^{-\frac{n}{2}-\frac{1}{2}}(\sigma^2)^{-\frac{n}{2}}(\sigma_0^2)^{-\frac{1}{2}}\exp(-\frac{1}{2}\frac{1}{\sigma^2/n}s^2)\int_{\mathbb{R}} \exp(A(\mu))\,\mathrm{d}\mu$$

$$= (2\pi)^{-\frac{n}{2}-\frac{1}{2}}(\sigma^2)^{-\frac{n}{2}}(\sigma_0^2)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\frac{1}{\sigma^2/n}s^2\right)(2\pi\hat{\sigma}^2)^{1/2}\exp\left(-\frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n} + \sigma_0^2}\right)$$

It is

$$\mathrm{B}_{01}(y) = \frac{f_0(y)}{f_1(y)} = \frac{\left(\frac{\sigma^2}{n}\right)^{-\frac{1}{2}}}{\left(\frac{\sigma^2}{n} + \sigma_0^2\right)^{-\frac{1}{2}}}\frac{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\theta_0)^2}{\frac{\sigma^2}{n}}\right)}{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\mu_0)^2}{\frac{\sigma^2}{n}+\sigma_0^2}\right)}$$

It is

$$\mathsf{P}_{\Pi}(\mathrm{H}_0|y) = \left(1 + \frac{1-\pi_0}{\pi_0}\mathrm{B}_{01}(y)^{-1}\right)^{-1} = \left(1 + \frac{1-\pi_0}{\pi_0}\frac{\left(\frac{\sigma^2}{n} + \sigma_0^2\right)^{-\frac{1}{2}}}{\left(\frac{\sigma^2}{n}\right)^{-\frac{1}{2}}}\frac{\exp\left(-\frac{1}{2}\frac{(yx-\mu_0)^2}{\frac{\sigma^2}{n}+\sigma_0^2}\right)}{\exp\left(-\frac{1}{2}\frac{(\bar{y}-\theta_0)^2}{\frac{\sigma^2}{n}}\right)}\right)^{-1}$$

B Calculations

You may use the following identity:

$$(y-\Phi\beta)^\top(y-\Phi\beta) + (\beta-\mu)^\top V^{-1}(\beta-\mu) = (\beta-\mu^*)^\top (V^*)^{-1}(\beta-\mu^*) + S^*;$$

$$S^* = \mu^\top V^{-1}\mu - (\mu^*)^\top (V^*)^{-1}(\mu^*) + y^\top y; \qquad V^* = \left(V^{-1} + \Phi^\top\Phi\right)^{-1}; \qquad \mu^* = V^*\left(V^{-1}\mu + \Phi^\top y\right)$$

Solution. For simplicity, we suppress the indexing denoting the sub-set of the regressors. It is

$$f(y) = \int \mathrm{N}\left(y|\Phi\beta, I\sigma^2\right)\mathrm{N}\left(\beta|\mu, V\sigma^2\right)\mathrm{IG}\left(\sigma^2|a,k\right)\mathrm{d}\beta\mathrm{d}\sigma^2$$

$$= \int \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}(y-\Phi\beta)^\top(y-\Phi\beta)\right) \times \left(\frac{1}{2\pi}\right)^{\frac{d}{2}}\left(\frac{1}{|\sigma^2 V|}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2\sigma^2}(\beta-\mu)^\top V^{-1}(\beta-\mu)\right)$$

$$\times \frac{k^a}{\Gamma(a)}\left(\frac{1}{\sigma^2}\right)^{a+1}\exp\left(-\frac{k}{\sigma^2}\right)\mathrm{d}\beta\mathrm{d}\sigma^2 = \ldots$$

$$
\begin{aligned}
f(y) =& \left(\frac{1}{2\pi}\right)^{\frac{n+d}{2}} \left(\frac{1}{|V|}\right)^{\frac{1}{2}} \frac{k^a}{\Gamma(a)} \\
& \times \int \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{d}{2}+a+1} \exp\left(-\frac{1}{2\sigma^2}(y-\Phi\beta)^\top(y-\Phi\beta) - \frac{1}{2\sigma^2}(\beta-\mu)^\top V^{-1}(\beta-\mu) - \frac{k}{\sigma^2}\right) \mathrm{d}\beta \mathrm{d}\sigma^2
\end{aligned}
$$

$$
\begin{aligned}
=& \left(\frac{1}{2\pi}\right)^{\frac{n+d}{2}} \left(\frac{1}{|V|}\right)^{\frac{1}{2}} \frac{k^a}{\Gamma(a)} \\
& \times \int \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{d}{2}+a+1} \exp\left(-\frac{1}{\sigma^2}\left[\frac{(y-\Phi\beta)^\top(y-\Phi\beta) + (\beta-\mu)^\top V^{-1}(\beta-\mu)}{2} + k\right]\right) \mathrm{d}\beta \mathrm{d}\sigma^2
\end{aligned}
$$

$$
\begin{aligned}
=& \left(\frac{1}{2\pi}\right)^{\frac{n+d}{2}} \left(\frac{1}{|V|}\right)^{\frac{1}{2}} \frac{k^a}{\Gamma(a)} \\
& \times \int \left[\left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{d}{2}+a+1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{1}{2}S + k\right)\right) \left[\int \exp\left(-\frac{1}{2}\frac{1}{\sigma^2}(\beta - v)^\top (V^*)^{-1} (\beta - \mu^*)\right) \mathrm{d}\beta\right]\right] \mathrm{d}\sigma^2
\end{aligned}
$$

$$
= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}+\frac{d}{2}} \left(\frac{1}{|V|}\right)^{\frac{1}{2}} \frac{k^a}{\Gamma(a)} \int \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}+\frac{d}{2}+a+1} \exp\left(-\frac{1}{\sigma^2}\left(\frac{1}{2}S + k\right)\right) \times \left[(2\pi)^{\frac{d}{2}} \left(\sigma^2\right)^{\frac{d}{2}} |V^*|^{\frac{1}{2}}\right] \mathrm{d}\sigma^2
$$

$$
= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{|V^*|}{|V|}\right)^{\frac{1}{2}} \frac{k^a}{\Gamma(a)} \frac{\Gamma\left(\frac{n}{2}+a\right)}{\left(\frac{1}{2}S + k\right)^{\frac{n}{2}+a}} = \quad \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left(\frac{|V^*|}{|V|}\right)^{\frac{1}{2}} \frac{k^a}{\Gamma(a)} \frac{\Gamma\left(\frac{n}{2}+a\right)}{(k^*)^{\frac{n}{2}+a}}
$$

Where

$$
S = \mu^\top V^{-1}\mu - (\mu^*)^\top (V^*)^{-1} (\mu^*) + y^\top y
$$

$$
k^* = k + \frac{1}{2}S; \qquad V^* = \left(V^{-1} + \Phi^\top\Phi\right)^{-1}; \qquad \mu^* = V^*\left(V^{-1}\mu + \Phi^\top y\right)
$$

For simplicity, we use the indexing $\cdot_0$ and $\cdot_1$ instead of $\cdot_{\mathscr{M}_0}$ and $\cdot_{\mathscr{M}_1}$ in what follows. So the Bayes factor is

$$
B_{01}(y) = \frac{f_0(y)}{f_1(y)} = \sqrt{\frac{|V_1|}{|V_0|}}\sqrt{\frac{|V_0^*|}{|V_1^*|}} \left(\frac{k_0^*}{k_1^*}\right)^{-\frac{n}{2}-a}
$$

and

$$
\mathsf{P}_\Pi(\mathrm{H}_0|y) = \left(1 + \frac{1-\pi_0}{\pi_0}\sqrt{\frac{|V_0|}{|V_1|}}\sqrt{\frac{|V_1^*|}{|V_0^*|}} \left(\frac{k_1^*}{k_0^*}\right)^{-\frac{n}{2}-a}\right)^{-1}
$$