LMU
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Patrick Schwaferts
Thomas Augustin

Updating Consistency in Bayes Factors

Technical Report Number 236, 2021
Department of Statistics
University of Munich

http://www.statistik.uni-muenchen.de

Updating Consistency in Bayes Factors

Patrick Schwaferts Thomas Augustin

Ludwig-Maximilians-Universität Munich
Department of Statistics
Methodological Foundations of Statistics and its Applications
Ludwigsstraße 33, 80539 Munich, Germany

Abstract

When it comes to extracting information from data by means of Bayes rule, it should not matter if all available data are considered at once or if Bayesian updating is performed subsequently with partitions of the data. This property is called updating consistency. However, in the context of Bayes factors, a prominent Bayesian tool that is used for comparing hypotheses, some researchers illustrated that updating consistency might not be given. Therefore, this technical report addresses the updating consistency of Bayes factors and shows its existence. In that, it serves as mathematical basis for the evaluation of the origin of putative updating inconsistencies. In addition, results about updating mixture priors are brought into the terminology commonly employed in the context of Bayes factors, as these were used in the elaboration about updating consistency. The depicted results imply that a necessary condition for updating consistency is to consider and report not only the Bayes factor value alone but also the posterior distributions as outcome of the analysis.

Keywords: Bayesian Statistics, Bayes Factor, Sequential Updating, Updating Consistency, Mixture Prior, Spike-and-Slab Prior

1 Introduction

Within the context of Bayesian statistics, the knowledge about a phenomenon of interest that is available prior to an investigation is typically formalized as a (subjective) prior probability distribution. Once a respective investigation has been performed, the obtained data are used to update this initial prior distribution via Bayes rule, yielding a posterior distribution. This posterior is said to reflect all relevant knowledge which is available after the investigation (see Figure 1, top-left). In that, the posterior distribution might be treated as prior distribution for a subsequent statistical analysis of a newly obtained data set (based on the same investigational setup). Naturally, the final posterior distribution after sequentially updating twice (see Figure 1, top-right) should be identical to the posterior distribution that is obtained by merging both data sets first and then updating the initial prior distribution at once (see Figure 1, bottom). If so, the Bayesian updating procedure is called "consistent" [cp. Rüger, 1998, p. 190].

However, the most prominent Bayesian method for hypothesis comparisons employed in psychological research - the so called Bayes factor [see e.g. Kass and Raftery, 1995, Gönen

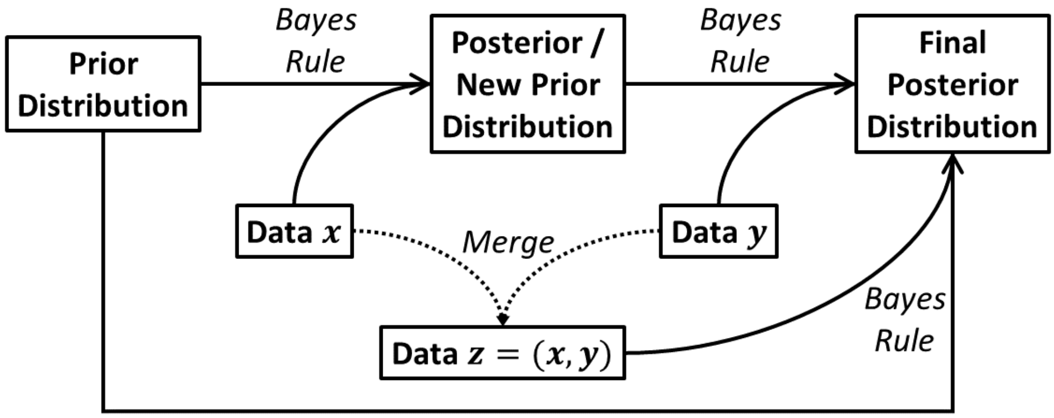

Figure 1: Consistent Bayesian Updating.

et al., 2005, Rouder et al., 2009] - might be characterized by an inconsistent Bayesian updating, as already indicated by Rouder and Morey [2011]. In contrast to specifying only one prior distribution that reflects the available knowledge prior to the investigation, an analysis with Bayes factors allows the specification of a prior distribution for each employed hypothesis, which is said to reflect its content [see e.g. Vanpaemel, 2010, Vanpaemel and Lee, 2012, Morey et al., 2016, Rouder et al., 2018a]. Although more than just a single prior distribution is employed, together with a distribution on the hypotheses themselves it is possible to merge all these hypothesis-based priors to an overall mixture distribution [see e.g. Rouder et al., 2018b]. By considering this mixture prior distribution, its updating might be assessed w.r.t. consistency, such that the origin of putative updating inconsistencies in the context of Bayes factors might be evaluated.

Accordingly, Bayes factors shall be outlined in Section 2 before considering the updating of the corresponding mixture prior in Section 3. These considerations are used to show that updating with Bayes factors is consistent (Section 4), but also that inconsistent updating might occur easily (Section 4.3). Implications about the minimal requirement of what is considered as outcome of an analysis with Bayes factors of a single data set are depicted in Section 5.

This technical report intends to depict the mathematical background of updating consistency in the context of Bayes factors in greater detail. Special emphasize will be given to explain mathematical transformations step by step with numerous references to previous definitions and equations. In addition, as all data, parameter, and hypotheses are random quantities, which are related to each other, Bayes rule is always applied meticulously, allowing clarity about which quantities are conditioned on.

2 Bayes Factors

Assume the observed data $\boldsymbol{x} = (x_1, \ldots, x_n)$ are modeled as realizations of independent and identically distributed (iid) random quantities $X_i \overset{iid}{\sim} P_{X_i|\theta}$ with parametric density $f(x_i|\theta)$

for all $i = 1, \dots, n$ and a parameter value $\theta \in \Theta$, such that $X \sim P_{X|\theta}$ with density

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \,. \tag{1}$$

Statistical hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{vs.} \quad H_1 : \theta \in \Theta_1 \tag{2}$$

contrast two subsets Θ_0 and Θ_1 of the parameter space Θ. Frequently, the null hypothesis is sharp and consists of only a single parameter value θ_0, i.e. $\Theta_0 = \{\theta_0\}$.

In the Bayesian setting, there is a prior distribution (P°_{θ}, see below) on the parameter θ. In the context of Bayes factors, however, a prior distribution is typically provided separately for each hypothesis: The prior distribution $P^{(1)}_{\theta}$ with density

$$\pi_1(\theta) := \pi(\theta|H_1) \tag{3}$$

is restricted to the (parameter subset specified within the) alternative hypothesis H_1, and the prior distribution $P^{(0)}_{\theta}$ with density

$$\pi_0(\theta) := \pi(\theta|H_0) \tag{4}$$

is restricted to the (parameter subset specified within the) null hypothesis H_0. If the null hypothesis is sharp, the corresponding prior distribution $P^{(0)}_{\theta}$ is degenerate with all probability mass on θ_0 .

In addition to $P^{(1)}_{\theta}$ and $P^{(0)}_{\theta}$, a prior distribution on the hypotheses themselves needs to be specified by

$$\rho := p(H_0) \quad \text{and} \quad p(H_1) = 1 - p(H_0) = 1 - \rho \,, \tag{5}$$

yielding the so called prior odds $p(H_1)/p(H_0)$.

The density of $P_{X|\theta}$ is assumed to be related to the hypotheses H_1 and H_0 only via the parameter value θ, i.e.

$$f(\boldsymbol{x}|H_1, \theta) = f(\boldsymbol{x}|H_0, \theta) = f(\boldsymbol{x}|\theta) \,. \tag{6}$$

The marginal density of the data $\boldsymbol{x}$ might be calculated w.r.t. each hypothesis

$$f(\boldsymbol{x}|H_1) \overset{marg.}{=} \int f(\boldsymbol{x}|H_1, \theta) \cdot \pi(\theta|H_1)\, d\theta \overset{\substack{eq.\\(6)\\(3)}}{=} \int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta \tag{7}$$

$$f(\boldsymbol{x}|H_0) \overset{marg.}{=} \int f(\boldsymbol{x}|H_0, \theta) \cdot \pi(\theta|H_0)\, d\theta \overset{\substack{eq.\\(6)\\(4)}}{=} \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta \tag{8}$$

and the Bayes factor based on data $\boldsymbol{x}$ w.r.t. the hypotheses H_0 and H_1 is defined as the ratio of these marginal densities

$$BF^{\boldsymbol{x}}_{10} := \frac{f(\boldsymbol{x}|H_1)}{f(\boldsymbol{x}|H_0)} \overset{\substack{eq.\\(7)\\(8)}}{=} \frac{\int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta}{\int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta} \,, \tag{9}$$

with a said interpretation of the data $\boldsymbol{x}$ being $BF_{10}^{\boldsymbol{x}}$ times as much evidence for H_1 than for H_0 [see e.g. Morey et al., 2016]. In that regard, consider the discussion in Section 5. Analogously, its inverse

$$BF_{01}^{\boldsymbol{x}} := \frac{1}{BF_{10}^{\boldsymbol{x}}} \tag{10}$$

should quantify the evidence within the data favoring H_0 over H_1.

The prior odds can be updated by the Bayes factor to the posterior odds

$$\frac{p(H_1|\boldsymbol{x})}{p(H_0|\boldsymbol{x})} \overset{\substack{Bayes\\rule}}{=} \frac{\frac{f(\boldsymbol{x}|H_1)\cdot p(H_1)}{f(\boldsymbol{x})}}{\frac{f(\boldsymbol{x}|H_0)\cdot p(H_0)}{f(\boldsymbol{x})}} \overset{\substack{eq.\\(9)}}{=} BF_{10}^{\boldsymbol{x}} \cdot \frac{p(H_1)}{p(H_0)}\,. \tag{11}$$

In that, the posterior probability of H_0 denoted by

$$\rho_{|\boldsymbol{x}} := p(H_0|\boldsymbol{x}) \tag{12}$$

can be calculated as

$$\begin{aligned}
&\frac{p(H_1|\boldsymbol{x})}{p(H_0|\boldsymbol{x})} = BF_{10}^{\boldsymbol{x}} \cdot \frac{p(H_1)}{p(H_0)} \quad \overset{\substack{eq.\\(5)\\(12)}}{\Leftrightarrow} \quad \frac{1-\rho_{|\boldsymbol{x}}}{\rho_{|\boldsymbol{x}}} = BF_{10}^{\boldsymbol{x}} \cdot \frac{1-\rho}{\rho} \\
\Leftrightarrow\quad & 1-\rho_{|\boldsymbol{x}} = BF_{10}^{\boldsymbol{x}} \cdot \frac{1-\rho}{\rho} \cdot \rho_{|\boldsymbol{x}} \quad \Leftrightarrow \quad 1 = BF_{10}^{\boldsymbol{x}} \cdot \frac{1-\rho}{\rho} \cdot \rho_{|\boldsymbol{x}} + \rho_{|\boldsymbol{x}} \\
\Leftrightarrow\quad & 1 = \rho_{|\boldsymbol{x}} \left[BF_{10}^{\boldsymbol{x}} \frac{1-\rho}{\rho} + 1 \right] \quad \Leftrightarrow \quad \rho_{|\boldsymbol{x}} = \frac{1}{BF_{10}^{\boldsymbol{x}} \frac{1-\rho}{\rho} + 1} \\
\Leftrightarrow\quad & \rho_{|\boldsymbol{x}} = \frac{\rho}{BF_{10}^{\boldsymbol{x}}(1-\rho)+\rho}\,.
\end{aligned} \tag{13}$$

3 Updating of Mixture Priors

Instead of treating the priors under both hypotheses separately, they might be merged to a single mixture prior distribution

$$P_\theta^\circ := \rho \cdot P_\theta^{(0)} + (1-\rho) \cdot P_\theta^{(1)}\,, \tag{14}$$

which has the density

$$\pi^\circ(\theta) = \rho \cdot \pi_0(\theta) + (1-\rho)\cdot \pi_1(\theta)\,. \tag{15}$$

With $P_\theta^{(0)}$ being degenerate this mixture prior is also referred to as spike-and-slab prior [see e.g. Rouder et al., 2018b], consisting of a spike-part $P_\theta^{(0)}$ and a slab-part $P_\theta^{(1)}$.

Theorem 1 (Updating of Mixture Priors)**.** *Updating the prior mixture distribution* P_θ° *using data* $\boldsymbol{x}$ *leads to the posterior distribution* $P_{\theta|\boldsymbol{x}}^\circ$ *with density*

$$\pi^\circ(\theta|\boldsymbol{x}) = \rho_{|\boldsymbol{x}} \cdot \pi_0(\theta|\boldsymbol{x}) + (1-\rho_{|\boldsymbol{x}}) \cdot \pi_1(\theta|\boldsymbol{x})\,, \tag{16}$$

where $\pi_0(\theta|\boldsymbol{x})$ *as well as* $\pi_1(\theta|\boldsymbol{x})$ *are posterior densities of* θ*, which arise from updating the prior densities* $\pi_0(\theta)$ *as well as* $\pi_1(\theta)$ *separately, i.e.*

$$\pi_0(\theta|\boldsymbol{x}) := \pi(\theta|H_0,\boldsymbol{x}) \overset{\substack{Bayes\\rule}}{=} \frac{f(\boldsymbol{x}|H_0,\theta)\cdot\pi(\theta|H_0)}{f(\boldsymbol{x}|H_0)} \overset{\substack{eq.\\(6)\\(4)\\(8)}}{=} \frac{f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)}{\int f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta}\,, \tag{17}$$

$$\pi_1(\theta|\boldsymbol{x}) := \pi(\theta|H_1, \boldsymbol{x}) \overset{\substack{Bayes\\rule}}{=} \frac{f(\boldsymbol{x}|H_1, \theta) \cdot \pi(\theta|H_1)}{f(\boldsymbol{x}|H_1)} \overset{\substack{eq.\\(6)\\(3)\\(7)}}{=} \frac{f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)}{\int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta} . \tag{18}$$

Proof. Before calculating the density

$$\pi^\circ(\theta|\boldsymbol{x}) \overset{\substack{Bayes\\rule}}{=} \frac{f(\boldsymbol{x}|\theta) \cdot \pi^\circ(\theta)}{f(\boldsymbol{x})} \tag{19}$$

of the posterior distribution $P^\circ_{\theta|\boldsymbol{x}}$ with

$$f(\boldsymbol{x}) \overset{marg.}{=} \int f(\boldsymbol{x}|\theta) \cdot \pi^\circ(\theta)\, d\theta\, , \tag{20}$$

consider the following first:

$$\begin{aligned}
f(\boldsymbol{x}) &\overset{\substack{eq.\\(20)}}{=} \int f(\boldsymbol{x}|\theta) \cdot \pi^\circ(\theta)\, d\theta \\
&\overset{\substack{eq.\\(15)}}{=} \int f(\boldsymbol{x}|\theta) \left[\rho \cdot \pi_0(\theta) + (1-\rho) \cdot \pi_1(\theta)\right] d\theta \\
&= \int \left[\rho \cdot f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\right] + \left[(1-\rho) \cdot f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\right] d\theta \\
&= \rho \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta + (1-\rho) \int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta
\end{aligned} \tag{21}$$

This can be transformed in two different ways:

$$\begin{aligned}
& f(\boldsymbol{x}) \overset{\substack{eq.\\(21)}}{=} \rho \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta + (1-\rho) \int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta \\
\Leftrightarrow \quad & f(\boldsymbol{x}) - \rho \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta = (1-\rho) \int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta \\
\Leftrightarrow \quad & \frac{f(\boldsymbol{x})}{\int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta} - \rho = (1-\rho) \frac{\int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta}{\int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta} \\
\overset{\substack{eq.\\(9)}}{\Leftrightarrow} \quad & \frac{f(\boldsymbol{x})}{\int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta} - \rho = (1-\rho) BF^{\boldsymbol{x}}_{10} \\
\Leftrightarrow \quad & f(\boldsymbol{x}) = \left[(1-\rho) BF^{\boldsymbol{x}}_{10} + \rho\right] \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta
\end{aligned} \tag{22}$$

or

$$\begin{aligned}
& f(\boldsymbol{x}) \overset{\substack{eq.\\(21)}}{=} \rho \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta + (1-\rho) \int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta \\
\Leftrightarrow \quad & f(\boldsymbol{x}) - (1-\rho) \int f(\boldsymbol{x}|\theta) \cdot \pi_1(\theta)\, d\theta = \rho \int f(\boldsymbol{x}|\theta) \cdot \pi_0(\theta)\, d\theta
\end{aligned}$$

$$\overset{}{\Leftrightarrow} \quad \frac{f(\boldsymbol{x})}{\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta} - (1-\rho) = \rho\,\frac{\int f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta}{\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}$$

$$\overset{\substack{eq.\\(9)\\(10)}}{\Leftrightarrow} \quad \frac{f(\boldsymbol{x})}{\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta} - (1-\rho) = \rho BF_{01}^{\boldsymbol{x}}$$

$$\Leftrightarrow \quad f(\boldsymbol{x}) = \left[\rho BF_{01}^{\boldsymbol{x}} + (1-\rho)\right]\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta\,. \tag{23}$$

In addition, consider

$$1-\rho_{|\boldsymbol{x}} \overset{\substack{eq.\\(13)}}{=} 1 - \frac{\rho}{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho} = \frac{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho-\rho}{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho} = \frac{(1-\rho)BF_{10}^{\boldsymbol{x}}}{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho}$$

$$= \frac{BF_{10}^{\boldsymbol{x}}(1-\rho)}{BF_{10}^{\boldsymbol{x}}\left[(1-\rho)+\frac{1}{BF_{10}^{\boldsymbol{x}}}\rho\right]} = \frac{(1-\rho)}{(1-\rho)+\frac{1}{BF_{10}^{\boldsymbol{x}}}\rho} \overset{\substack{eq.\\(10)}}{=} \frac{(1-\rho)}{(1-\rho)+BF_{01}^{\boldsymbol{x}}\rho}$$

$$= \frac{(1-\rho)}{\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)}\,. \tag{24}$$

Now, the posterior density $\pi^{\circ}(\theta|\boldsymbol{x})$ can be calculated as

$$\pi^{\circ}(\theta|\boldsymbol{x}) \overset{\substack{Bayes\\rule}}{=} \frac{f(\boldsymbol{x}|\theta)\cdot\pi^{\circ}(\theta)}{f(\boldsymbol{x})}$$

$$\overset{\substack{eq.\\(15)}}{=} \frac{f(\boldsymbol{x}|\theta)\left[\rho\cdot\pi_0(\theta)+(1-\rho)\cdot\pi_1(\theta)\right]}{f(\boldsymbol{x})}$$

$$= \rho\frac{f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)}{f(\boldsymbol{x})} + (1-\rho)\frac{f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)}{f(\boldsymbol{x})}$$

$$\overset{\substack{eq.\\(22)\\(23)}}{=} \frac{\rho}{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho}\cdot\frac{f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)}{\int f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta} + \frac{1-\rho}{\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)}\cdot\frac{f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)}{\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}$$

$$\overset{\substack{eq.\\(17)\\(18)}}{=} \frac{\rho}{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho}\,\pi_0(\theta|\boldsymbol{x}) + \frac{1-\rho}{\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)}\,\pi_1(\theta|\boldsymbol{x})$$

$$\overset{\substack{eq.\\(13)\\(24)}}{=} \rho_{|\boldsymbol{x}}\cdot\pi_0(\theta|\boldsymbol{x}) + (1-\rho_{|\boldsymbol{x}})\cdot\pi_1(\theta|\boldsymbol{x})\,,$$

□

Certainly, this is not a new result as e.g. Mitchell and Beauchamp [1988] already employed spike-and-slab priors (which are a special case of mixture priors) and e.g. Rouder et al.

[2018b] depicted the priors in the context of Bayes factors by an overall spike-and-slab prior. However, these considerations explicitly utilize a notation typically employed in analyses with Bayes factors and are needed for further elaboration on the updating consistency of Bayes factors.

4 Updating Consistency

4.1 Framework

In order to assess Bayes factors w.r.t. updating consistency two different data sets are necessary. Accordingly, in addition to $\boldsymbol{x}$ and X as in Section 2, consider a second data set $\boldsymbol{y} = (y_1, \ldots, y_m)$ being independent of the previous one and modeled analogously, i.e. $Y_j \overset{iid}{\sim} P_{Y_j|\theta}$ with the same parametric density $f(y_j|\theta)$ for all $j = 1, \ldots, m$. Therefore, $Y \sim P_{Y|\theta}$ with density

$$f(\boldsymbol{y}|\theta) = \prod_{j=1}^{m} f(y_j|\theta)\,. \tag{25}$$

Analogue to equation (6), the density of $P_{Y|\theta}$ is also assumed to be related to the hypotheses H_1 and H_0 only via the parameter value θ, i.e.

$$f(\boldsymbol{y}|H_1, \theta) = f(\boldsymbol{y}|H_0, \theta) = f(\boldsymbol{y}|\theta)\,. \tag{26}$$

Define $Z := (X, Y)$ and $\boldsymbol{z} := (\boldsymbol{x}, \boldsymbol{y})$. As

$$f(\boldsymbol{z}|\theta) \overset{\substack{X,Y \\ ind.}}{=} f(\boldsymbol{y}|\theta) \cdot f(\boldsymbol{x}|\theta)\,, \tag{27}$$

the density of $\boldsymbol{z}$ is related to the hypotheses only via the parameter value θ as well:

$$f(\boldsymbol{z}|H_1, \theta) = f(\boldsymbol{z}|H_0, \theta) = f(\boldsymbol{z}|\theta)\,. \tag{28}$$

Analogue to the marginal densities of $\boldsymbol{x}$ (equations (7) and (8)), those of $\boldsymbol{y}$ and $\boldsymbol{z}$ are calculated as

$$f(\boldsymbol{y}|H_1) \overset{marg.}{=} \int f(\boldsymbol{y}|H_1, \theta) \cdot \pi(\theta|H_1)\, d\theta \overset{\substack{eq. \\ (26) \\ (3)}}{=} \int f(\boldsymbol{y}|\theta) \cdot \pi_1(\theta)\, d\theta \tag{29}$$

$$f(\boldsymbol{y}|H_0) \overset{marg.}{=} \int f(\boldsymbol{y}|H_0, \theta) \cdot \pi(\theta|H_0)\, d\theta \overset{\substack{eq. \\ (26) \\ (4)}}{=} \int f(\boldsymbol{y}|\theta) \cdot \pi_0(\theta)\, d\theta \tag{30}$$

$$f(\boldsymbol{z}|H_1) \overset{marg.}{=} \int f(\boldsymbol{z}|H_1, \theta) \cdot \pi(\theta|H_1)\, d\theta \overset{\substack{eq. \\ (28) \\ (3)}}{=} \int f(\boldsymbol{z}|\theta) \cdot \pi_1(\theta)\, d\theta \tag{31}$$

$$f(\boldsymbol{z}|H_0) \overset{marg.}{=} \int f(\boldsymbol{z}|H_0, \theta) \cdot \pi(\theta|H_0)\, d\theta \overset{\substack{eq. \\ (28) \\ (4)}}{=} \int f(\boldsymbol{z}|\theta) \cdot \pi_0(\theta)\, d\theta\,. \tag{32}$$

and the marginal densities of $\boldsymbol{y}$ w.r.t. to the posterior distributions of θ given the first data set $\boldsymbol{x}$ are

$$\begin{aligned} f(\boldsymbol{y}|H_1,\boldsymbol{x}) &\overset{marg.}{=} \int f(\boldsymbol{y}|H_1,\theta,\boldsymbol{x}) \cdot \pi(\theta|H_1,\boldsymbol{x})\, d\theta \\ &\overset{\substack{X,Y\\ ind.}}{=} \int f(\boldsymbol{y}|H_1,\theta) \cdot \pi(\theta|H_1,\boldsymbol{x})\, d\theta \\ &\overset{\substack{eq.\\(26)\\(18)}}{=} \int f(\boldsymbol{y}|\theta) \cdot \pi_1(\theta|\boldsymbol{x})\, d\theta \end{aligned} \tag{33}$$

$$\begin{aligned} f(\boldsymbol{y}|H_0,\boldsymbol{x}) &\overset{marg.}{=} \int f(\boldsymbol{y}|H_0,\theta,\boldsymbol{x}) \cdot \pi(\theta|H_0,\boldsymbol{x})\, d\theta \\ &\overset{\substack{X,Y\\ ind.}}{=} \int f(\boldsymbol{y}|H_0,\theta) \cdot \pi(\theta|H_0,\boldsymbol{x})\, d\theta \\ &\overset{\substack{eq.\\(26)\\(17)}}{=} \int f(\boldsymbol{y}|\theta) \cdot \pi_0(\theta|\boldsymbol{x})\, d\theta\,. \end{aligned} \tag{34}$$

The corresponding Bayes factor values are

$$BF_{10}^{\boldsymbol{y}} := \frac{f(\boldsymbol{y}|H_1)}{f(\boldsymbol{y}|H_0)} \overset{\substack{eq.\\(29)\\(30)}}{=} \frac{\int f(\boldsymbol{y}|\theta)\cdot\pi_1(\theta)\, d\theta}{\int f(\boldsymbol{y}|\theta)\cdot\pi_0(\theta)\, d\theta} \tag{35}$$

$$BF_{10}^{\boldsymbol{z}} := \frac{f(\boldsymbol{z}|H_1)}{f(\boldsymbol{z}|H_0)} \overset{\substack{eq.\\(31)\\(32)}}{=} \frac{\int f(\boldsymbol{z}|\theta)\cdot\pi_1(\theta)\, d\theta}{\int f(\boldsymbol{z}|\theta)\cdot\pi_0(\theta)\, d\theta} \tag{36}$$

$$BF_{10}^{\boldsymbol{y}|\boldsymbol{x}} := \frac{f(\boldsymbol{y}|H_1,\boldsymbol{x})}{f(\boldsymbol{y}|H_0,\boldsymbol{x})} \overset{\substack{eq.\\(33)\\(34)}}{=} \frac{\int f(\boldsymbol{y}|\theta)\cdot\pi_1(\theta|\boldsymbol{x})\, d\theta}{\int f(\boldsymbol{y}|\theta)\cdot\pi_0(\theta|\boldsymbol{x})\, d\theta}\,. \tag{37}$$

4.2 Consistent Updating

Theorem 2 (Subsequent Updating with Bayes Factors)**.** *Based on the framework above, updating the prior odds $p(H_1)/p(H_0)$ using both $\boldsymbol{x}$ and $\boldsymbol{y}$ subsequently yields the posterior odds*

$$\frac{p(H_1|\boldsymbol{y},\boldsymbol{x})}{p(H_0|\boldsymbol{y},\boldsymbol{x})} = BF_{10}^{\boldsymbol{y}|\boldsymbol{x}} \cdot BF_{10}^{\boldsymbol{x}} \cdot \frac{p(H_1)}{p(H_0)}\,. \tag{38}$$

Proof.

$$\begin{aligned} \frac{p(H_1|\boldsymbol{y},\boldsymbol{x})}{p(H_0|\boldsymbol{y},\boldsymbol{x})} &\overset{\substack{Bayes\\ rule}}{=} \frac{f(\boldsymbol{y}|H_1,\boldsymbol{x})}{f(\boldsymbol{y}|H_0,\boldsymbol{x})}\frac{p(H_1|\boldsymbol{x})}{p(H_0|\boldsymbol{x})} \\ &\overset{\substack{Bayes\\ rule}}{=} \frac{f(\boldsymbol{y}|H_1,\boldsymbol{x})}{f(\boldsymbol{y}|H_0,\boldsymbol{x})}\frac{f(\boldsymbol{x}|H_1)}{f(\boldsymbol{x}|H_0)}\frac{p(H_1)}{p(H_0)} \end{aligned}$$

$$\overset{\substack{eq.\\(37)\\(9)}}{=} BF_{10}^{\boldsymbol{y}|\boldsymbol{x}} \cdot BF_{10}^{\boldsymbol{x}} \cdot \frac{p(H_1)}{p(H_0)} .$$

□

Theorem 3 (Consistent Updating with Bayes Factors)**.** *Based on the framework above, updating the prior odds $p(H_1)/p(H_0)$ with the corresponding Bayes factor values is consistent, i.e.*

$$\frac{p(H_1|\boldsymbol{z})}{p(H_0|\boldsymbol{z})} = \frac{p(H_1|\boldsymbol{y},\boldsymbol{x})}{p(H_0|\boldsymbol{y},\boldsymbol{x})} . \tag{39}$$

Proof. At first, consider

$$\begin{aligned} BF_{10}^{\boldsymbol{x}} &= BF_{10}^{\boldsymbol{x}} \frac{\rho + (1-\rho)BF_{10}^{\boldsymbol{x}}}{\rho + (1-\rho)BF_{10}^{\boldsymbol{x}}} = \frac{BF_{10}^{\boldsymbol{x}}\left[\rho + (1-\rho)BF_{10}^{\boldsymbol{x}}\right]}{BF_{10}^{\boldsymbol{x}}\left[\frac{\rho}{BF_{10}^{\boldsymbol{x}}} + (1-\rho)\right]} = \frac{\rho + (1-\rho)BF_{10}^{\boldsymbol{x}}}{\frac{\rho}{BF_{10}^{\boldsymbol{x}}} + (1-\rho)} \\ &\overset{\substack{eq.\\(10)}}{=} \frac{(1-\rho)BF_{10}^{\boldsymbol{x}} + \rho}{\rho\, BF_{01}^{\boldsymbol{x}} + (1-\rho)} . \end{aligned} \tag{40}$$

Now, the Bayes factor value $BF_{10}^{\boldsymbol{z}}$ might be decomposed:

$$\begin{aligned} BF_{10}^{\boldsymbol{z}} &\overset{\substack{eq.\\(36)}}{=} \frac{\int f(\boldsymbol{z}|\theta)\cdot\pi_1(\theta)\,d\theta}{\int f(\boldsymbol{z}|\theta)\cdot\pi_0(\theta)\,d\theta} \\ &\overset{\substack{X,Y\\ind.}}{=} \frac{\int f(\boldsymbol{y}|\theta)\cdot f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}{\int f(\boldsymbol{y}|\theta)\cdot f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta} \\ &= \frac{\frac{1}{f(\boldsymbol{x})}\int f(\boldsymbol{y}|\theta)\cdot f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}{\frac{1}{f(\boldsymbol{x})}\int f(\boldsymbol{y}|\theta)\cdot f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta} \\ &\overset{\substack{eq.\\(22)\\(23)}}{=} \frac{\frac{1}{\left[\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)\right]\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}\int f(\boldsymbol{y}|\theta)\cdot f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}{\frac{1}{\left[(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho\right]\int f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta}\int f(\boldsymbol{y}|\theta)\cdot f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta} \\ &= \frac{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho}{\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)} \cdot \frac{\int f(\boldsymbol{y}|\theta)\cdot\frac{f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)}{\int f(\boldsymbol{x}|\theta)\cdot\pi_1(\theta)\,d\theta}\,d\theta}{\int f(\boldsymbol{y}|\theta)\cdot\frac{f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)}{\int f(\boldsymbol{x}|\theta)\cdot\pi_0(\theta)\,d\theta}\,d\theta} \\ &\overset{\substack{eq.\\(17)\\(18)}}{=} \frac{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho}{\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)} \cdot \frac{\int f(\boldsymbol{y}|\theta)\cdot\pi_1(\theta|\boldsymbol{x})\,d\theta}{\int f(\boldsymbol{y}|\theta)\cdot\pi_0(\theta|\boldsymbol{x})\,d\theta} \\ &\overset{\substack{eq.\\(37)}}{=} \frac{(1-\rho)BF_{10}^{\boldsymbol{x}}+\rho}{\rho BF_{01}^{\boldsymbol{x}}+(1-\rho)} \cdot BF_{10}^{\boldsymbol{y}|\boldsymbol{x}} \\ &\overset{\substack{eq.\\(40)}}{=} BF_{10}^{\boldsymbol{x}} \cdot BF_{10}^{\boldsymbol{y}|\boldsymbol{x}} . \end{aligned} \tag{41}$$

Therefore:

$$\frac{p(H_1|\boldsymbol{z})}{p(H_0|\boldsymbol{z})} \overset{\substack{Bayes\\ rule}}{=} \frac{f(\boldsymbol{z}|H_1)}{f(\boldsymbol{z}|H_0)} \cdot \frac{p(H_1)}{p(H_0)} \overset{\substack{eq.\\ (36)}}{=} BF_{10}^{\boldsymbol{z}} \cdot \frac{p(H_1)}{p(H_0)}$$

$$\overset{\substack{eq.\\ (41)}}{=} BF_{10}^{\boldsymbol{y}|\boldsymbol{x}} \cdot BF_{10}^{\boldsymbol{x}} \cdot \frac{p(H_1)}{p(H_0)} \overset{\substack{eq.\\ (38)}}{=} \frac{p(H_1|\boldsymbol{y}, \boldsymbol{x})}{p(H_0|\boldsymbol{y}, \boldsymbol{x})} .$$

□

4.3 Inconsistent Updating

Remark that in order to update consistently with Bayes factors, the Bayes factor value $BF_{10}^{\boldsymbol{y}|\boldsymbol{x}}$ of the second data set $\boldsymbol{y}$ need to be based on the posterior distributions $\pi_1(\theta|\boldsymbol{x})$ and $\pi_0(\theta|\boldsymbol{x})$ that incorporate the information of the previous data set $\boldsymbol{x}$.

However, using $BF_{10}^{\boldsymbol{y}}$ instead of $BF_{10}^{\boldsymbol{y}|\boldsymbol{x}}$ is erroneous and yields odds

$$\frac{p(H_1|\boldsymbol{y}, \boldsymbol{x}^{(!)})}{p(H_0|\boldsymbol{y}, \boldsymbol{x}^{(!)})} := BF_{10}^{\boldsymbol{y}} \cdot BF_{10}^{\boldsymbol{x}} \cdot \frac{p(H_1)}{p(H_0)} , \tag{42}$$

which are in general different to the posterior odds obtained by updating consistently, i.e.

$$\frac{p(H_1|\boldsymbol{y}, \boldsymbol{x}^{(!)})}{p(H_0|\boldsymbol{y}, \boldsymbol{x}^{(!)})} \overset{\substack{in\\ gen.}}{\neq} \frac{p(H_1|\boldsymbol{y}, \boldsymbol{x})}{p(H_0|\boldsymbol{y}, \boldsymbol{x})} . \tag{43}$$

This is due to ignoring the information about θ within the first data set $\boldsymbol{x}$ while calculating the Bayes factor value based on the second data set $\boldsymbol{y}$, and the superscript (!) indicates this loss of information.

Bayes factor updating inconsistencies might occur e.g. in the following scenario: Two different research teams are interested in the same research question and utilize the same hypotheses and employ the same prior distributions on the parameter of interest. Both teams conduct a scientific investigation with identical design and calculate a Bayes factor value independently of each other. As each value is said to describe the change in belief within the hypotheses, it is tempting (e.g. in a meta-analysis of both investigations) to utilize both Bayes factor values to calculate the final belief (posterior odds) within the initially stated hypotheses. This, however, is exactly the error displayed in equation (42).

5 Outcome of Analyses with Bayes Factors

In order to avoid updating inconsistencies, both the Bayes factor value $BF_{10}^{\boldsymbol{x}}$ and the posterior distributions $\pi_1(\theta|\boldsymbol{x})$ as well as $\pi_0(\theta|\boldsymbol{x})$ are required to perform the analysis (with Bayes factors) of the second data set $\boldsymbol{y}$ once the first data set $\boldsymbol{x}$ is available.

Accordingly, considering solely the Bayes factor value $BF_{10}^{\boldsymbol{x}}$ as the outcome of the first analysis (of data $\boldsymbol{x}$) is not sufficient. Also the updated posterior distributions $\pi_1(\theta|\boldsymbol{x})$ and $\pi_0(\theta|\boldsymbol{x})$ need to be considered and reported. This appears to be obvious in the face of the posterior mixture distribution described in theorem 1, which cannot be described by the Bayes factor value $BF_{10}^{\boldsymbol{x}}$ alone. This is summarized in the following theorem.

Theorem 4 (Outcome of Analyses with Bayes Factors)**.** *A necessary condition for updating consistency in Bayes factors is to consider and report both the Bayes factor value* $BF_{10}^{\boldsymbol{x}}$ *and the posterior distributions* $\pi_1(\theta|\boldsymbol{x})$ *as well as* $\pi_0(\theta|\boldsymbol{x})$ *as outcomes of the analysis (of the data set* $\boldsymbol{x}$*).*

These considerations about updating inconsistency in Bayes factors might also be relevant e.g. in the following case: It is argued that, in the context of Bayes factors, the shape of the prior distributions ($\pi_1(\theta)$ and $\pi_0(\theta)$) reflects the content of the hypotheses [see e.g. Vanpaemel, 2010, Vanpaemel and Lee, 2012, Morey et al., 2016, Rouder et al., 2018a], but by incorporating the information of the data $\boldsymbol{x}$ into these distributions by means of Bayes rule, these distributions change to $\pi_1(\theta|\boldsymbol{x})$ and $\pi_0(\theta|\boldsymbol{x})$, which might then reflect different contents. Although erroneous, it is tempting to treat the Bayes factor value $BF_{10}^{\boldsymbol{x}}$ as quantification of the evidence within the data $\boldsymbol{x}$ w.r.t. to the hypotheses that are described by the initial prior distributions $\pi_1(\theta)$ and $\pi_0(\theta)$, as these hypotheses were formulated to answer the research question of interest. By doing so, the change within the distributions of θ is discarded and inconsistent updating might occur.

6 Summary

With theorem 1 results about updating mixture distributions are brought into the notation typically involved in the context of Bayes factors. Theorem 2 describes the final posterior odds after considering two separate data sets subsequently and theorem 3 argues that this updating procedure is consistent. As elaborated in Section 4.3, updating inconsistencies occur by discarding information and an exemplary situation was provided, in which this might happen unintentionally. Theorem 4 provides a minimum requirement on what to consider and report as outcome of a statistical analysis with Bayes factors, and a context in which this might oppose other recommendations about Bayes factors was illustrated in Section 5. However, a thorough discussion of the occurrence and consequences of updating inconsistencies in applied Bayes factors was not intended within this technical report and is still pending. Yet, this report enables this discussion by providing the necessary mathematical foundations.

References

M. Gönen, W. O. Johnson, Y. Lu, and P. H. Westfall. The Bayesian two-sample t test. *The American Statistician*, 59(3):252–257, 2005.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

T. J. Mitchell and J. J. Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

R. D. Morey, J.-W. Romeijn, and J. N. Rouder. The philosophy of Bayes factors and the quantification of statistical evidence. *Journal of Mathematical Psychology*, 72:6–18, 2016.

J. N. Rouder and R. D. Morey. A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, 18(4):682–689, 2011.

J. N. Rouder, P. L. Speckman, D. Sun, R. D. Morey, and G. Iverson. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2): 225–237, 2009.

J. N. Rouder, J. M. Haaf, and F. Aust. From theories to models to predictions: A Bayesian model comparison approach. *Communication Monographs*, 85(1):41–56, 2018a.

J. N. Rouder, J. M. Haaf, and J. Vandekerckhove. Bayesian inference for psychology, part iv: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, 25(1): 102–113, 2018b.

B. Rüger. *Test-und Schätztheorie: Band I: Grundlagen.* De Gruyter Oldenbourg, 1998.

W. Vanpaemel. Prior sensitivity in theory testing: An apologia for the Bayes factor. *Journal of Mathematical Psychology*, 54(6):491–498, 2010.

W. Vanpaemel and M. D. Lee. Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, 19(6):1047–1056, 2012.