Proceedings of Machine Learning Research 103:338–345, 2019 ISIPTA 2019

Imprecise Hypothesis-Based Bayesian Decision Making With Simple Hypotheses

Patrick Schwaferts PATRICK.SCHWAFERTS@STAT.UNI-MUENCHEN.DE
Thomas Augustin THOMAS.AUGUSTIN@STAT.UNI-MUENCHEN.DE
Institut für Statistik, Ludwig-Maximilians Universität München (LMU), Munich, Germany

Abstract

Applied real-world decisions are frequently guided by the outcome of hypothesis-based statistical analyses. However, most often relevant information about the phenomenon of interest is available only imprecisely, and misleading results might be obtained, in particular, by either ignoring relevant information or pretending a level of knowledge that is not given. In order to be able to include (partial) information authentically in the imprecise form it is available, this paper tries to extend the framework of hypothesis-based Bayesian decision making with simple hypotheses to be able to deal with imprecise information about the three relevant quantities: hypotheses, prior beliefs, and loss function. Although straightforward at first glance, it appears that by specifying the hypotheses imprecisely, Bayesian updating of the prior beliefs might be inconsistent. In that, this paper provides the basic mathematical formulation to further extend imprecise hypothesis-based Bayesian decision theory to more elaborate contexts, such as those involving composite imprecise hypotheses, and in addition highlights the necessity of paying particular attention to the depicted updating issues.

Keywords: Hypotheses, Likelihood Ratio, Imprecise Probabilities, Bayesian Decision Theory, Sequential Updating, Inconsistency, Statistics in Psychological Research

1. Introduction

In the face of the currently discussed reproducibility crisis in psychological research (Ioannidis, 2005), Bayesian statistics is gaining popularity (e.g. Van De Schoot et al., 2017) also in this area. Classical hypotheses tests are argued to be replaced by the so called Bayes factor (e.g. Kass and Raftery, 1995; Gönen et al., 2005; Rouder et al., 2009), a Bayesian quantity for hypothesis comparisons, which might be seen as a generalization of the likelihood ratio to include prior information about the parameter of interest by employing prior distributions on it. If these distributions are degenerate, i.e. have all mass on a single parameter value, the Bayes factor equals the likelihood ratio.

In addition to the prior distributions on the parameter, a Bayesian analysis in the context of statistical hypotheses requires prior probabilities of these hypotheses, which might be interpreted as subjective *belief* in the respective hypotheses and get updated by the data. It is the Bayes factor, which quantifies the change in these subjective probabilities (e.g. Morey et al., 2016), and therefore the Bayes factor is interpreted as quantification of the *evidence* in the data w.r.t. the hypotheses. The posterior probabilities of the hypotheses might then be used to guide a *decision* together with an appropriately specified loss function in the context of Bayesian decision theory (see e.g. Berger, 1995; Huntley et al., 2014).

In that, the changing focus onto Bayesian statistics within psychological research might be seen as a step towards rising awareness of the distinction between evidence, belief and decision in the context of an analysis of statistical hypotheses (see e.g. Lavine and Schervish (1999) and especially Royall (2004)).

Naturally, statistical hypotheses depend on the real-world research question, which might not always be unambiguously formalized mathematically. Prior probabilities of the hypotheses are subjective in nature and only rarely accessible as precise numerical values. The loss function depends on a putative real-world decision problem such that a precise specification of the loss function might not be given by the researcher.

Yet, certain potentially incomplete information about hypotheses, prior beliefs and the loss function might be available, such that both ignoring these information or specifying the respective quantities in an overly precise way might yield misleading results or decisions. In order to avoid untrustworthy results, it is thus necessary to allow researchers to include information into a statistical analysis specifically in the imprecise form it is available. Therefore, this paper intends to formulate the simplest case (using simple hypotheses) of hypothesis-based Bayesian decision theory in a way to include partial information about hypotheses, prior beliefs and the loss function. This might be seen as a fundamental, but necessary step to extend the imprecise probability framework (Walley, 1991) to the Bayes factor analyses that are recently applied in psychological research, working at the interface between statistical developments and empirical sciences.

As mentioned above, there are two different types of prior distributions inherent to a Bayes factor analysis:

(hypothesis-based) priors on the parameters and a prior on the hypotheses that is used to further guide decisions. In that, the Bayes factor analysis might be generalized within the framework of imprecise probabilities at these two distinct parts. Allowing the priors on the parameter to be specified imprecisely is restricted to the Bayes factor analysis itself and given an account by Ebner et al. (2019). However, considerations about allowing imprecise priors on the hypotheses and imprecise quantities relevant for a corresponding decision might apply to more situations than typically addressed in a Bayes factor analysis. Therefore, it will be given a separate account within this paper and discussed by referring to the likelihood ratio, which might be seen as both special case (using degenerate priors on the parameters) and foundational basis (see e.g. Royall, 2004) of the Bayes factor.

The present paper is structured as follows. Section 2 collects the basic ingredients of the classical case of Bayesian decision making based on two precise simple hypotheses. This framework will in Section 3 be powerfully extended to the situation where single hypotheses are interval-valued and the loss functions and prior odds are imprecise. Section 4 warns that, however, in this context some inconsistency issues may arise under updating and assess them in greater detail. Section 5 provides a numerical example as illustration and Section 6 concludes with a brief outlook.

2. Precise Hypothesis-Based Bayesian Decision Making

Assume a parametric statistical model, such that observed data $x = (x_1, \ldots, x_n)$ are modeled as realizations of independent and identically distributed random variables X_i, $i = 1, \ldots, n$, with parametric probability density $f(x_i|\theta)$, $\theta \in \mathscr{D}_\theta$, which specifies the joint density as

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

All considered parameter vales θ are comprised within the parameter space $\mathscr{D}_\theta$ and, for the sake of simplicity (especially w.r.t. notation), the parameter is assumed to be a single real-valued scalar here. Generalizations to multidimensional parameters are possible, but are left to further research.

Further assume two precise simple hypotheses

$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1\,, \tag{2}$$

where θ_0 and θ_1 are precise hypothesized parameter values, which implies that one of these two values is considered to be true. In a Bayesian context there is a subjective prior distribution on the hypotheses ($p(H_0)$ and $p(H_1) = 1 - p(H_0)$), forming the prior odds

$$\pi := \frac{p(H_0)}{p(H_1)}\,. \tag{3}$$

The prior odds can be updated by the observed data x via Bayes rule to the posterior odds

$$\frac{p(H_0|x)}{p(H_1|x)} = \frac{\frac{f(x|\theta_0)\cdot p(H_0)}{f(x)}}{\frac{f(x|\theta_1)\cdot p(H_1)}{f(x)}} = LR^x(\theta_0, \theta_1) \cdot \pi\,, \tag{4}$$

where

$$LR^x(\theta_0, \theta_1) = \frac{f(x|\theta_0)}{f(x|\theta_1)} \tag{5}$$

is the likelihood ratio and frequently referred to as Bayes factor (see e.g. Liu and Aitkin, 2008), as both hypotheses in equation (2) might be formulated by degenerate probability distributions with all probability mass on θ_0 and θ_1, respectively.

In order to guide a decision between two actions a_0 and a_1, a loss function

$$\begin{aligned} L : \mathscr{H} \times \mathscr{A} &\to \mathbb{R}_0^+ \\ (H, a) &\mapsto L(H, a) \end{aligned} \tag{6}$$

with $\mathscr{H} = \{H_0, H_1\}$ and $\mathscr{A} = \{a_0, a_1\}$ need to be specified, quantifying the "badness" of choosing a if H is true. The expected posterior loss

$$\begin{aligned} \rho : \mathscr{A} &\to \mathbb{R}_0^+ \\ a &\mapsto p(H_0|x)L(H_0, a) + p(H_1|x)L(H_1, a) \end{aligned} \tag{7}$$

can be used to find the optimal action(s)

$$a^* = \underset{a \in \mathscr{A}}{\operatorname{argmin}}\, \rho(a)\,. \tag{8}$$

Assume that, as is common practice in empirical research, the decision problem is formulated in regret form, where a_0 is associated with H_0 and a_1 with H_1 such that the correct decisions are evaluated to have zero loss, i.e. $L(H_0, a_0) = L(H_1, a_1) = 0$. Then it is only necessary to specify the ratio

$$k := \frac{L(H_0, a_1)}{L(H_1, a_0)} \tag{9}$$

in order to calculate the ratio of expected posterior losses

$$r := \frac{\rho(a_1)}{\rho(a_0)} = \pi \cdot LR^x(\theta_0, \theta_1) \cdot k \tag{10}$$

to determine

$$a^* = \begin{cases} a_0 & \text{if } r > 1 \\ a_1 & \text{if } r < 1 \end{cases}\,. \tag{11}$$

For $r = 1$ any action might be chosen.

3. Imprecise Hypothesis-Based Bayesian Decision Making

Within applied research, it is typically extremely difficult to specify the quantities θ_0, θ_1, π and k, which are necessary to determine a^*, as precise values. This is due to the fact that commonly some (potentially imprecise) information is available and several choices of precise values for these quantities are in accordance with it. Both ignoring the available relevant information and arbitrarily choosing among those plausible values, can hardly be an optimal strategy. Therefore, these quantities shall be specified imprecisely as an interval of values. Following Dubois' distinction (cp. Dubois, 1986, Section 1.4), these intervals have to be interpreted as conjunctive sets: they must be treated as a generalization of a single value and thus as an entity of its own. In that, as the interval of values replaces the respective precise value, the distribution is parametrically constructed by an interval (e.g. Augustin et al., 2014, Section 7.3.2). Also note that all four quantities θ_0, θ_1, π and k might be specifiable independently of each other, which allows subsequent calculations to be straightforward.

3.1. Imprecise Simple Hypotheses

Instead of a precise parameter value θ, the (imprecise) density of the data x is now dependent on an imprecise interval-valued parameter $\Theta = [\underline{\Theta}, \overline{\Theta}]$, i.e.

$$f(x|\Theta) = \{f(x|\theta)|\theta \in \Theta\} \tag{12}$$

with $\underline{\Theta}$ and $\overline{\Theta}$ being precise valued bounds on the parameter that are considered as defining the imprecise parameter Θ. Accordingly, the parameter space of Θ is now the set of all closed parameter intervals

$$\mathscr{D}_\Theta = \{[\underline{\Theta}, \overline{\Theta}]|\underline{\Theta} \in \mathscr{D}_\theta, \overline{\Theta} \in \mathscr{D}_\theta, \underline{\Theta} \leq \overline{\Theta}\}\,. \tag{13}$$

Consider imprecise, but simple hypotheses

$$H_0 : \Theta = \Theta_0 \quad \text{vs.} \quad H_1 : \Theta = \Theta_1\,, \tag{14}$$

where

$$\Theta_0 = [\underline{\Theta_0}, \overline{\Theta_0}]\,, \tag{15}$$

$$\Theta_1 = [\underline{\Theta_1}, \overline{\Theta_1}] \tag{16}$$

and $\underline{\Theta_0}$ (or $\underline{\Theta_1}$) is the lower bound as well as $\overline{\Theta_0}$ (or $\overline{\Theta_1}$) the upper bound for the simple hypothesized parameter value Θ under H_0 (or H_1). Although specified as intervals within this paper, simple imprecise hypotheses might also be generalized to hypothesize (convex) sets of parameters in general. Note that these hypotheses are not composite, as they consist only of one single, but imprecisely specified value. In contrast, composite hypotheses, for instance specified as

$$H_0 : \theta \in [\underline{\Theta_0}, \overline{\Theta_0}] \quad \text{vs.} \quad H_1 : \theta \in [\underline{\Theta_1}, \overline{\Theta_1}]\,, \tag{17}$$

would contain all precise parameter values within the respective intervals. That is exactly the crucial difference in interpreting composite and simple imprecise hypotheses. While the latter states that there is only one single parameter value which represents the hypothesis, yet there is not enough information available to precisely specify this single value, the former states that all the different parameter values, as a whole, represent the hypothesis. In that, composite hypotheses bound the unknown parameter value of a precise sampling model, while an imprecise parameter specifies an imprecise sampling model (e.g. Augustin et al., 2014, Section 7.2.5). As an outlook, composite imprecise hypotheses would be subsets of $\mathscr{D}_\Theta$ containing more than one parameter interval.

The Bayesian account to composite hypotheses is to employ a prior distribution on the hypothesized values and to calculate the respective marginal density of the observed data (as in a typical Bayes factor analysis (e.g. Morey et al., 2016)). While this prior is on the *parameter values* themselves within a precise composite hypothesis, it is on *parameter intervals* within an imprecise composite hypothesis. A simple imprecise hypothesis might therefore be described by a degenerate distribution with all mass on the respective parameter interval.

Accordingly, the fundamental technical difference between precise composite and simple imprecise hypotheses within the Bayesian framework is that only former requires the specification of a prior distribution on the hypothesized parameter values. In that, former might be incorporated within the Bayesian analysis by means of a marginal density and latter by means of the imprecise-valued density as in equation (12).

3.2. Imprecise Likelihood Ratio, Imprecise Prior Odds, and Imprecise Loss Function

Given data x, instead of a precise likelihood ratio, there is an interval-valued likelihood ratio

$$LR^x = [\underline{LR^x}, \overline{LR^x}]\,, \tag{18}$$

with

$$\underline{LR^x} = \min_{\substack{\theta_0 \in \Theta_0 \\ \theta_1 \in \Theta_1}} LR^x(\theta_0, \theta_1)\,, \tag{19}$$

$$\overline{LR^x} = \max_{\substack{\theta_0 \in \Theta_0 \\ \theta_1 \in \Theta_1}} LR^x(\theta_0, \theta_1)\,. \tag{20}$$

Note that within this paper a precise likelihood ratio value is denoted with its dependence on θ_0 and θ_1, whereas a interval-valued likelihood ratio is denoted without this dependence.

In addition, the prior odds

$$[\underline{\pi}, \overline{\pi}] \tag{21}$$

might be interval-valued with $\underline{\pi}$ being the lower bound and $\overline{\pi}$ being the upper bound of the subjectively specified prior odds, leading to the imprecisely defined posterior odds

$$[\underline{LR^x} \cdot \underline{\pi}, \overline{LR^x} \cdot \overline{\pi}]. \tag{22}$$

The loss function might also be specified imprecisely by

$$[\underline{k}, \overline{k}], \tag{23}$$

where, analogously, $\underline{k}$ is the lower bound and $\overline{k}$ is the upper bound for stating, in generalization of (9), how much "worse" a_1 would be under H_0 than a_0 would be under H_1, if deciding correctly has 0 "badness" (for a more general account on robust loss functions see Dey and Michaes (2000)).

In contrast to the precise case, the ratio of expected posterior losses r, which was used to determine the optimal action, is not precise anymore:

$$[\underline{r}, \overline{r}], \tag{24}$$

where

$$\underline{r} = \underline{\pi} \cdot \underline{LR^x} \cdot \underline{k} \tag{25}$$

$$\overline{r} = \overline{\pi} \cdot \overline{LR^x} \cdot \overline{k} \tag{26}$$

can be calculated from the respective lower and upper bounds of π, LR^x and k, as all these quantities are positive, and they vary independently. If one of these quantities is still precise, its lower and upper bounds are equal, for instance for a precise k it holds that $k = \underline{k} = \overline{k}$.

The optimal action is

$$a^* = \begin{cases} a_0 & \text{if } \underline{r} \geq 1 \\ a_1 & \text{if } \overline{r} \leq 1 \end{cases}, \tag{27}$$

however, for $\underline{r} < 1 < \overline{r}$, the decision cannot be guided unambiguously and more information is required. This might be accomplished by collecting more data, such that the imprecise likelihood ratio interval will become smaller, or by obtaining more information about the decision problem, such that θ_0, θ_1, π or k might be specified more accurately, i.e. by smaller intervals. With this additional information, the resulting imprecise ratio of expected posterior losses $[\underline{r}, \overline{r}]$ might become smaller and with sufficient information might exclude 1, allowing the determination of the optimal action a^*. This will be illustrated by an example in Section 5.

Certainly, not being able to determine an optimal action in the context of a given data set might at first glance seem to be a disadvantage of the imprecise framework. However, this might only occur if some of the available information is imprecise, such that specifying precise values for the necessary quantities is arbitrary, can be characterized as overprecision and might yield potentially misleading, enforced decisions. Nevertheless, if necessary, enforcing a decision is still possible for $\underline{r} < 1 < \overline{r}$, yet the researcher is now aware of its spuriousness, which might have been masked due to the overprecision within the precise case.

4. Potential Bayesian Updating Issues with Imprecise Hypotheses

Although within the last section simple hypotheses were allowed to be imprecisely specified, this might be accompanied by Bayesian updating inconsistencies that appear while sequentially considering two separate data sets. On that note, (e.g. Seidenfeld, 1994; Huntley et al., 2014) already emphasized the importance of being cautions with sequential decision problems in the context of imprecise probabilities.

4.1. Precise Case

Consider the presence of a second data set $y = (y_1, \ldots, y_m)$ being modeled analogously to x, i.e.

$$f(y|\theta) = \prod_{i=1}^{m} f(y_i|\theta), \tag{28}$$

and denote $z = (x, y)$ as the merged data set with

$$f(z|\theta) = \prod_{i=1}^{n+m} f(z_i|\theta) = f(y|\theta) \cdot f(x|\theta). \tag{29}$$

Therefore, with precise simple hypotheses as in equation (2) it holds that

$$LR^z(\theta_0, \theta_1) = LR^y(\theta_0, \theta_1) \cdot LR^x(\theta_0, \theta_1) \tag{30}$$

and the posterior odds after seeing all the data z

$$LR^z(\theta_0, \theta_1) \cdot \pi = LR^y(\theta_0, \theta_1) \cdot LR^x(\theta_0, \theta_1) \cdot \pi \tag{31}$$

(as well as the ratio of expected posterior losses r) do not depend on whether the data was merged or not.

4.2. Imprecise Case

However, in the context of the imprecise hypotheses from equation (14), define

$$(\underline{\theta_0^x}, \underline{\theta_1^x}) := \underset{\substack{(\theta_0,\theta_1): \\ \theta_0 \in \Theta_0, \theta_1 \in \Theta_1}}{\operatorname{argmin}} LR^x(\theta_0, \theta_1), \tag{32}$$

$$(\underline{\theta_0^y}, \underline{\theta_1^y}) := \underset{\substack{(\theta_0,\theta_1): \\ \theta_0 \in \Theta_0, \theta_1 \in \Theta_1}}{\operatorname{argmin}} LR^y(\theta_0, \theta_1), \tag{33}$$

$$(\underline{\theta_0^z}, \underline{\theta_1^z}) := \underset{\substack{(\theta_0,\theta_1): \\ \theta_0 \in \Theta_0, \theta_1 \in \Theta_1}}{\operatorname{argmin}} LR^z(\theta_0, \theta_1) \tag{34}$$

as the respective tuples of hypothesized parameter values, which lead for each data set to the respective minimal likelihood ratio. As in general

$$\underline{\theta_0^x} \neq \underline{\theta_0^y} \neq \underline{\theta_0^z}, \tag{35}$$

$$\underline{\theta_1^x} \neq \underline{\theta_1^y} \neq \underline{\theta_1^z}, \tag{36}$$

it follows that

$$f(z|\underline{\theta_0^z}) \neq f(y|\underline{\theta_0^y}) \cdot f(x|\underline{\theta_0^x}), \tag{37}$$

$$f(z|\underline{\theta}_1^z) \neq f(y|\underline{\theta}_1^y) \cdot f(x|\underline{\theta}_1^x) \tag{38}$$

and accordingly

$$\underline{LR}^z \neq \underline{LR}^y \cdot \underline{LR}^x \,. \tag{39}$$

Analogue considerations lead to

$$\overline{LR}^z \neq \overline{LR}^y \cdot \overline{LR}^x \,, \tag{40}$$

and an example of this inequality is provided within Section 5.

Therefore, in general, the imprecise posterior odds after considering the merged data

$$[\underline{LR}^z \cdot \underline{\pi}\,, \overline{LR}^z \cdot \overline{\pi}] \neq [\underline{LR}^y \cdot \underline{LR}^x \cdot \underline{\pi}\,, \overline{LR}^y \cdot \overline{LR}^x \cdot \overline{\pi}] \tag{41}$$

differ from those after subsequently considering both data sets separately, which treats the posterior odds after the first data set x as prior odds for the second data set y.

Accordingly, it might seem that the imprecise ratio of expected posterior losses and the resulting decision might depend on whether the data was merged or not. In that, the Bayesian updating procedure for the odds on the hypotheses might be characterized as 'inconsistent' in terms of Rüger (1998, p. 190)'s work on the foundations of statistics.

4.3. Evaluation

Evaluating these updating inconsistencies in greater detail, two characteristics emerge.

First, although the interval-valued likelihood ratio LR^x of the data set x might be outlined by its bounds $\underline{LR}^x$ and $\overline{LR}^x$, consistent updating dictates to also consider the dependence of the likelihood ratio values within LR^x on the parameter values θ_0 and θ_1 as the result of the analysis.

This can be seen based on the following considerations. The interval-valued likelihood ratio LR^x of equation (18) consists of all likelihood ratio values obtained with parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$, i.e.

$$LR^x = \{LR^x(\theta_0, \theta_1) | \theta_0 \in \Theta_0, \theta_1 \in \Theta_1\} \,. \tag{42}$$

In this regard, the values within the interval-valued likelihood ratio of the merged data z might be decomposed using equation (30) to

$$\begin{aligned} LR^z &= \{LR^z(\theta_0, \theta_1) | \theta_0 \in \Theta_0, \theta_1 \in \Theta_1\} & (43)\\ &= \{LR^y(\theta_0, \theta_1) \cdot LR^x(\theta_0, \theta_1) | \theta_0 \in \Theta_0, \theta_1 \in \Theta_1\} \,. & (44) \end{aligned}$$

It appears that for each value within LR^z the complete data set has to be evaluated using the same parameter values θ_0 and θ_1. However, for calculating e.g. $\underline{LR}^y \cdot \underline{LR}^x$, the first part of the data x was evaluated with different parameter values ($\underline{\theta}_0^x$ and $\underline{\theta}_1^x$) than the second part of the data y (evaluated with $\underline{\theta}_0^y$ and $\underline{\theta}_1^y$). Accordingly, the value $\underline{LR}^y \cdot \underline{LR}^x$ might not be contained within LR^z and updating might be inconsistent.

To enable consistent updating, from the first analysis of data set x, all values within the interval-valued likelihood ratio LR^x together with their dependence on θ_0 and θ_1, not only the bounds $\underline{LR}^x$ and $\overline{LR}^x$, are necessary to calculate the final interval-valued likelihood ratio LR^z in a subsequent analysis of both data sets x and y using equation (44).

Second, the values $\underline{LR}^y \cdot \underline{LR}^x$ and $\overline{LR}^y \cdot \overline{LR}^x$ might be considered as approximation of the interval LR^z by providing outer bounds, i.e.

$$LR^z = [\underline{LR}^z, \overline{LR}^z] \subseteq [\underline{LR}^y \cdot \underline{LR}^x, \overline{LR}^y \cdot \overline{LR}^x] \,. \tag{45}$$

This becomes apparent by considering the lower bound $\underline{LR}^z$, which is obtained with parameter values $\underline{\theta}_0^z \in \Theta_0$, $\underline{\theta}_1^z \in \Theta_1$. Applying equation (30) leads to

$$\underline{LR}^z = LR^z(\underline{\theta}_0^z, \underline{\theta}_1^z) = LR^y(\underline{\theta}_0^z, \underline{\theta}_1^z) \cdot LR^x(\underline{\theta}_0^z, \underline{\theta}_1^z) \tag{46}$$

and as $\underline{LR}^x$ and $\underline{LR}^y$ are minima, it also holds that

$$\begin{aligned} \underline{LR}^x &\le LR^x(\underline{\theta}_0^z, \underline{\theta}_1^z) & (47)\\ \underline{LR}^y &\le LR^y(\underline{\theta}_0^z, \underline{\theta}_1^z) \,, & (48) \end{aligned}$$

so that together (as all likelihood ratios are positive)

$$\underline{LR}^y \cdot \underline{LR}^y \le LR^y(\underline{\theta}_0^z, \underline{\theta}_1^z) \cdot LR^x(\underline{\theta}_0^z, \underline{\theta}_1^z) = \underline{LR}^z \,. \tag{49}$$

Analogue considerations lead to

$$\overline{LR}^y \cdot \overline{LR}^y \ge \overline{LR}^z \,, \tag{50}$$

finally allowing the approximation in equation (45).

5. Example

A short fictitious example shall serve as illustration (replicable with the R code in the electronic appendix).

Person A provides a huge amount of allegedly fair coins and offers a bet to person B for 1€: Person A will randomly take one of the coins and flip it. If tails, then person B will get back 4€. Naturally, person B is suspicious about the coins being fair and eventually obtains the permission to examine some coins. Based on the outcome of that sample, person B will have to decide whether to accuse person A of cheating (action a_1) or not (action a_0).

Modelling the coin flips as independent Bernoulli experiments with parameter p for the probability of heads, person B considers the possibility of the coins being fair with the precise null hypothesis $H_0 : p = 0.5$. However, person B is unsure about the parameter p if person A is cheating. Due to the offer of person A, p might be at least 0.75, but on the other hand, if p might be too high, say $p > 0.9$, it might be too suspicious. Person B regards those parameter values $[0.75, 0.9]$ as plausible, but is not able to further describe the plausibility of each of these parameter value. Furthermore, person B considers the possibility that different coins might have (slightly) different probabilities of heads and, therefore, chooses as alternative hypothesis the imprecise simple hypothesis $H_1 : p = [0.75, 0.9]$.

The loss $L(H_1, a_0)$ of not doing anything if the coins are truly biased is not too high, as the price of the bet is only 1€. Accusing person A of cheating if the coins are actually fair ($L(H_0, a_1)$), however, might result in a rather unpleasant situation. Naturally, both these losses are on a different scale, but need to be expressed in relation to each other. As this is rather difficult, Person B figures out that k might be somewhere between 8 and 20, being unable to further specify this value.

In a situation before checking the coins, person B is also not exactly sure what to belief about the coins. Certainly, with the offer of person A, the alternative hypothesis is at least as plausible as the null hypothesis. However, the coins look normal and so the null hypothesis is not absolutely implausible. After some consideration, person B determines that the prior odds are captured by $\pi = [1, 4]$.

Now, person B flips $n = 10$ coins, yielding heads $x = 9$ times. Based on this observation and the specifications given, person B calculates the interval-valued likelihood ratio

$$\mathcal{LR}^x = [0.025, 0.052] \tag{51}$$

and the ratio of expected posterior losses

$$[0.202, 4.162]\,, \tag{52}$$

which does not unambiguously favor one of the actions, as it contains the value 1.

Additional information is necessary to do so and person B flips another $m = 10$ coins, yielding heads $y = 5$ times. The corresponding interval-valued likelihood ratio is

$$\mathcal{LR}^y = [4.214, 165.4]\,. \tag{53}$$

Combining those interval-valued likelihood ratios yields

$$[\underline{\mathcal{LR}}^y \cdot \underline{\mathcal{LR}}^y, \overline{\mathcal{LR}}^y \cdot \overline{\mathcal{LR}}^y] = [0.105, 8.601]\,, \tag{54}$$

but knowing of the updating inconsistencies, person B treats this interval only as an approximation, resulting in an approximation of the ratio of expected posterior losses by

$$[0.843, 688.1]\,. \tag{55}$$

Still the value 1 is included within the interval and this approximation does not allow an unambiguous decision.

In order to account for the updating inconsistencies, person B merges both data sets $z = 9 + 5 = 14$ with $n + m = 20$, leading to the interval-valued likelihood ratio

$$\mathcal{LR}^z = [0.219, 4.169]\,, \tag{56}$$

which is truly different to and included by the interval in equation (54). The resulting ratio of expected posterior losses is

$$[1.754, 333.5]\,, \tag{57}$$

which finally favors to not accuse person A of cheating (action a_0).

By providing the data (n, m, x and y), sufficient information is available for subsequent analyses to consider the dependency of respective likelihood ratio values on the parameter θ_1.

Person B specified the relevant quantities as best as possible to the partially available knowledge and the analysis of the first data set indicated a lack of information for guiding the decision. A precise account of the situation, on the other hand, might have pretended a precision, which is not available. For example, person B might have arbitrarily chosen $H_1 : p = 0.8$, $k = 8$ and $\pi = 1$ of those possible values that are in accordance with the available knowledge, leading to a precise likelihood ratio of $\mathcal{LR}^x(0.5, 0.8) = 0.036$ and a ratio of expected posterior losses of $r = 0.29$ that favoured a_1.

Accordingly, person B would have accused person A of cheating, although the available information are rather ambiguous. Even worse, person B would not even be aware of the lack of information, as it was masked by the false precision of the arbitrarily chosen values.

6. Concluding Remarks

This paper elaborated on how to include partial information about simple hypotheses, prior beliefs and the loss function in the context of hypothesis-based Bayesian decision theory and depicted inconsistencies within the procedure of Bayesian updating that might arise from the use of imprecise simple hypotheses.

Typically, there is only one data set for the statistical analysis of an empirical study, so that the updating inconsistencies as depicted in Section 4 might not become visible. Furthermore, for guiding the decision based on a single data sets, within the context employed in this paper, only the bounds of the interval-valued likelihood ratio are necessary. Nevertheless, properly reporting the results of the analysis also requires to include the dependence of the likelihood ratio values within the interval-valued likelihood ratio on the parameter values. Naturally, as an alternative, the data can be made publicly accessible, so that all relevant information necessary for subsequent analyses might be extracted directly from the data.

Although two data sets were considered to outline the updating inconsistencies, this cannot be regarded as an unnatural approach, as one of the central characteristics of Bayesian learning is to employ a posterior distribution obtained from previous data as prior distribution for a subsequent analysis. Certainly, this reflects the natural way to accumulate information.

In addition, remark that within this paper, the (imprecise) prior odds are updated first to obtain the posterior odds before determining a potentially optimal decision. However, a different procedure might be possible as well. For each hypothesis a decision strategy might be calculated, which maps the potentially observed data to the optimal action. In that, a decision strategy might be chosen first based on the prior odds and then the optimal decision might be determined based on the observed data. While this equivalence of prior risk optimality and posterior loss optimality holds in the traditional case of precise probabilities and loss functions, it is no longer satisfied in more general settings (see explicitly Augustin (2003) and more generally the references in Section 4).

Sometimes, an applied researcher is not primarily interested in guiding a decision, but just in investigating a real-world phenomenon. In this case, a hypothesis-based statistical analysis might be superfluous and descriptive statistics seem to be sufficient (see also the literature about “new statistics”, e.g. Cumming, 2014). Nevertheless, all information should be provided, such that other researchers are able to guide a decision.

Furthermore, the only quantities treated imprecisely within this paper were the hypotheses, the prior odds and the loss function, however, also the data themselves might be available imprecisely, representing ambiguity in the data values. Although, most commonly, data values in psychological research represent scores that are designed to be precise, extending this framework to allow imprecise data looks very promising, as the data are independent of the other imprecisely specified quantities.

In summary, this paper addressed the imprecise generalization of hypothesis-based Bayesian decision making using simple hypothesis and, therefore, employed the likelihood ratio. A Bayes factor analysis typically employs composite hypotheses as well and might therefore be considered as more complex than the context depicted here. Yet, even within this simple context updating inconsistencies might occur, emphasizing the importance of investigating them in greater detail particularly with regard to their presence in analyses using Bayes factors.

Appendix A. R Code

R code to replicate the example is provided electronically.

Acknowledgments

We want thank all reviewers for their valuable comments. In addition, PS wants to thank the LMU Mentoring Program, which supports young researchers.

References

Thomas Augustin. On the suboptimality of the generalized Bayes rule and robust Bayesian procedures from the decision theoretic point of view — a cautionary note on updating imprecise priors. In Jean-Marc Bernard, Teddy Seidenfeld, and Marco Zaffalon, editors, *ISIPTA ’03: Proceedings of the Third International Symposium on Imprecise Probabilities and their Applications*, pages 31–45, Lugano, Waterloo, 2003. Carleton Scientific.

Thomas Augustin, Gero Walter, and Frank P. Coolen. Statistical inference. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. John Wiley & Sons, 2014.

James O. Berger. *Statistical decision theory and Bayesian analysis. 2nd edition*. Springer, 1995.

Geoff Cumming. The new statistics: Why and how. *Psychological Science*, 25(1):7–29, 2014.

Dipak K. Dey and Athanasios C. Michaes. Ranges of posterior expected losses and ε-robust actions. In David Ríos Insua and Fabrizio Ruggeri, editors, *Robust Bayesian Analysis*, pages 145–159. Springer, 2000.

Didier Dubois. Belief structure, possibility theory and decomposable confidence measures on finite sets. *Computers and Artificial Intelligence*, 5:403–416, 1986.

Luisa Ebner, Patrick Schwaferts, and Thomas Augustin. Robust Bayes factor for independent two-sample comparisons under imprecise prior information. conditionally accepted subject to minor revision for: Jasper de Bock, Cassio P. de Campos, Gert de Cooman, Erik Quaeghebeur, and Gregory Wheeler, editors, *Proceedings of the 11th International Symposium on Imprecise Probability: Theory and Applications (ISIPTA ’19, Ghent), Proceedings in Machine Learning Research*, 2019.

Mithat Gönen, Wesley O. Johnson, Yonggang Lu, and Peter H. Westfall. The Bayesian two-sample t test. *The American Statistician*, 59(3):252–257, 2005.

Nathan Huntley, Robert Hable, and Matthias C. M. Troffaes. Decision making. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 190–206. John Wiley & Sons, 2014.

John P. A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):e124, 2005.

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

Michael Lavine and Mark J. Schervish. Bayes factors: what they are and what they are not. *The American Statistician*, 53(2):119–122, 1999.

Charles C. Liu and Murray Aitkin. Bayes factors: Prior sensitivity and model generalizability. *Journal of Mathematical Psychology*, 52(6):362–375, 2008.

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

Jeffrey N. Rouder, Paul L. Speckman, Dongchu Sun, Richard D. Morey, and Geoffrey Iverson. Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2):225–237, 2009.

Richard Royall. The likelihood paradigm for statistical evidence. In Mark L. Taper and Subhash R. Lele, editors, *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, pages 119–152. University of Chicago Press, 2004.

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

Teddy Seidenfeld. When normal form and extensive form solutions differ. In Dag Prawitz, Brian Skyrms, and Dag Westerstahl, editors, *Logic, Methodology and Philosophy of Science IX (Uppsala, 1991)*, pages 451–463. Elsevier, 1994.

Rens Van De Schoot, Sonja D. Winter, Oisín Ryan, Mariëlle Zondervan-Zwijnenburg, and Sarah Depaoli. A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2):217–239, 2017.

Peter Walley. *Statistical Reasoning With Imprecise Probabilities*. Chapman & Hall, 1991.