Proceedings of Machine Learning Research 147:280–288, 2021 ISIPTA 2021

Imprecise Hypothesis-based Bayesian Decision Making with Composite 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

Statistical analyses with composite hypotheses are omnipresent in empirical sciences, and a decision-theoretic account is required in order to formally consider their practical relevance. A Bayesian hypothesis-based decision-theoretic analysis requires the specification of a prior distribution, the hypotheses, and a loss function, and determines the optimal decision by minimizing the expected posterior loss of each hypothesis. However, specifying such a decision problem unambiguously is rather difficult as, typically, the relevant information is available only partially. In order to include such incomplete information into the analysis and to facilitate the use of decision-theoretic approaches in applied sciences, this paper extends the framework of hypothesis-based Bayesian decision making with composite hypotheses into the framework of imprecise probabilities, such that imprecise specifications for the prior distribution, for the composite hypotheses, and for the loss function are allowed. Imprecisely specified composite hypotheses are sets of parameter sets that are able to incorporate blurring borders between hypotheses into the analysis. The imprecisely specified prior distribution gets updated via generalized Bayes rule, such that imprecise probabilities of the (imprecise) hypotheses can be calculated. These lead – together with the (imprecise) loss function – to a set-valued expected posterior loss for finding the optimal decision. Beneficially, the result will also indicate whether or not the available information is sufficient to guide the decision unambiguously, without pretending a level of precision that is not available.

Keywords: Decision Theory, Bayesian Statistics, Composite Hypotheses, Imprecise Probabilities

1. Introduction

There is an increased awareness within the applied sciences that it is important to consider the practical relevance of an effect in addition to its statistical significance (see e.g. Kirk, 1996). Implemented in hypothesis-based methodologies, this relates to specifying the hypotheses reasonably w.r.t. their practical implications (see e.g. methods using regions of practical equivalence (Kruschke, 2015, 2018) or equivalence tests (Lakens, 2017; Lakens et al., 2018)). Practical implications naturally depend on what results are used for. Aptly put by Berger and Wolpert (1988, p. 55): "But no matter what is meant by inference, if it is to be of any value, then somehow it must be used, or acted upon, and this does indeed lead back to the decision-theoretic framework." However, a decision-theoretic analysis (see e.g. Berger, 1985) is typically avoided in applied sciences (cp. e.g. the recommendation in Rouder et al., 2018, p. 110). This might be explained by the fact that many required quantities are very difficult to specify unambiguously, as relevant information is typically available only partially.

In order to facilitate such a decision-theoretic analysis, this paper intends to extend hypothesis-based analyses into a decision-theoretic framework that allows for impartial information to be included properly, building on the framework of imprecise probabilities (see e.g. Augustin et al., 2014a; Walley, 1991). Therefore, previous elaborations on imprecise hypothesis-based Bayesian decision making (Schwaferts and Augustin, 2019) shall be extended to composite hypotheses.

The paper starts by presenting the precise framework of hypothesis-based Bayesian decision making with composite hypotheses in Section 2, which serves as basis for its extension to the framework of imprecise probabilities in Section 3 by allowing imprecisely specified prior distributions (Section 3.1), composite hypotheses (Section 3.2), and loss functions (Section 3.3). A schematic example is provided in Section 4, followed by a discussion about scalability (Section 5.1) and the conditional perspective (Section 5.2), as well as by a brief outlook in Section 6.

2. Precise Hypothesis-Based Bayesian Decision Making

Within the context of decision making (for an extensive overview see Berger, 1985), the observed data x are commonly assumed to be parametrically distributed with density $f(x|\theta)$ and parameter $\theta \in \Theta$. Although generalizations to multidimensional parameters are possible, the parameter θ is assumed to be a single real-valued scalar within this paper to keep the notation simple.

In the Bayesian setting, there is a prior distribution π_θ on the parameter θ with density $\pi(\theta)$. In the presence of the observed data x, this prior distribution gets updated via Bayes rule to the posterior distribution $\pi_{\theta|x}$ with density

$$\pi(\theta|x) = \frac{f(x|\theta) \cdot \pi(\theta)}{f(x)}, \tag{1}$$

where

$$f(x) = \int_{\Theta} f(x|\theta) \cdot \pi(\theta)\, d\theta \tag{2}$$

is the marginal density of the data x, assumed to be strictly positive for all x.

Within a Bayesian analysis, results are derived from this posterior distribution $\pi_{\theta|x}$ exclusively, e.g. by considering mean, median, or certain credibility intervals (see e.g. Kruschke, 2015). If a research question contrasts different theoretical positions, these need to be formalized as statistical hypotheses, which are then evaluated using the posterior distribution. Formally, composite hypotheses

$$h_0 : \theta \in \Theta_0 \quad \text{vs.} \quad h_1 : \theta \in \Theta_1 \tag{3}$$

are subsets $\Theta_0, \Theta_1 \subset \Theta$ of the parameter space.

For the given prior distribution on θ and the observed data x, the posterior probabilities of the hypotheses are

$$p(h_0|x) := p(\Theta_0|x) = \int_{\Theta_0} \pi(\theta|x)\, d\theta, \tag{4}$$

$$p(h_1|x) := p(\Theta_1|x) = \int_{\Theta_1} \pi(\theta|x)\, d\theta, \tag{5}$$

where we assume non-degenerated cases with $p(h_0|x) > 0$ and $p(h_1|x) > 0$.

Frequently, contrasting statistical hypotheses (and the corresponding theoretical positions) is related to an applied research question or some practical implications, being formalized by a decision problem. Consider the case of a decision between two actions a_0 and a_1 (as only two hypotheses are considered within this paper). A loss function

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

with $\mathscr{H} = \{h_0, h_1\}$ being the hypothesis space and $\mathscr{A} = \{a_0, a_1\}$ being the action space, quantifies the "badness" of deciding for $a \in \mathscr{A}$ if $h \in \mathscr{H}$ is true.

Typically, deciding for a_1 if h_1 is true and for a_0 if h_0 is true is considered to be a correct decision, such that – without loss of generality – the loss function can be stated in regret form, in which deciding correctly has zero loss, i.e. $L(h_1,a_1) = L(h_0,a_0) = 0$. The remaining values refer to the type-I-error ($L(h_0,a_1)$) and the type-II-error ($L(h_1,a_0)$) and are assumed to be non-zero. In that, it is possible here to specify the loss function L by one single quantity

$$k := \frac{L(h_0,a_1)}{L(h_1,a_0)}, \tag{7}$$

which specifies how bad the type-I-error is compared to type-II-error (if deciding correctly has zero loss).

As the hypotheses (equation (3)) represent sets of parameters, the loss function L – which was defined on the hypothesis space $\mathscr{H}$ and the action space $\mathscr{A}$ within this paper (equation (6)) – might also be depicted w.r.t. the parameter space Θ and the action space $\mathscr{A}$, formally

$$L_\theta : \Theta \times \mathscr{A} \to \mathbb{R}_0^+ : (\theta, a) \mapsto L_\theta(a). \tag{8}$$

An example of such a loss function in regret form w.r.t. the hypotheses is illustrated in the context of the example (Section 4) in Figure 1 (top).

With the expected posterior loss $\rho : \mathscr{A} \to \mathbb{R}_0^+$:

$$a \mapsto \rho(a) = L(h_1,a) \cdot p(h_1|x) + L(h_0,a) \cdot p(h_0|x), \tag{9}$$

the ratio of expected posterior losses

$$r := \frac{\rho(a_1)}{\rho(a_0)} = \frac{L(h_0,a_1) \cdot p(h_0|x)}{L(h_1,a_0) \cdot p(h_1|x)} \tag{10}$$

$$= k \cdot \frac{p(h_0|x)}{p(h_1|x)} \tag{11}$$

allows to determine the set $\mathscr{A}^*$ of optimal actions (in the context of the conditional Bayes decision principle Berger, 1985, p. 16)

$$\mathscr{A}^* = \begin{cases} \{a_0\} & \text{if } r < 1 \\ \{a_1\} & \text{if } r > 1 \\ \{a_0, a_1\} & \text{if } r = 1 \end{cases} . \tag{12}$$

Both actions are optimal, if $r = 1$. In this case, one might arbitrarily select one of the actions, as both actions have exactly the same expected posterior loss and can, therefore, be considered as practically equal.

3. Imprecise Hypothesis-Based Bayesian Decision Making

Within the framework of hypothesis-based Bayesian decision theory, imprecision shall be investigated for the prior distribution, the composite hypotheses, and the loss function, as an unambiguous precise specification of those quantities appears to bear the most difficulties for applied scientists (in contrast to precise data, likelihoods and parameters). Nevertheless, an extension towards imprecise data, imprecise likelihoods (see e.g. Walley, 1991, ch. 8) and imprecise parameters (for imprecise parameters in the context of simple hypotheses Schwaferts and Augustin, 2019) seems very powerful and shall be considered in future developments.

3.1. Imprecise Prior Distribution

A Bayesian analysis requires the specification of a prior distribution on the parameter. However, it is often impossible to determine one single precise prior distribution describing adequately the extent and homogeneity of the knowledge at hand (e.g. Augustin et al., 2014b, Section 7.2). This impossibility is further illustrated by the fact that even the interpretation of a prior distribution is not unambiguously agreed on, with interpretations as knowledge (e.g. in Jaynes, 2003) or information (e.g. in Berger, 1985) or degrees of belief (e.g. in Jeffreys, 1961) or uncertainty (e.g. in Kruschke, 2015) about the parameter before observing the data. Instead, following the framework of imprecise probabilities (Walley, 1991), a set of prior distributions is considered to be more suitable. This set shall be denoted by Π_θ and referred to as imprecise prior distribution. It constitutes a quantity of its own and represents the prior situation. This imprecise prior gets updated after observing the data x to an imprecise posterior distribution

$$\Pi_{\theta|x} = \left\{ \pi_{\theta|x} \,\middle|\, \pi_\theta \in \Pi_\theta \right\} , \tag{13}$$

where each posterior distribution $\pi_{\theta|x}$ is obtained via Bayes rule (equation (1)) from one of the prior distributions $\pi_\theta \in \Pi_\theta$. Updating such a set of prior probabilities element by element is very natural in the context of Bayesian sensitivity analyses and so-called robust Bayesian approaches (e.g. Rios Insua and Ruggeri, 2000). Moreover, it can be justified by Walley's general coherence theory (Walley, 1991, Chapter 6ff.), where equation (13) is deduced from Walley's generalized Bayes rule.[1]

The resulting posterior distribution $\Pi_{\theta|x}$ represents the posterior situation (given the prior Π_θ and the data x) and underlies all further derivations in a Bayesian analysis.

3.2. Imprecise Composite Hypotheses

If two theoretical positions should be contrasted with each other, these need to be formalized as statistical hypotheses. However, determining which parameter values correspond to which theoretical position might not be unambiguous for all parameter values θ. To account for this, a hypothesis should comprise not only a single set of parameters (as in equation (3)) but a set of parameter sets. Denote these sets of parameter sets as

$$[\Theta]_0 := \{\Theta_0 \subset \Theta \,|\, \Theta_0 \text{ reasonable under } H_0\} \tag{14}$$
$$[\Theta]_1 := \{\Theta_1 \subset \Theta \,|\, \Theta_1 \text{ reasonable under } H_1\} , \tag{15}$$

where $[\Theta]_0$ contains all parameter sets Θ_0 that are reasonable for one hypothesis and $[\Theta]_1$ contains all parameter sets Θ_1 that are reasonable for the other hypothesis.

These sets are considered as entities on their own and formalize the theoretical positions that should be contrasted with each other, considering the available information as is.

Accordingly, the respective imprecisely specified hypotheses are

$$H_0 : \theta \in [\Theta]_0 \quad \text{vs.} \quad H_1 : \theta \in [\Theta]_1 . \tag{16}$$

1. See, in particular, the corresponding lower envelope theorem (Walley, 1991, Section 6.4.2) and (Walley, 1991, Section 7.8.1) on the coherence of envelopes of standard Bayesian inference.

This notation can be read as: The imprecise hypothesis H_0 states that the parameter θ is of a set Θ_0, which itself is only vaguely defined by $[\Theta]_0$ (H_1 analogously).

The crucial difference between precise hypotheses (equation (3)) and imprecise hypotheses (equation (16)) is that in the precise case, a certain parameter value θ might be assigned to *either* one hypothesis, the other hypothesis, both hypotheses (such that hypotheses are overlapping), or no hypothesis (such that this parameter value is not considered at all), while in the imprecise case, the assignment of a certain parameter value θ to the hypotheses might be *any* combination of these four options. Also note that the imprecise parameter sets $[\Theta]_0$ and $[\Theta]_1$ are different from

$$\bigcup_{\Theta_0 \in [\Theta]_0} \Theta_0 = \{\theta \in \Theta \,|\, \theta \in \Theta_0, \Theta_0 \in [\Theta]_0\} \tag{17}$$
$$\bigcup_{\Theta_1 \in [\Theta]_1} \Theta_1 = \{\theta \in \Theta \,|\, \theta \in \Theta_1, \Theta_1 \in [\Theta]_1\} , \tag{18}$$

which would represent two – most likely overlapping – precise hypotheses, and not two imprecisely specified hypotheses. Precise overlapping composite hypotheses imply that there is certainty that some parameter values $\theta \in \Theta$ are contained in both hypotheses, while the imprecise composite hypotheses state that there is uncertainty to which hypothesis some parameter values might be attributed. In that, latter hypotheses inherit far less requirements on the available information for their specification.

Although these formulations of imprecisely specified hypotheses might be employed in statistical analyses without being embedded into the decision theoretic context, this paper focuses on their use for guiding decisions. Then, imprecisely specified hypotheses might also be expressed by an imprecision in the parameter-based loss function L_θ (equation (8)), as illustrated in Figure 1 (center).

In order to obtain the (imprecise) posterior probabilities $P(H_0|x)$ and $P(H_1|x)$ of the imprecise hypotheses H_0 and H_1 using the imprecise posterior distribution $\Pi_{\theta|x}$, one might consider each combination of the distributions $\pi_{\theta|x} \in \Pi_{\theta|x}$ and the parameter sets $\Theta_0 \in [\Theta]_0$ or $\Theta_1 \in [\Theta]_1$, respectively, using equations (4) and (5):

$$P(H_0|x) = \left\{ p(h_0|x) \,\middle|\, \Theta_0 \in [\Theta]_0, \pi_{\theta|x} \in \Pi_{\theta|x} \right\} \tag{19}$$
$$P(H_1|x) = \left\{ p(h_1|x) \,\middle|\, \Theta_1 \in [\Theta]_1, \pi_{\theta|x} \in \Pi_{\theta|x} \right\} . \tag{20}$$

The (imprecise) ratio between these two imprecise quantities is

$$\left[\frac{P(H_0|x)}{P(H_1|x)} \right] := \left\{ \frac{p(h_0|x)}{p(h_1|x)} \,\middle|\, p(h_i|x) \in P(H_i|x), i = 0,1 \right\} \tag{21}$$

with supremum

$$\overline{P} := \sup \left[\frac{P(H_0|x)}{P(H_1|x)} \right] \tag{22}$$

and infimum

$$\underline{P} := \inf \left[\frac{P(H_0|x)}{P(H_1|x)} \right] . \tag{23}$$

3.3. Imprecise Loss Values

Typically, the value k (see equation (7)), which completely specifies the loss function (as in Section 2), is difficult to specify unambiguously as a precise value due to insufficient information. An imprecise loss function, however, allows to consider a set K of reasonable values for k (illustrated in Figure 1, bottom).

The imprecise ratio of expected posterior losses is a set

$$R := \left\{ r = k \cdot \frac{p(h_0|x)}{p(h_1|x)} \,\middle|\, k \in K, \frac{p(h_0|x)}{p(h_1|x)} \in \left[\frac{P(H_0|x)}{P(H_1|x)}\right] \right\} \tag{24}$$

that considers all obtainable ratios r of expected posterior losses (equation (11)) that arise within the imprecisely specified setting.

With the supremum $\overline{K} := \sup K$ and infimum $\underline{K} := \inf K$ of K, the imprecise ratio of expected posterior losses R is bounded by

$$\overline{R} := \sup R = \overline{K} \cdot \overline{P} \tag{25}$$
$$\underline{R} := \inf R = \underline{K} \cdot \underline{P}, \tag{26}$$

as all these quantities are non-negative.

Now, the set $\mathscr{A}^*$ of optimal actions is

$$\mathscr{A}^* = \begin{cases} \{\,\} & \text{if } \underline{R} < 1 < \overline{R} \\ \{a_0\} & \text{if } 1 \leq \underline{R}, 1 < \overline{R} \\ \{a_1\} & \text{if } \underline{R} < 1, \overline{R} \leq 1 \\ \{a_0, a_1\} & \text{if } \underline{R} = \overline{R} = 1 \end{cases} . \tag{27}$$

The case with $\underline{R} = \overline{R}$ depicts the precise case (as in Section 2) generalized within the imprecise framework. For $\underline{R} < 1 < \overline{R}$, the available information is not sufficient to unambiguously declare one of the actions as optimal. Therefore, the set $\mathscr{A}^*$ of optimal actions is empty and the decision should be withheld. If so, further information about the imprecisely specified quantities might be obtained, such that they can be specified more precisely (i.e. by smaller sets), or additional data might be collected, allowing to obtain a less imprecise ratio of expected posterior losses R to obtain an optimal action unambiguously.

In a sense, one might say to be "indecisive" if $\mathscr{A}^* = \{\,\}$ or if $\mathscr{A}^* = \{a_0, a_1\}$, as both cases do not yield a single optimal action. However, these cases are fundamentally different. In the first case, there is not enough information to declare one action as superior, in the second case, there is enough information to state that both actions should be rated as practically equal. In that, we want to emphasize that an action is to be considered optimal within this paper, if there is enough information available to declare it as superior or practically equal to the other action (or actions; for the more general case see Section 5.1).

4. Example

Does drug Z help to treat the symptoms of disease D, measured on scale S? Respective actions are

a_0: do not administer drug Z to patients with disease D

a_1: administer drug Z to patients with disease D

To assess this question, a team of investigators plans to run an experiment, in which a number $n = 100$ of patients with disease D are treated with drug Z and the change s_j ($j = 1, \ldots, 100$) of their symptoms on a metric scale S is measured.

Previous investigations showed that treating patients with disease D with a placebo did not increase the symptoms and the standard deviation of the change in symptoms on scale S was 15. Therefore (and for the sake of simplicity within this example), the changes s_j ($j = 1, \ldots, 100$) are modelled as (independent and identically) normally distributed random quantities $S_j \overset{iid}{\sim} N(\delta, (15)^2)$, where the effect parameter $\delta \in \Delta = \mathbb{R}$ represents the difference in change of symptoms compared to the zero-change of the placebo.

A prior distribution for δ is difficult to determine. Although the investigators agree that a normal distribution $\delta \sim N(\mu, \sigma^2)$ might be reasonable (with $\mu \in M$ and $\sigma \in \Sigma$, such that the hyperparameter space is $M \times \Sigma$), they disagree slightly about the specification of the hyperparameters. After some discussions they determine that the set

$$\Pi_\delta = \left\{ N(\mu, \sigma^2) \,\middle|\, \mu \in [3, 12], \sigma \in [10, 20] \right\} \tag{28}$$

covers all different opinions about the hyperparameters, and decide to use it as imprecise prior distribution.

Further, determining whether (a_1) or not (a_0) to administer drug Z to treat disease D for which effect values δ is not easy. The investigators try to consider the consequences of different scenarios, but these cannot be outlined unambiguously as this is a new field of research and some essential information is still pending from other investigations. They can agree that it is safe to assume that effects δ smaller than $\underline{\delta} = 5$ do not justify administering the drug Z due to adverse effects, and that effects δ larger than $\overline{\delta} = 8$ do justify administering the drug Z as the benefits outweigh the adverse effects. For effect values within $[\underline{\delta}, \overline{\delta}] = [5, 8]$ the situation is less obvious. Therefore, they decide to use the imprecise composite hypotheses

$$H_0 : \delta \in [\Delta]_0 \quad \text{vs.} \quad H_1 : \delta \in [\Delta]_1 \tag{29}$$

with

$$[\Delta]_0 := \left\{ \Delta_0 = (-\infty, \tilde{\delta}] \,\middle|\, \tilde{\delta} \in [5, 8] \right\} \tag{30}$$
$$[\Delta]_1 := \left\{ \Delta_1 = [\tilde{\delta}, \infty) \,\middle|\, \tilde{\delta} \in [5, 8] \right\} . \tag{31}$$

Weighting potential adverse effects against possible benefits of the drug Z is challenging, due to impartial information about the adverse effects. After considering different cases, the investigators determine that a loss function with $k \in K = [\underline{K}, \overline{K}] = [3, 15]$ might represent the current knowledge about the consequences of drug Z quite well.

This specification of K is sufficient for the remaining analyses, however, as illustration, consider that the loss function might also be expressed more extensively by relating it directly to the parameter δ (compare equation (8)) via the imprecise specifications (Figure 1, bottom)

$$L_\delta(a_0) = \begin{cases} \{0\} & \text{if } \delta < 5 \\ \{0,1\} & \text{if } 5 \leq \delta \leq 8 \\ \{1\} & \text{if } 8 < \delta \end{cases} \tag{32}$$

and

$$L_\delta(a_1) = \begin{cases} [3,15] & \text{if } \delta < 5 \\ \{0\} \cup [3,15] & \text{if } 5 \leq \delta \leq 8 \\ \{0\} & \text{if } 8 < \delta \end{cases} . \tag{33}$$

Also note, that by using this depiction, the loss function $L_\delta(a)$ might also be interpreted – for each action $a \in \mathscr{A}$ – as an imprecise gamble (illustrations in Figure 1), bridging to the mathematical foundation of the framework of imprecise probabilities in the spririt of Walley (1991) (see also e.g. Quaeghebeur, 2014; Miranda and de Cooman, 2014).

Now, the investigators perform the experiment, obtain the data $s = (s_1, \ldots, s_{100})$, and estimate the effect size (with the in-sample mean) to be $\hat{\delta} = m(s) = 10.03$.

Updating the imprecise prior Π_δ using the generalized Bayes rule (in this case element-wise using the Bayesian normal-normal model with known sample variance) results in the imprecise posterior distribution

$$\Pi_{\delta|s} = \left\{ N(\mu, \sigma^2) \,\middle|\, (\mu, \sigma) \in \mathscr{F} \right\}, \tag{34}$$

where $\mathscr{F}$ is the set (within the hyperparameter space $M \times \Sigma$) as displayed in Figure 2. It can be seen that, compared to the prior distribution Π_δ, the posterior distribution $\Pi_{\delta|s}$ is extremely narrowed down. This is because there is no prior-data-conflict (see e.g. Walter and Augustin, 2009) and the study is highly informative with $n = 100$ patients, such that the data s might easily overwhelm the initial prior uncertainty expressed by Π_δ.

With the stated hypotheses (equation (29)), the imprecise posterior probabilities $P(H_0|s)$ and $P(H_1|s)$ are bounded by

$$0.0003 \leq P(H_0|s) \leq 0.103 \tag{35}$$
$$0.897 \leq P(H_1|s) \leq 0.9997\,, \tag{36}$$

leading to ratios

$$0.0003 = \underline{P} \leq \frac{P(H_0|s)}{P(H_1|s)} \leq \overline{P} = 0.115\,. \tag{37}$$

Together with K the imprecise ratio R of expected posterior losses is characterized by

$$\underline{R} = 0.0009 \quad \text{and} \quad \overline{R} = 1.725\,. \tag{38}$$

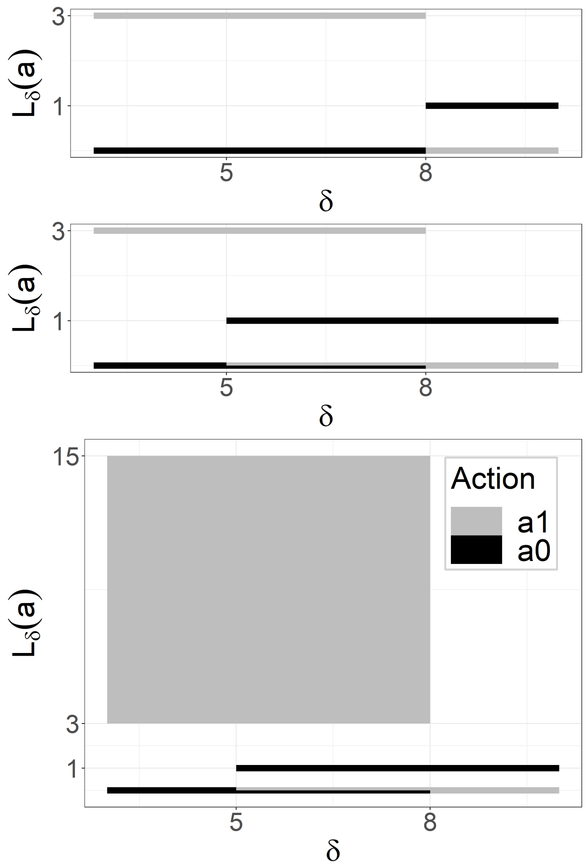

Figure 1: Loss Function. Loss values for both actions a_0 (black) and a_1 (gray) are depicted for varying effect values δ. The top plot illustrates a precise loss function in regret form with $\tilde{\delta} = 8$ (boundary between hypotheses) and $k = 3$. The center plot illustrates how using imprecisely specified hypotheses with $\tilde{\delta} \in [5, 8]$ can also be expressed by an imprecisely specified loss function, although $k = 3$ is still precise (please note the overlapping lines at $L_\delta(a) = 0$). The bottom plot adds an imprecisely specified $k \in [3, 15]$, leading to the loss function of the example (Section 4). Both the top and center plot are included to illustrate the extension of the precise case into the imprecise framework.

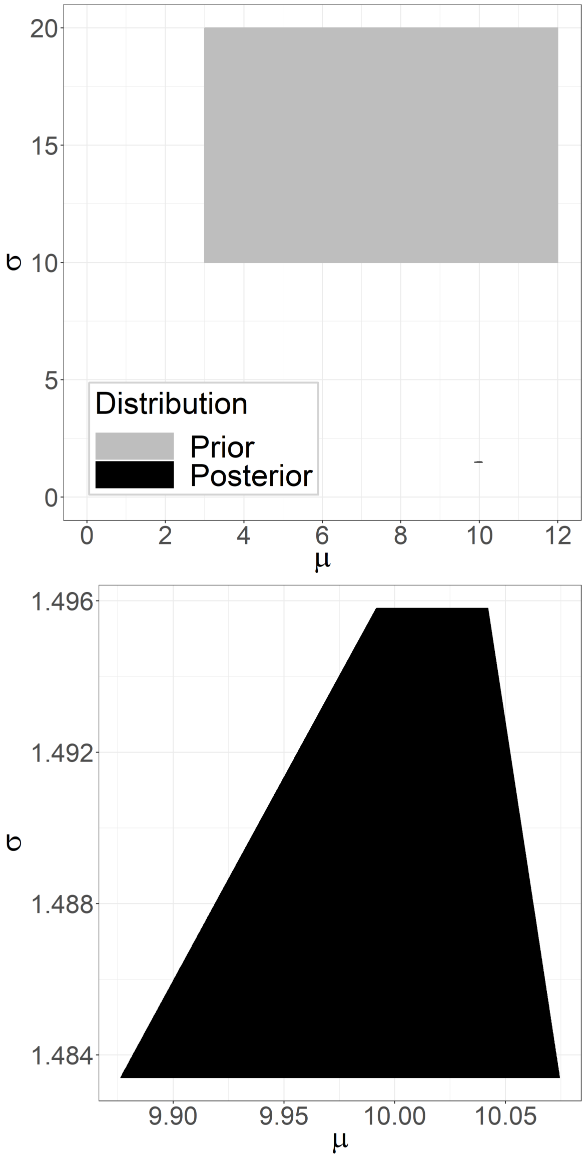

Figure 2: Prior and Posterior Distribution. The gray area (top plot) depicts the set of hyperparameters $(\mu,\sigma) \in M \times \Sigma$ for the normal distributions $N(\mu,\sigma^2)$ that define the imprecise prior distribution Π_δ. This set is bounded by the values $(3,10)$, $(3,20)$, $(12,10)$, and $(12,20)$. Updating the respective distributions leads to normal distributions with parameters $(9.87,1.48)$, $(9.99,1.49)$, $(10.07,1.48)$, and $(10.04,1.49)$, respectively, being extremal elements of the set $\mathscr{F}$ of hyperparamters that define the imprecise posterior distribution. This set is depicted as black area in the top plot bottom-right and enlarged in the bottom plot. Please note the substantial difference in scales between both plots and that this posterior set is not convex (despite its appearance).

As $\underline{R} < 1 < \overline{R}$, the investigators cannot state an unambiguous conclusion with the obtained data s and the available vague information.

Some time later, other investigations about the adverse effects of drug Z were finalized showing that the adverse effects are rather mild. This allows the investigators to specify the loss function more precisely by $K = [3,5]$ and the imprecise ratio R of expected posterior losses narrows down to lie between

$$\underline{R} = 0.0009 \quad \text{and} \quad \overline{R} = 0.575\,, \tag{39}$$

now permitting to state a_1 as the optimal action, as $\overline{R} < 1$. Accordingly, the investigators recommend to administer drug Z to treat patients with disease D.

R code to replicate the example is provided electronically.

5. Discussion

5.1. Scalability

The elaborations within this paper were restricted towards a framework that uses only two hypotheses and, therefore, only two actions. While this is currently the most used framework for hypothesis-based analyses in applied sciences, the considerations within this paper might naturally be scaled towards using multiple hypotheses and actions.

In fact, the depicted case (with only two hypotheses and actions) that determines the optimal action(s) (equation (27)) by considering the ratio R of expected posterior losses (equation (24)) might be considered as a special case in the context of interval dominance in the imprecise decision theoretic framework. This shall be illustrated within the context of the example (Section 4).

As outlined in equations (32) and (33), the loss function might also be viewed w.r.t. the parameter δ, such that the conditions in these multi-case equations are determined by the hypothesis specifications and the values by the actual consequences of the decision. Naturally, additional hypotheses and actions can easily be incorporated using this formulation (although the applied scientists might have more values to specify).

These conditions need to be evaluated w.r.t. the posterior distribution $\Pi_{\delta|s}$, determining the (imprecise) posterior probabilities to be bounded by

$$0.0003 \leq P(\delta < 5|s) \quad \leq 0.0005 \tag{40}$$
$$0.0806 \leq P(5 \leq \delta \leq 8|s) \leq 0.1024 \tag{41}$$
$$0.8970 \leq P(8 < \delta|s) \quad \leq 0.9189\,. \tag{42}$$

The expected posterior loss $\rho_\delta : \mathscr{A} \to \mathbb{R}_0^+$:

$$a \mapsto \rho_\delta(a) = \int_\Delta L_\delta(a) \cdot \pi(\delta|s)\, d\delta \tag{43}$$

is now based on the parameter-based loss function L_δ (in contrast to the hypothesis-based loss function L as in equation (6)), and needs to consider that L_δ is an imprecise quantity and that $\pi(\delta|s)$ denotes the probability densities of the imprecise posterior $\Pi_{\delta|s}$. Respective values are bounded by

$$0.897 \leq \rho_\delta(a_0) \leq 1.021 \quad (44)$$

$$0.0009 \leq \rho_\delta(a_1) \leq 1.544\,. \quad (45)$$

Interval dominance (see e.g. Huntley et al., 2014) compares interval-valued expected posterior losses. Then, an action is declared dominated if there exists another action that has an interval-valued expected posterior loss being strictly less. Such a dominated action can be ruled out. In line with the considerations about optimal actions in Section 3.3, we consider an action as optimal (in the context of interval dominance) if it dominates or practically equals[2] every other action. If there are other actions that cannot be dominated by an action, information is lacking to treat this action as superior and it should not be considered as optimal. In the case considered here, $\rho_\delta(a_0)$ lies within the range of $\rho_\delta(a_1)$, so neither expected posterior loss interval dominates the other one. Thus, no action proves itself as superior.

If, as in the example, additional information were gathered to narrow k down to be within $[3,5]$, the expected posterior loss $\rho_\delta(a_1)$ of action a_1 is then bounded by

$$0.0009 \leq \rho_\delta(a_1) \leq 0.514\,, \quad (46)$$

now being completely below the expected posterior loss $\rho_\delta(a_0)$ of action a_0. Action a_1 dominates action a_0 and is thus considered to be optimal. Apparently, these are the same results as in Section 4 and R code to replicate these numbers is provided electronically.

This illustrates that the framework depicted within this paper can be considered as a special case of the imprecise Bayesian decision theoretic framework using the concept of interval dominance and the conditional Bayes principle, which might be easily extended to additional hypotheses and actions. By being restricted to this special case of only two hypotheses and actions, a simple regret form of the loss function can be used, allowing to determine the optimal action easily via the ratio of expected posterior losses. This allows to extend hypothesis-based analysis into the imprecise decision theoretic framework without exceedingly complicated mathematical formulas, a fact that might be welcomed by applied scientists.

5.2. Conditional Perspective

A rigorous conditional perspective was taken within this Bayesian decision-theoretic approach: The prior gets updated first to the posterior before considering the decision problem and finding the optimal action based on this posterior distribution. However, there is also a different, i.e. unconditional, decision-theoretic approach, which starts by finding an optimal decision function (mapping all possible data sets to optimal actions) by minimizing the prior risk. This approach takes all potentially observable data sets into account, and focuses on the actually observed data only as a second step, evaluating the decision function at the concretely observed sample. Within the precise case, both approaches yield eventually the same optimal action (cp. e.g. Berger, 1985, p. 159). This is, however, not necessarily true within the imprecise case (Augustin, 2003),[3] see also (Seidenfeld, 1994) in a related game theoretic context, – a fact that breathes new life into an old debate between the conditional and the frequentist point of view. The existence of this decision-theoretic dynamic inconsistency in the context of point-wise updating set-valued distributions requires the applied scientists to reason about whether to use the conditional or the unconditional perspective in their analyses. However, the conditional perspective is generally considered to be the preferred point of view within Bayesian statistics, as it does not consider other potential data sets that were not observed (Jeffreys, 1961). The argumentation by Berger (1985, p. 160, notation adapted, italics preserved)

> Note that, from the conditional perspective together with the utility development of the loss, the *correct* way to view the situation is that of minimizing $[\rho(a)]$. One should condition on what is known, namely $[x]$ (...), and average the utility over what is unknown, namely θ. The desire to minimize [the prior risk] would be deemed rather bizarre from this perspective.

carries over to the rigorous generalization developed here, at least as long as single experiments with a single decision are considered.

6. Outlook

Within this paper, the hypothesis-based Bayesian decision-theoretic framework with composite hypotheses was extended to include imprecise specifications of the prior distribution, the hypotheses, and the loss function. These three quantities are expected to be the most difficult to specify in current applied sciences, if a hypothesis-based decision-theoretic Bayesian analysis is intended. Therefore, their imprecise extension might provide a useful framework for applied scientists.

2. As depicted in Section 2, an action practically equals another action if their expected posterior losses are precise and with identical value.

3. The proof of the equivalence of prior risk and posterior loss essentially relies on the interchangeability of integrals over the parameter space and the sample space, which is no longer valid if also maxima/minima of distributions have to be considered in between. This suggests that similar dynamic inconsistencies may occur as soon as imprecise losses are considered.

This approach might also be seen as an extension to the Bayes factor (Gönen et al., 2005; Kass and Raftery, 1995; Rouder et al., 2018), a quantity involved in updating prior probabilities of hypotheses to their posterior probabilities (as in Section 2, equations (4) and (5)) and interpreted as quantification of the evidence within the data w.r.t. the hypotheses. This extension covers the ability to include actions and a (hypothesis-based) loss function into the statistical analysis (such that the practical implication of the study can be considered on a formal level) and the opportunity to treat impartial information about the prior (see also Ebner et al., 2019), the hypotheses, and the loss function as it is, without requiring a level of precision that is not available.

Appendix A. R Code

R code to replicate the example is provided electronically.

Acknowledgments

The authors thank the reviewers for their valuable comments and open sharing of ideas that allowed to significantly improve the quality of this work.

Competing Interests

The authors declare to have no competing interests.

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. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.5.9683.

Thomas Augustin, Frank P. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley, Chichester, 2014a. doi: 10.1002/9781118763117.

Thomas Augustin, Gero Walter, and P. Coolen, Frank. Statistical inference. In Thomas Augustin, Frank P. Coolen, Gert de Cooman, and Matthias Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. Wiley, 2014b. doi: 10.1002/9781118763117.ch7.

James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition, 1985. doi: 10.1007/978-1-4757-4286-2.

James O. Berger and Robert L. Wolpert. The likelihood principle. *Lecture Notes-Monograph Series*, 6:iii–160.2 (discussion: 160.3–199), 1988. URL http://www.jstor.org/stable/4355509.

Luisa Ebner, Patrick Schwaferts, and Thomas Augustin. Robust Bayes factor for independent two-sample comparisons under imprecise prior information. In Jasper De Bock, Cassio P. de Campos, Gert de Cooman, Erik Quaeghebeur, and Gregory Wheeler, editors, *Proceedings of the Eleventh International Symposium on Imprecise Probability: Theories and Applications*, volume 103 of *Proceedings of Machine Learning Research*, pages 167–174. PMLR, 2019. URL http://proceedings.mlr.press/v103/ebner19a.html.

Mithat Gönen, Wesley O. Johnson, Yonggang Lu, and Peter H. Westfall. The Bayesian two-sample t test. *The American Statistician*, 59:252–257, 2005. doi: 10.1198/000313005X55233.

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. doi: 10.1002/9781118763117.ch8.

Edwin T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. doi: 10.1017/CBO9780511790423.

Harold Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, third edition, 1961.

Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995. doi: 10.2307/2291091.

Roger E. Kirk. Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5):746–759, 1996. doi: 10.1177/0013164496056005002.

John K. Kruschke. *Doing Bayesian Data Analysis: A Tutorial With R, JAGS, and Stan*. Academic Press, New York, 2015. doi: 10.1016/B978-0-12-405888-0.09999-2.

John K. Kruschke. Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2):270–280, 2018. doi: 10.1177/2515245918771304.

Daniël Lakens. Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4):355–362, 2017. doi: 10.1177/1948550617697177.

Daniël Lakens, Anne M. Scheel, and Peder M. Isager. Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2):259–269, 2018. doi: 10.1177/2515245918770963.

Enrique Miranda and Gert de Cooman. Lower previsions. In Thomas Augustin, Frank P. Coolen, Gert de Cooman, and Matthias Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 28–55. Wiley, 2014. doi: 10.1002/9781118763117.ch2.

Erik Quaeghebeur. Desirability. In Thomas Augustin, Frank P. Coolen, Gert de Cooman, and Matthias Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 1–27. Wiley, 2014. doi: 10.1002/9781118763117.ch1.

David Rios Insua and Fabbrizzio Ruggeri, editors. *Robust Bayesian Analysis. Lecture Notes in Statistics 152*. Springer, New York, 2000. doi: 10.1007/978-1-4612-1306-2.

Jeffrey N. Rouder, Julia M. Haaf, and Joachim Vandekerckhove. Bayesian inference for psychology, part iv: Parameter estimation and Bayes factors. *Psychonomic Bulletin & Review*, 25(1):102–113, 2018. doi: 10.3758/s13423-017-1420-7.

Patrick Schwaferts and Thomas Augustin. Imprecise hypothesis-based Bayesian decision making with simple hypotheses. In Jasper De Bock, Cassio P. de Campos, Gert de Cooman, Erik Quaeghebeur, and Gregory Wheeler, editors, *Proceedings of the Eleventh International Symposium on Imprecise Probability: Theories and Applications*, volume 103 of *Proceedings of Machine Learning Research*, pages 338–345. PMLR, 2019. URL http://proceedings.mlr.press/v103/schwaferts19a.html.

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 (Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science, Uppsala, 1991)*, pages 451–463. Elsevier, 1994. doi: 10.1016/S0049-237X(06)80056-X.

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

G. Walter and T. Augustin. Imprecision and prior-data conflict in generalized Bayesian inference. *Journal of Statistical Theory and Practice*, 3:255–271, 2009. doi: 10.1080/15598608.2009.10411924.