LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR STATISTIK

Patrick Schwaferts
Thomas Augustin

Bayesian Decisions using Regions of Practical Equivalence (ROPE): Foundations

Technical Report Number 235, 2020
Department of Statistics
University of Munich

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

Bayesian Decisions using Regions of Practical Equivalence (ROPE): Foundations

Patrick Schwaferts Thomas Augustin

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

Abstract

Kruschke [2018] proposes the so called HDI+ROPE decision rule about accepting or rejecting a parameter null value for practical purposes using a region of practical equivalence (ROPE) around the null value and the posterior highest density interval (HDI) in the context of Bayesian statistics. Further, he mentions the so called ROPE-only decision rule within his supplementary material, which is based on ROPE, but uses the full posterior information instead of the HDI.

Of course, if it is about formalizing and guiding decisions then statistical decision theory is the framework to rely on, and this technical report elaborates the decision theoretic foundations of both decision rules.

It appears that the foundation of the HDI+ROPE decision rule is rather artificial compared to the foundation of the ROPE-only decision rule, such that latter might be characterized as being closer to the underlying practical purpose than former. Still, the ROPE-only decision rule employs a truly arbitrary, and thus debatable, choice of loss values.

Keywords: Bayesian Decision Theory, Region of Practical Equivalence, ROPE, HDI+ROPE, ROPE-only, Imprecise Probabilities

1 Introduction

When it comes to applying statistics, there is an increased awareness that black-and-white thinking might lead to severe issues within the process of science, and thus binary decisions should be treated with caution [see e.g. Kruschke, 2018]. Reporting estimates together with the uncertainty about them might be seen as a fruitful alternative [see e.g. Cumming, 2014]. However, sometimes a decision is necessary and the use of statistical decision theory [see e.g. Berger, 1995, Robert, 2007] suggests itself. In that regard, every proposed or employed decision rule might be assessed on the basis of its decision theoretic foundation.

Kruschke [e.g. 2015, 2018] proposes a decision rule based on posterior highest density intervals (HDI) and regions of practical equivalence (ROPE).

A $(1-\alpha)$ highest density interval for a certain distribution of a parameter (prior or posterior) is an interval[1] that contains all parameter values with the highest probability densities and integrates to a probability of $1-\alpha$. Kruschke [2018] employs $(1-\alpha) = 0.95$ and uses the posterior distribution when referring to a highest density interval (HDI; also referred to as highest posterior density (HPD) interval), which will be adopted within this technical report.

A region of practical equivalence (ROPE) refers to a certain parameter value of interest, which might also be called "null value" and frequently (but not necessarily) the parameter value of interest is zero. A ROPE for a null value is a "range of parameter values that are equivalent to the null value for practical purposes" [Kruschke, 2018, p. 272]. Accordingly, "the limits of the ROPE depend on the practical purpose of the ROPE. If the purpose is to assess the equivalence of drug-treatment outcomes, then the ROPE limits depend on the real-world costs and benefits of the treatment and the ability to measure the outcome" [Kruschke, 2015, p. 338].

Once the ROPE is specified (before observing the data) and the HDI is calculated (after observing the data), the decision rule by Kruschke [2018, p. 272], referred to as HDI+ROPE decision rule, is as follows:

- If the HDI falls completely inside the ROPE, then accept the null value for practical purposes.
- If the HDI falls completely outside the ROPE, then reject the null value for practical purposes.
- Else, withhold a decision.

In addition to the HDI+ROPE decision rule, Kruschke [2018, supp. p. 5] mentions another exemplary decision rule within his supplementary material[2] that is based on the ROPE alone and considers the posterior distribution instead of the HDI. Referred to as ROPE-only decision rule, it states:

- If more than 95% of the posterior distribution fall within the ROPE, then accept the null value for practical purposes.
- If less than 5% of the posterior distribution fall within the ROPE, then reject the null value for practical purposes.
- Else, withhold a decision.

Within his supplementary material, Kruschke [2018, supp. p. 3–5] delineates preliminary ideas about the decision theoretic foundation of the HDI+ROPE decision rule. In addition, a foundation of the ROPE-only decision rule is also pending. In that, the purpose of this

[1] Certainly, it might be possible that a HDI is a set of parameters, which is not an interval, however, in accordance with Kruschke [2018], these cases are not considered within this technical report.

[2] This technical report is based on the supplementary material Version 1 of February 25, 2018, available at the Open Science Framework with the url `https://osf.io/jwd3t/` and downloaded at August 18, 2020.

technical report is to elaborate the decision theoretic foundations of both decision rules more profoundly.

Therefore, Bayesian decision theory is briefly recalled in Section 2, before outlining the foundations of the ROPE-only decision rule (Section 3) and of the HDI+ROPE decision rule (Section 4). A concluding discussion in Section 5 compares the foundations of both decision rules w.r.t. their interpretation and connection to the underlying real-world decision.

2 Recall of Bayesian Decision Theory

The observed data $\boldsymbol{x} \in \mathcal{X}$, where the sample space $\mathcal{X}$ comprises all potential data sets, are modeled parametrically as realization of the random quantity X with density $f(\boldsymbol{x}|\theta)$, where $\theta \in \Theta$ is a real-valued parameter and Θ the parameter space.

Within a Bayesian context, there is a prior distribution with density $\pi(\theta)$ on the parameter θ, which gets updated via Bayes formula to the posterior distribution with density $\pi(\theta|\boldsymbol{x})$ once the data $\boldsymbol{x}$ are observed.

In the context of an applied decision, one of different potential actions $\mathsf{a} \in \mathcal{A}$ should be selected, where $\mathcal{A}$ denotes the action space.

Deciding for a certain action $\mathsf{a} \in \mathcal{A}$ if a certain parameter value $\theta \in \Theta$ is true has consequences and the "badness" of these consequences is formally captured by a loss[3] function

$$L : \Theta \times \mathcal{A} \to \mathbb{R}_0^+ \,. \tag{1}$$

Naturally, the meaning of this "badness" is comparative and can only be judged w.r.t. the loss values of other actions and parameter values, yet their comparative meaning should reflect the characteristics within the applied real-world decision. However, it might be rather difficult[4] to specify an exact loss function that matches those characteristics, which are usually accessible only vaguely. As a solution, the loss function might be simplified using ROPE (see Section 3.1) and specified in an imprecise manner (see Section 3.2).

As the posterior density $\pi(\theta|\boldsymbol{x})$ is available within a Bayesian analysis, it is possible to calculate the expected posterior loss of each action $\mathsf{a} \in \mathcal{A}$

$$\rho : \mathcal{A} \to \mathbb{R}_0^+ \tag{2}$$

by

$$\rho(\mathsf{a}) = \int_\Theta L(\theta, \mathsf{a}) \pi(\theta|\boldsymbol{x}) d\theta \,. \tag{3}$$

[3]Sometimes decision theory is depicted with a utility function instead of a loss function, which quantifies the "utility" instead of the "badness" of the respective consequences.

[4]Of course, there are situations in which a loss function might be specified exactly, as e.g. some special cases in economy in which the loss might be related to monetary outcomes or obtained from preferences [see e.g. Berger, 1995, ch. 2.2]. In many research situations, however, the necessary information to do so might not be available.

Intuitively and following the conditional Bayes principle [see e.g. Berger, 1995], the action a with minimal expected posterior loss $\rho(\mathsf{a})$ should be chosen and is called (posterior loss[5]) Bayes action.

Taken together, all three quantities $\pi(\theta)$, $\boldsymbol{x}$, and L are required to find the (optimal) Bayes action.

Before observing the data, only the prior density $\pi(\theta)$ and the loss function L are available. Therefore, it is possible to consider each potentially observable data set $\boldsymbol{x} \in \mathcal{X}$ and evaluate which action $\mathsf{a} \in \mathcal{A}$ would be the corresponding Bayes action. This is formally captured by a decision rule

$$\delta : \mathcal{X} \to \mathcal{A}\,. \tag{4}$$

In the context of the conditional Bayes principle, the optimal decision rule has the following shape

$$\delta(\boldsymbol{x}) = \arg\min_{\mathsf{a}\in\mathcal{A}} \rho(\mathsf{a}) = \arg\min_{\mathsf{a}\in\mathcal{A}} \int_{\Theta} L(\theta, \mathsf{a})\pi(\theta|\boldsymbol{x})d\theta \tag{5}$$

and states the Bayes action for each potential data set.

Of course, it is possible to formulate other decision rules, but these might not find the Bayes action for each data set. In that, refer to the decision rule in equation (5), which matches every data set with the corresponding Bayes action, as Bayes rule[6] (not to be confused with Bayes formula for calculating $\pi(\theta|\boldsymbol{x})$ from $f(\boldsymbol{x}|\theta)$ and $\pi(\theta)$).

Note that a Bayes action and a Bayes rule always refer to a certain loss function. With a different loss function a different decision rule might be a Bayes rule and a different action might be a Bayes action for a given data set.

[5] Within this technical report, the term "Bayes action" always refers to a posterior loss Bayes action.

[6] This depiction of a Bayes rule as minimizing the expected posterior loss is based on one of the fundamental theorems of Bayesian decision theory [c.p. e.g. Berger, 1995, p. 159 Result 1]. In general, the definition of a Bayes rule might involve the minimization of the prior risk (which considers all potentially observable data sets) and this theorem states equivalence with minimizing the expected posterior loss. In anticipation of Section 3, this theorem, however, might not necessarily hold within the framework of imprecise probabilities in general and counterexamples involve imprecisely specified probabilities [Augustin, 2003]. Yet the involvement of the framework of imprecise probabilities within this technical report comprises only an imprecisely stated loss function (and no imprecisely specified probabilities, see Section 3.2), so that an equivalence analogue to this fundamental theorem should hold within the context depicted here. This should be addressed in further research. In any case, a Bayesian analysis typically sticks to a conditional point of view that conditions on the actually observed data and does not consider other potential data sets, which were not observed. In that, Berger [1995, p. 160, notation adapted, italics preserved] reasons:

> 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 [$\rho(\mathbf{a})$]. One should condition on what is known, namely [$\boldsymbol{x}$] (...), 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.

In summary, even if this fundamental theorem might not hold within the context employed here to depict the foundation of the ROPE-only rule (see Section 3), the depicted approach still appears to be reasonable.

3 Foundations of the ROPE-only decision rule

3.1 ROPE as Simplification

As made obvious by the quotes about the ROPE in Section 1, the ROPE cannot be separated from the underlying practical purposes. Implied by both the HDI+ROPE decision rule and the ROPE-only decision rule, the practical purpose is to decide between two actions a_0 and a_1. The first action a_0 is in accordance with the null value $\theta_0 \in \Theta$ and the second is in discordance with the null value θ_0. Accordingly, indicated by subscript P for "practical purpose", the action space $\mathcal{A}_P = \{\mathsf{a}_0, \mathsf{a}_1\}$ comprises these two actions of the practical purpose.

Kruschke [2018, p. 272] refers to these actions as "accept the null value for practical purposes" (a_0) and "reject the null value for practical purposes" (a_1). However, we want to refrain from using this terminology, because it tempts to ignore the actual real-world decision and to derive conclusions about actions that might not even be specified. Instead, we highly recommend to explicitly state the actions of interest, such that the real-world decision of interest might be formalized properly.

The corresponding loss function $L_P : \Theta \times \mathcal{A}_P \to \mathbb{R}_0^+$ quantifies the "badness" of each of those two practical actions under each parameter. With this loss function it would be possible to determine the Bayes action for the observed data set, however, the exact shape of this loss function L_P is hardly accessible in real life. Therefore, a way to deal with this issue is necessary and a first approach might be to simplify this loss function. Considerations in the context of ROPE lead to such a simplification.

By construction, under the null value θ_0 the loss of a_0 is smaller than the loss of a_1, i.e. $L_P(\theta_0, \mathsf{a}_0) < L_P(\theta_0, \mathsf{a}_1)$, as former action is in accordance and latter action in discordance with the null value.

If not specifying the exact values of the loss function L_P, the researcher might (or should) still be able to determine the appropriate action for each parameter value $\theta \in \Theta$. In that, there is a set Θ_0 of parameter values for which a_0 is appropriate (containing the null value θ_0) and a set $\Theta_1 = \Theta \setminus \Theta_0$ of the remaining parameter values for which a_1 is appropriate. The first set Θ_0 is the ROPE and usually an interval.

However, different parameter values within these sets, respectively, might still have different loss values. As these exact values are still hardly accessible in real life, a possible simplification is to treat each parameter value within the ROPE Θ_0 as "equivalent to the null value for practical purposes" [Kruschke, 2018, p. 272], i.e. assuming identical loss values for parameters within Θ_0:

$$\forall \theta_i, \theta_j \in \Theta_0 \, \forall \mathsf{a} \in \mathcal{A}_P : L_P(\theta_i, \mathsf{a}) = L_P(\theta_j, \mathsf{a}) \,. \tag{6}$$

In addition to the paramete values within the ROPE Θ_0, also the parameter values outside the ROPE, i.e. within Θ_1, might be treated as equivalent for practical purposes by

Table 1: Simplified loss function for the actions of the practical purpose using a regret form.

$L_P(\theta, \mathsf{a})$	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$\mathsf{a} = \mathsf{a}_0$	0	k_0
$\mathsf{a} = \mathsf{a}_1$	k_1	0

employing identical loss values:

$$\forall \theta_i, \theta_j \in \Theta_1 \; \forall \mathsf{a} \in \mathcal{A}_P : L_P(\theta_i, \mathsf{a}) = L_P(\theta_j, \mathsf{a}) \,. \tag{7}$$

In that, this simplified loss function needs only four values to be specified and, without loss of generality, a regret form might be employed, in which a_0 and a_1 have zero loss if $\theta \in \Theta_0$ and $\theta \in \Theta_1$, respectively. The remaining two loss values shall be denoted by (see Table 1)

$$\begin{aligned} k_0 &:= L_P(\theta, \mathsf{a}_0) \quad \forall \theta \in \Theta_1 \\ k_1 &:= L_P(\theta, \mathsf{a}_1) \quad \forall \theta \in \Theta_0 \,. \end{aligned}$$

Using this simplification, the expected posterior loss of each action $\mathsf{a} \in \mathcal{A}_P$ is

$$\begin{aligned} \rho(\mathsf{a}_0) &= \int_\Theta L_P(\theta, \mathsf{a}_0) \pi(\theta|\boldsymbol{x}) d\theta \\ &\overset{\substack{eq.\\(6)\\(7)}}{=} 0 \cdot \int_{\theta \in \Theta_0} \pi(\theta|\boldsymbol{x}) d\theta + k_0 \cdot \int_{\theta \in \Theta_1} \pi(\theta|\boldsymbol{x}) d\theta \\ &= k_0 \cdot p(\theta \in \Theta_1 | \boldsymbol{x}) \end{aligned}$$

and analogously

$$\rho(\mathsf{a}_1) = k_1 \cdot p(\theta \in \Theta_0 | \boldsymbol{x}) \,. \tag{8}$$

With $k := k_1/k_0$, the ratio of expected posterior losses is

$$\varrho(k) := \frac{\rho(\mathsf{a}_1)}{\rho(\mathsf{a}_0)} = \frac{k_1 \cdot p(\theta \in \Theta_0|\boldsymbol{x})}{k_0 \cdot p(\theta \in \Theta_1|\boldsymbol{x})} = k \cdot \frac{p(\theta \in \Theta_0|\boldsymbol{x})}{p(\theta \in \Theta_1|\boldsymbol{x})} \tag{9}$$

and the corresponding Bayes action is

$$\mathsf{a}_{\mathsf{Bayes}}(k) = \begin{cases} \mathsf{a}_0 & \text{if } \varrho(k) > 1 \\ \mathsf{a}_1 & \text{if } \varrho(k) < 1 \end{cases} \,. \tag{10}$$

If $\varrho(k) = 1$, then either action might be chosen.

The term $p(\theta \in \Theta_0|\boldsymbol{x})/p(\theta \in \Theta_1|\boldsymbol{x})$ can be calculated simply from the posterior density $\pi(\theta|\boldsymbol{x})$, however, k need to be specified w.r.t. to the practical purpose.

3.2 Framework of Imprecise Probabilities

Specifying a precise value for k, which defines the simplified loss function, might still be difficult for applied scientists and ideas from the framework of imprecise probabilities [Walley, 1991] come in handy. In addition, the foundations of the ROPE-only decision rule can be depicted elegantly within this framework.

In general, this framework is based on the fact that there is more to uncertainty than can be captured within precise probability values [e.g. Ellsberg, 1961, Levi, 1980, Walley, 1991, Etner et al., 2012]. As a solution, sets or intervals of probability values, so called imprecise probabilities, are employed instead of single precise probability values. These intervals are treated as an entity of its own [c.p. Walley, 1991] and numerous sources on how to calculate with imprecise probabilities are available [see e.g. Augustin et al., 2014, for a depiction of the state of the art within different fields of application at that time]. Naturally, this framework is appropriate whenever some relevant but potentially vague information about probabilities is available, yet it is not enough to unambiguously specify exact probability values. For example, within the Bayesian context, a researcher might be unable to specify the exact shape of a prior distribution and several different distributions are in accordance with the available prior knowledge. By comprising all these potential distributions within a set of distributions, the researcher obtains an imprecise prior distribution, which reflects the available knowledge and uncertainty as is, without pretending a level of precision that is not available [see also the framework of robust Bayesian statistics, e.g. Ríos Insua and Ruggeri, 2012].

Similarly, in the context of a real-world decision, some potentially vague information about potential consequences is supposed to be available. Yet, an applied scientist is usually unable to unambiguously specify a precise loss function as several different loss functions might be in accordance with the available (vague) information. An arbitrary specification of a loss function will result in an arbitrary decision. Not employing a loss function at all, on the other hand, leads to a decision that lacks a relation to the underlying real-world situation and is therefore arbitrary as well. In that, it seems obvious that partially available information about the loss function has to be included into the analysis in the form it is available.

Thus, analogue to imprecisely specified probabilities, the loss function might be specified imprecisely. In the context of the simplified loss function as depicted in Section 3.1, instead of a precise value k, an open[7] interval of values $K = \left(\underline{K}, \overline{K}\right)$ might be employed, where $\underline{K}$ and $\overline{K}$ denote the lower and upper bound, respectively, for stating how much "worse" a_1 would be if $\theta \in \Theta_0$ than a_0 would be if $\theta \in \Theta_1$ (if deciding correctly has zero "badness"). As every value $k \in K$ defines a different (simplified) loss function, the interval K defines a set of loss functions. For each of those loss functions, i.e. for every $k \in K$, it is possible

[7]Of course, K might also be specified by a closed interval $\left[\underline{K}, \overline{K}\right]$ and in many situations this might be more reasonable. However, in order to derive the ROPE-only decision rule, as stated by Kruschke [2018] within his supplementary material, K needs to be an open interval.

to determine the Bayes action $\mathsf{a}_{\mathsf{Bayes}}(k)$, once the data are available and the posterior distribution of θ is calculated. If the Bayes action $\mathsf{a}_{\mathsf{Bayes}}(k)$ is the same for all $k \in K$, then this action should be selected, else information is lacking to unambiguously guide a decision and a decision should be withheld.

Formally, an interval-valued ratio of expected posterior losses

$$\big(\varrho(\underline{K}), \varrho(\overline{K})\big) \tag{11}$$

is obtained by considering the interval-valued K, leading to the Bayes action

$$\mathsf{a}_{\mathsf{Bayes}}(K) = \begin{cases} \mathsf{a}_0 & \text{if } \varrho(\underline{K}) > 1 \\ \mathsf{a}_1 & \text{if } \varrho(\overline{K}) < 1 \end{cases} . \tag{12}$$

For $\varrho(\underline{K}) \leq 1 \leq \varrho(\overline{K})$, the decision should be withheld.

3.3 An Arbitrary Choice

By setting K arbitrarily to $K = (1/19, 19)$, the ROPE-only decision rule is obtained, because – according to the imprecise decision theoretic framework, especially considering equation (12) – action a_0 ("accept the null value for practical purposes") is optimal if

$$\begin{aligned} & \varrho(1/19) > 1 \\ \Leftrightarrow\quad & \frac{1}{19} \cdot \frac{p(\theta \in \Theta_0 | \boldsymbol{x})}{p(\theta \in \Theta_1 | \boldsymbol{x})} > 1 \\ \Leftrightarrow\quad & p(\theta \in \Theta_0 | \boldsymbol{x}) > 19 \cdot p(\theta \in \Theta_1 | \boldsymbol{x}) \\ \Leftrightarrow\quad & p(\theta \in \Theta_0 | \boldsymbol{x}) > 19 \cdot (1 - p(\theta \in \Theta_0 | \boldsymbol{x})) \\ \Leftrightarrow\quad & 20 \cdot p(\theta \in \Theta_0 | \boldsymbol{x}) > 19 \\ \Leftrightarrow\quad & p(\theta \in \Theta_0 | \boldsymbol{x}) > 0.95 \end{aligned}$$

and, analogously, action a_1 ("reject the null value for practical purposes") is optimal if

$$\begin{aligned} & \varrho(19) < 1 \\ \Leftrightarrow\quad & p(\theta \in \Theta_0 | \boldsymbol{x}) < 0.05\,, \end{aligned}$$

which reflect exactly those conditions defining the ROPE-only decision rule.

In any other case, i.e. for $0.05 \leq p(\theta \in \Theta_0 | \boldsymbol{x}) \leq 0.95$, both the imprecise decision theoretic framework using $K = (1/19, 19)$ and the ROPE-only decision rule recommend to withhold a decision.

4 Foundations of the HDI+ROPE decision rule

4.1 Action Space and Decision Rule

The general idea of the decision theoretic foundation of the HDI+ROPE decision rule was described within the supplementary material by Kruschke [2018, supp. p. 3–5]. However, some aspects depicted there are merely preliminary[8], so this technical report intends to outline this foundation more profoundly. In line with this idea and the considerations depicted by Rice et al. [2008] (which are also referred to by Kruschke [2018]), the corresponding action in the context of the HDI+ROPE decision rule comprises two aspects:

- the determination of the HDI and
- the assessment of the relation between the HDI and the ROPE (inside, outside, or overlap).

The action space w.r.t. the first aspect – indicated with subscript I for "interval" – contains all possible closed parameter intervals

$$\mathcal{A}_I = \{[a,b] | a, b \in \Theta, a < b\} \tag{13}$$

and the objective is to decide for the element within $\mathcal{A}_I$ that is the HDI.

The action space w.r.t. the second aspect – indicated with subscript R for "relation" – contains all three possible relations between a parameter interval and a predefined ROPE:

$$\mathcal{A}_R = \{\mathsf{r}_0, \mathsf{r}_1, \mathsf{r}_2\} \tag{14}$$

with

r_0: The parameter interval falls completely within the ROPE.

r_1: The parameter interval falls completely outside the ROPE.

r_2: The parameter interval and the ROPE overlap.

In conjunction, the overall action space is $\mathcal{A}_I \times \mathcal{A}_R$ and the corresponding decision rule maps the sample space $\mathcal{X}$ to this action space:

$$\delta_{HDI+ROPE} : \mathcal{X} \to \mathcal{A}_I \times \mathcal{A}_R \,. \tag{15}$$

[8] Within equation (1) on page 4 within Kruschke [2018]'s supplementary material, the argument s of the function $\mathbf{1}(s)$ is sometimes a set, yet it should be a statement. The explanation of one of the terms states "cost of reject if HDI overlaps ROPE" [Kruschke, 2018, supp. p. 4 eq. (1)], yet the term might rather refer to a cost of rejection if the HDI is within the ROPE. As outlined within this technical report (see esp. equation (30)), the cost of deciding correctly should be identical for each relation between HDI and ROPE, which is not necessarily the case in Kruschke's formula (1).

In that regard, Kruschke [2018] states that his ideas are "merely suggestive" [supp. p. 4] and his "goal is [only] to point out that formal expressions are possible for the loss implicit to the intuitive HDI+ROPE rule" [supp. p. 5]. In that, the elaborations within this technical report are based on this initial work by Kruschke [2018].

The exact shape of this decision rule

$$\delta_{HDI+ROPE}(\boldsymbol{x}) = \begin{pmatrix} \delta_I(\boldsymbol{x}) \\ \delta_R(\delta_I(\boldsymbol{x})) \end{pmatrix} \tag{16}$$

can be depicted using the functions

$$\delta_I : \mathcal{X} \to \mathcal{A}_I\,, \tag{17}$$

which maps the data $\boldsymbol{x}$ to the corresponding HDI, and

$$\delta_R : \mathcal{A}_I \to \mathcal{A}_R\,, \tag{18}$$

which maps an interval in parameter space $[a, b] \in \mathcal{A}_I$ to its correct relation with a predefined ROPE Θ_0 by

$$\delta_R([a,b]) = \begin{cases} \mathsf{r}_0 & \text{if } [a,b] \cap \Theta_0 = [a,b] \\ \mathsf{r}_1 & \text{if } [a,b] \cap \Theta_0 = \emptyset \\ \mathsf{r}_2 & \text{if } [a,b] \cap \Theta_0 \neq [a,b] \wedge [a,b] \cap \Theta_0 \neq \emptyset \end{cases}\,. \tag{19}$$

4.2 Loss Function

4.2.1 Determination of HDI

It is possible to state a loss function for which the determination of the HDI δ_I is a Bayes rule, namely [see e.g. Schervish, 1995, Rice et al., 2008]

$$L_I : \Theta \times \mathcal{A}_I \to \mathbb{R}_0^+ : L_I(\theta, [a,b]) = (b-a) + c \cdot \mathbf{1}(\theta \notin [a,b])\,, \tag{20}$$

where $\mathbf{1}(s) = 1$ if the statement s is true and $\mathbf{1}(s) = 0$ if s is false. The value c denotes a constant which determines the minimum density of a parameter to be included within the HDI (see below).

The expected posterior loss w.r.t. this loss function is

$$\begin{aligned} \rho_I([a,b]) &= \int_\Theta L_I(\theta, [a,b])\pi(\theta|\boldsymbol{x})d\theta \\ &= \int_\Theta \left[(b-a) + c \cdot \mathbf{1}(\theta \notin [a,b])\right]\pi(\theta|\boldsymbol{x})d\theta \\ &= (b-a) + c\int_\Theta \mathbf{1}(\theta \notin [a,b])\pi(\theta|\boldsymbol{x})d\theta \\ &= (b-a) + c\int_{\Theta\setminus[a,b]} \pi(\theta|\boldsymbol{x})d\theta \end{aligned}$$

and minimizing this expected posterior loss over the action space $\mathcal{A}_I$ yields as Bayes action the interval $[a, b]$ that contains all parameters with posterior density larger than c^{-1} [see e.g. Schervish, 1995, Rice et al., 2008]. By setting c appropriately, the 95%-HDI is obtained as Bayes action for a given data set $\boldsymbol{x}$ and the decision rule δ_I is a Bayes rule w.r.t. the loss function L_I.

4.2.2 Relation between HDI and ROPE

It is also possible to state a loss function L_R for which the assessment of the relation between a parameter interval $[a,b]$ and a predefined ROPE Θ_0 is a Bayes rule.

As outlined in Section 2, a loss function is defined on the parameter space Θ and on the action space, which is $\mathcal{A}_R$ (as defined in equation (14)) within this context. However, the employed loss function will depend on the ROPE Θ_0 and the parameter interval $[a,b]$ as well. Although the ROPE might be treated as given, this is not the case for the parameter interval $[a,b]$, especially when considering the overall decision rule $\delta_{HDI+ROPE}$ (as in the following Section 4.2.3). Accordingly, this dependence of L_R on $[a,b] \in \mathcal{A}_I$ needs to be taken into account, so that

$$L_R : \Theta \times \mathcal{A}_R \times \mathcal{A}_I \to \mathbb{R}_0^+ : (\theta, \mathsf{r}, [a,b]) \mapsto L_R^{[a,b]}(\theta, \mathsf{r})\,. \tag{21}$$

Considering δ_R in isolation, as within this subsection, also $[a,b]$ might be treated as given.

Although this loss function is technically defined using the parameter space Θ, this dependence is not necessary:

$$\forall \mathsf{r} \in \mathcal{A}_R : \forall \theta_i, \theta_j \in \Theta : L_R^{[a,b]}(\theta_i, \mathsf{r}) = L_R^{[a,b]}(\theta_j, \mathsf{r}) =: L_R^{[a,b]}(\mathsf{r})\,. \tag{22}$$

A candidate of this loss function is depicted in Table 2, formally stated as

$$\begin{aligned}
L_R^{[a,b]}(\mathsf{r}) = {} & c_1 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_0) \cdot \mathbf{1}([a,b] \cap \Theta_0 = [a,b]) + \\
& c_1 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_1) \cdot \mathbf{1}([a,b] \cap \Theta_0 = \emptyset) + \\
& c_1 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_2) \cdot \mathbf{1}([a,b] \cap \Theta_0 \neq [a,b] \wedge [a,b] \cap \Theta_0 \neq \emptyset) + \\
& c_2 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_0) \cdot \mathbf{1}([a,b] \cap \Theta_0 = \emptyset) + \\
& c_2 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_0) \cdot \mathbf{1}([a,b] \cap \Theta_0 \neq [a,b] \wedge [a,b] \cap \Theta_0 \neq \emptyset) + \\
& c_2 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_1) \cdot \mathbf{1}([a,b] \cap \Theta_0 = [a,b]) + \\
& c_2 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_1) \cdot \mathbf{1}([a,b] \cap \Theta_0 \neq [a,b] \wedge [a,b] \cap \Theta_0 \neq \emptyset) + \\
& c_2 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_2) \cdot \mathbf{1}([a,b] \cap \Theta_0 = [a,b]) + \\
& c_2 \cdot \mathbf{1}(\mathsf{r} = \mathsf{r}_2) \cdot \mathbf{1}([a,b] \cap \Theta_0 = \emptyset)\,,
\end{aligned} \tag{23}$$

where c_1, c_2 are arbitrary positive constants with $c_1 < c_2$ and, again, $\mathbf{1}(s) = 1$ if the statement s is true and $\mathbf{1}(s) = 0$ if s is false.

As L_R does not depend on the parameter θ, the expected posterior loss ρ_R of each action

Table 2: Loss function for finding the relation r between a parameter interval and a ROPE (see equation 23).

$L_R^{[a,b]}(\mathsf{r})$		Interval to ROPE		
		within	outside	overlap
Decision	$\mathsf{r} = \mathsf{r}_0$	c_1	c_2	c_2
	$\mathsf{r} = \mathsf{r}_1$	c_2	c_1	c_2
	$\mathsf{r} = \mathsf{r}_2$	c_2	c_2	c_1

$\mathsf{r} \in \mathcal{A}_R$ w.r.t. this loss function is the loss value itself:

$$\begin{aligned}\rho_R(\mathsf{r}) &= \int_\Theta L_R^{[a,b]}(\theta, \mathsf{r})\pi(\theta|\boldsymbol{x})d\theta \\ &= \int_\Theta L_R^{[a,b]}(\mathsf{r})\pi(\theta|\boldsymbol{x})d\theta \\ &= L_R^{[a,b]}(\mathsf{r}) \int_\Theta \pi(\theta|\boldsymbol{x})d\theta \\ &= L_R^{[a,b]}(\mathsf{r}) \end{aligned} \tag{24}$$

Minimizing ρ_R over the action space $\mathcal{A}_R$ results in the relation r that is obtained by the decision rule δ_R in equation (19). In that, this decision rule δ_R is a Bayes rule w.r.t. the loss function L_R for all parameter intervals $[a,b] \in \mathcal{A}_I$.

The loss function L_R was defined using only two different values c_1 and c_2. In that, two restrictions are imposed on the loss function L_R:

(I) Deciding correctly has the same loss c_1 independent of which relation is true.

(II) Deciding falsely has the same loss c_2 independent of which relation is true and which incorrect relation was chosen.

Of course, it would be possible to employ a loss function without these restrictions that uses e.g. nine different loss values instead of only two. However, the first restriction (I) will be necessary for combining both loss functions L_I and L_R, because then, independent of the parameter value $\theta \in \Theta$, for every potential interval $[a,b] \in \mathcal{A}_I$ the decision rule δ_R yields a relation $\mathsf{r} \in \mathcal{A}_R$ with the identical loss c_1, i.e.

$$\forall\theta \in \Theta \ \forall[a,b] \in \mathcal{A}_I : L_R^{[a,b]}(\theta, \delta_R([a,b])) = c_1\,, \tag{25}$$

a fact that will be referred to later.

The second restriction (II) is employed both out of convenience and to emphasize an important characteristic: Assume that up to six different values larger than c_1 would be employed instead of c_2 within the loss function L_R in equation (23). Still, the expected posterior loss $\rho_R(\mathsf{r})$ of each action r is the loss value $L_R^{[a,b]}(\mathsf{r})$ itself (equation (24)) and, again, the minimization leads to the action obtained by δ_R. In that, the decision rule

δ_R is a Bayes rule w.r.t. this loss function independent of the exact values that are used instead of c_2. The exact specification of these values does not contribute to guiding the decision about the relation $\mathsf{r} \in \mathcal{A}_R$, only the fact that they are larger than c_1. Therefore, if important information is incorporated within these values this information will not be used for guiding the decision, so a single value c_2 might be employed out of convenience.

4.2.3 Overall

Adding both loss functions lead to

$$L_{HDI+ROPE} : \Theta \times (\mathcal{A}_I \times \mathcal{A}_R) \to \mathbb{R}_0^+ \tag{26}$$

with

$$L_{HDI+ROPE}(\theta, ([a,b], \mathsf{r})) = L_I(\theta, [a,b]) + L_R^{[a,b]}(\theta, \mathsf{r}) \tag{27}$$

for which the overall decision rule $\delta_{HDI+ROPE}$ is a Bayes rule.

This can be seen by considering the expected posterior loss

$$\begin{aligned}
\rho_{HDI+ROPE}([a,b], \mathsf{r}) &= \int_\Theta L_{HDI+ROPE}(\theta, ([a,b], \mathsf{r}))\pi(\theta|\boldsymbol{x})d\theta \\
&= \int_\Theta \left[L_I(\theta, [a,b]) + L_R^{[a,b]}(\mathsf{r}) \right] \pi(\theta|\boldsymbol{x})d\theta \\
&= \int_\Theta L_I(\theta, [a,b])\pi(\theta|\boldsymbol{x})d\theta + \int_\Theta L_R^{[a,b]}(\mathsf{r})\pi(\theta|\boldsymbol{x})d\theta \\
&= \rho_I([a,b]) + L_R^{[a,b]}(\mathsf{r})\,.
\end{aligned} \tag{28}$$

The corresponding Bayes action

$$\underset{([a,b],\mathsf{r})\in\mathcal{A}_I\times\mathcal{A}_R}{\arg\min} \quad \rho_I([a,b]) + L_R^{[a,b]}(\mathsf{r})\,, \tag{29}$$

is obtained by minimizing this expected posterior loss.

The first part $\rho_I([a,b])$ does not depend on $\mathsf{r} \in \mathcal{A}_R$ and as outlined in in the previous Section 4.2.2, for all possible parameter intervals $[a,b]$, the second part $L_R^{[a,b]}(\mathsf{r})$ can be minimized by choosing its correct relation $\mathsf{r} \in \mathcal{A}_R$ with the predefined ROPE, which is obtained by the decision rule δ_R. Therefore, for any parameter interval $[a,b]$, the optimal relation is $\mathsf{r} = \delta_R([a,b])$.

The optimal parameter interval $[a,b] \in \mathcal{A}_I$ can now be obtained as

$$\begin{aligned}
&\underset{[a,b]\in\mathcal{A}_I}{\arg\min}\ \rho_I([a,b]) + L_R^{[a,b]}(\delta_R([a,b])) \\
\overset{\substack{eq.\\(25)}}{=}\ &\underset{[a,b]\in\mathcal{A}_I}{\arg\min}\ \rho_I([a,b]) + c_1 \\
=\ &\underset{[a,b]\in\mathcal{A}_I}{\arg\min}\ \rho_I([a,b])\,,
\end{aligned} \tag{30}$$

which is the 95%-HDI – for an appropriate choice of c (see Section 4.2.1).

Taken together, as stated at the beginning of this subsection, the Bayes action $([a, b], \mathsf{r})$ w.r.t. $L_{HDI+ROPE}$ is obtained by $\delta_{HDI+ROPE}$.

The first restriction (I) mentioned in the previous Section 4.2.2 (equation (25)) is employed for finding the optimal parameter interval in the overall case, i.e. within equation (30). Without this restriction (I), $L_R^{[a,b]}(\delta_R([a, b]))$ would not be a constant c_1, but a value that depends on the interval $[a, b]$. In that, the minimization in equation (30) could yield an interval, which is not the HDI, a fact that is referred to as "paradoxical behavior" by Kruschke [2018, supp. p. 4] (who also refers to Casella et al. [1993] in this context).

4.3 Final Decision

In contrast to the foundation of the ROPE-only decision rule, the loss function $L_{HDI+ROPE}$ does not allow a reasonable employment of the framework of imprecise probabilities. This is because L_I need to be as it is in order to obtain the HDI and L_R uses only the fact that c_1 is smaller than c_2. As depicted in Section 4.2.2, any additional information within these constants is not being used. Therefore, no potentially vague information can be captured within $L_{HDI+ROPE}$. As a consequence, the framework of imprecise probabilities cannot be employed within this context to elegantly formalize withholding a decision between a_0 and a_1. Therefore, the action space of the final decision comprises all a_0, a_1 and the action to withhold the decision.

Of course, in the context of the HDI+ROPE decision rule, there is a bijective mapping between $\mathcal{A}_R$ and the action space for this final decision:

$$\mathsf{r}_0 \mapsto \mathsf{a}_0 \qquad \mathsf{r}_1 \mapsto \mathsf{a}_1 \qquad \mathsf{r}_2 \mapsto \text{withhold decision}\,. \tag{31}$$

Accordingly, this last step does not need a separate decision theoretic account, as the final actions might be employed instead of the three relations $\mathsf{r} \in \mathcal{A}_R$.

Nevertheless, from a content point of view, this final step should be treated separately from the determination of the relation between the HDI and the ROPE. As outlined within this Section 4, the HDI+ROPE decision rule is primarily focusing on technical aspects of how to obtain the HDI and determine its relation with a predefined ROPE, and it is this final step that tries to build the connection to the underlying real-world decision of interest.

5 Discussion

The decision theoretic foundations of both Kruschke's HDI+ROPE decision rule [Kruschke, 2015, 2018] and the ROPE-only decision rule [Kruschke, 2018, supp. p. 5] are outlined within this technical report. Both decision rules are depicted as Bayes rules w.r.t. certain loss functions. In that, different loss functions are considered: First, although inaccessible, there is an underlying "true" loss function characterizing the real-world decision of interest.

Second, in the context of considerations about the ROPE (see Section 3.1), this "true" loss function is simplified, such that it might be specified by only a single number. Still, this simplified loss function characterizes the real-world decision of interest. Third, there is a loss function w.r.t. to finding the HDI for a given data set and, fourth, a rather artificial loss function might be employed in the context of determining the relation between a parameter interval and a pre-defined ROPE. Fifth, the loss function in the context of the HDI+ROPE decision rule is a combination of the previous two.

Naturally, by considering these different loss functions, different decision rules are characterized as Bayes rules. Put aptly by Rice et al. [2008, p. 3], "for the precise 'question' asked by loss function L and the stated modeling assumptions, one can think of the Bayes rules as providing the 'best' answer". In that, these five loss functions are asking:

- How should I decide in the real-world decision problem?
- Given the simplification, how should I decide in the real-world decision problem?
- Which interval is the HDI of the posterior distribution?
- How is the relation of the HDI and the ROPE?
- Which interval is the HDI of the posterior distribution and how is the relation of it with the ROPE?

The first question is of interest but cannot be answered, because the loss function is inaccessible. The second question does relate to the real-world decision of interest and might be used as a proxy for the first question (given the employed simplification is reasonable), as the corresponding simplified loss function still contains information w.r.t. to the real-world decision of interest. By allowing this loss function to be specified imprecisely, relevant information might be incorporated into the analysis as it is available. In this context, the ROPE-only decision rule is optimal when resorting to an arbitrary choice of interval-valued loss functions.

The third question does not address the real-world decision of interest at all. Although the fourth question contains the ROPE, the corresponding loss function considers only the bounds of the ROPE and not respective loss values that are in accordance with the real-world decision of interest (as within the second (simplified) loss function). In that, the fourth question relates to the real-world decision of interest only marginally and primarily addresses a rather technical interval comparison. Therefore, as a combination of the previous two, the fifth question does not primarily ask about the real-world decision problem, yet is implicitly used as a proxy for it when employing the HDI+ROPE decision rule.

In summary, the ROPE-only decision rule might be characterized as being closer to the real-world decision of interest than the HDI+ROPE decision rule. This might also be seen by the fact, that both the "true" underlying loss function and the posterior distribution are essential to derive the optimal decision in a Bayesian framework, yet the HDI+ROPE decision rule uses less of these information than the ROPE-only decision rule: First, former simplifies the posterior distribution by sorting back to the less-informative HDI. Second, former employs only the bounds of the ROPE and latter also information about the loss-magnitude.

Of course, the arbitrary choice about the loss value interval within the simplified loss function in the context of the ROPE-only decision rule (see Section 3.3) has to be criticized. The corresponding interval should be chosen based on the real-world decision of interest. As it is to expect that at least some information about this loss value in the simplified loss function is available[9], the framework of imprecise probabilities offers an elegant way to include this essential but vague information.

References

T. 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 J.-M. Bernard, T. Seidenfeld, and M. Zaffalon, editors, *ISIPTA '03: Proceedings of the Third International Symposium on Imprecise Probabilities and their Applications*, pages 31–45, Lugano, Waterloo, 2003. Carleton Scientific.

T. Augustin, F. P. A. Coolen, G. De Cooman, and e. Troffaes, Matthias C. M. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

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

G. Casella, J. T. G. Hwang, and C. Robert. A paradox in decision-theoretic interval estimation. *Statistica Sinica*, 3(1):141–155, 1993.

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

D. Ellsberg. Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 75:643–669, 1961.

J. Etner, M. Jeleva, and J.-M. Tallon. Decision theory under ambiguity. *Journal of Economic Surveys*, 26(2):234–270, 2012.

J. K. Kruschke. *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan*. Academic Press, 2015.

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

I. Levi. *The enterprise of knowledge: an essay on knowledge, credal probability, and chance*. MIT Press, Cambridge, 1980.

[9] Guiding a decision without any information at all about the consequences is truly arbitrary. If so, it is indispensable to put effort into obtaining this information first.

K. M. Rice, T. Lumley, and A. A. Szpiro. Trading bias for precision: decision theory for intervals and sets. Technical report, 2008. Retrieved from `https://biostats.bepress.com/uwbiostat/paper336/`.

D. Ríos Insua and F. Ruggeri, editors. *Robust Bayesian analysis*. Springer Science & Business Media, 2012.

C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer, 2007.

M. J. Schervish. *Theory of statistics*. Springer, 1995.

P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman & Hall, 1991.