A computation method in robust Bayesian decision theory

Christophe Abraham

UMR Analyses des systèmes et biométrie, Montpellier SupAgro, France

A R T I C L E   I N F O

Article history:
Available online 15 April 2008

Keywords:
Global robustness
Class of utility functions
Optimal decisions
Discretization
Monte Carlo optimization
Simulated annealing

A B S T R A C T

We propose a method for computing the range of the optimal decisions when the utility function runs through a class $U$. The class $U$ has constraints on the values and the shape of the utility functions. A discretization method enables to easily approximate the optimal decision associated with a particular utility function $u \in U$. The range of optimal decisions is computed by a Monte Carlo optimization method. An example is provided with numerical results.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

In many practical situations, we have to deal with making decisions under uncertainty. Examples include decision making in natural resource management (choosing the amount of trees or fishes to remove, choosing the height of a dam to prevent flood damage), in medicine (choosing a dose of a medical treatment which balances between efficiency and toxicity effects), etc.

Bayesian decision theory provides a framework for making decision under uncertainty. The elements of a Bayesian decision analysis are: an available decision $d$ in a set of decisions $D$, a state of nature (or parameter) $h$ in a set $H$, a prior distribution $p$ on $H$, an observation $x$ with density $p_h$, and an utility function $U(h,d)$. The prior distribution $p$ represents the prior knowledge on the parameter $h$. This prior knowledge is updated by Bayes's theorem to provide the posterior distribution. Thus, the posterior distribution combines the prior information ($p$) with the information provided by the observation. The utility function $U$ is a function from $D \times H$ to $\mathbb{R}$, it quantifies the utility of choosing the decision $d$ when the value of the parameter is $h$: choosing $d_1$ when the value of the parameter is $\theta_1$ is preferred to choosing $d_2$ when the value of the parameter is $\theta_2$ if and only if $U(\theta_1, d_1) > U(\theta_2,d_2)$. Sometimes, a loss function is used instead of a utility function; the two approaches are equivalent as the loss can be defined as the opposite of the utility. For an account on Bayesian decision theory, we refer the reader to [1–3]. To be concrete, let us consider the following example.

Example 1. Suppose we have to choose the height $d$ of a dam. Let $x = (x_1, \ldots, x_n)$ be some previous peak water levels with density $p_h(x)$. The distribution of the peak water levels is determined by $\theta$ and the posterior $\pi_h$ represents the updated information on $\theta$. The number $U(d, \theta)$ quantifies the utility of constructing a dam $d$ meters high when the value of the parameter is $\theta$.

According to Bayesian decision theory, an optimal decision (also called a Bayes action or a Bayes alternative) is any decision $d$ which maximizes the posterior expected utility. In other words, a decision $d_0$ is optimal if

E-mail address: abraham@supagro.inra.fr

0888-613X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2008.03.021
\[ U(\theta, d) = \int_a U(\theta, d) \pi_\theta(d\theta) = \sup_{d \in \mathcal{D}} \int_a U(\theta, d) \pi_\theta(d\theta). \] (1.1)

The elicitation of the prior distribution is a difficult task as well as the elicitation of the utility function. In practice, many prior distributions and many utility functions can approximately fit the prior knowledge and the preferences of the decision maker. Thus, it is of interest to know whether any two candidates for the prior distribution or the utility function yield nearly equivalent decisions or, on the contrary, very different ones. This is the concern of Bayesian robustness. In global robustness (a special kind of Bayesian robustness), the (single) prior and/or utility is replaced by a class of priors and/or utilities. The sensitivity of the analysis is then measured by the range of posterior quantities when the prior and/or the utility run through the class. For an account on the theory, we refer the reader to [4,5] and the references therein.

In this paper, we are interested in global robustness with respect to (w.r.t.) the utility function. Papers on utility robustness are rare compared with the large literature on prior robustness. References can be found in the works cited above. Several measures of robustness have been considered to investigate the sensitivity of the analysis w.r.t. the utility. Examples of such measures can be found in [1,6–9]. From a computational point of view, a challenging measure is the range of optimal decisions when the utility function ranges over a set \( \mathcal{U} \). Assume \( \mathcal{D} \subset \mathbb{R} \) and set

\[ d^+_u = \sup_{u \in \mathcal{U}} d^+_u \quad \text{and} \quad d^-_u = \inf_{u \in \mathcal{U}} d^-_u, \] (1.2)

where \( d^+_u \) and \( d^-_u \) are, respectively, the largest and the lowest optimal decision defined by (1.1) if the utility is \( U \) (note that the optimal decisions associated with \( U \) need not be unique in general). This measure has been considered by some authors but the computation of \( d^+_u \) and \( d^-_u \) has not been achieved only for particular classes of utilities. In [10,11] classes with bounded derivatives w.r.t. the decision \( d \) are considered. For such classes, it can be proved that \( d^+_u \) and \( d^-_u \) are the optimal decisions for two particular utility functions explicitly known, hence an easy computation of \( d^+_u \) and \( d^-_u \) is possible. By the same way, \( d^+_u \) and \( d^-_u \) could be easily computed for parametric classes of utility functions of the type \( \{U_w, w \in \mathcal{W}\} \) at least when the dimension of \( \mathcal{W} \) is not too high. These classes are computationally convenient but do not fit exactly the information provided by the usual assessment methods as it is shown in Section 2.

Some authors are interested in computing the nondominated set instead of the Bayes actions set. In [12], situations where the nondominated set coincides with \( [d^-_u, d^+_u] \) are given (this is true, in particular, when the utilities are strictly concave functions of \( d \)).

It is worth pointing out that stochastic optimization techniques were not fully exploited in Bayesian global robustness. Truly, classical classes of priors like \( \epsilon \)-contamination neighborhoods [13] or interval of measures [14] provide to some extend explicit solutions for the optimization problems involved. More recently, optimizations techniques such as simulated annealing or Monte Carlo algorithms are used in the context of imprecise probabilities [15–17].

We provide in this paper a method for solving optimization problems of the form (1.2). The class of utility functions have constraints both on values and shape (or derivatives). We emphasize that these two kinds of constraints are antagonistic from an optimization point of view. We use a simulated annealing algorithm for computing the largest and lowest optimal decisions. The advantage of using Monte Carlo optimization methods lies in the fact that the constraints can be easily included in the proposed distribution.

In Section 2, we briefly review the construction of a utility function from a usual assessment method and show that such a construction yields a class \( \mathcal{H} \) of utility functions (with particular constraints) rather than a single utility function. Section 3 presents a discretization method that enables to approximate the optimal decision for a given utility function \( u \in \mathcal{H} \) with few numerical computations. Some results on the accuracy of the approximations are also given. In Section 4, we use the results of Section 3 to solve (1.2) by a Monte Carlo optimization method. Every section is illustrated by an example with numerical computations. Auxiliary results and proofs (except for Proposition 4) are gathered in Section 5. Section 6 contains a short discussion.

2. Construction of an utility function

Let us give a brief exposition of the construction of a utility function \( U \). Each pair \( (\theta, d) \) is associated with a random consequence \( R \) whose distribution \( \xi_{\theta,d} \) depends on \( (\theta, d) \). For instance, in Example 1, we can set \( R = d - H \) where the random variable \( H \) represents the peak water level for the next year. Let \( \mathcal{H} \) be the set of consequences and assume that there exists a function \( u \) on \( \mathcal{H} \) such that \( u(r) \) is the utility of \( r \). Following [1], \( U \) can be constructed as follows:

\[ U(\theta, d) = \int_u u(r) \xi_{\theta,d}(dr). \] (2.1)

The choice of \( u \) is not straightforward. Its assessment is done by several comparisons between sure consequences \( r \) and random consequences of the form \( \alpha r_1 + (1-\alpha) r_2 \) where \( \alpha \) takes the value \( \alpha _1 \) with probability \( \alpha \) and \( \alpha _2 \) with probability \( 1-\alpha \). The range of values of \( u \) can be fixed in the following way. Assume that \( r_1 \) and \( r_2 \) are, respectively, the worst and the best possible consequences in \( \mathcal{H} \). Set arbitrarily \( u(r_1) = 0 \) and \( u(r_2) = 1 \) (the utility values are not crucial from a theoretical point of view). If the decision maker is able to provide a number \( \alpha \) such that the sure consequence \( r \) is equivalent to the random consequence \( \alpha r_1 + (1-\alpha) r_2 \), then by definition of \( u \), it happens that
\[ u(r) = xu(r^*) + (1-x)u(r_*), \]

A complete description of the construction of \( u \) can be found in Chapter 2 of [1] and in [18].

In practice, such a number \( x \) cannot be known exactly and it is more realistic to assume that it can only be bounded in an interval. Thus, instead of a single utility function \( u \), the utility assessment method only provides us with a class \( \mathcal{V} \) of utility functions of the form

\[ \mathcal{V} = \{ u : \mathcal{S} \to \mathbb{R} : a_k \leq u(r_k) \leq b_k, k = 1, \ldots, K \}, \]

where \( a_k \) and \( b_k \) are known numbers. In addition to the constraints on the value of \( u \) on a finite set of consequences, some information on the regularity or/and the shape of \( u \) is usually known. For instance, when \( r \) is a monetary reward, it is well known that a risk averse decision maker (i.e. a decision maker who prefers a sure reward rather than a random one with the same expectation) has a concave utility function. Denote by \( \mathcal{S} \) the class of continuous functions \( u \) (from \( \mathcal{S} \) to \( \mathbb{R} \)) with some given constraints on shape (increasing, unimodal, concave, etc.). The complete information on the function \( u \) is described by the class \( \mathcal{U} = \mathcal{V} \cap \mathcal{S} \). To shorten notation, we use the same letter \( \mathcal{U} \) for the class of functions \( U \) (Section 1) and for the class of functions \( u \) (Section 2).

Example 1 (Cont.). Define the consequence associated with \((\theta, d)\) by \( R = d - H \) where \( H \) is the peak water level of a potential flood. Note that \( R \) is a random variable as \( H \) is random with density depending on \( \theta \). For simplicity, assume that the peak level \( H \) is bounded by 1 and take \( \mathcal{S} = [0, 1] \) and \( \mathcal{U} = [-1, 1] \). For constructing \( u \), we need to provide an interval \([a, b]\) of possible values of \( u(r_k) \) on a finite subset \( \{r_1, \ldots, r_K\} \subset \mathcal{S} \). It is proper to set \( u(0) = 1 \) (best consequence) and \( u(-1) = 0 \) (worst consequence). Note that negative consequences correspond to floods. Taking into account that it is believed that a low flood \((r = -0.25)\) is not very crucial compared with a moderate flood \((r = -0.5)\), we obtain the values given in Table 1. In addition, it is reasonable to impose that \( u \) is continuous, increasing on \([-1, 0]\), decreasing on \([0, 1]\) with a unique mode at 0. This completes the construction of \( \mathcal{U} \).

From now on we make the assumption that the consequence \( R \) is of the form \( R = d - Y \) where \( Y \) is a random variable whose distribution \( \xi_\theta \) depends on \( \theta \) but not on \( d \). Denote by \( \mathcal{U} \) the set of values of \( Y \). Note that this assumption is appropriate in many real life situations like those given at the beginning of the paper. It is also interesting to remark that no assumption is required on \( \theta \). In particular, the prior and the posterior distributions can be very complicated and multidimensional like in practical hierarchical models for example. Thus, (2.1) reduces to

\[ U(\theta, d) = \int_\theta u(d - y) \xi_\theta(dy), \]

and, by (1.1), optimal decisions are maximizers of

\[ \int_\theta \int_y u(d - y) \xi_\theta(dy) \pi_u(\theta) = \int_y u(d - y) \mu(dy), \]

where \( \mu \) is the marginal distribution of \( Y \) defined by

\[ \mu(A) = \int_\theta \int_A \xi_\theta(dy) \pi_u(\theta). \]

We assume throughout the paper that such a distribution \( \mu \) does exist.

3. The discretization method

Let us recall the technical assumptions of the previous sections.

(H1) \( \mathcal{U} = \mathcal{V} \cap \mathcal{S} \) is made up of continuous functions from \( \mathcal{S} \) into \( \mathbb{R} \) with constraints on values \( \mathcal{V} \) and constraints on shape \( \mathcal{S} \).

(H2) The random consequence of choosing \( d \in \mathcal{S} \) is of the form \( R = d - Y \) where \( Y \) is a random variable with a distribution \( \xi_\theta \) independent of \( d \).

If, for all \( u \in \mathcal{U} \), we let

\[ u^*(d) = \int_y u(d - y) \mu(dy), \]

Table 1
Lower and upper bounds \((a_k, b_k)\) for the utility \( u(r_k) \) of Example 1

<table>
<thead>
<tr>
<th>( r_k )</th>
<th>(-1)</th>
<th>(-0.5)</th>
<th>(-0.25)</th>
<th>(0)</th>
<th>(0.5)</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_k )</td>
<td>(0)</td>
<td>(0.1)</td>
<td>(0.6)</td>
<td>(1)</td>
<td>(0.6)</td>
<td>(0.4)</td>
</tr>
<tr>
<td>( b_k )</td>
<td>(0)</td>
<td>(0.3)</td>
<td>(0.9)</td>
<td>(1)</td>
<td>(0.8)</td>
<td>(0.6)</td>
</tr>
</tbody>
</table>
then, an optimal decision is a maximizer of \( u^\omega \), that is any element of \( M(u) \) with
\[
M(u) = \{ d \in \mathcal{D}, u^\omega(d) = \sup_{t \in \mathcal{D}} u^\omega(t) \}.
\]

It is worth pointing out that, in addition to the decision problems described in Section 2, (3.1) corresponds to a large class of problems in Bayesian analysis including, for example, estimation problems \((y, \mu)\) the posterior distribution) or prediction problems \((\mu, \mu)\) is the predictive distribution).

The numerical problem is the following: we have to compute
\[
d_u^* = \inf_{u \in \mathcal{U}} d_u^-,\]
where \( d_u^- = \inf M(u) \). The problem for \( d_u^* \) is similar. Our strategy consists in running a stochastic exploration of \( \mathcal{U} \) with attractions around functions \( u \) associated with low value of \( d_u^- \). At first sight, this is nearly impossible for several reasons:

- \( \mathcal{U} \) is an infinite dimensional space,
- the computation of \( d_u^- \) is needed for many candidates \( u \), consequently numerical results have to be rapidly achieved,
- the constraints of \( \mathcal{U} \) must be fulfilled.

We solve these problems as follows. First \( \mathcal{U} \) is approximated by a finite dimensional space \( \mathcal{U}_j \) by means of a first discretization of \( u \). Then, for each \( u \in \mathcal{U}_j \), a second discretization enables fast computations of \( d_u^- \). Finally, the exploration of \( \mathcal{U}_j \) is running by means of a simulated annealing algorithm: the proposal distribution ensures that the constraints on \( u \) are fulfilled and the acceptance probability guarantees the attraction of the chain at low values of \( d_u^- \).

From now on, we assume that \( \mathcal{D}, \mathcal{U} \) and \( \mathcal{U} \) are compact intervals of \( \mathbb{R} \) (the non-compact case is discussed below). For simplicity, we make the following assumption:

\[(H_3) \; \mathcal{D} = \mathcal{U} = [0, 1] \quad \text{and} \quad \mathcal{U} = [-1, 1].\]

Thus, \( \mathcal{U} \) is made up of continuous functions from \([-1, 1]\) to \( \mathbb{R} \). Let \( j \) be a positive integer. For \( l \in \mathbb{Z} \), define the function \( F_l : \mathbb{R} \to \mathbb{R} \) such that \( F_l(t) = 0 \) if \( t \leq (l-1)/j \), \( F_l(t) = (t-(l-1)/j) \) for \( t \in [(l-1)/j, l/j] \) and \( F_l(t) = 1/j \) for \( t \geq l/j \). Thus, \( F_l \) is continuous, constant outside \( I_l = [l/j, (l+1)/j] \) and linear with slope 1 on \( I_l \). By construction, \( \mathcal{U}_j \) is made up of functions \( u_j \in \mathcal{U} \) of the form
\[
u_j = \sum_{l=-j}^{j} \gamma_l F_l,
\]
where \( \gamma = (\gamma, \ldots, \gamma) \in \mathbb{R}^{2j+1} \). Note that \( u_j \) is continuous, linear on \( I_l \) with slope \( \gamma_l \) on \( I_l \) for \(-j < l < j\) and that \( u_j(0) = 0 \).

For all \( u \in \mathcal{U}_j \), take \( u_j \in \mathcal{U}_j \), such that \( u_j(l/j) = u(l/j) \) for all \( l \in \{j, \ldots, j\} \). The function \( u_j \) can be viewed as an approximation of \( u \). In the sequel, an element \( \mathcal{U}_j \) will be denoted by \( u_j \), \( u_j \), or simply \( u \) according to the context.

**Proposition 2.** Under \( H_1 - H_3 \), for all \( u \in \mathcal{U} \) such that \( M(u) = \{ d_u \} \), there exists a sequence \( (u_j) \) with \( u_j \in \mathcal{U}_j \) such that \( u_j \to d_u \) as \( j \to \infty \) where \( t_j \in M(u_j) \). Furthermore, for \( j \) large enough, \( t_j \) is the unique maximizer of \( u_j \) (in other words, \( M(u_j) = \{ t_j \} \)).

**Proposition 2** shows that a maximizer \( d_u \) of \( u^\omega \) for \( u \in \mathcal{U} \) can be approximated by a maximizer of some \( u_j \in \mathcal{U}_j \) when \( d_u \) is the unique maximizer of \( u^\omega \). Without the uniqueness condition, it is only possible to prove that \( t_j \) is in any neighborhood of \( M(u) \) when \( j \) is large enough. We believe that this condition is not crucial in practice as the class \( \mathcal{U} \) is usually large enough so that any maximizer \( d_u \in M(u) \) can be approximated by a (unique) maximizer \( d_u \) such that \( \{ d_u \} = M(u') \) for some \( u' \in \mathcal{U} \). Unfortunately, this seems to be difficult to prove in a general setting.

According to **Proposition 2**, \( \inf_{u \in \mathcal{U}_j} d_u^- \) approximates to \( d_u^- \). Let us first find the zeros of \( Du^\omega \) for \( u \), of the form (3.2), where \( D \) denotes the derivative operator. It is required that
\[(H_4) \; \mu \text{ has a density } \phi_u \text{ with respect to the Lebesgue measure on } [0, 1].\]

Then, by **Proposition 12** and \( H_4 \), we have that
\[
\begin{align*}
u^\omega(d) &= \sum_{l=-j}^{j} \gamma_l \int_{0}^{1} F_l(d - y) \mu(dy), \\
Du^\omega(d) &= \sum_{l=-j}^{j} \gamma_l \int_{0}^{1} \mathbb{1}_I(d - y) \mu(dy)
\end{align*}
\]
for all \( d \in \mathcal{D} \), where \( \mathbb{1}_I \) denotes the indicator function of \( I_l \). For all \( t \in \mathbb{R} \), define
\[
\Phi^\omega(t) = \int_{0}^{t} \mu(dy).
\]
Then, it is easily seen that

$$Du^u_l(d) = \sum_{j=1}^{l} \gamma_j \left[ \phi_u \left( d - \frac{l-1}{j} \right) - \phi_u \left( d - \frac{j}{l} \right) \right].$$

(3.4)

Now let us proceed to a second discretization to approximate $\phi_u$ by setting

$$\phi_j = \sum_{l=1}^{j} \mu_l F_i$$

(3.5)

with $\mu_l$ such that $\phi_u(l/j) = \phi_j(l/j)$ for all $l \in \{0/j, \ldots, j/j\}$. As $\mu_l$ is the slope of $\phi_j$ on $I_l$, it is easily seen that $\mu_l = j [\phi_u(l/j) - \phi_u((l-1)/j)]$.

**Proposition 3** below enables to control the accuracy of the approximation of $\phi_u$ by $\phi_j$. For $u_i \in \mathbb{U}_i$, let $\hat{Du}_j$ be the approximation of $Du^u_l$ defined by replacing $\phi_u$ by $\phi_j$ in (3.4). For all $\eta > 0$ and a function $f$, set

$$\omega(f, \eta) = \sup_{|t-t'|<\eta} |f(t) - f(t')|.$$  

**Proposition 3.** Under $H_1-H_4$, for all $u_i \in \mathbb{U}_i$, we have

$$\sup_{d \in [-1,1]} |Du^u_l(d) - \hat{Du}_j(d)| \leq (2 + 1/j) \omega(\phi_u, 2/j) \max_{l \in \{1, \ldots, j\}} |\gamma_l|.$$  

Note that this result suggests to consider an additional assumption of the form $H'$. There exists $B > 0$ such that $|u(t) - u(t')| \leq B|t - t'|$ for all $u \in \mathbb{U}$ and all $t, t' \in [-1, +1]$.

As mentioned at the end of the paper in Example 1, such an assumption is realistic and is needed in practice to discard some utilities with strong variations. Under $H'$ and the assumptions of **Proposition 3**, $\max_{l \in \{1, \ldots, j\}} |\gamma_l| \leq B$, and if $\phi_u$ is continuous on $[0, 1]$, $\omega(\phi_u, 2/j) \to 0$ as $j \to \infty$ and we have

$$\sup_{u \in \mathbb{U}_i} \sup_{d \in [-1,1]} |Du^u_l(d) - \hat{Du}_j(d)| \to 0, \quad \text{as } j \to \infty.$$ 

**Proposition 4.** Under $H_1-H_4$, for all $u_i \in \mathbb{U}_i$, we have

$$\hat{Du}_j = \sum_{l=1}^{j} a_l(\gamma) F_i,$$

where $a_l(\gamma) = \sum_{i=j}^{j} A_{i}i(\gamma)$ and $A_{i+1} = \mu_{i+1}1_{[j+1, j]}(l) - \mu_{i}1_{[j+1, j]}(l)$.  

**Propositions 3 and 4** enable to approximate numerically the solution of $Du^u_l(d) = 0$ for any $\gamma \in \mathbb{R}^{2^i+1}$. Actually, $Du^u_l$ is approximated by $\hat{Du}_j$, a piecewise linear function on $[-1,1]$ with slope $a_l(\gamma)$ on $I_j (-j < l \leq j)$ and initial value $u_j(-1) = \gamma_{-j}$. If we note that $a_l(\gamma)$ is simply a linear function of $\gamma$, we realize that a fast computation of the zeros of $\hat{Du}_j$ is now possible.

**Proof.** Note that, from (3.5) and Lemma 5, we have that

$$\phi_j \left( d - \frac{l-1}{j} \right) - \phi_j \left( d - \frac{j}{l} \right) = \sum_{i=1}^{l} \mu_l F_i \left( d - \frac{l-1}{j} \right) - \sum_{i=1}^{j} \mu_l F_i \left( d - \frac{j}{l} \right) = \sum_{i=1}^{l} \mu_l F_{i,j-1} (d) - \sum_{i=1}^{j} \mu_l F_{i,j} (d)$$

$$= \sum_{j=1}^{l} \mu_l F_{j+1} (d) - \sum_{i=1}^{j} \mu_l F_{i,j} (d) = \sum_{j=1}^{l} A_{j}F_i (d).$$

Thus, from (3.4) and the definition of $\hat{Du}_j$, we conclude that

$$\hat{Du}_j (d) = \sum_{j=1}^{l} \gamma_j \left[ \phi_j \left( d - \frac{l-1}{j} \right) - \phi_j \left( d - \frac{j}{l} \right) \right] = \sum_{j=1}^{l} \sum_{i=1}^{j} \gamma_i A_{j}F_i (d).$$

The proof of **Lemma 5** is left to the reader.

**Lemma 5.** For all $l$ and $i$ in $\mathbb{Z}$ and all $d \in \mathbb{R}$, we have

$$F_i (d - i/j) = F_{i,i} (d).$$
Let \( Z(\gamma) = \{ d \in \mathcal{D}, \nabla \hat{u}^\mu(d) = 0 \} \). As \( \nabla \hat{u}^\mu \) is piecewise linear, the computation of \( Z(\gamma) \) is easy. It remains to select from \( Z(\gamma) \) the lowest global maximum of \( u_t^\mu \). For doing this, we need to compare \( u_t^\mu(0), u_t^\mu(1) \) and \( u_t^\mu(d) \) for all \( d \in Z(\gamma) \). Proposition 7 enables to compute rapidly these numbers. For all \( t \in \mathbb{R} \), define
\[
\Psi^\mu(t) = \int_0^t y\mu(dy)
\]
and approximate again \( \Psi^\mu \) by \( \Psi^\mu_j \) by setting:
\[
\Psi^\mu_j = \sum_{i=1}^j \Psi_j F_i
\]
in such a way that \( \Psi^\mu(l/j) = \Psi^\mu_j(l/j) \) for all \( l \in \{0, \ldots, j\} \) (that is \( \Psi_j = j[\Psi^\mu(l/j) - \Psi^\mu((l-1)/j)] \)). Recall that
\[
u_t^\mu(d) = \sum_{i=1}^l \gamma_l I_l(d),
\]
where
\[
I_l(d) = \int_0^1 F_i(d - y)\mu(dy).
\]
From the definition of \( F_i \), it is easily seen that
\[
I_l(d) = \left( d - \frac{l - 1}{j} \right) \left[ \phi^\mu \left( d - \frac{l - 1}{j} \right) - \phi^\mu \left( d - \frac{l - 1}{j} \right) \right] - \left[ \psi^\mu \left( d - \frac{l - 1}{j} \right) - \psi^\mu \left( d - \frac{l - 1}{j} \right) \right] + \frac{1}{j} \phi^\mu \left( d - \frac{l - 1}{j} \right).
\]
Define \( \hat{\nu}_t \) by replacing \( \nu_t^\mu \) and \( \nu_t^\mu_j \) by \( \phi^\mu \) and \( \psi^\mu \), respectively, in the definition of \( I_l \) and let
\[
\hat{\nu}_t^\mu(d) = \sum_{i=1}^l \gamma_l \hat{I}_l.
\]

**Proposition 6.** Under \( H_1 - H_4 \), for all \( u_t \in \mathcal{U} \) (defined by (3.2)) and all \( d \in \mathcal{D} \), we have
\[
\sup_{d \in [0,1]} |\nu_t^\mu(d) - \hat{\nu}_t^\mu(d)| \leq (2 + 1/j) B_\max |\gamma|,
\]
where \( B_\max = (2\omega(\phi^\mu, 2/j) + 2\omega(\psi^\mu, 2/j) + \max_{i \in \{0, \ldots, j\}} \mu(I_l)) \) and \( \psi^\mu(y) = y\phi^\mu(y) \) for all \( y \in [0, 1] \).

Thus, if in addition it is assumed that \( H^2 \) is fulfilled and that \( \phi^\mu \) is continuous on \( [0, 1] \), we deduce from Proposition 6 that
\[
\sup_{u_t \in \mathcal{U}_t, d \in [0,1]} |\nu_t^\mu(d) - \hat{\nu}_t^\mu(d)| \to 0
\]
as \( j \to \infty \).

Proposition 6 shows that \( \hat{u}_t \) approximates to \( u_t \). In Proposition 7, we derive an interesting formula for the computation of \( \hat{u}_t(d) \).

**Proposition 7.** Under \( H_1 - H_4 \), for all \( u_t \in \mathcal{U}_t \) and all \( d \in \mathcal{D} \), we have
\[
\hat{\nu}_t^\mu(d) = \sum_{i=1}^l \gamma_l \left( d - \frac{1}{j} \right) A_{t,l} - A_{t,l}^* + A_{t,l}^\mu
\]
where
\[
A_{t,l}^\mu = \sum_{i=1}^l \gamma_l \left( d - \frac{1}{j} \right) A_{t,l} - A_{t,l}^* + A_{t,l}^\mu
\]
and where \( A_{t,l} \) is defined by replacing \( \mu \) by \( \Psi \) in the definition of \( A_{t,l} \) and where \( A_{t,l}^* = \mu_{t,l}/j_1 1_{1+,1+,j}(l) \).

We finish this section with a short discussion about the assumptions \( H_1 - H_4 \). Assumption \( H_1 \) is mainly needed for theoretical results on approximation: the uniform approximation of \( u_t \in \mathcal{U} \) by some \( u_t \in \mathcal{U}_t \) and the approximation of a maximizer of \( u_t^\mu \) by a maximizer of \( u_t^\mu \). Assumptions \( H_2 - H_4 \) are used for computational issues. From \( H_2 = H_4 \), we can obtain two major formulas: Lemma 5 and (3.3). From these formulas, we deduce the convenient expression of \( \nabla \hat{u}^\mu \) of Proposition 4 and then, we can compute the approximations of the Bayes actions. The compactness assumption \( H_3 \) is useful for theoretical results on approximation of \( u_t \) and it is also very important from a practical point of view. Let us show that it is necessary for the practical computation of the zeros of \( \nabla \hat{u}_t \) with Proposition 4. Assume that \( \mathcal{D} = \mathcal{U} = \mathcal{D} = \mathbb{R} \). Then, any function \( u_t \in \mathcal{U} \) can be approximated by a function \( u_t \) of the form \( u_t = \sum_{i=1}^\infty \nu_t F_i \), and, similarly, \( \Phi^\mu \) can be approximated by a function \( \Phi^\mu_t \) of the form \( \Phi^\mu_t = \sum_{i=1}^\infty \lambda_t F_i \). Then, attention shows that the result of Proposition 4 becomes
\[
\widehat{Du}_c^c = \sum_{i=0}^{\infty} a_i^c(\gamma) F_i,
\]

with
\[
a_i^c(\gamma) = \sum_{n=0}^{\infty} \gamma_n (\mu_{i,n+1} - \mu_{i,n}),
\]

and it is clearly not easy to compute rapidly the zeros of \(\widehat{Du}_c^c\).

### 4. Monte Carlo optimization

We are now in position to compute \(d^c_u\) and \(d^c_w\) using a Monte Carlo optimization method. Denote by \(d^c_u\) the lowest Bayes action derived from the utility function \(u \in \Psi_u\). Several stochastic methods which propose an exploration of \(\Psi_u\) with attraction at low values of \(d^c_u\) can be contemplated. For a recent account on the theory, we refer the reader to [19]. The simulated annealing algorithm is particularly well adapted to the optimization issue of this paper. It can be described as follows.

At iteration \(n\), the algorithm is at \(\gamma(n)\).

1. Simulate \(\gamma\) from the proposal distribution \(P_{\gamma(n)}\).
2. Accept \(\gamma(n+1) = \gamma\) with probability \(\rho_n = \exp\left(-\beta_n(d^c_{\gamma(n)} - d^c_{\gamma(n)})\right) \wedge 1\); take \(\gamma(n+1) = \gamma(n)\) otherwise.

It is worth pointing out that the numbers \(A_{ij}, A'_{ij}\) and \(A''_{ij}\) involved in the computation of \(d^c_u\) are computed only once. Thus, the computation of \(d^c_u\) at each iteration of the chain only requires few operations.

The inverse of \(\beta_n\) is the so-called temperature parameter. In Section 5.1, the proof of the convergence of \(d^c_{\gamma(n)}\) is provided when \(\beta_n = C^{-1} \log(n + e)\) where \(C\) is a constant. The proposal distribution \(P_{\gamma(n)}\) (described in details in Section 5.1) depends on the constraints of \(\Psi\) and ensures that the proposition \(u\) remains inside \(\Psi\). Roughly, \(u\) is equal to \(u_{\gamma(n)}\), except on a set of the form \((l - 1)/j, (l + 1)/j\) where \(l\) is chosen at random. The particular proposal distribution of Example 1 is described below.

**Example 1 (Cont.).** Recall that \(u\) is continuous, linear on \(I_j\), \(l \in \{-j + 1, \ldots, j\}\) with \(u_l(-1) = 0\) and \(u_l(0) = 1\). Thus, \(u\) is uniquely determined by its value at \(I_j, l \in L = \{-j, \ldots, j\} \setminus \{j, 0\}\). We first sample a position \(p\) from a uniform distribution on \(L\) and then, given \(p\), we simulate a new value for \(u_l\) at \(p/j\) taking into account the constraint of \(\Psi\). Precisely, given \(p\) and \(u_{\gamma(n)}\), it is easy to calculate explicitly an interval of admissible values for \(u_{\gamma(n)}(p/j)\) and to sample \(u_{\gamma(n)}(p/j)\) uniformly on this interval, hence a proposal for \(\gamma\). Thus, at iteration \(n\), \(u_{\gamma(n)}\) is simply a local perturbation of \(u_{\gamma(n)}\). In Section 5, we show the convergence of the algorithm for \(\beta_n = C^{-1} \log(n + e)\) with \(C > 2(2j - 1)\) (in practice, the same numerical results were obtained with \(\beta_n = C^{-1} \log(n + e)\) and with \(\beta_n = n\)). In Fig. 1, we plot the initial utility function \((n = 1)\), the final utility function \((n = 3000)\) and two intermediate utility functions \((n = 3000)\). Each constraint is indicated by two small circles possibly connected by a vertical line. In Fig. 2, we plot the associated derivatives \(Du_{\gamma(n)}\). Fig. 3 provides the utility functions for the optima \(d^c_u\) (achieved at \(n = (359, 186)\) and \(d^c_w\) (achieved at \(n = (332, 451)\)). We can note that the optima were achieved for functions \(u\) with strong slope variations. It is reasonable to think that such functions do not fit the actual preferences of the decisions maker and should be discarded. This can be done easily by constraining the slopes \(\gamma_i\) (or the variations \(\gamma_i - \gamma_{i-1}\)) to stay in a given interval. Such an additional condition would simply affect the calculations of the interval of admissible values for \(u_{\gamma(n)}(p/j)\). The values of \(d^c_{u,n}\) and \(d^c_{w,n}\) are plotted in Figs. 4 and 5. Finally, the approximation of \([d^c_u, d^c_w]\) is \([0.2408, 0.6624]\). We choose the precision \(\beta = 20\) and a Beta(5.7) distribution for \(\mu\). The distribution \(\mu\) is only involved in the computation of \(A_{ij}, A'_{ij}\) and \(A''_{ij}\). These numbers do not change with the iteration number \(n\) and can be computed from MCMC methods if needed. The numerical computations are done with a standard personal computer using the R software [20]. It takes less than 2 min to run the entire chain.

![Fig. 1. Utility functions \(u_{\gamma(n)}\) for \(n = 1\) (full line) and \(n = 3000.10^5, 4 \times 10^5\) (dashed lines) in the research of \(d^c_u\).](image-url)
5. Auxiliary results and proofs

5.1. Convergence of the simulated annealing algorithm

Let us first review some results about the convergence of the simulated annealing algorithm. Let \((E, \mathcal{E})\) be a measurable space and \(V : E \rightarrow [0, \infty)\) a function such that \(\text{osc}(V) < \infty\) with

\[
\text{osc}(V) = \sup \{|V(x) - V(y)| : x, y \in E\}.
\]

Let \(K\) be a transition kernel on \(E \times \mathcal{E}\) and denote by \((X_n)\) the Markov chain of the simulated annealing algorithm with transition kernel \(K\). Thus, we have the following algorithm:

![Fig. 2](image1.png) Derivatives \(D u_n^c\) for \(n = 1\) (full line) and \(n = 3000, 10^5, 4 \times 10^5\) (dashed lines) in the research of \(d_n^c\).

![Fig. 3](image2.png) Utility functions \(u_n^c\) for \(d_n^c\) (full line) and \(d_n^u\) (dashed line).

![Fig. 4](image3.png) Optimum \(d_n^c\) and \(d_n^u\) vs. \(n\).
Proposition 8. Assume that:

1. For all $x$, define $X_n$ by $X_n = \arg\min_{x \in \mathcal{X}} f(x)$.
2. For all $x$ and $u$, define $Y_n = \arg\min_{y \in \mathcal{Y}} \{ \rho(u) \}$.

Example 1 (Cont.). Take $u \in \mathcal{U}$, then $u(-1) = 0$ and $u(0) = 1$. Since $u \in \mathcal{U}$ is uniquely determined by its values $u(l/j)$ for $l \in L$ for some $L \subseteq \{-j, \ldots, +j\}$. Write $u_l$ the vector with components $u_l$, $l \in L$, and denote by $E$ the subset of $R^p$ ($p \leq j + 1$) such that $u \in \mathcal{U}$ if and only if $u_l \in E$. Assume that $E$ is a bounded subset of $R^p$.

For a proof of Proposition 8, we refer the reader to [21].

We now proceed with the study of the convergence of the algorithm defined in Section 4. Note that any function $u$ in $\mathcal{U}$ is completely defined by its values $u(l/j)$ for all $l \in L$ for some $L \subseteq \{-j, \ldots, +j\}$. Write $u_l$ the vector with components $u(l/j)$, $l \in L$, and denote by $E$ the subset of $R^p$ ($p \leq j + 1$) such that $u \in \mathcal{U}$ if and only if $u_l \in E$. Assume that $E$ is a bounded subset of $R^p$ (this assumption is required to use the uniform distribution on $E$, if $E$ is not bounded, other distributions can be used).

Take $\beta_n = C^{-1} \log(n + e)$ with $C > p \log(V)$. Then, for all $\varepsilon > 0$, $\Pr(X_n \in V') \to 1$ as $n \to \infty$ where

$$
V' = \{ x \in E : V(x) \leq \exp\{\varepsilon\} \}
$$

$$
\varepsilon = \sup\{ \alpha \geq 0 : f(x) \leq 1 \}.
$$

For a proof of Proposition 8, we refer the reader to [21].

We now proceed with the study of the convergence of the algorithm defined in Section 4. Note that any function $u$ in $\mathcal{U}$ is completely defined by its values $u(l/j)$ for all $l \in L$ for some $L \subseteq \{-j, \ldots, +j\}$. Write $u_l$ the vector with components $u(l/j)$, $l \in L$, and denote by $E$ the subset of $R^p$ ($p \leq j + 1$) such that $u \in \mathcal{U}$ if and only if $u_l \in E$. Assume that $E$ is a bounded subset of $R^p$ (this assumption is required to use the uniform distribution on $E$, if $E$ is not bounded, other distributions can be used).

Example 1 (Cont.). Take $u \in \mathcal{U}$, then $u(-1) = 0$ and $u(0) = 1$. Since $u \in \mathcal{U}$ is uniquely determined by its values $u(l/j)$ for $l \in L = \{-j, \ldots, +j\}$ for $p = 2j - 1$. Since $u$ is increasing on $[-1, 0]$ and decreasing on $[0, 1]$, $E = \{(x_{-1}, x_{-1}, x_1, x_2, \ldots, x_p) \in [0, 1]^{2j} : 0 \leq x_{-1} \leq \cdots \leq x_{-1} \leq 1 \geq x_1 \geq \cdots \geq x_p\}$.

For all $x \in E$ and all $l \in \{1, \ldots, p\}$, define $E_l = \{x_l \in \mathcal{X} : (x_1, x_2, \ldots, x_l, x_{l+1}) \in E\}$.

For all $l \in \{1, \ldots, p\}$, denote by $dx_l$ the Lebesgue measure on $\mathbb{R}$ and set

$$
\mathcal{K} = \int_E dx_1 \ldots dx_p, \quad \kappa_{x,l} = \int_{E_l} dx_l.
$$
The Markov chain associated with the algorithm of Section 4 is based on the transition kernel $K$ defined below. For all $x, y \in E$, let

$$K(x, dy) = p^{-1} \sum_{l=1}^{p} \tilde{K}_i(x, dy)$$

with

$$\tilde{K}_i(x, dy) = \left[ \prod_{k=r}^{l} \delta_{x_k}(dy_k) \right] K(x_{-l}, dy_l),$$

$$K(x_{-l}, dy) = \kappa^{-1}_l 1_{E_l}(y_l) dy_l,$$

where $y = (y_1, \ldots, y_p), x = (x_1, \ldots, x_p)$ and $x_{-l} = (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_p)$. Note that, for $x$ fixed, $\kappa_l(x_{-l}, dy_l)$ is the uniform distribution on $E_{x_{-l}}$. Let $\lambda(dx)$ be the uniform distribution on $E$, that is

$$\lambda(dx) = \kappa^{-1} 1_{E}(x) \prod_{l=1}^{p} dx_l.$$ 

(5.1)

**Proposition 9.** The measure $K(x, dy)\lambda(dx)$ is symmetric, that is $K(x, dy)\lambda(dx) = K(y, dx)\lambda(dy)$.

**Proof.** It is sufficient to prove that $\tilde{K}_i(x, dy)\lambda(dx) = \tilde{K}_i(y, dx)\lambda(dy)$ for all $l \in \{1, \ldots, p\}$. Note that

$$\tilde{K}_i(x, dy)\lambda(dx) = F(x, y)m(dx, dy)$$

with

$$F(x, y) = \kappa^{-1}_x 1_{E_i}(y_i) 1_{E}(x),$$

$$m(dx, dy) = \left[ \prod_{k=r}^{l} \delta_{x_k}(dy_k) \right] dy_l dx_l.$$

Note that, $m$-almost surely, $x_k = y_k$ for all $k \neq l$. Thus, $m$-almost surely, we have that $E_{x_{-l}} = E_{y_{-l}}, \kappa_{x_{-l}} = \kappa_{y_{-l}}$ and

$$1_{E_{x_{-l}}}(y_{x_{-l}}) 1_{E}(x) = 1_{E_{y_{-l}}}(y_{y_{-l}}) 1_{E}(y).$$

Thus, $F(x, y) = F(y, x)$ $m$-almost surely. We conclude by noting that $m(dx, dy) = m(dy, dx)$. \qed

Let $S_p$ be the set of the permutations of $\{1, \ldots, p\}$ and let $\sigma \in S_p$. Denote by $\tilde{K}_\sigma(x, dy)$ the composition of the transition kernels $\tilde{K}_i(x, dy)$ with $\sigma$. It is easy to see that,

$$\tilde{K}_\sigma(x, dy) = K_{\sigma(1)}(x_{\sigma(1)}, \ldots, x_{\sigma(p)}), dy_{\sigma(1)}), K_{\sigma(2)}(y_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(p)}), dy_{\sigma(2)}), \ldots, K_{\sigma(p)}(y_{\sigma(1)}, \ldots, y_{\sigma(p-1)}), dy_{\sigma(p)} \propto 1_{D_{x,\sigma}}(y) \prod_{l=1}^{p} dy_l,$$

where $D_{x,\sigma}$ is the support of $\tilde{K}_\sigma(x, dy)$. Thus, for $x$ fixed, $\tilde{K}_\sigma(x, dy)$ is the uniform distribution on $D_{x,\sigma}$. Denote by $\kappa^{-1}_{x,\sigma}$ the constant such that $K_{\sigma}(x, dy) = \kappa^{-1}_{x,\sigma} 1_{D_{x,\sigma}}(y) \prod_{l=1}^{p} dy_l$.

**Proposition 10.** If $\bigcup_{\sigma \in S_p} D_{x,\sigma} \supset E$ for all $x \in E$, then we have that

$$K^p(x, dy) \geq p^{-p} \lambda(dy) \quad \text{for all } x \in E.$$

**Proof.** With the notation $x = x^\emptyset$ and $y = x^\emptyset$, we have

$$K^p(x, dy) = K^p(x^\emptyset, dx^\emptyset) = \int \cdots \int K(x^\emptyset, dx^1) \cdots K(x^{p-1}, dx^p)$$

$$= p^{-p} \sum_{i=-1}^{p} \sum_{i=-1}^{p} \int \cdots \int \tilde{K}_i(x, dx^1) \cdots \tilde{K}_p(x^{p-1}, dx^p) \geq p^{-p} \sum_{\sigma \in S_p} \tilde{K}_\sigma(x^\emptyset, dx^\emptyset).$$

We conclude by noting that

$$\sum_{\sigma \in S_p} \tilde{K}_\sigma(x, dy) = \sum_{\sigma \in S_p} \kappa^{-1}_{x,\sigma} 1_{D_{x,\sigma}}(y) dy$$

$$\geq \kappa^{-1} \left[ \sum_{\sigma \in S_p} 1_{D_{x,\sigma}}(y) \right] dy \quad \text{(since } D_{x,\sigma} \subseteq E, \kappa_{x,\sigma} \leq \kappa)$$

$$\geq \kappa^{-1} 1_E(dy) \quad \text{(since } \bigcup_{\sigma \in S_p} D_{x,\sigma} \supset E)$$

$$= \lambda(dy). \quad \square$$
Example 1 (Cont.). For all \( x \in E \), denote by \( u_x \) the function \( u \in \Psi \) such that \( u(l/j) = x_l \) for all \( l \in L \). Denote by \( d^*_c \) the smallest maximizer of \( u^*_c = \int y u_x(y) \mu(dy) \). By taking \( V(x) = d^*_c + 1 \), the simulated algorithm of Section 5.1 coincides with the algorithm of Section 4. It is left to the reader to check that, for all \( x, y \in E \), there exists a sequence \( x = x^{(1)}, x^{(2)}, \ldots, x^{(p-1)}, x^{(p)} = y \) with \( p' \leq p \) and \( x^{(i)} \in E \) such that \( x^{(i)} = x^{(i+1)} \) for all \( l \) but one (in other words, there exists \( s \in S_p \) such that \( y \in D_{s \gamma} \)), hence the condition of Proposition 10. Note that \( V(x) \in [0,2] \) for all \( x \in E \). By Propositions 9 and 10, conditions 1 and 2 of Proposition 8 are fulfilled with \( x \) defined by \((5.1)\), \( p = 2j - 1 \), \( \epsilon = (2j - 1)^{-1/2} - 1 \) and \( \gamma = \lambda \). Thus, the convergence is proved for \( C > 2p \).

5.2. Auxiliary results

Lemma 11. For all continuous function \( u: [-1,1] \rightarrow \mathbb{R} \), let \( u_j \) defined as in Section 3 by \( u_j \in \Psi \) and \( u_j(l/j) = u(l/j) \) for all \( l \in \{-j, \ldots, j\} \). We have

\[
\sup_{t \in [-1,1]} |u(t) - u_j(t)| \leq o(u, 1/j),
\]

\[
\sup_{t \in [0,1]} |u^*(t) - u^*_j(t)| \leq o(u, 1/j),
\]

where \( o(u, 1/j) \rightarrow 0 \) as \( j \rightarrow \infty \).

Proof. Let \( l \in \{-j, \ldots, j\} \). Recall that \( l_i = \lfloor (l - 1)/j \rfloor /j \) for \( l \in \mathbb{Z} \). By the continuity of \( u \) and \( u_j \), there exists \( t_0 \in l_i \) such that

\[
\sup_{t \in l_i} |u(t) - u_j(t)| = |u(t_0) - u_j(t_0)| \leq \max\{|u(t_0) - u_{j-1}(l/j)|, |u(t_0) - u_j((l-1)/j)|\}
\]

\[
= \max\{|u(t_0) - u(l/j)|, |u(t_0) - u((l-1)/j)|\} \leq o(u, 1/j).
\]

Since \( u(-1) = u_{(-1)} = 0 \), we have

\[
\sup_{t \in [-1,1]} |u(t) - u_j(t)| = \max_{l_i \in [-j-1, j]} \sup_{t \in l_i} |u(t) - u_j(t)| \leq o(u, 1/j),
\]

hence the first inequality. The second inequality follows from, for all \( t \in [0,1] \),

\[
|u^*(t) - u^*_j(t)| \leq \int_0^t |u(t - y) - u_j(t - y)| \mu(dy) \leq o(u, 1/j).
\]

We conclude the proof by noting that \( u \) is uniformly continuous. \( \square \)

Proposition 12. Let \( d \in \mathbb{R} \) and \( l \in \mathbb{Z} \). If \( \mu(d - (l-1)/j) = \mu(d - l/j) = 0 \), then

\[
D \int_0^1 F_l(d - y) \mu(dy) = \int_0^1 1_{[d-1/2, d-1/2]}(d - y) \mu(dy).
\]

Proof. Fix \( d \in \mathbb{R} \) and \( l \in \mathbb{Z} \). Let \( (t_n) \) be any sequence such that \( t_n \rightarrow 0 \) with \( t_n \neq 0 \). For all \( y \in \Psi \) and all integer \( n \), define

\[
f_n(y) = 1/t_n |F_l(d + t_n - y) - F_l(d - y)|.
\]

First, we note that \( f_n \) are measurable functions of \( y \) and that \( f_n(y) \rightarrow 1_{[d-1/2, d-1/2]} \) as \( j \rightarrow \infty \), for all \( y \in \Psi \setminus \{d - (l-1)/j, d - l/j\} \). Then, by noting that, for all \( a, b \in \mathbb{R} \)

\[
|F_l(a) - F_l(b)| \leq |a - b|,
\]

we have that \( f_n(y) \leq 1 \) for all \( y \in \Psi \). Thus, by the Dominated-Convergence theorem, we conclude that

\[
\lim_n \int_0^1 f_n(y) \mu(dy) = \int_0^1 1_{[d-1/2, d-1/2]}(d - y) \mu(dy).
\]

hence the result. \( \square \)

Lemma 13. Under \( H_\alpha \), we have

\[
\sup_{t \in [0,1]} |A(t)| \leq 2/j0o(\phi^*, 2/j),
\]

where

\[
A(t) = |\Phi^*(t + 1/j) - \phi^*(t)| - |\Phi^*_j(t + 1/j) - \phi^*_j(t)|.
\]

Proof. Let \( t \in [0,1] \) and take \( k \in \{-j, \ldots, j\} \) such that \( t \in I_k \) (that is \( (k-1)/j < t \leq k/j \)). From the definition of \( \Phi^*_j \) for all \( t \in I_k \), we have

\[
\Phi^*_j(t) = \Phi^*_j((k-1)/j) + j[t - (k-1)/j] [\Phi^*(k/j) - \Phi^*_j((k-1)/j)].
\]
Then, it is easy to check that
\[ A(t) = c_k(t) - j(t - k/j + 1/j)c_k(k/j), \]
where
\[ c_k(t) = [\Phi^\infty(t + 1/j) - \Phi^\infty(t)] - [\Phi^\infty(k/j) - \Phi^\infty((k - 1)/j)]. \]
Since \( t \in I_k \), we have that
\[ |A(t)| \leq 2 \sup_{c_k} |c_k(t)| \]
and we conclude the proof by noting that
\[ |c_k(t)| = \left| \int_{t}^{t+1/j} \Phi^\infty(t) \, dt - \int_{k}^{k+1} \Phi^\infty(t) \, dt \right| \leq 1/j \max \left\{ \sup_{c_k} \phi^\infty(t) - \inf_{c_k} \phi^\infty(t), \inf_{c_k} \phi^\infty(t) - \sup_{c_k} \phi^\infty(t) \right\} \leq 1/j \sup_{t \in (k-1)/j} |\phi^\infty(t) - \phi^\infty(t)| \leq 1/j \alpha(\phi^\infty, 2/j). \]

5.3. Proofs

**Proof** (Proposition 2). Take \( u \in \mathcal{U} \) and \( d_u \) such that \( M(u) = \{d_u\} \). First, note that, by the Dominated-Convergence Theorem, it is easy to see that \( u^\infty \) is continuous. As \( u^\infty \) is continuous on the compact set \( \mathcal{U} \), there exists \( d_0 \neq d_u \) such that

\[ \sup_{|d-d_0|>\varepsilon} u^\infty(d) = u^\infty(d_0) < u^\infty(d_u). \]  

(5.2)

From (5.2), we deduce that for all \( \varepsilon > 0 \), there exists \( \eta > 0 \) such that

\[ u^\infty(d_u) - u^\infty(d) < \eta \Rightarrow |d_u - d| < \varepsilon. \]  

(5.3)

By Lemma 11, take \((u_j)\) such that \( \sup_{d \in (0,1]} |u_j^\infty(d) - u^\infty(d)| \rightarrow 0 \) as \( j \rightarrow \infty \). Take \( \varepsilon > 0 \). Since, by definition \( u_j^\infty(t_j) \Downarrow u_j^\infty(d_u) \), we have

\[ u^\infty(d_u) - u^\infty(t_j) = u^\infty(d_u) - u_j^\infty(t_j) + u_j^\infty(t_j) - u^\infty(t_j) \leq 2 \sup_{d \in (0,1]} |u_j^\infty(d) - u^\infty(d)|. \]

Thus, by (5.3), \( |d_u - t_j| < \varepsilon \) for \( j \) large enough. \( \Box \)

**Proof** (Proposition 3). It is easily seen that

\[ D u^\infty(d) - \hat{D} u^\infty(d) = \sum_{l=-j}^{l+d} A_l(d - (l - 1)/j), \]

where \( A_l \) is defined in Lemma 13. We conclude by using Lemma 13. \( \Box \)

**Proof** (Proposition 6). Let

\[ C_l(t) = 1/|\Phi^\infty(t) - \Phi^\infty(t)| \]

and define \( B_l \) by replacing \( \Phi \) by \( \Psi \) in the definition of \( A_l \) in Lemma 13. From (3.7), it is easy to check that

\[ I_l(d) - \tilde{T}_l(d) = \left( d - \frac{l-1}{j} \right) A_l(d - l/j) + B_l(d - l/j) + C_l(d - l/j). \]

By Lemma 13, \( |A_l(t)| \) is uniformly bounded by \( 2/j \alpha(\psi^\infty, 2/j) \) and, by replacing \( \Phi \) by \( \Psi \) in the proof of Lemma 13, it is easy to check that

\[ \sup_{t \in (0,1]} |B_l(t)| \leq 2/j \alpha(\psi^\infty, 2/j). \]

By noting that \( \phi^\infty \) and \( \phi_j^\infty \) are increasing and that \( \Phi^\infty(l/j) = \Phi_j^\infty(l/j) = \Phi^\infty((l-1)/j) \), we have

\[ \sup_{t \in (0,1]} |\phi^\infty(t) - \phi_j^\infty(t)| = \max_{l \in [1,j]} \sup_{t \in [l-1,j]} |\phi^\infty(t) - \phi_j^\infty(t)| \leq \max_{l \in [1,j]} \mu(l), \]

(5.4)

hence a bound for \( |C_l(d - l/j)| \). Finally, as \( |d - l/j| \leq 1 \), we have, for all \( d \in [0,1] \), that \( |I_l(d) - \tilde{T}_l(d)| \leq B_l \) and

\[ |u_j^\infty(d) - \hat{u}_j^\infty(d)| \leq \frac{2j+1}{j} B_j. \]  

\( \Box \)
Proof (Proposition 7). Recall that

\[ \hat{u}_l(d) = \sum_{i=0}^{j} \gamma_i \hat{u}_l(d), \]

where \( \hat{u}_l(d) \) is defined by replacing \( \phi^o \) and \( \psi^o \) by \( \phi^o \) and \( \psi^o \) in (3.7). From the proof of Proposition 4, we have that

\[ \phi^o_j(d - \frac{i-1}{j}) - \phi^o_j(d - \frac{i}{j}) = \sum_{k=0}^{j-1} A_{jk} F_l(d). \]

As in the proof of Proposition 4, it can be proved that

\[ \psi^o_j(d - \frac{i-1}{j}) - \psi^o_j(d - \frac{i}{j}) = \sum_{k=0}^{j-1} A_{jk} F_l(d). \]

We finish the proof by noting that

\[ \phi^o_j(d - \frac{i}{j}) = \sum_{k=0}^{i} \mu_k F_l(d - i/j) = \sum_{k=0}^{j-1} 1_{l + 1} \mu_k F_l(d). \]

\[ \square \]

6. Discussion

We begin the paper by noting that the usual methods for the construction of an utility function provide a class \( U \) of utility functions rather than a single one. This class \( U \) is made up of continuous functions with some constraints on the values and the shape. We develop a method for computing the lowest and greatest Bayes actions when the utility function \( u \) ranges over \( U \). This method is applied in an example where the decision \( d \in [0, 1] \) is the height of a dam. In the example, the numerical computations take less than 2 minutes and it is obtained that the optimal height ranges over \( [0.2408, 0.6624] \) when \( u \) ranges over a class \( U \).

In this paper, it is assumed that the consequence \( R \) of choosing a decision \( d \in \mathcal{D} \) is of the form \( R = d - Y \) where \( Y \in \mathcal{U} \) is a random variable with a distribution independent of \( d \) (H2). Let us show that the results of the paper can be adapted if this assumption is dropped. Roughly, we can summarize the method as follows. The optimal decision for the utility function \( u \) is the maximizer of

\[ u^*(d) = \int_{\mathcal{U}} u(d - y) \mu(dy). \]

The first discretization gives \( u(r) \approx \sum_{l=1}^{j} \gamma_l \hat{u}_l(r) \) (see Section 3 for the definition of \( F \)) and, by the second discretization, we have

\[ u^*(d) \approx \sum_{l=1}^{j} \sum_{i=1}^{j} \gamma_l \mu F_l(d - i/j). \]

As the two discretizations are done with the same precision \( (1/j) \), by Lemma 5, \( u^*(d) \) can be approximated by a piecewise linear function of \( d \):

\[ u^*(d) \approx \sum_{l=1}^{j} \sum_{i=1}^{j} \gamma_l \mu F_l(d - i/j) = \sum_{l=1}^{j} \sum_{i=1}^{j} \gamma_l \mu F_{l+i}(d) \]

and the computation of the Bayes actions set can be done by using the simulated annealing algorithm. Assume now that H2 is dropped. The distribution of \( R \) does depend on \( d \) and \( u^*(d) \) is now of the form:

\[ u^*(d) = \int u(r) \mu_d(dr). \]

By the first discretization, we have

\[ u^*(d) \approx \sum_{l=1}^{j} \gamma_l \int_{\mathcal{U}} F_l(r) \mu_d(dr). \]

Now, by the discretization of \( \mathcal{D} \), we have, for \( d \in I_l \),

\[ u^*(d) \approx \sum_{l=1}^{j} \gamma_l \int_{\mathcal{U}} F_l(r) \mu_{d_l}(dr) = \sum_{l=1}^{j} \gamma_l c_{d_l}. \]

where \( c_{d_l} = \int_{\mathcal{U}} F_l(r) \mu_{d_l}(dr) \). Thus, it is possible to compute the set \( I_l \) that contains the optimal decision for \( u \). As the numbers \( c_{d_l} \) do not depend on \( u \), we still can approximate the lowest and greatest optimal decisions by using the simulated annealing algorithm.
Acknowledgements

The author wishes to thank the referees for many helpful comments which have led to improvements in the presentation of the paper.

References