Matthias C. M. Troffaes. Finite approximations to coherent choice. *International Journal of Approximate Reasoning*, 50(4):655-665, April 2009.

# Finite Approximations To Coherent Choice

Matthias C. M. Troffaes

*Department of Mathematical Sciences, Durham University*
*Science Laboratories, South Road*
*Durham, DH1 3LE, United Kingdom*

**Abstract**

This paper studies and bounds the effects of approximating loss functions and credal sets on choice functions, under very weak assumptions. In particular, the credal set is assumed to be neither convex nor closed. The main result is that the effects of approximation can be bounded, although in general, approximation of the credal set may not always be practically possible. In case of pairwise choice, I demonstrate how the situation can be improved by showing that only approximations of the extreme points of the closure of the convex hull of the credal set need to be taken into account, as expected.

*Key words:* decision making, E-admissibility, maximality, numerical analysis, lower prevision, sensitivity analysis

## 1 Introduction

Classical decision theory tells a decision maker to choose that option which maximises his expected utility. A generalisation of this principle is compelling when the probabilities and utilities relevant to the problem are not well known. Choice functions are one such generalisation, and select a set of optimal options: instead of pointing to a single solution based on possibly wrong assumptions, choice functions provide a set of optimal options. The decision maker can then investigate further if the set is too large, or not, if for instance the optimal set is a singleton, or if a single option from the set stands out from the rest by other arguments.

*Email address:* `matthias.troffaes@gmail.com` (Matthias C. M. Troffaes).

However, in modelling decision problems, we often afford ourselves the luxury of infinite spaces and infinite sets, making those problems sometimes hard to solve analytically. In such cases we must resort to computers, and these cannot handle random variables on infinite spaces, let alone arbitrary infinite sets of probabilities. Hence, in that case we must approximate our infinite sets by finite ones. By taking the finite sets sufficiently large, hopefully the approximation reflects the true result accurately. This paper confirms this intuition when modelling choice functions induced by arbitrary (not necessarily convex) sets of probabilities and a single cardinal utility, extending similar results known in classical decision theory [1,2].

The paper is organised as follows. Section 2 introduces notation, and briefly reviews the theory of coherent choice functions and their role in decision theory. In Section 3 the building blocks for a theory of approximation are introduced, along with some useful results on what they imply for loss functions, sets of probabilities, and expected utility. The main part of the paper begins in Section 4, studying and bounding the effects of approximation on coherent choice functions. Section 5 improves the results of the previous section for pairwise choice. Section 6 concludes the paper. Some essential but technical results on approximating the standard simplex in $\mathbb{R}^n$ are deferred to an appendix.

## 2 Choice Functions

Let $\Omega$ denote an arbitrary set of states. Bounded random quantities on $\Omega$, i.e. bounded maps from $\Omega$ to $\mathbb{R}$, are also called *gambles* [3], and will be denoted by $f$, $g$, … $\mathcal{L}(\Omega)$ denotes the set of all gambles on $\Omega$. Finitely additive probability measures, or briefly *probability charges* [4], are denoted by $P$, $Q$, … and $\mathcal{P}(\Omega)$ denotes the set of all probability charges on the power set $\wp(\Omega)$ of $\Omega$.

In a decision problem, we desire to choose an optimal option $d$ from a set $D$ of options. Choosing $d$ induces an uncertain reward $r$ from a set $R$ of rewards, with probability charge $\mu_d(\cdot|w)$ over $\wp(R)$, depending on the outcome of the uncertain state $w \in \Omega$. For each $w \in \Omega$, $\mu_d(\cdot|w)$ is a *lottery* over $R$, and as a function of $w$, $\mu_d(\cdot|\cdot)\colon w \mapsto \mu_d(\cdot|w)$ is a *horse lottery* or *act*.

If we model our belief about states and rewards by a probability charge $P$ on $\wp(\Omega)$ and a state dependent utility function $U(\cdot|w)$ on $R$, then utility theory [5,6,7] tells us to choose a decision $d$ which maximises the expected utility, or prevision:

$$\begin{aligned} E(d) &= \int_\Omega \left( \int_R U(r|w)\,\mathrm{d}\mu_d(r|w) \right) \mathrm{d}P(w) \\ &= \int_\Omega f_d(w)\,\mathrm{d}P(w) \end{aligned}$$

where $f_d(w) = \int_R U(r|w)\,\mathrm{d}\mu_d(r|w)$ is the gamble associated with decision $d$, and the integrals are Dunford integrals [4]. For simplicity, in this paper, we assume $U(r|w)$ to be bounded, i.e.

$$\sup_{r,w} U(r|w) - \inf_{r,w} U(r|w) < +\infty$$

Among other things, this ensures that relative approximation can be defined, as in Section 3, without technical complications.

A decision which maximises expected utility is called a *Bayes decision* for the decision problem $(\Omega, D, P, U)$.

However, if we are not sure about the probability of all events and the utility of all rewards, a more reliable design is to use a family $(P_\alpha, U_\alpha)_{\alpha\in\aleph}$ of probability-utility pairs (where $\aleph$ is an arbitrary index set), and to elicit from $D$ those options which maximise expected utility with respect to at least one of the pairs $(P_\alpha, U_\alpha)$. First, for each $\alpha \in \aleph$, let

$$E_\alpha(d) = \int_\Omega f_d^\alpha(w)\,\mathrm{d}P_\alpha(w)$$

where $f_d^\alpha(w) = \int_R U_\alpha(r|w)\,\mathrm{d}\mu_d(r|w)$ is the gamble associated with decision $d$ and model $\alpha \in \aleph$. Then we define:

**Definition 1** *A decision $d \in D$ is called an* optimal decision *for the decision problem $(\Omega, D, (P_\alpha, U_\alpha)_{\alpha\in\aleph})$ if $d$ belongs to the set*

$$\begin{aligned} \operatorname{opt}(\Omega, D, (P_\alpha, U_\alpha)_{\alpha\in\aleph}) &= \{d \in D\colon (\exists\alpha \in \aleph)(\forall e \in D)(E_\alpha(d) \geq E_\alpha(e))\} \\ &= \left\{ d \in D\colon (\exists\alpha \in \aleph) \left( E_\alpha(d) = \sup_{e\in D} E_\alpha(e) \right) \right\} \end{aligned}$$

As such, the operator opt selects a *set* of optimal decisions, namely all decisions which are Bayes with respect to $(\Omega, D, P_\alpha, U_\alpha)$ for at least one $\alpha \in \aleph$. Such an operator is called a *choice function* or *optimality operator* [8,9].

In case $(P_\alpha, U_\alpha)_{\alpha\in\aleph} = \mathcal{M} \times \mathcal{U}$ for some convex sets $\mathcal{M}$ and $\mathcal{U}$, optimality as defined above is also called *E-admissibility* [10, Sec. 4.8].

There are many ways to define a choice function starting from a set $(P_\alpha, U_\alpha)_{\alpha\in\aleph}$ (see [10,11,3,12,9]). The one in Definition 1 satisfies an interesting set of axioms [12,13], and is the subject of a representation theorem in case utility is precise and state independent (i.e. if $U_\alpha(r|w)$ depends neither on $\alpha$ nor on $w$) and $\Omega$ is finite (for infinite $\Omega$ the representation theorem is subject to additional constraints, which preclude merely finitely additive probabilities over $\Omega$) [13].

For the sake of simplicity, we shall only be concerned about decision problems with precise and state independent utility functions, i.e. when $(P_\alpha, U_\alpha)_{\alpha\in\aleph} = \mathcal{M} \times \{U\}$ with $U\colon R \to \mathbb{R}$ a bounded state independent utility over $R$ and

$$\mathcal{M} = \{P_\alpha \colon \alpha \in \aleph\}$$

The set $\mathcal{M}$ is called a *credal set* as it represents our belief about $w \in \Omega$. We can identify $\mathcal{M}$ itself as index set, and write

$$E_P(d) = \int_\Omega f_d(w)\,\mathrm{d}P(w)$$

with $f_d(w) = \int_R U(r)\,\mathrm{d}\mu_d(r|w)$, for any $P \in \mathcal{M}$.

Finally, defining the loss function $L\colon D\times\Omega \to \mathbb{R}$ as $L(d,w) = -f_d(w)$, the expected value $E_P(d)$ is uniquely determined by $P$ and $L$ alone: we need not be concerned explicitly with $R$, $\mu_d(r|w)$, and $U(r)$.

## 3 Approximate Gambles, Probabilities, and Previsions

Let $\mathcal{A} = \{A_1, \dots, A_n\}$ denote a finite partition of $\Omega$. As we approximate $\Omega$ by the finite set $\mathcal{A}$, we also need to approximate decisions, gambles, and probability charges on $\Omega$.

Let $\epsilon \geq 0$. For a gamble $f$ in $\mathcal{L}(\Omega)$ and a gamble $\hat{f}$ in $\mathcal{L}(\mathcal{A})$, we shall write $f \sim_\epsilon \hat{f}$ if

$$\max_{A\in\mathcal{A}} \sup_{w\in A} \left| f(w) - \hat{f}(A) \right| \leq [\sup f - \inf f]\epsilon$$

Note that $f \sim_\epsilon \hat{f}$ implies $af + b \sim_\epsilon a\hat{f} + b$, for any real numbers $a$ and $b$, $a > 0$. Therefore, the relation $\sim_\epsilon$ is invariant with respect to positive linear transformations of utility: it only depends on our preferences over lotteries, and not on our particular choice of utility scale.

For a probability charge $P$ in $\mathcal{P}(\Omega)$, and a probability charge $\hat{P}$ in $\mathcal{P}(\mathcal{A})$, we shall write $P \sim_\epsilon \hat{P}$ if

$$\sum_{A \in \mathcal{A}} \left| P(A) - \hat{P}(A) \right| \leq \epsilon$$

Note that this implies $|P(A) - \hat{P}(A)| \leq \epsilon$ for any $A \in \wp(\mathcal{A})$. Also note the differences between the definitions of $\sim_\epsilon$ for gambles and bounded charges.

For a loss function $L$ on $D \times \Omega$ and a loss function $\hat{L}$ on $D \times \mathcal{A}$ we write $L \sim_\epsilon \hat{L}$ if for all $d \in D$

$$f_d \sim_\epsilon \hat{f}_d$$

(with $f_d(w) = -L(d, w)$ and $\hat{f}_d(A) = -\hat{L}(d, A)$).

For a subset $\mathcal{M}$ of $\mathcal{P}(\Omega)$ and a subset $\hat{\mathcal{M}}$ of $\mathcal{P}(\mathcal{A})$, we write $\mathcal{M} \sim_\epsilon \hat{\mathcal{M}}$ if for every $P$ in $\mathcal{M}$ there is a $\hat{P}$ in $\hat{\mathcal{M}}$ such that $P \sim_\epsilon \hat{P}$, and for every $\hat{P}$ in $\hat{\mathcal{M}}$ there is a $P$ in $\mathcal{M}$ such that $P \sim_\epsilon \hat{P}$.

A few useful results about approximations are stated in the next lemmas.

**Lemma 2** *Assume that $D$ is finite. Then, for every loss function $L$ on $D \times \Omega$ and every $\epsilon > 0$, there is a finite partition $\mathcal{A}$ of $\Omega$ and a loss function $\hat{L}$ on $D \times \mathcal{A}$ such that $L \sim_\epsilon \hat{L}$ and $|\mathcal{A}| \leq (1 + 1/\epsilon)^{|D|}$.*

**PROOF.** Consider any $d$ in $D$, and let $R_d = \sup f_d - \inf f_d$. Because $f_d$ is bounded, we can embed the range of $f_d$ in $k$ intervals $I_1, \ldots, I_k$ of length $R_d\epsilon$, say

$$[\inf f_d, \inf f_d + R_d\epsilon),\ [\inf f_d + R_d\epsilon, \inf f_d + 2R_d\epsilon),$$
$$\ldots, [\inf f_d + (k-1)R_d\epsilon, \inf f_d + kR_d\epsilon)$$

with $k$ such that $\sup f_d \in I_k$. Therefore, $\inf f_d + (k-1)R_d\epsilon \leq \sup f_d < \inf f_d + kR_d\epsilon$ and hence $k - 1 \leq 1/\epsilon < k$. Observe that $k$ is independent of $d \in D$.

The sets $A_1, \ldots, A_k$ defined by

$$A_j = f_d^{-1}(I_j)$$

form a finite partition $\mathcal{A}_d = \{A_j \colon A_j \neq \emptyset\}$ of cardinality $|\mathcal{A}_d| \leq k \leq 1 + 1/\epsilon$ and the gamble $\hat{f}_d \in \mathcal{L}(\mathcal{A}_d)$ defined by

$$\hat{f}_d(A_i) = \inf_{w \in A_i} f_d(w)$$

Table 1
Upper bound on $\log_{10}(|\mathcal{A}|)$, i.e. the logarithm of the cardinality of the finite partition $\mathcal{A}$ for various values of precision $\epsilon > 0$ and number of decisions (see Lemma 2).

| | $\epsilon$: | | | | |
|---|---|---|---|---|---|
| | 0.2 | 0.1 | 0.05 | 0.02 | 0.01 |
| $\lvert D\rvert$: 2 | 1.6 | 2.1 | 2.6 | 3.4 | 4.0 |
| 4 | 3.1 | 4.2 | 5.3 | 6.8 | 8.0 |
| 8 | 6.2 | 8.3 | 10.6 | 13.7 | 16.0 |
| 16 | 12.5 | 16.7 | 21.2 | 27.3 | 32.1 |
| 32 | 24.9 | 33.3 | 42.3 | 54.6 | 64.1 |

satisfies

$$\begin{aligned}\sup_{w \in A_j} \left| f_d(w) - \hat{f}_d(A_j) \right| &= \sup_{f_d(w) \in I_j} \left| f_d(w) - \inf_{f_d(w) \in I_j} f_d(w) \right| \\ &\leq \sup I_j - \inf I_j = R_d \epsilon\end{aligned}$$

for all $A_j \in \mathcal{A}_d$; hence $f_d \sim_\epsilon \hat{f}_d$. Defining $\hat{L}(d, A) = -\hat{f}_d(A)$ for all $d \in D$, we have $L \sim_\epsilon \hat{L}$.

The finite collection of partitions $\{\mathcal{A}_d : d \in D\}$ has a smallest common refinement $\mathcal{A}$. Since each $\mathcal{A}_d$ has no more than $1 + 1/\epsilon$ elements, $\mathcal{A}$ has no more than $(1 + 1/\epsilon)^{|D|}$ elements. Indeed, two partitions of cardinalities $k_1$ and $k_2$ respectively have a smallest common refinement of cardinality no more than $k_1 k_2$. By induction, $n$ partitions of cardinalities $k_1, \ldots, k_n$ have a smallest common refinement of cardinality no more than $\prod_{j=1}^n k_j$ and hence,

$$|\mathcal{A}| \leq (1 + 1/\epsilon)^{|D|}$$

Table 1 lists upper bounds on the size of the partition, to ensure $L \sim_\epsilon \hat{L}$, for various values of $\epsilon$ and $|D|$, according to Lemma 2.

Let $\binom{a}{b}$ be the binomial coefficient, defined for all real numbers $a \geq b \geq 0$ by

$$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}$$

with $\Gamma$ the Gamma function.

**Lemma 3** *For every subset $\mathcal{M}$ of $\mathcal{P}(\Omega)$, every $\delta > 0$, and every finite partition $\mathcal{A}$ of $\Omega$, there is a finite subset $\hat{\mathcal{M}}$ of $\mathcal{P}(\mathcal{A})$ such that $\mathcal{M} \sim_\delta \hat{\mathcal{M}}$ and $|\hat{\mathcal{M}}| \leq \binom{|\mathcal{A}|(1+1/\delta)}{|\mathcal{A}|-1}$.*

**PROOF.** Consider any $P$ in $\mathcal{M}$. Let $n = |\mathcal{A}|$ and let the elements of $\mathcal{A}$ be $A_1$, ..., $A_n$. Consider the vector $\underline{x} = (P(A_1), \dots, P(A_n))$ in $\Delta^n$. Let $N$ be the smallest natural number such that $N \geq n/\delta$.

By Lemma 13 in the appendix, there is a vector $\underline{y}$ in $\Delta^n_N$ such that

$$|\underline{x} - \underline{y}|_1 < n/N \leq \delta$$

Define $\hat{P}$ in $\mathcal{P}(\mathcal{A})$ by

$$\hat{P}(A_i) = y_i$$

for all $i \in \{1, \dots, n\}$—by finite additivity, $\hat{P}$ is well defined on $\wp(\mathcal{A})$. By construction, $P \sim_\delta \hat{P}$ because

$$\sum_{i=1}^{n} \left| P(A_i) - \hat{P}(A_i) \right| = |\underline{x} - \underline{y}|_1 < \delta$$

Approximating each $P$ in $\mathcal{M}$ in this manner, the set

$$\hat{\mathcal{M}} = \{\hat{P} \colon P \in \mathcal{M}\}$$

is finite as each of its elements corresponds to an element of the finite set $\Delta^n_N$, and therefore $|\hat{\mathcal{M}}| \leq |\Delta^n_N|$. By Lemma 12 in the appendix,

$$\begin{aligned} |\hat{\mathcal{M}}| &\leq \binom{N+n-1}{N} = \binom{N+n-1}{n-1} \\ &\leq \binom{n/\delta + 1 + n - 1}{n-1} = \binom{|\mathcal{A}|(1+1/\delta)}{|\mathcal{A}|-1} \end{aligned}$$

The second inequality follows from the fact that $\binom{a}{b}$ is strictly increasing in $a$, for fixed $b$ (for integer $a$ and $b$ this follows immediately from Pascal's triangle; the general case follows from the properties of the Gamma function).

Table 2 lists upper bounds on the cardinality of $\hat{\mathcal{M}}$ on a logarithmic scale, for some values of $|\mathcal{A}|$ and $\delta$. The cardinality grows enormously fast with increasing $|\mathcal{A}|$ and $1/\delta$. Within the range of Table 2, an exponential trend is obvious. The

Table 2
Upper bound on $\log_{10}(|\hat{\mathcal{M}}|)$, i.e. the logarithm of the cardinality of the finite set of probability charges $\hat{\mathcal{M}}$, for various values of precision $\delta > 0$ and cardinality of the partition $|\mathcal{A}|$ (see Lemma 3).

| | $\delta$: | | |
|---|---|---|---|
| | 0.2 | 0.1 | 0.05 |
| $\lvert\mathcal{A}\rvert$: 4 | 3.3 | 4.1 | 5.0 |
| 8 | 7.9 | 9.8 | 11.8 |
| 12 | 12.5 | 15.5 | 18.7 |
| 16 | 17.1 | 21.3 | 25.6 |
| 20 | 21.8 | 27.1 | 32.6 |
| 24 | 26.4 | 32.9 | 39.5 |
| 28 | 31.1 | 38.6 | 46.5 |
| 32 | 35.8 | 44.4 | 53.4 |
| $\log_{10}(\lvert\mathcal{A}\rvert)$: 0.7 | 4.4 | 5.5 | 6.7 |
| 1.4 | 27.6 | 34.3 | 41.3 |
| 2.1 | 144.6 | 179.5 | 215.5 |
| 2.8 | 731.3 | 906.8 | 1088.2 |
| 3.5 | 3666.1 | 4544.7 | 5452.8 |
| 4.2 | 18341.5 | 22735.9 | 27277.5 |
| 4.9 | 91719.7 | 113693.0 | 136402.5 |

table shows that the influence of $|\mathcal{A}|$ is much larger than the influence of $\delta$: more precisely, doubling $|\mathcal{A}|$ increases $|\hat{\mathcal{M}}|$ by far more than halving $\delta$.

Next, we study the effect on the expectation if both gambles and probabilities are approximated. Let us use the notation $E_P(f) = \int_\Omega f(w)\,\mathrm{d}P(w)$. In the lemma below, assume $0 < \epsilon < 1/2$.

**Lemma 4** *For every finite partition $\mathcal{A}$ of $\Omega$, every $f \in \mathcal{L}(\Omega)$, $\hat{f} \in \mathcal{L}(\mathcal{A})$, $P \in \mathcal{P}(\Omega)$, and $\hat{P} \in \mathcal{P}(\mathcal{A})$, the following implications hold. If $f \sim_\epsilon \hat{f}$ and $P \sim_\delta \hat{P}$ then*

$$\left|E_P(f) - E_{\hat{P}}(\hat{f})\right| \leq [\sup f - \inf f](\epsilon + \delta(1 + 2\epsilon))$$

*and*

$$\left|E_P(f) - E_{\hat{P}}(\hat{f})\right| \leq [\sup \hat{f} - \inf \hat{f}] \left(\frac{\epsilon}{1-2\epsilon} + \delta\right)$$

**PROOF.** Let $R = \sup f - \inf f$, $\hat{R} = \sup \hat{f} - \inf \hat{f}$, and write $\inf_A f$ for $\inf_{w \in A} f(w)$ and $\sup_A f$ for $\sup_{w \in A} f(w)$. Then

$$\left|E_P(f) - E_{\hat{P}}(\hat{f})\right| = \left|\sum_{A \in \mathcal{A}} \left(\int_A f \, \mathrm{d}P - \hat{f}(A)\hat{P}(A)\right)\right|$$

and since $P(A) \inf_A f \leq \int_A f \, \mathrm{d}P \leq P(A) \sup_A f$, there is an $r_A \in [\inf_A f, \sup_A f]$ such that $P(A) r_A = \int_A f \, \mathrm{d}P$, and hence

$$= \left|\sum_{A \in \mathcal{A}} \left(r_A P(A) - \hat{f}(A)\hat{P}(A)\right)\right|$$

but, because $|f(w) - \hat{f}(A)| \leq R\epsilon$ for all $w \in A$, and $\inf_A f \leq r_A \leq \sup_A f$, it must also hold that $|r_A - \hat{f}(A)| \leq R\epsilon$, so $\left|\sum_{A \in \mathcal{A}} \left(r_A P(A) - \hat{f}(A) P(A)\right)\right| \leq \sum_{A \in \mathcal{A}} \left|r_A - \hat{f}(A)\right| P(A) \leq \sum_{A \in \mathcal{A}} R\epsilon P(A) = R\epsilon$, whence

$$\begin{aligned}
&\leq \left|\sum_{A \in \mathcal{A}} \left(\hat{f}(A)P(A) - \hat{f}(A)\hat{P}(A)\right)\right| + R\epsilon \\
&= \left|\sum_{A \in \mathcal{A}} \hat{f}(A) \left(P(A) - \hat{P}(A)\right)\right| + R\epsilon
\end{aligned}$$

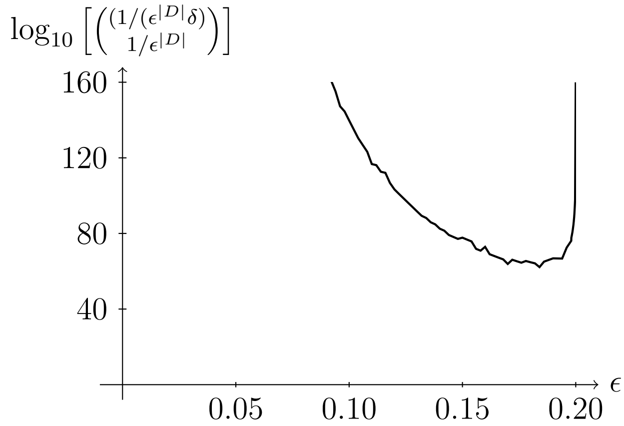
and because $\sum_{A \in \mathcal{A}} (P(A) - \hat{P}(A)) = 0$,

$$\begin{aligned}
&= \left|\sum_{A \in \mathcal{A}} (\hat{f}(A) - \inf \hat{f}) \left(P(A) - \hat{P}(A)\right)\right| + R\epsilon \\
&\leq \sum_{A \in \mathcal{A}} (\hat{f}(A) - \inf \hat{f}) \left|P(A) - \hat{P}(A)\right| + R\epsilon \\
&\leq (\sup \hat{f} - \inf \hat{f}) \sum_{A \in \mathcal{A}} \left|P(A) - \hat{P}(A)\right| + R\epsilon \\
&\leq \hat{R}\delta + R\epsilon
\end{aligned}$$

and since $R(1 + 2\epsilon) \geq \hat{R} \geq R(1 - 2\epsilon)$

$$\leq \begin{cases} R(1+2\epsilon)\delta + R\epsilon = R(\epsilon + \delta(1+2\epsilon)) \\ \hat{R}\delta + \hat{R}\epsilon/(1-2\epsilon) = \hat{R}\left(\epsilon/(1-2\epsilon) + \delta\right) \end{cases}$$

Fig. 1. Upper bound on $\log_{10}|\hat{\mathcal{M}}|$ for various values of $\epsilon$, with $\epsilon+\delta=0.2$ and $|D|=2$.




Let us now investigate what is the most optimal choice for $\epsilon>0$ and $\delta>0$. The cardinality of $\hat{\mathcal{M}}$ is of largest concern as it grows enormously fast with increasing cardinality of the finite partition $\mathcal{A}$ and with increasing precision $1/\delta$ (see Table 2). Therefore, as a first step, let us see how we can minimise $|\hat{\mathcal{M}}|$, assuming a fixed relative error $\epsilon+\delta$ on the expectation (see Lemma 4)—omitting higher order terms in $\epsilon$ and $\delta$ to simplify the analysis.

We wish to minimise the upper bound (neglecting lower order terms)

$$\binom{(1/(\epsilon^{|D|}\delta)}{1/\epsilon^{|D|}}$$

on $|\hat{\mathcal{M}}|$ along the $\epsilon$–$\delta$-curve $\gamma(\epsilon,\delta)=\epsilon+\delta=\gamma_*$. Figure 1 demonstrates a typical case: the $\epsilon$–$\delta$-ratio has a large impact on the upper bound of $|\hat{\mathcal{M}}|$. In particular, the curve grows extremely large for small $\epsilon$, because a small $\epsilon$ corresponds to a large partition $\mathcal{A}$, and the cardinality of the partition has a huge impact on the cardinality of $\mathcal{M}$ as shown in Table 2.

## 4 Approximate Choice

Let us now consider again the decision problem $(\Omega, D, \mathcal{M}, L)$ with state space $\Omega$, decision space $D$, credal set $\mathcal{M}$, and loss function $L$, and reflect upon how the results in the previous section could be of use in finding the optimal decisions $\mathrm{opt}(\Omega, D, \mathcal{M}, L)$. Can we still find the optimal decisions after approximating the loss function $L$ and the set of probabilities $\mathcal{M}$?

As we admit a relative error on gambles and probabilities, and therefore also on previsions, we should admit a relative error on the choice function as well. Let $R_D$ be defined by (recall that $f_d(w) = -L(d, w)$)

$$R_D = \sup_{d \in D} [\sup f_d - \inf f_d]$$

**Definition 5** *Let $\epsilon \geq 0$. A decision $d$ in $D$ is called an* $\epsilon$-optimal decision *for the decision problem $(\Omega, D, \mathcal{M}, L)$ if it belongs to the set*

$$\mathrm{opt}^{\epsilon}(\Omega, D, \mathcal{M}, L) = \left\{ d \in D \colon (\exists P \in \mathcal{M}) \left( \sup_{e \in D} E_P(e) - E_P(d) \leq \epsilon R_D \right) \right\}$$

Note that

$$\mathrm{opt}^{\epsilon}(\Omega, D, \mathcal{M}, aL + b) = \mathrm{opt}^{\epsilon}(\Omega, D, \mathcal{M}, L)$$

for any real numbers $a$ and $b$, $a > 0$. In other words, $\mathrm{opt}^{\epsilon}(\Omega, D, \mathcal{M}, L)$ is invariant with respect to positive linear transformations of utility: $\epsilon$-optimality does not depend on our choice of utility scale.

Clearly,

$$\mathrm{opt}(\Omega, D, \mathcal{M}, L) \subseteq \mathrm{opt}^{\epsilon}(\Omega, D, \mathcal{M}, L)$$

because

$$\mathrm{opt}^{\epsilon}(\Omega, D, \mathcal{M}, L) \subseteq \mathrm{opt}^{\delta}(\Omega, D, \mathcal{M}, L)$$

whenever $\epsilon \leq \delta$, and

$$\mathrm{opt}^{0}(\Omega, D, \mathcal{M}, L) = \mathrm{opt}(\Omega, D, \mathcal{M}, L)$$

In approximating a decision problem $(\Omega, D, \mathcal{M}, L)$, we start with a finite partition $\mathcal{A}$, consider a (possibly finite) set $\hat{\mathcal{M}}$ such that $\mathcal{M} \sim_{\delta} \hat{\mathcal{M}}$, and approximate the loss $L(d, w)$ by a loss $\hat{L}(d, A)$ such that $L \sim_{\epsilon} \hat{L}$.

**Theorem 6** *Consider two decision problems $(\Omega, D, \mathcal{M}, L)$ and $(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L})$. If $L \sim_{\epsilon} \hat{L}$ and $\mathcal{M} \sim_{\delta} \hat{\mathcal{M}}$ then, for any $\gamma \geq 0$,*

$$\mathrm{opt}^{\gamma}(\Omega, D, \mathcal{M}, L) \subseteq \mathrm{opt}^{\frac{\gamma}{1-2\epsilon} + 2(\frac{\epsilon}{1-2\epsilon} + \delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) \tag{1}$$

*and*

$$\mathrm{opt}^{\gamma}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) \subseteq \mathrm{opt}^{\gamma(1+2\epsilon) + 2(\epsilon + \delta(1+2\epsilon))}(\Omega, D, \mathcal{M}, L) \tag{2}$$

**PROOF.** We prove Eq. (1). Let $d \in \mathrm{opt}^{\gamma}(\Omega, D, \mathcal{M}, L)$. Then

$$\sup_{e \in D} E_P(f_e) - E_P(f_d) \leq \gamma R_D \tag{3}$$

for some $P \in \mathcal{M}$. Let $\hat{P}$ be such that $P \sim_\delta \hat{P}$. Because, by Lemma 4,

$$\begin{aligned}\left|\sup_{e \in D} E_{\hat{P}}(\hat{f}_e) - \sup_{e' \in D} E_P(f_{e'})\right| &\leq \sup_{e \in D} \left|E_{\hat{P}}(\hat{f}_e) - E_P(f_e)\right| \\ &\leq \sup_{e \in D} [\sup \hat{f}_e - \inf \hat{f}_e](\epsilon/(1-2\epsilon) + \delta) \\ &= (\epsilon/(1-2\epsilon) + \delta)\hat{R}_D \end{aligned} \tag{4}$$

it follows that

$$\sup_{e \in D} E_{\hat{P}}(\hat{f}_e) - E_{\hat{P}}(\hat{f}_d) \leq \sup_{e \in D} E_P(f_e) - E_{\hat{P}}(\hat{f}_d) + (\epsilon/(1-2\epsilon) + \delta)\hat{R}_D$$

and again by Lemma 4,

$$\leq \sup_{e \in D} E_P(f_e) - E_P(f_d) + 2(\epsilon/(1-2\epsilon) + \delta)\hat{R}_D$$

and by Eq. (3),

$$\begin{aligned} &\leq \gamma R_D + 2(\epsilon/(1-2\epsilon) + \delta)\hat{R}_D \\ &\leq [\gamma/(1-2\epsilon) + 2(\epsilon/(1-2\epsilon) + \delta)]\hat{R}_D \end{aligned}$$

hence, $d \in \mathrm{opt}^{\gamma/(1-2\epsilon)+2(\epsilon/(1-2\epsilon)+\delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L})$.

Next, we prove Eq. (2). Let $d \in \mathrm{opt}^{\gamma}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L})$. Then

$$\sup_{e \in D} E_{\hat{P}}(\hat{f}_e) - E_{\hat{P}}(\hat{f}_d) \leq \gamma \hat{R}_D \tag{5}$$

Because, by Lemma 4,

$$\begin{aligned}\left|\sup_{e \in D} E_{\hat{P}}(\hat{f}_e) - \sup_{e' \in D} E_P(f_{e'})\right| &\leq \sup_{e \in D} \left|E_{\hat{P}}(\hat{f}_e) - E_P(f_e)\right| \\ &\leq \sup_{e \in D} [\sup f_e - \inf f_e](\epsilon + \delta(1+2\epsilon)) \\ &= (\epsilon + \delta(1+2\epsilon))R_D \end{aligned} \tag{6}$$

we have that

$$\sup_{e \in D} E_P(f_e) - E_P(f) \leq \sup_{e \in D} E_{\hat{P}}(\hat{f}_e) - E_P(f) + (\epsilon + \delta(1+2\epsilon))R_D$$

and again by Lemma 4,

$$\leq \sup_{e \in D} E_{\hat{P}}(\hat{f}_e) - E_{\hat{P}}(\hat{f}_e) + 2(\epsilon + \delta(1 + 2\epsilon))R_D$$

and by Eq. (5)

$$\leq \gamma \hat{R}_D + 2(\epsilon + \delta(1 + 2\epsilon))R_D$$
$$\leq [\gamma(1 + 2\epsilon) + 2(\epsilon + \delta(1 + 2\epsilon))]R_D$$

so $d \in \mathrm{opt}^{\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathcal{M}, L)$.

If we ignore higher order terms in $\gamma$, $\epsilon$, and $\delta$, then the above theorem says that when moving from an original decision problem to an approximate decision problem, or the other way around, with relative error $\epsilon$ in gambles and relative error $\delta$ in probabilities, the relative error in optimality increases by $2(\epsilon + \delta)$. For example, for small $\epsilon$ and $\delta$ the following holds, up to a small error: if $L \sim_\epsilon \hat{L}$ and $\mathcal{M} \sim_\delta \hat{\mathcal{M}}$, then

$$\mathrm{opt}(\Omega, D, \mathcal{M}, L) \subseteq \mathrm{opt}^{2(\epsilon+\delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) \subseteq \mathrm{opt}^{4(\epsilon+\delta)}(\Omega, D, \mathcal{M}, L)$$

So, the approximate problem with relative error $2(\epsilon + \delta)$ will contain all solutions to the original problem with no relative error, and will, so to speak, not contain any solutions to the original problem with relative error over $4(\epsilon + \delta)$. Because of this property, $\mathrm{opt}^{2(\epsilon+\delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L})$ seems a logical choice when solving decision problems in practice.

## 5 Pairwise Choice

Table 2 reveals that the size of the credal set is a serious computational bottleneck. Therefore, it is worth investigating how the size of $\hat{\mathcal{M}}$ can be reduced, without compromising the accuracy $\delta > 0$. One way to this end is to restrict to pairwise comparisons, i.e. using maximality (see Walley [3, Sec. 3.7–3.9]).

### *5.1 Maximality*

**Definition 7** *A decision $d \in D$ is called a* maximal decision *for the decision problem $(\Omega, D, \mathcal{M}, L)$ if $d$ belongs to the set*

$$\max(\Omega, D, \mathcal{M}, L) = \{d \in D \colon (\forall e \in D)(\exists P \in \mathcal{M})\,(E_P(d) \geq E_P(e))\}$$

Denote by $\operatorname{co}(\mathcal{M})$ the convex hull of $\mathcal{M}$. Obviously it holds that

$$\max(\Omega, D, \mathcal{M}, L) = \max(\Omega, D, \operatorname{co}(\mathcal{M}), L)$$

because for any $\lambda \in [0,1]$ and any two $P$ and $Q$ in $\mathcal{M}$, the inequalities $E_P(d) \geq E_P(e)$ and $E_Q(d) \geq E_Q(e)$ imply the inequality

$$E_{\lambda P+(1-\lambda)Q}(d) \geq E_{\lambda P+(1-\lambda)Q}(e)$$

This does not hold for optimality as defined in Definition 1: assuming $\Omega$ finite, for any two distinct subsets $\mathcal{M}$ and $\mathcal{M}'$ of $\mathcal{P}(\Omega)$, we can always find a set $D$ and a loss function $L$ such that $\operatorname{opt}(\Omega, D, \mathcal{M}, L) \neq \operatorname{opt}(\Omega, D, \mathcal{M}', L)$ (see Kadane, Schervish, and Seidenfeld [12, Thm. 1, p. 53]).

To understand why the above notion of optimality is called maximality, consider the strict partial ordering $>$ on $D$ defined by

$$e > d \iff (\forall P \in \mathcal{M})\,(E_P(e) > E_P(d))$$

for any $d$ and $e$ in $D$, that is, $e$ is strictly preferred to $d$ if $e$ is strictly preferred to $d$ with respect to every $P \in \mathcal{M}$. Then,

$$\max(\Omega, D, \mathcal{M}, L) = \{d \in D \colon (\forall e \in D)(e \not> d)\}$$

so $\max(\Omega, D, \mathcal{M}, L)$ elects those decisions $d$ which are undominated with respect to $>$. Therefore, maximality can be expressed through pairwise preferences only—again in contrast to $\operatorname{opt}(\Omega, D, \mathcal{M}, L)$ as for instance demonstrated by Kadane, Schervish, and Seidenfeld [12, Sec. 4, p. 51].

However, because

$$\operatorname{opt}(\Omega, D, \mathcal{M}, L) \subseteq \max(\Omega, D, \mathcal{M}, L)$$

we may interpret $\max(\Omega, D, \mathcal{M}, L)$ as an approximation to $\operatorname{opt}(\Omega, D, \mathcal{M}, L)$, an approximation which discards all preferences but the pairwise ones.

Let us admit a relative error on the choice function max as well. Recall, $R_D = \sup_{d \in D}[\sup f_d - \inf f_d]$.

**Definition 8** *Let $\epsilon \geq 0$. A decision $d$ in $D$ is called an $\epsilon$-*maximal decision *for the decision problem $(\Omega, D, \mathcal{M}, L)$ if it belongs to the set*

$$\begin{aligned} &\max{}^{\epsilon}(\Omega, D, \mathcal{M}, L) \\ &\qquad = \{d \in D \colon (\forall e \in D)(\exists P \in \mathcal{M})(E_P(e) - E_P(d) \leq \epsilon R_D)\} \end{aligned}$$

### *5.2 Approximating Extreme Points*

It turns out that we can restrict our attention to the extreme points of the closure of the convex hull of $\mathcal{M}$, with respect to the topology of pointwise convergence on members of $\mathcal{L}(\Omega)$. This topology is characterised by the following notion of convergence: for every directed set $(A, \leq)$ and every net $(P_\alpha)_{\alpha \in A}$, we have that $\lim_\alpha P_\alpha = P$ if

$$\lim_\alpha E_{P_\alpha}(f) = E_P(f) \text{ for all } f \in \mathcal{L}(\Omega)$$

Without further mention, I will assume this topology on $\mathcal{P}(\Omega)$. See for instance [14] for more information regarding nets [14, Chapter 7] and this topology [14, §28.15].

There is a nice connection between the closure of $\mathcal{M}$, denoted by $\mathrm{cl}(\mathcal{M})$, and $\epsilon$-optimality and $\epsilon$-maximality.

**Lemma 9** *Assume that $R_D > 0$ and let $\epsilon \geq 0$. For any decision problem $(\Omega, D, \mathcal{M}, L)$, the following equality holds:*

$$\max{}^{\epsilon}(\Omega, D, \mathrm{cl}(\mathcal{M}), L) = \bigcap_{\delta > 0} \max{}^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L) \tag{7}$$

*and if additionally $D$ is finite, then the following equality holds as well:*

$$\mathrm{opt}^{\epsilon}(\Omega, D, \mathrm{cl}(\mathcal{M}), L) = \bigcap_{\delta > 0} \mathrm{opt}^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L) \tag{8}$$

**PROOF.** We start with proving Eq. (7).

Assume $d \in \max^{\epsilon}(\Omega, D, \mathrm{cl}(\mathcal{M}), L)$. Consider any $e \in D$. By assumption, there is a $P \in \mathrm{cl}(\mathcal{M})$ such that $E_P(e) - E_P(d) \leq R_D\epsilon$. Because $P \in \mathrm{cl}(\mathcal{M})$, there is a net $(P_\alpha \in \mathcal{M})_{\alpha \in A}$ such that $\lim_\alpha E_{P_\alpha}(f) = E_P(f)$ for all gambles $f$. It follows that

$\lim_\alpha E_{P_\alpha}(e) - \lim_\alpha E_{P_\alpha}(d) \leq R_D\epsilon$. This implies that for every $\delta > 0$, there is an $\alpha \in A$ such that $E_{P_\alpha}(e) - E_{P_\alpha}(f) \leq (\epsilon+\delta)R_D$. So, for every $\delta > 0$, there is a $P \in \mathcal{M}$ such that $E_P(e) - E_P(f) \leq (\epsilon+\delta)R_D$. Whence, because this holds for any $e \in D$, $d \in \max^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L)$ for all $\delta > 0$, and therefore, $d \in \bigcap_{\delta>0} \max^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L)$.

Conversely, assume $d \in \bigcap_{\delta>0} \max^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L)$. Consider any $e \in D$. Then, for all $\delta > 0$, there is a $P_\delta \in \mathcal{M}$ such that $E_{P_\delta}(e) - E_{P_\delta}(f) \leq (\epsilon+\delta)R_D$. Hence, for all $n \in \mathbb{N}$, there is a $P_n \in \mathcal{M}$ such that

$$E_{P_n}(e) - E_{P_n}(d) \leq 1/n + \epsilon R_D \tag{9}$$

For any $m \in \mathbb{N}$, consider the following closed subset of $\mathcal{P}(\Omega)$:

$$\mathcal{R}_m = \mathrm{cl}(\{P_n \colon n \geq m\})$$

The collection $\{\mathcal{R}_m \colon m \in \mathbb{N}\}$ satisfies the finite intersection property. By the Banach-Alaoglu-Bourbaki theorem [14, §28.29(UF26)] $\mathcal{P}(\Omega)$ is compact, and hence

$$\mathcal{R} = \cap_{m\in\mathbb{N}}\mathcal{R}_m$$

is non-empty as well [14, §17.2].

Take any $R \in \mathcal{R}$. Since each $P_n \in \mathcal{M}$, it follows that each $\mathcal{R}_m \subseteq \mathrm{cl}(\mathcal{M})$, and hence $R \in \mathrm{cl}(\mathcal{M})$. If we can show that $E_R(e) - E_R(d) \leq \epsilon R_D$, then $d \in \max^\epsilon(\Omega, D, \mathrm{cl}(\mathcal{M}), L)$ is established.

Indeed, fix $m \in \mathbb{N}$. Because $R \in \mathcal{R}_m$, there is a net $(P_{n_\alpha})_{\alpha\in A}$ in $\{P_n \colon n \geq m\}$—so $n_\alpha \geq m$, but $n_\alpha$ is not necessarily an increasing function of $\alpha$—such that $\lim_\alpha E_{P_{n_\alpha}}(f_e - f_d) = E_R(f_e - f_d)$. Hence, for each $\gamma > 0$, there is an $\alpha \in A$ such that $E_R(e) - E_R(d) \leq E_{P_{n_\alpha}}(e) - E_{P_{n_\alpha}}(d) + \gamma$, and therefore by Eq. (9), $E_R(e) - E_R(d) \leq 1/n_\alpha + \epsilon R_D + \gamma$. Because this inequality holds for every $m$ and every $\gamma > 0$, and $n_\alpha \geq m$, it follows that $E_R(e) - E_R(d) \leq \epsilon R_D$.

Let us now prove Eq. (8), under the additional assumption that $D$ is finite. The proof goes along similar lines as the one for Eq. (7).

Assume $d \in \mathrm{opt}^\epsilon(\Omega, D, \mathrm{cl}(\mathcal{M}), L)$. By assumption, there is a $P \in \mathrm{cl}(\mathcal{M})$ such that $E_P(e) - E_P(d) \leq R_D\epsilon$ for every $e \in D$. Because $P \in \mathrm{cl}(\mathcal{M})$, there is a net $(P_\alpha \in \mathcal{M})_{\alpha\in A}$ such that $\lim_\alpha E_{P_\alpha}(f) = E_P(f)$ for all gambles $f$. In particular, there is a net $(P_\alpha \in \mathcal{M})_{\alpha\in A}$ such that $\lim_\alpha E_{P_\alpha}(e) - \lim_\alpha E_{P_\alpha}(d) \leq R_D\epsilon$ for every $e \in D$. So, for every $e \in D$ and $\delta > 0$, there is an $\alpha_{e,\delta} \in A$ such that $E_{P_\alpha}(e) - E_{P_\alpha}(f) \leq (\epsilon+\delta)R_D$ for all $\alpha \geq \alpha_{e,\delta}$. Because $D$ is finite, there is an $\alpha_\delta$ such that $\alpha_\delta \geq \alpha_{e,\delta}$ for all $e \in D$. Hence, for every $\delta > 0$, there is a $\alpha_\delta \in A$ such that $E_{P_{\alpha_\delta}}(e) -$

$E_{P_{\alpha_\delta}}(f) \leq (\epsilon + \delta)R_D$ for every $e \in D$. Whence, because $P_{\alpha_\delta} \in \mathcal{M}$, it follows that $d \in \mathrm{opt}^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L)$ for all $\delta > 0$, and therefore, $d \in \bigcap_{\delta>0} \mathrm{opt}^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L)$.

Conversely, assume $d \in \bigcap_{\delta>0} \mathrm{opt}^{\epsilon+\delta}(\Omega, D, \mathcal{M}, L)$. Then, for all $\delta > 0$, there is a $P_\delta \in \mathcal{M}$ such that $E_{P_\delta}(e) - E_{P_\delta}(f) \leq (\epsilon + \delta)R_D$ for every $e \in D$. Hence, for all $n \in \mathbb{N}$, there is a $P_n \in \mathcal{M}$ such that for every $e \in D$

$$E_{P_n}(e) - E_{P_n}(d) \leq 1/n + \epsilon R_D \tag{10}$$

Now choose any $R$ in

$$\mathcal{R} = \cap_{m\in\mathbb{N}} \mathrm{cl}(\{P_n \colon n \geq m\})$$

Similarly as before, it can be established that $\mathcal{R}$ is non-empty and that $R \in \mathrm{cl}(\mathcal{M})$. If we can show that $E_R(e) - E_R(d) \leq \epsilon R_D$ for all $e \in D$, then $d$ indeed belongs to $\mathrm{opt}^\epsilon(\Omega, D, \mathrm{cl}(\mathcal{M}), L)$ and the desired result is established.

Indeed, because $R \in \mathrm{cl}(\{P_n \colon n \geq m\})$, for every $e \in D$, there is a net $(P_{n_{\alpha,e}})_{\alpha\in A}$ in $\{P_n \colon n \geq m\}$—so $n_{\alpha,e} \geq m$—such that $\lim_\alpha E_{P_{n_{\alpha,e}}}(f_e - f_d) = E_R(f_e - f_d)$. Hence, for every $e \in D$ and every $\gamma > 0$, there is an $\alpha \in A$ such that $E_R(e) - E_R(d) \leq E_{P_{n_{\alpha,e}}}(e) - E_{P_{n_{\alpha,e}}}(d) + \gamma$, and therefore by Eq. (10), $E_R(e) - E_R(d) \leq 1/n_{\alpha,e} + \epsilon R_D + \gamma$. Because this inequality holds for every $m$ and every $\gamma > 0$, and $n_{\alpha,e} \geq m$, it follows that $E_R(e) - E_R(d) \leq \epsilon R_D$ for every $e \in D$.

In particular, assuming $R_D > 0$, if for any $\delta > \epsilon > 0$

$$\max^\epsilon(\Omega, D, \mathcal{M}, L) = \max^\delta(\Omega, D, \mathcal{M}, L)$$

then

$$\max^\epsilon(\Omega, D, \mathcal{M}, L) = \max^\epsilon(\Omega, D, \mathrm{cl}(\mathcal{M}), L)$$

A similar result holds for the $\mathrm{opt}^\epsilon$ operator for finite $D$.

As a special case, Lemma 9 implies an interesting connection between maximality and $\epsilon$-maximality:

**Corollary 10** *Assume that $R_D > 0$. For any decision problem $(\Omega, D, \mathcal{M}, L)$, the following equality holds:*

$$\max(\Omega, D, \mathrm{cl}(\mathcal{M}), L) = \bigcap_{\epsilon>0} \max^\epsilon(\Omega, D, \mathcal{M}, L)$$

Again, a similar result holds for optimality and $\epsilon$-optimality, in case $D$ is finite.

In the following theorem, assume that $0 < \epsilon < 1/2$.

**Theorem 11** *Consider two decision problems* $(\Omega, D, \mathcal{M}, L)$ *and* $(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L})$. *Assume that* $R_D > 0$. *If* $L \sim_\epsilon \hat{L}$ *and* $\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M}))) \sim_\delta \hat{\mathcal{M}}$ *then, for any* $\gamma \geq 0$,

$$\max{}^\gamma(\Omega, D, \mathcal{M}, L) \subseteq \bigcap_{\eta>0} \max{}^{\eta+\frac{\gamma}{1-2\epsilon}+2(\frac{\epsilon}{1-2\epsilon}+\delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) \tag{11}$$

$$\max{}^\gamma(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) \subseteq \bigcap_{\eta>0} \max{}^{\eta+\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathcal{M}, L) \tag{12}$$

**PROOF.** First, note that

$$\begin{aligned}\max{}^\gamma(\Omega, D, \mathcal{M}, L) &= \max{}^\gamma(\Omega, D, \mathrm{co}(\mathcal{M}), L) \\ &\subseteq \max{}^\gamma(\Omega, D, \mathrm{cl}(\mathrm{co}(\mathcal{M})), L)\end{aligned}$$

and by convexity of $\mathrm{cl}(\mathrm{co}(\mathcal{M}))$ [14, §26.23] and the Krein-Milman theorem [15, p. 74], the closed convex hull of $\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M})))$ is $\mathrm{cl}(\mathrm{co}(\mathcal{M}))$, so

$$= \max{}^\gamma(\Omega, D, \mathrm{cl}(\mathrm{co}(\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M}))))), L)$$

and now by Corollary 10,

$$\begin{aligned}&= \cap_{\eta>0} \max{}^{\gamma+\eta}(\Omega, D, \mathrm{co}(\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M})))), L) \\ &= \cap_{\eta>0} \max{}^{\gamma+\eta}(\Omega, D, \mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M}))), L)\end{aligned}$$

Now apply the same argument as in the proof of Theorem 6 to recover Eq. (11).

To establish Eq. (12), again use the same argument as in the proof of Theorem 6,

$$\begin{aligned}\max{}^\gamma(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) &\subseteq \max{}^{\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M}))), L) \\ &\subseteq \max{}^{\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathrm{cl}(\mathrm{co}(\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M}))))), L)\end{aligned}$$

and again by the Krein-Milman theorem [15, p. 74], the closed convex hull of $\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M})))$ is $\mathrm{cl}(\mathrm{co}(\mathcal{M}))$, so

$$\begin{aligned}&= \max{}^{\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathrm{cl}(\mathrm{co}(\mathcal{M})), L) \\ &= \bigcap_{\eta>0} \max{}^{\eta+\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathrm{co}(\mathcal{M}), L) \\ &= \bigcap_{\eta>0} \max{}^{\eta+\gamma(1+2\epsilon)+2(\epsilon+\delta(1+2\epsilon))}(\Omega, D, \mathcal{M}, L)\end{aligned}$$

Again, if we ignore higher order terms in $\gamma$, $\epsilon$, and $\delta$, then the above theorem says that when moving from the original decision problem to the approximate decision problem, with relative error $\epsilon$ in gambles and relative error $\delta$ in probabilities, the relative error in maximality increases by $2(\epsilon + \delta)$. Hence, for small $\epsilon$ and $\delta$ the following holds, up to a small error: if $L \sim_\epsilon \hat{L}$ and $\mathrm{ext}(\mathrm{cl}(\mathrm{co}(\mathcal{M}))) \sim_\delta \hat{\mathcal{M}}$, then

$$\max(\Omega, D, \mathcal{M}, L) \subseteq \max^{2(\epsilon+\delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L}) \subseteq \max^{4(\epsilon+\delta)}(\Omega, D, \mathcal{M}, L)$$

Again, $\max^{2(\epsilon+\delta)}(\mathcal{A}, D, \hat{\mathcal{M}}, \hat{L})$ seems a logical choice when calculating maximal decisions in practice.

## 6 Conclusion and Remarks

With this paper, I hope to have consolidated at least part of our every day intuition when approximating decision problems involving sets of probabilities, for instance when those problems have to be solved by computer.

One result is quite depressing: Lemma 2 and Lemma 3 seem to tell us that except in the simplest cases, any approximation will need too many resources to be of any practical value, as demonstrated by Table 1 and Table 2.

Fortunately, not all is lost. If we resort to pairwise comparison, we may restrict ourselves to the extreme points of the closure of the convex hull of the credal set, which can be *much* smaller than the original credal set. Closing the credal set only has an arbitrary small effect on maximality, and in part for this reason, it turns out that approximating extreme points suffices when restricting to pairwise preference.

I wish to emphasise that the bounds on the cardinalities of the approximating partition and the approximating credal set are only upper bounds under very weak assumptions. These bounds are only attained in extreme situations. In many cases the credal set and the loss function have additional structure which may allow for much lower upper bounds.

In case the problem has sufficient structure, an alternative approach is to develop algorithms which do not need to traverse the complete credal set (or an approximation thereof) to compute the optimal solution. The imprecise Dirichlet model has already been given considerable attention in this direction [16].

Obermeier and Augustin [17] have described a method to approximate decision problems by applying Luceños' adaptive discretisation method to either all elements of the credal set (so the partition varies with the distribution), or on a reference distribution of that set. This type of approximation aims to preserve the first $r$ moments of a distribution. Although precise convergence results and bounds on the precision of this approximation have not yet been proven, examples have shown that this method can yield good results in practice.

Finally, another approach could consist of sampling elements from the credal set, for instance through Monte-Carlo techniques, and solve a classical decision problem for each of these elements. If the sample $s$ from $\hat{\mathcal{M}}$ is large enough, then—since $\bigcup_{P \in s} \operatorname{opt}(\mathcal{A}, D, P, L) = \operatorname{opt}(\mathcal{A}, D, s, L)$—hopefully

$$\operatorname{opt}(\mathcal{A}, D, \mathcal{M}, L) = \bigcup_{P \in s} \operatorname{opt}(\mathcal{A}, D, P, L)$$

The question how large a sample we need to ensure convergence is definitely worth further investigation.


### Acknowledgements

I am grateful to Teddy Seidenfeld for the many helpful discussions on issues related to this paper, and also for encouraging me to extend my view on approximations to choice functions. I thank Max Jensen for his help in characterising the discretisation of the simplex in $\mathbb{R}^n$, presented in the appendix. I also thank all three referees for their constructive comments and useful suggestions which have improved the presentation of this paper. The research reported in this paper has been supported in part by the Belgian American Educational Foundation.


## A Discretisation Of The Standard Simplex In $\mathbb{R}^n$

In this appendix a simple discretisation of $\Delta^n$, the standard simplex in $\mathbb{R}^n$, is studied—these results are not new and are in fact related to well known notions from combinatorics, in particular multisets [18]. The standard simplex $\Delta^n$ is defined as

$$\Delta^n = \{\underline{x} \in \mathbb{R}^n \colon \underline{x} \geq 0,\ |\underline{x}|_1 = 1\}$$

where $|\cdot|_1$ denotes the 1-norm, i.e. $|\underline{x}|_1 = \sum_{i=1}^n |x_i|$.

For any non-zero natural number $N$, let $\Delta^n_N$ denote the following finite subset of $\Delta^n$:

$$\Delta^n_N = \{\underline{m}/N \colon \underline{m} \in \mathbb{N}^n,\ |\underline{m}|_1 = N\}$$

(above, $\mathbb{N}$ is the set of natural numbers including 0).

**Lemma 12** *The cardinality of* $\Delta_N^n$ *is* $\binom{N+n-1}{N}$.

**PROOF.** There is an obvious one-to-one and onto correspondence between $\Delta_N^n$ and all multisets of cardinality $N$ with elements taken from $\{1,\ldots,n\}$—for any $\underline{m}/N \in \Delta_N^n$, interpret $m_i$ as the multiplicity of $i$. The number of all such multisets is precisely $\binom{N+n-1}{N}$ (see Stanley [18]).

**Lemma 13** *For every* $\underline{x}$ *in* $\Delta^n$ *there is a* $\underline{y}$ *in* $\Delta_N^n$ *such that*

$$|\underline{x}-\underline{y}|_1 < n/N$$

**PROOF.** For each $i \in \{1,\ldots,n\}$, let $m_i$ be the unique natural number such that $x_i \in [m_i/N, (m_i+1)/N)$, or equivalently, let $m_i$ be the largest natural number such that $m_i/N \le x_i$. Define $M = \sum_{i=1}^n m_i$. Then, $M \le N < M+n$ since $M/N = |\underline{m}/N|_1 \le |\underline{x}|_1 = 1$ and $(M+n)/N = |(\underline{m}+1)/N|_1 > |\underline{x}|_1 = 1$. Define

$$e_i = \begin{cases} 1 & \text{if } i \in \{1,\ldots,N-M\} \\ 0 & \text{if } i \in \{N-M+1,\ldots,n\} \end{cases}$$

and let $\underline{y} = (\underline{m}+\underline{e})/N$. Note that $\underline{y} \in \Delta_N^n$ because $|\underline{y}|_1 = |\underline{m}+\underline{e}|_1/N = (M + (N-M))/N = 1$. Finally,

$$|\underline{x}-\underline{y}|_1 = \sum_{i=1}^{N-M} |x_i - \tfrac{m_i+1}{N}| + \sum_{i=N-M+1}^{n} |x_i - \tfrac{m_i}{N}| < n/N$$

as $|x_i - \frac{m_i+1}{N}| \le 1/N$ and $|x_i - \frac{m_i}{N}| < 1/N$.

## References


[1] P. C. Fishburn, A. H. Murphy, H. H. Isaacs, Sensitivity of decisions to probability estimation errors: A reexamination, Operations Research 16 (2) (1968) 254–267.

[2] D. A. Pierce, J. L. Folks, Sensitivity of Bayes procedures to the prior distribution, Operations Research 17 (2) (1969) 344–350.

[3] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[4] K. Bhaskara Rao, M. Bhaskara Rao, Theory of Charges, a Study of Finitely Additive Measures, Academic Press, London, 1983.

[5] J. von Neumann, O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, 1944.

[6] F. J. Anscombe, R. J. Aumann, A definition of subjective probability, Annals of Mathematical Statistics 34 (1) (1963) 199–205.

[7] B. de Finetti, Theory of Probability: A Critical Introductory Treatment, Wiley, New York, 1974–5, two volumes.

[8] G. de Cooman, M. C. M. Troffaes, Dynamic programming for deterministic discrete-time systems with uncertain gain, International Journal of Approximate Reasoning 39 (2–3) (2005) 257–278.

[9] M. C. M. Troffaes, Decision making under uncertainty using imprecise probabilities, International Journal of Approximate Reasoning 45 (2007) 17–29.

[10] I. Levi, The Enterprise of Knowledge. An Essay on Knowledge, Credal Probability, and Chance, MIT Press, Cambridge, 1980.

[11] T. Seidenfeld, M. J. Schervish, J. B. Kadane, A representation of partially ordered preferences, The Annals of Statistics 23 (1995) 2168–2217.

[12] J. B. Kadane, M. J. Schervish, T. Seidenfeld, A Rubinesque theory of decision, in: A festschrift for Herman Rubin, Vol. 45 of IMS Lecture Notes – Monograph Series, Inst. Math. Statist., Beachwood, Ohio, 2004, pp. 45–55.

[13] T. Seidenfeld, M. Schervish, J. Kadane, Coherent choice functions under uncertainty, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), ISIPTA'07: Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, Charles University, Faculty of Mathematics and Physics, Prague, 2007, pp. 385–394.

[14] E. Schechter, Handbook of Analysis and Its Foundations, Academic Press, San Diego, 1997.

[15] R. Holmes, Geometric Functional Analysis and Its Applications, Springer, New York, 1975.

[16] M. Hutter, Robust estimators under the imprecise dirichlet model, in: J.-M. Bernard, T. Seidenfeld, M. Zaffalon (Eds.), ISIPTA '03 – Proceedings of the Third International Symposium on Imprecise Probabilities and Their Applications, Carleton Scientific, 2003, pp. 274–289.

[17] M. Obermeier, T. Augustin, Luceños' discretization method and its application in decision making under ambiguity, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), ISIPTA'07: Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, Charles University, Faculty of Mathematics and Physics, Prague, 2007, pp. 327–336.

[18] R. P. Stanley, Enumerative Combinatorics, Cambridge University Press, 1997.