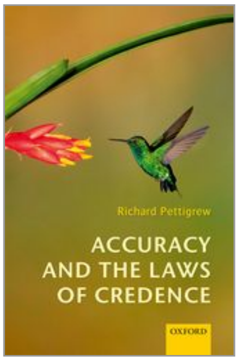

Accuracy and the Laws of Credence
Richard Pettigrew

Print publication date: 2016
Print ISBN-13: 9780198732716
Published to Oxford Scholarship Online: May 2016
DOI: 10.1093/acprof:oso/9780198732716.001.0001

Hurwicz, regret, and $\mathscr{C}$-maximin

Richard Pettigrew

DOI:10.1093/acprof:oso/9780198732716.003.0014

Abstract and Keywords

This chapter considers a number of different risk-sensitive decision principles, such as the Hurwicz criterion, Minimax regret, and C-maximin. The chapter describes the principles of credal rationality that they entail. It considers how the Principle of Indifference relates to the Principal Principle.

Keywords: Risk aversion, risk, Hurwicz criterion, Minimax regret, C-Maximin

In the previous chapter, we considered the credal principles that follow from Maximin and Maximax in conjunction with Veritism and Brier Alethic Accuracy. But these are just two of the many decision-theoretic principles that incorporate sensitivity to considerations of risk. In this chapter, we consider a number of other such principles and explore their consequences for credences.[1]

13.1 The Hurwicz criterion

In the previous chapter, we considered Williams James' epistemic conservative and epistemic radical. We described the way that adherents to extreme versions of both positions would go about setting their initial credences: the extreme conservative would adopt Maximin and, as a result, they would take the uniform distribution as their credence function as recommended by the Principle of Indifference; the extreme radical would adopt Maximax and, in consequence, they would set their credences to match the omniscient credences at some world. However, we also noted that there are intermediate positions. There are those who are epistemic conservatives, primarily pursuing the goal *Shun error*!, but who also recognize the importance of the other goal, *Believe truth*!, and accord it at least some weight in their epistemic decisions. Likewise, there are epistemic radicals, who will accord some weight to the goal *Shun error*!, even though their primary aim is *Believe truth*!. And of course there are those who are epistemic neutrals, perfectly poised between the two extremes, who accord equal weight to both goals.

In decision-theoretic terms, these moderate epistemic conservatives are risk-averse, trying hardest to minimize the badness of the worst-case scenario; but they are not maximally risk-averse, since they also pay some attention to the badness of the best-case scenario and try to minimize that. Inversely, epistemic radicals who are not maximally risk-seeking, pay most attention to the best-case, but also pay some attention to the worst-case. Thus, each position on the spectrum from extreme epistemic conservativism to extreme epistemic radicalism corresponds to an instance (p.169) of the following decision-theoretic principle schema, which is parametrized by a number $0 \leq \lambda \leq 1$ (Hurwicz, 1951, 1952).

Hurwicz$_\lambda$ criterion Suppose $\mathscr{O}$ is a set of options, $\mathscr{W}$ is the set of possible worlds, and $\mathfrak{U}$ is a utility function. If $o \in \mathscr{O}$, let

$$H^{\mathrm{U}}_{\lambda}(o) := \lambda \max_{w \in \mathrm{W_F}} \mathrm{U}(o, w) + (1-\lambda) \max_{w \in \mathrm{W_F}} \mathrm{U}(o, w)$$

Equivalently,

$$-H^{\mathrm{U}}_{\lambda}(o) := \lambda \min_{w \in \mathrm{W_F}} -\mathrm{U}(o, w) + (1-\lambda) \min_{w \in \mathrm{W_F}} -\mathrm{U}(o, w)$$

Suppose $o, o' \in \mathscr{O}$. Then, if

(i)

$$H^{\mathrm{U}}_{\lambda}(o) < H^{\mathrm{U}}_{\lambda}(o')$$

and

(ii) there is no o'' in ?? such that

$$H^{\mathrm{U}}_{\lambda}(o') < H^{\mathrm{U}}_{\lambda}(o'')$$

then

(iii) o is irrational for an agent at the beginning of her epistemic life with utility function $\mathfrak{U}$.

Thus, according to this principle, an option is irrational if it doesn't maximize H^{U}_{λ}, providing there is an option that does maximize that quantity.

Again, since an inaccuracy measure is a disutility function, it will be clearer if we focus on the formulation of this principle on which it exhorts agents to minimize $-H^{\mathrm{U}}_{\lambda}$, rather than maximize H^{U}_{λ}. Thus, if $\mathfrak{I}$ is an inaccuracy measure, Hurwicz$_{\lambda}$ exhorts you to choose a credence function that minimizes

$$-H^{\mathfrak{I}}_{\lambda}(c) = \lambda \min_{w \in \mathcal{W}_{\mathcal{F}}} \mathfrak{I}(c, w) + (1-\lambda) \min_{w \in \mathcal{W}_{\mathcal{F}}} \mathfrak{I}(c, w)$$

The principle codifies greater risk-aversion for lower values of λ: as λ decreases, the worst-case scenario—the scenario in which the option achieves its minimum utility— is given greater weight. Indeed, Hurwicz$_0$ is equivalent to Maximin; and Hurwicz$_1$ is equivalent to Maximax. Thus, epistemic conservatives will apply Hurwicz$_{\lambda}$—in conjunction with Veritism and Brier Alethic Accuracy—for $k < \frac{1}{2}$, while epistemic radicals will apply it for $k > \frac{1}{2}$. We know the consequences when $\lambda = 1$ (only omniscient credence functions are not irrational) and $\lambda = 0$ (only the uniform distribution is not irrational). But what if $0 < \lambda < 1$? The answer is given by the following theorem, which proves a conjecture due to Jason Konek.[2]

(p.170) **Theorem 13.1.1** *Suppose $\mathfrak{D}$ is an additive Bregman divergence and $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$. So, $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure. Suppose $\mathcal{F}$ is a finite, rank-complete set of propositions and $\mathcal{W}_{\mathcal{F}} = \{w_1, \ldots, w_n\}$. There are two cases*:

(i) *Suppose $\lambda \geq \frac{1}{n}$. For each $i = 1, \ldots, n$, define the probabilistic credence function c^{λ}_i on $\mathcal{F}$ as follows*:

$$c^{\lambda}_i(w_j) := \begin{cases} \lambda & \text{if } j = i \\ \frac{1-\lambda}{n-1} & \text{if } j \neq i \end{cases}$$

If $c \neq c^{\lambda}_i$ for all $i = 1, \ldots, n$, then

$$-H^{\mathfrak{I}}_{\lambda}(c^{\lambda}_i) < -H^{\mathfrak{I}}_{\lambda}(c)$$

for all $i = 1, \ldots, n$.

(ii) *Suppose $\lambda \leq \frac{1}{n}$. Then, as above, define the uniform distribution $c^{\dagger}$ on $\mathcal{F}$ to be the probabilistic credence function such that*:

$$c^{\dagger}(w_j) = \frac{1}{n}$$

Then, if $c \neq c^{\dagger}$

$$-H^{\mathfrak{I}}_{\lambda}(c^{\dagger}) < -H^{\mathfrak{I}}_{\lambda}(c)$$

Thus, for $\lambda \geq \frac{1}{n}$, the credence function c^{λ}_i is a mixture of the omniscient credence function of world w_i —that is, v_{w_i}. —and the uniform distribution—that is, $c^{\dagger}$.[3] The greater λ is—that is, the greater weight is given to the best-case scenario—the more heavily the mixture is weighted towards v_{w_i}. The lower λ is—that is, the greater weight is given to the worst-case scenario—the more weight is assigned to $c^\dagger$. At $\lambda = \frac{1}{n}$, it reaches $c^\dagger$. Figure 13.1 illustrates the credence functions c_i^λ that this principle permits for the case in which the rank-complete set $\mathscr{F}$ is the three-cell partition $\{X_1, X_2, X_3\}$ and $\lambda = 0, \frac{1}{2}, \frac{3}{4}, 1$. And the following table gives the credences that each c_i^λ assigns to each world w_j:

	w_1	w_2	$\dots$	w_n
c_1^λ	λ	$\frac{1-\lambda}{n-1}$	$\dots$	$\frac{1-\lambda}{n-1}$
c_2^λ	$\frac{1-\lambda}{n-1}$	λ	$\dots$	$\frac{1-\lambda}{n-1}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
c_n^λ	$\frac{1-\lambda}{n-1}$	$\frac{1-\lambda}{n-1}$	$\dots$	λ

Thus, we have the following credal principle, which follows from Hurwicz$_\lambda$, along with Veritism and Brier Alethic Accuracy: (p.171)

Risk Spectrum$_\lambda$ Suppose $\mathscr{F}$ is a finite, rank-complete set of propositions with $\mathscr{W}_\mathscr{F} = \{w_1, \dots, w_n\}$.

(I) Suppose $\lambda \geq \frac{1}{n}$. If an agent has an initial credence function c_0 defined on $\mathscr{F}$, then rationality requires that $c_0 = c_i^\lambda$ for some $i = 1, \dots, n$.

(II) Suppose $\lambda \leq \frac{1}{n}$. If an agent has an initial credence function c_0 defined on $\mathscr{F}$, then rationality requires that $c_0 = c^\dagger$.

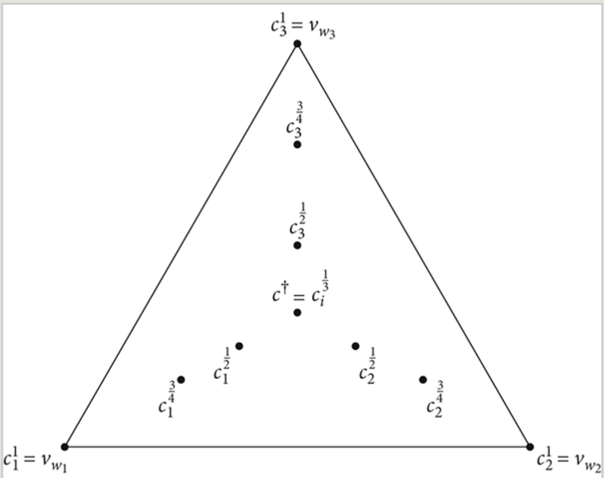

Figure 13.1 This shows the permissible credence functions relative to Hurwicz$_\lambda$ for $\lambda = 0, \frac{1}{2}, \frac{3}{4}, 1$.

(I$_{RS_\lambda}$) **Veritism**

(II$_{RS_\lambda}$) **Brier Alethic Accuracy**

(III$_{RS_\lambda}$) **Hurwicz$_\lambda$**

(IV$_{RS_\lambda}$) **Theorem 13.1.1**

Therefore,

(V$_{RS_\lambda}$) **Probabilism + Risk Spectrum$_\lambda$**

As with all our arguments so far, this one will still go through if we replace the second premise with the claim that inaccuracy must be measured by inaccuracy measure $\mathfrak{I}$, providing $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure. And, as with our arguments in the previous chapter, it will also go through if we replace the second premise with Supervaluationism about Inaccuracy Measures or Epistemicism about Inaccuracy Measures.

13.2 Risking regret

Risk-sensitive principles in decision theory—such as Maximin, Maximax, and the Hurwicz$_\lambda$ criterion—have been studied in decision theory because real agents often (p.172) exhibit behaviour that, it is claimed, can be explained only by saying that they are choosing in line with such a principle. The Allais paradox is the most famous example (Allais, 1953). However, some behaviour is best explained not by saying that the agent is risk-averse in the sense that she seeks to make her worst-case *utility* as *high* as possible, but rather by saying that she is regret-averse, in the sense that she seeks to make her worst-case *regret* as *low* as possible (Savage, 1961). For this purpose, the regret that attaches to an option at a world is the difference between the utility of that option at that world and the highest utility of any option at that world: you regret a choice to the extent that an alternative choice would have gained you something better. Thus, if $\mathscr{O}$ is a set of options, $\mathscr{W}$ is the set of possible worlds, and $\mathfrak{U}$ is a utility function, we define the regret that attaches to o at w to be:

$$R_{\mathrm{U}}(o, w) := \left[\max_{o' \in \mathrm{O}} \mathrm{U}(o', w)\right] - \mathrm{U}(o, w)$$

For instance, suppose I must choose whether or not to buy the lottery ticket that is in my hand. If I do and it loses, I am down £1; if I do and it wins, I am up £13,999,999. If I don’t, I am up nothing and down nothing. Thus, the decision problem is represented by the following table:

	Win	*Lose*
Play	13,999,999	−1
Don’t Play	0	0

Maximin demands *Don’t Play*, since the worst-case scenario for that option is better than the worst-case scenario for the alternative, *Play*. However, if instead of maximizing worst-case utility, I wish to minimize worst-case regret, I will choose *Play*, since its worst-case regret occurs in *Lose* and has value £1, whereas the worst-case regret for *Don’t Play* occurs in *Win* and has value £13,999,999.

Here’s the decision-theoretic principle we’ve been considering.

Minimax Regret Suppose $\mathscr{O}$ is the set of options, $\mathscr{W}$ is the set of possible worlds, and $\mathfrak{U}$ is a utility function. Suppose o, o' in $\mathscr{O}$. Then, if

(i)

$$\max_{w \in W_F} R_U(o', w) < \max_{w \in W_F} R_U(o, w)$$

and

(ii) there is no o'' in $\mathscr{O}$ such that

$$\max_{w \in W_F} R_U(o'', w) < \max_{w \in W_F} R_U(o', w)$$

then

(iii) o is irrational, for an agent at the beginning of her epistemic life.

(p.173) Note that

$$R_U(o, w) = -U(o, w) - \min_{o' \in O} - U(o', w)$$

As usual, this is the easiest formulation to keep in mind when applying Minimax Regret to the case of credence functions and inaccuracy measures, since the latter are disutility functions. Thus, if the options are credence functions and the utility function is the negative $-\mathfrak{I}$ of an inaccuracy measure $-\mathfrak{I}$, Minimax Regret asks us to pick c in a way that minimizes

$$\max_{w \in W_F} R_{-I}(c, w) = \max_{w \in W_F} \left[I(c, w) - \min_{c'} I(c', w) \right]$$

However, for each additive and continuous strictly proper inaccuracy measure $\mathfrak{I}$, and for every world w,

$$\min_{c'} I(c', w) = 0$$

and it attains that minimum at $c = v_w$. Thus,

$$\max_{w \in W_F} R_{-I}(c, w) = \max_{w \in W_F} I(c, w)$$

Thus, minimizing worst-case regret is tantamount to minimizing worst-case inaccuracy. Thus, Minimax Regret entails the same requirements of rationality as Maximin: both entail PoI. Thus, we have:

($\mathrm{I}^{\star}_{PoI}$) **Veritism**

($\mathrm{II}^{\star}_{PoI}$) **Brier Alethic Accuracy**

($\mathrm{III}^{\star}_{PoI}$) **Minimax Regret**

($\mathrm{IV}^{\star}_{PoI}$) **Theorems 12.4.1** and **I.B.2**

Therefore,

($\mathrm{V}^{\star}_{PoI}$) **Probabilism + PoI**

And similarly if we replace Brier Alethic Accuracy with the claim that the only legitimate measure of inaccuracy is $\mathfrak{I}$, where $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure.

13.3 Risk and chances

The credal principles that we have so far sought to justify in this part of the book are all very well if the agent whose rationality we are assessing does not have any opinions about objective chances. But, if they do, then problems can arise combining the risk- sensitive principles—such as PoI and Risk Spectrum$_\lambda$—with the chance-credence principles explored in Part II—such as the Temporal Principle.

(p.174) For instance, suppose Anna only has opinions about the outcomes of the toss of a biased coin, as well as opinions about the coin's bias, where the bias gives it either a 60% or a 90% chance of landing heads on a given toss. Thus, there are just two possible chance functions, ch_1 and ch_2:

$$ch_1(Heads) = 0.6 \quad ch_1(\overline{Heads}) = 0.4$$

and

$$ch_2(Heads) = 0.9 \quad ch_2(\overline{Heads}) = 0.1$$

And there are two current chance hypotheses corresponding to them: T_{ch_1} and T_{ch_2}. Thus, Anna's credence function is defined on

$$\mathcal{F} = \{Heads, \overline{Heads}, T_{ch_1}, T_{ch_2}\}$$

And we will suppose that both possible chance functions are immodest: that is, $ch_i(T_{ch_i}) = 1$.

Now, if c is Anna's credence function, the Temporal Principle demands

$$ch_1(Heads) = 0.6 \leq c(Heads) \leq 0.9 = ch_2(Heads)$$

since T_{ch_1} and T_{ch_2} form a partition. On the other hand, the Principle of Indifference demands

$$c(Heads) = c(\overline{Heads}) = 0.5$$

Thus, Anna cannot satisfy both principles.

In fact, this should not surprise us. The Temporal Principle is derived from a chance dominance principle that demands of an agent that she choose in line with the dictates of the objective chances whenever they speak with one voice. The Principle of Indifference, on the other hand, is derived from an extremely risk-averse principle, which demands that an agent choose by minimizing the badness of the worst-case scenario. It is easy to see that those principles will sometimes make conflicting demands. If the possible objective chances all find the worst-case scenario sufficiently unlikely to occur, it will not be afforded the weight demanded of it by Maximin or some other principle of extreme risk aversion.

Consider, for instance, two lotteries between which Josh must choose. The first costs £l,000 to enter and pays out £1,000,000 with either a $\frac{1}{100}$ chance or a $\frac{1}{1{,}000}$ chance. The second costs £l to enter and pays out £l0,000 with a $\frac{1}{1{,}000{,}000}$ chance or a $\frac{1}{2{,}000{,}000}$ chance. Maximin looks only to the worst-case scenario for each option, which is obviously the situation in which they lose: to minimize the badness in this situation, it demands that Josh choose the second lottery. On the other hand, the different possible chance functions—of which there are four—all speak with one voice in expecting the first lottery to do better than the second for Josh. So a chance dominance principle will (p.175) demand that he choose that. Again, a risk averse principle and a chance dominance principle will make conflicting demands.

How, then, are we to find a risk-averse decision-theoretic principle that is compatible with the versions of chance dominance used to justify the Temporal Principle and other chance-credence principles in Part II? The idea is this: Instead of choosing an option that maximizes the *utility* obtained in the world at which it has lowest *utility*, we choose an option that maximizes the *expected utility* as evaluated by the chance function that assigns it lowest *expected utility*. This principle is known as $\mathscr{C}$-maximin, where $\mathscr{C}$ is the set of possible chance functions.

> $\mathscr{C}$-**Maximin** Suppose $\mathscr{O}$ is a set of options, $\mathscr{W}$ is the set of possible worlds, $\mathscr{C}$ is the set of possible chance functions, and $\mathfrak{U}$ is a utility function. Suppose *o, o′* in $\mathscr{O}$ Then, if
>
> (i)
>
> $$\min_{ch \in C} \mathrm{Exp}_{U}(o|ch) < \min_{ch \in C} \mathrm{Exp}_{U}(o'|ch)$$
>
> and
>
> (ii) there is no *o″* in $\mathscr{O}$ such that
>
> $$\min_{ch \in C} \mathrm{Exp}_{U}(o'|ch) < \min_{ch \in C} \mathrm{Exp}_{U}(o''|ch)$$
>
> then
>
> (ii) *o* is irrational, for an agent at the beginning of her epistemic life.

Now, $\mathscr{C}$-maximin is clearly a risk-averse principle, since it pays attention to a worst- case scenario, albeit a different worst-case scenario from that to which Maximin tells us to pay attention. Choosing a credence function in line with $\mathscr{C}$-maximin requires the agent to minimize the maximal *expected inaccuracy* she might incur; it does not require her to minimize the maximal *inaccuracy* she might incur. Furthermore, $\mathscr{C}$-maximin is compatible with the various versions of chance dominance to which we appealed in Part II.

Which credal principles follow from $\mathscr{C}$-maximin together with an account of inaccuracy measures? The following theorem gives conditions under which it entails the same principle as Maxmin itself, namely, PoI:

Theorem 13.3.1 *Suppose $\mathscr{C}$ is the set of possible chance functions. Suppose $\mathfrak{D}$ is an additive Bregman divergence and $\mathfrak{I}(c, w) = \mathfrak{D}(v_w, c)$. So, $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure. Then, if*

(i) *$c^\dagger$ is in $\mathscr{C}^+$; and*
(ii) *for all ch, ch' in $\mathscr{C}$, $\mathfrak{D}(ch, c^\dagger) = \mathfrak{D}(ch', c^\dagger)$*

then, for all $c \neq c^\dagger$,

$$\max_{ch \in C} \mathrm{Exp}_{\mathfrak{I}}(c^\dagger | ch) < \max_{ch \in C} \mathrm{Exp}_{\mathfrak{I}}(c | ch)$$

(p.176) Now, these conditions do not always hold. For instance, they do not hold in our toy example of Anna and the biased coin. In that case, $c^\dagger$ lies closer to ch_1 than it lies to ch_2. So Theorem 13.3.1 does not apply. Indeed, in this situation, $c^\dagger$ is ruled irrational by $\mathscr{C}$-maximin since it is chance dominated—there is an alternative credence function that both ch_1 and ch_2 agree is expected to be more accurate than $c^\dagger$ is expected to be. Which credence functions, then, minimize the maximum objective expected inaccuracy in that situation? It turns out that it depends on which inaccuracy measure we use.

Let's see how that works. Note that a credence function over $\{T_{ch_1}, T_{ch_2}, Heads, \overline{Heads}\}$ that satisfies the Temporal Principle is determined only by its credence in T_{ch_1} (or by its credence in T_{ch_2})). After all:

- $c(T_{ch_2}) = 1 - c(T_{ch_1})$
- $c(Heads) = c(T_{ch_1})ch_1(Heads) + c(T_{ch_2})ch_2(Heads)$
- $c(\overline{Heads}) = c(T_{ch_1})ch_1(\overline{Heads}) + c(T_{ch_2})ch_2(\overline{Heads})$

Now we have:

Proposition 13.3.2

(1) *Let $\mathfrak{L}$ be the logarithmic inaccuracy measure.*[4] *then*

$$\max_{ch \in C} \mathrm{Exp}_{\mathrm{L}}(c | ch)$$

is minimized for c such that $c(T_{ch_1}) \approx 0.5861$ and thus c(Heads) ≈ 0.72417.

(2) *Let* $\mathfrak{LA}$ *be the additive logarithmic inaccuracy measure.*[5] *then*

$$\max_{ch \in C} \mathrm{Exp}_{LA}(c|ch)$$

is minimized for c such that $c(T_{ch_1}) \approx 0.5733$ *and thus c*(Heads) ≈ 0.72801.

(3) *Let* $\mathfrak{B}$ *be the Brier score.*[6] *then*

$$\max_{ch \in C} \mathrm{Exp}_{B}(c|ch)$$

is minimized for c such that $c(T_{ch_1}) = 0.5688$ *and thus c*(Heads) ≈ 0.72936.

The following is an intriguing fact: The credence function in Proposition l3.3.2(1) that minimizes worst-case expected inaccuracy relative to $\mathfrak{L}$ also maximizes Shannon entropy amongst credence functions that lie in $\mathscr{C}^+$, the convex hull of $\mathscr{C}$. Recall: the Shannon entropy of a probabilistic credence function c is defined as follows:

$$H(c) := -\sum_{w \in W_F} c(w) \ln c(w)$$

(p.177) As noted above, $H(c) = \mathrm{Exp}_{\mathfrak{L}}(c|c)$. That is, the Shannon entropy of a credence function is its expected inaccuracy by its own lights when inaccuracy is measured by the so-called logarithmic inaccuracy measure $\mathfrak{L}$.

Now, noting that alternative (and equivalent) definition of Shannon entropy, we can now define a measure of entropy corresponding to any inaccuracy measure $\mathfrak{I}$:

$$H_I(c) := \mathrm{Exp}_I(c \mid c)$$

Now, it turns out that the credence function in Proposition 13.3.2(2) that minimizes worst-case expected inaccuracy relative to the additive logarithmic inaccuracy measure is also the credence function in $\mathscr{C}^+$ that maximizes $H_{\mathfrak{LA}}$, the entropy corresponding to the additive logarithmic inaccuracy measure $\mathfrak{LA}$. And similarly the credence function in Proposition 13.3.2(3) maximizes $H_{\mathfrak{B}}(c)$. Indeed, these are instances of an extremely general phenomenon that was first noticed by Topsøe (1979) in the case of $H_{\mathfrak{L}}$ and $\mathfrak{L}$: for many sets $\mathscr{C}$ of possible chance functions, the same credence function maximizes $H_{\mathfrak{L}}$ on $\mathscr{C}$ and minimizes worst-case expected inaccuracy by the lights of chance functions in $\mathscr{C}$, when inaccuracy is measured by $\mathfrak{L}$. Grünwald & Dawid (2004) extended these results to many other inaccuracy measures, including all those we countenance here. I refer the interested reader particularly to the result of Section 6 of that paper for the relevant results.

(p.178) Appendix IV: The mathematical results

IV.A Proof of Theorem 13.1.1

Theorem 13.1.1 Suppose $\mathfrak{D}$ is an additive Bregman divergence and $\mathfrak{I}(c, w) = \mathfrak{D}\{v_w, c)$. So, $\mathfrak{I}$ is an additive and continuous strictly proper inaccuracy measure. Suppose $\mathscr{F}$ is a finite, rank-complete set of propositions and $\mathscr{W}_{\mathscr{F}} = \{w_1, \ldots, w_n\}$. There are two cases:

(i) Suppose $\lambda \geq \frac{1}{n}$. For each $i = 1, \ldots, n$, define the probabilistic credence function c_i^λ on $\mathscr{F}$ as follows:

$$c_i^\lambda(w_j) := \begin{cases} \lambda & if\ j = k \\ \frac{1-\lambda}{n-1} & if\ j \neq k \end{cases}$$

If $c \neq c_i^\lambda$ for all $i = 1, \ldots, n$, then

$$-H_\lambda^I(c_i^\lambda) < -H_\lambda^I(c)$$

for all $i = 1, \ldots, n$.

(ii) Suppose $\lambda \leq \frac{1}{n}$. Then, as above, define the uniform distribution $c^\dagger$ on $\mathscr{F}$ to be the probabilistic credence function such that:

$$c^\dagger(w_j) = \frac{1}{n}$$

Then, if $c \neq c^\dagger$

$$-H_\lambda^I(c^\dagger) < -H_\lambda^I(c)$$

Before we begin in earnest, we require the following lemma. It says that a probabilistic credence function defined on a finite, rank-complete set of propositions is more accurate at one world than at another iff it assigns more probability to the first than to the second. This shouldn't be surprising.

Lemma IV.A.1 *Suppose $\mathscr{F}$ is a rank-complete set, suppose $\mathscr{W}_{\mathscr{F}} = \{w_1, \ldots, w_n\}$, and suppose c is a probabilistic credence function defined on $\mathscr{F}$. Then, if $c(w_i) \geq c(w_j)$, then* (p.179)

$$I(c, w_i) \leq I(c, w_j)$$

In particular, c has minimal inaccuracy at a world to which it assigns maximal credence.

Proof. First, we note that:

$$I(c, w_i) = \sum_{X \in F} \mathfrak{s}(v_{w_i}(X), c(X))$$
$$= \sum_{\substack{X \in F \\ w_i \in X \\ w_j \in X}} \mathfrak{s}(1, c(X)) + \sum_{\substack{X \in F \\ w_i \in X \\ w_j \notin X}} \mathfrak{s}(1, c(X))$$
$$+ \sum_{\substack{X \in F \\ w_i \notin X \\ w_j \in X}} \mathfrak{s}(0, c(X)) + \sum_{\substack{X \in F \\ w_i \notin X \\ w_j \notin X}} \mathfrak{s}(0, c(X))$$

and

$$I(c, w_j) = \sum_{X \in F} s(v_{w_i}(X), c(X))$$

$$= \sum_{\substack{X \in F \\ w_i \in X \\ w_j \in X}} s(1, c(X)) + \sum_{\substack{X \in F \\ w_i \in X \\ w_j \notin X}} s(0, c(X))$$

$$+ \sum_{\substack{X \in F \\ w_i \notin X \\ w_j \in X}} s(1, c(X)) + \sum_{\substack{X \in F \\ w_i \notin X \\ w_j \notin X}} s(0, c(X))$$

Now, if $\mathscr{F}$ is rank-complete, then

$$\{X \in F : w_i \notin X \ \&\ w_j \in X\} = \{X\overline{w_i} \vee w_j \in F : w_i \in X \ \&\ w_j \notin X\}$$

So:

$$\sum_{\substack{w_i \in X \\ w_j \notin X}} s(1, c(X))$$

$$= \sum_{\substack{w_i \in X \\ w_j \notin X}} s(1, c(X\overline{w_i}) + c(w_i)$$

$$\leq \sum_{\substack{w_i \in X \\ w_j \notin X}} s(1, c(X\overline{w_i}) + c(w_j)) \text{ since } c(w_i) \geq c(w_j)$$

$$= \sum_{\substack{w_i \notin X \\ w_j \in X}} s(1, c(X)) \text{ Since } F \text{ is rank-complete}$$

Moreover,

$$\{X \in F : w_i \in X \ \&\ w_j \notin X\} = \{X\overline{w_i} \vee w_j\, F : w_i \notin X \ \&\ w_j \in X\}$$

(p.180) So:

$$\sum_{\substack{w_i \notin X \\ w_j \in X}} s(0, c(X))$$

$$= \sum_{\substack{w_i \notin X \\ w_j \in X}} s(0, c(X\overline{w_j}) + c(w_j)$$

$$\leq \sum_{\substack{w_i \notin X \\ w_j \in X}} s(0, c(X\overline{w_j}) + c(w_i)) \text{ since } c(w_i) \geq c(w_j)$$

$$= \sum_{\substack{w_i \in X \\ w_j \notin X}} s(0, c(X)) \text{ Since } F \text{ is rank-complete}$$

Thus,

$$I(c, w_i) \leq I(c, w_j)$$

as required. □

IV.A.1 Proof of Theorem 13.1.1(I)

We are now ready to prove Theorem 13.1.1(I). We prove it in two parts. First, we state the following straightforward lemma, which follows because $\mathscr{F}$ is rank-complete.

Lemma IV.A.2 *For all $i, j = 1, \ldots, n$*

$$-H_\lambda^{I}(c_i^\lambda) = -H_\lambda^{I}(c_j^\lambda)$$

Next, we prove the following lemma:

Lemma IV.A.3 *Suppose $c \neq c_j^\lambda$ for all j. And suppose that c assigns maximal credence to world w_i: that is, $c(w_i) \geq c(w_j)$ for all j. Then*

$$-H_\lambda^{I}(c_i^\lambda) < -H_\lambda^{I}(c).$$

Proof of Lemma IV.A.3. First, note that, since $\mathfrak{I}$ is strictly proper, we have the following:

$$\mathrm{Exp}_I\left(c_i^\lambda | c_i^\lambda\right) := \sum_{j=1}^{n} c_i^\lambda(w_j) I(c_i^\lambda, w_j) < \sum_{j=1}^{n} c_i^\lambda(w_j) I(c, w_j) = : \mathrm{Exp}_I(c | c_i^\lambda)$$

Now note that, since $\lambda \geq \frac{1}{n}$, we have $\lambda \geq \frac{1-\lambda}{n-1}$, and thus, by Lemma IV.A.1, $\mathfrak{J}(c_i^\lambda, w)$ is minimized (as a function of w) at w_i and maximized at all other worlds. Thus

$$\begin{aligned}
&\mathrm{Exp}_I(c_i^\lambda | c_i^\lambda) \\
&= \sum_{j=1}^{n} c_i^\lambda(w_j) I(c_i^\lambda, w_j) \\
&= \tfrac{1-\lambda}{n-1} I(c_i^\lambda, w_1) + \ldots + \lambda I(c_i^\lambda, w_i) + \ldots + \tfrac{1-\lambda}{n-1} I(c_i^\lambda, w_n) \\
&= \tfrac{1-\lambda}{n-1} \max_{w \in W} I(c_i^\lambda, w) + \ldots + \lambda \min_{w \in W} I(c_i^\lambda, w) + \ldots + \tfrac{1-\lambda}{n-1} \max_{w \in W} I(c_i^\lambda, w) \\
&= (1-\lambda) \max_{w \in W} I(c_i^\lambda, w) + \lambda \min_{w \in W} I(c_i^\lambda, w) \\
&= -H_\lambda^{I}(c_i^\lambda)
\end{aligned}$$

(p.181) Furthermore, by definition, for all j

$$I(c, w_j) \leq \max_{w \in W} I(c, w)$$

Thus,

$$\begin{aligned}
&\mathrm{Exp}_I(c | c_i^\lambda) \\
&= \sum_{j=1}^{n} c_i^\lambda(w_j) I(c, w_j) \\
&= \tfrac{1-\lambda}{n-1} I(c, w_1) + \ldots + \lambda I(c, w_i) + \ldots + \tfrac{1-\lambda}{n-1} I(c, w_n) \\
&\leq \tfrac{1-\lambda}{n-1} \max_{w \in W} I(c, w) + \ldots + \lambda \min_{w \in W} I(c, w) + \ldots + \tfrac{1-\lambda}{n-1} \max_{w \in W} I(c, w) \\
&= (1-\lambda) \max_{w \in W} I(c, w) + \lambda \min_{w \in W} I(c, w) \\
&= -H_\lambda^{I}(c)
\end{aligned}$$

Thus,

$$-H_\lambda^{I}(c_i^\lambda) = \mathrm{Exp}_I(c_i^\lambda | c_i^\lambda) < \mathrm{Exp}(c | c_i^\lambda) \leq -H_\lambda^{I}(c)$$

as required. □

Putting together Lemma IV.A.2 and Lemma IV.A.3, we obtain Theorem 13.1.1(I).

IV.A.2 Proof of Theorem 13.1.1(II)

Next, we prove Theorem 13.1.1(II). Suppose $\lambda < \frac{1}{n}$. Now first, we note that

$$\begin{aligned} \text{Exp}_I(c^\dagger|c^\dagger) &= I(c^\dagger, w) \text{ for all } w \in \mathrm{W_F} \\ &= \lambda \min_{w \in \mathrm{W_F}} I(c^\dagger, w) + (1-\lambda) \max_{w \in \mathrm{W_F}} I(c^\dagger, w) \\ &= -H_\lambda^I(c^\dagger) \end{aligned}$$

Next, we have that

$$\begin{aligned} \text{Exp}_I(c|c^\dagger) &= \frac{1}{n} \sum_{w \in \mathrm{W_F}} I(c, w) \\ &\leq \lambda \min_{w \in \mathrm{W_F}} I(c, w) + (1-\lambda) \max_{w \in \mathrm{W_F}} I(c, w) \\ &= -H_\lambda^I(c) \end{aligned}$$

But, since $\mathfrak{I}$ is strictly proper, we have

$$-H_\lambda^I(c^\dagger) = \text{Exp}_I(c^\dagger|c^\dagger) < \text{Exp}_I(c|c^\dagger) \leq -H_\lambda^I(c)$$

as required. This gives us Theorem 13.1.1(II).

(p.182) This completes our proof of Theorem 13.1.1.

IV.B Proof of Theorem 13.3.1

Theorem 13.3.1 Suppose $\mathscr{C}$ is the set of possible chance functions. Suppose $\mathfrak{I}$ is an inaccuracy measure generated by additive Bregman divergence $\mathfrak{D}$. Then, if

(i) $c^\dagger$ is in $\mathscr{C}^+$; and
(ii) for all ch, ch' in $\mathscr{C}$, $\mathfrak{D}(ch, c'^\dagger) = \mathfrak{D}(ch', c^\dagger)$ then, for all $c \neq c^\dagger$,

$$\max_{ch \in \mathrm{C}} \text{Exp}_I(c^\dagger|ch) < \max_{ch \in \mathrm{C}} \text{Exp}_I(c|ch)$$

Proof. By Theorem I.B.4, since $\mathfrak{I}$ is generated by an additive Bregman divergence $\mathfrak{D}$, we have:

(i) $\mathfrak{D}(ch, c) = \text{Exp}_\mathfrak{I}(c|ch) - \text{Exp}_\mathfrak{I}(ch|ch)$
(ii) $\mathfrak{D}(ch, c^\dagger) = \text{Exp}_\mathfrak{I}(c^\dagger \mid ch) - \text{Exp}_\mathfrak{I}(ch|ch) = \mathfrak{I}(c^\dagger, w) - \text{Exp}_\mathfrak{I}(ch|ch)$ (for any w).

Thus, if we let $X = \mathfrak{I}(c^\dagger, w)$, which is well-defined since it takes the same value for any w, and $Y = \mathfrak{D}(ch, c^\dagger)$, which is also well-defined since by hypothesis this takes the same value for any ch, we have

$$\text{Exp}_I(c|ch) = \mathrm{D}(ch, c) + X - Y.$$

Now, suppose $c \in \mathscr{C}^+$ and $c \neq c^\dagger$. Then there is $ch \in \mathscr{C}$ such that

$$\mathrm{D}(ch, c) > \mathrm{D}(ch, c^\dagger)$$

Then

$$\mathrm{Exp}_I(c|ch) > \mathrm{Exp}_I(c^\dagger|ch)$$

Since $\mathrm{Exp}_{\mathfrak{I}}(c^\dagger|ch)$ takes the same value for all ch, we then have

$$\max_{ch \in C} \mathrm{Exp}_I(c^\dagger|ch) < \max_{ch \in C} \mathrm{Exp}_I(c|ch)$$

as required.

Notes:

([1]) This chapter draws on some material from (Pettigrew, to appear b).

([2]) Konek (to appear) extends this theorem to the case of imprecise credences.

([3]) More precisely, $c_i^\lambda = \lambda v_{w_i} + (1-\lambda)c_i$, where c_i is the following probabilistic credence function: $c_i(w_j) = 0$, if $i = j$; and $c_i(w_j) = \frac{1}{n-1}$, if $i \neq j$.

([4]) Thus: $\mathfrak{L}(c, w) := -ln\ c(w)$.

([5]) Thus: $\mathfrak{LA}(c, w) := \sum_{X \in \mathscr{F}} \mathfrak{l}(v_w(X), c(X))$,where $\mathfrak{l}(1, x) := -ln\ x$ and $\mathfrak{l}(0, x) := -ln\ (1 - x)$.

([6]) Thus: $\mathfrak{B}(c, w) = \sum_{X \in \mathscr{F}} \mathfrak{q}(v_w(X), c(X))$,where $\mathfrak{q}(1, x) = (1 - x)^2$ and $\mathfrak{q}(0, x) = x^2$.