On Concave Functions over Lotteries*

Roberto Corrao† MIT
Drew Fudenberg‡ MIT
David Levine§ EUI and WUSTL

February 13, 2023

1 Introduction

This note discusses functions over lotteries that are concave and continuous, but not necessarily superdifferentiable. Machina [1984] and Maccheroni [2002]'s analyses of utility functions over lotteries and the statement of Theorem 2 in Frankel and Kamenica [2019]'s proposed measure of information and uncertainty all claim or imply that if a function over lotteries is concave and continuous, then it can be written as the minimum of affine functions. However, Section 3 gives an example of a concave and continuous function that cannot be written as the minimum of affine functions, because there is no tangent hyperplanes that dominates the functions at the boundary.[1] Section 4 shows that concavity and upper semi-continuity are equivalent to a representation as the infimum of affine functions, and that these assumptions imply continuity for functions on finite-dimensional lotteries. Therefore, in finite-dimensional simplices, concavity and continuity are equivalent to the "infimum" representation.[2] The "mimimum"

*We thank Xiaosheng Mu and Luciano Pomatto for helpful comments and NSF grant SES 1951056 and the Gordon B. Pye Dissertation Fellowship for financial support.

†Department of Economics, MIT, rcorrao@mit.edu

‡Department of Economics, MIT, drew.fudenberg@gmail.com

§Department of Economics, EUI and WUSTL, david@dklevine.com

[1] The example uses a finite-dimensional probability simplex; the issue with boundary points is even more important in the infinite-dimensional setting.

[2] The main result of this section relies on a version of the Hahn Banach Theorem which, for completeness, we state and prove in the Apppendix.

representation is equivalent of the existence of local utilities (i.e., supporting affine functions) at every lottery, a property that is equivalent to superdifferentiability.[3]

2 Preliminaries

We study concave functions V on the space $\mathcal{F}$ of probability measures on a compact metric space X. Let $C(X)$ denote the set of continuous functions over X, and endow X with the Borel sigma-algebra. We give $\mathcal{F}$ the topology of weak convergence, so it is metrizable and compact.

The following statement is made in different, yet equivalent, forms in Machina [1984], Maccheroni [2002], and Frankel and Kamenica [2019].

Claim: *If a function $V : \mathcal{F} \to \mathbb{R}$ is concave and continuous, then it can be written as the minimum of a set of affine continuous functions, that is,*

$$V(F) = \min_{w \in \mathcal{W}} \int w(x) dF(x)$$

for some set $\mathcal{W} \subseteq C(X)$.

We provide a counterexample to this claim. Notably, the set X considered in the counterexample is finite, disproving the claim even in this simpler case. The example has a utility function over lotteries with three outcomes that is continuous and can be represented by a concave function. Moreover, there is a deterministic outcome that is preferred to all other lotteries with the property that the independence axiom is satisfied with respect to that certain outcome. However, the associated preferences over lotteries cannot be represented as the minimum of affine functions. This directly contradicts Theorem 1 in Maccheroni [2002], and as we explain below, it is not consistent with the statements of Theorem 2 in Machina [1984] and Theorem 2 in Frankel and Kamenica [2019].[4]

[3] This is also implied by the duality results in Dworczak and Kolotilin [2022].

[4] In Machina [1984] the claim is made for convex, as opposed to concave, functions, but the error is the same.

3 A Counterexample and Its Implications

3.1 The Snowcone

Suppose that X has 3 elements so that $\mathcal{F}$ is the simplex

$$\Delta = \left\{(p, q) \in [0, 1]^2 : p + q \leqslant 1\right\} \subset \mathbb{R}^2_+.$$

For every $(\tilde{p}, \tilde{q}) \in \mathbb{R}^2_+$ we let $(r(\tilde{p}, \tilde{q}), \theta(\tilde{p}, \tilde{q})) \in \mathbb{R}_+ \times [0, \pi/2]$ denote the corresponding polar coordinates, and conversely let $(\tilde{p}(r, \theta), \tilde{q}(r, \theta)) \in \mathbb{R}^2_+$ denote the point in the positive orthant corresponding to the polar coordinates (r, θ). In polar coordinates, the simplex can be expressed as

$$\mathcal{P}_\Delta = \left\{(r, \theta) \in \mathbb{R}_+ \times [0, \pi/2] : r \leqslant \frac{\sin(\pi/4)}{\sin(3\pi/4 - \theta)}\right\}.$$

It is convenient to denote points $(p, q) \in \Delta$ by F and use the notation $(r(F), \theta(F)) \in \mathcal{P}_\Delta$ and $F(r, \theta) \in \Delta$.

We now construct a utility function over the simplex by first constructing one of its "indifference curves," specifically the one corresponding to $V(F) = -1$. Toward this goal, consider a continuous function $\nu : [0, \pi/2] \to [0, 1]$ such that $\nu(\theta) \neq 0$ and $(\nu(\theta), \theta) \in \mathcal{P}_\Delta$ for all $\theta \in [0, \pi/2]$, and define the function $\tilde{V} : \mathbb{R}^2_+ \to \mathbb{R}$ as

$$\tilde{V}(\tilde{p}, \tilde{q}) = -\frac{r(\tilde{p}, \tilde{q})}{\nu(\theta(\tilde{p}, \tilde{q}))}.$$

Observe that $\tilde{V}$ is continuous and linearly homogeneous:

$$\tilde{V}(\gamma F) = -\frac{r(\gamma F)}{\nu(\theta(\gamma F))} = -\frac{\gamma r(F)}{\nu(\theta(F))} = \gamma \tilde{V}(F)$$

for all $\gamma \geqslant 0$ and $F \in \Delta$. Let $V : \Delta \to \mathbb{R}$ denote the restriction of $\tilde{V}$ over Δ. We interpret V as a utility function over lotteries. Because $r(F), \nu(\theta(F)) \in \mathbb{R}_+$ for every $F \in \Delta$, this utility function is negative and takes on a maximum at the certain outcome $p = q = 0$, which has utility 0. Finally, V assigns utility -1 to each point in the graph of ν, which is the set $\{F(\nu(\theta), \theta) \in \Delta : \theta \in [0, \pi/2]\}$. For this reason, we call ν the "indifference curve" of V.

To fix ideas, we construct one such indifference curve in a graphical way. Consider

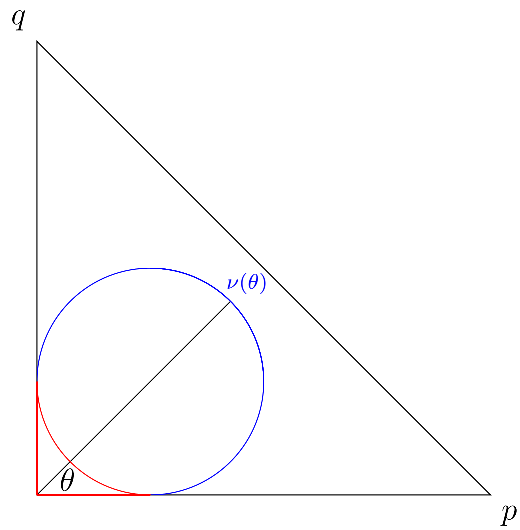

Figure 1: Construction of the indifference curves of the utility function.

the "snowcone" created by taking a circle tangent to the axes as in Figure 1. For each angle θ, the ray through θ intersects the circle at most twice; we take $\nu(\theta)$ equal to the length of the ray from the origin to the farther point of intersection, that is the one in the upper part of the circle, so that the indifference curve ν corresponds to the arc of the circle joining the two tangent points (see the blue arc in Figure 1). Note that this indifference curve is tangent to both the axes.

We say that the indifference curve ν is convex if the utility of any convex combination of two points on the indifference curve lies below the indifference curve: For all $F, \tilde{F}$ such that $V(F) = V(\tilde{F}) = -1$, it holds $V(\lambda F + (1-\lambda)\tilde{F}) \geqslant -1$. Notice that the "snowcone" indifference curve in Figure 1 is convex because the unit sphere is a convex set.

Proposition 1. *If V has a convex indifference curve, it is concave.*

A monotone decreasing utility function, like V, is concave when its indifference curves are convex. It is sufficient to assume convexity of one such indifference curve and then use positive homogeneity to get convexity of all indifference curves, as shown in Appendix A.

Next, we observe that V is affine with respect the most preferred lottery, which $p = q = 0$. To see this, consider the lottery $(1-\lambda)F + \lambda 0$. By homogeneity $V((1-\lambda)F + \lambda 0) = (1-\lambda)V(F)$, so F is preferred to $\tilde{F}$ if and only if $(1-\lambda)F + \lambda 0$ is preferred to $(1-\lambda)\tilde{F} + \lambda 0$.

To sum up, for any convex indifference curve ν we have constructed a utility function V over lotteries that is concave, continuous, and has a deterministic outcome that is preferred to all other lotteries, where the independence axiom is satisfied with respect to that outcome. Moreover, along the axes $p = 0$ and $q = 0$ preferences are strictly declining in the other probability. We showed by example that this indifference curve can be tangent to the axis, and since it does not include zero, it must intersect the axis. Since there is strict preference along the y axis, every straight line passing through that point must contained points strictly preferred to the intersection point. Hence the collection of affine linear functions that dominate a utility representation of these preferences cannot have a minimum at that point.

3.2 Implications of the Example

Machina's induced preferences

Machina [1984] analyzes preferences over lotteries $\mathcal{F} = \Delta([0,1])$ that carry delayed risk. Concretely, consider an agent choosing first $F \in \mathcal{F}$ and then, before the outcome from F has been realized, an action $y \in Y$ from a set of feasible actions. Even if the agent has expected utility preferences over pairs of outcomes and actions, the induced preferences over lotteries is

$$V(F) = \max_{y \in Y} \int u(x,y) dF(x) \tag{1}$$

where u is the utility of the agent and where we assume that the maximum is attained.

Theorem 2 in Machina [1984] states a converse of this fact, that is, if $V : \mathcal{F} \to \mathbb{R}$ is continuous and convex, then there exists a space of actions Y and a utility function $u(x,y)$ such that V can be represented as in equation 1 with the maximum being attained for every F. In particular, the set of actions is

$$Y = \left\{ y \in C([0,1]) : \forall F \in \mathcal{F}, V(F) \geqslant \int y(x) dF(X) \right\}.$$

and $u(x,y) = w(y)$. However the utility function $-V(F)$, where V is defined as in the snowcone example above, is continuous and convex, but does not have a representation in the form of a maximum as Machina asserts.

Maccheroni's model

Maccheroni [2002] studies preferences on lotteries over an arbitrary outcome space that admit representations V that is continuous, convex, and such that it satisfies the independence axiom with respect to the most favorite outcome whose existence is assumed. The main result claims that these preferences admit a maxmin expected utility representation where the decision maker evaluates each lottery according to the worst possible expected utility taken from a set of utilities over outcomes. In particular, it is claimed that for every lottery, one such worst-scenario utility exists and the minimum is attained. However, the snowcone example exhibits a preference over lotteries on three outcomes that satisfies all the axioms considered in Maccheroni [2002] but for which the minimum is not attained for lotteries at the boundaries of the simplex, contradicting the claim.

Measures of uncertainty and information

A similar statement to that of Machina is contained in Theorem 2 in Frankel and Kamenica [2019], which defines a measure of the uncertainty in a decision problem as follows. Fix a finite state space X and let $F \in \mathcal{F}$ denote an arbitrary belief over X. A *decision problem* is a pair (u, Y) of an action space Y and a utility function $u(x, y)$ such that, for every belief F, the maximum $\max_{y \in Y} \int u(x, y) dF(x)$ is attained. They say that $V : \mathcal{F} \to \mathbb{R}$ is a *valid measure of uncertainty* if there exists a decision problem (u, Y) such that

$$V(F) = \int \max_{y \in Y} u(x, y) dF(x) - \max_{y \in Y} \int u(x, y) dF(x)$$

for every F. Their Theorem 2 then states that V is a valid measure of uncertainty *if and only if* it is concave and such that $V(\delta_x) = 0$ for all $x \in X$, but the "if" direction is o contradicted by the snowcone example.

4 Concave Functions and Adversarial Representations

4.1 The Adversarial Representation

We say that V has an *adversarial representation* if

$$V(F) \equiv \inf_{y \in Y} \int u(x, y) dF(x)$$

where Y is a separable metric space and, for every $y \in Y$, $u(\cdot, y)$ is continuous over X.

Theorem 1. *V has an adversarial representation if and only if it is upper semi-continuous and concave.*

The issue with the disproved claim is that there may be no tangent hyperplanes at the boundary points of the simplex.[5] The idea of the theorem is that we can fix this by taking separating hyperplanes that aren't tangent to the concave function, but pass through a point above and near it. Where the function has infinite slope as we take closer points we get steeper separating hyperplanes, which is why we must use the inf rather than the min.

Proof. First we show that adversarial implies concave. Consider $F, \tilde{F}$ with $0 \leqslant \lambda \leqslant 1$ and y^n such that $\int u(x, y^n) dF(x) \to V(\lambda F + (1 - \lambda)\tilde{F})$. Consider that $V(F) \leqslant \int u(x, y^n) dF(x)$ and $V(\tilde{F}) \leqslant \int u(x, y^n) d\tilde{F}(x)$ so that $\lambda V(F) + (1-\lambda) V(\tilde{F}) \leqslant \int u(x, y^n) d(\lambda F + (1 - \lambda)\tilde{F})(x) \to V(\lambda F + (1 - \lambda)\tilde{F})$.

To show continuity, let $F^n \to F$,and choose y^m such that $V(F) > \int u(x, y^m) dF(x) - 1/m$. Then

$$V(F^n) \leqslant \int u(x, y^m) dF^n(x),$$

so

$$\lim V(F^n) \leqslant \lim \int u(x, y^m) dF^n(x) = \int u(x, y^m) dF(x) < V(F) + 1/m.$$

Hence V is upper semi-continuous.

[5]This problem is especially pervasive in the infinite-dimensional case, where the set of Borel probability measures over a compact metric space has empty (relative) interior when endowed with the topology of weak convergence.

To prove the other direction we use a version of the separating hyperplane theorem. Theorem 3 in the Appendix shows that for each $v > V(F)$ there is a continuous function $w(x)$ such that $\int w(x)d\tilde{F} > V(\tilde{F})$ for all $\tilde{F} \in \mathcal{F}$. Now take Y to be the subset of continuous functions in the sup norm for which $\int y(x)d\tilde{F}(x) > V(\tilde{F})$ for all $\tilde{F}$. This is a separable metric space, since it is an open subset of separable space of all continuous function in that norm. Since for every (v, F) there exists a $y \in Y$ such that $v \geqslant \int y(x)dF(x) > V(F)$, we see that $V(F) \equiv \inf_{y\in Y} \int u(x, y)dF(x)$. □

4.2 Concavity and continuity over finite-dimensional simplices

Notice that $V(F)$ arising from an adversarial representation need not be continuous: Theorem 1 only delivers concavity and upper semi-continuity. Even in finite-dimensional spaces, there are concave functions that fail to be lower semi-continuous, as shown by an example in Rockafellar [1970] Chapter 10.[6] However, in studying preference over lotteries, the convex sets on which utility is defined are typically taken to be probability simplices, and the restriction of concave and upper semi-continuous functions to a finite-dimensional subset of $\mathcal{F}$ is continuous, as we show below. Thus when we consider an arbitrary finite $X_0 \subseteq X$, the restriction of V to the space of lotteries over X_0 is continuous.

Theorem 2. *Consider the space $\overline{\mathcal{F}}$ of all convex combinations of N lotteries $\overline{F}^1, ..., \overline{F}^N$, and suppose that $V(F)$ is concave. Then $V(F)$ restricted to $\overline{\mathcal{F}}$ is lower semi-continuous, and in particular if $V(F)$ is upper semi-continuous then it is continuous.*

When the probability distributions $\overline{F}^1, ..., \overline{F}^N$ coincide with the point-mass measures over N outcomes, the set $\overline{\mathcal{F}}$ is a probability simplex. The result holds for "generalized simplicies" that are formed by linearly combining N arbitrary lotteries.

Proof. (*Adapted from Chapter 10 in Rockafellar [1970].*) There is a subset of $\{\overline{F}^1, ..., \overline{F}^N\}$ that consists of extremal points and whose convex hull is equal to $\overline{\mathcal{F}}$, so w.l.o.g. can assume that $\overline{F}^1, ..., \overline{F}^N$ are extremal and in particular affinely independent. Hence we may think of points being identified with vectors p on the n-dimensional simplex and we write $\overline{p}^i$ for the basis vectors.

Now consider $\tilde{p} \in \overline{\mathcal{F}}$. Our goal is to prove that for any sequence $p^n \to \tilde{p}$ we have $\liminf V(p^n) \leqslant V(\tilde{p})$. Suppose $\tilde{p}$ is not extremal and consider some particular

[6] Here is a version of that example: the linearly homogeneous function $u(p, q) = -(q + p)^2 p^{-1}$ for $p > 0$. Here, the boundary of the domain is non-linear, which causes a failure of lower semi-continuity.

p^n and define $\lambda \equiv \max\{\lambda \geqslant 0 | \lambda\tilde{p} \leqslant p^n\}$. If $\lambda = 0$ choose i such that $p_i^n = 0$ and $\tilde{p}_i > 0$ otherwise choose i such that $\lambda\tilde{p}_i = p_i^n$ and $\tilde{p}_i > 0$. Consider then the set $\{\overline{p}^1, ..., \overline{p}^N, \tilde{p}\} - \overline{p}_i$. Since $\tilde{p}_i > 0$ this set is affinely independent. We claim in addition that p^n is a convex combination of these vectors. If $\lambda = 0$ since $p_i^n = 0$ we have p^n already a convex combination of $\{\overline{p}^1, ..., \overline{p}^N\} - \overline{p}_i$. Otherwise since $p_j^n - \lambda\tilde{p}_j \geqslant 0$ we may write $p^n = \lambda\tilde{p} + \sum_j (p_j^n - \lambda\tilde{p}_j)\overline{p}^j$ since this is the same as $p_j^n = \lambda\tilde{p}_j + (p_j^n - \lambda\tilde{p}_j)$.

Consider affinely independent sets of the form $(\tilde{p}, \tilde{p}^1, ..., \tilde{p}^{n-1})$. We showed that if $\tilde{p}$ is not extremal then p^n there exists a set of this form, so p^n lies in the convex hull of the set, and if $\tilde{p}$ is extremal this is true by taking the remaining $n-1$ vectors to be the remaining basis vectors. Since there are at most n sets of this form it follows that there is a subseqence p^m converging to $\tilde{p}$ that lies entirely in such a set. Clearly $\liminf V(p^n) \leqslant \liminf V(p^m)$, so it suffices to prove $\liminf V(p^m) \leqslant V(\tilde{p})$.

Consider then that we can write $p^m = \gamma^m\tilde{p} + \sum_{i=1}^{n-1} \gamma_i^m \tilde{p}^i$ and that $\gamma^m \to 1, \gamma_k^m \to 0$. Then

$$V(p^m) = V(\gamma^m\tilde{p} + \sum_{i=1}^{n-1} \gamma_i^m V(\tilde{p}^i) \geqslant \gamma^m V(\tilde{p}) + \sum_{i=1}^{n-1} \gamma_i^m V(\tilde{p}^i) \to V(\tilde{p})$$

which was our goal. □

4.3 Local utilities and minima

We say that a continuous function $w(x)$ on a compact set X is a *local utility function* for V at $F \in \mathcal{F}$ if $\int w(x) d\tilde{F}(x) \geqslant V(\tilde{F})$ for all $\tilde{F} \in \mathcal{F}$ and $\int w(x) dF(x) = V(F)$. We say that V has a local expected utility if it has a local utility function at each $F \in \mathcal{F}$.[7] We let $\mathcal{W}_V(F) \subseteq C(X)$ denote the set of local expected utilities of V at F and, when V has a local expected utility, we denote $\mathcal{W}_V = \bigcup_{F \in \mathcal{F}} \mathcal{W}_V(F)$.

Proposition 2. *V has a local expected utility if and only if there exists a set $\mathcal{W} \subseteq C(X)$ such that $V(F) = \min_{w \in \mathcal{W}} \int w(x) dF(x)$. In this case, one such set is $\mathcal{W} = \mathcal{W}_V$.*

The proof of this result is simple and relegated to Appendix A. It is not hard to verify that V has a local utility at F if and only if it is superdifferentiable at F in the sense that there exists a linear function L such that $V(\tilde{F}) \leqslant V(F) + L(\tilde{F} - F)$ for all $\tilde{F} \in \mathcal{F}$. Thus Proposition 2 implies that the "minimum" representation is equivalent to superdifferentiability, as also shown in Dworczak and Kolotilin, 2022. We do not

[7]It is easy to that if V has a local expected utility, then it is concave (see Corrao, Fudenberg, Levine, et al. [2022]).

know of a characterization of superdifferentibility of V in terms of more primitive functional conditions or axioms, but there are stronger conditions and axioms that imply superdifferentiability and the "minimum" representation. Notable examples of this are Chatterjee and Krishna [2011], Evren [2014], Sarver [2018], and Ke and Zhang [2020].

References

Aliprantis, C. D. and K. C. Border (2006). *Infinite Dimensional Analysis*. Springer.

Bogachev, V. I. and O. G. Smolyanov (2017). *Topological vector spaces and their applications*. Springer.

Chatterjee, K. and R. V. Krishna (2011). "A nonsmooth approach to nonexpected utility theory under risk." *Mathematical Social Sciences*, 62, 166–175.

Corrao, R., D. Fudenberg, D. Levine, et al. (2022). *Adversarial forecasters, surprises and randomization*. Tech. rep. David K. Levine.

Dworczak, P. and A. Kolotilin (2022). "The persuasion duality." *Available at SSRN 3474376*.

Evren, Ö. (2014). "Scalarization methods and expected multi-utility representations." *Journal of Economic Theory*, 151, 30–63.

Frankel, A. and E. Kamenica (2019). "Quantifying information and uncertainty." *American Economic Review*, 109, 3650–80.

Ke, S. and Q. Zhang (2020). "Randomization and ambiguity aversion." *Econometrica*, 88, 1159–1195.

Maccheroni, F. (2002). "Maxmin under risk." *Economic Theory*, 19, 823–831.

Machina, M. J. (1984). "Temporal risk and the nature of induced preferences." *Journal of Economic Theory*, 33, 199–231.

Rockafellar, R. T. (1970). *Convex analysis*. Vol. 18. Princeton university press.

Sarver, T. (2018). "Dynamic mixture-averse preferences." *Econometrica*, 86, 1347–1382.

A Appendix: Proofs

Proof of Proposition 1. First, we show that, for all $F \in \Delta$, it holds $\frac{1}{-V(F)}F \in \Delta$ and $V\left(\frac{1}{-V(F)}F\right) = -1$. Indeed, $r\left(\frac{1}{-V(F)}F\right) = \frac{1}{-V(F)}r(F) = \nu(\theta(F))$ and $\theta\left(\frac{1}{-V(F)}F\right) = \theta(F)$, implying that $\left(r\left(\frac{1}{-V(F)}F\right), \theta\left(\frac{1}{-V(F)}F\right)\right) \in \mathcal{P}_\Delta$ by the properties of ν, and hence that $\frac{1}{-V(F)}F \in \Delta$. Moreover,

$$V\left(\frac{F}{-V(F)}\right) = V\left(\frac{F\nu(\theta(F))}{r(F)}\right) = \frac{\nu(\theta(F))}{r(F)}V(F) = -\frac{\nu(\theta(F))}{r(F)}\frac{r(F)}{\nu(\theta(F))} = -1,$$

yielding the second part of the claim. Second, observe that for all $\gamma \in [0,1]$ and $F, \tilde{F} \in \Delta$ we have

$$V\left(\gamma\frac{F}{-V(F)} + (1-\gamma)\frac{\tilde{F}}{-V(\tilde{F})}\right) \geqslant -1$$

by the first claim and the convexity of the indifference curve $\nu(\theta)$. Third, fix $\lambda \in [0,1]$, and $F, \tilde{F} \in \Delta$, and define

$$\gamma = \frac{\lambda V(F)}{\lambda V(F) + (1-\lambda)V(\tilde{F})}$$

and observe that

$$1-\gamma = \frac{(1-\lambda)V(\tilde{F})}{\lambda V(F) + (1-\lambda)V(\tilde{F})}$$

and that both γ and $(1-\gamma)$ are in $[0,1]$ since $0 \leqslant \lambda \leqslant 1$ and $V \leqslant 0$. Then

$$\begin{aligned} -1 \leqslant V\left(\gamma\frac{F}{-V(F)} + (1-\gamma)\frac{\tilde{F}}{-V(\tilde{F})}\right) &= V\left(-\frac{\lambda F + (1-\lambda)\tilde{F}}{\lambda V(F) + (1-\lambda)V(\tilde{F})}\right) \\ &= -\frac{1}{\lambda V(F) + (1-\lambda)V(\tilde{F})}V\left(\lambda F + (1-\lambda)\tilde{F}\right). \end{aligned}$$

Because $\lambda V(F) + (1-\lambda)V(\tilde{F})$ is negative, $V\left(\lambda F + (1-\lambda)\tilde{F}\right) \geqslant \lambda V(F) + (1-\lambda)V(\tilde{F})$, yielding concavity of V. □

Proof of Proposition 2. If V has a local expected utility, then for each $F \in \mathcal{F}$ $\int \hat{w}(x)dF(x) = V(F)$ for all $\hat{w} \in \mathcal{W}_V(F) \subseteq \mathcal{W}_V$, and $\inf_{w\in\mathcal{W}_V}\int w(x)dF(x) \geqslant V(F)$, so the "only if' part follows. Conversely, let V be such that $V(F) = \min_{w\in\mathcal{W}}\int w(x)dF(x) = \min_{w\in\mathcal{W}}\int w(x)dF(x)$ for some set $\mathcal{W} \subseteq C(X)$. Because the minimum is attained, for every F, there exists $w_F \in \mathcal{W}$ such that $V(F) = \min_{w \in \mathcal{W}} \int w(x)dF(x) = \int w_F(x)dF(x)$, so that $w_F \in \mathcal{W}_V(F) \neq \varnothing$. This in turn implies that V has a local expected utility. □

B Appendix: The Separating Hyperplane Theorem

Infering properties such as differentiability and concavity from a utility function rests on the separating hyperplane theorem, and one source of error has been mis-applying the theorem to infinite- dimensional lotteries. Here we give a careful proof of the separating hyperplane theorem that applies in this setting. Our starting point is the Hahn decomposition Theorem as stated in Aliprantis and Border [2006]. For ease of reference we state that result in the form in which we use it.

Aliprantis and Border [2006] Theorem 5.79. *If the hypograph of $V(F)$, that is the set in $\mathbb{R} \times \mathcal{F}$ consisting of $L = \{(v, F)|v \leqslant V(F)\}$, is closed and convex, then for each singleton set (v, F) with $v > V(F)$ by we may find a continuous linear functional separating (v, F). This means that there are numbers c_0, z and continuous function $w_1(x)$ such that for $\tilde{v}, \tilde{F} \in L$ we have $c_0\tilde{v} + \int w_1(x)d\tilde{F}(x) < z$ and $c_0 v + \int w_1(x)dF(x) > z$.*[8]

Theorem 3. *For each $v > V(F)$ there exists a continuous function $w(x)$ such that $\int w(x)d\tilde{F} > V(\tilde{F})$ for all $\tilde{F} \in \mathcal{F}$.*

Proof. We analyze the space of signed measures $H \in \mathcal{M}$ on the Borel σ-sigma algebra of X. The Hahn decomposition theorem says that for any signed measure H the space X can be partitioned into two Borel sets A, B such that for Borel $E \subseteq A$ we have $H(E) \geqslant 0$ and for $E \subseteq B$ we have $H(E) \leqslant 0$. The Jordan decomposition further states that there are two positive (ordinary measures) H^+, H^- (uniquely defined) such that for any Hahn decomposition $H^+(B) = 0, H^-(A) = 0$ and $H = H^+ - H^-$. With this in mind, for any continuous function $w : Z \to \mathbb{R}$ and any signed measure we may define $\int w(x)dH(s) = \int w(x)dH^+(x) - \int w(x)dH^-(s)$ where these are ordinary integrals with respect to a signed measure. We may define the total variation $|H| = \int dH^+(x) + \int dH^-(x)$.

[8] Crucially, this proposition is false if V can take real values, as shown by Bogachev and Smolyanov [2017].

Denote the space of bounded continuous functions in the sup norm on X by $C(X)$. On $\mathcal{M} \times C(X)$ we define the operation $< H, w >\equiv \int w(x)dH(x)$. If this is continuous and linear in each argument and $< H, w >= 0$ for all $w \in C(X)$ if and only if $H = 0$ and for all $H \in \mathcal{M}$ if and only if $w = 0$ then $\mathcal{M}, C(X)$ is a dual pair. Continuity follows immediately from $< H, w > \int w(x)dH(x) \leqslant \|w\| \, \|H\|$ and this implies $< H, w >$ is jointly continuous in the product topology on $\mathcal{M} \times C(X)$.

It follows that $\mathcal{H}$ is a locally convex topological space in the weak topology induced by $C(Z)$, and that its continuous linear functionals have the form $\int c(z)dH(z)$ for $c \in C(Z)$. Hence this topology relativizes to the subset of probability measures as the topology of weak convergence. That $< H, w >= 0$ for all H if $w = 0$ is obvious and only if $w = 0$ follows from considering that the Dirac delta functions $\delta(x)$ are in H and $\int w(x)d\delta(\hat{x})(x) = w(\hat{x})$. That $< H, w >= 0$ for all w if $H = 0$ is obvious but only if requires some work.

For any $H \neq 0$, we want to find a continuous function $w(x)$ such that $\int w(x)dH(x) \neq 0$. Let A, B be a corresponding Hahn partition of X, and write the Jordan decomposition $H = H^+ - H^-$. Assume without loss of generality that $H^+ \neq 0$. Because X is compact H^+, H^- are regular measures and in particular $H^+(A) = \sup_{K \subset A} H^+(K)$ and $H^-(B) = \sup_{K \subset B} H^-(K)$ where the supremum is over all compact subsets. Hence we can fix compact $K^+ \subset A$ such that $|H^+(A) - H^+(K^+)| \leqslant (1/3)H^+(A)$ and a compact $K^- \subset B$ such that $|H^-(A) - H^-(K)| \leqslant (1/3)H^+(A)$. As K^-, K^+ are disjoint and X is metric and K^+, K^- are close we may use Ursyohn's Lemma to find a continuous function $0 \leqslant w(x) \leqslant 1$ which is equal to 1 on K^+ and 0 on K^-. Now write the integral

$$\begin{aligned}
\int w(x)dH(x) &= \left[\int_{K^+} w(x)dH^+(x) + \int_{X-K^+} w(x)dH^+(x)\right] \\
&\quad - \left[\int_{K^-} w(x)dH^-(x) + \int_{K^+} w(x)dH^-(x) + \int_{X-K^+-K^-} w(x)dH^-(x)\right] \\
&= H^+(K^+) + \int_{X-K^+} w(x)dH^+(x) - \int_{X-K^+-K^-} w(x)dH^-(x) \\
&\geqslant (2/3)H^+(A) - (1/3)H^+(A) \geqslant (1/3)H^+(A) > 0.
\end{aligned}$$

Since $\mathcal{M}, C(X)$ is a dual pair $\mathcal{M}$ is locally convex with respect to the weak topology induces by $C(X)$ in the sup norm: relativized to the probability measures this is the same as the topology of weak convergence.

The hypograph of $V(F)$ is closed because V is upper semi-continuous. Hence by

Theorem for each compact (singleton) set (v, F) with $v > V(F)$ there are numbers c_0, z and continuous function $w_1(x)$ such that for $\tilde{v}, \tilde{F} \in L$ we have $c_0\tilde{v} + \int w_1(x)d\tilde{F}(x) < z$ and $c_0v+\int w_1(x)dF(x) > z$. Applying the first to $\tilde{v}, F \in L$ we have $c_0\tilde{v}+\int w_1(x)dF(x) < z$ so that $c_0\tilde{v} < c_0v$ implying since $v > \tilde{v}$ that $c_0 > 0$. Define $w(x) = -(w_1(x) - z)/c_0$. Observing that $(V(\tilde{F}), \tilde{F}) \in L$ the first inequality says $\int w(x)d\tilde{F} > V(\tilde{F})$ for all $\tilde{F}$ while the second implies $v \geqslant \int w(x)dF(x)$. □