Disentangling Strict and Weak Choice in Random Expected Utility Models

Evan Piermont* and Roee Teper†

July 16, 2018

Abstract

We put forth a model of random choice in which precise choice frequencies of objects are identified only up to the frequency they are chosen by strict maximization. The resulting primitive is a sub-additive capacity (i.e., set function). We provide simple restrictions on this primitive that are both necessary and sufficient for consistency with a random expected utility function. Thus, our model embeds both deterministic choice, regular random expected utility a la Gul and Pesendorfer (2006), and any combination between. We discuss several economic environments where such a primitive could be observed.

1 Introduction

When the characteristics of agents are observed only in coarse manner, a population of observationally identical decision makers might take distinct actions. For example, agents who stake their action on different private information. From the analyst's perspective, choice appears to be random. *Random Utility Models* (RUM)—a set of utility functions and a probability measure thereover—are a powerful and tractable tool in the analysis of such a scenario. The probability of observing x from the decision problem D is the probability of a utility function u such that $x = \arg\max_{z \in D} u(z)$.

The modern decision theoretic foundation for RUMs was introduced by Gul and Pesendorfer (2006) (also referred to as GP). However, properly dealing with indifference has beleaguered this literature. Consider the case where with positive probability $u(x) = u(y)$; the probability x is chosen from $D = \{x, y\}$ is undefined by the RUM. GP and Frick et al. (2017) deal with this issue by considering only *regular* (sometimes referred to as *proper*) RUMs such that ties occur with probability 0; Lu (2016a) extends this methodology to

*Royal Holloway, University of London, Department of Economics, evan.piermont@rhul.ac.uk

†University of Pittsburgh, Department of Economics, rteper@pitt.edu

allow ties with probability 0 or probability 1; Apesteguia et al. (2017) rule out ties by fiat, considering only linearly ordered preferences.

In this paper, *we put forth a model of random choice in which precise choice frequencies are identified only up to the frequency they are chosen by strict maximization.* We show that such a model (over a linear space) is always representable by a random (linear) utility model without *any* restrictions on the measure over utility functions. In other words, the set of random expected utility models and the set of choice rules considered in this paper are in bijection.

The primitive is a *choice capacity* over a linear space (such as von-Neumann-Morgenstern lotteries): Let D be a set of von-Neumann-Morgenstern lotteries. A choice capacity is a set function $\rho_D : 2^D \to [0,1]$, that need not be additive. In what follows, we explore how a choice capacity can reveal *more* about agents' (random) preferences.[1] The interpretation is that $\rho_D(A)$ reflects maximal frequency with which elements of $A \subseteq D$ are chosen when the decision problem is D.[2] When choice out of D is contingent on tie breaking, then the objective probability of some subsets will only be set identified—the analyst can only claim that the empirical frequency of elements in A being chosen lies within an interval. In such cases, we assume that ρ reflects the upper bound of the interval; because of this, ρ need not be additive. The following example clarifies and explores this idea.

Example. Let $\{a, b\}$ be a set of prizes, with $P \cong [0,1]$ the set of lotteries thereon. The set of expected utility indices that are realized with positive probability are given by

$$u_1 = [1,0], u_2 = [-1,0], \text{ and } u_3 = [0,0].$$

Let ξ be the uniform measure over these utility indices. Consider the decision problem $D = \{a, b\} \subset P$ (where we identify outcomes with the degenerate lotteries thereon). Let $\rho^\circ_D(a)$ be the (objective, but not necessarily observable) probability that a is chosen from D. Notice, if u_1 is realized then a definitively chosen, so $\rho^\circ_D(a) \geqslant \frac{1}{3}$; if u_2 then a is definitively not chosen, so $\rho^\circ_D(a) \leqslant \frac{2}{3}$. When u_3 is realized, the probability a is chosen depends on the tie breaking rule; it is not identified by ξ. If a single tie breaking rule is not consistently employed, this probability may not converge at all. It will be, however, set identified: we know only that $\rho^\circ_D(a) \in [\frac{1}{3}, \frac{2}{3}]$. Further consider $\rho^\circ_D(\{a, b\})$, the probability that either a or b is chosen. Of course, $\rho^\circ_D(\{a, b\})$ can be point identified as 1, as it does not depend on how ties are treated. Since ρ_D reflects the upper bounds of ρ°_D, we see that $\rho_D(\{a, b\}) = 1 < \rho_D(a) + \rho_D(b) = \frac{2}{3} + \frac{2}{3}$, so that ρ_D is not additive.

We provide the conditions on a choice capacity to ensure it maximizes a probability

[1] Our model is flexible enough to encompass all matter of different phenomena, ranging from degenerate choice (nesting the classical choice model) to regular random utility functions a la GP, and any combination thereof.

[2] Maximal in the sense that an element of A is chosen unless some other element *strictly* dominates it. We could just as well start with the minimal frequency, $\rho^l_D(A) \coloneqq 1 - \rho_D(D \backslash A)$, reflecting the frequency with which A is chosen strictly. Everything in our analysis would follow.

distribution over utilities. That is, to ensure the existence of a probability over utilities, ξ, such that

$$\rho_D(A) = \xi\big(\{u \mid \arg\max_{z \in D} u(z) \cap A \neq \varnothing\}\big).$$

Without any restrictions on how ties are broken, $\rho_D(A)$ reflects the maximal probability that an element of A is chosen when the choice problem is D and preferences are realized according to ξ. When ξ is regular, ρ is a measure, and all of the GP axioms hold. Thus, our innovation concerns dealing directly with how the non-additivity can enter ρ.

Notice, accommodating for tie-breaking in random choice models by introducing choice capacities is not merely a technical point, but carries economic content. Private information acquisition is a natural example for random choice: given a common prior, randomness enters because different agents observe different signals. However, conditional on a private signal, an agent will necessarily be indifferent between some alternatives, and may be forced to break ties. Lu (2016a) takes into account ties that happen with probability 0 or 1. This excludes many natural and commonly employed signal structures, such as a common prior and a finite number of private signals. By allowing indifference to obtain with arbitrary probability, our model allows for random choice based on any information structure.

Lastly, there is the matter of observability of choice capacities. While the primitive in Lu (2016a) is a classical distribution over alternatives, this issue is briefly touched upon. It is suggested that non-additivity of probabilities stems from non-measurability of choice problems; he takes as part of his primitive an algebra of 'measurable sets'—those sets where ρ is identified by strict maximization—and extends the choice rule to non-measurable sets via the outer measure. Such a methodology would clearly also work in our environment. But, we here note that choice capacities can be *directly* observed in an array of economic environments, without exogenously imposing the set of measurable sets.

In Section 5, we discuss how this could be achieved in various contexts: choice in an online market place, choice with status quo bias or in the presence of a default option, choice across different populations, and from non-convergent statistics.

2 Notation

Primitives. A finite subset of $\mathbb{R}^n$ is referred to as a *menu* or *decision problem*. Let $\mathcal{D}$ denote the set of all decision problems with D a generic element. For any vector, $x \in \mathbb{R}^n$ let x^i denote its i^{th} component.

The primitive of the theory is a choice capacity (CC): $\rho = \{\rho_D\}_{D \in \mathcal{D}}$ where for each D, ρ_D is a *capacity* over D. Specifically ρ_D is a *grounded, normalized, and monotone set function*: i.e., $\rho_D : 2^D \to [0,1]$ such that $\rho_D(\varnothing) = 0$, $\rho_D(D) = 1$, $\rho_D(A \cup B) \geqslant \rho_D(A)$. When it is not confusing to do so, we will abuse notation, letting $\rho_D(A) = \rho_D(A \cap D)$ so as to extend ρ_D to A that are not subsets of D.

GP show (in their Appendix B) that it is without loss of generality to consider only choice problems that are elements of the n dimensional simplex. This lends the interpretation that there is a set of $n+1$ consumption prizes, and decision problems are sets of lotteries thereover—the resulting representation is interpreted as a probability distribution on vNM indices. The advantage of the more general framework is that it allows other interpretations without any change to the primitive. Indeed, we could interpret each dimension as a 'state-of-the-world,' and a decision problem as a collection of Anscombe-Aumann acts (whose outcomes are in utils). Here, the resulting representation is interpreted as an information representation, a la Lu (2016a), a probability distribution over beliefs regarding the state space.

Random Linear Representations. While a CC corresponds to the observable behavior of a population of agents, we interpret the choices as resulting from the maximization of preference. Here, we take a preference to be a linear function over the n dimensions, which, of course, can be represented by a vector in R^n. When interpreting our primitive as choices over lotteries, the linear function corresponds to a utility index over the $n+1$ prizes.[3] When considering our primitive to be choices over Anscombe-Aumann acts, the linear function corresponds to the relative likelihood of each of the n states.[4]

For $(u, x) \in R^n \times R^n$, we write $u(x)$ to denote the inner product of the vectors u and x. For a given decision problem D, let $M(D, u)$ denote the set of vectors that maximize u over the domain D: $M(D, u) = \arg\max_{x \in D} u(x)$. In dual fashion, for $A \subset R^n$, define $N(D, A)$ to be the set of utilities such that something in A is maximal over D according to u: $N(D, A) = \{u \in \mathbb{R}^n \mid A \cap M(D, u) \neq \varnothing\}$. When $A = \{x\}$ is a singleton, then $N(D, \{x\})$ is the normal cone to D at x. The idea being that if an agent entertains preference u when facing problem D, her selection will be in the set $M(D, u)$. Taking this as given, if we observe the agent choose $x \in A$ from decision problem D, it must be that her preference was in $N(D, A)$.

Of course, the potential randomness of ρ indicates that the underlying preference may not be constant. Towards this, we define a *Random Linear Representation* (RLR). Let Ω be the smallest algebra on $\mathbb{R}^n$ that contains $N(D, A)$ for all (D, A) (where we set $N(D, A) = \varnothing$ if $A \cap D = \varnothing$). Then a *Random Linear Representation* is a probability measure over $(\mathbb{R}^n, \Omega)$.

Definition. *Let ξ be a RLR. Then say that ρ maximizes ξ if $\rho_D(A) = \xi(N(D, A))$ for all (D, A).*

GP define the maximization by a random choice rule (i.e., and additive CC), in an analogy the the definition above, but impose consistency between ρ and ξ only over singleton sets—of course, additivity ensures ρ_D can be extended to arbitrary sets. Without additivity,

[3]Per usual, we can normalize the utility of the $(n+1)^{th}$ prize to 0, so that the set of utility functions considered is in bijection with R^n.

[4]Notice, to make sense of this interpretation, we need to ensure that beliefs can be normalized, hence, the linear function must be a strictly positive vector. This requires additional axioms; see Lu (2016b).

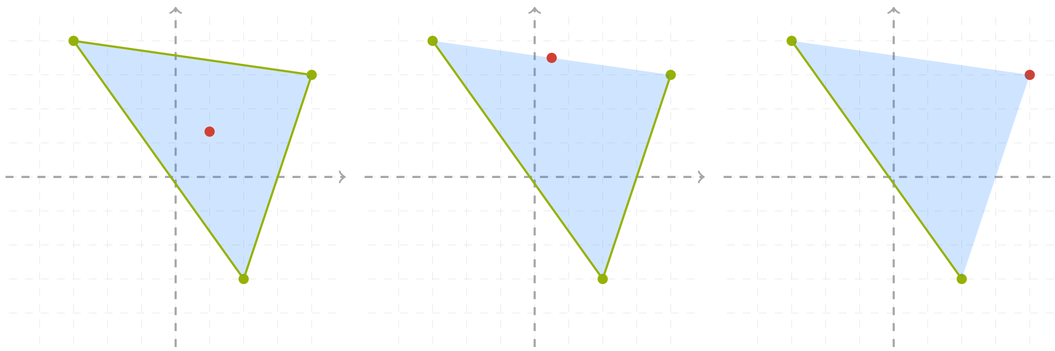

Figure 1: $\mathrm{pi}(D, A)$ for different sets A and the same D. The set A is the red point, $\mathrm{pi}(D, A)$ is in blue, and $D \backslash \mathrm{pi}(D, A)$ is in green.

we must impose consistency directly over all subsets of the decision problem. Despite this, if ρ maximizes ξ, then it is uniquely determined by its value on singleton sets.

Operations on Convex Sets. For a set $A \subset R^n$, let $\mathrm{conv}(A)$ and $\mathrm{int}(A)$ denote the convex hull and the interior of A, respectively. Moreover, if A is convex the let $\mathrm{ext}(A)$ collect the extreme points of A and $\mathrm{ri}(A)$ denote the relative interior of A. When it is not confusing to do so, we will write $\mathrm{ri}(A)$ and $\mathrm{ext}(A)$ to mean $\mathrm{ri}(\mathrm{conv}(A))$ and $\mathrm{ext}(\mathrm{conv}(A))$ for non-convex A.

Let $\mathrm{pi}(D, A) = \{x \in \mathrm{conv}(D) \mid x = \alpha a + (1-\alpha)y, a \in A, y \in \mathrm{conv}(D), \alpha \in (0, 1)\}$ denote the projective interior of A in D. This is the set of vectors in D which can be written as a convex combination placing positive weight on elements of A. It is straightforward to show that $\mathrm{pi}(D, A)$ is the union of the interiors of all faces intersecting A.

3 Axiomatics

3.1 Some Intuition

There are two key axioms. The first replaces the GP extremeness axiom and which we inventively call *Extremeness*. Let $D = \{x, y, \frac{1}{2}x + \frac{1}{2}y\}$. Trivially, for expected utility function u, $u(\frac{1}{2}x + \frac{1}{2}y) \leqslant \max\{u(x), u(y)\}$; so, the convex combination of two lotteries will be chosen if and only if the two lotteries are both chosen themselves.

The standard extremeness axiom states that the extreme points of D are chosen with probability 1. In GP, extremeness and additivity collectively ensure *non*-extreme points must chosen with probability 0; we can conclude that the probability that x is chosen is the same as the probability that either x or $\frac{1}{2}x + \frac{1}{2}y$ is chosen since the latter is not an extreme point. We arrive at the conclusion of the above paragraph. In our model, the conclusion still holds but the argument does not work since non-extreme points *can* be chosen, albeit only in the event of a tie. Our extremeness axiom is correspondingly more direct: it must be that $\rho_D(\{x\}) = \rho_D(\{x, \alpha x + (1-\alpha)y\})$ whenever $\alpha \in (0,1)$ and $y \in D$. This statement implies the canonical extremeness.

The second novel axiom, *Convex-Modularity*, regards the limits of how non-additive ρ can be. Consider again the same D as above, notice that

$$\rho_D(\{x\}) = \xi\big(\{u \mid u(x) > u(y)\}\big) + \xi\big(\{u \mid u(x) = u(y)\}\big), \text{ and}$$
$$\rho_D(\{y\}) = \xi\big(\{u \mid u(y) > u(x)\}\big) + \xi\big(\{u \mid u(x) = u(y)\}\big).$$

Therefore, $\rho_D(\{x\}) + \rho_D(\{y\}) \neq \rho_D(\{x,y\}) = 1$ if and only if $\xi(\{u \mid u(x) = u(y)\} > 0$. Moreover, by the arguments laid out above, we will have that $\rho_D(\{\frac{1}{2}x + \frac{1}{2}y\})$ is exactly the probability of a tie occurring. Simple accounting reveals the following modularity relation

$$\rho_D(\{x,y\}) + \rho_D(\{\frac{1}{2}x + \frac{1}{2}y\}) = \rho_D(x) + \rho_D(y).$$

Our Convex-Modularity axiom states that ρ must satisfy a generalized form of the above relation: the sum of the probability of two sets is the probability of the union plus the probability of the convex combination. Notice this ensures the capacity is sub-additive.

3.2 The Axioms

Axiom 1—Strong Monotonicity. *Let $D \subset D'$ be two decision problems, and let $A \subset D$. Then*

$$\rho_D(A) \geqslant \rho_{D'}(A),$$

with equality whenever $\text{ext}(D) = \text{ext}(D')$.

The usual monotonicity condition states that adding additional elements to a choice set cannot *increase* the likelihood of a (previously available) element being chosen. This is essentially a form of independence of irrelevant alternatives. Strong monotonicity, in addition, states that if we do not change the extreme points of a menu, then there can be *no* change in the likelihoods of choosing given elements. Notice, in the case where only extreme points are ever chosen, this additional dictate is implied by the usual monotonicity axiom.

Axiom 2—Extremeness. *Let $A \subseteq D$. Then*

$$\rho_D(\text{pi}(D, A)) = \rho_D(A).$$

The conventional extremeness axioms set $\rho_D(\text{ext}(D)) = 1$; this is implied by our version by setting $A = \text{ext}(D)$ so that $\text{pi}(D, A) = D$: $\rho_D(A) = \rho_D(D) = 1$.

Axiom 3—Convex-Modularity. *Let $A, B \subseteq D$ be such that $\frac{1}{2}A + \frac{1}{2}B \subseteq D$. Then*

$$\rho_D(A \cup B) = \rho_D(A) + \rho_D(B) - \rho_D(\frac{1}{2}A + \frac{1}{2}B)$$

Convex-Modularity indicates that the gap between $\rho_D(A \cup B)$ and $\rho_D(A) + \rho_D(B)$ is determined by the convex combinations of the menus. Given our interest in linear utilities, the choice of $\frac{1}{2}A + \frac{1}{2}B$ indicates indifference between A and B; hence any 'non-additivity' of ρ stems directly from indifferences.

Axiom 4—Linearity. *Let $A \subseteq D$. Then*

$$\rho_{\lambda D+z}(\lambda A + z) = \rho_D(A).$$

for all $\lambda > 0$ and $z \in \mathbb{R}^n$.

Linearity is standard.

Axiom 5—Mixture Continuity. *For $D, D' \in \mathcal{D}$, $\rho_{\lambda D+\lambda' D'}$ is continuous in λ, λ' for $\lambda, \lambda' \geqslant 0$.*

Mixture continuity is also standard. These five axioms are necessary and sufficient for the existence of a random linear representation.

4 Representation Results

Theorem 4.1. *The CC ρ satisfies Monotonicity, Extremeness, Convex-Modularity, Linearity, and Mixture-Continuity if and only if it maximizes a finitely additive RLR ξ.*

The proof of Theorem 4.1 explicitly constructs the measure ξ. As a preliminary, we show two key facts. The first is that ρ is completely determined by its value over singletons. Convex-Modularity places strict limits on the flexibility gained by allowing ρ to be non-additive; if $\rho_D(x)$, $\rho_D(y)$ and $\rho_D(\frac{1}{2}x + \frac{1}{2}y)$ are identified, then so too is $\rho_D(\{x, y\})$; Monotonicity allows us to add the necessary mixtures. Inductively, this determines all choice probabilities. The second fact, replicating sentiments from GP, is that whenever $N(D, \{x\}) = N(D', \{x'\})$ then $\rho_D(x) = \rho_{D'}(x')$.

Armed with these two observations, we construct the measure ξ. For technical reasons, we first identify the measure of the *relative interior* of each $N(D, \{x\})$, then appeal the extension theorems to complete the construction. We proceed inductively on the dimension of the relative interior. To illustrate this we will consider the menu shown in Figure 2. There is a single normal cone of dimension 0, namely $\mathbf{0}$. Since $x \in \text{int}(\text{conv}(D))$ we have $N(D, x) = \mathbf{0}$, so we can set $\xi(\mathbf{0}) = \rho_D(x)$.

Then, since $N(D', \{y\})$ is 1 dimensional, it's boundary is a 0 dimensional convex cone (hence $\mathbf{0}$). As such we can set $\xi(\text{ri}(N(D, \{y\}))) = \rho_D(y) - \rho_D(x)$. Moving up a level, we see that $N(D, \{w\})$ is 2 dimensional and its boundary consists is the union of all three previously identified sets. Therefore $\xi(\text{ri}(N(D, \{w\}))) = \rho_D(w) - \rho_D(y) - \rho_D(z) + \rho_D(x)$. Notice we must add back $\rho_D(x)$ as $N(D, \{x\}) = N(D, \{y\}) \cap N(D, \{z\})$ and was therefore subtracted off twice in prior steps. That this process is well defined and results in a measure representing all choice frequencies is a direct consequence of the above two observations.

5 The Many faces of ρ

In this section we briefly describe several different data generating processes that lead to (the identification of) a choice capacity.

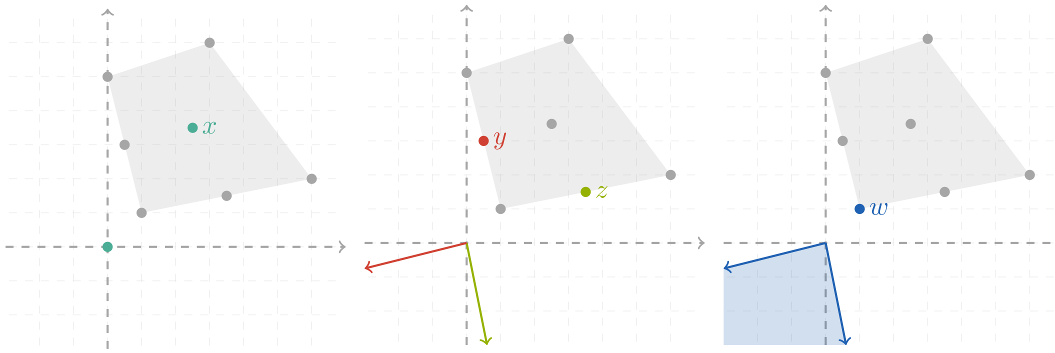

Figure 2: Each panel regards the same fixed menu D (in gray). The first panel $N(D, \{x\})$; the second $N(D, \{y\})$ (in red) and $N(D, \{z\})$ (in green); the third $N(D, \{w\})$.

5.1 Set Valued Choice

In some environments, a modeler might directly observe the entire set of maximizers associated with a decision problem. In other words, the data available to the modeler is the frequency with which each subset of D is chosen—a measure m_D over 2^D. For example, in digital markets where adding possible alternatives to an online cart is costless, consumers often add many potentially acceptable items to their carts, then make a choice from this set later (see Kukar-Kinney and Close (2010)). It is reasonable to view the cart as the acceptable set, and the final decision as a tie breaking procedure. If we take the observed measures $\{m_D\}_{D\in\mathcal{D}}$ as our primitive, we can construct a choice capacity, $\{\rho^m_D\}_{D\in\mathcal{D}}$, as follows:

$$\rho^m_D(A) = \sum_{\substack{B\in 2^D,\\ B\cap A\neq\varnothing}} m_D(B). \tag{5.1}$$

This construction is consistent, as given by the following result.

Theorem 5.1. *Take $\{m_D\}_{D\in\mathcal{D}}$. If ρ^m, as defined in (5.1), maximizes ξ, then*

$$m_D(A) = \xi\big(\{u \mid \arg\max_{z\in D} u(z) = A\}\big),$$

for all $D \in \mathcal{D}$.

So, understanding when a measure over subsets of alternatives arises from a RUM is as simple as constructing a choice capacity according to (5.1) and checking the axiomatic characterization below.

5.2 Status Quo

Often there is an exogenous default implemented in the case of indifference. For example, if the set of acceptable options includes the status quo, then the status quo is implemented. If our primitive observable data is a choice rule defined over a set *and* an observed status quo alternative, then variation in the default can identify a random choice capacity. Assume that this observable data is being generated by a RUM, such that whenever the status quo element is a maximizer, it is definitively chosen (i.e., irrespective of how indifference is broken in other choice problems). In particular: assume for each $x \in X$, and each choice problem D we observe an (additive) random choice rule, ρ^x_D from D under status quo x. Then we can recover a choice capacity as follows:

$$\rho^{sq}_D(\{x\}) = \rho^x_D(\{x\}), \tag{5.2}$$

Although 5.2 defines ρ^{sq}_D only when the choice is a singleton, it is sufficient to identify a unique choice capacity that satisfies our axioms. This result is formally captured by Lemma 3. This identified choice capacity will maximize the same RUM as the the random choice rules with status quo.

Theorem 5.2. *If $\{\rho^x_D\}_{x\in X, D\in\mathcal{D}}$ maximizes ξ for each x, then so too does the (unique) ρ^{sq} consistent with $\{\rho^x_D\}_{x\in X, D\in\mathcal{D}}$ as given by (5.2).*

Thus, even when the tie breaking rules are inconsistent or depend on the status quo alternative, identification is still immediate by filtering through our results. There are, of course, many other lexicographic costs for which the same argument could be applied in direct analogy. Take as another example: choice from a list. When there are multiple acceptable options, then the earliest such option is taken.

5.3 Non-Convergent Processes

Consider a population of agents, each of which has to choose a portfolio of risky assets each period: let $\rho^t_D(A)$ denote the empirical frequency with which elements of A were chosen in period t. The choice of each individual in each period depends on her personal risk preference, which is believed to be stationary in t. However, in the case of indifference many other parameters might influence tie breaking: current trends in investing, individual information regarding the state of the economy, transaction costs, etc. Thus, the dynamics of the population choices depends on the stochastic process of these parameters. Assume the modeler is interested in providing estimates to the distribution of the risk preference and thus wants to filter out the noise introduced by changes in other variables.

Depending on the stochastic process in the background, the empirical estimates of choice frequencies in period t, $\rho^t_D(A)$, may not converge. Let $\rho_D(A)$ be the *limsup* of this frequency. If the choice-determining stochastic process is eventually stationary, then the frequency of choosing each option converges, and ρ is a classical additive RUM. Otherwise, ρ is a choice capacity. Nonetheless, our above analysis provides a method of understanding the distribution of risk preferences when ρ is identified only as a capacity.

5.4 Sets of Random Choice Rules

In many situations, a modeler will collect data from distinct populations and compare statistical properties of the different populations. If our primitive observable data is a *set* of (additive) random choice rules, then we can construct a choice capacity by taking the upper-bound across the measures.[5] When will this choice capacity satisfy our axiomatic restrictions? As was suggested by our initial motivation for the non-additivity of ρ, this is the case exactly when the differences across the set of random choice rules arise entirely from differential tie breaking procedures.

Thus, the following result justifies our initial interpretation and reveals a considerable amount about the underlying anatomy of the capacities ρ. The structure imposed on ρ_D—which was both necessary and sufficient for a RLR—implies that it is a *coherent upper probability*: there exists a set of probability measures $M \subseteq \Delta(D)$ such that

$$\rho_D(A) = \sup_{m \in M} m(A),$$

for all $A \subseteq D$. Even more revealing is that the set M has a tight characterization. Each m in M is the composition of ξ and a tie breaking rule. Towards making this definite, for any RLR, ξ, define the set of measures

$$M(\xi, D) = \left\{ \int_{R^n} \tau_u(A)\xi(du) \mid \tau_u \in \Delta(\mathbb{R}^n),\ \mathrm{supp}(\tau_u) = \arg\max_{y \in D} u(y) \right\},$$

where $\mathrm{supp}(\tau)$ is the support of the measure τ.

Theorem 5.3. *Let ρ maximize ξ. Then $\rho_D = \sup_{m \in M(\xi,D)} m$ for all D.*

In other words, the capacity ρ is the upper bound of all measures over D which are constructed by first choosing a utility u according to ξ, and subsequently choosing among the maximizers in D according to some tie breaking procedure. Theorem 5.3 follows very directly from work done on belief and plausibility functions (Dempster, 1967; Wasserman, 1990). Indeed, the characterization of ρ_D via $M(\xi, D)$ can be made generally for plausibility functions and our notion of maximization according to ξ is essentially the *definition* of a plausibility function.

A Proofs

A.1 An Onslaught of Definitions

Unfortunately, we need to define a bunch of objects. If A is a convex set and $\mathrm{ext}(A)$ is finite then A is a called at polytope. For a polytope A, let $F \subset A$ be called a face if whenever $\alpha x + (1-\alpha)y \in F$ then also $x, y \in F$. Let $F(A)$ denote the set of all (non-empty) faces of A and $F^0(A) = \{\mathrm{ri}(F) \mid F \in F(A)\}$. It is well known that $F(A)$ is finite and $F^0(A)$ is a partition of A (theorems 19.1 and 18.2 of Rockafellar (1970), respectively). A face $F \in F(A)$ is called exposed if it is the intersection of A with a supporting hyperplane, or, equivalently, if $F = M(A, u)$ for some $u \in \mathbb{R}^n \backslash \mathbf{0}$. Every proper face (i.e., $F \in F(A)$, $F \neq A$) is an exposed face (Corollary 2.4.2 Schneider (2014)).

[5]This is, in fact, the content of the *limsup* operation in the pervious subsection

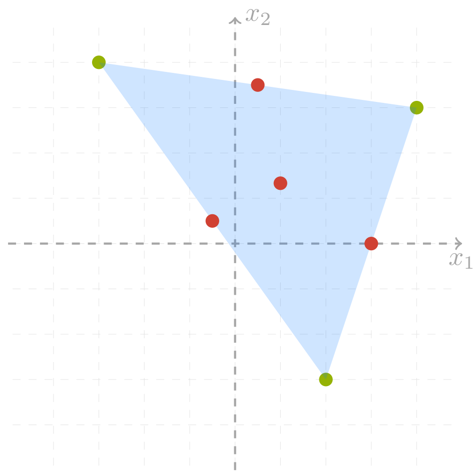

Figure 3: The set A of affinely independent coordinates is given by the green points. The red points are the additional elements of A^*. See, it even looks like a little $*$.

If $\lambda K \subseteq K$ for all $\lambda \geqslant 0$ then K is called a *cone*. We say a cone K is generated by A if $K = \{\lambda x \mid x \in A, \lambda \geqslant 0\}$. A cone K is *polyhedral* if it is generated by a polytope; let $\mathcal{K}$ denote all such cones. Let $\mathcal{K}^*$ denote the set of pointed polyhedral cones, those cones with $\mathbf{0} \in \mathrm{ext}(K)$. The face of a polyhedral cone is a polyhedral cone. By proposition 4 of Gul and Pesendorfer (2006), Ω is the algebra generated by $\mathcal{K}^*$.

Let $CB = \{x \in \mathbb{R}^n \mid |x^i| = 1 \text{ for some } i, x^j = 0, j \neq i\}$ denote an n dimensional cube. It is true that $\bigcup_{x \in CB} N(CB, x) = \mathbb{R}^n$.

If $A = \{x_1, \ldots x_k\}$ is a set of affinely-independent points then let

$$A^* = \bigcup_{I \subseteq \{1 \ldots k\}} \sum_{i \in I} \frac{x_i}{|I|}$$

The set A^* is a decision problem that has A as the set of extreme points, and contains a point in the relative interior of every face of the decision problem.

A.2 Lemmas

Lemma 1. *Let D be a polytope and $A \subseteq D$. Then $pi(A, D) = \bigcup_{\{F \in F(D) | F \cap A \neq \varnothing\}} \mathrm{ri}(F)$.*

Proof. Let $x \in \mathrm{pi}(D, A)$. That $A \subseteq \bigcup_{\{F \in F(D) | F \cap A \neq \varnothing\}} \mathrm{ri}(F)$ follows from the fact that $F^0(D)$ partitions D. So, take $x = \alpha a + (1 - \alpha) y$ with $a \in A$, $y \in D$ and $\alpha \in (0, 1)$. Again, since $F^0(A)$ partitions D, we have that $x \in \mathrm{ri}(F)$ for some $F \in F(D)$. Moreover, since F is a face, by definition $\{a, y\} \subset F$, so $F \cap A \neq \varnothing$.

Towards the other inclusion, let $x \in \mathrm{ri}(F)$ for some $F \in F(D)$ with $F \cap A \neq \varnothing$. Let $a \in F \cap A$. Since $\mathrm{ri}(F)$ is convex and relatively open, and $a \in \mathrm{cl}(\mathrm{ri}(F))$, we can write $x = \alpha a + (1 - \alpha) y$ for some $y \in \mathrm{ri}(F)$. Hence $x \in \mathrm{pi}(D, A)$. ∎

Lemma 2. *If $N(D,A) \subseteq N(D,B)$, then $A \in \mathrm{pi}(D,B)$.*

Proof. Let $x \in A$. If $x \in \mathrm{ri}(D)$ then the claim holds immediately ($\mathrm{ri}(D) \subseteq \mathrm{pi}(D,x)$ for all x, per Lemma 1). So assume $x \in \mathrm{ri}(F)$ for some $F \in F(D)$, $F \neq D$. Now every proper face of D is an exposed face, so let u be such that $M(D,u) = F$. Thus, $u \in N(D,A) \subseteq N(D,B) = \bigcup_{y \in B} N(D,y)$; we have $u \in N(D,y)$ for some $y \in B$. This indicates, $y \in M(D,u) = F$ and therefore, by Lemma 1, we have $x \in \mathrm{ri}(F) \subseteq \mathrm{pi}(D,B)$. ∎

We now show that if D is a choice problem then the entirety of ρ_D depends only on the value of singletons in D^*. In other words, if we know the value of $\rho_{D^*}(x)$ for all $x \in D^*$ then we know all choice probabilities out of D.

Lemma 3. *ρ_D is uniquely determined by $\{\rho_{D^*}(x) \mid x \in D^*\}$.*

Proof. We will prove $\rho_D(A)$ is identified by induction on the cardinality $A \subseteq D$. Let $A = \{x\}$. Since D and D^* have the same extreme points, Monotonicity states that $\rho_D(x) = \rho_{D^*}(x)$.

Now, assume this was the case for all sets with n or fewer elements, and let $|A| = n+1$. Then $A = B \cup \{x\}$ for some B with $|B| = n$. Monotonicity states that $\rho_D(A) = \rho_{D^*}(A)$. Moreover, notice that $(\frac{1}{2}B + \frac{1}{2}x) \subseteq D^*$ by construction. Therefore, appealing to Convex-Modularity delivers,

$$\rho_D(A) = \rho_{D^*}(A) = \rho_{D^*}(B) + \rho_{D^*}(x) - \rho_{D^*}(\frac{1}{2}B + \frac{1}{2}x);$$

each set in question is identified by the inductive hypothesis. ∎

Lemma 4. *If $N(D,\{x\}) = N(D',\{x'\})$ then $\rho_D(x) = \rho_{D'}(x')$.*

Proof. Lemma 1 of Gul and Pesendorfer (2006). ∎

Lemma 5. *Let $A_1 \dots A_k \subseteq D$. Then*

$$\rho_{D^*}(\bigcup_{i \leqslant k} A_i) = \sum_{I \subseteq \{1\dots k\}} (-1)^{|I|+1} \rho_{D^*}\left(\sum_{i \in I} \frac{A_i}{|I|}\right).$$

Proof. Follows inductively from successive applications of Convex-Modularity. Notice that for some face F and strictly positive $\alpha_1 \dots \alpha_k$ summing to 1, we have $\alpha_1 x_1 + \dots + \alpha_k x_k \in \mathrm{ri}(F)$ if and only if $\frac{1}{k}x_1 + \dots + \frac{1}{k}x_k \in \mathrm{ri}(F)$. ∎

Lemma 6. *Let $K_0, K_1 \dots K_k \in \mathcal{K}^*$ be such that $\mathrm{ri}(K_0) = \bigcup_{i=1\dots k} \mathrm{ri}(K_i)$ and $\mathrm{ri}(K_i) \cap \mathrm{ri}(K_j) = \varnothing$ for $i \neq j \neq 0$. Then (i) $K_0 = \bigcup_{i=1\dots k} K_i$ and (ii)*

$$\sum_{\substack{F \in F(K_0),\\ F \neq K_0}} \mathbb{1}(\mathrm{ri}(F)) = \sum_{1 \leqslant i \leqslant k} \sum_{\substack{F \in F(K_i),\\ F \neq K_i}} \mathbb{1}(\mathrm{ri}(F)) + \sum_{\substack{I \subseteq \{1\dots k\}\\ |I| \geqslant 2}} (-1)^{|I|+1} \mathbb{1}(\bigcap_I K_i)$$

where $\mathbb{1}$ is the indicator function $\mathbb{R}^n \to \mathbb{R}$ taking a value of 1 on the indicated set and 0 elsewhere.

Proof. Twoards (i) Since, $\mathrm{ri}(K_i) \subset \mathrm{ri}(K_0)$ it follows directly that $K_i \subset K_0$ for all i. Thus we need only show that for all $x \in K_0$, $x \in K_i$ for some k. Take $\{x_m\}_{m\in\mathbb{N}} \subset \mathrm{ri}(K_0)$ converging to x. Then there is a subsequence (without relabeling) such that $x_m \in K_i$ for all m (since there are only finitely many K_i). But this subsequence converges to x, so by the fact that K_i is closed, $x \in K_i$ and we are done. Claim (ii) follows directly. ∎

A.3 Proof of Theorem 4.1

Necessity. The necessity of Monotonicity, Linearity, and Mixture-Continuity are essentially the same as in GP. We will use the facts that $N(D, A \cup B) = N(D, A) \cup N(D, B)$ and $N(D, A) \cap N(D, B) = N(D, \alpha A + (1 - \alpha)B)$ for $\alpha \in (0, 1)$.

From the definition of pi and these properties of normal cones, we have that

$$N(D, A) \subseteq N(D, \mathrm{pi}(D, A)) = \bigcup_{y\in D} (N(D, A) \cap N(D, y)) \subseteq N(D, A).$$

Extremeness follows.

These same properties also imply $N(D, A \cup B)$ is equal to the (piecewise disjoint) expression

$$\Big(N(D, A)\backslash N(D, \frac{1}{2}A + \frac{1}{2}B)\Big) \cup \Big(N(D, B)\backslash N(D, \frac{1}{2}A + \frac{1}{2}B)\Big) \cup N(D, \frac{1}{2}A + \frac{1}{2}B),$$

indicating that Convex-Modularity must hold.

Sufficiency. First, define $\Omega^0 = \{\mathrm{ri}(K) | K \in \mathcal{K}\}$. GP show that Ω^0 is a semi-ring and that Ω can be reclaimed by taking finite disjoint unions over Ω^0. Thus it suffices to define a finitely additive ξ over Ω^0, as it will extend uniquely to Ω.

We will construct ξ inductively on the dimension of K. Let $\xi(\varnothing) = 0$. Let K be 0 dimensional so that $K = \mathrm{ri}(K) = \{\mathbf{0}\}$. We have that $\mathbf{0} = N(CB^*, \mathbf{0})$. Set

$$\xi(\mathbf{0}) = \rho_{CB*}(\mathbf{0}).$$

Lemma 4 ensures this is well defined. Now assume that this process has been completed for all K with dimension k or less.

Consider a K of dimension $k+1$. By Proposition 4 of GP, $K = N(D, x)$ for some (D, x), with Lemma 4 ensuring it does not matter which such (D, x) we choose. Set

$$\xi(\mathrm{ri}(K)) = \rho_D(x) - \sum_{F\in F(K), F\neq K} \xi(\mathrm{ri}(F)), \tag{A.1}$$

where the latter is previously set by the inductive hypothesis and the fact that for all $F \in F(K)$, $F \neq K$, F is a polyhedral cone such that $dim(F) < k + 1$ (corollary 18.1.3 of Rockafellar (1970)).

Lemma 7. *ξ is finitely additive.*

Proof. We will prove the claim by induction on the dimension of the sets in question. When $dim(K) = 0$ there is a single convex cone, to wit, $\mathbf{0}$, so the claim hold trivially. Assume that ξ is finitely additive over any sets of whose union is of dimension m or less. Let

$K_0, K_1 \dots K_k \in \mathcal{K}^*$ be such that $\text{ri}(K_0) = \bigcup_{i=1\dots k} \text{ri}(K_i)$ and $\text{ri}(K_i) \cap \text{ri}(K_j) = \varnothing$ for $i \neq j \neq 0$, with K_0 of dimension $m+1$. From Lemma 6, $K_0 = \bigcup_{i=1\dots k} K_i$.

For the first half of the claim, we will follow the general logic of GP's lemma 4. By Proposition 4 of GP, we can find $D_i \in \mathcal{D}$ and $x_i \in D_i$ such that $K_i = N(D_i, x_i)$ for $i = 0 \dots m$. Let $D = D_0 + D_1 + \dots + D_k$. For $y \in \bigcup_{j=0}^{k} D_j$, construct the sets:

$$Z(y) = \left\{ z = (z^0 \dots z^k) \in \prod_{j=0}^{k} D_j \mid z^j = y, \text{ for some } j \right\}$$

and

$$G(y) = \left\{ y' \in D \mid y' = \sum_{j=0}^{k} z^j, z \in Z(y) \right\}.$$

Using Mixture-Continuity, GP show that $\rho_D(G(y)) = \rho_{D_i}(y)$.

Now, by construction $N(D, G(x_0)) = N(D, \bigcup_{i \leqslant k} G(x_i))$. By Lemma 2 this implies $G(x_0) \subseteq \text{pi}(D, \bigcup_{i \leqslant k} G(x_i))$ and $\bigcup_{i \leqslant k} G(x_i) \subseteq \text{pi}(D, G(x_0)$. Therefore, by Extremeness, $\rho_D(G(x_0)) = \rho_D(\bigcup_{i \leqslant k} G(x_i))$. Now this implies, by the construction of ξ, via (A.1),

$$\begin{aligned} \xi(\text{ri}(K_0)) &= \rho_{D_0}(x_0) - \sum_{\substack{F \in F(K_0), \\ F \neq K_0}} \xi(\text{ri}(F)) \\ &= \rho_D(G(x_0)) - \sum_{\substack{F \in F(K_0), \\ F \neq K_0}} \xi(\text{ri}(F)) \\ &= \rho_D(\bigcup_{i \leqslant k} G(x_i)) - \sum_{\substack{F \in F(K_0), \\ F \neq K_0}} \xi(\text{ri}(F)). \end{aligned} \tag{A.2}$$

Appealing to Lemma 5, we can rewrite (A.2):

$$\xi(\mathrm{ri}(K_0)) = \sum_{I\subseteq\{1\ldots k\}} (-1)^{|I|+1}\rho_{D*}\left(\sum_{i\in I}\frac{G(x_i)}{|I|}\right) - \sum_{\substack{F\in F(K_0),\\ F\neq K_0}} \xi(\mathrm{ri}(F))$$

$$= \sum_k \rho_D(G(x_i)) + \sum_{\substack{I\subseteq\{1\ldots k\}\\ |I|\geqslant 2}} (-1)^{|I|+1}\rho_{D*}\left(\sum_{i\in I}\frac{G(x_i)}{|I|}\right) - \sum_{\substack{F\in F(K_0),\\ F\neq K_0}} \xi(\mathrm{ri}(F)) \tag{A.3}$$

$$= \sum_k \xi(\mathrm{ri}(K_i)) + \sum_k \sum_{\substack{F\in F(K_i),\\ F\neq K_i}} \xi(\mathrm{ri}(F)) +$$

$$+ \sum_{\substack{I\subseteq\{1\ldots k\}\\ |I|\geqslant 2}} (-1)^{|I|+1}\rho_{D*}\left(\sum_{i\in I}\frac{G(x_i)}{|I|}\right) - \sum_{\substack{F\in F(K_0),\\ F\neq K_0}} \xi(\mathrm{ri}(F)) \tag{A.4}$$

$$= \sum_k \xi(\mathrm{ri}(K_i)) + \sum_k \sum_{\substack{F\in F(K_i),\\ F\neq K_i}} \xi(\mathrm{ri}(F)) +$$

$$+ \sum_{\substack{I\subseteq\{1\ldots k\}\\ |I|\geqslant 2}} (-1)^{|I|+1}\xi\left(\bigcap_I K_i\right) - \sum_{\substack{F\in F(K_0),\\ F\neq K_0}} \xi(\mathrm{ri}(F)). \tag{A.5}$$

Notice the equality between (A.3) and (A.4) appeals again to the construction of ξ, via (A.1), and between (A.4) and (A.5) appeals to the well know fact that $N(D, \alpha x + (1-\alpha y)) = N(D,x) \cap N(D,y)$ for $\alpha \in (0,1)$.

Finally, notice that Lemma 6 indicates that

$$\sum_{1\leqslant i\leqslant k} \sum_{\substack{F\in F(K_i),\\ F\neq K_i}} \mathbb{1}(\mathrm{ri}(F)) + \sum_{\substack{I\subset\{1\ldots k\}\\ |I|\geqslant 2}} (-1)^{|I|+1}\mathbb{1}\left(\bigcap_I K_i\right) - \sum_{\substack{F\in F(K_0),\\ F\neq K_0}} \mathbb{1}(\mathrm{ri}(F)) = 0$$

so, by the inductive hypothesis that ξ is additive over such a domain,[6] we can conclude that the sum of all but the first term of (A.5) equals 0, so that $\sum_k \xi(\mathrm{ri}(K_i)) = \xi(\mathrm{ri}(K_0))$. ∎

Extend ξ from Ω^0 to Ω in the usual way. Since $\bigcup_{y\in CB} N(CB,y) = \mathbb{R}^n$ and $\rho_{CB}(CB) = 1$ we have that ξ is a finitely additive measure.

Lemma 8. *ρ maximizes ξ.*

Proof. Consider (D,x) with $dim(D) = n$ so that $N(D,x) = K \in \mathcal{K}$. Recall, that (i) $F^0(K)$ partitions K and (ii) $K \in F(K)$. Therefore, we have

$$\rho_D(x) = \xi(\mathrm{ri}(K)) + \sum_{F\in F(K), F\neq K} \xi(\mathrm{ri}(F)) = \sum_{A\in F^0(K)} \xi(A) = \xi(N(D,x)).$$

[6]Notice that all sets in question are subsets of the boundary of cones themselves of dimension at most $m+1$. Further, while the value of ξ was not explicitly defined on the cones of dimension m or less, the fact that such objects are partitioned into relative interiors of faces, and the inductive hypothesis of additivity, indicates that ξ is implicitly defined over such objects.

By Lemma 3, the entirety of ρ_D is determined by ρ's value on singletons, hence ρ maximizes ξ on all n dimensional problems.

Let D of dimension less than n, and consider $D+\alpha CB$, the dimension of the later object is n. We have

$$\begin{aligned}\rho_{D+\alpha CB}(x+\alpha CB) &= \xi(N(D+\alpha CB, x+\alpha CB)) \\ &= \xi(\bigcup_{y\in CB} N(D+\alpha CB, x+\alpha y)) \\ &= \xi(\bigcup_{y\in CB} (N(D,x)\cap N(CB,y)) \\ &= \xi(N(D,x)).\end{aligned}$$

The first equality follows the fact that ρ maximizes ξ for n dimensional problems, the second from the definition of $N(D,A)$, the third from properties of normal cones, and the final equality from the fact that $\bigcup_{y\in CB} N(CB,y) = \mathbb{R}^n$. Appealing to Mixture-Continuity—letting α tend to 0—we conclude that $\rho_D(x) = \xi(N(D,x))$, as desired. ∎

A.4 Proofs of other Theorems

Proof of Theorem 5.1. We can rewrite (5.1) as

$$\rho_D(A) = 1 - \sum_{\substack{B\in 2^D,\\ B\subseteq A^c}} m_D(B).$$

The proof in by structural induction in the number of elements in A. Let $A=\{x\}$. Then

$$m_D(\{x\}) = 1-\rho^m_D(A^c) = 1-\xi\big(\{u \mid \arg\max_{z\in D} u(z)\cap A^c \neq \varnothing\}\big),$$

which, of course, is exactly the probability of drawing a u that is maximized uniquely by x. Now, assume this holds for all A with n or fewer elements. Then,

$$m_D(A) = 1-\rho^m_D(A^c) - \sum_{\substack{B\in 2^D,\\ B\subsetneq A}} m_D(B),$$

which by our inductive hypothesis and the definition of ξ is equal to

$$1-\xi\big(\{u \mid \arg\max_{z\in D} u(z)\cap A^c \neq \varnothing\}\big) - \sum_{\substack{B\in 2^D,\\ B\subsetneq A}} \xi\big(\{u \mid \arg\max_{z\in D} = B\}\big)$$

And, as desired, this last line is the 1 minus the probability that the set of maximizers is larger than A, minus the probability the set is smaller than A. ∎

Proof of Theorem 5.2. First, by Lemma 3, (5.2) suffices to identify all of ρ, so there is a unique choice capacity. Second, when x is the status quo, it will be chosen whenever a u is realized that is maximized at x. But the probability of this is exactly $\xi\big(\{u \mid \arg\max_{z\in D} u(z)\cap\{x\}\neq\varnothing\}\big)$. ∎

Proof of Theorem 5.3. The definition of ρ maximizing ξ indicates that ρ is a plausibility function according to Dempster (1967). The theorem in question follows as a direct corollary of Theorem 2.1 of Wasserman (1990). ∎

References

Jose Apesteguia, Miguel A Ballester, and Jay Lu. Single-crossing random utility models. *Econometrica*, 85(2):661–674, 2017.

Arthur P Dempster. Upper and lower probabilities induced by a multivalued mapping. *The annals of mathematical statistics*, pages 325–339, 1967.

Mira Frick, Ryota Iijima, and Tomasz Strzalecki. Dynamic random utility. 2017.

Faruk Gul and Wolfgang Pesendorfer. Random expected utility. *Econometrica*, 74(1):121–146, 2006.

Monika Kukar-Kinney and Angeline G Close. The determinants of consumers? online shopping cart abandonment. *Journal of the Academy of Marketing Science*, 38(2):240–250, 2010.

Jay Lu. Random choice and private information. *Econometrica*, 84(6):1983–2027, 2016a.

Jay Lu. Supplement to ?random choice and private information? *Econometrica*, 84(6): 1983–2027, 2016b.

R Tyrrell Rockafellar. *Convex Analysis*, volume 28. Princeton University Press, 1970.

Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory.* Number 151. Cambridge university press, 2014.

Larry Alan Wasserman. Prior envelopes based on belief functions. *The Annals of Statistics*, pages 454–464, 1990.