Axiomatization of the Mixed Logit Model

Kota Saito*
California Institute of Technology

June 17, 2018

Abstract

A mixed logit function, also known as a random-coefficient logit function, is an integral of logit functions. Necessary and sufficient conditions are provided under which a random choice function can be represented as a mixed logit function. The axioms are based on the social surplus function proposed by McFadden (1978, 1981).

Keywords: Random choice, mixed logit, random coefficients.

1 Introduction

The purpose of this paper is to provide axiomatizations of the mixed logit model, also known as the random-coefficient logit model. The mixed logit model is one of the most widely used models in the analysis of discrete choice for studying aggregated demand across consumers, especially in the empirical literature on marketing, industrial organization, and public economics.

In this paper, the observed behavior is described by a random choice function ρ, which assigns to each choice set D a probability distribution over D. The number $\rho(D, x)$ is the probability that an alternative x is chosen from a choice set D.

*This paper was first presented at the University of Tokyo on July 29, 2017. An earlier version of this paper was posted on September 15, 2017. See http://www.hss.caltech.edu/content/axiomatizations-mixed-logit-model. I appreciate the valuable discussions I had with Brendan Beare, Kim Border, Yimeng Li, Jay Lu, and Matt Shum. Jay Lu also read an earlier version of the manuscript and offered helpful comments. This research is supported by Grant SES1558757 from the National Science Foundation.

As in empirical analysis, an alternative x is identified by a real vector of explanatory variables of the alternative. The random choice function describes aggregate choices across a population of individuals. The aggregate choices are random due to unobserved heterogeneity across the individuals.

The function ρ is called a *mixed logit function* if there exists a probability measure m such that

$$\rho(D, x) = \int \frac{\exp(\beta \cdot p(x))}{\sum_{y \in D} \exp(\beta \cdot p(y))} dm(\beta), \tag{1}$$

where $\beta \cdot p(x)$ is a polynomial of x. The probability distribution m captures the unobserved heterogeneity across the population of individuals. Each logit function in the support of m describes aggregate choices in a subpopulation. This paper provides not only an axiomatization of the model (1) but also its special case with linear p (i.e., $\beta \cdot p(x) = \beta \cdot x$) and its general case with an arbitrary function $u(\cdot)$ in place of a polynomial $\beta \cdot p(\cdot)$.

The first axiom in the paper may be seen as a normative one. To define a normative requirement, I consider a representative agent whose random choice is described by ρ. Then I compare the representative agent's random choice with deterministically rational choices as benchmarks. As a criterion for the comparison, I adopt the concept of the *social surplus function* proposed by McFadden (1978, 1981). Given a utility function u of the representative agent, McFadden (1978, 1981) defines the *social surplus*, denoted by $G(\rho : u)$, as the expected utility of the representative agent whose random choice is described by ρ.[1]

From our viewpoint as outside observers, the utility function u of the representative agent is unobservable. The axiom requires that no matter which utility function the outside observer uses, *the social surplus obtained by the representative agent's choice should be larger than the minimum social surplus obtained by deterministically rational choices.* The requirement of the axiom is weak in the sense that the axiom does *not* require that the agent's random choice dominate the deterministically rational choices; the axiom only requires that the agent's random choice should be better than the *worst* deterministically rational choices in terms of the social surplus. Theorem 1 in section 3 states that, under an assumption about the set of alternatives that can be shown to hold generically, this axiom is necessary and sufficient for a random choice function to be represented as a mixed logit function.

I also provide an alternative axiom, which may be considered as a descriptive

[1]See (3) in section 3 for the definition of G.

one. Since the social surplus is the expected utility obtained by random choice ρ, it follows that the social surplus function $G(\rho : u)$ is linear in ρ. This implies that the social surplus of a mixed logit function is an average surplus over a population of individuals. Consequently, the social surplus must be larger than the smallest surplus among the subpopulations (i.e., $G(\rho : u) = \int G(\rho_l : u)dm \geq \inf_{\rho_l} G(\rho_l : u)$). Corollary 1 in section 3.1 states that, under the generic property on the set of alternatives, a slightly stronger condition is not only necessary but also sufficient for a random choice function to be represented as a mixed logit function.

In the course of proving the axiomatizations, I have obtained several results which could be of interest by themselves. First, generically speaking, any interior random utility function can be represented as a convex combination of logit functions with polynomials of at most degree d if and only if d is larger than a threshold. The threshold can be calculated explicitly from the number of explanatory variables and the number of all alternatives.[2] See Proposition 1 in section 3 and Corollary 4 in section 4 for details.

Second, in Proposition 2 in section 3, I show that the affine hull of the set of random utility functions contains the set of random choice functions. As I show in Corollary 6 in section 5, this result together with Proposition 1 implies that any interior random *choice* function is generically represented as an affine combination of two mixed logit functions.

No axiomatic characterizations for the mixed logit model have yet been provided, to my knowledge. However, other generalizations of the logit model have been axiomatized recently. Gul et al. (2014) axiomatize a model called the *complete attribute rule*, which is similar to the nested logit model. By using a model of rational inattention, Matějka and McKay (2015) provide a novel characterization of a generalization of the logit model. In a dynamic setup, Fudenberg and Strzalecki (2015) axiomatize a generalization of the discounted logit model which incorporates a parameter to capture an agent's costs and benefits of choosing from larger choice sets. By using this parameter, their model can succinctly capture both a preference for flexibility as well as the phenomenon of *choice aversion*. Echenique and Saito (2015), Ahumada and Ulku (2017), Horan (2018), and Cerreia-Vioglio et al. (2018) axiomatize generalizations of the logit model which allow zero-probability choices. Moreover, Cerreia-Vioglio et al. (2018) axiomatize another generalization of the

[2] To calculate the threshold explicitly, let K be the number of explanatory variables and X be the set of all alternatives. Then the threshold is a minimal positive integer d such that $\binom{d+K}{K} \geq |X|$.

logit model in which the systematic part of utility is time-independent but the shock component is time-dependent. This dependence is crucial in quantal response equilibrium theory as well as in neuroscience.

While the papers mentioned above provide different generalizations of the logit model, other recent models of random choice have been proposed that are not variations of the logit model. For example, Gul and Pesendorfer (2006), Lu (2016), and Lu and Saito (2016) axiomatize variations of the random utility model. Fudenberg et al. (2015) and Cerreia-Vioglio et al. (2017) axiomatize models in which an agent deliberately randomizes choice. See Strzalecki (2018) for a recent extensive survey on the literature of random choice.

In the next section, I introduce the models formally. In section 3, I provide the axiomatizations of the mixed logit model. In section 4, I provide results on the denseness of the mixed logit model in the random utility model. In section 5, I state corollaries and lemmas which I obtain in the course of proving the axiomatizations.

2 Model

The set of all alternatives is denoted by X. In the analysis of discrete choice, the number of alternatives in a choice set is finite. Since the number of choice sets is usually finite, X is finite. An alternative x can be identified by a real vector of explanatory variables of x.[3] For example, if an alternative is a consumption good, the alternative can be identified by its price and various measures of its quality. Hence, X is a finite subset of $\mathbf{R}^K$, where K is the number of the explanatory variables. For each $x \in X$ and $k \in \{1, \ldots, K\}$, I write $x(k)$ to denote the k-th element of x.[4]

Let $\mathcal{D} \subset 2^X \setminus \{\emptyset\}$. $\mathcal{D}$ is the set of choice sets. Notice $\mathcal{D}$ can be a proper subset of $2^X \setminus \emptyset$.

Definition 1. *A function $\rho : \mathcal{D} \times X \to [0, 1]$ is called a* random choice function *if $\sum_{x \in D} \rho(D, x) = 1$ and $\rho(D, x) = 0$ for any $x \notin D$. The set of random choice functions is denoted by $\mathcal{P}$.*

For each $(D, x) \in \mathcal{D} \times X$, the number $\rho(D, x)$ is the probability that an alternative x is chosen from a choice set D. A random choice function ρ is an element of

[3] An empirical researcher can include 1 as an explanatory variable if he wants to use a constant term.

[4] In section 3.2, where I axiomatize a generalization of the mixed logit model, the set X does not have to be a subset of $\mathbf{R}^K$.

$\mathbf{R}^{\mathcal{D}\times X}$. The random choice function is the observable choice data in this paper. I interpret the random choice function ρ as aggregate choices across individuals. The choices are random due to unobserved heterogeneity across individuals.

Definition 2. *A random choice function ρ is called a* mixed logit function *if there exists a positive integer d and a probability measure m such that for all $(D, x) \in \mathcal{D} \times X$, if $x \in D$, then*

$$\rho(D, x) = \int \frac{\exp(\beta \cdot p_d(x))}{\sum_{y \in D} \exp(\beta \cdot p_d(y))} dm(\beta), \tag{2}$$

where $\beta \cdot p_d(x)$ is a polynomial of at most degree d. Given a positive integer d, the function ρ defined by (2) with a measure m is called a mixed logit function with polynomials of at most degree *d.*

The set of mixed logit functions is denoted by $\mathcal{P}_{ml}$. Given a positive integer d, the set of mixed logit functions with polynomials of at most degree d is denoted by $\mathcal{P}_{ml}(d)$. Thus $\mathcal{P}_{ml} = \bigcup_{d \in \mathbf{Z}_+} \mathcal{P}_{ml}(d)$, where $\mathbf{Z}_+$ is the set of positive integers.

When m is degenerate (that is, $m = \delta_\beta$ for some β) in (2), then ρ is called a *logit function.* Given a positive integer d, ρ defined by (2) with a degenerate measure m is called a *logit function with polynomials of at most degree d.* The set of logit functions is denoted by $\mathcal{P}_l$. Given a positive integer d, the set of logit functions with polynomials of at most degree d is denoted by $\mathcal{P}_l(d)$. Note that $\mathcal{P}_l = \bigcup_{d \in \mathbf{Z}_+} \mathcal{P}_l(d)$.

Given a positive integer d and $x \in X$, the vector $p_d(x)$ consists of *monomials* of at most degree d (i.e., higher order terms such as $x(k)^n$ where $n \leq d$, and interaction terms such as $\prod_{k=1}^{K} x(k)^{n_k}$, where $\sum_{k=1}^{K} n_k \leq d$).[5] In some results of the paper I consider the linear case in which $d = 1$ (i.e., $p_d(x) = x$). However, in an empirical analysis, such a linear relationship may not hold. For example, if an empirical researcher is modeling demand for a product in terms of consumers' income, one may find that the income elasticity of demand is not a linear function of the level of income. In that case, the researcher may want to include higher order terms of income. Moreover, the effect of one explanatory variable often depends on another explanatory variable. For example, the effect of income on the elasticity of demand may depend on age groups. One way to deal with such dependencies is to include an interaction term among explanatory variables, such as income and an index of age group.

[5] For example, if $K = 2$ and $d = 2$, then $p(x) = (x(1), x(2), x(1)^2, x(1)x(2), x(2)^2)$ where $x = (x(1), x(2))$.

The next result simplifies our analysis.

Remark 1. *For any positive integer d, $\mathcal{P}_{ml}(d) = co.\mathcal{P}_l(d)$ (i.e., the set of mixed logit functions with polynomials of at most degree d is the convex hull of the set of logit functions with polynomials of at most degree d).*

The remark implies that $\mathcal{P}_{ml} = \text{co}.\mathcal{P}_l$ (i.e., the set of mixed logit functions equals the set of convex combinations of logit functions). Thus, to axiomatize the mixed logit model, it is necessary and sufficient to axiomatize the convex hull of the set of logit functions.[6]

In section 3.2, I axiomatize a generalization of the mixed logit model defined with an arbitrary function $u \in \mathbf{R}^X$ in place of polynomials as follows.

Definition 3. *A random choice function ρ is called a* general mixed logit function *if there exists a probability measure m on $\mathbf{R}^X$ such that for all $(D, x) \in \mathcal{D} \times X$, if $x \in D$, then*

$$\rho(D, x) = \int \frac{\exp(u(x))}{\sum_{y \in D} \exp(u(y))} dm(u). \tag{3}$$

When m is degenerate (that is, $m = \delta_u$ for some u), then ρ is called a general logit function.

For the axiomatization of the general model above, X need *not* be a subset of finite dimensional real space as long as the number of elements in X is finite.

In the following, I introduce several definitions. Let Π be the set of bijections between X and $\{1, \ldots, |X|\}$, where $|X|$ is the number of elements of X. If $\pi(x) = i$, then I interpret x to be the $|X| + 1 - i$-th best element of X with respect to π. If $\pi(x) > \pi(y)$, then x is better than y with respect to π. An element π of Π is called a *strict preference ranking* (or simply, a *ranking*) over X. For all $(D, x) \in \mathcal{D} \times X$ such that $x \in D$, if $\pi(x) > \pi(y)$ for all $y \in D \setminus \{x\}$, then I often write $\pi(x) \geq \pi(D)$.

There are $|X|!$ elements in Π. I denote the set of probability measures over Π by $\Delta(\Pi)$. Since Π is finite, it follows that $\Delta(\Pi) = \left\{(\nu_1, \ldots, \nu_{|\Pi|}) \in \mathbf{R}_+^{|\Pi|} \,\middle|\, \sum_{i=1}^{|\Pi|} \nu_i = 1\right\}$, where $\mathbf{R}_+$ is the set of nonnegative real numbers.

Definition 4. *A random choice function ρ is called a* random utility function *if there exists a probability measure $\nu \in \Delta(\Pi)$ such that for all $(D, x) \in \mathcal{D} \times X$, if*

[6] The result holds as long as the set of alternatives is finite. As mentioned, in the analysis of discrete choice, the number of alternatives in a choice set is finite by definition. The number of choice sets is usually finite, so the set of alternatives is usually finite.

$x \in D$, then

$$\rho(D,x) = \nu(\pi \in \Pi | \pi(x) \geq \pi(D)).$$

The probability measure ν is said to rationalize ρ. *The set of random utility functions is denoted by $\mathcal{P}_r$.*

A random utility function is a probability distribution over the strict preference rankings over X.[7]

Finally, I review essential mathematical concepts. A *polyhedron* is an intersection of finitely many closed half spaces. A *polytope* is a bounded polyhedron. Equivalently, a polytope is the convex hull of finitely many points.

The closure of a set C is denoted by cl.C. The *affine hull* of a set C is the smallest affine set that contains C, and it is denoted by aff.C. The convex hull of a set C is denoted by co.C.

The *relative interior* of a convex set C is an interior of C in the relative topology with respect to aff.C. The relative interior of C is denoted by rint.C. If C is not empty, then (i) rint.C is not empty, and (ii) rint.$C = \{x \in C |$for all $y \in C$ there exists $\alpha \in \mathbf{R}$ such that $\alpha > 1$ and $\alpha x + (1-\alpha)y \in C\}$. (See Theorem 6.4 in Rockafellar (2015) for the proof.)

3 Axiomatization of the Mixed Logit Model

In this section, I provide axiomatizations of the mixed logit model. The first axiom in the paper may be seen as a normative one. To define a normative requirement, I consider a representative agent whose random choice follows ρ, and thereby compare the representative agent's random choice with *deterministically rational choices* as benchmarks.

A random choice function ρ' is said to be *deterministically rational* if there exists a strict preference ranking $\pi \in \Pi$ such that

$$\rho'(D,x) = \begin{cases} 1 & \text{if } \pi(x) \geq \pi(D); \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

[7] While the function above is often called a random ranking function, a random utility function is often defined differently–by using the existence of a probability measure μ over utilities such that for all $(D,x) \in \mathcal{D} \times X$, if $x \in D$, then $\rho(D,x) = \mu(u \in \mathbf{R}^X | u(x) \geq u(D))$. Block and Marschak (1960)(Theorem 3.1) prove that the two definitions are equivalent.

This random choice function ρ' is denoted by ρ^π. The function ρ^π gives probability one to the best alternative x in a choice set D according to the strict preference ranking π. Remember that this is the standard way to define the rationality of a deterministic choice function.

As a criterion for the comparison between the representative agent's random choice and the deterministically rational choices, I adopt the *social surplus* function proposed by McFadden (1978, 1981). To introduce this concept, consider a representative agent whose random choice follows ρ.[8] Let $u(D, x)$ be the representative agent's utility when he chooses x from D (i.e., x is the best alternative in D). Since $\rho(D, x)$ is the probability that x is chosen from D, the expected welfare of the representative agent who chooses from D is

$$\sum_{x \in D} \rho(D, x) u(D, x).$$

This is the social surplus of choice set D.

Notice that the function u depends on choice set D through the conditioning event that x is the best alternative in D.[9] Moreover, the representative agent's utility could depend on choice set D because the utility itself could be menu-dependent. (McFadden (2001), Swait et al. (2002), and Rooderkerk et al. (2011) all address the importance of context dependence for the analysis of discrete choice.) Since the set $\mathcal{D}$ of choice sets may contain multiple elements, I generalize the social surplus as follows:[10]

$$G(\rho : u) \equiv \sum_{D \in \mathcal{D}} \sum_{x \in D} \rho(D, x) u(D, x). \tag{5}$$

From our viewpoint, as outside observers, the utility function of the representative agent is unobservable. However, it is natural to assume that u belongs to the

[8] See Rust (1994) and Chiong et al. (2016) for papers which use the social surplus functions to study the welfare of a representative agent in the analysis of dynamic discrete choice.

[9] To understand the dependency of u on the set D, consider the case of random utility. Then $u(D, x) = E\big[w_x + \varepsilon_x \big| w_x + \varepsilon_x \geq w_y + \varepsilon_y \text{ for all } y \in D\big]$, where w_x is the representative agent's systematic part of the utility of x and ε_x is the additive shock to the utility. The function u depends on D through the conditioning event.

[10] A careful reader may wonder why the outside observer needs to sum up utilities over $\mathcal{D}$ uniformly. It is possible to introduce weights r over $\mathcal{D}$ and define $G(\rho : u, r) = \sum_{D \in \mathcal{D}} r(D) \sum_{x \in D} \rho(D, x) u(D, x)$. For each $D \in \mathcal{D}$, $r(D)$ can be interpreted as the outside observer's subjective belief that the representative agent chooses from D. The axioms in this paper can be generalized easily to include the weights r.

following set:

$$\mathcal{U} = \left\{ u \in \mathbf{R}_+^{\mathcal{D}\times X} \middle| \text{(i) } u(D,x) = 0 \text{ if } x \notin D; \text{(ii) } u(D,\cdot) \text{ is not constant on some } D \right\},$$

where $\mathbf{R}_+$ is the set of nonnegative real numbers. A utility $u(D,x)$ is nonnegative. Moreover if x is not available in D, then $u(D,x)$ is zero, as required in condition (i). If a utility function does not satisfy condition (ii), then the social surplus is the same for any random choice function. Such a utility function is not useful to evaluate random choice functions in terms of the social surplus.[11]

The next axiom requires that no matter which utility function $u \in \mathcal{U}$ the outside observer uses, *the social surplus obtained by the representative agent's choice should be larger than the minimum social surplus obtained by the deterministically rational choices.*

Axiom 1. *(Aggregated Stochastic Rationality) For any $u \in \mathcal{U}$,*

$$G(\rho : u) > \min_{\pi\in\Pi} G(\rho^{\pi} : u). \tag{6}$$

Aggregated Stochastic Rationality may be seen as a normative axiom. The normative requirement of the axiom is weak in the sense that the axiom does *not* require that the agent's random choice dominate the deterministically rational choices; the axiom only requires that the agent's random choice should be better than the *worst* deterministically rational choices.

The next theorem shows that Aggregated Stochastic Rationality characterizes the mixed logit model. For the sufficiency of the axiom, I need to assume a condition on the set of alternatives.[12]

Definition 5. *The set X of alternatives is said to be* in general position *if (i) X is affinely independent or (ii) there exists $k \in \{1,\dots,K\}$ such that $x(k) \neq y(k)$ for all $x, y \in X$.*

Remember that X is a subset of K-dimensional real space. With respect to condition (i), note that if $|X| \leq K+1$, then, generically speaking, X is affinely

[11] To see this point, suppose that u does not satisfy condition (ii). Then for each $D \in \mathcal{D}$ there exists $v_D \in \mathbf{R}$ such that $u(D,x) = v_D$ for any $x \in D$. For any random choice function $\rho \in \mathcal{P}$, $\sum_{x\in D}\rho(D,x) = 1$ for each $D \in \mathcal{D}$. It follows that $G(\rho : u) = \sum_{D\in\mathcal{D}} v_D$.

[12] For the axiomatization of the general mixed logit model, I do not need the condition. See Corollary coro:general in Section 3.2.

independent. That is, even if X is not affinely independent, adding a small perturbation to X makes it affinely independent. (For example when $K = 2$ and $X = 3$, the only case in which X is not affinely independent is when the points are collinear.) Note that condition (i) is similar in spirit to no perfect multicollinearity, which is considered to hold generically.[13]

On the other hand if $|X| > K+1$, then X cannot be affinely independent.[14] For this case, I require that X satisfy condition (ii). Condition (ii) means that there exists an index k of an explanatory variable at which the alternatives (i.e., $x(k)$s) are all distinct. Condition (ii) is also satisfied generically in the sense that adding a small perturbation to X makes it satisfy condition (ii). Since the observed data are inevitably perturbed by measurement error, it is likely that X is in general position.

Theorem 1. *Suppose that X is in general position. A random choice function ρ satisfies Aggregated Stochastic Rationality if and only if ρ is a mixed logit function.*

The sufficiency part of the proof can be sketched as follows. (See the appendix for the complete proof.) First, I state two propositions which are necessary for the axiomatization.

Proposition 1. *For any positive integer d, the set of mixed logit functions with polynomials of at most degree d is the relative interior of the set of random utility functions (i.e., $\mathcal{P}_{ml}(d) = rint.\mathcal{P}_r$) if and only if $\{p_d(x)|x \in X\}$ is affinely independent.*

In section 4, I will provide details of Proposition 1. The next proposition characterizes the affine hull of the set $\mathcal{P}_r$ of random utility functions.

Proposition 2. *The affine hull of $\mathcal{P}_r$ is*

$$\left\{q \in \mathbf{R}^{\mathcal{D}\times X} \middle| (i) \sum_{x\in D} q(D,x) = 1 \text{ for any } D \in \mathcal{D};\ (ii)\ q(D,x) = 0 \text{ for any } D \in \mathcal{D}, x \notin D\right\}.$$

Proposition 2 implies that the set of random choice functions is a subset of the affine hull of the set of random utility functions (i.e., $\mathcal{P} \subset \text{aff.}\mathcal{P}_r$).[15]

[13]No perfect multicollinearity means that any explanatory variable cannot be represented as an affine combination of the other explanatory variables. Formally, no perfect multicollinearity requires that $\{x(k)\}_{k=1}^{K}$ is affinely independent, where $x(k) = (x_1(k), \ldots, x_{|X|}(k)) \in \mathbf{R}^{|X|}$. Note that, on the other hand, condition (i) means $X \equiv \{x_i\}_{i=1}^{|X|}$ is affinely independent, where $x_i \in \mathbf{R}^K$.

[14]To see this remember that if a set is affinely independent, then the maximal number of elements contained by the set is the dimension of the set plus one.

[15]Note that it is a proper subset because the affine hull contains a vector whose element is negative.

Given the two propositions above, Theorem 1 can be proved as follows. It can be shown that the set $\mathcal{P}_r$ of random utility functions is a polytope. That is, $\mathcal{P}_r = \text{co.}\{\rho^\pi | \pi \in \Pi\}$. Moreover, it follows that there exist a set $\{t_i\}_{i=1}^n \subset \mathbf{R}^{\mathcal{D}\times X}\backslash\{0\}$ and a set $\{\alpha_i\}_{i=1}^n \subset \mathbf{R}$ such that

$$\mathcal{P}_r = \cap_{i=1}^n \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\} \cap \text{aff.}\mathcal{P}_r. \tag{7}$$

As mentioned earlier, Proposition 2 implies that $\mathcal{P}_r \subset \mathcal{P} \subset \text{aff.}\mathcal{P}_r$. This implication and (7) show that $\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i \geq \alpha_i\}$. It follows that $\text{rint.}\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\}$. Proposition 1 implies that for any positive integer d, $\mathcal{P}_{ml}(d) = \text{rint.}\mathcal{P}_r$ if and only if the set $\{p_d(x) | x \in X\}$ is affinely independent. Remark 2 in section 4 states that $\{p_d(x) | x \in X\}$ is affinely independent for some integer d if X is in general position. Hence, I obtain $\mathcal{P}_{ml}(d) = \cap_{i=1}^n \{\rho \in \mathcal{P} | \rho \cdot t_i > \alpha_i\}$.

For each $i \in \{1, \ldots, n\}$, I can find a utility vector $u_i \in \mathcal{U}$ and $\beta_i \in \mathbf{R}$ such that $\rho \cdot t_i > \alpha_i$ if and only if $G(\rho : u_i) > \beta_i$. Therefore, $\mathcal{P}_r = \cap_{i=1}^n \{\rho \in \mathcal{P} | G(\rho : u_i) \geq \beta_i\}$ and $\mathcal{P}_{ml}(d) = \cap_{i=1}^n \{\rho \in \mathcal{P} | G(\rho : u_i) > \beta_i\}$. Since $\rho^\pi \in \mathcal{P}_r$ for any $\pi \in \Pi$, it follows that $G(\rho^\pi : u_i) \geq \beta_i$ for all $i \in \{1, \ldots, n\}$. Hence, Aggregated Stochastic Rationality implies that $G(\rho : u_i) > \beta_i$ for all $i \in \{1, \ldots, n\}$. So, $\rho \in \cap_{i=1}^n \{\rho \in \mathcal{P} | G(\rho : u_i) > \beta_i\} = \mathcal{P}_{ml}(d)$.

3.1 Alternative Axiom

In this section, I provide an alternative axiomatization of the mixed logit model. The necessity of the alternative axiom can be understood heuristically as follows.[16] Given a utility function $u \in \mathcal{U}$, the social surplus function $G(\rho : u)$ is linear in ρ. Hence, if ρ is a mixed logit function (i.e., $\rho = \int \rho_l dm$, where ρ_l is a logit function), then for any $u \in \mathcal{U}$,

$$G(\rho : u) = \int G(\rho_l : u) dm \geq \inf_{\rho_l \in \mathcal{P}_l} G(\rho_l : u), \tag{8}$$

where the equality holds by the Fubini theorem and the linearity of G in ρ. This condition (8) is a necessary condition for ρ to be a mixed logit function. Indeed, one can show that the last inequality holds strictly. This condition with the strict inequality turns out to be sufficient as well.

[16] The following argument is not the proof of the necessity since the axiom requires strict inequality. The following argument guarantees weak inequality only.

Axiom 2. *(Aggregated Logit Rationality) For any $u \in \mathcal{U}$,*

$$G(\rho : u) > \inf_{\rho_l \in \mathcal{P}_l} G(\rho_l : u). \tag{9}$$

To interpret the axiom, remember that the probability measure m captures the unobservable heterogeneity across a population of individuals. Moreover, each logit function ρ_l in the support of m captures aggregate choices in a subpopulation. Thus Aggregated Logit Rationality means that the social surplus must be larger than the smallest surplus among the subpopulations. This is because, as shown in equation (8), the social surplus equals an average surplus across the total population.

By slightly modifying the proof of Theorem 1, I obtain the following result:

Corollary 1. *Suppose that X is in general position. A random choice function ρ satisfies Aggregated Logit Rationality if and only if ρ is a mixed logit function.*

The proof of the corollary is in the appendix.

3.2 Special case with linear p and general case with arbitrary function

In this section, I provide axiomatizations of a special case and a general case of the mixed logit model. The special case I examine here is that of mixed logit functions with polynomials of degree 1 (i.e., $p_d(x) = x$).

Corollary 2. *Suppose that X is affinely independent. A random choice function ρ satisfies Aggregated Stochastic Rationality if and only if ρ is a mixed logit function with polynomials of degree* 1*.*

The general case I examine next is that which applies to mixed logit functions with a general function in place of polynomials. Because of the generality, I do *not* need any conditions on X except for the finiteness. The set X does not have to be a subset of finite-dimensional real space. Needless to say, it does not have to be in general position.

Corollary 3. *A random choice function ρ satisfies Aggregated Stochastic Rationality if and only if ρ is a general mixed logit function.*

The proofs of these two corollaries are in the appendix.

The axiomatizations above are based on Aggregated Stochastic Rationality. By modifying Aggregated Logit Rationality, one can easily provide alternative axiomatizations of the special case and the general case of the mixed logit model.

4 Denseness in the Random Utility Model

In this section, I discuss Proposition 1, which states that the set of mixed logit functions with polynomials of at most degree d is dense in the set of random utility functions if and only if $\{p_d(x)|x \in X\}$ is affinely independent. Proposition 1 implies the following result:

Corollary 4. *Let d be a positive integer.*
(i) If $|X| \leq \binom{d+K}{K}$, then any interior random utility function is generically represented as a convex combination of logit functions with polynomials of at most degree d.[17]
(ii) If $|X| > \binom{d+K}{K}$, then there is a random utility function which cannot be approximated by mixed logit functions with polynomials of at most degree d.

To see how Proposition 1 implies Corollary 4, note that for any $x \in X$ and any positive integer d, $p_d(x)$ is $\binom{d+K}{K}-1$-dimensional real vector. By the same argument after Definition 5, if $|X| \leq \binom{d+K}{K}$, then generically speaking $\{p_d(x)|x \in X\}$ is affinely independent. If $|X| > \binom{d+K}{K}$ then, $\{p_d(x)|x \in X\}$ is not affinely independent. Therefore Proposition 1 implies Corollary 4.

Corollary 4 is related to Theorem 1 of McFadden and Train (2000). In their Theorem 1, McFadden and Train (2000) state that under some technical conditions, any random utility function can be approximated by mixed logit functions. McFadden and Train (2000) admit "One limitation of Theorem 1 is that it provides no practical indication of how to choose parsimonious mixing families, or how many terms are needed to obtain acceptable approximations..."(p. 452) This means that in order to achieve better approximation, they need to use arbitrarily higher order polynomials.

Corollary 4 overcomes the limitation. Corollary 4 gives a precise condition on the degree d of the polynomial. It is necessary and sufficient that the degree d be large enough to satisfy $|X| \leq \binom{d+K}{K}$. There are two additional advantages to

[17]This implies that any noninterior random utility function can be approximated by a convex combination of logit functions with polynomials of at most degree d.

Corollary 4 in comparison with Theorem 1 of McFadden and Train (2000). First, the result by McFadden and Train (2000) guarantees only an approximation, while result (i) in Corollary 4 guarantees the exact equality for the case of interior random utility functions. Second, to achieve the exact equality, Corollary 4 states that it is enough to use a finite convex combination of logit functions, rather than an integral over logit functions.

On the other hand, the setup of McFadden and Train (2000) is more general than mine. They allow X to be infinite, while I assume X is finite following a classical setup in the decision theory literature of random choice.[18] Consequently, Corollary 4 does not imply their Theorem 1.[19] McFadden and Train's (2000) proof of Theorem 1 and and my proof of Proposition 1 and Corollary 4 are very different. While their proof crucially depends on the Weierstrass approximation theorem, my proof does not depend on the approximation theorem but the geometric structure of the set of random utility functions, as I will explain below. Assuming the finiteness of the set of alternatives in the proof of McFadden and Train (2000) does not yield Corollary 4.

I prove Proposition 1 by using the lemmas below. First I introduce a definition.

Definition 6. *For any positive integer d, a ranking $\pi \in \Pi$ is linearly representable by polynomials of at most degree d if there exists a real vector β such that for all $x, y \in X$, $\pi(x) > \pi(y)$ if and only if $\beta \cdot p_d(x) > \beta \cdot p_d(y)$.*

Notice that the above definition requires that all points be ordered according to a given ranking π. Hence, the definition is stronger than the concept of *shattering* in machine learning. In the standard setup, *shattering* only requires that points be separated into two groups.[20]

Lemma 1. *For any positive integer d, the set of mixed logit functions with polynomials of at most degree d is the relative interior of the set of random utility functions (i.e., $\mathcal{P}_{ml}(d) = rint.\mathcal{P}_r$) if and only if any ranking $\pi \in \Pi$ is linearly representable by polynomials of at most degree d.*

[18] Moreover, as explained, in the analysis of discrete choice, the number of alternatives in a choice set is finite by definition. The number of choice sets is usually finite, so the set of alternatives is usually finite.

[19] McFadden and Train (2000) also allow for a random choice function to be dependent on the observed attributes of individuals. To make the discussion above clearer, I assumed that the set of the individuals is homogeneous. I can easily include the set of the observed attributes in my model by allowing a primitive random choice function to be dependent on the individuals' observed attributes.

[20] I am grateful to Prof. Brendan Beare, who informed me about the concept of shattering.

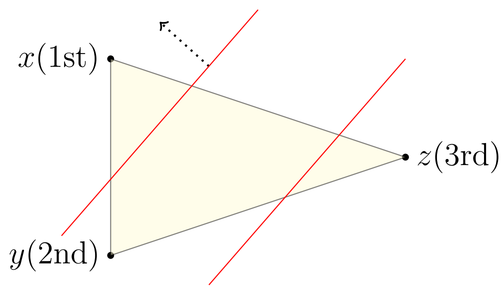

Figure 1: The set $X = \{x, y, z\}$ is affinely independent. Any ranking is linearly representable with polynomials of degree $d = 1$. For example, the ranking $\pi(x) > \pi(y) > \pi(z)$ is linearly representable with polynomials of degree $d = 1$ by $\beta \in \mathbf{R}^2$, which defines the parallel hyperplanes.

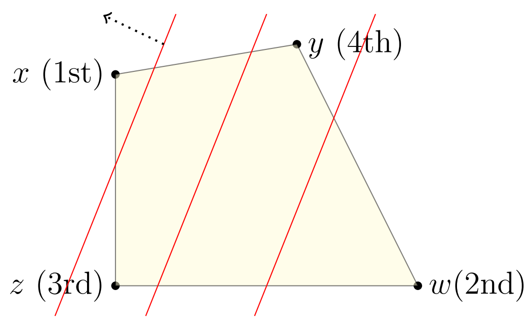

Figure 2: The set $X = \{x, y, z, w\}$ is affinely dependent. The ranking $\pi(x) > \pi(w) > \pi(y) > \pi(z)$ is not linearly representable with polynomials of degree $d = 1$. As the figure shows, no matter how one chooses $\beta \in \mathbf{R}^2$ and draws parallel hyperplanes, it does not hold that $\beta \cdot x > \beta \cdot w > \beta \cdot z > \beta \cdot y$.

To check whether any ranking $\pi \in \Pi$ is linearly representable, the following lemma is useful.

Lemma 2. *For any positive integer d, the set $\{p_d(x)|x \in X\}$ is affinely independent if and only if any ranking $\pi \in \Pi$ is linearly representable by polynomials of at most degree d.*

Lemmas 1, 2 imply Proposition 1. To understand Lemma 2 geometrically, see figures 1 and 2. In the figures, I assume that $K = 2$ and $d = 1$ (i.e., $p_d(x) = x$). Hence, $\{p_d(x)|x \in X\}$ is affinely independent if and only if X is affinely independent.

Although $\{p_d(x)|x \in X\}$ is generically affinely independent when $|X| \leq \binom{d+K}{K}$, a careful reader may wonder when $\{p_d(x)|x \in X\}$ is always (and not just generically) affinely independent. The next remark provides an answer to the question.

Remark 2. *If X is in general position, then $\{p_d(x)|x \in X\}$ is affinely independent for some d.*

5 Discussion

This paper provides axiomatizations of the mixed logit model. In the course of proving the axiomatizations, I have obtained several results which could be of interest by themselves. In this section, I present three such results. The first result (Lemma 3) provides an additional result on the denseness in the random utility model. The second result (Corollary 5) provides a necessary and sufficient condition under which a random utility function can be written as a linear random-coefficient model. The last result (Corollary 6) states that any interior random choice function is generically represented as an affine combination of two mixed logit functions.

Lemma 3. *Let $\mathcal{Q}$ be a subset of $rint.\mathcal{P}_r$. Then $rint.\mathcal{P}_r = co.\mathcal{Q}$ if and only if for any $\pi \in \Pi$, there exists a sequence $\{\rho_n\}_{n=1}^{\infty}$ of $\mathcal{Q}$ such that $\rho_n \to \rho^\pi$ as $n \to \infty$.*

Lemma 3 gives a necessary and sufficient condition under which any interior random utility function can be represented as a convex combination of elements of $\mathcal{Q}$. The condition of Lemma 3 is satisfied when (i) $\mathcal{Q}$ is the set of logit functions and (ii) the degree of polynomials is high enough. The condition is also be satisfied by some other classes of random utility functions, such as the set of probit functions. Hence, Lemma 3 implies that the convex hull of the set of probit functions is dense in the set of random utility functions.

The next result provides a representation of a random utility function.

Corollary 5. *For any random utility function ρ, there exists $\mu \in \Delta(\mathbf{R}^K)$ such that*

$$\rho(D, x) = \mu(\{\beta|\beta \cdot x \geq \beta \cdot y \text{ for all } y \in D\})$$

if and only if X is affinely independent.

In the empirical literature of the random-coefficient model, researchers have analyzed various ways to introduce the randomness of coefficients (i.e., β). In this literature, assuming the linear model is sometimes considered to be restrictive. Corollary 5 states, however, that one can focus on the linear model with no loss of generality if and only if X is affinely independent.

The last result shows representations of a random choice function.

Corollary 6. *(i) Suppose that X is in general position. For any interior random choice function ρ, then there exist a real number α and a pair (ρ_1, ρ_2) of convex combinations of logit functions such that $\rho = \alpha\rho_1 + (1-\alpha)\rho_2$.*[21]
(ii) For any random choice function ρ, there exist a real number α and a pair (ρ_1, ρ_2) of random utility functions such that $\rho = \alpha\rho_1 + (1-\alpha)\rho_2$.

Remember that random choice functions do not have any mathematical structures except that $\rho(\cdot, D)$ is a probability distribution over D, while logit functions and random utility functions have rich mathematical structures. Nevertheless, in Corollary 6, statement (i) says that an interior random choice function is generically represented as an affine combination of two convex combinations of logit functions; statement (ii) says that a random choice function is always (and not just generically) represented as an affine combination of two random utility functions. Dogan and Yildiz (2018) obtained a result which is similar to statement (ii) independently.[22]

To see how Corollary 6 holds, remember that Proposition 2 implies that $\mathcal{P} \subset \text{aff.}\mathcal{P}_r$. That is, for any $\rho \in \mathcal{P}$, there exist $\{\lambda_i\}_{i=1}^n \subset \mathbf{R}$ and $\{\rho'_i\}_{i=1}^n \subset \mathcal{P}_r$, such that $\rho = \sum_{i=1}^n \lambda_i \rho'_i$ and $\sum_{i=1}^n \lambda_i = 1$. Define $\alpha = \sum_{i:\lambda_i>0} \lambda_i$ and $\beta = \sum_{i:\lambda_i<0} \lambda_i$, so that $\alpha + \beta = 1$. Moreover, $\rho_1 \equiv \sum_{i:\lambda_i>0}(\lambda_i/\alpha)\rho'_i$ and $\rho_2 \equiv \sum_{i:\lambda_i<0}(-\lambda_i/-\beta)\rho'_i$ are random utility functions. It follows that $\rho = \sum_{i=1}^n \lambda_i \rho'_i = \alpha\rho_1 + \beta\rho_2 = \alpha\rho_1 + (1 - \alpha)\rho_2$. This establishes statement (ii). Given statement (ii), if ρ is an interior random choice function, then it is without loss of generality to assume that ρ_1 and ρ_2 are interior random utility functions. Hence, statement (i) follows from Proposition 1 and Remark 1.

The appendices follow.

[21] Note that ρ is an interior random choice function if ρ is random choice function and $\rho(D, x) > 0$ for any $D \in \mathcal{D}$ and $x \in D$.

[22] Statement (ii) of Corollary 6 is mentioned in a footnote (footnote 7) in an earlier version of this paper posted on September 15, 2017. See http://www.hss.caltech.edu/content/axiomatizations-mixed-logit-model. I wish to acknowledge Jay Lu for the discussion that led to statement (ii). To obtain Theorem 1 of Dogan and Yildiz (2018) from statement (ii), suppose that $\nu_1, \nu_2 \in \Delta(\Pi)$ represent ρ_1 and ρ_2, respectively. Define $\{\succ_i\} \equiv \text{supp.}\nu_1$. For each ranking $\succ$ on X, define an "inverse" ranking $\succ^{-1}$ by flipping the order of $\succ$ (i.e., $x \succ^{-1} y$ if and only if $y \succ x$). Define $\{\rhd_j\} \equiv \{\succ \mid \succ^{-1} \in \text{supp.}\nu_2\}$. Then $\{\rhd_j^{-1}\} \equiv \text{supp.}\nu_2$. For each $\succ_i$, define $\lambda(\succ_i) = \alpha_1\nu_1(\succ)$. For each $\rhd_j$, define $\lambda(\rhd_j) = |\beta|\nu_2(\rhd_j^{-1})$. Then $\rho(D, x) = \alpha\rho_1(D, x) - |\beta|\rho_2(D, x) = \alpha\nu_1(\succ_i |x \succ_i y \text{ for all } y \in D) - |\beta|\nu_2(\succ_j |x \succ_j y \text{ for all } y \in D) = \lambda(\succ_i |x \succ_i y \text{ for all } y \in D) - \lambda(\rhd_j | y \rhd_j x \text{ for all } y \in D)$.

A Proof of Lemmas and Remarks

In the following, I prove Remarks 1, 2 and Lemmas 2, 3. In section B, I will prove Lemma 1. First I state several lemmas that I use in the rest of the appendix.

Lemma 4. *The set $\mathcal{P}_r$ of random utility functions is a polytope. Moreover, $\mathcal{P}_r = co.\{\rho^\pi | \pi \in \Pi\}$, and there exist hyperplanes $\{H_i\}_{i=1}^n$ in $\mathbf{R}^{\mathcal{D}\times X}$ such that $aff.\mathcal{P}_r \not\subset H_i^-$ and $\mathcal{P}_r = (\cap_{i=1}^n H_i^-) \cap aff.\mathcal{P}_r$, where H_i^- is the closed lower-half space of H_i for each $i \in \{1, \dots, n\}$.*

Proof. Choose any $\rho \in \mathcal{P}_r$ to show $\rho \in$ co.$\{\rho^\pi | \pi \in \Pi\}$. There exists $\nu \in \Delta(\Pi)$ that rationalizes ρ. Define $\lambda_\pi = \nu(\pi)$ for each $\pi \in \Pi$. Define $\rho' = \sum_{\pi \in \Pi} \lambda_\pi \rho^\pi$ to show $\rho = \rho'$. For each $(D, x) \in \mathcal{D} \times X$, $\rho(D, x) = \nu(\pi \in \Pi | \pi(x) \geq \pi(D)) = \sum_{\pi \in \Pi} \nu(\pi) 1(\pi(x) \geq \pi(D)) = \rho'(D, x)$. Then $\rho = \rho' \in$ co.$\{\rho^\pi | \pi \in \Pi\}$. So $\mathcal{P}_r \subset$ co.$\{\rho^\pi | \pi \in \Pi\}$. The argument can be reversed to obtain the converse. By the definition of polytope and Theorem 9.4 of Soltan (2015), the desired hyperplanes exist. □

I will use the following version of theorem of alternatives in several places.

Lemma 5. *Let A be an $r \times n$ real matrix, B be an $l \times n$ real matrix, and E be an real $m \times n$ matrix. Exactly one of the following alternatives is true.*

1. *There is $u \in \mathbf{R}^n$ such that $A \cdot u = 0$, $B \cdot u \geq 0$, $E \cdot u \gg 0$.*
2. *There is $\theta \in \mathbf{R}^r$, $\eta \in \mathbf{R}^l$, and $\pi \in \mathbf{R}^m$ such that $\theta \cdot A + \eta \cdot B + \pi \cdot E = 0$; $\pi \gg 0$ and $\eta \geq 0$.*

See Theorem 1.6.1 of Stoer and Witzgall (2012) for the proof.

A.1 Proof of Remark 1

To prove the remark, I will prove the following general result as a claim. The claim is trivial when the set C is closed. In Remark 1, I use the claim with $C = \mathcal{P}_l$, where the set $\mathcal{P}_l$ is not closed.

Claim: For any set $C \subset \mathbf{R}^K$, let $\Delta(C)$ denote the set of probability measures over C. Then, co.$C = \left\{ \int x dm(x) | m \in \Delta(C) \right\}$, where $\int x dm(x)$ denotes K-dimensional vector whose k-th element is $\int x(k) dm(x)$ for any $k \in \{1, \dots, K\}$.

Proof. By definition, I immediately obtain co.$C \subset \{\int xdm(x)|m \in \Delta(C)\}$. In the following, I will show that

$$\left\{\int xdm(x)|m \in \Delta(C)\right\} \subset \text{co}.C. \tag{10}$$

First I will show that

$$\left\{\int xdm(x)|m \in \Delta(C)\right\} \subset \text{cl.co}.C. \tag{11}$$

To prove this statement, suppose by way of contradiction that $\int xdm(x) \notin$ cl.co.C for some $m \in \Delta(C)$. Then by the strict separating hyperplane theorem (Corollary 11.4.2 of Rockafellar (2015)), there exist $t \in \mathbf{R}^K \setminus \{0\}$ and $\alpha \in \mathbf{R}$ such that $(\int xdm(x)) \cdot t = \alpha > x \cdot t$ for any $x \in$ cl.co.C. This is a contradiction because $\alpha = (\int xdm(x)) \cdot t = \int (x \cdot t)dm(x) < \int \alpha dm(x) = \alpha$.

I now will show (10) by the induction on the dimension of co.C.

Induction Base: If $\dim \text{co}.C = 1$, then (10) holds obviously. If $\dim \text{co}.C = 2$, then there must exist y, z such that co.C is the line segment between y and z. In the following, I assume that the line segment does not contain both y and z but the proof for the other cases are similar. Then for any $x \in$ co.C, there exists unique $\alpha(x) \in (0, 1)$ such that $x = \alpha(x)y + (1 - \alpha(x))z$. Notice that the function α is continuous in x and hence measurable. Moreover, the function α is integrable because α is bounded and nonnegative. Choose any $m \in \Delta(C)$. Then $\int \alpha(x)dm(x)$ exists. Moreover, since $0 < \alpha(x) < 1$, it follows from the monotonicity of integral that $0 < \int \alpha(x)dm(x) < 1$. Denote the value of the integral by $\beta \in (0, 1)$. Then, $\int xdm(x) = \int \alpha(x)y + (1 - \alpha(x))zdm(x) = \beta y + (1 - \beta)z \in$ co.C, as desired.

Choose an integer $l \geq 3$.

Induction Hypothesis: Now suppose that (10) holds for any C such that $\dim C \leq l$.

Induction Step: For any C such that $\dim C = l + 1$, (10) holds. To prove the step, choose any $m \in \Delta(C)$. By (11), I have $\int xdm(x) \in$ cl.co.C.

First consider the case where $\int xdm(x) \in$ rint.cl.co.C. Then since rint.cl.co.$C =$ rint.co.C (by Theorem 6.3 of Rockafellar (2015)), so $\int xdm(x) \in$ co.C, as desired.

Next consider the case where $\int xdm(x) \notin$ rint.cl.co.C. Then, $\int xdm(x) \in \partial$cl.co.$C \equiv$ cl.co.$C \setminus$ rint.co.C. There exists a supporting hyperplane H of cl.co.C at $\int xdm(x)$. Then, there exist $t \in \mathbf{R}^K \setminus \{0\}$ and $\alpha \in \mathbf{R}$ such that $H = \{x|x \cdot t = \alpha\}$ and $\int xdm(x) \cdot t = \alpha > x \cdot t$ for any $x \in$ cl.co.$C \cap H^c$. This implies that $m(H) = 1$.

Hence, $m(H \cap C) = 1$. Since H is a supporting hyperplane and cl.co.$C \not\subset H$, I obtain $\dim(H \cap \text{aff.}C) \leq l$. Hence, $\dim(H \cap C) \leq l$. Therefore, the induction hypothesis shows that $\int xdm(x) \in \text{co.}(H \cap C) \subset \text{co.}C$, as desired. □

The claim above implies Remark 1. The result is not true in an infinite dimensional space.[23]

A.2 Proof of Lemma 2

For any ranking $\pi \in \Pi$ and a positive integer d, consider the following condition: if $\sum_{i=1}^{|X|-1} \lambda_i(p_d(\pi^{-1}(|X|+1-i)) - p_d(\pi^{-1}(|X|-i))) = 0$ and $\lambda_i \geq 0$ for all $i \in \{1, \ldots, |X|-1\}$, then $\lambda_i = 0$ for all $i \in \{1, \ldots, |X|-1\}$. I call this condition as Condition $(*)$.

Step 1: For each $\pi \in \Pi$ and a positive integer d, Condition $(*)$ holds if and only if π is linearly representable by polynomials at most degree d (i.e., there exists β such that for any $x, y \in X$, $\pi(x) > \pi(y) \Longleftrightarrow \beta \cdot p_d(x) > \beta \cdot p_d(y)$).

Proof. Fix $\pi \in \Pi$.

$$\begin{aligned}
&\exists \beta \left[\beta \cdot p_d(\pi^{-1}(|X|)) > \beta \cdot p_d(\pi^{-1}(|X|-1)) > \cdots > \beta \cdot p_d(\pi^{-1}(2)) > \beta \cdot p_d(\pi^{-1}(1))\right] \\
&\Longleftrightarrow \exists \beta \left[\beta \cdot (p_d(\pi^{-1}(|X|)) - p_d(\pi^{-1}(|X|-1))) > 0, \ldots, \beta \cdot (p_d(\pi^{-1}(2)) - p_d(\pi^{-1}(1))) > 0\right] \\
&\Longleftrightarrow \nexists \lambda \in \mathbf{R}^{|X|-1} \left[\textstyle\sum_{i=1}^{|X|-1} \lambda_i(p_d(\pi^{-1}(|X|+1-i)) - p_d(\pi^{-1}(|X|-i))) = 0, \lambda \geq 0, \text{ and } \lambda \neq 0\right] \\
&\Longleftrightarrow \text{Condition}(*),
\end{aligned}$$

where the second to the last equivalence is by Lemma 5. □

Step 2: For a given positive integer d, the set $\{p_d(x) | x \in X\}$ is affinely independent if and only if Condition $(*)$ holds for the given positive integer d and any $\pi \in \Pi$.

Proof. I first show that the only if part. Fix any $\pi \in \Pi$. Without loss of generality assume that $\pi(x_i) = |X| + 1 - i$ for all $i \in \{1, \ldots, |X|\}$. Suppose that $\sum_{i=1}^{|X|-1} \lambda_i(p_d(\pi^{-1}(|X|+1-i)) - p_d(\pi^{-1}(|X|-i))) \equiv \sum_{i=1}^{|X|-1} \lambda_i(p_d(x_i) - p_d(x_{i+1})) = 0$ and $\lambda_i \geq 0$ for all i. Define $\mu_1 = \lambda_1$, $\mu_i = \lambda_i - \lambda_{i-1}$ for all $i \in \{2, \ldots, |X|-1\}$, and $\mu_{|X|} = -\lambda_{|X|-1}$. Then $\sum_{i=1}^{|X|-1} \lambda_i(p_d(x_i) - p_d(x_{i+1})) = \lambda_1 p_d(x_1) + \sum_{i=2}^{|X|-1}(\lambda_i -$

[23] Let $\{e_i\}_{i=1}^{\infty}$ be the base of the infinite dimensional real space. Define $C = \{e_i\}_{i=1}^{\infty}$. Define a measure m on C such that $m(e_i) = (1/2)^i$ for each i. Then, $\sum_{i=1}^{\infty} m(e_i) = 1$, so that m is a probability measure on C. $\int xdm$ cannot be represented as any convex combination of elements of C. For any $y \in \text{co.}C$, there exists i such that $y(e_i) = 0$.

$\lambda_{i-1})p_d(x_i) + (-\lambda_{|X|-1})p_d(x_{|X|}) = \mu_1 p_d(x_1) + \sum_{i=2}^{|X|-1} \mu_i p_d(x_i) + \mu_{|X|}p_d(x_{|X|}) = \sum_{i=1}^{|X|} \mu_i p_d(x_i)$. Since $\sum_{i=1}^{|X|-1} \lambda_i(p_d(x_i) - p_d(x_{i+1})) = 0$, I have $\sum_{i=1}^{|X|} \mu_i p_d(x_i) = 0$. Moreover, $\sum_{i=1}^{|X|} \mu_i = \lambda_1 + \sum_{i=2}^{|X|-1}(\lambda_i - \lambda_{i-1}) + (-\lambda_{|X|-1}) = 0$. If $\{p_d(x)|x \in X\}$ is affinely independent, then $\mu_i = 0$ for all $i \in \{1, \ldots, |X|\}$. Hence, $\lambda_i = 0$ for all $i \in \{1, \ldots, |X| - 1\}$.

Next I will show the if part. Choose any real numbers $\{\mu_i\}_{i=1}^{|X|}$ such that $\sum_{i=1}^{|X|} \mu_i p_d(x_i) = 0$ and $\sum_{i=1}^{|X|} \mu_i = 0$ to show $\mu_i = 0$ for all $i \in \{1, \ldots, |X|\}$. Order μ_i by its value. Without loss of generality assume that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{|X|}$. If $\mu = 0$, then the proof is finished. If $\mu \neq 0$ then $\mu_1 > 0$. For each $x_i \in X$, define $\pi(x_i) = |X| + 1 - i$. Then $\pi \in \Pi$.

Define $\lambda_1 = \mu_1$ and $\lambda_i = \sum_{j=1}^{i} \mu_j$ for all $i \in \{2, \ldots, |X|-1\}$. Then $\lambda \neq 0$ because $\mu_1 > 0$. I will show that $\lambda_i \geq 0$ for all $i \in \{1, \ldots, |X|-1\}$. Suppose by way of contradiction that $\lambda_i < 0$ for some i. Then $\mu_i < 0$ because $\mu_1 \geq \cdots \geq \mu_i$. Since $0 > \mu_i \geq \mu_j$ for all $j \geq i$, I have $\sum_{j=i+1}^{|X|} \mu_j < 0$. It follows that $\sum_{j=1}^{|X|} \mu_j = \lambda_i + \sum_{j=i+1}^{|X|} \mu_j < 0$. This contradicts that $\sum_{i=1}^{|X|} \mu_i = 0$. Therefore, $\lambda_i \geq 0$ for all $i \in \{1, \ldots, |X| - 1\}$. Moreover $\sum_{i=1}^{|X|-1} \lambda_i(p_d(\pi^{-1}(|X| + 1 - i)) - p_d(\pi^{-1}(|X| - i))) = \sum_{i=1}^{|X|-1} \lambda_i(p_d(x_i) - p_d(x_{i+1})) = \lambda_1 p_d(x_1) + \sum_{i=2}^{|X|-1}(\lambda_i - \lambda_{i-1})p_d(x_i) + (-\lambda_{|X|-1})p_d(x_{|X|}) = \mu_1 p_d(x_1) + \sum_{i=2}^{|X|-1} \mu_i p_d(x_i) + (-\sum_{i=1}^{|X|-1} \mu_i)p_d(x_{|X|}) = \sum_{i=1}^{|X|} \mu_i p_d(x_i) = 0$, where the second to the last equality holds because $\sum_{i=1}^{|X|} \mu_i = 0$. Therefore, by Condition $(*)$, $\lambda_i = 0$ for all $i \in \{1, \ldots, |X| - 1\}$. Hence, $\mu_i = 0$ for all $i \in \{1, \ldots, |X|\}$. □

A.3 Proof of Lemma 3

Let $\mathcal{Q}$ be any subset of rint.$\mathcal{P}_r$. I will show that rint.$\mathcal{P}_r$ = co.$\mathcal{Q}$ if and only if for any $\pi \in \Pi$ there exists a sequence $\{\rho_n\}_{n=1}^{\infty}$ of $\mathcal{Q}$ such that $\rho_n \to \rho^\pi$ as $n \to \infty$.

Step 1: I will show the if part of the statement. Suppose by way of contradiction that there exists $\rho \in$ rint.$\mathcal{P}_r \setminus$ co.$\mathcal{Q}$. Because co.$\mathcal{Q} \neq \emptyset$, I obtain rint.co.$\mathcal{Q} \neq \emptyset$. Since $\rho \notin$ co.$\mathcal{Q}$, then by the proper separating hyperplane theorem (Theorem 11.3 of Rockafellar (2015)), there exist $t \in \mathbf{R}^{\mathcal{D}\times X} \setminus \{0\}$ and $a \in \mathbf{R}$ such that $\rho \cdot t \geq a \geq \rho' \cdot t$ for any $\rho' \in$ co.$\mathcal{Q}$, and $a > \rho'' \cdot t$ for some $\rho'' \in$ co.$\mathcal{Q}$.

I obtain a contradiction by two substeps. Define $\hat{\mathcal{P}}_r = \{\hat{\rho} \in \mathcal{P}_r | t \cdot \hat{\rho} > t \cdot \rho\}$.

Step 1.1: $\hat{\mathcal{P}}_r \neq \emptyset$. To prove the step, remember that there exists $\rho'' \in$ co.$\mathcal{Q}$ such that $\rho'' \cdot t < \rho \cdot t$. Moreover, since $\mathcal{Q} \subset \mathcal{P}_r$ and $\mathcal{P}_r$ is convex, it follows that $\rho'' \in$ co.$\mathcal{Q} \subset \mathcal{P}_r$. Since $\rho \in$ rint.$\mathcal{P}_r$, there exists $\lambda > 1$ such that $\lambda\rho + (1-\lambda)\rho'' \in \mathcal{P}_r$. Moreover, $(\lambda\rho + (1-\lambda)\rho'') \cdot t = \lambda\rho \cdot t + (1-\lambda)\rho'' \cdot t = \rho \cdot t + (\lambda - 1)(\rho \cdot t - \rho'' \cdot t) > \rho \cdot t$, where the last inequality holds because $\lambda > 1$ and $\rho'' \cdot t < \rho \cdot t$. So $\lambda\rho + (1-\lambda)\rho'' \in \hat{\mathcal{P}}_r$, and $\hat{\mathcal{P}}_r \neq \emptyset$.

Step 1.2: There exists $\rho' \in \text{co}.\mathcal{Q}$ such that $\rho' \cdot t > \rho \cdot t$. To prove the step, choose any $\hat{\rho} \in \hat{\mathcal{P}}_r$. By Lemma 4, there exist nonnegative numbers $\{\hat{\lambda}_\pi\}_{\pi \in \Pi}$ such that $\hat{\rho} = \sum_{\pi \in \Pi} \hat{\lambda}_\pi \rho^\pi$ and $\sum_{\pi \in \Pi} \hat{\lambda}_\pi = 1$.

By the supposition of the lemma, for any $\pi \in \Pi$, there exists a sequence $\{\rho'_n\}_{n=1}^\infty$ of $\mathcal{Q}$ such that $\rho'_n \to \rho^\pi$ as $n \to \infty$. Therefore, for any $\pi \in \Pi$ and any positive number ε, there exists $\rho'_\pi \in \{\rho'_n\}_{n=1}^\infty$ such that $\|\rho'_\pi - \rho^\pi\| < \varepsilon$. Define $\rho' = \sum_{\pi \in \Pi} \hat{\lambda}_\pi \rho'_\pi$. Then $\rho' \in \text{co}.\mathcal{Q}$ and $\|\rho' - \hat{\rho}\| = \left\| \sum_{\pi \in \Pi} \hat{\lambda}_\pi (\rho'_\pi - \rho^\pi) \right\| \leq \sum_{\pi \in \Pi} \hat{\lambda}_\pi \|\rho'_\pi - \rho^\pi\| \leq \sum_{\pi \in \Pi} \hat{\lambda}_\pi \varepsilon = \varepsilon$. Therefore, $|t \cdot \rho' - t \cdot \hat{\rho}| \leq \|t\| \|\rho' - \hat{\rho}\| \leq \|t\| \varepsilon$. Since $t \cdot \hat{\rho} > t \cdot \rho$, then by choosing ε small enough, I obtain $t \cdot \rho' > t \cdot \rho$.

Step 2: I will show the only inf part of the statement. Since $\text{rint}.\mathcal{P}_r = \text{co}.\mathcal{Q}$,

$$\mathcal{P}_r = \text{cl}.\mathcal{P}_r = \text{cl}.\text{rint}.\mathcal{P}_r = \text{cl}.\text{co}.\mathcal{Q} = \text{co}.\text{cl}.\mathcal{Q}, \tag{12}$$

where the first equality holds because $\mathcal{P}_r$ is closed, the second equality holds by Theorem 6.3 of Rockafellar (2015), and the last equality holds because $\mathcal{Q}$ is bounded and by Theorem 17.2 of Rockafellar (2015). Since $\mathcal{P}_r = \text{co}.\text{cl}.\mathcal{Q}$, for any $\pi \in \Pi$, there exist positive numbers $\{\lambda_i\}_{i=1}^m$ such that $\sum_{i=1}^m \lambda_i = 1$ and a convergent sequence $\{\rho_n^i\}_{n=1}^\infty$ of $\mathcal{Q}$ for each $i \in \{1, \ldots, m\}$ such that $\sum_{i=1}^m \lambda_i \rho_n^i \to \rho^\pi$ as $n \to \infty$. Since ρ^π is a vertex of $\mathcal{P}_r$, $\rho_n^i \to \rho^\pi$ as $n \to \infty$ for all i.

A.4 Proof of Remark 2

If X is affinely independent, then the result holds with $d = 1$. Consider the case where X is not affinely independent. Suppose by way of contradiction that $\{p_d(x) | x \in X\}$ is not affinely independent with $d = |X|$. Let $X = \{x_1, \ldots, x_{|X|}\}$. Without loss of generality, assume that there exists $\alpha \in \mathbf{R}^{|X|-1}$ such that $x_1^n = \sum_{i=2}^{|X|} \alpha_i x_i^n$ for all $n \in \{1, \ldots, |X|\}$ and $\sum_{i=2}^{|X|} \alpha_i = 1$. For each $k \in \{1, \ldots, K\}$,

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_2(k) & x_3(k) & \cdots & x_{|X|}(k) \\ & & \vdots & \\ x_2^{|X|}(k) & x_3^{|X|}(k) & \cdots & x_{|X|}^{|X|}(k) \end{bmatrix} \begin{bmatrix} \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_{|X|} \end{bmatrix} = \begin{bmatrix} 1 \\ x_1(k) \\ \vdots \\ x_1^{|X|}(k) \end{bmatrix}. \tag{13}$$

Fix $k \in \{1, \ldots, K\}$. By Lemma 5, the existence of $\alpha \in \mathbf{R}^{|X|-1}$ satisfying (13) implies the nonexistence of $\theta \in \mathbf{R}^{|X|+1}$ satisfying the following equations

$$\begin{bmatrix} 1 & x_1(k) & \cdots & x_1^{|X|}(k) \\ 1 & x_2(k) & \cdots & x_2^{|X|}(k) \\ & & \vdots & \\ 1 & x_{|X|}(k) & \cdots & x_{|X|}^{|X|}(k) \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_{|X|} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{14}$$

The rectangle matrix in (14) is a Vandermonde matrix. Since X is in general position, the rank of the matrix is $|X|$. Hence, $\theta \in \mathbf{R}^{|X|+1}$ satisfying (14) must exist. This is a contradiction.

B Proof of Lemma 1 and Proposition 1

By Remark 1, it suffices to show that for any positive integer d, $\mathcal{P}_{ml}(d) = \text{rint}.\mathcal{P}_r$ if and only if $\{p_d(x) | x \in X\}$ is affinely independent. To show the result, I prove two lemmas.

Lemma 6. *co.$\mathcal{P}_l \subset$ rint.$\mathcal{P}_r$.*

Proof. First I show that for any $\rho \in \mathcal{P}_l$, there exists $\nu \in \Delta(\Pi)$ such that ρ is rationalized by ν. Moreover $\nu(\pi) > 0$ for all $\pi \in \Pi$.

To show the statement, remember that for any $\rho \in \mathcal{P}_l$, there exists a real vector β and a positive integer d such that $\rho(D, x) = \exp(\beta \cdot p_d(x)) / \sum_{y \in D} \exp(\beta \cdot p_d(y))$. By Block and Marschak (1960), $\rho \in \mathcal{P}_r$, so there exists $\nu \in \Delta(\Pi)$ such that ν rationalizes ρ. Moreover, in their construction of ν, they obtain that for any $\pi \in \Pi$,

$$\nu(\pi) = \prod_{k=1}^{|X|} \frac{\exp(\beta \cdot p_d(x_k))}{\sum_{l=k}^{|X|} \exp(\beta \cdot p_d(x_l))} > 0,$$

where $X = \{x_1, x_2, \ldots, x_{|X|}\}$ and $\pi(x_1) > \pi(x_2) > \cdots > \pi(x_{|X|})$. Therefore $\rho = \sum_{\pi \in \Pi} \nu(\pi) \rho^\pi$ and $\sum_{\pi \in \Pi} \lambda_\pi = 1$. Since $\nu(\pi) > 0$ for all $\pi \in \Pi$, it follows from Theorem 6.9 in Rockafellar (2015) that $\rho \in \text{rint.co.}\{\rho^\pi | \pi \in \Pi\} = \text{rint}.\mathcal{P}_r$, where the last equality holds by Lemma 4. $\square$

Lemma 7. *For any ranking $\pi \in \Pi$, π is linearly representable by polynomials of at most degree d if and only if there exists a sequence $\{\rho_n\}_{n=1}^{\infty}$ of $\mathcal{P}_l(d)$ such that $\rho_n \to \rho^\pi$ as $n \to \infty$.*

Proof. Assume that a ranking π is linearly representable by polynomials at most degree d. Without loss of generality, assume that $X = \{x_1, \ldots, x_{|X|}\}$ and $\pi(x_1) > \pi(x_2) > \cdots > \pi(x_{|X|})$. Then there exists β such that $\beta \cdot p_d(x_1) > \beta \cdot p_d(x_2) > \cdots > \beta \cdot p_d(x_{|X|})$. For any positive integer k and any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$,

$$\begin{aligned}&\rho_{k\beta}(D, x)\\ &\equiv \frac{\exp(k\beta \cdot p_d(x))}{\sum_{y \in D} \exp(k\beta \cdot p_d(y))}\\ &= \frac{1}{\sum_{y \in D: \pi(y) > \pi(x)} \exp(k\beta \cdot (p_d(y) - p_d(x))) + 1 + \sum_{y \in D: \pi(y) < \pi(x)} \exp(k\beta \cdot (p_d(y) - p_d(x)))}.\end{aligned}$$

For any $y \in D$, $\pi(y) > \pi(x)$ if and only if $\beta \cdot (p_d(y) - p_d(x)) > 0$. Therefore, as $k \to \infty$, if $\pi(x) \geq \pi(D)$, then $\rho_{k\beta}(D, x) \to 1$; if $\pi(x) < \pi(D)$, then $\rho_{k\beta}(D, x) \to 0$. Hence, $\rho_{k\beta} \to \rho^\pi$ as $k \to \infty$.

To show the converse, fix a positive integer d and a sequence $\{\beta_n\}_{n=1}^{\infty}$ such that $\rho_{\beta_n} \to \rho^\pi$ as $n \to \infty$, where for any $D \in \mathcal{D}$ and $x \in D$, $\rho_{\beta_n}(D, x) \equiv \exp(\beta_n \cdot p_d(x)) / \sum_{y \in D} \exp(\beta_n \cdot p_d(y))$. For any $D \in \mathcal{D}$ and $x \in D$, note that

$$\rho_{\beta_n}(D, x) = \frac{1}{1 + \sum_{y \in D \setminus x} \exp(\beta_n \cdot (p_d(y) - p_d(x)))}.$$

Let $\pi(x) \geq \pi(D)$. Since $\rho_{\beta_n} \to \rho^\pi$ as $n \to \infty$, it must hold that $\beta_n \cdot (p_d(y) - p_d(x)) \to -\infty$ as $n \to \infty$ for all $y \in D \setminus \{x\}$. Therefore, for each $D \in \mathcal{D}$ there exists $\overline{n}(D)$ such that for all $n > \overline{n}(D)$ and all $y \in D \setminus \{x\}$, $\beta_n \cdot p_d(x) > \beta_n \cdot p_d(y)$.

Without loss of generality assume that $X = \{x_1, \ldots, x_{|X|}\}$ and $\pi(x_1) > \pi(x_2) > \cdots > \pi(x_{|X|})$. Let $n > \max\left\{\overline{n}(X), \overline{n}(\{x_i\}_{i=2}^{|X|}), \ldots, \overline{n}(\{x_i\}_{i=|X|-1}^{|X|})\right\}$. Then, $\beta_n \cdot p_d(x_1) > \beta_n \cdot p_d(x_2) > \cdots > \beta_n \cdot p_d(x_{|X|-1}) > \beta_n \cdot p_d(x_{|X|})$. Therefore, π is linearly representable by polynomials of at most degree d. □

By Lemma 6, I can apply Lemma 3 with $\mathcal{Q}$ being the set of logit functions with polynomials of at most degree d. Then Lemmas 3, 7 imply Lemma 1. Lemmas 1, 2 imply Proposition 1.

C Proof of Proposition 2

To prove Proposition 2, I prove one more lemma.

Lemma 8. *For any $t \in \mathbf{R}^{\mathcal{D}\times X}$, $\rho^{\pi} \cdot t = \rho^{\pi'} \cdot t$ for all $\pi, \pi' \in \Pi$ if and only if $t(D,x) = t(D,y)$ for all $D \in \mathcal{D}$ and $x, y \in D$.*

Proof. For notational convenience, for any $\pi \in \Pi$ and $D \in \mathcal{D}$ with $D = \{x_1, \dots, x_{|D|}\}$, I write $\rho^{\pi}(D) = \big(\rho^{\pi}(D, x_1), \dots, \rho^{\pi}(D, x_{|D|})\big)$. The if part of the statement is easy to prove. Assume $t(D,x) = t(D,y)$ for all $D \in \mathcal{D}$ and $x, y \in D$. Define $t(D) = t(D,x)$ for any $x \in D$. Then for any $\pi \in \Pi$, $\rho^{\pi} \cdot t = \sum_{D\in\mathcal{D}} \sum_{x\in D} \rho^{\pi}(D,x)t(D,x) = \sum_{D\in\mathcal{D}} t(D) \sum_{x\in D} \rho^{\pi}(D,x) = \sum_{D\in\mathcal{D}} t(D)$.

To show the only if part, let k be the minimal integer such that $|D| \geq k+1$ for any $D \in \mathcal{D}$.

Claim: *For any $D \in \mathcal{D}$ such that $|D| = k+1$ and any $x, y \in D$, $t(D,x) = t(D,y)$.*

To prove the claim, denote D by $\{x, y, w_1, \dots, w_{k-1}\}$. (If $k \leq 1$, then w_is are not included in D and remove w_is in the following proof.) Choose any $\pi, \pi' \in \Pi$ such that for any $z \in X \setminus \{x, y, w_1, \dots, w_{k-1}\}$ and any $i \in \{1, \dots, k-1\}$, $\pi(z) = \pi'(z)$, $\pi(z) > \pi(x) > \pi(y) > \pi(w_i)$, $\pi'(z) > \pi'(y) > \pi'(x) > \pi'(w_i)$, and $\pi(w_i) = \pi'(w_i)$.

To show the claim, I will show the following two facts: (a) For any $E \in \mathcal{D}$, $\rho^{\pi}(E) \neq \rho^{\pi'}(E)$ if and only if $\{x,y\} \subset E$ and $\pi(x) \geq \pi(E)$; (b) If $E \in \mathcal{D}$, $\{x,y\} \subset E$ and $\pi(x) \geq \pi(E)$, then $\rho^{\pi}(E,x) = 1$, $\rho^{\pi}(E,z) = 0$ for any $z \in D \setminus \{x\}$ and $\rho^{\pi'}(E,y) = 1$, $\rho^{\pi'}(E,z) = 0$ for any $z \in E \setminus \{y\}$.

It is easy to see statement (b) and the only if part of statement (a). To show the if part of statement (a), assume $\{x,y\} \not\subset E$ or $\pi(x) < \pi(z)$ for some $z \in E$. First consider the case where $\{x,y\} \not\subset E$. If both x, y do not belong to E, then $\rho^{\pi}(E) = \rho^{\pi'}(E)$ because the ranking over $X \setminus \{x,y\}$ is the same for π and π'. If only one of them, say x, belongs to E, then $\rho^{\pi}(E) = \rho^{\pi'}(E)$ because the ranking over $X \setminus \{y\}$ is the same for π and π'.

Next consider the case where $\pi(x) < \pi(z)$ for some $z \in E$. Then by the definition of π, I obtain $z \in X \setminus \{x, y, w_1, \dots, w_{k-1}\}$. Therefore, $\pi'(y) < \pi'(z)$. Hence, $\rho^{\pi}(E,z) = 1 = \rho^{\pi'}(E,z)$ and $\rho^{\pi}(E,z') = 0 = \rho^{\pi'}(E,z')$ for all $z' \in E \setminus \{z\}$.

Now, I will prove the claim. Since $t \cdot \rho^{\pi} = t \cdot \rho^{\pi'}$,

$$\begin{aligned} 0 &= \textstyle\sum_{(E,z)\in\mathcal{D}\times X} t(E,z)(\rho^{\pi}(E,z) - \rho^{\pi'}(E,z)) \\ &= \textstyle\sum_{(E,z)\in\mathcal{D}\times X:\{x,y\}\subset E,\pi(x)\geq\pi(E)} t(E,z)(\rho^{\pi}(E,z) - \rho^{\pi'}(E,z)) && (\because \text{(a)}) \\ &= \textstyle\sum_{E\in\mathcal{D}:\pi(x)\geq\pi(E),\{x,y\}\subset E} t(E,x) - t(E,y) && (\because \text{(b)}) \\ &= t(D,x) - t(D,y) + \textstyle\sum_{E\in\mathcal{D}:\pi(x)\geq\pi(E),\{x,y\}\subset E,|E|\leq k}(t(E,x) - t(E,y)). \end{aligned}$$

The second term is zero because there is no $D \in \mathcal{D}$ such that $|D| \leq k$. So $t(D,x) = t(D,y)$. This completes the proof of the claim.

The general case can be proved by the induction on $|D|$. Choose any D such that $|D| = k' + 1$, where $k' > k$. Choose any $x, y \in D$. As an induction hypothesis, suppose that for any $E \in \mathcal{D}$, if $|E| \leq k'$ then $t(E,x) = t(E,y)$ for any $x, y \in E$. By the same argument (with k' in place of k) in the proof of the claim, I have

$$0 = t(D,x) - t(D,y) + \sum_{E\in\mathcal{D}:\pi(x)\geq\pi(E),\{x,y\}\subset E,|E|\leq k'} (t(E,x) - t(E,y)).$$

Since the second term is zero by the induction hypothesis, $t(D,x) = t(D,y)$. □

Now I will prove Proposition 2.

The set $\{q \in \mathbf{R}^{\mathcal{D}\times X} | \text{(i) and (ii)}\}$ is affine. So it suffices to show that for any affine set A, if $\mathcal{P}_r \subset A$, then $\{q \in \mathbf{R}^{\mathcal{D}\times X} | \text{(i) and (ii)}\} \subset A$. Since the set is affine, then by Rockafellar (2015), there exist a positive integer L, $L \times (|\mathcal{D}| \times |X|)$ matrix B, and $L \times 1$ vector b such that $A = \{q \in \mathbf{R}^{\mathcal{D}\times X} | Bq = b\}$. For any $l \in \{1, \dots, L\}$, $B_l(D,x)$ denotes $(l,(D,x))$ entry of B. (Remember that B has a column vector for each $(D,x) \in \mathcal{D} \times X$.) So $Bq = b$ means that for any $l \in \{1, \dots, L\}$,

$$\sum_{D\in\mathcal{D}} \sum_{x\in X} B_l(D,x) q(D,x) = b_l. \tag{15}$$

By assuming $\mathcal{P}_r \subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | Bq = b\}$, I will show that if q satisfies (i) and (ii), then (15) holds for any $l \in \{1, \dots, L\}$.

Step 1: $B_l(D,x) = B_l(D,y)$ for any $l \in \{1, \dots, L\}$, $D \in \mathcal{D}$, and $x, y \in D$. To prove Step 1, fix any l. For any $\pi \in \Pi$, $\rho^{\pi} \in \mathcal{P}_r \subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | Bq = b\}$. Hence, (15) holds with $q = \rho^{\pi}$ for any $\pi \in \Pi$. Thus $\rho^{\pi} \cdot B_l = \rho^{\pi'} \cdot B_l$ for any $\pi, \pi' \in \Pi$. By Lemma 8, this implies that $B_l(D,x) = B_l(D,y)$ for any $D \in \mathcal{D}$, and $x, y \in D$.

By Step 1, I can define $B_l(D) = B_l(D,x)$ for any $x \in D$.

Step 2: If q satisfies (i) and (ii), then $Bq = b$, or $\sum_{D\in\mathcal{D}}\sum_{x\in X} B_l(D,x)q(D,x) = b_l$ for any $l \in \{1,\dots,L\}$. To prove Step 2, choose any $\pi \in \Pi$ and $l \in \{1,\dots,L\}$. Since $\rho^\pi \in \mathcal{P}_r \subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | Bq = b\}$, then by (15),

$$b_l = \sum_{D\in\mathcal{D}}\sum_{x\in X} B_l(D,x)\rho^\pi(D,x) = \sum_{D\in\mathcal{D}} B_l(D), \tag{16}$$

where the second equality holds by $\rho^\pi(D,z) = 1$ if $\pi(z) \geq \pi(D)$ and $\rho^\pi(D,z) = 0$ otherwise.

Finally by using these equalities, for each $l \in \{1,\dots,L\}$, I obtain the following equations:

$$\begin{aligned}
\textstyle\sum_{D\in\mathcal{D}}\sum_{z\in X} B_l(D,z)q(D,z) &= \textstyle\sum_{D\in\mathcal{D}}\sum_{z\in D} B_l(D,z)q(D,z) && (\because \text{ (ii)})\\
&= \textstyle\sum_{D\in\mathcal{D}}\sum_{z\in D} B_l(D)q(D,z) && (\because \text{Step 1})\\
&= \textstyle\sum_{D\in\mathcal{D}} B_l(D)\sum_{z\in D} q(D,z) &&\\
&= \textstyle\sum_{D\in\mathcal{D}} B_l(D) && (\because \text{ (i)})\\
&= b_l. && (\because (16))
\end{aligned}$$

This establishes that aff.$\mathcal{P}_r = \{q \in \mathbf{R}^{\mathcal{D}\times X} |\text{(i) and (ii)}\}$.

D Proof of Theorem 1

Before proving the theorem, note that for any random choice function $\rho \in \mathcal{P}$ and any $u \in \mathcal{U}$, it holds that $G(\rho : u) = \rho \cdot u$. To see this notice that $G(\rho : u) \equiv \sum_{D\in\mathcal{D}}\sum_{x\in D} \rho(D,x)u(D,x) = \sum_{D\in\mathcal{D}}\sum_{x\in X} \rho(D,x)u(D,x) \equiv \rho \cdot u$, where the second equality holds because $\rho(D,x) = 0$ if $x \notin D$. In the following, I will use this equality freely.

To show the necessity of Aggregated Stochastic Rationality, fix any $u \in \mathcal{U}$. Since $u(D,\cdot)$ is not constant for some $D \in \mathcal{D}$. Lemma 8 shows that $G(\rho^\pi : u) = \rho^\pi \cdot u \neq \rho^{\pi'} \cdot u = G(\rho^{\pi'} : u)$ for some $\pi, \pi' \in \Pi$. Fix $\rho \in \mathcal{P}_{ml}$. By Remark 1 and Lemma 6, $\rho \in \mathcal{P}_{ml} = \text{co}.\mathcal{P}_l \subset \text{rint}.\mathcal{P}_r$. Hence, ρ is rationalized by full support $\nu \in \Delta(\Pi)$. Then, $G(\rho : u) = \sum_{\pi\in\Pi} \nu(\pi)G(\rho^\pi : u) > \min_{\pi\in\Pi} G(\rho^\pi : u)$.

Now I will show the sufficiency of Aggregated Stochastic Rationality. First I will show $\mathcal{P}_{ml} = \cap_{i=1}^n \{\rho' \in \mathcal{P} | \rho' \cdot t_i > \alpha_i\}$ for some $\{t_i\}_{i=1}^n \subset \mathbf{R}^{\mathcal{D}\times X} \setminus \{0\}$ and $\{\alpha_i\}_{i=1}^n \subset \mathbf{R}$. By Lemma 4, there exist $\{t_i\}_{i=1}^n \subset \mathbf{R}^{\mathcal{D}\times X} \setminus \{0\}$ and $\{\alpha_i\}_{i=1}^n \subset \mathbf{R}$ such that $\mathcal{P}_r = \cap_{i=1}^n \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\} \cap \text{aff}.\mathcal{P}_r$ and aff.$\mathcal{P}_r \not\subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\}$ for all $i \in \{1,\dots,n\}$. Since rint.$\mathcal{P}_r \neq \emptyset$, then by Theorem 6.5 of Rockafellar (2015),

$\text{rint}.\mathcal{P}_r = \cap_{i=1}^n \text{rint}.\{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\} \cap \text{aff}.P_r = \cap_{i=1}^n \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i > \alpha_i\} \cap \text{aff}.P_r$. By Proposition 2, $\mathcal{P}_r \subset \mathcal{P} \subset \text{aff}.\mathcal{P}_r$. Thus

$$\begin{aligned}
\mathcal{P}_r &= \mathcal{P}_r \cap \mathcal{P} && (\because \mathcal{P}_r \subset \mathcal{P}) \\
&= \cap_{i=1}^n \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\} \cap \text{aff}.P_r \cap \mathcal{P} \\
&= \cap_{i=1}^n \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\} \cap \mathcal{P} && (\because \mathcal{P} \subset \text{aff}.\mathcal{P}) \\
&= \cap_{i=1}^n \{\rho' \in \mathcal{P} | \rho' \cdot t_i \geq \alpha_i\}.
\end{aligned}$$

Hence

$$\mathcal{P}_r = \cap_{i=1}^n \{\rho' \in \mathcal{P} | \rho' \cdot t_i \geq \alpha_i\} \tag{17}$$

and

$$\text{rint}.\mathcal{P}_r = \cap_{i=1}^n \{\rho' \in \mathcal{P} | \rho' \cdot t_i > \alpha_i\}. \tag{18}$$

Since X is in general position, it follows from Proposition 1 and Remark 2 that $\mathcal{P}_{ml}(d) = \text{rint}.\mathcal{P}_r$ for some positive integer d. Hence

$$\mathcal{P}_{ml}(d) = \cap_{i=1}^n \{\rho' \in \mathcal{P} | \rho' \cdot t_i > \alpha_i\}. \tag{19}$$

Fix any $i \in \{1, \ldots, n\}$. I will show that there exist $\pi, \pi' \in \Pi$ such that $\rho^\pi \cdot t_i \neq \rho^{\pi'} \cdot t_i$. Suppose, by way of contradiction, that for all $\pi, \pi' \in \Pi$, $\rho^\pi \cdot t_i = \rho^{\pi'} \cdot t_i$. Let $\alpha'_i \equiv \rho^\pi \cdot t_i$ for some $\pi \in \Pi$. Since $\rho^\pi \in \mathcal{P}_r$ and (18) holds, I have $\alpha'_i \geq \alpha_i$. Then, $\text{aff}.\mathcal{P}_r = \text{aff.co}.\{\rho^\pi | \pi \in \Pi\} = \text{aff}.\{\rho^\pi | \pi \in \Pi\} \subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i = \alpha'_i\} \subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\}$. This is a contradiction to the fact that $\text{aff}.\mathcal{P}_r \not\subset \{q \in \mathbf{R}^{\mathcal{D}\times X} | q \cdot t_i \geq \alpha_i\}$ for all $i \in \{1, \ldots, n\}$. By Lemma 8, the existence of $\pi, \pi' \in \Pi$ such that $\rho^\pi \cdot t_i \neq \rho^{\pi'} \cdot t_i$ implies that $t_i(D, \cdot)$ is nonconstant on some $D \in \mathcal{D}$.

Now I will define u_i for each t_i. First, for each $i \in \{1, \ldots, n\}$, define $\beta_i = \max_{(D,x)\in\mathcal{D}\times X \text{ s.t. } t_i(D,x)<0}(-t_i(D, x))$. For any $(D, x) \in \mathcal{D} \times X$ such that $x \in D$, define $u_i(D, x) = t_i(D, x) + \beta_i$. Then $u_i(D, x) \geq 0$. For any $(D, x) \in \mathcal{D} \times X$ such that $x \notin D$, define $u_i(D, x) = 0$. Moreover, $u_i(D, \cdot)$ is nonconstant on some $D \in \mathcal{D}$ because $t_i(D, \cdot)$ is nonconstant on some $D \in \mathcal{D}$. It follows that $u_i \in \mathcal{U}$.

For all $i \in \{1, \ldots, n\}$, I will show $\min_{\pi\in\Pi} G(\rho^\pi : u_i) \geq \alpha_i + \beta_i|\mathcal{D}|$. To see this note that for any $\pi \in \Pi$, $\rho^\pi \cdot t_i \geq \alpha_i$ by (17). So $G(\rho^\pi : u_i) = \rho^\pi \cdot u_i = \rho^\pi \cdot t_i + \sum_{D\in\mathcal{D}} \sum_{x\in D} \beta_i \rho^\pi(D, x) = \rho^\pi \cdot t_i + \beta_i|\mathcal{D}| \geq \alpha_i + \beta_i|\mathcal{D}|$. Thus $\min_{\pi\in\Pi} G(\rho^\pi : u_i) \geq \alpha_i + \beta_i|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$.

By a similar calculation, I have $G(\rho : u_i) = \rho \cdot u_i = \rho \cdot t_i + \sum_{D\in\mathcal{D}} \sum_{x\in D} \beta_i \rho(D, x) = \rho \cdot t_i + \beta_i|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$.

Remember that Aggregated Stochastic Rationality requires $G(\rho : u_i) > \min_{\pi\in\Pi} G(\rho^\pi : u_i)$ for each $i \in \{1, \ldots, n\}$. Hence, by the above inequalities, $\rho\cdot t_i+\beta_i|\mathcal{D}| > \alpha_i+\beta_i|\mathcal{D}|$, so that $\rho \cdot t_i > \alpha_i$ for all $i \in \{1, \ldots, n\}$. Therefore, I have $\rho \in \cap_{i=1}^n\{\rho' \in \mathcal{P}|\rho' \cdot t_i > \alpha_i\} = \mathcal{P}_{ml}(d)$ by (19).

E Proof of Corollaries

E.1 Proof of Corollary 1

To see the necessity of Aggregated Logit Rationality, fix any $u \in \mathcal{U}$ and $\rho \in \mathcal{P}_{ml}$. Then by Remark 1 and Lemma 6, $\mathcal{P}_{ml} = \text{co}.\mathcal{P}_l \subset \text{rint}.\mathcal{P}_r$. Hence, $\rho \in \text{rint}.\mathcal{P}_r$. Notice that

$$\inf_{\rho'\in\mathcal{P}_l} G(\rho' : u) = \inf_{\rho'\in\text{co}.\mathcal{P}_l} G(\rho' : u) = \inf_{\rho'\in\text{rint}.\mathcal{P}_r} G(\rho' : u) = \min_{\rho'\in\mathcal{P}_r} G(\rho' : u) < G(\rho : u),$$

where the first equality holds because $G(\rho' : u)$ is linear in ρ', the second equality holds because $\text{co}.\mathcal{P}_l = \text{rint}.\mathcal{P}_r$ (by Remark 1 and Proposition 1 and the assumption that X is in general position), the third equality holds because $G(\rho' : u)$ is continuous in ρ' and $\mathcal{P}_r$ is compact, and the last strict inequality holds because G is linear in ρ', $\mathcal{P}_r$ is closed, and $\rho \in \text{rint}.\mathcal{P}_r$.

The sufficiency part of the proof is the same as the proof of Theorem 1 except the last part. For all $i \in \{1, \ldots, n\}$, I will show $\inf_{\rho_l\in\mathcal{P}_l} G(\rho_l : u_i) \geq \alpha_i + \beta_i|\mathcal{D}|$. To see this note that for any $\rho_l \in \mathcal{P}_l$, $\rho_l \cdot t_i > \alpha_i$ by (18). So $G(\rho_l : u_i) = \rho_l \cdot u_i = \rho_l \cdot t_i + \sum_{D\in\mathcal{D}}\sum_{x\in D}\beta_i\rho_l(D, x) = \rho_l \cdot t_i + \beta_i|\mathcal{D}| > \alpha_i + \beta_i|\mathcal{D}|$. Thus $\inf_{\rho_l\in\mathcal{P}_l} G(\rho_l : u_i) \geq \alpha_i + \beta_i|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$.

Moreover as in the proof of Theorem 1, $G(\rho : u_i) = \rho \cdot t_i + \beta_i|\mathcal{D}|$ for all $i \in \{1, \ldots, n\}$. If ρ satisfies Aggregated Logit Rationality, then $G(\rho : u_i) > \inf_{\rho_l\in\mathcal{P}_l} G(\rho_l : u_i)$ for all $i \in \{1, \ldots, n\}$. Hence $\rho \cdot t_i + \beta_i|\mathcal{D}| > \alpha_i + \beta_i|\mathcal{D}|$, so that $\rho \cdot t_i > \alpha_i$ for all $i \in \{1, \ldots, n\}$. Therefore, $\rho \in \cap_{i=1}^n\{\rho' \in \mathcal{P}|\rho' \cdot t_i > \alpha_i\} = \mathcal{P}_{ml}$ by (19).

Finally, I will provide an alternative proof for the sufficiency part. Suppose by the way of contradiction that $\rho \notin \mathcal{P}_{ml}$. By Remark 1, $\rho \notin \text{co}.\mathcal{P}_l$. By a separating hyperplane theorem, there exists $t \in \mathbf{R}^{\mathcal{D}\times X}$ such that

$$\rho \cdot t \leq \rho' \cdot t \text{ for all } \rho' \in \mathcal{P}_l \text{ and } \rho \cdot t < \rho'' \cdot t \text{ for some } \rho'' \in \mathcal{P}_l.$$

Define $\beta = \max_{(D,x)\in\mathcal{D}\times X \text{ s.t. } t(D,x)<0}(-t(D,x))$. For any $(D,x) \in \mathcal{D}\times X$ such that $x \in D$, define $u(D,x) = t(D,x)+\beta$. Then $u(D,x) \geq 0$. For any $(D,x) \in \mathcal{D}\times X$ such that $x \notin D$, define $u(D,x) = 0$.

Then $G(\rho : u) = \rho \cdot t + \beta|\mathcal{D}| \leq \rho' \cdot t + \beta|\mathcal{D}| = G(\rho' : u)$ for all $\rho' \in \mathcal{P}_l$ and $G(\rho : u) = \rho \cdot t + \beta|\mathcal{D}| < \rho'' \cdot t + \beta|\mathcal{D}| = G(\rho'' : u)$ for some $\rho'' \in \mathcal{P}_l$. This implies that

$$G(\rho : u) \leq \inf_{\rho'\in\mathcal{P}_l} G(\rho' : u).$$

Moreover, there must exist $D \in \mathcal{D}$ such that $u(D,x) \neq u(D,y)$ for some $x, y \in D$. Otherwise for each $D \in \mathcal{D}$ there exists v_D such that $u(D,x) = v_D$ for all $x \in D$. Then for any $\hat{\rho} \in \mathcal{P}$, $G(\hat{\rho} : u) = \sum_{D\in\mathcal{D}} v_D$ because $\sum_{x\in D}\hat{\rho}(D,x) = 1$. This contradicts with the fact that $G(\rho : u) < G(\rho'' : u)$.

Hence, $G(\rho : u) \leq \inf_{\rho'\in\mathcal{P}_l} G(\rho' : u)$ and $u \in \mathcal{U}$. This contradicts with Aggregated Logit Rationality.

E.2 Proof of Corollaries 2 and 3

By Proposition 1, the relative interior of the set of random utility functions is the set of mixed logit functions with polynomials of degree $d = 1$ (i.e., $p_d(x) = x$) if and only if X is affinely independent. Hence Corollary 2 holds.

By modifying the proof of Proposition 1, it is easy to show that for each $\pi \in \Pi$ there exists a sequence $\{\rho_n\}$ of general logit functions such that $\rho_n \to \rho^\pi$. Hence by Lemma 3, the relative interior of the set of random utility functions is the set of general mixed logit functions. (This result holds without any condition on X except for the finiteness.) Hence Corollary 3 holds.

References

Ahumada, A. and L. Ulku (2017): "Luce rule with limited consideration," Working paper.

Block, H. D. and J. Marschak (1960): "Random orderings and stochastic theories of responses," *Contributions to probability and statistics*, 2, 97–132.

Cerreia-Vioglio, S., D. Dillenberger, P. Ortoleva, and G. Riella (2017): "Deliberately stochastic," Working paper.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and A. Rustichini

——— (2018): "Multinomial logit processes and preference discovery: inside and outside the black box," Working paper.

Chiong, K. X., A. Galichon, and M. Shum (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7, 83–115.

Dogan, S. and K. Yildiz (2018): "Choice through a unified lens: The prudential model," Working paper.

Echenique, F. and K. Saito (2015): "General Luce model," Working paper.

Fudenberg, D., R. Iijima, and T. Strzalecki (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

Fudenberg, D. and T. Strzalecki (2015): "Dynamic logit with choice aversion," *Econometrica*, 83, 651–691.

Gul, F., P. Natenzon, and W. Pesendorfer (2014): "Random choice as behavioral optimization," *Econometrica*, 82, 1873–1912.

Gul, F. and W. Pesendorfer (2006): "Random Expected Utility," *Econometrica*, 74, pp. 121–146.

Horan, S. (2018): "Extended Luce rules," Working paper.

Lu, J. (2016): "Random Choice and Private Information," *Econometrica*, 84, 1983–2027.

Lu, J. and K. Saito (2016): "Random intertemporal choice," Working paper.

Matějka, F. and A. McKay (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–98.

McFadden, D. (1978): "Modeling the choice of residential location," *Transportation Research Record*.

——— (1981): "Econometric models of probabilistic choice," *Structural analysis of discrete data with econometric applications*, 198272.

——— (2001): "Economic choices," *American economic review*, 91, 351–378.

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 447–470.

Rockafellar, R. T. (2015): *Convex analysis*, Princeton university press.

Rooderkerk, R. P., H. J. Van Heerde, and T. H. Bijmolt (2011): "Incorporating context effects into a choice model," *Journal of Marketing Research*, 48, 767–780.

Rust, J. (1994): "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4, 3081–3143.

Soltan, V. (2015): *Lectures on convex sets*, World Scientific.

Stoer, J. and C. Witzgall (2012): *Convexity and optimization in finite dimensions I*, vol. 163, Springer Science & Business Media.

Strzalecki, T. (2018): "Hotelling Lectures: Stochastic Choice," URL: `https://scholar.harvard.edu/files/tomasz/files/lisbon32-post.pdf`.

Swait, J., W. Adamowicz, M. Hanemann, A. Diederich, J. Krosnick, D. Layton, W. Provencher, D. Schkade, and R. Tourangeau (2002): "Context dependence and aggregation in disaggregate choice analysis," *Marketing Letters*, 13, 195–205.