Average choice

David Ahn — UC Berkeley
Federico Echenique — Caltech
Kota Saito — Caltech

November 13, 2015

Abstract

This is an investigation of stochastic choice when only the average of the choices is observable. For example when one observes aggregate sales numbers from a store, but not the frequency with which each item was purchased. The focus of our work is on the Luce model, also known as the Logit model. We show that a simple path independence property of average choice uniquely characterizes the Luce model. We also characterize the linear Luce mode, using similar tools. A linear version of the Luce model is used most frequently in empirical work by applied economists.

Our characterization is based on the property of *path independence*, which runs counter to early impossibility results on path independent choice. From an empirical perspective, our results provide a small-sample advantage over the tests of Luce's model that rely on estimating choice frequences.

1. Introduction

We investigate the meaning of stochastic choice theory for data on average choices. Aggregate or average choices are often available when exact choice frequencies are unobservable. Such aggregate statistics are not only easy to obtain, but also easier to estimate efficiently than are choice frequencies. Taking average choice as primitive, we introduce a novel characterization of the ubiquitous Luce, or *Logit* model of discrete choice.

The crucial property for the Luce model is a weak version of *path independence.* Classic path independence requires choices to be recursive moving from subsets to larger sets. In particular, first making two selection from two disjoint sets and then choosing among the pair of selections leads to the final outcome as immediately selecting from the union of these sets. For example in a supermarket we may choose Camembert cheese from the cheese section. Then choose from a simplified representation of the supermarket in which Camembert is the only cheese available. Our notion of path independence is weaker, requiring the average choice to be any convex combination of the average choices from a partition of the set, rather than requiring the average choice to be the average choice from the average choice.

Our result is surprising because path independence was thought to be too strong a property to impose on average choice. Path independence was first proposed by Plott (1973), who used it to characterize rational choice in an abstract environment. Kalai and Megiddo (1980) and Machina and Parks (1981) consider path independence for average choice, essentially the same primitive as we assume in our paper. These papers present impossibility theorems. They conclude that no continuous average choice can satisfy Plott path independence. We consider a weakening of path independence, where the average choice from a set must be a convex combination (a lottery) of the average choices from a partition of the set in two. It does not need to coincide with the average choice one would make when faced only with the two average choices. See Section 4.2 for a comparison of the two axioms.

Our weakening of path independence not only avoids the impossibility result, but together with continuity it pins down the Luce model uniquely. Our main result says that the unique average choice that satisfies continuity and path independence is the Luce average rule.

We proceed to expand on why average choices are a reasonable primitive for stochastic choice, then explain our main contributions in some detail, and finally discuss some related literature.

Average choices. We may often observe aggregate data on random choice, and not the actual choice probabilities. There are two interpretations of the model of random choice, and in each case it makes sense to think that one may be able to observe average choice instead of choice probabilities.

The first interpretation of random choice is that it reflects the choice frequencies of a population of individual agents. While each individual has determinate choice behavior, the randomness of the choice function records population-level heterogeneity. We claim that one may often actually observe the aggregate quantities chosen and not the choice frequencies. For example, the ratio of wine sales to total revenue at a supermarket can be immediately read from a balance sheet without observing each shopper's receipt. The average choice economizes on describing each product's distribution of sales (it does not require knowing how many shoppers bought over a dozen bottles of wine) and on describing each product's correlation with others (it does not require knowing how frequently cheese is purchased with wine).

The store example makes sense if we interpret random choice as the behavior of a population of individuals, each one with an idiosyncratic choice behavior. We may instead focus on a single individual, who may have different choice behaviors at different points in time. But even for a single individual it is reasonable to think that we can observe average choice and not choice frequencies. For example we may ask an individual how much wine, cheese and meat they purchased last month. This is much easier to answer compared to the detailed record of each trip to the store that would be required to compile information on choice frequencies.

Finally, even if one has information on choice frequencies it is *statistically* desirable to use a test based on average choice. Studies of stochastic choice take population choice probabilities as known. But in actual applications of the theory one only observes a finite sample of choices. It is therefore important to understand how one would test for the axioms that characterize the theory, given a finite-sample approximation to the population stochastic choice probabilities.

The average is statistically very well behaved while the main alternative to our characterization uses relative probabilities; these have much poorer qualities as a test statistic. The alternative to our test is Luce's axiomatization, based on testing for the independence of irrelevant alternatives (IIA). Testing IIA requires estimating relative choice probabilities: these are statistically more complicated objects than the average. In small samples they are hard to estimate robustly, and even with a large sample their estimators can have an arbitrarily large variance. We discuss these issues in detail in Section 4.4.

In fact, Luce (1959) already makes this point. Luce makes a very similar calculation to our calculations in Section 4.4, and concludes that a reliable test based on his IIA axiom woud require very large datasets: Sample sizes of the order of several thousands. This would render the tests inapplicable in most experimental setups.[1] Our axiomatization, based on average choice, is in that sense much better behaved because averages can be estimated more efficiently than fractions of probabilities.

Contributions. Our main contribution is an axiomatic characterization of the Luce model when the primitive is average choice. The Luce model is very important in economics, and very heavily used by applied economists and practitioners.

Our axiomatization is based on path independence, as described in the beginning of the introduction. The axiomatization is useful as a statistical test for the Luce model because data on average choices are a) available when data on choice frequencies are not, and b) more robustly estimated given data on choices.

The axiomatization is also useful because of what we learn about the Luce mode. Path independence of average choice has been studied before, and the connection to Luce's IIA went unnoticed. As mentioned in the beginning of the introduction, the literature concluded that path independence of average choice was too strong a property, and lead to an impossibility result. So the fact that a weak version of path independence lies behind the Luce model is probably surprising, and uncovers a new facet of the Luce model. In the paper (Section 4.1) we discuss the relation between IIA and path independence, and why one cannot prove our main result by simply showing that path independence must imply the IIA.

Our second contribution is to provide the first axiomatization of the *linear* Luce model. The Luce model (or Logit model) is ubiquitous in applied economics: It is taught to every first-year graduate student; and it is the basic go-to specification for the demand side of most structural economic models. But the version of the model that is actually empirically estimated assumes that utility is linear in object attributes. As far as we know, the full empirical content of this model has not been investigated

[1] Luce suggests that one would use psychophysical experiments to obtain the kinds of sample sizes needed to test IIA.

until now. For example, in McFadden (1974), the influential paper that lays out the Logit model as an empirical strategy, linearity is postulated as an axiom (McFadden's Axiom 4).

We show that an average choice version of the von-Neumann Morgenstern independence axiom ensures that the Luce utility function is ordinally equivalent to a linear function. After discussing this axiom, we turn to a kind of calibration axiom that ensures that the Luce utility is affinely equivalent to a linear utility. These axioms, together with continuity and path independence, constitute the content of the linear Luce model for average choice.

Related Literature. The papers that are closely related to ours are: Luce (1959), Plott (1973), Kalai and Megiddo (1980), and Machina and Parks (1981). These have been discussed above. In particular, the papers by Kalai and Megiddo (1980) and Machina and Parks (1981) use the same primitive as we do, and argue that path independence gives rise to an impossibility. We weaken path independence and find that it gives rise to the Luce rule. Section 4.2 explains the Plott path independence axiom in detail, and gives a simplified explanation behind the impossibility results.

The literature on path independent choice is large. Notable contributions include the work of Blair (1975), Parks (1976), and Ferejohn and Grether (1977). This work follows Plott in assuming an environment of abstract choice, without the linear structure that is part of our primitives.

The linear Luce model is the standard econometric specification of the Luce, or Logit model. A discussion can be found in any econometric textbook: see for example Greene (2003). The Logit model as an empirical strategy is laid out in McFadden (1974). We are apparently the first to investigate the behavioral content of the linearity assumption in the Luce model.

2. Average choice and the Luce model

2.1. Notation

Definitions of mathematical terms are in Section 5.1.

If $A \subseteq \mathbf{R}^n$, then conv$A$ denotes the the set of all convex combinations of elements in A, termed the *convex hull* of A. Let conv$^0 A$ denote the relative interior of conv$A$.

A binary relation on a set is called a *preference relation* if it is complete and transitive (in other words, if it is a *weak order*).

Let Y be a set. Then $f : Y \to \mathbf{R}_+$ is a *finite support measure* on Y if there is a finite subset $A \subseteq Y$ such that $f(x) = 0$ for $x \notin A$. In this case we say that the set A is a *support* for f. A finite support measure may have more than one support. If f in addition satisfies that $\sum_{x \in A} f(x) = 1$ then we say that it is a *finite support probability measure* on X. When Y is a finite set, we use the term *lottery* to refer to a finite support probability measure.

Let $\Delta(Y)$ denote the set of all finite support probability measures; and for $A \subset Y$, let $\Delta(A)$ denote the set of finite support measures that have support A. We shall only discuss finite support probability measures, so we some times refer to them as probability measures.

2.2. Primitives

Let X be a compact and convex subset of $\mathbf{R}^n$, with $n \geq 3$. An important special case is when X is a set of lotteries, in that case there is a finite set P of *prizes*, and $X = \Delta(P)$ is the set of all lotteries over P.

Let $\mathcal{A}$ be the set of all finite subsets of X.

An *average choice* is a function

$$\rho^* : \mathcal{A} \to X,$$

such that, for all $A \in \mathcal{A}$, $\rho^*(A)$ is in conv$^0 A$, the relative interior of the convex hull of A.

The primitive of our study is an average choice.

2.3. Luce model

A *stochastic choice* is a function

$$\rho : \mathcal{A} \to \Delta(X)$$

such that $\rho(A) \in \Delta(A)$.

A stochastic choice $\rho : \mathcal{A} \to \Delta(X)$ a *continuous Luce rule* if there is a continuous function $u : X \to \mathbf{R}_{++}$ such that

$$\rho(x, A) = \frac{u(x)}{\sum_{y \in A} u(y)}.$$

An average choice ρ^* is *continuous Luce rationalizable* if there is a continuous Luce rule ρ such that

$$\rho^*(A) = \sum_{x \in A} x\rho(x, A)$$

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

2.4. A characterization of continuous Luce rationalizable average choices.

Our main axiom is a version of the *path independence* axiom proposed by Plott (1973). Plott's axiom says that choice is recursive, in the sense that choice from $A \cup B$ is obtained by choosing first from A and B separately, and then from a set consisting of the two chosen elements. The idea is that all that matters about A and B in $A \cup B$ are its "best" elements, the element chosen from A and B. Plott worked in an environment of abstract choice. For average choice, the same environment as we use in our paper, the papers by Kalai and Megiddo (1980) and Machina and Parks (1981) argue that Plott's axiom is too strong. They show that Plott's path independence gives an impossibility result. Our result is that a natural weakening of Plott's axiom does not give an impossibility but instead pins down the Luce model.

Our notion of path independence is weaker than Plott's, and says that the average choice from $A \cup B$ is a convex combination of the average choice from A and the average choice from B. Formally, our axiom is:

Path independence: *If $convA \cap convB = \emptyset$ then*

$$\rho^*(A \cup B) \in conv^0\{\rho^*(A), \rho^*(B)\}.$$

Our second axiom imposes a continuity property on average choice.

Continuity: *Let $x \notin A$. For any sequence x_n in X, if $x = \lim_{n\to\infty} x_n$, then*

$$\rho^*(A \cup \{x\}) = \lim_{n\to\infty} \rho^*(A \cup \{x_n\}).$$

(*understood as weak* convergence*).

Continuity is a technical axiom but it raises some interesting issues related to choosing among multiple copies of identical, or close to identical, objects. We discuss these in Section 4.

Theorem 1. *An average choice is continuous Luce rationalizable iff it satisfies continuity and path independence.*

We can gain some intuition for Theorem 1 from Figure 1. Essentially what happens is that path independence pins down $\rho^*(A)$ uniquely from the value of $\rho^*(A')$ when $A' \subsetneq A$.[2] We build up a rationalizing Luce model from the average choice at sets of small cardinality, and then use path independence to show that ρ^* must always coincide with the average choice generated from that Luce model.

Figure 1 shows how $\rho^*(A)$ is uniquely determined. In the figure, the points x and y are extreme points of conv(A), where $A = \{x, y, z, w, q, r\}$. Path independence says that $\rho^*(A)$ must lie on the line segment joining x and $\rho^*(A \setminus \{x\})$ and on the line segment joining y and $\rho^*(A \setminus \{y\})$. The main step in the proof (Lemma 6) establishes that, outside of some "non-generic" sets A, the intersection of the line segments x—$\rho^*(A \setminus \{x\})$ and y—$\rho^*(A \setminus \{y\})$ is a singleton. This means that path indepence pins down $\rho^*(A)$ uniquely from $\rho^*(A \setminus \{x\})$ and $\rho^*(A \setminus \{y\})$.

Figure 1 illustrates why the line segments x—$\rho^*(A \setminus \{x\})$ and y—$\rho^*(A \setminus \{y\})$ have a singleton intersection. We choose x and y to lie on a proper face of conv$A$, so there is a hyperplane supporting A at that face. If there were two points in the intersection of x—$\rho^*(A \setminus \{x\})$ and y—$\rho^*(A \setminus \{y\})$, then all four points x, $\rho^*(A \setminus \{x\})$, y and $\rho^*(A \setminus \{y\})$ would lie in the same line. This implies that $\rho^*(A \setminus \{x\})$ and $\rho^*(A \setminus \{y\})$ lie on the hyperplane as well. But since these two average choices are in the relative interior of the respective sets, it implies that A also lies in the hyperplane. This is

[2] See also the discussion in Section 4.1, where we compare path independence and Luce's IIA, and argue that we could not use IIA to the same effect as path independence.

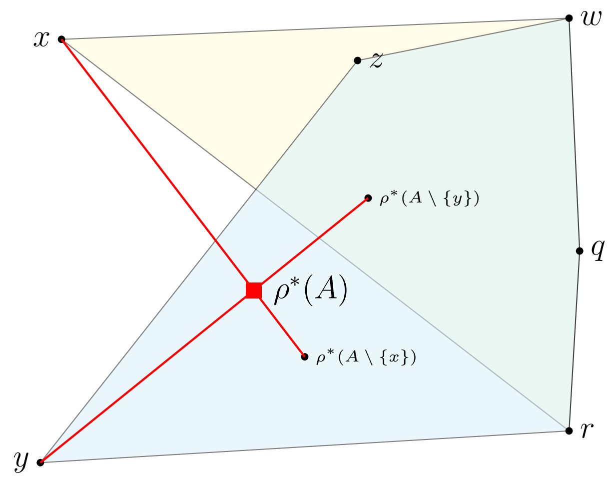


Figure 1: The set $A = \{x, y, z, w, q, r\}$: $\rho^*(A)$ is determined from path independence, $\rho^*(A \setminus \{x\})$, and $\rho^*(A \setminus \{y\})$

what we mean by A being "not generic:" it means that conv$A$ does not have low dimension.

3. Linear Luce

We have mentioned the popularity of the Luce model among practitioners and empirical economists. Interestingly, the version of the Luce model they estimate is linear in object attributes. Here we turn our attention to such a linear model. Specifically, we consider continuous Luce models in which:

- $u(x) = f(v \cdot x)$ with $v \in \mathbf{R}^n$ and f a continuous strictly monotonic function;
- or $u(x) = v \cdot x + \beta$, with $v \in \mathbf{R}^n$ and $\beta \in \mathbf{R}$.

The first model we term the *linear Luce model.* The second is called the *strictly affine Luce model.*

Our results on linear Luce and the strictly affine Luce model are the first axiomatic characterizations of any version of Luce model that assumes linearity in object attributes, the kind of Luce model that is usually estimated empirically.

3.1. Linear Luce

A stochastic choice $\rho : \mathcal{A} \to \Delta(X)$ a *linear Luce rule* if there is $v \in \mathbf{R}^n$ and a monotone and continuous function $f : \mathbf{R} \to \mathbf{R}_{++}$ such that

$$\rho(x, A) = \frac{f(v \cdot x)}{\sum_{y \in A} f(v \cdot y)}.$$

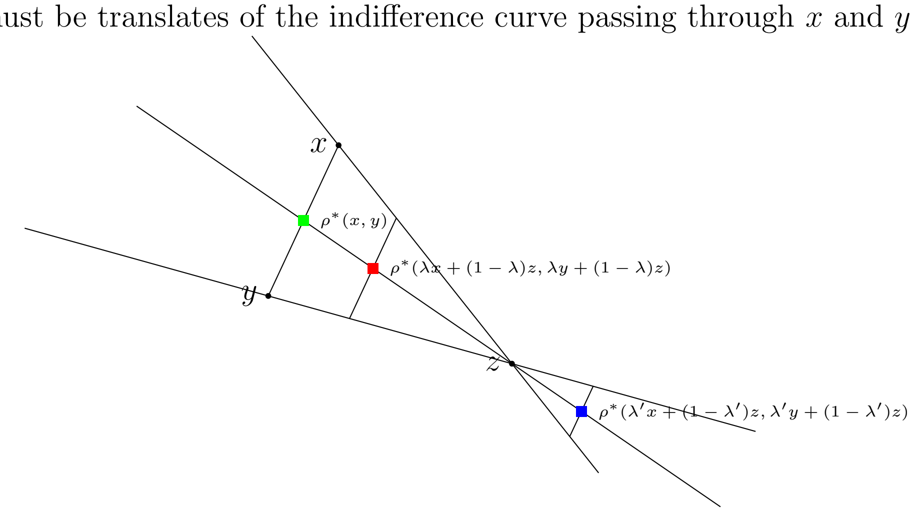
Consider the following axiom, which captures the kind of independence property that is normally associated with von Neumann-Morgenstern expected utility theory.

Independence: $\rho^*(\{x, y\}) = (1/2)x + (1/2)y$ *iff for all* $\lambda \in \mathbf{R}$ *and all* z *s.t.* $\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X$,

$$\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) = \lambda \rho^*(\{x, y\}) + (1-\lambda)z$$

Note that if ρ^* is continuous Luce rationalizable with utility u, then $u(x) = u(y)$ iff $\rho^*(\{x, y\}) = (1/2)x + (1/2)y$. So the meaning of the independence axiom is that $u(x) = u(y)$ iff for all $\lambda \in \mathbf{R}$ and all z s.t. $\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X$, $\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) = \lambda \rho^*(\{x, y\}) + (1-\lambda)z$.

The next diagram is a geometric illustration of independence. Note that the average choice $\rho^*(\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z)$ must be translated with λ from $\rho^*(x, y)$. Since we have $u(x) = u(y)$, this will force a rationalizing rule to translate indifference curves in λ in a similar fashion, so the indifference curves passing through the points $\lambda x + (1-\lambda)z$ and $\lambda y + (1-\lambda)z)$, and through $\lambda' x + (1-\lambda')z$ and $\lambda' y + (1-\lambda)z')$, must be translates of the indifference curve passing through x and y.



The independence axiom ensures that u satisfies von Neumann-Morgenstern independence. The immediate implication of the axiom (Lemma 8) says that u satisfies a weak version of independence, but this version can be shown to suffice for the result (Lemma 9). The result is as follows.

Theorem 2. *An average choice is continuous linear Luce rationalizable iff it satisfies independence, continuity and path independence.*

3.2. Strictly affine Luce model

A stochastic choice $\rho : \mathcal{A} \to \Delta(X)$ is a *strictly affine Luce rule* if there is $v \in \mathbf{R}^n$ and $\beta \in \mathbf{R}$ such that

$$\rho(x, A) = \frac{v \cdot x + \beta}{\sum_{y \in A}(v \cdot y + \beta)}.$$

In the linear model, u is ordinally equivalent to a linear function of x. The so-called strictly affine model has u being an affine transformation of a linear model. Independence is not sufficient to obtain this model and we need a stronger axiom; an axiom that allows us to measure the consequences of u being affine on average choice. This axiom is termed calibration:

Calibration:

$$\rho^*(\{\lambda x+(1-\lambda)y, \lambda y+(1-\lambda)x\}) = \rho^*(\{x,y\})+2\lambda(1-\lambda)[(x-\rho^*(\{x,y\}))+(y-\rho^*(\{x,y\}))]$$

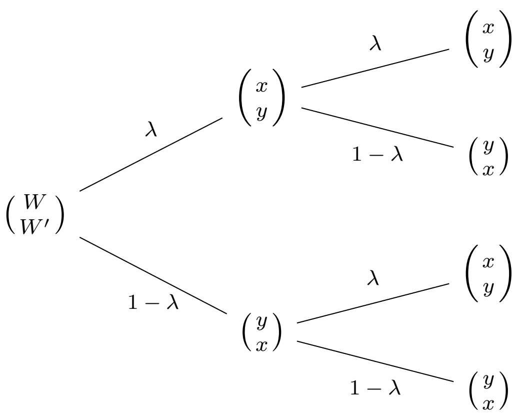
Theorem 3. *An average choice is strictly affine Luce rationalizable iff it satisfies calibration, independence, continuity and path independence.*

The calibration axiom looks messy but it has a straightforward interpretation. Suppose that ρ^* is Luce rationalizable with a utility function u. Calibration is going to force u to be affine. Specifically that if W is a random vector, then $u(\mathbf{E}W) = \mathbf{E}u(W)$. We use $\mathbf{E}$ to denote mathematical expectation. Now, u is not a primitive of our model, so we must detect the consequences of $u(\mathbf{E}W) = \mathbf{E}u(W)$ for average choice.

It turns out that the consequences are easy to read from "symmetric" combinations:

$$\lambda x + (1 - \lambda)y \text{ and } \lambda y + (1 - \lambda)x$$

Think of the vectors $w = \lambda x + (1-\lambda)y$ and $w' = \lambda y + (1-\lambda)x$ as two perfectly correlated lotteries. Write W for the random vector that takes x with probability λ and y with probability $1-\lambda$. Similarly define a random vector W' from w'. So $w = \mathbf{E}W$ and $w' = \mathbf{E}W'$. The following diagram illustrates the two lotteries.



When ρ^* is Luce rationalizable by a function u, we have that

$$\rho^*(\{\lambda x + (1-\lambda)y, \lambda y + (1-\lambda)x\}) = \frac{u(\mathbf{E}W)\mathbf{E}W + u(\mathbf{E}W)\mathbf{E}W}{u(\mathbf{E}W) + u(\mathbf{E}W)}$$

Because we focus on symmetric lotteries, when u is a strictly affine Luce model, $u(\mathbf{E}W) + u(\mathbf{E}W') = \mathbf{E}u(W) + \mathbf{E}u(W') = u(x) + u(y)$ is independent of λ. So the calibration we seek is reflected in the numerator.

Recall that we want to calculate the consequences of $u(\mathbf{E}W) = \mathbf{E}u(W)$ and $u(\mathbf{E}W') = \mathbf{E}u(W')$. So consider the formula for ρ^* where we use expected utility in place of the utility of the expectation. In this formula for ρ^* we randomize twice. First we randomize according to W and W' to determine $u(W)$. Then we randomize according to W and W' to determine the value of W and W' that is weighted by $u(W)$ or $u(W')$. This means that with probability $\lambda^2 + (1-\lambda)^2$ we draw $u(x)x + u(y)y$ and with probability $2\lambda(1-\lambda)$ we draw the "mismatched" $u(x)y + u(y)x$. That is:

$$\begin{aligned}\rho^*(\{\lambda x + (1-\lambda)y, \lambda y + (1-\lambda)x\}) = (\lambda^2 + (1-\lambda)^2)\frac{u(x)x + u(y)y}{u(x) + u(y)} \\ + (2\lambda(1-\lambda))\frac{u(y)x + u(x)y}{u(x) + u(y)}.\end{aligned}$$

The behavioral version of the latter formula is precisely the calibration axiom.

The work in proving Theorem 3 lies in showing that the rather special consequence of u being affine expressed in the calibration axiom is sufficient to characterize affine u.

4. Discussion

An axiomatic characterization serves two kinds of purposes. One is that the ideas behind the axioms give some qualitative meaning to a quantitative model of choice. Path independence captures a procedural approach to choice: this idea is most clearly captured by Plott's version of path independence. Plott's notion is too strong in our environment: we discuss these issues in Section 4.2.

Our second axiom, continuity, takes the form it takes because of a well-known phenomenon in stochastic choice. This is discussed in 4.3.

The other use of an axiomatic characterization is that it provides a test for the theory. Luce's model is the most widely used model of discrete choice. While you read this paper, dozens of economists are hard at work estimating hundreds of Luce models (Logit models) of choice. But Luce's characterization of the model relies on stochastic choice as primitive, we argue that there are practical empirical advantages to testing Luce's model using average choice. These are discussed in Section 4.4.

In the first place, however, we turn to a theoretical discussion of Luce's and our axiomatization of the Luce model.

4.1. Path independence and Luce's IIA.

Luce (1959) takes a stochastic choice ρ as primitive, and shows that Luce's rue emerges from the axiom:

$$\frac{\rho(x,\{x,y\})}{\rho(y,\{x,y\})} = \frac{\rho(x,A)}{\rho(y,A)}$$

for any $A \ni x, y$. It is interesting to consider the relation between IIA and path independence.

Figure 2 exhibits a violation of path independence. The average choice $\rho^*(\{x,y,z\})$ is not on the line connecting z with $\rho^*(\{x,y\}$. Instead, $\rho^*(\{x,y,z\})$ is above the line. If we write $\rho^*(\{x,y,z\})$ as a weighted average of the vectors x, y and z, then this

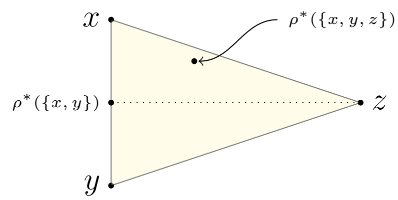


Figure 2: Path independence and Luces' IIA.

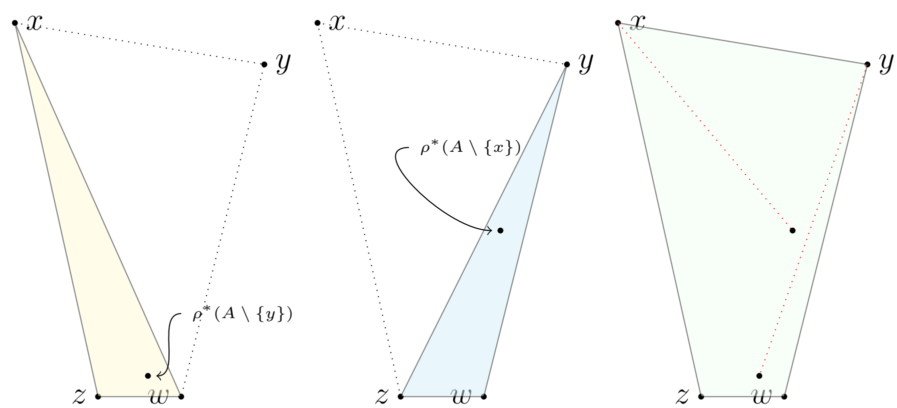


Figure 3: Violation of path independence; $A = \{x, y, z, w\}$.

means that the relative weight places on x, relative to the weight place on y, is clearly higher than the relative weight places on x in $\rho^*(\{x, y\})$. This would be a violation of Luce's IIA, when the weights are choice probabilities.

So it is easy to see, at least in some special cases, how path independence is related to IIA. It is not, however, so easy to prove Theorem 1 using IIA. Consider the violation of path independence in Figure 3.

Figure 3 exhibits a violation of path independence. The figure has the average choice from some subsets of $A = \{x, y, z, w\}$, and these choices are incompatible with path independence. It is easy to draw violations of path independence, but Figure 3 illustrates that the axiom can be violated given *only* information about the average choice from $\{x, z, w\}$ and $\{y, z, w\}$. This information does not allow us to reduce the problem to a violation IIA because we cannot relate the probability weights in a larger set to the probability weights in the smaller set.

The diagram on the left shows where $\rho^*(\{x, z, w\})$ is in $\text{conv}\{x, z, w\}$, while the diagram in the middle depicts $\rho^*(\{y,z,w\})$ in $\text{conv}\{y,z,w\}$. It should be clear, as illustrated in the diagram on the right of Figure 3, that path independence cannot be satisfied because the lines x–$\rho^*(A \setminus \{x\})$ and y–$\rho^*(A \setminus \{y\})$ do not cross.

The example in Figure 3 because it shows how the average choice at sets of smaller cardinality pin down uniquely the average choice at sets of larger cardinality. Luce's IIA cannot be used in the same way. To check for IIA, as suggested by Figure 2, requires the probability weights to be unique, but this is only possible for affinely independent sets of vectors. So it is only possible for sets A of small cardinality. Our proof of Theorem 1 also works inductively over the cardinality of A, but path independence (in contrast to IIA) does pin down average choice in a way that makes Luce's model unavoidable.

4.2. Plott path independence

Path indepence was introduced to choice theory by Plott (1973). In our model, Plott's notion of path independence translates into the axiom:

Plott path independence: *If* $convA \cap convB = \emptyset$ *then*

$$\rho^*(A \cup B) = \rho^*\left(\{\rho^*(A), \rho^*(B)\}\right).$$

The idea is that all that matters for choosing from $A \cup B$ are the choices from A and from B. Once we know $\rho^*(A)$ and $\rho^*(B)$ there is no more relevant information in the original sets A and B.

It is important to see how our path independence axiom weakens Plott's. We require only that $\rho^*(A \cup B)$ be a convex combination of $\rho^*(A)$ and $\rho^*(B)\}$, but the weights in this convex combination does not need to coincide with the weights in the choice from the set $\{\rho^*(A), \rho^*(B)\}$. So the elements of $A \cup B$ are allowed to influence the average choice $\rho^*(A \cup B)$ by way of affecting the weights placed on $\rho^*(A)$ and $\rho^*(B)\}$.

There is an extensive literature following the work of Plott. The most relevant for our purposes is Kalai and Megiddo (1980) and Machina and Parks (1981), who argue that Plott path independence is too strong an axiom in our framework. It is particularly simple to see Kalai's, Megiddo's, Machina's, and Park's point when we talk about Luce rationalizable average choice.

Proposition 4. *If an average choice is continuous Luce rationalizable, then it cannot satisfy Plott path independence.*

Proof. Let ρ^* be an average choice that is continuous Luce rationalizable, and let $u : X \to \mathbf{R}_{++}$ be the continuous utility function in the Luce model that rationalizes ρ^*. Suppose towards a contradiction that ρ^* satisfies Plott path independence.

Fix any affinely independent $x, y, z \in X$. Then Plott path independence implies that

$$\rho^*(\{x,y,z\}) = \rho^*(\rho^*(\{x,y\}), \{z\}) = \frac{u(\rho^*(\{x,y\}))\rho^*(\{x,y\}) + u(z)z}{u(\rho^*(\{x,y\})) + u(z)}.$$

But at the same time, $\rho^*(\{x,y,z\})$ also equals

$$\frac{u(x)x + u(y)y + u(z)z}{u(x) + u(y) + u(z)}.$$

Affine independence then implies that

$$\frac{u(z)}{u(x) + u(y) + u(z)} = \frac{u(z)}{u(\rho^*(\{x,y\})) + u(z)}.$$

Then $u(z) > 0$ gives that

$$u(\rho^*(\{x,y\})) = u\left(\frac{u(x)x + u(y)y}{u(x) + u(y)}\right) = u(x) + u(y).$$

One can choose y arbitrarily close to x such that x, y, z are affinely independent. Therefore, continuity of u would imply that $u(x) = 2u(x)$, which is impossible as $u(x) > 0$. $\square$

4.3. *Debreu's example*

The continuity axiom we are using is perhaps not the first such axiom one would think of (it certainly was not the first we thought of). A stronger axiom would demand that ρ^* is continuous, but this is incompatible with the Luce rule.

It is interesting to see why, as it relates the the famous "blue bus – red bus" example of Debreu (1960) and Tversky (1972). Suppose that an agent has to choose a mode of transportation. Choosing x means taking a bus while choosing y means taking a cab.[3] Alternative z is also a bus, but of a different color from x. Debreu argues that the presence of z should matter for the relative probability of choosing x over y, as the agent would consider whether to take a bus or a taxi and then be indifferent over which bus to take.

To see the connection to Debreu's example, suppose that ρ^* is continuous Luce rationalizable, with a continuous function u. Let z_n be a sequence in X converging to $x \in X$. Then

$$\rho^*(\{x, y\}) \neq \frac{2u(x)x + u(y)y}{2u(x) + u(y)} = \lim_{n\to\infty} \rho^*(\{x, y, z_n\}).$$

Thus ρ^* must be discontinuous.

The lack of continuity of ρ^* is reminiscent of the blue bus-red bus example. The lack of continuity results because the Luce model demands that z_n be irrelevant for the relative choice of x over y, even when z_n becomes very similar to x. In that sense, the lack of continuity of ρ^* is related to the blue bus-red bus phenomenon.

4.4. *Sampling error*

Theoretical studies of stochastic choice assume that choice probabilities are observed perfectly. But we intend our axioms to be useful as empirical tests, so that one can empirically decide whether observed data are consistent with the theory.

In an empirical study, however, choice probabilities must be estimated from sample frequencies. In the individual interpretation of the stochastic choice model (see the discussion in the Introduction), one agent would make a choice repeatedly from a set of available alternatives. This allows us to estimate the stochastic choice, but not to perfectly observe it. In the population interpretation (again, see the Introduction) we can observe the choices of a group of agents. The group may be large, and the fraction with which an agent makes a choice may be close to the population fraction

[3]Actually Debreu talked about two different recordings of the same Beethoven symphony vs. a suite by Debussy; but the literature has preferred buses and cabs.

of that choice, but there is likely some important randomness due to sampling. Here we want to argue that our axiomatization is better suited to dealing with such errors.

In his discussion of the Luce mode, Luce (1959) makes the point that testing his axiom, IIA, requires a large sample size. We follow up on Luce's remark and compare the efficiency of the IIA test statistic with the test statistic neded to test for path independence, the average of the choices.

Fix a set of alternatives A. Suppose that the population choices from A are given by $p \in \Delta(A)$, which comes from some continuous Luce rule $u : X \to \mathbf{R}_{++}$. We do not observe p but instead a sample $X_1, \ldots, X_n$ of choices, with $X_i \in A$ for $i = 1, \ldots, n$. The X_i are independent, and distributed according to p. The probability p is estimated from the empirical distribution:

$$p_x^n = \frac{|i : X_i = x|}{n}.$$

We have two possibilities to test the Luce model:

1. Test the model using Luce's original axiomatization, by computing relative probabilities

   $$\frac{p_x^n}{p_y^n},$$

   for $x, y \in A$.

2. Test the model using our axiomatization, by computing average choice:

   $$\mu^n = \sum_{x \in A} x p_x^n$$

We argue that option (2) is better for two reasons. One is that in small samples it is more reliable to use the average than to use relative probabilities. The other is that, even in a large sample, the test statistic in (1) can have a very large variance relative to the test statistic in (2). Specifically, for any M there is an instance of the Luce model in which the asymptotic variance of the statistic in (1) is M times the asymptotic variance of the statistic in (2).

We proceed to formalize the second claim.

Standard calculations (see Appendix B) yield:

$$\sqrt{n}\left(\frac{p_x^n}{p_y^n}-\frac{p_x}{p_y}\right)\xrightarrow{d} N\left(0,\sqrt{\frac{p_x}{np_y^2}\left[1+\frac{p_x}{p_y}\right]}\right)$$

On the other hand, $\sqrt{n}(\mu^n-\mu)\xrightarrow{d} N(0,\Sigma)$, where

$$\Sigma=(\sigma_{l,h}) \text{ and } \sigma_{l,h}=\sum_{x\in A}\sum_{y\in A\setminus x}-x_l y_h p_x p_y+\sum_{x\in A}x_l x_h p_x(1-p_x).$$

It is obvious that the entries of Σ are bounded and the asymptotic variance of p_x^n/p_y^n can be taken to be as large as desired by choosing a Luce model in which $u(x)/u(y)$ is large. So for any M there is a Luce model for which the asymptotic variance of p_a^n/p_b^n relative to $\max\{\sigma_{l,h}\}$, the largest element of Σ, is greater than M.

5. Proof of Theorem 1

5.1. *Definitions from convex analysis*

Let $x, y \in \mathbf{R}^n$. The *line* passing through x and y is the set $\{x+\theta(y-x) : \theta\in\mathbf{R}\}$. The *line segment* joining x and y is the set $\{x+\theta(y-x) : \theta\in[0,1]\}$. A subset of $\mathbf{R}^n$ is *affine* if it contains the line passing through any two of its members. A subset of $\mathbf{R}^n$ is *convex* if it contains the line segment joining any two of its members.

If A is a subset of $\mathbf{R}^n$, the *affine hull* of A is the intersection of all affine subsets of $\mathbf{R}^n$ that contain A. An *affine combination* of members of A is any finite sum $\sum_{i=1}^{l}\lambda_i x_i$ with $\lambda_i\in\mathbf{R}$ and $x_i\in A$ $i=1,\ldots,l$ and $\sum_{i=1}^{l}\lambda_i=1$. The affine hull of A is equivalently the collection of all affine combinations of its members. If A is a subset of $\mathbf{R}^n$, the *convex hull* of A is the intersection of all convex subsets of $\mathbf{R}^n$ that contain A. A set A is *affinely independent* if none of its members can be written as an affine combination of the rest of the members of A.

A point $x\in A$ is *relative interior for* A if there is a neighborhood N of x in $\mathbf{R}^n$ such that the intersection of N with the affine hull of A is contained in A. The *relative interior* of A is the set of points $x\in A$ that are relative interior for A.

A *polytope* is the convex hull of a finite set of points. The *dimension* of a polytope P is $l-1$ if l the largest cardinality of an affinely independent subset of A. A vector

$x \in A$ is an *extreme point* of a set A if it cannot be written as the convex combination of the rest of the members of A. A *face* of a convex set A is a convex subset $F \subseteq A$ with the property that if $x, y \in A$ and $(x+y)/2 \in F$ then $x, y \in F$. If F is a face of a polyope P, then F is also a polytope and there is $\alpha \in \mathbf{R}^n$ and $\beta \in \mathbf{R}$ such that $P \subseteq \{x \in \mathbf{R}^n : \alpha \cdot x \leq \beta\}$ and $F = \{x \in P : \alpha \cdot x = \beta\}$. If a polytope has dimension l then it has faces of dimension $0, 1, \ldots, l$ (Corollary 2.4.8 in Schneider (2013)).

5.2. Proof

The following axiom seems to be weaker than path independence, but it is not.

One-point path independence:

$$\rho^*(A) \in conv^0\{x, \rho^*(A \setminus \{x\})\}.$$

We use one-point path independence in the proof of sufficiency in Theorem 1. Since path independence is satisfied by any continuous Luce rationalizable average choice, one-point path independence and path independence are equivalent (at least under the hypothesis of continuity).

We now proceed with the formal proof. The first lemma establishes necessity, and is also needed in the proof of sufficiency.

Lemma 5. *If ρ^* is continuous Luce rationalizable, then it satisfies continuity and path independence.*

Proof. Continuity of the average choice is a direct consequence of the continuity of u. Path independence is also simple: Note that when $\text{conv}A \cap \text{conv}B = \emptyset$ then A and B are disjoint. So it follows that

$$\begin{aligned}\left(\sum_{x\in A} u(x) + \sum_{y\in B} u(x)\right)\rho^*(A\cup B) &= \sum_{x\in A} u(x)x + \sum_{x\in B} u(x)x \\ &= \rho^*(A)\sum_{x\in A} u(x) + \rho^*(B)\sum_{x\in B} u(x);\end{aligned}$$

whence

$$\rho^*(A\cup B) \in \text{conv}^0\{\rho^*(A), \rho^*(B)\},$$

as $\sum_{x\in A} u(x) > 0$ and $\sum_{x\in B} u(x) > 0$. □

The proof of the sufficiency of the axioms relies on two key ideas. One is that when A is affinely independent, $\rho^*(A)$ has a unique representation as a convex combination of the elements of A. The other is the following lemma, which is used to show that path independence determines average choice uniquely.

Lemma 6. *Suppose that ρ^* satisfies one-point path independence. Let A be a finite set with $|A| \geq 3$, and let $x, y \in A$ with $x \neq y$. If x and y are extreme points of A and there is a proper face F of conv(A) with dim(F) ≥ 1 and $x, y \in F$, then:*

$$conv^0(\{x, \rho^*(A \setminus \{x\})\}) \cap conv^0(\{y, \rho^*(A \setminus \{y\})\})$$

is a singleton.

Proof. First, $\rho^*(A) \in \text{conv}^0\{x, \rho^*(A \setminus \{x\})\}$ and $\rho^*(A) \in \text{conv}^0\{y, \rho^*(A \setminus \{y\})\}$, as ρ^* satisfies one-point path independence. So

$$\emptyset \neq \text{conv}^0\{x, \rho^*(A \setminus \{x\})\} \cap \text{conv}^0\{y, \rho^*(A \setminus \{y\})\}.$$

Since $x, y \in F$ there is a vector p and a scalar α with $F \subseteq \{z \in X : p \cdot z = \alpha\}$ –one of the hyperplanes supporting conv$A$– and such that conv$A \subseteq \{z \in X : p \cdot z \geq \alpha\}$. We shall prove that if there is

$$z^1, z^2 \in \text{conv}^0\{x, \rho^*(A \setminus \{x\})\} \cap \text{conv}^0\{y, \rho^*(A \setminus \{y\})\},$$

with $z^1 \neq z^2$, then $A \subseteq F$, contradicting that F is a proper face of conv(A).

Now, $z^1, z^2 \in \text{conv}^0\{x, \rho^*(A \setminus \{x\})\}$ implies that there is $\theta_x, \theta_{\rho^*(A\setminus\{x\})} \in \mathbf{R}$ such that $x = z^2 + \theta_x(z^1 - z^2)$ and $\rho^*(A \setminus \{x\}) = z^2 + \theta_{\rho^*(A\setminus\{x\})}(z^1 - z^2)$. Similarly, we have $y = z^2 + \theta_y(z^1 - z^2)$ and $\rho^*(A \setminus \{y\}) = z^2 + \theta_{\rho^*(A\setminus\{y\})}(z^1 - z^2)$.

As a consequence then of $p \cdot x = \alpha = p \cdot y$ we have

$$p \cdot z^2 + \theta_x p \cdot (z^1 - z^2) = p \cdot z^2 + \theta_y p \cdot (z^1 - z^2).$$

Since $\theta_x \neq \theta_y$ (as $x \neq y$) we obtain that $p \cdot (z^1 - z^2) = 0$. So

$$\alpha = p \cdot x = p \cdot z^2 + p \cdot \theta_x(z^1 - z^2) = p \cdot z^2.$$

Then

$$p \cdot \rho^*(A \setminus \{x\}) = p \cdot (z^2 + \theta_{\rho^*(A\setminus\{x\})}(z^1 - z^2)) = p \cdot z^2 = \alpha.$$

But $\rho^*(A \setminus \{x\}) = \sum_{z \in A\setminus\{x\}} \lambda_z z$ for some $\lambda_z > 0$, as $\rho^*(A \setminus \{x\})$ is in the relative interior of $\text{conv}(A \setminus \{x\})$. Then $p \cdot z \geq \alpha$ for all $z \in A$ implies that $p \cdot z = \alpha$ for all $z \in A \setminus \{x\}$.

This means that $A \subseteq F$, contradicting that F is a proper face of $\text{conv}A$. $\square$

The proof of sufficiency proceeds by first constructing a stochastic choice ρ, then arguing that it is a Luce rule that rationalizes ρ^*. The first step is to define $\rho(A)$ for A with $|A| = 2$. Then we define $\rho(A)$ for A with $|A| = 3$, and finally we define ρ on all of $\mathcal{A}$ by constructing a continuous Luce rule, and using this rule to define ρ. The average choice defined from the Luce rule must be path independent, so Lemma 6 is used to show that the average choice from the constructed Luce rule coincides with ρ^*.

Let $x, y \in X$, $x \neq y$. Since $\rho^*(\{x, y\})$ is in the relative interior of $\{x, y\}$ there is a unique $\theta \in (0, 1)$ with $\rho^*(\{x, y\}) = \theta x + (1 - \theta)y$. Define $\rho(x, \{x, y\}) = \theta$ and $\rho(y, \{x, y\}) = 1 - \theta$. Note that $\rho(x, \{x, y\})$ is a continuous function of x and y.

Thus we have defined $\rho(A)$ for A with $|A| = 2$. Now turn to A with $|A| = 3$. Let $A = \{x, y, z\}$. Consider the case when the vectors x, y and z are affinely independent. Such collections of three vectors exist because $|P| \geq 3$. Then there are unique $\rho(x, A)$ $\rho(y, A)$ and $\rho(z, A)$ (non-negative and adding up to 1) such that $\rho^*(A) = x\rho(x, A) + y\rho(y, A) + z\rho(z, A)$. Since $\rho^*(A)$ is in the relative interior of $\text{conv}(A)$ in fact, $\rho(x, A), \rho(y, A), \rho(z, A) > 0$.

By path independence, there is θ such that

$$\rho^*(A) = \theta z + (1 - \theta)\rho^*(A \setminus \{z\}) = \theta z + (1 - \theta)[x\rho(x, \{x, y\}) + y\rho(y, \{x, y\})].$$

The vectors x, y and z are affinely independent, so the weights $\rho(x, A)$, $\rho(y, A)$ and $\rho(z, A)$ are unique, which implies that $\rho(x, A) = (1 - \theta)\rho(x, \{x, y\})$ and $\rho(y, A) = (1 - \theta)\rho(y, \{x, y\})$. Hence

$$\frac{\rho(x, A)}{\rho(y, A)} = \frac{\rho(x, \{x, y\})}{\rho(y, \{x, y\})}. \tag{1}$$

Consider now the case when the vectors x, y and z are not affinely independent. Choose a sequence $\{z_n\}$ such that

a) x, y and z_n are affinely independent for all n,

b) $z = \lim_{n\to\infty} z_n$,

c) and $\rho(\{x, y, z_n\})$ converges.

To see that it is possible to choose such a sequence, note that x, y and z_n are affinely independent if and only if $x - z_n$ and $y - z_n$ are not collinear. Now $x - z$ and $y - z$ are collinear, so there is $\theta \in \mathbf{R}$ with $x - z = \theta(y - z)$. For each n the ball with center z and radius $1/n$ has full dimension, so the intersection of this ball with the complement in X of the line passing through x and z (which is also the line passing through y and z) is nonempty. By choosing z_n in this ball, but outside of the line passing through x and z, we obtain a sequence that converges to z Then we have that

$$\frac{(x - z_n)_i}{(y - z_n)_i} = \frac{\theta(y - z)_i + (z - z_n)_i}{(y - z) + (z - z_n)_i} = \frac{\theta + \frac{(z-z_n)_i}{(y-z)_i}}{1 + \frac{(z-z_n)_i}{(y-z)_i}},$$

a ratio that is not a constant function of i, as z_n is not on the line passing through y and z. Finally, by going to a subsequence if necessary, we can ensure that condition c) holds because the simplex is compact. Define $\rho(\{x, y, z\})$ to be the limit of $\rho(\{x, y, z_n\})$.

We have shown that Equation (1) holds for sets of three affinely independent vectors. So $\rho(x, A)/\rho(y, A) = \rho(x, \{x, y\})/\rho(y, \{x, y\})$, as $\rho(x, \{x, y, z_n\})/\rho(y, \{x, y, z_n\}) = \rho(x, \{x, y\})/\rho(y, \{x, y\})$ for all n. In particular, this means that $\rho(x, A), \rho(y, A) \in (0, 1)$, as $\rho(x, \{x, y\}), \rho(y, \{x, y\}) \in (0, 1)$.

Again, the fact that Equation (1) holds for sets of three affinely independent vectors implies:

$$\frac{\rho(z_n, \{x, y, z_n\})}{\rho(y, \{x, y, z_n\})} = \frac{\rho(z_n, \{z_n, y\})}{\rho(y, \{z_n, y\})} \to \frac{\rho(z, \{z, y\})}{\rho(y, \{z, y\})}$$

as $n \to \infty$. Then $\rho(z, A)/\rho(y, A) = \rho(z, \{z, y\})/\rho(y, \{z, y\})$. In particular, $\rho(z, A) \in (0, 1)$ because $\rho(z, \{z, y\}) \in (0, 1)$.

Thus we have established that, for any three distinct vectors $x, y, z \in X$ (affinely independent or not), Equation (1) holds with $A = \{x, y, z\}$.

Now we turn to the definition of ρ on $\mathcal{A}$. The definition proceeds by induction. We use (1) to define a utility function $u : X \to \mathbf{R}_{++}$ and a Luce rule. Then we show by induction, and using Lemma 6, that ρ rationalizes ρ^*.

Fix $y_0 \in X$. Let $u(y_0) = 1$. For all $x \in X$, let $u(x) = \frac{\rho(x,\{x,y_0\})}{\rho(y_0,\{x,y_0\})}$. Note that u is a continuous function, as $x \mapsto (\rho(x, \{x, y\}), \rho(y, \{x, y\}))$ is continuous, and that $u > 0$. By Equation (1), we obtain

$$\begin{aligned} \frac{u(x)}{u(y)} &= \frac{\rho(x, \{x, y_0\})}{\rho(y_0, \{x, y_0\})} \frac{\rho(y_0, \{y, y_0\})}{\rho(x, \{y, y_0\})} \\ &= \frac{\rho(x, \{x, y\})}{\rho(y, \{x, y\})} \\ &= \frac{\rho(x, \{x, y, z\})}{\rho(y, \{x, y, z\})}, \end{aligned}$$

for all x and y.

Let $\rho(A)$ be the Luce rule defined by u. This definition coincides with the definition of ρ we have given for A with $|A| \leq 3$ because

$$\frac{u(x)}{u(y)} = \frac{\rho(x, \{x, y\})}{\rho(y, \{x, y\})} \text{ and } \frac{u(x)}{u(y)} = \frac{\rho(x, \{x, y, z\})}{\rho(y, \{x, y, z\})}$$

for all $x, y, z \in X$.

Note also that, by definition of ρ,

$$\begin{aligned} \rho^*(\{x, y\}) &= x\rho(x, \{x, y\}) + y\rho(y, \{x, y\}) \\ \text{and } \rho^*(\{x, y, z\}) &= x\rho(x, \{x, y, z\}) + y\rho(y, \{x, y, z\}) + z\rho(z, \{x, y, z\}) \end{aligned} \tag{2}$$

for all $x, y, z \in X$.

Let

$$\bar{\rho}(A) = \sum_{x \in A} x\rho(x, A).$$

We shall prove that $\bar{\rho}(A) = \rho^*(A)$ for all $A \in \mathcal{A}$, which finishes the proof of the theorem.

The proof proceeds by induction on the size of A. We have established (see Equation (2)) that $\bar{\rho}(A) = \rho^*(A)$ for $|A| \leq 3$. Suppose then that $\bar{\rho}(A) = \rho^*(A)$ for all $A \in \mathcal{A}$ with $|A| \leq k$. Let $A \in \mathcal{A}$ with $|A| = k + 1$. We shall prove that $\bar{\rho}(A) = \rho^*(A)$.

Suppose first that $\dim(\text{conv}A) \geq 2$. Then by Corollary 2.4.8 in Schneider (2013) there is a proper face F of $\text{conv}(A)$ with dimension ≥ 1. Let $x, y \in A$ be two extreme points of $\text{conv}(A)$, such that $x, y \in F$. Such x and y exist because F has dimension at least 1. Note that $\bar{\rho}(A \setminus \{x\}) = \rho^*(A \setminus \{x\})$ $\bar{\rho}(A \setminus \{y\}) = \rho^*(A \setminus \{y\})$ by the inductive hypothesis. By Lemma 5, $\bar{\rho}$ satisfies path independence. So Lemma 6 implies that

$$\mathrm{conv}^0(\{x, \bar{\rho}(A \setminus \{x\})\}) \cap \mathrm{conv}^0(\{y, \bar{\rho}(A \setminus \{y\})\}) = \mathrm{conv}^0(\{x, \rho^*(A \setminus \{x\})\}) \cap \mathrm{conv}^0(\{y, \rho^*(A \setminus \{y\})\})$$

is a singleton. Since $\bar{\rho}$ and ρ^* both satisfy path independence, $\bar{\rho}(A) = \rho^*(A)$.

In second place we consider the case when $\dim(A) = 1$. Let $z \in A$ be an extreme point of $\mathrm{conv}A$. Let $\{z_n\}$ be a sequence with $z_n \to z$, such that $A \setminus \{z\} \cup \{z_n\}$ has dimension ≥ 2. This is possible because X has dimension larger than 3. The we obtain that $\bar{\rho}(A) = \rho^*(A)$ by continuity of ρ^* and $\bar{\rho}$.

6. Proof of Theorems 2 and 3

Lemma 7. *If ρ^* is continuous linear Luce rationalizable, then it satisfies independence.*

Proof. Let ρ^* be continuous linear Luce rationalizable, with $u(x) = f(v \cdot x)$. We check that it satisfies independence. Note first that:

$$\begin{aligned}\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) &= \mu(\lambda x + (1-\lambda)z) + (1-\mu)(\lambda y + (1-\lambda)z) \\ &= \lambda(\mu x + (1-\mu)y) + (1-\lambda)z,\end{aligned}$$

where

$$\mu = \frac{f(v \cdot (\lambda x + (1-\lambda)z))}{f(v \cdot (\lambda x + (1-\lambda)z)) + f(v \cdot (\lambda y + (1-\lambda)z))}.$$

Suppose first that $u(x) = u(y)$. Then $v \cdot x = v \cdot y$, as f is monotone increasing. Then $f(v \cdot (\lambda x + (1-\lambda)z)) = f(v \cdot (\lambda y + (1-\lambda)z))$. This means that $\mu = 1/2$. So $\mu x + (1-\mu)y = \rho^*(\{x, y\})$, because $u(x) = u(y)$ implies that $\rho^*(\{x, y\}) = \frac{1}{2}x + \frac{1}{2}y$.

Conversely, suppose that for all λ and all z such that $\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X$

$$\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) = \lambda \rho^*(\{x, y\}) + (1-\lambda)z,$$

but (towards a contradiction) that $u(x) \neq u(y)$. Say that $u(x) > u(y)$. Choose z and λ such that $\mu > 1/2$, where μ is such that $\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) = \mu(\lambda x + (1-\lambda)z) + (1-\mu)(\lambda y + (1-\lambda)z)$ (which can be done by the continuity of u). Then $\rho^*(\{x, y\}) \neq \mu x + (1-\mu)y$, a contradiction. □

Let ρ^* be continuous Luce rationalizable with utility function u. Write $x \sim y$ when $u(x) = u(y)$.

Lemma 8. *If ρ^* satisfies independence then it satisfies the following property:*

$$x \sim y \text{ iff } \forall \lambda \in \mathbf{R} \forall z \in X (\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X \Longrightarrow \lambda x + (1-\lambda)z \sim \lambda y + (1-\lambda)z)$$

Proof. Let $x \sim y$, $\lambda \in (0,1)$ and $z \in X$. Let μ be such that

$$\begin{aligned}\rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) &= \mu(\lambda x + (1-\lambda)z) + (1-\mu)(\lambda y + (1-\lambda)z) \\ &= \lambda(\mu x + (1-\mu)y) + (1-\lambda)z.\end{aligned}$$

Then independence implies that

$$\lambda(\mu x + (1-\mu)y) + (1-\lambda)z = \rho^*(\{\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z\}) = \lambda \rho^*(\{x,y\}) + (1-\lambda)z,$$

and thus $\mu x + (1-\mu)y = \rho^*(\{x,y\})$ Since $x \sim y$ we must have $\mu = 1/2$. Hence

$$\frac{1}{2} = \frac{u(\lambda x + (1-\lambda)z)}{u(\lambda x + (1-\lambda)z) + u(\lambda y + (1-\lambda)z)}.$$

Therefore $\lambda x + (1-\lambda)z \sim \lambda y + (1-\lambda)z$.

Conversely, if $\lambda x + (1-\lambda)z \sim \lambda y + (1-\lambda)z$ for all λ and all z then $x \sim y$ by continuity of u and the fact that $\lambda x + (1-\lambda)z, \lambda y + (1-\lambda)z \in X$. □

The property in Lemma 8 is weaker than the standard von-Neuman Morgenstern independence property (restricted to $\lambda \in (0,1)$). Using Lemma 8, however, we can establish the stronger independence property, as stated in the next lemma. Then the proof of Theorem 2 follows from the expected utility theorem: the preference relation represented by u is a weak order, it satisfies continuity (as u is continuous), and independence.

Lemma 9. *If ρ^* is continuous Luce rationalizable, and it satisfies independence, then it satisfies the following property*

$$u(x) \geq u(y) \text{ iff } \forall \lambda \in [0,1] \forall z \in X \ u(\lambda x + (1-\lambda)z) \geq u(\lambda y + (1-\lambda)z).$$

Proof. Note that by Lemma 8, independence implies that

$$x \sim y \text{ iff } x + \theta(z - x) \sim y + \theta(z - y)$$

for all scalars θ and for all z.

Suppose towards a contradiction that $u(x) \geq u(y)$ but that $u(\lambda x + (1-\lambda)z) < u(\lambda y + (1-\lambda)z)$. By continuity of u there is $\lambda^* < \lambda$ such that $u(\lambda x + (1-\lambda)z) = u(\lambda y + (1-\lambda)z)$.

Let $\lambda^* x + (1-\lambda^*)z = x'$. Then $x = x' - \frac{1-\lambda^*}{\lambda^*}(z - x')$. Similarly, $y = y' - \frac{1-\lambda^*}{\lambda^*}(z - y')$ where $\lambda^* y + (1-\lambda^*)z = y'$. Then independence and $u(x') = u(y')$ implies that $u(x) = u(y)$. Then $u(\lambda x + (1-\lambda)z) < u(\lambda y + (1-\lambda)z)$ is a violation of independence. □

6.1. Proof of Theorem 3

By Theorem 2, a strictly affine Luce rationalizable rule satisfies independence, continuity and path independence. Lemma 10 establishes that calibration is also necessary for the rule to be strictly affine Luce rationalizable.

For necessity, we know that path independence, continuity and independence imply that an average choice rule is linear Luce rationalizable. Lemma 11 finishes the proof by establishing that calibration is sufficient for the linear Luce rule to be strictly affine.

Lemma 10. *The strictly affine Luce model satisfies calibration*

Proof. Let $w = \lambda x + (1-\lambda)y$ and $w' = \lambda y + (1-\lambda)x$. Note that

$$v \cdot w + v \cdot w' = v \cdot x + v \cdot y.$$

Therefore,

$$\begin{aligned}&\rho^*(\{\lambda x + (1-\lambda)y, \lambda y + (1-\lambda)x\}) \\ &= \frac{(v \cdot w + \beta)(\lambda x + (1-\lambda)y) + (v \cdot w' + \beta)(\lambda y + (1-\lambda)x)}{(v \cdot x + \beta) + (v \cdot y + \beta)}.\end{aligned}$$

Note that

$$(v \cdot w + \beta)(\lambda x + (1-\lambda)y) + (v \cdot w' + \beta)(\lambda y + (1-\lambda)x)$$

is:

$$
\begin{aligned}
(\lambda v\cdot x+(1-\lambda)v\cdot y+\beta)(\lambda x+(1-\lambda)y)+(\lambda v\cdot y+(1-\lambda)v\cdot x+\beta)(\lambda y+(1-\lambda)x)&\\
=((\lambda^2+(1-\lambda)^2)v\cdot x+2\lambda(1-\lambda)v\cdot y)x&\\
+((\lambda^2+(1-\lambda)^2)v\cdot y+2\lambda(1-\lambda)v\cdot x)y&\\
+\beta(\lambda x+(1-\lambda)y+\lambda y+(1-\lambda)x)&\\
=((\lambda^2+(1-\lambda)^2)v\cdot x+2\lambda(1-\lambda)v\cdot y)x+((\lambda^2+(1-\lambda)^2)v\cdot y+2\lambda(1-\lambda)v\cdot x)y&\\
+\beta(x+y).&
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\rho^*(\{\lambda x+(1-\lambda)y,\lambda y+(1-\lambda)x\})\\
=&(\lambda^2+(1-\lambda)^2)\frac{v\cdot xx+v\cdot yy}{(v\cdot x+\beta)+(v\cdot y+\beta)}+2\lambda(1-\lambda)\frac{v\cdot yx+v\cdot xy}{(v\cdot x+\beta)+(v\cdot y+\beta)}\\
&+\frac{\beta(x+y)}{(v\cdot x+\beta)+(v\cdot y+\beta)}\\
=&(\lambda^2+(1-\lambda)^2)\frac{(v\cdot x+\beta)x+(v\cdot y+\beta)y}{(v\cdot x+\beta)+(v\cdot y+\beta)}+2\lambda(1-\lambda)\frac{(v\cdot y+\beta)x+(v\cdot x+\beta)y}{(v\cdot x+\beta)+(v\cdot y+\beta)}\\
=&(\lambda^2+(1-\lambda)^2)\rho^*(\{x,y\})+2\lambda(1-\lambda)(x+y-\rho^*(\{x,y\}))\\
=&\rho^*(\{x,y\})+2\lambda(1-\lambda)(x+y-2\rho^*(\{x,y\}))\\
=&\rho^*(\{x,y\})+2\lambda(1-\lambda)(x-\rho^*(\{x,y\}))+2\lambda(1-\lambda)(y-\rho^*(\{x,y\}))
\end{aligned}
$$

□

Lemma 11. *Let ρ^* be a linear Luce rationalizable, with utility function $u(x)=f(v\cdot .x)$. Suppose that ρ^* satisfies calibration. Then there is $\alpha>0$ and β such that $f(v\cdot x)=\alpha v\cdot x+\beta$ for all $x\in X$.*

Proof. Since ρ^* satisfies continuity, path independence, and independence, it is linear Luce rationalizable with some $v\in\mathbf{R}^n$ and function f.

By the calibration axiom, for all $x,y\in X$ and $\lambda\in[0,1]$,

$$\frac{f(v\cdot w)w+f(v\cdot w')w'}{f(v\cdot w)+f(v\cdot w')}=(\lambda^2+(1-\lambda)^2)\frac{f(v\cdot x)x+f(v\cdot y)y}{f(v\cdot x)+f(v\cdot y)}+2\lambda(1-\lambda)\frac{f(v\cdot x)y+f(v\cdot y)x}{f(v\cdot x)+f(v\cdot y)},$$

where $w = \lambda x + (1-\lambda)y$ and $w' = \lambda y + (1-\lambda)x$. The weight placed on x must be the same in both convex combinations, therefore:

$$\frac{\lambda f(v\cdot w) + (1-\lambda)f(v\cdot w')}{f(v\cdot w) + f(v\cdot w')} = \frac{(\lambda^2 + (1-\lambda)^2)f(v\cdot x) + 2\lambda(1-\lambda)f(v\cdot y)}{f(v\cdot x) + f(v\cdot y)},$$

Let $V = \{a \in \mathbb{R} | v\cdot x = a \text{ for some } x \in X\}$. Then, the above equation implies that for all $\lambda \in [0,1] \setminus \{\frac{1}{2}\}$ and $a, b \in V$,

$$\begin{aligned}\frac{\lambda f(\lambda a + (1-\lambda)b) + (1-\lambda)f(\lambda b + (1-\lambda)a)}{f(\lambda a + (1-\lambda)b) + f(\lambda b + (1-\lambda)a)} &= \frac{(\lambda^2 + (1-\lambda)^2)f(a) + 2\lambda(1-\lambda)f(b)}{f(a)+f(b)} \\ &= \lambda\frac{\lambda f(a) + (1-\lambda)f(b)}{f(a)+f(b)} \\ &+ (1-\lambda)\frac{(1-\lambda)f(a) + \lambda f(b)}{f(a)+f(b)}.\end{aligned}$$

Let

$$A = \frac{f(\lambda a + (1-\lambda)b)}{f(\lambda a + (1-\lambda)b) + f(\lambda b + (1-\lambda)a)} \text{ and } B = \frac{\lambda f(a) + (1-\lambda)f(b)}{f(a)+f(b)}.$$

Then $\lambda A + (1-\lambda)(1-A) = \lambda B + (1-\lambda)(1-B)$, so that $A = B$.

Step 1: For all $\lambda \in [0,1] \setminus \{\frac{1}{2}\}$ and $a, b \in V$, $f(\lambda a + (1-\lambda)b) > \lambda f(a) + (1-\lambda)f(b) \Leftrightarrow f(\lambda b + (1-\lambda)a) < \lambda f(b) + (1-\lambda)f(a)$.

Proof of Step 1: If $f(\lambda a + (1-\lambda)b) < \lambda f(a) + (1-\lambda)f(b)$, then $A = B$ implies that $f(\lambda a + (1-\lambda)b) + f(\lambda b + (1-\lambda)a) > f(a) + f(b)$. So $\lambda f(a) + (1-\lambda)f(b) + f(\lambda b + (1-\lambda)a) > f(a) + f(b)$. Therefore, $f(\lambda b + (1-\lambda)a) > \lambda f(b) + (1-\lambda)f(a)$.

Conversely, if $f(\lambda a + (1-\lambda)b) > \lambda f(a) + (1-\lambda)f(b)$, then $A = B$ implies that $f(\lambda a + (1-\lambda)b) + f(\lambda b + (1-\lambda)a) < f(a) + f(b)$. So $\lambda f(a) + (1-\lambda)f(b) + f(\lambda b + (1-\lambda)a) < f(a) + f(b)$. Therefore, $f(\lambda b + (1-\lambda)a) < \lambda f(b) + (1-\lambda)f(a)$.

Step 2: $f(\frac{1}{2}a + \frac{1}{2}b) = \frac{1}{2}f(a) + \frac{1}{2}f(b)$ for all $a, b \in V$.

Proof of Step 2: Suppose that $f(\frac{1}{2}a + \frac{1}{2}b) \neq \frac{1}{2}f(a) + \frac{1}{2}f(b)$ for some $a, b \in V$. Without loss of generality, assume $f(\frac{1}{2}a + \frac{1}{2}b) > \frac{1}{2}f(a) + \frac{1}{2}f(b)$. Define for all $\lambda \in [0,1]$,

$$d(\lambda) = f(\lambda a + (1-\lambda)b) - \lambda f(a) + (1-\lambda)f(b).$$

Then by Step 1, we must have for all $\lambda \in [0,1] \setminus \{\frac{1}{2}\}$,

$$d(\lambda) > 0 \Leftrightarrow d(1-\lambda) < 0.$$

and $d(\frac{1}{2}) > 0$. Since d is continuous, there exists a positve number ε such that $d(\frac{1}{2}+\varepsilon) > 0$ and $d(\frac{1}{2}-\varepsilon) > 0$. This is a contradiction. So we have $f(\frac{1}{2}a+\frac{1}{2}b) = \frac{1}{2}f(a)+\frac{1}{2}f(b)$ for all $a, b \in V$.

Step 3: $f(\lambda a + (1-\lambda)b) = \lambda f(a) + (1-\lambda)f(b)$ for all $a, b \in V$ and $\lambda \in (0,1)$.
Proof of Step 3: Suppose toward contradiction that there exist $a, b \in V$ and $\lambda \in (0,1)$ such that $f(\lambda a + (1-\lambda)b) \neq \lambda f(a) + (1-\lambda)f(b)$. Without loss of generality, assume that $a < b$.

Consider the case where $f(\lambda a+(1-\lambda)b) > \lambda f(a)+(1-\lambda)f(b)$. Define a function g as follows: for all $x \in V$

$$g(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a) - f(a).$$

Note that

$$\begin{aligned} g(\lambda a + (1-\lambda)b) &= f(\lambda a + (1-\lambda)b) - \frac{f(b)-f(a)}{b-a}(\lambda a + (1-\lambda)b - a) - f(a) \\ &> \lambda f(a) + (1-\lambda)f(b) - (1-\lambda)(f(b)-f(a)) - f(a) \\ &= 0. \end{aligned}$$

So $\sup_{x\in[a,b]} g(x) > 0$. Define $\gamma \equiv \sup_{x\in[a,b]} g(x)$.

Moreover, $g(a) = g(b) = 0$. By Step 2,

$$\begin{aligned} g\Big(\frac{x+y}{2}\Big) &= f\Big(\frac{x+y}{2}\Big) - \frac{f(b)-f(a)}{b-a}\Big(\frac{x+y}{2} - a\Big) - f(a) \\ &= \frac{g(x)+g(y)}{2}. \end{aligned}$$

Define $c = \inf\{x \in [a,b] | g(x) = \gamma\}$. Since g is continuous (because f is continuous), $g(c) = \gamma$. Since $\gamma > 0$, it must be that $c \in (a,b)$. For any $h > 0$ such that $c+h \in (a,b)$ and $c-h \in (a,b)$, we have that $g(c-h) < g(c)$ and $g(c+h) \leq g(c)$. Hence, by Step 2,

$$g(c) = \frac{g(c-h)+g(c+h)}{2} < \frac{g(c)+g(c)}{2} = g(c),$$

which is a contradiction.

We can obtain a contrdiction in the same way in the case where $f(\lambda a+(1-\lambda)b) < \lambda f(a)+(1-\lambda)f(b)$.

So we have proved that there is $\alpha > 0$ and β such that $f(v \cdot x) = \alpha v \cdot x + \beta$ for all $x \in X$. We redefine v as $v = \alpha v$. So we finished the proof of Theorem 3. □

A. Appendix

The following result is standard in convex analysis (see for example Theorem 1.1.14 in Schneider (2013)). We include a proof for convenience.

Lemma 12. *Let $A \subseteq \mathbf{R}^n$ be a finite set. Then x is in the relative interior of $conv(A)$ iff there exists $(\lambda_a)_{a\in A}$ such that $x = \sum_{a\in A} \lambda_a a$, $\sum_{a\in A} \lambda_a = 1$ and $\lambda_a > 0$ for all $a \in A$.*

Proof. Let $A^* = \{l \in A | \text{there exists no } \{r_i\} \subset A, \{\alpha_i\} \text{ s.t. } l = \sum \alpha_i r_i\}$. Since $\rho^*(A) \in conv(A)$, by Choquet theorem, there exists ρ such that $\rho^*(A) = \sum_{a\in A^*} a\rho(a, A)$. If $A = A^*$, there is nothing to prove. Consider the case $A \setminus A^* \neq \emptyset$.

Since A is finite, we can write $A^* = \{r^1, \ldots, r^m\}$ and $A \setminus A^* = \{a^1, \ldots, a^n\}$. By definition, for each a^j, there exist $\alpha^j \in [0, 1]^m$ such that $a^j = \sum_{i=1}^m \alpha^j(i) r^i$ and $\sum_{i=1}^m \alpha^j(i) = 1$. For each $j \in \{1, \ldots, n\}$, choose a positive number ε_j such that for each $i \in \{1, \ldots, m\}$

$$\sum_{j=1}^n \varepsilon_j \alpha^j(i) < \rho(r^i, A).$$

Define ρ' as follows: for each i, define $\rho'(r^i, A) = \rho(r^i, A) - \sum_{j=1}^n \varepsilon_j \alpha^j(i)$, for each a^j, define $\rho'(a^j, A) = \varepsilon_j$. By the deifnition of ε, $\rho'(a, A) > 0$ for all $a \in A$. To finish the proof, we need to show $\sum_{a\in A} \rho'(a, A) = 1$ (so that $\rho'(A) \in \Delta(A)$) and

$\sum_{a\in A} a(x)\rho'(a, A) = \rho^*(A)$.

$$\begin{aligned}
\sum_{a\in A}\rho'(a,A) &= \sum_{i=1}^{m}\rho'(r^i,A) + \sum_{j=1}^{n}\rho'(a^j,A) \\
&= \sum_{i=1}^{m}\Big(\rho(r^i,A) - \sum_{j=1}^{n}\varepsilon_j\alpha^j(i)\Big) + \sum_{j=1}^{n}\varepsilon_j \\
&= \sum_{i=1}^{m}\rho(r^i,A) - \sum_{i=1}^{m}\sum_{j=1}^{n}\varepsilon_j\alpha^j(i) + \sum_{j=1}^{n}\varepsilon_j \\
&= \sum_{i=1}^{m}\rho(r^i,A) - \sum_{j=1}^{n}\varepsilon_j\Big(\sum_{i=1}^{m}\alpha^j(i)\Big) + \sum_{j=1}^{n}\varepsilon_j \\
&= \sum_{i=1}^{m}\rho(r^i,A) \quad (\because \sum_{i=1}^{m}\alpha^j(i) = 1) \\
&= 1.
\end{aligned}$$

Hence, $\rho' \in \Delta(A)$. Moreover, for each $x \in P$,

$$\begin{aligned}
\sum_{a\in A} a(x)\rho'(a,A) &= \sum_{i=1}^{m} r^i(x)\rho'(r^i,A) + \sum_{j=1}^{n} a^j(x)\rho'(a^j,A) \\
&= \sum_{i=1}^{m} r^i(x)\rho'(r^i,A) + \sum_{j=1}^{n}\Big(\sum_{i=1}^{m}\alpha^j(i)r^i(x)\Big)\rho'(a^j,A) \\
&= \sum_{i=1}^{m} r^i(x)\Big(\rho(r^i,A) - \sum_{j=1}^{n}\varepsilon_j\alpha^j(i)\Big) + \sum_{j=1}^{n}\Big(\sum_{i=1}^{m}\alpha^j(i)r^i(x)\Big)\varepsilon_j \\
&= \sum_{i=1}^{m} r^i(x)\rho(r^i,A) \\
&= \sum_{a\in A^*} a(x)\rho(a,A) \\
&= \rho^*(x,A).
\end{aligned}$$

So $\sum_{a\in A} a(x)\rho'(a, A) = \rho^*(A)$. □

B. Calculations from Section 4.4

We have $\mathbf{E}p_x^n = p_x/n$, $\mathbf{V}p_x^n = p_x(1-p_x)/n$ and $\mathbf{Cov}(p_x^n, p_y^n) = -p_xp_y$ when $y \neq x$. Moreover, the central limit theorem implies that $\sqrt{n}(p_x^n - p_x) \xrightarrow{d} N(0, \sqrt{p_x(1-p_x)})$.

Then, using the delta method and the following calculation yields the result in the text:

$$\begin{aligned}
\mathbf{V}\left(\frac{p_x^n}{p_y^n}-\frac{p_x}{p_y}\right) &= \left(\frac{1}{p_y}\right)^2 \mathbf{V}p_x^n + \left(\frac{-p_x}{p_y^2}\right)^2 \mathbf{V}p_y^n + \frac{2}{p_y}\frac{-p_x}{p_y^2}(-p_x p_y) \\
&= \frac{1}{np_y^2}\left[p_x(1-p_x) + \frac{p_x^2}{p_y^2}p_y(1-p_y) + 2p_x\right] \\
&= \frac{p_x}{np_y^2}\left[(1+p_x) + \frac{p_x}{p_y}(1-p_y)\right] \\
&= \frac{p_x}{np_y^2}\left[1+\frac{p_x}{p_y}\right]
\end{aligned}$$

References

Blair, D. H. (1975): "Path-independent social choice functions: A further result," *Econometrica*, 43, 173.

Debreu, G. (1960): "Review of R. Duncan Luce, Individual choice behavior: A theoretical analysis," *American Economic Review*, 50, 186–188.

Ferejohn, J. A. and D. M. Grether (1977): "Weak path independence," *Journal of Economic Theory*, 14, 19 – 31.

Greene, W. H. (2003): *Econometric analysis*, Pearson Education India.

Kalai, E. and N. Megiddo (1980): "Path independent choices," *Econometrica*, 48, 781–784.

Luce, R. D. (1959): *Individual Choice Behavior a Theoretical Analysis*, John Wiley and sons.

Machina, M. J. and R. P. Parks (1981): "On Path Independent Randomized Choice," *Econometrica*, 1345–1347.

McFadden, D. (1974): "Conditional logit analysis of qualitative choice behavior," in *Frontiers in econometrics*, ed. by P. Zarembka, Academic Press New York, 105.

Parks, R. (1976): "Further results on path independence, quasitransitivity, and social choice," *Public Choice*, 26, 75–87.

Plott, C. R. (1973): "Path independence, rationality, and social choice," *Econometrica*, 41, 1075–1091.

Schneider, R. (2013): *Convex bodies: the Brunn–Minkowski theory*, 151, Cambridge University Press.

Tversky, A. (1972): "Elimination by aspects: A theory of choice." *Psychological review*, 79, 281.