Random Utility Models, IIA, Mattsson-Weibull-Li Distributions and a Conjecture by Luce and Suppes

P O Lindberg

April 5, 2012

Abstract

In recent papers, Mattsson and Weibull, and independently Li, have introduced a class of independent distributions, where the individual cdf's are powers of each other. For this class they have showed that the choice probabilities are of "Luce form", i.e. the form following from the Independence from Irrelevant Alternatives (IIA) assumption. We show a converse to this result. For a class of "independent" Random Utility (RU) Models, allowing non-uniform expansion, and where the cdf's in question are well-behaved in a certain sense, the only possibility is that the cdf's are of the above mentioned Mattsson-Weibull-Li (MWL) type. This result resolves an old conjecture by Luce and Suppes. The result demands unbounded choice sets. For finite models, we show on the contrary, that any finite independent RU model with Luce form choice probabilities, there is a version with random terms that are not MWL-distributed

1. ***Introduction***

Discrete (Probabilistic) Choice (DC) *Models* originated in mathematical psychology, to account for nondeterministic behavior of respondents in test situations, through works such as Thurstone (1927), Luce (1959), Marschak (1960). A special form of a DC Model is a *Random Utility Model* (RUM). These were introduced by Marschak (1960), who termed them Random Utility *Maximization* Models.
In a RUM, we have a (random) vector $\mathbf{X} = (X_1, ..., X_n)^T$ of utilities of *choice alternatives* in some given "universal" *choice set* $I = \{1, 2, ..., n\}$, and we postulate that the probability of a random decision-maker to choose alternative *i* from *I* is

$$P_I(i) = \Pr(X_i > X_j, j \in I, j \neq i). \quad (1)$$

One typically also is interested in choices from subsets $J \subseteq I$, and then postulates that the corresponding choice probabilities are

$$P_J(i) = \Pr(X_i > X_j, j \in J, j \neq i), \text{ for } i \in J \quad (2)$$

Later RUMs have receive wide usage used in many other areas, such as travel demand, e.g. Ben-Akiva and Lerman (1985), labor markets, e.g. Keane Todd and Wolpin (2010), housing markets, e.g. McFadden (1978), environmental economics, e.g. Hoyos (2010), It has even been awarded with a Nobel Prize in Economics, McFadden (2001).
In Discrete Choice, the well-known IIA (Independence from Irrelevant Alternatives, Luce, 1959) hypothesis, is known to lead to choice probabilities of the simple "*Luce form*", (4) below. Such a DC model is also called a *strict utility* model by Block and Marschak (1960).

In the recent papers, Li (2011, e.g. formula (6)) and Mattson and Weibull (2010, also formula (6)), the authors note that if we have a RUM with independent error terms X_i, each with a cdf of the form Φ^{α_i}, for some given "*base*" cdf Φ, then the choice probabilities are of Luce form. Lets us say that such RUMs are of *power form*, and that the error term $X = (X_1, ..., X_n)$ has a *Mattsson-Weibull-Li* (MWL) distribution. In his paper, Li goes on to show that different choices of the base cdf Φ give different sensitivities for increases in travel costs, in travel demand models, and suggest how that may be utilized. Some Φ give increasing sensitivity with increased cost, whereas others give decreasing. The standard Multi-Nomial Logit (McFadden, 1974) is shown to have constant sensitivity. Mattsson and Weibull (2010), on the other hand also show that a RUM of power form has the *invariance property*, i.e. that the distribution of achieved utility is independent of which alternative is chosen (cf. Robertson and Strauss (1981), and Lindberg, Eriksson and Mattsson (1995)). They, further go on to show that all RUMs with independent random terms and having the invariance property must in fact be of power form.

The observation that a RUM of power form has choice probabilities of Luce form may have been noted already by Luce and Suppes (1965). In stating (on p. 338-339) that any DC model with non-zero choice probabilities of Luce form in fact can be viewed as a RUM, they cite an unpublished proof attributed to E. Holman and A. A. J. Marley.

For a given DC model, with choice probabilities fulfilling (4) with weights w_i, they construct a RUM with independent error terms X_i, each having an exponential distribution on the negative half axis with parameter w_i, i.e. with cdf $F^{(i)}(x) = \Pr\{X_i \leq x\} = e^{w_i x}$. It is then elementary to verify that this RUM has choice probabilities of the desired form (4).

Luce and Suppes go on to note, that the form of the choice probabilities in invariant under monotone transformations of the random terms and they conjecture that the ensuing models "are the only reasonably well-behaved examples" (of RUMs with Luce form choice probabilities that is).

Now, it might be noted that the cdf's of the Holman-Marley example are of MWL form. Further transforming the random terms by any monotone transformation, we still get distributions of MWL form, so in some sense, the results of Li and Mattsson-Weibull are already there in Luce and Suppes, although in a disguised form.

The aim of the current paper is to prove a version of the Luce-Suppes conjecture, with some stress on "reasonably well-behaved". It turns out that we have to define out some possible freakish behavior of the cdf's in question. To be specific, we need the cdf's to be continuously differentiable, and such the derivatives (i.e. densities) of one cdf and of an appropriate power of the other do not both agree and disagree arbitrarily close to a certain point in their supports. We further need to allow for "non-uniform expansion" of the choice set. (Cf. the uniform expansion in Yellott (1977). Also note that McFadden(1974) uses non-uniform expansion in the proof of his Lemma 2.)

2. Warming up

2.1 *Setup*

We consider parameterized families $\Omega = \{X_\alpha\}_{\alpha \in A}$ of independent random variables X_α on $\Re$, each with a distribution whose support is an interval, D_α. We also assume that the corresponding cdf's have positive, continuous densities on their domains.

We further consider a Random Utility, DC setting. That is, given a (finite) selection $\{X_{\alpha_i}\}_{i \in I}$ from Ω, we are interested in the choice probabilities

$$P_J(i) = \Pr(X_{\alpha_i} > X_{\alpha_j}, j \in J, j \neq i) \qquad (3)$$

for arbitrary $i \in J \subseteq I$. Moreover, we are interested in when the choice probabilities (3) are of *Luce form*, i.e. when there are positive reals $w_i, i \in I$, such that

$$P_J(i) = \frac{w_i}{\sum_{j \in J} w_j}, \qquad (4)$$

for all $i \in J \subseteq I$.

Definition 1. With an (*independent*) *Luce System* we mean a parameterized family $\Omega = \{X_\alpha\}_{\alpha \in A}$ of independent random variables as above, such that the choice probabilities are on Luce form, i.e. for any set $I = \{1,2,\ldots,n\}$, any finite subset $\{\alpha_j\}_{j \in I} \subset A$, , and any $i \in J \subseteq I$

$$P_J(i) = \Pr(X_{\alpha_i} > X_{\alpha_j}, j \in J, j \neq i) = \frac{w_i}{\sum_{j \in J} w_j} \text{ for appropriate positive reals } w_i, i \in I.$$

It is further assumed that the w_i only depend on the α_i, i.e. $w_i = w(\alpha_i)$ for some function $w: A \to \Re_{++}$.

We further explicitly allow for repetitions among the X_{α_i}, i.e. it may be that $\alpha_i = \alpha_j$ for some pair $\{i, j\}$, i.e. that X_{α_i} and X_{α_j} have the same distribution. In such a case it is still assumed that X_{α_i} and X_{α_j} are independent. If the X_{α_i} are of power form then we will call $\Omega = \{X_\alpha\}_{\alpha \in A}$ an *MWL-system.* □

To allow for repetitions among the X_{α_i}, is to allow for *expansion* of the choice set. This could be done in a *uniform* way, à la Yellott (1977), where there the number of copies of each X_{α_i} is the same.

We will however allow for *non-uniform expansion* of the choice set, i.e. the number of copies of the different X_{α_i} are allowed to be different.

2.2 *Preamble*

Our first consideration concerns the domains of the random variables.

***Lemma* 1**. Let $\Omega = \{X_\alpha\}_{\alpha \in A}$ be a Luce system, but only closed under uniform expansion, then $d_\alpha =_{df} \sup\{D_\alpha\}$ is the same for all $\alpha \in A$.

Proof: Assume by contradiction that for some α_1, α_2 we have $d_{\alpha_1} > d_{\alpha_2}$.

Consider a choice experiment where we choose between X_{α_1} and X_{α_2}. Then

$P_{\{1,2\}}(1) = \dfrac{w_1}{w_1 + w_2}$, where $w_i = w(\alpha_i)$, i=1,2.

Since further D_{α_1} is the support of X_{α_1}, the probability

$\bar{p}_1 = \Pr(X_{\alpha_1} > d_{\alpha_2})$ is positive.

Consider now another choice experiment, where we have n-folded the first experiment, and assume that $\alpha_{2j-1} = \alpha_1$ and $\alpha_{2j} = \alpha_2$ for $j = 1,2,\ldots,n$. (I.e. a *uniform* expansion of the choice set.)

Then, for an odd i, and with $I = \{1,2,\ldots,2n\}$, since we have a Luce system,

$P_I(i) = \dfrac{w_1}{n(w_1 + w_2)}$, and hence , P_I (odd i chosen) $= \dfrac{nw_1}{n(w_1 + w_2)} = \dfrac{w_1}{w_1 + w_2}$.

Further, since $\bar{p}_1 = \Pr(X_{\alpha_1} > d_{\alpha_2}) > 0$, we have by independence that

$\Pr(\max_{j \text{ odd}} X_{\alpha_j} > d_{\alpha_2}) = 1 - \Pr(\max_{j \text{ odd}} X_{\alpha_j} \le d_{\alpha_2}) = 1 - (1 - \bar{p}_1)^n$ which tends to 1 as $n \to \infty$.

We get

$\dfrac{w_1}{w_1 + w_2} = P_I$ (odd i chosen) $\ge \Pr(\max_{j \text{ odd}} X_{\alpha_j} > d_{\alpha_2}) = 1 - (1 - \bar{p}_1) \to 1$,

a contradiction, since $w_2 > 0$. □

As a further preparation for the main result we first determine expression for the choice probabilities.

Lemma 2. Assume that $\{X_{\alpha_i}\}_{i \in I}, I = \{1,2,\ldots,n\}$ are independent random variables with supports D_i, and with cdf's and densities F_i and f_i, respectively. Then

$$P_I(i) = \Pr(X_{\alpha_i} > X_{\alpha_j}, j \in I, j \neq i) =$$

$$= \int_{x \in D_i} f_i(x) \Pr(X_{\alpha_j} < x, j \in I, j \neq i)dx = \int_{x \in D_i} f_i(x) \prod_{j \neq i} F_j(x)dx. \quad \square \qquad (5)$$

The proof of the Lemma is in the statement. This result is well-known, and is stated for reference.

We will use (5) to compare choice probabilities. A more elaborate version of Lemma 2 is the following

Lemma 3. With assumptions as in Lemma 2, and with $\hat{X} = \max_{i\in I} X_{\alpha_i}$, and $a < b$ in D_i,

$$\int_a^b f_i(x)\prod_{j\neq i} F_j(x)dx = \Pr(\hat{X} \in (a,b)) \cdot P_I(i \mid a < \hat{X} < b) \qquad (6)$$

In the MWL case, with $F_i = \Phi^{w_i}$ for some base cdf Φ we have

$$\int_a^b f_i(x)\prod_{j\neq i} F_j(x)dx = \Pr(\hat{X} \in (a,b)) \cdot P_I(i), \qquad (7)$$

implying that $P_I(i \mid a < \hat{X} < b) = P_I(i)$ in this case.

Proof: $\int_a^b f_i(x)\prod_{j\neq i} F_j(x)dx = \Pr(X_{\alpha_i} \in (a,b)\ \&\ X_{\alpha_i} > X_{\alpha_j}\ j \neq i) =$

$= \Pr(\hat{X} \in (a,b)\ \&\ X_{\alpha_i} > X_{\alpha_j}\ j \neq i) = P(\hat{X} \in (a,b)) \cdot P_I(i \mid a < \hat{X} < b)$.

In the MWL case the cdf of $\hat{X}$ is $\hat{F}(x) =_{df} \prod_i \Phi^{\alpha_i}(x) = \Phi^{\sum_i \alpha_i}(x)$. We get

$$\int_a^b f_i(x)\prod_{j\neq i} F_j(x)dx = \int_a^b \alpha_i \Phi'(x)\Phi^{\alpha_i - 1}(x)\prod_{j\neq i}\Phi^{\alpha_j}(x)dx = \int_a^b \alpha_i \Phi'(x)\Phi^{\sum_j \alpha_j - 1} =$$

$$= \frac{\alpha_i}{\sum_j \alpha_j}(\Phi^{\sum_j \alpha_j}(b) - \Phi^{\sum_j \alpha_j}(a)) = P_I(i) \cdot \Pr(\hat{X} \in (a,b)) \qquad \square$$

A direct consequence of this Lemma is

Corollary 1. Let $X_i, i = 1,2$ be independent random variables with cdf's F_i, continuous densities f_i and supports D_i. Assume $\bar{d} = \sup\{D_i\} < \infty$, and independent of *i* and, and assume $f_1(\bar{d}) > 0$. Then, as $a \uparrow \bar{d}$, for $i = 1,2$ and $j = 3 - i$,

$$\int_a^{\bar{d}} f_i(x)F_j(x)dx / (\bar{d} - a) \to f_i(\bar{d}), \qquad (8)$$

and as a consequence

$$P_{\{1,2\}}(1 \mid a < \hat{X} < \bar{d}) \to \frac{f_1(\bar{d})}{f_1(\bar{d}) + f_2(\bar{d})}. \qquad (9)$$

Proof: The integrand in (8) tends to $f_i(\bar{d})$ as $x \uparrow \bar{d}$. Thus (8) follows. Thus by (6),

$$\Pr(\hat{X} \in (a,\bar{d})) \cdot P_{\{1,2\}}(i \mid a < \hat{X} < \bar{d}) / (\bar{d} - a) \to f_i(\bar{d}). \qquad (10)$$

Here $\Pr(\hat{X} \in (a,\bar{d}))/(\bar{d} - a) \to \hat{f}(\bar{d}) > 0$, where $\hat{f}$ is the density of $\hat{X}$. Thus, by (10),

$P_{\{1,2\}}(i \mid a < \hat{X} < \bar{d}) \to f_i(\bar{d}) / \hat{f}(\bar{d})$, and (9) follows. $\square$

Remark: The condition $f_1(\bar{d}) > 0$ can always be achieved by transforming the X_i, according to the observation by Luce and Suppes, such that X_1 becomes e.g. uniform on [0,1]. $\square$

Suppose we have $X_i = X_{\alpha_i}, i = 1,2$ with cdf's F_i, and that

$p_1 = P_{\{1,2\}}(1) = \alpha$, and hence $p_2 = P_{\{1,2\}}(2) = 1-\alpha$.

This would be the case in the MWL-setting with $F_1 = \Phi^{\alpha}$ and $F_2 = \Phi^{1-\alpha}$ for some base

cdf Φ. In this case we would further have $F_1 = F_2^{\alpha/(1-\alpha)}$.

We will argue by contradiction and assume that $F_1 \neq G =_{df} F_2^{\alpha/(1-\alpha)}$, and as a consequence that $f =_{df} F_1' \neq g =_{df} G'$.

As a warm-up for the main result, and a preparation for the upcoming definition of regular pairs of independent random variables, we will prove that the densities of f_1 and g have to be equal at $\bar{d} = \sup\{D_i\}, i = 1,2$.

Lemma 4. Let $\Omega = \{X_1, X_2\}$ be a Luce system, but only closed under uniform expansion.

Assume $p_1 = P_{\{1,2\}}(1) = \alpha \in (0,1)$. Let the X_i have cdf's F_i, and let as above $G =_{df} F_2^{\alpha/(1-\alpha)}$, $f =_{df} F_1'$, and $g =_{df} G'$. Further assume $\bar{d} = \sup\{D_i\} < \infty, i = 1,2$. Then $f(\bar{d}) = g(\bar{d})$.

Proof: Assume by contradiction that $f(\bar{d}) \neq g(\bar{d})$. Assume WLOG $f(\bar{d}) > g(\bar{d})$. (The other case can be treated by switching X_1 and X_2). Then, by continuity of *f* and *g*, there is an $\varepsilon > 0$, such that $f(x) \geq (1+\varepsilon)g(x)$ on $[\bar{d} - \varepsilon, \bar{d}]$.

Let $X_{\bar{1}}$ have cdf *G* and be independent of X_1 and X_2.

Consider now two sets of experiments. In the first we have, as in Lemma 1, *n*-folded the single experiment of choosing between X_1 and X_2. Let $I_n = \{1,2,\ldots,2n\}$ be the corresponding index set, with odd indices corresponding to X_1. In the other we have *n*-folded choices between $X_{\bar{1}}$ and X_2. Let $\bar{I}_n = \{1,2,\ldots,2n\}$ be the corresponding index set, with odd indices corresponding to $X_{\bar{1}}$.

Let $\hat{X}_n = \max_{i \in I_n} X_i$, and $\hat{\hat{X}}_n = \max_{i \in \bar{I}_n} X_i$.

For an arbitrary $\delta > 0$, $\Pr\{\hat{X}_n < \bar{d} - \delta\} \to 0$ and $\Pr\{\hat{\hat{X}}_n < \bar{d} - \delta\} \to 0$ as $n \to \infty$.

Let $p_n =_{df} P_{I_n}(\text{odd } i \text{ chosen}) = \dfrac{n\alpha}{n\alpha + n(1-\alpha)} = \dfrac{\alpha}{\alpha + (1-\alpha)}$ by the Luce system assumption. Similarly, let $\bar{p}_n =_{df} P_{\bar{I}_n}(\text{odd } i) = \dfrac{n\alpha}{n\alpha + n(1-\alpha)} = \dfrac{\alpha}{\alpha + (1-\alpha)}$ by the properties of MWL distributions.

Thus,

$$p_n = \Pr\{\hat{X}_n < \bar{d} - \delta\} P_{I_n}(\text{odd } i \,|\, \hat{X}_n < \bar{d} - \delta) + \Pr\{\hat{X}_n \geq \bar{d} - \delta\} P_{I_n}(\text{odd } i \,|\, \hat{X}_n \geq \bar{d} - \delta) = \frac{\alpha}{\alpha + (1-\alpha)} \qquad (11)$$

Now let $\delta = \delta_k \downarrow 0$. For each *k*, choose $n = n_k$ so large that $\Pr\{\hat{X}_{n_k} < \bar{d} - \delta_k\} < 1/k$.

In the choice of an odd *i*, we can first choose the maximum in each pair 1 and 2, 3 and 4, etc., and then choose the maximal maximum. By symmetry all pairs have equal probability of being chosen. Thus the probability $P_{I_n}(\text{odd } i \mid \hat{X}_n \geq \bar{d}-\delta) = P_{\{1,2\}}(1 \mid \hat{X}_2 \geq \bar{d}-\delta)$. As $\delta=\delta_k \downarrow 0$, this probability tends to

$\frac{f_1(\bar{d})}{f_1(\bar{d})+f_2(\bar{d})}$, by (9) in Cor.1.

In total, going in the limit in (11), we get

$$\frac{\alpha}{\alpha+(1-\alpha)} = p_n \to \frac{f(\bar{d})}{f(\bar{d})+f_2(\bar{d})} = 1-\frac{f_2(\bar{d})}{f(\bar{d})+f_2(\bar{d})} \geq$$

$$\geq 1-\frac{f_2(\bar{d})}{(1+\varepsilon)g(\bar{d})+f_2(\bar{d})} = 1-\frac{f_2(\bar{d})/(1+\varepsilon)}{g(\bar{d})+f_2(\bar{d})/(1+\varepsilon)} = \frac{g(\bar{d})}{g(\bar{d})+f_2(\bar{d})/(1+\varepsilon)}$$

Doing the same analysis for the other set of experiments, we get

$\bar{p}_n \to \frac{\alpha}{\alpha+(1-\alpha)} = \frac{g(\bar{d})}{g(\bar{d})+f_2(\bar{d})}$, a contradiction. □

Now we want to do a more detailed analysis to arrive at the announced proof of the Luce-Suppes conjecture. To avoid freakish behavior, we will rule out situations where f_1-g behaves as $(x-z)\sin(\frac{1}{x-z})$ or similar in a neighborhood of the point *z*, where they start to differ. The definition below may seem ad hoc, but from the above analysis it is clear that f_1-g will play an important role. And moreover, it is the behavior of f_1-g, when they start to differ, that matters. On the other hand, one can say that the condition is not very demanding.

Definition 2. Let X_1 and X_2 be independent random variables with supports D_1 and D_2, with $\bar{d} = \sup D_1 = \sup D_2 < \infty$, and with densities f_1 and f_2. Assume $p_1 =_{df} P_{\{1,2\}}(1) = \alpha$.

Let $G =_{df} F_2^{\alpha/(1-\alpha)}$, and $g = G'$. Let $z = \sup\{x \in D_1 \bigcap D_2 \mid f(x) \neq g(x)\}$

We say that the pair $\{X_1, X_2\}$ is *irregular* if, for arbitrarily small punctuated ε-neighborhoods $O_\varepsilon(z) = \{x \in D_1 \bigcap D_2 \mid x \neq z\}$ of *z*, there are points in $O_\varepsilon(z)$ where f_1 and g agree, as well as points where they do not agree.

The pair $\{X_1, X_2\}$ is called *regular* if $\bar{d} = \sup D_1 = \sup D_2 < \infty$ and it is not irregular.

For a more general set $\Omega = \{X_\alpha\}_{\alpha \in A}$ of independent random variables, we say that Ω is regular if every pair $\{X_{\alpha_1}, X_{\alpha_2}\}, \alpha_i \in A$ is regular. □

Remark: The conditions on the supports can easily be achieved by transforming the X_i, as in the preceding remark. □

The following lemma shows that irregularity puts strong demands on $\phi = f_1 - g$

Lemma 5. Suppose that $\{X_1, X_2\}$, with supports $D_1 = D_2$ is irregular and let $\phi = f_1 - g$, where f_1 and g are as in Definition 2. Suppose that ϕ is *m* times continuously differentiable. Then all derivatives of ϕ, up to (and including) the *m*-th order, are 0 at the irregularity point *z*.

Proof: Let the domains of X_1 and X_2 be D_1 and D_2.

To start with observe that ϕ is 0 arbitrarily close to *z*. By continuity we must then have $\phi(z)=0$. (If $z\notin D_1\cap D_2$ we can enlarge the domain of ϕ by including *z*.) Assume inductively that we proved the result up to (and including) derivatives of order *k*<*m*. Then by the continuity of derivatives

$$\phi(z+\delta)=0+0\cdot\delta+...+0\cdot\delta^k+\phi^{(k+1)}(z)\delta^{k+1}+O(\delta^{k+2}) \text{ for } z+\delta\in D_1\cap D_2 \qquad (12)$$

Let $\{x_i\}_{i=1}^n$ be a sequence points tending to *z*, and with $\phi(x_i)=0$. Letting $\delta_i=x_i-\bar{x}\to 0$, inserting $\delta=\delta_i$ into (5), and dividing by δ_i^{k+1} we get

$$0=\phi^{(k+1)}(z)+O(\delta_i),$$

implying $\phi^{(k+1)}(z)=0$, so by induction the result follows. □

***Corollary* 2**. If the densities f_1 and g of X_1 and $X_{\bar{1}}$ are analytical, and hence also ϕ, then the pair $\{X_1,X_2\}$ is regular. □

3. Main result

We are now in a situation to attack the general case.

Lemma 6. Let $\Omega=\{X_1,X_2\}$ be a regular pair of independent random variables. Further assume Ω to be a Luce system. Then Ω is an MWL system.

Proof: Assume as above that $p_1=P_{\{1,2\}}(1)=\alpha$ and hence $p_2=P_{\{1,2\}}(2)=1-\alpha$. Let F_i be the cdf of X_i, i=1,2 respectively. As noted above, we would in the MWL-setting have $F_1=F_2^{\alpha/(1-\alpha)}$. Let D_i, i=1,2 be the domains of the F_i, and $d=\sup D_1=\sup D_2$.

As hinted will argue by contradiction and assume that $F=_{df}\ F_1\neq G=_{df}\ F_2^{\alpha/(1-\alpha)}$, and as a consequence that $f=_{df}\ F_1'\neq g=_{df}\ G'$. Let $\phi=f-g$. Further let $H=F_2$.

Let $X_{\bar{1}}$ have cdf *G*, and be independent of X_1 and X_2. Then $\overline{\Omega}=_{df}\ \{X_{\bar{1}},X_2\}$ is an MWL-system, with $\bar{p}_1=P_{\{\bar{1},2\}}(1)=\alpha$

Let x_4 be $\sup\{x|\phi(x)\neq 0\}$. Since Ω is regular, there is a smallest $x_1<x_4$, such that ϕ is nonzero on the interval $D=(x_1,x_4)$.

Assume first that $\phi>0$ on *D*. Choose arbitrarily $x_2=_{df}\ 2x_1/3+x_4/3$, and $x_3=_{df}\ x_1/3+2x_4/3$. Since $f>g>0$ on *D*, $\bar{\varepsilon}(x)=_{df}\ \dfrac{f(x)-g(x)}{g(x)}$ is continuous and positive on $[x_2,x_3]$. Let $\varepsilon=_{df}\ \inf_{x\in[x_2,x_3]}\{\bar{\varepsilon}(x)\}>0$. Then $f\geq(1+\varepsilon)g$ on $[x_2,x_3]$.

We will now compare two choice experiments. In the first we choose between X_1 and *n* independent copies of X_2. Let $I=\{1,2,...,n+1\}$ index the choices. By the Luce property of Ω, we have that

$$p_n=_{df}\ P_I(1)=\frac{\alpha}{\alpha+n(1-\alpha)}.$$

Let $\hat{X}_n=_{df}\ \max_{i\in I}X_i$. Then $\hat{X}_n$ has cdf FH^n.

In the other experiment we choose instead between $X_{\bar{1}}$ and n independent copies of X_2. Let $\bar{I} = \{\bar{1}, 2, ..., n+1\}$ index the choices in this case. Since $\overline{\Omega}$ is an MWL-system, it also fulfills he Luce property, and we get the same choice probabilities as for Ω,

$$\bar{p}_n =_{df} P_{\bar{I}}(1) = \frac{\alpha}{\alpha + n(1-\alpha)} = p_n . \qquad (13)$$

Here we let $\hat{\hat{X}}_n =_{df} \max_{i \in \bar{I}} X_i$. Then $\hat{\hat{X}}_n$ has cdf GH^n.

By (5)

$$p_n =_{df} P_I(1) = \int_{x \in D_1} f(x) H^n(x) dx = \Big(\int_{x<x_1} + \int_{x_1}^{x_2} + \int_{x_2}^{x_3} + \int_{x_3}^{x_4} + \int_{x_4}^{d} \Big) f(x) H^n(x) dx =_{df} I_1 + I_2 + I_3 + I_4 + I_5 .$$

For these integrals we get

$$I_1 = \int_{x<x_1} f(x) H^n(x) dx = \{\text{by Lemma 5}\} = \Pr(\hat{X}_n < x_1) \cdot P_I(1 | \hat{X}_n < x_1) ,$$

$$I_2 = \int_{x_1}^{x_2} f(x) H^n(x) dx \geq \int_{x_1}^{x_2} g(x) H^n(x) dx = \{\text{by Lemma 5}\} = \Pr(\hat{\hat{X}}_n \in (x_1, x_2)) \cdot P_{\bar{I}}(1) = \{\text{by 9}\} = \Pr(\hat{\hat{X}}_n \in (x_1, x_2)) \cdot p_n$$

$$I_3 = \int_{x_2}^{x_3} f(x) H^n(x) dx \geq \int_{x_2}^{x_3} (1+\varepsilon) g(x) H^n(x) dx = (1+\varepsilon) \Pr(\hat{\hat{X}}_n \in (x_2, x_3)) \cdot p_n$$

$$I_4 = \int_{x_3}^{x_4} f(x) H^n(x) dx \geq \int_{x_3}^{x_4} g(x) H^n(x) dx = \Pr(\hat{\hat{X}}_n \in (x_3, x_4)) \cdot p_n$$

$$I_5 = \int_{x_4}^{d} f(x) H^n(x) dx = \int_{x_4}^{d} g(x) H^n(x) dx = \Pr(\hat{\hat{X}}_n \in (x_4, d)) \cdot p_n .$$

Summing up we have

$$p_n \geq \Pr(\hat{X}_n < x_1) \cdot P_I(1 | \hat{X}_n < x_1) + p_n \cdot [\Pr(\hat{\hat{X}}_n \in (x_1, x_2) + (1+\varepsilon) \Pr(\hat{\hat{X}}_n \in (x_2, x_3)) + \Pr(\hat{\hat{X}}_n \in (x_3, x_4))$$
$$+ \Pr(\hat{\hat{X}}_n \in (x_4, d))\}] = \Pr(\hat{X}_n < x_1) \cdot P_I(1 | \hat{X}_n < x_1) + p_n \cdot [\Pr(\hat{\hat{X}}_n \in (x_1, d) + \varepsilon \Pr(\hat{\hat{X}}_n \in (x_2, x_3)], \text{ or}$$
$$p_n \cdot \Pr(\hat{\hat{X}}_n < x_1) \geq \Pr(\hat{X}_n < x_1) \cdot P_I(1 | \hat{X}_n < x_1) + p_n \varepsilon \Pr(\hat{\hat{X}}_n \in (x_2, x_3))$$

Since $\hat{X}_n$ has cdf FH^n, and $\hat{\hat{X}}_n$ has cdf GH^n, and $p_n = \dfrac{\alpha}{\alpha + n(1-\alpha)}$ we get

$$p_n \cdot G(x_1) H^n(x_1) \geq F(x_1) H^n(x_1) \cdot P_I(1 | \hat{X}_n < x_1) + p_n \varepsilon \cdot [G(x_3) H^n(x_3) - G(x_2) H^n(x_2)] \cdot$$

Dividing with $p_n G(x_3) H^n(x_3)$, we get

$$\frac{G(x_1) H^n(x_1)}{G(x_3) H^n(x_3)} \geq \frac{F(x_1) H^n(x_1)}{G(x_3) H^n(x_3)} P_I(1 | \hat{X}_n < x_1) / p_n + \varepsilon [1 - \frac{G(x_2) H^n(x_2)}{G(x_3) H^n(x_3)}]$$

Since $\frac{H(x_1)}{H(x_3)} < 1$, $\frac{H(x_2)}{H(x_3)} < 1$ and $p_n = O(\frac{1}{n})$, we have as $n \to \infty$,

$\frac{G(x_1)H^n(x_1)}{G(x_3)H^n(x_3)} \to 0$, $\frac{F(x_1)H^n(x_1)}{G(x_3)H^n(x_3)} P_I(1|\hat{X}_n < x_1)/p_n \to 0$, and $\frac{G(x_2)H^n(x_2)}{G(x_3)H^n(x_3)} \to 0$, and in the limit we get the relation $0 \geq 0 + \varepsilon$, a contradiction.

Now assume that $\phi < 0$ on *D*. In this case, we let ε be the largest ε, such that $f(x) \leq (1-\varepsilon)g(x)$ on $(x_2, x_3) =_{df} (x_1 + \varepsilon, x_4 - \varepsilon)$.

Performing, the same breaking into parts of $p_n =_{df} P_I(1) = \int_{x \in D_1} f(x)H(x)dx$, we get $\leq$-signs instead of $\geq$-signs $-\varepsilon$ instead of ε, and in the end we get the limiting relation

$0 \leq 0 - \varepsilon$. Still a contradiction, thereby proving the lemma. □

As a corollary we have

Theorem 1. Let $\Omega = \{X_\alpha\}_{\alpha \in A}$ be a regular system of independent random variables. Further assume Ω to be a Luce system. Then Ω is a Mattsson-Weibull-Li system.

4. Finite Choice Sets

For independent RUMs, we saw in section 2 that IIA plus non-uniform expansion implies MWL distribution for the error terms. I made this proof of the Luce-Suppes (LS) conjecture before I was aware of the conjecture itself.

Upon reading the conjecture, I was not quite sure what conjecture really meant. In proving my version of the conjecture, I had to use unlimited choice sets, i.e. no upper bound on the number of elements in the choice set. On the other hand, Luce and Suppes assume finite choice sets in their statement (Thm. 31) that IIA leads to Luce form choice probabilities.

So the question arises what is the case for finite choice sets? The current section is devoted to proving that for any finite "universal" choice set $I = \{1, 2, \ldots, n\}$ with Luce form choice probabilities for admissible subsets, there exist random utilities of non-MWL type, giving the given probabilities. This implies that if the LS conjecture was intended to subsume finite choice sets, that part of the conjecture is counter proved.

It is further not quite clear what subsets $J \subseteq I$ are allowed as choice sets. If we allow only *I* itself, then any RUM has Luce form choice probabilities, since trivially

$p_I(i) =_{df} \Pr(\text{choosing } i \text{ from the choice set } I) = \frac{p_I(i)}{\sum_{j \in I} p_I(j)}$, for all *i*,

which is of Luce form.

Thus, to be able to say anything interesting, we need to allow for some nontrivial subsets $J \subseteq I$ as admissible choice sets.

Denoting $p_i =_{df} p_I(i)$ for short, we thus demand

$$p_J(i) = \frac{p_i}{\sum_{j \in J} p_j}$$

for at least some nontrivial subsets $J \subseteq I$.

Theorem 2. Let $I = \{1,2,...,n\}$ be a finite "universal" choice set, and let $\{w_j\}_{j\in I}$ be arbitrary positive weights. Then there is a RUM with independent random terms of non-MWL form, such that

$$p_J(i) = \frac{w_i}{\sum_{j\in J} w_j} \text{ for all subsets } J \subseteq I \text{, and all } i \in J. \quad (14)$$

Proof: First note that we WLOG may assume that e.g. $w_1 = 1$. The by the Mattsson-Weibull-Li construction we can choose any base cdf Φ and let the random terms have cdf's Φ^{w_i}. This will give a RUM with the desired choice probabilities (14) for all subsets $J \subseteq I$, and all $i \in J$. In particular we can choose Φ such that $\Phi^{w_1}(x) = \Phi^1(x) = x$ on [0, 1].

We will now modify the cdf *x* of X_1 so that we keep the choice probabilities (14), but don't have an MWL distribution anymore.

For a given *I*, there are finitely many conditions (14). For each *J*, one condition is redundant, since the $p_J(i)$ sum to 1. Further, since we only change the cdf of X_1, the set of the other X_1, still has independent cdf's of MWL form. Thus, for choice sets J, not containing 1, the condition (14) will be automatically fulfilled. With some hindsight denote the number of non-redundant conditions by *M* -2.

Let $a_k = k / M$ for $k = 0,...,M$. We now split [0, 1] into *M* intervals $C_0 =_{df} [0, a_1]$ and $C_k =_{df} (a_k, a_{k+1}]$ for $k = 1,..., M-1$. We will then modify X_1 to $\overline{X}_1$ by changing the density from 1 to λ_k on each C_k. Thus the cdf of $\overline{X}_1$ becomes $\overline{F}(x) = \sum_{l=0}^{k} \lambda_l / M + \lambda_k (x - a_k)$ on C_k.

The total mass becomes $\sum_{l=0}^{M-1} \lambda_l / M$. We thus need to require

$$\sum_{l=0}^{M-1} \lambda_l / M = 1. \quad (15)$$

We will now try to modify the λ_k-values from 1 to some other positive numbers, and still satisfy the conditions (14).

For a given *J* containing 1, consider first *i*=1. We have

$$p_J(1) = \int_0^1 [\prod_{j\in J, j\neq 1} x^{w_j}(x)] d\overline{F}(x) = \sum_{k=0}^{M-1} \int_{a_k}^{a_{k+1}} x^{\sum_{j\in J, j\neq 1} w_j} \lambda_k dx := \mathbf{b}_{1,J}^T \boldsymbol{\lambda}, \quad (16)$$

The right hand side of (16) is linear in $\boldsymbol{\lambda} =_{df} (\lambda_0,...,\lambda_{M-1})^T$. Thus, for some appropriate vector $\mathbf{b}_{1,J}^T$ it equals $\mathbf{b}_{1,J}^T \boldsymbol{\lambda}$. Similarly, we get for $i \neq 1$

$$p_J(i) = \int_0^1 \overline{F}(x) [\prod_{j\in J, j\neq i,1} x^{w_j}(x)] w_i x^{w_i - 1}(x) dx = \sum_{k=0}^{M-1} \int_{a_k}^{a_{k+1}} [\sum_{l=0}^{k-1} \lambda_l / M + \lambda_k (x - a_k)] w_i x^{\sum_{j\in J, j\neq 1} w_j - 1} dx. \quad (17)$$

Similarly to (16), the right hand side of (17) is affine in $\boldsymbol{\lambda}$, and we will denote it by $\mathbf{b}_{i,J}^T \boldsymbol{\lambda} + \beta_{i,J}$.

In total we get the constraints

$$\mathbf{b}_{i,J}^T \boldsymbol{\lambda} + \beta_{i,J} = \frac{w_i}{\sum_{j\in J} w_j}, \text{ for subsets } J \text{ comtaining 1, and indices } i \in J \text{ (where now } \beta_{1,J} = 0)$$

plus the total mass constraint

$\mathbf{1}^T \boldsymbol{\lambda} = 1$ (18)

This is a linear system. We know it has the solution $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}} =_{df} \mathbf{1} =_{df} (1,1,\ldots,1)^T$. Further, there are by construction more variables (*M*), than constraints (*M*-2+1=*M*-1). Hence the solution set is a nonempty affine subspace of dimension at least 1. Thus there are other solutions arbitrary close to $\boldsymbol{\lambda} = \mathbf{1}$, and hence a solution, $\tilde{\boldsymbol{\lambda}}$ say, that is strictly positive, and different from $\bar{\boldsymbol{\lambda}}$.

It is further obvious that with $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}$, $\bar{F}(x)$ is not of the form x^{a_i} for any a_i. This proves the Thm. □

References

Ben-Akiva, M. and S.R. Lerman (1985), *Discrete Choice Analysis*, Cambridge, The MIT Press.

Block, H.D. and J. Marschak (1960), Random Orderings and Stochastic Theories of Responses, in I. Olkin (Ed) *Contributions to Probability and Statistics*, Stanford Univ. Press, Stanford, California.

Hoyos, D. (2010), The state of the art of environmental valuation with discrete choice experiments, *Ecological Economics* **69,** 1595-1603.

Keane, M., P. Todd and K.I. Wolpin (2010), The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, *Handbook of Labor Economics* , eds. Card and Ashenfelter, Volume 2, Elsevier, p. 332-461.

Li, B. (2011), The Multinomial Logit Model revisited: A Semi-parametric Approach in Discrete Choice Analysis, *Transportation Research B* **45** 461-473.

Lindberg, P.O., E.A. Eriksson and L.-G. Mattsson (1995), Invariance of Achieved Utility in Random Utility Models, *Env. Plan.* **27,** 121-142.

Luce, R.D. (1959), *Individual Choice Behavior*, Wiley, New York.

Luce, R.D. and P. Suppes (1965), Preferences, utility, and subjective probability, in R. Luce, R. Bush, and E. Galanter, eds., *Handbook of Mathematical Psychology*, Wiley, 249-410.

Marschak, J. (1960), Binary choice constraints on random utility indicators, in K. Arrow, ed., *Stanford Symposium on Mathematical Models in the Social Sciences*, Stanford Univ. Press, 312-329.

Mattsson, L.-G. and J. Weibull (2010), Extreme-Values Invariance and Selection Probabilities, Working Paper 101231, Stockholm School of Economics.

McFadden, D. (1974), Conditional logit analysis of choice behavior, in Zaremba (ed.), *Frontiers in Econometrics*, 75-96, Academic Press, New York.

McFadden, D. (1978), Modelling the Choice of Residential Location, in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull (eds.), *Spatial Interaction Theory and Planning Models*, 75-96, North Holland: Amsterdam.

McFadden, D. (2001), Economic Choices, *American Economic Review* **91**, 351-378.

Robertson, C. A. and D. J. Strauss (1981), A Characterization Theorem for Random Utility Variables, *J. Math. Psychology* **23**, 259-263.

Thurstone, L.L. (1927), A Law of Comparative Judgment, *Psychological Review* **34**, 273-286.

Yellott, J.I. (1977), The Relationship between Luce's Choice Axiom, Thurston's Theory of Comparative Judgment, and the Double Exponential Distribution, J. Math. Psych. **15**, 109-144