Ph. D. Thesis in Transportation Science

Contributions to Probabilistic Discrete Choice

Defendant

P O Lindberg, Div. TLA, Dept. Transp. Sc., KTH

Opponent

Tony Smith, Department of Electrical and Systems Engineering, University of Pennsylvania

Supervisors

Lars-Göran Mattsson and Anders Karlström, TLA

This is not the thesis I intended to write. I planned to write on something like:

Methods for Computing Traffic Equilibria, and their Applications to Traffic Planning Problems.

But fate had other plans.

In March 2010 L-G Mattsson gave me a copy of an article together with J Weibull, on Random Utility (RU) Models, later becoming paper 2 in this thesis.

I was very intrigued by their results. In particular I felt I should be able to prove a converse to one of the main results.

I tried to and fro under the year, but I didn't manage the fine details.

Then in December, D McFadden gave a seminar at KTH. After the seminar, I was invited to a discussion with him together with LGM and JW.

They presented a new version of their paper. Again I was very intrigued, and started to suggest additions and improvements (I hope). So in the end I was invited to become a coauthor. Shortly after that I managed to prove the desired converse.

In the process I started thinking on other questions on RU Models, and during the Spring I wrote (the main parts of) papers 3-6, ending in early May, when I got a paper back from refereeing.

Random Utility Models

The opponent has given an excellent overview of this area so I will not dwell on this.

Choice Axiom

In 1959 Luce published his influential book *Individual Choice Behavior*.

It concerns probabilistic choice between discrete objects.

Its base is the Choice Axiom, relating choice probabilities of two objects/alternatives x and y from different finite Choice Sets, S and T within some universal choice set U. First some notation:

$P_T(x) =_{df}$ Pr{choosing x from T}, a prob. distr. on T.

$P(x,y) =_{\text{df}} P_{\{x,y\}}(x)$, i.e. prob. choosing x between x and y.

Choice Axiom (CA, Luce, 1959, Axiom 1). Let T be a finite subset of U such that for every $S \subseteq T$, P_S is defined.

(i) If $P(x,y) \neq 0,1$ for all $x, y \in T$, then for $R \subseteq S \subseteq T$,

$$P_T(R) = P_S(R)P_T(S).$$

(ii) If $P(x,y) = 0$ for some $x, y \in T$, then for every $S \subseteq T$

$$P_T(S) = P_{T-\{x\}}(S - \{x\}).$$

(Here $S - \{x\}$ is interpreted as $\{y \in S | y \neq x\}$) □

The choice axiom has some fundamental consequences:

1. IIA (Independence from Irrelevant Alternatives)

Proposition (IIA, Luce 1959, Lemma 3). Under CA, for any finite $T \subseteq U$, such that $P(y,x) \neq 0,1$, for all $x, y \in T$ the IIA relation below holds for all $S \subseteq T$:

$$\frac{P(x,y)}{P(y,x)} = \frac{P_S(x)}{P_S(y)}. \qquad \text{(IIA)}$$

2. LFCP (Luce Form Choice Probabilities) or Strict Utility Model (SUM)

Theorem. (*Strict Utility Model* (SUM) or *Luce Form Choice Probabilities*, (LFCP), Luce, 1959, Theorem 4.) Under CA, for any finite $T \subseteq U$, such that $P(y,x) \neq 0,1$, for all $x, y \in T$, there is a function $v : T \to \Re_{++}$ (the positive reals), unique up to scaling by positive scalars, such that for any $S \subset T$,

$$P_S(x) = \frac{v(x)}{\sum_{y \in S} v(y)}. \qquad \text{(LFCP)}$$

The Papers.

The thesis contains 8 papers:

1. Lindberg, P.O., E.A. Eriksson and L.-G. Mattsson (1995), **Invariance of Achieved Utility in Random Utility Models**, *Env. Plan.* **27,** 121-142.
2. Mattsson L.-G., J.W. Weibull and P.O. Lindberg (2011), **Extreme Values, Invariance and Choice Probabilities**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, submitted to J. Ec. Theory

The rest are singly authored unsubmitted working papers at Dept Transp. Science.

3. **Random Utility Models, IIA, Mattsson-Weibull-Li Distributions and a Conjecture by Luce and Suppes.**
4. **Choice Probabilities in Random Utility Models – Necessary and Sufficient Conditions for a Much Used Formula and Implications for the Conditional Distribution of Achieved Utility**.
5. **A Simple Derivation of the Williams-Daly-Zachery Theorem**.
6. **The Choice Axiom Revisited**.
7. **A Note on Power Invariance in Random Utility Models**.
8. **Random Utility Invariance Revisited**.

Paper 1: *Invariance of Achieved Utility in Random Utility Models.*

This 1995 paper arose from detecting and correcting an error in the Robertson & Strauss (R&S) paper that first characterized invariance in ARU Models.

The RS error is in a characterization of functional dependence.

Given two functions *f* and *g* on $\Re^n$, let f_i denote the partial derivative w.r.t. x_i.

For some particular *f* and *g* R&S observe that

$$f_i g_j = f_j g_i.$$

From this they conclude that *f* and *g* are functionally related, i.e. that there is some monotonous function *h*, such that $f(\mathbf{x}) = h(g(\mathbf{x}))$.

This is however incorrect, as we demonstrate by counterexamples.

This error keeps recurring. It even occurs in Handbook of Mathematical Economics.

It also appears in a Thm. in a recent book by Joe on multivariate distributions. This Thm. was a potential base for proving our fundamental result on Power Invariance (Paper 7).

In the paper we go on to characterize ARU invariance through the representation

$$F_X(\mathbf{x}) = \rho(-S(-\mathbf{x})), \qquad \text{(LEM)}$$

for the cdf of the random term, as compared to the R&S representation

$$F_X(\mathbf{x}) = \varphi(-H(e^{-\mathbf{x}})) \qquad \text{(RS)}$$

The LEM representation has the advantage that ρ is the cdf of the maximum utility, S is (up to an additive constant) the expected achieved utility, and the partial derivatives S_i are the choice probabilities.

S is additively homogeneous, as compared to H that is linearly homogeneous.

Paper 3: *Random Utility Models, IIA, Mattsson-Weibull-Li Distributions and a Conjecture by Luce and Suppes.*

This is the paper inspired by the early version of the paper by LGM and JW.

LGM and JW (2010) studied independent r.v.'s X_i, each with a cdf of the form

$$\Phi^{\alpha_i}(x), \qquad \text{(MWL)}$$

for some powers $\alpha_i > 0$. (L for Li, 2011, who showed the same)

Then they showed that the corresponding choice probabilities are

$$P_S(i) = \frac{\alpha_i}{\sum_{j \in S} \alpha_j}, \qquad \text{(LFCP)},$$

i.e. of Luce form.

I prove a converse of this:

If we have an (infinite) set of independent random variables $\{X_i\}_{i \in G}$, and corresponding reals $\alpha_i > 0$, such that for any finite $T \subset G$, (LF) holds,

and G is closed under non-uniform expansion, i.e. we can introduce any number of independent copies of any X_i into the model,

and further the cdf's of the X_i are well behaved in the sense that the derivatives of certain powers of their cdf's do not both agree and disagree arbitrarily close to a certain point in their supports.

Then, then the cdf's must be of the Mattsson Weibull form (MWL) above. □

The proof studies in detail the integrals that define the choice probabilities, and going in the limit with the number of X_i, chosen appropriately, derives a contradiction unless the cdf's have the desired forms.

Going to the sources, it turned out that Luce and Suppes (1965) may have conjectured this result. Citing an unpublished proof by Holman and Marley

(that any set of choice objects fulfilling the LFCP is a RU Model)

they go on to note, that the form of the choice probabilities is 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).

Paper 2: *Extreme Values, Invariance and Choice Probabilities*.

This is the current version of the paper that inspired the current thesis. The paper introduces dependent generalizations of the above mentioned MWL class.

The cdf's of these distributions have the form

$$F_X(\mathbf{x}) = \exp(-G(-\mathbf{a} * \ln(\Phi(\mathbf{x})))),$$

where G is a linearly homogenous function fulfilling some normalization conditions, Φ is a univariate cdf (cumulative distribution function).

Further a univariate function of a vector (such as $\Phi(\mathbf{x})$) is the vector of function values, (i.e. $(\Phi(x_1), \Phi(x_2), ..., \Phi(x_n))$), and * stands for component-wise multiplication of vectors.

These distributions are generalizations of the GEV-distributions mentioned by the opponent.

Together with GEV distributions they exhibit closed form choice probabilities and invariance of achieved utility. The marginal distributions are $\Phi^{a_i}(x)$, which allows some freedom as compared to GEV which has Gumbel marginals.

Paper 4. *Choice Probabilities in Random Utility Models – Necessary and Sufficient Conditions for a Much Used Formula and Implications for the Conditional Distributions of Achieved Utility*.

This paper arose out of my dissatisfaction with the initial "engineering math" derivation of a central choice probability formula in paper 2 (and paper 1).

The formula that gives the choice probability states

$$p_i = \int_{-\infty}^{\infty} F_i(x, x, ..., x)dx, \qquad \text{(CPF)}$$

where F is the cdf of the random vector **X**, and F_i is its derivative w.r.t. x_i.

The original derivation was something like this:

$$p_i = \Pr\{X_i > X_j, j \neq i\} = \int_{-\infty}^{\infty} \Pr\{X_i \in (s, s + ds \;\&\; X_j < s)dx = \int_{-\infty}^{\infty} F_i(x, x, ..., x)dx.$$

As an alternative, we assumed that the distribution of **X** had densities, and integrated all the way up from the bottom, which seemed as overkill. Further all papers I found, did the same, or just stated the formula without proof. I asked around by probabilists and transport scientists (including my present opponent), but no one seemed to know under which conditions (CPF) was valid. So I undertook to study it.

The study becomes quite messy, first deriving many relations between the partial derivatives of F, and of $f(x) =_{df} F(x, x,.., x) = \Pr\{\hat{X} \le x) = \Pr\{\max_i X_i \le x)$.

For (CPF) to be of any practical use, we further need

$$\sum_i p_i = 1. \qquad (1)$$

Introduce $\bar{p}_i =_{df} \Pr\{X_i \ge X_j, j \ne i\}$, and let $p_i =_{df} \Pr\{X_i > X_j, j \ne i\}$

Then $\sum_i p_i = 1 = \sum_i \bar{p}_i$ if and only if $p_i = \bar{p}_i$ for all i. So this is a necessary condition to put on F and the (CPF) formula.

So let's say that F is *well-behaved* (in the choice probability sense) if (CPF) and (1) are valid.

In practice, we change (CPF) to a relaxed version, that does not require differentiability if F:

$$p_i = \int_{-\infty}^{\infty} F_i^+(x, x,..., x)dx, \qquad \text{(CPF+)}$$

where F_i^+ is the directional derivative of F w.r.t. x_i in the positive direction.

Looking closely at how F attributes mass to rectangles, one can see that in general

$$f^+(x) \geq \sum_{i=1}^{n} F_i^+(x\mathbf{1}),$$

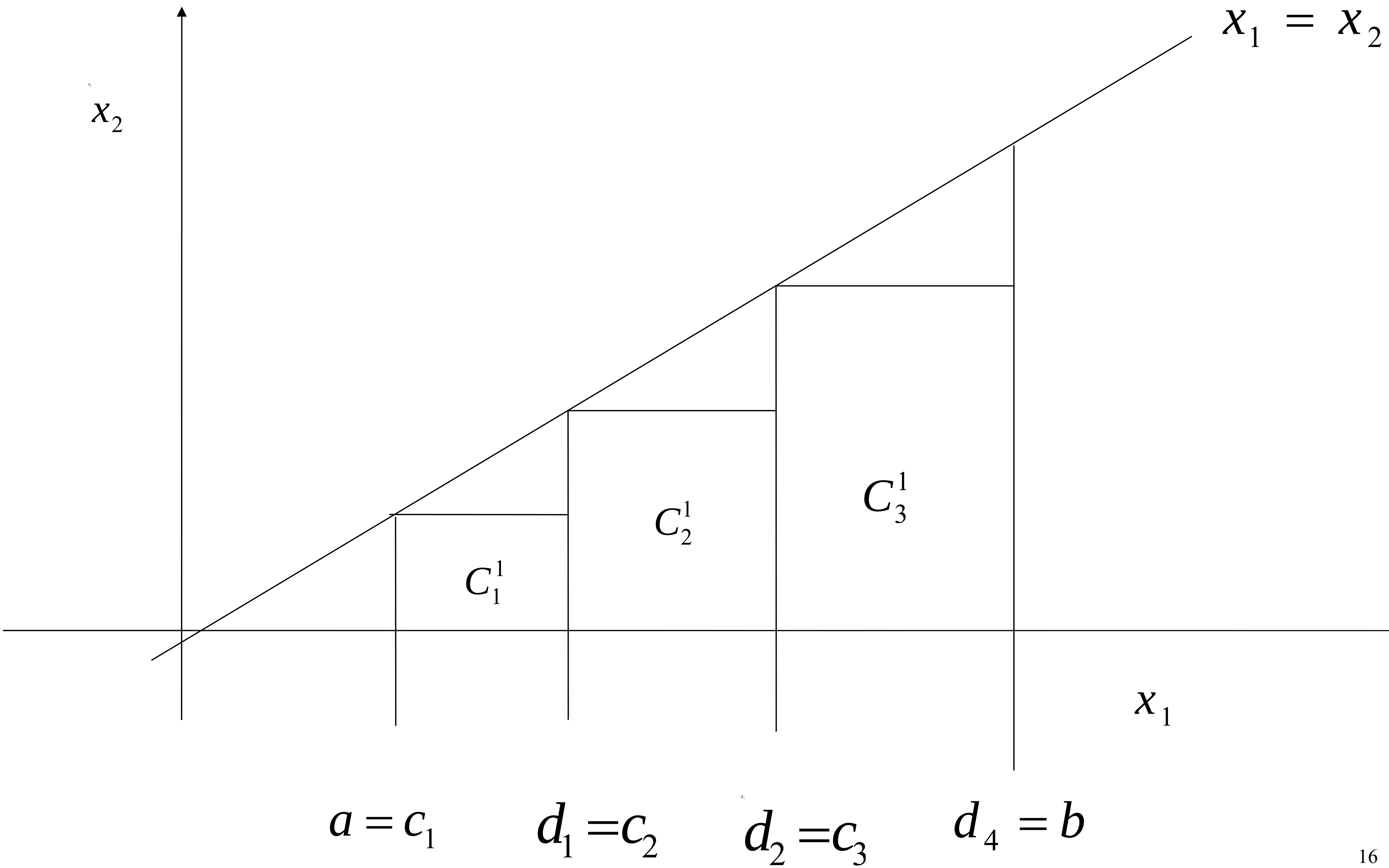
when the derivatives exists. (Here f^+ is the directional derivative of f)

Thus, to have $\sum_i p_i = 1$, we need $1 = \int_{-\infty}^{\infty} f^+(x, x, ..., x)dx$, which in turn requires f to be absolutely continuous.

Otherwise, f could in principle be freakish: non-decreasing, going from 0 to 1 on an interval [a, b], and differentiable a.e., but still have

$$\int_a^b f(x)dx = 0.$$

To arrive at our result we approximate probabilities of the form $\Pr\{X_i \in (a,b] \,\&\, X_i > X_j, j \neq i\}$ by under-estimations of the type in the figure next.



$x_1 = x_2$
x_2
x_1
C_1^1
C_2^1
C_3^1
$a = c_1$
$d_1 = c_2$
$d_2 = c_3$
$d_4 = b$

After hard work we arrive at results of the type

Theorem. Assume *f* absolutely continuous, and $\sum_i p_i = 1$.

Then $F_i^+(x, x, \ldots, x)$ exists a.e. and moreover $f'(x) = \sum_{i=1}^{n} F_i^+(x, x, \ldots, x)$ a.e. □

Which in turn gives

Corollary. The conditions

(i) *f* absolutely continuous, and

(ii) $\sum_i p_i = 1$,

are necessary and sufficient for *F* to be well-behaved. □

Summing up we arrive at more practical conditions

Theorem. For *F* to be well-behaved, it is sufficient that *F* is continuous and piecewise differentiable with locally bounded gradient along the diagonal. □

Conditional distribution of $\hat{X}_i$.

We introduced $\hat{X} = \max_i X_i$ before. Its cdf is $f(x)$.

Sometimes one wants to study the distribution of $\hat{X}$ conditioned on alternative i being chosen. Let us term this variable $\hat{X}_i$. The cdf of $\hat{X}_i$ is

$$F^{(i)}(x) =_{df} \Pr\{X_i \le x \,\&\, X_j \le X_i, j \ne i\}.$$

Using our machinery it follows rather straightforwardly that

$$\hat{F}^{(i)}(x) = \frac{1}{p_i}\int_{-\infty}^{x} F_i^{+}(s, s, \ldots, s)ds$$

This will be used in later papers.

Paper 5. *A simple derivation of the Williams-Daly-Zachery theorem.*

This paper concerns ARU models. Let's write them in the form

$$U_i(\mathbf{v}) = v_i + X_i, \qquad (2)$$

where the v_i are observable *population values* and the X_i are unobservable *individual values*, where the lack of knowledge is modeled by randomness. Let

$$\overline{P}_i(\mathbf{v}) = \Pr(v_j + X_j \leq v_i + X_i, j \neq i), \qquad \text{(CPV)}$$

be the choice probabilities as functions of the v_i.

The Williams-Daly-Zachery "theorem" says that under appropriate conditions,

$$\overline{P}_i(\mathbf{v}) = \frac{\partial u(\mathbf{v})}{\partial v_i}, \qquad \text{(WDZ)}$$

where $u(\mathbf{v})$ is the expected achieved utility as a function of $\mathbf{v}$.

I knew since "long" that if the choice probabilities (CPV) are continuous as functions of the v_i, then the WDZ theorem follows.

The previous paper, gave me the machinery to prove the continuity of $\overline{P}_i(\mathbf{v})$. But in the end I really didn't need the machinery.

To see how the continuity of $\overline{P}_i(\mathbf{v})$ enters, assume that the current v_i-value changes from v_i to $v_i + \Delta v_i$, and let Δp_i be the corresponding change in $\overline{P}_i(\mathbf{v})$.

We can divide the choice makers into groups according to their choices:

G_1, those who choose alternative *i* both before and hence also after the change of v_i,

G_c, those who change to alternative *i*, from some other alternative, and

G_2, those who don't choose alternative *i*, neither before nor after the change.

By analyzing the effects of the change in v_i for the different groups, we arrive at the following Lemma:

Lemma. Let Δu be the change in the expected achieved utility when v_i changes, Then

$p_i \Delta v_i \leq \Delta u \leq (p_i + \Delta p_i)\Delta v_i$. □

As a corollary we can then derive

Proposition. Assume that the choice probability $P_i(\mathbf{v})$ is continuous at $\mathbf{v} = \bar{\mathbf{v}}$, then the partial derivative of $u(\mathbf{v})$, the expected achieved utility, with respect to v_i exists at $\bar{\mathbf{v}}$, and fulfils the Williams-Daly-Zachery "relation"

$$\frac{\partial u(\mathbf{v})}{\partial v_i} = P_i(\mathbf{v}). \qquad \square$$

A more detailed analysis of the choice probabilities shows that the $P_i(\mathbf{v})$ are continuous as functions of single v_j if the probability of ties is zero, due to the monotonicity properties of the $P_i(\mathbf{v})$

And as a consequence, also due to monotonicity properties, the $P_i(\mathbf{v})$ are continuous as functions of **v**, if the probability of ties is zero.

Paper 6. *The Choice Axiom Revisited.*

This paper arose from reading Luce's *Individual Choice Behavior*, and then questioning his assumption.

Remember the Choice Axiom:

Choice Axiom (CA, Luce, 1959, Axiom 1). Let T be a finite subset of U such that for every $S \subseteq T$, P_S is defined.

(i) If $P(x, y) \neq 0, 1$ for all $x, y \in T$, then for $R \subseteq S \subseteq T$,

$$P_T(R) = P_S(R)P_T(S).$$

(ii) If $P(x, y) = 0$ for some $x, y \in T$, then for every $S \subseteq T$

$$P_T(S) = P_{T-\{x\}}(S - \{x\}).$$

(Here $S - \{x\}$ is interpreted as $\{y \in S \mid y \neq x\}$) □

Why is Luce restricting the applicability in (i)?

I thought the following Strong Choice Axiom (SCA) was more natural:

Strong Choice Axiom (SCA). For any finite choice set $T \subseteq U$, and any $R \subseteq S \subseteq T$,

$$P_T(R) = P_S(R)P_T(S).$$ □

It turned out that Luce originally had this form of his axiom, but rejected it for leading to consequences he thought counterintuitive.
SCA leads to the following result:
Lemma. Under SCA: For any finite $T \subseteq U$ containing x, y, we have
(i) If $P(x,y) > 0$, then $P_T(y) = 0$ if $P_T(x) = 0$
(ii) If $P(x,y) < 1$, then $P_T(y) > 0$ if $P_T(x) > 0$. □
I.e. if $P(x,y) \in (0,1)$, then $P_T(y)$ and $P_T(x)$, are either both zero, or else both positive.

Consider a pair x, y with $P(x,y) = 0$. By SCA, $P_T(x) = P_{\{x,y\}}(x)P_T(\{x,y\}) = 0$
Then we introduce a z, such that $T = \{x,y,z\}$.
By the Lemma, if $P(x,z) \in (0,1)$, then we must have $P_T(z) = 0$, which Luce found counterintuitive.
Now, start instead with x, z, with $P(x,z) \in (0,1)$ (and hence $P_{\{x,z\}}(x)$ and $P_{\{x,z\}}(z)$ positive) and then add y. Luce would like to have the possibility that $P_T(x)$ drops to 0, while $P_T(z)$ remains positive, which I find counterintuitive.

The Lemma implies (as can be seen), that the relation $P(x,y)\in(0,1)$ is an equivalence relation. Thus, U decomposes into equivalence sets C_α for α in some index set $\boldsymbol{A}$. Moreover, it can be seen that, if $x\in C_\alpha$, $y\in C_\beta$, and $P(x,y)=1$, then $P(u,v)=1$ for any $u\in C_\alpha$ and $v\in C_\beta$.

Thus we have an order relation $\alpha\succ\beta$ on $\boldsymbol{A}$ that turns out to be a strict total order. For a given α we can define a positive function $v(x)=v_\alpha(x)$ on C_α:

Theorem 1. (*Luce Form Choice Probabilities*, LFCP.) Assume IIA to hold. Let C_α be an equivalence set. Then there is a function $v_\alpha : C_\alpha \to \Re_{++}$ (the positive reals), unique up to scaling by positive scalars, such that for any finite $T\subseteq C_\alpha$, and any $x\in T$,

$$P_T(x)=\frac{v_\alpha(x)}{\sum_{y\in T} v_\alpha(y)}. \qquad \text{(LFCP)}$$

We know from above, that if $\alpha \succ \beta$, then $P(y,x)=0$ for any $x \in C_\alpha$ and $y \in C_\beta$, and hence $P_T(y)=0$ for any finite *T* containing *x* and *y*. For this reason we will *define* $v_\alpha(y) =_{df} 0$ for any $y \in C_\beta$ with $\alpha \succ \beta$. Then the theorem can be generalized:

Theorem 2. (*Extended Luce Form Choice Probabilities*, (ELFCP))

For any finite $T \subseteq U$, with $\overline{\alpha} =_{df} \max\{\alpha \mid T \cap C_\alpha \text{ nonempty}\}$, and with $v_{\overline{\alpha}}$ defined on $C_{\overline{\alpha}}$ according to Theorem 1 , and extended as in the definition above to all C_β for any $\beta \prec \overline{\alpha}$, then for any $x \in T$,

$$P_T(x) = \frac{v_{\overline{\alpha}}(x)}{\sum_{y \in T} v_{\overline{\alpha}}(y)}. \qquad \text{(ELFCP)}$$

Moreover, for any $S \subseteq T$ with $\sum_{y \in S} v_{\overline{\alpha}}(y) > 0$, and for any $x \in S$,

$$P_S(x) = \frac{v_{\overline{\alpha}}(x)}{\sum_{y \in S} v_{\overline{\alpha}}(y)} \qquad \square$$

This theorem gives a sort of lexicographic Luce form of choice probabilities.

Paper 7. *A Note on Power Invariance in Random Utility Models.*

This paper arose under the work with paper 2, together with LGM and JW.

Paper 2 introduces a class of cdf's of the form

$F(\mathbf{x}) = \exp(-H(-\mathbf{a} * \ln(\Phi(\mathbf{x})))),$ (MW)

Where $H : \Re^n_{++} \rightarrow \Re_{++}$ is a homogenous function, Φ is a univariate cdf, * denotes component-wise multiplication of vectors, and a function of a vector is the vector of function-values of the components, e.g. $\Phi(\mathbf{x}) =_{df} (\Phi(x_1), \Phi(x_2), \ldots \Phi(x_n))$. This class generalizes the GEV-class. It has some remarkable properties, which it shares with GEV (and MWL).

1. Closed form choice probabilities:

$$p_i = \frac{a_i H_i(\mathbf{a})}{H(\mathbf{a})}$$

2. It has the *invariance property*, i.e. the distribution of $\hat{X}_i$ is independent of *i*:

The cdf of $\hat{X}_i$ is

$\hat{F}^{(i)}(x) = \hat{F}(x) = \{\text{i.e. the distribution of } \hat{X}\} = (\Phi(x))^{H(\mathbf{a})}.$

The paper on *power invariance* arose as an attempt to motivate the class above.

Already the first mentioned paper, with the MWL class, by LGM and JW had the mentioned invariance properties. There one had an RU model with independent r.v.'s X_i, each with a cdf Φ^{α_i}.These above result then hold for any $\alpha_i > 0$.

How can this be generalized to dependent cases?
A unified way to treat dependence is through *copulas*.
A copula is a cdf *C* on the unit cube, with marginal distributions that are U[0,1].
Given a copula *C*, and univariate cdf's Φ_i, then the function F^{Φ} defined by
$F^{\Phi}(\mathbf{x}) =_{df} C(\mathbf{\Phi}(\mathbf{x})) =_{df} C(\Phi_1(x_1), \Phi_2(x_2), ..., \Phi_n(x_n))$ is a cdf with marginals Φ_i.
And conversely, any cdf *F* with marginals Φ_i can be written in this way, for some appropriate copula *C*.

Let the *independence Copula* be the copula $\overline{C}(\mathbf{x}) =_{df} x_1 \cdot x_2 \cdot ... \cdot x_n$.
If we plug in univariate cdf's Φ_i into this copula, we get a distributions with independent r.v.'s and cdf $F(\mathbf{x}) = \Phi_1(x_1) \cdot \Phi_2(x_2) \cdot ... \cdot \Phi_n(x_n)$.

By the result of LGM and JW, the independence copula $\overline{C}$ has the invariance property. But not only this. For any $\alpha_i > 0$, the cdf $\overline{C}_{\mathbf{a}}(\mathbf{x}) =_{df} x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_n^{\alpha_n}$ has the invariance property.
So let's say that a copula C has the *power invariance property*, if for any $\alpha_i > 0$, the cdf $F_{\mathbf{a}}(\mathbf{x}) =_{df} C(x_1^{\alpha_1}, x_2^{\alpha_2}, \ldots, x_n^{\alpha_n})$ has the invariance property.

Using results from a paper with the same authors as paper 1, I can then show:
Theorem. A continuously differentiable copula C, with non-vanishing gradient, has the power invariance property if and only if it can be written in the form

$$C(\mathbf{x}) = \exp(-H(-\ln \mathbf{x})), \qquad \text{(PI)}$$

for some differentiable linearly homogenous function $H : \Re_{++}^N \to (0, \infty)$.
on: Introducing powers of the x_i, we get
$C(\mathbf{x}^{\mathbf{a}}) = \exp(-H(-\ln \mathbf{x}^{\mathbf{a}})) = \exp(-H(-\mathbf{a} * \ln \mathbf{x}))$.
Then introducing marginal cdf's Φ, we get
$C(\Phi^{\mathbf{a}}) = \exp(-H(-\mathbf{a} * \ln \Phi(\mathbf{x})))$, i.e. the class introduced in paper 2.

Paper 8**.** *Random Utility Invariance Revisited*.

In this last paper, I reuse some results for the class (MW) above. Results, that did not make it into the paper.

I apply them to some old results by Robertson & Strauss (1981), concerning invariance in ARU models,

$$U_i = v_i + X_i. \qquad \text{(ARU)}$$

R&S show that if the model (ARU) has invariance for all v_i, then the cdf of $\mathbf{X}$, must be of the form

$$F_X(\mathbf{x}) = F_{RS}(\mathbf{x}) = \varphi(H(e^{-\mathbf{x}})), \qquad \text{(RS)}$$

for some monotone univariate φ and homogenous H (given that it is indeed a cdf).

The proof of R&S contains errors, which are corrected in paper 1.

So for which φ and H is F_X a cdf? This is the theme of paper 8.

Smith (1984) has made a corresponding analysis for the GEV case, where

$F_{GEV}(\mathbf{x}) = \exp(-H(e^{-\mathbf{x}})).$ (PCH)

He shows that for F_{GEV} to be a cdf it is necessary and sufficient that

$(-1)^{(n-1)} H_{i_1 i_2 \dots i_m,}(\mathbf{x}) \geq 0$ for all $i_1 < i_2 < \dots < i_m \leq n$, and all $\mathbf{x} \in \Re^n_{++}$.

The R&S case is more complicated since we have an unspecified function φ.

Initially, we make some normalizations, showing that we WLOG may assume:

Assumption. We assume that F_{RS} has the representation

$$F_{RS}(\mathbf{x}) =_{df} \phi(-H(e^{-\mathbf{x}})), \qquad (8)$$

where $\phi : (-\infty, 0] \to [0,1]$ is non-decreasing with $\phi(-\infty) = 0$ and $\phi(0) = 1$, and H is positive and linearly homogenous, with $H(\mathbf{u}_i) = 1$, where $\mathbf{u}_i$ is the i-th coordinate unit vector. We will say that such ϕ and H are RS-*appropriate*.

This implies that the cdf's of the marginals are $\phi(-e^{-\mathbf{x}})$

Simplifying further, we show that F_{RS} is a cdf if and only if $\overline{F}_{RS}(\mathbf{x}) =_{df} \phi(-H(-\mathbf{x}))$ is a cdf.

Classical conditions for a function F on $\Re^N$ to be a cdf are

Proposition. A function, $F : \Re^N \to \Re$ for which all strictly mixed partials exist, is a cdf

a. if conditions (i) to (iii) below are fulfilled, and

b. only if the conditions (i) to (iv) are met:

(i) $\lim_{x_k \to -\infty} F(\mathbf{x}) = 0$, for all k, (c1)

(ii) $\lim_{\min_k x_k \to \infty} F(\mathbf{x}) = 1$, (c2)

(iii) $\dfrac{\partial F(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_N} \geq 0$ for all $\mathbf{x}$, (c3)

(iv) $\dfrac{\partial F(\mathbf{x})}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_m}} \geq 0$ for all $\mathbf{x}$ and $i_1 < i_2 < \ldots < i_m, m \leq N$.(c4) □

Using these it is straightforward that a necessary condition is

$(-1)^{(1-1)} H_i(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \Re^n_{++}$, (PCH1)

The cdf φ lives on $(-\infty,0]$, but this doesn't imply that the support of φ equals $(-\infty,0]$. We first show that the supremum of the support must be 0, which is important for the next step. Since the marginals of F_{RS} are $\phi(-e^{-\mathbf{x}})$, the sup of their supports must be ∞.

We need to restrict the behavior of φ in the neighbourhood of 0. To this end we define φ to be RS-*well-behaved*, if for all $m \leq N$, the derivatives $\phi^{(m)}$ of ϕ fulfill the following regularity condition:

$\phi^{(m)}(s)$ is $o(\phi'(s)(-s)^{1-m})$ ("little o") as $s \uparrow 0$,

(i.e. for every $\varepsilon > 0$ there is a $\delta > 0$, such that $\left|\phi^{(m)}(s)\right| \leq \varepsilon\phi'(s)(-s)^{1-m}$ on $(-\delta,0)$).

Using well-behaved-ness, we can in a first (or maybe second) step show that

$(-1)^{(2-1)} H_{ij}(\mathbf{x}) \geq 0$, all i,j and $\mathbf{x} \in \Re^n_{++}$ (PCH2)

using the homogeneity of H.

Differentiating further, one can see that

$$\overline{F}_{i_1 i_2 \dots i_m}(\mathbf{x}) = \sum_{k=1}^{m} (-1)^{m-k} \phi^{(k)}(-H) S_k^m(H),$$

where the $S_k^m(H)$ are polynomials in the derivatives of H.

Further one can show that $S_k^m(H(-\mathbf{x}))$ is homogenous of degree (k- m). This gives:

If ϕ and H are n times differentiable, with ϕ RS-well-behaved, then a necessary condition for $\overline{F}_{RS}$ to be a cdf is that for any $i_1 < i_2 < \dots < i_m, m \le n$

$$(-1)^{m-1} H_{i_1 i_2 \dots i_m}(-\mathbf{x}) \ge 0, \text{ for any } \mathbf{x} \in \Re^n_{--}. \qquad \text{(PCHm)}$$

More Conditions on φ

The positivity condition (PCH2) can be used to show that in fact

$H_i(-\mathbf{x}) > 0$, for any $\mathbf{x} \in \Re^N_{--}$.

This can in turn be used to show that

The support S_ϕ of φ has no holes, i.e. it is either $(-\infty,0]$ or $(-c,0]$ for some $c > 0$.

It is rather natural to consider the following positivity conditions on φ:

$\varphi^{(m)}(x) \geq 0$ for all $x \in S_\phi$. (PCφ)

It is then straightforward to see that if φ and H fulfil their respective positivity conditions (PCφ) and (PCHm) then $\overline{F}_{RS}$ fulfills the positivity condition (c3) for a cdf.

The limit conditions (c1) and (c2) are seen to follow rather easily from that H is homogenous and RS-appropriate. In total:

If φ and H fulfil there respective positivity conditions (PCφ) and (PCHn) then $\vec{F}_{RS}$ and hence F_{RS} are cdf's.

Special cases of φ

Supp of φ *equals* $(-\infty,0]$.

What φ, except the GEV case $\varphi(x)=e^{-z}$ work on $(-\infty,0]$?

Test case $\varphi(x)=\dfrac{1}{1-x}=(1-x)^{-1}$.

Checking gives (after some calculations)

$\varphi^{(n)}(x)=n!(1-x)^{-(n-1)}\geq 0$.

Thus, together with an H, fulfilling the positivity condition (PCHn), this φ gives RS distributions.

The marginal can be seen to be logistic, i.e. with cdf

$\varphi(-e^{-x})=\dfrac{1}{1+e^{-x}}$.

Supp of φ *equals* $(-c,0]$.

Here we can try e.g. with $\varphi(x) = 1 + x$.

This φ obviously fulfils $\varphi^{(m)}(x) \geq 0$, and hence gives RS-distributions with any H fulfilling the positivity conditions (PCH$\underline{\text{m}}$).

The marginals are seen to be exponential. In particular they are non-negative.

Finding φ for a given marginal-

We know that the marginals of an RS-distribution have cdf $\varphi(-e^{-x})$.

Thus for a given desired marginal cdf, we can derive φ

Example. Erlang-*2*.

The Erlang-*2* distribution has density $f(x;2) = x^{2-1}e^{-x}/(2-1)!$,

and cdf $F(x;2) = 1 - e^{-x}\sum_{n=0}^{1} x^{-n}/n! = 1 - e^{-x}(1+x)$.

Thus we want $\varphi(-e^{-x}) == 1 - e^{-x}(1+x)$.

Let $y = -e^{-x} \in (-1,0)$, i.e. $x = -\ln(-y) \in (0,\infty)$. We get

$\varphi(y) = 1 - e^{-(-\ln(-y))}(1-\ln(-y)) = 1 + y(1-\ln(-y))$.

Checking derivatives we get, remembering $\varphi(y) = 1 + y(1 - \ln(-y))$

$$\varphi'(y) = (1 - \ln(-y)) + y(-\frac{1}{-y} \cdot (-1)) = \ln(-y)) \geq 0 \text{ on } (-1,0].$$

$$\varphi''(y) = -\frac{1}{-y} \cdot (-1)) = \frac{1}{-y} = (-y)^{-1} \geq 0,$$

$\varphi'''(y) = (-1)(-y)^{-2} \cdot (-1) = (-y)^{-2} \geq 0$, and in general

$$\varphi^{(n)}(y) = (n-2)!(-y)^{-(n-1)} \geq 0.$$

Thus, with this φ we get RS-distributions with Erlang-*2* marginals.