KTH Architecture and the Built Environment

Contributions to Probabilistic Discrete Choice

P O Lindberg, Department of Transport Science, KTH

Doctoral Thesis in Transport Science

With specialization in Transport and Location Analysis

May 2012

Division of Transport and Location Analysis

Department of Transport Science

Royal Institute of Technology (KTH)

SE-100 44 Stockholm, Sweden

TRITA-TSC-PHD 12-001
ISBN 978-91-85539-86-4
ISSN 1653-445X

Akademisk avhandling som med tillstånd av Kungliga Tekniska Högskolan i Stockholm framlägges till offentlig granskning för avläggande av teknologie doktorsexamen fredagen den 1 juni 201 kl 9 i sal D2, Lindstedtsvägen 5, Kungliga Tekniska Högskolan, Stockholm

Abstract

Probabilistic Discrete Choice, or *Discrete Choice* (DC) for short, is central tool in traffic and transport planning. In these areas DC is used to model central entities, such as choice of transportation mode, workplace, and housing. DC models are also used in many other areas, such as consumer theory, product differentiation and labor economics. This thesis contains 8 papers that contribute to some rather fundamental questions in DC theory.

The central model type in Probabilistic DC is *Random Utility* (RU) *Models*, in particular in the form of *Additive RU* (ARU) Models. It is the central modeling tool in the above mentioned applications.

All papers in this thesis, except paper 6, concern RU Models.

Four of my papers concern *invariance* in RU Models, i.e. the property that the statistical distribution of achieved utility is independent of the chosen alternative. This somewhat unexpected property is showed by most central models in DC, such as MNL (Multinomial Logit) and GEV (Generalized Extreme Value) models.

The oldest paper, from 1995, notes and corrects some errors in the Robertson-Strauss (RS) paper which first characterized invariance. Our paper also gives a different characterization of invariance.

Paper 2, together with Mattsson and Weibull, which was the starting point of this thesis, introduces a new class of distributions giving invariance, and containing inter alias distributions with Weibull and Fréchet marginals. We later term the "independent" distributions in this class independent Mattsson-Weibull-Li (MWL) distributions

The third invariance paper, paper 7, goes on to study *power invariance*, a stronger property than invariance. We show that the class of power invariant distribution spans the general class introduced in paper 2.

The fourth paper, paper 8, goes back to the RS characterization of invariance, and studies its consequences for the cdf's in questions.

The remaining three RU papers, and the non-RU paper concern other fundamental aspects of DC models.

In the first of these, paper 3, I give a (partial) proof of a conjecture by Luce and Suppes, namely that the independent MWL-distributions are the only ones that have choice probabilities of "Luce form"

$$P_I(i) =_{df} \Pr\{i \text{ chosen out of } I\} = \frac{w_i}{\sum_{j\in I} w_j}$$

for appropriate weights w_j. This proof is carried through under the assumption that "non-uniform expansions" of the choice set *I* are allowed.

In the second basic paper, paper 4, I derive (rather weak) necessary and sufficient conditions for a much used formula for the choice probabilities, to hold. These results are further used to derive the distributions of achieved utility, conditioned on the chosen alternative. This is useful in the above studies of invariance, as well as in the papers below.

The third basic paper, paper 5, derives the Williams-Daly-Zachary (WDZ) theorem under the same weak conditions as the above mentioned choice probability formula. The WDZ theorem is an analog, in the DC setting, to Roy's identity in consumer theory.

The fourth basic and non-RU paper, paper 6, goes back to the classical *Choice Axiom* by R.D. Luce. I modify the choice axiom to a (in my view) more natural form, which I term the *Strong Choice Axiom*. This form allows me to derive a more general, "lexicographic" version of the above Luce form choice probability formula. But in the end, a more detailed study shows that the new and the old versions of the choice axiom are in fact equivalent.

Svensk sammanfattning

Diskret Valhandlingsteori (eng. Discrete Choice, DC, Theory) är ett centralt verktyg inom trafik- och transportplanering. Inom dessa områden används DC för att modellera centrala storheter, som val av färdmedel, arbetsort och bostad. DC-modeller används också inom många andra områden, som konsumentteori, produktdifferentiering, och arbetsmarknadsekonomi. Denna avhandling består av 8 artiklar, som avhandlar tämligen fundamentala frågor inom valhandlingsteori.

Den centrala modelltypen inom DC är Slumpnyttemodeller (Random Utility, RU, Models), speciellt i form av additiva RU modeller. Dessa är det centrala modelleringsverktyget i ovannämnda tillämpningar.

Alla artiklar i denna avhandling, utom artikel 6, handlar om RU modeller.

Fyra av artiklarna berör invarians i RU modeller, d.v.s. egenskapen att sannolikhetsfördelningen för uppnådd nytta är oberoende av valt alternativ. Denna, något oväntade egenskap uppvisas av de mest central modellerna inom DC, som MNL- (Multinomial Logit) och GEV- (Generalized Extreme Value) modellerna.

Den äldsta artikeln, från 1995, noterar och korrigerar vissa fel i den artikel av Robertson och Strauss (RS) som först behandlade invarians. Vår artikel ger också en ny karakterisering av invarians.

Artikel 2, tillsammans med Mattsson och Weibull, vilken var startpunkten för denna avhandling, introducerar en ny klass av fördelningar som ger invarians, och som innehåller bl.a. fördelningar med Weibull- och Fréchet-fördelade marginalfördelningar. Vi kallar senare de oberoende fördelningarna i denna klass för oberoende Mattsson-Weibull-Li (MWL)-fördelningar.

Det tredje invariansartikeln, artikel 7, går vidare med att studera ”*power invariance*”, en starkare egenskap än invarians. Vi visar bl.a. att denna klass spänner upp den generella klass som införts i artikel 2.

Den fjärde invariansartikeln, artikel 8, går tillbaks till RS-karakteriseringen av invarians och studerar dess konsekvenser.

De tre resterande RU-artiklarna, samt icke-RU-artikeln, artikel 6, behandlar andra fundamentala frågor inom DC.

I den första av dessa, artikel 3, ger jag ett (partiellt) bevis för en förmodan av Luce och Suppes, nämligen att de oberoende MWL-fördelningarna är de enda som har valsannolikheter av Luce-form:

$$P_I(i) =_{df} \Pr\{i \text{ väljs ur } I\} = \frac{w_i}{\sum_{j\in I} w_j},$$

för lämpliga vikter w_j. Detta bevis görs under antagandet att icke-uniform expansion av valmängden I är tillåten.

I den andra grundläggande artikeln, artikel 4, härleder jag (ganska svaga) nödvändiga och tillräckliga villkor för att en ofta använd formel för valsannolikheterna ska gälla. Dessa resultat används vidare för härleda fördelningen för uppnådd nytta, betingat av valt alternativ. Denna fördelning används i ovanstående studier av invarians, liksom även i artiklarna nedan.

Den tredje grundläggande artikeln, artikel 5, härleder Williams-Daly-Zachary (WDZ)-teoremet under samma svaga villkor som valsannolikhetsformeln. WDZ-teoremet är en analog i DC-sammanhang till Roy’s identitet i konsumentteori.

Den fjärde grundläggande och icke RU artikeln, artikel 6, går tillbaks till det klassiska Choice Axiom av R.D. Luce. Jag modifierar axiomet till en (i mitt tycke) mer naturlig form, som jag kallar Strong Choice Axiom. Denna form möjliggör härledning a en mer generell, lexikografisk version av ovannämnda Luce form för valsannolikheterna. Men till slut, visar en mer detaljerad studie av de två varianterna av axiomen att de är ekvivalenta.

Acknowledgements

This thesis grew out my interest in a paper by Lars-Göran Mattsson and Jörgen Weibull; a paper which developed into paper 2 of this thesis. My work on this paper, lead me to study several fundamental aspects of Probabilistic Discrete Choice, eventually resulting in papers 3-8, and this thesis. I thank Lars-Göran and Jörgen for inviting me on this stimulating journey.

I further thank Lars-Göran, who in fact is my previous student and now professor at the division of Transport and Location Analysis, for letting me pursue doctorate studies in his group with him as my supervisor. Thus, by June 1, we hope to create a cycle in the advisor-student-tree of the mathematics genealogy project, http://genealogy.math.ndsu.nodak.edu

Finally, and foremost, I would again like to thank my wife Christina for putting up with my absentmindedness when I am contemplating on research problems, which can happen any time of the day (or night). And also for her loving support in general.

Stockholm, April 2012

P O Lindberg

List of Papers

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

II. 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

III. Lindberg, P.O., Random Utility Models, IIA, Mattsson-Weibull-Li Distributions and a Conjecture by Luce and Suppes, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.

IV. Lindberg, P.O., Choice Probabilities in Random Utility Models – Necessary and Sufficient Conditions for a Much Used Formula and Implications for the Conditional Distribution of Achieved Utility, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.

V. Lindberg, P.O., A Simple Derivation of the Williams-Daly-Zachery Theorem, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.

VI. Lindberg, P.O., The Choice Axiom Revisited, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.

VII. Lindberg, P.O., Power Invariance in Random Utility Models, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012

VIII. Lindberg, P.O., Random Utility Invariance Revisited, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.

Contents

1. Introduction

1.1 Anecdotic background of the thesis

This is not the thesis I intended to write. I planned to write on *Methods for Computing Traffic Equilibria, and their Applications to Traffic Planning Problems*, of which Mitradjieva and Lindberg (2011) is the first paper. But fate had other plans.

In March of 2010 L.-G. Mattsson (LGM) gave me a copy of an article (together with J Weibull, JW) 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 their main results (later becoming paper 3 in this thesis). 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. At the discussion, they presented a new version of their paper. Again I was very intrigued, and started over the next weeks to suggest additions and improvements (I hope) to the paper. So in the end I was invited to become a coauthor. At about the same time I managed to prove the converse of paper 3.

In the process I started thinking on other questions on *Random Utility* (RU) *Models*, and during the Spring I wrote (the main parts of) papers 3-6, ending in early May, when I got Mitradjieva and Lindberg (2011) back from the refereeing-process, and started accommodating the views of the referees.

Except for paper 3, which I kept trying to improve, and paper 2 with LGM and JW, I did not work on the papers until Mitradjieva and Lindberg (2011) was accepted in November, and I sent in a final version of that paper.

1.2 Random Utility Models

Most papers in this thesis concern aspects of *Random Utility* (RU) Models. RU Models are used to model probabilistic Discrete Choice. In a RU Model, 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 for a random decision-maker to choose alternative *i* from *I* is

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

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

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

More often than not the random utilities have further structure, mostly additive:

$$X_i = v_i + U_i. \qquad (3)$$

In this *Additive* RU (ARU) Model the utility X_i is decomposed into the sum of a deterministic *population value* v_i, assumed known to the analyst (such as the cost of alternative *i*) and an *individual value* U_i, assumed unknown to the analyst, and hence considered as random when studying choices by the choice makers.

RU Models, and in particular ARU Models, have receive wide usage used in 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).

1.3. Short introduction to the papers.

The papers of this thesis concern (Probabilistic) Discrete Choice, DC for short. Of these papers, all but paper 6, concern RU Models. Paper 6 concerns other aspects of DC, namely the Luce's famous Choice Axiom (Luce 1959). Four papers, paper 1, 2, 7 and 8, concern *invariance* in RU Models, i.e. that the distribution of achieved utility is independent of the chosen alternative.

Paper 3 gives necessary and sufficient conditions for a much used formula for the choice probabilities in RU Models to hold.

Paper 4 studies when RU Models give choice probabilities of the Luce form (LFCP) below.

Paper 5 gives a simple derivation of the so called Williams-Daly-Zachary (WDZ) Theorem, which says that the choice probabilities in an ARU Model are the partial derivatives of the expected achieved utility.

Paper 6 goes back to a classical theory by Luce, the *Choice Axiom*, and its consequences.

Paper 7 introduces and characterizes the property of *power invariance* for RU Models

Paper 8, finally, goes back to the characterization of Robertson and Strauss (1981) of invariance in RU Models, and investigates some of its consequences.

2. Theoretical Models

2.1 Origins

RU Models originated in Psychology, where one noted that subjects in test situations could be non-consistent, e.g. concerning which of two tones was the loudest.

Thurstone (1927) suggested a model where a given subject would choose alternative *i* (e.g. tone *i*) with a probability p_i, where

$$p_i =_{df} \Pr\{v_i + X_i > v_j + X_j, j \neq i\}.$$

Here the v_i were the true stimuli, and the X_i were "noise" terms, normally distributed with a common standard deviation σ.

Later, Berkson (e.g. Berkson, 1944) suggested using *logistic* instead of normal distributions for the differences of the noise terms. He termed the model *logit*, as compared to Thurstone's, which was then called *probit*. The logit assumption gave closed form choice probabilities,

$$p_i =_{df} \frac{1}{(1+e^{-(v_i - v_j)/\sigma})}.$$

Both these models concerned *binary* choice.

2.2 Choice Axiom

In 1959 Luce published his influential book *Individual Choice Behavior*, Luce (1959). 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*. To state the axiom, we first need some notation:

$P_T(x) =_{df}$ Pr{choosing *x* from *T*}, supposedly a probability distribution on *T*. For binary choice, one uses the notation

$P(x, y) =_{\mathrm{df}} P_{\{x,y\}}(x)$, i.e. probability of 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.

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

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

2. 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 $x \in T$,

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

2.3 Random Utility Models

Random Utility (RU) models were introduced by Marschak (1960). He introduced *random utilities* X_i for each choice alternative, and postulated

$$P_T(i) =_{df} \Pr\{X_i > X_j, j \neq i, j \in I.$$

He further showed indirectly, that for any Luce-system $\{v_i\}_{i \in I}$, there is a RU model that gives the same choice probabilities. Holman and Marley, in an unpublished paper, gave a constructive proof of this result, using exponential variables on the negative half-axis, as related by Luce and Suppes (1965).

2.4 Multinomial Logit

McFadden (middle 60's, but published later) constructed the *Multinomial Logit* (MNL) Model to describe choice of travel mode. He postulated that the choice probabilities were

$$P_T(i) =_{df} \frac{e^{-v_i}}{\sum_{j \in T} e^{-v_j}}, \qquad \text{(MNL)}$$

where v_i is the "cost" of alt. *i*, linear in its parameters

As is easily seen, this model specializes to the Logit model above for the binary case. He later showed that this model appears if (and also only if, under certain conditions) the random utilities have the form

$$X_i = v_i + U_i, \qquad (3)$$

where now the U_i are i.i.d. random error term with *Gumbel* distributions, i.e. with cdf (cumulative distribution function) $F_U(x) =_{df} \Pr\{U_i \le x\} = \exp(-e^{-x})$. To derive the only if case he assumes that the choice probabilities (MNL) are valid for all values of the v_i, and that one may introduce several independent "copies" of one alternative.

This is an example of an *Additive* RU (ARU) Model. In general, the U_i can have any distribution, independent or not.

For a description of MNL see e.g. McFadden (1974).

2.5 Red Bus Blue Bus and Nested Logit

The Luce model was early criticized for the Red Bus Blue Bus anomaly:

If choosing between a red and a blue bus, gives choice probabilities ½, and one then introduces another blue bus, the probability of choosing red bus drops to 1/3, which is counterintuitive.

The way out of this dilemma was *Nested Logit Models*, where one in the first stage chooses e.g. between all blue buses (or all red buses), and in the next stage chooses between the best red bus and blue bus alternative, and so on.

Nested models introduce dependence between the choice alternatives. An "ultimate" dependent model is the next one.

2.6 Generalized Extreme Value (GEV) Models

In the GEV models, introduced in McFadden (1978), the vector **U** of random errors is supposed to have the cdf

$$F(\mathbf{x}) = \exp(-G(-e^{-\mathbf{x}}))),$$

where *G* is a *linearly homogenous* function, i.e. fulfilling $G(\lambda\mathbf{x}) = \lambda G(\mathbf{x})$ for all $\lambda > 0$.

Examples of *G* giving nested models are e.g.

$$G(\mathbf{x}) = (\sum_{i \in B} x_i^{\rho})^{1/\rho} + (\sum_{i \in C} x_i^{\tau})^{1/\tau}$$

2.7 Invariance in RU Models.

With invariance in an RU Model we mean that the distribution of achieved utility is invariant across chosen alternatives. This property was first studied by Robertson and Strauss (1981), who ”proved” (see 3.2 below) that for an ARU Model to have the invariance property for all values of the v_i, it is necessary that the cdf of the random terms is of the form $F_U(\mathbf{x}) = \varphi(H(e^{-\mathbf{x}}))$, for some is homogenous function *H*.

3. Results and Discussions

3.0. *The papers*

The thesis consists of eight papers. In sort of chronological order:

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
3. Lindberg, P.O., **Random Utility Models, IIA, Mattsson-Weibull-Li Distributions and a Conjecture by Luce and Suppes**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.
4. Lindberg, P.O., **Choice Probabilities in Random Utility Models – Necessary and Sufficient Conditions for a Much Used Formula and Implications for the Conditional Distribution of Achieved Utility**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.
5. Lindberg, P.O., **A Simple Derivation of the Williams-Daly-Zachery Theorem**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.
6. Lindberg, P.O., **The Choice Axiom Revisited**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012.
7. Lindberg, P.O., **A Note on Power Invariance in Random Utility Models**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012
8. Lindberg, P.O., **Random Utility Invariance Revisited**, Working Paper, Dept Transp. Science, KTH Royal Institute of Technology, Stockholm, 2012

I concentrate on the papers where I am single author. For of each paper, I make a short discussion at the end of the summary.

3.1. Paper 1: *Invariance of Achieved Utility in Random Utility Models* (with Eriksson and Mattson)

This is my first paper in the DC area. As mentioned above, Robertson and Strauss (R&S) gave a characterization of cdf’s of the random terms in ARU Models, necessary and sufficient for them to have the invariance property for all values of the population values v_i.

This paper shows that the proof of R&S is incorrect, and sets out to correct it. To this end we use another equivalent characterization of the cdf $F_U(\mathbf{x})$, namely what we termed *RS form*: $F_U(\mathbf{x}) = \rho(-S(-\mathbf{x}))$, where S is *additively homogenous*, i.e. $S(\mathbf{x}+\lambda\mathbf{1}) = \lambda + S(\mathbf{x})$ for all $\lambda \in \Re$, and where $\mathbf{1} = (1,1,\ldots,1)$ is the *one-vector*. The benefit of this representation, is that ρ is the cdf of the maximum utility, whereas S is the expected achieved utility (up to an additive constant), which moreover fulfills the Williams-Daly-Zachary relation, that its partial derivatives $S_i(\mathbf{x})$ are the choice probabilities.

The paper further discusses characterizations of Choice Probability Structures (CPS's) corresponding to distributions of RS form.

As noted in paper 8, this paper is one of the few papers that really comment on the R&S paper. Our paper has received sadly few citations, perhaps due to the choice of journal. But I still believe that our RS form is worth investigating further.

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

(This paper is presented before Paper 2, for reasons of presentation.)

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

Mattsson and Weibull (2010) studied independent X_i, each with a cdf of the form

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

(L for Li (2011) who showed the same)

Then they showed that the corresponding choice probabilities were

$$P_T(i) = \frac{\alpha_i}{\sum_{j\in T}\alpha_j}, \qquad \text{(LF)},$$

i.e. of Luce form.

I prove a converse of this:

<u>If</u> we have an (infinite) set of random variables $\{X_i\}_{i\in G}$, with probability densities, such that for any finite $T \subset G$ there are $\alpha_i > 0$, such that (LF) holds,

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

<u>and further</u> the cdf's of the X_i are well behaved in the sense that the densities of certain powers of the densities do not both agree and disagree arbitrarily close to any point in their supports.

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

The proof in detail studies 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 the proof of Holman and Marley (p.338-339), they go on to note, that the form of the choice probabilities are 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).

I have found only found one reference to the Luce and Suppes conjecture, in Yellott (1977). Yellot mentions that Holman and Marley (H&L) have given a way to construct RU models that replicates any set of choice probabilities of Luce form. However, Yellott states that the H&L RU model uses Gumbel distributed random terms, in contrast to the exponential variables on the negative half-axes, which Luce and Suppes attribute to the example of H&L. This version is also confirmed by Marley (2012). Yellott notes that ARU models, with i.i.d. Gumbel distributed error terms have choice probabilities of Luce form. He goes on to prove that (under uniform expansion of the choice set) for ARU models no other i.i.d. distribution of the error term gives the same choice probabilities for the unexpanded and all expanded choice sets for all population values. Given the introduction, with the discussion of the Luce and Suppes conjecture, one might be lead to believe that solves the conjecture. But it does not, since it pertains to ARU models, and one uses that the population values can be varied freely.

A similar example, which does not hint to solve the L&S conjecture, is McFadden (1974). He proves, using non-uniform expansion of the choice set, that no other i.i.d. distribution for the error terms than Gumbel, gives the same (MNL) choice probabilities for all possible population values.

I find it interesting that the L&S conjecture has been (at least partially) resolved after so long time.

3.3. Paper 2: *Extreme values, invariance and choice probabilities* (with Mattson and Weibull)

This paper generalizes the MWL-class of Paper 3. In particular it studies RU Models whose random terms have cdf's of the form

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

where G is a linearly homogenous function, normalized so that $G(\mathbf{u}_i) = 1$ (where the $\mathbf{u}_i$ are the coordinate unit vectors), bold-face is used for vectors such as $\mathbf{x}$ and $\mathbf{a}$, * stands for component-wise multiplication of vectors, and a univariate function of a vector, such as $\Phi(\mathbf{x})$, is the vector of function-values of the components of the vector.)

The cdf according (MW) has marginal distributions with cdf's $\Phi^{a_i}(x)$. Letting Φ be a Gumbel cdf, (MW) gives GEV-distributions. But we may choose any Φ.

As an application of the independent case, when $G(\mathbf{x}) = \sum_i x_i$, it is in particular shown that Frechet, and Weibull distributed random terms give choice probabilities of Luce form.

The Gumbel, Fréchet and reversed Weibull distribution stand out in relation to cdf's of the form $\Phi^{a_i}(x)$. As shown in the paper, these distributions are characterized by that $\Phi(x)$ is a Gumbel (Fréchet or reversed Weibull) cdf if and only if Φ is defined on $\Re$ and for all *t*, $\Phi(x-t) = \Phi^{\tau}(x)$ (Φ is defined on $\Re$ and for all *t*, $\Phi(x/t) = \Phi^{\tau}(x)$ or Φ is defined on $\Re_-$ and for all *t*, $\Phi(x/t) = \Phi^{\tau}(x)$) for some $\tau > 0$.

For the general (MW) case, it is shown that RU Models with such random terms have the invariance property.

For the independent case, the converse is shown, i.e. that invariance implies (MW) distributions (given that all marginals have the same support).

The (MW) class might allow for more flexibility in modelling discrete choice. The special case of Weibull distributions has e.g. been shown by Fosgerau and Bierlaire (2009) to give better fits in some circumstances.

3.4. 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 papers 1 and 2.

The formula, that gives the choice probability, states

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

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:

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

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 coming opponent), but no one seemed to know under which conditions (CFP) 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 $f(x) =_{df} F(x, x, .., x) = \Pr\{\hat{X} \leq x) = \Pr\{\max_i X_i \leq x)$.

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

$$\sum_i p_i = 1. \qquad (4)$$

Introduce $\bar{p}_i = \Pr\{X_i \geq X_j, j \neq i\}$.

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 RU 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:

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

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

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

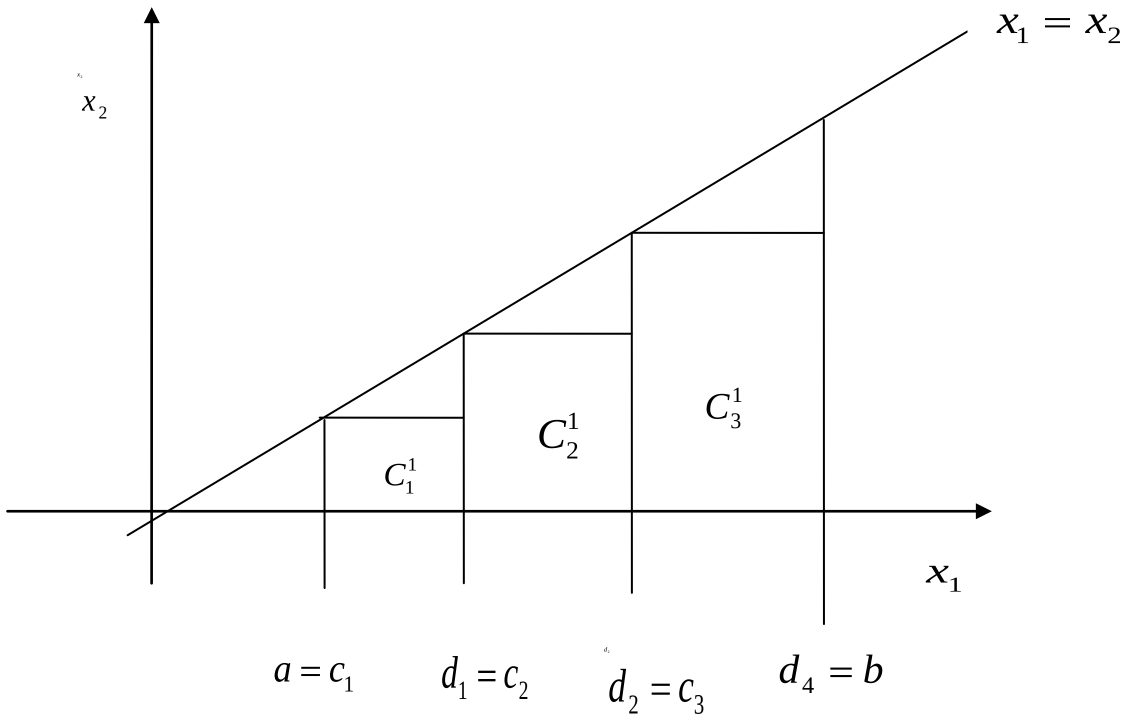
when the derivatives exists. (Here f^+ is the directional derivative of f)

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

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 goal 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 below.



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., and

$$\int_a^b F_i^+(x\mathbf{1})dx = \Pr\{A^i_{(a,b]}\} \qquad \square$$

This 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 the alternative being chosen. Let us term this variable $\hat{X}_i$. The cdf of $\hat{X}_i$ is

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

Using our machinery it follows rather straightforwardly that

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

$p_i = \int_{-\infty}^{x} F_i^{+}(s,s,\ldots,s)ds$

This will be used extensively in later papers.

3.5. 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

$$\overline{P}_i(\mathbf{v}) = \Pr(v_i + X_i \le v_i + X_i, j \ne i), \qquad \text{(CPV)}$$

be the choice probabilities as functions of the v_i.

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

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

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 $\overline{P}_i(\mathbf{v})$.

Assume first $\Delta v_i > 0$, so that alternative *i* becomes more attractive.

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.

The expected changes in achieved utility for individuals in G_1 and G_2 (conditioned on being in that group) are Δv_i and 0 respectively. Let ΔU_c be the corresponding change in G_c.

The probability masses of the groups are respectively p_i, $\Delta p_i \geq 0$ and $1 - p_i - \Delta p_i$.

Summing the effects on $\overline{U}(\mathbf{v})$ in the different groups we have for the change $\Delta\overline{U}$ in $\overline{U}(\mathbf{v})$,

$\Delta\overline{U} = p_i \Delta v_i + \Delta p_i \Delta U_c + 0$. (a)

Further ΔU_c obviously fulfils $0 \leq \Delta U_c \leq \Delta v_i$. (b)

Thus, by (a) and (b),

$p_i \Delta v_i \leq \Delta\overline{U} \leq p_i \Delta v_i + \Delta p_i \Delta v_i = (p_i + \Delta p_i)\Delta v_i$, i.e.

$$p_i \Delta v_i \leq \Delta\overline{U} \leq (p_i + \Delta p_i)\Delta v_i$$

The case $\Delta v_i < 0$ is treated symmetrically, and gives the same result.

As a corollary we can then derive

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

$$\nabla\overline{U}(\mathbf{v}) = P(\mathbf{v}).$$

The paper continues to show

- First, that the $P_i(\mathbf{v})$ are continuous in each variable v_j, if the probability of ties is zero,
- And then, the $P_i(\mathbf{v})$ are continuous in $\mathbf{v}$ if they are continuous in each variable.

The WDZ relations are of interest for their own sake. But they are also of importance to evaluate the societal effects of changes to systems (such as transportation systems), whose utilization are determined by discrete choicer models. as have been demonstrated by e.g. Williams (1977).

3.6 Paper 6: *The Choice Axiom Revisited.*

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

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 $R \subseteq S \subseteq T$,

$$P_T(R) = P_S(R) P_T(S).$$ □

Remark: Note that SCA follows from CA if $P(x,y) \neq 0,1$ for all $x, y \in U$.

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, $P_T(x) = 0$.

Then we introduce a z, such that $T = \{x, y, z\}$.

By SCA, $P_T(x) = P_{\{x,y\}}(x) P_T(\{x,y\}) = 0$

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 $v(x) = v_\alpha(x)$ on C_α :

Theorem (*Luce Form Choice Probabilities*, LFCP). 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 $x \in T$,

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

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 $\overline{\alpha} =_{df} \max(\alpha | T \cap C_\alpha \neq \Phi))$, and with $v_{\overline{\alpha}}$ defined on $C_{\overline{\alpha}}$ according to the Theorem, and extended as in the definition above to all C_β for any $\beta \prec \overline{\alpha}$, and for any $x \in T$,

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

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.

After having derived the ELFCP, I contemplated more on the relation between CA and SCA.

It was rather straightforward to see that SCA implies ELFCP. But it turns out that we can also derive SCA from ELFCP.

3.7. 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, as an attempt to motivate the class of distributions introduced there.

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

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

Where $H : \Re^n_{++} \to \Re_{++}$ 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 is a generalization of the GEV-class. It has some remarkable properties, which it shares with GEV.

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

$\hat{F}^{(i)}(x) = \hat{F}(x) = \{$ 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 MWL class in the first mentioned paper Mattsson and Weibull (2010) had these properties.

There one had an RU model with independent r.v.'s X_i, each with a cdf Φ^{α_i} for $\alpha_i > 0$.

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

$F^{\Phi}(\mathbf{x}) =_{df} C(\mathbf{\Phi}(\mathbf{x})) =_{df} C(\Phi_1(x_1), \Phi_2(x_2), \ldots, \Phi_n(x_n))$ 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 $\overline{C}(\mathbf{x}) =_{df} x_1 \cdot x_2 \cdot \ldots \cdot x_n$.

If we plug in univariate cdf's Φ_i into this copula, we get a distributions with independent r.v.'s and cdf $F(\mathbf{x}) = \Phi_1(x_1) \cdot \Phi_2(x_2) \cdot \ldots \cdot \Phi_n(x_n)$.

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 $\overline{F}_{\alpha}(\mathbf{x}) =_{df} x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_n^{\alpha_n}$ 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 $\overline{F}_{\alpha}(\mathbf{x}) =_{df} C(x_1^{\alpha_1}, x_2^{\alpha_2}, \ldots, x_n^{\alpha_n})$ has the invariance property.

Using results from Lindberg Eriksson and Mattsson (2002) (i.e. 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 $H : \Re_{++}^N \to (0,\infty)$.

Observation: 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 MW class introduced in paper 2.

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

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

I apply them to some old results by Robertson & Strauss (1979), concerning invariance in ARU models,

$X_i = v_i + U_i$. (ARU)

R&S show that if the model (ARU) has invariance for all v_i, then the cdf of **U**, must be of the form

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

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_U indeed a cdf? This is the theme of paper 8.

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

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

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

(i) $(-1)^{(n-1)} H_{i_1 i_2 \ldots i_m,}(\mathbf{x}) \geq 0$ for all $i_1 < i_2 < \ldots < i_m \leq N$, and all $\mathbf{x} \in \Re_{++}^N$, and

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

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 $\phi(0) = 1$ and $\phi(-\infty) = 0$, 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 $\mathbf{x}$, and (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

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

using the homogeneity of H.

Differentiating further, one can see that

$$\overline{F}_{i_1 i_2 \ldots i_n}(\mathbf{x}) = \sum_{k=1}^{n} (-1)^{n-k} \phi^{(k)}(-H) S^n_k(H),$$

Where the $S^n_k(H)$ are polynomials in the derivatives of H.

Further one can show that $S_k^n(H(-\mathbf{x}))$ is homogenous of degree (*k*- *n*). 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 $i_1 < i_2 < ... < i_n, n \le N$

$$(-1)^{n-1} H_{i_1 i_2 ... i_n}(-\mathbf{x}) \ge 0, \text{ for any } \mathbf{x} \in \Re^N_{--}. \qquad \text{(PCHn)}$$

More Conditions on φ

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

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

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:

$$\varphi^{(n)}(x) \ge 0 \text{ for all } x \in S_\phi. \qquad (\text{PC}\varphi)$$

It is then straightforward to see that if φ and *H* fulfil their respective positivity conditions (PCφ) and (PCH*n*) 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 their respective positivity conditions (PCφ) and (PCH*n*) then $\overline{F}_{RS}$ and hence F_{RS} are cdf's.

Special cases of φ

Supp of φ *equals* $(-\infty,0]$.

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

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

Checking gives (after some calculations)

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

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

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

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

Supp of φ equals $(-c,0]$.

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

This φ obviously fulfills $\varphi^{(n)}(x) \geq 0$, and hence gives RS-distributions with any H fulfilling the positivity conditions (PCH<u>n</u>).

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 $\varphi(y) = 1 + y(1 - \ln(-y))$

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

Choice probabilities.

Finally, let us derive choice probabilities for ARU model with RS-distribute error terms. The random utilities X_i in an ARU model

$X_i = v_i + U_i$

with error term **U,** have cdf $F(\mathbf{x}) = F_U(\mathbf{x} - \mathbf{v})$.

Thus if **U** has an RS-distribution,

$F(\mathbf{x}) = \phi(-H(e^{-(\mathbf{x}-\mathbf{v})})$.

In particular

$F_i(\mathbf{x}) = \phi'(-H(e^{-(\mathbf{x}-\mathbf{v})})H_i(e^{-(\mathbf{x}-\mathbf{v})})e^{-(x_i-v_i)}$, and

$F_i(x,x,...,x) = \phi'(-H(e^{-((x,x,...,x)-\mathbf{v})}))H_i(e^{-((x,x,...,x)-\mathbf{v})})e^{-(x-v_i)} = \{$using homogeneity$\}$

$= \phi'(-e^{-x}H(e^{\mathbf{v}}))H_i(e^{\mathbf{v}})e^{-(x-v_i)} = \{$with $a =_{df} H(e^{\mathbf{v}})$ $\} = \phi'(-ae^{-x})H_i(e^{\mathbf{v}})e^{v_i}e^{-x} =$

$=\{$with $a_i =_{df} H_i(e^{\mathbf{v}})e^{v_i}$ $\} = a_i\phi'(-ae^{-x})e^{-x}$, implying that $p_i = ca_i$ for some c. Thus $1 = \sum_i p_i = c\sum_i a_i$,giving $c = (\sum_i a_i)^{-1}$, and $p_i = a_i/(\sum_i a_i) = H_i(e^{\mathbf{v}})e^{v_i}/(\sum_i H_i(e^{\mathbf{v}})e^{v_i}) = \{$using homogeneity$\}$ $H_i(e^{\mathbf{v}})e^{v_i}/H_i(e^{\mathbf{v}})$

4. Conclusions

The eight papers of the thesis have contributed in different ways to fundamental questions in the area of probabilistic discrete choice. Among the results are stringent characterizations of ARU Models with invariance (paper 1), introduction of new classes of distributions with closed form choice probabilities and invariance (paper 2), proof of an old conjecture by Luce and Suppes (paper 3), stringent proofs of necessary and sufficient conditions for a much used formula for choice probabilities (paper 4), a simple derivation of the Williams-Daly-Zachary Theorem (paper 5), a simplifying revision of Luce's classical Choice Axiom (paper 6), introduction and characterization of the class power invariant copulas (paper 7), and finally a deepened investigation of the Robertson and Strauss characterization of invariance in Random Utility Models (paper 8).

5. References

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

Berkson, J. (1944), Application of the Logistic Function to Bio-Assay, *Journal of the American Statistical Association* **39**, 357-365

Fosgerau, M. and M. Bierlaire (2009): "Discrete choice models with multiplicative error terms", *Transportation Research B* **43**, 494-505.

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. (1977), 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 (2002), Homothetic Functions Revisited, *Ec. Theory.* **19** 417-427.

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.

Marley, A.A.J. (2012), Personal Communication.

Marschak, J. (1960), Binary Choice Constraints and Random Utility Indicator, in Arrow, Karlin, Suppes (Eds) Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford, CA.

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

McFadden, D. (1978), Modeling the Choice of Residential Location, in Karlqvist et al. (eds.), *Spatial Interaction Theory and Planning Models,* 75-96, North Holland, Amsterdam.

McFadden, D. (2001), Econometric Choices, Nobel Prize Lecture, American Economic Review, Vol. **91** 351-378

Mitradjieva, M. and P.O. Lindberg (2011), The Stiff is Moving - Conjugate Direction Frank-Wolfe Methods with Applications to Traffic Assignment, accepted for publication in *Transportation Science*

Robertson, C. A. and D. J. Strauss (1981), A Characterization Theorem for Random Utility Variables, *J. Math. Psychology* **23**, 184-189.

Smith, T. E. (1988), A Choice Probability Characterization of Generalized Extreme Value Models, *Applied Mathematics and Computation* **14**, 35-62.

Thurstone, L. L. (1927), A law of comparative judgment. *Psychological Review* **34**, 273-286.

Williams, H.C.W.L. (1977), On the Formulation of Travel Demand Models and Economic Measures of User benefit, *Environment & Planning* A **9**(3), 285-344.

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