Regional Science and Urban Economics 28 (1998) 75–90

An asymptotic foundation for logit models

M. Raouf Jaïbi[a], Thijs ten Raa[b,*]

[a]Département de Mathématiques, Faculté des Sciences de Tunis, Campus Universitaire, 1060 Tunis, Tunisia

[b]Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Received 25 April 1994; received in revised form 22 August 1996; accepted 31 January 1997

Abstract

An agent makes a discrete choice between alternatives. The alternatives are grouped, so that systematic utility differences occur between groups. Within groups, utility levels are random, based on a distribution F. For large groups, the behavior of the agent is shown to be governed by the logit model or its limiting cases. The cases are related to a classification of distributions by decay speed of the upper tail. The connections with the types of extreme value theory and elements of spatial economics are discussed.

Keywords: Logit model; Discrete choice

JEL classification: C25; R22; R10

1. Introduction

A common approach in spatial economics is the following. The probability of choosing a destination $i \in I = \{1, \ldots, m\}$ (a shopping center, say), is proportional to its attractivity, A_i (the number of shops at i, say), but decays with the distance to i, or, more precisely, the associated (travel) cost, c_i, according to

$$p_i = A_i \, e^{-\mu c_i} \Big/ \sum_{K \in I} A_k \, e^{-\mu c_k}$$

*Corresponding author. Tel.: (31-13) 4662365; fax: (31-13) 4663280; e-mail: tenraa@kub.nl

PII S0166-0462(97)00015-X

where μ is a nonnegative parameter trading off the two effects. (μ is estimated econometrically.) The formula is the so called logit model, modified by attractivities A_i. The purpose of this paper is to understand it in the context of net surplus or, for short, utility maximization. Utilities are indirect, with prices given.

It is well known that if the utilities of alternatives i are

$$u_i = \tilde{u}_i - c_i, \tilde{u}_i \sim F$$

with F the double exponential or so called Gumbel distribution (Resnick, 1987), common to all i, the above choice probabilities can be derived. This result is special. It is not difficult to prove that any other utility distribution would yield a different formula for the choice probabilities (Yellot, 1977). Yet the logit model is popular with empirical researchers and theorists continue to come to grips with it. Spatial economic theory contains two elements underpinning the above logit model. First, if there is no spatial differentiation (in preferences, technologies and endowments of an economy), then the Spatial Impossibility Theorem (Starrett, 1978) says that no competitive equilibrium has positive total transport cost. In the context of urban land use, this means that agents concentrate choices in one location or disperse them uniformly (Fujita, 1985, p. 138). A recent description of the phenomenon is in (Ciccone and Hall, 1996, p. 59): ‘‘How can states or countries be in equilibrium with different densities?’’ They proceed to argue that under neoclassical assumptions density should be equal everywhere or concentrated in a single county, depending on their returns to scale parameter. These cases correspond to the limiting cases of the logit model, $\mu = 0$ and $\mu = \infty$, respectively. Second, if there is spatial differentiation, distance is thought to effect choices in exponential decay fashion (Tomlin and Tomlin, 1968; Domencich and McFadden, 1975). This case corresponds to the proper logit model, with μ positive and finite. This paper will formulate and prove the two elements in a concise and general setting. Error distribution F will have no specific functional form and we will proceed to accommodate multiple observations in the framework described above.

Our analysis contributes to discrete choice theory (Train, 1980; Ben-Akiva and Lerman, 1985; Anderson et al., 1992, and references given there) by offering an asymptotic foundation to the logit model. It shows that the logit model emerges when the sample size becomes large, much as the Normal distribution emerges when normalized sums are taken over large numbers of terms. In a way one could say that our theorem can be used as a counterpart to a well known usage of the central limit theorem, signifying the logit model in the context of discrete choice rather than the normal distribution in the admittedly more widely applicable context of summation.

Section 2 presents the model. A decision maker observes different times within alternatives and then chooses the alternative where a utility realization is maximum. Not surprisingly, qualitative behavior of the upper tail of the utility distribution matters. Section 3 classifies distributions by the nature of the upper tail. Section 4 presents the main result: discrete choice is shown to be governed by the logit model or its two limiting cases when the numbers of observations are large. The pertinent classification of utility distributions by tail nature is new and, surprisingly, does not agree with the classification of extreme value theory. This is explained in Section 5. We relate to the applied literature in Section 6 and conclude with Section 7.

2. The model

Outcomes are 0, 1, . . . , m. Subset $\mathrm{I}=\{1, \dots, m\}$ is the set of alternatives, such as locations. 0 is reserved for the case of no choice by absence of information. It accommodates zero purchases, for example when there is some threshold opportunity value as discussed by Perloff and Salop (1985). If x_i observations are made in $i \in I$, net utility of alternative i is, assuming rational choice,

$$u_i = \max_{1 \leq j \leq x_i} u_{ij} - c_i$$

where u_{ij} are random utility values at i and c_i is a systematic effect associated with i, such as search or travel cost. By convention, $u_i = -\infty$ if $x_i = 0$. We assume that the random terms are drawn independently from cumulative distribution function (c.d.f.) F, which is common to alternatives $i \in I$.

Assumption 1: $\{u_{ij} | i \in I \text{ and } j \in \mathbb{N}\}$ is a family of independent and identically distributed random variables with c.d.f. F.

Assumption 2: C.d.f. F is continuous and has regular upper tail in the sense of Definition 2 of Section 3.

The assumption of continuity can be dispensed. It spares the trouble of making choice multi-valued when alternatives have equal net utility with positive probability, and letting these complications wash out in the asymptotic analysis, and it spares a separate treatment of distributions with bounded support. The assumption of regularity is a technical requirement. It excludes pathological distributions.

For any numbers of observations, $x = (x_1, \dots, x_m)$, the probabilities of choosing alternatives i and of no choice, are defined by

$$P_i(x) = \mathbb{P}\{u_i > u_k, \text{ all } k \neq i\} \text{ and } P_0(x) = 0$$

if $x \neq 0$, and by the exceptional case,

$$P_i(0) = 0 \text{ and } P_0(0) = 1.$$

Under Assumptions 1 and 2, $\mathbb{P}\{u_i = u_k > -\infty$, some $k \neq i\} = 0$, and, therefore, $\Sigma_{i=0}^{m} P_i(x) = 1$. In fact,

Lemma 1: Under Assumptions 1 and 2, for $i \in I$,

$$P_i(x) = x_i \int_u F(u)^{x_i - 1} \prod_{\substack{k \neq i \\ x_k > 0}} F(u + c_k - c_i)^{x_k} \, \mathrm{d}F(u).$$

Proof: See Appendix A.

Numbers of observations are now assumed to become larger and larger and to approach certain proportions.

Assumption 3: $\{x_i^n | i \in I$ and $n \in \mathbb{N}\}$ is a family of independent random integers with finite expectations, independent of $\{u_{ij} | i \in I$ and $j \in \mathbb{N}\}$ and such that for every i,

$$\lim_{n \to \infty} \mathbb{E}(x_i^n) = \infty,$$

$$\lim_{n \to \infty} \frac{\mathbb{E}(x_i^n)}{\sum_{k=1}^{m} \mathbb{E}(x_k^n)} = A_i \in [0, 1],$$

$$\frac{x_i^n}{\mathbb{E}(x_i^n)} \to 1 \text{ in probability as } n \to \infty.$$

A sufficient condition for the third part of Assumption 3 is in Appendix A. The net utility of alternative i is now

$$u_i^n = \max_{1 \leq j \leq x_i^n} u_{ij} - c_i,$$

and the probabilities of choosing alternatives i and of no choice are now defined by

$$P_i^n = \mathbb{P}\{u_i^n > u_k^n, \text{ all } k \neq i\},$$

$$P_0^n = \mathbb{P}\{x^n = 0\}.$$

Under Assumptions 1 and 2, $\Sigma_{i=0}^{m} P_i^n = 1$. In fact,

Lemma 2: Under Assumptions 1, 2 and 3, for $i \in I$,

$$P_i^n = \mathbb{E}[P_i(x^n)]$$

where $P_i(x^n)$ is given by Lemma 1.

Proof: See Appendix A.

The purpose of the paper is the determination of the *limiting choice probabilities,*

Definition 1: $p_i = \lim_{n\to\infty} P_i^n$, $i = 0,1,\ldots,m$. Observations contain all relevant information and are exogenous to the individual, with the distribution being a frequency device to the observer. The limit of infinite observation size is merely a mathematical approximation to the notion of a large, yet *fixed* vector of observations. It is for this reason that our model of discrete choice constitutes no more than an asymptotic theory.

3. Tails of distributions

Since the numbers of observations go to infinity in probability, only the upper tail of the utility distribution matters. We shall classify distributions by decay speed of the upper tail: slow, exponential or fast.

Definition 2: C.d.f. F has *regular upper tail,* if

$$\varphi(c) = \lim_{u\uparrow sup\{u|F(u)<1\}} \frac{1-F(u+c)}{1-F(u)}$$

is well defined for $c \geq 0$.

Lemma 3 will prove that φ must essentially be an exponential decay function, $\varphi(c)=\exp(-\mu c)$, including the limiting cases of no decay ($\mu=0$), defined by $\varphi(c)=1$, and of sudden decay ($\mu=\infty$), defined by $\varphi(0)=1$ and $\varphi(c)=0$ ($c>0$). With this convention,

Lemma 3: If $\varphi(c)$ is well defined for $c\geq 0$, $\varphi(c)=\exp(-\mu c)$ with μ zero, positive or infinite. If μ is zero or positive, $\varphi(c)$ is also well defined for $c<0$ and the equality holds.

Proof: See Appendix A.

Lemma 4 will prove that μ must essentially be the limit of the hazard rate of F. The hazard rate at u is the probability density of being (only) u, conditional on being at least u:

Definition 3: The *hazard rate* at u of c.d.f. F with support unbounded above and probability density f, is defined by

$$\rho(u) = \frac{f(u)}{1 - F(u)}.$$

Lemma 4: For a c.d.f. F of Definition 3,

$$\frac{1 - F(u + c)}{1 - F(u)} = \exp\left[-\int_{u}^{u+c} \rho(v)\, \mathrm{d}v \right],$$

and if ρ has limit μ, be it zero, positive or infinite, then it fulfils

$$\varphi(c) = \exp(-\mu c)$$

where the left-hand side is given by Definition 2.

Proof: See Appendix A.

We use μ, be it the limit of the hazard rate or, if the latter does not exist, the element on the right-hand side of Lemma 3 with the left-hand side given by Definition 2, to classify p.d.f.'s with regular upper tails.

Definition 4: A c.d.f. with regular upper tail features *slow*, *exponential* or *fast* decay (of the upper tail), if μ of Lemma 3, is zero, positive or infinite, respectively.

To classify a c.d.f. according to Definition 4, Lemma 4 is practical. A stronger sufficient condition is the following. If $\log(1-F)$ is convex (concave) near infinity, then F features slow or exponential (fast or exponential) decay. A weaker condition is the following. F with support unbounded above (see otherwise Example 2 below), features exponential (slow) decay, if and only if $1-F\circ\log$ is regularly (slowly) varying at infinity. A definition and representation of regular (slow) variation are given in, for example, (Feller, 1971, p. 282), but verification does not seem to simplify a direct check of Definitions 2 and 4.

3.1. Examples

(1) For β and λ positive, the Weibull distribution is defined by $F_\beta(u)=1-\exp(-\lambda^\beta u^\beta)$ on $u\geq 0$ and zero elsewhere (Galambos, 1987). For any β, it has finite moments of all orders. F_β has slow, exponential or fast decay, if β is less than, equal to, or greater than one, respectively.

(2) The Log-normal, Cauchy and Pareto distributions have slow decay. The exponential, Gumbel and Gamma distributions have exponential decay. The Normal distribution and distributions with support bounded above have fast decay.

(3) The distribution defined by $F(u)=1-\exp(-2u-\sin u)$ on $u\geq 0$ and zero elsewhere, has decreasing density. Yet, it has no regular upper tail.

4. Limiting choice probabilities

Limiting choice probabilities are determined by type of decay of the upper tail of the utility distribution. When decay is fast, choice is based on the systematic parts of the utilities of the alternatives, yielding the choice $I_{\min} = \{i \in I | c_i = \min_{k\in I} c_k\}$. When decay is exponential, there is a trade-off with uncertainty, according to the logit model. When decay is slow, uncertainty governs choice, according to the relative sample sizes.

Theorem: *Under Assumptions 1–3, referring to Definition 1, $p_0 = 0$, and, if the decay of the upper tail of utility distribution F is*

1. *slow, then $p_i = A_i$,*
2. *exponential, then $p_i = A_i\, e^{-\mu c_i} / \Sigma_{k\in I} A_k\, e^{-\mu c_k}$,*
3. *fast, and $\Sigma_{K\in I_{\min}} A_k > 0$, then $p_i = A_i / \Sigma_{k\in I_{\min}} A_k$ for $i \in I_{\min}$ and zero otherwise.*

Proof: See Appendix A.

Remark: If F features fast decay and $\Sigma_{k\in I_{\min}} A_k = 0$, examples show that $\Sigma_{i\in I_{\min}} p_i$ may be one or zero. If the fast decay is by virtue of bounded support, an ad hoc proof shows that $p_i = \lim_{n\to\infty} [(\mathbb{E}(x_i^n)/(\Sigma_{k\in I_{\min}} \mathbb{E}(x_k^n))]$ for $i \in I_{\min}$ and zero otherwise.

The proof in Appendix A is quite technical. We offer an alternative, heuristic derivation to provide some intuition. Let the utility maximizer attain net utility level u. Define p_i^u as the conditional probability that this level is attained in alternative i, then

$$p_i^u = \mathbb{P}\{u_i > u | \max_k u_k > u\} = \mathbb{P}\{u_i > u\} / \mathbb{P}\{\max_k u_k > u\}$$
$$= \mathbb{P}\{\max_{1\le j\le x_i} u_{ij} - c_i > u\} / \mathbb{P}\{\max_{\substack{1\le j\le x_k\\ 1\le k\le m}} u_{kj} - c_k > u\} = \frac{1 - F(u + c_i)^{x_i}}{1 - \prod_k F(u + c_k)^{x_k}}.$$

We emphasize that p_i^u is assumed to be a proxy for p_i, at least for large values of u. This is the real heuristic part of our argument. Due to the sampling, u will be close to its supremum value, hence $F(u + c_k)$ close to one. Hence we have the first-order approximations,

$$F(u + c_k)^{x_k} = \{1 - [1 - F(u + c_k)]\}^{x_k} \simeq 1 - x_k[1 - F(u + c_k)]$$

and

$$\prod_k F(u+c_k)^{x_k} \simeq 1 - \sum_k x_k[1 - F(u+c_k)].$$

Substituting,

$$p_i^u \simeq \frac{x_i[1-F(u+c_i)]}{\sum_k x_k[1-F(u+c_k)]}$$
$$= \frac{x_i}{\sum_j x_j} \frac{1-F(u+c_i)}{1-F(u)} \Big/ \sum_k \frac{x_k}{\sum_j x_j} \frac{1-F(u+c_k)}{1-F(u)}$$

Now let the numbers of observations (x) go to infinity with proportions A_i. Then u may be taken to tend to its supremum value $b = \sup\{u : F(u) < 1\}$ and the expression for p_i^u tends to

$$A_i \lim_{u \uparrow b} \frac{1-F(u+c_i)}{1-F(u)} \Big/ \sum_k A_k \lim_{u \uparrow b} \frac{1-F(u+c_k)}{1-F(u)}$$

By elementary Lemma 3, $\lim_{u \uparrow b} [(1-F(u+c_i))/(1-F(u))] = \exp(-\mu c_i)$, so that the logit model is derived heuristically.

5. Relationship with extreme value theory

Since in this paper we study the maximum of a sample of random utilities it is natural to think of extreme value theory (Galambos and Kotz, 1978). The subject of that theory is the distribution of the maximum of a large number of independent random variables. Since the maximum goes off to the upper bound of the support of the distribution, usually infinity, some normalization must be involved. The basic result is that fur large numbers of observations, the distribution of the normalized maximum is one of three types. (The first type is the Gumbel distribution.) Which case applies, is a matter of the shape of the underlying error distribution. Distributions which belong to the so called domain of attraction I yield a normalized maximum that is distributed according to the first type (Gumbel). Similarly, there are domains of attraction II and III. The domains of attraction are detailed in Feller (1971); Galambos (1987). The relationship with our classification of distributions, featuring exponential decay (E), slow decay (S), or fast decay (F), is as follows. Domain of attraction II is contained in class S. Domain of attraction III is contained in class F. Domain of attraction I intersects with all our classes. Domain of attraction I members Gumbel, Log-normal and Normal fall in classes E, S and F, respectively. That everything can happen with distributions of the domain of attraction I comes as a surprise. It has been an article of faith that frequent observation and maximization with such distributions yield approximately Gumbel distributed maxima in the alternatives and hence the logit model. An early presentation is by Cochrane (1975). (Leonardi, 1984, p. 120) seems to have been aware of the fact that not all members of domain of attraction I yield the logit model, but his analysis is in a restricted context and imprecise. Like Cochrane, he confines the analysis to distributions in the domain of attraction of the Gumbel distribution. Moreover, the mathematical errors were not corrected.

6. Implications for applied work

The choice probabilities introduced in Section 1 are basically logit, but feature two ingredients of economic geography: attractivities A_i and distance deterrent μ. These parameters offer convenient room for estimation and a good fit. However, since the economic geographic notions have not been established in an economic framework, they are typically absent from more rigorous studies. The main relevance of our result to applications is that is reconciles the hitherto considered ad hoc economic geographic gadgets with the maximum utility framework of the logit model.

Attractivities and distance deterrents are perfectly legitimate parameters to modify the basic logit model when a better fit is wanted.

The origin of the attractivities throws light on what are perhaps the two main shortcomings of the logit model: the requirement that alternatives are distinct and the property of independence of irrelevant alternatives. We shall take up these issues in turn. The classification of choice outcomes in a number of discrete alternatives plagues the practitioner. Following (Witlox, 1994, p. 23), the two classical examples to illustrate this problem are Debreu's (Debreu, 1960, pp. 186–188) case of the recordings of the same concerto with a live performance and McFadden's (McFadden, 1974, p. 113) case of the red and blue buses. The difficulty is that logit probabilities are influenced by the classification. The probability of taking the bus in one application is not the sum of the probabilities of taking the red or the blue bus in the other one. This problem emerges in the basic logit model because it does not account for relative numbers of observations. In this paper, attractivities are shown to be equal to the (limiting) shares of observations. Hence attractivities are influenced directly by the classification of choice outcomes. The incorporation of this effect in the specification of the model nullifies the bizarre effect of classification.

The property of independence of irrelevant alternatives states that the odds of two alternatives is independent of the number and characteristics of the other alternatives. Indeed, in the basic logit model, $p_i = e^{-\mu c_i}/\sum_{k\in I} e^{-\mu c_k}$, we have $p_i/p_j = e^{\mu(c_j - c_i)}$, independent of the other c_k and the numbers of the other alternatives. As (Witlox, 1994, p. 22) puts it, this means that the effect of similarities among alternatives is ignored. In the full model, however, the odds are multiplied by A_i/A_j, the size odds of the two alternatives relative to the entire sample, including the other alternatives. This ratio is affected by the introduction of a new alternative, if only because the A_i's must be adjusted so that they continue to sum to unity.

In short, the simultaneous accounting of numbers of observations with the classification of the alternatives takes care of the shortcomings of the basic logit model. Another way of appreciating the relevance of our theorem is by considering it as an aggregation result. The essence of the theorem is that if numerous, independent alternatives are grouped by level of systematic utility or cost, then the probability that a utility maximizer picks from a particular group is given by the generalized logit model of the theorem. The condition that must be fulfilled for this aggregation to hold, pertains to the classification in groups, but not to the error distribution. The error distribution may be any, but groups must be large. We therefore provide theoretical support to the use of the logit model for the determination of probabilities of aggregated events.

7. Conclusion

When alternatives have different systematic utilities, but commonly distributed random terms, the behavior of an agent who observes sufficiently often and maximizes utility, is one of three types. Either the agent concentrates choice in the alternatives which maximize the systematic part of utility, or (s)he chooses according to the logit model, or (s)he randomizes decision making completely. The type is determined by the thickness of the tail of the utility distribution, as measured by the rate of decay.

Random utility with, however, a thin tail (featuring fast decay), yields systematic utility maximization. A thick tail (featuring slow decay), yields behavior randomized by the relative sample sizes. Intermediate tails (featuring exponential decay), yield a trade-off between systematic and random utility maximization according to the logit model.

The first element of spatial economics, by which agents concentrate choice in one location or disperse it uniformly in the absence of spatial differentiation, is established if randomness in the tail is nearly absent (thin tail) or overwhelming (thick tail), respectively. The only case which remains (intermediate tail), yields decay of choice in exponential fashion, ascertaining the second element of spatial economics. Since concentration and complete dispersal of choice can be considered limiting cases of the logit model, the latter has been given an asymptotic foundation.

Acknowledgements

Simon Anderson, Masa Fujita, Laurens de Haan, Tony Smith and Casper de Vries kindly made helpful remarks. Hospitality at Université du Québec à

Montréal is gratefully acknowledged. The research was performed at Tilburg University with support provided by Fellowships of the Economics Research Foundation, the Netherlands Organization for Scientific Research, and of the Royal Netherlands Academy of Arts and Sciences, respectively. We wish to acknowledge intellectual debt to two spatial authors. Wilson (1970) determines (shopping) trip patterns by entropy maximization, without appeal to economic behavior. Smith (1978) disaggregates departures and arrivals data by the so called cost-efficiency principle (the assumption that trip patterns with low aggregated costs are more probable), without explaining the totals. It is interesting to note that both authors generate the exponential decay effect. These partial results stimulated us to pursue our inquiry.

Appendix A

Proof of Lemma 1: By Assumptions 1 and 2 and convention $u_k = -\infty$ for $x_k = 0$,

$$P_i(x) = \mathbb{P}\{u_i > u_k, \text{ all } k \neq i\} = \int_u \prod_{k \neq i} \mathbb{P}\{u_k < u\} \, \mathrm{d}\mathbb{P}\{u_i < u\}$$

$$= \int_u \prod_{\substack{k \neq i \\ x_k > 0}} \mathbb{P}\{u_k < u\} \, \mathrm{d}\mathbb{P}\{u_i < u\}$$

For $x_i > 0$, by Assumption 1,

$$\mathbb{P}\{u_i < u\} = \mathbb{P}\{\max_{1 \leq j \leq x_i} u_{ij} < u + c_i\} = \prod_{j=1}^{x_i} \mathbb{P}\{u_{ij} < u + c_i\} = F(u + c_i)^{x_i}$$

Substitution and change of variables, $v = u + c_i$, complete the proof for $x_i \neq 0$. Otherwise $P_i(x) = 0$, and the right-hand side is zero by probability theoretic convention $0.\infty = 0$. Q.E.D.

Remark on Assumption 3: A sufficient condition for the third part is

$$\lim_{n \to \infty} \frac{\mathbb{E}|x_i^n - \mathbb{E}(x_i^n)|}{\mathbb{E}(x_i^n)} = 0,$$

or, a fortiori,

$$\lim_{n \to \infty} \frac{Var(x_i^n)}{[\mathbb{E}(x_i^n)]^2} = 0.$$

Proof of Lemma 2: Continuing the proof of Lemma 1,

$$P_i^n = \int_u \prod_{k \neq i} \mathbb{P}\{u_k^n < u\} \, \mathrm{d}\mathbb{P}\{u_i^n < u\}$$

with, using independence Assumptions (3),

$$\mathbb{P}\{u_i^n < u\} = \sum_{p=0}^{\infty} \mathbb{P}\{x_i^n = p\}[F(u + c_i)]^p = \mathbb{E}[F(u + c_i)^{x_i^n}].$$

Consequently,

$$\mathrm{d}\mathbb{P}\{u_i^n < u\} = \mathbb{E}[x_i^n F(u + c_i)^{x_i^n - 1}] \, \mathrm{d}\mathbb{F}(u + c_i).$$

Substitution, independence of x_k^n $(k \neq i)$ and x_i^n (Assumption 3), Fubini's theorem, and change of variables complete the proof. Q.E.D.

Proof of Lemma 3: For c and $d \leq 0$,

$$\frac{1 - F(u + c + d)}{1 - F(u)} = \frac{1 - F(u + c)}{1 - F(u)} \cdot \frac{1 - F(u + c + d)}{1 + F(u + c)}.$$

Taking limits,

$$\varphi(c + d) = \varphi(c)\varphi(d).$$

Since φ is monotonic, the functional equation is known to characterize the exponential function (Feller, 1971). Consequently,

$$\varphi(c) = A \exp(-\mu c)$$

with $A = \varphi(0) = 1$ and $\mu = -\log \ \varphi(1)$. Since $0 \leq \varphi(1) \leq 1$, μ must be zero, positive, or infinite.If μ is zero or positive, for $c < 0$,

$$\begin{aligned}\varphi(c) &= \lim_{u \to \infty} \frac{1 - F(u + c)}{1 - F(u)} = \lim_{u \to \infty} \left[\frac{1 - F(u)}{1 - F(u + c)} \right]^{-1} \\ &= \lim_{u \to \infty} \left[\frac{1 - F(u - c)}{1 - F(u)} \right]^{-1} = \varphi(-c)^{-1} = [\exp(\mu c)]^{-1} = \exp(-\mu c).\end{aligned}$$

Q.E.D.

Proof of Lemma 4: Differentiating,

$$\frac{\mathrm{d}}{\mathrm{d}u} \log(1 - F)(u) = -\rho(u).$$

Integrating back between u and $u + c$ and taking exponents, we obtain the first part of the lemma. If ρ has limit μ, the second part is a corollary. Q.E.D.

Proof of Theorem: The strategy of proof is to minorate $\liminf_{n\to\infty} P_n^i$ by $P_i^\infty \geq 0$, adding up to unity. Since P_i^n add up for all n, they must tend to P_i^∞. (Otherwise some subsequence $P_{i_0}^{n_k}$ would go to $P_{i_0}^+ > P_i^\infty$. But then $1 = \lim_{k\to\infty} \Sigma_{i=0}^m P_i^{n_k} \geq \Sigma_{i=0}^m \liminf_{k\to\infty} P_i^{n_k} \geq \Sigma_{i\neq i_0} P_i^\infty + P_{i_0}^+ > \Sigma_{i=0}^m P_i^\infty = 1$.) We shall show that P_i^∞ add up over certain subsets, namely $I_{\min}$ in case of fast decay, and I otherwise. Since this will force the other limiting probabilities to zero, we may limit attention to $i \in I_{\min}$ for fast decay, and to $i \in I$ otherwise. We will minorate $P_i(x)$ of Lemma 1 in terms of $\varphi(c_k - c_i)$ of Lemma 3. If decay is fast, we limit attention to $i \in I_{\min}$, so that $c_k - c_i$ is nonnegative and φ applies. Otherwise, $i \in I$ and $c_k - c_i$ may be negative, which is in agreement with Lemma 3. With this limitation on i in mind, we may proceed.

Fact 1: Under Assumptions 1 and 2, for $x \neq 0$ and any $\varepsilon \in (0,1)$, there is a B_ε with $F(B_\varepsilon) < 1$ and, referring to Lemma 1,

$$P_i(x) \geq \frac{C_\varepsilon^{-1} x_i}{\sum_{k=1}^m x_k \varphi_k} (1 - \exp\{-C_\varepsilon[1 - F(B_\varepsilon)]x_i\})$$

where $C_\varepsilon = -\varepsilon^{-1} \log(1-\varepsilon)$ and $\varphi_k = \varphi(c_k - c_i) + \varepsilon$.

Proof of Fact 1: Define $\bar{\varphi} = \max_{k\in I}(c_k - c_i)$. Then $\bar{\varphi} \geq 1$. By Lemma 3 (substituting Definition 2), for any $\varepsilon \in (0,1)$, there is B_ε close to $\sup\{u | F(u) < 1\}$ in the sense that $\frac{\bar{\varphi}}{\bar{\varphi} + \varepsilon} \leq F(B_\varepsilon) < 1$ and $k \in I$,

$$\begin{aligned} 0 &\leq 1 - F(u + c_k - c_i) \leq v_k = [\varphi(c_k - c_i) + \varepsilon] \cdot [1 - F(u)] \\ &\leq (\bar{\varphi} + \varepsilon)[1 - F(B_\varepsilon)] \leq (\bar{\varphi} + \varepsilon)\left(1 - \frac{\bar{\varphi}}{\bar{\varphi} + \varepsilon}\right) = \varepsilon. \end{aligned}$$

By concavity of $\log(1-v)$, hence decreasingness of $\frac{\log(1-v)}{v}$, both in $v \in (0,1)$,

$$\frac{\log(1 - v_k)}{v_k} \geq \frac{\log(1-\varepsilon)}{\varepsilon} = -C_\varepsilon \text{ or } \log(1 - v_k) \geq -C_\varepsilon v_k$$

and

$$\begin{aligned} F(u + c_k - c_i)^{x_k} &= \exp\{x_k \log[1 - [1 - F(u + c_k - c_i)]]\} \\ &\geq \exp[x_k \log(1 - v_k)] \geq \exp(-C_\varepsilon x_k v_k). \end{aligned}$$

In particular, for $u \geq B_\varepsilon$,

$$F(u)^{x_i - 1} \geq F(u)^{x_i} \geq \exp(-C_\varepsilon x_i v_i).$$

By Lemma 1, substituting back $v_k = \varphi_k[1 - F(u)]$ and $v_i = \varphi_i[1 - F(u)]$, for $x \neq 0$,

$$P_i(x) \geq x_i \int_{B_\varepsilon}^{\infty} \exp\left\{-C_\varepsilon \sum_{k=1}^{m} x_k \varphi_k [1-F(u)]\right\} \mathrm{d}F(u)$$
$$= x_i \left(C_\varepsilon \sum_{k=1}^{m} x_k \varphi_k\right)^{-1} \left(1 - \exp\left\{-C_\varepsilon \sum_{k=1}^{m} x_k \varphi_k [1-F(B_\varepsilon)]\right\}\right).$$

Since $\Sigma_{k=1}^{m} x_k \varphi_k \geq x_i$, Fact 1 is obtained.

Fact 2: Under Assumption 3,

$$\lim_{n\to\infty} \mathbb{E}\left(\frac{x_i^n}{\sum_{k=1}^{m} x_k^n \varphi_k}\right) = A_i \Big/ \sum_{k=1}^{m} A_k \varphi_k \text{ and } \lim_{n\to\infty} \mathbb{E}[\exp\{-C_\varepsilon[1-F(B_\varepsilon)]x_i^n\}] = 0.$$

Proof of Fact 2: Define $y_i^n = x_i^n / \mathbb{E}(x_i^n)$ and (nonrandom) $z_i^n = \mathbb{E}(x_i^n)/\Sigma_{k=1}^{m} \mathbb{E}(x_k^n)$. By Assumption 3, $(y^n, z^n) = (y_1^n, \ldots, y_m^n, z_1^n, \ldots, z_m^n) \to (1, \ldots, 1, A_1, \ldots, A_m) = (1, A)$ in probability as $n\to\infty$. Because $x_i^n/\Sigma_{k=1}^{m} x_k^n \varphi_k = y_i^n z_i^n / \Sigma_{k=1}^{m} y_k^n z_k^n \varphi_k = h_i(y^n, z^n)$ is the value under a continuous function; they tend to $h_i(1, A) = A_i/\Sigma_{k=1}^{m} A_k \varphi_k$ in probability as $n\to\infty$. Moreover, because $x_i^n/\Sigma_{k=1}^{m} x_k^n \varphi_k \leq x_i^n / x_i^n \varphi_i = 1/(1+\varepsilon) < 1$, $h_i(y^n, z^n)$ are uniformly integrable in n, implying convergence of the means (Billingsley, 1968, p. 31). Similarly, $x_i^n \to \infty$ in probability as $n\to\infty$. Because $C_\varepsilon[1-F(B_\varepsilon)]$ is positive (Fact 1), $\exp\{-C_\varepsilon[1-F(B_\varepsilon)]x_i^n\}$ tend to zero in probability and are bounded by unity, implying convergence of the means. The proof of Fact 2 is complete.

Now we prove the theorem. By Lemma 2 and Fact 1,

$$P_i^n = \mathbb{E}[P_i(x^n) \geq C_\varepsilon^{-1} \mathbb{E}\left(\frac{x_i^n}{\sum_{k=1}^{m} x_k^n \varphi_k}(1 - \exp\{-C_\varepsilon[1-F(B_\varepsilon)]x_i^n\})\right).$$

Since $x_i^n/\Sigma_{k=1}^{m} x_k^n \varphi_k \leq 1/\varphi_i < 1$,

$$P_i^n \geq C_\varepsilon^{-1} \mathbb{E}\left(\frac{x_i^n}{\sum_{k=1}^{m} x_k^n \varphi_k}\right) - C_\varepsilon^{-1} \mathbb{E}[\exp\{-C_\varepsilon[1-F(B_\varepsilon)]x_i^n\}].$$

Now let $n\to\infty$. Apply Fact 2 and the definition of φ_k (Fact 1) to the right-hand side of the just derived inequality. Then

$$\liminf_{n\to\infty} P_i^n \geq C_\varepsilon^{-1} A_i \Big/ \sum_{k=1}^{m} A_k [\varphi(c_k - c_i) + \varepsilon].$$

Since this is true for any $\varepsilon \in (0,1)$ and $C_\varepsilon \to 1$ for $\varepsilon \downarrow 0$, it must be that

$$\liminf_{n\to\infty} P_i^n \geq A_i \Big/ \sum_{k=1}^{m} A_k \varphi(c_k - c_i) = P_i^\infty.$$

By Definition 4 and Lemma 3, if decay is

1. slow, $P_i^\infty = A_i / \Sigma_{k=1}^m A_k$ and add up over I,
2. exponential, $P_i^\infty = A_i \exp(-\mu c_i) / \Sigma_{k=1}^m A_k \exp(-\mu c_k)$ and add up over I,
3. fast, $P_i^\infty = A_i / \sum_{k \in I_{\min}} A_k$ and add up over $I_{\min}$, provided that $\sum_{k \in I_{\min}} A_k > 0$.

By the introduction of the proof, these are the limiting choice probabilities for the indicated sets, and the remaining must be zero. Q.E.D.

References

Anderson, S.P., de Palma, A., Thisse, J.-F., 1992. Discrete Choice Theory of Product Differentiation, MIT Press.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis–Theory and Application to Travel Demand. MIT Press Transportation Studies.

Billingsley, P., 1968. Convergence of Probability Measures. John Wiley and Sons, New York.

Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. American Economic Review 86, 54–70.

Cochrane, R.A., January 1975. A possible economic basis for the gravity model. Journal of Transport Economics and Policy, 34–49.

Debreu, G., 1960. Review of R.D. Luce, Individual choice behavior, a theoretical analysis. American Economic Review 50, 186–188.

Domencich, T.A., McFadden, D., 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland, Amsterdam.

Feller, W., 1971. An Introduction to Probability Theory and its Applications, Vol. 2, 2nd ed. John Wiley and Sons, New York.

Fujita, M., January 1985. Towards general equilibrium models of urban lend use. Revue Économique 1, 135–167.

Galambos, J., Kotz, S., 1978. Characterizations of probability distributions. Springer Lecture Notes in Mathematics, Vol. 675. New York.

Galambos, J., 1987. The Asymptotic Theory of Extreme Order Statistics, 2nd ed. John Wiley and Sons, New York.

Leonardi, G., 1984. The structure of random utility models in the light of the asymptotic theory of extremes. In: Florian, M. (Ed.), Transportation Planning Models.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P.C. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

Perloff, J.M., Salop, S.C., 1985. Equilibrium with product differentiation. Review of Economic Studies 52 (1), 107–120.

Resnick, S.I., 1987. Extreme Values, Regular Variation, and Point Processes. Springer, New York.

Smith, T.E., 1978. A cost-efficiency principle of spatial interaction behavior. Regional Science and Urban Economics 8, 313–337.

Starrett, D., 1978. Market allocation of location choice in a model with free mobility. Journal of Economic Theory 17, 21–37.

Tomlin, J.A., Tomlin, S.G., 1968. Traffic distribution and entropy. Nature 220, 974–976.

Train, K., 1980. A structured logit model of auto ownership and mode choice. Review of Economic Studies 48 (2), 357–370.

Wilson, A.G., 1970. Entropy in urban and regional modelling. Pion Press, London.

Witlox, F., 1994. Discrete choice modelling in a context of spatial choice behavior: A survey, SESO Report 311, Universiteit Antwerpen.

Yellot, J.I., 1977. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgement and the double exponential distribution. J. Math. Psychology 15, 109–146.