International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

The Structures of Random Utility Models in the Light of the Asymptotic Theory of Extremes

Leonardi, G.

IIASA Working Paper

WP-82-091

September 1982

THE STRUCTURE OF RANDOM UTILITY MODELS IN THE LIGHT OF THE ASYMPTOTIC THEORY OF EXTREMES*

Giorgio Leonardi

September 1982
WP-82-91

*Paper presented at the 1st Course on Transportation and Planning Models, International School of Transportation Planning, Amalfi, Italy, October 11-16, 1982.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

ABSTRACT

The paper reconsiders random utility choice models in the light of asymptotic theory of extremes. The theory is introduced and its main general results are outlined. A stochastic extremal search process is then built, which is shown to produce the Logit model as an asymptotic result under very general conditions. Further applications of the new approach are discussed and outlined.

CONTENTS

1. INTRODUCTION 1

2. THE BASIC RANDOM UTILITY MODEL AND ITS RELATIONSHIP WITH EXTREME VALUE THEORY 2

3. MAIN RESULTS FROM THE ASYMPTOTIC THEORY OF EXTREMES 11

4. ASYMPTOTIC DERIVATION OF THE MULTINOMIAL LOGIT MODEL 22

5. CONCLUDING REMARKS 42

REFERENCES 44

THE STRUCTURE OF RANDOM UTILITY MODELS IN THE LIGHT OF THE ASYMPTOTIC THEORY OF EXTREMES

1. INTRODUCTION

Random utility choice models have become quite popular in such applications as trip distribution analysis, modal choice, residential choice, and so on. However, in spite of the elegance of the theory behind them, some theoretical dissatisfaction is caused by the exceedingly restrictive assumptions currently used to derive specific forms for such models. Namely, the prevailing philosophy seems to be an emphasis on the ability for such models to capture features of individual behavior in a very disaggregated way, so that a one-to-one mapping is tacitly implied between a specific assumption at the disaggregate level (namely, the form of the distribution of the random utility terms) and the resulting observable probabilistic choice pattern.

The goal of this paper is to show that most of these assumptions are unjustified and that, under rather general conditions, observed choice patterns are quite insensitive to disaggregate assumptions. The approach which will be used in order to do so is to reformulate random utility choice theory in terms of asymptotic extreme value theory, a branch of mathematical statistics dealing with the properties of maxima (or minima) of sequences of random variables with a large number of terms.

In particular, it will be shown that, for a sufficiently large number of alternatives, a very wide family of distributions leads to the Logit models as an asymptotic approximation to choice behavior.

2. THE BASIC RANDOM UTILITY MODEL AND ITS RELATIONSHIP WITH EXTREME VALUE THEORY

Let a discrete set of objects, S, be given, and denote by $j \in S$ any of its elements. Let a real constant vector

$$V = (v_1, \ldots, v_m)$$

be given, where $m = |S|$ is the cardinality of S. Finally, let a probability distribution on $\mathbb{R}^m$ be defined by

$$F(X) = P_r\ (\tilde{X} < X)$$

where $X \in \mathbb{R}^m$ and

$$\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_m)$$

is a sequence of m real random variables.

For the triplet (S,V,F) the following problems will be considered:

a. Find the probability distribution

$$H_S(x) = P_r\ (\max_{j \in S} \tilde{u}_j < x) \tag{1}$$

for the maximum element in the sequence of random variables

$$\tilde{U} = (\tilde{u}_1, \ldots, \tilde{u}_m) \tag{2}$$

where $\tilde{u}_j = \tilde{x}_j + v_j$, $j = 1, \ldots, m$

b. Find the discrete probability distribution

$$p_j = P_r(\max_{k \in S-\{j\}} \tilde{u}_k - \tilde{u}_j < 0) \qquad j = 1, \ldots, m \tag{3}$$

for the element in S to which the maximum value in the sequence $\tilde{U}$ is associated.

c. Analyse the behavior of $H_S(x)$ and p_j, $j = 1,\dots,m$ when m becomes, in some sense, large.

Problems like a and c are typically addressed by the extreme value theory, a branch of mathematical statistics for which a well developed theory exists (Von Mises 1936, Gnedenko 1943, de Haan 1970, Galambos 1978).

Problem b can easily be recognized as the mathematical formulation of the main problem addressed by random utility choice models. Indeed, if S is the set of possible choices, v_j and $\tilde{x}_j$ are, respectively, the deterministic and random part of the utility associated with alternative j, $j = 1,\dots,m$, and F(X) is the joint distribution for the random parts, then equation (3) defines the probability of choosing each alternative $j \in S$. Equation (3) is the starting point to build all currently used random utility choice models (see, for instance, Manski 1973, Domencich and McFadden 1975, Williams 1977, Leonardi 1981).

Although it is evident that a theory which solves problem a also solves problem b as a by-product, most of the literature on random utility models seems to be unaware of extreme value theory. Moreover, problem c, which is the bulk of extreme value theory, has been ignored in random utility models, which are always built by introducing very specific assumptions on the distribution F(X).

On the other hand, the tools developed in extreme value theory enable many statements on the asymptotic behavior of $H_S(\mathbf{x})$ and p_j to be made, without requiring a specific form for F(X), provided the set S is, in some sense, large. Since many practical choice situations actually meet (at least approximately) the condition of S being large, some standard results for extreme value statistics might be used to provide random utility models with some general justifications.

The rest of the paper explores this possibility, with no claim of exhausting it.

Before we go on to do so, however, some simple general results will be stated, which hold independently from the form of $F(X)$ and the size of S. $F(X)$ will be assumed to be differentiable up to the mth order (i.e., to admit all partial densities) for every $X \in \mathbb{R}^m$ and to have finite first moments. Moreover, the vector V will be assumed to be bounded, i.e.,

$$-\infty < v_j < \infty \quad , \qquad\qquad j = 1,\ldots,m$$

The partial density with respect to the jth variable will be denoted by $F_j(X)$ and defined as

$$F_j(X) = \frac{\partial F(X)}{\partial x_j}$$

while the marginal density with respect to the jth variable will be denoted by $f_j(x)$.

As a preliminary remark, it is easily seen that, if the distribution for the sequence $\tilde{X}$ is $F(X)$, the distribution for the sequence $\tilde{U}$ defined in (2) is

$$G(X) = P_r(\tilde{X} + V < x) = P_r(\tilde{X} < X - V) = F(X - V) \qquad (4)$$

Moreover, since the event

$$\max_{j \in S} \tilde{u}_j < x$$

is equivalent to the event

$$\tilde{u}_j < x \quad \text{for all } j \in S$$

it follows that

$$H_S(x) = G(x, x,\ldots,x) = F(x-v_1,\ldots,x-v_m) \qquad (5)$$

Since it is evident from (5) that $H_S(x)$ is a function of V, the notation

$$H(x,V) \mathrel{\hat{=}} H_S(x) \qquad (6)$$

will be used.

A lemma will now be stated which summarizes the relationships between the extreme value distribution $H(x,V)$ and the choice probabilities p_j.

Lemma 2.1. Under the assumptions stated above for $F(X)$, and with p_j defined as in (3):

(i) $$p_j = -\int_{-\infty}^{\infty} \frac{\partial H(x,V)}{\partial v_j}\, dx \tag{7}$$

where $H(x,V)$ is the function defined in (6); moreover, $0 \leq p_j \leq 1$ and $\sum_{j=1}^{m} p_j = 1$

(ii) $$\phi(V) = \int_{-\infty}^{\infty} x \; dH(x,V) < \infty \tag{8}$$

and

$$p_j = \frac{\partial \phi(V)}{\partial v_j} \tag{9}$$

(iii) if the function $F(X)$ is replaced by the function

$$F^*(X) = \int_{-\infty}^{\infty} F(x_1+y,\ldots,x_m+y)\; dQ(y) \tag{10}$$

where $Q(y)$ is a univariate probability distribution with finite first moment α, the probabilities p_j are the same as those obtained from $F(X)$.

Proof. To prove statement (i), notice that by definition

$$F_j(X) = \lim_{y\to 0} \frac{P_r(x_j \leq \tilde{x}_j < x_j+y \;,\; \tilde{x}_k < x_k \;:\; k\in S - \{j\})}{y}$$

therefore

$$F_j(x-v_1,\ldots,x-v_m) = \lim_{y\to 0} \frac{P_r(x \leq \tilde{x}_j + v_j < x+y \;,\; \tilde{x}_k + v_k < x \;:\; k\in S-\{j\})}{y}$$

It follows that the total probability for the event

$$\tilde{u}_j > \max_{k \in S-\{j\}} \tilde{u}_k$$

or equivalently

$$\tilde{u}_j > \tilde{u}_k \qquad \text{for all } k \in S-\{j\}$$

where $(u_1,\ldots,u_m)$ is the sequence defined in (2), is given by

$$p_j = \int_{-\infty}^{\infty} F_j(x-v_1,\ldots,x-v_m)\; dx \tag{11}$$

On the other hand, it is evident from definition (6) that

$$F_j(x-v_1,\ldots,x-v_m) = -\frac{\partial H(x,V)}{\partial v_j}$$

and replacing this result in (11) equation (7) follows. To prove that p_j is a proper probability distribution, first note that from (11) it is obviously true that

$$p_j \geq 0 \qquad , \qquad j = 1,\ldots,m \tag{12}$$

moreover

$$\sum_{j=1}^{m} p_j = \int_{-\infty}^{\infty} \sum_{j=1}^{m} F_j(x-v_1,\ldots,x-v_m)\; dx$$

$$= \int_{-\infty}^{\infty} dF(x-v_1.,,,.x-v_m) = \int_{-\infty}^{\infty} dH(x,V) = 1$$

since $H(x,V)$ is a probability distribution. From this and (12) it follows that

$$p_j \leq 1 \qquad , \qquad j = 1,\ldots,m$$

This completes the proof of statement (i).

To prove (8) define the first moments

$$\mu_j = \int_{-\infty}^{\infty} x \ f_j(x) \ dx \quad , \qquad j = 1,\ldots,m$$

where $f_j(x)$ is the jth marginal density of $F(x)$. By assumption

$$\mu_j < \infty \quad , \qquad j = 1,\ldots,m$$

By definition

$$dH(x,V) = dF(x-v_1,\ldots,x-v_m)$$

and by the differentiability assumption [already used to prove statement (i)]

$$dF(x-v_1,\ldots,x-v_m) = \sum_{j=1}^{m} F_j(x-v_1,\ldots,x-v_m) \ dx$$

On the other hand, from the standard properties of probability distributions:

$$F_j(x-v_1,\ldots,x-v_m) \leq f_j(x-v_j)$$

Substitution into (8) yields

$$\phi(V) = \int_{-\infty}^{\infty} x \sum_{j=1}^{m} F_j(x-v_1,\ldots,x-v_m) \ dx \leq$$

$$\leq \sum_j \int_{-\infty}^{\infty} x \ f_j(x-v_j) \ dx = \sum_j (\mu_j + v_j)$$

and since both μ_j and v_j are finite (8) follows. To prove (9) take the derivatives of $\phi(V)$ and integrate by parts

$$\frac{\partial \phi(V)}{\partial v_j} = \int_{-\infty}^{\infty} x \ d \, \frac{\partial H(x,V)}{\partial v_j} = \frac{\partial H(x,V)}{\partial v_j} \, x \ \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\partial H(x,V)}{\partial v_j} \, dx \tag{13}$$

but since

$$\frac{\partial H(x,V)}{\partial v_j} = - F_j(x-v_1,\ldots,x-v_m)$$

and

$$F_j(\infty) = F_j(-\infty) = 0$$

the first term on the right-hand-side of (13) vanishes, and comparison of the second term with (7) establishes (9). This completes the proof of statement (ii).

Statement (iii) easily follows from (9). From definitions (5), (6), and (10)

$$\begin{aligned} H^*(x,V) &= F^*(x-v_1,\ldots,x-v_m) \\ &= \int_{-\infty}^{\infty} F(x+y-v_1,\ldots,x+y-v_m)\; dQ(y) \\ &= \int_{-\infty}^{\infty} H(x+y,V)\; dQ(y) \end{aligned}$$

and from definition (8) and the assumption on the moment of Q(y)

$$\begin{aligned} \phi^*(V) &= \int_{-\infty}^{\infty} x \; dH^*(x,V) \\ &= \int_{y=-\infty}^{\infty} \left[\int_{x=-\infty}^{\infty} x \; dH(x+y,V) \right] dQ(y) \\ &= \int_{y=-\infty}^{\infty} [\phi(V) - y]\; dQ(y) = \phi(V) - \alpha < \infty \end{aligned}$$

It follows that

$$\frac{\partial \phi^*(V)}{\partial v_j} = \frac{\partial \phi(V)}{\partial v_j}$$

which together with (9) proves statement (iii).

Discussion

Statement (i) is a basic result, establishing a general rule to obtain the choice probabilities from the extreme value distribution. That is, in a sense, a precise statement of the claim that a random utility choice model is a by-product of the more general extreme value distribution problem. Statement (ii) strengthens this result, giving with equation (9), a method which is operationally much more useful than representation (7). It also says something more on the properties of a random utility choice model. The function $\phi(V)$ defined by (8) is the expected value of the maximum in the sequence $\tilde{U}$, that is, in the random utility interpretation, the expected utility deriving from an optimal choice. The fact that the choice probabilities p_j are the partial derivatives of the expected utility $\phi(V)$ has an interesting economic interpretation. From the mathematical point of view, (9) states the *integrability conditions* for the vector function

$$P(V) = [p_1(V),\ldots,p_m(V)]$$

showing that the general integral

$$\int P(V)\ dV = \int \sum_{j=1}^{m} p_j(V)\ dv_j$$

exists, and is independent from the path of integration, and it is given by $\phi(V)$ up to an additive constant.

In economics, these conditions are equivalent to those proposed by Hotelling (1938) to ensure the existence and uniqueness of a consumer surplus for a vector demand function of many substitutable commodities. Of course, the original Hotelling formulation uses prices instead of utilities, but one can, with no loss of generality, write

$$v_j = - c_j \quad , \qquad\qquad j = 1,\ldots,m$$

and call c_j a price.

The fact that random utility models satisfy the Hotelling conditions has been noted and studied by many authors in the last 10 years, mainly in the field of transport demand analysis and land use plans evaluation. The approach has been pioneered by Neuburger (1971), for gravity-type models, although no random utility assumption was considered in Neuburger's paper. The link with random utility theory is extensively discussed in Williams (1977), Coelho (1980), Daly (1979) Ben Akiva and Lerman (1979). However, most of the above works are tied to very specific assumptions on the form of $F(x)$, and it does not seem to be explicitly recognized that statement (ii) is fairly general. More recent work free from specific assumptions is found in Leonardi (1981) and Smith (1982).

Statement (iii) is also very important, although almost obvious. It basically says that choice probabilities are unaffected by a shift in the utility scale. Moreover, it says that they remain unaffected by any mixture of shifts. For instance, if choices are made by a population which shifts the utility scale heterogeneously, according to the distribution $Q(y)$, this heterogeneity does not affect the choice behavior. Due to the way the function $H(x,V)$ is defined, it is evident that the shift can be indifferently considered as applied to the random terms $\tilde{x}_j$ or to the deterministic terms v_j. That is, it can easily be proved that

$$\phi(V - y) = \phi(V) - y \tag{14}$$

(see Leonardi 1981), and therefore

$$\int_{-\infty}^{\infty} \phi(V - y)\, dQ(y) = \phi(V) - \alpha$$

which leads to an alternative proof of (iii).

Besides other considerations, statement (iii) says that there is a lot of arbitrariness in fixing an origin for the utility scale, and additive constants (either deterministic or stochastic) can be ignored.

3. MAIN RESULTS FROM THE ASYMPTOTIC THEORY OF EXTREMES

As stated in Section 2, the main problem in extreme value theory is analyzing the behavior of the extreme value distribution when the number of elements in the sequence of random variables becomes large. More precisely, using the terminology introduced in Section 2, the general problem to be explored is to find sequences of normalizing constants (a_m) and (b_m), where $m = |S|$, such that, as $m \to \infty$

$$\lim H_S(a_m + b_m x) = H(x)$$

where $H(x)$ is a nondegenerate probability distribution. The main interest in exploring this problem lies in the fact that, as it will be seen, $H(x)$ can be expected to be largely independent from the specific form of $F(X)$, on which only some weak conditions need to be imposed. This is in striking constrast with the approach followed in building most random utility models, usually obtained by imposing very restrictive and specific assumptions on $F(X)$.

For instance, the Logit model is obtained by assuming $\tilde{X}$ is a sequence of independent identically distributed (i.i.d) random variables with common distribution function:

$$P_r(x_j < x) = \exp\left(-e^{-\beta x}\right) \tag{15}$$

Function (15) will actually be shown to play a fundamental role in extreme value theory, in the sense that almost every asymptotic extreme value distribution can be reduced to (15) by a suitable transformation. But it need not be assumed for each $j \in S$, provided $|S| \to \infty$ and some other requirements are met.

As a matter of fact, the main results in this paper are those on *convergence to the multinomial Logit* for a wide class of random utility models. This makes, in a sense, the Logit model an *aggregate* rather than a *disaggregate* one. Conversely, eventual convergence properties to the multinomial Logit for a wide class of distributions $F(X)$ would make it hardly justifiable to make any inference on the specific form of the actual $F(X)$, since the mapping between $F(X)$ and $H(x)$ is many-to-one.

Most well-known results on extreme value statistics concern sequences of i.i.d. random variables. The independency assumption could be replaced by asymptotic independency for many results, but this will not be pursued here. Some general results from the theory developed for i.i.d. random variables will be sufficient to give an asymptotic justification to the logit model. The following terminology will be used

$F(x) = P_r(\tilde{x} < x)$ is the univariate distribution of any random variable $\tilde{x}$ in the sequence considered

$F_n(X) = \prod_{j=1}^{n} F(x_j)$ is the joint distribution for a sequence of n terms

$H_n(x) = F^n(x)$ is the distribution of the maximum term in a sequence of n terms [this follows from (5) as a special case]

$\alpha(F) = \inf\{x:F(x) > 0\}$ is the lower endpoint of $F(x)$

$\omega(F) = \sup\{x:F(x) < 1\}$ is the upper endpoint of $F(x)$

$H_n(x)$ will be said to *belong to the domain of attraction* of some nondegenerate distribution $H(x)$ if sequences of normalizing constraints a_n and b_n can be found such that

$$\lim_{n\to\infty} H_n(a_n + b_n x) = H(x) \tag{16}$$

The following will be referred to as Condition 3.1 and Condition 3.2.

Condition 3.1. $\omega(F) = \infty$ and there is a constant $\beta > 0$ such that, for all $x > 0$

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\beta}$$

Condition 3.2. For some finite a

$$\int_a^{\omega(F)} [1 - F(x)]\, dx < \infty$$

and for all real x

$$\lim_{t \to \omega(F)} \frac{1 - F[t + xR(t)]}{1 - F(t)} = e^{-x}$$

where $R(t)$, $\alpha(F) < t < \omega(F)$ is the function

$$R(t) = [1 - F(t)]^{-1} \int_t^{\omega(F)} [1 - F(x)]\, dx$$

The main results for i.i.d. random variables are stated without proof in the following reference Lemma.

Lemma 3.1. Any nondegenerate limit in (16) satisfies the functional equation

$$H^m(A_m + B_m x) = H(x) \qquad m \geq 1 \qquad (17)$$

Equation (17) has only three solutions:

$$H_1(x) = \begin{cases} \exp(-x^{-\beta}) & x > 0 \, , \ \beta > 0 \\ 0 & x \leq 0 \end{cases} \qquad (18)$$

$$H_2(x) = \begin{cases} 1 & x \geq 0 \\ \exp[-(-x)^{\beta}] & x < 0 \, , \ \beta > 0 \end{cases} \qquad (19)$$

$$H_3(x) = \exp(-e^{-x}) \qquad -\infty < x < \infty \qquad (20)$$

$F(x)$ belongs to the domain of attraction of:

$H_1(x)$ if, and only if, condition 3.1 holds

$H_2(x)$ if, and only if, $\omega(F) < \infty$ and for the modified distribution

$F^*(x) = F[\omega(F) - \frac{1}{x}]$ $\qquad x > 0$

condition 3.1 holds

$H_3(x)$ if, and only if, condition 3.2 holds.

The sequences of normalizing constants a_n and b_n in (16) can be computed as:

(i) $a_n = 0$, $b_n = \inf\{x: 1 - F(x) \leq 1/n\}$ if $F(x)$ belongs to the domain of attraction of $H_1(x)$

(ii) $a_n = \omega(F)$, $b_n = \omega(F) - \inf\{x: 1 - F(x) \leq 1/n\}$ if $F(x)$ belongs to the domain of attraction of $H_2(x)$

(iii) $a_n = \inf\{x: 1 - F(x) \leq 1/n\}$, $b_n = R(a_n)$ if $F(x)$ belongs to the domain of attraction of $H_3(x)$.

Discussion

Lemma 3.1 is a collection of results scattered in the literature. They can be found with the proofs in the comprehensive book by Galambos (1978), although they date back somewhat earlier. Not mentioning the work on extremes done by Bernoulli and Poisson, the first systematic results on the three possible limits are due to Fisher and Tippet (1928) and von Mises (1936), although these authors limited their considerations to absolutely continuous $F(x)$. The current level of the theory, not requiring continuity, is due mainly to Gnedenko (1943) and de Haan (1970).

The most important qualitative result is perhaps the exhaustive list of possible limits. Lemma 3.1 does not say that a nondegenerate limit in (15) always exists. If, however, it does, then it can only be either $H_1(x)$, $H_2(x)$, or $H_3(x)$. $H_3(x)$ is, of course, of the same form as (15), therefore, one might expect a tight relationship between Condition 3.2 and a possible limiting Logit form for the choice probabilities. It would be nice to have strong behavioral justifications for choosing Condition 3.2, rather than Condition 3.1. However, there is no self-evidence in the way the two conditions are formulated. Some insight is given by assuming $F(x)$ is differentiable. In this case, if one defines the hazard rate

$$\rho(x) = \frac{F'(x)}{1 - F(x)} = \frac{d}{dx} \log [1 - F(x)]$$

it can be shown (von Mises 1936) that Condition 3.1 is equivalent to

$$\lim_{x \to \infty} x \, \rho(x) = \beta \tag{21}$$

while Condition 3.2 is equivalent to

$$\lim_{x \to \omega(F)} \frac{d}{dx} \left[\frac{1}{\rho(x)} \right] = 0 \tag{22}$$

In other words, with (21) the hazard rate is, in the limit, inversely proportional to x, while with (22) the hazard rate is, in the limit, constant. One might recall the meaning of the hazard rate:

$$\rho(x)\,dx = P_r(x \leq \tilde{x} < x + dx \mid \tilde{x} \geq x)$$

that is, $\rho(x)dx$ is the (infinitesimal) probability that the random variable $\tilde{x}$ is nearly x, conditional to the event that it is not less than x. In terms of utilities, it would seem plausible to assume a diminishing returns effect, by which the probability of finding higher utilities does not increase for high utility levels. This would lead to discard assumption (21), and hence Condition 3.1, as unrealistic. There is, however, a stronger argument to justify working under Condition 3.2 only.

Assume, further to Condition 3.1, $\alpha(F) \geq 0$, and define the new random variable

$$\tilde{z} = \log \tilde{x} \tag{23}$$

with distribution $G(\tilde{x})$. It is clear that

$$G(x) = P_r\ (\log \tilde{x} < x) = P_r(\tilde{x} < e^x) = F(e^x)$$

and the hazard rate of G(x) is

$$\rho^*(x) = \frac{G'(x)}{1 - G(x)} = \frac{e^x\ F'(e^x)}{1 - F(e^x)}$$

therefore, defining $y = e^x$

$$\lim_{x\to\infty} \rho^*(x) = \lim_{y\to\infty} \frac{y\ F'(x)}{1 - F(y)} = \lim_{y\to\infty} y\rho(y) = \beta$$

because of (21). Hence, the transformed variable (23) has an asymptotically constant hazard rate, meets condition (22), and belongs to the domain of attraction of $H_3(x)$. Thus a simple logarithmic transformation maps a large subset of the domain of attraction of $H_1(x)$ [namely, the F(x) for which $\alpha(F) \geq 0$] into a subset of the domain of attraction of $H_3(0)$. Similar considerations apply to the domain of attraction of $H_2(x)$. By Lemma 3.1, F(x) belongs to the domain of attraction of $H_2(x)$ if

$$F^*(x) = F\left[\omega(F) - \frac{1}{x}\right] \tag{24}$$

belongs to the domain of attraction of $H_1(x)$; hence

$$G^*(x) = F^*(e^x)$$

belongs to the domain of attraction of $H_3(x)$. Transformation (24) is interesting in its own right. One can write:

$$F^*(x) = P_r\left[\tilde{x} < \omega(F) - \frac{1}{x}\right] = P_r\left(\frac{1}{\omega(F) - \tilde{x}} < x\right)$$

hence $F^*(x)$ is the distribution of the random variable:

$$\frac{1}{\omega(F) - \tilde{x}}$$

Applying now transformation (23) to $F^*(x)$

$$G^*(x) = P_r \{-\log[\omega(F) - \tilde{x}] < x\}$$

therefore, the distribution of the random variable

$$\tilde{u} = -\log\,[\omega(F) - \tilde{x}] \tag{25}$$

belongs to the domain of attraction of $H_3(0)$. The one given by (25) is an interesting utility function. If $\omega(F)$ is interpreted as an *ideal level* for the variable $\tilde{x}$, then for $\tilde{x} \to \omega(F)$

$$\tilde{u} \to \infty$$

the maximum satisfaction. If one additionally assumes:

$$\alpha(F) = 0$$

$$\omega(F) = 1$$

so that $0 \leq \tilde{x} \leq 1$

$$0 \leq \tilde{u} \leq \infty$$

therefore (25) can be used to map a normalized weight into a nonnegative real number.

To summarize, given a sequence of i.i.d. random variables whose maximum has an asymptotic nondegenerate distribution $H_k(x)$, $k = 1,2,3,$ a suitable transformation can always reduce them to a sequence whose asymptotic distribution for the maximum is $H_3(x)$. The transformations are:

I. $\tilde{u} = \log \tilde{x}$ if $k = 1$

II. $\tilde{u} = -\log\,[w(F) - \tilde{x}]$ if $k = 2$

III. $\tilde{u} = \tilde{x}$ if $k = 3$

and their shape is as shown in Figure 1.

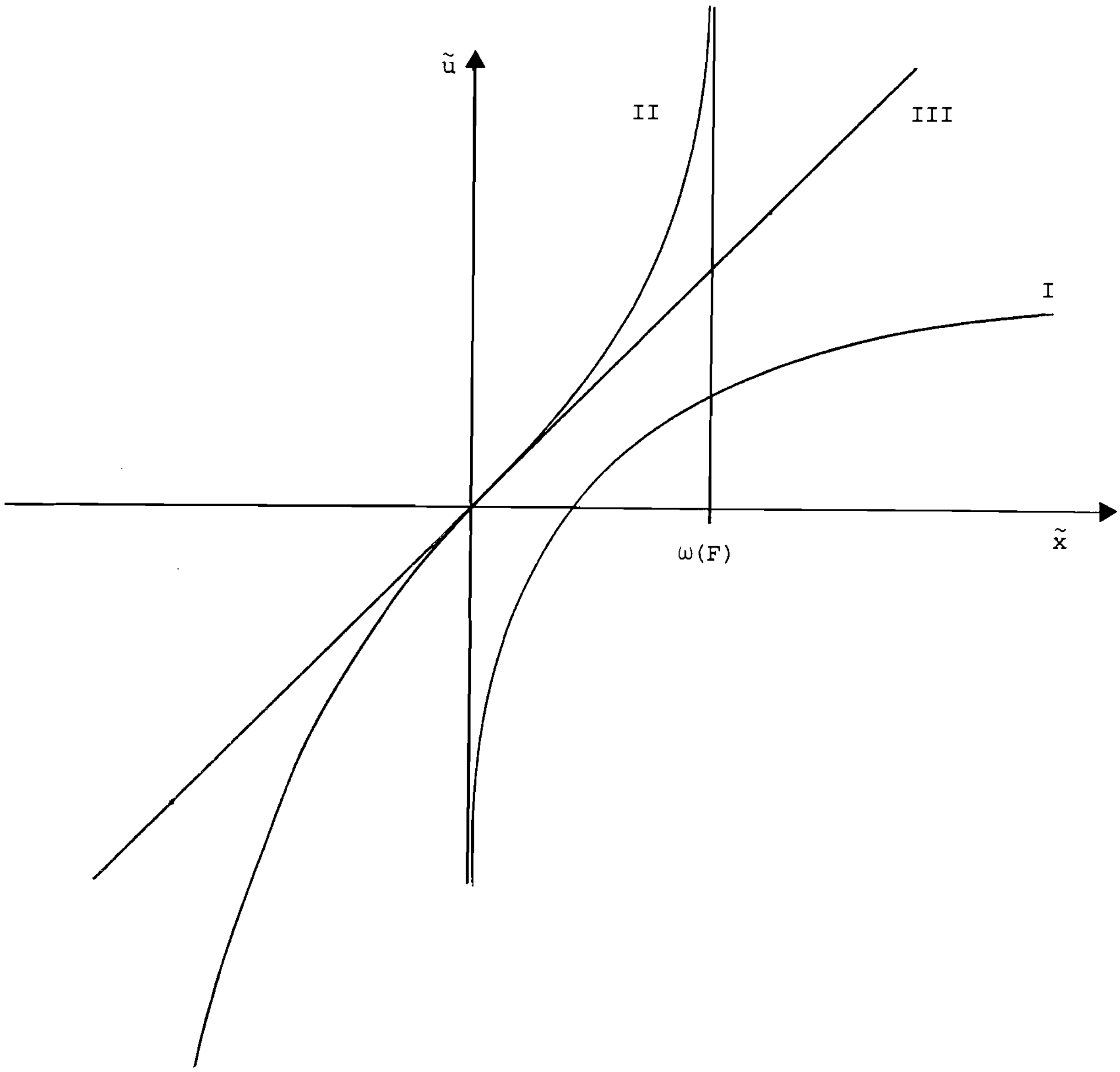

Figure 1. Graph of the three possible transformations required to generate extremes converging to

$$H_3(x) = \exp(-e^{-x})$$

Besides the above theoretical considerations, one might ask how some of the most commonly used distributions behave in the limit. The answer is that most of them belong to the domain of attraction of $H_3(x)$, like the exponential, the normal, the lognormal, the gamma, and the logistic. Some less usual distributions belong to the domain of attraction of $H_1(x)$, such as Cauchy and Pareto distribution, or of $H_2(x)$, such as the uniform and the beta distribution. However, there are qualitative differences in the way of reaching the limit, even within the same domain of attraction, which raise some theoretical and empirical problems. The crucial problem in applications is estimating the normalizing constants a_n and b_n. In a random utility choice model, the constant a_n is not critical, no matter how large it becomes as $n \to \infty$, because of statement (iii) in Lemma 2.1. The constant b_n, on the contrary, is critical, since it is related to the dispersion of the extreme value distribution. Indeed, if

$$\lim_{n\to\infty} H_n(a_n + b_n x) = H_3(x) = \exp(-e^{-x})$$

it follows that

$$\lim_{n\to\infty} H_n(x) = \lim_{n\to\infty} \exp\left[-e^{-\frac{1}{b_n}(x - a_n)}\right] \tag{26}$$

and the right hand side of (26) would be used as an approximation of $H_n(x)$, for large n. It is clear that, if b_n does not depend much on n, it makes sense to estimate it as a constant and consider the limiting approximation a stable low. If, on the other hand, the dependence of b_n from n is not negligible, any empirical estimation of its value would strongly depend on the conditions under which the observations were made, and the limiting law in (26) would be a poor forecasting tool. The nature of this problem is clarified by two examples. As a first example, assume

$$F(x) = 1 - e^{-\beta x}$$

an exponential distribution. Hence it belongs to the domain of attraction of $H_3(x)$. By applying the rules given in Lemma 3.1 it can easily be shown that

$$b_n = 1/\beta \qquad \text{for all } n \geq 1$$

therefore, the limiting approximation would be:

$$H_n(x) \sim \exp\left[-e^{-\beta(x - a_n)}\right]$$

a very stable law. As a second example, assume

$$F(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

a standard normal distribution. It also belongs to the domain of attraction of $H_3(x)$, but the sequence b_n in this case can be shown to be

$$b_n = (2 \log n)^{-1/2}$$

therefore the coefficient in the exponential (26) would be:

$$\beta_n = \frac{1}{b_n} = (2 \log n)^{1/2}$$

The sequence β_n increases with n, although very slowly, and $\lim_{n\to\infty} \beta_n = \infty$, although an unbelievably large n is required to get a large value of β_n (for instance, $\beta_n = 10$ for $n \sim 5.10^{21}$, which far exceeds any reasonable number of alternatives one could find in this world). This sequence is shown in Figure 2. Since the dependency on n does not disappear, any empirical estimate for β_n would depend on the number n of alternatives available at the time. Therefore, with changing size of the choice set, the value for β_n would change. This means, if the population is normal, a constant β is a poor approximation for forecasting purposes.

The fact that β_n increases with n seems counterintuitive, since it implies that the dispersion in the limiting distribution decreases, while many empirical observations on choice behavior (for instance, urban trips) seem to suggest that dispersion increases with the number of alternatives.

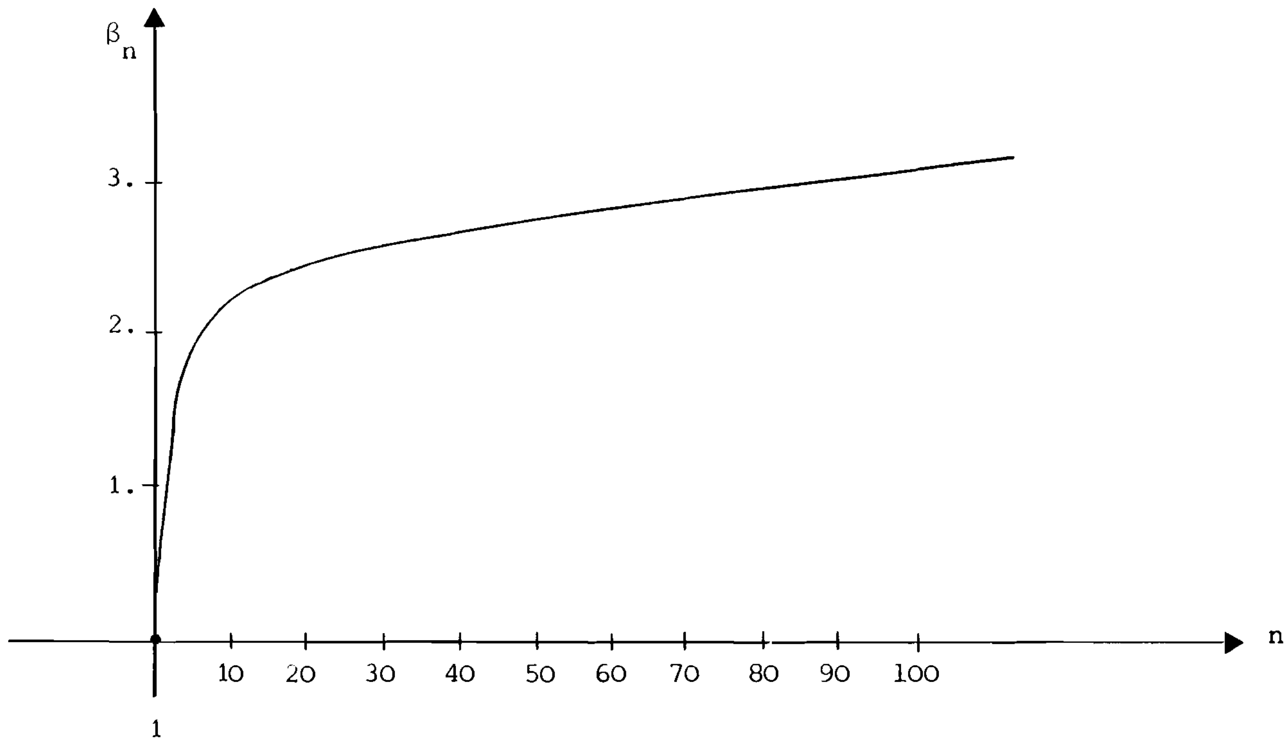

Figure 2. Graph of the sequence $\beta_n = (2 \log n)^{1/2}$

4. ASYMPTOTIC DERIVATION OF THE MULTINOMIAL LOGIT MODEL

Lemma 3.1 will now be used to derive a first result on the asymptotic convergence of choice probabilities to the multinomial Logit model. In order to do so, an additional assumption on the structure of the choice set, the measured utilities and the choice behavior is required. It will be referred to as Condition 4.1.

Condition 4.1. Let S be the total choice set and associate with each $\sigma \in S$ a real number $v(\sigma)$, the measured utility of σ. Assume S can be partitioned into $m \geq 1$ subsets $S_1, S_2, \ldots, S_m$ such that

$$v(\sigma) = v_j \qquad \text{for all } \sigma \in S_j; \; -\infty < v_j < \infty \; ; \; j = 1, \ldots, m$$

$$|S_j| = \infty \qquad \cdot \qquad j = 1, \ldots, m$$

Assume a sample of size n is drawn from S according to n Bernoulli trials, such that, if $\sigma(k) \in S$ is the alternative drawn at trial k,

$$P_r[\sigma(k) \in S_j] = w_j > 0 \quad , \quad \sum_{j=1}^{m} w_j = 1, \; k = 1, \ldots, m \quad .$$

In short, Condition 4.1 assumes the alternatives can be partitioned into very large subsets homogeneous with respect to the measured utilities. The information on alternatives is obtained by independent trials. Clustering alternatives into homogeneous subsets is more often than not a natural way to formulate a choice problem. For instance, in models for travel demand, a geographic area is usually divided into smaller zones, each zone containing many possible trip destinations, and the same average travel cost is assigned to the trips from a given origin to any destination within the same zone. In residential mobility and migrations, alternatives are clustered into regions, and the same average attributes are assigned to any alternative within the same region. Aggregating alternatives into homogeneous subsets is actually a need in modeling spatial choice problems, since the task of listing them one by one is impossible and unrealistic.

All the ingredients are now ready to prove the following.

Theorem 4.1. Assume the random utility terms are i.i.d. random variables with distribution $F(x)$ belonging to the domain of attraction of $H_3(x)$ and let Condition 4.1 hold. Assume further $\omega(F) = \infty$ and there is some constant x_1, such that $F'(x)$, $F''(x)$ exist for $x > x_1$. Define σ_n such that

$$v(\sigma_n) + \tilde{x}(\sigma_n) = \max_{1 \leq k \leq n} v[\sigma(k)] + \tilde{x}[\sigma(k)]$$

(ties broken arbitrarily)

where $\tilde{x}(\sigma)$ is a random variable with distribution $F(x)$, and

$$P_n(j) = P_r(\sigma_n \in S_j) \qquad \begin{array}{l} j = 1,\ldots,m \\ n \geq 1 \end{array}$$

Then:

(i) if $\lim_{t\to\infty} \frac{F'(t)}{1 - F(t)} = \beta < \infty$

$$\lim_{n\to\infty} P_n(j) = \frac{w_j \, e^{\beta v_j}}{\sum_{j=1}^{m} w_j \, e^{\beta v_j}} \tag{27}$$

(ii) if $\lim_{t\to\infty} \frac{F'(t)}{1 - F(t)} = \infty$

$$\lim_{n\to\infty} P_n(j) = \begin{cases} 1 \,, & v_j = \max_{1\leq k\leq m} v_k \\ 0 \,, & \text{otherwise} \end{cases} \tag{28}$$

Proof. Consider the following stochastic process in discrete time. The system will be said to be in state (j,x) at time n if $\sigma_n \in j$ and $v(\sigma_n) + \tilde{x}(\sigma_n) = x$. The above process is easily seen to be a homogeneous Markov chain with mixed state space. Define the transition probabilities

$$p(j,y:i,x) = P_r\left\{\sigma_{n+1} = j, [v(\sigma_{n+1}) + \tilde{x}(\sigma_{n+1})] < y \mid \sigma_n = i\ [v(\sigma_n) + \tilde{x}(\sigma_n)] = x\right\}$$

For $j \neq i$, a transition from i to j occurs only if an alternative from S_j is drawn, which according to Condition 4.1 happens with probability w_j, and it has a total utility in the interval $[x,y)$ which happens with probability

$$F(y - v_j) - F(x - v_j) \qquad x \leq y$$

If $y < x$ no transition occurs. Therefore,

$$p(j,y:i,x) = \begin{cases} w_j[F(y - v_j) - F(x - v_j) , & x \leq y \\ 0 , & x > y \end{cases} \tag{29}$$

for $j \neq i$.

The system can remain in state i in two mutually exclusive ways:

I. An alternative in S_i is drawn (different from the current one) with total utility in the interval $[x,y)$; the probability for this event is given by (29), for $j = i$.

II. An alternative is drawn from any S_j, $1 \leq j \leq m$, but it has a total utility less than x; this happens with probability

$$\sum_{j=1}^{m} w_j\ F(x - v_j)$$

Adding up events I and II one gets:

$$p(i,y;i,x) = \begin{cases} w_i\ F(y - v_i) + \sum\limits_{j\neq i} w_j\ F(x - v_j) , & x \leq y \\ 0 & , x > y \end{cases} \tag{30}$$

The state probabilities

$$p_n(j,y) = P_r[\sigma_n = j \ , \ v(\sigma_n) + x(\sigma_n) < y]$$

satisfy the Kolmogoroff equations:

$$P_{n+1}(j,y) = \sum_{i=1}^{m} \int_{-\infty}^{y} p(j,y;i,x) \, dP_n(i,x) \qquad (31)$$

$$j = 1,\ldots,m \qquad -\infty < y < \infty \qquad n \geq 1$$

with the initial conditions

$$P_1(j,y) = w_j F(y - v_j) \qquad j = 1,\ldots,m \quad -\infty < y < \infty \qquad (32)$$

After substitution from (29) and (30) and some rearrangements, equation (31) becomes

$$P_{n+1}(j,y) = w_j \, F(y - v_j) \sum_{i=1}^{m} P_n(i,y)$$

$$- \, w_j \int_{-\infty}^{y} F(x - v_j) \, d \sum_{i=1}^{m} P_n(i,x)$$

$$+ \int_{-\infty}^{y} \left[\sum_{i=1}^{m} w_i \, F(x - v_i) \right] d \, P_m(j,x) \qquad (33)$$

Now let the following functions be introduced

$Q_n(y) = \sum_{i=1}^{m} P_n(i,y)$ the distribution of the maximum total utility after n trials

$G(x) = \sum_{i=1}^{m} w_i \, F(x - v_i)$ the distribution of the total utility after one trial

The equation (33) can be reformulated as:

$$P_{n+1}(j,y) = w_j\, F(y - v_j)\, Q_n(y)$$

$$- w_j \int_{-\infty}^{y} F(x - v_j)\, d\, Q_n(x)$$

$$+ \int_{-\infty}^{y} G(x)\, d\, P_n(j,x)$$

and since from the rule of integration by parts

$$\int_{-\infty}^{y} Q_n(x)\, dF(x - v_j) = F(y - v_j)\, Q_n(y)$$

$$- \int_{-\infty}^{y} F(x - y_j)\, d\, Q_n(x)$$

one finally gets:

$$P_{n+1}(j,y) = w_j \int_{-\infty}^{y} Q_n(x)\, d\, F(x - v_j)$$

$$+ \int_{-\infty}^{y} G(x)\, d\, P_n(j,x) \qquad j = 1,\ldots,m \quad n \geq 1 \quad -\infty < y < \infty \qquad (34)$$

Summing both sides of (34) over $j = 1,\ldots,m$ and using the rule of integration by parts again, the following equation relating $Q_n(y)$ and $G(y)$ is obtained:

$$Q_{n+1}(y) = Q_n(y)\, G(y) \qquad n \geq 1 \quad -\infty < y < \infty \qquad (35)$$

with the initial condition

$$Q_1(y) = G(y)$$

The solution of (35) is obviously

$$Q_n(y) = G^n(y) \qquad (36)$$

a result which could have been obtained directly. Substitution of (36) into (34) and induction over n with the initial conditions (32) easily yield a closed-form solution for the state probabilities:

$$P_{n+1}(j,y) = (n + 1)\, w_j \int_{-\infty}^{y} G^n(x)\, d\, F(x - v_j) \qquad (37)$$

Now clearly,

$$P_n(j) = P_r(\sigma_n \in S_j) = P_n(j,\infty)$$

therefore

$$P_{n+1}(j) = (n + 1)\, w_j \int_{-\infty}^{\infty} G^n(x)\, d\, F(x - v_j) \qquad (38)$$

Hence, the asymptotic behavior of $P_n(j)$ depends on the asymptotic behavior of $G^n(x)$. Let it first be proved that $G(x)$ belongs to the domain of attraction of $H_3(x)$, i.e., it satisfies Condition 3.2. Since $F(x)$ satisfies this condition, there is some finite a for which:

$$\int_a^{\infty} [1 - F(x)]\, dx < \infty$$

and of course this implies

$$\int_b^{\infty} [1 - F(x)]\, dx < \infty$$

for any $b > a$. Now

$$\int_b^{\infty} [1 - G(x)]\, dn = \sum_{j=1}^{m} w_j \int_b^{\infty} [1 - F(x - v_j)]\, dx$$

$$= \sum_{j=1}^{m} w_j \int_{b-v_j}^{\infty} [1 - F(x)]\, dx$$

and if one chooses $b = a + \max_{1 \le j \le m} v_j$, all the above integrals converge. Hence $G(x)$ satisfies the first part of Condition 3.2. For the second part, define

$$R(t) = [1 - F(t)]^{-1} \int_t^{\infty} [1 - F(x)]\, dx$$

and

$$\bar{R}(t) = [1 - G(t)]^{-1} \int_t^{\infty} [1 - G(x)]\, dx$$

From the definition of G(x) and R(t) it follows:

$$\bar{R}(t) = \left\{ \sum_j w_j \, [1 - F(t - v_j)] \right\}^{-1} \sum_{j=1}^{m} w_j \int_t^{\infty} [1 - F(x - v_j)]\, dx$$

$$= \sum_j W_j(t)\, R(t - v_j)$$

where

$$W_j(t) = \frac{w_j [1 - F(t - v_j)]}{\sum_{j=1}^{m} w_j [1 - F(t - v_j)]} \qquad j = 1,\ldots,m$$

For the weights $W_j(t)$ it is true that

$$W_j(t) \geq 0 \quad , \quad \sum_{j=1}^{m} W_j(t) = 1 \qquad \text{for all real } t$$

hence $\bar{R}(t)$ is a weighted arithmetic mean of $R(t - v_j)$, $j = 1,\ldots,m$, and

$$\min_{1 \leq j \leq m} R(t - v_j) \leq \bar{R}(t) \leq \max_{1 \leq j \leq m} R(t - v_j)$$

Since the v_j are finite

$$\lim_{t \to \infty} \min_{1 \leq j \leq m} R(t - v_j) = \lim_{t \to \infty} \max_{1 \leq j \leq m} R(t - v_j) = \lim_{t \to \infty} R(t)$$

and one concludes that

$$\lim_{t \to \infty} \bar{R}(t) = \lim_{t \to \infty} R(t) = \lim_{t \to \infty} R(t - v_j) \, , \; j = 1,\ldots,m \qquad (39)$$

Using the weights $W_j(t)$ defined above, it is easily shown that

$$\frac{1 - F[t + x\bar{R}(t)]}{1 - G(t)} = \sum_{j=1}^{m} W_j(t) \frac{1 - F[t - v_j + x\bar{R}(t)]}{1 - F(t - v_j)}$$

Using (39) and Condition 3.2 for $F(x)$:

$$\lim_{t\to\infty} \frac{1 - F[t - v_j + x\bar{R}(t)]}{1 - F(t - v_j)} = \lim_{t\to\infty} \frac{1 - F[t - v_j + x\bar{R}(t - v_j)]}{1 - F(t - v_j)} = e^{-x}$$

Therefore

$$\lim_{t\to\infty} \frac{1 - G[t + x\bar{R}(t)]}{1 - G(t)} = \sum_{j=1}^{m} W_j(t)\, e^{-x} = e^{-x}$$

hence $G(x)$ also satisfies the second part of Condition 3.2, and belongs to the domain of attraction of $H_3(x)$. According to Lemma 3.1, normalizing constants a_n and b_n exist such that:

$$\lim_{n\to\infty} G^n(a_n + b_n x) = \exp(-e^{-x}) \tag{40}$$

and they can be computed as

$$a_n = \inf [x : 1 - G(x) \leq 1/n] \tag{41}$$

$$b_n = \bar{R}(a_n) \tag{42}$$

Due to the definition of $G(x)$ and the continuity assumption on $F(x)$ for $x > x_1$, for large n rule (41) reduces to finding the only root of the equation

$$\sum_{j=1}^{m} w_j \, [1 - F(a_n - v_j)] = \frac{1}{n} \tag{43}$$

which holds for all $a_n > x_1 + \max_j v_j$, an inequality surely met when $n \to \infty$, since $\lim_{n\to\infty} a_n = \infty$.

Using the limit (40) one can approximate the integral on the right hand side of (38) for large n:

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} G^n(x)\, dF(x - v_j) = \lim_{n\to\infty} \int_{-\infty}^{\infty} G^n(a_n + b_n x)\, dF(a_n + b_n x - v_j)$$

$$= \lim_{n\to\infty} \int_{\infty}^{\infty} \exp\,(-e^{-x})\, dF(a_n + b_n x - v_j) \tag{44}$$

Since $\lim_{n\to\infty} a_n = \infty$, from equation (42) and result (39) it follows

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} \bar{R}(a_n) = \lim_{n\to\infty} R(a_n - v_j) \qquad 1 \leq j \leq m \tag{45}$$

Therefore, from Condition 3.2:

$$\lim_{n\to\infty} [1 - F(a_n - v_j + b_n x) = \lim_{n\to\infty} \left\{1 - F[a_n - v_j + xR(a_n - v_j)]\right\}$$

$$= \lim_{n\to\infty} \frac{1 - F[a_n - v_j + xR(a_n - v_j)]}{1 - F(a_n - v_j)} [1 - F(a_n - v_j)]$$

$$= e^{-x} \lim_{n\to\infty} [1 - F(a_n - v_j)]$$

or

$$\lim_{n\to\infty} F(a_n - v_j + b_n x) = 1 - e^{-x} \lim_{n\to\infty} [1 - F(a_n - v_j)]$$

Replacing this result into (44)

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} G^n(x)\, dF(x - v_j) = \left[\int_{-\infty}^{\infty} e^{-x} \exp\,(-e^{-x})\, dx\right] \lim_{n\to\infty} [1 - F(a_n - v_j)]$$

$$= \left[\int_{-\infty}^{\infty} dH_3(x)\right] \lim_{n\to\infty} [1 - F(a_n - v_j)]$$

$$= \lim_{n\to\infty} [1 - F(a_n - v_j)] \tag{45}$$

since

$$\int_{-\infty}^{\infty} d\ H_3(x) = 1$$

Substitution of result (45) into (38) finally yields:

$$\lim_{n\to\infty} P_n(j) = w_j \lim_{n\to\infty} n\ [1 - F(a_n - v_j)]$$

and since from equation (43)

$$n = \frac{1}{\sum_{j=1}^{m} w_j [1 - F(a_n - v_j)]}$$

it follows:

$$\lim_{n\to\infty} P_n(j) = \lim_{n\to\infty} \frac{w_j [1 - F(a_n - v_j)]}{\sum_{j=1}^{m} w_j [1 - F(a_n - v_j)]} \tag{46}$$

Since $F(x)$ is assumed to belong to the domain of attraction of $H_3(x)$, property (22) holds for the hazard rate:

$$\rho(x) = \frac{F'(x)}{1 - F(x)}$$

Clearly property (22) implies that either

$$\lim_{x\to\infty} \rho(x) = \beta < \infty$$

or

$$\lim_{x\to\infty} \rho(x) = \infty$$

On the other hand, it is true in general that for any probability distribution $F(x)$ which is continuous for $x > x_1$:

$$[1 - F(x)] = [1 - F(x_1)] \exp\left[- \int_{x_1}^{x} \rho(y)\, dy\right]$$

Therefore, if $a_n - v_j > x_1$:

$$[1 - F(a_n - v_j)] = [1 - F(a_n)] \exp\left[- \int_{a_n - v_j}^{a_n} \rho(y)\, dy\right]$$

$$= [1 - F(a_n)] \exp\left[\int_0^{v_j} \rho(a_n - x)\, dx\right] \qquad (47)$$

By using the mean value theorem for integrals, there is some ξ, $\xi \in (0, v_j)$, such that

$$\int_0^{v_j} \rho(a_n - x)\, dx = v_j\, \rho(a_n - \xi) \qquad (48)$$

substituting (48) into (47) yields the estimate:

$$[1 - F(a_n - v_j)] = [1 - F(a_n)]\, e^{v_j\, \rho(a_n - \xi)}, \qquad \xi \in (0, v_j)$$

and (46) becomes

$$\lim_{n\to\infty} P_n(j) = \lim_{n\to\infty} \frac{w_j\, e^{v_j\, \rho(a_n - \xi)}}{\sum_{j=1}^{m} w_j\, e^{v_j\, \rho(a_n - \xi)}}, \qquad \xi \in (0, v_j) \qquad (49)$$

Consider the case

$$\lim_{x\to\infty} \rho(x) = \beta < \infty$$

Then

$$\lim_{n\to\infty} \rho(a_n - \xi) = \beta$$

and

$$\lim_{n\to\infty} P_n(j) = \frac{w_j\, e^{\beta v_j}}{\sum_{j=1}^{m} w_j\, e^{\beta v_j}}$$

Now consider the case

$$\lim_{x\to\infty} \rho(x) = \infty$$

and define

$$v_{j*} = \max_{1\leq j\leq m} v_j$$

Equation (49) can be written for j* as

$$\lim_{n\to\infty} P_n(j*) = \lim_{n\to\infty} \frac{w_{j*}}{w_{j*} + \sum_{k\neq j*} w_k \, e^{(v_k - v_{j*})\rho(a_n - \xi)}}$$

and since $v_k - v_{j*} < 0$, $k = 1,\dots,m$ for all $k \neq j$ and $\lim_{n\to\infty} \rho(a_n - \xi) = \infty$:

$$\lim_{n\to\infty} P_n(j*) = 1$$

and this, of course, inplies

$$\lim_{n\to\infty} P_n(j) = 0 \qquad \text{for all } j \neq j*$$

The proof of Theorem 4.1 is completed.

Together with the above limiting results for the choice probabilities, one would like to also have an asymptotic approximation for the expected utility, as defined in (8) and make sure that property (9) holds in the limit. This is provided by the following corollary.

Corollary 4.1. Let the assumptions of Theorem 4.1 hold with $x_1 = -\infty$, and define the function

$$\phi_n(V) = \int_{\infty}^{\infty} y \, d\,Q_n(y)$$

where $Q_n(y)$ is given by equation (36). Then:

(i) $$\frac{\partial \phi_n(V)}{\partial v_j} = P_n(j)$$

(ii) $$\lim_{n\to\infty} \phi_n(V) = \begin{cases} \frac{1}{\beta} \log \sum_{j=1}^{m} w_j e^{\beta v_j} + \lim_{n\to\infty} C_n, & \text{if } \lim_{t\to\infty} \rho(t) = \beta < \infty \\ \max_{1\le j\le m} v_j + \lim_{\to\infty} C_n, & \text{if } \lim_{t\to\infty} \rho(t) = \infty \end{cases}$$

where $\rho(t) = F'(t)/[1 - F(t)]$ and C_n is a sequence asymptotically independent from V.

Proof. From equation (36)

$$Q_n(y) = G^n(y)$$

where G(y) is defined as

$$G(y) = \sum_{j=1}^{m} w_j F(y - v_j)$$

Therefore, since F(x) is differentiable for every real x,

$$\frac{\partial \phi_n(V)}{\partial v_j} = \int_{-\infty}^{\infty} y \, d\left[n G^{n-1}(y) \frac{G(y)}{v_j}\right]$$

$$= -n w_j \int_{-\infty}^{\infty} y \, d[G^{n-1}(y) F'(y - v_j)]$$

Integration by parts yields

$$\frac{\partial \phi_n(V)}{\partial v_j} = n w_j \int_{-\infty}^{\infty} G^{n-1}(y) \, dF(y - v_j) \tag{50}$$

and comparison of (50) with (38) proves statement (i).

To prove the first case considered in statement (ii), one simply observes that:

$$\frac{\partial}{\partial v_j}\left[\frac{1}{\beta} \log \sum_{j=1}^{m} w_j e^{\beta v_j}\right] = \frac{w_j e^{\beta v_j}}{\sum_{j=1}^{m} w_j e^{\beta v_j}} \tag{51}$$

which is the same as (27). On the other hand, by using the asymptotic approximation

$$\lim_{n\to\infty} \phi_n(V) = \lim_{n\to\infty} \int_{-\infty}^{\infty} y \, d\, G^n(y) = \lim_{n\to\infty} \int_{-\infty}^{\infty} (a_n + b_n x) \, d\, G^n(a_n + b_n x)$$

$$= \lim_{n\to\infty} \left[a_n + b_n \int_{-\infty}^{\infty} x \, e^{-x} \exp(-e^{-x}) \, dx \right]$$

$$= \lim_{n\to\infty} (a_n + b_n \gamma)$$

where γ is Euler's constant. Of course, the sequence $a_n + b_n \gamma$ tends to infinity as $n \to \infty$, therefore, taking equation (51) into account, the limiting form for $\phi_n(V)$ must be of the form

$$\lim_{n\to\infty} \phi_n(V) = \frac{1}{\beta} \log \sum_{j=1}^{n} w_j \, e^{\beta v_j} + \lim_{n\to\infty} C_n$$

where $C_n \to \infty$ as $n \to \infty$ and $\lim_{n\to\infty} \frac{\partial C_n}{\partial v_j} = 0$, $j = 1,\ldots,m$.

The second case considered in statement (ii) follows obviously from equation (28).

Discussion of Theorem 4.1 and Corollary 4.1.

The somewhat lengthy proof of Theorem 4.1 is actually an excuse to introduce the prototype of a basic stochastic search model, namely, the Markov Chain defined by equations (33). The main departure of a search-based random utility model from standard ones is the assumption on the knowledge of the choice set, which in a search behavior is always limited, although increasing with the number of trials, while in the classic random utility model it is unlimited from the start. While the assumption of Bernoulli trials might seem restrictive, the main results obtained actually carry over to a wider class of sampling processes, provided they satisfy a strong law of large numbers, in the sense that, if $H_j(n)$ is the number of units in S_j sampled after n trials, $\lim_{n\to\infty} H_j(n)/n = w_j$, a constant.

The Markov Chain of equations (33) is related to the Extremal Process, whose theory was started by Dwass (1964) and Lamperti (1964). A treatment of discrete-time extremal processes is found in Shorrock (1974). The theory for such processes is rapidly developing, and a closer look at it by social scientists is surely worth the effort.

Equations (27) and (28) raise again the problem of qualitative differences in the limiting behavior among distributions belonging to the domain of attraction of $H_3(x)$. The limit given by (27) is a standard multinomial logit, while the one given by (28) is equivalent to deterministic utility maximizing. However, while using (27) would give fairly good approximations to the actual behavior when the assumption $\beta < \infty$ holds, the limiting form (28) would give a totally useless approximation to the actual behavior, since such a limit is usually reached so slowly that no real situation will ever be close enough to it. Even when the hazard rate tends to infinity, much better approximations are still obtained by using a multinomial logit, but in this case the estimated β parameter is not independent from n.

In the nondegenerate case (27), the choice probabilities also depend on the w_j, the sampling probabilities in the Bernoulli trials. Since the model would remain unaffected if the w_j were all multiplied by a constant, only the knowledge of weights proportional to the sampling probabilities is needed. A natural and simple assumption would be:

$$w_j \propto |S_j|$$

but any other assumption is possible. If, for instance, the actor has a prior knowledge or guess on the convenience of each S_j, it might be reasonable to assume:

$$w_j \propto g(v_j)$$

where $g(x)$ is some nonnegative, nondecreasing function. A possible meaningful generalization suggests itself in this case. The assumption of sampling probabilities remaining constant during the search is unrealistic. It is more reasonable to assume the actor will modify them during the search, depending on the outcomes of the trials. In other words, a *learning* and *adaptation* mechanism would be introduced. Such a more general search model will be the subject of further work.

Although the formulation of a choice process as a search one is natural in many ways, one might find the assumption on the clustering structure of the choice set too restrictive, and insist on having a different measured utility for each alternative. In this case Condition 4.1 will fail to hold and Theorem 4.1 is useless. The following reference lemma will be useful to deal with this case. It is a well known result, due to Mejzler (1950), therefore it will be stated without proof.

Lemma 4.1. Let $\tilde{x}_1,\ldots,x_n$ be a sequence of independent random variables with distributions

$$F_j(x) = P_r(\tilde{x}_j < x) \qquad j = 1,\ldots,n$$

Let there be a sequence Z_n such that

$$\lim_{n\to\infty} \sum_{j=1}^{n} [1 - f_j(Z_n)] = A \qquad 0 < A < \infty$$

Furthermore, let there be a constant B such that

$$n[1 - F_j(Z_n)] \leq B \qquad \text{for all n and j}$$

Then

$$\lim_{n\to\infty} \prod_{j=1}^{n} F_j(Z_n) = e^{-A} .$$

Lemma 4.1 keeps the independency assumption, but does not require identical distributions. It can be applied to the special structure of random utility models, where the distribution for the total utility of each alternative j is $F(x - v_j)$. The result is stated in the next theorem.

Theorem 4.2. Let $\tilde{u}_1,\ldots,u_m$ be a sequence of independent random variables with distributions

$$P_r(\tilde{u}_j < x) = F(x - v_j) \qquad j = 1,\dots,n$$

where $-\infty < v_j < \infty$, $j = 1,\dots,n$ and $F(x)$ is a twice-differentiable univariate distribution with $\alpha(F) = -\infty$, $\omega(F) = \infty$, and such that

$$\lim_{t\to\infty} \frac{1 - F(t + x)}{1 - F(t)} = e^{-\beta x} \qquad 0 < \beta < \infty \tag{52}$$

Define

$$P_n(j) = P_r(\tilde{u}_j > \max_{\substack{1\le k\le n \\ k\ne j}} \tilde{u}_k)$$

Then

$$\lim_{n\to\infty} p_n(j) = \lim_{n\to\infty} \frac{e^{\beta v_j}}{\sum_{j=1}^{n} e^{\beta v_j}} \tag{53}$$

Proof. It is first remarked that assumption (52) is equivalent to stating that $F(x)$ belongs to the domain of attraction of $H_3(x)$ and

$$\lim_{t\to\infty} \frac{F'(t)}{1 - F(t)} = \beta$$

It will now be shown that the sequence $\tilde{u}_1,\dots,\tilde{u}_n$ satisfies the conditions of Lemma 4.1. Consider the sequence

$$Z_n = a_n + x \qquad -\infty < x < \infty$$

where a_n is the root of the equation

$$\sum_{j=1}^{n} [1 - F(a_n - v_j)] = 1 \tag{54}$$

Since $1 - F(x)$ is a continuous monotone decreasing function (54) implies:

$$\lim_{n\to\infty} a_n = \infty$$

Therefore, from assumption (52)

$$\lim_{n\to\infty} [1 - F(a_n - v_j - x)] = \lim_{n\to\infty} \frac{1 - F(a_n - v_j + x)}{1 - F(a_n - v_j)} [1 - F(a_n - v_j)$$

$$= e^{-\beta x} \lim_{n\to\infty} [1 - F(a_n - v_j)$$

and

$$\lim_{n\to\infty} \sum_{j=1}^{n} [1 - F(a_n - v_j + x)] = e^{-\beta x} \lim_{n\to\infty} \sum_{j=1}^{n} [1 - F(a_n - v_j)] = e^{-\beta x} \tag{55}$$

because of equation (54). Thus the first condition of Lemma 4.1 is met by the sequence $a_n + x$, with

$$A = e^{-\beta x} \quad .$$

Moreover, the quantities

$$n[1 - F(a_n - v_j + x)] \tag{56}$$

are easily seen to be bounded for all n and j. Indeed, for $n < \infty$, expression (56) is finite, since $0 \leq 1 - F(x) \leq 1$ for all x. For $n \to \infty$, the following asymptotic approximation can be used:

$$\lim_{n\to\infty} [1 - F(a_n - v_j)] = \lim_{n\to\infty} \frac{1 - F(a_n - v_j)}{1 - F(a_n)} [1 - F(a_n)]$$

$$= e^{\beta v_j} \lim_{n\to\infty} [1 - F(a_n)]$$

which follows from assumption (52). Substitution of this approximation into equation (54) yields

$$\lim_{n\to\infty} [1 - F(a_n)] \sum_{j=1}^{n} e^{\beta v_j} = 1 \tag{57}$$

from which the inequality follows:

$$\lim_{n\to\infty} [1 - F(a_n)] \leq \frac{1}{\min_j e^{\beta v_j}} < \infty$$

Again using the asymptotic approximation,(56) becomes

$$\lim_{n\to\infty} n\ [1 - F(a_n - v_j + x)] = e^{-\beta(x-v_j)} \lim_{n\to\infty} n\ [1 - F(a_n)]$$

$$\leq \frac{e^{\beta v_j}}{\min_j e^{\beta v_j}} e^{-\beta x} < \infty$$

Thus Lemma 4.1 applies and

$$\lim_{n\to\infty} \prod_{j=1}^{n} F(a_n - v_j + x) = \exp\ (-e^{-\beta x})\ , \quad -\infty < x < \infty$$

The limiting expected value for $\max_{1\leq j\leq n} \tilde{u}_j$ is given by

$$\lim_{n\to\infty} \phi_n(V) = \lim_{n\to\infty} \int_{\infty}^{\infty} (a_n + x)\ d\ [\exp\ (-e^{-\beta n})]$$

$$= \lim_{n\to\infty} a_n + \gamma/\beta$$

where γ is Euler's constant. The sequence a_n is, of course, a function of $V = (v_1,\ldots,v_n)$ while the term γ/β is a constant. Therefore, applying property (9) in Lemma 2.1

$$\lim_{n\to\infty} p_n(j) = \lim_{n\to\infty} \frac{\partial \phi_n(V)}{\partial v_j} = \lim_{n\to\infty} \frac{\partial a_n}{\partial v_j}$$

From (57) it follows that

$$\lim_{n\to\infty} \log\,[1 - F(a_n)] = - \lim_{n\to\infty} \log \sum_{j=1}^{n} e^{\beta v_j}$$

and applying the derivation rule for implicit functions

$$- \lim_{n\to\infty} \frac{F'(a_n)}{1 - F(a_n)} \frac{\partial a_n}{\partial v_j} = - \beta \lim_{n\to\infty} \frac{e^{\beta v_j}}{\sum_{j=1}^{n} e^{\beta v_j}}$$

But from assumption (52):

$$\lim_{n\to\infty} \frac{F'(a_n)}{1 - F(a_n)} = \beta$$

and one concludes that

$$\lim_{n\to\infty} p_n(j) = \lim_{n\to\infty} \frac{e^{\beta v_j}}{\sum_{j=1}^{n} e^{\beta v_j}}$$

Discussion

Theorem 4.2 seems more general than Theorem 4.1, but it is actually less useful. The assumption on the clustering structure of the choice set has been dropped, but the price paid for this is the unrealistic assumption of unlimited knowledge of the choice set. This assumption is needed in order to make result (53) independent from the way the alternatives are ordered. When the right hand side of (53) is used as an approximation for a *finite* (although large) n, one has therefore to make an arbitrary decision on which alternatives to drop from the sequence. This always produces an unpredictably biased model. Moreover, the limiting choice probabilities (53) are, in a sense, always ill-defined and difficult to handle empirically, since they are of the order of magnitude 1/n and assume very small values as $n \to \infty$. This, of course, does not happen when clusters are introduced, since limits (27) and (28) do not depend on n.

It should be remarked that Theorem 4.2 has been proved under the assumption of a bounded limiting hazard rate. The case of an unbounded limiting hazard rate would have led to a result similar to (28), adding nothing really new to what has been said already about this degenerate behavior. As an additional historical remark, it must be mentioned that, although Theorem 4.2 does not appear in the literature, an analogous result has been proved by Juncosa (1949). The problem addressed by Juncosa seems different since it deals with lifetime heterogeneity, the distribution for the minimum, rather than the maximum, is looked for, and the distribution is assumed to satisfy Condition 3.1 (suitably restated for minima) rather than 3.2. However, since a problem of minimum can be restated as a problem of maximum by a change in sign, and since Condition 3.1 can be mapped into Condition 3.2 by a logarithmic transformation (this has been shown in the discussion following Lemma 3.1), part of Theorem 4.2 can actually be derived from the result of Juncosa.

5. CONCLUDING REMARKS

This paper has explored with some success the usefulness of reinterpreting random utility models in the light of asymptotic theory of extremes. This has, in a sense, turned the usual philosophy upside-down, producing the Logit model as an *aggregate* rather than a disaggregate result. An unsuspected robustness of the Logit formula has also been found, since the assumptions on the individual distributions required to derive Theorem 4.1 are by far weaker than the ones found in the literature.

Another point of departure from the traditional approach is the use of a dynamic search formulation, rather than the usual static choice one. The stochastic process of search used in the proof of Theorem 4.1 is an interesting result in itself, and can be considered as the simplest prototype of a family of such models, which needs further exploration in the future.

The effectiveness of the method outlined in the paper (combining asymptotic theory of extremes and stochastic search processes) has thus been shown, although its power has yet to be explored in full. Natural further steps in future research suggest themselves, such as considering forms of dependency in the sequence of random terms and learning mechanisms changing the knowledge of the choice set and the dispersion of the distribution during the search.

REFERENCES

Ben-Akiva, M., and S.R. Lerman (1979) Disaggregate Travel and Mobility Choice Models and Measures of Accessibility. In *Behavioral Travel Modeling* edited by D. Hensher and P. Stopher. London: Croom Helm.

Coelho, J.D. (1980) *Optimizacão, interaccão espacial e teoria do comportamento*. Nota N.5. Lisbon, Portugal: Centro de Estatistica e Aplicacões, Dept. Matemática Aplicada, Facultade de Ciências de Lisboa.

Daly, A.J. (1979) Some Developments in Transport Demand Modeling. Pages 319-333 in *Behavioral Travel Modeling* edited by D. Hensher and P. Stopher. London:Croom Helm.

Domencich, T., and D. McFadden (1975) *Urban Travel Demand: A Behavioral Analysis*. Amsterdam: North Holland.

Dwass, M. (1964) Extremal Processes. *Annals of Mathematical Statistics* 35:1718-1725.

Fisher, R.A., and L.H.C. Tippet (1928) Limiting Forms of the Frequency Distributions of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society* 24:180-190.

Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*. New York: John Wiley & Sons.

Gnedenko, B.V. (1943) Sur la distribution limite du terme maximum d'une série aléatorie. *Annalia Mathematica* 44:423-453.

Haan, L., de (1970) On Regular Variation and its Application to the Weak Convergence of Sample Extremes. *Mathematical Centre Tracts* 32 (Amsterdam).

Hotelling, H. (1938) The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates. *Econometrica* 6:242-269.

Juncosa, M.L. (1949) On the Distribution of the Minimum in a Sequence of Mutually Independent Random Variables. *Duke Mathematical Journal* 16:609-618.

Lamperti, J. (1964) On Extreme Order Statistics. *Annals of Mathematical Statistics* 35:1726-1737.

Leonardi, G. (1981) *The Use of Random-Utility Theory in Building Location-Allocation Models.* WP-81-28. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Manski, C.F. (1973) *Structure of Random Utility Models.* Ph.D. dissertation. Cambridge, Mass.: Massachusetts Institute of Technology, Department of Economics.

Mejzler, D.G. (1950) On the Limit Distribution of the Maximal Term of a Variational Series. *Dopovidi Akad. Hauk. Ukrain SSR1:*3-10 (in Ukrainian).

Mises, R., von (1936) La distribution de la plus grande de n valeurs. Reprinted in *Selected Papers II, American Mathematical Society*, Providence, R.I. 1954:271-294.

Neuburger, H.L.I. (1971) User Benefit in the Evaluation of Transport and Land Use Plans. *Journal of Transport Economics and Policy* 5:52-75.

Shorrock, R.W. (1974) On Discrete Time Extremal Processes. *Advances in Applied Probability* 6:580-592.

Smith, T.E. (1982) *Random Utility Models of Spatial Choice: A Structural Analysis.* Paper presented at the Workshop on Spatial Choice Models, April, 1982, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Williams, H.C.W.L. (1977) On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A* 9:285-344.