Mathematical Statistics
Stockholm University

Extremal Behaviour, Weak Convergence and Argmax Theory for a Class of Non-Stationary Marked Point Processes

Research Report 2013:5

ISSN 1650-0377

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report ,
http://www.math.su.se/matstat

Extremal Behaviour, Weak Convergence and Argmax Theory for a Class of Non-Stationary Marked Point Processes

Hannes Malmberg*
Ola Hössjer†

June 27, 2013

Abstract

We formulate a random utility model where we choose from n options $1, \cdots, n$. The options have associated independent and identically distributed (i.i.d) random variables $\{X_i, U_i\}_{i=1}^n$, where X_i are the characteristics of option i and U_i is its associated utility.

We use the connection between point processes and extreme value theory to analyze the statistical properties of choice characteristics X of the object with the highest utility as $n \to \infty$. We derive analytic expressions of the asymptotic distribution of choice characteristics for a range of distributional assumptions on the utilities U_i.

In our discussion section, we suggest an extension of our method to allow us to further relax our distributional assumptions. We also show how our theoretical model can be used to explain empirical patterns relating to commuting time distributions.

*Department of Mathematics, Div. of Mathematical Statistics, Stockholm University.
†Department of Mathematics, Div. of Mathematical Statistics, Stockholm University.

1 Introduction

This paper deals with statistical models of choice behavior. First, it shows that if we model choice characteristics as stochastic, random utility models from economics can be understood in terms of the statistical theory of concomitant statistics of extremes. Secondly, the paper makes a contribution to the analysis of the concomitant of extremes by showing how point process theory can be used to derive tractable results for a range of distributional assumptions.

The economic and psychological theory of choice was initiated by Luce (1959) who posited a number of axioms for probabilistic choice, from which he derived the logit model for choice probabilities. The axiomatic approach was later partially subsumed under an approach based on utility maximization with unobservable characteristics/preferences (McFadden, 1980). In this literature, subjects are assumed to value choice options according to

$$U_i = h(x_i) + \varepsilon_i, \tag{1}$$

where x_i is a vector of (non-random) choice characteristics of option $i = 1, \cdots, n_0$. It can be shown that in this model, the probability of selecting alternative i is $e^{h(x_i)} / \sum_{j=1}^{n_0} e^{h(x_j)}$ if the ε_i's are Gumbel distributed. This approach is called the random utility approach to probabilistic choice and has been extended to more functional forms, distributional assumptions and applications since McFadden's initial contribution (Ben-Akiva and Lerman, 1985, Anderson et al., 1992, Train, 2009).

Mathematically, random utility theory is closely related to the theory of concomitants of extreme order statistics (David and Galambos, 1974, Nagaraja and David, 1994, Ledford and Tawn, 1998). This theory deals with the asymptotic behavior of the object

$$X_{[n:n]} = X_{I_n}$$

where $(X_1, U_1), \cdots (X_n, U_n)$ is a sequence of i.i.d. random variables where the U_i's are real-valued, the X_i's belong to a general space, and $I_n = \arg\max_{1 \leq i \leq n} U_i$. The main difference difference from (1) is that not only U_i, but also X_i, is random.

In this paper we use the theory of point process to analyze concomitant extreme order statistics. For more information on point processes in general, see for example Cox and Isham (1980) and Jacobsen (2005). In particular, we modify the methodologies presented in Resnick (2007), where the general connection between point processes and extreme value theory is analyzed.

We focus on the problem for a range of specifications of the distribution of $U|X = x$. We treat the problem when $X_{[n:n]}$ converges weakly to a specified non-degenerate distribution (not all mass at one point mass or infinity), but we also show in the discussion section how we can use the same theory to analyze the convergence rates to different types of degenerate distributions.

By extending the theory of concomitants, the results in this paper provide a framework for looking at random utility models in the limiting case when the number of alternatives tends to infinity.

The paper is similar in aim to Malmberg (2012) and Malmberg and Hössjer (2012). However, those papers used asymptotic properties of deterministic point process, and analyzed the problem using continuity properties of random fields. The novel approach in this paper is to instead use point process theory to analyze the question, and this method turns out to allow a simplification of the theory compared to our previous papers.

2 Model

Consider a sequence of independent and identically distributed pairs of random variables $\{(X_i, U_i)\}_{i=1}^{\infty}$, where $X_i \in \Omega \subseteq \mathbb{R}^d$ and $U_i \in \mathbb{R}$. We define $U_{n:i}$ as the i^{th} order statistic of $\{U_1, \cdots, U_n\}$. For each n, we define the location $X_{[n:i]}$ to be the X-value associated with $U_{n:i}$ for a sample of size n.

As mentioned in Section 1, we can think of U_i as the utility of alternative i and X_i as its observable characteristics. We are interested in the limiting properties of the optimal choice, and thus we study the asymptotic behavior of the sequence of probability measures

$$C_n(\cdot) = P\left(X_{[n:n]} \in \cdot\right). \tag{2}$$

We will represent the distribution of (X, Y) as

$$P((X, U) \in A \times B) = \int_A \mu(x; B) d\Lambda(x),$$

where $F_X = \Lambda$ is the marginal distribution of X over Ω, and $\mu(x; \cdot)$ is the conditional probability measure of U_i given $X_i = x$. We make the following assumption on μ:

Assumption 1 *For the collection $\mu = \{\mu(x;\cdot); x \in \Omega\}$, there exists a function*

$$p : \Omega \to (0, \infty),$$

and a one-dimensional family of probability measures $\{Q(s;\cdot) : s \in \mathbb{R}, s > 0\}$ with such that

$$\mu(x;\cdot) = Q(p(x);\cdot).$$

$Q(s;\cdot)$ is monotonic, i.e. $Q(t;\cdot)$ stochastically dominates $Q(s;\cdot)$ whenever $t > s$. Furthermore, there exist sequences a_n, b_n, independent of x, and a distribution function G_α with $\alpha \in \mathbb{R}$, such that

$$Q(s;(-\infty, \frac{u - b_n}{a_n}))^n \to G_\alpha(u)^s \tag{3}$$

as $n \to \infty$, where G_α is a distribution function of one of the following three forms:

$$G_\alpha(u) = \begin{cases} \exp(-(-u)^{-\alpha})^{\{I(u<0)\}}, & \alpha < 0, \\ \exp(-\exp(-u)), & \alpha = 0, \\ I(u > 0)\exp(-u^{-\alpha}), & \alpha > 0, \end{cases}$$

and $I(\cdot)$ is the indicator function.

In effect, our assumption is an assertion that all $\mu(x;\cdot)$ belong to the same extreme value family α, and that their relative size can be described by the one dimensional parameter $p(x)$.

2.1 Method

The sequence $\{(X_i, U_i)\}_{i=1}^n$ may be viewed as a random collection of points in $\Omega \times \mathbb{R}$, and described as a sequence of point processes ξ_n. We will show that after a suitable transformation, this sequence of point processes ξ_n converges to a Poisson point process ξ in a sense which will be specified later. As

$$C_n(A) = P(X_{[n:n]} \in A) = P\left(\sup_{i:X_i \in A} U_i > \sup_{i:X_i \notin A} U_i\right)$$

is a functional on our point process ξ_n, the problem reduces to determining whether this functional is continuous. In this case, we can use the limiting point process ξ to calculate our results.

We will start with an introduction to point processes – in particular sufficient conditions for convergence. After this, we will apply the point process machinery to our setup, and characterize the limit of our point process. Once this is done, we will define random fields taking point processes as inputs, and derive the asymptotic behavior of C_n from continuity properties of these random fields.

3 Extremal Point Process Convergence

3.1 Background on Point Processes and Convergence Results

This section contains background results and a notational machinery for point processes. See Chapter 3 of Resnick (2007) for a more detailed treatment.

Throughout this discussion, the generic point process will take values in a set E, with an associated σ-algebra $\mathcal{E}$. For the purpose of our discussion, we will take E to be a subset of a $d+1$-dimensional Euclidean space with the associated Borel σ-algebra $\mathcal{B}(\Omega)$. A *point mass* is a set function, defined by

$$\delta_z(F) = \begin{cases} 1 & \text{if } z \in F \\ 0 & \text{if } z \notin F \end{cases},$$

where $F \subseteq E$, $F \in \mathcal{E}$. A *point measure* is a measure $m(\cdot)$ such that there exists a countable collections of points $\{z_k\}$ and numbers $\{w_k\} \geq 0$, such that

$$m(\cdot) = \sum_{z_k} w_k \delta_{z_k}(\cdot).$$

We will confine our attention to the case $w_k \equiv 1$.

Let $\mathcal{M}_P(E)$ be the set of point measures on E, and let it have the minimal σ-algebra which makes

$$\{m \in \mathcal{M}_P(E) : m(F) \in B\}$$

measurable for all $F \in \mathcal{E}$, $B \subseteq \mathcal{B}(R)$ where $m(F)$ is the point measure m evaluated at the set F and $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. We define a *point process* to be a probability distribution over $\mathcal{M}_P(E)$.

If N is an arbitrary point process, we define the Laplace transform ψ associated with N as

$$\psi_N(f) = E \exp\left\{-\sum_{z\in N} f(z)\right\} = \int_{N' \in \mathcal{M}_P^+(E)} \exp\left\{-\sum_{z \in N'} .f(z)\right\} dP.$$

Here P is a probability measure over the set $\mathcal{M}_P(E)$. Moreover, the class of functions f for which we are interested in ψ_N is usually the continuous non-negative functions on E with a compact support. We write $C_K^+(E)$ to denote this set.

Definition 1 *If we have a sequence of point processes N_n, $n \geq 0$, we say that N_n converges weakly to N_0, written $N_n \Rightarrow_p N_0$, if*

$$\psi_{N_n}(f) \to \psi_{N_0}(f)$$

for all $f \in C_K^+(E)$.

We use the notation $\Longrightarrow$ for weak convergence of vector valued random variables in Euclidean space, in contrast to $\Rightarrow_p$ for point process convergence.

Definition 2 *Let X be a metric space. We call $F \subseteq X$ relatively compact if its closure $\bar{F}$ in X is compact.*

Definition 3 *Let μ be a measure on a metric space X. We say that a sequence of measures μ_n converges vaguely to μ, written*

$$\mu_n \Rightarrow_v \mu$$

if

$$\mu_n(F) \to \mu(F)$$

for all relatively compact F with $\mu(\partial F) = 0$, where ∂F is the boundary of the set F.

Definition 4 *A Poisson random measure N on E with intensity measure μ is a point process with Laplace functional*

$$\Psi_N(f) = e^{-\int_E \left(1-e^{-f(x)}\right) d\mu(x)}$$

The following two results are known from point process theory (see, for example Resnick, 2007).

Proposition 1 *Definition 4 uniquely defines a point process N. This point process has the property that for any $F \in \mathcal{E}$, and any non-negative integer k, we have*

$$P(N(F) = k) = \begin{cases} e^{-\mu(F)}(\mu(F))^k/k! & \text{if } \mu(F) < \infty \\ 0 & \text{if } \mu(F) = \infty \end{cases},$$

and that for any $k \geq 1$, if $F_1, \cdots, F_k$ are mutually disjoint sets in $\mathcal{E}$, then $\{N(F_i)\}$ are independent random variables.

Proposition 2 *For each n, suppose $\{Z_{n,j} : 1 \leq j \leq n\}$ are i.i.d. random variables and that*

$$nP(Z_{n,1} \in \cdot) \Rightarrow_v \mu.$$

Then

$$N_n = \sum_{i=1}^{n} \delta_{Z_{n,j}} \Rightarrow_p N$$

where N is a Poisson random measure on E with intensity μ.

3.2 Point Process Convergence in our Setup

We will consider a sequence of transformations

$$g_n(u) = (u - b_n)/a_n$$

of offer values, where g_n is chosen to ensure extreme value convergence for all x as in Assumption 1.

Let $\delta_{(x,u)}$ denote a one point distribution at (x, u) and define the extremal marked point process (cf. Resnick 2007)

$$\xi_n = \sum_{i=1}^{n} \delta_{(X_i, g_n(U_i))} \tag{4}$$

for a sample of size n. This is a random measure on $(\Omega \times \mathbb{R}, \mathcal{B}(\Omega \times \mathbb{R}))$.

Before stating our theorem, we prove a preliminary lemma on boundary sets of product spaces.

Lemma 1 *If $(X \times U, \Lambda \times \nu)$ is a product measure space of two metric spaces, and if $F \subseteq X \times U$ satisfies*

$$(\Lambda \times \nu)(\partial F) = 0,$$

then

$$\nu(\partial F_x) = 0 \quad \Lambda - a.e.$$

where $F_x = \{u \in U : (x, u) \in F\}$ is the cross-section of F at the point x.

Proof. We note that if we write

$$X \times \partial F_X = \{(x,u) \in X \times U : u \in \partial F_x\},$$

we have

$$X \times \partial F_X \subseteq \partial F$$

(as each ball around a point $(x,u) \in X \times \partial F_X$ contains both a point within and outside F). Thus, as

$$(\Lambda \times \nu)(X \times \partial F_X) = \int_X \nu(\partial F_x) d\Lambda(x) \leq \Lambda(\partial F) = 0$$

we get that $\nu(\delta F_x) = 0$ Λ-almost everywhere. □

We can now formulate our main result:

Theorem 1 *Let G_α and p be as in Assumption 1. Suppose that the image of every compact set under $p : \Omega \to (0, \infty)$ is bounded. Then, as $n \to \infty$, it holds that*

$$\xi_n \Rightarrow_p \xi,$$

where ξ_n is given by (4), and ξ is a Poisson Random Measure on $(\Omega \times \mathbb{R}, \mathcal{B}(\Omega \times \mathbb{R}))$ with mean intensity $\Lambda_p \times \nu_\alpha$, where

$$\Lambda_p(A) = \int_A p(x)\Lambda(dx)$$

for all $A \in \mathcal{B}(\Omega)$ and

$$\nu_\alpha([u,\infty)) = -\log(G_\alpha(u)) = \begin{cases} I(u<0)(-u)^{-\alpha}, & \text{if } \alpha < 0 \text{ and } u < 0\ , \\ \exp(-u), & \text{if } \alpha = 0, \\ u^{-\alpha}, & \text{if } \alpha > 0 \ \text{ and } u > 0. \end{cases}$$

This theorem is similar to Proposition 3.21 in Resnick's book. The difference is that he considers a sequence of point processes $\xi_n = \sum_{j=1} \delta_{(jn^{-1}, g_n(X_j))}$ where $\{X_j\}$ is a sequence of independent and identically distributed random variables. Thus, the difference is that we model the first coordinate as a random variable, and let the distribution of the second coordinate depend on this first coordinate. This creates some technical issues, which however turn out not to affect the main result.

Proof. Before starting, we note that we have $G_\alpha(u) = 0$ for $\alpha > 0$ and $u \leq 0$. Whenever $\alpha > 0$, it is implicit in the proof that $u > 0$. Using the proof of Proposition 2, it suffices to show that

$$nP((X_1, g_n(U_1)) \in \cdot) \Rightarrow_v \Lambda_p \times \nu_\alpha,$$

i.e. that

$$nP((X_1, g_n(U_1)) \in F) \to (\Lambda_p \times \nu_\alpha)(F),$$

for all $F \subseteq \Omega \times \mathbb{R}$ which are relatively compact sets with respect to $\mathcal{B}(\Omega \times \mathbb{R})$ and satisfy

$$(\Lambda_p \times \nu_\alpha)(\partial F) = 0.$$

Henceforth, let F be an arbitrary set with these properties. Now, we note that

$$nP((X_1, g_n(U_1)) \in F) = \int_\Omega nP(g_n(U_1) \in F_x | X_1 = x) d\Lambda(x),$$

where F_x is the x-cross section of F. Thus, our task is to show that

$$\int_\Omega nP(g_n(U_1) \in F_x | X_1 = x) d\Lambda(x) \to \int_\Omega p(x) \nu_\alpha(F_x) d\Lambda(x).$$

We do this first by showing that the integrand converges almost everywhere to the desired quantity, and then we show that the sequence of integrands satisfy regularity conditions allowing us to infer convergence of integrals from pointwise convergence.

We observe that for every x,

$$nP(g_n(Y_1) \in \cdot | X_1 = x) \Rightarrow_v p(x)\nu_\alpha(\cdot). \tag{5}$$

Indeed, it is true that if

$$x_n^n \to a, \tag{6}$$

we have

$$n(1 - x_n) \to -\log(a). \tag{7}$$

Thus, by the reasoning above and Assumption (1), we have

$$nP(g_n(U_1) \geq u | X_1 = x) \to -p(x) \log(G_\alpha(u)) = p(x)\nu_\alpha([u, \infty)). \tag{8}$$

In order to deduce (5) from (8), we can note that if we have a measure γ with

$$\gamma([u, \infty)) < +\infty$$

for some u, then vague convergence of γ_n to γ is equivalent to

$$\gamma_n([u,\infty)) \to \gamma([u,\infty)), \tag{9}$$

for all u such that $\gamma(\{u\}) = 0$. This can be seen by noting that if (9) is true, then the sequence $P_{nu}(\cdot) = \gamma_n(\cdot \cap [u,\infty))/\gamma_n([u,\infty))$ of probability measures converges weakly for all continuity points u of $\gamma([u,\infty))$ to $P_u(\cdot) = \gamma(\cdot \cap [u,\infty))/\gamma[u,\infty))$, and hence $P_{nu}(F) \to P_u(F)$ for all such u, from which (5) follows.

Now, using Lemma 1, we know that

$$\nu_\alpha(\partial F_x) = 0 \quad \Lambda_p - a.e.$$

which means that

$$p(x)\nu_\alpha(\partial F_x) = 0 \quad \Lambda_p - a.e$$

as $p(x) > 0$ implies that $p(x)\nu_\alpha$ and ν_α are equivalent for all $x \in \Omega$. Thus, we can use (5) to conclude that

$$nP(g_n(U_1) \in F_x | X_1 = x) \to p(x)\nu_\alpha(F_x) \quad \Lambda_p - a.e.$$

Therefore, we have established pointwise convergence of the integrand almost everywhere.

Now, we seek to show that $nP(g_n(U_1) \in F_x | X_1 = x)$ is uniformly bounded over n and Ω to ensure that pointwise convergence almost everywhere implies convergence in integrals. To do so, we try to define a maximal random variable which dominates $nP(g_n(U_1) \in F_x | X_1 = x)$ for all n and x.

This works as $p(x)$ indexes the distributions by stochastic dominance. We write

$$\pi_\Omega : (x,u) \mapsto x$$

and

$$\pi_U : (x,u) \mapsto u$$

for the projection on Ω and $\mathbb{R}$ respectively. In this case, we know that $\pi_\Omega(F)$ and $\pi_U(F)$ are relatively compact sets of Ω and $\mathbb{R}$ respectively, and we define

$$\bar{p} = \sup_{x \in \pi_\Omega(F)} p(x).$$

We can now define the maximum random variable as having the law

$$\bar{U}(F) \sim Q_\alpha(\bar{p}; \cdot).$$

By the monotonicity assumption of Q_α made in Assumption (1), we know that $\bar{U}(F)$ stochastically dominates $U_1|X_1 = x$ for all $x \in \pi_\Omega(F)$.

Furthermore, we can define $\underline{\mathrm{u}}$ as the smallest u-value attained on the whole set $\pi_U(F)$, which again is finite by the assumption of F being relatively compact. Combining these two definitions gives us

$$\begin{array}{rcl} nP(g_n(U_1) \in F_x | X_1 = x) & \leq & nP(g_n(U_1) \geq \underline{\mathrm{u}} | X_1 = x) \\ & \leq & nP(g_n(\bar{U}(F)) \geq \underline{\mathrm{u}} | X_1 = x) \\ & = & nP(g_n(\bar{U}(F)) \geq \underline{\mathrm{u}}) \\ & \rightarrow & \max_{x \in P_x(F)} p(x)\nu_\alpha([\underline{\mathrm{u}}, \infty)) \\ & < & +\infty \end{array}$$

which means that $nP(g_n(U_1) \in F_x | X_1 = x)$ is uniformly bounded. Using the bounded convergence theorem, we get

$$\begin{array}{rcl} nP((X_1, g_n(U_1)) \in F) & = & \int_\Omega nP(g_n(U_1) \in F_x | X_1 = x) d\Lambda(x) \\ & \rightarrow & \int_\Omega \nu_\alpha(F_x) p(x) d\Lambda(x) \\ & = & (\Lambda_p \times \nu_\alpha)(F) \end{array}$$

which completes the proof. □

4 Convergence of Functionals of Random Fields

Recall that our task is to study the limiting behavior of C_n as defined in (2). The key to connect this limit to point processes is the observation that as as g_n is strictly increasing for all n, we have:

$$\begin{array}{rcl} C_n(A) & = & P(X_{[n:n]} \in A) \\ & = & P(M_{\xi_n}(A) > M_{\xi_n}(A)) \end{array}$$

where M_{ξ_n} is the random field defined as

$$M_{\xi_n}(A) = \max_{\substack{X_i \in A \\ 1 \leq i \leq n}} g_n(U_i), \qquad A \in \mathcal{B}(\Omega),$$

where $\mathcal{B}(\Omega)$ is the Borel sigma algebra over Ω, and ξ_n is the point process from (4). This formulation of the argmax-measure C_n in terms of random fields defined over point processes allows us to generalize the notion of argmax to the limiting case where the number of offers goes to infinity. We will study the limiting behaviour of finite dimensional distributions of M_{ξ_n} and this will allow us to calculate the limit of C_n.

Write

$$\xi = \sum_{i=1}^{\infty} \delta_{(X_i^\infty, U_i^\infty)}$$

for a realization of the limiting Poisson point process ξ derived in Theorem 1. We then can define,

$$M_\xi(A) = \max_{i; X_i^\infty \in A} Y_i^\infty,$$

and

$$C(A) = P(M_\xi(A) > M_\xi(A^c)).$$

Proposition 3 *If $\Lambda_p(\Omega) < \infty$, we have*

$$C(A) = \Lambda_p(A)/\Lambda_p(\Omega).$$

Proof: Suppose first that $\Lambda_p(A^c) = 0$ or $\Lambda_p(A) = 0$. In this case, it is clear that we have $C(A) = 1$ or $C(A) = 0$ respectively as required by the formula for $A \in \mathcal{B}(\Omega)$. Indeed, using the convention that the supremum of an empty set is minus infinity, if $\Lambda_p(A) = 0$, then $M_\xi(A) = -\infty$ almost surely. As $M_\xi(A^c) > -\infty$ almost surely, we will get $C(A) = 0$. A similar reasoning applies to A^c.

Furthermore, since ξ is a Poisson random measure with mean measure $\Lambda_p \times \nu_\alpha$, we note that if $\Lambda_p(\Omega) < \infty$ we have that $M_\xi(A)$ and $M_\xi(A^c)$ are two independent, proper random variables with

$$P(M_\xi(A) \le y) = P(\xi(A \times [y, \infty)) = 0) = e^{-\Lambda_p(A)\nu_\alpha([y,\infty))} \tag{10}$$

$$P(M_\xi(A^c) \le y) = P(\xi(A^c \times [y, \infty)) = 0) = e^{-\Lambda_p(A^c)\nu_\alpha([y,\infty))}. \tag{11}$$

Using standard results from proportional hazards theory (Cox and Oakes, 1984, Fleming and Harrington, 1991), we get that

$$P(M_\xi(A) > M_\xi(A^c)) = \frac{\Lambda_p(A)}{\Lambda_p(A) + \Lambda_p(A^c)} = \Lambda_p(A)/\Lambda_p(\Omega)$$

and our proof is complete. □

From this result, we automatically get that C is a probability measure as it is a normalized version of Λ_p which is a finite measure.

In order to prove weak convegence of C_n, we need some additional results and notation. We will use that

$$\nu_1 \ll \mu_1 \text{ and } \nu_2 << \mu_2 \Rightarrow \nu_1 \times \nu_2 \ll \mu_1 \times \mu_2, \tag{12}$$

where $\ll$ means "absolutely continuous with respect to".

We will also use that if ξ_n are point processes, ξ is a Poisson process, and

$$\xi_n \Rightarrow_p \xi,$$

then

$$P(\xi_n(F) = 0) \to P(\xi(F) = 0) \tag{13}$$

for all $F \in \mathcal{E}$ with $\mu(\partial F) = 0$, where μ is the intensity measure of ξ.

Lastly, we recall that if S_n is any sequence of random variables taking values in $\mathbb{R}^k$, we have that $S_n \Rightarrow S$ if and only if

$$G_{S_n}(s_1, \cdots, s_k) \to G_S(s_1, \cdots, s_k) \tag{14}$$

for all points of continuity of G_S where G_S denotes the distribution function of the random variable S.

Theorem 2 *If* $\Lambda_p(\Omega) < \infty$, *we have*

$$C_n(\cdot) \Rightarrow C(\cdot) = \frac{\Lambda_p(\cdot)}{\Lambda_p(\Omega)}. \tag{15}$$

Proof: Assume we have A with $C(\partial A) = 0$. We aim to prove that $C_n(A) \to C(A)$. By Proposition 3, C and Λ_p are equivalent, and we have $\Lambda_p(\partial A) = 0$. Noting that the result is clearly true whenever $\Lambda_p(A) = 0$ or $\Lambda_p(A^c) = 0$, we can assume that both are different from 0. By (10) and (11), this means that $(M_\xi(A), M_\xi(A^c))$ is a proper random variable on $\mathbb{R}^2$, and we will show that $(M_{\xi_n}(A), M_{\xi_n}(A^c))$ jointly converge weakly to this random variable. Indeed, consider

$$\begin{aligned} P(M_{\xi_n}(A) \leq x_1, M_{\xi_n}(A^c) \leq x_2) &= P\left(\xi_n(A \times (x_1, \infty) \cup A^c \times (x_2, \infty)) = 0\right) \\ &\to P\left(\xi(A \times (x_1, \infty) \cup A^c \times (x_2, \infty)) = 0\right) \\ &= P(M_\xi(A) \leq x_1, M_\xi(A^c) \leq x_2) \\ &= F_{M_\xi(A), M_\xi(A^c)}(x_1, x_2). \end{aligned}$$

The convergence step uses (13) and that

$$\partial\left(A \times (x_1, \infty) \cup A^c \times (x_2, \infty)\right) \subset \partial A \times (\min(x_1, x_2), \infty) = F$$

and we have $(\Lambda_p \times \nu_\alpha)(F) = 0$ as $\Lambda_p(\partial A) = 0$, where $\Lambda_p \times \nu_\alpha$ is the intensity measure of ξ.

Hence, it follows from (14) that

$$(M_{\xi_n}(A), M_{\xi_n}(A^c)) \Rightarrow (M_\xi(A), M_\xi(A^c)).$$

Defining

$$D = \{(a,b) \in \mathbb{R}^2 : a > b\}$$

and using (12) , with $\nu_1 \sim M_\xi(A)$, $\nu_2 \sim M_\xi(A^c)$, and μ_1, μ_2 Lebesgue measure in $\mathbb{R}$, to conclude that.

$$P((M_\xi(A)M_\xi(A^c)) \in \partial D) = 0$$

we get

$$\begin{aligned} C_n(A) &= P(M_{\xi_n}(A) > M_{\xi_n}(A^c)) \\ &= P((M_{\xi_n}(A), M_{\xi_n}(A^c)) \in D) \\ &\to P((M_\xi(A)M_\xi(A^c)) \in D) \\ &= C(A) \end{aligned}$$

and the proof is complete. □

5 Examples

Here we provide some examples to illustrate our theory.

Example 1 (Exponential and mixture models.) A class of distributions that satisfy Assumption 1 are

$$\mu_\alpha(x;\cdot) \sim \begin{cases} P\left((2 \times 1_{\{V_1 < p(x)\}} - 1)(1 - V_2^{-1/\alpha}) \in \cdot\right) & \alpha < 0, \\ P\left(\log(p(x)/V_1 \in \cdot\right), & \alpha = 0, \\ P\left((2 \times 1_{\{V_1 < p(x)\}} - 1)V_2^{-1/\alpha} \in \cdot\right), & \alpha > 0, \end{cases}$$

where $V_1, V_2 \sim U(0,1)$ are two independent and uniformly distributed random variables on $(0,1)$. A bit less formal, we may write

$$\mu_\alpha(x) \sim \begin{cases} -(1-p(x))\text{Beta}(1,-\alpha) + p(x)\text{Beta}(1,-\alpha) & \alpha < 0, \\ \text{Exp}(\log(p(x)), 1), & \alpha = 0, \\ -(1-p(x))\text{Pareto}(\alpha,1) + p(x)\text{Pareto}(\alpha,1), & \alpha > 0, \end{cases}$$

where $\text{Beta}(a,b)$ refers to a Beta distribution with density $Cx^{a-1}(1-x)^{b-1}$ on $(0,1)$, $\text{Exp}(a,b)$ is a shifted exponential distribution with location paramter a and scale parameter b, having distribution function $1 - e^{-(x-a)/b}$ for $x \geq$

a, Pareto(α, b) is a Pareto distribution with shape parameter α and scale parameter b, corresponding to a distribution function $1 - (x/b)^{-\alpha}$ for $x \geq b$. We have chosen the parameter α for the distributions μ_α in a way so that they lie in the domain of convergence of G_α in (1).

Example 2 (An example from the commuting literature) Focusing on $\alpha = 0$ in the previous example, we have an interesting special case. Suppose that the population is distributed uniformly on $B(0, R)$, a disk in $\mathbb{R}^2$. The utility associated with each point is

$$U|X = x \sim \text{Exp}(-c||x||, 1),$$

where $||x||$ is the Euclidean distance from the origin. This is a good benchmark model for commuting choices. In this case, Λ has a uniform distribution on $B(0, R)$, and $p(x) = \text{Exp}(-c||x||)$. Thus, we get

$$C(A) = \frac{\int_A e^{-c||x||} dx}{\int_{B(0,R)} e^{-c||x||} dx}.$$

The particular direction of commuting is often not as interesting as the distribution of distances. The probability that we commute less than r is given by

$$C(\{x : ||x|| \leq r) = \frac{\int_0^r s e^{-cs} ds}{\int_0^R s e^{-cs} ds},$$

which we recognize as a truncated Gamma$(1, 1)$-distribution.

There is suggestive evidence that commuting patterns follow a gamma distribution over short distances. We provide an example in Figure 1 with a histogram over commuting distances with a super-imposed gamma distribution with parameters provided by moment fitting. The moment-fitted density provides a reasonable fit for the left half of the data.

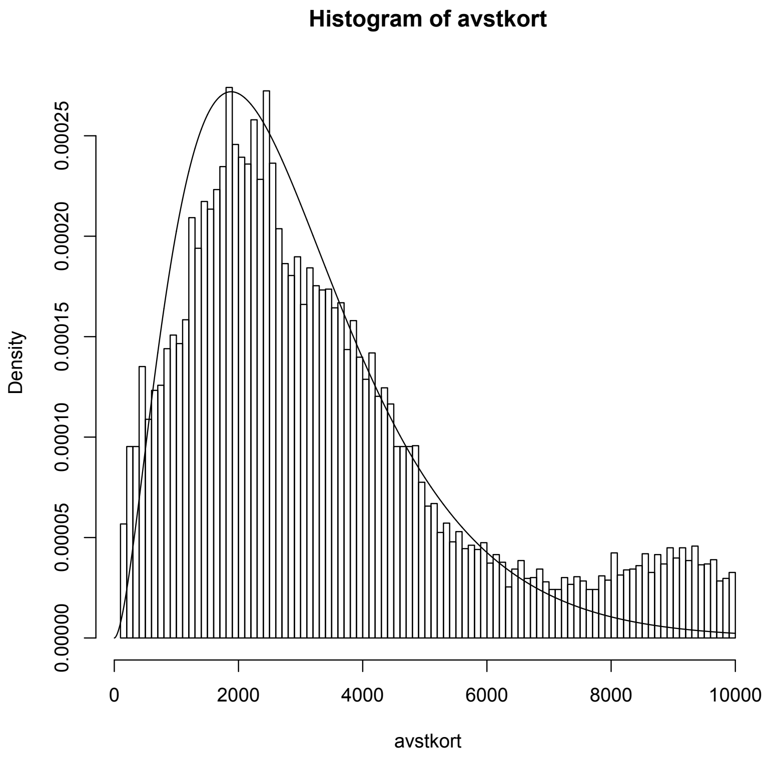
Example 3 (The logit model: a special case) Let Λ be uniformly distributed on the finite support $\{x_1, .., x_{n_0}\}$. Let utilities be given by

$$U_j|X_j = \text{Exp}(-c\|X_j\|, 1) \tag{16}$$

This corresponds to $p(x_i) = e^{h(x_i)}$ and we get

$$C(\{x_i\}) = \frac{e^{h(x_i)}}{\sum_{j=1}^{n_0} e^{h(x_j)}}$$

Figure 1: Histogram over commuting distances in Kungsholmen, Stockholm



just as in the logit model (1). We interpret the offers in (1) as standardized maximal offers:

$$\max_{j:1\leq j\leq n, X_j=x_i} U_j - \log(n) \tag{17}$$

derived from (16) as $n \to \infty$. From extreme value theory, we deduce that (17) has an asymptotic Gumbel distribution plus $h(x_i)$, and this provides additional justification of (1).

6 Discussion

6.1 Mathematical extensions

We have derived a way to calculate the asymptotic behavior of $C_n = X_{[n:n]}$, and have done so for a number of assumptions on the joint distribution of (X_i, U_i). However, in order to extend our results to a wider class of distributional assumptions, we must relax our requirement that $X_{[n:n]}$ should converge to a non-degenerate distribution. For example, when X and U are distributed bivariate normally with positive correlation, $X_{[n:n]} \to \infty$ almost surely.

In these cases, it can nevertheless be possible to find a sequence of functions h_n such that

$$h_n(X_{[n:n]}) \Rightarrow S$$

for a non-degenerate random variable S. In this case, we would have

$$X_{[n:n]} \overset{d}{\approx} h_n^{-1}(S)$$

for large n, where $\overset{d}{\approx}$ means that the two random variables have approximately the same distribution.

This would extend the empirical application of our results. Of course, we will not know the exact n in practice, but if S belongs to a class of distributions invariant under n, we know which distribution class our result can be expected to belong to. Furthermore, the asymptotic behaviour of h_n can be used to assess how different moments of $X_{[n:n]}$ will develop as $n \to \infty$, thus giving us a way of predicting the effect of for example increased population density on commuting choices.

We have done some exploratory studies on this extension, and there are indications that for a much larger class of distribution than studied in the paper, it is possible to find sequences h_n and g_n such that

$$\sum_{i=1}^{n} \delta_{(h_n(X_i), g_n(U_i))} \Rightarrow_p \xi$$

for some non-degenerate Poisson process ξ. With this result, it is possible to apply analogous result to those in this paper to analyze the asymptotic behavior of $X_{[n:n]}$ more generally.

6.2 Empirical applications

In Example 2, we showed that with linear transport costs and uniform population distribution on a two-dimensional disc, the resulting distribution of the distance from the origin of the optimal choice is asymptotically a truncated gamma distribution when utilities are exponentially distributed and have a deterministic additive term.
This is agreement with an observed empirical regularity that commuting distances seem to follow a gamma distribution for short distances. The extension outlined in Section 6.1 seeks to show that this result is true not only for exponentially distributed utilities, but whenever utilities belong to the Gumbel domain of attraction (i.e. that their extreme values converge to a Gumbel distribution). If this can be shown, the empirical regularity with gamma distributed commuting distances will have a foundation in utility maximization and probabilistic choice.

References

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1992). *Discrete choice theory of product differentiation.* MIT press.

Ben-Akiva, M. and Lerman, S. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. Chapman & Hall/CRC, London.

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*, volume 21. Chapman & Hall/CRC, London.

David, H. and Galambos, J. (1974). The asymptotic theory of concomitants of order statistics. *Journal of Applied Probability*, pages 762–770.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, volume 8. Wiley Online Library, New York.

Jacobsen, M. (2005). *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhäuser, Boston.

Ledford, A. W. and Tawn, J. A. (1998). Concomitant tail behaviour for extremes. *Advances in Applied Probability*, 30(1):197–215.

Luce, R. D. (1959). *Individual Choice Behavior a Theoretical Analysis*. John Wiley and sons, New York.

Malmberg, H. (2012). Argmax over continuous indices of random variables – an approach using random fields. Master's thesis, Stockholm University.

Malmberg, H. and Hössjer, O. (2012). Argmax over continuous indices of random variables – an approach using random fields. Technical report, Division of Mathematical Statistics, Department of Mathematics, Stockholm University. Submitted.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, 53(3):13–29.

Nagaraja, H. N. and David, H. A. (1994). Distribution of the maximum of concomitants of selected order statistics. *The Annals of Statistics*, 22(1):478–494.

Resnick, S. I. (2007). *Extreme values, regular variation, and point processes*. Springer, New York.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge, 2nd edition.