Licenciate Thesis – Random Choice over a Continuous Set of Options*

Hannes Malmberg†

May 15, 2013

Abstract

Random choice theory has traditionally modeled choices over a finite number of options. This thesis generalizes the literature by studying the limiting behavior of choice models as the number of options approach a continuum.

The thesis uses the theory of random fields, extreme value theory and point processes to calculate this limiting behavior. For a number of distributional assumptions, we can give analytic expressions for the limiting probability distribution of the characteristics of the best choice. In addition, we also outline a straightforward extension to our theory which would significantly relax the distributional assumptions needed to derive analytical results.

Some examples from commuting research are discussed to illustrate potential applications of the theory.

*Filosofie licenciatavhandling. Avhandlingen kommer att presenteras onsdagen den 5/6 2013, kl 15.15 i rum 5:306, Matematiska institutionen, Stockholms universitet, Kräftriket

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

Acknowledgements

First, I would like to thank my supervisor and co-author Ola Hössjer for all enjoyable mathematical discussions at the white-board and over email which have pushed this project towards completion. With his help, I have learned mathematical research as this project has progressed from a Bachelor Thesis to a Master Thesis, and lastly to this Licenciate Thesis.

I also would like to thank my second supervisor Dmitrii Silvestrov. Dmitrii's courses Probability Theory IV and Stochastic Processes IV led to a quantum leap in my understanding of probability theory. Moreover, during my research, his thorough reading and constructive criticism have repeatedly forced me to clarify and restructure my ideas.

I would also like to thank Pieter Trapman for reading and commenting on the manuscript, as well as my very close friend Zihan Hans Liu for never failing to expand my mathematical horizons when I talk to him.

List of Papers

This thesis consists of two papers

1. MALMBERG, H., HÖSSJER, O.: Argmax over Continuous Indices of Random Variables – An Approach Using Random Fields, submitted to *Applied Probability Trust*.

2. MALMBERG, H., HÖSSJER, O.: Extremal Behaviour, Weak Convergence and Argmax Theory for a Class of Non-Stationary Marked Point Processes, submitted to *Extremes*.

In both papers, the authors collaborated on developing the general structure of the ideas, and discussed to overcome problems arising during the progress of the works. H. Malmberg developed most of the exact statements and provided the proofs. O. Hössjer read the papers a large number of times, and following this readthroughs, theoretical extensions were developed in joint discussions.

Introduction

1 Background

Imagine a person who has decided to start a new job, and who is about to choose a place of residence. Two counteracting tendencies exist. On the one hand, living close to the job is preferable as costs associated with transport increase with distance. On the other hand, the area per radial segment increases the further away you go from your workplace. There is more area between 100 m and 110 m from your job than between 0 m and 10 m. Thus, the probability of finding a good house in a given radial segment increases with distance from the job. How do these two tendencies interact to shape the statistical behavior of residential choice?

The residential choice problem belongs to a class of problems which this thesis addresses. We develop a framework for discussing questions of choice where the set of choice options is continuous, and where there is a random element in the choice process.

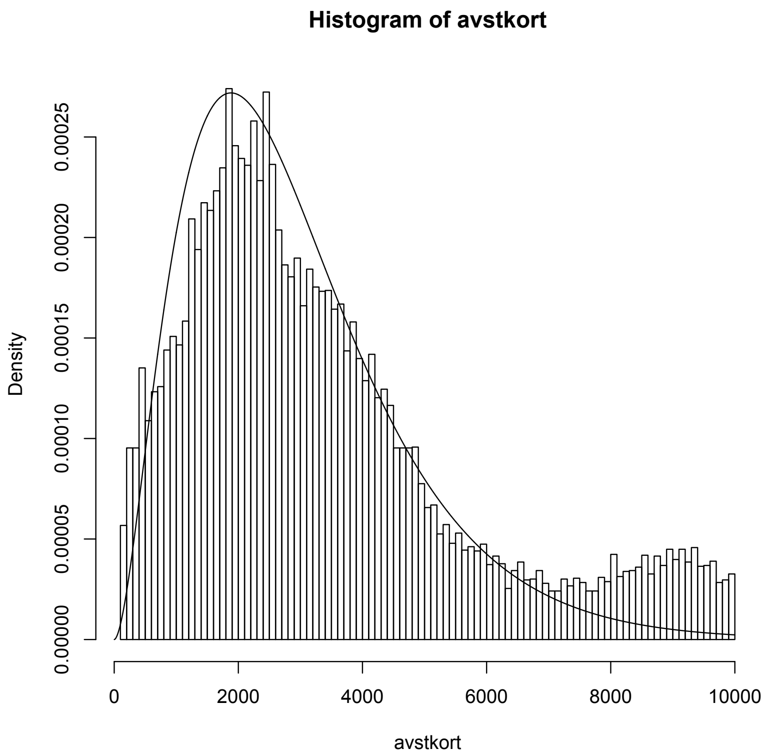
In this introduction, we will give a brief overview of motivating empirical regularities and previous theory, as well as introducing our setup and describing our results. We will conclude with a discussion on potential future developments. To allow us to focus on the mathematical intuition, some technical detail will be left out in the kappa. Interested readers are referred to the papers for formal definitions and proofs.

1.1 Empirical Motivations

When discussing how people make choices over continuous variables such as residential location, two important observations stand out.

First, people make very different choices, and all variation cannot be explained by observed individual characteristics. This suggests that a statistical approach is appropriate. Secondly, there are some statistical regularities which warrants a search for an explanatory model. Figure 1 summarizes some salient features. The figure shows commuting distances in the Swedish labor market from Kungsholmen, Stockholm. We note that the distribution is unimodal with a skew to the right. It has the property of being approximately Gamma distributed over short distances, with a somewhat thicker tail than a gamma distribution. The pattern in Figure 1 can be found in other similar applications, such as when we measure the distance traveled to school.

Figure 1: Histogram over commuting distances in Kungsholmen, Stockholm



The fact that people make different choices, but that these choices display regular features when aggregated, suggests that there is a value in attempting to develop a statistical theory to explain the underlying choice process. This is the aim of this thesis.

2 Theoretical preliminaries

2.1 Random choice theory

In our model setting, an agent makes a zero-one choice concerning every point in a continuous space – in that sense we model discrete choices over a continuum of choice options. We follow the tradition in economics and model these discrete choice problems as random. This stems from an aim to predict the *proportion* of people selecting a specific option (or collection of options in our case), which differs from traditional demand analysis where we want to explain *how much* consumers buy of a particular good.

The probabilistic theory of choice started with Luce (1959) who posited a collection of axioms 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 choices according to the expression

$$U_i = h(x_i) + \varepsilon_i \quad i = 1, \cdots, n_0, \tag{1}$$

where x_i are the (non-random) characteristics of option i. 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 independently Gumbel-distributed. Thus, we can derive logit probabilities from the assumption of utility maximization by making appropriate distributional assumptions.

This approach to probabilistic choice is called random utility theory 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).

Our thesis can be viewed as an extension of this framework in two directions. First, the x_i's are random variables in our setup. Furthermore, we let the number of choices n_0 go to infinity, and study the continuous limit of a sequence of discrete choice models.

2.2 Extreme Value Theory

Extreme value theory is a branch of mathematics studying the asymptotic properties of the sequence of random variables

$$M_n = \max_{1 \leq i \leq n} Z_i,$$

where $\{Z_i\}_{i=1}^{\infty}$ is a sequence of random variables. The foundational theorem in the literature deals with the case when the sequence of random variables Z_i are independent and identically distributed (Fisher and Tippett, 1928, Gnedenko, 1941). This theorem states that if there exist sequences of real numbers a_n and b_n such that for all y, we have

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq y\right) \to G(y), \tag{2}$$

then the function G belongs to one of the following parametric families $\mathcal{F}$ of distributions functions, with functional forms:

$$G_0(z) = \exp\left\{-\exp\left(-\left(\frac{z-b}{a}\right)\right)\right\}, \text{ for } z \in \mathbb{R} \tag{3}$$

$$G_{-\alpha}(z) = \begin{cases} 0 & z \leq b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\} & z > b \end{cases} \tag{4}$$

$$G_{\alpha}(z) = \begin{cases} \exp\left\{-\left(-\left(\frac{z-b}{a}\right)\right)^{\alpha}\right\} & z \leq b \\ 1 & z > b \end{cases} \tag{5}$$

where a, b and α are constants, of which a and α are constrained to be positive. Whereas α determines $\mathcal{F}$, a and b are the scale and location parameters of $\mathcal{F}$. These three functional forms are the Gumbel (3), Weibull (4), and Frechet (5) families respectively.

The theorem means that insofar the maximum of a collection of random variable converges after a suitable sequence of affine transformations, the resulting distribution will belong to a small class of distributions. Different distributions of Z_i will yield different limiting distributions G, and by modifying a_n and b_n it is easy to see that any combination of a and b can be attained as limit in (2) within the family $\mathcal{F}$.

The earlier extreme value theory has been extended in a number of different directions. Most similar to our project in Paper 1 is the work on relaxing the assumption of the Z_i's being identically distributed while retaining the assumption of independence (Weissman, 1975, Horowitz, 1980). Our approach in Paper 1 also connects to the study of Gumbel random fields. For a recent treatment of the subject of on Gumbel random fields, see Robert (2013).

We refer to Leadbetter et al. (1983) and Resnick (2007) for a more comprehensive treatment of extreme value theory.

2.3 Concomitants of extreme order statistics

The research area in statistics which is most closely related to our problem (and to random utility theory) is 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 $I_n = \arg\max_{1 \leq i \leq n} U_i$. A difference from (1) is that not only U_i, but also X_i, is random.

3 Description of Papers

The two papers in the thesis answer a similar question with similar results. The difference between them is the strategies they employ to perform the main step of the derivation. Thus, all but two subsections in this section will be common for both papers.

3.1 Problem formulation

For presentational clarity, we describe a single problem setup although it differs somewhat between the papers. The setup presented here is used in Paper 2, and although there are some differences to the setup used in Paper 1, these differences are sufficiently non-fundamental so that the methodology of Paper 1 can be explained using the setup in Paper 2.

There is a set $\Omega \subseteq \mathbb{R}^k$ of choice characteristics, and a distribution Λ on Ω giving the relative prevalence of different characteristics. For each characteristic $x \in \Omega$, there is a conditional probability distribution of utility

$$\mu(\cdot; x) = P(U \in \cdot | X = x).$$

We can define the bivariate distribution of characteristics and utilities on rectangular sets $A \times B$ with $A \subseteq \Omega$ and $B \subseteq \mathbb{R}$ as

$$P((X, U) \in A \times B) = \int_A \mu(B; x) d\Lambda(x). \tag{6}$$

Our basic building block will be a sequence of choice alternatives

$$(X_1, U_1), \cdots, (X_n, U_n), \tag{7}$$

which are independently and identically distributed according to the bivariate distribution given in (6). For each fixed n, we define

$$I_n = \arg\max_{1 \leq i \leq n} U_i$$

as the index of the variable with the highest utility, and $X_{I_n} = X_{[n:n]}$ as the characteristics vector of this random variable. For a fixed n, the probability distribution of the characteristics vector of the selected alternative is

$$C_n(\cdot) = P(X_{I_n} \in \cdot). \tag{8}$$

In both papers, we look for the asymptotic properties of the sequence C_n. In particular, we will look for a probability measure C on Ω such that

$$C_n \Rightarrow C, \tag{9}$$

where $\Rightarrow$ stands for weak convergence of probability measures on $(\Omega, \mathcal{B}(\Omega))$, where $\mathcal{B}(\Omega)$ is the σ-algebra of Borel subsets of Ω (for an extensive treatment of weak convergence, see Billingsley, 1971)

Later in the introduction, we will sometimes write C_n and C to denote random variables, the laws of which are given by (8) and (9). It will be clear from the context when this has been done.

3.2 Random fields

In both papers, we study the asymptotic behavior of (8) by studying an intermediate mathematical object, namely the random field M_n, which in Paper 1 is defined by

$$M_n(A) = \sup_{1 \leq i \leq n, X_i \in A} U_i \quad A \subseteq \Omega \tag{10}$$

with the convention that the supremum of the empty set is $-\infty$. We can parse this definition. This random field takes subsets A of the characteristics space Ω as arguments, and returns a real number. Supplied with the argument A, the random field returns the value of the best offer having a characteristic vector in the set $A \subseteq \Omega$.

We will define an important functional from the set of random fields to the set of probability measures on Ω by

$$F(\cdot; M) = P(M(\cdot) > M(\cdot^c)), \tag{11}$$

where $\cdot^c$ stands for the complement of $\cdot$ in Ω. Intuitively, $F(A; M)$ gives the probability that the best offer, corresponding to max field M, belongs to A.

We connect the random fields (10) and the functional (11) to our problem by observing that

$$C_n(\cdot) = P(M_n(\cdot) > M_n(\cdot^c)) = F(\cdot; M_n). \tag{12}$$

Equation (12) suggests that we can study the asymptotic behavior of C_n by studying the asymptotic behavior of M_n. Indeed, we seek to find a random field M such that M_n, or a monotone transformation of M_n, converges to M in an appropriate way. The correct sense of convergence is one which makes (11) continuous as a function from the set of random fields to the set of probability measures, where the topology on probability measures is that of weak convergence. If we can demonstrate such continuity, we can conclude that

$$F(\cdot; M_n) \Rightarrow F(\cdot; M) \tag{13}$$

on $(\Omega, \mathcal{B}(\Omega))$, whenever M_n converges to M. Combining (13) with (12), we then get

$$C_n(\cdot) \Rightarrow F(\cdot; M),$$

and insofar $F(\cdot; M)$ is known, we have characterized the limit of C_n.

The method outlined above is the approach used in both papers to solve our problem. The distinctiveness of the two papers lie in that they use different ways to find M, and different ways to demonstrate that the sense of convergence of finite sample max fields to M ensures continuity of F.

3.3 Paper 1

In Paper 1, we define a notion of convergence $\overset{m}{\to}$ on the space of random fields. We say that

$$M_n \overset{m}{\to} M$$

if there exists a sequence of strictly increasing functions g_n such that for each $A \subseteq \Omega$ satisfying some regularity conditions,

$$g_n(M_n(A)) \Rightarrow M(A).$$

In Paper 2, we show that this notion of convergence makes (11) a continuous from the set of random fields to the set of probability measures on Ω.

Having made assumptions on the bivariate distribution of (X_i, U_i), we use extreme value theory to derive M. The sequence of strictly increasing functions g_n is defined as

$$g_n(z) = \frac{z - b_n}{a_n},$$

where a_n and b_n are the normalizing sequences described in Section 2.2. This method works as the number of offers having characteristics in A increases to infinity as $n \to \infty$. Because their associated utilities are conditionally independent, we can apply extreme value theory. The main caveat lies in that traditional extreme value theory assumes that the U_i are identically distributed, whereas the distribution of the U_i's varies with characteristics in our setup. A large part of the theoretical work in Paper 1 involves dealing with this variation in the distribution of U_i.

3.4 Paper 2

In Paper 2, we place a second intermediate object between the sequence (7) and the quantity (8). In this paper, we observe that (7) can be viewed as a point process on the product space $\Omega \times \mathbb{R}$. Defining the set function

$$\delta_{(x,u)}(F) = \begin{cases} 1 & \text{if } (x,u) \in F \\ 0 & \text{if } (x,u) \notin F \end{cases},$$

with F a Borel subset of $\Omega \times \mathbb{R}$, we define our point process as

$$\xi_n(\cdot) = \sum_{i=1}^{n} \delta_{(X_i, g_n(U_i))}(\cdot). \tag{14}$$

where

$$g_n(z) = \frac{z - b_n}{a_n}$$

will be a sequence of strictly increasing functions, coinciding with the normalization from Section 2.2 for the distribution $U|X = x$, assuming that the same normalization g_n can be used for all $x \in \Omega$.

In this setup, we can define the random field

$$M_{\xi_n}(A) = \sup\{u \in \mathbb{R} : \xi_n(A \times [u, \infty)) = 0\} \quad A \subseteq \Omega, \tag{15}$$

that is, the largest number u such that the point process ξ_n has a point in the set $A \times [u, \infty)$. As the positions of points are random, $M_{\xi_n}(A)$ will be a random variable for each $A \subseteq \Omega$, and thus, M_{ξ_n} is a random field with subsets of Ω as arguments. Again, we have

$$C_n(\cdot) = P(M_{\xi_n}(\cdot) > M_{\xi_n}(\cdot^c)). \tag{16}$$

Notice that the max fields in Papers 1 and 2 are slightly different, since $M_{\xi_n}(A) = g_n(M_n(A))$, although they only differ by the monotone transformation g_n. The fact that M_n and M_{ξ_n} are related by a montone transformation also means that (16) is equivalent to the definition in (8).

In Paper 2, we study the limiting behavior of (16) by studying the limiting behavior of (14). Building on the connection between extreme value theory and point processes described in Resnick (2007), we find a Poisson process ξ such that

$$\xi_n \overset{p}{\Rightarrow} \xi,$$

where $\overset{p}{\Rightarrow}$ denotes convergence in a point process sense. We show that point process convergence implies that M_{ξ_n} converges to M_ξ, defined in (15), in such a way that

$$F(\cdot; M_{\xi_n}) \Rightarrow F(.; M_\xi).$$

and we can thus conclude that

$$C_n(\cdot) \Rightarrow P(M_\xi(\cdot) > M_\xi(\cdot^c))$$

as required. Insofar $P(M_\xi(\cdot) > M_\xi(\cdot^c))$ is easy to characterize, we have succeeded with our aim of finding the asymptotic behavior of C_n.

3.5 Examples

We apply our theory by making assumptions on the distribution of characteristics Λ and the conditional probability measures of utility $\mu(\cdot; x)$ for $x \in \Omega$. In this section, we will write $\Gamma(k, \lambda)$ to denote a gamma distribution with density function $\lambda^k x^{k-1} e^{-x\lambda}/\Gamma(k)$, where $\Gamma(\cdot)$ denotes the gamma function.

We assume that the distribution of characteristics Λ on Ω has a density function $\lambda(x)$, and that utility is given by

$$U_i = m(x_i) + \epsilon_i$$

with $m(x)$ being a regression function and $\epsilon_i \sim \Gamma(1, 1)$, so that errors are exponentially distributed. In this case, the limit

$$\lim_n C_n(\cdot),$$

in the sense of weak convergence, has a density function

$$f(x) = \frac{e^{m(x)}\lambda(x)}{\int_\Omega e^{m(y)}\lambda(y)dy} \quad x \in \Omega.$$

This means that the process of taking the best choice in the limit has the effect of exponentially tilting the initial distribution with $e^{m(x)}$, thereby attaching more weight to points with high deterministic utility component.

In particular, let us return to our example in Section 1 on residential location choice, with $\Omega = \mathbb{R}^2$, $x = (x_1, x_2)$ and linear cost, that is

$$m(x) = -c||x||,$$

where $||\cdot||$ is the Euclidean distance from the origin. In this case, we get the choice density

$$\frac{e^{-c||x||}\lambda(x)}{\int_{\mathbb{R}^2} e^{-c||s||}\lambda(s)ds}.$$

If, furthermore, we have constant (improper) population density $\lambda \equiv 1$, the density over the distance from job $||x||$ is

$$c^2||x||e^{-c||x||},$$

which is a $\Gamma(2, c)$ distribution. This result agrees with the empirical pattern of a unimodal commuting length distribution with a skew to the right.

4 Discussion

In this thesis, we build two frameworks to analyze the asymptotic behavior of the choice random variable C_n. However, there are theoretical challenges remaining. In particular, we have only found tractable results for a small number of distributional assumptions. Going ahead, the main priority is to enable us to relax these strict assumptions.

The problem arises because the limiting distribution of C_n is very dependent on the tail behavior of the stochastic utility component. Essentially, if the tail of ϵ_i is too thin, the best choice will, in the limit, always be determined by the deterministic component $m(x)$ of utility. Simply put, with thin tails, the stochastic element can never beat the deterministic element. In these cases C_n will converge to a degenerate distribution, supported on the set of characteristics

$$\arg\max_{x\in\Omega} m(x)$$

with a maximal deterministic utility component.

We can illustrate this by going back to our standard residential choice example with $\Omega = \mathbb{R}^2$ and $m(x) = -c||x||$. Assume that the stochastic component of utility is normally distributed. In that case, $C_n \to (0, 0)$ almost surely as $n \to \infty$. That is, with probability 1, we will live right next to our job.

The idea going forward is that although C_n converges to a degenerate distribution, the way in which it converges is interesting. We never have an infinite amount of residential offers, but merely a large finite number. Thus, we would like to extend our framework to look for sequences h_n such that

$$h_n(C_n)$$

converges to a non-degenerate random variable.

We believe this is a promising avenue of research, and tentative results suggest that it for example might be possible to get a gamma distribution for all stochastic disturbances ϵ_i which lie in the extremal convergence domain of the Gumbel distribution, which is a very large family of random variables.

If this aim can be achieved, the theory will have become a very general tool to map distributional assumptions in random utility theory to outcome predictions of the characteristics of the best 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.

Billingsley, P. (1971). *Weak convergence of measures.* SIAM, Philadelphia.

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

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press.

Gnedenko, B. (1941). Limit theorems for the maximal term of a variational series. In *Doklady Akad. Nauk SSSR.*

Horowitz, J. (1980). Extreme values from a nonstationary stochastic process: an application to air quality analysis. *Technometrics*, 22(4):469–478.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes.* Springer Verlag.

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.

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.

Robert, C. Y. (2013). Some new classes of stationary max-stable random fields. *Statistics & Probability Letters*, 83(6):1496–1503.

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

Weissman, I. (1975). Extremal processes generated by independent non-identically distributed random variables. *The Annals of Probability*, pages 172–177.