Stockholms universitet

Argmax over Continuous Indices of Random Variables – An Approach Using Random Fields

Hannes Malmberg

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2012:2
Matematisk statistik
Februari 2012

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2012:2**
http://www.math.su.se

Argmax over Continuous Indices of Random Variables – An Approach Using Random Fields

Hannes Malmberg*

Februari 2012

Abstract

In commuting research, it has been shown that it is fruitful to model choices as optimization over a discrete number of random variables. In this essay we pass from the discrete to the continuous case, and consider the limiting distribution as the number of offers grow to infinity. The object we are looking for is an argmax measure, describing the probability distribution of the location of the best offer.

Mathematically, we have $\Omega \subseteq R^k$ and seek a probability distribution over Ω. The first argument of the argmax measure is Λ, a probability distribution over Ω, which describes the relative intensity of offers received from different parts of space. The second argument is a measure index $\mu : \Omega \to \mathcal{P}^R$ which associates every point in Ω with a distribution over R, and describes how the type of random offers varies over space.

To define an argmax measure, we introduce a concept called *point process argmax measure*, defined for deterministic point processes. The general argmax measure is defined as the limit of such processes for triangular arrays converging to the distribution Λ.

Introducing a theoretical concept called a max-field, we use continuity properties of this field to construct a method to calculate the argmax measure. The usefulness of the method is demonstrated when the offers are exponential with a deterministic additive disturbance term – in this case the argmax measure can be explicitly calculated.

In the end simulations are presented to illustrate the points proved. Moreover, it is shown that several research developments exist to extend the theory developed in the paper.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:per.hannes.malmberg@googlemail.com. Supervisor: Ola Hössjer, Dmitrii Silvestrov.

Contents

Acknowledgements

This project would not have been possible without the generous help offered me by my supervisors Ola Hossjer and Dmitrii Silvestrov. The questions, comments, criticism and suggestions that I have got from them have been invaluable. The discussions with them has given me a much more global view of the theory of Mathematical Statistics since this project started. They have also shared their experience of writing mathematical papers to a beginner in the field, and thus made sure that the initial ideas were brought from a mental sketch to a completed project.

Another special thanks to my very good friend Zihan Hans Liu. His contribution to my mathematical development is hard to overestimate. With him, essentially every mathematical discussion leads to a new perspective on the subject one is thinking about. The barely two hours we discussed this project inspired precise ideas for future research questions – some of which are included in the section on further research.

1 Introduction

The question which led to this essay came from commuting research. The paper is a formalization of ideas which were first developed in the author's Bachelor Thesis (Malmberg, 2011). The question was whether commuting choices could be explained as optimal choices over a very large number of competing offers. This question led to an inquiry into a mathematical formalism of maximization over a potentially infinite number of random offers.

There are two inputs to such a theory. Firstly, we need the quality of offers associated with each point – that is, the distribution of the value of an offer associated with a particular point in space. Secondly, we need a population distribution, which gives us the relative intensity with which offers are received from different locations.

To put it more formally, let $\Omega \subseteq \mathbb{R}^k$ be a borel measurable set (unless otherwise stated, this will be the interpretation of Ω throughout the paper), and let $\mathcal{P}^{\mathbb{R}}$ denote the set of probability measures on $\mathbb{R}$. We index a set of distributions by Ω,

$$\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$$

Such an indexation can for example state that the distribution of offers become shifted to the left the further away from the origin we are, due to travelling costs. Secondly, we have a population distribution Λ on Ω, giving us the relative number of offers we can expect from different locations.

The task is to define an object corresponding to the idea of the probability distribution of the location of the best offer, when we have a relative intensity of offers given by Λ, and a relative quality of offers given by μ.

We build the theory by first noting that the probability distribution of the location of the best offer is well-defined when we have a deterministic point process. We construct the definition for a general probability distribution as a limiting case from the distributions determined from deterministic point processes.

In this paper we will show that this limiting process has interesting mathematical properties, and that for particular choices of μ (exponential distributions with deterministic additive disturbance), this limit is also very explicit and interpretable. In the process of answering our posed question, a number of theoretical tools are developed and results are derived that are interesting in their own right.

In the end, we show that that the theory can potentially be extended in a number of interesting directions, and we sketch research questions and conjectures for some of these developments.

1.1 Relation to previous theory

The closest relative to the theory developed in this paper is random utility theory in economics. This is the branch of economics that has dealt most with commuting decisions, and the theory postulates that we value options according to a deterministic component and a stochastic disturbance term. (Manski & McFadden, 1981)

The key difference between discrete choice theory and the theory outlined in this paper is that this paper extends the discrete choice paradigm into a random utility theory for choices over a continuous indexing set. In many real life applications there is a very large number of offers which means that a continuity assumption can be justified. Furthermore, it is often the case that after having created a technical machinery, continuous theory allows for much neater and clearer mathematical results which are more easily interpretable. Indeed, the explicitness of the limits derived in this paper suggests that this optimism is warranted.

2 Defining the argmax measure

In this section, we provide the definition of the argmax measure with respect to μ and Λ. We will first define the relevant concepts needed to state the definition.

Definition 1. *Let $\Omega \subseteq \mathbb{R}^k$ and let*

$$\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$$

where $\mathcal{P}^{\mathbb{R}}$ is the space of probability measures on on $\mathbb{R}$. Then μ is called an absolutely continuous measure index on Ω if $\mu(x)$ is an absolutely continuous probability measure for each $x \in \Omega$.

Remark 1. *Unless otherwise stated, μ will always refer to a absolutely continuous measure index.*

The basic building block of our theory will be the argmax measure associated with a deterministic point processes. We provide the following definition.

Definition 2. *Let*

$$N^n = \{x_{n1}, x_{n2}, ..., x_{nn}\}$$

be a deterministic point process on Ω (with the x_{ni}'s not necessarily distinct). Then we define the point process argmax-measure $\tilde{T}_\mu^{N^n}$ as

$$\tilde{T}_\mu^{N^n}(A) = \mathbb{P}\left(\sup_{x_{ni} \in N^n \cap A} Y_{ni} > \sup_{x_{ni} \in N^n \cap A^c} Y_{ni}\right)$$

where $Y_{ni} \sim \mu(x_{ni})$ *are independent random variables, for all borel measurable sets* A*.*

Remark 2. *The paper contains a number of objects which can take both probability distributions and deterministic point processes as arguments. These will have analogous, but not identical, properties. We will use the convention of putting a* $\sim$ *on top of objects taking deterministic point processes as arguments.*

Each deterministic point process has a corresponding probability distribution that is obtained by placing equal weight on all the points in the process. We introduce the following notation.

Definition 3. *For a given deterministic point process* $N^n = \{x_{n1}, x_{n2}, ..., x_{nn}\}$*, the corresponding probability distribution, denoted* P^{N^n} *is given by*

$$P^{N^n}(A) = \frac{\#\{A \cap N^n\}}{n}$$

for all borel measurable sets A*.*

We have now introduced all the concepts we need to define the argmax measure. Using the point process argmax measure $\tilde{T}_\mu^{N^n}$ defined for N^n,we define the general argmax measure for Λ as the limit of $\tilde{T}_\mu^{N^n}$ when the probability measure P^{N^n} converges weakly to Λ. The formal definition of the key concept of the paper is given below.

Definition 4 (General argmax measure)**.** *Let* $\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$ *be a continuous measure index and let* Λ *be a probability distribution on the Borel* σ*-algebra of* Ω*. Suppose now that there exists a probability measure* T_μ^Λ *such that*

$$\tilde{T}_\mu^{N^n} \Rightarrow_\Lambda T_\mu^\Lambda$$

for all point processes N^n *satisfying*

$$P^{N^n} \Rightarrow \Lambda$$

where $\Rightarrow$ *denotes weak convergence and* $\Rightarrow_\Lambda$ *means that the probability measure converges for all sets* A *with* $\Lambda(\partial A) = 0$*.*[1] *Then we call* T_μ^Λ *the argmax measure with respect to measure index* μ *and probability distribution* Λ*.*

[1]Instead of all sets for which $T_\mu^\Lambda(\partial A) = 0$, which is the definition of weak convergence.

3 The max-field and the argmax measure

In this section we will derive the main methodology for finding the argmax measure. Before proceeding, we will first provide a sketch of the formal argument to clarify what will be achieved in this section.

When analysing the limiting behaviour of $\tilde{T}_\mu^{N^n}$, the key object of study is the following random field (for a more general treatment of random fields, see for example Khoshnevisan (2002)):

$$\tilde{M}_\mu^{N^n} = \{\tilde{M}_\mu^{N^n}(A) = \sup_{x_i \in N^n \cap A} Y_i : A \subseteq \Omega \ A \text{ is measurable, and } Y_i \sim \mu(x_i)\}$$

Firstly, there is an immediate connection between this random field and the point process argmax measure in the sense that $\tilde{T}_{\mu,\Lambda}^n$ can be recovered from $\tilde{M}_\mu^{N^n}$. Indeed

$$\tilde{T}_\mu^{N^n}(A) = \mathbb{P}\left(\tilde{M}_\mu^{N^n}(A) > \tilde{M}_\mu^{N^n}(A^c)\right) \tag{1}$$

Leveraging on this connection, we will turn around the problem and make $\tilde{M}_\mu^{N^n}$ the primary object of study. The reason of doing this, an idea that we will develop in the following section, is that the relation (1) exhibits a certain type of continuity which can help us calculate the argmax measure. That is, if there exists a random field M such that

$$\tilde{M}^{N^n} \to M \tag{2}$$

(where the exact sense in which convergence occurs will be specified formally in a later section), then we can show that

$$\tilde{T}_\mu^{N^n} \Rightarrow_\Lambda F(\ldots; M) \tag{3}$$

where $F(\ldots, M)$ is a probability measure defined by

$$F(A; M) = \mathbb{P}\left(M(A) > M(A^c)\right) \tag{4}$$

where we clearly see the connection between (1) and (4).

Now, the important point is that the convergence in (2) sometimes can be established when the only fact assumed of N^n is that

$$P^{N^n} \Rightarrow \Lambda \tag{5}$$

If this is the case, it means that $\tilde{T}_\mu^{N^n} \Rightarrow_\Lambda F(\ldots; M)$ for all N^n having property (5) . Then, the conditions of Definition 4 are satisfied and have shown that $F(\ldots, M) = T_\mu^\Lambda$, the argmax measure. Therefore, if we can prove that the argument outlined above works, we have constructed a method for calculating the argmax measure.

There are two important questions that need to be answered to prove these results. First, can we provide conditions on a general random field M to ensure that $F(\dots, M)$ as defined in (4) actually is a probability measure? Secondly, what is the type of convergence that we have to require in (2) to ensure that (2) implies (3)? These questions will be the topic of the following section.

3.1 Definition of max-fields and connection to argmax-measures

Our first task is to establish a set of sufficient conditions for a random field to generate a probability measure as defined in (4). This is done by the following definition and lemma.

Definition 5. *Let $\mathcal{S}$ be the Borel σ-algebra on Ω, and let $M : \mathcal{S} \to L$, where L is a space of random variables. Let Λ be a probability measure on Ω. We call M an max-field with respect to Λ if the following six properties hold:*

1. *The max measures of disjoint sets are independent;*
2. *If $I = A \cup B$ we have $M(I) = \max\{M(A), M(B)\}$;*
3. *$|M(A)| < \infty$ almost surely if $\Lambda(A) > 0$;*
4. *If $A_1 \supseteq A_2 \dots$, and $\Lambda(A_n) \to 0$, then $M(A_n) \to -\infty$ almost surely;*
5. *$\Lambda(A) = 0 \Rightarrow M(A) = -\infty$ almost surely.*
6. *$M(A)$ is absolutely continuous for all $A \in \mathcal{S}$ with $\Lambda(A) > 0$;*

Remark 3. *Property 6 actually implies property 3 so the list is not minimal. However, property 3 has an independent role in a number of proofs, and is therefore included to more clearly illustrate the properties needed by the max-field.*

The assumptions in the Definition 5 have been chosen to enable us to prove the following lemma.

Lemma 1. *Let $\mathcal{S}$ be the class of measurable subsets of Ω and suppose we have a max-field M with respect to some probability measure Λ. Consider the set function F defined on $\mathcal{S}$ by*

$$F(A; M) = P(M(A) > M(\Omega \setminus A))$$

This set function is a probability measure on the σ-algebra $\mathcal{S}$, which is absolutely continuous with respect to Λ.

Proof. We start by proving that the set function $F(A; M)$ is absolutely continuous with respect to Λ. Indeed, assume that $\Lambda(A) = 0$. In this case, $M(A) = -\infty$ almost surely using property 5 (whenever we refer to numbered properties and do not say otherwise, we are referring to the max-field conditions of Definition 5). Also $\Lambda(\Omega \setminus A) = 1$, and therefore $M(\Omega \setminus A) > -\infty$ almost surely by property 3. Therefore, we get that

$$F(A; M) = \mathbb{P}\left(M(A) > M(\Omega \setminus A)\right) = 0$$

as required.

To prove that $F(\ldots; M)$ is a probability measure, we first note that

$$F(A; M) \in [0, 1]$$

for all measurable $A \subseteq \Omega$. Furthermore, $M(\Omega) > -\infty$ almost surely, and $M(\emptyset) = -\infty$ almost surely. Hence

$$F(\Omega; M) = \mathbb{P}\left(M(\Omega) > M(\emptyset)\right) = 1$$

The only step remaining is to prove countable additivity.

The first step is to establish that F is finitely additive on $\mathcal{S}$. I.e., that if $A = \cup_{i=1}^{n} A_i$, where the A_i's are disjoint, we have

$$F(A; M) = \sum_{i=1}^{n} F(A_i; M)$$

To prove finite additivity, introduce a new notation for the residual set

$$A_{n+1} = \Omega \setminus \bigcup_{i=1}^{n} A_i$$

we can now introduce the events

$$B_i = \mathbb{P}\left(M(A_i) > M\left(\Omega \setminus A_i\right)\right) \text{ for } i = 1, 2, ..., n+1$$

We note that because of absolute continuity, $\mathbb{P}(B_i \cap B_j) = 0$ for all $i \neq j$.

Therefore

$$\begin{aligned}F(A;M) =& \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right)\\ =& \sum_{i=1}^{n} \mathbb{P}(B_i)\\ =& \sum_{i=1}^{n} F(A_i;M)\end{aligned}$$

as required.

Having proved finite additivity, we now need to prove countable additivity. It suffices to show that if we have a decreasing chain of subsets

$$A_1 \supseteq A_2 \supseteq A_3 \dots.$$

such that $\cap_n A_n = \emptyset$, then $F(A_n;M) \to 0$. The technical assumptions we have made on the max measure M make the result simple. Indeed, if $\cap_n A_n = \emptyset$, we have that $\Lambda(A_n) \to 0$, which means that

$$M(A_n) \to -\infty$$

almost surely. But by property 3, $M(\Omega \setminus A_n) > -\infty$ almost surely if $\Lambda(A_n) < 1$, and hence

$$F(A_n;M) = P(M(A_n) \geq M(\Omega \setminus A_n)) \to 0$$

as $n \to \infty$ which completes the proof. □

3.2 Convergence of max-field implies convergence of argmax measure

The second question we need to address is the sense in which convergence in max-fields implies convergence in argmax measure. This is addressed by the following theorem.

Theorem 1. *Fix a measure index*

$$\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$$

and let N^n be a deterministic point process with $P^{N^n} \Rightarrow \Lambda$. Let

$$\tilde{M}_{\mu}^{N^n}(A) = \max_{x_{ni} \in A \cap N^n} Y_{ni}$$

with $Y_{ni} \sim \mu(x_{ni})$ are independent random variables. be an empirical max-field. Let M_μ^Λ be a max-field with respect to Λ and let $F(\dots; M_\mu^\Lambda)$ be defined by

$$F(A; M_\mu^\Lambda) = \mathbb{P}\left(M_\mu^\Lambda(A) > M_\mu^\Lambda(\Omega \setminus A)\right)$$

Now, suppose there exists a sequence of strictly increasing functions g_n, such that

$$\tilde{M}_\mu'^{N^n}(A) = g_n\left(\tilde{M}_\mu^{N^n}(A)\right) \Rightarrow M_\mu^\Lambda(A)$$

for all measurable $A \subseteq \Omega$ with $\Lambda(\partial A) = 0$. Then,

$$\tilde{T}_\mu^{N^n} \Rightarrow_\Lambda F(\dots; M_\mu^\Lambda)$$

where $\tilde{T}_\mu^{N^n}$ are the point process argmax measures associated with N^n.

Proof. By Lemma 1, $F(\dots, M_\mu^\Lambda)$ defines a probability measure. Now, let $A \subseteq \Omega$ be measurable with $\Lambda(\partial A) = 0$ We go through three different cases to prove our result.

Case 1. $\Lambda(A) \in (0, 1)$.
By assumption, there exists a sequence of strictly increasing functions g_n such that:

$$g_n(\tilde{M}_\mu^{N^n}(A)) \Rightarrow M_\mu^\Lambda(A)$$
$$g_n(\tilde{M}_\mu^{N^n}(A^c)) \Rightarrow M_\mu^\Lambda(A^c)$$

hold simultaneously. As $g_n(\tilde{M}_\mu^{N^n}(A))$ and $g_n(\tilde{M}_\mu^{N^n}(A^c))$ are independent for all n, this means that

$$g_n(\tilde{M}_\mu^{N^n}(A)) - g_n(\tilde{M}_\mu^{N^n}(A^c)) \Rightarrow M_\mu^\Lambda(A) - M_\mu^\Lambda(A^c)$$

We note that $M_\mu^\Lambda(A) - M_\mu^\Lambda(A^c)$ is absolutely continuous. Indeed, by the property of max-fields $M_\mu^\Lambda(A)$ and $M_\mu^\Lambda(A^c)$ are absolutely continuous and independent, and therefore their difference is absolutely continuous. We can therefore deduce that

$$\begin{aligned}
\tilde{T}_\mu^{N^n}(A) =& \mathbb{P}(\tilde{M}_\mu^{N^n}(A) > \tilde{M}_\mu^{N^n}(A^c)) \\
=& \mathbb{P}(g_n(\tilde{M}_\mu^{N^n}(A)) > g_n(\tilde{M}_\mu^{N^n}(A^c))) \\
=& \mathbb{P}(g_n(\tilde{M}_\mu^{N^n}(A)) - g_n(\tilde{M}_\mu^{N^n}(A^c)) > 0) \\
\to& \mathbb{P}(M_\mu^\Lambda(A) - M_\mu^\Lambda(A^c) > 0) \\
=& F(A; M_\mu^\Lambda)
\end{aligned}$$

where the convergence step uses absolute continuity to conclude that 0 is a point of continuity of $M_\mu^\Lambda(A) - M_\mu^\Lambda(A^c)$. Therefore, we we have shown that

$$\tilde{T}_\mu^{N^n}(A) \to F(A; M_\mu^\Lambda)$$

for all measurable subsets A with $\Lambda(\partial A) = 0$ and $\Lambda(A) \in (0, 1)$.

Case 2. $\Lambda(A) = 0$
Suppose that $\Lambda(A) = 0$. In Lemma 1, we showed that $F(\dots; M_\mu^\Lambda)$ was absolutely continuous with respect to Λ when M_μ^Λ was a max-field, and hence we know that $F(A; M_\mu^\Lambda) = 0$. Furthermore, as M_μ^Λ is a max-field with respect to Λ, we know that $M_\mu^\Lambda(A) = -\infty$ almost surely, and that $M_\mu^\Lambda(A^c) > -\infty$ almost surely. Thus, we know that

$$g_n\left(M_\mu^{N^n}(A)\right) \Rightarrow -\infty$$
$$g_n\left(M_\mu^{N^n}(A^c)\right) \Rightarrow M_\mu^\Lambda(A^c) > -\infty \text{ a.s}$$

We can find K such that $\mathbb{P}(M_\mu^\Lambda(A^c) > K) = 1 - \epsilon$, and then find N such that for all $n \geq N$

$$\mathbb{P}\left(g_n\left(M_\mu^{N^n}(A)\right) < K\right) > 1 - \epsilon \qquad \mathbb{P}\left(g_n\left(M_\mu^{N^n}(A^c)\right) > K\right) > 1 - 2\epsilon$$

Then, for all $n \geq N$,

$$\mathbb{P}(M_\mu^{N^n}(A) > M_\mu^{N^n}(A^c)) < 3\epsilon$$

and we have proved that

$$\tilde{T}_\mu^{N^n}(A) \to 0 = F(A; M_\mu^\Lambda)$$

as required.

Case 3. $\Lambda(A) = 1$
In this case, we can use that $\partial A = \partial A^c$ to conclude that $\Lambda(\partial A^c) = 0$. Furthermore, we have that

$$\tilde{T}_\mu^{N^n}(A) = 1 - \tilde{T}_\mu^{N^n}(A^c)$$

and we can apply the reasoning from case 2 to argue that

$$\tilde{T}_\mu^{N^n}(A) \to 1$$

Lastly, $F(A; M_\mu^\Lambda) = 1$ as

$$\begin{aligned} F(A; M_\mu^\Lambda) =& F(A; M_\mu^\Lambda) + F(A^c; M_\mu^\Lambda) \\ =& F(\Omega; M_\mu^\Lambda) \\ =& 1 \end{aligned}$$

where the first step uses absolute continuity of $F(\ldots; M_\mu^\Lambda)$ and the second step uses finite additivity of $F(\ldots, M_\mu^\Lambda)$.

Having gone through the three cases, we have showed that for all A with $\Lambda(\partial A) = 0$, it holds that

$$\tilde{T}_\mu^{N^n}(A) \to F(A; M_\mu^\Lambda)$$

and we conclude that

$$\tilde{T}_\mu^{N^n} \Rightarrow_\Lambda F(\ldots; M_\mu^\Lambda)$$

□

Corollary 1. *If there exists a max-field M_μ^Λ such that for all N^n with*

$$P^{N^n} \Rightarrow \Lambda$$

it holds that

$$\tilde{M}_\mu^{N^n}(A) \to M_\mu^\Lambda(A)$$

for all measurable $A \subseteq \Omega$ with $\Lambda(\partial A) = 0$. Then, the argmax measure T_μ^Λ exists and is given by

$$T_\mu^\Lambda(A) = F(A; M_\mu^\Lambda) = \mathbb{P}(M_\mu^\Lambda(A) > M_\mu^\Lambda(\Omega \setminus A))$$

.

Proof. A direct consequence of Theorem 1 and Definition 4. □

4 Argmax measure for exponential offers

The result in Corollary 1 shows that the methods developed in the previous section gives a calculation method for the argmax measure that is workable insofar it is possible to find a max-field M_μ^Λ to which $\tilde{M}_\mu^{N^n}$ is converging for all N^n with $P^{N^n} \Rightarrow \Lambda$.

In this section we make a particular choice for μ and show that under this measure index, such a convergence can indeed be established. As in the previous section, we will start by giving a brief sketch of the argument that subsequently will be developed in full.

In this section we will assume that

$$\mu(x) = m(x) + Exp(1)$$

where the notation should be interpreted as $\mu(x)$ being the law of a random variable Y with the property that

$$Y - m(x) \sim Exp(1)$$

From the previous section we know that the object of interest is the empirical max-field $\tilde{M}_\mu^{N^n}$ which in this case is given for measurable $A \subseteq \Omega$ as

$$\tilde{M}_\mu^{N^n}(A) = \sup_{x_{ni} \in A \cap N^n} Y_{ni}$$

with $Y_{ni} - m(x_{ni}) \sim Exp(1)$ being independent random variables.

Our task is to find the limiting behaviour of this random field. We note that for all A with $\Lambda(A) > 0$, $|A \cap N^n| \to \infty$ as $n \to \infty$ which means that we have a problem involving taking the maximum over a large number of independent random variables. Thus, the natural choice is to apply extreme value theory.

We will have three parts in this section. First we state a general result in extreme value theory, and its specific counterpart related to exponential offers. The second subsection develops the extreme value theory to deal with the fact that $m(x)$ is varying, and we calculate a max-field M_μ^Λ to which $\tilde{M}_\mu^{N^n}$ is converging (after a sequence of monotone transformations). The third subsection uses M_μ^Λ and applies Corollary 1 to calculate the argmax measure T_μ^Λ.

4.1 A primer on extreme value theory

The following theorem a key result in extreme value theory.

Theorem 2 (Fisher-Tippet-Gnedenko Theorem (Extreme Value Theorem))**.** *Let $\{X_n\}$ be a sequence of independent and identically distributed random variables and let $M^n = \max\{X_1, X_2, \ldots, X_n\}$. If there exist sequences $\{a_n\}$ and $\{b_n\}$ with $a_n > 0$ such that:*

$$\lim_{n\to\infty} P\left(\frac{M^n - b_n}{a_n} \leq x\right) = H(x)$$

then $H(x)$ belongs to either the Gumbel, the Frechet, or the Weibull family. (Leadbetter et al., 1983)

Remark 4. *Under a wide range of distributions of X_n, convergence does occur, and for most common distributions the convergence is to the Gumbel(μ, β)-distribution, which has the form Gumbel$(x) = \exp\left(-\exp(-\frac{x-\mu}{\beta})\right)$ for some parameters μ, β.*

We can give a more precise statement when the random variables have an exponential distribution.

Proposition 1. *Let* $\{X_i\}_{i=1}^n$ *be a sequence of i.i.d. random variables with* $X_i \sim Exp(1)$*. Then:*

$$\max_{1\leq i\leq n} X_i - \log(n) \Rightarrow Gumbel(0,1)$$

Proof. Let F be the distribution function of $Exp(1)$, and consider $G_n(x) = F(x + \log(n))^n$. Then

$$\begin{aligned} G_n(x) =& F(x+\log(n))^n \\ =& \left(1 - e^{-x-\log(n)}\right)^n \\ =& \left(1 - \frac{e^{-x}}{n}\right)^n \\ \to& e^{-e^{-x}} \end{aligned}$$

as required (we use the fact that for real valued random variables, pointwise convergence in distribution function implies weak convergence). □

4.2 Limiting max-field with varying $m(x)$

Ordinary extreme value theory assumes that random variables are independently and identically distributed. In our case we do not have identically distributed random variables, as the additive term $m(x)$ varies over space. Hence, we need to establish how the convergence in Proposition 1 works when we take the maximum over independent random variables with varying $m(x)$. We will prove a lemma characterizing this convergence, but we first need a definition and a proposition from the theory of Prohorov metrics which we will use in our proof.

Definition 6. *Let* (M, d) *be a metric space and let* $\mathcal{P}(M)$ *denote the set of all probability measures on* M*. For an arbitrary* $A \subseteq M$*, we define*

$$A^\epsilon = \bigcup_{x\in A} B_\epsilon(x)$$

Given this definition, the Prohorov metric on $\mathcal{P}(M)$ *is defined as (see for example Billingsley (2004))*

$$\pi(\mu,\nu) = \inf\{\epsilon > 0 : \mu(A) \leq \nu(A^\epsilon) + \epsilon \text{ for all measurable } A \subseteq M\}$$

The Prohorov metric has the following important property (the result is included to make the paper mathematically self-contained)

Proposition 2. *If two random variables X and Y taking values in M have the property that $d(X,Y) < \epsilon$ almost surely, then*

$$\pi(\mu_X, \mu_Y) \leq \epsilon$$

where μ_X and μ_Y are the laws of X and Y.

Proof. We note that if X belongs to A, then apart from a set of measure 0, Y will be less than ϵ away. Thus

$$\{X \in A\} \subseteq \{Y \in A^\epsilon\} \cup V$$

where V has measure 0. Hence,

$$\mu_X(A) \leq \mu_Y(A^\epsilon) < \mu_Y(A^\epsilon) + \epsilon$$

As this property holds for all measurable sets A, $\pi(\mu_X, \mu_Y) \leq \epsilon$. □

We now have the mathematical preliminaries to give a full characterization of the limit when $m(x)$ varies.

Theorem 3. *Let $\tilde{M}_\mu^{N^n}(A) = \max_{x_{ni} \in A \cap N^n} Y_{ni}$ where $Y_{ni} - m(x) \sim Exp(1)$ independently. Suppose that Λ is a probability measure on the Borel σ-algebra on Ω and that the following properties hold*

1. *m is bounded*
2. *$P^{N^n} \Rightarrow \Lambda$*
3. *$\int_\Omega \lambda(x) e^{m(x)} \nu(dx) < \infty$*
4. *For all $a < b$, we have that $\Lambda(\partial m^{-1}[a,b)) = 0$*[2]

Then

$$\tilde{M}_\mu'^{N^n}(A) = \tilde{M}_\mu^{N^n}(A) - \log(n) \Rightarrow M_\mu^\Lambda(A)$$

for all A with $\Lambda(\partial A) = 0$, where

$$M_\mu^\Lambda(A) = \log\left(\int_A \lambda(x)\exp(m(x))\nu(dx)\right) + Gumbel(A)$$

Here, $Gumbel(A)$ is a standard Gumbel random variable, where the A-notation denotes that $Gumbel(A)$ and $Gumbel(B)$ are independent for all $A \cap B = \emptyset$. $\nu(dx)$ indicates that the integral is of Lebesgue type with respect to the Lebesgue measure of $\Omega \subseteq \mathbb{R}^k$.

[2]The last point says that the boundary measure of the pre-images of intervals under m has Λ-measure 0 (this holds for example for all continuous functions with finitely many turning points).

Proof. We seek to show that $d(M'^{N^n}_{\mu}(A), M(A)) \to 0$ where d is the Prohorov metric. To prove this, we first make the following preliminary observations.

As m is bounded, we know that we can find K such that

$$-K \leq m(x) < K$$

for all $x \in \Omega$. Then, for any $\delta > 0$, we can define

$$m^{\delta}(x) = \sum_{n=0}^{\lceil 2K/\delta \rceil - 1} (-K + n\delta) I(-K + n\delta \leq m(x) < -K + (n+1)\delta)$$

This function has the property that

$$\sup_{x \in \Omega} |m^{\delta}(x) - m(x)| \leq \delta$$

Analogously, we define

$$\begin{aligned}
&\mu^{\delta}(x) = m^{\delta}(x) + Exp(1) \\
&\tilde{M}'^{N^n}_{\mu^{\delta}}(A) = \tilde{M}^{N^n}_{\mu^{\delta}}(A) - \log(n) \\
&M^{\Lambda}_{\mu^{\delta}}(A) = \log\left(\int_A \lambda(x) \exp(m^{\delta}(x)) dx\right) + Gumbel(A)
\end{aligned}$$

We will also use the notation

$$\begin{aligned}
Y(x) &= m(x) + Exp_x(1) \\
Y^{\delta}(x) &= m^{\delta}(x) + Exp_x(1)
\end{aligned}$$

Here, $Exp_x(1)$ denotes a random variable having distribution $Exp(1)$, and we remember the fact that we have allowed the point process to have non-distinct value, and if we have $x_{ni} = x_{nj}$ for $x_{ni}, x_{nj} \in N^n$ with $i \neq j$ then $Y(x_{ni})$ and $Y(x_{nj})$ are still independent (as the x_{ni}'s in the expression only *index* which probability measure on $\mathbb{R}$ we are supposed to use). This slight notational imprecision does not yield a problem the times the notation is used.

We can use the triangle inequality for the Prohorov metric to conclude that

$$\begin{aligned}
&d(M'^{N^n}_{\mu}(A), M^{\Lambda}_{\mu}(A)) \\
&\leq d(M'^{N^n}_{\mu}(A), M'^{N^n}_{\mu^{\delta}}(A)) + d(M'^{N^n}_{\mu^{\delta}}(A), M^{\Lambda}_{\mu^{\delta}}(A)) + d(M^{\Lambda}_{\mu^{\delta}}(A), M^{\Lambda}_{\mu}(A))
\end{aligned}$$

and we have reduced the problem to show that for an arbitrary $\epsilon > 0$ we can bring the right hand side below ϵ for all sufficiently large n.

To prove the result, we will first use Proposition 2 to bound the first and the last of the three terms. Indeed,

$$\begin{aligned}
|M'^{N^n}_{\mu^\delta}(A) - M'^{N^n}_{\mu}(A)| =& |\max_{x \in A \cap N^n} Y^\delta(x) - \max_{x \in A \cap N^n} Y(x)| \\
=& |\max_{x \in A \cap N^n} m^\delta(x) + Exp_x(1) - \max_{x \in A \cap N^n} m(x) + Exp_x(1)| \\
\leq& \sup_{x \in A \cap N^n} |m^\delta(x) - m(x)| \\
\leq& \sup_{x \in \Omega} |m^\delta(x) - m(x)| \\
\leq& \delta
\end{aligned}$$

On the second line, $Exp_x(1)$ refers to the same random variable in both the max-expressions
Similarly, we have that

$$\begin{aligned}
|M^\Lambda_{\mu^\delta}(A) - M^\Lambda_{\mu}(A)| &= \left|\log\left(\int_A \lambda(x) e^{m^\delta(x)} \nu(dx)\right) - \log\left(\int_A \lambda(x) e^{m(x)} \nu(dx)\right)\right| \\
&= \left|\log\left(\frac{\int_A \lambda(x) e^{m^\delta(x)} \nu(dx)}{\int_A \lambda(x) e^{m(x)} \nu(dx)}\right)\right| \\
&= \left|\log\left(\frac{\int_A \lambda(x) e^{m(x)} e^{m^\delta(x) - m(x)} \nu(dx)}{\int_A \lambda(x) e^{m(x)} \nu(dx)}\right)\right| \\
&\leq \left|\log\left(e^{\sup_{x \in \Omega} |m^\delta(x) - m(x)|} \frac{\int_A \lambda(x) e^{m(x)} \nu(dx)}{\int_A \lambda(x) e^{m(x)} \nu(dx)}\right)\right| \\
&\leq \delta
\end{aligned}$$

Thus, we have shown that we can make the first and last term arbitrarily small. The remaining step is to show that for any δ, we have that

$$d(M'^{N^n}_{\mu^\delta}(A), M^\Lambda_{\mu^\delta}(A)) \to 0$$

as $n \to \infty$. We will use the fact that convergence in the Prohorov metric is equivalent to weak convergence, and show that weak convergence holds. First we give two preliminary results, and include the proofs to make the exposition self-contained.

Claim 1. *Let $X^i_n \Rightarrow X^i$ for $i = 1, 2.., k$, where X^i is absolutely continuous, and $P(X^i > -\infty) = 1$ for all i. Furthermore, let $Z^i_n \Rightarrow -\infty$ for $i = 1, .., m$.*

Then

$$M^{X_n,Z_n} = \max\left\{X_n^1,..,X_n^k,Z_n^1,..,Z_n^m\right\} \Rightarrow \max\{X^1,..,X^k\}$$

Proof of claim. We prove this by convergence of distribution functions. Indeed,

$$\begin{aligned} F_{M^{X_n,Z_n}}(x) =& P(\max\left\{X_n^1,..,X_n^k,Z_n^1,..,Z_n^m\right\} \leq x) \\ =& F_{\max\{X_n^1,..,X_n^k\}}(x) P(\max\left\{X_n^1,..,X_n^k\right\} \geq \max\left\{Z_n^1,..,Z_n^m\right\}) \\ +& F^-_{\max\{Z_n^1,..,Z_n^m\}}(x) P(\max\left\{X_n^1,..,X_n^k\right\} < \max\left\{Z_n^1,..,Z_n^m\right\}) \\ \to& F_{\max\{X^1,..,X^k\}}(x) \times 1 + 0 \\ =& F_{\max\{X^1,..,X^k\}}(x) \end{aligned}$$

which proves the result. □

Claim 2. *If $X_i \sim Gumbel(\mu_i, 1)$ for $i = 1,..,k$, then*

$$\tilde{X} = \max_{1\leq i\leq k} X_i \sim Gumbel\left(\log\left(\sum_{i=1}^k e^{\mu_i}\right), 1\right)$$

Proof. We prove this directly by the distribution function.

$$\begin{aligned} F_{\tilde{X}}(x) =& \prod_{1\leq i\leq k} F_{X_i}(x) \\ =& \prod_{1\leq i\leq k} e^{-e^{-x+\mu_i}} \\ =& e^{-\sum_{i=1}^k e^{-x+\mu_i}} \\ =& e^{-e^{-x+\log\left(\sum_{i=1}^k e^{\mu_i}\right)}} \end{aligned}$$

and we recognize the expression on the last line as the distribution function of $Gumbel\left(\log\left(\sum_{i=1}^k e^{\mu_i}\right), 1\right)$. □

We now return to the original problem of showing that

$$\begin{aligned} M'^{N^n}_{\mu^\delta}(A) =& \max_{x\in A\cap N^n}\left\{m^\delta(x) + Exp_x(1)\right\} - \log(n) \\ &\Rightarrow \log\left(\int_A \lambda(x)e^{m^\delta(x)}\nu(dx)\right) + Gumbel(A) \end{aligned}$$

Introduce the following notation

$$k = \lceil \frac{2K}{\delta} \rceil$$
$$A_i = \{x \in \Omega : m(x) \in [-K + i\delta, -K + (i+1)\delta)\} \quad i = 0, 1.., k-1$$
$$J = \{i \in \{0, 1, 2.., k-1\} : \Lambda(A_i) > 0\}$$

We note that

$$\begin{aligned} \tilde{M}'^{N^n}_{\mu^\delta}(A) &= \max_{0 \le i \le k-1} \left\{ \max_{x \in A_i \cap N^n} m^\delta(x) + Exp_x(1) - \log(n) \right\} \\ &= \max_{0 \le i \le k-1} \tilde{M}'^{N^n}_{\mu^\delta}(A_i) \end{aligned}$$

As

$$A_i = m^{-1}[-K + \delta i, -K + \delta(i+1))$$

the conditions given in the statement of the theorem ensures that $\Lambda(\partial A_i) = 0$. Therefore, as we have assumed that $P^{N^n} \Rightarrow \Lambda$, we know that

$$P^{N^n}(A_i) \to \Lambda(A_i) = \int_A \lambda(x)\nu(dx)$$

We will find the limiting behaviour for $\tilde{M}'^{N^n}_{\mu^\delta}(A)$ for both $i \in J$ and $i \notin J$.

We start with $i \notin J$. In this case, we want to show that $\tilde{M}'^{N^n}_{\mu^\delta}(A) \Rightarrow -\infty$. Using that $m^\delta(x)$ is constant over A_i, we get

$$\tilde{M}'^{N^n}_{\mu^\delta}(A) = m^\delta(x) + \max_{x \in N^n \cap A} Exp_x(1) - \log(n)$$

If $|N^n \cap A|$ is bounded as $n \to \infty$, this expression clearly tends to $-\infty$. If $|N^n \cap A|$ is not bounded above we can rewrite the expression as follows

$$\tilde{M}'^{N^n}_{\mu^\delta}(A) = m^\delta(x) + \max_{x \in N^n \cap A} \{Exp_x(1) - \log(|N^n \cap A|)\} + \log\left(\frac{|N^n \cap A|}{n}\right)$$

The expression in the curly brackets converges to a Gumbel distribution and the last expression in the logarithm tends to $-\infty$, so again the whole expression tends to $-\infty$.

We now look at $i \in J$. If we write $S_n^i = |A_i \cap N^n|$, then

$$
\begin{aligned}
\tilde{M}'^{N^n}_{\mu}(A_i) &= \max_{x \in A_i \cap N^n} \left\{ m^\delta(x) + Exp_x(1) \right\} - \log(n) \\
&= (K + \delta i) + \max_{x \in A_i \cap N^n} \left\{ Exp_x(1) - \log(S_n^i) \right\} + \log\left(\frac{S_n^i}{n}\right) \\
&\Rightarrow (K + \delta i) + Gumbel(A_i) + \int_A \lambda(x)\nu(dx) \\
&= \log\left(e^{K+\delta i} \int_A \lambda(x)\nu(dx)\right) + Gumbel(A_i) \\
&= \log\left(\int_A \lambda(x) e^{m^\delta(x)} \nu(dx)\right) + Gumbel(A_i)
\end{aligned}
$$

Combining our findings with the results in Claim 1 and 2, we get that.

$$
\begin{aligned}
M'^{N^n}_{\mu^\delta}(A) &= \max_{0 \le i \le k-1} M'^{N^n}_{\mu^\delta}(A_i) \\
&= \max\{\max_{i \in J} M'^{N^n}_{\mu^\delta}(A_i), \max_{i \notin J} M'^{N^n}_{\mu^\delta}(A_i)\} \\
&\Rightarrow \max_{i \in J}\left(\log\left(\int_{A_i} \lambda(x) e^{m^\delta(x)} \nu(dx)\right) + Gumbel(A_i)\right) \\
&= \log\left(\sum_{i \in J} e^{\log\left(\int_{A_i} \lambda(x) e^{m^\delta(x)} \nu(dx)\right)}\right) + Gumbel(A) \\
&= \log\left(\sum_{i \in J} \int_{A_i} \lambda(x) e^{m^\delta(x)} \nu(dx)\right) + Gumbel(A) \\
&= \log\left(\sum_{i \in J} \int_{A_i} \lambda(x) e^{m^\delta(x)} \nu(dx) + \sum_{i \notin J} \int_{A_i} \lambda(x) e^{m^\delta(x)} \nu(dx)\right) \\
&\quad + Gumbel(A) \\
&= \log\left(\int_A \lambda(x) e^{m^\delta(x)} \nu(dx)\right) + Gumbel(A)
\end{aligned}
$$

where the second to last line uses that $\Lambda(A_i) = 0$ for $i \notin J$. Therefore, we have shown that

$$\tilde{M}'^{N^n}_{\mu^\delta}(A) \Rightarrow M^\Lambda_{\mu^\delta}(A))$$

we can use the equivalence of weak convergence and convergence in Prohorov metric to conclude that

$$d(\tilde{M}'^{N^n}_{\mu^\delta}(A), M^\Lambda_{\mu^\delta}(A)) \to 0$$

as $n \to \infty$.

By picking $\delta < \epsilon/3$ and S such that $d(M'^{\delta,N^n}_{\mu}(A), M^{\delta}(A)) < \epsilon/3$ for all $n \geq S$, we get

$$\begin{aligned}
&d(M'^{N^n}_{\mu}(A), M^{\Lambda}_{\mu}(A)) \\
&\leq d(M'^{N^n}_{\mu}(A), M'^{N^n}_{\mu^{\delta}}(A)) + d(M'^{N^n}_{\mu^{\delta}}(A), M^{\Lambda}_{\mu^{\delta}}(A)) + d(M^{\Lambda}_{\mu^{\delta}}(A), M^{\Lambda}_{\mu}(A)) \\
&< \epsilon
\end{aligned}$$

and we are done. □

Proposition 3. *The random field defined by*

$$M(A) = \log\left(\int_A \lambda(x) e^{m(x)} \nu(dx)\right) + \mathit{Gumbel}(A)$$

fulfills the conditions of Definition 5 when m and λ satisfy the conditions of Theorem 3.

Proof. We note that property 1 clearly holds as the $M(A)$ and $M(B)$ are measurable with respect to independent σ-algebras. Property 2 holds by the properties of the Gumbel distribution. To prove that property 3 holds, it suffices to prove that property 6 holds as the latter implies the former. And, indeed,

$$\log\left(\int_A \lambda(x) e^{m(x)} \nu(dx)\right) + \mathit{Gumbel}(A)$$

is absolutely continuous when $\int_{A_n} \lambda(x) e^{m(x)} \nu(dx) > 0$, which in turn is true when

$$\Lambda(A) = \int_A \lambda(x) dx > 0$$

Property number 4 holds as

$$\log\left(\int_{A_n} \lambda(x) e^{m(x)} \nu(dx)\right) \to -\infty$$

as $\int_{A_n} \lambda(x) dx \to 0$.

Lastly, property 5 holds. If $\Lambda(A) = 0$, it is clear that $\int_A \lambda(x) e^{m(x)} \nu(dx) = 0$, and we get $M(A) = -\infty$ almost surely. □

4.3 Argmax density with exponential offers

In Corollary 1, it was shown that the limiting behaviour of $\tilde{M}_\mu^{N^n}$ determines the argmax measure. Thus, we can use the limit derived in Theorem 3 together with Corollary 1 to derive the argmax measure associated with μ and Λ.

Theorem 4. *Let $\mu(x) = m(x) + Exp(1)$ and let Λ be a probability measure with density $\lambda(x)$. Suppose that $\lambda(x)$ and $m(x)$ jointly satisfy the conditions in Theorem 3. Then, argmax measure T_μ^Λ exists, and has density*

$$T_\mu^\Lambda(x) = C\lambda(x)\exp(m(x))$$

where

$$C = \left(\int_\Omega \lambda(x)e^{m(x)}\nu(dx)\right)^{-1}$$

is a normalizing constant.

Proof. We first note that Theorem 3 imples that for all point processes N^n with $P^{N^n} \Rightarrow \Lambda$, it holds that

$$\tilde{M}_\mu^{N^n}(A) - \log(n) \Rightarrow M_\mu^\Lambda(A)$$

for all measurable A with $\Lambda(\partial A) = 0$, where M_μ^Λ is defined as in Theorem 3. As Proposition 3 states that M_μ^Λ is a max-field, we can apply Corollary 1 and conclude that the argmax measure T_μ^Λ exists and is given by

$$T_\mu^\Lambda(A) = \mathbb{P}\left(M_\mu^\Lambda(A) > M_\mu^\Lambda(A^c)\right)$$

We can now derive the density of the argmax-measure. Let A be measurable and let

$$G(x) = e^{-e^{-x}}$$

denote the distribution function of a standard Gumbel distribution. For notational brevity, let

$$L(A) = \left(\log(\int_A \lambda(x)e^{m(x)}d\nu(x)\right)$$

We then get:

$$\begin{aligned}
&T^{\Lambda}_{\mu}(A) \\
&= P\left(M^{\Lambda}_{\mu}(A) > M^{\Lambda}_{\mu}(\Omega \setminus A)\right) \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(M(A) \in dr\right) \mathbb{P}\left(M(\Omega \setminus A) < r\right) \\
&= \int_{-\infty}^{\infty} G'\left(r + L(A)\right) G\left(r + L(\Omega \setminus A)\right) dr \\
&= \int_{-\infty}^{\infty} e^{-r+L(A)} e^{-e^{-r+L(A)}} e^{-e^{-r+L(\Omega \setminus A)}} dr \\
&= e^{L(A)} \int_{-\infty}^{\infty} \exp(-r) \exp\left(-e^{-r+L(\Omega)}\right) dr \\
&= C\left(\int_A \lambda(x) e^{m(x)} \nu(dx)\right)
\end{aligned}$$

As this holds for all measurable sets A, $C\lambda(x)e^{m(x)}$ is the density of the argmax measure with respect to the Lebesgue measure. □

5 Verification by Simulation

What we have done is to show that if offers are exponentially distributed with an additive, position dependent, term, we have solved the distribution for the argmax for all density distributions. We will show how it works for some examples by simulation. What we do is that we generate 1000 random points x_i on $\mathbb{R}$ or $\mathbb{R}^2$ according to some probability distribution, then we generate 1000 points $y_i(x_i)$ by $y_i = m(x_i) + Exp(1)$ for some predefined function m. We then return $\arg\max_{x_i} y(x_i)$. We repeat the procedure $10,000$ times and draw a histogram with the result together with the theoretically predicted density. For one case, we also plot the values $y(x_i)$ and check these against the theoretically predicted distribution.

The first graph comes from the commuting example. We sample from a uniform distribution over a disc on $\mathbb{R}^2$ with radius 100. We let

$$m(x) = -0.05 \times \sqrt{x_1^2 + x_2^2}$$

be a function to describe travel costs and the density $\lambda(x)$ is given by

$$\lambda(x) = \frac{1}{100^2\pi} \text{ for } ||x|| < 100$$

We record the distance to the origin of the best offer, and plot the result. We note that the argmax density in two dimensions in this case is given by

$$t_\mu^\Lambda(x) = C\lambda(x)\exp(-cr)$$

where $r = \sqrt{x_1^2 + x_2^2}$, $r \in (0, 100)$, $c = 0.05$ and C is a normalizing constant. When we integrate to get the density of the distance, we get that it it is

$$t_\mu'^\Lambda(r) = 2\pi C r\lambda(x)\exp(-cr) = \frac{2r\exp(-cr)}{100^2}$$

We also plot the best value, which in general case should distributed as

$$Gumbel(\log\left(\int_\Omega \lambda(x)e^{m(x)}d\nu(x)\right), 1) = Gumbel(\log\left(\frac{1}{C}\right), 1)$$

By the definition of C in our case, we get that

$$C = \left(\int_0^{100} \frac{2s\exp(-cs)}{100^2}dr\right)^{-1}$$

Hence, the best value is distributed as

$$Gumbel\left(\log\left(\int_0^{100} \frac{s}{0.5\times 100^2}e^{-cs}ds\right), 1\right)$$

After this, we display the generality of the result by using very different sampling intensities λ and mean-value functions m in one dimension. We summarize the results in the three graphs below which show the results for uniform, Weibull and lognormal sampling densities with mean-value functions $m(x) = |x|$, $m(x) = \sqrt{x+1}$ and $m(x) = -x^2$. In all cases, our theoretical predictions bear out. This illustrates the generality of the results proved in the Section 4.

Figure 1: Argmax distribution of distance to origin for a uniform sample on $D((0,0),100) \subseteq \mathbb{R}^2$ with $m(x) = -0.05 \times \sqrt{x_1^2 + x_2^2}$ (line theoretical result))

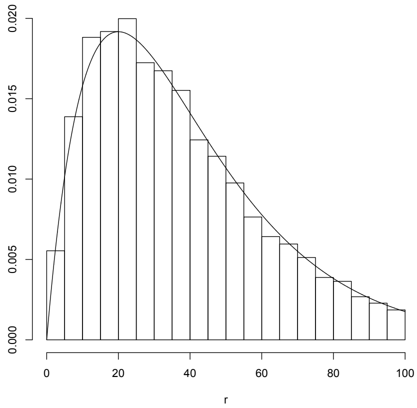
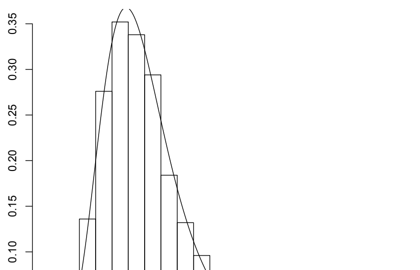


Figure 2: Distribution of best value on $D((0,0),100) \subseteq \mathbb{R}^2$ with $m(x) = -0.05 \times \sqrt{x_1^2 + x_2^2}$ (line theoretical result))



6 Conclusion

In this paper we set out to define and prove results about the concept of an argmax measure over a continuous index of probability distributions. A reasonable definition has been provided, and we have expanded the toolbox

Figure 3: Argmax distribution for sampling intensity $U[-1,1]$ and $m(x) = |x|$ (line theoretical result)

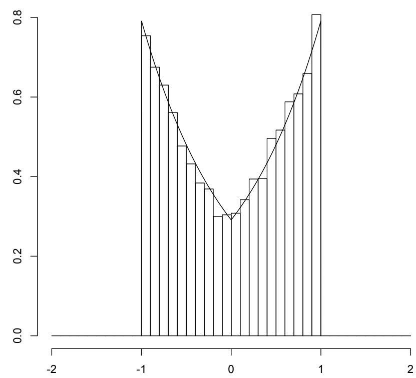
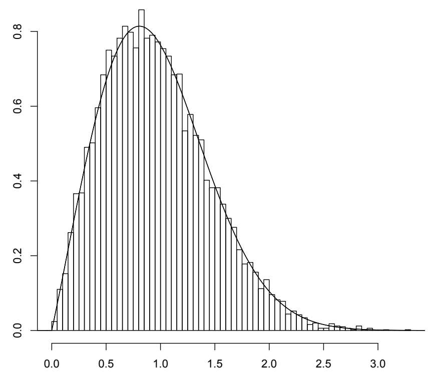


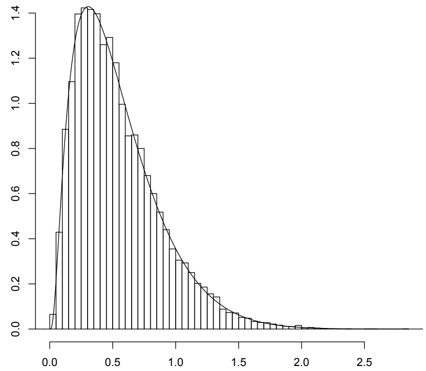
Figure 4: Argmax measure with sampling intensity Weibull(2,1) and $m(x) = \sqrt{x+1}$ (line theoretical result)



available to address these types of problems by introducing the max-field calculation method. The usefulness of the developed method is shown when it is applied to the case of taking the argmax over what can be described as exponential white noise.

Many of the results have been proved under quite restrictive assumptions

Figure 5: Argmax measure for a standard lognormal distribution with $m(x) = -x^2$ (line theoretical result)



of independence, deterministic point processes, and exponential distributions. This means that there are plenty of potential generalizations and extensions of the theory available on the basis of the work done in this paper.

We have identified three different avenues of further research. Firstly, it is possible to have more general distributional assumptions regarding independence and type of distributions. Secondly, it is possible to construct a theory where we do not have deterministic point processes as our basic building blocks, but rather allow there to be stochasticity in the selection of points, leading to a doubly stochastic problem. Lastly, extending on the second point there is a slightly more radical reformulation of the theory, where the process of taking the argmax is viewed as a measure transformation, starting with a probability measure Λ and getting a new probability distribution T_μ^Λ. The problem can then be studied from the viewpoint of continuity properties of this rather abstract map.

We will conclude the paper with a short exposition of these different ideas. We sketch what we believe could be good ways of proceeding along these different lines, and along the way highlight the problems that have to be addressed if such a theory were to be developed.

6.1 More general stochasticity in offered values

In this paper we have defined the max-field by

$$\tilde{M}_\mu^{N^n} = \left\{ M_\mu^{N^n}(A) = \sup_{x_{ni} \in A \cap N^n} Y_{ni} \quad : A \text{ is measurable} \right\}$$

where $Y_{ni} \sim \mu(x_{ni})$ are independent random variables. When we made explicit calculations we assumed that $\mu(x)$ was a perturbed exponential distribution for all $x \in \Omega$.

It is important that the specific distributional assumptions and the independence assumption can be relaxed. We want the unconditional distribution of offers to be much more general than just exponential (wage distributions are for example much closer to lognormal) and most random processes in space have some sort of spatial autocorrelation which would violate the independence assumption.

The possibilities for generalization also look good. The results given in Theorem 1 and Corollary 1 do not use any distributional assumption or independence assumption. The only important thing is to find an appropriate max-field M_μ^Λ to which $\tilde{M}_\mu^{N^n}$ converges in the sense outlined in Theorem 1.

The property of the exponential distribution that we used in our proofs was that the extreme value behaviour of the exponential distribution is especially nice. If we retain the independence assumption but with a more general μ, the generality of extreme value theory still applies, and we can therefore hope to get a convergence to a max-field.

There are somewhat larger obstacles if we drop the independence assumption. The behaviour of the random variable $\tilde{M}_\mu^{N^n}(A)$ can indeed be good for individual sets A as $n \to \infty$ (especially if we let A become infinitesimal, think for example of the Brownian motion where the distribution of the supremum over a very small interval $[t, t+\delta)$ is likely to be close to $N(0, t)$ under suitable regularity conditions). On the other hand, the resulting random field M_μ^Λ would not necessarily be a max-field, as $M(A)$ and $M(B)$ would not necessarily be independent for disjoint sets A and B. Thus, in order to drop the independence assumption, we have to carefully go through all proofs involving the max-field, check the exact way in which the independence assumption is used, and see whether it can be weakened in our proofs, or see if there are alternative ways of proving the same results.

6.2 Stochasticity in the choice of test points

Our current theory uses a definition of argmax measure that depends on a certain limiting behaviour for all deterministic point processes N^n having certain properties (that is, $P^{N^n} \Rightarrow \Lambda$). An alternative way would be to have some sort of stochasticity in the selection of points, for example we could define

$$N^n = \{X_{n1}, X_{n2}, ..., X_{nn}\}$$

where $X_{ni} \sim \Lambda$ independently and identically. This sort of procedure would induce a form of double stochasticity which would increase the complexity of the problem, but also make it slightly more intuitive. In the current paper, the pre-limiting and limiting objects are of very different types. The pre-limiting probability distribution is conditioned on a particular deterministic point process, and it should converge to a single probability distribution for all such deterministic point processes. It is for example not generally the case that $\tilde{T}_\mu^{N^n}$ is the same object as $T_\mu^{P^{N^n}}$. This difference in what they are is what necessitates the $\sim$-notation distinguishing between objects taking point processes as arguments, and objects taking probability distributions as arguments.

If the point processes are generated by a random process, we have a single random probability measure that should converge almost surely to another probability measure, a case which would have more symmetry in the treatment of limiting and prelimiting objects. We will sketch how such a theory could look. In the i.i.d. case, for example, we could define the following two objects:

Definition 7. *Let $\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$ be a measure index and let Λ be a probability measure on Ω. We define a couple sample $\{X_i, Y_i\}_{i=1}^n$ with respect to Λ and μ by drawing an i.i.d. sample $X_1, .., X_n$ from Λ and generate another independent sample $Y_1, .., Y_n$ with*

$$Y_i \sim \mu(X_i)$$

Definition 8. *Suppose we have a measure index $\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$ and a probability distribution Λ on Ω. We then define the sample argmax transformation $T_{\mu,\Lambda}^n$ to be the law associated with the random variable X_I^n where*

$$I = \arg \max_{1 \le i \le n} Y_i$$

and $\{X_i, Y_i\}_{i=1}^n$ is a coupled sample with respect to Λ and μ.

Having done this definition, the problem could now be to find a random variable T_μ^Λ such that

$$T_\mu^\Lambda = \lim_{n \to \infty} T_{\mu,\Lambda}^n$$

almost surely. It does introduce a number of problems in exactly what sense this problem can be reduced to the problem we have already proved. However, if this could be done, the theory would be representable in a more compact way.

Lastly, we could even depart from letting the triangular array N^n be the result of i.i.d. draws from Λ. An interesting thing would be to see if well-behaved limits could be defined even if N^n was generated by a Markov Chain or some other more general stochastic process. It is likely that, possibly under somewhat stronger regularity conditions, there will be convergence for a larger class of processes than just i.i.d.-processes.

6.3 Argmax as a measure transformation

One way of re-conceptualize the problem of taking the argmax is to see it as a measure transformation indexed by a measure index. Indeed, for a fixed μ, what our theory does is to start with one probability distribution Λ, and end up with another probability distribution T_μ^Λ. In a more abstract setting, the study of the argmax-measure would be the study of this map. Section 6.2. hints at this map being the limit of the measure transformation $T_{\mu,\Lambda}^n$ which is obtained by taking a finite sample of points $X_{n1}, .., X_{nn}$ from Λ and then draw the associated offers $Y_{n1}, .., Y_{nn}$ and return the X_{ni} associated with the highest Y_{ni}. As the sample size grow, the empirical distribution will almost surely converge to the true distribution, so the problem can potentially be cast in continuity terms.

Also, an interesting question would then be to find correspondences between the measure index μ and the effect our mapping has on the initial distribution. We can for example note that in the exponential case our new distribution has a density proportional to $\lambda(x)e^{m(x)}$ and taking the argmax therefore corresponds to exponential tilting of the initial distribution. It might be possible to explore more connections between μ and the effect of T_μ^Λ.

References

[1] Billingsley, P, *Convergence of Probability Measures*, Wiley-Interscience, 1999

[2] Gumbel, E.J., *Statistics of Extremes (new edition)*, Courier Dover Publications, 2004

[3] Khoshnevisan, D, *Multiparameter Processes: An Introduction to Random Fields*, Springer, 2002

[4] Leadbetter, M. R.; Lindgren, G. & Rootzen, H, *Extremes and Related Properties of Random Sequences and Processes*, Springer Verlag, 1983.

[5] Malmberg, H *Spatial Choice Processes and the Gamma Distribution*, Stockholm University, 2011

[6] Manski, C.F. & McFadden, D., *Structural Analysis of Discrete Data with Econometric Applications*, Section III, MIT Press, 1981, *http://elsa.berkeley.edu/ mcfadden/discrete.html*

[7] Resnick, S.I. *Adventures in Stochastic Processes*, Birkhuser Boston, 1992