Mathematical Statistics
Stockholm University

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

Hannes Malmberg
Ola Hössjer

Research Report 2012:7

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 **2012:7**,
http://www.math.su.se/matstat

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

Hannes Malmberg*
Ola Hössjer†

September 27, 2012

Abstract

In commuting research, it is common to model choices as optimization over a discrete number of random variables. In this paper we extend this theory from the discrete to the continuous case, and consider the limiting distribution of the location of the best offer as the number of offers tends to infinity.

Given a set $\Omega \subset \mathbb{R}^d$ of possible offers we seek a distribution over Ω, the argmax measure of the best offer. It depends on Λ, the sampling distribution of offer locations, and a measure index μ, which assigns to each point $x \in \Omega$ a probability distribution of offers.

This problem is closely related to argmax theory of marked point processes, altough we consider deterministic sequences of points in space, to allow for greater generality. We first define a finite sample argmax measure and then give conditions under which it converges as the number of offers tends to infinity.

To this end, we introduce a max-field of best offers and use continuity properties of this field to calculate the argmax measure. We demonstrate the usefulness of the method by giving explicit formulas for the limiting argmax distribution for a large class of models, including exponential independent offers with a deterministic, additive disturbance term. Finally, we illustrate the theory by simulations.

Key words: Argmax distribution, commuting, extreme value theory, exponential offers, marked point processes, max field.

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

1 Introduction

In commuting research, choices are offered at various points in space and are assumed to have random value. It is of interest to determine which offer is optimal as well as deriving the statistical properties of the choice, both in terms of the value of the offer and its position in space. This is related to random utility theory, the 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, see Manski and McFadden (1981).

In this paper, we focus in particular on the positional distribution of the best offer, continuing studies initiated in Malmberg (2011, 2012). To this end, we create a mathematical formalism of maximization over a potentially infinite number of random offers. To put it more formally, let $\Omega \subseteq \mathbb{R}^k$ be a Borel measurable set, and let $\mathcal{P}^{\mathbb{R}}$ denote the set of probability measures on $\mathbb{R}$. We index a set of distributions by $\mu : \Omega \to \mathcal{P}^{\mathbb{R}}$, where $\mu(x)$ is the distribution of offers at location $x \in \Omega$. 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 the probability distribution of the location of the best offer when the relative intensity of offers is provided by Λ, and the relative quality of offers by μ. We build the theory by first defining the probability distribution of the location of the best offer for finite samples and then define a limiting distribution when the number of offers tends to infinity.

It turns out that the distribution has very interesting mathematical properties, and that for particular choices of μ, including exponential distributions with deterministic additive disturbances, this limit is also very explicit and interpretable. In the process of answering our posed question, some 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.

There is quite an extensive literature on ranom utility theory for finite choice sets, see for instance Marley and Colonius (1992), Mattsson et al. (2011) and references therein. Our paper represents a generalization non-finite choice sets. In this, it has similar aims as Ben-Akiva et al. (1985), Resnick and Roy (1989), and Dagsvik (1994). The main difference is that our approach does not depend on the "independence of irrelevant alternatives" assumption on the final distribution of choices. Instead, we let the numbers of alternatives grow where each alternative has a random utility with arbitrary distribution. In our case, the continuous logit model is a special case when this arbitrary distribution is an exponential distribution with determinstic, spatially varying, shifts.

The location of offers can be viewed as the realization of a point process, cf. Cox and Isham (1980) and Diggle (2003). The value associated to each offer is a mark, and hence the joint sequence of locations and values of all offers becomes a marked point process (Jacobsen, 2006). Our results are closely related to an asymptotic theory for the argmax (or the position of the largest record) of a marked point processes as the intensity of the point process tends to infinity. The limiting argmax distribution coincides with Λ when the offer distribution $\mu(x) \equiv H$ is independent of location. On the other hand, when $\mu(x)$ varies with x we get a non-stationary sequence of marks, which, under certain conditions, yields an associated non-trivial limiting argmax distribution.

The theory which is most closely related mathematically to the one presented in this paper is the theory of concomitants of extreme order statistics, see for example Ledford and Tawn (1998) and the references therein. The main difference is that we consider limits of deterministic point processes in contrast to large samples from explicitly bivariate distributions. Moreover, we treat the case when μ corresponds to homoscedastic regression in particular detail. To the best of our knowledge ours is the first attempt to apply an approach using random fields to the analysis of concomitants of extreme order statistics.

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 introduce some 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, for each $x \in \Omega$, $\mu(x)$ is an absolutely continuous probability measure on $\mathbb{R}$ with respect to Lebesgue measure.

Unless otherwise stated, μ refers to an absolutely continuous measure index and Ω is a subset of $\mathbb{R}^k$.

We will now introduce the basic building block of our theory: the argmax measure associated with a deterministic set of points. Throughout the discussion, elements of point sequences $N^n = \{x_{n1}, x_{n2}, ..., x_{nn}\}$ will be multi-sets, i.e. the $x_{n,i}$'s are not necessarily distinct for identical n.

Definition 2 *An indexed random vector Y^{N^n} with respect to μ is a random vector on $\mathbb{R}^n$ with independent components, where each component has marginal distribution $\mu(x_{ni})$.*

Unless there is ambiguity, we omit the superscript N^n.

Definition 3 *The point process argmax-measure $\tilde{T}_\mu^{N^n}$ is defined as*

$$\tilde{T}_\mu^{N^n}(A) = \mathbb{P}\left(\max_{1\leq i\leq n: x_{ni}\in N^n\cap A} Y_{ni} \geq \max_{1\leq i\leq n} Y_{ni}\right) = \mathbb{P}(X_n \in A), \tag{1}$$

for all Borel measurable sets $A \subseteq \Omega$, and

$$X_n = \arg\max_{x_{ni}\in N^n} Y_{ni} \tag{2}$$

is the almost surely unique argmax of $\{Y_{ni}\}$.

We use the convention of putting a $\sim$ on top of objects having (deterministic) empirical distributions as arguments, and drop $\sim$ for their large sample limits. We will write $\mathcal{Q}^\Omega$ to denote the set of finite multisets on Ω. With this notation, $\tilde{T}_\mu$ is a function from $\mathcal{Q}^\Omega$ to $\mathbb{R}_+^{\mathcal{B}(\Omega)}$, the family of non-negative set functions on the Borel sigma algebra on Ω. We use the family of non-negative set functions as we want to be able to consider mappings which possibly take values which are not probability measures.
Even though N^n is a deterministic set of points, it can typically be thought of as the realization of a point process. If so, we condition on the randomness associated with that process. In any case, it is convenient to define the empirical distribution function

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

for all Borel sets $A \subset \Omega$.

Definition 4 *For a probability distribution* Λ*, we define the point sequence domain of convergence as*

$$\mathcal{N}^{\Lambda} = \left\{ \{N^n\} : P^{N^n} \Rightarrow \Lambda \right\}$$

i.e. the class of point sequences whose empirical distributions converge to Λ*.*

We have now introduced the concepts needed to define the argmax measure.

Definition 5 (Limiting argmax measure.) *A probability measure* T_{μ}^{Λ} *such that*

$$\tilde{T}_{\mu}^{N^n} \Rightarrow T_{\mu}^{\Lambda} \tag{3}$$

for all $\{N^n\}_{n \in \mathbb{N}} \in \mathcal{N}^{\Lambda}$ *will be called an argmax measure with respect to* μ *and* Λ*. Here (and everywhere else in the paper),* $\Rightarrow$ *refers to weak convergence.*

3 Calculating the argmax measure

In this section, we will develop a method for calculating the argmax measure. For each N^n, we attach a particular random field $\tilde{M}^{N^n}$. Thereafter, we derive asymptotic properties of $\tilde{T}^{N^n}$ by considering the asymptotic behavior of $\tilde{M}^{N^n}$. We will first introduce random fields and define the relevant terms. Thereafter, a notion of convergence in random fields is introduced, and we prove a result connecting this convergence with the convergence to an argmax measure.

3.1 Random fields in an argmax context

We write a random field over the sigma algebra of Ω as

$$M : \mathcal{S} \times \mathcal{B}(\Omega) \to \mathbb{R}$$

where $\mathcal{S}$ is a generic sample space and $\mathcal{B}(\Omega)$ denotes the Borel σ-algebra on Ω. Thus, for fixed s, $M(s, \cdot)$ is a set function on $\mathcal{B}(\Omega)$ and for fixed $A \in \mathcal{B}(\Omega)$, $M(\cdot, A)$ is a random variable taking values in $\mathbb{R}$. We sometimes write $M(A)$ as short-hand for $M(\cdot, A)$ and we write $M(s, A)$ for a particular realization of the random variable $M(\cdot, A)$. We will write $\mathcal{M}^{\mathcal{B}(\Omega)}$ to denote the set of all random fields over $\mathcal{B}(\Omega)$. See for instance Khoshnevisan (2002) for more details on random fields.

Our most important random field will be

$$\tilde{M}^{N^n} = \left\{ \tilde{M}^{N^n}(A) = \max_{1 \le i \le n : x_{ni} \in N^n \cap A} Y_{ni}, A \in \mathcal{B}(\Omega) \right\} \tag{4}$$

with the convention that the maximum over the empty set is $-\infty$. We note that this is a function from $\mathcal{Q}^{\Omega}$ to $\mathcal{M}^{\mathcal{B}(\Omega)}$. The following operator on random fields is also important.

Definition 6 *The pseudo-argmax measure $F : \mathcal{M}^{\mathcal{B}(\Omega)} \to \mathbb{R}_+^{\mathcal{B}(\Omega)}$ is defined by*

$$F(A, M) = \mathbb{P}\left(M(A) \ge M(\Omega)\right)$$

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

We note that $F(\cdot, M)$ is a set function in $\mathbb{R}_+^{\mathcal{B}(\Omega)}$. It is clear from our definitions that

$$F(\cdot; \tilde{M}^{N^n}) = \tilde{T}^{N^n}$$

which is illustrated in the following commutative diagram.

$$\begin{tikzcd}
Q^{\Omega} \arrow[r, "\tilde{M}"] \arrow[dr, "\tilde{T}"'] & \mathcal{M}^{\mathcal{B}(\Omega)} \arrow[d, "F"] \\
 & \mathbb{R}_+^{\mathcal{B}(\Omega)}
\end{tikzcd}$$

We will use this commutative property to derive convergence in $\tilde{T}$ in Definition 3 from convergence in $\tilde{M}$.

3.2 Max-fields

When considering the asymptotic properties of $\tilde{M}^{N^n}$, we have to worry about two things. Firstly, although we know that $F(\cdot, \tilde{M}^{N^n})$ is a probability measure on Ω, we do not know that this is true for a any candidate limiting random field. Thus, we need a set of conditions on M to ensure that $F(\cdot; M)$ is a probability measure. Secondly, to discuss limiting behavior we need a notion of convergence, and it should have the property that F is continuous under this definition with respect to the weak topology on $\mathbb{R}_+^{\mathcal{B}(\Omega)}$.

Definition 7 *Let $M : \mathcal{S} \times \mathcal{B}(\Omega) \to \mathbb{R}$ be a random field over $\mathcal{B}(\Omega)$. We call M an (independence) max-field if the following seven properties hold:*

1. *$M(A)$ and $M(B)$ are independent random variables whenever $A \cap B = \emptyset$;*

2. *If $I = A \cup B$ then $M(I) = \max\{M(A), M(B)\}$;*

3. *$|M(A)| < \infty$ almost surely or $M(A) = -\infty$ almost surely;*

4. *If $A_1 \supseteq A_2 \ldots$, and $\bigcap_n A_n = \emptyset$, then $M(A_n) \to -\infty$ almost surely;*

5. *If $M(\emptyset) = -\infty$;*

6. *If $M(A) = -\infty$ almost surely, $M(\Omega \setminus A) > -\infty$ almost surely;*

7. *If $M(A) > -\infty$ almost surely, $M(\cdot, A)$ is an absolutely continuous probability measure on $\mathbb{R}$ with respect to Lebesgue measure*

The assumptions in Definition 7 have been chosen to enable us to prove the following lemma.

Lemma 1 *If the random field M is a max-field, then the pseudo argmax measure $F(\cdot; M)$ is a probability measure over $\mathcal{B}(\Omega)$.*

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

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

for all $A \in \mathcal{B}(\Omega)$. Furthermore, $M(\emptyset) = -\infty$ and $M(\Omega) > -\infty$ by property 6 and 7. Hence,

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

We need to demonstrate countable additivity. As a first step, we establish finite additivity. We introduce a new notation for the residual set $A_{n+1} = \Omega \setminus \bigcup_{i=1}^{n} A_i$, and the events

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

It is evident that $F(A_i; M) = 0$ if $M(A_i) = -\infty$ so let us assume they are not. By absolute continuity, the B_i's are almost surely disjoint. Hence,

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

For countable additivity, it suffices to show that if $A_1 \supseteq A_2 \supseteq A_3 \dots$ such that $\cap_n A_n = \emptyset$, then $F(A_n; M) \to 0$. However, by Definition 7, $M(A_n) \to -\infty$ almost surely. Furthermore,

$$\max\{M(A_n), M(\Omega \setminus A_n)\} = M(\Omega) > -\infty$$

almost surely. Hence,

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

and the proof is complete. $\square$

3.3 Derivation of calculation methods

We define a notion of convergence on $\mathcal{M}^{\mathcal{B}(\Omega)}$ under which the pseudo argmax-measure map

$$F : \mathcal{M}^{\mathcal{B}(\Omega)} \to \mathbb{R}^{\mathcal{B}(\Omega)}$$

is continuous with respect to the weak topology on $\mathbb{R}^{\mathcal{B}(\Omega)}$. This gives us a method to calculate the argmax measure.

Definition 8 *A sequence of max-fields M_n on $\mathcal{B}(\Omega)$ is said to m-converge to the max-field M ($M_n \stackrel{m}{\to} M$) if there exists a sequence $g_n : \mathbb{R} \to \mathbb{R}$ of strictly increasing functions such that*

$$g_n(M_n(A)) \Rightarrow M(A). \tag{5}$$

for all A with

$$F(\partial A, M) = 0$$

Theorem 1 *Let $\{M_n\}$ and M be max-fields such that*

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

Then

$$F(\cdot, M_n) \Rightarrow F(\cdot, M)$$

where $F(\cdot, M)$ is the pseudo argmax-measure.

Before proving the theorem, we state an important corollary illustrating how it can be used.

Corollary 1 *Suppose there exists a max-field M^Λ such that for all $N^n \in \mathcal{N}^\Lambda$*

$$\tilde{M}^{N^n} \overset{m}{\to} M^\Lambda$$

Then, the argmax measure T^Λ exists and is given by

$$T^\Lambda = F(\cdot; M^\Lambda). \tag{6}$$

Proof of corollary. We note that

$$\tilde{T}^{N^n} = F(\cdot; \tilde{M}^{N^n})$$

and apply Theorem 1 to conclude that

$$\tilde{T}^{N^n} \Rightarrow F(\cdot; M^\Lambda)$$

for all $\{N^n\} \in \mathcal{N}^\Lambda$. By Definition 5, T^Λ is the argmax measure. □

Proof of Theorem 1. Let $A \subseteq \Omega$ be measurable with $F(\partial A; M) = 0$. We seek to show that $F(A; M_n) \to F(A; M)$, and consider three cases.

Case 1. $M(A), M(A^c) > -\infty$ a.s.. By the assumption of m-convergence and $F(\partial A; M) = 0$, we can find a sequence of strictly increasing functions g_n such that

$$\begin{aligned} g_n(\tilde{M}_\mu^{N^n}(A)) &\Rightarrow M(A) \\ g_n(\tilde{M}_\mu^{N^n}(A^c)) &\Rightarrow M(A^c) \end{aligned}$$

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

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

By Definition 7, $M(A)$ and $M(A^c)$ are absolutely continuous with respect to lesbesgue measure and independent, and therefore their difference is absolutely continuous. Hence,

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

where we use absolute continuity to conclude that 0 is a point of continuity of $M(A) - M(A^c)$. Therefore, we get

$$F(A; M_n) \to F(A; M)$$

Case 2. $M(A) = -\infty$ *a.s.* From Definition 7, $M(A^c) > -\infty$ almost surely, which means that $F(A; M) = 0$. Furthermore,

$$\begin{array}{rcl} g_n\left(M_n(A)\right) & \Rightarrow & -\infty. \\ g_n\left(M_n(A^c)\right) & \Rightarrow & M(A^c) > -\infty. \end{array}$$

We can find K such that $\mathbb{P}(M(A^c) > K) = 1 - \epsilon$, and n_0 such that for all $n \geq n_0$, $\mathbb{P}\left(g_n\left(M_n(A)\right) < K\right) > 1 - \epsilon$ and $\mathbb{P}\left(g_n\left(M_n(A^c)\right) > K\right) > 1 - 2\epsilon$. Then, for all $n \geq n_0$, $\mathbb{P}(M_n(A) > M_n(A^c)) < 3\epsilon$. As ϵ was arbitrary, we get

$$F(A; M_n) \to 0 = F(A; M).$$

Case 3. $M(A^c) = -\infty$. We use $F(A, M_n) = 1 - F(A^c, M_n)$ to conclude from Case 2 that

$$F(A, M_n) \to 1.$$

Furthermore, $F(A; M) = 1$ as

$$\begin{array}{rcl} F(A; M) & = & P(M(A) > M(A^c)) \\ & = & 1. \end{array}$$

and we get that

$$F(A; M_n) \to F(A; M)$$

in this case as well

□

4 Argmax measure for homoscedastic regression models

The result in Corollary 1 shows that the tools developed in the previous section give a method for calculating the argmax measure that is workable insofar it is possible to find a max-field M_μ^Λ to which $\tilde{M}_\mu^{N^n}$ m-converges for all $N^n \in \mathcal{N}^\Lambda$.

In this section we make a particular choice

$$Y_{ni} = m(x_{ni}) + \epsilon_{ni}, \tag{7}$$

for $i = 1, \ldots, n$, where $m : \Omega \to \mathbb{R}$ is a given deterministic regression function and $\{\epsilon_{ni}\}$ are independent and identically distributed (i.i.d.) error terms with a common distribution function H. This is a homoscedastic regression model, corresponding to a measure index

$$\mu(x) = H(\cdot - m(x)). \tag{8}$$

In order to to find the limiting behavior of the empirical max-field $\tilde{M}_\mu^{N^n}$ defined in (4), we note that for all A with $\Lambda(A) > 0$, $|A \cap N^n| \to \infty$ as $n \to \infty$, which means that maximum is taken over a large number of independent random variables. Thus, the natural choice is to apply extreme value theory.

We will divide the exposition into four subsections. First we state a classical result in extreme value theory for $m \equiv 0$, and its specific counterpart related to offers $H \sim \mathrm{Exp}(s)$ having an exponential distribution with mean s. The second subsection develops the extreme value theory for exponential offers with varying $m(x)$, in order to calculate a max-field M_μ^Λ to which $\tilde{M}_\mu^{N^n}$ m-converges for an appropriate sequence g_n of monotone transformations. Then Corollary 1 is applied in order to calculate the argmax measure T_μ^Λ. The fourth subsection considers more briefly other distributions H than the exponential.

4.1 Some extreme value theory

The following theorem is a key result in extreme value theory, see for instance Fisher and Tippett (1928), Gnedenko (1943), Leadbetter et al. (1983), Gumbel (2004) and Resnick (2008).

Theorem 2 (Fisher-Tippet-Gnedenko Theorem.) *Let $\{Y_n\}$ be a sequence of independent and identically distributed (i.i.d.) random variables and let $M^n = \max\{Y_1, Y_2, \dots, Y_n\}$. If there exist sequences $\{a_n\}$ and $\{b_n\}$ with $a_n > 0$ such that*

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

for all $x \in \mathbb{R}$, then G belongs to either the Gumbel, the Frechet, or the Weibull family.

Under a wide range of distributions of Y_n, convergence does occur, and for most common distributions the convergence is to the Gumbel(γ, β) law, whose distribution function has the form

$$G(x; \gamma, \beta) = \exp\left(-\exp(-\frac{x-\gamma}{\beta})\right)$$

for some parameters γ and β and $x \in \mathbb{R}$. We can give a more precise statement of Gumbel convergence with $a_n = 1$ and $b_n = \log(n)$ when the random variables Y_i have a standard exponential distribution, see for instance Resnick (2008) for a proof.

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

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

4.2 Exponential offers

It turns out that the argmax theory for homoscedastic regression models depends crucially on the error distribution H, and the exponential distribution is an important boundary between more light and heavy tailed distributions. Therefore, we treat $H \sim \text{Exp}(s)$ separately in this subsection.

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

Ordinary extreme value theory assumes that random variables are independently and identically distributed. However, in our case we do not have identically distributed random variables, as the additive term $m(x)$ varies over space (for references on the theory of extremes with non-identically distributed random variables, see for example Weissman (1975), Horowitz (1980) and Hüsler (1986)). Thus, we prove a result characterizing the max-field with $H \sim \text{Exp}(s)$ and $m(x)$ varying.

Theorem 3 *Let $\tilde{M}_\mu^{N^n}(A)$ be as defined in (4), with $Y_{ni} - m(x_{ni}) \sim Exp(s)$ independently for $i = 1, \ldots, n$ and $s > 0$. Suppose Λ is a probability measure on the Borel σ-algebra on Ω and that the following properties hold:*

1. *m is bounded*
2. *$\{N^n\}_{n\geq 1} \in \mathcal{N}^\Lambda$*
3. *$\Lambda(\bar{D}_m) = 0$, where $D_m = \{x \in \Omega : m(x) \text{ is discontinuous at } x\}$ and $\bar{D}_m = closure(D_m)$.*

Then (5) holds with $g_n(y) = y/s - \log(n)$, i.e.

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

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

$$M_\mu^\Lambda(A) = \log\left(\int_A e^{m(x)/s}\Lambda(dx)\right) + \mathit{Gumbel}(A). \tag{9}$$

The notation Gumbel(A) refers to a standard Gumbel(0,1) random variable, with Gumbel(A) and Gumbel(B) independent for all $A \cap B = \emptyset$.

Proof. After a standardization $Y_{ni} \leftarrow Y_{ni}/s$, we may without loss of generality assume $s = 1$.

Let $A \subset \Omega$ with $\Lambda(\partial A) = 0$. We note that we have weak convergence of P^{N^n} to Λ when both measures are restricted to $A \cap \bar{D}_m^c$, and that on this set m is a continuous bounded function. Thus, by the properties of weak convergence (cf. e.g. Billingsley 1999), we get

$$\begin{array}{lll} \frac{1}{n} & \sum_{1\leq i\leq n: x_{in}\in A} & e^{m(x_{in})} \\ & = & \int_{A\cap\bar{D}_m^c} e^{m(x)} dP^{N^n}(x) + \frac{1}{n}\sum_{1\leq i\leq n: x_{in}\in A\cap\bar{D}_m} e^{m(x_{in})} \\ & \to & \int_{A\cap\bar{D}_m^c} e^{m(x)} d\Lambda(x) + 0 \\ & = & \int_A e^{m(x)} d\Lambda(x) \end{array} \tag{10}$$

The last sum on the first line tends to 0 as we can write

$$\bar{m} = \sup_{x\in\Omega} m(x) \tag{11}$$

and get

$$\begin{array}{lll} \frac{1}{n}\sum_{1\leq i\leq n: x_{in}\in A\cap\bar{D}_m} e^{m(x_{in})} & \leq & \frac{1}{n}\sum_{1\leq i\leq n: x_{in}\in A\cap\bar{D}_m} e^{\bar{m}} \\ & = & \frac{1}{n} n P^{N^n}(A\cap\bar{D}_m)e^{\bar{m}} \\ & \to & \Lambda(A\cap\bar{D}_m)e^{\bar{m}} \\ & \leq & \Lambda(\bar{D}_m)e^{\bar{m}} \\ & = & 0, \end{array}$$

where in the second last step we utilized that

$$\Lambda\left(\partial(A\cap\bar{D}_m)\right) \leq \Lambda(\partial A) + \Lambda(\partial\bar{D}_m) \leq \Lambda(\partial A) + \Lambda(\bar{D}_m) = 0 + 0 = 0,$$

since $\bar{D}_m$ is a closed set. We can use (10) to derive the max-field directly. With $g_n(y) = y - \log(n)$ we get that if $Z_n = \log(\mathbb{P}(g_n\left(\tilde{M}_\mu^{N^n}(A)\right) \leq y))$ it holds that

$$\begin{array}{lll} Z_n & = & \log(\mathbb{P}(\tilde{M}_\mu^{N^n}(A) \leq y + \log(n))) \\ & = & \sum_{1\leq i\leq n; x_{ni}\in A} \log(1 - \exp(-y - \log(n) + m(x_{ni}))) \\ & = & -\exp(-y)\frac{1}{n}\sum_{1\leq i\leq n; x_{ni}\in A} \exp(m(x_{ni}) + e(n) \\ & \to & -\exp(-y)\int_A \exp(m(x))\Lambda(dx) \\ & = & -\exp(-y + \log\left(\int_A \exp(m(x))\Lambda(dx)\right)) \end{array}$$

where we recognize the last line as the logarithm of a Gumbel distribution function with an additive term $\log\left(\int_A \exp(m(x))\Lambda(dx)\right)$ as required. Thus, we have proved our result provided we can verify that the error term $e(n) \to 0$.

To show this we note that

$$e(n) = \sum_{1\leq i\leq n; x_{ni}\in A} \log(1-\exp(-y-\log(n)+m(x_{ni}))) + \exp(-y-\log(n)+m(x_{ni}))$$

Indeed, using the well-known result that

$$|\log(1-x)+x| \leq \frac{x^2}{1-x}$$

we get that

$$|e(n)| \leq \sum_{1\leq i\leq n; x_{ni}\in A} \frac{\exp(-2y-2\log(n)+2m(x_{ni}))}{1-\exp(-y-\log(n)+m(x_{ni}))} \to 0$$

and we have proved our result. $\square$

Proposition 2 *The random field defined by*

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

is a max-field in the sense of Definition 7 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 can be shown to hold by the properties of the Gumbel distribution. Property 3 holds as m is bounded. Property 4Âăand 5Âăhold as $\lim_{x\to 0}\log(x) = -\infty$. Property 6 can be verified directly from the expression of M, and Property 7 is true as the Gumbel distribution is absolutely continuous. $\square$

4.2.2 Argmax distribution

In Corollary 1, it was shown that the limiting behavior of $\tilde{M}_{\mu}^{Nn}$ determines the argmax measure. Thus, we can use the limit derived in Theorem 3 together with Proposition 2 and Corollary 1 to derive the argmax measure associated with μ and Λ.

Theorem 4 *Let $\mu(x) = m(x) + Exp(s)$ and let Λ be a probability measure on Ω. Suppose that Λ and m jointly satisfy the conditions in Theorem 3. Then the argmax measure T_μ^Λ exists and is given by the exponentially tilted distribution*

$$T_\mu^\Lambda(A) = C \int_A e^{m(x)/s} \Lambda(dx), \tag{12}$$

where

$$C = \left(\int_\Omega e^{m(x)/s} \Lambda(dx) \right)^{-1} \tag{13}$$

is a normalizing constant. In particular, if Λ has a density function λ with respect to Lebesgue measure ν on Ω, then T_μ^Λ has the density function

$$t_\mu^\Lambda(x) = C\lambda(x) \exp(m(x)/s) \tag{14}$$

for $x \in \Omega$, i.e. $T_\mu^\Lambda(A) = \int_A t_\mu^\Lambda(x)\nu(dx)$ for all Borel sets $A \subset \Omega$.

Proof. After standardizing data $Y_{ni} \leftarrow Y_{ni}/s$, we may, without loss of generality, assume that $s = 1$. Proposition 2 states that M_μ^Λ, defined as in Theorem 3. is a max-field, and in order to find its pseudo argmax measure we let $G(x) = G(x; 0, 1) = e^{-e^{-x}}$ denote the distribution function of a standard Gumbel distribution and put $L(A) = \log \left(\int_A e^{m(x)} d\Lambda(x) \right)$. Then

$$\begin{aligned}
F(A; M_\mu^\Lambda) &= \mathbb{P}\left(M_\mu^\Lambda(A) > M_\mu^\Lambda(\Omega \setminus A) \right) \\
&= \int_{-\infty}^{\infty} \mathbb{P}\left(M(A) \in dr \right) \mathbb{P}\left(M(\Omega \setminus A) < r \right) \\
&= \int_{-\infty}^{\infty} G'\left(r - L(A) \right) G\left(r - L(\Omega \setminus A) \right) dr \\
&= \int_{-\infty}^{\infty} e^{-r+L(A)} e^{-e^{-r+L(A)}} e^{-e^{-r+L(\Omega \setminus A)}} dr \\
&= e^{L(A)} \int_{-\infty}^{\infty} \exp(-r) \exp\left(-e^{-r+L(\Omega)} \right) dr \\
&= C \int_A e^{m(x)} \Lambda(dx)
\end{aligned}$$

for all Borel sets A.

Then note that Theorem 3 implies that

$$\tilde{M}_\mu^{N^n}(A) - \log(n) \Rightarrow M_\mu^\Lambda(A) \tag{15}$$

holds for $\{N^n\}_{n \geq 1} \in \mathcal{N}^\Lambda$ and all Borel sets A with $\Lambda(\partial A) = 0$. It can be shown that if $\Lambda(\partial A) > 0$, we have $F(\partial A, M_\mu^\Lambda) > 0$. Consequently, $\tilde{M}_\mu^{N^n} \xrightarrow{m} M_\mu^\Lambda$. Finally, Corollary 1 implies that the argmax measure $T_\mu^\Lambda = F(\cdot; M_\mu^\Lambda)$ exists and is given by (12) $\square$

Theorem 4 is remarkably simple and explicit. It turns out that this is due to the memoryless property of the exponential distribution. Indeed, suppose $\{x_{ni}\}_{i=1}^n$ is an i.i.d. sample from Λ, with n large. Recall definition (11) of

$\bar{m}$, put $I = \arg\max_{1\le i\le n} Y_{ni}$ and assume for simplicity $s = 1$. Then, for any $i = 1,\dots,n$,

$$\begin{aligned}
\mathbb{P}(I = i) &\approx \mathbb{P}(Y_{ni} \ge \bar{m})\mathbb{P}(I = i|Y_{ni} \ge \bar{m}) \\
&\approx e^{-(\bar{m}-m(x_{ni}))}/\left(n\mathbb{P}(m(X) + \epsilon \ge \bar{m})\right) \\
&\propto e^{m(x_{ni})}/\left(n\int_\Omega e^{m(x)}P^{N_n}(dx)\right) \\
&\approx e^{m(x_{ni})}/\left(n\int_\Omega e^{m(x)}\Lambda(dx)\right).
\end{aligned}$$

In the first step we utilized that $\max_{1\le i\le n} Y_{ni} \ge \bar{m}$ holds with probability close to 1 when n is large, and in the second step approximated the number of i for which $Y_{ni} = m(x_{ni}) + \epsilon_{ni} \ge \bar{m}$ as

$$\sum_{i=1}^{n} 1_{\{m(x_{ni})+\epsilon_{ni}\ge\bar{m}\}} \approx n\mathbb{P}(m(X) + \epsilon \ge \bar{m}),$$

where $\{x_{ni}, \epsilon_{ni}\}_{i=1}^n$ is an i.i.d. sample from $\Lambda \times \text{Exp}(1)$. Finally, we used the memoryless property of the exponential distribution to deduce that all indeces i with $Y_{ni} \ge \bar{m}$ have the same conditional probability of being the argmax, i.e. $I = i$.

4.3 Non-exponential offers

In the previous subsection, we found that with m fixed, exponentially distributed offers gave us a one-parameter family of argmax distributions, indexed by $s > 0$. We will now provide arguments for other error distributions and find that the exponential case provides the borderline between more light- and heavy-tailed distributions. Loosely speaking, for light-tailed distribution, it is only the extremal behavior of m that determines the asymptotic argmax distribution, whereas m has no asymptotic impact for heavy-tailed distributions.

4.3.1 Light-tailed error distributions

Formally, the light-tailed case corresponds to the class of distribution for which the moment generating function of the disturbance function is finite for the whole real line. For simplicity, we assume that the support of the continuous distribution H has an upper bound

$$K = \sup\{x;\, H(x) < 1\} < \infty,$$

and that m is not constant. Applying the identity transformation $g_n(y) = y$, we deduce that

$$\tilde{M}_\mu^{N^n}(A) \Rightarrow M_\mu^\Lambda(A) = K + \sup_{x\in A} m(x).$$

The limiting max field M_μ^Λ is a degenerate in the sense that $M_\mu^\Lambda(A)$ has a one point distribution, so that the absolute continuity Property 7 of Definition 7 is violated. Therefore we cannot use Theorem 1 in order to deduce the argmax measure, but have to employ a more direct argument.

Given any $\varepsilon > 0$, we let $h(x) = H'(x)$ and define the measure

$$\Lambda_\varepsilon(A) = C \int_A \frac{h\left(K - \varepsilon + \bar{m} - m(x)\right)}{H[K - \varepsilon + \bar{m} - m(x), K]} \Lambda(dx),$$

with $h(x) = 0$ if $x > K$, $\bar{m}$ as in (11), the convention $H([K', K]) = 0$ when $K' > K$, and $C = C(\varepsilon)$ a normalizing constant chosen so that $\Lambda_\varepsilon(\Omega) = 1$. Assume further that a limit measure $\Lambda_{\max}$ exists, supported on the set

$$\Omega_{\max} = \{x \in \Omega; m(x) = \bar{m}\}$$

where m is maximal, such that

$$\Lambda_\varepsilon \Rightarrow \Lambda_{\max} \text{ as } \varepsilon \to 0. \tag{16}$$

It is reasonable to assume that Λ_ε should approximate the conditional distribution of X_n given that $Y_{n:n} = \max_{1 \le i \le n} Y_{ni} = \bar{m} + K - \varepsilon$. (A more formal argument is provided below).Hence (16) suggests that

$$T_\mu^\Lambda = \Lambda_{\max}, \tag{17}$$

since $Y_{n:n}$ tends in probability to $\bar{m} + K$ as n grows. In order to establish (17) according to Defintion 5, we need a slightly stronger condition though than (16), as the following theorem reveals:

Theorem 5 *For any $\varepsilon > 0$, put*

$$P_\varepsilon^{N^n}(A) = C_n \int_A \frac{h\left(K - \varepsilon + \bar{m} - m(x)\right)}{H[K - \varepsilon + \bar{m} - m(x), K]} P^{N^n}(dx),$$

where $C_n = C_n(\varepsilon)$ is a normalizing constant assuring that $P_\varepsilon^{N^n}(\Omega) = 1$, and

$$Q_n(\varepsilon) = \int_0^\varepsilon \frac{d\varepsilon'}{C_n(\varepsilon')} = \int_\Omega H\left([K - \varepsilon + \bar{m} - m(x), K]\right) P^{N^n}(dx).$$

Assume that

$$P_{Q_n^{-1}(c/n)}^{N^n} \Rightarrow \Lambda_{\max} \text{ as } n \to \infty \tag{18}$$

uniformly for all $c \in (0, \bar{c}]$, for any $\bar{c} > 0$, with Q_n^{-1} the inverse function of Q_n. Then (17) holds.

Proof: According to Definition 5, we need to prove $\tilde{T}_\mu^{N^n} \Rightarrow \Lambda_{\max}$ for any $\{N^n\}_{n\geq 1} \in \mathcal{N}^\Lambda$. Let $Z_n = \bar{m} + K - Y_{n:n}$. We first note that

$$\begin{array}{rcl} P(X_n = x_{ni}|Z_n) & = & h(\bar{m} + K - m(x_{ni}) - Z_n) \prod_{j\neq i} H(\bar{m} + K - m(x_{nj}) - Z_n) \\ & \propto & \frac{h(\bar{m}+K-m(x_{ni})-Z_n)}{H(\bar{m}+K-m(x_{ni})-Z_n)} \end{array}$$

where X_n is defined as in (2). By conditioning on Z_n we notice that

$$\tilde{T}_\mu^{N^n}(A) = \int_0^\infty P_\varepsilon^{N^n}(A) F_{Z_n}(d\varepsilon). \tag{19}$$

Z_n furthermore has the property that

$$nQ_n(Z_n) \Rightarrow Exp(1)$$

Indeed, for $x > 0$, we can use the monotonicity of Q_n to deduce that

$$\begin{array}{rcl} P(nQ_n(Z_n) \leq x) & = & P\left(Z_n \leq Q_n^{-1}(x/n)\right) \\ & = & 1 - \prod_{i=1}^n (1 - H[K + \bar{m} - m(x_{ni} - Q_n^{-1}(x/n), K]) \\ & \to & 1 - e^{-x} \end{array}$$

where the last step uses the well known fact

$$\prod_{i=1}^n (1 - a_{n,i}) \to e^{-a}$$

if

$$\lim_{n\to\infty} \sum_{i=1}^n a_{n,i} = a$$

and $\lim_{n\to\infty} \max a_{n,i} = 0$. These conditions hold in our case as

$$\begin{array}{rcl} \sum_{i=1}^n H[K + \bar{m} - m(x_{ni} - Q_n^{-1}(x/n), K] & = & nQ_n(Q_n^{-1}(x/n)) \\ & = & x \end{array}$$

and $\lim_{n\to\infty} \max H[K + \bar{m} - m(x_{ni} - Q_n^{-1}(x/n), K] = 0$ assuming that H has no point mass on K.
Thus, $nQ_n(Z_n) \Rightarrow Exp(1)$, and we conclude the proof by performing a change of variable $c = nQ_n(\epsilon)$ on (19) to get

$$\tilde{T}_\mu^{N^n}(A) = \int_0^\infty P_{Q_n^{-1}(c/n)}^{N^n}(A) F_{nQ_n(Z_n)}(dc). \tag{20}$$

Letting $e(c,n) = |\Lambda_{\max}(A) - P_{Q_n^{-1}(c/n)}^{N^n}(A)|$ which tends uniformly to 0 on $[0, \bar{c})$ for any $\bar{c}$, we get that

$$|\tilde{T}_\mu^{N^n}(A) - \Lambda_{\max}(A)| \leq \sup_{c\in[0,\bar{c})} e(c,n) P(nQ_n(Z_n) \in [0,\bar{c})) + P(nQ_n(Z_n) \notin [0,\bar{c}))$$

which can be made arbitrarily small. Thus, our proof is completed.

$\square$

4.3.2 Heavy-tailed error distributions

It can be shown that the class of heavy-tailed distributions corresponds to those for which the moment-generating function is undefined for positive values. For simplicity, we consider the class of Pareto distributions with shape parameter $\alpha > 0$ and scale parameter 1, i.e.

$$H(x) = \text{Pareto}(x; \alpha, 1) = 1 - x^{-\alpha}$$

for $x \geq 1$. Then Theorem 2 holds with $b_n = 0$, $a_n = n^{1/\alpha}$, and

$$G(x) = \text{Frechet}(x; \alpha, 1, 0) = \exp(-x^{-\alpha})$$

for $x > 0$ has a Frechet distribution with shape parameter α, scale parameter 1 and location parameter 0. Since a_n increases with n at polynomial rate, it turns out that any local variation of the bounded function m has no impact on the asymptotic max field, as the following result reveals:

Theorem 6 *Let $\tilde{M}_\mu^{N^n}(A)$ be as defined in (4), with $Y_{ni} - m(x_{ni}) \sim$ Pareto$(\alpha, 1)$ independently for $i = 1, \ldots, n$. Suppose Λ is a probability measure on the Borel σ-algebra on Ω and that properties 1-3 of Theorem 3 hold.*

Then (5) holds with $g_n(y) = y/n^{1/\alpha}$, i.e.

$$\tilde{M}_\mu^{N^n}(A)/n^{1/\alpha} \Rightarrow M_\mu^\Lambda(A) = \Lambda(A)^{1/\alpha} \mathit{Frechet}_\alpha(A) \tag{21}$$

for all A with $\Lambda(\partial A) = 0$. Moreover the argmax measure exists and is given by

$$T_\mu^\Lambda = \Lambda. \tag{22}$$

In the notation, $\text{Frechet}_\alpha(A)$ refers to a $\text{Frechet}(\alpha, 1, 0)$ distributed random variable for any Borel set $A \subset \Omega$, which is independent of $\text{Frechet}_\alpha(B)$ for B such that $A \cap B = \emptyset$.

Proof. We begin by (21). Let A be a measurable set with $\Lambda(\partial A) = 0$. Then, if $F_{n,A}$ is the distribution function of $\tilde{M}_\mu^{N^n}(A)/n^{1/\alpha}$ we have

$$\begin{aligned} \log F_{n,A}(y) &= \log\left(\prod_{x_{n,i} \in A} P(Y_{n,i} + m(x_{n,i}) \leq n^{1/\alpha} y)\right) \\ &= \sum_{1 \leq i \leq n, x_{n,i} \in A} \log\left(1 - (n^{1/\alpha} y - m(x_{n,i}))^{-\alpha}\right) \\ &= \sum_{1 \leq i \leq n} \log\left(1 - \frac{(y - n^{-1/\alpha} m(x_{n,i}))^{-\alpha}}{n}\right)^n \frac{I(x_{n,i} \in A)}{n} \\ &= \sum_{1 \leq i \leq n} f(n,i) h(n,i) \end{aligned} \tag{23}$$

As m is bounded, $f(n,i) \to -y^{-\alpha}$ uniformly over i. Therefore, we get

$$\begin{aligned} \lim_{n\to\infty} \log F_{n,A}(y) &= \lim_{n\to\infty} \textstyle\sum_{1\le i\le n} f(n,i)h(n,i) \\ &= -y^{-\alpha} \lim_{n\to\infty} \textstyle\sum_{1\le i\le n} \frac{I(x_{n,i}\in A)}{n} \\ &= -y^{-\alpha}\Lambda(A) \end{aligned} \tag{24}$$

where the last step uses weak convergence of P^{N^n} to Λ. After exponentiation we recognize the right-hand side, as required, as the distribution function of $\Lambda(A)^{1/\alpha}\text{Frechet}_\alpha(A)$.

It remains to prove (22). To this end, we notice that pseudo argmax measure of M_μ^Λ equals

$$\begin{aligned} F(A; M_\mu^\Lambda) &= P\left(\Lambda(A)^{1/\alpha}\text{Frechet}_\alpha(A) > \Lambda(A^c)^{1/\alpha}\text{Frechet}_\alpha(A^c)\right) \\ &= P\left(\Lambda(A)\text{Frechet}_1(A) > \Lambda(A^c)\text{Frechet}_1(A^c)\right) \\ &= \Lambda(A), \end{aligned} \tag{25}$$

where the last line follows from the properties of the Frechet distribution. Indeed, if $X, Y \sim \text{Frechet}_1$ independently,

$$\begin{aligned} P\left(\Lambda(A)X > \Lambda(A^c)Y\right) &= \textstyle\int_0^\infty \frac{\Lambda(A)}{y^2}\text{Exp}(-\Lambda(A)y^{-1})\text{Exp}\left(-\Lambda(A^c)y^{-1}\right)dy \\ &= \Lambda(A)\textstyle\int_0^\infty 1/y^2\exp(-y^{-1})dy \\ &= \Lambda(A), \end{aligned} \tag{26}$$

Since $F(\cdot; M_\mu^\Lambda) = \Lambda$, it follows from (21) that $\tilde{M}_\mu^{N^n} \overset{m}{\to} M_\mu^\Lambda$. Hence, by Corollary 1, $T_\mu^\Lambda = F(\cdot; M_\mu^\Lambda) = \Lambda$ exists. $\square$

5 Examples

We will investigate the accuracy of the asymptotic results for the homoscedastic regression model by simulation, generating n random points x_i on $\mathbb{R}$ or $\mathbb{R}^2$ according to some probability distribution Λ. Then we generate

$$Y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

for some predefined function m, where $\{\epsilon_i\}_{i=1}^n$ are i.i.d. random variabels with distribution H. We then return the max $M = Y_{n:n} = \max_{1\le i\le n} Y_i$ and the argmax $X = x_{n:n}$. We repeat the procedure 10,000 times and draw either histograms or QQ-plots for X and/or M together with their theoretically predicted densities. In the first two examples we consider exponential offers with $H \sim \text{Exp}(1)$.

Example 1 (Optimal exponential offers in one dimension.) We display density plots of X for three one-dimensional examples, when either $\Lambda \sim U(-1,1)$ and $m(x) = |x|$ (Figure 1), $\Lambda \sim \text{Weibull}(2,1)$ and $m(x) = \sqrt{x+1}$ (Figure 2) and $\Lambda \sim \text{LogN}(0,1)$ and $m(x) = -x^2$ (Figure 3). In all cases our theoretical prediction (14) bears out. This illustrates the generality of the results proved in Section 4. $\square$

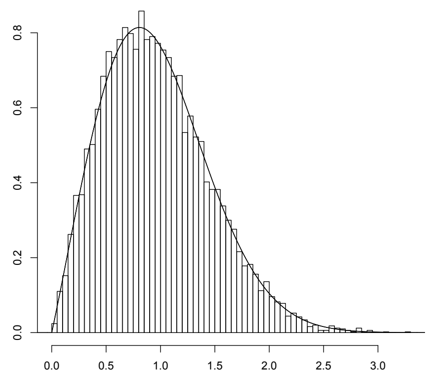

Figure 1: Histogram of argmax distribution when the sampling distribution is $\Lambda \sim \text{Weibull}(2,1)$, $m(x) = \sqrt{x+1}$, $s = 1$ and $n = 1000$. The solid curve is the asymptotic density (14).

Example 2 (Commuting with exponential offers.) The second illustration is the commuting example that motivated this work, as discussed in the introduction. We sample from a uniform distribution over a disc $\Omega = B_{100}(0,0) \subset \mathbb{R}^2$ of radius 100, i.e.

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

with $||x|| = \sqrt{x_1^2 + x_2^2}$ the Euclidean distance. We let

$$m(x) = -0.05 \times ||x||$$

be a function that describes travel costs and record the distance $||X||$ to the origin of the best offer, see Figure 4. We note that the argmax density (14)

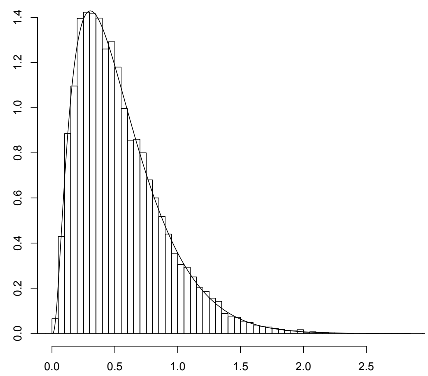

Figure 2: Histogram of argmax distribution when the sampling distribution Λ is a standard lognormal distribution, $m(x) = -x^2$, $s = 1$ and $n = 1000$. The solid curve is the asymptotic density (14).

of X is

$$t_\mu^\Lambda(x) = C\lambda(x)\exp(-cr)1_{\{||x||<100\}}$$

for $c = 0.05$ and a normalizing constant C. By integrating, we get a truncated gamma density

$$f_{||X||}(r) = 2\pi r t_\mu^\Lambda(x) = \frac{2r\exp(-cr)}{100^2}1_{\{0<r<100\}} \tag{27}$$

for the distance to the best offer. We also plot the density of the best value M in Figure 5, corresponding to the distribution $M_\mu^\Lambda(\Omega)$ in (9), which simplifies to

$$\text{Gumbel}\left(\log\left(\int_0^{100}\frac{2re^{-cr}}{100^2}dr\right), 1\right), \tag{28}$$

using the fact that $\int_\Omega \lambda(x)e^{m(x)}d\nu(x) = C^{-1} = 100^{-2}\int_0^{100} 2r\exp(-cr)dr$. It is seen that the finite sample distributions of $\|X\|$ and M are well approximated by their asymptotic limits. $\square$

Example 3 (General offer distributions.) We will now consider more general error distributions H. In particular, we will contrast the behavior when H is light-tailed and heavy-tailed respectively. In both cases we let

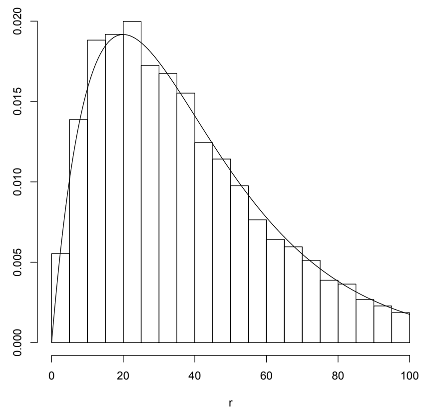

Figure 3: Histogram of distance to origin $||X||$ for argmax of a uniform sample on $B_{100}(0,0)$ when $m(x) = -0.05 \times ||x||$, $s = 1$ and $n = 1000$. The solid curve is the asymptotic truncated gamma density (27).

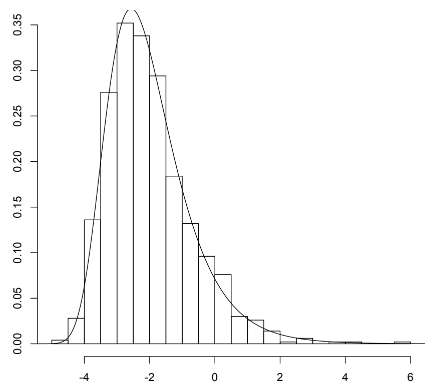

Figure 4: Histogram of best value M on $B_{100}(0,0)$, when $m(x) = -0.05 \times ||x||$, $s = 1$ and $n = 1000$. The solid curve is the asymptotic density corresponding to (28).

$\Omega = [0,1]$ and Λ the uniform distribution on $[0,1]$. For the light-tailed case we consider a uniform $H \sim U(0,s)$ for various choices of s and

$$m(x) = 1_{\{x \in [0.5,1]\}}. \tag{29}$$

According to the theory for light-tailed distribution, the limiting argmax distribution (17) should be uniformly distributed on $[0.5, 1]$.

For the heavy-tailed case we consider $H \sim \text{Pareto}(\alpha, 1)$ and $m(x) = 0.5x$. In this case the theory (22) predicts that despite the varying m, the limiting argmax distribution should equal $\Lambda \sim U(0, 1)$.

The results are displayed in Figures 5 and 6, with different plots for different n. The various colors show different parameters of the H-distribution and illustrate that the rate of convergence depends negatively on the spread of the distribution H.

$\square$

Example 4 (A counterxample.) Let $\Omega = \{0, 1\}$, $\mu(0) \sim U(-1, 0)$, $\mu(1) \sim U(0, 1)$ and $\Lambda = \delta_0$, the point mass at 0. In this case an argmax distribution T_μ^Λ does not exist. Indeed, consider two different multi-sets N^n and $\bar{N}_n$, with empirical distributions

$$\begin{aligned} P^{N^n} &= \left(1 - \tfrac{1}{\sqrt{n}}\right)\delta_0 + \tfrac{1}{\sqrt{n}}\delta_1, \\ P^{\bar{N}^n} &= \delta_0 \end{aligned}$$

respectively that both converge weakly to Λ. However, it is easy to see, either directly, or through max fields, that $\tilde{T}_\mu^{N^n} = \delta_1$ and $\tilde{T}_\mu^{\bar{N}^n} = \delta_0$ for all n. Hence, according to Definition 5, T_μ^Λ does not exist. The problem arises since $\mu(0)$ and $\mu(1)$ have disjoint supports. More generally, it suffices that $\mu(0)$ and $\mu(1)$ have different supports to the right for Definition 5 to fail. $\square$

6 Discussion and Extensions

In this paper we set out to define and prove limit 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 available to address these types of problems by introducing the max-field concept. The usefulness of the developed method is shown when applied to a regression model with homoscedastic error terms. We found that the limiting argmax distribution is nontrivial for exponential white noise, which provides a borderline between more light- and heavy distributions.

There are plenty of potential generalizations and extensions of the theory available on the basis of the work done in this paper, as discussed in Malmberg (2012). Firstly, it is possible to construct a theory where the locations

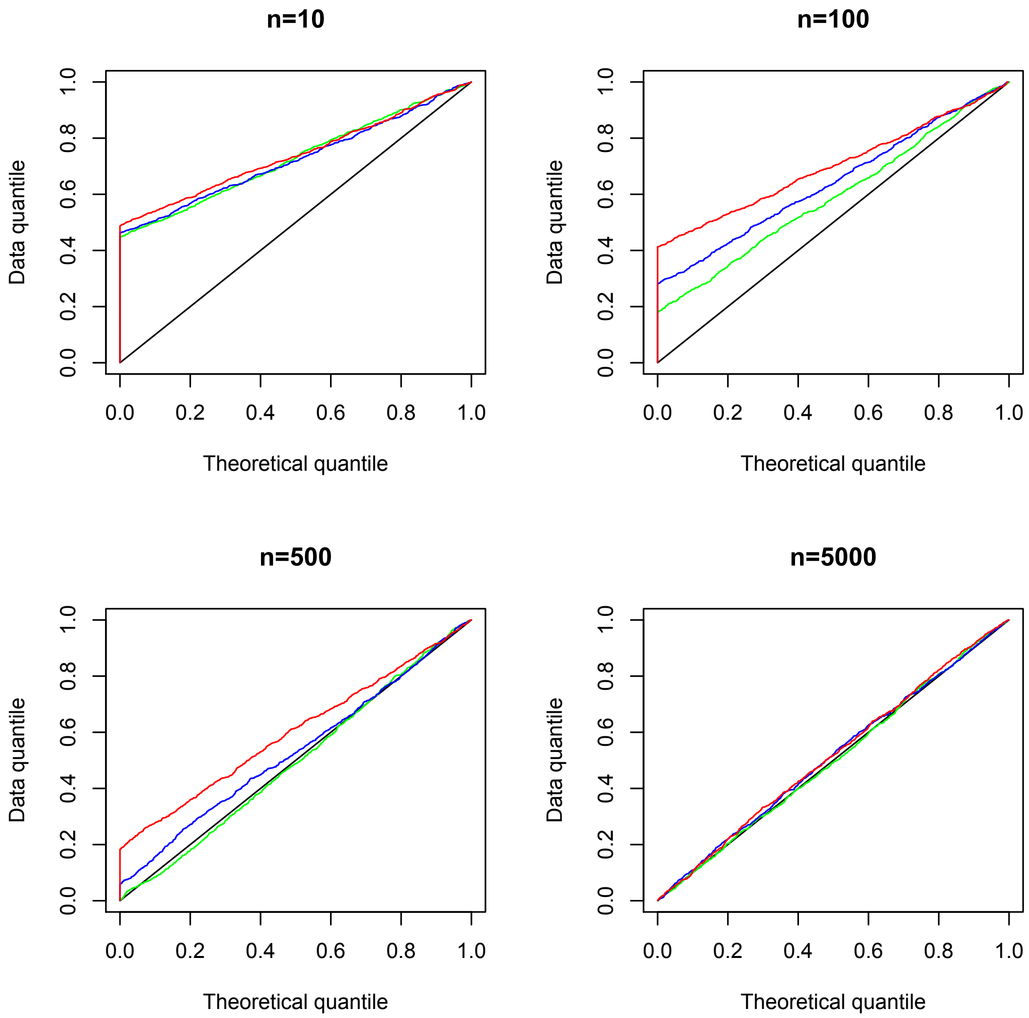

Figure 5: QQ-plots for varying n when $\Lambda \sim U[0,1]$, $m(x)$ is given by (29) and $H \sim U[0,s]$, with $s = 1$ green, $s = 2$ blue, and $s = 5$ red. The asymptotic argmax distribution, given by (17), equals $U(0.5, 1)$.

x_{ni} of offers are not deterministic, but rather allow there to be stochasticity in the selection of points, leading to a doubly stochastic problem. When $\{x_{ni}\}_{i=1}^{n}$ is a point process, this yields an argmax theory of marked point process as the intensity of the underlying point process tends to infinity. In particular, when $(x_{ni}, Y_{ni}) = (x_i, Y_i)$ is an i.i.d. sequence of pairs of random variables, the argmax distribution for a sample of size n is the concomitant

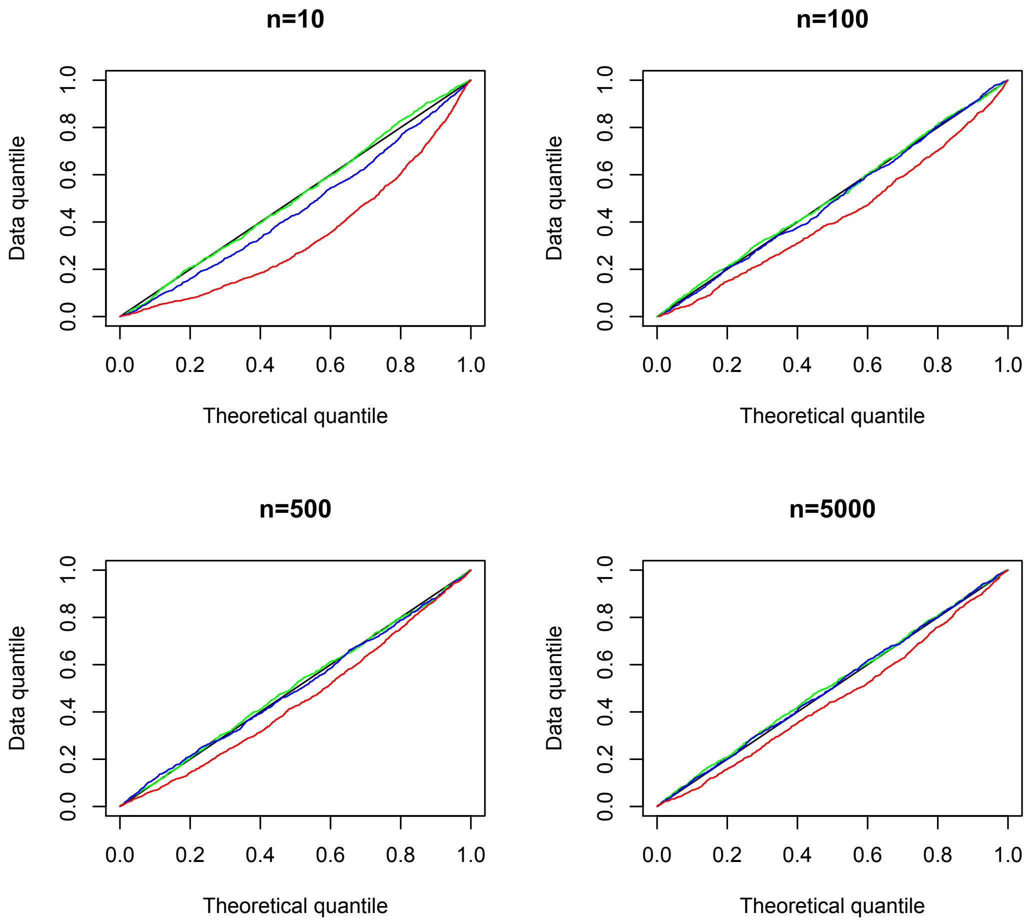

Figure 6: QQ-plots for varying n when $\Lambda \sim U[0,1]$, $m(x) = 0.5x$ and $H \sim \text{Pareto}(\alpha, 1)$, with $\alpha = 0.5$ green, $\alpha = 1$ blue, and $\alpha = 5$ red. The asymptotic argmax distribution (22) thus equals $U(0,1)$.

of the extreme order statistic among $Y_1, \ldots, Y_n$.

Secondly, it is of interest to derive explicit argmax limits for other measure indeces μ than homoscedastic regression models. Generally, if all $\{\mu(x)\}_{x\in\Omega}$ are similar enough, their difference will asymptotically have no impact, so that $T_\mu^\Lambda = \Lambda$, as in (22). On the other hand, if $\{\mu(x)\}_{x\in\Omega}$ differ a lot and can be linearly stochastically ordered, only the stochastically largest distributions will contribute to the limiting argmax distribution, i.e. $T_\mu^\Lambda(A) = \Lambda_{\max}$, a measure supported on the set

$$\Omega_{\max} = \{x \in \Omega; \mu(x) \text{ stochastically largest}\},$$

as in (17). The challenge is to find other non-trivial argmax distributions between these two extremes. One such argmax distribution is provided by (12). Another example is derived from a mixture class

$$\mu(x) = (1 - p(x))U(-1, 0) + p(x)U(0, 1) \tag{30}$$

of probability measures, where $p : \Omega \to [0, 1]$ gives the location dependent mixture between two uniform distributions. It can be seen, when $\{x_{ni}\}_{i=1}^n$ is an i.i.d. sample from Λ, that $b_n = 1$, $a_n = 1/n$ and $g_n(y) = n(y - 1)$ gives a max field

$$M_\mu^\Lambda(A) \sim -\text{Exp}\left((\int_A p(x)\Lambda(dx))^{-1} \right)$$

and argmax law

$$T_\mu^\Lambda(A) = C \int_A p(x)\Lambda(dx)$$

that is a weighted distribution with weight function p, cf. Patil (2002).

Thirdly, it would be interesting to generalize the point process approach described in Chapter 4 of Resnick (2008) for $d = 1$ and stationary mark distributions ($\mu(x) \equiv H$ for some H). This entails establishing weak convergence of the sequence of point processes $\xi_n = \sum_{i=1}^n \delta_{(x_{ni}, g_n(Y_{ni}))}$ as $n \to \infty$ to an appropriate Poisson Random Measure $\xi = \sum_{i=1}^\infty \delta_{(x_i, Y_i)}$ on $\Omega \times \mathbb{R}$. Once this is done, the limiting max field and argmax distributions are

$$M_\mu^\Lambda(A) = \max_{i; x_i \in A} Y_i.$$

and

$$T_\mu^\Lambda(A) = \mathbb{P}(X \in A),$$

respectively, with $X = \arg\max_{x_i} Y_i$.

Fourthly, the max fields are related to extremal processes, as described for instance in Chapter 4 of Resnick (2008). Indeed, when $d = 1$ and $\Omega = (0, 1]$, we may define

$$\mathcal{M}_\mu^\Lambda(t) = M_\mu^\Lambda\left((0, t]\right)$$

and

$$\tilde{\mathcal{M}}_\mu'^{N_n}(t) = g_n\left(\tilde{M}_\mu^{N_n}\left((0, t]\right)\right)$$

for $0 < t \leq 1$. Suppose offer locations are equispaced ($x_{ni} = i/n$), marks stationary ($\mu(x) \equiv H$) and $g_n(y) = (y - b_n)/a_n$, with a_n and b_n the normalizing constants of Theorem 2 when $Y_i \sim H$. Then $\mathcal{M}_\mu^\Lambda$ is an extremal process generated by H and functional weak convergence $\tilde{\mathcal{M}}_{\mu}^{\prime N_n} \Rightarrow \mathcal{M}_{\mu}^{\Lambda}$ can be established on $D(0,1]$, the space of right continuous functions on $(0,1]$ with left hand limits, embedded with the Skorohood topology. Theorem 1 can be viewed as an analogous (marginal) convergence result for max fields corresponding to more general sampling dimensions d, sampling distributions Λ and possibly nonstationary measure indeces with varying $\mu(x)$.

Fifthly, it is possible to allow for dependent offers. For instance, one may consider a triangular array

$$Y_{ni} = m(x_{ni}) + Z(x_{ni}) + \epsilon_{ni}, \quad i = 1, \ldots, n,$$

of offers, with m a deterministic mean function, $Z : \Omega \to \mathbb{R}$ a zero mean random field and $\{\epsilon_{ni}\}_{i=1}^{n}$ zero or constant mean white noise. Such models are are frequently encountered in spatial statistics (Cressie, 1993). In this paper we have focused on models with $Z \equiv 0$ and homoscedastic error terms $\epsilon_{ni} \sim H$, although (30) is another possible choice, with $m(x) = p(x) - 1/2$ and heteroscedasticity, since the variance of (30) depends on x. Conversely, $Z \neq 0$ and $\epsilon_{ni} \equiv 0$ leads to models with no nugget effect and an argmax theory or random fields, since, as $n \to \infty$, the distribution of the argmax X_n should be close to that of

$$X_\infty := \arg\max_{x\in\Omega} \left(m(x) + Z(x)\right),$$

provided $\text{supp}(\Lambda) = \Omega$ and that $m + Z$ is sufficiently regular (for instance continuous), with a unique maximum almost surely.

Acknowledgements

The authors wish to thank Dmitrii Silvestrov for reading through several versions of this manuscript and for providing valuable comments. Ola Hössjer's research was financially supported by the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine.

References

[1] Ben-Akiva, M., N. Litinas and K. Tsunokawa (1985): Continuous spatial choice: The continuous logit model and distributions of trips and urban densities. *Transportation Research*, **A19**, 119-145.

[2] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed., Wiley, Hoboken, NJ.

[3] Cox, D.R. and Isham, V. (1980). *Point processes*, Chapman and Hall, London.

[4] Cressie, N. (1993). *Statistics for spatial data*, revised version, Wiley, New York.

[5] Dagsvik, J. K. (1994): Discrete and continuous choice, max-stable processes and Independence from Irrelevant Attributes. *Econometrica*, **62**, 1179-1205.

[6] Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed., Arnold, London.

[7] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distributions of largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180-190.

[8] Gnedenko, B.V. (1943). Sur la distribution limité du terme d'une série aléatoire. *Ann. Math.* **44**, 423-453.

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

[10] Horowitz, J. (1980). Extreme values from a non-stationary stochastic processes: an application to air quality analysis. process:. *Technometrics* **22**(4), 469-482.

[11] Hüsler, J. (1986). Extreme values of non-stationary random sequences. *J. Appl. Prob.* **23**, 937-950.

[12] Jacobsen, M. (2006). *Point process theory and applications. Marked point and piecewise deterministic processes.* Birkhäuser, Boston.

[13] Khoshnevisan, D. (2002). *Multiparameter Processes: An Introduction to Random Fields*, Springer.

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

[15] Ledford, A.W. and Tawn, J.A. (1998) Concomitant tail behaviour for extremes. *Adv. Appl. Prob.*, **30**(1), 197-215

[16] Malmberg, H. (2011). *Spatial Choice Processes and the Gamma Distribution*, Bachelor thesis 2011:3, Division of Mathematical Statistics, Stockholm University.

[17] Malmberg, H. (2012). *Argmax over Continuous Indeces of Random Variables - An Approach Using Random Fields*, Master thesis 2012:2, Division of Mathematical Statistics, Stockholm University.

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

[19] Marley, A.A.J. and Colonius, H. (1992). The "Horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *J. Math. Psychology* **36**, 1-20.

[20] Mattsson, L-G., Weibull, J.W. and Lindberg, P-O. (2011). Extreme values, invariance and choice probabilities. Manuscript.

[21] Patil, G. (2002). Weighted distributions. In *Encyclopedia of Environmetrics* **4**, 2369-2377.

[22] Resnick, S. and R. Roy (1991): Random USC functions, max-stable processes and continuous choice. *Ann. Appl. Prob.*, **1**, 267-292.

[23] Resnick, S.I. (2008). *Extreme values, regular variation and point processes*, Springer, New York.

[24] Weissman, I. (1975). Extremal processes generated by independent non-identically distributed random variables, *Ann. Prob.* **3**(1), 172-177.