Chapter 18
Probabilistic Choice with an Infinite Set of Options: An Approach Based on Random Sup Measures

Hannes Malmberg and Ola Hössjer

Abstract This chapter deals with probabilistic choice when the number of options is infinite. The choice space is a compact set $S \subseteq \mathbb{R}^k$ and we model choice over S as a limit of choices over triangular sequences $\{x_{n1}, \ldots, x_{nn}\} \subseteq S$ as $n \to \infty$. We employ the theory of random sup measures and show that in the limit when $n \to \infty$, people behave as though they are maximising over a random sup measure. Thus, our results complement Resnick and Roy's [18] theory of probabilistic choice over infinite sets. They define choice as a maximisation over a stochastic process on S with upper semi-continuous (usc) paths. This connects to our model as their random usc function can be defined as a sup-derivative of a random sup measure, and their maximisation problem can be transformed into a maximisation problem over this random sup measure. One difference remains though: with our model the limiting random sup measures are independently scattered, without usc paths. A benefit of our model is that we provide a way of connecting the stochastic process in their model with finite case distributional assumptions, which are easier to interpret. In particular, when choices are valued additively with one deterministic and one random part, we explore the importance of the tail behaviour of the random part, and show that the exponential distribution is an important boundary case between heavy-tailed and light-tailed distributions.

H. Malmberg (✉) · O. Hössjer
Stockholm University, Stockholm, Sweden
e-mail: hannes.malmberg@iies.su.se

O. Hössjer
e-mail: ola@math.su.se

D. Silvestrov and A. Martin-Löf (eds.), *Modern Problems in Insurance Mathematics*, EAA Series, DOI: 10.1007/978-3-319-06653-0_18,
© Springer International Publishing Switzerland 2014

18.1 Introduction

Random utility theory and probabilistic choice have for a long time been the standard tools for modelling discrete choice within the behavioural sciences [12, 15, 21].

This chapter presents a model of probabilistic choice when the set of options is infinite. The set of possible choices is $S \subseteq \mathbb{R}^k$, a compact set, and the aim is to find a choice probability distribution over S as a function of model parameters. Possible applications include choice of residential location and commuting distances.

We first define a choice problem when the set of possible choices is a finite set $N^n = \{x_{n1}, x_{n2}, \ldots, x_{nn}\} \subseteq S$. In this case, we model the different alternatives as yielding utility

$$Y_{ni} \sim \mu(x_{ni}) \quad i = 1, \ldots, n \tag{18.1}$$

independently, where $\mu(\cdot)$ associates each point in S with a probability distribution on $\mathbb{R}$. This setup is standard in the probabilistic choice literature and captures the fact that some factors affecting choice are not observable to the analyst [14]. The optimal choice is x_{I_n} where

$$I_n = \arg \max_{1 \leq i \leq n} Y_{ni}.$$

This value is unique almost surely provided $\mu(x_{ni})$ is a continuous distribution for all $i = 1, \ldots, n$. The argmax distribution on S is given by

$$\tilde{T}_{\mu}^{N^n}(\cdot) = P(X_{I_n} \in \cdot).$$

The continuous choice distribution is defined as the limit in distribution of $\tilde{T}^N(\cdot)$ as $n \to \infty$ and the empirical distribution of $N_n = \{x_{n,1}, x_{n,2}, \ldots, x_{n,n}\}$ converges to some pre-specified distribution Λon S.

To calculate the asymptotic behaviour, we will use the theory of random sup measures. In this, we will draw upon [16]. We show that in the limit when $n \to \infty$, people behave as though they are maximising over a random sup measure M. The earlier chapter closest to our approach is [18]. They analyse probabilistic choice over continuous sets when the random utility function is a random upper semi-continuous (usc) function. However, their random usc function can be viewed as a sup-derivative of an underlying random sup measure. Maximising over the random usc function is equivalent to maximising over their underlying random sup measure in a sense which will be defined in Sect. 18.3. Thus, we show that the limiting behaviour of probabilistic choice in our model has some similarities to the model developed in [18]. There are some important differences in the structure of the random sup measures as well, most importantly that ours are independently scattered (c.f. [20]), without a sup derivative. Instead we model slowly varying trends over S by means of a deterministic component in $\mu(\cdot)$. This allows us to translate specific distributional assumptions on the microlevel to the shape of our random sup field. Thus, we can explore how statistical properties such as tail behaviour of utility disturbances affect choice behaviour.

The paper [18] focuses on random usc functions such that the underlying random sup measure is a max-stable measure. In this, it follows a broader literature within probabilistic choice. For other applications of max-stable processes and random choice, see [2] and [3].

The structure of this chapter is as follows. In Sect. 18.2 we give the formal definition of the argmax measure. In Sect. 18.3 we provide the necessary background on sup measures and random sup measures to clarify the relation between this chapter and earlier research. In Sect. 18.4 we prove the relevant theorems to verify that our approach works. In Sect. 18.5 we solve the model in the special case when μ is the sum of a deterministic term $m(\cdot)$ varying over S and a random disturbance ε_i. In particular, Sect. 18.5.2 deals with the case when ε_i has an exponential distribution, and Sect. 18.5.3 considers other distributional assumptions. Section 18.6 provides a conclusion.

18.2 Defining the Argmax Measure

In this section, we provide the definition of the argmax measure. We will first introduce some relevant concepts needed to state the definition.

Definition 18.1 Let $S \subseteq \mathbb{R}^k$ be a compact set and let

$$\mu : S \to \mathscr{P}$$

where $\mathscr{P}$ is the space of probability measures on $\mathbb{R}$. Then μ is called an absolutely continuous measure index on S if, for each $x \in S$, $\mu(x)$ 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 S is a compact subset of $\mathbb{R}^k$. We write $\mathscr{P}^S$ for the set of absolutely continuous measure indices on S.

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 18.2 An indexed random vector $Y^{N^n} = (Y_{n1}, \ldots, Y_{nn})$ 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 18.3 The point process argmax measure $\tilde{T}_\mu^{N^n}$ is defined as

$$\tilde{T}_\mu^{N^n}(A) = \mathbb{P}\left(\max_{1 \le i \le n : x_{ni} \in N^n \cap A} Y_{ni} \ge \max_{1 \le i \le n} Y_{ni}\right) = \mathbb{P}(X_n \in A), \tag{18.2}$$

for all Borel measurable sets $A \subseteq S$, where

$$X_n = \arg \max_{x_{ni} \in N^n} Y_{ni} \tag{18.3}$$

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 $\mathscr{Q}^S$ to denote the set of finite multisets on S. With this notation, $\tilde{T}_\mu^{N^n}$ is a function from $\mathscr{Q}^S$ to $[0, 1]^{\mathscr{B}(S)}$, the family of set functions on the Borel sigma algebra $\mathscr{B}(S)$ on S taking values in $[0, 1]$. We deliberately avoid identifying this set with the set of probability distributions on S, as we will not always know a priori that the relevant set function will be countably additive.

Even though N^n is a deterministic set of points, it can typically be thought of as the realisation 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 $A \subseteq S$.

Definition 18.4 For a probability distribution Λ, we define the point sequence domain of convergence as

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

i.e. the class of point sequences whose empirical distributions converge weakly to Λ on $\mathscr{B}(S)$.

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

Definition 18.5 (*Limiting argmax measure*) A probability measure T_μ^Λ such that

$$\tilde{T}_\mu^{N^n} \Rightarrow T_\mu^\Lambda \tag{18.4}$$

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

18.3 Sup Measures, Random Sup Measures, and Upper-Semi Continuous Functions

In this section, we provide the necessary background on random fields, sup measures, random sup measures, and upper semi-continuous functions. See [9] for an introduction to random fields. A more careful treatment of random sup measures and upper semi-continuous functions can be found in [16], Sects. 2–4, and older works on the topic include [22] and [23].

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

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

where Ω is a generic sample space and $\mathscr{B}(S)$ denotes the Borel σ-algebra on S. Thus, for a fixed ω, $M(\omega, \cdot)$ is a set function on $\mathscr{B}(S)$ and for fixed $A \in \mathscr{B}(S)$, $M(\cdot, A)$ is a random variable taking values in $\mathbb{R}$. We sometimes write $M(A)$ as short-hand for $M(\cdot, A)$ and we write $M(\omega, A)$ for a particular realisation of the random variable $M(\cdot, A)$. We will write $\mathscr{P}^{\mathscr{B}(S)}$ to denote the set of all random fields over $\mathscr{B}(S)$.

Suppose that we have a random field such that for each fixed ω, $M(\omega, \cdot) = m_\omega(\cdot)$ satisfies

$$m_\omega\left(\bigcup_\alpha A_\alpha\right) = \bigvee_\alpha m_\omega(A_\alpha) \tag{18.5}$$

for any arbitrary collection $\{A_\alpha\} \subseteq \mathscr{B}(S)$. Then we call the random field M a random sup measure.

For each random sup measure M, we may define the *sup-derivative* d' by

$$Y(x) = d'M(x) = \inf\{M(G); G \in \mathscr{B}(S), x \in G\} \tag{18.6}$$

which is a stochastic process on S with upper semi-continuous paths (recall that a function f is upper semi-continuous if $\{x : f(x) < y\}$ is open for all y). There is a close connection between random sup measures and stochastic processes with usc paths. Indeed, for $A \in \mathscr{B}(S)$

$$\bigvee_{x \in A} Y(x) = \bigvee_{x \in A} \bigwedge_{G \ni x} M(G) = \bigwedge_{G \supset A} M(G) = M(A).$$

In [18], the sup-derivative is used to derive a random usc function from an underlying random sup measure.

18.3.1 Calculating Argmax on Random Sup Measures

In this section, we consider how to select the element having the largest value for a random sup measure, and the relation between the distribution of maximisers of a random sup measure and those of its sup derivative, provided that the latter exists.

The probability distribution of the largest element X of a random sup measure is given by

$$P(X \in A) = P(M(A) > M(S \setminus A)). \tag{18.7}$$

We can see that this agrees with a definition based on the distribution of the location of the maximal element of the sup derivative of M. Indeed, assuming that the maximiser of a random usc function is unique, and given by

$$X' = \arg\max_x Y(x) \tag{18.8}$$

we have

$$\begin{aligned} P(X' \in A) &= P\left(\bigvee_{x\in A} Y(x) > \bigvee_{x\in A^c} Y(x)\right) \\ &= P\left(M(A) > M(S \setminus A)\right) \\ &= P(X \in A). \end{aligned}$$

This means that maximising over a random usc function as in [18] can be seen as maximising over a random sup measure in line with Eq. (18.7), which is the method we will use in this chapter. This in turn shows the strong connections between the mathematics of optimal choice in a framework based on limiting behaviour and one based on maximisation over a random usc function.

18.4 Calculating the Argmax Measure

In this section, we will develop a method for calculating the argmax measure relying on continuity properties of random fields.

For each multiset N^n, we construct a random field

$$\tilde{M}_\mu^{N^n} = \left\{ \max_{1\le i\le n : x_{ni}\in N^n\cap A} Y_{ni},\, A \in \mathscr{B}(S) \right\} \tag{18.9}$$

where $\sup \emptyset = -\infty$. This random field is also a random sup measure as defined in Definition 18.3, since it satisfies (18.5).

We connect this random sup measure to the argmax measures by the *pseudo-argmax measure*.

Definition 18.6 The pseudo-argmax measure $F : \mathscr{P}^{\mathscr{B}(S)} \to [0, 1]^{\mathscr{B}(S)}$ is defined by

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

for all $A \in \mathscr{B}(S)$.

We note that

$$F\left(\cdot; \tilde{M}_\mu^{N^n}\right) = \tilde{T}_\mu^{N^n} \tag{18.10}$$

and that $F(\cdot; \tilde{M}^{N^n})$ is a probability measure. We seek to show that $F(\cdot; M)$ is continuous in M in an appropriate sense, and then use (18.10) to derive the limiting behaviour of $\tilde{T}_{\mu}^{N^n}$ from the asymptotic behaviour of $\tilde{M}^{N^n}$.

In particular, we will:

1. define a notion of convergence $\stackrel{m}{\rightarrow}$ on the set of random fields, *and*
2. define a class of random fields, absolutely continuous independently scattered random sup measures (acisrsm), *such that,*
3. if $M_n \stackrel{m}{\rightarrow} M$, and M is an acisrsm, then $F(\cdot; M_n) \Rightarrow F(\cdot; M)$ and $F(\cdot; M)$ is a probability measure.

We need to explicitly show that $F(\cdot; M)$ is a probability measure, as the set of probability measures is not closed under weak convergence.

Definition 18.7 A sequence of random fields M_n in $\mathscr{P}^{\mathscr{B}(S)}$ is said to *m-converge* to the random field M ($\stackrel{m}{\rightarrow}$) if there exists a sequence $g_n : \mathbb{R} \to \mathbb{R}$ of strictly increasing functions such that

$$g_n(M_n(A)) \Rightarrow M(A) \tag{18.11}$$

for all A with

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

A general notion of convergence for random sup measures is presented in [16]. However, for us this notion is unnecessarily strong as it requires convergence of all finite dimensional distributions of the random field. Instead we make a slightly weaker requirement of two dimensional convergence for a set A and its complement.

Definition 18.8 Let $M : \Omega \times \mathscr{B}(S) \to \mathbb{R} \cup \{-\infty\}$ be a random field over $\mathscr{B}(S)$. We call M an absolutely continuous independently scattered random sup measure (henceforth acisrsm) if the following 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 $P\left(\limsup_n M(A_n) \geq M(S)\right) = 0$;
5. $M(\emptyset) = -\infty$;
6. If $M(A) = -\infty$ almost surely, then $M(S \setminus A) > -\infty$ almost surely;
7. If $M(A) > -\infty$ almost surely, then $M(\cdot, A)$ is an absolutely continuous probability measure on $\mathbb{R}$ with respect to Lebesgue measure.

An acisrsm is similar to the "independently scattered random sup measures" introduced in [20], although that article restricts its attention to the case with Frechét-distributed marginals. Typically, Property 5 implies that the right-hand side of (18.6) is $-\infty$ for all $x \in S$. We can either interpret this as a sup derivative $Y(x)$ that is not well-defined, or that it equals $-\infty$. In either case, we cannot use the sup derivative in order to derive an argmax distribution, as in Sect. 18.3.1.

Theorem 18.1 *Let $\{M_n\}$ be a sequence of random fields and let M be an acisrsm such that*

$$M_n \stackrel{m}{\rightarrow} M.$$

Then

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

and $F(\cdot, M)$ is a probability measure.

Proof Let $A \subseteq S$ be measurable with $F(\partial A; M) = 0$. We seek to show that $F(A; M_n) \rightarrow F(A; M)$, 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

$$\begin{aligned} g_n(M_n(A)) &\Rightarrow M(A) \\ g_n(M_n(A^c)) &\Rightarrow M(A^c) \end{aligned}$$

hold simultaneously. As $g_n(M_n(A))$ and $g_n(M_n(A^c))$ are independent for all n, this means that

$$g_n(M_n(A)) - g_n(M_n(A^c)) \Rightarrow M(A) - M(A^c).$$

By Definition 18.8, $M(A)$ and $M(A^c)$ are absolutely continuous with respect to Lebesgue measure and independent, and therefore their difference is absolutely continuous. Hence,

$$\begin{aligned} F(A; M_n) &= \mathbb{P}(M_n(A) > M_n(A^c)) \\ &= \mathbb{P}(g_n(M_n(A)) > g_n(M_n(A^c))) \\ &= \mathbb{P}(g_n(M_n(A)) - g_n(M_n(A^c)) > 0) \\ &\rightarrow \mathbb{P}(M(A) - M(A^c) > 0) \\ &= F(A; M), \end{aligned}$$

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

$$F(A; M_n) \rightarrow F(A; M).$$

Case 2 $M(A) = -\infty$ a.s. From Definition 18.8, $M(A^c) > -\infty$ almost surely, which means that $F(A; M) = 0$. Furthermore,

$$\begin{aligned} g_n\,(M_n(A)) &\Rightarrow -\infty. \\ g_n\,(M_n(A^c)) &\Rightarrow M(A^c) > -\infty. \end{aligned}$$

We can find K such that $\mathbb{P}(M(A^c) > K) = 1 - \varepsilon$, and n_0 such that for all $n \geq n_0$, $\mathbb{P}\,(g_n\,(M_n(A)) < K) > 1 - \varepsilon$ and $\mathbb{P}\,(g_n\,(M_n(A^c)) > K) > 1 - 2\varepsilon$. Then, for all $n \geq n_0$, $\mathbb{P}(M_n(A) > M_n(A^c)) < 3\varepsilon$. 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

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

and we get that

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

in this case as well.

It remains to show that $F(\cdot; M)$ is a probability measure. Countable additivity is the only non-trivial property.

We first establish finite additivity. We introduce a new notation for the residual set $A_{n+1} = S \setminus \bigcup_{i=1}^n A_i$, and the events

$$B_i = \{M(A_i) > M\,(S \setminus A_i)\} \text{ for } i = 1, 2, \ldots, n.$$

It is evident that if there is an i such that $M(A_i) = -\infty$, then $F(A_i; M) = 0$, so let us assume that this not true for any i. By absolute continuity, the B_i's are almost surely disjoint. Hence,

$$\begin{aligned} F\left(\bigcup_{i=1}^n A_i; M\right) &= P(\max_{1\le i\le n} M(A_i) > M(A_{n+1})) \\ &= \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) \\ &= \sum_{i=1}^n \mathbb{P}(B_i) \\ &= \sum_{i=1}^n F(A_i; M). \end{aligned}$$

For countable additivity, it suffices to show that if $A_1 \supseteq A_2 \supseteq A_3 \ldots$ such that $\cap_n A_n = \emptyset$, then $F(A_n; M) \to 0$. However, by Definition 18.8, Property 4,

$$F(A_n; M) = P(M(A_n) > M(S \setminus A_n)) \le P(M(A_n) \ge M(S)) \to 0,$$

and the proof is complete. □

The following corollary establishes the connection between the theorem and the calculation of the argmax measure.

Corollary 18.1 *Suppose there exists an acisrsm M_μ^Λ such that for all $N^n \in \mathscr{N}^\Lambda$*

$$\tilde{M}_\mu^{N^n} \stackrel{m}{\to} M_\mu^\Lambda.$$

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

$$T_\mu^\Lambda = F\left(\cdot; M_\mu^\Lambda\right). \tag{18.12}$$

Proof We note that

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

and apply Theorem 18.1 to conclude that

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

for all $\{N^n\} \in \mathcal{N}^\Lambda$, and that $F(\cdot; M_\mu^\Lambda)$ is a probability measure. By Definition 18.5, T_μ^Λ is the argmax measure. □

18.5 The Argmax Measure for Homoscedastic Regression Models

The result in Corollary 18.1 shows that the methodology developed in the previous section gives a way to calculate the argmax measure which is workable insofar as it is possible to find an acisrsm 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

$$Y_{ni} = m(x_{ni}) + \varepsilon_{ni}, \tag{18.13}$$

for $i = 1, \ldots, n$, where $m : S \to \mathbb{R}$ is a given deterministic regression function and $\{\varepsilon_{ni}\}$ 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

$$\mu(x) = H(\cdot - m(x)). \tag{18.14}$$

In order to find the limiting behaviour of the empirical acisrsm $\tilde{M}_\mu^{N^n}$ defined in (18.9), we note that for all open sets A with $\Lambda(A) > 0$, $|A \cap N^n| \to \infty$ as $n \to \infty$, which means that the 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 three subsections. First we state a classical result in extreme value theory for $m \equiv 0$, and its specific counterpart related to offers $H \sim \text{Exp}(s)$ 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 an acisrsm M_μ^Λ to which $\tilde{M}_\mu^{N^n}$ m-converges for an appropriate sequence g_n of monotone transformations. Then Corollary 18.1 is applied in order to calculate the argmax measure T_μ^Λ. The third subsection considers other distributions H than the exponential.

18.5.1 Some Extreme Value Theory

The following theorem is a key result in extreme value theory, see for instance [4–6, 10, 17].

Theorem 18.2 (Fisher–Tippet–Gnedenko Theorem) *Let $\{Y_n\}$ be a sequence of independent and identically distributed (i.i.d.) random variables and define the random variable $M^n = \max\{Y_1, Y_2, \ldots, Y_n\}$. 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 Fréchet, or the Weibull families of distributions.

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

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

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 $b_n = s\log(n)$ when the random variables Y_i have an exponential distribution with mean s, see for instance [17] for a proof.

Proposition 18.1 *Let $\{Y_i\}_{i=1}^n$ be a sequence of i.i.d. random variables with $Y_i \sim$ $\mathrm{Exp}(s)$. Then*

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

18.5.2 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 $H \sim \mathrm{Exp}(s)$ separately in this subsection.

18.5.2.1 Limiting Acisrsm 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

[7, 8, 24]). Thus, we prove a result characterising the acisrsm with $H \sim \mathrm{Exp}(s)$ and $m(x)$ varying.

Theorem 18.3 *Let $\tilde{M}_\mu^{N^n}(A)$ be as defined in (18.9), with $Y_{ni} - m(x_{ni}) \sim \mathrm{Exp}(s)$ independently for $i = 1, \ldots, n$ and $s > 0$. Suppose Λ is a probability measure on the Borel σ-algebra on S and that the following properties hold:*

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

Then (18.11) holds with $g_n(y) = y/s - \log(n)$, i.e.

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

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

$$M_\mu^\Lambda(A) \sim \log\left(\int_A \mathrm{e}^{m(x)/s}\Lambda(\mathrm{d}x)\right) + \mathrm{Gumbel}(A). \tag{18.15}$$

The Gumbel(A) means that the marginal distribution of $M_\mu^\Lambda(A)$ is a constant plus a standard Gumbel(0,1) random variable. We give the marginal distributions of the limiting random field, and do not specify the full set of the finite dimensional distributions, as the only property on the joint distributions we will need is that the random variables $M_\mu^\Lambda(A)$ and $M_\mu^\Lambda(A^c)$ are independent. This is true as $g_n\left(\tilde{M}_\mu^{N^n}(A)\right)$ and $g_n\left(\tilde{M}_\mu^{N^n}(A^c)\right)$ are independent for all pre-limiting random variables . For more discussion on Gumbel random fields, see [19].

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

Let $A \subset S$ 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 [1]), we get

$$\begin{aligned}
&\frac{1}{n}\sum_{1\leq i\leq n: x_{in}\in A} \mathrm{e}^{m(x_{in})} \\
&\quad = \int_{A\cap\bar{D}_m^c} \mathrm{e}^{m(x)}\mathrm{d}P^{N^n}(x) + \frac{1}{n}\sum_{1\leq i\leq n: x_{in}\in A\cap\bar{D}_m} \mathrm{e}^{m(x_{in})} \\
&\quad \to \int_{A\cap\bar{D}_m^c} \mathrm{e}^{m(x)}\mathrm{d}\Lambda(x) + 0 \\
&\quad = \int_A \mathrm{e}^{m(x)}\mathrm{d}\Lambda(x).
\end{aligned} \tag{18.16}$$

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

$$\bar{m} = \sup_{x \in S} m(x) \tag{18.17}$$

and get

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

where in the third last step we utilised that

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

since $\bar{D}_m$ is a closed set. We can use (18.16) to derive the acisrsm directly. With $g_n(y) = y - \log(n)$ we get that if $Z_n = \log\left(\mathbb{P}\left(g_n\left(\tilde{M}_\mu^{N^n}(A)\right) \le y\right)\right)$ it holds that

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

where we recognise the last line as the logarithm of a Gumbel distribution function with an additive term $\log\left(\int_A \exp(m(x)) \Lambda(\mathrm{d}x)\right)$ 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

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

Indeed, using the well-known result that

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

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. □

Proposition 18.2 *The random field defined by*

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

is an acisrsm in the sense of Definition 18.8 when m and Λ *satisfy the conditions of Theorem 18.3.*

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 4 and 5 hold 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. □

18.5.2.2 Argmax Distribution

In Corollary 18.1, 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 18.3 together with Proposition 18.2 and Corollary 18.1 to derive the argmax measure associated with μ and Λ.

Theorem 18.4 *Let* $\mu(x) = m(x) + \mathrm{Exp}(s)$ *and let* Λ *be a probability measure on S. Suppose that* Λ *and m jointly satisfy the conditions in Theorem 18.3. Then the argmax measure* T_μ^Λ *exists and is given by the exponentially tilted distribution*

$$T_\mu^\Lambda(A) = C \int_A e^{m(x)/s} \Lambda(dx), \tag{18.18}$$

where

$$C = \left(\int_S e^{m(x)/s} \Lambda(dx)\right)^{-1} \tag{18.19}$$

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

$$t_{\mu}^{\Lambda}(x) = C\lambda(x)\exp(m(x)/s) \tag{18.20}$$

for $x \in S$, *i.e.* $T_{\mu}^{\Lambda}(A) = \int_A t_{\mu}^{\Lambda}(x)\nu(\mathrm{d}x)$ *for all Borel sets* $A \subset S$.

Proof After standardising data $Y_{ni} \leftarrow Y_{ni}/s$, we may, without loss of generality, assume that $s = 1$. Proposition 18.2 states that M_{μ}^{Λ}, defined as in Theorem 18.3, is an acisrsm, and in order to find its pseudo argmax measure we let $G(x) = G(x; 0, 1) = \mathrm{e}^{-\mathrm{e}^{-x}}$ denote the distribution function of a standard Gumbel distribution and put $L(A) = \log\left(\int_A \mathrm{e}^{m(x)}\mathrm{d}\Lambda(x)\right)$. Then

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

for all Borel sets A.

Then note that Theorem 18.3 implies that

$$\tilde{M}_{\mu}^{N^n}(A) - \log(n) \Rightarrow M_{\mu}^{\Lambda}(A) \tag{18.21}$$

holds for $\{N^n\}_{n\geq 1} \in \mathscr{N}^{\Lambda}$ 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, $\tilde{M}_{\mu}^{N^n} \overset{m}{\to} M_{\mu}^{\Lambda}$. Finally, Corollary 18.1 implies that the argmax measure $T_{\mu}^{\Lambda} = F(\cdot; M_{\mu}^{\Lambda})$ exists and is given by (18.18). □

Theorem 18.4 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 (18.17) of $\bar{m}$, put $I_n = \arg\max_{1\leq i\leq n} Y_{ni}$ and assume for simplicity $s = 1$. Then, for any $i = 1, \ldots, n$,

$$\begin{aligned}
\mathbb{P}(I_n = i) &\approx \mathbb{P}(Y_{ni} \geq \bar{m})\mathbb{P}(I_n = i | Y_{ni} \geq \bar{m}) \\
&\approx \mathrm{e}^{-(\bar{m}-m(x_{ni}))} / \left(n\mathbb{P}(m(X) + \varepsilon \geq \bar{m})\right) \\
&= \mathrm{e}^{m(x_{ni})} \Bigg/ \left(n\int_S \mathrm{e}^{m(x)}\Lambda(\mathrm{d}x)\right).
\end{aligned}$$

In the first step we utilised that $\max_{1\leq i\leq n} Y_{ni} \geq \bar{m}$ holds with probability close to 1 when n is large, and in the second step approximated the number of i for which

$Y_{ni} = m(x_{ni}) + \varepsilon_{ni} \geq \bar{m}$ as

$$\sum_{i=1}^{n} 1_{\{m(x_{ni})+\varepsilon_{ni} \geq \bar{m}\}} \approx n\mathbb{P}(m(X) + \varepsilon \geq \bar{m}),$$

where $\{x_{ni}, \varepsilon_{ni}\}_{i=1}^{n}$ 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 indices i with $Y_{ni} \geq \bar{m}$ have the same conditional probability of being the argmax, i.e. $I_n = i$.

18.5.3 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 solutions 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 distributions, it is only the extremal behaviour of m that determines the asymptotic argmax distribution, whereas m has no asymptotic impact for heavy-tailed distributions.

18.5.3.1 Light-Tailed Error Distributions

Formally, the light-tailed case corresponds to the class of distributions 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 degenerate in the sense that $M_{\mu}^{\Lambda}(A)$ has a one point distribution, so that the absolute continuity Property 7 of Definition 18.8 is violated. Therefore we cannot use Theorem 18.1 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

$$\Lambda_{\varepsilon}(A) = C \int_{A} \frac{h\,(K - \varepsilon + \bar{m} - m(x))}{H([K - \varepsilon + \bar{m} - m(x), K])} \Lambda(\mathrm{d}x),$$

with $h(x) = 0$ if $x > K$, $\bar{m}$ as in (18.17), the convention $H([K', K]) = 0$ when $K' > K$, and $C = C(\varepsilon)$ a normalising constant chosen so that $\Lambda_\varepsilon(S) = 1$. Assume further that a limit measure $\Lambda_{\max}$ exists, supported on the set

$$S_{\max} = \{x \in S; m(x) = \bar{m}\}$$

where m is maximal, such that

$$\Lambda_\varepsilon \Rightarrow \Lambda_{\max} \text{ as } \varepsilon \to 0. \tag{18.22}$$

It is reasonable to assume that Λ_ε should approximate the conditional distribution of X_n in (18.3) 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 (18.22) suggests that

$$T_\mu^\Lambda = \Lambda_{\max}, \tag{18.23}$$

since $Y_{n:n}$ tends in probability to $\bar{m} + K$ as n grows. Notice that $\Lambda_{\max}$ has a one point distribution when $S_{\max} = x_{\max}$ consists of one single element. This accords with (18.8), since the sup derivative $Y(x) = K + m(x)$ of M_μ^Λ exists.

In order to establish (18.23) according to Definition 18.5, we need a slightly stronger condition though than (18.22), as the following theorem reveals:

Theorem 18.5 *For any $\varepsilon > 0$, put*

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

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

$$Q_n(\varepsilon) = \int_S H\,([K - \varepsilon + \bar{m} - m(x), K])\, P^{N^n}(\mathrm{d}x).$$

Assume that

$$P^{N^n}_{Q_n^{-1}(c/n)} \Rightarrow \Lambda_{\max} \text{ as } n \to \infty \tag{18.24}$$

uniformly for all $c \in (0, \bar{c}]$, for any $\bar{c} > 0$, with Q_n^{-1} the inverse function of Q_n. Then (18.23) holds.

Proof According to Definition 18.5, we need to prove $\tilde{T}_\mu^{N^n} \Rightarrow \Lambda_{\max}$ for any $\{N^n\}_{n \ge 1} \in \mathscr{N}^\Lambda$. Let $Z_n = \bar{m} + K - Y_{n:n}$. We first note that

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

where X_n is defined as in (18.3). By conditioning on Z_n we notice that

$$\tilde{T}_{\mu}^{N^n}(A) = \int_0^{\infty} P_{\varepsilon}^{N^n}(A) F_{Z_n}(\mathrm{d}\varepsilon). \tag{18.25}$$

Furthermore, Z_n has the property that

$$nQ_n(Z_n) \Rightarrow \mathrm{Exp}(1).$$

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

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

where the last step uses the well known fact

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

if $a_{n,i} \ge 0$,

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

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

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

and $\lim\limits_{n\to\infty} \max H[K + \bar{m} - m(x_{ni}) - Q_n^{-1}(x/n), K] = 0$ assuming that H has no point mass on K.

Thus, $nQ_n(Z_n) \Rightarrow \mathrm{Exp}(1)$, and we conclude the proof by performing a change of variable $c = nQ_n(\epsilon)$ on (18.25) to get

$$\tilde{T}_{\mu}^{N^n}(A) = \int\limits_0^{\infty} P_{Q_n^{-1}(c/n)}^{N^n}(A) F_{nQ_n(Z_n)}(\mathrm{d}c). \tag{18.26}$$

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 $\bar{c}$ such that $\Lambda_{\max}(\partial A) = 0$, 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. □

18.5.3.2 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 Fréchet distribution with shape parameter α, scale parameter 1 and location parameter 0. Since a_n increases with n at a 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 18.6 *Let $\tilde{M}_{\mu}^{N^n}(A)$ be as defined in (18.9), with $Y_{ni} - m(x_{ni}) \sim \text{Pareto}(\alpha, 1)$ independently for $i = 1, \ldots, n$. Suppose Λ is a probability measure on the Borel σ-algebra on S and that properties 1–3 of Theorem 18.3 hold.*

Then (18.11) holds with $g_n(y) = y/n^{1/\alpha}$, i.e.

$$\tilde{M}_{\mu}^{N^n}(A)/n^{1/\alpha} \Rightarrow M_{\mu}^{\Lambda}(A) = \Lambda(A)^{1/\alpha}\text{Frechet}_{\alpha}(A) \tag{18.27}$$

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

$$T_{\mu}^{\Lambda} = \Lambda. \tag{18.28}$$

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

Proof We begin with (18.27). 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

$$\begin{aligned}\log F_{n,A}(y) &= \log\left(\prod_{x_{n,i}\in A} P(Y_{n,i} + m(x_{n,i}) \le n^{1/\alpha} y)\right)\\ &= \sum_{1\le i\le n, x_{n,i}\in A} \log\left(1 - (n^{1/\alpha} y - m(x_{n,i}))^{-\alpha}\right)\\ &= \sum_{1\le i\le n} \log\left(1 - \frac{(y-n^{-1/\alpha} m(x_{n,i}))^{-\alpha}}{n}\right)^n \frac{I(x_{n,i}\in A)}{n}\\ &= \sum_{1\le i\le n} f(n,i)h(n,i).\end{aligned} \tag{18.29}$$

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

$$\begin{aligned}\lim_{n\to\infty} \log F_{n,A}(y) &= \lim_{n\to\infty} \sum_{1\le i\le n} f(n,i)h(n,i)\\ &= -y^{-\alpha} \lim_{n\to\infty} \sum_{1\le i\le n} \frac{I(x_{n,i}\in A)}{n}\\ &= -y^{-\alpha}\Lambda(A),\end{aligned} \tag{18.30}$$

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

It remains to prove (18.28). To this end, we notice that the pseudo argmax measure of M_μ^Λ equals

$$\begin{aligned}F(A; M_\mu^\Lambda) &= P\left(\Lambda(A)^{1/\alpha}\text{Frechet}_\alpha(A) > \Lambda(A^c)^{1/\alpha}\text{Frechet}_\alpha(A^c)\right)\\ &= P\left(\Lambda(A)\text{Frechet}_1(A) > \Lambda(A^c)\text{Frechet}_1(A^c)\right)\\ &= \Lambda(A),\end{aligned} \tag{18.31}$$

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

$$\begin{aligned}P\left(\Lambda(A)X > \Lambda(A^c)Y\right) &= \int_0^\infty \frac{\Lambda(A)}{y^2}\text{Exp}(-\Lambda(A)y^{-1})\text{Exp}\left(-\Lambda(A^c)y^{-1}\right)\mathrm{d}y\\ &= \Lambda(A)\int_0^\infty 1/y^2 \exp(-y^{-1})\mathrm{d}y\\ &= \Lambda(A),\end{aligned} \tag{18.32}$$

Since $F(\cdot; M_\mu^\Lambda) = \Lambda$, it follows from (18.27) that $\tilde{M}_\mu^{N^n} \xrightarrow{m} M_\mu^\Lambda$. Hence, by Corollary 1, $T_\mu^\Lambda = F(\cdot; M_\mu^\Lambda) = \Lambda$ exists. $\square$

18.6 Conclusion

In this chapter we introduced a definition for an argmax measure on an infinite compact set $S \subseteq \mathbb{R}^k$.

We showed that there is a close relation between an argmax measure defined as a limit of choices with a finite number of options, and an argmax definition based on selecting maximisers of a random sup measure [18].

Our limit-based definition also allows us to explore the consequence of different distributional assumptions in a homoscedastic regression model, with one deterministic component and another random disturbance component. This analysis showed that a model with an exponentially distributed disturbance term is an important intermediate case. The class of heavy-tailed disturbances that we studied correspond to Fréchet random sup measures where the deterministic component is unimportant. Light-tailed distributions (with compact support) correspond to non-random sup measures where only the deterministic component $m(\cdot)$ matters for the argmax measure.

It is possible to extend the approaches of this chapter and link it more closely to theory of point processes and extreme values [17] as well as the theory of concomitants of extreme order statistics [11]. Indeed, one can construct a theory where the locations x_{ni} of offers are not deterministic, 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 when 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 of the extreme order statistic among $Y_1, \ldots, Y_n$. Some work in this direction is presented in [13].

Acknowledgments 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. Billingsley, P.: Convergence of Probability Measures, 2nd edn. Wiley, Hoboken (1999)
2. Cosslett, S.R.: Extreme-value stochastic processes: a model of random utility maximization for a continuous choice set. Preprint, Department of Economics, Ohio State University (1988)
3. Dagsvik, J.K.: The generalized extreme value random utility model for continuous choice. Technical Report, Tilburg University, Center for Economic Research (1989)
4. Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 24, pp. 180–190. Cambridge University Press (1928)
5. Gnedenko, B.: Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. **44**(3), 423–453 (1943)
6. Gumbel, E.J.: Statistics of Extremes. DoverPublications.com, New York (1958)
7. Horwitz, J.: Extreme values from a non stationary stochastic process: an application to air quality analysis (with discussion). Technometrics **22**, 469–482 (1980)

8. Hüsler, J.: Extreme values of non-stationary random sequences. J. Appl. Probab. **23**, 937–950 (1986)
9. Khoshnevisan, D.: Multiparameter Processes: An Introduction to Random Fields. Springer, New York (2002)
10. Leadbetter, M.R., Lindgren, G., Rootzén, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, Berlin (1982)
11. Ledford, A.W., Tawn, J.A.: Concomitant tail behaviour for extremes. Adv. Appl. Probab. **30**(1), 197–215 (1998)
12. Luce, R.D.: Individual Choice Behavior: A Theoretical Analysis. Wiley, New York (1959)
13. Malmberg, H., Hössjer, O.: Extremal behaviour, weak convergence and argmax theory for a class of non-stationary marked point processes (2013). (Submitted)
14. Manski, C.F., McFadden, D., et al.: Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge (1981)
15. McFadden, D.: Econometric models for probabilistic choice among products. J. Bus. **53**(3), S13–S29 (1980)
16. O'Brien, G.L., Torfs, P.J., Vervaat, W.: Stationary self-similar extremal processes. Probab. Theory Relat. Fields **87**(1), 97–119 (1990)
17. Resnick, S.I.: Extreme Values, Regular Variation, and Point Processes. Springer, New York (2007)
18. Resnick, S.I., Roy, R.: Random usc functions, max-stable processes and continuous choice. Ann. Appl. Probab. **1**, 267–292 (1991)
19. Robert, C.Y.: Some new classes of stationary max-stable random fields. Stat. Probab. Lett. **83**, 1496–1503 (2013)
20. Stoev, S.A., Taqqu, M.S.: Extremal stochastic integrals: a parallel between max-stable processes and α-stable processes. Extremes **8**(4), 237–266 (2005)
21. Train, K.: Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge (2009)
22. Vervaat, W.: Stationary self-similar extremal processes and random semicontinuous functions. Depend. Probab. Stat. **11**, 457–473 (1986). (Oberwolfach, 1985)
23. Vervaat, W.: Random Upper Semicontinuous Functions and Extremal Processes, pp. 1–43. Department of Mathematical Statistics, Centrum for Wiskunde and Information, Amsterdam (R8801), (1988)
24. Weissman, I.: Extremal processes generated by independent nonidentically distributed random variables. Ann. Probab. **3**, 172–177 (1975)