Electronic Journal of Statistics
Vol. 5 (2011) 421–439
ISSN: 1935-7524
DOI: 10.1214/11-EJS613

A continuous mapping theorem for the smallest argmax functional

Emilio Seijo and Bodhisattva Sen

Department of Statistics
Columbia University
1255 Amsterdam Avenue
New York, NY. 10027.
e-mail: emilio@stat.columbia.edu; bodhi@stat.columbia.edu

Abstract: This paper introduces a version of the argmax continuous mapping theorem that applies to M-estimation problems in which the objective functions converge to a limiting process with multiple maximizers. The concept of the smallest maximizer of a function in the d-dimensional Skorohod space is introduced and its main properties are studied. The resulting continuous mapping theorem is applied to three problems arising in change-point regression analysis. Some of the results proved in connection to the d-dimensional Skorohod space are also of independent interest.

Keywords and phrases: Change-point, compound Poisson process, Cox proportional hazards model, multiple maximizers, Skorohod spaces with multidimensional parameter.

Received February 2011.

Contents

1. Introduction

Many estimators in statistics are defined as the maximizers of certain stochastic processes, called objective functions. This procedure for computing estimators is known as M-estimation and is quite common in modern statistics. A standard way to find the asymptotic distribution of a given M-estimator, is to obtain the limiting law of the (appropriately normalized) objective function and then apply the so-called argmax continuous mapping theorem (see Theorem 3.2.2, page 286 of Van der Vaart and Wellner (1996) for a quite general version of this result). Chapter 3.2 in Van der Vaart and Wellner (1996) gives an excellent account of M-estimation problems and applications of the argmax continuous mapping theorem.

Despite its proven usefulness in a wide range of applications, there are some M-estimation problems that cannot be solved by an application of the usual argmax continuous mapping theorem. This is particularly true when the objective functions converge in distribution to the law of some process that admits multiple maximizers. This situation arises frequently in problems concerning change-point estimation in regression settings. In these problems, the estimators are usually maximizers of processes that converge in the limit to two-sided, compound Poisson processes that have a complete interval of maximizers. See, for instance, Kosorok (2008) (Section 14.5.1, pages 271–277), Lan et al. (2009), Kosorok and Song (2007), Pons (2003) and Seijo and Sen (2010). This issue has been noted before by several authors, such as Ferger (2004).

The main goal of this paper is to derive a version of the argmax continuous mapping theorem specially taylored for situations like the one described in the previous paragraph. A distinctive feature of the argmax continuous mapping theorem in this setup is that it requires the weak convergence, not only of the objective functions, but also of some associated *pure jump processes*. Although this requirement has been overlooked by some authors in the past (we discuss these omissions in Section 5), its necessity can be easily seen; see Section 4 for an example.

To illustrate the situations on which our results are applicable, we start with the following simple problem that arises in least squares change-point regression. Detailed accounts of this type of models can be found in Kosorok (2008) (Section 14.5.1, pages 271–277), Lan et al. (2009) and Seijo and Sen (2010). In its simplest form the model considers a random vector $X = (Y, Z)$ satisfying the following relation:

$$Y = \alpha_0 \mathbf{1}_{Z \le \zeta_0} + \beta_0 \mathbf{1}_{Z > \zeta_0} + \epsilon, \tag{1}$$

where Z is a continuous random variable, $\alpha_0 \neq \beta_0 \in \mathbb{R}$, $\zeta_0 \in [c_1, c_2] \subset \mathbb{R}$ and ϵ is a continuous random variable, independent of Z with zero expectation and finite variance $\sigma^2 > 0$. The parameter of interest is ζ_0, the change-point. Given a random sample from this model, the *least squares estimator* $\hat{\theta}_n$ of $\theta_0 =$

$(\zeta_0, \alpha_0, \beta_0) \in \Theta := [c_1, c_2] \times \mathbb{R}^2$ is obtained by maximizing the criterion function

$$M_n(\theta) := -\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \alpha \mathbf{1}_{Z_i \le \zeta} + \beta \mathbf{1}_{Z_i > \zeta}\right)^2,$$

i.e.,

$$\hat{\theta}_n := (\hat{\zeta}_n, \hat{\alpha}_n, \hat{\beta}_n) = \operatorname*{sargmax}_{\theta\in\Theta}\{M_n(\theta)\}, \tag{2}$$

where sargmax denotes the maximizer with the smallest ζ value. This distinction is made as there is no unique maximizer for ζ, in fact, for any α, β, $M_n(\cdot, \alpha, \beta)$ is constant on every interval $[Z_{(j)}, Z_{(j+1)})$, where $Z_{(j)}$ stands for the j-th order statistic. It can be shown, see either Kosorok (2008) (Section 14.5.1, pages 271–277) or Seijo and Sen (2010), that $n(\hat{\zeta}_n - \zeta_0)$ converges in distribution to the smallest maximizer a two-sided, compound Poisson process. The convergence results in this paper, Theorems 3.1 and 3.2, can, in particular, be applied to derive the asymptotic distribution of this estimator (see Section 5.1).

Our results will be applicable to M-estimation problems for which the objective function takes arguments in some compact rectangle $K \subset \mathbb{R}^d$, $d \ge 1$. We focus on functions belonging to the Skorohod space $\mathcal{D}_K$ as defined in Neuhaus (1971). The elements of $\mathcal{D}_K$ are functions with finite "quadrant limits" (generalized one-sided limits) and are "continuous from above" (generalization of right-continuity) at each point in K. In Section 2 we describe the Skorohod space $\mathcal{D}_K$ in details and state some fundamental properties of the sargmax functional. Some of the results developed in this connection can also be of independent interest. In Section 3 we prove a version of the continuous mapping theorem for the sargmax functional for elements of $\mathcal{D}_K$ which are cádlág in the first component and jointly continuous on the last $d-1$. In Section 4 we describe an example that illustrates the necessity of the convergence of the associated pure jump processes in the results of Section 3. Finally, in Section 5 we apply the theorems of Section 3 to the change-point regression problem described above and to the estimation of a change-point in time and in a covariate in the Cox-proportional hazards model.

2. The Skorohod space $\mathcal{D}_K$

2.1. Definition and basic properties

We start by recalling the Skorohod space as discussed in Neuhaus (1971). To simplify notation, we write the coordinates of any vector in $\mathbb{R}^d$ with upper indices. We consider a compact rectangle $K = [a, b] = [a^1, b^1] \times \cdots \times [a^d, b^d]$ for some $a < b \in \mathbb{R}^d$ with the inequality holding componentwise. For any space $\mathbb{R}^m$ we will write $|\cdot|$ for the Euclidian norm (although the $\mathbb{L}^\infty$-norm is used in Neuhaus (1971), the results in there hold if one uses the Euclidian norm instead). For $k \in \{1,\dots,d\}$, $t \in [a^k, b^k]$ and $s \in \{a^k, b^k\}$ we write:

$$
\begin{aligned}
I_k(s,t) &:= \begin{cases} [a^k, t) & \text{if } s = a^k, \\ (t, b^k] & \text{if } s = b^k. \end{cases} \\
J_k(s,t) &:= \begin{cases} [a^k, t) & \text{if } s = a^k \text{ and } t < b^k, \\ [a^k, b^k] & \text{if } s = a^k \text{ and } t = b^k, \\ \emptyset & \text{if } s = b^k \text{ and } t = b^k, \\ [t, b^k] & \text{if } s = b^k \text{ and } t < b^k. \end{cases}
\end{aligned}
$$

and for any $\rho \in \mathcal{V} := \prod_{k=1}^d \{a^k, b^k\}$, $x = (x^1, \dots, x^d) \in \mathbb{R}^d$,

$$
\begin{aligned}
Q(\rho, x) &:= \prod_{k=1}^d I_k(\rho^k, x^k), \\
\tilde{Q}(\rho, x) &:= \prod_{k=1}^d J_k(\rho^k, x^k).
\end{aligned}
$$

Remark. Some properties of the sets $\tilde{Q}(\rho, x)$ are:

(a) $\tilde{Q}(\rho, x) \cap \tilde{Q}(\gamma, x) = \emptyset$ for every $\gamma \neq \rho \in \mathcal{V}$ and every $x \in K$.
(b) $K = \bigcup_{\rho \in \mathcal{V}} \tilde{Q}(\rho, x)$ for every $x \in K$.

Hence, $\{\tilde{Q}(\rho, x)\}_{\rho \in \mathcal{V}}$ forms a partition of K. We are now in a position to define the so-called quadrant limits, the concept of continuity from above and the Skorohod space.

Definition 2.1 (**Quadrant Limits and Continuity from Above**)**.** Consider a function $f : \mathbb{R}^d \to \mathbb{R}$, $\rho \in \mathcal{V}$ and $x \in K$. We say that a number l is the ρ-limit of f at x if for every sequence $\{x_n\}_{n=1}^\infty \subset Q(\rho, x)$ satisfying $x_n \to x$ we have $f(x_n) \to l$. In this case we write $l = f(x + 0_\rho)$. When $\rho = b$ we may write $f(x + 0_+) := f(x + 0_b)$. With this notation, f is said to be continuous from above at x if $f(x + 0_+) = f(x)$.

Definition 2.2 (**The Skorohod Space**)**.** We define the Skorohod space $\mathcal{D}_K$ as the collection of all functions $f : K \to \mathbb{R}$ which have all ρ-limits and are continuous from above at every $x \in K$.

Remark. It is easily seen that if $f \in \mathcal{D}_K$, $\rho \in \mathcal{V}$, $x \in K$ and $\{x_n\}_{n=1}^\infty \subset \tilde{Q}(\rho, x)$ is a sequence with $x_n \to x$, then $f(x_n) \to f(x + 0_\rho)$. This follows from the continuity from above as $Q(\rho, x) \cap Q(b, \xi) \neq \emptyset$ for every $\xi \in \tilde{Q}(\rho, x)$.

Before stating some of the most important properties of $\mathcal{D}_K$ we will introduce some further notation. Consider the partitions $\mathcal{T}_j = \{a^j = t_{j,0} < t_{j,1} < \cdots < t_{j,r_j} = b^j\}$ for $j = 1, \dots, d$. We define the rectangular partition $\mathcal{R}(\mathcal{T}_1, \dots, \mathcal{T}_d)$ determined by $\mathcal{T}_1, \dots, \mathcal{T}_d$ as the collection of all rectangles of the form

$$
R = \prod_{k=1}^d \left[t_{k,j_k-1}, t_{k,j_k}\right\rangle, j_k \in \{1, \dots, r_k\}, \; k = 1, \dots, d,
$$

where $\rangle$ stands for ")" or "]" if $t_{k,j_k} < b^k$ or $t_{k,j_k} = b^k$, respectively. With the aid of this notation, we can now state two important lemmas.

Lemma 2.1. *Let $f \in \mathcal{D}_K$. Then, for every $\epsilon > 0$ there is $\delta > 0$ and partitions $\mathcal{T}_j$ of $[a^j, b^j]$, $j = 1, \dots, d$, such that for any $R \in \mathcal{R}(\mathcal{T}_1, \dots, \mathcal{T}_d)$ and any $\theta, \vartheta \in R$ with $|\theta - \vartheta| < \delta$ the inequality $|f(\theta) - f(\vartheta)| < \epsilon$ holds. Furthermore, we can take the partitions in such a way that $\sup_{\theta,\vartheta \in R}\{|\theta - \vartheta|\} < \delta$ for every $R \in \mathcal{R}(\mathcal{T}_1, \dots, \mathcal{T}_d)$.*

Lemma 2.2. *Every function in $\mathcal{D}_K$ is bounded on K.*

Lemmas 2.1 and 2.2 are, respectively, Lemma 1.5 and Corollary 1.6 in Neuhaus (1971). Their proofs can be found there.

Let $K_1 = [a^1, b^1]$ and $K_2 = [a^2, b^2] \times \cdots \times [a^d, b^d]$, so $K = K_1 \times K_2$. We will be dealing with functions which are cádlág on the first coordinate and continuous on the remaining $d-1$. For this purpose we will turn our attention to the space $\widetilde{\mathcal{D}}_K \subset \mathcal{D}_K$ of all functions $f \in \mathcal{D}_K$ such that $f(t, \cdot) : K_2 \to \mathbb{R}$ is continuous $\forall$ $t \in K_1$ and $f(\cdot, \xi) : K_1 \to \mathbb{R}$ is cádlág $\forall$ $\xi \in K_2$.

Remark. It is worth noting that *all elements in $\mathcal{D}_K$ are componentwise cádlág*, so it is really the continuity in the last $d-1$ coordinates what makes $\widetilde{\mathcal{D}}_K$ a proper subspace of $\mathcal{D}_K$.

Lemma 2.3. *Let $f \in \widetilde{\mathcal{D}}_K$ and $\epsilon > 0$. Then, there is $\delta > 0$ such that*

$$\sup_{\substack{|\xi-\eta|<\delta \\ \xi,\eta \in K_2}} \{|f(t,\xi) - f(t,\eta)|\} \le \epsilon \quad \forall\ t \in K_1.$$

Proof. From Lemma 2.1 we can find $\delta_0 > 0$ and partitions $\mathcal{T}_j$ of $[a^j, b^j]$, $j = 1, \dots, d$ such that the conclusions of the lemma hold true with ϵ replaced by $\frac{\epsilon}{3}$. We take the partitions in such a way that whenever θ and ϑ belong to the same rectangle, the distance between them is less than δ_0. Let $s \in \mathcal{T}_1$. Since K_2 is compact and $f(s, \cdot)$ is continuous, we can find δ_s such that for any $\xi, \eta \in K_2$ with $|\xi - \eta| < \delta_s$ we get $|f(s,\xi) - f(s,\eta)| < \frac{\epsilon}{3}$. Let $\delta = \min_{s \in \mathcal{T}_1}\{\delta_s\}$ and pick $t \in K_1$ and $\xi, \eta \in K_2$ with $|\xi - \eta| < \delta$. Take the largest $s \in \mathcal{T}_1$ with $s \le t$. Then, $|s - t| < \delta_0$ and hence

$$|f(t,\eta) - f(t,\xi)| \le |f(t,\xi) - f(s,\xi)| + |f(s,\eta) - f(s,\xi)| + |f(t,\eta) - f(s,\eta)| < \epsilon.$$

The proof is then finished by taking the supremum over ξ and η and noticing that the choice of δ was independent of t. □

2.2. The Skorohod topology

So far we have not yet defined a topology on $\mathcal{D}_K$, so we turn our attention to this issue now. We will start by defining the Skorohod metric as given in Neuhaus (1971). Then, we will define a second metric on $\widetilde{D}_K$ and show that it is equivalent to the corresponding restriction of the Skorohod metric. This second metric will be more natural for the structure of $\widetilde{D}_K$ and will prove useful in the proof of the continuous mapping theorem for the smallest argmax functional. In order to define both of these metrics and state some of their properties, we will need some additional notation.

Consider a closed interval $I \subset \mathbb{R}$ and the class Λ_I of all functions $\lambda : I \to I$ which are surjective (onto) and strictly monotone increasing. Define the function $|||\cdot|||_I : \Lambda_I \to \mathbb{R}$ by the formula $|||\lambda|||_I = \sup_{s\neq t}\{\left|\log\left(\frac{\lambda(t)-\lambda(s)}{t-s}\right)\right|\}$. We write $\Lambda_K := \Lambda_{[a^1,b^1]} \times \cdots \times \Lambda_{[a^d,b^d]}$ and for $\lambda := (\lambda_1, \dots, \lambda_d) \in \Lambda_K$, $|||\lambda|||_K := \max_{1\leq k\leq d}\{|||\lambda_k|||_{[a^k,b^k]}\}$. In a similar fashion, we define $\Lambda_{K_2} := \Lambda_{[a^2,b^2]} \times \cdots \times \Lambda_{[a^d,b^d]}$ and for $\lambda \in \Lambda_{K_2}$, $|||\lambda|||_{K_2} := \max_{2\leq k\leq d}\{|||\lambda_k|||_{[a^k,b^k]}\}$. Note that for $(\lambda_1, \lambda) \in \Lambda_K = \Lambda_{K_1} \times \Lambda_{K_2}$ we have $|||(\lambda_1, \lambda)|||_K = |||\lambda_1|||_{K_1} \vee |||\lambda|||_{K_2}$. We will use the sup-norm notation also: for a function $f : A \to \mathbb{R}$ we write $\|f\|_A = \sup_{x\in A}\{|f(x)|\}$.

Definition 2.3 (**The Skorohod metric**)**.** We define the Skorohod metric $d_K : \mathcal{D}_K \times \mathcal{D}_K \to \mathbb{R}$ as follows:

$$d_K(f, g) = \inf_{\lambda\in\Lambda_K} \{|||\lambda|||_K + \|f - g \circ \lambda\|_K\}.$$

With this definition we can now state the following fundamental result about the Skorohod space.

Lemma 2.4. *The Skorohod metric is a metric. If $\mathcal{D}_K$ is endowed with the topology defined by d_K, then it becomes a Polish space.*

For a proof of the last result, we refer the reader to Section 2 in Neuhaus (1971). We now proceed to define another metric, $\widetilde{d}_K$, on $\mathcal{D}_K$ by the formula:

$$\widetilde{d}_K(f, g) = \inf_{\lambda\in\Lambda_{[a^1,b^1]}} \left\{ |||\lambda|||_{[a^1,b^1]} + \sup_{(t,\xi)\in K_1\times K_2} \{|f(t, \xi) - g(\lambda(t), \xi)|\} \right\}.$$

To properly describe the properties of $\widetilde{d}_K$ we need the ball notation for metric spaces: given a metric space $(\mathtt{X}, d)$, $r > 0$ and $x \in \mathtt{X}$ we write $B^d_r(x)$ for the open ball of radius r and center at x with respect to the metric d. Additionally, the following lemma will prove to be useful.

Lemma 2.5. *Let $I \subset \mathbb{R}$ be any compact interval. Then, for $\epsilon > 0$ there is $\delta > 0$ such that for any $\lambda \in \Lambda_I$ with $|||\lambda|||_I < \delta$ we also have*

$$\sup_{s\in I}\{|\lambda(s) - s|\} < \epsilon.$$

Proof. Assume that $I = [u, v]$. It suffices to choose $\delta < \frac{1}{4} \wedge \frac{\epsilon}{2|v-u|}$. To see this, observe that for any $\tau \in (0, \frac{1}{4})$, $\tau < 2\tau - 4\tau^2 \leq \log(1 + 2\tau)$ and for any $\tau > -1$, $\log(1 + \tau) \leq \tau$. It follows that for $\lambda \in \Lambda_I$ with $|||\lambda|||_I < \delta$ and any $s \in I$, $\log(1 - 2\delta) < -\delta \leq \log\frac{\lambda(s)-u}{s-u} \leq \delta < 2\delta - 4\delta^2 \leq \log(1 + 2\delta)$ and thus, $|\lambda(s) - s| < 2(s - u)\delta \leq 2|u - v|\delta$. In the previous inequalities we have made implicit use of the fact that $\lambda(u) = u$. □

The next lemma contains some of the most relevant properties of $\widetilde{d}_K$.

Lemma 2.6. *The following statements are true:*

(i) $\widetilde{d}_K$ *is a metric on* $\mathcal{D}_K$.
(ii) $d_K(f,g) \le \widetilde{d}_K(f,g) \le \|f-g\|_K \ \forall\ f,g \in \mathcal{D}_K$.
(iii) *If* $f \in \widetilde{\mathcal{D}}_K$, *then for every* $r>0$ *there is* $\delta > 0$ *such that* $B_\delta^{d_K}(f) \subset B_r^{\widetilde{d}_K}(f)$. *Moreover, the metrics* d_K *and* $\widetilde{d}_K$ *generate the same topology on* $\widetilde{\mathcal{D}}_K$.
(iv) *If* f *is continuous, then for every* $r>0$ *there is* $\delta>0$ *such that* $B_\delta^{\widetilde{d}_K}(f) \subset B_r^{\|\cdot\|_K}(f)$. *Moreover, the metrics* d_K *and* $\widetilde{d}_K$ *and* $\|\cdot\|_K$ *generate the same topology on the space of continuous functions on* K.
(v) $(\widetilde{\mathcal{D}}_K, d_K)$ *is a Polish space.*

Proof. It is straightforward to see that (ii) holds. The proof of (i) follows along the lines of the proof of the analogous results for the classical Skorohod metric (see Chapter 3 of Billingsley (1968)). For the sake of brevity we omit these arguments. For (iii) we use Lemma 2.3. Let $f \in \widetilde{\mathcal{D}}_K$, $r>0$ and take $\delta_1 > 0$ such that the conclusions of Lemma 2.3 hold with $\frac{r}{3}$ replacing ϵ. Also, consider $\delta_2 > 0$ such that $|||\lambda|||_{K_2} < \delta_2$ implies $\sup_{\xi\in K_2}\{|\lambda(\xi)-\xi|\} < \delta_1$ (whose existence is a consequence of Lemma 2.5 applied to each of the intervals $[a^2,b^2],\dots,[a^d,b^d]$). Let $\delta = \delta_2 \wedge \frac{r}{3}$ and take $g \in B_\delta^{d_K}(f)$. Find $(\lambda_1,\lambda) \in \Lambda_K = \Lambda_{K_1}\times\Lambda_{K_2}$ such that $|||(\lambda_1,\lambda)|||_K < \delta$ and $\|g - f\circ(\lambda_1,\lambda)\|_K < \frac{r}{3}$. Then, for any $(t,\xi) \in K_1\times K_2$ we have:

$$\begin{aligned}|g(t,\xi) - f(\lambda_1(t),\xi)| &\le |g(t,\xi) - f(\lambda_1(t),\lambda(\xi))| + |f(\lambda_1(t),\lambda(\xi)) - f(\lambda_1(t),\xi)| \\ &< \frac{r}{3} + \frac{r}{3},\end{aligned}$$

where the second term in the sum of the right-hand side of the first inequality in the preceding display is less than $\frac{r}{3}$ because of Lemma 2.3 since $|||\lambda|||_{K_2} < \delta_2$. Taking supremum over $(t,\xi)\in K$ and considering that $|||\lambda_1|||_{K_1} < \frac{r}{3}$ we get that $\widetilde{d}_K(f,g) < r$. Thus, $B_\delta^{d_K}(f) \subset B_r^{\widetilde{d}_K}(f)$. Taking (ii) into account we can conclude that $\widetilde{d}_K$ and d_K are equivalent metrics on $\widetilde{\mathcal{D}}_K$.

We now turn out attention to (iv). Let $r>0$. Then, there is $\delta_1>0$ such that $|f(x)-f(y)| < \frac{r}{2}$ whenever $|x-y|<\delta_1$. Also, there is $\delta_2>0$ such that $|||\lambda|||_{K_1} < \delta_2$ implies $\sup_{t\in K_1}\{|\lambda(t)-t|\} < \delta_1$. Let $\delta = \delta_2 \wedge \frac{r}{2}$ and let $g \in \mathcal{D}_K$ with $\widetilde{d}_K(f,g) < \delta$ and $\lambda \in \Lambda_{K_1}$ such that $|||\lambda\,|||_{K_1} + \|g(\cdot,\cdot) - f(\lambda(\cdot),\cdot)\|_{K_1\times K_2} < \delta$. Then, for any $(t,\xi)\in K_1\times K_2$ we have

$$|f(t,\xi) - g(t,\xi)| \le |f(t,\xi) - f(\lambda(t),\xi)| + |f(\lambda(t),\xi) - g(t,\xi)| < r.$$

Thus, $B_\delta^{\widetilde{d}_K}(f) \subset B_r^{\|\cdot\|_K}(f)$.

To prove (v) it suffices to show that $\widetilde{\mathcal{D}}_K$ is a closed subset of $\mathcal{D}_K$, as the latter space is known to be Polish (see Neuhaus (1971)). Let $(f_n)_{n=1}^\infty$ be a sequence in $\widetilde{\mathcal{D}}_K$ such that $f_n \xrightarrow{d_K} f$ for some $f \in \mathcal{D}_K$. We will show that $f(t,\cdot)$ is continuous for every t and that will imply that $f \in \widetilde{\mathcal{D}}_K$ since f is automatically componentwise cádlág. Let $(t,\xi) \in K_1 \times K_2 = K$ and $\epsilon > 0$. Consider $n \in \mathbb{N}$ large enough so that $d_K(f, f_n) < \frac{\epsilon}{3}$ and take $\delta_1 > 0$ such that the conclusions of Lemma 2.3 hold true for f_n and $\frac{\epsilon}{3}$. Let $(\lambda_{n,1}, \lambda_n) \in \Lambda_{K_1} \times \Lambda_{K_2}$ such that $|||(\lambda_{n,1}, \lambda_n)|||_K + \|f - f_n \circ (\lambda_{n,1}, \lambda_n)\|_K < \frac{\epsilon}{3}$. Since λ_n is continuous, there is $\delta > 0$ such that $|\xi - \eta| < \delta$ implies $|\lambda_n(\xi) - \lambda_n(\eta)| < \delta_1$. It follows that $|f_n(\lambda_{n,1}(t), \lambda_n(\xi)) - f_n(\lambda_{n,1}(t), \lambda_n(\eta))| < \frac{\epsilon}{3}$ whenever $|\xi - \eta| < \delta$. Hence,

$$\begin{aligned}|f(t,\xi) - f(t,\eta)| \;\le\; & |f(t,\xi) - f_n(\lambda_{n,1}(t), \lambda_n(\xi))| + |f(t,\eta) - f_n(\lambda_{n,1}(t), \lambda_n(\eta))| \\ & + |f_n(\lambda_{n,1}(t), \lambda_n(\xi)) - f_n(\lambda_{n,1}(t), \lambda_n(\eta))| \\ <\; & \epsilon, \quad \forall\, \xi, \eta \in K_2 \text{ such that } |\xi - \eta| < \delta.\end{aligned}$$

It follows that $f(t, \cdot)$ is continuous for every $t \in K_1$. Hence, $f \in \widetilde{\mathcal{D}}_K$ and $\widetilde{\mathcal{D}}_K$ is closed. □

Remark. Observe that the previous lemma implies that for a convergent sequence in $\mathcal{D}_K$ with a limit in $\widetilde{\mathcal{D}}_K$ convergence in the $\widetilde{d}_K$ and d_K metrics are equivalent. When the limit is continuous, convergence in any of these metrics is equivalent to convergence in the sup-norm topology.

2.3. The sargmax functional on $\mathcal{D}_K$

We now turn our attention to the smallest argmax functional on $\mathcal{D}_K$.

Definition 2.4 (The sargmax Functional). A function $f \in \mathcal{D}_K$ is said to have a maximizer at a point $x \in K$ if any of the quadrant-limits of x equals $\sup_{\xi \in K}\{f(\xi)\}$. For any $f \in \mathcal{D}_K$ we can define the *smallest argmax* of f over the compact rectangle K, denoted by $\mathrm{sargmax}_{x \in K}\{f(x)\}$, as the unique element $x = (x^1, \ldots, x^d) \in K$ satisfying the following properties:

(i) x is a maximizer of f over K,
(ii) if $\xi = (\xi^1, \ldots, \xi^d)$ is any other maximizer, then $x^1 \le \xi^1$,
(iii) if ξ is any maximizer satisfying $x^j = \xi^j \ \forall\ j = 1, \ldots, k$ for some $k \in \{1, \ldots, d-1\}$, then $x^{k+1} \le \xi^{k+1}$.

We say that x is the largest maximizer of f, denoted by $\mathrm{largmax}_{\xi \in K}\{f(\xi)\}$, if it is a maximizer that satisfies *(ii)* and *(iii)* above with the inequalities reversed.

The first question that one might ask is whether or not the sargmax is well defined for all functions in the Skorohod space. Before attempting to give an answer, we will use our notation to clarify the concept of a maximizer: a point $x \in K$ is a maximizer of $f \in \mathcal{D}_K$ if

$$\max_{\rho \in \mathcal{V}}\{f(x + 0_\rho)\} = \sup_{\xi \in K}\{f(\xi)\}.$$

We can now prove a result concerning the set of maximizers of a function in $\mathcal{D}_K$.

Lemma 2.7. *The set of maximizers of any function in $\mathcal{D}_K$ is compact.*

Proof. Let $f \in \mathcal{D}_K$. Since the set of maximizers of f is a subset of the compact rectangle K, it suffices to show that any convergent sequence of maximizers converges to a maximizer. Let $(x_n)_{n=1}^{\infty}$ be a sequence of maximizers with limit x. For each x_n we can find ξ_n with $|x_n - \xi_n| < \frac{1}{n}$ and such that $|f(\xi_n) - \max_{\rho \in \mathcal{V}}\{f(x_n + 0_\rho)\}| < 1/n$. Then we have that $\xi_n \to x$ and $|f(\xi_n) - \sup_{\xi \in K}\{f(\xi)\}| < 1/n$ $\forall$ $n \in \mathbb{N}$. Since K is the disjoint union of $\{\tilde{Q}(\rho, x)\}_{\rho \in \mathcal{V}}$, it follows that there is $\rho_* \in \mathcal{V}$ and a subsequence $(\xi_{n_k})_{k=1}^{\infty}$ such that $\xi_{n_k} \in \tilde{Q}(\rho_*, x)$ $\forall$ $k \in \mathbb{N}$. Therefore, the remark stated right after the definition of the Skorohod space implies that $f(\xi_{n_k}) \to f(x + 0_{\rho_*})$ and, consequently, $f(x + 0_{\rho_*}) = \sup_{\xi \in K}\{f(\xi)\}$. □

The previous lemma can be used to show that the sargmax functional is well defined on $\mathcal{D}_K$.

Lemma 2.8. *For each $f \in \mathcal{D}_K$ there is a unique element in $x \in K$ such that $x = sargmax_{\xi \in K}\{f(\xi)\}$.*

Proof. Let $f \in \mathcal{D}_K$. Since the set of maximizers of f is compact, if we can show that it is nonempty then the compactness will imply that there is a unique element $x \in K$ satisfying properties (i), (ii) and (iii) of Definition 2.4. Hence, it suffices to show that f has at least one maximizer. For this purpose, for each $n \in \mathbb{N}$ choose x_n such that $\sup_{\xi \in K}\{f(\xi)\} < f(x_n) + \frac{1}{n}$. Since K is compact, there is $x \in K$ and a subsequence $(x_{n_k})_{k=1}^{\infty}$ such that $x_{n_k} \to x$. Just as in the proof of the previous lemma, we can find $\rho_* \in \mathcal{V}$ and a further subsequence $(x_{n_{k_s}})_{s=1}^{\infty}$ such that $x_{n_{k_s}} \in \tilde{Q}(\rho_*, x)$ $\forall$ $s \in \mathbb{N}$. It follows that $f(x_{n_{k_s}}) \to f(x + 0_{\rho_*})$ and hence $\sup_{\xi \in K}\{f(\xi)\} = f(x + 0_{\rho_*})$. Therefore, the set of maximizers is nonempty and the sargmax is well defined. □

We finish this section with a continuity theorem for the sargmax functional on continuous functions.

Lemma 2.9. *Let $W \in \mathcal{D}_K$ be a continuous function which has a unique maximizer $x^* \in K$. Then, the smallest argmax functional is continuous at W (with respect to d_K, $\widetilde{d}_K$ and the sup-norm metric).*

Proof. Let $(W_n)_{n=1}^{\infty}$ be a sequence converging to W in the Skorohod topology. Let $\epsilon > 0$ be given and G be the open ball of radius ϵ around x^* and let $\delta := (W(x^*) - \sup_{x \in K \setminus G}\{W(x)\})/2 > 0$. By Lemma 2.6 we have $\|W_n - W\|_K < \delta$ for all large n (d_K, $\widetilde{d}_K$ and $\|\cdot\|_K$ generate the same local topology on W). Then

$$W(x^*) = 2\delta + \sup_{x \in K \setminus G}\{W(x)\} > \delta + \sup_{x \in K \setminus G}\{W_n(x)\}.$$

But $\|W_n - W\|_K < \delta$ also implies that $\sup_{x \in K}\{W_n(x)\} > W(x^*) - \delta$. The combination of these two facts shows that if $\|W_n - W\|_K < \delta$, then any maximizer of W_n must belong to G. Thus, $|\operatorname{sargmax}_{x \in K}\{W_n(x)\} - x^*| < \epsilon$ for n large enough. □

3. A continuous mapping theorem for the sargmax functional on functions with jumps

Lemma 2.9 shows that the sargmax functional is continuous on continuous functions with unique maximizers. However, its raison d'être is to fix a unique maximizer on a function having multiple maximizers. Thus, a continuous mapping theorem on functions with jumps and possibly multiple maximizers is desired. We will show a version of the continuous mapping theorem on a suitable subset of our space $\widetilde{\mathcal{D}}_K$.

To state and prove our version of the continuous mapping theorem for the sargmax functional, we need to introduce some notation. We start with the space $\mathcal{D}_K^0$ consisting of all functions $\psi : K_1 \times K_2 \to \mathbb{R}$ which can be expressed as:

$$\psi(t,\xi) = V_0(\xi)\mathbf{1}_{a_{-1}\le t<a_1} + \sum_{k=1}^{\infty} V_k(\xi)\mathbf{1}_{a_k\le t<a_{k+1}} + \sum_{k=1}^{\infty} V_{-k}(\xi)\mathbf{1}_{a_{-k-1}\le t<a_{-k}} \quad (3)$$

where $(\cdots < a_{-k-1} < a_{-k} < \cdots < a_0 = 0 < \cdots < a_k < a_{k+1} < \cdots)_{k\in\mathbb{N}}$ is a sequence of jumps and $(V_k)_{k\in\mathbb{Z}}$ is a collection of continuous functions. Note that $\mathcal{D}_K^0 \subset \widetilde{\mathcal{D}}_K$. Observe that the representation in (3) is not unique. However, knowledge of the function ψ and of the jumps $(a_k)_{k\in\mathbb{Z}}$ completely determines the continuous functions $(V_k)_{k\in\mathbb{Z}}$.

Our theorem will require not only Skorohod convergence of the elements of $\mathcal{D}_K^0$, but also convergence of their associated *pure jump functions*. To define properly these jump functions, we introduce the space $\mathcal{S}$ all piecewise constant, cádlág functions $\tilde{\psi} : \mathbb{R} \to \mathbb{R}$ such that $\tilde{\psi}(0) = 0$; $\tilde{\psi}$ has jumps of size 1; and $\tilde{\psi}(-t)$ and $\tilde{\psi}(t)$ are nondecreasing on $(0,\infty)$. For any closed interval $I \subset \mathbb{R}$ we introduce the space $\mathcal{S}_I := \{f|_I : f \in \mathcal{S}\}$. We endow the spaces $\mathcal{S}_I$ with the usual Skorohod topology d_I. Observe that the fact that all elements of $\mathcal{S}$ are cádlág and have jumps of size one implies that *any function in $\mathcal{S}_I$ has a finite number of jumps on I.*

We associate with every $\psi \in \mathcal{D}_K^0$, expressed as in (3), a pure jump function $\tilde{\psi} \in \mathcal{S}$ whose sequence of jumps is exactly the a_k's, i.e.,

$$\tilde{\psi}(t) = \sum_{k=1}^{\infty} \mathbf{1}_{a_k\le t} + \sum_{k=1}^{\infty} \mathbf{1}_{a_{-k}>t}. \quad (4)$$

We will show that Skorohod-convergence of functions in $\mathcal{D}_K^0$ and Skorohod convergence of their associated pure jump functions implies convergence of the corresponding sargmax and largmax functionals.

The following convergence result is a generalization of both, Lemma 3.1 of Lan et al. (2009) and Lemma A.3 in Seijo and Sen (2010).

Theorem 3.1. *Assume that $d \ge 2$ and let $(\psi_n, \tilde{\psi}_n)_{n=1}^{\infty}$, $(\psi_0, \tilde{\psi}_0)$ be functions in $\mathcal{D}_K^0 \times \mathcal{S}_{K_1}$ such that ψ_n satisfies (3) for the sequence of jumps of $\tilde{\psi}_n$ for any $n \ge 0$. Assume that $(\psi_n, \tilde{\psi}_n) \to (\psi_0, \tilde{\psi}_0)$ in $\mathcal{D}_K^0 \times \mathcal{S}_{K_1}$ (with the product topology). Suppose, in addition, that ψ_0 can be expressed as (3) for the sequence of* *jumps* $(\cdots < a_{-k-1} < a_{-k} < \cdots < a_0 = 0 < \cdots < a_k \ < a_{k+1} < \cdots)_{k\in\mathbb{N}}$ *of* $\tilde{\psi}_0$ *and some continuous functions* $(V_j)_{j\in\mathbb{Z}}$*, each having a unique maximizer on* K_2*, with the property that for any finite subset* $A \subset \mathbb{Z}$ *there is only one* $j \in A$ *for which*

$$\max_{m\in A}\left\{\sup_{\xi\in K_2}\{V_m(\xi)\}\right\} = \sup_{\xi\in K_2}\{V_j(\xi)\}\,. \tag{5}$$

Finally, assume that ψ_0 *has no jumps at the extreme points of* K_1*. Then,*

(i) $\underset{x\in K}{\textit{sargmax}}\{\psi_n(x)\} \to \underset{x\in K}{\textit{sargmax}}\{\psi_0(x)\}$ *as* $n\to\infty$*;*
(ii) $\underset{x\in K}{\textit{largmax}}\{\psi_n(x)\} \to \underset{x\in K}{\textit{largmax}}\{\psi_0(x)\}$ *as* $n\to\infty$*.*

The result is also true when $d = 1$ *under the same assumptions, but taking the sequence* $(V_j)_{j\in\mathbb{Z}}$ *to be a sequence of constants such that for any finite subset* $A \subset \mathbb{Z}$ *there is a unique* $j \in A$ *such that* $\max_{m\in A}\{V_m\} = V_j$*.*

Proof. We focus on the case when $d > 1$ as the one-dimensional case is just Lemma 3.1 of Lan et al. (2009). Without loss of generality, assume that $K_1 = [-C, C]$ for some $C > 0$.

We can write ψ_n in the form (3) with $(\cdots < a_{n,-k-1} < a_{n,-k} < \cdots < a_{n,0} = 0 < \cdots < a_{n,k} < a_{n,k+1} < \cdots)_{k\in\mathbb{N}}$ being the sequence of jumps of ψ_n and $V_{n,j}$ being the continuous functions. Consequently, $\tilde{\psi}_n$, the pure jump function associated with ψ_n, can be expressed as (4) with jumps at $(a_{n,k})_{k\in\mathbb{Z}}$.

Let N_r and N_l be the number of jumps of $\tilde{\psi}_0$ in $[0, C]$ and $[-C, 0)$ respectively. Let $\epsilon > 0$ be sufficiently small such that all the points of the form $a_j \pm \epsilon$ are continuity points of ψ_0, for $-N_l \le j \le N_r$. Since convergence in the Skorohod topology of $\tilde{\psi}_n$ to $\tilde{\psi}_0$ implies point-wise convergence for continuity points of $\tilde{\psi}_0$ (see page 121 of Billingsley (1968)), and all of them are integer-valued functions, we see that $\tilde{\psi}_n(a_j - \epsilon) = j - 1$ and $\tilde{\psi}_n(a_j + \epsilon) = j$ for any $1 \le j \le N_r$, and $\tilde{\psi}_n(C) = N_r$ for all sufficiently large n. Thus, for all but finitely many n's we have that $\tilde{\psi}_n$ has exactly N_r jumps between 0 and C and that the location of the j-th jump to the right of 0 satisfies $|a_{n,j} - a_j| < \epsilon$. Since $\epsilon > 0$ can be made arbitrarily small, we get that all the jumps $a_{n,j}$ converge to their corresponding a_j for all $1 \le j \le N_r$. The same happens to the left of zero: for all but finitely many n's, $\tilde{\psi}_n$ has exactly N_l jumps in $[-C, 0)$ and the sequences of jumps $(a_{n,-j})_{n=1}^{\infty}$, $1 \le j \le N_l$, converge to the corresponding jumps a_{-j}.

Let $V^* = \sup\{V_j(\xi) : \xi \in K_2, -N_l \le j \le N_r\}$. Our assumptions on the V_j's imply that this supremum is actually achieved at some unique vector $\xi^* \in K_2$ and that there is a unique "flat stretch" at which this supremum is attained (the last assertion follows form (5)).

Suppose, without loss of generality, that the maximum value is achieved in an interval of the form $[a_k, a_{k+1} \wedge C)$ for a unique $k \in \{1, \ldots, N_r\}$. Now, write $b_0 = 0$; $b_j = \frac{a_j + C\wedge a_{j+1}}{2}$ for $1 \le j \le N_r$; and $b_j = \frac{a_j + (-C)\vee a_{j-1}}{2}$ for $-N_l \le j \le -1$. Note that the b_j's (for any value of $\xi \in K_2$) are continuity points of both ψ_0 and $\tilde{\psi}_0$.

Let $\kappa = \min_{-N_l \le j \le N_r+1}(C \wedge a_j - (-C) \vee a_{j-1})$ be the length of the shortest stretch. Take $0 < \eta, \delta < \kappa/4$. Considering the convergence of the jumps of ψ_n to those of ψ_0, there is $N \in \mathbb{N}$ such that for any $n \ge N$, the following two statements hold:

(a) Consider $\rho > 0$ such that if $|||\lambda|||_{K_1} < \rho$, then

$$\sup\{|s - \lambda(s)| : s \in [-C, C]\} < \delta.$$

The existence of such ρ follows from Lemma 2.5. By the convergence of ψ_n to ψ_0 in the Skorohod topology, there exists $\lambda_n \in \Lambda_{K_1}$ such that $|||\lambda_n|||_{K_1} < \rho$ and

$$\sup_{(t,\xi)\in K_1\times K_2} \{|\psi_n(\lambda_n(t), \xi) - \psi_0(t, \xi)|\} < \eta.$$

(b) For any $1 \le j \le N_r$ (respectively, $j = 0$, $-N_l \le j \le -1$), b_j lies somewhere inside the interval $(a_{n,j} + \delta, C \wedge a_{n,j+1} - \delta)$ (respectively $(a_{n,-1} + \delta, a_{n,1} - \delta)$, $((-C) \vee a_{n,j-1} + \delta, a_{n,j} - \delta)$). This follows from what was proven in the first two paragraphs of this proof.

From (a) we see that $|\lambda_n(b_j) - b_j| < \delta$ for all $-N_l \le j \le N_r$. But (b) and the size of δ in turn imply that b_j and $\lambda_n(b_j)$ belong to the same "flat stretch" of ψ_n and thus $\psi_n(\lambda_n(b_j), \xi) = \psi_n(b_j, \xi) = V_{n,j}(\xi)$ for all $\xi \in K_2$ and all $-N_l \le j \le N_r$. Considering again (b) and the second inequality in (a), we conclude that $\|V_{n,j} - V_j\|_{K_2} < \eta$ for all $-N_l \le j \le N_r$ and all $n \ge N$. Hence, all the sequences $(V_{n,j})_{n=1}^{\infty}$ converge uniformly in K_2 to their corresponding V_j. Consequently:

$$\begin{aligned}
\max_{\substack{-N_l \le j \le N_r \\ j \ne k}} \left\{ \sup_{\xi \in K_2} V_{n,j}(\xi) \right\} &\longrightarrow \max_{\substack{-N_l \le j \le N_r \\ j \ne k}} \left\{ \sup_{\xi \in K_2} V_j(\xi) \right\}, \\
\max_{\xi \in K_2} \{V_{n,k}(\xi)\} &\longrightarrow \max_{\xi \in K_2} \{V_k(\xi)\} = V_k(\xi^*), \\
\operatorname*{argmax}_{\xi \in K_2} \{V_{n,k}(h_1, h_2)\} &\longrightarrow \operatorname*{argmax}_{\xi \in K_2} \{V_k(\xi)\} = \xi^*, \\
\overline{\lim_{n\to\infty}} \max_{\substack{-N_l \le j \le N_r \\ j \ne k}} \left\{ \sup_{\xi \in K_2} V_{n,j}(\xi) \right\} &< \underline{\lim_{n\to\infty}} \max_{\xi \in K_2} \{V_{n,k}(\xi)\}.
\end{aligned}$$

The above, together with (5) and the fact that $a_{n,k} \to a_k$ and $a_{n,k+1} \to a_{k+1}$, imply that

$$\begin{aligned}
&\operatorname*{sargmax}_{x\in K}\{\psi_n(x)\} \to (\xi^*, a_k) = \operatorname*{sargmax}_{x\in K}\{\psi_0(x)\} \\
&\operatorname*{largmax}_{x\in K}\{\psi_n(x)\} \to (\xi^*, a_{k+1}) = \operatorname*{largmax}_{x\in K}\{\psi_0(x)\}
\end{aligned}$$

as $n \to \infty$. □

We now present a version of the previous result but for random elements in $\mathcal{D}_K^0$. To prove it, we will use Lemma 4.2 in Prakasa Rao (1969). In the remaining of the paper we will use the symbol $\rightsquigarrow$ to represent weak convergence.

Lemma 3.1. *Consider the random vectors* $\{W_{n\epsilon}, W_n, W_\epsilon\}_{\epsilon\geq 0}^{n\in\mathbb{N}}$ *and* W. *Suppose that the following conditions hold:*

(i) $\lim_{\epsilon\to 0}\overline{\lim}_{n\to\infty}\mathbf{P}(W_{n\epsilon}\neq W_n) = 0$,
(ii) $\lim_{\epsilon\to 0}\mathbf{P}(W_\epsilon\neq W) = 0$,
(iii) $W_{n\epsilon}\rightsquigarrow W_\epsilon$ *(as* $n\to\infty$*) for every* $\epsilon > 0$.

Then, $W_n \rightsquigarrow W$.

In the next theorem we will be taking the sargmax and largmax functionals over rectangles that may not be compact. When this happens, we say that these functionals are *well defined* if there is an element in the corresponding rectangle satisfying conditions $(i)-(iii)$ defining the smallest and largest argmax functionals (see Definition 2.4). If we are given a rectangle $\Theta\subset\mathbb{R}^d$ which can be written as the Cartesian product of possibly unbounded closed intervals, we will denote by $\mathcal{D}_\Theta$ the collection of functions $f:\Theta\to\mathbb{R}$ whose restrictions to all compact rectangles $K\subset\Theta$ belong to $\mathcal{D}_K$.

Theorem 3.2. *Assume that* $K = K_1\times K_2$ *is a closed rectangle in* $\mathbb{R}^d$ *and that* $0\in K_1^\circ$. *Let* $(\Omega,\mathcal{F},\mathbf{P})$ *be a probability space and let* $(\Psi_n,\Gamma_n)_{n=1}^\infty$, (Ψ_0,Γ_0) *be random elements taking values in* $\mathcal{D}_K^0\times\mathcal{S}_{K_1}$ *such that* Ψ_n *satisfies (3) for the sequence of jumps of* Γ_n *for any* $n\geq 0$*, almost surely. Moreover, suppose that, with probability one, we have that:* Ψ_0 *satisfies (5);* Γ_0 *has no fixed time of discontinuity; the sargmax and largmax functionals over* K *are finite for* Ψ_0 *(this assumption is essential as* K *is not necessarily compact). If the following hold:*

(i) For every compact subinterval $B_1\subset K_1$ *and compact sub-rectangle* $B := B_1\times B_2\subset K$ *we have* $(\Psi_n,\Gamma_n)\rightsquigarrow(\Psi_0,\Gamma_0)$ *on* $\mathcal{D}_B\times\mathcal{D}_{B_1}$;
(ii) $\left(\mathop{sargmax}_{\theta\in K}\{\Psi_n(\theta)\}, \mathop{largmax}_{\theta\in K}\{\Psi_n(\theta)\}\right) = O_{\mathbf{P}}(1)$;

then we also have

$$\left(\mathop{sargmax}_{\theta\in K}\{\Psi_n(\theta)\}, \mathop{largmax}_{\theta\in K}\{\Psi_n(\theta)\}\right)\rightsquigarrow\left(\mathop{sargmax}_{\theta\in K}\{\Psi_0(\theta)\}, \mathop{largmax}_{\theta\in K}\{\Psi_0(\theta)\}\right).$$

Proof. Consider $C>0$ and let

$$\begin{aligned}\phi_n &:= \left(\mathop{\mathrm{sargmax}}_{\theta\in K}\{\Psi_n(\theta)\}, \mathop{\mathrm{largmax}}_{\theta\in K}\{\Psi_n(\theta)\}\right)\\ \phi_{n,C} &:= \left(\mathop{\mathrm{sargmax}}_{\theta\in[-C,C]^d\cap K}\{\Psi_n(\theta)\}, \mathop{\mathrm{largmax}}_{\theta\in[-C,C]^d\cap K}\{\Psi_n(\theta)\}\right),\end{aligned}$$

for all $n\geq 0$. To prove the result, we will apply Theorem 3.1 and Lemma 3.1. Using the notation of the latter, set $\epsilon = \frac{1}{C}$, $W_{n\epsilon} = \phi_{n,C}$ for $n\geq 1$, $W_\epsilon = \phi_{0,C}$, $W_n = \phi_n$ for $n\geq 1$ and $W = \phi_0$. From (ii) we see that $\lim_{\epsilon\to 0}\overline{\lim}_{n\to\infty}\mathbf{P}(W_{n\epsilon}\neq W_n) = 0$. Our assumptions on Ψ_0 and Γ_0 imply that $\lim_{\epsilon\to 0}\mathbf{P}(W_\epsilon\neq W) = 0$. Finally, Theorem 3.1 and an application of Skorohod's Representation Theorem (see either Theorem 1.8, page 102 in Ethier and Kurtz (2005) or Theorems 1.10.3 and 1.10.4, pages 58 and 59 in Van der Vaart and Wellner (1996)) show that $W_{n\epsilon} \rightsquigarrow W_\epsilon$ and hence, from Lemma 3.1, we conclude that $\phi_n \rightsquigarrow \phi_0$. □

4. On the necessity of the convergence of the associated pure jump processes

Condition (i) in Theorem 3.2 involves the joint convergence of the processes whose maximizers are being considered and their associated pure jump processes. One may ask whether or not this condition is actually necessary for the weak convergence of the corresponding smallest maximizers. A simple counterexample shows that such a condition is indeed essential to guarantee the desired weak convergence under the assumptions of Theorem 3.2.

Let Ψ be a two-sided, right-continuous Poisson process and $T_{\pm 1} := \pm \inf\{t > 0 : \Psi(\pm t) > 0\}$. Consider the following $\mathcal{D}_{\mathbb{R}}$-valued random elements: $\Psi_0 := -\Psi$ and $\Psi_n = \Psi_0 + \frac{1}{n}\mathbf{1}_{\left[\frac{1}{2}T_{-1}, \frac{1}{2}T_1\right)}$. Then, $\Psi_n \rightsquigarrow \Psi$ in $\mathcal{D}_I$ for every compact interval I (in fact, the weak convergence holds in $\mathcal{D}_{\mathbb{R}}$ with the corresponding Skorohod topology). However,

$$\left(\underset{\mathbb{R}}{\operatorname{sargmax}}\{\Psi_n\}, \underset{\mathbb{R}}{\operatorname{largmax}}\{\Psi_n\}\right) = \frac{1}{2}\left(\underset{\mathbb{R}}{\operatorname{sargmax}}\{\Psi_0\}, \underset{\mathbb{R}}{\operatorname{largmax}}\{\Psi_0\}\right),$$

for all $n \in \mathbb{N}$. It is easily seen that all the conditions of Theorem 3.2 hold, with the exception of (i). Hence, the weak convergence of the processes Ψ_n alone is not enough to guarantee weak convergence of the corresponding maximizers.

5. Applications

5.1. Stochastic design change-point regression

We start by analyzing the example of the least squares change-point estimator given by (2) in the Introduction. Assume that we are given an i.i.d. sequence of random vectors $\{X_n = (Y_n, Z_n)\}_{n=1}^{\infty}$ defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ having a common distribution $\mathbb{P}$ satisfying (1) for some parameter $\theta_0 := (\zeta_0, \alpha_0, \beta_0) \in \Theta := [c_1, c_2] \times \mathbb{R}^2$. Suppose that Z has a uniformly bounded, strictly positive density f (with respect to the Lebesgue measure) on $[c_1, c_2]$ such that $\inf_{|z-\zeta_0|\le\eta} f(z) > \kappa > 0$ for some $\eta > 0$ and that $\mathbb{P}(Z < c_1) \wedge \mathbb{P}(Z > c_2) > 0$. For $\theta = (\zeta, \alpha, \beta) \in \Theta$, $x = (y, z) \in \mathbb{R}^2$ write

$$m_\theta(x) := -\left(y - \alpha\mathbf{1}_{z\le\zeta} - \beta\mathbf{1}_{z>\zeta}\right)^2,$$

and $\mathbb{P}_n$ for the empirical measure defined by $X_1, \ldots, X_n$. Note that $M_n(\theta) := -\mathbb{P}_n[m_\theta]$ and recall the definition of $\hat{\theta}_n$.

The asymptotic properties of this estimator are well-known and have been deduced by several authors. They are available, for instance, in Kosorok (2008) or Seijo and Sen (2010). It follows from Proposition 3.2 in Seijo and Sen (2010) that $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_{\mathbf{P}}(1)$, $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_{\mathbf{P}}(1)$ and $n(\hat{\zeta}_n - \zeta_0) = O_{\mathbf{P}}(1)$.

For $h = (h_1, h_2, h_3) \in \mathbb{R}^3$, let $\vartheta_{n,h} := \theta_0 + \left(\frac{h_1}{n}, \frac{h_2}{\sqrt{n}}, \frac{h_3}{\sqrt{n}}\right)$ and

$$\hat{E}_n(h) := n\mathbb{P}_n\left[m_{\vartheta_{n,h}} - m_{\theta_0}\right].$$

A consequence of the rate of convergence result in Seijo and Sen (2010) is that with probability tending to one, we have

$$\hat{h}_n := \operatorname*{sargmax}_{h\in\mathbb{R}^3} \hat{E}_n(h) = \left(n(\hat{\zeta}_n - \zeta_0), \sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\right).$$

Write $\hat{J}_n$ for the pure jump process associated with $\hat{E}_n$. It is shown in Lemma 3.3 of Seijo and Sen (2011) that

(a) $(\hat{E}_n, \hat{J}_n) \rightsquigarrow (E^*, J^*)$ in $\mathcal{D}_K \times \mathcal{S}_I$,

on every compact rectangle $K = I \times A \times B \subset \mathbb{R}^3$ for some process $E^* \in \mathcal{D}_{\mathbb{R}^3}$ with an associated pure jump process J^*. Then, an application of Theorem 3.2 shows that

$$\hat{h}_n = \left(n(\hat{\zeta}_n - \zeta_0), \sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0)\right) \rightsquigarrow \operatorname*{sargmax}_{h\in\mathbb{R}^3}\{E^*(h)\}.$$

It must be noted that the results in Seijo and Sen (2010) are stated in terms of a triangular array of random vectors that satisfy some regularity conditions. Even in such generality, Proposition 3.3 in Seijo and Sen (2010) can be derived from Theorem 3.2.

We would like to point out that the derivation of the asymptotic distribution of this estimator can also be found in Kosorok (2008). The arguments there can be modified to obtain the result from an application of Theorem 3.2.

5.2. Estimation in a Cox regression model with a change-point in time

Define $\Theta := (0,1) \times \mathbb{R}^{p+2q}$ for given $p, q \in \mathbb{N}$. For $\theta = (\tau, \xi) = (\tau, \alpha, \beta, \gamma) \in \Theta = (0,1) \times \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^q$ consider a survival time T^0, a censoring time C and covariate cáglád (left-continuous with right-hand side limits) $\mathbb{R}^{p+q}$-valued process $Z = (Z_1, Z_2)$ where the sample paths of Z_1 and Z_2 live in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Assume that C and Z have laws G and H, respectively. Note that G is a distribution on the nonnegative real line and H a probability measure on the space of left continuous processes with right-hand side limits. In our Cox model with a change-point in time we make the additional assumption that, conditionally on Z, the hazard function of the survival time is given by:

$$\begin{aligned}\lambda(t|Z) &:= \lim_{\Delta t\downarrow 0} \frac{\mathbf{P}\left(t \le T^0 < t + \Delta t | T^0 \ge t;\ Z(s),\ 0 \le s \le t\right)}{\Delta t} \\ &= \lambda(t) e^{\alpha\cdot Z_1(t) + (\beta + \gamma \mathbf{1}_{t>\tau})\cdot Z_2(t)}\end{aligned}$$

where λ is the *baseline hazard function* and $\cdot$ denotes the standard inner product on Euclidian spaces. We write $\mathbb{P}_{\theta,\lambda,G,H}$ for the law of (T^0, C, Z). We would like to point out that we assume that G and the finite dimensional distributions of Z are all *continuous*.

Suppose that there is a random sample

$$(T_1^0, C_1, Z_{1,1}, Z_{2,1}), \ldots, (T_n^0, C_n, Z_{1,n}, Z_{2,n}) \overset{i.i.d.}{\sim} \mathbb{P}_{\theta_0,\lambda_0,G_0,H_0}$$

from which we are only able to observe $Z_{1,j}$, $Z_{2,j}$, $\Delta_j := \mathbf{1}_{T_j^0 \leq C_j}$ and $T_j := T_j^0 \wedge C_j$ for $j = 1, \ldots, n$. The goal is to estimate the change-point $\tau_0 \in (0,1)$ given these observations.

A standard method of estimation in this setting is via Cox's partial likelihood, in which case the likelihood and log-likelihood functions are given by

$$\begin{aligned} L_n(\tau,\alpha,\beta,\gamma) &:= \prod_{\substack{1\leq k\leq n \\ T_k^0 \leq C_k}} \frac{e^{\alpha\cdot Z_{1,k}(T_k^0) + (\beta+\gamma\mathbf{1}_{T_k^0>\tau})\cdot Z_{2,k}(T_k^0)}}{\sum_{\{1\leq j\leq n:\ T_k^0 \leq T_j^0 \wedge C_j\}} e^{\alpha\cdot Z_{1,j}(T_k^0) + (\beta+\gamma\mathbf{1}_{T_k^0>\tau})\cdot Z_{2,j}(T_k^0)}}, \\ l_n(\theta) &:= \log\left(L_n(\tau,\xi)\right) = \log\left(L_n(\tau,\alpha,\beta,\gamma)\right). \end{aligned}$$

In this case, the maximum partial likelihood estimator of the change-point and the covariate multipliers is given by

$$\hat{\theta}_n = (\hat{\tau}_n, \hat{\xi}_n) = (\hat{\tau}_n, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n) := \underset{\theta\in\Theta}{\mathrm{sargmax}}\{l_n(\theta)\}.$$

Pons (2002) derived the asymptotics for this estimator. For $u = (u^1, u^2, \ldots, u^{1+p+2q}) = (u^1, v) \in \mathbb{R}^{1+p+2q}$ define $\theta_{n,u} = \left(\tau_0 + \frac{u^1}{n}, \xi_0 + \frac{v}{\sqrt{n}}\right)$. Then, under some regularity conditions, Theorem 2 in Pons (2002) shows that

$$\left(n(\hat{\tau}_n - \tau_0), \sqrt{n}(\hat{\xi}_n - \xi_0)\right) = \underset{u\in\mathbb{R}^{1+p+2q}:\ \theta_{n,u}\in\Theta}{\mathrm{sargmax}} \{l_n(\theta_{n,u}) - l_n(\theta_0)\} = O_{\mathbf{P}}(1).$$

It can also be inferred from Proposition 3 and Theorem 3 of the same paper that $\Psi_n := l_n(\theta_{n,u}) - l_n(\theta_0) \rightsquigarrow \Psi$ on $\mathcal{D}_K$ for every compact rectangle $K \subset \mathbb{R}^{1+p+2q}$, where Ψ is a stochastic process of the form

$$\Psi(u^1, v) = Q(u^1) + v\cdot\tilde{W} - \frac{1}{2} v\tilde{I}\cdot v, \tag{6}$$

with Q being a two-sided, compound Poisson process, $\tilde{W}$ a Gaussian random variable independent of Q and $\tilde{I}$ some positive definite matrix on $\mathbb{R}^{(p+2q)\times(p+2q)}$. For a detailed description of Q, $\tilde{W}$ and $\tilde{I}$ we refer the reader to Section 4 of Pons (2002).

If one defines Γ_n and Γ to be the pure jump processes associated with Ψ_n and Ψ, respectively, it can be shown, using similar techniques as in the proof of Theorem 3 of Pons (2002), that $(\Psi_n, \Gamma_n) \rightsquigarrow (\Psi, \Gamma)$ on $\mathcal{D}_B \times \mathcal{D}_{B_1}$ for every compact subinterval $B_1 \subset \mathbb{R}$ and compact rectangle $B := B_1 \times B_2 \subset \mathbb{R}^{1+p+2q}$. Hence, Theorem 3.2 can be applied in this situation to conclude that

$$\Big(n(\hat{\tau}_n - \tau_0), \sqrt{n}(\hat{\xi}_n - \xi_0)\Big) \rightsquigarrow \operatorname*{sargmax}_{u\in\mathbb{R}^{1+p+2q}} \{\Psi(u)\}.$$

It must be noted that the proof of Theorem 4 in Pons (2002) makes no mention of the pure jump processes Γ_n and Γ. On the second sentence of this proof, the author claims that the asymptotic distribution follows just from the weak convergence of the processes Ψ_n. As we saw in Section 4 this fact alone is not enough to conclude the weak convergence of the smallest maximizers. Thus, the argument given in this section completes the mentioned proof in Pons (2002).

5.3. Estimating a change-point in a Cox regression model according to a threshold in a covariate

We will now discuss another application from survival analysis. Consider again a Cox regression model but now with a covariate process of the form $Z = (Z_1, Z_2, Z_3)$ where Z_1 and Z_2 are as in Section 5.2 and Z_3 is a continuous random variable in $\mathbb{R}$. We will denote the survival and censoring times as in Section 5.2. We are now concerned with a hazard function of the form

$$\lambda(t|Z) = \lambda(t)e^{\alpha\cdot Z_1(t)+\beta\cdot Z_2(t)\mathbf{1}_{Z_3\le\zeta}+\gamma\cdot Z_2(t)\mathbf{1}_{Z_3>\zeta}},$$

for $\alpha \in \mathbb{R}^q$, $\beta,\gamma \in \mathbb{R}^q$ and some $\zeta \in I$ where I is a closed interval entirely contained in the interior of the support of Z_3. We now consider the parameter space $\Theta := I \times \mathbb{R}^{p+2q}$ and we write $\theta = (\zeta,\xi) := (\zeta,\alpha,\beta,\gamma) \in \Theta$. The partial likelihood and log-likelihood functions are now given by

$$\begin{aligned} &L_n(\zeta,\alpha,\beta,\gamma)\\ &\quad:= \prod_{\substack{1\le k\le n\\ T_k^0\le C_k}} \frac{e^{\alpha\cdot Z_{1,k}(T_k^0)+\beta\cdot Z_{2,k}(T_k^0)\mathbf{1}_{Z_{3,k}\le\zeta}+\gamma\cdot Z_{2,k}(T_k^0)\mathbf{1}_{Z_{3,k}>\zeta}}}{\sum_{\{1\le j\le n:\ T_k^0\le T_j^0\wedge C_j\}} e^{\alpha\cdot Z_{1,j}(T_k^0)+\beta\cdot Z_{2,j}(T_k^0)\mathbf{1}_{Z_{3,j}\le\zeta}+\gamma\cdot Z_{2,j}(T_k^0)\mathbf{1}_{Z_{3,j}>\zeta}}},\\ &l_n(\theta) := \log\left(L_n(\zeta,\xi)\right) = \log\left(L_n(\zeta,\alpha,\beta,\gamma)\right). \end{aligned}$$

As before, we assume that the observations come from a model with some specific value $\theta_0 \in \Theta$. Following the notation of Section 5.2, for $u = (u^1, u^2, \ldots, u^{1+p+2q}) = (u^1, v) \in \mathbb{R}^{1+p+2q}$ define $\theta_{n,u} = \big(\zeta_0 + \frac{u^1}{n}, \xi_0 + \frac{v}{\sqrt{n}}\big)$. Then, under some regularity conditions, Theorem 2 in Pons (2003) shows that

$$\Big(n(\hat{\zeta}_n - \zeta_0), \sqrt{n}(\hat{\xi}_n - \xi_0)\Big) = \operatorname*{sargmax}_{u\in\mathbb{R}^{1+p+2q}:\ \theta_{n,u}\in\Theta} \{l_n(\theta_{n,u}) - l_n(\theta_0)\} = O_{\mathbf{P}}(1).$$

Lemma 5 and Theorem 3 in Pons (2003) show that $\Psi_n := l_n(\theta_{n,u}) - l_n(\theta_0) \rightsquigarrow \Psi$ on $\mathcal{D}_K$ for every compact rectangle $K \subset \mathbb{R}^{1+p+2q}$, where Ψ is another stochastic process of the form (6) but with different two-sided, compound Poisson process Q, Gaussian random variable $\tilde{W}$ and positive definite matrix $\tilde{I}$. The details can be found in Section 4 of Pons (2003).

Letting Γ_n and Γ to be the pure jump processes associated with Ψ_n and Ψ, respectively, it can be shown that $(\Psi_n, \Gamma_n) \rightsquigarrow (\Psi, \Gamma)$ on $\mathcal{D}_B \times \mathcal{D}_{B_1}$ for every compact subinterval $B_1 \subset \mathbb{R}$ and compact rectangle $B := B_1 \times B_2 \subset \mathbb{R}^{1+p+2q}$. Hence, another application of Theorem 3.2 shows that

$$\left(n(\hat{\tau}_n - \tau_0), \sqrt{n}(\hat{\xi}_n - \xi_0)\right) \rightsquigarrow \operatorname*{sargmax}_{u \in \mathbb{R}^{1+p+2q}} \{\Psi(u)\}.$$

As in Pons (2002), the argument to derive the asymptotic distribution given in the proof of Theorem 5 lacks a proper discussion of the convergence of the associated pure jump processes. Therefore, the analysis just given can be seen as a complement to the proof of Theorem 5 in Pons (2003).

More general models involving right censoring for survival times and a change-point based on a threshold in a covariate can be found in Kosorok and Song (2007). There, the change-point estimator also achieves a n^{-1} rate of convergence. The asymptotic distribution of this estimator also corresponds to the smallest maximizer of a two-sided, compound Poisson process and can be deduced from an application of Theorem 3.2. We would like to point out that the above authors omit a discussion about the associated pure jump processes. They claim the desired stochastic convergence follows from an application of Theorem 3.2.2 in Van der Vaart and Wellner (1996) (see the last paragraph of the proof of Theorem 5 in page 985 of Kosorok and Song (2007)), but this theorem cannot be applied as the maximizer of a compound Poisson process is not unique. Thus, a proper application of Theorem 3.2 would complete the argument in Kosorok and Song (2007).

References

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley, New York, NY, USA. MR0233396

Ethier, S. and Kurtz, T. (2005). *Markov Processes, Characterization and Convergence*. John Wiley & Sons, New York, NY, USA. MR0838085

Ferger, D. (2004). A continuous mapping theorem for the argmax-functional in the non-unique case. *Statist. Neerlandica*, 48:83–96. MR2042258

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York, NY, USA. MR2724368

Kosorok, M. and Song, R. (2007). Inference under right censoring for transformation models with a change-point based on a covariate threshold. *Ann. Statist.*, 35:957–989. MR2341694

Lan, Y., Banerjee, M., and Michailidis, G. (2009). Change-point estimation under adaptive sampling. *Ann. Statist.*, 37:1752–1791. MR2533471

Neuhaus, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.*, 42:1285–1295. MR0293706

Pons, O. (2002). Estimation in a cox regression model with a change-point at an unknown time. *Statistics*, 36:101–124. MR1910255

Pons, O. (2003). Estimation in a cox regression model with a change-point according to a threshold in a covariate. *Ann. Statist.*, 31:442–463. MR1983537

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhya*, 31:23–36. MR0267677

Seijo, E. and Sen, B. (2010). Change-point in stochastic design regression and the bootstrap. To appear in *The Annals of Statistics*.

Seijo, E. and Sen, B. (2011). Supplement to “change-point in stochastic design regression and the bootstrap”.

Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York, NY, USA. MR1385671