PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 105, Number 3, March 1989

CONTINUITY PROPERTIES OF OPTIMAL STOPPING VALUE

JOHN ELTON

(Communicated by William D. Sudderth)

ABSTRACT. The optimal stopping value of a sequence (finite or infinite) of integrable random variables is lower semicontinuous for the topology of convergence in distribution, when restricted to a collection with uniformly integrable negative parts. It is continuous for finite sequences which are adapted by a continuous invertible "triangular" function to independent sequences, such as partial averages; this is our main result. The proof depends on conditional weak convergence, uniform on compact sets, for such processes. A topological result on the inverses of triangular functions on iteratively connected domains may be of independent interest (§3).

1. LOWER SEMICONTINUITY OF VALUE

The optimal stopping *value* of a sequence (finite or infinite) $X_1, X_2, \ldots$ of integrable random variables is defined by

$$V(X_1, X_2, \ldots) = \sup_{\tau} \mathsf{E}(X_\tau),$$

where the supremum is taken over nonanticipating a.s. finite stop rules τ. For finite sequences, the supremum is attained; see [CRS].

Collections of random sequences will be given the usual topology of convergence in distribution, i.e., weak convergence of the corresponding probability distributions; see [B]. This is a metrizable topology, where random sequences with the same distribution are identified. Finite sequences of random variables will be referred to as random vectors. Convergence in distribution for random N-vectors can be characterized as follows:

$$\mathbf{X}^n \xrightarrow{\mathscr{D}} \mathbf{Z} \quad \text{iff } \mathsf{E}f(\mathbf{X}^n) \to \mathsf{E}f(\mathbf{X})$$

for all $f \in C(\mathbf{R}^N)$, the bounded continuous functions on $\mathbf{R}^N$.

Received by the editors September 22, 1987, and, in revised form, July 1, 1988; presented to the Society, January 8, 1988, at the 839th meeting in Atlanta, Georgia.

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60G40; Secondary 26B99, 54C30, 54D05, 62L15, 90C99.

Key words and phrases. Optimal stopping value, weak convergence, conditional distribution, triangular function, adapted, iteratively connected.

Partially supported by Darpa Applied Mathematical Sciences Program.

A real function F on a metric space is *lower semicontinuous* if $x_n \to x \Rightarrow F(x) \le \underline{\lim} F(x_n)$. A useful property is that if the space is compact, such an F attains its minimum.

Theorem 1.1. *Let $\mathscr{C}$ be a collection of integrable random N-vectors whose components have negative parts which are uniformly integrable (u.i.). Then V is lower semicontinuous on $\mathscr{C}$.*

Passing to limits, one obtains

Corollary 1.2. *If $\mathscr{C}$ is a collection of sequences of integrable random variables such that for each N, the initial segments of length N have u.i. negative parts, then V is lower semicontinuous on $\mathscr{C}$.*

Proof of Theorem 1.1. We need to show that if $\mathbf{X}^n \in \mathscr{C}$ and $\mathbf{X}^n \xrightarrow{\mathscr{D}} \mathbf{X} \in \mathscr{C}$, then $V(\mathbf{X}) \le \underline{\lim} V(\mathbf{X}^n)$.

For some stop rule τ

$$V(\mathbf{X}) = \mathbf{E}(X_\tau) = \sum_{k=1}^{N} \int_{\tau=k} X_k \, dP.$$

For a vector $\mathbf{x} = (x_1, \dots, x_N) \in \mathbf{R}^N$, we shall denote the projection of $\mathbf{x}$ onto the first k coordinates by

$$\mathbf{x}_k = (x_1, \dots, x_k).$$

We hope this distinction between $\mathbf{x}_k$ and x_k causes no confusion.

Since $\{\tau = k\} \in \sigma(\mathbf{X}_k)$, the σ-field generated by $X_1, \dots, X_k$, we have

$$\{\tau = k\} = \{\mathbf{X}_k \in A_k\}$$

for some $A_k \in \mathscr{B}_k$, the Borel sets in $\mathbf{R}^k$, $k = 1, \dots, N$. Furthermore, since the events $\{\tau = k\}$ are a partition, we may assume WLOG that $A_k \subset \bigcap_{j=1}^{k-1} \widetilde{A}_j \times \mathbf{R}^{k-j}$ for $1 < k < N$, and $A_N = \bigcap_{j=1}^{N-1} \widetilde{A}_j \times \mathbf{R}^{N-j}$ ($\sim$ denotes complement).

Let $\varepsilon > 0$, and choose $\delta > 0$ such that $\int_S |X_k| \, dP < \varepsilon$ whenever $P(S) < \delta$, $k = 1, \dots, N$. By a standard approximation argument, for each k there exists $B_k \in \mathscr{B}_k$ such that $P(\mathbf{X}_k \in A_k \Delta B_k) < \delta/N$ and $P(\mathbf{X}_k \in \partial B_k) = 0$.

Define $C_1 = B_1$, and $C_k = B_k \cap (\bigcap_{j=1}^{k-1} \widetilde{B}_j \times \mathbf{R}^{k-j})$ for $k = 2, \dots, N-1$, and $C_N = \bigcap_{j=1}^{N-1} \widetilde{B}_j \times \mathbf{R}^{N-j}$. One can easily show $P(\mathbf{X}_k \in A_k \Delta C_k) \le P(\mathbf{X}_k \in A_k \Delta B_k) + \sum_{j=1}^{k-1} P(\mathbf{X}_j \in A_j \Delta B_j) \le \delta$, so

$$V(\mathbf{X}) \le \sum_{k=1}^{N} \int_{\mathbf{X}_k \in C_k} X_k \, dP + N\varepsilon.$$

Next, choose λ so large that

$$\int_{X_k^n \le -\lambda} |X_k^n| \, dP < \varepsilon \quad \text{for all } n, k \text{ (by the u.i. hypothesis)},$$

and also $\int_{X_k \geq \lambda} X_k\, dP < \varepsilon$, $k = 1, \ldots, N$.

The set of discontinuities of the function $I_{\mathscr{C}_k}$ is precisely ∂C_k, so by Theorem 5.2 (iii), p. 31 of [B], since $P(\mathbf{X}_k \in \partial C_k) = 0$,

$$\int_{\mathbf{X}_k^n \in C_k} f(X_k^n)\, dP \to \int_{\mathbf{X}_k \in C_k} f(X_k)\, dP,$$

where f is the bounded and continuous function defined by $f(t) = t$ if $|t| \leq \lambda$, $f(t) = \lambda \operatorname{sgn}(t)$ if $|t| > \lambda$. Thus

$$\underline{\lim} \int_{\mathbf{X}_k^n \in C_k} X_k^n\, dP \geq \int_{\mathbf{X}_k \in C_k} X_k\, dP - 2\varepsilon.$$

Finally, from the definition of the C_k's, for each n the events

$$\{\mathbf{X}_k^n \in C_k\}, \qquad k = 1, \ldots, N$$

are a partition, so the stop rules

$$\tau_n = k \Leftrightarrow \mathbf{X}_k^n \in C_k$$

show that for each n,

$$V(\mathbf{X}^n) \geq \sum_{k=1}^{N} \int_{\mathbf{X}_k^n \in C_k} X_k^n\, dP.$$

Thus $\underline{\lim}\, V(\mathbf{X}^n) \geq V(\mathbf{X}) - 3N\varepsilon$. □

Example 1.3. This example shows that V is *not in general continuous*, even when restricted to uniformly bounded exchangeable pairs.

Let (X_1^n, X_2^n) have range

$$\{(0, \tfrac{1}{2}), (\tfrac{1}{2}, 0), (\tfrac{1}{2} + \tfrac{1}{n}, 1), (1, \tfrac{1}{2} + \tfrac{1}{n})\},$$

each point taken with probability $\frac{1}{4}$. The obvious stop rule gives

$$V(X_1^n, X_2^n) = \tfrac{1}{4}(\tfrac{1}{2} + \tfrac{1}{2} + 1 + 1) = \tfrac{3}{4}.$$

But $(X_1^n, X_2^n) \xrightarrow{\mathscr{D}} (X_1, X_2)$ which has range $\{(0, \frac{1}{2}), (\frac{1}{2}, 0), (\frac{1}{2}, 1), (1, \frac{1}{2})\}$, each point having probability $\frac{1}{4}$. But

$$V(X_1, X_2) = \tfrac{1}{4} \times \tfrac{1}{2} + \tfrac{1}{2} \times \tfrac{1}{2} + \tfrac{1}{4} \times 1 = \tfrac{5}{8},$$

so $V(X_1^n, X_2^n) \nrightarrow V(X_1, X_2)$.

Example 1.4. This example shows that even in the case of i.i.d. pairs of random variables, the conclusion in Theorem 1.1 can fail without the u.i. hypothesis.

Let

$$X_k^n = \begin{cases} 0 & \text{with probability } 1 - 1/\sqrt{n}, \\ -n & \text{with probability } 1/\sqrt{n}, \end{cases}$$

$k = 1, 2$, and X_1^n, X_2^n independent. Obviously $(X_1^n, X_2^n) \xrightarrow{\mathscr{D}} (0, 0)$, which has value 0. But $\mathsf{E}(X_2^n \mid X_1^n) = -\sqrt{n}$, so $V(X_1^n, X_2^n) = 0(1 - 1/\sqrt{n}) + (1/\sqrt{n})(-\sqrt{n}) = -1$. So V is not lower semicontinuous on the collection $\{\mathbf{X}^n\}$.

Application to prophet inequalities. Let $X_1, \ldots, X_N$ be integrable random variables. Inequalities which compare $\mathsf{E}\max(\mathbf{X}) = \mathsf{E}(\max\{X_1, \ldots, X_N\})$ to $V(\mathbf{X})$ for a class of random N-vectors (or infinite sequences) have been called *prophet inequalities* because the first is the expected return of a prophet (using complete foresight) and the second is the return of a gambler using nonanticipating stop rules; see e.g., [KS_1, KS_2, HK_1, HK_2, HK_3, Ker, K, CK]. One type of inequality is an upper bound on the difference between the prophet's expected return and the gambler's optimal expected return for random vectors in the class, as e.g., in [HK_1, HK_2]. More generally, one may consider, for each y, the greatest lower bound v on $V(\mathbf{X})$ for those $\mathbf{X}$ in the class $\mathscr{C}$ satisfying $\mathsf{E}\max(\mathbf{X}) = y$. If the "prophet region" $\mathscr{R}(\mathscr{C}) = \{(V(\mathbf{X})), \mathsf{E}\max(\mathbf{X})) : \mathbf{X} \in \mathscr{C}\} \subset \mathbf{R}^2$ is convex, the graph of v as a function of y is the upper boundary curve of $\mathscr{R}(\mathscr{C})$. See for example [HK_3, Ker, EK].

We shall show that in many cases these bounds are attained.

Corollary 1.5. *If $\mathscr{C}$ is a tight, closed collection of integrable random $\mathbb{N}$-vectors whose negative parts are u.i., then*

$$\sup\{\mathsf{E}\max(\mathbf{X}) - V(\mathbf{X}) : \mathbf{X} \in \mathscr{C}\}$$

is attained on $\mathscr{C}$, and so is

$$\inf\{V(\mathbf{X}) : \mathbf{X} \in \mathscr{C}, \mathsf{E}\max(\mathbf{X}) = y\},$$

for each y in $\mathsf{E}\max(\mathscr{C})$.

Proof. $\mathscr{C}$ is compact by Prohorov's Theorem [B]. Also $\{\mathbf{X} \in \mathscr{C} : \mathsf{E}\max(\mathbf{X}) = y\}$ is closed, hence compact, since $\mathsf{E}\max$ is clearly continuous. Now $\mathsf{E}\max - V$ is upper semicontinuous and V is lower semicontinuous by Theorem 1.1, so the results follow. □

Example 1.6. Let $\mathscr{C} = \{\mathbf{X} : \mathbf{X}$ is a random N-vector with exchangeable components with values in $[0,1]\}$. It was shown in [EK] that $\mathscr{R}(\mathscr{C})$ is a convex set. $\mathscr{C}$ is obviously tight, and a limit in distribution of a sequence of exchangeable N-vectors is exchangeable, so Corollary 1.5 applies. Thus $\mathscr{R}(\mathscr{C})$ is a closed set as well. The form of the $\mathbf{X}$ attaining the upper boundary has not yet been found except for $N = 2$.

2. Continuity of value

We present now a common situation in which V is actually continuous. A simple example is partial averages of finite sequences of independent, u.i. random variables; more generally we consider sequences continuously adapted to independent sequences.

Call a function $\mathbf{a} : \mathbf{R}^N \to \mathbf{R}^N$ *triangular* if for $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbf{R}^N$,

$$\mathbf{a}(\mathbf{x}) = (a_1(x_1), \ldots, a_k(\mathbf{x}_k), \ldots, a_N(\mathbf{x})),$$

where $\mathbf{x}_k$ denotes $(x_1, \ldots, x_k)$ as in the proof of Theorem 1.1. That is, the kth output depends only on the first k inputs. Thus a finite sequence $(Y_1, \ldots, Y_N)$

of random variables is adapted to $(X_1, \dots, X_N)$ (in the sense that Y_k is measurable $\sigma(X_1, \dots, X_k)$ for all k) iff $\mathbf{Y} = \mathbf{a}(\mathbf{X})$ for some Borel-measurable triangular function $\mathbf{a}$.

Call a function $\mathbf{a}$ *linearly bounded* if there exists c such that $|\mathbf{a}(\mathbf{x})| \leq c(|\mathbf{x}| \vee 1)$ for all $\mathbf{x}$ in the domain of $\mathbf{a}$ (we use the norm $|x| = \max_{1 \leq i \leq N} |x_i|$ on $\mathbf{R}^N$). Note that if $\mathbf{a}$ is one-to-one, then $\mathbf{a}^{-1}$ is linearly bounded iff there exists $d > 0$ such that $|\mathbf{a}(\mathbf{x})| \vee 1 \geq d|\mathbf{x}|$ for all $\mathbf{x}$ in domain of $\mathbf{a}$.

The following result requires a one-to-one, continuous triangular function whose inverse is also triangular. In the appendix we show that this is often redundant, which is perhaps of independent interest.

Theorem 2.1. *Let $\mathscr{C}$ be a u.i. collection of random N-vectors with independent components. Let $\mathbf{a}$ be a one-to-one function on range$(\mathscr{C}) \subset \mathbf{R}^N$ such that both $\mathbf{a}$ and its inverse are continuous, linearly bounded, and triangular. Let $\mathscr{A} = \{\mathbf{a}(\mathbf{X}) : \mathbf{X} \in C\}$. Then V is continuous on $\mathscr{A}$.*

Remarks. (1) Examples of processes continuously adapted to independent sequences abound. The canonical examples are independent sequences themselves and partial sums and averages of independent random variables. Other examples are the "burglar problem" [CRS, p. 44], and extreme order statistics with cost of sampling [P].

(2) The continuity assumption on $\mathbf{a}$ *is* needed. It is easy to give an example, with uniformly bounded random variables and $N = 2$, for which all the hypotheses of Theorem 2.1 except the continuity are satisfied, and the conclusion fails.

(3) It is not obvious to the author what might be reasonable necessary conditions for continuity of V. The proof of Theorem 2.1 uses conditional weak convergence (Lemma 2.5), which relies on independence, but only in a special way, so perhaps nothing as strong as Lemma 2.5 is really needed.

Proof. Choose $c \geq 1$ to work for both $\mathbf{a}$ and its inverse $\mathbf{b}$, in the definition of linearly bounded.

Let $\mathbf{X}^n = (X_1^n, \dots, X_N^n) \in \mathscr{C}$, and let $\mathbf{A}^n = (A_1^n, \dots, A_N^n) = \mathbf{a}(\mathbf{X}^n)$. Suppose $\mathbf{A}^n \xrightarrow{\mathscr{D}} \mathbf{A} \in \mathscr{R}$. We need to show that $V(\mathbf{A}^n) \to V(\mathbf{A})$. Since $\mathbf{a}$ has a continuous inverse, $\mathbf{A}^n \xrightarrow{\mathscr{D}} \mathbf{A}$ iff $\mathbf{X}^n \xrightarrow{\mathscr{D}} \mathbf{X}$, where $\mathbf{A} \doteq \mathbf{a}(\mathbf{X})$.

The proof will involve weak convergence of conditional distributions, for which the independence is crucial.

Let $\mu_k^n(y_{k-1}; B) = P(A_k^n \in B \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1})$, $\mathbf{y}_{k-1} \in \mathbf{R}^{k-1}$, for $k = 2, \dots, N$; and $\mu_1^n(B) = P(A_1^n \in B)$. Here and in the rest of the proof, we understand such statements involving n to include $n = \infty$; i.e., $\mu_k(\mathbf{y}_{k-1}; B) = P(A_k \in B | \mathbf{A}_{k-1} = \mathbf{y}_{k-1})$ also. It will be important to choose these conditional distributions in a canonical way which is possible because of the invertible relationship between $\mathbf{A}^n$ and $\mathbf{X}^n$. Specifically, we will always take

$$\mu_k^n(\mathbf{y}_{k-1}; B) = P(A_k^n \in B \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1})$$

to be

$$P(a_k(\mathbf{X}_k^n) \in B \mid \mathbf{X}_{k-1}^n = \mathbf{b}_{k-1}(\mathbf{y}_{k-1})) = P(a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}), X_k^n) \in B),$$

by independence of X_k^n from $\mathbf{X}_{k-1}^n$.

Define functions ψ_k^n and ϕ_k^n on $\mathbf{R}^k$ using backward induction, by

$$\psi_{k-1}^n(\mathbf{y}_{k-1}) = \int \phi_k^n(\mathbf{y}_k)\mu_k^n(\mathbf{y}_{k-1}; dy_k),$$

$k = 1, \ldots, N$ (ψ_0^n is just a number), where

$$\phi_k^n(\mathbf{y}_k) = y_k \vee \psi_k^n(\mathbf{y}_k).$$

To start the induction, define $\psi_N^n(\mathbf{y}_N) = y_N = \phi_N^n(\mathbf{y})$.

By the *backward induction principle* of optimal stopping [CRS, Chap. 3 and 4], these are just the *conditional values*:

$$\psi_k^n(\mathbf{y}_k) = V(A_{k+1}^n, \ldots, A_N^n \mid \mathbf{A}_k^n = \mathbf{y}_k),$$
$$\psi_0^n = V(\mathbf{A}^n).$$

We shall show $\psi_0^n \to \psi_0$ by showing for any $k = 1, \ldots, N$ that $\psi_k^n \to \psi_k$ uniformly on compact sets $\Rightarrow \psi_{k-1}^n \to \psi_{k-1}$ uniformly on compact sets. Since $\psi_N^n = \psi_N$, this will prove the result by backward induction.

We break the proof up into several easy lemmas.

Lemma 2.2. *If K is a compact subset of $\mathbf{R}^{k-1}$,*

$$\lim_{\lambda\to\infty} \sup_{n,\mathbf{y}_{k-1}\in K} \int_{|y_k|>\lambda} |\mathbf{y}_k|\mu_k(\mathbf{y}_{k-1}; dy_k) = 0.$$

Proof. $K \subset [-M, M]^{k-1}$ for some $M \geq 1$. The integral above is

$$\mathbf{E}(|\mathbf{A}_k^n| I_{|A_k^n|>\lambda} \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1}).$$

For $\mathbf{y}_{k-1} \in K$,

$$|\mathbf{A}_k^n| = |\mathbf{a}_k(\mathbf{X}_k^n)| = |\mathbf{a}_k(\mathbf{b}_{k-1}(\mathbf{A}_{k-1}^n), X_k^n)|$$
$$\leq c(cM \wedge |X_k^n|).$$

Assume $\lambda > c^2 M$; then since $|A_k^n| \leq |\mathbf{A}_k^n|$,

$$|A_k^n| > \lambda \Rightarrow |X_k^n| > \lambda/c.$$

So for all $\mathbf{y}_{k-1} \in K$,

$$\mathbf{E}(|\mathbf{A}_k^n| I_{|A_k^n|>\lambda} \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1}) \leq \mathbf{E}(c|X_k^n| I_{|X_k^n|>\lambda/c} \mid \mathbf{X}_{k-1}^n = \mathbf{b}_{k-1}(\mathbf{y}_{k-1}))$$
$$= c\mathbf{E}(|X_k^n| I_{|X_k^n|} > \lambda/c),$$

by independence and canonical choice of conditional distribution of A_k^n stated above. This converges to 0 as $\lambda \to \infty$ uniformly in n, by the u.i. assumption. □

Lemma 2.3. *There exists* $\alpha < \infty$ *such that*

$$|\phi_k^n(\mathbf{y}_k)| \le \alpha|\mathbf{y}_k| + \alpha \quad \textit{for all } k, n, \mathbf{y}_k \in \mathbf{R}^k.$$

Proof.

$$\begin{aligned}|\psi_k^n(\mathbf{y}_k)| &= V(A_{k+1}^n, \dots, A_N^n \mid \mathbf{A}_k^n = \mathbf{y}_k)\\ &\le \mathbf{E}(|A_{k+1}^n| \vee \cdots \vee |A_N^n| \mid \mathbf{A}_k^n = \mathbf{y}_k)\\ &\le \mathbf{E}(|\mathbf{A}^n| \mid \mathbf{A}_k^n = \mathbf{y}_k)\\ &= \mathbf{E}(|\mathbf{a}(\mathbf{b}_k(\mathbf{A}_k^n), X_{k+1}^n, \dots, X_N^n)| \mid \mathbf{A}_k^n = \mathbf{y}_k)\\ &= \mathbf{E}(|\mathbf{a}(\mathbf{b}_k(\mathbf{y}_k), X_{k+1}^n, \dots, X_N^n)|) \text{ by independence}\\ &\le \mathbf{E}(c(c(|\mathbf{y}_k| \vee 1) \vee |X_{k+1}^n| \vee \cdots \vee |X_N^n|))\\ &\le c^2(|\mathbf{y}_k| \vee 1) + cN\beta,\end{aligned}$$

where $\beta \ge E|X_j^n|$ for all j, n (possible by the u.i. assumption). This obviously implies the result. □

Lemma 2.4. ψ_k^n *is continuous, for all* k *and* n.

Proof. By backward induction. ψ_N^n is obviously continuous. Assume ψ_k^n is continuous, so that ϕ_k^n is also. Now

$$\begin{aligned}\psi_{k-1}^n(\mathbf{y}_{k-1}) &= \mathbf{E}(\phi_k^n(\mathbf{A}_{k-1}^n, a_k(\mathbf{b}_{k-1}(\mathbf{A}_{k-1}^n), X_k^n)) \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1})\\ &= \mathbf{E}(\phi_k^n(\mathbf{y}_{k-1}, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}), X_k^n)))\end{aligned}$$

by independence. But the integrand is

$$\le \alpha(|\mathbf{y}_{k-1}| + c^2(|\mathbf{y}_{k-1}| \vee 1) + c|X_k^n|) + \alpha,$$

by Lemma 2.3. Now the result clearly follows by Lebesgue's dominated convergence theorem and the continuity of ϕ_k^n, $\mathbf{a}$, and $\mathbf{b}$. □

The following lemma gives conditional weak convergence, uniform on compact sets.

Lemma 2.5. *Let* K *be a compact subset of* $\mathbf{R}^{k-1}$, *and* $f \in C(\mathbf{R}^k)$. *Then*

$$\mathbf{E}(f(\mathbf{A}_k^n) \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1}) \to \mathbf{E}(f(\mathbf{A}_k) \mid \mathbf{A}_{k-1} = \mathbf{y}_{k-1})$$

uniformly over $\mathbf{y}_{k-1} \in K$.

Proof. $K \subset [-M, M]^{k-1}$ for some $M \ge 1$. Now

$$\begin{aligned}\mathbf{E}(f(\mathbf{A}_k^n) \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1}) &= \mathbf{E}(f(\mathbf{A}_{k-1}^n, a_k(\mathbf{b}_{k-1}(\mathbf{A}_{k-1}^n), X_k^n)) \mid \mathbf{A}_{k-1}^n = \mathbf{y}_{k-1})\\ &= \mathbf{E}(f(\mathbf{y}_{k-1}, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}), X_k^n))) \quad \text{by independence.}\end{aligned}$$

Let $\varepsilon > 0$. Choose λ so large that $P(|X_k^n| > \lambda) < \varepsilon\|f\|$ for all n (by u.i.). Let $L = c(cM \vee \lambda)$. Since $h(\mathbf{y}_{k-1}, x) := f(\mathbf{y}_{k-1}, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}), x))$ is uniformly continuous on $K \times [-L, L]$, there is $\delta > 0$ such that for each $x \in [-L, L]$, the oscillation of $h(\cdot, x)$ over any set of radius $\le \delta$ contained in K is less than ε. Choose a finite δ-net $\mathscr{N}$ for K. Since $X_k^n \xrightarrow{\mathscr{D}} X_k$, we can choose n_0 so that $n \geq n_0 \Rightarrow$

$$|\mathbf{E}f(\mathbf{y}_{k-1}^*, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}^*), X_k^n)) - \mathbf{E}f(\mathbf{y}_{k-1}^*, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}^*), X_k))| < \varepsilon$$

for all $\mathbf{y}_{k-1}^* \in \mathscr{N}$. Now for $\mathbf{y}_{k-1} \in K$, choose $\mathbf{y}_{k-1}^* \in \mathscr{N}$ such that $|\mathbf{y}_{k-1} - \mathbf{y}_{k-1}^*| < \delta$. It follows by the triangle inequality that for $n \geq n_0$,

$$|\mathbf{E}f(\mathbf{y}_{k-1}, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}), X_k^n)) - \mathbf{E}f(\mathbf{y}_{k-1}, a_k(\mathbf{b}_{k-1}(\mathbf{y}_{k-1}), X_k))| < 7\varepsilon. \quad \square$$

Proof of Theorem 2.1. Fix $k \geq 1$. Assume $\psi_k^n \to \psi_k$ uniformly on compacts, so $\phi_k^n \to \phi_k$ uniformly on compacts also. We shall show that the same is true of ψ_{k-1}^n.

Let K be a compact set in $\mathbf{R}^{k-1}$. Then

$$\psi_{k-1}^n(\mathbf{y}_{k-1}) - \psi_{k-1}(\mathbf{y}_{k-1}) = \int [\phi_k^n(\mathbf{y}_k) - \phi_k(\mathbf{y}_k)]\mu_k^n(\mathbf{y}_{k-1}; dy_k) + \int \phi(\mathbf{y}_k)[\mu_k^n(\mathbf{y}_{k-1}; dy_k) - \mu_k(\mathbf{y}_{k-1}; dy_k)].$$

Let $\varepsilon > 0$. Choose $\lambda > 1$ so large that

$$\int_{|y_k|>\lambda} |\mathbf{y}_k|\mu_k^n(\mathbf{y}_{k-1}; dy_k) < \varepsilon \qquad \text{for all } n, \mathbf{y}_{k-1} \in K,$$

by Lemma 2.2. Choose n_0 so that $n \geq n_0 \Rightarrow |\phi_k^n - \phi_k| < \varepsilon$ on $K \times [-\lambda, \lambda]$. By Lemma 2.3, $|\phi_k^n(\mathbf{y}_k)| \leq 2\alpha|\mathbf{y}_k|$ for $|\mathbf{y}_k| > \lambda$, so the first integral in $(*)$ has magnitude $\leq \varepsilon + 4\alpha\varepsilon$ for $\mathbf{y}_{k-1} \in K$, $n \geq n_0$.

Let $\tau: \mathbf{R} \to \mathbf{R}$ be truncation at $\alpha\lambda + \alpha$, i.e.,

$$\tau(x) = \begin{cases} x & \text{if } |x| \leq \alpha\lambda + \alpha; \\ \alpha\lambda + \alpha & \text{if } x > \alpha\lambda + \alpha; \\ -\alpha\lambda - \alpha & \text{if } x < -\alpha\lambda - \alpha. \end{cases}$$

Then

$$\lim_{n\to\infty} \int \tau \circ \phi_k(\mathbf{y}_k)[\mu_k^n(\mathbf{y}_{k-1}; dy_k) - \mu_k(\mathbf{y}_{k-1}; dy_k)] = 0,$$

with the convergence *uniform* over $\mathbf{y}_{k-1} \in K$, by Lemma 2.5. But $|\phi_k(\mathbf{y}_k)| \leq \alpha|\mathbf{y}_k| + \alpha$, so

$$|\phi_k(\mathbf{y}_k)| > \alpha\lambda + \alpha \Rightarrow |\mathbf{y}_k| > \lambda.$$

It follows that the second integral in $(*)$ has magnitude $< 4\alpha\varepsilon$ for sufficiently large n, uniformly over $\mathbf{y}_{k-1} \in K$. Thus $\psi_{k-1}^n \to \psi_{k-1}$ uniformly on K, and the proof is complete. $\square$

Corollary 2.6. *V is a continuous function on any u.i. collection of random N-vectors with independent components, and also on the collection consisting of the N-vectors of partial averages or partial sums of those vectors.*

Remark. The analogy of this for *infinite* sequences of independent random variables fails. As a trivial example, consider $\mathbf{X}^n = (0, 0, \ldots, 0, 1, 1, \ldots, 1, \ldots)$, with zeros in the first n components and 1 thereafter. This even has *constant* components. Now $\mathbf{X}^n \to \mathbf{X} = (0, 0, \ldots)$. But $V(\mathbf{X}) = 0$ and $V(\mathbf{X}^n) = 1$ for all n.

3. Appendix on inverses of triangular functions

We shall show that a one-to-one continuous triangular function (see §2 for definition) on an open set in $\mathbf{R}^N$ with a certain connectivity property will automatically have a triangular inverse. We use invariance of domain.

For all l, all $k \geq l$, let P_l be the projection of $\mathbf{R}^k$ onto the first l coordinates: $P_l(x_1, \ldots, x_k) = (x_1, \ldots, x_l)$.

Definition 3.1. $\mathscr{D} \subset \mathbf{R}^N$ is *iteratively connected* if $P_1(\mathscr{D})$ is connected and for all $1 < k < N$, for all $(x_1, \ldots, x_{k-1}) \in P_{k-1}\mathscr{D}$,

$$\{t : (x_1, \ldots, x_{k-1}, t) \in P_k\mathscr{D}\}$$

is connected.

Remark. For $N = 2$, the only requirement is that $P_1\mathscr{D}$ is connected.

Examples 3.2. (1) Any convex set is iteratively connected. (2) A cylinder in $\mathbf{R}^3$ with elements parallel to the x_3-axis and a horseshoe-shaped cross-section aligned along the x_1-axis is *not* iteratively connected, even though it is simply connected. (3) $(0, 1) \times [(0, 1) \cup (2, 3)] \subset \mathbf{R}^2$ is iteratively connected but not connected. This is only because no condition is made on the highest dimension. Except for that, iterative connectivity is much stronger than connectivity.

Proposition 3.3. *If $\mathscr{D}$ is an open, iteratively connected subset of $\mathbf{R}^N$, and $\mathbf{a}: D \to \mathbf{R}^N$ is one-to-one, continuous and triangular, then $\mathbf{a}^{-1}$ is triangular also.*

Proof. Write $\mathbf{a}(x_1, \ldots, x_N) = (a_1(x_1), a_2(x_1, x_2), \ldots, a_N(x_1, \ldots, x_N))$. Clearly, it is enough to show that for $1 \leq k \leq N$, for all $(x_1, \ldots, x_{k-1}) \in P_{k-1}\mathscr{D}$, $a_k(x_1, \ldots, x_{k-1}, \cdot)$ is one-to-one on $G(x_1, \ldots, x_{k-1}) := \{t : (x_1, \ldots, x_{k-1}, t) \in P_k\mathscr{D}\}$; for then x_k can be determined from $a_k(x_1, \ldots, x_k)$, once $x_1, \ldots, x_{k-1}$ are known. When $k = N$, the fact that $\mathbf{a}$ is one-to-one obviously implies $a_N(x_1, \ldots, x_{N-1}, \cdot)$ is one-to-one, so we only need consider $k < N$. When $k = 1$, we mean by the above to simply show that a_1 is one-to-one on $P_1\mathscr{D}$.

Suppose for some $1 \leq k < N$ and $\mathbf{x}_{k-1} \in P_{k-1}\mathscr{D}$, $a_k(\mathbf{x}_{k-1}, \cdot)$ is *not* one-to-one on $G = G(x_1, \ldots, x_{k-1})$. G is connected by hypothesis, and is clearly open since $\mathscr{D}$ is. It is easy to see by the intermediate value theorem that there exist $t^0 \in G$ and sequences t^n, u^n in G with $t^n \neq u^n$ and $a_k(\mathbf{x}_{k-1}, t^n) = a_k(\mathbf{x}_{k-1}, u^n)$ for all n, and $t^n \to t^0$ and $u^n \to t^0$.

Define a new function $\mathbf{f}: \mathscr{D} \to \mathbf{R}^N$ by

$$\mathbf{f}(\mathbf{x}) = (\mathbf{x}_k, a_{k+1}(\mathbf{x}_{k+1}), \ldots, a_{\mathbf{N}}(\mathbf{x})).$$

This is one-to-one on $\mathscr{D}$, since

$$(a_{k+1}(\mathbf{x}_k, \cdot), a_{k+2}(\mathbf{x}_k \cdot, \cdot), \ldots, a_N(\mathbf{x}_k, \ldots))$$

must be one-to-one for each $\mathbf{x}_k \in P_k\mathscr{D}$ or else $\mathbf{a}$ would not be one-to-one.

By Brouwer's invariance of domain theorem [AB, p. 156], $\mathbf{f}(\mathscr{D})$ is open in $\mathbf{R}^N$. Since $t^0 \in G$, there exists $\mathbf{x}^0 \in \mathscr{D}$ such that $\mathbf{x}^0_{k-1} = \mathbf{x}_{k-1}$ and $x^0_k = t^0$. So there exist open sets $A \subset \mathbf{R}^{k-1}$, $B \subset \mathbf{R}^{N-k}$, and $\alpha < \beta$ such that $\mathbf{f}(\mathbf{x}^0) \in A \times (\alpha, \beta) \times B \subset \mathbf{f}(\mathscr{D})$. Since $f_k(\mathbf{x}^0) = t^0$, we must have $t^0 \in (\alpha, \beta)$, so for some n, $t^n \in (\alpha, \beta)$ and $u^n \in (\alpha, \beta)$ also. Let $t = t^n$ and $u = u^n$.

Let $\mathbf{z} = (f_{k+1}(\mathbf{x}^0), \dots, f_N(\mathbf{x}^0)) = (a_{k+1}(\mathbf{x}^0_{k+1}), \dots, a_N(\mathbf{x}^0)) \in B$. Also, $\mathbf{f}_{k-1}(\mathbf{x}^0) = \mathbf{x}_{k-1} \in A$. Thus

$$(\mathbf{x}_{k-1}, t, \mathbf{z}) \in A \times (\alpha, \beta) \times B \subset \mathbf{f}(D)$$

and $(\mathbf{x}_{k-1}, u, \mathbf{z}) \in \mathbf{f}(\mathscr{D})$ also. So there exist $\mathbf{v}$, $\mathbf{w}$ in $\mathscr{D}$ such that $\mathbf{f}(\mathbf{v}) = (\mathbf{x}_{k-1}, t, \mathbf{z})$ and $\mathbf{f}(\mathbf{w}) = (\mathbf{x}_{k-1}, u, \mathbf{z})$. From the definition of $\mathbf{f}$, we get $v_k = t$, $w_k = u$, $\mathbf{v}_{k-1} = \mathbf{w}_{k-1} = \mathbf{x}_{k-1}$, and $(a_{k+1}(\mathbf{v}_{k+1}), \dots, a_{\mathscr{N}}(\mathbf{v})) = (a_{k+1}(\mathbf{w}_{k+1}), \dots, a_N(\mathbf{w})) = \mathbf{z}$. Since $a_k(\mathbf{x}_{k-1}, t) = a_k(\mathbf{x}_{k-1}, u)$, this implies $\mathbf{a}(\mathbf{v}) = \mathbf{a}(\mathbf{w})$. But $\mathbf{v} \neq \mathbf{w}$, so $\mathbf{a}$ is not one-to-one, a contradiction. □

Examples 3.4. (1) Let $\operatorname{sgn} x = 1$ if $x \geq 0$, -1 if $x < 0$. The discontinuous triangular function $(|x_1|, e^{x_2} \operatorname{sgn} x_1)$ is one-to-one on $\mathbf{R}^2$ but does not have a triangular inverse.

(2) The same function as in (1) is continuous when restricted to the disconnected open set $[(-\infty, -1) \cup (1, \infty)] \times \mathbf{R}$, but does not have a triangular inverse.

(3) The continuous triangular function (x_1^2, x_2) is one-to-one on the nonopen, iteratively connected set $\{x_1 = x_2\}$, but does not have a triangular inverse.

(4) For a more interesting example, we describe a triangular function which is one-to-one and continuous on an open, simply connected (but not iteratively connected) bounded domain in $\mathbf{R}^3$, but does not have a triangular inverse.

Let $\mathscr{D} = W \times (0, 1)$ where

$$W = [(-1, 3) \times (1, 2)] \cup [(-1, 3) \times (-2, -1)] \cup [(2, 3) \times [-1, 1]]$$

which is a "horseshoe" in $\mathbf{R}^2$. Note $\mathscr{D}$ is open and simply connected.

Let $a_1(x_1) = x_1$. Let

$$a_2(x_1, x_2) = \begin{cases} x_2, & 1 \leq x_1 < 3 \text{ or } x_2 > 0; \\ x_2 + 3(1 - |x_1|), & |x_1| < 1 \text{ and } x_2 < 0. \end{cases}$$

Let

$$a_3(x_1, x_2, x_3) = \begin{cases} (1 + x_3)3, & 2 \leq x_1 < 3; \\ (1 + x_3)(3 + 2(2 - x_1) \operatorname{sgn} x_2), & 1 \leq x_1 < 2; \\ (1 + x_3)(3 + \operatorname{sgn} x_2), & |x_1| < 1. \end{cases}$$

Let $\mathbf{a} = (a_1, a_2, a_3)$ have domain $\mathscr{D}$. It is easily checked that $\mathbf{a}$ is continuous. Now $(a_1(x_1), a_2(x_1, x_2))$ is *not* one-to-one on W, since $a_2(0, \frac{3}{2}) = a_2(0, -\frac{3}{2}) = \frac{3}{2}$. Thus $\mathbf{a}$ cannot have a triangular inverse. But $\mathbf{a}$ *is* one-to-one on $\mathscr{D}$: the only case which needs to be checked is when $|x_1| < 1$, but then $\operatorname{sgn} x_2$ can be determined from whether a_3 lies in $(2, 4)$ or $(4, 8)$ and then x_2 can be determined from a_2 and x_1.

Acknowledgments

We would like to thank Ted Hill and Bob Kertz for helpful suggestions and, also the referee. We thank Martin Jones for suggesting the problem.

References

[AB] E. Artin and H. Braun, *Introduction to algebraic topology*, Charles Merrill, Columbus, Ohio, 1969.

[B] P. Billingsley, *Convergence of probability measures*, Wiley, New York, 1968.

[CRS] Y. Chow, H. Robbins, and D. Siegmund, *Great expectations: The theory of optimal stopping*, Houghton Mifflin, Boston, 1971.

[CkD.] Cox and R. Kertz, *Prophet regions and sharp inequalities for p-th absolute moments of martingales*, J. Multivariate Anal. **18** (1986), 242–273.

[EK] J. Elton and R. Kertz, *Comparison of stop rule and maximum expectations for finite sequences of exchangeable random variables*, (preprint).

[HK_1] T. Hill and R. Kertz, *Additive comparisons of stop rule and supremum expectations of uniformly bounded random variables*, Proc. Amer. Math. Soc. **83** (1981), 582–585.

[HK_2] ——, *Comparisons of stop rule and supremum expectations of i.i.d. random variables*, Ann. Prob. **10** (1982), 336–345.

[HK_3] ——, *Stop rule inequalities for uniformly bounded sequences of random variables*, Trans. Amer. Math. Soc. **278** (1983), 197–207.

[K] D. Kennedy, *Optimal stopping of independent random variables and maximizing prophets*, Ann. Probab. **13** (1985), 566–571.

[Ker] R. Kertz, *Stop rule and supremum expectations of i.i.d. random variables: a complete comparison by conjugate duality*, J. Multivariate Anal. **19** (1986), 88–112.

[KS_1] U. Krengel and L. Sucheston, *Semiamarts and finite values*, Bull. Amer. Math. Soc. **83** (1977), 745–747.

[KS_2] ——, *On semiamarts, amarts, and processes with finite value*, Adv. Prob. Related. Topics **4** (1978), 197–266.

[P] J. Pickands, *Extreme order statistics with cost of sampling*, Adv. in Appl. Probab. **15** (1983), 783–797.

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332