J. Appl. Prob. **35,** 693–706 (1998)
Printed in Israel

OPTIMAL MULTIVARIATE STOPPING RULES

DAVID ASSAF,* ** AND
ESTER SAMUEL-CAHN,* *The Hebrew University of Jerusalem*

Abstract

For fixed i let $\boldsymbol{X}(i) = (X_1(i), \ldots, X_d(i))$ be a d-dimensional random vector with some known joint distribution. Here i should be considered a time variable. Let $\boldsymbol{X}(i), i = 1, \ldots, n$ be a sequence of n independent vectors, where n is the total horizon. In many examples $X_j(i)$ can be thought of as the return to partner j, when there are $d \geq 2$ partners, and one stops with the ith observation. If the jth partner alone could decide on a (random) stopping rule t, his goal would be to maximize $EX_j(t)$ over all possible stopping rules $t \leq n$. In the present 'multivariate' setup the d partners must however cooperate and stop at the *same* stopping time t, so as to maximize some agreed function $h(\cdot)$ of the individual expected returns. The goal is thus to find a stopping rule t^* for which $h(EX_1(t), \ldots, EX_d(t)) = h(E\boldsymbol{X}(t))$ is maximized. For continuous and monotone h we describe the class of optimal stopping rules t^*. With some additional symmetry assumptions we show that the optimal rule is one which (also) maximizes EZ_t where $Z_i = \sum_{j=1}^{d} X_j(i)$, and hence has a particularly simple structure. Examples are included, and the results are extended both to the infinite horizon case and to the case when $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ are dependent. Asymptotic comparisons between the present problem of finding $\sup h(E\boldsymbol{X}(t))$ and the 'classical' problem of finding $\sup Eh(\boldsymbol{X}(t))$ are given. Comparisons between the optimal return to the statistician and to a 'prophet' are also included. In the present context a 'prophet' is someone who can base his (random) choice g on the full sequence $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$, with corresponding return $\sup h(E\boldsymbol{X}(g))$.

Keywords: Cooperative stopping rule; finite and infinite horizon; maximization of functions of expected value; optimal choice set; prophet inequalities

AMS 1991 Subject Classification: Primary 60G40

1. Introduction

Consider a firm which holds $d \geq 2$ different stocks, and is seeking a loan against these holdings. The firm pledges to sell the stocks within a given time period, say a year, all stocks to be sold at one time. The loan granted will be some known function $h(\cdot)$ of d variables, here of the expected values of the stocks, when sold. We shall formalize this model. Let $\boldsymbol{X}(i) = (X_1(i), \ldots, X_d(i))$ be the (random) value of the d stocks at time i, where $i = 1, \ldots, n$. We assume that the joint distribution of all variables is known, and that all have finite expectations. n is the total horizon. The time of selling, t, is a random *stopping time*, $t \leq n$, and the decision to sell at time i can clearly depend only on the values observed in the past and at present, i.e.

Received 1 October 1996; revision received 24 March 1997.
* Postal address: Department of Statistics, Hebrew University, Jerusalem 91905, Israel.
** Email address: msdave@pluto.mscc.huji.ac.il
This research was supported by Grant no. 94-00186 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel.

on $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)$, but not on future values. Thus the random value of the stocks at the time of sale is $\boldsymbol{X}(t) = \sum_{i=1}^{n} \boldsymbol{X}(i) I(t = i)$, and we denote by $E\boldsymbol{X}(t)$ the vector of expectations, $E\boldsymbol{X}(t) = (EX_1(t), \ldots, EX_d(t))$, i.e. the vector of the expected values of the d stocks at the time sold.

If T_n denotes the set of all stopping rules for which $P(t \leq n) = 1$, and the goal is to maximize the loan, we seek

$$\sup_{t \in T_n} h(E\boldsymbol{X}(t)). \tag{1.1}$$

The present paper is devoted to solving (1.1), that is, to finding the value in (1.1), showing that under mild conditions it is actually attainable, and describing the rules for which it is attained.

It should be noted that the problem described is entirely different from that of finding a rule $t \leq n$ for which

$$\sup_{t \in T_n} Eh(\boldsymbol{X}(t)) \tag{1.2}$$

is attained. The latter problem entails nothing new, since if we set $Y_i = h(\boldsymbol{X}(i))$, then (1.2) is equivalent to seeking an optimal stopping rule for the 'classical' one dimensional problem $\sup_{t \in T_n} EY_t$ and can be solved by backward induction. To distinguish (1.1) from (1.2) we call the former the 'multivariate stopping problem', as in the title of this paper. There is no direct backward induction technique for the multidimensional problem. Sometimes the term 'cooperative stopping problem' may be more suitable (see Assaf and Samuel-Cahn (1998)). The following is an example.

A doctors' office with $d \geq 2$ partners plans to hire a receptionist. n candidates apply for the job, and are interviewed sequentially. The candidate must be told immediately after the interview whether he or she is hired. The utility of the ith candidate is a random vector $\boldsymbol{X}(i) = (X_1(i), \ldots, X_d(i))$ with known joint distribution, where $X_j(i)$ is the utility attached by the jth doctor to this candidate, $j = 1, \ldots, d$. The jth doctor's ideal would be to find a stopping rule so as to maximize $EX_j(t)$. (In this particular example the d variables will typically be dependent.) Since all doctors will engage the same receptionist, the jth doctor must, however, cooperate with the others, and they agree to maximize some known agreed function h of each of their expected utilities, upon stopping. Thus (1.1) describes their agreed goal.

In Section 2 we find the general solution to the maximization problem (1.1) for the case where the vectors $\boldsymbol{X}(i)$, $i = 1, \ldots, n$ are independent. In the presence of a symmetric structure the solution becomes surprisingly simple and is given explicitly. The optimal rule for this case is a 'one-dimensional' rule which maximizes EZ_t, where $Z_i = \sum_{j=1}^{d} X_j(i)$, $i = 1, \ldots, n$. In Section 3 examples are given, and Section 4 is devoted to additional remarks and generalizations. These include asymptotic comparisons, the infinite horizon problem and generalizations to the dependent case. The last section is devoted to 'prophet inequalities' in this multivariate setting.

Multivariate stopping rules have been considered in the literature earlier. Most of the papers have been in a game-theoretic setting. For example, Mamer (1987) establishes, under certain assumptions, the existence of a Nash equilibrium in non-randomized stopping times. In the setup described there, the players are not all required to stop at the same time. The approach of Yasuda *et al.* (1982) is closer to the approach taken here. In that paper, at every time instant i, each partner (player) votes whether to stop, (1) or not, (0), after seeing $X(i)$ (and depending on

$X(i)$ only). The group decision on whether or not to stop is a monotone function (on $\{0, 1\}^d$) of the individual stopping decisions. Under certain conditions the existence of an equilibrium stopping strategy and associated 'values' (gains) are described.

Bassan and Scarsini (1995) discuss a related setting, where the 'value to society' is some increasing function of the sum of the utilities of the d partners. The problem there is that of how much information should be released at each stage by a 'chief', whose goal it is to maximize the value to society. The model is therefore related to ours, but different inasmuch as it does not involve optimal stopping.

2. Main results

We begin by describing the set of all (randomized) stopping rules, T_n, for the sequence $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$. A stopping rule $t \in T_n$ consists of a sequence of 'critical functions', $\{\phi_i\}_{i=1}^n$, where $0 \le \phi_i \le 1$, and $\phi_i = \phi_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i))$, $\phi_n \equiv 1$. Here $\phi_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i))$ is measurable with respect to the σ-field generated by $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)$, and denotes the *conditional* probability with which one stops at the ith observation (i.e. $t = i$) when $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)$ is observed, and one has not stopped earlier. Non-randomized rules are obtained by letting ϕ_i take values 0 and 1 only. For $t \in T_n$ we thus have

$$\begin{aligned} E\boldsymbol{X}(t) &= E\sum_{i=1}^{n} \boldsymbol{X}(i) I(t = i) \\ &= \sum_{i=1}^{n} E\left[\prod_{k=1}^{i-1}\{1 - \phi_k(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(k))\}\right]\phi_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i))\boldsymbol{X}(i). \end{aligned} \tag{2.1}$$

In the case where the $\boldsymbol{X}(i)$'s are independent, $\phi_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i))$ can, without loss of generality, be replaced by $\tilde{\phi}_i(\boldsymbol{X}(i)) = E[\phi_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)) \mid \boldsymbol{X}(i)]$, and we henceforth assume that this replacement has taken place in all cases of independence, and thus omit the tilde, and write $\phi_i \equiv \phi_i(\boldsymbol{X}(i))$.

Let $S_n = \{E\boldsymbol{X}(t), t \in T_n\}$. We begin by showing that S_n is a compact set in $\mathbb{R}^d$.

Theorem 2.1. *Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ be independent. Then S_n is compact.*

Proof. Clearly S_n is bounded since by assumption $E|X_j(i)| < \infty$ for all i and j, hence for any $t \in T_n$

$$-\infty < -\sum_{i=1}^{n} E|X_j(i)| \le EX_j(t) \le \sum_{i=1}^{n} E|X_j(i)| < \infty, \qquad j = 1, \ldots, d.$$

We shall prove the fact that S_n is closed, by induction on n. For $n = 1$ one has $S_1 = \{E\boldsymbol{X}(1)\}$ and the statement is trivially true. Now assume that the statement is true for any $n - 1$ random vectors. Let S_{n-1}^* be the set corresponding to S_{n-1}, but based on the $n - 1$ random vectors $\boldsymbol{X}(2), \ldots, \boldsymbol{X}(n)$, and let T_{n-1}^* be the corresponding set of stopping rules. By the induction hypothesis S_{n-1}^* is compact. Let $\boldsymbol{s}_n \in S_n$. By (2.1) $\boldsymbol{s}_n$ can be written as

$$\begin{aligned} \boldsymbol{s}_n &= E[\phi_1(\boldsymbol{X}(1))\boldsymbol{X}(1) + \{1 - \phi_1(\boldsymbol{X}(1))\}\boldsymbol{s}_{n-1}^*] \\ &= E[\phi_1(\boldsymbol{X}(1))(\boldsymbol{X}(1) - \boldsymbol{s}_{n-1}^*)] + \boldsymbol{s}_{n-1}^*, \end{aligned} \tag{2.2}$$

where $\boldsymbol{s}_{n-1}^* \in S_{n-1}^*$ and is based on the rule $t^* \in T_{n-1}^*$ with $\phi_{i-1}^*(\boldsymbol{X}(i)) = \phi_i(\boldsymbol{X}(i))$, $i = 2, \ldots, n$. Now consider a sequence $\{\boldsymbol{s}_n^{(i)}\}$ of elements in S_n, such that $\lim_{i\to\infty} \boldsymbol{s}_n^{(i)} = \boldsymbol{s}_n^0$. We must show that $\boldsymbol{s}_n^0 \in S_n$. Let $\phi_1^{(i)}$ and $\boldsymbol{s}_{n-1}^{(i)}$ be the components of $\boldsymbol{s}_n^{(i)}$ corresponding to representation (2.2). By compactness of S_{n-1}^* the sequence $\boldsymbol{s}_{n-1}^{(i)}$ has a converging subsequence $\boldsymbol{s}_{n-1}^{(i_j)}$ with limit point $\boldsymbol{s}_{n-1}^0 \in S_{n-1}^*$, where $\boldsymbol{s}_{n-1}^0$ corresponds to a stopping rule $t_{n-1}^0 \in T_{n-1}^*$. Consider the set $\{E[\phi(\boldsymbol{X}(1))(\boldsymbol{X}(1) - \boldsymbol{s}_{n-1}^0)] : \phi \in \Gamma]\}$, where Γ denotes the set of critical functions $\phi = \phi(\boldsymbol{x})$. By Theorem 5(iv) in Chapter 3 of Lehmann (1991) this set is compact. (This is essentially the weak compactness theorem.) Thus $E[\phi_1^{(i_j)}(\boldsymbol{X}(1))(\boldsymbol{X}(1) - \boldsymbol{s}_{n-1}^0)]$ $j = 1, 2, \ldots$ has a converging subsequence which tends to $E[\phi_1^0(\boldsymbol{X}(1))(\boldsymbol{X}(1) - \boldsymbol{s}_{n-1}^0)]$ for some critical function ϕ_1^0. Now consider the rule $\tilde{t} \in T_n$ with $\tilde{\phi}_1 = \phi_1^0$ and continuation corresponding to $t_{n-1}^0 \in T_{n-1}^*$. Again using (2.2), it is seen that its value is $E[\phi_1^0(\boldsymbol{X}(1))(\boldsymbol{X}(1) - \boldsymbol{s}_{n-1}^0)] + \boldsymbol{s}_{n-1}^0$. But it follows by representation (2.2) and by the construction that this value must be $\boldsymbol{s}_n^0$. Thus $E\boldsymbol{X}(\tilde{t}) = \boldsymbol{s}_n^0$, which shows that S_n is compact.

Theorem 2.2. *S_n is convex.*

Proof. Let $t \in T_n$ be given. A representation of t equivalent to the one given in the beginning of this section is by means of the sequence of functions

$$p_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)) = \left[\prod_{k=1}^{i-1}\{1 - \phi_k(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(k))\}\right]\phi_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)),\ i = 1, \ldots, n,$$

where $p_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i))$ denotes the unconditional probability of stopping at time i when $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i)$ is observed. With this representation (2.1) is written as

$$E\boldsymbol{X}(t) = \sum_{i=1}^{n} Ep_i(\boldsymbol{X}(1), \ldots, \boldsymbol{X}(i))\boldsymbol{X}(i).$$

Now let $\boldsymbol{s}_r \in S_n$, $r = 1, 2$, with corresponding sequences $\{p_i^{(r)}\}_1^n$, $r = 1, 2$. Then $\lambda\boldsymbol{s}_1 + (1-\lambda)\boldsymbol{s}_2$ for $0 \le \lambda \le 1$ is the value for the stopping rule t with $p_i(\cdot) = \lambda p_i^{(1)}(\cdot) + (1-\lambda)p_i^{(2)}(\cdot)$, $i = 1, \ldots, n$ (where it is quite straightforward to show that $\{p_i(\cdot)\}_1^n$ corresponds to a stopping rule).

Now consider the function $h(\ )$ for which we seek $\sup_{t \in T_n} h(EX_1(t), \ldots, EX_d(t))$. We assume it is well defined for all $\boldsymbol{X}(i)$.

Theorem 2.3. *Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ be independent. If $h(\cdot)$ is upper semicontinuous there exists a rule $t^* \in T_n$ such that*

$$\sup_{t \in T_n} h(EX_1(t), \ldots, EX_d(t)) = h(EX_1(t^*), \ldots, EX_d(t^*)),$$

i.e. the supremum is attained.

Proof. Immediate by compactness.

(Clearly upper semicontinuity is only a sufficient condition for the maximum to be attained. There are examples where one can show directly that the maximum is attained, without this assumption.)

We now approach the practical problem of calculating $\max_{t \in T_n} h(E\boldsymbol{X}(t))$, and, perhaps more importantly, of finding a rule t^* for which this maximum is attained. Simple backward induction cannot be used, since $h(E\boldsymbol{X}(t))$ is a d-variate function.

Consider the case where $h(\cdot)$ is upper semicontinuous and monotone in each coordinate. It may be monotone increasing in some coordinates and monotone decreasing in the others. Denote the class of such functions by H. Our main purpose in the present section is to show that for $h \in H$ a rule t^* maximizing $h(E\boldsymbol{X}(t))$ has a particularily simple form. It is an optimal rule for a suitably defined sequence of one-dimensional independent random variables $Z_1, \dots, Z_n$, i.e. it maximizes EZ_t. Since the structure of optimal rules for a finite sequence of independent one-dimensional random variables is particularly simple and well known (see e.g. Chow *et al.* (1971), p. 50), this reflects on the form of t^*. Recall that the rules maximizing EZ_t for independent Z_i are obtained through the sequence of constants c_i, $i = 1, \dots, n$ (called 'thresholds') where c_i are defined recursively by

$$c_n = -\infty, \qquad c_i = E(Z_{i+1} \vee c_{i+1}), \qquad i = n-1, \dots, 1. \tag{2.3}$$

Then t is optimal for the Z-sequence if and only if it is equivalent to

$$\phi_i(Z_i) = \begin{cases} 1 & \text{for } Z_i > c_i \\ \text{arbitrary} & \text{for } Z_i = c_i \\ 0 & \text{for } Z_i < c_i, \qquad i = 1, \dots, n. \end{cases} \tag{2.4}$$

On the 'arbitrary' part, $\{Z_i = c_i\}$ we have let the randomization depend on Z_i only, i.e. it is a constant (as we wrote $\phi_i(Z_i)$), though in fact one may also consider external randomization. Though any rule of the form (2.4) maximizes EZ_t it is customary in (2.4) to consider the non-randomized rule with $\phi_i(Z_i) = 1$ on $\{Z_i = c_i\}$, as this rule stops as early as possible. For the solution of the multivariate case, which is described in the next theorem, however, randomization cannot be ignored. Furthermore, on the set $\{Z_i = c_i\}$ it will not suffice to let the randomization be a constant $0 \le \gamma \le 1$. See Example 3.2.

Let $\langle \cdot, \cdot \rangle$ denote inner product.

Theorem 2.4. *Suppose $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ are independent. Let $h \in H$. There exists a vector of constants $\boldsymbol{a} = (a_1, \dots, a_d)$, to be described explicitly, such that if we define $Z_i = \langle \boldsymbol{a}, \boldsymbol{X}(i) \rangle$, $i = 1, \dots, n$, then an optimal rule $t^* \in T_n$ for maximizing $h(E\boldsymbol{X}(t))$ is of the form*

$$\phi_i(\boldsymbol{X}(i)) = \begin{cases} 1 & \text{if } Z_i > c_i \\ 0 & \text{if } Z_i < c_i, \qquad i = 1, \dots, n \end{cases} \tag{2.5}$$

where the c_i's are defined for the Z-sequence, in (2.3). On $\{Z_i = c_i\}$ randomization depending on the observed $\boldsymbol{X}(i)$ value is generally needed, unless this set has probability zero.

Proof. Let $\boldsymbol{s}^* \in S_n$ be a point where $\max_{t \in T_n} h(E\boldsymbol{X}(t))$ is attained, with corresponding rule t^*. By the monotonicity of h it follows that $\boldsymbol{s}^*$ can without loss of generality be taken to be a boundary point of S_n. By the convexity of S_n there exists a supporting hyperplane to S_n through $\boldsymbol{s}^*$. Denote its equation by $\langle \boldsymbol{a}, \boldsymbol{x} \rangle = b$, where $\langle \boldsymbol{a}, E\boldsymbol{X}(t) \rangle \le b$ for all $t \in T_n$, with equality for $t = t^*$. Let $Z_i = \langle \boldsymbol{a}, \boldsymbol{X}(i) \rangle$ for the $\boldsymbol{a}$ so defined. Let T_n^Z denote the set of (randomized) stopping rules based on the Z-sequence only. Clearly $T_n^Z \subset T_n$. Note that since Z_i is a function of $\boldsymbol{X}(i)$ the value EZ_t is well defined for all $t \in T_n$. Thus

$$\begin{aligned} \max_{t \in T_n^Z} EZ_t &\le \sup_{t \in T_n} EZ_t = \sup_{t \in T_n} E\langle \boldsymbol{a}, \boldsymbol{X}(t) \rangle \\ &= \sup_{t \in T_n} \langle \boldsymbol{a}, E\boldsymbol{X}(t) \rangle = \max_{\boldsymbol{s} \in S_n} \langle \boldsymbol{a}, \boldsymbol{s} \rangle = \langle \boldsymbol{a}, \boldsymbol{s}^* \rangle = EZ_{t^*}. \end{aligned} \tag{2.6}$$

We next show that the inequality in (2.6) can be replaced by equality. This is achieved by showing that for each $t \in T_n$ there exists a $t^Z \in T_n^Z$ for which $EZ_t = EZ_{t^Z}$. Let t be defined through the sequence $\phi_i(\boldsymbol{X}(i))$, $i = 1, \ldots, n$. Define $\phi_i^Z(Z_i)$, corresponding to the desired $t^Z \in T_n^Z$ through $\phi_i^Z(z) = E[\phi_i(\boldsymbol{X}(i)) \mid \langle \boldsymbol{a}, \boldsymbol{X}(i)\rangle = z]$. Then clearly $EZ_t = EZ_{t^Z}$. Hence equality holds throughout in (2.6), which shows that the $t^{*Z} \in T_n^Z$ corresponding to t^* is optimal for $EZ_t, t \in T_n^Z$. Thus t^{*Z} must be of the form (2.4), which implies that t^* must a.s. satisfy (2.5).

Remark 2.1. Note that if the maximum of $h(\boldsymbol{s})$ over S_n is attained at a unique point $\boldsymbol{s}^*$, and if there is a unique supporting hyperplane $\langle \boldsymbol{a}, \boldsymbol{x}\rangle = b$ through $\boldsymbol{s}^*$, then a t^* corresponding to $\boldsymbol{s}^*$ *necessarily* satisfies (2.5) for $Z_i = \langle \boldsymbol{a}, \boldsymbol{X}(i)\rangle$. The randomization part need not necessarily be unique (see Example 3.2).

Suppose the distribution of $\boldsymbol{X}(i)$ is the same as that of $\boldsymbol{X}_\Pi(i) = (X_{\Pi(1)}(i), \ldots, X_{\Pi(d)}(i))$ for all $d!$ permutations Π of $\{1, \ldots, d\}$. We call such an $\boldsymbol{X}(i)$ exchangeable.

The most useful result in the present section is Corollary 2.1. It gives quite broad sufficient conditions for the Z_i's of Theorem 2.4 to have the simple form $Z_i = \sum_{j=1}^d X_j(i)$, where an optimal rule for these Z_i's does not require randomization.

For any given $\boldsymbol{x} = (x_1, \ldots, x_d)$ write $\bar{x} = \sum_{j=1}^d x_j/d$. Let $\hat{H}$ denote the subset of H of all $h(\,)$ which are monotone non-decreasing in each coordinate, and satisfy

$$h(x_1, \ldots, x_d) \le h(\bar{x}, \ldots, \bar{x}) \qquad \text{for all } \boldsymbol{x}. \tag{2.7}$$

Corollary 2.1. *Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ be independent and suppose $\boldsymbol{X}(i)$ is exchangeable, for all $i = 1, \ldots, n$. Let $h \in \hat{H}$, and $Z_i = \sum_{j=1}^d X_j(i)$. Then any rule satisfying (2.4) for these Z_i's maximizes $h(E\boldsymbol{X}(t))$.*

Proof. The exchangeability of each of the $\boldsymbol{X}(i)$'s implies that S_n is permutation invariant. Suppose $\boldsymbol{s}^* = (s_1^*, \ldots, s_d^*)$ is such that $\max_{\boldsymbol{s} \in S_n} h(\boldsymbol{s}) = h(\boldsymbol{s}^*)$. Then $\boldsymbol{s}_\Pi^* \in S_n$ for all permutations Π, and $\Sigma \boldsymbol{s}_\Pi^*/d!$ also belongs to S_n, by convexity, where the summation is over all $d!$ permutations. But $\Sigma \boldsymbol{s}_\Pi^*/d! = (\bar{s}^*, \ldots, \bar{s}^*) = \bar{\boldsymbol{s}}^*$, and by (2.7) the inequality $h(\boldsymbol{s}^*) \le h(\bar{\boldsymbol{s}}^*)$ holds, thus there necessarily exists a vector in S_n with all coordinates equal, where the maximum of $h(\boldsymbol{s})$ is attained. This point can again be taken to be a boundary point. A supporting hyperplane through this point is $\sum_{j=1}^d x_j = d\bar{s}^*$. Now by the exchangeability, any rule t maximizing the value for the $Z_i = \sum_{j=1}^d X_j(i)$ sequence, necessarily satisfies $EX_1(t) = \cdots = EX_d(t)$, and thus this value must be $\bar{s}^*$. (Note, in particular, that in the above case no randomization is needed.)

Remark 2.2. The class of functions satisfying (2.7) is quite broad. For example, it includes as a subset the set of all Schur concave h's, which, in turn, contains all symmetric and concave h's. (See for example Marshall and Olkin (1979), Chapter 3.)

3. Examples

In all examples in the present section we assume that $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ are independent.

Example 3.1. Let $X_j(i) \ge 0$ for $j = 1, \ldots, d$, $i = 1, \ldots, n$ and let

$$h(\boldsymbol{x}) = h(x_1, \ldots, x_d) = \prod_{j=1}^d x_j.$$

Then h is monotone increasing and (2.7) holds, since here (2.7) is equivalent to the statement that the geometric mean is less than or equal to the arithmetic mean. If $\boldsymbol{X}(i)$ is exchangeable, $i = 1, \ldots, n$, it follows from Corollary 2.1 that there exists a rule maximizing $h(E\boldsymbol{X}(t))$, which is of the (non-randomized) form $t = \inf\{i : Z_i \geq c_i\}$ with $Z_i = \sum_{j=1}^d X_j(i)$, and the c_i's are the values given in (2.3). Note that the 'stopping regions' are half spaces (i.e. linear), even though the function h is a product! Further results and aspects of this example are studied in detail in Assaf and Samuel-Cahn (1998). The approach and proof in that paper is particular to the product function.

Example 3.2. The following simple example shows that in Theorem 2.4 it may not suffice to consider non-randomized rules. Furthermore, randomization depending only on Z does not suffice. Rather, the randomization needed to obtain an optimal value and rule depends on the actually observed $\boldsymbol{X}$-value. Let $n = 2$ and $d = 2$ and let $\boldsymbol{X}(1)$ and $\boldsymbol{X}(2)$ be i.i.d. where $(X_1(1), X_2(1))$ takes on each of the two values (1, 2) and (2, 1) with probability 1/2. When $n = 2$, t depends on $\phi_1(\boldsymbol{X}(1))$ only, and we shall write $\phi_1((1, 2)) = \gamma_1$, $\phi_1((2, 1)) = \gamma_2$ for simplicity. Direct evaluation for any such rule shows

$$S_2 = \{(\tfrac{3}{2} - \tfrac{1}{4}\gamma_1 + \tfrac{1}{4}\gamma_2, \tfrac{3}{2} + \tfrac{1}{4}\gamma_1 - \tfrac{1}{4}\gamma_2); \quad 0 \leq \gamma_1 \leq 1,\ 0 \leq \gamma_2 \leq 1\}.$$

Equivalently S_2 consists of the straight line-segment $\{x_2 = 3 - x_1, \frac{5}{4} \leq x_1 \leq \frac{7}{4}\}$. Let

$$h(x_1, x_2) = x_1^{1.2} x_2.$$

The *unique* maximum of $h(\cdot)$ over $\boldsymbol{x} \in S_2$ is attained at $(s_1^*, s_2^*) = (\frac{18}{11}, \frac{15}{11})$ with corresponding γ-values satisfying $\gamma_2 - \gamma_1 = \frac{6}{11}$. Thus one can take $\gamma_2 = \frac{6}{11}$, $\gamma_1 = 0$, but one of the γ_j differs necessarily from 0 or 1. Clearly the randomization depends on the value of $\boldsymbol{X}(1)$. Note that the only supporting hyperplane to S_2 through (s_1^*, s_2^*) satisfies $x_1 + x_2 = 3$, i.e. $\boldsymbol{a} = (1, 1)$ and $Z_1 = Z_2 = 3$. Thus no randomization based on the Z_i's can suffice. Also note that γ_1, γ_2 are not unique (see Remark 2.1).

Example 3.3. For each i, let $\boldsymbol{X}(i)$ be exchangeable, $i = 1, \ldots, n$, and consider

$$h(x_1, \ldots, x_d) = \min_{1 \leq j \leq d} x_j.$$

Clearly $h \in \hat{H}$ and hence Corollary 2.1 applies. It thus follows that rules maximizing

$$\min_{1 \leq j \leq d} EX_j(t)$$

can be taken to depend on the sequence of $Z_i = \sum_{j=1}^d X_j(i)$, $i = 1, \ldots, n$ only, and are exact *same* rules as those maximizing the product function, described in Example 3.1 (provided $X_j(i) \geq 0$, $j = 1, \ldots, d$, $i = 1, \ldots, n$)!

Example 3.4. Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ be arbitrary (independent) and consider $h(x_1, \ldots, x_d) = \max_{1 \leq j \leq d} x_j$. Here h is symmetric and increasing but condition (2.7) fails. Nevertheless the solution to our maximization problem is easily obtained. Let t_j be a rule maximizing $EX_j(t)$ for the sequence $X_j(1), \ldots, X_j(n)$ only, ignoring the other random variables. By the independence structure of $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ it follows that when $X_j(i)$ is observed (if one has not stopped earlier) and the objective is to maximize $EX_j(t)$, then $X_1(i), \ldots, X_{j-1}(i)$, $X_{j+1}(i), \ldots, X_d(i)$ can be considered as 'nuisance variables' in determining whether or not to stop at time i, whatever the dependence structure within the vector $\boldsymbol{X}(i)$ is. Thus $t_j \in T_n$ satisfies

$$\max_{t \in T_n} EX_j(t) = EX_j(t_j).$$

Hence the rule maximizing $\max_{1 \le j \le d} EX_j(t)$ is just t_{j^*} where

$$j^* = \arg\max\{EX_j(t_j) : 1 \le j \le d\}.$$

If the maximum is not unique, any fixed j^* for which it is attained will do. The above statement is equivalent to the fact that the supporting hyperplane to S_n though a maximizing $\boldsymbol{s}^*$ for this problem is parallel to the other axes (thus $Z_i = X_{j^*}(i)$), and that this hyperplane has largest intercept.

This leads to a ‘strange’ phenomenon. Suppose $\boldsymbol{X}(i)$ is exchangeable, $i = 1, \ldots, n$, and to fix ideas let $d = 2$ and $X_j(i)$ be i.i.d. uniform on [0, 1]. Then clearly any choice of j^* will do. This choice must however be made *a priori*, and the optimal stopper cannot ‘change his mind’ along the way. Suppose for example that the choice $j^* = 1$ is made, and that one has not stopped before time i, at which $(X_1(i) = 0.001, X_2(i) = 0.999)$ is observed. Acting optimally the stopper must ignore the 0.999 and continue to stage $i + 1$! Basing the ‘choice of j^*’ on the first observation $\boldsymbol{X}(1)$ cannot improve matters either. For example, with $n = 2,\ d = 2$ and uniform [0, 1] observations $\max_t EX_j(t) = 5/8,\ j = 1, 2$, while picking the random j^* for which $\max\{X_1(1), X_2(1)\}$ is observed, and stopping when this maximum exceeds 1/2 yields $EX_1(t) = EX_2(t) = 9/16$.

This example illuminates the difference between the goal of maximizing $Eh(\boldsymbol{X}(t))$ and $h(E\boldsymbol{X}(t))$. The optimal rule for the former would clearly be to stop when

$$Y_i = \max(X_1(i), X_2(i)) \ge c_i,$$

where the c_i-sequence is defined through (2.3) for Z_i there replaced by Y_i.

4. Remarks and generalizations

Remark 4.1. *Asymptotic comparison of $h(E\boldsymbol{X}(t))$ and $Eh(\boldsymbol{X}(t))$.* The difference between the problem of maximizing $Eh(\boldsymbol{X}(t))$ and that of maximizing $h(E\boldsymbol{X}(t))$ sometimes becomes even more apparent when one lets the horizon, n, tend to infinity. For simplicity, assume that $\boldsymbol{X}, \boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ are i.i.d., having a compact support $B \subset \mathbb{R}^d$, and that h is continuous. Then clearly

$$\lim_{n\to\infty} \sup_{t_n \in T_n} Eh(\boldsymbol{X}(t_n)) = \max_{\boldsymbol{x}\in B} h(\boldsymbol{x}) \tag{4.1}$$

For this situation, rates of convergence in (4.1) can be obtained by using the results of Kennedy and Kertz (1991) (see also Assaf and Samuel-Cahn (1998)).

The corresponding limit for $h(E\boldsymbol{X}(t))$ is given in the following result.

Proposition 4.1. *Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ be i.i.d. with compact support B, and let $C(B)$ be its convex hull. If h is continuous then*

$$\lim_{n\to\infty} \sup_{t_n \in T_n} h(E\boldsymbol{X}(t_n)) = \max_{\boldsymbol{x}\in C(B)} h(\boldsymbol{x}). \tag{4.2}$$

Thus, for example, when B is convex the two limits (4.1) and (4.2) will be equal, but clearly the value in (4.2) is generally larger than that of (4.1). The following simple example enhances the difference.

Example 4.1. Let $d = 2$, and let $\boldsymbol{X} = (X_1, X_2)$ take the three values (0, 0), (10, 0) and (0, 20) with probability 1/3 each. Let $h(\boldsymbol{x}) = x_1^9 x_2$. Since $h(\boldsymbol{X})$ is 0 a.s., the value in (4.1) is 0. Here $C(B)$ is a triangle in the plane, and the maximum value of any monotone increasing h is attained at some point(s) on the edge $\{x_2 = 20 - 2x_1,\ 0 \le x_1 \le 10\}$. For $h(\boldsymbol{x}) = x_1^9 x_2$ the maximum is attained at $(x_1, x_2) = (9, 2)$ with value $9^9 \times 2$, which is the value of (4.2).

Proof of Proposition 4.1. Clearly $E\boldsymbol{X}(t_n) \in C(B)$, for all $t_n \in T_n$ and all n, thus the left hand side in (4.2) is no larger than the right hand side. To show equality, it suffices to show that for every $\varepsilon > 0$ there exist n and t_n such that

$$h(E\boldsymbol{X}(t_n)) > \max_{\boldsymbol{x} \in C(B)} h(\boldsymbol{x}) - \varepsilon. \tag{4.3}$$

Let $\boldsymbol{x}^* \in C(B)$ satisfy

$$h(\boldsymbol{x}^*) = \max_{\boldsymbol{x} \in C(B)} h(\boldsymbol{x}) \tag{4.4}$$

and let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m)}$ $(1 \le m \le d+1)$ be m distinct points in B for which $\boldsymbol{x}^* = \sum_{k=1}^m \alpha_k \boldsymbol{x}^{(k)}$, with $0 < \alpha_k \le 1$ and $\sum_{k=1}^m \alpha_k = 1$. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ set $|\boldsymbol{x} - \boldsymbol{y}| = \max_{1 \le j \le d} |x_j - y_j|$ and write $\boldsymbol{x} \le \boldsymbol{y}$ iff $x_j \le y_j$ for all $j = 1, \ldots, d$. Let $0 < \delta < \frac{1}{2} \min_{k \ne \ell} |\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(\ell)}|$ and write $p_k = P(|\boldsymbol{X} - \boldsymbol{x}^{(k)}| < \delta)$. The upper bound on δ is chosen so as to make the events $\{|\boldsymbol{X} - \boldsymbol{x}^{(k)}| < \delta\}$ disjoint, $k = 1, \ldots, m$. Thus $\sum_{k=1}^m p_k \le 1$. Since $\boldsymbol{x}^{(k)} \in B$ where B is the support of $\boldsymbol{X}$, it follows that $p_k > 0$, $k = 1, \ldots, m$. Let c be such that $\gamma_k = c\alpha_k / p_k$ satisfies $0 < \gamma_k \le 1$ for $k = 1, \ldots, m$ and $0 < c \le 1$. For $n \ge 2$ define t_n as follows. For $i = 1, \ldots, n-1$ let $A_k(i)$ denote the event $\{|\boldsymbol{X}(i) - \boldsymbol{x}^{(k)}| < \delta\}$, and define

$$\phi_i(\boldsymbol{X}(i)) = \begin{cases} \gamma_1 & \text{if } A_1(i) \text{ occurs} \\ \vdots & \qquad \vdots \\ \gamma_m & \text{if } A_m(i) \text{ occurs} \\ 0 & \text{if none of the events } A_k(i) \text{ occur}, \qquad k = 1, \ldots, m. \end{cases} \tag{4.5}$$

Note that (4.5) implies that the conditional probability that t_n stops at time i, conditional on not having stopped earlier, is c, and, for $1 \le i \le n-1$

$$P(|\boldsymbol{X}(i) - \boldsymbol{x}^{(k)}| < \delta \mid t_n = i) = \alpha_k, \qquad k = 1, \ldots, m.$$

Let $\Delta = (\delta, \ldots, \delta)$. Then

$$\begin{aligned} \left\{ \sum_{k=1}^m \alpha_k (\boldsymbol{x}^{(k)} - \Delta) \right\} (1 - (1-c)^{n-1}) + (1-c)^{n-1} E\boldsymbol{X} < E\boldsymbol{X}(t_n) \\ < \left\{ \sum_{k=1}^m \alpha_k (\boldsymbol{x}^{(k)} + \Delta) \right\} (1 - (1-c)^{n-1}) + (1-c)^{n-1} E\boldsymbol{X} \end{aligned} \tag{4.6}$$

and for every fixed δ and $c = c(\delta)$ the left and right hand sides of (4.6) tend to $\boldsymbol{x}^* - \Delta$ and $\boldsymbol{x}^* + \Delta$ respectively, as $n \to \infty$. Since h is continuous at $\boldsymbol{x}^*$ one can determine δ such that $|h(\boldsymbol{x}) - h(\boldsymbol{x}^*)| < \varepsilon$ for all $|\boldsymbol{x} - \boldsymbol{x}^*| < 2\delta$. This implies, by (4.6), that (4.3) holds for all n sufficiently large.

Similar results hold for more general situations, such as non-identically distributed random variables.

It should be remarked that though the value in (4.1) is always smaller than or equal to the value in (4.2), a reverse inequality may hold for finite n, when considering the corresponding optimal rules. See for example Table 1 in Assaf and Samuel-Cahn (1998).

Remark 4.2. *Complexity of computation*. The main difficulty in finding an optimal t for the general case, is determining the d constants of the vector $\boldsymbol{a} = (a_1, \dots, a_d)$. The thresholds $c_1, \dots, c_{n-1}$ in (2.4) are then given in (2.3). Thus ignoring the additional complexity involved in the possible need for randomization, one only has to determine $d + n - 1$ constants. This is rather surprising when considering the abundance of rules in T_n.

Remark 4.3. *Infinite horizon*. All results can be extended to the case of infinite horizon, where one deals with an infinite sequence $\boldsymbol{X}(1), \boldsymbol{X}(2), \dots,$ of independent random vectors. In the definitions of the set T_∞ of all stopping rules for this case, the requirement $P(t \le n) = 1$ is replaced by $P(t < \infty) = 1$, i.e. one must stop with probability 1. For bounded $h \in H$ and $\boldsymbol{X}(i)$ such that $E(\sup_i \boldsymbol{X}(i)) < \infty$ the optimal t is now optimal for the one-dimensional infinite sequence of $Z_i = \langle \boldsymbol{a}, \boldsymbol{X}(i) \rangle$, $i = 1, 2, \dots,$ and with the previous exchangeability assumptions and $h \in \hat{H}$, for $Z_i = \sum_{j=1}^d X_j(i)$. The main differences are that: (i) in this case one cannot use backward induction, and must instead resort to the general theory of optimal stopping for the 'one dimensional' case with infinite horizon, and (ii) the existence of an optimal rule is not ensured. See for instance Chow *et al.* (1971), Chapter 5.

Remark 4.4. *The dependent case*. Most of the results given in Section 2 and in Remark 4.3 can be generalized to the arbitrarily dependent case, i.e. where $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ are dependent random variables, with finite expectations. S_n is again compact and convex, and the proofs are only slightly more involved. The parallel to Theorem 2.4 now is that there exists a vector of constants $\boldsymbol{a}$ such that if we define $Z_i = \langle \boldsymbol{a}, \boldsymbol{X}(i) \rangle$ then the rule t^* optimal for $h(E\boldsymbol{X}(t))$ is a rule optimal for the sequence $Z_1, \dots, Z_n$. Note however, that with arbitrary dependence structure the optimal rule for the Z_i's is *not* a threshold rule, but rather is defined by backward induction as in Theorem 3.2 on p. 50 of Chow *et al.* (1971). Parallel to Corollary 2.1 we have that if $h \in \hat{H}$, and the $\boldsymbol{X}(i)$'s are exchangeable, Z_i is again defined through $Z_i = \sum_{j=1}^d X_j(i)$, where here exchangeability refers to permutation of the d coordinates in the joint distribution of $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$. For details, see Glickman (1999).

Remark 4.5. *The 'optimal choice set'*. The 'optimal choice set' problem is defined as follows. For a given random variable $\boldsymbol{X}$, and a given α, $0 < \alpha < 1$, let Γ_α denote the set of all critical functions $\phi(\boldsymbol{x})$, $0 \le \phi \le 1$, such that $E\phi(\boldsymbol{X}) = \alpha$. For given h the problem is that of finding

$$\sup_{\phi \in \Gamma_\alpha} h(E(\boldsymbol{X}\phi(\boldsymbol{X}))) \tag{4.7}$$

For monotone h the optimal solution is always a half-space, i.e. of the form

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \langle \boldsymbol{a}, \boldsymbol{x} \rangle > b \\ 0 & \text{if } \langle \boldsymbol{a}, \boldsymbol{x} \rangle < b, \end{cases} \tag{4.8}$$

where randomization may be required on $\langle \boldsymbol{a}, \boldsymbol{x} \rangle = b$. This can be shown by exactly the same method as the proofs in Section 2. For the particular case where $\boldsymbol{X}$ is exchangeable and $h \in \hat{H}$ one has $\boldsymbol{a} = (1, \dots, 1)$. When $\boldsymbol{X}$ is discrete randomization will usually be required.

5. Multivariate prophet considerations

In the usual univariate setting, for a given sequence $Y_1, \dots, Y_n$ of random variables the optimal stopper seeks $V_{\mathrm{s}}^n = \sup_{t \in T_n} EY_t$, while the prophet can observe the entire Y-sequence before he makes his choice, and thus the value to the prophet is $V_{\mathrm{p}}^n = E(\max(Y_1, \dots, Y_n))$. Classical prophet inequalities compare V_{s}^n and V_{p}^n either in terms of their difference or ratio. See Hill and Kertz (1992) for an excellent review. The best-known prophet inequality is the one which states that for any independent non-negative Y_i's the inequality $V_{\mathrm{p}}^n < 2V_{\mathrm{s}}^n$ always holds, and the constant 2 cannot be improved upon, for all $n \geq 2$.

In the present multivariate setting we must first define the collection of all available strategies to the prophet. Since he observes $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ prior to selection, the collection of non-randomized decision rules G_n^0 for the prophet, is the set of measurable functions g defined by

$$G_n^0 = \{g;\, g : (\mathbb{R}^d)^n \to \{1, \dots, n\}\}.$$

The interpretation is that for a given g the prophet selects the index $g(\boldsymbol{X}(1), \dots, \boldsymbol{X}(n))$ after observing $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$. The set of randomized rules, G_n, for the prophet, is defined in a similar manner, except that instead of selecting one particular index after observing $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$, the prophet may now select, for each $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$, a probability distribution on $\{1, \dots, n\}$. The (random) return to the prophet is $\boldsymbol{X}(g) = \sum_{i=1}^n \boldsymbol{X}(i) I(g(\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)) = i)$, with corresponding expected value $E\boldsymbol{X}(g)$. Analogously to the definition of S_n define

$$P_n = \{E\boldsymbol{X}(g);\, g \in G_n\}.$$

Analogously to Theorems 2.1 and 2.2 one can show that P_n is compact and convex. In the multivariate setting, for a given h, the value to the prophet is therefore

$$\sup_{g \in G_n} h(E\boldsymbol{X}(g)) = \sup_{\boldsymbol{x} \in P_n} h(\boldsymbol{x}) \tag{5.1}$$

and any g for which this supremum is attained is called optimal for the prophet.

The solution for the prophet is obtained analogously to that for the statistician. When $h \in H$ the supremum in (5.1) is attained at some boundary point $\boldsymbol{x}^*$ of P_n. Denote the supporting hyperplane to P_n at $\boldsymbol{x}^*$ by $\langle \boldsymbol{b}, \boldsymbol{x} \rangle = c$. Then an optimal rule for the prophet is

$$g^*(\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)) = \arg\max\{\langle \boldsymbol{b}, \boldsymbol{X}(i) \rangle;\, i = 1, \dots, n\}$$

if the maximum of $\langle \boldsymbol{b}, \boldsymbol{X}(i) \rangle$ is unique. Randomization among all integers for which the maximum is attained may be required, in the case of non-uniqueness.

As an example where randomization for the prophet is required, consider again Example 3.2. There are 4 equally likely outcomes for $\boldsymbol{X}(1), \boldsymbol{X}(2)$, hence $2^4 = 16$ non-randomized rules. It is easily seen that for this example all 16 non-randomized rules yield one of the three points (3/2, 3/2), (5/4, 7/4) or (7/4, 5/4). Actually here $P_2 = S_2$ and the optimal value of $E\boldsymbol{X}(g)$ for the h considered is attained at $E\boldsymbol{X}(g^*) = (18/11, 15/11)$, and thus corresponds to a randomized rule, g^*.

For a given h one will usually have $\boldsymbol{a} \neq \boldsymbol{b}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ denote the coefficients of the hyperplanes for the statistician and the prophet, respectively. A notable exception is the following analogue to Corollary 2.1.

Theorem 5.1. *Let $\boldsymbol{X}(i)$ be exchangeable, $i = 1, \dots, n$, and suppose $h \in \hat{H}$. Let $Z_i = \sum_{j=1}^{d} X_j(i)$. Then an optimal strategy for the prophet is*

$$g^*(\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)) = \arg\max\{Z_i, i = 1, \dots, n\},$$

where randomization among all indices for which the maximum of Z_i is achieved may be required, when this maximum is not unique.

The following example exhibits another aspect of the difference between $Eh(\boldsymbol{X}(t))$ and $h(E\boldsymbol{X}(t))$, and obtains $\sup_g h(E\boldsymbol{X}(g))$ and $\sup_g Eh(\boldsymbol{X}(g))$, which is the prophet value in the classical setting.

Example 5.1. Let $n = d = 2$, and let $\boldsymbol{X}(1)$ and $\boldsymbol{X}(2)$ be independent and identically distributed, taking on each of the following values with probability 1/8; (0, 12), (12, 0), (1, 10), (10, 1), (2, 8), (8, 2), (3, 6), (6, 3). Take $h(\boldsymbol{x}) = x_1 x_2$. Note that for this example the ordering of the points by the magnitude of the sum $x_1 + x_2$ (needed when maximizing EZ_t of Corollary 2.1) is exactly the opposite ordering to that obtained when ordering them by the magnitude of $x_1 x_2$ (needed when maximizing $Eh(\boldsymbol{X}(t))$). Since $EZ_i = 21/2$ the optimal t^* will stop at time 1 iff $\boldsymbol{X}(1)$ is one of the values (0, 12), (12, 0), (1, 10), (10, 1), and thus $EX_1(t^*) = EX_2(t^*) = 11/2$, yielding $\sup_t h(E\boldsymbol{X}(t)) = (11/2)^2$. For this example, when maximizing $Eh(\boldsymbol{X}(t))$ the optimal $\tilde{t}$ stops at time 1 if and only if $\boldsymbol{X}(1)$ is one of the values (2, 8), (8, 2), (3, 6), (6, 3) i.e. exactly when t^* does not stop! The value is $Eh(\boldsymbol{X}(\tilde{t})) = 14$. Here the prophet value is $\sup_g h(E\boldsymbol{X}(g)) = E(\max_{i=1,2} Z_i)^2/4 = (89/16)^2$, while the prophet value for the classical problem is $E(\max\{X_1(1)X_2(1), X_1(2)X_2(2)\}) = 59/4$.

We next focus on the possibility of constructing prophet inequalities in the present multivariate setting. To abbreviate notation set

$$V_{\text{s}}^{(n)}(h) = \sup_{t \in T_n} h(E\boldsymbol{X}(t)) \quad \text{and} \quad V_{\text{p}}^{(n)}(h) = \sup_{g \in G_n} h(E\boldsymbol{X}(g)).$$

Then a ratio prophet inequality in the present setting would claim the existence of some constant c_n, such that

$$V_{\text{p}}^{(n)}(h) \leq c_n V_{\text{s}}^{(n)}(h) \tag{5.2}$$

holds for a sufficiently large class of $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ and a sufficiently large class of functions h.

We shall first show that (5.2) cannot hold uniformly in non-negative h, even for independent non-negative random variables, and $d = 1$, i.e. for the univariate case.

Example 5.2. Consider $d = 1, n = 2$ and the classical example where $X(1) \equiv \mu$ and $X(2)$ equals 1 and 0 with probabilities μ and $1-\mu$ respectively, where $0 < \mu < 1$. Here $EX(t) = \mu$ for all $t \in T_2$ and $\sup_{g \in G_2} EX(g) = 2\mu - \mu^2$, and thus these values (and corresponding strategies) for the statistician and prophet will be optimal for all increasing functions h. (These random variables are used in showing that the constant '2' mentioned in the first paragraph of the present section, cannot be improved upon; cf. Hill and Kertz (1981).) Now let $h(x) = \max(0, x^k)$. Then $V_{\text{p}}^{(2)}(h)/V_{\text{s}}^{(2)}(h) = (2\mu - \mu^2)^k/\mu^k \to 2^k$ as $\mu \to 0$, and thus the constant c_2 in (5.2) for this h is at least $2^k \to \infty$ as $k \to \infty$.

The next example shows, again for $d = 1$, that even for a fixed, continuous, monotone h (5.2) cannot hold for all non-negative independent $X(i)$, even for $n = 2$.

Example 5.3. Let $h(x) = \mathrm{e}^x$, $d = 1$ and $n = 2$, and take $X(1) \equiv \mu$, $X(2)$ equals 2μ and 0, each with probability 1/2. Again $EX(t) = \mu$ for all t while $\sup_g EX(g) = 3\mu/2$, and $V_\mathrm{p}^{(2)}(h)/V_\mathrm{s}^{(2)}(h) = \mathrm{e}^{\mu/2} \to \infty$ as $\mu \to \infty$.

Under more restrictive assumptions one can still obtain interesting ratio prophet inequalities. Assume for example that $\boldsymbol{X}(i)$ are exchangeable, $i = 1, \dots, n$ with $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ independent and non-negative and that $h \in \hat{H}$. Let t^* and g^* be optimal policies for the statistician and prophet, based on the sequence of the $Z_i = \sum_{j=1}^d X_j(i)$ only, as described in Corollary 2.1 and Theorem 5.1 respectively. For such t^* we have seen that $EX_1(t^*) = \cdots = EX_d(t^*)$ and similarly one has $EX_1(g^*) = \cdots = EX_d(g^*)$.

Now by the univariate prophet inequality we have $EZ(g^*) < 2EZ(t^*)$, i.e.

$$E\left(\sum_{j=1}^d X_j(g^*)\right) < 2E\left(\sum_{j=1}^d X_j(t^*)\right)$$

which, since all terms in each of the summands are equal, implies

$$EX_j(g^*) < 2EX_j(t^*) \qquad \text{for } j = 1, \dots, d. \tag{5.3}$$

(Note that (5.3) holds even though neither t^* nor g^* need be optimal for the jth coordinate sequence $X_j(1), \dots, X_j(n)$.) Since h is monotone non-decreasing, (5.3) implies

$$V_\mathrm{p}^{(n)}(h) = h(E\boldsymbol{X}(g^*)) \leq h(2E\boldsymbol{X}(t^*)). \tag{5.4}$$

Though the right hand side is not expressed in terms of $V_\mathrm{s}^{(n)}(h)$, it is closely related to this value. For example for $\tilde{h}(\boldsymbol{x}) = \prod_{j=1}^d x_j$, as in Example 3.1, (5.4) yields the prophet inequality

$$V_\mathrm{p}^{(n)}(\tilde{h}) < 2^d V_\mathrm{s}^{(n)}(\tilde{h}) \tag{5.5}$$

for all exchangeable non-negative $\boldsymbol{X}(i)$, $i = 1, \dots, n$, with independent $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$, which is a direct generalization of the classical prophet inequality obtained for $d = 1$. For the example $\hat{h}(\boldsymbol{x}) = \min\{x_1, \dots, x_d\}$ considered in Example 3.3, (5.4) yields the classical constant '2' for all d, i.e.

$$V_\mathrm{p}^{(n)}(\hat{h}) < 2V_\mathrm{s}^{(n)}(\hat{h}) \qquad \text{for all } d. \tag{5.6}$$

Acknowledgement

The authors are grateful to Dr. Uwe Saint-Mont for his thorough reading and valuable comments on an earlier version of the paper.

References

Assaf, D. and Samuel-Cahn, E. (1998). Optimal cooperative stopping rules for maximization of the product of expected stopped values. *Stat. Prob. Lett.* **38**, 89–99.

Bassan, B. and Scarsini, M. (1995). On the value of information in multi-agent decision theory. *J. Math. Econ.* **24**, 557–576.

Chow, Y. S., Robbins, H. and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin, Boston.

Glickman, H. (1999). Cooperative stopping rules in multivariate problems. Ph.D. Thesis, The Hebrew University of Jerusalem.

HILL, T. P. AND KERTZ, R. P. (1981). Ratio comparisons of supremum and stop rule expectations. *Z. Wahrscheinlichkeitsth.* **56**, 283–285.

HILL, T. P. AND KERTZ, R. P. (1992). *A survey of prophet inequalities in optimal stopping theory.* Contemporary Mathematics **125**, American Mathematical Society, Providence, RI, pp. 191–207.

KENNEDY, D. P. AND KERTZ, R. P. (1991). The asymptotic behavior of the reward sequence in the optimal stopping of i.i.d. random variables. *Ann. Prob.* **19**, 329–341.

LEHMANN, E. L. (1991). *Testing Statistical Hypotheses*, 2nd edn. Wadsworth & Brooks / Cole, Pacific Grove, CA.

MAMER, J. W. (1987). Monotone stopping games. *J. Appl. Prob.* **24**, 386–401.

MARSHALL, A. W. AND OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

YASUDA, M., NAKAGAMI, J. AND KURANO, M. (1982). Multivariate stopping problems with a monotone rule. *J. Operat. Res. Soc. Japan* **25**, 334–349.