

NOTE: A TIME-SEQUENTIAL GAME FOR SUMS OF RANDOM VARIABLES

Minoru Sakaguchi
Osaka University

(Received August 22, 1979)

Abstract A two-person zero-sum time-sequential game for sums of iid random variables, in which decision for the players at each stage is either to accept or reject sequentially presented r.v.'s, each r.v. is selected only when both players accept it, and each player has a limited number of rejections he can exercise. The algorithm for deriving the optimal strategies is given and a numerical calculation is provided for the case of uniform distribution. The bilateral-game version, in which decisions are made by each side one by one, is also discussed and is shown to possess "reversibility" of the optimal strategies. These results clarify some earlier works in this area.

1. Introduction and Purpose

In a recent paper [2] Brams and Davis consider the following problem : Prosecution and defense (players I and II, respectively) must choose J jurors. Potential jurors with a priori probabilities of voting for conviction are assumed to be chosen randomly from a population and to come up one at a time for decision. Let X_i , $i=1,2, \dots$, be iid random variables, each assuming a real value $0 \leq x_i \leq 1$, interpreted as the i -th potential juror's probability of voting for conviction. As each potential juror comes up, players I and II must decide, simultaneously and independently of the other player's choice, whether to accept or challenge him according to the observed value he possesses. If both players accept, the i -th potential juror becomes a member of the jury ; if either player challenges, he is rejected, in which case a new potential juror comes up for challenge. The number of peremptory challenges for each side is limited, as A for I and B for II. The process continues until J jurors are selected. The problem is to find the strategy for I and II which maximizes for I, and minimizes for II, the expected payoff of

the game $E \prod X_{i'}$, where i' 's represent all the jurors so far selected. In [2] the algorithm for deriving optimal strategies and values in this time-sequential zero-sum game, is given, and some numerical calculations of them are provided.

The purpose of this note is to present the solution algorithm of the alternative version of the game in which the payoff is defined by the sum $\sum X_{i'}$, instead of the product $\prod X_{i'}$. Although the difference of the two versions is seemingly slight, the study of our version is not valueless since :

- (1) This class of time-sequential games for sums of random variables is often found in our real life in which player II gives a fixed number, J , of offers to player I in a given horizon. The pairs of players II-I, are, for example, father-sons, husband-wife, bank-entrepreneur, government-corporation, etc.
- (2) It provides an important generalization of the now classical work by Gilbert and Mosteller [4] and gives a foresight to the further work in the line of works [5] and [6].

2. Solution of the Zero-Sum Game for the Sum

Let $X_{i'}$, $i'=1,2, \dots$, be iid random variables that can be observed one by one sequentially. Players I and II are asked to select a set of r observations. The common cdf $F(x)$ of each $X_{i'}$ is assumed to be known by both players. Let $X_{i'}$, $i'=1, \dots, r$, be the set of r selected observations, and suppose that the objective of I(II)'s sequential decisions is to maximize (minimize) the expected value $E(\sum_{i'=1}^r X_{i'})$, of the sum of the observations they have selected.

Let $\sigma = (j, a, b)$ denote the state of the process in which j r.v.'s remain to be selected and the players I and II have currently a and b times of rejection, respectively, still remaining. Also let $\alpha = (j, a-1, b)$, $\beta = (j, a, b-1)$, $\gamma = (j-1, a, b)$ and $\delta = (j, a-1, b-1)$. Then these states represent the four possible states that can be reached (in principle at least) from σ by the decision pair rej-acc, acc-rej, acc-acc, and rej-rej, respectively. Let $V(\sigma)$ be the value of the game in state σ . Also let $V(\sigma|x)$ be the conditional value of the game in state σ given that the observation x is currently observed. Then we have

$$(2.1) \quad V(\sigma) = EV(\sigma|X) \quad ,$$

$$(2.2) \quad V(\sigma|x) = \text{val} \left\{ \begin{array}{l} \text{rej} \\ \text{acc.} \end{array} \left[\begin{array}{cc} \text{rej} & \text{acc} \\ V(\delta) & V(\alpha) \\ V(\beta) & x+V(\gamma) \end{array} \right] \right\} \quad ,$$

where val A denotes the value of the (two-person zero-sum) matrix game A. Now the algorithm embodied in the following theorem allows us to calculate the optimal strategies for the players and the value, $V(\sigma)$, recursively. First we prepare a lemma.

Lemma 1. If both players play optimally, they will never both reject the same observation.

Proof. Let $c_1 = V(\alpha) - V(\gamma)$ and $c_2 = V(\beta) - V(\gamma)$. Then we have $c_1 \leq c_2$, since we may suppose that, for any history of the process, the relation expressed by the inequalities

$$V(j, a-1, b) \leq V(j, a-1, b-1) \leq V(j, a, b-1), \text{ i.e. } V(\alpha) \leq V(\delta) \leq V(\beta).$$

are valid. That is, player I stands advantageous if one of II's available rejection is transferred to I for his own use. (This assumption seems to be reasonable, but unfortunately the conjecture that this can be proven by induction is very roundabout to be ascertained. For this point, see [5; p.907], [3] and [6; p.504])

Basing on these inequalities it follows that the 2×2 matrix in the right-hand side of (2.2) has saddle points at

- (1,2) element, i.e rej-acc, if $0 \leq x \leq c_1$;
- (2,2) element, i.e acc-acc, if $c_1 < x \leq c_2$;
- (2,1) element, i.e acc-rej, if $x > c_2$.

Hence we conclude that if both players play optimally, they will never both reject the same observation.

Theorem 1. (i) $V(\sigma)$ satisfies the recursive relations

$$(2.3) \quad V(\sigma) = V(\beta) - \int_{c_1}^{c_2} F(x)dx,$$

where $c_1 = V(\alpha) - V(\gamma)$ and $c_2 = V(\beta) - V(\gamma)$. The optimal strategy in state σ for I(II) is to reject the observation if and only if $x \leq c_1$ ($\geq c_2$).

(ii) The boundary conditions for (2.3) are given by :

- For $j = 0$, $V(0, a, b) = 0$
- For $a = 0$, $V(j, 0, b) = v^{**}(j, b)$
- For $b = 0$, $V(j, a, 0) = v^*(j, a)$,

where the two functions $v^*(j, a)$ and $v^{**}(j, b)$ are determined by

$$(2.5) \quad v^*(j, a) = v^*(j, a-1) + T_F(v^*(j, a-1) - v^*(j-1, a)) \quad , \\ (j \geq 1 ; \quad v^*(0, a) = 0, v^*(j, 0) = jEX) .$$

$$(2.6) \quad v^{**}(j, b) = v^{**}(j-1, b) + EX - T_F(v^{**}(j, b-1) - v^{**}(j-1, b)) . \\ (j \geq 1 ; \quad v^{**}(0, b) = 0, v^{**}(j, 0) = jEX) .$$

and the function $T_F(z)$ is defined by $E_F[(X - z)^+]$.

Note that $v^*(v^{**})$ represents the value of the one-person game in which only player I(II) has some number of rejections left.

Proof Applying Lemma 1 to (2.1) - (2.2) we have

$$(2.7) \quad V(\sigma) = \int_{-\infty}^{\infty} V(\sigma | x) dF(x) = F(c_1)V(\alpha) + \int_{c_1}^{c_2} (x+V(\gamma))dF(x) + \bar{F}(c_2)V(\beta) .$$

where $\bar{F}(c_2) = 1 - F(c_2)$. We can simplify the middle term on the right-hand

side by integration by parts : $\int_{c_1}^{c_2} x dF(x) = c_2 F(c_2) - c_1 F(c_1) - \int_{c_1}^{c_2} F(x) dx$.

After substituting this expression into (2.7), we obtain the equation for $V(\sigma)$ given by (2.3).

To prove the second part of the theorem, consider the cases where the first column or row of the matrix game in (2.2) disappears.

Then we have

$$v^*(j, a) = E[v^*(j, a-1) \vee (X+v^*(j-1, a))] \\ \text{and} \quad v^{**}(j, b) = E[v^{**}(j, b-1) \wedge (X+v^{**}(j-1, b))] ,$$

respectively. Here we have used the notations $u \vee v = \max(u, v)$ and $u \wedge v = \min(u, v)$. We obtain (2.5) from the first equation, and (2.6) from the second, since, for any real numbers k and ℓ , we have

$$\int (k \vee (x+\ell)) dF(x) = k + T_F(k-\ell) ,$$

and

$$\int (k \wedge (x+\ell)) dF(x) = \ell + EX - T_F(k-\ell) .$$

This completes the proof of the theorem.

Table 1 gives the values of $V(\sigma)$, for $j = 1(1)3$ and $a, b = 0(1)7$, computed from (2.3) ~ (2.6), in the case of uniform distribution : $F(x) = x, 0 \leq x \leq 1$, and $T_F(z) = \frac{1}{2}(1-z)^2$. The values on the upper half of the table are omitted since we have, for any $a, b = 0, 1, 2, \dots$,

$$V(j, a, b) + V(j, a', b') = j, \quad \text{if } a' = b \text{ and } b' = a .$$

Table 1. Game Values $V(j, a, b)$ for Uniform Distribution*

b	0	1	2	3	4	5	6	7
a=0	.5000 1.0000 1.5000							
1	.6250 1.1953 1.7417	.5000 1.0000 1.5000						
2	.6953 1.3203 1.9091	.5786 1.1340 1.6757	.5000 1.0000 1.5000					
3	.7417 1.4091 2.0341	.6340 1.2337 1.8115	.5580 1.1031 1.6393	.5000 1.0000 1.5000				
4	.7751 1.4761 2.1318	.6757 1.3115 1.9203	.6031 1.1856 1.7534	.5462 1.0842 1.6159	.5000 1.0000 1.5000			
5	.8004 1.5287 2.2105	.7084 1.3741 2.0099	.6393 1.2534 1.8487	.5842 1.1544 1.1742	.5385 1.0712 1.5994	.5000 1.0000 1.5000		
6	.8203 1.5712 2.2756	.7347 1.4258 2.0851	.6692 1.3102 1.9299	.6159 1.2142 1.7988	.5712 1.1325 1.6858	.5331 1.0619 1.5871	.5000 1.0000 1.5000	
7	.8364 1.6064 2.3303	.7565 1.4692 2.1492	.6943 1.3586 1.9999	.6429 1.2657 1.8724	.5994 1.1858 1.7617	.5619 1.1162 1.6643	.5290 1.0547 1.5776	.5000 1.0000 1.5000

* In each cell the number on the top, middle, and bottom corresponds to $j = 1, 2, 3$, respectively.

This fact is due to symmetry existing in both of the setting of our game and the uniform distribution of the random variable.

Remark 1. The computer calculations reported in Table 1 could be made more interesting by the addition of the numerical results for larger values of j , a and b . We can nevertheless have some observations which reflect some characteristic of the underlying time-sequential game supported by the data contained in Table 1. We shall list some of these observations as follows :

01. If j is small and $a = b$, the threshold points c_1 and c_2 approach $1/2$, as a becomes larger.

02. If $a = kb$, with $k > 1$, player I is more favored as a becomes larger.

03. If $a = kb$, with $k > 1$, then as k becomes larger, the threshold points c_1 and c_2 increase, and the acc-acc interval $(c_1, c_2]$ moves to the right with diminishing width.

04. If $a = kb$, with $k > 1$, then, as a becomes greater, ratios $V(2, a, b)/V(1, a, b)$ and $V(3, a, b)/V(1, a, b)$ become greater, i.e., the greater advantage a multi-choice game is to player I than a single-choice game.

Remark 2. Our problem is a generalization of the problem of optimal selection first discussed by Gilbert and Mosteller [4 : Sec. 5c], in which the objective of the single decision-maker (i.e., player I) is to maximize the sum of the r accepted values among a set of n sequentially arriving random variables. With $(j, a) = (r, n-r)$, (2.5) becomes

$$v^*(r, n-r) = v^*(r, n-r-1) + T_F(v^*(r, n-r-1) - v^*(r-1, n-r)), \\ (1 \leq r \leq n-1).$$

It is easy to show that $v^*(r, n-r) = \sum_{j=1}^r \mu_{n,j}$, and the optimal decision for I in state $\sigma = (r, n-r, 0)$ is to accept if and only if $x \geq \mu_{n-1,r}$, where the numbers $\mu_{n,r}$ are determined by

$$\mu_{n,r} = \mu_{n-1,r} + T_F(\mu_{n-1,r}) - T_F(\mu_{n-1,r-1}), \\ (2 \leq r \leq n-1; \sum_{j=1}^n \mu_{n,j} = n EX),$$

$$\mu_{n-1} = \mu_{n-1,1} + T_F(\mu_{n-1,1}), \quad (n \geq 2; \mu_{1,1} = EX).$$

Optimal values $v^*(r, n-r)$ for the uniform, exponential and normal distributions, for $r = 1(1)3$, $n = 1(1)10$, are given in Table 14 Of [4].

Similar arguments will be made from the side of the minimizer (i.e., player II). With $(j, b) = (r, n-r)$, (2.6) becomes

$$v^{**}(r, n-r) = v^{**}(r-1, n-r) + EX - T_F(v^{**}(r, n-r-1) - v^{**}(r-1, n-r)), \\ (1 \leq r \leq n-1),$$

It is again easy to show that $v^{**}(r, n-r) = \sum_{j=1}^r v_{n,j}$, and the optimal decision for II in state $\sigma = (r, 0, n-r)$ is to accept if and only if $x \leq v_{n-1,r}$, where the numbers $v_{n,r}$ are determined by

$$v_{n,r} = v_{n-1,r-1} + T_F(v_{n-1,r-1}) - T_F(v_{n-1,r}), \\ (2 \leq r \leq n-1, \sum_{j=1}^n v_{n,j} = n EX)$$

$$v_{n,1} = EX - T_F(v_{n-1,1}) \quad (n \geq 2 ; v_{1,1} = EX).$$

Remark 3. The limiting behaviors as $j+a+b \rightarrow \infty$, holding $j : a : b$ fixed, of the value $V(\sigma)$ and the threshold points c_1 and c_2 's would be worth to be studied in further research. For the one-person game version Albright and Derman [1] proved the result that as j and a approach infinity, holding $j/(j+a) = 1-\theta$ and $a/(j+a) = \theta$ fixed, where $0 < \theta < 1$, the average expected reward per choice under the optimal strategy approaches $\int_{F^{-1}(\theta)}^{\infty} x dF(x)$, where $F^{-1}(\theta)$ denotes the 100θ % quantile of the cdf $F(x)$. Hence the purpose of our future study mentioned above is to extend the above result in [1] to the two-person zero-sum game version.

3. The Bilateral Game and Reversibility

Consider the two-person bilateral-game version in which each round consists of two consecutive moves, first by player I, and then by II. In face of the observation x , player I must first decide whether to accept or reject it, and if he rejects it the next r.v. will be observed. If I accepts the observation, then II must next decide whether to accept the current observation or not, and if he rejects it the next r.v. is observed. Only when both players accept the observation, it is selected as one element of the set of r observations.

Let $U(\sigma)$ denote the value of the two-person bilateral game in state $\sigma = (j, a, b)$. Then by considering the consequences of the two possible decisions firstly by I, and then, if I accepts, by II, we can write

$$(3.1) \quad U(\sigma) = E[\max\{U(\alpha), \min(U(\beta), X+U(\gamma))\}] ,$$

since if I firstly accepts the observation X , then nextly II, under his optimal decision,

$$\left\{ \begin{array}{l} \text{rejects} \\ \text{accepts} \end{array} \right\} , \text{ according as } U(\beta) \left\{ \begin{array}{l} < \\ \geq \end{array} \right\} X + U(\gamma) .$$

Defining $d_1 \equiv U(\alpha) - U(\gamma)$ and $d_2 \equiv U(\beta) - U(\gamma)$, we may assume that the inequality $d_1 \leq d_2$ is valid just as we have done so in the proof of Lemma 1.

Subtracting $U(\gamma)$ from both sides of (3.1) we have

$$U(\sigma) - U(\gamma) = E[\max\{d_1, \min(d_2, X)\}]$$

$$= d_1 F(d_1) + \int_{d_1}^{d_2} x dF(x) + d_2 \bar{F}(d_2)$$

where $\bar{F}(d_2) = 1 - F(d_2)$. Applying integration by parts to the middle term of the right-hand side, we will finally obtain

$$(3.2) \quad U(\sigma) = U(\beta) - \int_{d_1}^{d_2} F(x) dx,$$

which is identical to (2.3). A check of the above arguments gives the fact that the optimal strategy in state σ for I(II) is to reject the observation if and only if $x \leq d_1$ ($> d_2$).

Let $u^*(j, a)$ ($u^{**}(j, b)$) represent the value of the one-person game in which only player I(II) has some number of rejections left. Then we can easily find that the boundary conditions for (3.2) are given again by (2.4) ~ (2.6), with V and v replaced by U and u , respectively. Moreover if we consider the case where player II makes the first decisions, we easily find, by the similar arguments to those used above, that the relations (2.3) ~ (2.6) still hold. Hence we have shown :

Theorem 2 The order in which the two sides exercise their rejections is irrelevant to both of them.

This fact, called "reversibility", is proven under the more general setting of the underlying decision process in [3 ; Theorems 2 and 5], so that our result serves as a reconfirmation of the main result of [3].

Acknowledgement. The author is thankful to Dr. Kensaku Kikuta for his computer work in preparing Table 1, and to the referee for his helpful comment.

References

- [1] C. Albright and C. Derman : Asymptotic optimal policies for the stochastic sequential assignment problem. *Manag. Sci.*, 19 (1972), 46-51.
- [2] S.J. Brams and M.D. Davis : Optimal jury selection : A game-theoretic model for the exercise of peremptory challenges. *Opns. Res.*, 26 (1978), 966-991
- [3] M.H. DeGroot and J.B. Kadane : Optimal challenges for selection. Carnegie-Mellon University Techn. Rep. no.143 (1978), pp.27.
- [4] J.P. Gilbert and F. Mosteller : Recognizing the maximum of a sequence, *J. Amer. Stat. Assoc.*, 61 (1966), 35-73.

- [5] R. Roth, J.B. Kadane and M.H. DeGroot : Optimal peremptory challenges in trials by juries : A bilateral sequential process, *Opns. Res.*, 25 (1977), 901-919.
- [6] M. Sakaguchi : A bilateral sequential game for sums of bivariate random variables. *J. Oper. Res. Soc. Japan*, 21 (1978), 486-508.

Minoru SAKAGUCHI
Department of Applied Mathematics
Faculty of Engineering Science
Osaka University
Toyonaka, Osaka, Japan.