Gry niesymetryczne ze stopowaniem
On some game with unbounded horizon

Marek Skarupski, Krzysztof Szajowski

15.07.2016

Wrocław University of Technology

Abstract

In this talk we investigate the game with two players who observe objects: random variables from uniform distribution. They observe them sequentially and decide whether to stop or continue observations. Three models are presented. The first model is focused on the case when the number of objects is random and comes from geometrical distribution. The second one admits various information about applicants available to the players. They are looking for equilibrium. In the third model the game is not symetric. The player with determined type is playing against the player of unknown type. Different types of the players means his various, random ability to adopt the observation.

Contents

Introduction

Suppose we observe sequentially $X_1, ..., X_N$ i.i.d. What is more we can assume that they are uniformly distributed on interval $[0, 1]$. We also consider a special case where N is a random variable geometrically distributed, i.e.

$$P(N = k) = p_k = pq^{k-1}, \quad 0 < p < 1; q = 1 - p, k = 1, 2, ... \tag{1}$$

Payoff function

$$\begin{aligned}w_1(n,x) =&P(\text{selecting the best on stage } (n,x))\\ =&P(X_{n+1} < x, ..., X_N < x, N = k | N \geq n)\\ =&\sum_{k=n}^{\infty} x^{k-n} \frac{p_k}{\pi_n} = p\sum_{k=n}^{\infty}(qx)^{k-n} = \frac{p}{1-qx}\end{aligned}$$

where $\pi_k = \sum_{j=k}^{\infty} p_j = q^{k-1}$. Therefore the probability of winning in the future

$$\begin{aligned}w_2(n,x) &= \int_x^1 \sum_{m=n+1}^{\infty} P(\text{selecting the best on stage } (m,y))p((n,x);(m,dy))\\ &= \int_x^1 \sum_{m=n+1}^{\infty} x^{m-n-1}\frac{\pi_m}{\pi_n}\frac{p}{1-qy}dy = -\frac{p}{1-qx}\log\left(\frac{p}{1-qx}\right)\end{aligned}$$

where $p((A),(B))$ is the transition probability from state A to state B.

$$p((n,x),(m,y)) = yx^{m-n-1}\frac{\pi_m}{\pi_n}, y > x.$$

Payoff matrix

		Player II	
		S	F
Player I	S	$\varphi(x)$	$\varphi(x)$
	F	$-\varphi(x)$	$v(x)$

where v denote the payoff earned by continuing observations in an optimal manner, and

$$\varphi(n,x) = w_1(n,x) - w_2(n,x) = \frac{p}{1-qx}\left(1 + \log\left(\frac{p}{1-qx}\right)\right) \qquad (2)$$

Equilibrium point

We will omit letter n in notation since the function does not depend on this parameter.

For big values of x let us say $x > x_0$ we have two conditions (since player I is maximizing his payoff, and player two is minimizing his loose) to (S,S) be a Nash equilibrium:

1. $\varphi(x) \geq -\varphi(x)$
2. $\varphi(x) \leq \varphi(x)$.

Second condition is satisfied all the time. From first we get, that $\varphi(x) \geq 0$. It leads to the condition

$$x \geq \frac{1-pe}{1-p} := x_0. \tag{3}$$

Equilibrium point

Let $x \geq x_0$. The player II is in the game alone (since the player I with priority has stopped). He will maximize his payoff. The optimal payoff provided that we start from the state $(n, X_n = x)$ leads to the recursive equation:

$$v(x) = q\left(xv(x) + \int_x^1 w(y)dy\right), \tag{4}$$

where $w(x) = \max\{\varphi(x), v(x)\}$. In this case

$$v(x) = \frac{q}{1-qx}\int_x^1 \frac{p}{1-qy}dy \tag{5}$$

It leads to the condition $\varphi(x) > 0$, but in this region it is satisfied all the time.

Equilibrium point

Let $x \le x_0$. Suppose that we are in the state $(n, X_n = x)$. The player I prefers to take an action F. We derive the best response for this strategy for the player II. If he stops he will earn $\varphi(x)$. If he continues observation in an optimal manner his future payoff is:

$$v(x) = \frac{q}{1-qx}\int_x^{x_0} v(y)dy - \frac{q}{1-qx}\int_{x_0}^1 \varphi(y)dy. \tag{6}$$

By taking the derivative of both sides we get that

$$v'(x) = 0 \Longrightarrow v(x) = const. \tag{7}$$

And from continuity $v(x) = v(x_0) = -\frac{q}{1-qx}\int_{x_0}^1 \varphi(y)dy = 0.5e^{-1}$.

Because for $x \le x_0$ we have $\varphi(x) \le v(x)$. Therefore the best response of the player II for strategy of the player I is also not to stop and continue the observation.

Optimal strategy

Optimal strategy

In the best choice game with random geometrical horizon with priority of player I the optimal strategy is

- (S,S) for $x \geq x_0$
- (F,F) for $x < x_0$.

The value of the game for player I is

- $v(x) = -0.5e^{-1}, x < x_0$
- $v(x) = \frac{p}{1-qx}\left(0.5\log^2\left(\frac{p}{1-qx}\right) + \log\left(\frac{p}{1-qx}\right)\right), x \geq x_0$

Question

How define the game model when Player I knows the a priori distribution of the number of applicant but the Player II has no such knowledge?

Description of the model

Full-information and no-information players

Consider a game in which two players want to choose the best object overall. They observe N objects sequentially. They get a profit only if the player chooses the best object and the rival will not find the better one. In other case he gets the award. If both players wants to stop on the current object the nature chooses it by the fair coin toss. Suppose that:

1. The player I is full information, i.e. he observes sequentially $X_1, ..., X_N$ i.i.d., sees its value, and also can calculate the rank of the current object.
2. The player II is no information, i.e. he observes only the relative ranks of the current objects:

$$Y_n = \#\{1 \leq i \leq n : X_i \leq X_n\}. \tag{8}$$

Player I payoff

The reward for the player I for stopping on nth object of the value $X_n = x$ is

$$s_{1,n}(x) = x^{N-n} \tag{9}$$

and for continuing observation is given by

$$\begin{aligned} c_{1,n}(x) &= \sum_{k=n+1}^{N} 1 \cdot p((n,x),(k,(x,1])) \\ &= \sum_{k=n+1}^{N} x^{k-n-1} \int_x^1 dx = (1-x) \sum_{k=n+1}^{N} x^{k-n-1} \\ &= 1 - x^{N-n}. \end{aligned} \tag{10}$$

Player II payoff

The reward for the player II for stopping on nth object is

$$s_{2,n} = \frac{n}{N} \tag{11}$$

and for continuing observations

$$c_{2,n} = \sum_{k=n+1}^{N} \frac{n}{k(k-1)} \frac{k}{N} = \frac{n}{N} \sum_{k=n+1}^{N} \frac{1}{k-1}. \tag{12}$$

Payoff matrix

Suppose that we are in some moment n and the value of the current object is x and both players want to stop. If the player I gets the object (with probability 0.5) he gets reward $s_{1,n}(x)$. With probability 0.5 the player II gets the object so I must continue the observations and gets reward $c_{1,n}(x)$ or award $-s_{1,n}(x)$. The payoff matrix for player I is given by

$$v_{1,n}(x) = \begin{array}{c|c|c} & S & F \\ \hline S & \frac{1-x^{N-n}}{2} & 2x^{N-n}-1 \\ \hline F & 1-2x^{N-n} & v_{1,n+1}(x) \end{array}$$

Similar consideration gives the matrix for the player II:

$$v_{2,n} = \begin{array}{c|c|c} & S & F \\ \hline S & \frac{n}{2N}\sum_{k=n+1}^{N}\frac{1}{k-1} & \frac{n}{N}\sum_{k=n+1}^{N}\frac{1}{k-1} - \frac{n}{N} \\ \hline F & \frac{n}{N} - \frac{n}{N}\sum_{k=n+1}^{N}\frac{1}{k-1} & v_{2,n+1} \end{array}.$$

Equilibrium point

Since both players want to maximixe their profits we have the following conditions) to (S,S) be a Nash equilibrium:

$$\frac{1 - x^{N-n}}{2} \geq 1 - 2x^{N-n} \tag{13a}$$

$$\frac{n}{2N} \sum_{k=n+1}^{N} \frac{1}{k-1} \geq \frac{n}{N} \sum_{k=n+1}^{N} \frac{1}{k-1} - \frac{n}{N} \tag{13b}$$

From 13a we get

$$x \geq 3^{-\frac{1}{N-n}} := x_n^*. \tag{14}$$

From 13b we get that

$$n^* = \max\{n : \sum_{k=n+1}^{N} \frac{1}{k-1} > 2\}. \tag{15}$$

Optimal behaviour

Suppose that we are in region $n > n^*$ and $x_n < x_n^*$, the current state is $(n, X_n = x)$ and the player I is in the game alone (since player II has stopped). Player I will get profit only if X_n is the global maximum. Stopping in this moment he will get $s_{1,n}(x)$ while waiting he will get

$$\tilde{w}_{1,n}(x) = \sum_{k=n+1}^{N} x^{k-n-1} \int_x^1 (\tilde{w}_{1,k}(y) \vee s_{1,k}(y))dy - s_{1,n}(x). \qquad (16)$$

For 16 we have the following boundary condition: because in moment N is optimal to stop whatever is the observation we get that doing one step more (i.e. not selecting any object) we get

$$\tilde{w}_{1,N} = -1, \forall x \in [0,1]$$

and

$$-1 \geq 1, \forall x \in [0,1]$$

so the threshold is

$$\tilde{x}_N = 0$$

Optimal behaviour

The bellow table presents some numerical results:

i	$\tilde{x}_{N-i}$	x^*_{N-i}
0	0	0
1	0.33333	0.33333
2	0.57735	0.57735
3	0.70738	0.69336
4	0.78271	0.75984
5	0.83032	0.80274
6	0.86252	0.83268
7	0.88546	0.85475
8	0.90249	0.87169
9	0.91555	0.88509
10	0.92582	0.89596
11	0.93408	0.90495
12	0.94085	0.91251
13	0.94648	0.91896
14	0.95122	0.92453
15	0.95527	0.92938

Optimal behaviour

Conjecture

$\tilde{x}_n \geq x_n^*$ for every $n > n^*$.

Conjecture

$\tilde{x}_n$ is non-increasing sequence for $n > n^*$

Optimal behaviour

Conjecture

$\tilde{x}_n \geq x_n^*$ *for every* $n > n^*$.

Conjecture

$\tilde{x}_n$ *is non-increasing sequence for* $n > n^*$

Now consider the opposite situation: let $n < n^*$, player I has already stopped, and the current state is $(n, X_n = x)$. Player II sees only the rank of it and it is obviously 1. Then a gain function for player II is $s_{2,n}$. Because he is alone in the game by doing one step more he gets

$$Ts_{2,n} = \sum_{k=n+1}^{N} p_{n,k} s_{2,k} - s_{2,n} = \sum_{k=n+1}^{N} \frac{n}{k(k-1)} \frac{k}{N} - \frac{n}{N}, \qquad (17)$$

where an operator $T(\cdot)$ is called the averaging operator.

Optimal behaviour

To find an optimal stopping rule we check when $Ts_{2,n} \leq s_{2,n}$, i.e. when the expected value of doing one step more is less or equal to pay-off in current state. We get condition that stopping rule is

$$\tilde{n} = \max\{n : \sum_{k=n+1}^{N} \frac{1}{k-1} > 2\}. \tag{18}$$

Note that $\tilde{n} = n^*$.

Discussion

Player II is indeed "no information": even if the rival stops he does not get any additional information. He has to behave only on the basis of the knowledge of the relative ranks.
Different situation is when the player II gets some additional information, e.x. the threshold of the rival or the value of the chosen object. This gives him an additional information about the future distribution of the relative ranks.

Bayesian approach to the incomplete information games

In particular games the incomplete information is related to a situation in which some players do not know the other players' characteristics precisely. It can be a player may not know exact the payoff function of others.We assume that player knows about this beliefs and the opponent knows that the player knows.

Let $(X_n, \mathcal{F}_n, \mathbf{P}_x)_{n=0}^N$ be a homogeneous Markov process defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with the state space $(\mathbb{E}, \mathcal{B})$. At each moment $n = 1, 2, ..., N$ the decision makers are able to observe the Markov chain. Each player has his utility function $g_i : \mathbb{E} \to \mathbb{R}$, $i = 1, 2$, and at each moment n each decides separately to accept or reject a realization x_n of X_n. We assume that g_i are measurable and bounded. If it happens that both players have selected the same moment n to accept x_n, then a random assignment mechanism is applied.

Model description

Let $\mathcal{S}^N$ be the aggregation of Markov times with respect to $(\mathcal{F}_n)_{n=0}^N$. We admit that $\mathbf{P}_x(\tau \leq N) < 1$ for some $\tau \in \mathcal{S}^N$ (i.e. there is a positive probability that the Markov chain will not be stopped). The elements of $\mathcal{S}^N$ are possible strategies for the players with the restriction that player 2 and player 1 cannot stop at the same moment. If both players declare willingness to accept the same object, the random device decides who is endowed.
Denote $\mathcal{S}_k^N = \{\tau \in \mathcal{S}^N : \tau \geq k\}$. Let Λ_k^N and M_k^N be copies of $\mathcal{S}_k^N$ ($\mathcal{S}^N = \mathcal{S}_0^N$).

Model description

The sets of strategies for player 1 and 2 are:

$$\tilde{\Lambda}^N = \{(\lambda, \{\sigma_n^1\}) : \lambda \in \Lambda^N, \sigma_n^1 \in \Lambda_{n+1}^N \text{ for every } n\} \tag{19}$$

and

$$\tilde{M}^N = \{(\mu, \{\sigma_n^2\}) : \mu \in M^N, \sigma_n^2 \in M_{n+1}^N \text{ for every } n\} \tag{20}$$

Model description

Denote $\tilde{\mathcal{F}}_n = \sigma(\mathcal{F}_n, \xi_1, \xi_2, \ldots, \xi_n)$ and let $\tilde{\mathcal{S}}^N$ be the set of stopping times with respect to $(\tilde{\mathcal{F}}_n)_{n=0}^N$, where ξ_n are taken from uniform distribution and we can treat it as a *lottery variable*.

Model description

Denote $\tilde{\mathcal{F}}_n = \sigma(\mathcal{F}_n, \xi_1, \xi_2, \ldots, \xi_n)$ and let $\tilde{\mathcal{S}}^N$ be the set of stopping times with respect to $(\tilde{\mathcal{F}}_n)_{n=0}^N$, where ξ_n are taken from uniform distribution and we can treat it as a *lottery variable*.
Define the random sequence $\{\alpha_n\}_{n=0}^N$ which is observable by both players measurable with respect to the filtration $(\tilde{\mathcal{F}}_n)_{n=0}^N$ (such that for every n we have $\mathcal{F}_n \subset \tilde{\mathcal{F}}_n$. When the lottery have to decide about assignment then the sample ξ_n from uniform distribution is taken. If $\xi_n \leq \alpha_n$, then the player 1 is benefited.

Model description

Denote $\tilde{\mathcal{F}}_n = \sigma(\mathcal{F}_n, \xi_1, \xi_2, \ldots, \xi_n)$ and let $\tilde{\mathcal{S}}^N$ be the set of stopping times with respect to $(\tilde{\mathcal{F}}_n)_{n=0}^N$, where ξ_n are taken from uniform distribution and we can treat it as a *lottery variable*.
Define the random sequence $\{\alpha_n\}_{n=0}^N$ which is observable by both players measurable with respect to the filtration $(\tilde{\mathcal{F}}_n)_{n=0}^N$ (such that for every n we have $\mathcal{F}_n \subset \tilde{\mathcal{F}}_n$. When the lottery have to decide about assignment then the sample ξ_n from uniform distribution is taken. If $\xi_n \leq \alpha_n$, then the player 1 is benefited.
Similarly to the random priority model the behavior of the applicant after the proposal from the Player 1 is coded by the random sequence $\{\beta_n\}_{n=0}^N$. This show a chance that the proposal of Player 1 will be accepted. The sequence $\{\beta_n\}_{n=0}^N$ is observable by the Player 1.

Payoff matrix

Consider a two person an incomplete information nonzero-sum game with random priority, which is related to the secretary problem. We assume that both players observe Markov chain $W_t,\ t = 1, 2, ...$ and their utility functions are $g_j(r) = f(r),\ j = 1, 2,\ r \in \mathbb{E}$. Let lottery $\bar{\alpha}$ be constant, i.e. $\alpha_i = \alpha,\ i = 1, 2, ..., N$. Denote $\tilde{c}(r) = \tilde{c}_{BA}(r)$ the perspective expected gain in the best choice problem in which the stopping was not allowed before the r-th applicant had been evaluated and
$r_a = \inf\{1 \leq r \leq N : \sum_{i=r+1}^{N} \frac{1}{i-1} \leq 1\}$ and
$\tau_r^* = \inf\{s > r : Y_s = 1, s \geq r_a\}$.

Payoff matrix

We have the following payoff matrix:

$$\begin{pmatrix} (\alpha f(n,X_n)+(1-\alpha)\tilde{c}(n,X_n), & (\beta f(n,X_n)+(1-\beta)\tilde{c}(n,X_n), \\ (1-\alpha)f(n,X_n)+\alpha\tilde{c}(n,X_n)) & (1-\hat{\beta})f(n,X_n)+\hat{\beta}\tilde{c}(n,X_n)) \\ & \\ (\tilde{c}(n,X_n),f(n,X_n)) & (\tilde{v}_1(n,X_n,\overline{\beta_n}),\tilde{v}_2^{B_2}(n,X_n,)) \end{pmatrix}$$

where $\overline{\beta_n}=(\beta_n,...,\beta_N)$ and

$$\tilde{v}_1(r,\alpha)=\sum_{i=r+1}^{N} p(r,i)(\alpha f(i,X_i)+(1-\alpha)\tilde{c}(i,X_i)) \tag{21a}$$

$$\tilde{v}_2^{B_2}(r,\alpha)=\sum_{i=r+1}^{N} p(r,i)((1-\alpha)f(i,X_i)+\alpha\tilde{c}(i,X_i)) \tag{21b}$$

Equilibrium point

The equilibrium payoffs are $(v_1^{\beta}, v_2^{B_2}) = (v_1(1,\alpha,\beta), v_2^{B_2}(1,\alpha))$. When $N \to \infty$ such that $\frac{r}{N} \to x$ we obtain

$$\hat{v}_1(x,\alpha,\beta) = \lim_{N\to\infty} v_1(r,\alpha,\beta) = \begin{cases} \hat{w}_1(x,x,\alpha) & \text{if } x \geq a, \\ \hat{w}_1(x,a,\alpha,\beta) & \text{if } b(\beta) \leq r < a, \\ \hat{w}_1(b(\beta),a,\alpha,\beta) & \text{if } 0 \leq r < b(\beta), \end{cases}$$

and

$$\hat{v}_2^{B_2}(x,\alpha) = \lim_{N\to\infty} v_2^{B_2}(r,\alpha) = E_{\boldsymbol{\beta}} \begin{cases} \hat{w}_2(x,x,\alpha) & \text{if } x \geq a, \\ \hat{w}_2(x,a,\alpha,\boldsymbol{\beta}) & \text{if } b(\boldsymbol{\beta}) \leq r < a, \\ \hat{w}_2(b(\boldsymbol{\beta}),a,\alpha,\boldsymbol{\beta}) & \text{if } 0 \leq r < b(\boldsymbol{\beta}), \end{cases}$$

Equilibrium point

The equilibrium payoffs are $(v_1^{\beta}, v_2^{B_2}) = (v_1(1,\alpha,\beta), v_2^{B_2}(1,\alpha))$. When $N \to \infty$ such that $\frac{r}{N} \to x$ we obtain

$$\hat{v}_1(x,\alpha,\beta) = \lim_{N\to\infty} v_1(r,\alpha,\beta) = \begin{cases} \hat{w}_1(x,x,\alpha) & \text{if } x \geq a, \\ \hat{w}_1(x,a,\alpha,\beta) & \text{if } b(\beta) \leq r < a, \\ \hat{w}_1(b(\beta),a,\alpha,\beta) & \text{if } 0 \leq r < b(\beta), \end{cases}$$

and

$$\hat{v}_2^{B_2}(x,\alpha) = \lim_{N\to\infty} v_2^{B_2}(r,\alpha) = E_{\boldsymbol{\beta}} \begin{cases} \hat{w}_2(x,x,\alpha) & \text{if } x \geq a, \\ \hat{w}_2(x,a,\alpha,\boldsymbol{\beta}) & \text{if } b(\boldsymbol{\beta}) \leq r < a, \\ \hat{w}_2(b(\boldsymbol{\beta}),a,\alpha,\boldsymbol{\beta}) & \text{if } 0 \leq r < b(\boldsymbol{\beta}), \end{cases}$$

where

$$\hat{w}_1(x,y,\alpha,\beta) = \beta x \ln\frac{y}{x} - \frac{1-\beta}{2} x \ln\frac{y}{x} \ln(x\cdot y) + \frac{x}{y}\tilde{v}_1(y,\alpha),$$

$$\hat{w}_2(x,y,\alpha,\beta) = \hat{w}_1(x,y,1-\alpha,1-\beta)$$

and

$$\tilde{v}_1(x,\alpha) = -\alpha x \ln(x) + \frac{1-\alpha}{2} x \ln^2(x).$$

Decision points: numerical results

We have $a = \exp(-1)$ and $b(\beta)$ is the positive root of equation:

$$\hat{w}_1(x, a, \alpha, \beta) = \beta x - (1 - \beta)x \ln(x).$$

The value of the game for the first player and the second player is

$$(\hat{v}_1(\alpha, \beta), \hat{v}_2^{B_2}(\alpha)) = (\hat{w}_1(b(\beta), a, \alpha, \beta), \boldsymbol{E}_{\boldsymbol{\beta}} \hat{w}_2(b(\boldsymbol{\beta}), a, \alpha, \boldsymbol{\beta}). \qquad (22)$$

Decision points: numerical results

(α,β)	$b(\beta)$	$\hat{v}_1(\alpha,\beta)$
$(\alpha_0, 0.55)$	.2538	.2962
$(\alpha_0, 0.75)$	.2729	.293275
$(0.4, 0.55)$	2804	.2966
$(0.4, 0.75)$	.2524	.2762

The $\hat{v}_2^{B_2}(\alpha)$ is the expected value: $\boldsymbol{E}_{\boldsymbol{\beta}}\hat{w}_2(b(\boldsymbol{\beta}), a, \alpha, \boldsymbol{\beta})$.

References

- A. Kurushima, K. Ano (2010) *Full-information duration problem and its generalizations*, RIMS Kôkyûroku , No. 1682, 50-54
- Corless, R. M.; Gonnet, G. H.; Hare, D. E. G.; Jeffrey, D. J.; Knuth, D. E. (1996) *On the Lambert W function*, Advances in Computational Mathematics 5: 329-359
- T.S. Ferguson, J.P. Hardwick, M. Tamaki (1991) *Maximizing the duration of owning a relatively best object*, in Strategies for Sequential Search and Selection in Real Time, Contemporary Mathematics 125, F.T. Bruss, T.S. Ferguson and S.M. Samuels eds., 37-57
- Z. Porosiński, M. Skarupski, K. Szajowski (2016) *Duration problem: basic concept and some extensions*, Math. Appl. 44(1), 87–112.
- M. Tamaki (2013) *Optimal stopping rule for the full-information duration problem with random horizon*. Sûrikaisekikenkyûsho Kôkyûroku, (1864):12–19, 2013-11. Accepted in Adv. Appl. Probab.