Economics Letters 117 (2012) 303–305

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Non-probabilistic decision making with memory constraints

Alexander Vostroknutov

P.O. Box 616, Maastricht University, 6200 MD, Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 20 March 2012
Received in revised form
23 May 2012
Accepted 1 June 2012
Available online 8 June 2012

JEL classification:
D83
D81
C02

Keywords:
Adaptive learning
Constrained memory
Bandit problems

ABSTRACT

The single decision maker chooses one of the actions repeatedly. She chooses the action with the highest weighted average of the past payoffs. In the long run either the action with highest expected payoff or the action with highest minimal payoff is chosen depending on how weights evolve.

1. Introduction

The environment in which economic agents make decisions can be very complex. This accounts for the tendency of the agents to simplify their decisions. We study the behavior of a single decision maker who does not deliberately randomize while choosing among a finite set of actions repeatedly (Sarin and Vahid, 1999). Instead, she chooses the action with the highest subjective assessment, which is represented by the weighted average of the past payoffs. The decision maker has no information about the environment apart from the payoffs she receives.

We extend the model of Sarin and Vahid (1999) by analyzing the long run behavior of the decision maker for very general weight structures. In particular, weights can change from period to period with little restrictions. In addition, we allow for the possibility that the decision maker makes mistakes and chooses the action not originally intended. We show that the long run behavior of the decision maker *either* is characterized by only two very distinctive regimes, independent of the relative payoffs obtained from different actions, *or* cannot be characterized because of the dependence on the relative payoffs.

2. The model

The decision maker faces the same decision problem repeatedly. Each period she is choosing one of the I actions from a finite set $A = \{a_1, a_2, \ldots, a_I\}$. After the action is chosen, the state of the world $\omega_t \in \Omega$ is realized (t stands for the time period). Ω is assumed finite. There is some probability measure defined on Ω. For each t, ω_t is identically and independently distributed in each time period. After the state of the world is chosen, the decision maker receives her payoff from choosing action i according to the utility function $u_i : \Omega \to \mathbb{R}_+$.

Before each period t the decision maker calculates subjective assessments $\alpha(t) = (\alpha_1(t), \alpha_2(t), \ldots, \alpha_I(t))$ for each action. Then, with probability $1-\varepsilon$ she chooses the action with the *highest subjective assessment* and with small probability ε she makes a mistake, in which case she chooses some other action. All other actions have equal probability of being chosen. Initial assessments before period 1 are denoted $\alpha(0) = (\alpha_1(0), \alpha_2(0), \ldots, \alpha_I(0))$ and assumed to lie in between maximum and minimum possible utilities for each action.

The subjective assessment for an action is updated whenever that action is chosen. Consider some fixed infinite triangular array of weights

$$
\begin{array}{l}
1 \\
\lambda_{10}\ \lambda_{11} \\
\lambda_{20}\ \lambda_{21}\ \lambda_{22} \\
\vdots \\
\lambda_{k0}\ \lambda_{k1}\ \ldots\ \lambda_{kk} \\
\vdots,
\end{array}
$$

E-mail addresses: aevk79@gmail.com, a.vostroknutov@maastrichtuniversity.nl.

0165-1765/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2012.06.004

where $\sum_{j=0}^{k} \lambda_{kj} = 1$, $\forall k \in \mathbb{N}$. Before period t, after action i was chosen k times, the assessment is $\alpha_i(t) = \sum_{j=1}^{k} \lambda_{k(k-j)} u_{ij} + \lambda_{kk}\alpha_i(0)$, where u_{ij} is the utility from action i after it was chosen for the jth time. Here the most recent utility observation u_{ik} is assigned the weight λ_{k0} and the oldest one, $\alpha_i(0)$, the weight λ_{kk}. The assessments of actions that were not chosen in period $t-1$ stay the same and carry over to period t.

We are interested in the long run behavior of the decision maker as $t \to \infty$, which depends on the evolution of weights over time. Consider the vector space $\mathbb{R}^{\mathbb{N}}$ of all infinite sequences of real numbers $x = (x_1, x_2, \ldots)$ with the sup norm $\|x\| = \sup_{t\in\mathbb{N}} |x_t|$. The metric $d(x, y) = \|x - y\|$, thus, generates the topology on $\mathbb{R}^{\mathbb{N}}$. Any row $\lambda_k = (\lambda_{k0}, \lambda_{k1}, \ldots, \lambda_{kk})$ can be naturally associated with an element $(\lambda_{k0}, \lambda_{k1}, \ldots, \lambda_{kk}, 0, 0, \ldots)$ of $\mathbb{R}^{\mathbb{N}}$. Abusing notation, let us write $\lambda_k \in \mathbb{R}^{\mathbb{N}}$. Consider two cases:[1]

Case I: $\lim_{k\to\infty} \lambda_k = 0$. This case consists of the triangular arrays that put vanishing weights on infinitely many past observations as $t \to \infty$. The common example of such triangular array is the average of all past observations: $\lambda_k = \left(\frac{1}{k+1}, \frac{1}{k+1}, \ldots, \frac{1}{k+1}\right)$.

Case II: $\lim_{k\to\infty} \lambda_k = \lambda_\infty$. Arrays in this case put positive weights only on some observations in the past. We assume that for all $k, j \geq K \lambda_{kj} = 0$: only K latest payoffs are remembered. Any limit λ_∞ has $\sum_{j=0}^{K} \lambda_{\infty j} = 1$ and $\sum_{j=K}^{\infty} \lambda_{\infty j} = 0$.

3. Results

Proposition 1. *If weights evolve according to Case* I *then the decision maker asymptotically plays the action with maximal expected payoff with probability* $1 - \varepsilon$.

Proof. Let us show that the assessments of all actions converge to the expected value of payoffs. Consider some action i and time periods $t_1, t_2, \ldots$ when this action is played ($t_1 < t_2 < \cdots$). We are interested in the behavior of the assessment $\alpha_i(t_n)$ as $n \to \infty$ (since the decision maker makes mistakes we can be sure that any action is played infinitely often). For each update period t_n we know that

$$\alpha_i(t_n) = \lambda_{n0} u_i(\omega_{t_n}) + \lambda_{n1} u_i(\omega_{t_{n-1}}) + \cdots + \lambda_{nn}\alpha_i(0),$$

where $u_i(\omega_t)$ is the realization of the payoff in period t. To prove the statement we use the theorem of Fristedt and Gray (1997, Theorem 25, p. 311). First we show that the assumptions of the theorem are satisfied and then explicitly find the limiting distribution of $\alpha_i(t)$.

Consider the triangular array of random variables:

$$\begin{array}{llll}
\alpha_i(0) & & & \\
\lambda_{10} u_i(\omega_{t_1}) & \lambda_{11}\alpha_i(0) & & \\
\lambda_{20} u_i(\omega_{t_2}) & \lambda_{21} u_i(\omega_{t_1}) & \lambda_{22}\alpha_i(0) & \\
\vdots & & & \\
\lambda_{n0} u_i(\omega_{t_n}) & \lambda_{n1} u_i(\omega_{t_{n-1}}) & \ldots & \lambda_{nn}\alpha_i(0) \\
\vdots & & &
\end{array}$$

This array is row-wise independent.[2] Indeed, each time the decision maker plays a_i she receives independent realization of $u_i(\cdot)$. Moreover, this array is uniformly asymptotically negligible: this is easy to see since, by the definition of Case I weights, for any fixed $\delta > 0$ we can always find the row of weights small enough for $\sup_k P[|\lambda_{nk} u_i(\omega_{t_{n-k}})| > \delta] = 0$ to be true for some n and any row that follows. This is the consequence of the assumption that $u_i(\cdot)$ can take on values only in the bounded interval of $\mathbb{R}$.

Now let us verify the conditions of the theorem. We claim that the Lévy measure $\nu(x, \infty] = 0$, $\forall x > 0$ satisfies the first condition. For any $x > 0$ we can find n big enough so that for all $k \leq n$ we have $P[\lambda_{nk} u_i(\omega_{t_{n-k}}) > x] = 0$, hence the probability measure corresponding to any random variable $\lambda_{\ell k} u_i(\omega_{t_{\ell-k}})$ where $\ell \geq n$ and $k \leq \ell$ is zero on the interval $[x, \infty)$. Thus, the limit of the sums of these measures in each row is zero.

Denote by Q^i_{nk} the distribution of $\lambda_{nk} u_i(\omega_{t_{n-k}})$ and consider the integral

$$\int_{(0,\delta]} x\, Q^i_{nk}(dx).$$

Since $\lim_{n\to\infty} \lambda_{nk} = 0$ there exists n big enough so that $Q^i_{nk}[\delta, \infty) = 0$. Therefore for all $\ell \geq n$

$$\int_{(0,\delta]} x\, Q^i_{\ell k}(dx) = \lambda_{\ell k} E[u_i(\cdot)]$$

and

$$\sum_{k=1}^{\ell} \int_{(0,\delta]} x\, Q^i_{\ell k}(dx) = E[u_i(\cdot)]$$

So the second condition of the theorem is clearly satisfied:

$$\lim_{\delta\searrow 0} \limsup_{n\to\infty} \sum_{k=1}^{n} \int_{(0,\delta]} x\, Q^i_{nk}(dx)$$
$$= \lim_{\delta\searrow 0} \liminf_{n\to\infty} \sum_{k=1}^{n} \int_{(0,\delta]} x\, Q^i_{nk}(dx) = E[u_i(\cdot)].$$

Now, the theorem tells us that the assessment converges to a random variable which corresponds to the pair $(E[u_i(\cdot)], 0)$ via the Lévy–Khinchin Representation Theorem. This random variable has a moment generating function $\exp(-E[u_i(\cdot)])$ which obviously corresponds to the delta distribution at $E[u_i(\cdot)]$. This finishes the proof of the statement above.

In the model the decision maker makes mistakes, so all the actions are played infinitely often. Therefore, as it was shown, assessments of all actions converge to the expected value. Since the decision maker chooses the action with the highest assessment, she will eventually choose the one with maximum expected payoff. □

For the Case II result we need some definitions first. Since λ_∞ puts weights only on the last K utilities, there are only finitely many values of the assessments with weights λ_∞. Let M_i denote the finite set of such assessments for action i, and $M = \times_{j=1}^{I} M_j$ denote the finite set of all possible assessments combinations.[3] Let P^ε denote the Markov chain defined by the choice procedure on the set M. Let $a_{\text{maxmin}} \in A$ denote the action with the maximal minimum utility.

Proposition 2. *If weights evolve according to Case* II *then the decision maker asymptotically plays* a_{maxmin} *most of the time.*[4]

Proof. According to Theorem 3.1 of Sarin (2000), the Markov chain P^0 converges to the choice of a_{maxmin} with probability 1. For $\varepsilon > 0$, P^ε is a *regular perturbed Markov process*.[5] This is clear since ε

[1] From here on, all limits of sequences λ_k in $\mathbb{R}^{\mathbb{N}}$ are taken in the topology generated by $d(x, y) = \|x-y\|$. 0 is naturally understood as the sequence $(0, 0, \ldots)$.

[2] For the definitions of the terms used in this proof see Fristedt and Gray (1997).

[3] To avoid uninteresting assumptions about tie breaking, assume that no two sets M_i and M_j contain the same numbers.

[4] See proof for the exact meaning of "most of the time".

[5] See definition in Young (1998, p. 54).

enters only as a linear term $c\varepsilon$ for some constant c in transition probabilities of P^ε. According to Theorem 3.1 of Young (1998) the stationary distribution μ^ε of P^ε converges to one of the stationary distributions of P^0 as $\varepsilon \to 0$. Therefore, since all stationary distributions of P^0 involve playing a_{maxmin} with probability 1, for small $\varepsilon > 0$ the decision maker will play a_{maxmin} most of the time.

It is left to show that the same asymptotic behavior is inherent to the process with Case II weights. Consider the finite set $\overline{M} = \cup_{j=1}^{l} M_j \subset \mathbb{R}$. Since all actions are chosen infinitely many times and weights converge in the sup norm, there will be a time period t after which all possible assessments for each action i made with current weights λ_i^t will lie in some open balls around the corresponding assessments in $\overline{M}$. Moreover, these balls will all be pairwise disjoint. This implies that the Markov chain $P^{\varepsilon,t}$ with the state space M^t, generated using current weights $(\lambda_i^t)_{i\in A}$, will put exactly the same probabilities on transitions as P^ε with the state space M. Thus, after period t, the behavior of the decision maker with changing weights will be indistinguishable from the behavior of the decision maker with weights λ_∞. □

4. Discussion

The actions chosen by the decision maker in the long run differ depending on the weights the decision maker attaches to the past payoffs. If she cares only about recent payoffs then the maximin action is chosen in the long run. If the decision maker cares about payoffs received a long time ago then she converges to the maximal expected payoff action. It is relatively easy to see that exact predictions like this are impossible to make for sequences of weights that evolve differently from Cases I and II. In particular, such sequences would put a positive weight on some payoffs and distribute the rest of the weight among infinitely many payoffs as time goes to infinity. For example, if some weight, say $\frac{1}{2}$, is attached to the last payoff and the rest of the weight is equally distributed among all other past payoffs, the assessment of the action i will converge to the random variable $\frac{1}{2}(u_i(\cdot) + E[u_i(\cdot)])$. In this case the long run behavior of the decision maker will depend on relative values of the utilities obtained from different actions and cannot be characterized in general.

Since the model gives very specific predictions for Cases I and II it would be interesting to test experimentally if actual human behavior conforms to Cases I or II, or is neither and depends on relative utilities. This can shed some light on how aggregation of past experiences is done in the brain.[6]

The model can be extended in several ways. First, the technique used to prove the convergence result can be applied almost without changes to the case of infinite state space Ω. The only condition that matters is bounded support of the distributions of payoffs. Second, it could be interesting to investigate the case when the action space is large and the decision maker experiments only in the vicinity of the action she plays. Third, one might check how this model performs in games (see e.g. Huck and Sarin, 2004; Young, 2009).

Acknowledgments

I would like to thank Tilman Börgers for invaluable discussions and Andrew Chesher, Beth Allen, and Aldo Rustichini for helpful comments.

References

Fristedt, Bert, Gray, Lawrence, 1997. A Modern Approach to Probability Theory. Birkhäuser, Boston.
Huck, Steffen, Sarin, Rajiv, 2004. Players with limited memory. Contributions to Theoretical Economics 4 (1).
Sarin, Rajiv, 2000. Decision rules with bounded memory. Journal of Economic Theory 90 (1), 151–160.
Sarin, R., Vahid, F., 1999. Payoff assessments without probabilities: a simple dynamic model of choice. Games and Economic Behavior 28 (2), 294–309.
Sarin, R., Vahid, F., 2001. Predicting how people play games: a simple dynamic model of choice. Games and Economic Behavior 34 (1), 104–122.
Young, H. Peyton, 1998. Individual Strategy and Social Structure: An Evolutionary Theory of Institutions. Princeton University Press, Princeton, N.J.
Young, H. Peyton, 2009. Learning by trial and error. Games and Economic Behavior 65 (2), 626–643.

6 Sarin and Vahid (2001) show that the Case II model does remarkably well at predicting behavior in some simple games.