

RESEARCH

Open Access



# Hidden Markov Bayesian Game with Application to Chinese Education Game

Wen He<sup>1</sup> and Jinwu Gao<sup>2\*</sup>

\*Correspondence: [jgao@ruc.edu.cn](mailto:jgao@ruc.edu.cn)

<sup>2</sup>School of Information, Renmin University of China, Beijing 100872, China  
Full list of author information is available at the end of the article

## Abstract

In real game situations, players are often not perfectly informed about true states but can observe signals, and decision-making may involve several periods. In order to formulate such situations, this paper uses a hidden Markov model to describe the state process, thus introducing a repeated game with hidden Markovian states, called *hidden Markov Bayesian game*. For the new model, a notion of Nash equilibrium is presented and an algorithm is developed to facilitate obtaining the equilibrium quickly. An analysis of the Chinese education game shows that the observed signals play an important role in analyzing players' behavior.

**Keywords:** Bayesian game, Repeated game, Hidden Markov model, Nash equilibrium

## Introduction

Game models have been successfully applied in various areas, including operation research, computer science, and social science. They are able to formulate and analyze complex interactions between decision-makers, such as investors in financial markets, firms along a supply chain, and countries with conflicting interests. Realistic features—e.g., dynamic decisions, imperfect information, and behavioral elements—are usually incorporated by game models to reflect real-life situations.

According to decision-making structure (e.g., static or dynamic) and information structure (e.g., perfect or imperfect information), game models can be classified into different types. Various equilibrium notions that refine the Nash equilibrium have been invented in order to describe players' strategies in the different types of games [1]. For dynamic games with sequential or simultaneous moves, the subgame perfect equilibrium is usually used to characterize players' rational choices [2]; the grim trigger strategy is also a reasonable alternative for players in repeated games [3]. Regarding information aspects, Bayesian games consider imperfect or incomplete information scenarios where players have beliefs on uncertain factors and aim to maximize their respective expected utilities [4]. By Harsanyi's approach [5] of introducing nature as a player, games of incomplete information can be converted to games of imperfect information.

In this paper, we propose a repeated game with hidden Markovian states, called *hidden Markov Bayesian game*. Starting with the repeated game, the proposed model considers situations where at each period, players are not aware of the actual state but can observe signals that are driven by the state. Particularly, the state process over the periods

is modeled by a hidden Markov model (HMM); in an HMM [6], states follow a Markov process, and beliefs on them can be updated over time based on the observed sequence of signals.

Specifically, the proposed game model proceeds as follows. At the beginning of each period, a nature state is realized and each player observes a signal that may provide information about the state. Then, players simultaneously choose actions. Then, the next period begins, and the procedure is repeated. Note that their payoffs cannot be realized until the end of the final period (this assumption is consistent with a lot of previous literature [7]). We also define a Nash equilibrium for the model. At the Nash equilibrium, each player chooses a strategy to maximize the expectation of his total payoff over all periods, given his signal observations and the other players' strategies. Moreover, we design an algorithm to quickly calculate the probability distributions of hidden states under all possible signal sequences. These probability distributions are essential to obtaining the model's equilibria.

Furthermore, the proposed model is able to capture many real-life situations and produce insights. For example, we apply it to controversial Chinese education issues. From the perspective of the static game with perfect information, schools, overwhelmed by examination-oriented education, are trapped in the classic prisoner's dilemma. But our model is capable to shed light on the dynamic and informational aspects. The results show that schools tend to prefer examination-oriented education if the education policy keeps changing. Also, schools lacking information might keep choosing their previous way of education.

The main feature that differentiates our model from existing game models is that we introduce signals into our model and then describe nature states as an HMM. In stochastic games [8], players choose actions, for multiple periods, not knowing the actual states. But different from our model, stochastic games do not consider signals that players may receive. In dynamic games with incomplete information [9], a player can view another player's action as a signal in order to update his belief on the unobservable state (often referring to some player's type). But signals in our model are defined to be some information observed outside the game (which cannot be maneuvered by opponents) rather than opponents' actions.

The HMM provides a theoretical tool to describe signal process systems and has been widely applied to speech recognition [10], activity sequences [11], etc. For example, Netzer et al. [12] proposes an HMM to model latent customer-firm relationship states and customer buying behavior (signals). Dias and Ramos [13] gave a dynamic clustering of energy by an extended hidden Markov approach. Ansari et al. [14] and Kujala et al. [15] employed hidden Markov approaches to analyze some game situations. Since our model considers both nature states and observable signals over multiple periods, applying the HMM in a game context—although different from existing approaches—is appropriate in our setting.

The rest of this paper is organized as follows. In the "Preliminaries" section, we provide a brief review on the concepts of finitely repeated games, Bayesian games, and hidden Markov models. In the "Model: Hidden Markov Bayesian Game" section, we introduce our model and defined its Nash equilibrium. The algorithm is provided in the "Algorithm" section. We apply our model to education issues in the "Application" section and concluded in the "Conclusions" section.

## Preliminaries

### Finitely Repeated Game

In a repeated game, players make actions simultaneously over multiple periods [16]. It can be further categorized into two types based on the number of periods, namely, finitely and infinitely. Our model is based on the framework of the finitely repeated game. A finitely repeated game is defined as follows:

**Definition 1.** A finitely repeated game is a tuple  $\langle N, (A_i), (u_i), \delta, T \rangle$ , where

$N$  is a finite set of  $n$  players

$A_i$  is a set of actions available for  $i \in N$ , action profile is denoted as  $A = \times_{i \in N} A_i$

$u_i$  is the payoff function of  $i \in N$  at every period,  $u_i : A \rightarrow \mathcal{R}$

$\delta$  is the discount factor,  $0 \leq \delta \leq 1$

$T$  is the number of the periods.

The discount factor  $\delta$  reflects time's effect on the payoff. For example, the payoff at period  $t$  should be multiplied by  $\delta^{t-1}$  when calculating the total payoff over all the periods.

### Bayesian Game

Our model carries out a Bayesian game at each period. A key component in the specification of a Bayesian game is the set of states [16]. At the start, a state is realized but is not observable to players. Players have prior beliefs on the states and can update their beliefs after observing signals based on Bayes' rule. At the (Bayesian) Nash equilibrium, players maximize their expected payoffs. A Bayesian game can be defined as [17] follows:

**Definition 2.** A Bayesian game is a tuple  $\langle S, V, \pi, N, (A_i), (u_i), (\tau_i) \rangle$ , where

$N$  is a finite set of  $n$  players

$S$  is a finite set of states

$V$  is a finite set of signals

$\tau_i$  is player  $i$ 's signal function, at state  $w$ , player  $i$  receive  $\tau_i(w) \in V$  with some probability distribution

$\pi$  is a prior belief about the states

$A_i$  is a set of actions available for  $i \in N$

$u_i$  is the payoff function of  $i \in N$ .

### Hidden Markov Model

Our model extends the Bayesian game to finitely periods. It models the discrete stochastic process of the state by HMM. The state in the HMM is a Markov chain, which undergoes changes according to the state transition probability distribution. Besides, under each state, players observe signals according to observation signal probability distribution. An HMM is defined as [6] follows:

**Definition 3.** An HMM is a tuple  $\lambda = \langle S, V, R, B, \pi \rangle$ , where

$S$  is a finite set of states

$V$  is a finite set of signals

$R$  is the state transition probability distribution,  $R = (r_{ij})$ , where  $r_{ij}$  indicates the probability of state  $i$  changing to state  $j$

$B$  is the observation signal probability distribution in state  $j$ ,  $B = (b_j(k))$ , where  $b_j(k)$  indicates the probability of signal  $k$  being observed in state  $j$

$\pi$  is the initial state distribution  $\pi = (\pi_i)$ , where  $\pi_i$  is the probability of state  $i$  before the model starts.

**Model: Hidden Markov Bayesian Game**

**Definition**

Our model is a finitely repeated game with hidden Markovian states and can be classified as a repeated game with imperfect information. At the beginning of each period, a nature state is realized, but players do not observe this state. Rather, each player observes a signal which is driven by the state. Then, players simultaneously choose actions, and the next period begins. The procedure is repeated for a finite number of periods.

We further assume that players are not informed about their payoffs until the end of the final period. Thus, when each player is choosing actions, he knows all the players’ previous actions and his own observed signal sequence but knows nothing about other players’ signals and his payoffs in previous periods. Then, based on the definitions presented in the “Finitely Repeated Game,” “Bayesian Game,” and “Hidden Markov Model” sections, our model can be defined as follows:

**Definition 4.** A hidden Markov Bayesian game is a tuple  $\langle \lambda, N, (A_i), (u_i), \delta, T \rangle$ , where

$N$  is a finite set of  $n$  players

$\lambda$  is an HMM of the states and  $\lambda = \langle S, V, R, B, \pi \rangle$

$A_i$  is a set of actions available for  $i \in N$ , action profile is denoted as  $A = \times_{i \in N} A_i$

$u_i$  is the payoff function of  $i \in N$  at every period,  $u_i : A \times S \rightarrow \mathcal{R}$ , and  $u_i(a, w)$  is said to be  $i$ ’s payoff under action profile  $a$  and the state  $w$

$\delta$  is the discount factor,  $0 \leq \delta \leq 1$

$T$  is the number of the periods.

Obviously, if the state process  $\lambda$  is a constant, meaning that there is only one state and players have perfect information, this model becomes a normal repeated game. That is,  $\langle c, N, (A_i), (u_i), \delta, T \rangle$  ( $c$  is a constant) is a  $T$ -period repeated game with the discount factor  $\delta$ . Furthermore, if there is only one period involved, i.e.,  $T = 1$ , the model  $\langle \lambda, N, (A_i), (u_i), \delta, 1 \rangle$  becomes a traditional Bayesian game. Thus, both finitely repeated games and traditional Bayesian games can be considered as special cases of the model.

Consider period  $t$ ,  $1 \leq t \leq T$ . The signal sequence player  $i \in N$  has observed so far is denoted as

$$O_i^t = (o_1, o_2, \dots, o_t), \tag{1}$$

where  $o_j$  is the signal player  $i$  observes at period  $j$ ,  $1 \leq j \leq t$ . Then, player  $i$  would update his belief (probability distribution) on the hidden state  $w_t$  as

$$\Pr(w_t | O_i^t). \tag{2}$$

Besides the belief of the states, player  $i \in N$  chooses the action also according to all players’ previous actions, which are denoted by

$$h^t = (a^0, a^1, a^2, \dots, a^{t-1}), \tag{3}$$

where  $a^j$  represents the action profile of all the players at period  $j$ ,  $1 \leq j \leq t - 1$ , and  $a^0 = \phi$ . Then, in this period  $t$ , player  $i$  chooses his action according to the signal sequence  $O_i^t$  he has observed and all players' previous actions  $h^t$ .

In addition, each player is not sure about the true state and thinks every state is possible. Thus, at this period  $t$ , the expected payoff of player  $i$  when he chooses the action  $a_i^t$  is expressed as

$$U_{i,t}(a_i^t, a_{-i}^t) = \sum_{w_t \in S} \Pr(w_t | O_i^t) u_i(a_i^t, a_{-i}^t, w_t), \tag{4}$$

where  $a_{-i}^t$  denotes the action profile of all the players except player  $i$ , and  $\Pr(w_t | O_i^t)$  can be precisely calculated by the algorithm we proposed later.

**Nash Equilibrium**

Assume that before the whole game begins, players have determined their strategies. Thus, player  $i$ 's strategy  $\sigma_i$  is a function assigning an action to all possible  $O_i^t$  and  $h^t$ , that is,

$$\sigma_i(h^t, O_i^t) \in A_i, \tag{5}$$

where  $i \in N$ ,  $1 \leq t \leq T$ . Let  $\sigma$  be players' strategy profile. Then, given  $\sigma$ , player  $i$ 's total discounted payoff (in expected value) over all the  $T$  periods is expressed as

$$\bar{U}_i(\sigma) = \sum_{t=1}^T \delta^{t-1} U_{i,t}(\sigma_1(h^t, O_1^t), \sigma_2(h^t, O_2^t), \dots, \sigma_n(h^t, O_n^t)), \tag{6}$$

where  $U_{i,t}$  is defined in (4).

Let  $\varphi_i$  be the set of player  $i$ 's all strategies. Then, the definition of Nash equilibrium of our model can be given.

**Definition 5.** A Nash equilibrium of a hidden Markov Bayesian game  $\langle \lambda, N, (A_i), (u_i), \delta, T \rangle$  is a strategy profile  $\sigma$ , with the property that for every player  $i \in N$ , we have

$$\bar{U}_i(\sigma_i, \sigma_{-i}) \geq \bar{U}_i(\sigma'_i, \sigma_{-i}) \text{ for all } \sigma'_i \in \varphi_i. \tag{7}$$

Under Nash equilibrium, player  $i$  ( $i \in N$ ) would choose a strategy  $\sigma_i$  that maximizes his total discounted payoff  $\bar{U}_i(\sigma_i, \sigma_{-i})$ , given the other players' strategies  $\sigma_{-i}$ . Then, during the game, player  $i$  would choose his action as  $a_i^t = \sigma_i(h^t, O_i^t)$ , where  $h^t$  is the history about all players' previous actions and  $O_i^t$  is the signal sequence that player  $i$  has observed so far. We should note that this equilibrium definition is based on our assumption that players update beliefs on the nature state only according to the signal sequence they observed, as (2). Our model does not consider cases where the beliefs can be also updated according to other players' actions, for example, a player's belief may be influenced by the actions of another player having more accurate observation. Due to this assumption, the belief system of our model becomes quite simple.

To describe players' rational strategies, Nash equilibrium alone may not be sufficient and there may be a need to refine it following the concept of subgame perfection [1]. This is because a strategy profile that constitutes Nash equilibrium for the whole game may not represent players' rational choices at a subgame. Here, we investigate the subgame perfect equilibrium (SPE) for a special case of our model and leave the task of defining the perfect Bayesian equilibrium (PBE) to future research. Consider the case where all the players observe identical signals, that is, players share the same HMM. Then, the start of

any period  $t \leq T$  would form a well-defined subgame because every player has the same belief on the nature state. The SPE for this special case of model is defined as follows:

**Definition 6.** For a hidden Markov Bayesian game where all the players observe identical signals, a subgame perfect equilibrium is a strategy profile  $\sigma$ , with the property that at the start of every period  $t = 1, \dots, T$ , given the previously occurred signal sequence  $(o_1, o_2, \dots, o_{t-1})$  and actions  $h^{t-1}$ , for every player  $i \in N$ , we have

$$\bar{U}_i(\sigma_i, \sigma_{-i}) \geq \bar{U}_i(\sigma'_i, \sigma_{-i}) \text{ for all } \sigma'_i \in \varphi_i. \tag{8}$$

**An Example**

As we know, the return of an investment usually depends on economic conditions, as well as decisions of competitors. But true economic conditions are not completely observable to investors. Besides, in many cases, the investment requires multiple decision periods, and its return can only be collected at the end of the last period. Such an investment problem can be modeled as a *hidden Markov Bayesian game*. Here, we present an example with three decision periods and two players. At each period, players have two alternative actions: invest ( $Y$ ) or not invest ( $N$ ). The economic condition has two states: good ( $s_1$ ) or bad ( $s_2$ ). Players may use the real-estate price as a signal, which can be high ( $v_1$ ) or low ( $v_2$ ), based on which they can update their beliefs on the economic condition.

In this example, the state transition probability distribution is  $R = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$ . The observation signal probability distribution is (players observe same signals)  $B = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$ . And the initial state distribution is  $\pi = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ . The discount factor is  $\delta = 1$ , that is, we do not consider the discount factor's effect in this example.

The players' payoffs under  $s_1$  and  $s_2$  are represented respectively by Tables 1 and 2. Obviously, these two players' interests are consistent. The following strategy profile is one Nash equilibrium (also an SPE) of this game (the method to obtain Nash equilibrium will be discussed in the next section):

- Player 1 chooses  $N$  if the signal sequence that he observes is  $(v_2), (v_2, v_2), (v_1, v_2, v_2), (v_2, v_1, v_2)$ , or  $(v_2, v_2, v_2)$  and chooses  $N$  if observing other signal sequences, irrespective of previous actions.
- Player 2 has the same strategy as player 1.

**Algorithm**

An essential part of obtaining the Nash equilibrium of our model is to calculate the expected payoff defined in (4). To calculate the expected payoff, we must know the hidden (nature) state  $w_t$ 's probability distribution given any observed signal sequence  $(o_1, o_2, \dots, o_t)$  at each period  $t = 1, \dots, T$ , that is,

$$\Pr(w_t = s_t | o_1, o_2, \dots, o_t), \tag{9}$$

**Table 1** Under good economic condition

		Player 2	
		Y	N
Player 1	Y	10,10	8,0
	N	0,8	0,0

**Table 2** Under bad economic condition

		Player 2	
		Y	N
Player 1	Y	-5, -5	-6, 0
	N	0, -6	0, 0

where  $1 \leq i \leq N$ . For example, we need to obtain  $\Pr(w_1 = s_1|o_1)$  for period 1 and  $\Pr(w_2 = s_1|o_1, o_2)$  for period 2. In the proposed algorithm, we have derived (9) for all the periods, starting from  $t = 1$ , then  $t = 2, \dots$ , and at last, the final period  $t = T$ .

We begin with briefly reviewing the forward variable definition and the iterative algorithm given by Rabiner [6]. For an HMM  $\lambda = \langle S, V, R, B, \pi \rangle$ , its forward variable for period  $t$  is defined as

$$\Pr(o_1, o_2, \dots, o_t, w_t = s_i) \tag{10}$$

which can be calculated by the following algorithm:

1. Initialization, period 1:

$$\Pr(o_1, w_1 = s_i) = \pi_i b_i(o_1). \tag{11}$$

2. Induction, period  $\tau + 1$ , where  $1 \leq \tau \leq t - 1$ :

$$\Pr(o_1, o_2, \dots, o_{\tau+1}, w_{\tau+1} = s_i) = b_i(o_{\tau+1}) \sum_{j=1}^N r_{ji} \Pr(o_1, o_2, \dots, o_{\tau}, w_{\tau} = s_j). \tag{12}$$

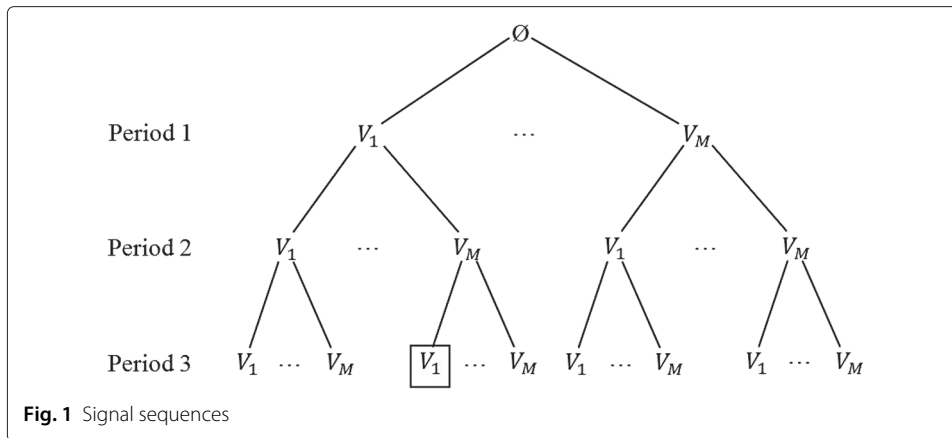
Iteratively running the induction step gives the value of  $\Pr(o_1, o_2, \dots, o_t, w_t = s_i)$ . Then, (9) can be calculated as:

$$\Pr(w_t = s_i|o_1, o_2, \dots, o_t) = \frac{\Pr(o_1, o_2, \dots, o_t, w_t = s_i)}{\sum_{i=1}^N \Pr(o_1, o_2, \dots, o_t, w_t = s_i)}. \tag{13}$$

However, to find the equilibrium, we must know the hidden state's distribution for any period and for any possible signal sequence. Obviously, there is an enumeration method: calculate the distribution at each period given each signal sequence using the above algorithm until all possible cases are enumerated. But this method might waste a significant amount of computing resources. For example, when calculating  $\Pr(w_3 = s_1|v_1, v_2, v_2)$ , we need to obtain  $\Pr(v_1, v_2, w_2 = s_i)$  during induction, but this value has been obtained when we calculate  $\Pr(w_2 = s_1|v_1, v_2)$ . Such repeat calculation is not necessary. With this concern, we develop an algorithm that can output all the distributions quickly without any repeat calculation.

Given signal set  $V = v_1, v_2, \dots, v_M$ , there are  $M + M^2 + \dots + M^T = \frac{M^{T+1}-M}{M-1}$  possible signal sequences over the  $T$  periods, including  $(v_1), (v_1, v_M), (v_1, v_M, v_2), \dots$ . We arrange these sequences as a tree, as Fig. 1 shows, where each node represents a signal sequence. These nodes, from top to bottom and left to right, are indexed as  $X^0, X^1, X^2, X^3 \dots$ . For example, the node in a square represents signal sequence  $(v_1, v_M, v_1)$  at period  $t = 3$  and is indexed as  $X^{2M^2+1}$ .

In the new algorithm, the forward variable  $\Pr(X^j, w_t = s_i)$  ( $t$  is implied by  $X^j, 1 \leq j \leq \frac{M^{T+1}-M}{M-1}$ ) is stored in the node  $X^j$ . For convenience, we denote  $\Pr(X^j, w_t = s_i)$  as  $\xi_i(j)$  and denote the hidden state's distribution  $\Pr(w_t = s_i|X^j)$  as  $\eta_i(j)$ . From (12), we know that  $\xi_i(j)$



only depends on its parent node (i.e.,  $\xi_i(\text{par}(X^j))$ ). Due to this property, the algorithm is developed as follows:

- for  $j = 1, 2, \dots, M$

$$\xi_i(j) = \pi_i b_i(o_j), \quad 1 \leq i \leq N$$

- for  $j = M + 1, M + 2, \dots, \frac{M^{T+1}-M}{M-1}$

$$\xi_i(j) = b_i(\text{late}(j)) \sum_{k=1}^N r_{ki} \xi_k(\text{par}(j)), \quad 1 \leq i \leq N$$

(late( $j$ ) indicates the last signal in  $X^j$ , e.g.,  $\text{late}(v_1, v_M, v_2) = v_2$ )

- for  $j = M + 1, M + 2, \dots, \frac{M^{T+1}-M}{M-1}$

$$\text{Output : } \eta_i(j) = \frac{\xi_i(j)}{\sum_{k=1}^N \xi_k(j)}, \quad 1 \leq i \leq N$$

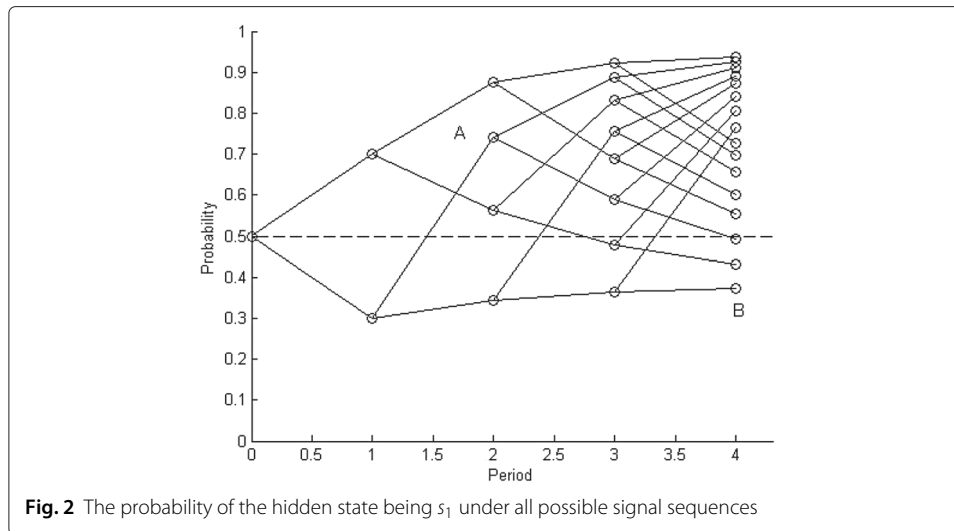
Then, we have the distribution  $\eta_i(j)$  given all possible signal sequence  $X^j$ .

With this algorithm, we examine the HMM used in the given example of the “An Example” section. The probability distributions of the hidden state under all possible signal sequences are shown in Fig. 2. In this figure, each node (the small circle) represents a signal sequence, just as Fig. 1. Figure 2 can be thought of as Fig. 1 turning 90 degree. For example, node *A* indicates signal sequence  $(v_2, v_1)$  observed at period 2, with hidden state  $w_2$ 's distribution being  $\Pr(w_2 = s_1 | v_2, v_1) = 0.740$  (implying  $\Pr(w_2 = s_2 | v_2, v_1) = 1 - 0.740 = 0.260$ ). Hereafter, results of probability distributions will be presented in such figures.

### Application

In this section, we apply our model to the highly controversial education issues, referred to as *Chinese education game*. From the perspective of the static game with perfect information, schools, overwhelmed by examination-oriented education, are trapped in the classic prisoner's dilemma. However, the background is changing over time. The education game should contain multiple periods, and participants make decisions dynamically under imperfect information. Our model emphasizes the dynamic and informational aspects, and investigates on questions: why is examination-oriented education so popular and under which conditions will this dilemma be overcome?





### Chinese Education Game

To simplify the problem, we assume there are two schools: school 1 and 2; each of them has two choices: be examination-oriented ( $Y$ ), or not ( $N$ ). The schools' payoff closely relates to factors such as university admission criteria, government policies, and parents' expectations. We refer to these factors as background, which has two states: negative ( $s_1$ ) and positive ( $s_2$ ). Schools' payoff matrices under the two states are shown in Tables 3 and 4. Under negative background  $s_1$ , schools' strategy profile ( $N, N$ ) is optimal for both schools, but it is not stable. In fact, ( $Y, Y$ ) (both schools being examination-oriented) is the only Nash equilibrium, which is exactly the prisoner's dilemma. On the contrary, under the positive state  $s_2$ , the ( $N, N$ ) becomes the only equilibrium.

However, the background might not be completely observed by schools, that is,  $s_1$  and  $s_2$  are hidden states. The background cannot be measured, but it can be estimated by some indicators (signals). Let  $v_1$  and  $v_2$  be the signals, where  $v_i$  ( $i = 1, 2$ ) means  $s_i$  is more likely to happen. We assume that the background evolves as a Markov chain over time. Thus, the background process is actually an HMM. Relevant parameters are set as follows (also shown in Fig. 3):

- Prior belief on the background (hidden states):  $\pi = \{0.8, 0.2\}$
- State transition probability distribution  $R = \begin{pmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \end{pmatrix}$  (The setting implies the background is more likely to stay on  $s_1$  meaning turning positive is more difficult than turning negative.)
- Observation signal probability distribution  $B = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$  (That is, schools have a probability of 0.7 to observe the background correctly.)

**Table 3** Under negative background  $s_1$

		School 2	
		Y	N
School 1	Y	2,2	4,1
	N	1,4	3,3

**Table 4** Under positive background  $s_2$

		School 2	
		Y	N
School 1	Y	2,2	4,3
	N	3,4	5,5

Then, we are ready to apply the *hidden Markov Bayesian game* to analyze this problem. (The discounting factor is set as  $\delta = 1$ ).

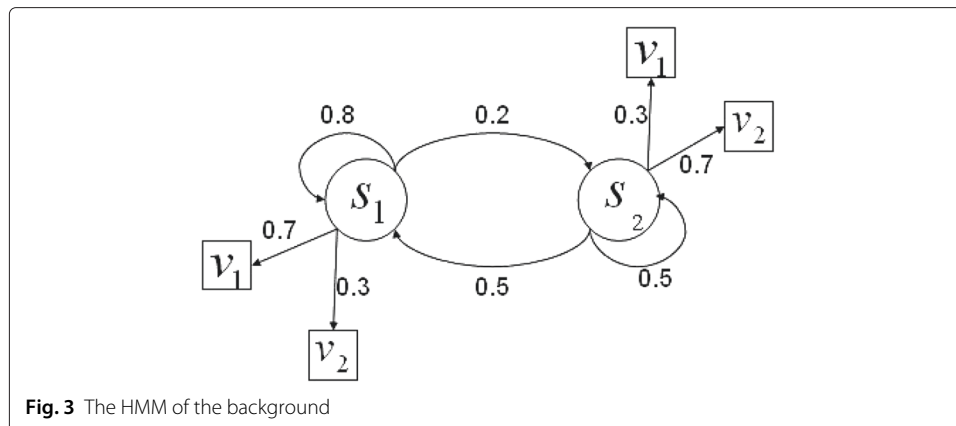
**Equilibrium**

Let  $p$  denote the probability of the background being negative ( $s_1$ ). For one period, the Nash equilibrium can be easily obtained, which is  $(N, N)$  if  $p < 1/2$ ,  $(Y, Y)$  or  $(N, N)$  if  $p = 1/2$ , and  $(Y, Y)$  otherwise. So, both schools will not choose being examination-oriented only if  $p < 1/2$ .

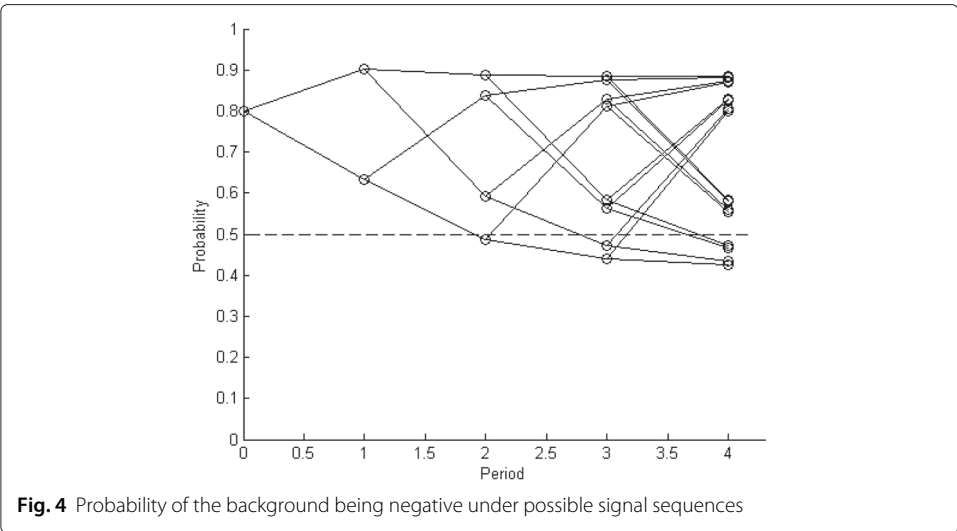
For multiple periods, we calculate the probability distribution of background by the algorithm proposed in the last section. The result is provided in Fig. 4. Figure 4 shows an interesting pattern: at period  $t$  ( $t \geq 2$ ), if schools observe signal  $v_2$  at both periods  $t$  and  $t - 1$  (at each node, the two nodes connected to it on the right side are its child nodes; the upward child means observing signal  $v_2$ , the downward one means observing  $v_1$ ), the probability of background being negative will be smaller than 0.5; otherwise, this probability will be greater than 0.5. In other words, schools will believe background is more likely to be positive only when the signal sequence is  $(\dots, v_2, v_2)$ . Therefore, the Nash equilibrium (also a subgame perfect equilibrium) of the education game is the following strategy profile:

- School 1 chooses not being examination oriented ( $N$ ) when observing signal  $v_2$  in the latest two consecutive periods (i.e., the sequence is  $(\dots, v_2, v_2)$ ); school 1 chooses being examination oriented ( $Y$ ) otherwise.
- School 2 uses the same strategy as school 1.

This equilibrium might suggest that schools tend to prefer examination-oriented education if the education policy keeps changing. It also implies the importance of a stable positive background to schools' education choices.



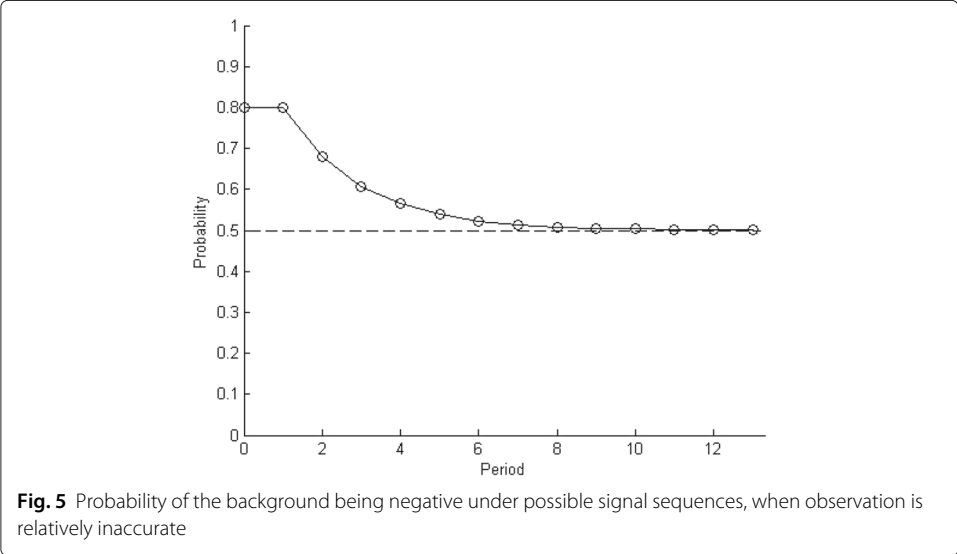
**Fig. 3** The HMM of the background



**The Effect of Observation Accuracy**

The observation signal distribution  $B$  represents the relationship between signals and the hidden states. It also represents the observation accuracy. Our previous setting of  $B$  indicates a case of relatively accurate observation. Now, we reset  $B' = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$  to indicate the case of relatively inaccurate observation, which may be caused by schools lacking information or the bad communication between schools and other parties. Under  $B'$ , we calculate the background distribution with our algorithm (see Fig. 5).

In Fig. 5, the curve converges to 0.5, indicating that when observation is inaccurate, schools' estimates on the hidden background are indifferent to the signals. No matter what signals they observe, their belief are the same, and they will keep choosing examination-oriented education that is the only equilibrium. The result suggests the importance of information or observation accuracy, because schools lacking information might not change their previous education way.



## Conclusions

In this paper, we introduce a repeated game with hidden Markovian states. The motivation is to formulate and analyze situations where players lack full information but can observe signals, and decisions are made simultaneously for several periods. We also define the Nash equilibrium and develop an algorithm to facilitate finding the equilibrium.

We then apply the model to controversial education issues. The results show that schools tend to prefer examination-oriented education if the education policy keeps changing. Also, schools lacking information might keep choosing their previous way of education. Such insights highlight the impact of the observed signal sequence in players' behavior, which may not be derived from other game models.

To implement the model for practical applications, we may estimate parameters of the embedded HMM using a Markov chain Monte Carlo method [12]. For the purpose of behavior analysis (with hypothetical parameter settings), we can make simplifications to the model for tractability, such as assuming players all observe identical signals.

For future research, one may extend our model to incorporate other features, for example, only one player can observe signals or the number of periods is not fixed. One may analyze situations where beliefs can be updated based on previous actions of a more informed player. The study of other equilibria including the mixed-strategy equilibrium and the perfect Bayesian equilibrium should be important. It would also be interesting to apply the model to other cases and generate insights.

## Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No. 61074193 and 61374082).

## Author details

<sup>1</sup>School of Business Administration, Hunan University, Changsha 410082, China. <sup>2</sup>School of Information, Renmin University of China, Beijing 100872, China.

Received: 30 September 2015 Accepted: 4 March 2016

Published online: 11 March 2016

## References

1. Fudenberg, D, Tirole, J: *Game Theory*. MIT Press, Cambridge (1991)
2. Fudenberg, D, Levine, D: Subgame-perfect equilibria of finite-and infinite-horizon games. *J. Econ. Theory*. **31**(2), 251–268 (1983)
3. Mailath, GJ, Samuelson, L: *Repeated Games and Reputations: Long-Run Relationships* (Vol. 2). Oxford university press, Oxford (2006)
4. Dekel, E, Fudenberg, D, Levine, DK: Learning to play Bayesian games. *Games. Econ. Behavior*. **46**(2), 282–303 (2004)
5. Harsanyi, JC: Games with incomplete information played by "Bayesian" players, Part I–III. Part I. The basic model. *Manag. Sci.* **14**, 159–182 (1967)
6. Rabiner, LR: A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*. **77**(2), 257–286 (1989)
7. Megiddo, N: On repeated games with incomplete information played by non-Bayesian players. *Int. J. Game Theory*. **9**(3), 157–167 (1980)
8. Shapley, LS: Stochastic games. *PNAS*. **39**(10), 1095–1100 (1953)
9. Aumann, RJ, Maschler, M, Stearns, RE: *Repeated Games with Incomplete Information*. The MIT press, Cambridge (1995)
10. Bishop, CM: *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York (2006)
11. Liu, F, Janssens, D, Cui, J: Characterizing activity sequences using profile Hidden Markov models. *Expert Syst. Appl.* **42**(13), 5705–5722 (2015)
12. Netzer, O, Lattin, JM, Srinivasan, V: A hidden Markov model of customer relationship dynamics. *Market. Sci.* **27**(2), 185–204 (2008)
13. Dias, JG, Ramos: Dynamic clustering of energy markets: an extended hidden Markov approach. *Expert Syst. Appl.* **41**(17), 7722–7729 (2014)
14. Ansari, A, Montoya, R, Netzer, O: Dynamic learning in behavioral games: a hidden Markov mixture of experts approach. *QME-Quant. Mark. Econ.* **10**(4), 475–503 (2012)
15. Kujala, JV, Richardson, U, Lyytinen, H: A Bayesian-optimal principle for learner-friendly adaptation in learning games. *J. Math. Psychol.* **54**(2), 247–255 (2010)
16. Osborne, MJ: *An Introduction to Game Theory*. Shanghai University of Finance and Economics Press, Shanghai (2005)
17. Chiu, SY, Pu, YJ: *Game Theory and Economic Modeling*, in Chinese. Renmin University Press, Beijing (2010)