Jiangjing Zhou, Ovanes Petrosian, Bayesian learning in fish wars: dynamic estimation of unknown states and private information, *Mat. Teor. Igr Pril.*, 2024, Volume 16, Issue 2, 92–112

Математическая Теория Игр и её Приложения, т.16, в.2, с. 92–112

УДК 519.83, 004.42
ББК 22.18

БАЙЕСОВСКОЕ ОБУЧЕНИЕ В «РЫБНЫХ ВОЙНАХ»: ДИНАМИЧЕСКАЯ ОЦЕНКА НЕИЗВЕСТНЫХ СОСТОЯНИЙ И ЧАСТНОЙ ИНФОРМАЦИИ*

Цзянцзин Чжоу
Ованес Петросян
Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., 7/9
e-mail: st092028@student.spbu.ru, petrosian.ovanes@yandex.ru

В статье исследуется новый вариант «рыбной войны» – игры, в которой участники оценивают неизвестные параметры окружающей среды и частную информацию противников на основе полученных сигналов. Строится динамическая игровая модель, в которой эволюция популяции рыб зависит от неизвестного параметра ϵ, а функция выигрыша каждого игрока включает его частную информацию δ. С помощью байесовских методов обучения демонстрируется, как участники могут динамически обновлять свои оценки указанных неизвестных параметров. Доказано, что эти оценки сходятся к истинным значениям параметров с течением времени. Построено равновесие Нэша с байесовским обучением в качестве решения данной игры. Приведены результаты численного моделирования, иллюстрирующие совпадение представлений игроков и сравнение их стратегий управления в различных случаях.

---

* Исследование выполнено при поддержке гранта Санкт-Петербургского государственного университета № 94-06-2114.

Ключевые слова: динамическое байесовское обучение, игра, рыбная война, частная информация, неизвестные параметры.

Поступила в редакцию: 26.01.24 *После доработки*: 18.03.24 *Принята к публикации*: 03.06.24

1. Введение

Исторически теория игр обеспечивает надежный аппарат для понимания стратегических взаимодействий [11, 10, 18]. Однако, традиционные модели часто предполагают наличие идеальной информации – условие, редко встречающееся на полях экологических и экономических сражений. «Рыбная война» [6, 3], игра, описывающая конкурентную добычу общего ресурса, служит идеальной основой для изучения стратегической динамики в условиях неопределенности. В работе [17] изучены неопределенности в динамике экологических процессов и в значениях коэффициентов дисконтирования. В связи с этим, актуальной задачей является моделирование неопределенностей в динамике экологической системы, а также влияния неопределенности коэффициентов дисконтирования на оценку долгосрочной политики. Учитывая эти неопределенности, крайне важно включить их в модели. Примерами такого подхода являются работы [16, 21, 22].

Недавние исследования, как подчеркивается в статьях [20, 4, 1], показывают важность байесовского обучения для решения задач оценки неизвестных параметров в различных областях. В статье [20] продемонстрирована эффективность байесовского обучения при оценке неопределенных функций спроса. В статье [4] дано математическое описание неизвестной динамики в нелинейных системах. В статье [1] разработаны почти оптимальные алгоритмы управления для моделей массового обслуживания с неизвестными марковскими процессами принятия решений.

Во статьях [13–15, 19] акцент сделан на исследование неопределенности в уравнениях динамики. В частности, в [13] исследована динамика обучения в контексте международной борьбы с загрязнением в условиях экологической неопределенности. В статье [14] изучено влияние обучения на будущие выигрыши и выделены двойные источники риска – структурная неопределенность, присущая уравнениям состояния, и неопределенность, возникающая из-за ожидания обучения. Процессы обучения, связанные со стохастической динамикой, управляющей эволюцией государственного капитала, рассмотрены в публикации [15], с акцентом на то, как технологические инвестиции влияют на будущие показатели акций. В [19] исследована двухшаговая игра управления выбросами загрязняющих веществ с двумя участниками. В этом случае потенциальный ущерб от выбросов характеризуется диапазоном значений, каждое из которых связано с определенными вероятностями.

В рамках игрового подхода к моделированию «рыбной войны», представленного ниже, основное внимание уделяется неопределенностям в функциях выигрыша относительно коэффициентов дисконтирования игроков. Для решения этой проблемы используется динамический байесовский механизм обучения. При таком подходе игроки постоянно корректируют свои оценки неизвестных параметров в свете новой информации (см., например, [9]). Этот метод является инновационным поскольку преодолевает ключевое ограничение, часто встречающееся в классических байесовских играх [8], а именно, предположение о статических представлениях игрока. При динамически меняющихся представлениях, модель точнее отражает реальную ситуацию и требует корректировки на основе новых данных и опыта.

Основная цель данного исследования – изучить неопределенности, присущие уравнениям динамики и функциям выигрыша в рамках игровых моделей. В частности, неопределенность в уравнениях динамики возникает из-за неизвестных реализаций случайных величин на каждом временном шаге, в то время как неопределенность в функциях выигрыша объясняется асимметричной информацией. В работе демонстрируется эффективность использования байесовского обучения для оценки этих неопределенностей. Как показано ниже, с течением времени оценки игроков все больше приближаются к истинным значениям указанных неизвестных параметров. Такой подход обеспечивает устранение неопределенностей, связанных как с динамическим характером игровой среды, так и с асимметричным распределением информации между игроками. С помощью теоретического анализа и численного моделирования подтверждается эффективность байесовского обучения в повышении точности оценок и стратегий игроков в неопределенных динамических игровых условиях.

Настоящее исследование вносит двойной вклад: расширяется теоретическое понимание байесовского обучения в рамках теории игр, а также предлагается практический инструмент для игроков в играх с неполной информацией. Более того, эта работа демонстрирует эволюционную природу формирования стратегий в динамических системах с неопределенностями.

Статья имеет следующую структуру. В разделе 2 описывается основная модель, в разделе 3 – байесовский механизм обучения. Раздел 4 посвящен сравнительному анализу равновесия Нэша. Результаты численного моделирования представлены в разделе 5. Наконец, в заключении обобщаются результаты данного исследования применительно к принятию стратегических решений в экологических играх.

2. Модель

Рассмотрим игровую модель «рыбной войны» [2, 5, 12], а именно, динамическую структуру стратегического взаимодействия при добыче общего рыбного ресурса. Модель включает уравнение динамики и функции выигрыша участников.

Динамика развития популяции рыб описывается уравнением

$$S_{t+1} = (E(\epsilon|\cdot)S_t - \sum_{i=1}^{2} u_i(t,\cdot))^{\alpha},$$

где:

- S_t – размер популяции рыб в момент времени (на шаге) t,
- $u_i(t,\cdot)$ – улов игрока i на шаге t, $i = 1, 2$,
- ϵ – параметр, отражающий экологические и биологические факторы роста популяции рыб,
- $E(\epsilon|\cdot)$ – условное математическое ожидание случайной переменной ϵ на основе текущих представлений игрока,
- α – константа, отражающая характер динамики популяции.

Параметр ϵ рассматривается как случайная величина, что вносит неопределенность в модель. В самом деле, реальные темпы роста популяции рыб зависят от различных неизвестных факторов окружающей среды. Реализация ϵ на каждом шаге неизвестна игрокам до выбора их стратегий. Учитывая неопределенность ϵ, игроки должны оценивать значение этого неизвестного параметра на каждом шаге прежде чем принимать свои решения. Эта оценка имеет решающее значение, поскольку она напрямую влияет на их стратегию (вылов).

В реальных ситуациях указанные факторы часто непредсказуемы и подвержены различным внешним воздействиям, что делает ϵ по своей сути неопределенным. В частности, в данной модели неопределенность ϵ характеризуется неизвестным средним и дисперсией. Эта неопределенность представляет собой естественную непредсказуемость окружающей среды, включающую климатические условия, качество воды и другие экологические переменные. Их трудно измерить точно, и они подвержены изменениям с течением времени.

Предположение о том, что математическое ожидание (МО) и дисперсия случайной величины ϵ неизвестны, является расширением классических моделей. (Как правило, эти параметры считаются известными или могут быть оценены с высокой точностью.) В рамках данной модели, отсутствие точной информации о ϵ требует от игроков более осторожного и адаптивного подхода. Они должны постоянно обновлять свои представления и стратегии на основе наблюдаемых результатов, что более реалистично: лица, принимающие решения, часто действуют в условиях значительной неопределенности относительно параметров окружающей среды. Такая модель отражает суть реальных процессов принятия решений взадачах управления природными ресурсами, где заинтересованные стороны часто сталкиваются с неполной информацией об управляемых системах.

Определим выигрыш для каждого игрока в виде

$$J_i = \sum_{t=0}^{\infty} \delta_i^t \log u_i(t, \cdot), \quad i = 1, 2,$$

где δ_i – коэффициент дисконтирования, отражающий временные предпочтения игрока i. Таким образом, целью каждого игрока является максимизация его совокупной дисконтированной полезности от рыбного промысла на бесконечном горизонте.

Коэффициент дисконтирования δ_i каждого игрока является его частной информацией, недоступной другим игрокам. Предположение о том, что временные предпочтения каждого игрока – частная информация, является ключевым аспектом модели. Это предположение основано на реалистичном представлении о том, что игроки имеют уникальные (в своем роде) личные оценки будущей прибыли. При добыче ресурсов (напр., в «рыбной войне»), игроки могут иметь разное терпение, отношение к риску, а также разные стратегии долгосрочного планирования. Эти различия отражены в коэффициенте дисконтирования δ_i.

В рамках рассматриваемой игровой модели вводится понятие функции ожидаемого выигрыша. Эта функция имеет решающее значение для игроков, принимающих стратегические решения на основе своих представлений о неизвестных параметрах других игроков. В частности, когда стратегия игрока зависит от неизвестного параметра противника, функция ожидаемого выигрыша позволяет максимизировать выигрыши с помощью оценивания этого параметра.

Определим функцию ожидаемого выигрыша для игрока $i = 1, 2$, имеющего некоторое представление о параметре соперника δ_j, следующим образом:

$$EJ_i(u_i, u_j; x_0) = \max_{u_i} \sum_{t=0}^{\infty} \sum_{\delta_j \in D_j} \left[p_i^t(\delta_j | y_j(t-1))(\delta_i^t \log u_i(t, S, \delta_i, \cdot)) \right],$$

где:

- $p_i^t(\delta_j|y_j(t-1))$ – представление игрока i о типе δ_j игрока j на (текущем) шаге t, основанное на сигнале $y_j(t-1)$ на (предыдущем) шаге $t-1$,

- $u_i(t, S, \delta_i, \cdot)$ – решение, принимаемое игроком i на шаге t, которое зависит от его типа δ_i.

Функция ожидаемого выигрыша представляет собой усовершенствованную модель принятия решений, в которой игроки включают свои ожидания и оценки относительно стратегий оппонентов в свое стратегическое планирование. Анализируя сигналы и наблюдения, игроки могут формировать свои представления δ_i о типах оппонентов, а также делать выводы о неизвестных параметрах ϵ.

3. Байесовский подход к обучению

В данном разделе рассматриваются теоретические основы байесовского обучения в рамках нашей игровой модели, чтобы показать, как игроки динамически уточняют свои оценки неизвестных параметров. Применяя этот подход, игроки могут ориентироваться в условиях неопределенности и неполной информации, характерных для стратегических взаимодействий в задачах управления ресурсами. В результате процессы принятия решений в большей степени используют имеющуюся информацию и становятся потенциально оптимальными.

В области байесовской оценки нормальных распределений начнем с нормально распределенной случайной величины ϵ. Задача заключается в изучении неизвестных параметров распределения, $\theta' = (\mu, \sigma^2)$, по серии шагов t в диапазоне от 0 до ∞, где априорные представления об указанных параметрах образуют вектор $\xi_t(\theta')$. В этой байесовской схеме основное внимание уделяется точности нормального распределения, определяемой в виде $\lambda = \frac{1}{\sigma^2}$ (положительное число, равное обратной величине дисперсии). Высокая точность соответствует низкой дисперсии, и наоборот. Для байесовской оценки представляют интерес параметры μ и λ, которые объединяются в вектор параметров $\theta = (\mu, \lambda)$ для описания распределения. Здесь μ – МО, а точность λ характеризует разброс распределения, будучи обратно пропорциональной дисперсии.

Обозначим $\xi_t(\theta)$ – априорное совместное распределение параметров μ и λ байесовской схемы оценивания на шагах $t = 0, 1, 2, ...$ Начальное распределение имеет вид

$$\begin{aligned}\xi_0(\theta) &= \mathcal{N}\left(\mu|\mu_0, (\kappa_0\lambda)^{-1}\right) Ga\left(\lambda|\alpha_0,\ \text{rate } = \beta_0\right)\\ &= \frac{1}{Z_{NG}\left(\mu_0, \kappa_0, \alpha_0, \beta_0\right)} \lambda^{\frac{1}{2}} \exp\left(-\frac{\kappa_0\lambda}{2}\left(\mu-\mu_0\right)^2\right) \lambda^{\alpha_0-1} e^{-\lambda\beta_0}\\ &= \frac{1}{Z_{NG}} \lambda^{\alpha_0-\frac{1}{2}} \exp\left(-\frac{\lambda}{2}\left[\kappa_0\left(\mu-\mu_0\right)^2 + 2\beta_0\right]\right),\end{aligned}$$

где $Z_{NG}\left(\mu_0, \kappa_0, \alpha_0, \beta_0\right) = \frac{\Gamma(\alpha_0)}{\beta_0^{\alpha_0}}\left(\frac{2\pi}{\kappa_0}\right)^{\frac{1}{2}}$; Ga – гамма-распределение; μ_t представляет собой априорную оценку или представление о МО μ на шаге t; κ_t – параметр, связанный с точностью нормального распределения для μ. (Более высокие значения κ_t соответствуют более высокой точности.) Для гамма-распределения с параметром точности λ, величины α_t и β_t являются параметрами формы и скорости распределения.

После наблюдения сигнала $x_t \in X_t$ на каждом шаге t применяется байесовский вывод для вычисления апостериорного распределения $\hat{\xi}_t(\theta)$ параметра θ. Это распределение имеет следующий вид:

$$
\begin{aligned}
&\hat{\xi}_t(\theta) \propto \phi(x_t|\mu,\lambda)\xi_t(\theta)\\
&\propto \quad \lambda^{\frac{1}{2}} e^{-\left(\kappa_t\lambda(\mu-\mu_t)^2\right)/2}\lambda^{\alpha_t-1}e^{-\beta_t\lambda}\times\lambda^{1/2}e^{-\frac{\lambda}{2}(x_t-\mu)^2}\\
&\propto \quad \lambda^{\frac{1}{2}}\lambda^{\alpha_t+1/2-1}e^{-\beta_t\lambda}e^{-(\lambda/2)[\kappa_t(\mu-\mu_t)^2+(x_t-\mu)^2]}
\end{aligned}
$$

$$
\propto\mathcal{N}\left(\mu \mid \frac{\kappa_t\mu_t+x_t}{\kappa_t+1},((\kappa_t+1)\lambda)^{-1}\right)\times Ga\left(\lambda \mid \alpha_t+\frac{1}{2},\beta_t+\frac{\kappa_t\left(x_t-\mu_t\right)^2}{2\left(\kappa_t+1\right)}\right).
$$

Таким образом, после наблюдения сигнала x_t, представления о параметре θ меняются с априорного $\xi_t(\theta)$ на апостериорное $\hat{\xi}_t(\theta)$ распределение (обновляются на основе новых фактических данных).

После введения байесовского механизма обновления представлений для одного шага, предположим, что апостериорное представление игрока на шаге t становится его априорным представлением на шаге $t+1$. Это предположение позволяет получить следующую итеративную процедуру эволюции представлений.

Предложение 3.1.

$$
\begin{aligned}
\mu_{t+1} &= \frac{\kappa_t\mu_t+x_t}{\kappa_t+1},\\
\kappa_{t+1} &= \kappa_t+1,\\
\alpha_{t+1} &= \alpha_t+\frac{1}{2},\\
\beta_{t+1} &= \beta_t+\frac{\kappa_t\left(x_t-\mu_t\right)^2}{2(\kappa_t+1)},
\end{aligned}
$$

где исходное представление задано $\mu_0, \kappa_0, \alpha_0$ и β_0. Игроки итеративно уточняют эти представления на основе новых наблюдений и накопленной информации, обеспечивая динамическое и последовательное обновление представлений с течением времени.

Данное утверждение играет ключевую роль в байесовской схеме обучения. Оно формализует механизм, с помощью которого участники игры обновляют свои представления о критических параметрах на каждом шаге. Этот итеративный процесс, см. уравнение выше, гарантирует, что система представлений каждого игрока не является статической, а непрерывно меняется по мере поступления новой информации. В частности, согласно правилам обновления для μ_{t+1}, κ_{t+1}, α_{t+1} и β_{t+1}, игроки усваивают новые наблюдения (x_t) для уточнения своих оценок МО и точности распределения.

Важно отметить, что по предположению все игроки получают согласованные сигналы о неизвестных параметрах в уравнениях динамики и являются рациональными. Следовательно, при одних и тех же сигналах оценки неизвестных параметров различными игроками совпадают: оценки зависят исключительно от принятых сигналов.

Чтобы доказать сходимость представлений, введем понятие случайной величины M_t, отражающей представление об МО распределения на шаге t. При стремлении t к бесконечности M_t сходится к истинному МО μ случайной величины ϵ; см. теорему ниже. Эта сходимость подразумевает, что независимо от вариаций в наблюдениях оценка M_t со временем стабилизируется и согласуется с неизвестным истинным средним μ.

Поскольку представления на шаге $t = 0$ предопределены или известны заранее, то исследуем эволюцию представлений при $t \geq 1$, где они динамически обновляются с течением времени.

Теорема 3.1. *Оценка неизвестного МО стремится к истинному значению с течением времени:*

$$\lim_{t\to\infty} M_t = \mu,$$

где M_t – случайная величина, описывающая представление μ_t о неизвестном МО μ на шаге $t \geq 1$.

Доказательство. Рассмотрим бесконечную последовательность X_0, $X_1, X_2, \ldots$, состоящую из независимых одинаково распределенных интегрируемых по Лебегу случайных величин. Последовательность M_t, являющаяся функцией этих случайных величин, определяется следующим образом:

$$M_t = \frac{\kappa_0\mu_0}{\kappa_0 + t} + \frac{1}{\kappa_0 + t}\left(X_0 + X_1 + \cdots + X_{t-1}\right).$$

Согласно закону больших чисел, среднее последовательности независимых одинаково распределенных случайных величин сходится к МО при бесконечно большом числе наблюдений. Поэтому получаем:

$$\lim_{t\to\infty} \frac{X_0 + X_1 + \cdots + X_{t-1}}{t} = \mu.$$

Следовательно, пределом M_t при стремлении t к бесконечности является МО μ:

$$\lim_{t\to\infty} M_t = \mu.$$

Таким образом, байесовское обучение эффективно: представления с течением времени согласуются с истинными значениями параметров. $\square$

Учитывая, что коэффициент дисконтирования является константой, тип игрока j инвариантен на шагах t и $t+1$ и может быть прямо задан как δ_j, т.е.

$$\delta_j(t+1) = \delta_j(t) = \delta_j, \quad \forall t \in \{0, 1, 2, ..., \infty\}.$$

В рассматриваемой модели игрок i стремится оценить параметр δ_j игрока j. Предполагается, что этот параметр постоянен, а игрок i получает сигнал

$$y_j = \delta_j + w_j,$$

где w_j – белый шум с дисперсией R_j. Чтобы оценить этот неизвестный параметр, игрок i использует фильтр Калмана (специальную форму байесовского обучения [7] для линейных систем с гауссовским шумом).

Такой сигнал характерен в случае наблюдения возмущенной величины истинного параметра, содержащей белый шум. Белый шум часто используется в теории обработки сигналов и теории управления из-за его простоты и математической разрешимости, поскольку он представляет собой случайные некоррелированные флуктуации, влияющие на сигнал.

Использование фильтра Калмана оправдано благодаря его эффективности при оценке состояния линейной системы с гауссовским шумом. С этой целью фильтры Калмана широко применяются в различных приложениях, поскольку они позволяют получить оценку истинного состояние системы (в данном случае, параметр δ_j) из зашумленных наблюдений. Фильтр работает в два этапа: прогнозирование и обновление. На этапе прогнозирования оценка текущего состояния используется для получения оценки для следующего шага. На этапе обновления следующее измерение используется для уточнения этой оценки, чтобы более лучше оценить состояние системы. Этот процесс согласован с байесовскими принципами, где текущая оценка состояния рассматривается как априорная, а следующая оценка (после наблюдения новых данных) – как апостериорная.

В нашей модели фильтр Калмана используется для итеративного обновления оценки параметра δ_j. Этот метод дает особые преимущества при зашумленных наблюдениях. Фильтр Калмана описывается следующим образом:

Предложение 3.2.

$$\begin{aligned} \hat{\delta}_i(t+1) &= \hat{\delta}_i(t) + \frac{P_i(t)}{P_i(t)+R_i} \cdot \left(y_i(t+1) - \hat{\delta}_i(t)\right), \\ P_i(t+1) &= \frac{P_i(t)R_i}{P_i(t)+R_i}, \end{aligned} \tag{3.1}$$

где $\hat{\delta}_i(0) = \delta_i^0$ *и* $P_i(0) = P_i^0$ *(начальные условия).*

Уравнение (3.1) представляет собой механизм обновления фильтра Калмана. Здесь $\hat{\delta}_i(t+1)$ – апостериорная оценка δ_i на шаге $t+1$, которая вычисляется с использованием предыдущей оценки $\hat{\delta}_i(t)$, дисперсии оценки $P_i(t)$ и нового наблюдения $y_i(t+1)$. Член $P_i(t+1)$ является апостериорной оценкой дисперсии. Начальные условия включают $\hat{\delta}_i(0) = \delta_i^0$ (начальная оценка) и $P_i(0) = P_i^0$ (начальная дисперсия). Этот итеративный процесс позволяет непрерывно уточнять оценку на основе новой информации, воплощая суть байесовского обучения при наличии зашумленных данных.

Игрок i оценивает коэффициент дисконтирования игрока j поразному на основе полученных сигналов. При этом оценка игроком i коэффициента дисконтирования игрока j не зависит от оценки игроком j его собственного коэффициента дисконтирования. Другими словами, $\hat{\delta}_i(t)$ не зависит от $\hat{\delta}_j(t)$. Однако $\hat{\delta}_i(t)$ влияет на стратегические решения игрока $j - u_j$.

Предложение 3.3. *В рассматриваемой модели фильтр Калмана (3.1) эффективно обновляет оценку неизвестного параметра δ_i при заданных начальной оценке δ_i^0, начальной дисперсии P_i^0, а также дисперсии сигнала R_i.*

Фильтр предоставляет механизм для уточнения этой оценки с течением времени, используя новые наблюдения на каждом шаге.

Теорема 3.2. *Для случайных наблюдений $Y_i(0), Y_i(1), ..., Y_i(t-1)$, $t \geq 1$, случайная величина $\hat{T}_i(t)$, представляющая апостериорное значение δ_i на шаге t, сходится к неизвестному истинному значению δ_i^*:*

$$\lim_{t\to\infty} \hat{T}_i(t) = \delta_i^*, \quad \forall i \in N.$$

Доказательство. Рассмотрим итеративное уравнение (3.1) для оценки фильтра Калмана:

$$\hat{T}_i(t) = \left[\prod_{h=1}^{t} \frac{R_i}{R_i + P_i(h-1)}\right] \delta_i^0 + \sum_{m=1}^{t} \left[\prod_{h=m+1}^{t} \left(\frac{R_i}{R_i + P_i(h-1)}\right)\right] \cdot \left(\frac{P_i(m-1)}{P_i(m-1) + R_i}\right) \cdot Y_i(m), \tag{3.2}$$

где $P_i(h)$ – дисперсия оценки δ_i на шаге h, R_i – дисперсия сигналов, а $Y_i(m)$ – сигнал, полученный на шаге m.

Для упрощения вычислений, сначала найдем произведение

$$\prod_{h=1}^{t} \frac{R_i}{R_i + P_i(h-1)} = \frac{R_i}{R_i + P_i^0 t}. \tag{3.3}$$

Тогда сумму можно выразить в виде

$$\sum_{m=1}^{t} \left[\frac{R_i + P_i^0 m}{R_i + P_i^0 t}\right] \cdot \left(\frac{P_i(m-1)}{P_i(m-1) + R_i}\right) \cdot Y_i(m) = \frac{P_i^0 t}{R_i + P_i^0 t} \left[\frac{\sum_{m=1}^{t} Y_i(m)}{t}\right]. \tag{3.4}$$

Подставляя уравнения (3.3) и (3.4) в уравнение (3.2), получим

$$\hat{T}_i(t) = \frac{R_i}{R_i + P_i^0 t}\delta_i^0 + \frac{P_i^0 t}{R_i + P_i^0 t}\left[\frac{\sum_{m=1}^{t} Y_i(m)}{t}\right]. \tag{3.5}$$

По закону больших чисел, имеем

$$\lim_{t\to\infty} \frac{\sum_{m=1}^{t} Y_i(m)}{t} = \delta_i^*.$$

Следовательно, из уравнения (3.5)

$$\lim_{t\to\infty} \hat{T}_i(t) = \delta_i^*. \tag{3.6}$$

□

Таким образом, оцененная случайная величина $\hat{T}_i(t)$ сходится к истинному значению δ_i^* с течением времени, подчеркивая эффективность фильтра Калмана при уточнении оценки неизвестных параметров в динамических моделях.

4. Анализ равновесия Нэша с байесовским обучением

Исследуем, как игроки определяют свои стратегии, используя обновленные представления о неизвестных параметрах уравнения динамики и функций выигрыша. Этот ключевой аспект модели предполагает детальное изучение процесса принятия решений в условиях неопределенности, когда стратегия каждого игрока является реакцией не только на текущее состояние игры, но и на его ожидания и представления о неизвестных параметрах и будущих действиях других игроков.

Равновесие Нэша с динамическим байесовским обучением является важнейшей концепцией решения для байесовских игр. Оно достигается, когда все игроки выбирают свои оптимальные стратегии, учитывая свои эволюционирующие представления и поступающую информацию. Формализуем эту концепцию следующим образом.

Определение 4.1. *Равновесием Нэша с динамическим байесовским обучением в игре называется набор стратегий (u_1^*, u_2^*) таких, что для каждого игрока $i = 1, 2$ и для каждого типа $\tau_i \in T_i$ стратегия u_i^**

дает ожидаемый выигрыш не меньше, чем любая альтернативная стратегия u'_i:

$$EJ_i(u_i^*, u_{-i}^*; S_0) \geq EJ_i(u'_i, u_{-i}^*; S_0), \quad \forall u'_i \in U_i,\, \forall \tau_i \in T_i,\, \forall i = 1, 2.$$

В этом исследовании предлагается новая концепция равновесия, «равновесие Нэша с байесовским обучением», отличающаяся от традиционных равновесий Нэша или Байеса-Нэша. В рамках этого нового подхода, представления и оценки игроков относительно неизвестных параметров учитываются в игре. Следовательно, равновесие Нэша с байесовским обучением предоставляет комплексный механизм анализа игр, в которых игроки должны адаптировать свои стратегии на основе меняющихся представлений о неизвестных игровых элементах.

Определив равновесие Нэша с динамическим байесовским обучением, найдем теперь равновесие Нэша в модели «рыбной войны». Необходимо показать, как рациональные игроки с обновленными представлениями друг о друге и состоянии игры определяют свои оптимальные стратегии.

Теорема 4.1. *В игровой модели рыбной войны, равновесная по Нэшу стратегия с байесовским обучением для игрока i, обозначаемая u_i, имеет вид*

$$u_i(t, S, \delta_i; \hat{\epsilon}(t), \hat{\delta}_j(t)) = \frac{\hat{\delta}_j(t)(1 - \delta_i \alpha)}{\delta_i + \hat{\delta}_j(t) - \delta_i \hat{\delta}_j(t)\alpha} \hat{\epsilon}(t) S_t, \tag{4.1}$$

где:

- $\hat{\delta}_j(t)$ *– оценка игроком i параметра δ_j игрока j на шаге t;*
- $\hat{\epsilon}(t)$ *– коллективная рациональная оценка неизвестного параметра в уравнении динамики.*

Доказательство. Чтобы построить равновесие Нэша с байесовским обучением, рассмотрим динамику наилучшего ответа для каждого типа игроков. Пусть $V_i(S, t, \delta_i)$ обозначает функцию цены игрока i, которая удовлетворяет уравнению Гамильтона–Якоби–Беллмана:

$$V_i(S,t,\delta_i) = \max_{u_i}\left\{\delta_i^t \log u_i + \sum_{\delta_j\in D_j} p_i^t(\delta_j|y_j(\cdot))V_i((\hat{\epsilon}(t)S-u_i-u_j)^\alpha, t+1, \delta_i)\right\}. \tag{4.2}$$

Игровая модель имеет линейную структуру относительно состояния. Поэтому предположим, что функция цены является линейной и может быть представлена в виде

$$V_i(S,t,\delta_i) = A_i\log S + B_i. \tag{4.3}$$

По определению математического ожидания и фильтра Калмана, получаем

$$\sum_{\delta_j\in D_j}[\delta_j\cdot p_i^t(\delta_j|y_j(t-1))] = \hat{\delta}_j(t).$$

Максимизируя свою функцию выигрыша, игрок i рассматривает тип игрока j как свою оценку j, обозначаемую $\hat{\delta}_j$.

Подставляя выражение (4.3) в уравнение (4.2), находим условие равновесия первого порядка:

$$\frac{1}{u_i(t,S,\delta_i,\cdot)} - \alpha\delta_i A_i\frac{1}{\hat{\epsilon}(t)S-u_i(t,S,\delta_i,\cdot)-u_j(t,S,\hat{\delta}_j)} = 0. \tag{4.4}$$

Решая уравнение (4.4) относительно u_i при условии линейной стратегии наилучшего ответа $u_i^* = \gamma_i S$, определяем коэффициенты функции цены:

$$A_i = \frac{1}{1-\delta_i\alpha}.$$

Соответствующая стратегия имеет вид

$$u_i(t,S,\delta_i,\hat{\delta}_j) = \frac{1-\delta_i\alpha}{\delta_i}\frac{\hat{\delta}_j}{1-\hat{\delta}_j\alpha}u_j(t,S,\hat{\delta}_j). \tag{4.5}$$

Наконец, подставляя формулу (4.5) обратно в уравнение (4.4), приходим к равновесию Нэша с байесовским обучением (4.1).

Данный результат отражает суть равновесия Нэша с байесовским обучением: равновесная стратегия игрока i зависит от его собственного типа δ_i, его оценки типа противника $\hat{\delta}_j(t)$ и коллективной оценки неопределенного параметра $\hat{\epsilon}(t)$. □

Доказанная теорема описывает равновесную стратегию в модели «рыбной войны». А именно, на решение каждого игрока на любом шаге влияет его тип, его оценка типа противника, а также коллективная оценка неизвестных параметров, характеризующих состояние игры. Таким образом, равновесие Нэша возникает как функция этих факторов, демонстрируя сложный механизм взаимодействия стратегического мышления, обновления представлений и адаптивного принятия решений в игре.

5. Численное моделирование

В настоящем разделе приведены результаты численного моделирования для процесса принятия стратегических решений в экологических играх. На первом из них показана сходимость оценок экологической неопределенности игроков к истинным условиям окружающей среды (процесс обучения и адаптации); на втором – точность восприятия игроками коэффициентов дисконтирования друг друга (важно для понимания временного стратегического планирования); на третьем – различия равновесных по Нэшу стратегий в случаях полной информации и неопределенности. Таким образом, продемонстрировано динамическое взаимодействие стратегий игроков, механизмов обучения и непредсказуемости окружающей среды.

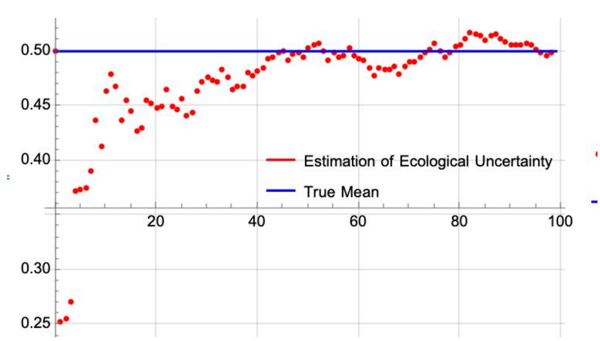

Рисунок 1. Оценка неопределенности окружающей среды в динамических моделях

Рис.1 иллюстрирует процесс оценивания экологической неопределенности в рамках динамической модели. Со временем оценки игрока приближаются к истинному МО, т.е. улучшается его понимание переменных окружающей среды в игре. Эта сходимость показывает, что игрок эффективно использует новую информацию для уточнения своей модели (стратегическое обучение в игре).

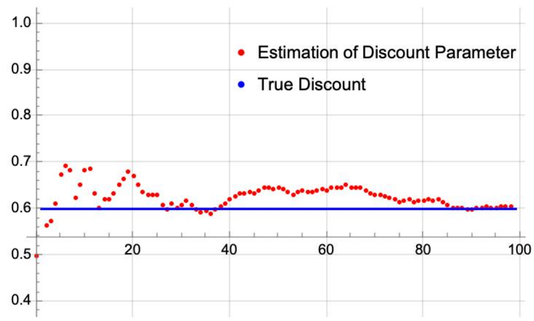

Рисунок 2. Оценка игроком i коэффициента дисконтирования игрока j

На рис. 2 представлена оценка игроком i коэффициента дисконтирования игрока j. Разрыв между расчетным и истинным значениями постепенно сокращается: происходит уточнение прогнозов игрока i относительно временных предпочтений игрока j. По мере развития игры и проведения дополнительных наблюдений игрок i корректирует свои представления, в результате получаются оценки, которые более точно отражают фактическое дисконтирование их партнера.

На рис. 3 представлен сравнительный анализ равновесных по Нэшу стратегий в двух случаях: полная информация и неопределенность. Красной линией отмечены стратегии, сформированные при полной информации, демонстрирующие быстрое снижение (агрессивные первоначальные стратегии, которые быстро стабилизируются). Синей линией отмечены стратегии при неопределенности (постепенное снижение, более осторожное принятие решений). Более того, с течением времени стратегии игроков в условиях неопределенности постепенно приближаются к стратегиям в случае полной информации. Игроки уточняют свои оценки неизвестных параметров, чтобы приблизиться к истинным значениям.

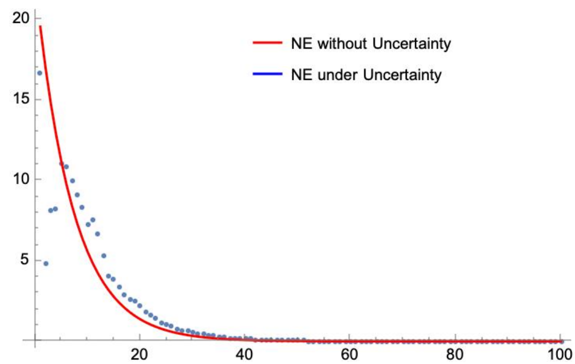

Рисунок 3. Равновесия Нэша: случаи с неопределенностью и без нее в игровых моделях

6. Заключение

Подведем итоги данного исследования. Теоретически доказано, что благодаря байесовскому обучению оценки игроков приближаются к истинным значениям неизвестных параметров в модели «рыбной войны». Это согласуется с основным байесовским принципом: получая больше информации, игроки могут пересмотреть свои представления для более точного отражения реальных ситуаций.

Проведено численное моделирование для эмпирической поддержки теоретических результатов. Моделирование продемонстрировало эффективность байесовского обучения на практике: стратегии игроков меняются с течением времени по мере того, как они получают новые сигналы и соответствующим образом корректируют свои оценки. Подтверждена достоверность теоретических результатов, а также показано, как они могут быть применены в реальных ситуациях принятия стратегических решений.

Перспективным направлением будущих исследований является использование указанных идей в других экономических и экологических задачах, а также оценка надежности байесовских подходов в различных условиях. Например, экспериментальная проверка, включение в модели более сложной (нелинейной) динамики, а также изучение того, как общение между игроками влияет на сходимость представлений и стратегий.

СПИСОК ЛИТЕРАТУРЫ

1. Adler S., Subramanian V. Bayesian learning of optimal policies in Markov decision processes with countably infinite state-space// Advances in Neural Information Processing Systems. 2024.

2. Breton M., Keoula M. Y. A great fish war model with asymmetric players // Ecological economics. 2014. V. 97, 209–223.

3. Cave J. Long-term competition in a dynamic game: the cold fish war // The RAND Journal of Economics. 1987. 596–610.

4. Chakrabarty A., Benosman M. Safe learning-based observers for unknown nonlinear systems using Bayesian optimization// Automatica. 2021. V. 133, 109860.

5. Fischer R. D., Mirman L. J. Strategic dynamic interaction: fish wars // Journal of Economic Dynamics and Control, 1992. V. 16(2), 267–287.

6. Fischer R. D., Mirman L. J. The compleat fish wars: Biological and dynamic interactions // Journal of Environmental Economics and Management. 1996. V. 30(1), 34–42.

7. Grewal M. S., Andrews A. P. Kalman filtering: Theory and Practice with MATLAB // John Wiley & Sons, 2014.

8. Harsanyi J. Games with Incomplete Information Played by Bayesian Players, I–III. Part III. The Basic Probability Distribution of the Game // Management Science. 1968. V. 14(7), 486–502.

9. Huang L., Zhu Q. Analysis and computation of adaptive defense strategies against advanced persistent threats for cyber-physical systems // In Decision and Game Theory for Security: 9th International Conference, GameSec 2018, Seattle, WA, USA, October 29–31, 2018, Proceedings 9, pages 205–226.

10. Haurie A., Krawczyk J.B., Zaccour G. Games and dynamic games // World Scientific Publishing Company, Singapore, 2012.

11. Kumar P., Schuppen J. V. On Nash equilibrium solutions in stochastic dynamic games // IEEE Transactions on Automatic Control. 1980. V. 25(6), 1146–1149.

12. Levhari D., Mirman L. J. The great fish war: an example using a dynamic Cournot–Nash solution // In Fisheries Economics. 2020. V. II, 49–61.

13. Masoudi S., Santugini M., Zaccour G. A dynamic game of emissions pollution with uncertainty and learning // Environmental and Resource Economics. 2016. V. 64, 349–372.

14. Mirman L.J., Santugini M., Koulovatianos C. Optimal growth and uncertainty: Learning // Journal of Economic Theory. 2009. V. 144(1), 280–295.

15. Mirman L. J., Santugini M. Learning and technological progress in dynamic games // Dynamic Games and Applications. 2014. V. 4, 58–72.

16. Pindyck R. S. Irreversibilities and the timing of environmental policy // Resource and energy economics. 2000. V. 22(3), 233–259.

17. Pindyck R. S. Uncertainty in environmental economics // Review of environmental economics and policy, 2007.

18. Schofield N. Instability of simple dynamic games // The Review of Economic Studies. 1978. V. 45(3), 575–594.

19. Ulph, A., Maddison, D. Uncertainty, learning and international environmental policy coordination // Environmental and Resource Economics. 1997. V. 9, 451–466.

20. Xiao B., Yang W. A Bayesian learning model for estimating unknown demand parameter in revenue management// European Journal of Operational Research. 2021. V. 293(1), 248–262.

21. Yeung D. W., Petrosyan L. A. A cooperative stochastic differential game of transboundary industrial pollution // Automatica. 2008. V. 44(6), 1532–1544.

22. Zeeuw A. D., Zemel A. Regime shifts and uncertainty in pollution control // Journal of Economic Dynamics and Control. 2012. V. 36(7), 939–950.

BAYESIAN LEARNING IN FISH WARS: DYNAMIC ESTIMATION OF UNKNOWN STATES AND PRIVATE INFORMATION

Zhou Jiangjing, Saint Petersburg State University (st092028@student.spbu.ru),
Ovanes Petrosian, Saint Petersburg State University, Dr.Sc., professor (petrosian.ovanes@yandex.ru)

Abstract: This paper investigates a unique variant of the «fish war» game, where participants are required to estimate unknown environmental parameters and the private information of adversaries based on received signals. We develop a dynamic game model where the evolution of the fish population is influenced by an unknown parameter, ϵ, and each player's payoff function incorporates their private information, δ. Utilizing Bayesian learning methods, we demonstrate how participants can update their estimates of these unknown parameters over time. We prove that these estimates converge to true values as time progresses. The paper further presents a Nash Equilibrium with Bayesian learning, providing a solution to this specialized game. Numerical simulations are included to illustrate the convergence of beliefs among players and to compare their control strategies under various scenarios.

Keywords: dynamic Bayesian learning, «fish war» game, private information, unknown parameters.