Информационные процессы, Том 15, № 1, 2015, стр. 1–19

МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ВЫЧИСЛИТЕЛЬНЫЕ МЕТОДЫ

Алгоритм следования за возмущенным лидером и его применения для построения игровых стратегий[1]

В. В. Вьюгин

Институт проблем передачи информации им. А.А.Харкевича, Российская академия наук, Москва, Россия
e-mail: vyugin@iitp.ru

Поступила в редколлегию 02.02.2015

Аннотация—Рассматривается модификация алгоритма Ханнана следования за возмущенным лидером для случая неограниченных выигрышей и потерь. Получены оценки ошибки предсказания в терминах понятий объема v_t и флуктуации игры $\mathrm{fluc}(t) = \Delta v_t / v_t$, где $\Delta v_t = v_t - v_{t-1}$. Мы доказываем асимптотическую состоятельность в среднем данного варианта алгоритма при $\mathrm{fluc}(t) = o(t)$. Рассматриваются приложения алгоритма для построения игровых стратегий. Определяются игровые стратегии, которые используют различие между "микро" и "макро" волатильностью дискретного временного ряда (цен финансового инструмента) для получения арбитража. Смешивание этих экспертных стратегий производится на основе разработанного варианта алгоритма Ханнана.

КЛЮЧЕВЫЕ СЛОВА: предсказания с использованием экспертных стратегий, алгоритм следования за возмущенным лидером, адаптивный параметр обучения, состоятельность по Ханнану, алгоритмическая торговая стратегия.

1. ВВЕДЕНИЕ

В данной работе в качестве приложений основных результатов мы рассматриваем игровые стратегии, которые приводят к арбитражу, в том случае когда имеется различие между "микро" и "макро" волатильностью дискретного временного ряда (цен финансового инструмента). Каждая из стратегий использует определенное предположение о характере поведения временного ряда. В целом, сумма выигрышей и потерь этих стратегий равна нулю. Целью работы является построение алгоритма, который дает выигрыш не намного меньший чем выигрыш наилучшей экспертной стратегии минус некоторая ошибка – регрет. Для смешивания различных игровых стратегий мы используем наши новые результаты из теории предсказания с использованием экспертных стратегий.

Для построения игровых стратегий мы используем результаты работ [3] и [14] из финансовой математики. В этих работах анализируются две игровые стратегии, одна из которых приводит к выигрышу в случае когда цены финансового инструмента следуют фрактальному броуновскому движению. Первая стратегия приводит к выигрышу в случае "большой" волатильности временного ряда, вторая стратегия приводит к выигрышу в случае "малой" волатильности временного ряда. Мы построим алгоритм, который смешивает эти стратегии в одну игровую стратегию. Мы покажем, что этот алгоритм получает арбитраж в случае фрактального броуновского движения цены финансового инструмента.

Характерной особенностью данных игровых стратегий является то, что на одном шаге игры выигрыш или проигрыш может быть как угодно велик по абсолютной величине. Кроме этого,

[1] Исследование выполнено в ИППИ РАН за счет гранта Российского научного фонда (проект № 14-50-00150).

форма этого выигрыша (или потерь) не может быть описана с помощью конкретной функции выигрыша (функции потерь), такой как, абсолютная, логарифмическая, квадратичная и др.

Основным техническим результатом данной статьи является построение алгоритма, приспособленного к неограниченным пошаговым потерям и выигрышам. Как правило, известные алгоритмы машинного обучения используют предположение об ограниченности потерь на каждом шаге или используют специальный вид функции потерь.

Рассматривается следующая постановка задачи предсказания с использованием экспертных стратегий (Prediction with Expert Advice). Задан набор N экспертных стратегий (или просто экспертов), $i = 1, \dots N$. Эксперты принимают решения (или действия) последовательно на шагах $t = 1, 2, \dots, T$. На каждом таком шаге t каждый эксперт $i = 1, \dots N$ получает результат своего действия в виде своего выигрыша или потери (проигрыша) s_t^i, в зависимости знака этого числа. В дальнейшем мы будем называть величины s_t^i выигрышами независимо от их знака.[2]

В начале каждого шага t *Статистик* (*Learner*), наблюдает суммарные выигрыши $s_{1:t-1}^i = s_1^i + \dots + s_{t-1}^i$ каждого эксперта $i = 1, \dots N$ на предыдущих шагах и назначает неотрицательный вес w_t^i (сумма этих весов равна 1) каждому эксперту i. После этого *Статистик* получает выигрыш равный взвешенной сумме $\tilde{s}_t = \sum_{i=1}^{N} w_t^i s_t^i$ выигрышей всех экспертов на шаге t.

Суммарный выигрыш *Статистика* на первых T шагах равен величине $\tilde{s}_{1:T} = \sum_{t=1}^{T} \tilde{s}_t$.

Данную постановку можно также переформулировать в вероятностных терминах следующим образом. На каждом шаге t *Статистик* случайным образом выбирает эксперта i в соответствии со своим внутренним распределением вероятностей $P\{I_t = i\} = w_t^i$, $i = 1, \dots N$. В конце шага t *Статистик* получает тот же самый выигрыш s_t^i что и i-й эксперт. Суммарный выигрыш *Статистика* на шаге t равен случайной величине $s_{1:t} = s_{1:t-1} + s_t^i$.

Пусть $E(s_t)$ – математическое ожидание выигрыша *Статистика* на шаге t при рассмотренной выше рандомизации. Эта величина совпадает с взвешенной суммой $E(s_t) = \sum_{i=1}^{N} w_t^i s_t^i$ $(= \tilde{s}_t)$ выигрышей экспертов. Суммарный выигрыш за первые T шагов равен математическому ожиданию $E(s_{1:T}) = \sum_{t=1}^{T} E(s_t)$. Ясно, что $\tilde{s}_{1:T} = E(s_{1:T})$ для всех T.

Нашей целью является построения такого самообучающегося алгоритма, чтобы *Статистик* получил выигрыш $\tilde{s}_{1:T} = E(s_{1:T})$, который асимптотически был бы не намного меньше, чем выигрыш наилучшего эксперта.

В классической постановке теории предсказаний с использованием экспертных стратегий обычно рассматриваются неотрицательные потери вместо выигрышей и предполагается, что потери экспертов на каждом шаге ограничены, например, $0 \le s_t^i \le 1$ для всех i и t. В нашей постановке потери – это выигрыши с противоположным знаком, при этом они могут быть неограниченными. Мы также не будем использовать какой либо специальный вид этих выигрышей (потерь).

В данной работе мы развиваем метод "следования за возмущенным лидером" для случая неограниченных выигрышей и потерь. Данный метод предложен Ханнаном [6] в 1957 г. и долгое время был мало известен среди специалистов в области построения самообучающихся алгоритмов. Калаи и Вемпала заново переоткрыли этот метод в 2003 г. и опубликовали его формулировку и простое доказательство в статье [8]. Они назвали этот алгоритм алгоритмом

[2] Природа выигрыша не имеет значения – это некоторый ход игрока в приведенной далее в разделе 3 игре. В частности, не предполагается, что величина выигрыша является случайной величиной.

следования за возмущенным лидером – FPL (Following the Perturbed Leader). Хуттер и Поланд [7] развили этот алгоритм для счетного класса экспертов, ввели адаптивно меняющийся параметр обучения и т.д. Однако, во всех этих работах рассматривался случай, когда потери экспертов на одном шаге ограничены: $0 \le s_t^i \le 1$ для всех i и t.

Заметим, что предположения об ограниченности пошаговых потерь экспертов также используются в других смешивающих алгоритмах, например, в алгоритмах типа взвешенного большинства – Weighted Majority (WM) Литтлстоуна и Вармута [9], в алгоритме хеджирования "hedge" Фройнда и Шапире [5]. Алгоритм FPL имеет аналогичную ошибку предсказания как и алгоритмы типа WM с точностью до множителя $\sqrt{2}$. Преимуществом алгоритма FPL является простота, с которой используется параметр обучения, что позволяет легко ввести адаптивные варианты этого параметра (см. замечание в [7]).

В большинстве статей в области предсказания с использованием экспертных стратегий либо рассматриваются ограниченные пошаговые потери либо предполагается, что пошаговые потери выражаются в виде функций потерь специального вида. Постановки, в которых одновременно не используется специальный вид пошаговых потерь и эти потери не ограничены, мало отражены в специальной литературе; мы можем только привести работы [2, 11], в которых рассматриваются полиномиальные границы на пошаговые потери экспертов.

В данной работе предлагается модификация алгоритма Калаи и Вемпала [8] следования за возмущенным лидером для случая пошаговых выигрышей $s_t^i \in (-\infty, +\infty)$ не ограниченных априори. Этот алгоритм используется адаптивно изменяющиеся веса, зависящие от выигрышей экспертов в прошлом.

Мы вводим новое понятие *объема игры*

$$v_t = \sum_{j=1}^{t} \max_i |s_j^i| + v_0,$$

где v_0 – константа, а также понятие *флуктуации* игры

$$\mathrm{fluc}(t) = \Delta v_t / v_t,$$

где $\Delta v_t = v_t - v_{t-1}$ для $t \ge 1$.

В разделе 2 в качестве приложения основного результата рассматривается игра двух экспертов с нулевой суммой, которые выигрывают (или проигрывают) в случае, когда микро и макро волатильности временного ряда цен финансового инструмента (акции) различаются. Данные выигрыши и проигрыши не могут быть ограничены априори. Эта игра служит мотивирующим примером для разработки варианта алгоритма следования за возмущенным лидером в случае неограниченных выигрышей и потерь.

В разделе 3 рассмотривается случай N экспертов, выигрыши которых неограничены $s_t^i \in (-\infty, +\infty)$. Будет предложен вероятностный алгоритм, основанный на выборе экспертов. Будет доказано, что этот алгоритм оптимален в смысле приведенной ниже оценки (1) при достаточно широких предположениях.

В разделе 4 обсуждается дерандомизированная версия этого алгоритма для случая двух экспертов из раздела 2.

Теорема 2 (раздел 3) утверждает, что если $\mathrm{fluc}(t) \le \gamma(t)$ для всех t, где $\gamma(t)$ – неубывающая вычислимая функция такая, что $0 < \gamma(t) \le 1$ для всех t, то алгоритм следования за возмущенным лидером с адаптивными весами, построенный в разделе 3, имеет нижнюю оценку математического ожидания суммарного выигрыша

$$E(s_{1:T}) \ge \max_{i=1,\dots N} s_{1:T}^i - 2\sqrt{8(1+\ln N)} \sum_{t=1}^{T} (\gamma(t))^{1/2} \Delta v_t. \qquad (1)$$

Если $\gamma(t) \to 0$ и $v_t \to \infty$ при $t \to \infty$, то этот алгоритм асимптотически состоятелен в среднем в следующем модифицированном смысле

$$\liminf_{T\to\infty} \frac{1}{v_T} E(s_{1:T} - \max_{i=1,\dots N} s^i_{1:T}) \geq 0, \tag{2}$$

где $s_{1:T}$ – суммарный выигрыш этого алгоритма на шагах $\leq T$.

Теорема 1 из раздела 3 показывает, что если условие $\Delta v_t = o(v_t)$ не выполнено, то суммарный выигрыш произвольного вероятностного алгоритма может быть значительно меньше, чем выигрыш одного из экспертов.

2. АРБИТРАЖНЫЕ СТРАТЕГИИ

Рассмотрим некоторый дискретный временной ряд. Например, это могут быть значения цен некоторого финансового инструмента (акции) на фондовым рынке. Тогда эксперты могут покупать и продавать акции при этих значениях цен в соответствии со своими стратегиями.

Пусть K, M и T – положительные целые числа и пусть временной интервал $[0, KT]$ разделен на большое число KM подинтервалов. Пусть также $S(t)$ – функция, представляющая цену акции в момент времени равный t. Рассмотрим дискретный временной ряд

$$S_0 = S(0), \quad S_1 = S\left(\frac{T}{KM}\right), \quad S_2 = S\left(\frac{2T}{KM}\right), \quad \dots, \quad S_{KM} = S(T). \tag{3}$$

В этой статье понятие волатильности является неформальным понятием. Мы будем говорить, что сумма

$$\sum_{i=0}^{K-1} (S_{(i+1)T} - S_{iT})^2$$

является мерой “макро” волатильности временного ряда (3), а сумма

$$\sum_{t=0}^{KT-1} (\Delta S_t)^2,$$

где $\Delta S_t = S_{t+1} - S_t$, $t = 1, \dots KT$, является мерой “микро” волатильности этого же временного ряда. В этой работе для простоты рассматривается случай $K = 1$. Однако приложение, рассмотренное в разделе 4, легко может быть обобщено на случай $K > 1$.

Неформально, мы рассмотрим две стратегии игры, при которых первая стратегия выигрывает, когда $(S_T - S_0)^2 \gg \sum_{t=0}^{T-1} (\Delta S_t)^2$, а вторая стратегия выигрывает, при $(S_T - S_0)^2 \ll \sum_{t=0}^{T-1} (\Delta S_t)^2$. Имеется также область неопределенности в случае когда оба соотношения $\gg$ и $\ll$ не выполнены.

Идея таких стратегий основана на работе [3] (см. также [12], [4]), в которой для случая непрерывного времени были определены стратегии игры на финансовом рынке, состоящего из денежного счета и акции цена которой следует фрактальному броуновскому движению. В работе [14] эти стратегии были сформулированы для дискретного времени. Мы определим смешанную стратегию на основе двух этих стратегий. Заметим, что в случае непрерывного времени не существует области неопределенности.

Рассмотрим игру инвестора на финансовом рынке. Инвестор может использовать длинную и короткие позиции при этой игре.[3]

[3] Игра в длинной позиции предполагает, что в начале периода инвестор покупает некоторое количество акций, а в конце периода продает их. Разность в цене акции определяет его выигрыш или потери. Короткая позиция

В начале шага t инвестор имеет в своей собственности (например, в результате покупки на этом шаге или на предыдущих шагах) C_t акций по цене S_{t-1} за каждую. Если $C_t < 0$, то это означает, что инвестор имеет долг, который он должен вернуть в конце торгового периода в виде данного числа акций $|C_t|$. В конце торгового периода (это может быть минутный, часовой или интервал другой длительности) становится известна цена S_{t+1} акции и инвестор подсчитывает свой выигрыш или потери в размере $s_t = C_t \Delta S_t$ за период от $[t, t+1)$.

Имеет место следующее тождество

$$(S_T - S_0)^2 = \left(\sum_{t=0}^{T-1} \Delta S_t \right)^2 = \sum_{t=0}^{T-1} 2(S_t - S_0)\Delta S_t + \sum_{t=0}^{T-1} (\Delta S_t)^2. \tag{4}$$

Тождество (4) подсказывает нам две игровые стратегии для двух инвесторов. В начале шага t Инвестор 1 и Инвестор 2 владеют числом акций

$$C_t^1 = 2C(S_t - S_0), \tag{5}$$
$$C_t^2 = -C_t^1, \tag{6}$$

соответственно, где C – некоторая положительная константа.

Эти стратегии имеют выигрыш (потери) за шаг t, равный $s_t^1 = 2C(S_t - S_0)\Delta S_t$ и $s_t^2 = -s_t^1$, соответственно. Стратегия (5) имеет за первые T шагов выигрыш

$$s_{1:T}^1 = \sum_{t=1}^{T} s_t^1 = 2C \left((S_T - S_0)^2 - \sum_{t=1}^{T-1} (\Delta S_t)^2 \right).$$

Стратегия (6) на первых T шагах имеет выигрыш $s_{1:T}^2 = -s_{1:T}^1$.

Число акций C_t^1 при стратегии (5) или число акций $C_t^2 = -C_t^1$ при стратегии (6) может быть как положительным так и отрицательным. Выигрыши s_t^1 и $s_t^2 = -s_t^1$ априори не ограничены и также могут быть как положительными так и отрицательными: $s_t^i \in (-\infty, +\infty)$.

Введем *Статистика*, который может выбирать или комбинировать эти две стратегии (5) и (6). Разумеется за подобные переходы от одной стратегии к другой *Статистику* придется платить. Величина этой платы называется ошибкой предсказания или регретом. Основным предметом изучения эффективности смешивающих алгоритмов будет минимизация такого регрета.

Выбор оптимальной стратегии является непростой задачей даже в случае ограниченных выигрышей как показывает следующий пример. *Статистик* может получить значительный проигрыш, если он будет просто выбирать наиболее удачливого в прошлом эксперта. Пусть текущие выигрыши двух экспертов на шагах $t = 0, 1, \ldots, 6$ будут следующими: $s_{0,1,2,3,4,5,6}^1 = (1/2, -1, 1, -1, 1, -1, 1)$ и $s_{0,1,2,3,4,5,6}^2 = (0, 1, -1, 1, -1, 1, -1)$.

Тогда наивный алгоритм "следования за лидером" будет всегда делать неверные предсказания. его выигрыш (а точнее потери) составит $s_{1:6} = -5.5$.

Подобная трудность преодолевается в разделе 3 с помощью рандомизации суммарных выигрышей экспертов.

означает, что инвестор продает в начале периода некоторое количество акции, а в конке периода покупает их. В этом случае, соответствующая разность в цене акции также определяет его выигрыш или потери.

3. АЛГОРИТМ СЛЕДОВАНИЯ ЗА ВОЗМУЩЕННЫМ ЛИДЕРОМ С АДАПТИВНЫМИ ВЕСАМИ

Пусть на каждом шаге t игры выигрыш i-го эксперта составляет $s_t^i \in (-\infty, +\infty)$, а его его суммарный выигрыш равен

$$s_{1:t}^i = s_{1:t-1}^i + s_t^i,$$

где $i = 1, \ldots, N$.

Вероятностный алгоритм на шаге t получает на входе суммарные выигрыши $s_{1:t-1}^i$ экспертов $i = 1, \ldots N$ и выдает для каждого i вероятности $P\{I_t = i\}$ следования за каждым из i экспертов. Рассмотрим наиболее общий протокол работы самообучающегося вероятностного алгоритма, выбирающего экспертов.

Вероятностный алгоритм выбора эксперта. На шаге $t = 1, \ldots, T$:

Статистик наблюдает суммарные выигрыши экспертов $i = 1, \ldots, N$ за предыдущие шаги: $s_{1:t-1}^i$ и вычисляет вероятности $P\{I_t = i\}$ выбора каждого эксперта i.

Статистик выбирает эксперта i с вероятностью $P\{I_t = i\}$.

Эксперты $i = 1, \ldots, N$ получают свои выигрыши s_t^i на шаге t и вычисляют свои суммарные выигрыши $s_{1:t}^i = s_{1:t-1}^i + s_t^i$.

Статистик получает выигрыш, s_t равный выигрышу выбранного эксперта i, а именно, $s_t = s_t^i$.

Статистик вычисляет свой суммарный выигрыш за первые t шагов: $s_{1:t} = s_{1:t-1} + s_t = s_{1:t-1} + s_t^i$ (полагаем $s_{1,0} = 0$).

На этом описание шага работы алгоритма заканчивается.

Количественной оценкой эффективности работы этого вероятностного алгоритма является математическое ожидание его ошибки предсказания (expected regret)

$$E(\max_{i=1,\ldots N} s_{1:T}^i - s_{1:T}),$$

где случайная величина $s_{1:T}$ есть суммарный выигрыш нашего алгоритма, $s_{1:T}^i$ есть суммарные выигрыши экспертов $i = 1, \ldots, N$ на шагах $t = 1, \ldots, T$, буква E обозначает математическое ожидание по вероятностному распределению, которое порождается вероятностями $P\{I_t = i\}$.

Изучим асимптотическую состоятельность произвольного самообучающегося вероятностного алгоритма в случае когда его пошаговые выигрыши неограничены. Вероятностный алгоритм называется *асимптотически состоятельным в среднем*, если

$$\liminf_{T\to\infty} \frac{1}{T} E(s_{1:T} - \max_{i=1,\ldots N} s_{1:T}^i) \geq 0. \tag{7}$$

Заметим, что при $0 \leq s_t^i \leq 1$ суммарный выигрыш любого экспертного алгоритма за T шагов не превосходит T. В случае неограниченных пошаговых выигрышей это свойство уже не всегда будет выполнено. Поэтому в этом случае не имеет смысла делить среднюю ошибку предсказания (7) на время T.

Модифицируем определение (7) нормированной средней ошибки предсказания следующим образом. Определим *объем* игры на шаге t

$$v_t = \sum_{j=1}^{t} \max_i |s_j^i| + v_0,$$

где v_0 – неотрицательная константа. По определению, $0 \leq v_{t-1} \leq v_t$ для всех $t \geq 1$.

Вероятностный самообучающийся алгоритм называется *асимптотически состоятельным в среднем* (в модифицированном смысле) в игре с N экспертами, если

$$\liminf_{T\to\infty} \frac{1}{v_T} E(s_{1:T} - \max_{i=1,\dots N} s^i_{1:T}) \geq 0. \tag{8}$$

Обозначим $\Delta v_t = v_t - v_{t-1}$ при $t \geq 1$. Функция

$$\mathrm{fluc}(t) = \frac{\Delta v_t}{v_t} = \frac{\max_i |s^i_t|}{v_t}, \tag{9}$$

называется *флуктуацией* игры на шаге t (полагаем $0/0 = 0$).

По определению $0 \leq \mathrm{fluc}(t) \leq 1$ для всех $t \geq 1$.

Игра называется *невырожденной* если $v_t \to \infty$ при $t \to \infty$.

Следующее простое утверждение показывает, что произвольный вероятностный самообучающийся алгоритм не может быть асимптотически оптимальным в некоторой игре, для которой $\mathrm{fluc}(t) \not\to 0$ при $t \to \infty$. Для простоты мы рассмотрим случай двух экспертов.

Теорема 1. *Пусть $\epsilon > 0$. Для любого самообучающегося вероятностного алгоритма существуют два эксперта, для которых*

$$\mathrm{fluc}(t) \geq 1 - \epsilon,$$
$$\frac{1}{v_t} E(\max_{i=1,2} s^i_{1:t} - s_{1:t}) \geq \frac{1}{2}(1-\epsilon)$$

для всех t, где $s_{1:t}$ – суммарный выигрыш этого алгоритма. Соответствующая игра является невырожденной.

Доказательство. Определим по $0 < \epsilon < 1$ и по заданному вероятностному алгоритму пошаговые выигрыши: s^1_t – Эксперта 1 и s^2_t – Эксперта 2, с помощью рекурсии по шагам $t = 1, 2, \dots$. Пусть $s^1_{1:t-1}$ и $s^2_{1:t-1}$ – суммарные выигрыши этих экспертов на шагах $\leq t-1$ и пусть v_{t-1} – соответствующий объем игры.

Определим $v_0 = 1$ и $M_t = 4v_{t-1}/\epsilon$ для $t \geq 1$.

При $t \geq 1$ определим $s^1_t = 0$ и $s^2_t = M_t$, если $P\{I_t = 1\} > \frac{1}{2}$, и определим $s^1_t = M_t$ и $s^2_t = 0$ в противном случае. Пусть при $t \geq 1$, v_t есть объем игры.

Пусть s_t – выигрыш на шаге t нашего вероятностного алгоритма и пусть $s_{1:t}$ – его суммарный выигрыш на этом шаге, прием $t \geq 1$. По определению при $t \geq 1$,

$$E(s_t) = s^1_t P\{I_t = 1\} + s^2_t P\{I_t = 2\} \leq \frac{1}{2} M_t.$$

Очевидно, что $E(s_{1,1}) = E(s_1)$ и $E(s_{1:t}) = E(s_{1:t-1}) + E(s_t)$ для всех $t \geq 2$. Тогда $E(s_{1:t}) \leq \frac{1}{2}(1+\epsilon/2)M_t$ для всех $t \geq 1$. Также очевидно, что $v_t = v_{t-1} + M_t = M_t(1+\epsilon/4)$ и $M_t \leq \max\limits_i s^i_{1:t}$ для всех $t \geq 1$.

Таким образом, нормированная средняя ошибка (регрет) вероятностного алгоритма ограничена снизу

$$\frac{1}{v_t} E(\max_i s^i_{1:t} - s_{1:t}) \geq \frac{\frac{1}{2}(1-\epsilon/2)M_t}{M_t(1+\epsilon/4)} \geq \frac{1}{2}(1-\epsilon).$$

для всех t. Для флуктуации игры имеет место оценка:

$$\mathrm{fluc}(t) = \frac{\Delta v_t}{v_t} = \frac{M_t}{M_t(1+\epsilon/4)} > 1-\epsilon$$

для всех t. Теорема доказана. $\triangle$

Пусть $\xi^1,\ldots\xi^N$ – последовательность случайных величин независимых и одинаково распределенных согласно экспоненциальному закону с плотностью

$$p(x) = \begin{cases} \exp\{-x\}, \text{ при } x \geq 0, \\ 0, \text{ при } x < 0. \end{cases}$$

Пусть $\gamma(t)$ – невозрастающая функция, принимающая вещественные значения такая, что $0 < \gamma(t) < 1$ для всех t; например, $\gamma(t) = (t+c)^{-\delta}$, где $c > 0$, $\delta > 0$ и $t \geq 1$. Определим

$$\alpha_t = \frac{1}{2}\left(1 - \frac{\ln\dfrac{1+\ln N}{8}}{\ln\gamma(t)}\right) \text{ и} \tag{10}$$

$$\mu_t = (\gamma(t))^{\alpha_t} = \sqrt{\frac{8}{1+\ln N}}(\gamma(t))^{1/2}. \tag{11}$$

для всех t.[4] Предположим, что любая верхняя граница $\gamma(t)$ для флуктуаций удовлетворяет неравенству $\gamma(t) \leq \min\{A, A^{-1}\}$ для всех t, где $A = 8/(1+\ln N)$. За счет выбора v_0 можно добиться того, чтобы это условие было выполнено для флуктуаций, а затем можно соответствующим образом перестроить функцию $\gamma(t)$.

Мы рассмотрим алгоритм FPL (Follow the Perturbed Leader) следования за возмущенным лидером с адаптивным параметром обучения

$$\epsilon_t = \frac{1}{\mu_t v_{t-1}}, \tag{12}$$

где μ_t определено по (11), а объем v_{t-1} зависит от действий экспертов на шагах $< t$. По определению $v_t \geq v_{t-1}$ и $\mu_t \leq \mu_{t-1}$ для $t = 1, 2, \ldots$. Также, если $\gamma(t) \to 0$ при $t \to \infty$, то $\mu_t \to 0$ при $t \to \infty$.

Мы допускаем без потери общности, что $s_0^i = 0$ для всех i и $\epsilon_0 = \infty$.

Алгоритм FPL **следования за возмущенным лидером**.

На шаге $t = 1, \ldots T$ определим

$$I_t = \mathrm{argmax}_{i=1,2,\ldots N}\left\{s^i_{1:t-1} + \frac{1}{\epsilon_t}\xi^i\right\}, \tag{13}$$

где ϵ_t определено по (12), и получим выигрыш $s_t^{I_t}$ эксперта I_t.[5]

Считаем это значение выигрышем алгоритма FPL: $s_t = s_t^{I_t}$.

На этом описание алгоритма закончено.

[4] Выбор оптимального значения параметра α_t будет разъяснен далее. Это значение будет получено путем минимизации соответствующего члена суммы (38) ниже. Определение (10) теряет смысл при $\gamma(t) = 1$. Выражение для μ_t не теряет смысла и при $\gamma(t) = 1$, поэтому можно использовать значения $(\gamma(t))^{\alpha_t}$ и $(\gamma(t))^{1-\alpha_t}$ определенные по (11).

[5] Если максимум достигается для более чем одного значения i, выберем наименьшее из них. Заметим также, что номер I_t наилучшего эксперта является случайной величиной.

Пусть $s_{1:T} = \sum_{t=1}^{T} s_t^{I_t}$ – суммарный выигрыш FPL алгоритма за первые T шагов.

В следующей теореме получена верхняя оценка для ошибки FPL алгоритма. Она показывает. что если игра невырожденная, $\Delta v_t = o(v_t)$ при $t \to \infty$ и существует алгоритмическая верхняя оценка для этой сходимости, то алгоритм FPL с адаптивным параметром μ_t является асимптотически состоятельным в среднем.

Теорема 2. *Пусть $\gamma(t)$ – невозрастающая вычислимая функция такая, что $0 < \gamma(t) \le 1$ и выигрыши экспертов удовлетворяют условию*

$$\mathrm{fluc}(t) \le \gamma(t) \tag{14}$$

для всех t. Тогда средний выигрыш алгоритма FPL *с адаптивным параметром обучения (12) удовлетворяет*

$$E(s_{1:T}) \ge \max_i s_{1:T}^i - 2\sqrt{8(1+\ln N)}\sum_{t=1}^{T}(\gamma(t))^{1/2}\Delta v_t. \tag{15}$$

для всех T.

Если игра невырожденная и $\gamma(t) \to 0$ при $t \to \infty$, то этот алгоритм асимптотически состоятелен в среднем

$$\liminf_{T\to\infty} \frac{1}{v_T} E(s_{1:T} - \max_{i=1,\dots N} s_{1:T}^i) \ge 0. \tag{16}$$

Замечание. Допустим, что выигрыши экспертов ограничены, например, $s_t^i \in [-1, 1]$ для всех i и t. Тогда $v_t \le t$ и в (16) можно заменить v_T на более привычную величину T.

Доказательство теоремы. Доказательство следует схеме подобных доказательств из статей [8] и [7]; оригинальным является выбор адаптивного параметра обучения и соответствующие ему более сложные оценки. Анализ оптимальности алгоритма FPL использует вспомогательный алгоритм IFPL (Infeasible FPL).

Алгоритм IFPL.

На шаге $t = 1, \dots, T$ алгоритма определим

$$\epsilon_t' = \frac{1}{\mu_t v_t}, \tag{17}$$

где v_t – объем игры на шаге t и μ_t определена по (11). Определим также

$$J_t = \mathrm{argmax}_{i=1,\dots,N}\left\{ s_{1:t}^i + \frac{1}{\epsilon_t'}\xi^i \right\}.$$

Получим выигрыш $s_t^{J_t}$ эксперта J_t и считаем его выигрышем IFPL алгоритма. На этом описание алгоритма закончено.

Заметим, что IFPL алгоритм использует для своего предсказания величины ϵ_t' и $s_{1:t}^i$, $i = 1, \dots, N$, которые неизвестны в начале шага t. Поэтому IFPL алгоритм не реализуем (infeasible); он служит вспомогательным средством необходимым для доказательства нужных свойств FPL алгоритма.

Имеем для произвольного t

$$I_t = \mathrm{argmax}_i\{s_{1:t-1}^i + \frac{1}{\epsilon_t}\xi^i\},$$

$$J_t = \mathrm{argmax}_i\{s_{1:t}^i + \frac{1}{\epsilon_t'}\xi^i\}.$$

Математические ожидания пошагового и суммарного выигрышей алгоритмов FPL и IFPL на шаге t обозначаем

$$l_t = E(s_t^{I_t}) \text{ и } r_t = E(s_t^{J_t}),$$
$$l_{1:T} = \sum_{t=1}^{T} l_t \text{ и } r_{1:T} = \sum_{t=1}^{T} r_t,$$

соответственно, где $s_t^{I_t}$ – выигрыш алгоритма FPL на шаге t и $s_t^{J_t}$ – выигрыш алгоритма IFPL на шаге t, буква E обозначает математическое ожидание.

Лемма 1. *Математические ожидания суммарных выигрышей* FPL *и* IFPL *алгоритмов удовлетворяют неравенству*

$$l_{1:T} \geq r_{1:T} - 8\sum_{t=1}^{T}(\gamma(t))^{1-\alpha_t}\Delta v_t \tag{18}$$

для всех T.

Доказательство. Для произвольного $j = 1, \ldots, n$ и фиксированных вещественных чисел $c_1, \ldots, c_N$ определим

$$m_j = \max_{i \neq j}\{s_{1:t-1}^i + \frac{1}{\epsilon_t}c_i\},$$
$$m'_j = \max_{i \neq j}\{s_{1:t}^i + \frac{1}{\epsilon'_t}c_i\} = \max_{i \neq j}\{s_{1:t-1}^i + s_t^i + \frac{1}{\epsilon'_t}c_i\}.$$

Пусть $m_j = s_{1:t-1}^{j_1} + \frac{1}{\epsilon_t}c_{j_1}$ и $m'_j = s_{1:t-1}^{j_2} + s_t^{j_2} + \frac{1}{\epsilon'_t}c_{j_2}$. По определению величины m'_j и, так как $j_1 \neq j$, имеем

$$\begin{aligned} m'_j &= s_{1:t-1}^{j_2} + s_t^{j_2} + \frac{1}{\epsilon'_t}c_{j_2} \geq \\ &\geq s_{1:t-1}^{j_1} - \Delta v_t + \frac{1}{\epsilon'_t}c_{j_1} = \\ &= s_{1:t-1}^{j_1} - \Delta v_t + \frac{1}{\epsilon_t}c_{j_1} + \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t}\right)c_{j_1} = \\ &= m_j + \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t}\right)c_{j_1} - \Delta v_t. \end{aligned} \tag{19}$$

Здесь мы использовали неравенство $s_t^{j_2} \geq -\Delta v_t$ для всех t.

Сравним условные вероятности

$$P\{I_t = j | \xi^i = c_i, i \neq j\} \text{ и } P\{J_t = j | \xi^i = c_i, i \neq j\}.$$

Имеет место следующая цепь равенств и неравенств:

$$\begin{aligned}
&P\{I_t = j|\xi^i = c_i, i \neq j\} = \\
&= P\{s^j_{1:t-1} + \frac{1}{\epsilon_t}\xi^j \geq m_j|\xi^i = c_i, i \neq j\} = \\
&= P\{\xi^j \geq \epsilon_t(m_j - s^j_{1:t-1})|\xi^i = c_i, i \neq j\} = \\
&= P\{\xi^j \geq \epsilon'_t(m_j - s^j_{1:t-1}) + \\
&+(\epsilon_t - \epsilon'_t)(m_j - s^j_{1:t-1})|\xi^i = c_i, i \neq j\} = \\
&= P\{\xi^j \geq \epsilon'_t(m_j - s^j_{1:t-1}) + \\
&+(\epsilon_t - \epsilon'_t)(s^{j_1}_{1:t-1} - s^j_{1:t-1} + \frac{1}{\epsilon_t}c_{j_1})|\xi^i = c_i, i \neq j\} \geq \qquad (20)
\end{aligned}$$

$$\begin{aligned}
&\geq P\{\xi^j \geq \epsilon'_t(m'_j - s^j_{1:t} + s^j_t + \Delta v_t - \left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon_t}\right)c_{j_1}) + \\
&+(\epsilon_t - \epsilon'_t)(s^{j_1}_{1:t-1} - s^j_{1:t-1}) + (\epsilon_t - \epsilon'_t)\frac{1}{\epsilon_t}c_{j_1}|\xi^i = c_i, i \neq j\} = \qquad (21) \\
&= P\{\xi^j \geq \epsilon'_t(m'_j - s^j_{1:t}) + \\
&+(\epsilon_t - \epsilon'_t)(s^{j_1}_{1:t-1} - s^j_{1:t-1}) + \epsilon'_t(s^j_t + \Delta v_t)|\xi^i = c_i, i \neq j\} = \\
&= P\{\xi^j > \frac{1}{\mu_t v_t}(m'_j - s^j_{1:t}) + \\
&+\left(\frac{1}{\mu_t v_{t-1}} - \frac{1}{\mu_t v_t}\right)(s^{j_1}_{1:t-1} - s^j_{1:t-1}) + \frac{s^j_t + \Delta v_t}{\mu_t v_t}|\xi^i = c_i, i \neq j\} \geq \\
&\geq P\{\xi^j > \frac{1}{\mu_t v_t}(m'_j - s^j_{1:t}) + \\
&+\frac{\Delta v_t}{\mu_t v_t}\frac{(s^j_{1:t-1} - s^{j_1}_{1:t-1})}{v_{t-1}} + \frac{2\Delta v_t}{\mu_t v_t}|\xi^i = c_i, i \neq j\} \geq \qquad (22) \\
&\geq \exp\left\{-\frac{\Delta v_t}{\mu_t v_t}\left|\left(2 + \frac{s^{j_1}_{1:t-1} - s^j_{1:t-1}}{v_{t-1}}\right)\right|\right\} P\{J_t = 1|\xi^i = c_i, i \neq j\}. \qquad (23)
\end{aligned}$$

Здесь при переходе (22)–(23) мы использовали неравенство

$$P\{\xi > a + b\} \geq e^{-|b|}P\{\xi > a\},$$

которое имеет место для любой случайной переменной ξ, распределенной в соответствии с экспоненциальным распределением, и для любых a и b.

Неравенство (20)–(21) следует из (19) и $\epsilon_t \geq \epsilon'_t$ для всех t.

Соответствующее отношение в экспоненте (23) ограничено:

$$\left|\frac{s^{j_1}_{1:t-1} - s^j_{1:t-1}}{v_{t-1}}\right| \leq 2, \qquad (24)$$

так как $\left|\frac{s^i_{1:t-1}}{v_{t-1}}\right| \leq 1$ для всех t и i.

Таким образом, получаем

$$P\{I_t = j|\xi^i = c_i, i \neq j\} \geq$$
$$\geq \exp\left\{-\frac{4}{\mu_t}\frac{\Delta v_t}{v_t}\right\} P\{J_t = j|\xi^i = c_i, i \neq j\} \geq \quad (25)$$
$$\geq \exp\{-4(\gamma(t))^{1-\alpha_t}\} P\{J_t = j|\xi^i = c_i, i \neq j\}. \quad (26)$$

Так как неравенство (26) между условными вероятностями имеет место для всех c_i, оно имеет место также и безусловно

$$P\{I_t = j\} \geq \exp\{-4(\gamma(t))^{1-\alpha_t}\} P\{J_t = j\}. \quad (27)$$

для всех $t = 1, 2, \ldots$ и $j = 1, \ldots, N$.

Так как $s_t^j + \Delta v_t \geq 0$ для всех j и t, мы получаем из (27)

$$\begin{aligned} l_t + \Delta v_t = E(s_t^{I_t} + \Delta v_t) &= \sum_{j=1}^{N} (s_t^j + \Delta v_t) P(I_t = j) \geq \\ &\geq \exp\{-4(\gamma(t))^{1-\alpha_t}\} \sum_{j=1}^{N} (s_t^j + \Delta v_t) P(J_t = j) = \\ &= \exp\{-4(\gamma(t))^{1-\alpha_t}\} (E(s_t^{J_t}) + \Delta v_t) = \\ &= \exp\{-4(\gamma(t))^{1-\alpha_t}\} (r_t + \Delta v_t) \geq \\ &\geq (1 - 4(\gamma(t))^{1-\alpha_t})(r_t + \Delta v_t) = \\ &= r_t + \Delta v_t - 4(\gamma(t))^{1-\alpha_t}(r_t + \Delta v_t) \geq \\ &\geq r_t + \Delta v_t - 8(\gamma(t))^{1-\alpha_t} \Delta v_t. \end{aligned} \quad (28)$$

Для вывода последней строки (28) мы использовали неравенство $|r_t| \leq \Delta v_t$ для всех t, а также неравенство $\exp\{-4r\} \geq 1 - 4r$ для всех r.

Вычитаем Δv_t из обоих частей неравенств (28) и суммируем полученные неравенства по $t = 1, \ldots, T$, получим

$$l_{1:T} \geq r_{1:T} - 8 \sum_{t=1}^{T} (\gamma(t))^{1-\alpha_t} \Delta v_t$$

для всех T. Лемма 1 доказана. $\triangle$

В следующей лемме дана нижняя оценка среднего выигрыша "нереализуемого" алгоритма IFPL.

Лемма 2. *Математическое ожидание выигрыша алгоритма* IFPL *с адаптивным параметром обучения* (17) *ограничено снизу*

$$r_{1:T} \geq \max_i s_{1:T}^i - (1 + \ln N) \sum_{t=1}^{T} (\gamma(t))^{\alpha_t} \Delta v_t \quad (29)$$

для всех T.

Доказательство. Схема доказательства аналогична схеме из [7], с тем лишь исключением, что у нас последовательность ϵ_t' не является монотонной.

Обозначим временно внутри этого доказательства $s_t = (s_t^1, \dots, s_t^N)$ – вектор пошаговых выигрышей и $s_{1:t} = (s_{1:t}^1, \dots, s_{1:t}^N)$ – вектор суммарных выигрышей на шаге t экспертных алгоритмов. Также, пусть $\xi = (\xi^1, \dots, \xi^N)'$ – вектор независимых случайных величин, одинаково распределенных в соответствии с экспоненциальным распределением вероятностей.

Напомним, что $\epsilon'_t = 1/(\mu_t v_t)$ и $v_0 = 0$, $\epsilon_0 = \infty$.

Определим $\tilde{s}_{1:t} = s_{1:t} + \dfrac{1}{\epsilon'_t}\xi$ при $t = 1, 2, \dots$. Временно введем вспомогательные выигрыши $\tilde{s}_t = s_t + \xi\left(\dfrac{1}{\epsilon'_t} - \dfrac{1}{\epsilon'_{t-1}}\right)$. Для любого вектора s и единичного вектора d определим

$$M(s) = \operatorname{argmax}_{d \in D}\{d \cdot s\},$$

где $D = \{(0, \dots, 1), \dots (1, \dots, 0)\}$ – множество всех N единичных векторов размерности N и "$\cdot$" – скалярное произведение двух векторов. Покажем, что

$$\sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot \tilde{s}_t \geq M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T}. \tag{30}$$

При $T = 1$ это утверждение очевидно. Для индуктивного перехода от $T-1$ к T надо доказать, что

$$M(\tilde{s}_{1:T}) \cdot \tilde{s}_T \geq M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T} - M(\tilde{s}_{1:T-1}) \cdot \tilde{s}_{1:T-1}.$$

Это нераенство следует из равенства $\tilde{s}_{1:T} = \tilde{s}_{1:T-1} + \tilde{s}_T$ и неравенства

$$M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T-1} \leq M(\tilde{s}_{1:T-1}) \cdot \tilde{s}_{1:T-1}.$$

Перепишем (30) следующим образом:

$$\sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot s_t \geq M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T} - \sum_{t=1}^{T} M(\tilde{s}_{1:t}) \cdot \xi\left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}}\right). \tag{31}$$

По определению M имеем

$$\begin{aligned} M(\tilde{s}_{1:T}) \cdot \tilde{s}_{1:T} &\geq M(s_{1:T}) \cdot \left(s_{1:T} + \frac{\xi}{\epsilon'_T}\right) = \\ &= \max_{d \in D}\{d \cdot s_{1:T}\} + M(s_{1:T}) \cdot \frac{\xi}{\epsilon'_T}. \end{aligned} \tag{32}$$

Имеем

$$\sum_{t=1}^{T}\left(\frac{1}{\epsilon'_t} - \frac{1}{\epsilon'_{t-1}}\right) M(\tilde{s}_{1:t}) \cdot \xi = \sum_{t=1}^{T}(\mu_t v_t - \mu_{t-1} v_{t-1}) M(\tilde{s}_{1:t}) \cdot \xi. \tag{33}$$

Мы будем использовать неравенство

$$0 \leq E(M(\tilde{s}_{1:t}) \cdot \xi) \leq E(M(\xi) \cdot \xi) = E(\max_i \xi^i) \leq 1 + \ln N. \tag{34}$$

Доказательство этого неравенства использует оценку, полученную в лемме 1 из [7]. Для экспоненциально распределенных случайных величин ξ^i, $i = 1, \dots, N$, выполнено

$$P\{\max_i \xi^i \geq a\} = P\{\exists i(\xi^i \geq a)\} \leq \sum_{i=1}^{N} P\{\xi^i \geq a\} = N \exp\{-a\}. \tag{35}$$

Так как для каждой неотрицательной случайной переменной η выполнено

$$E(\eta) = \int_0^\infty P\{\eta \geq y\}dy,$$

по (35) получаем

$$E(\max_i \xi^i - \ln N) = \int_0^\infty P\{\max_i \xi^i - \ln N \geq y\}dy \leq$$

$$\int_0^\infty N \exp\{-y - \ln N\}dy = 1.$$

Таким образом, $E(\max_i \xi^i) \leq 1 + \ln N$.

По (34) математическое ожидание величины (33) имеет верхнюю оценку

$$\sum_{t=1}^T E(M(\tilde{s}_{1:t}) \cdot \xi)(\mu_t v_t - \mu_{t-1} v_{t-1}) \leq (1 + \ln N) \sum_{t=1}^T \mu_t \Delta v_t.$$

Здесь было использовано неравенство $\mu_t \leq \mu_{t-1}$ для всех t.

Так как $E(\xi^i) = 1$ для всех i, математическое ожидание последнего члена (32) равно

$$E\left(M(s_{1:T}) \cdot \frac{\xi}{\epsilon'_T}\right) = \frac{1}{\epsilon'_T} = \mu_T v_T. \tag{36}$$

Комбинируя оценки (31)–(33) и (36), получим

$$r_{1:T} = E\left(\sum_{t=1}^T M(\tilde{s}_{1:t}) \cdot s_t\right) \geq$$

$$\geq \max_i s^i_{1:T} + \mu_T v_T - (1 + \ln N) \sum_{t=1}^T \mu_t \Delta v_t \geq$$

$$\geq \max_i s^i_{1:T} - (1 + \ln N) \sum_{t=1}^T \mu_t \Delta v_t. \tag{37}$$

Лемма доказана. $\triangle$.

Закончим доказательство теоремы.

Неравенство (18) леммы 1, а также неравенство (29) леммы 2 влекут неравенство

$$E(s_{1:T}) \geq \max_i s^i_{1:T} - \sum_{t=1}^T (8(\gamma(t))^{1-\alpha_t} + (1 + \ln N)(\gamma(t))^{\alpha_t}) \Delta v_t. \tag{38}$$

для всех T.

Оптимальное значение (10) параметра α_t можно легко получить путем минимизации по α_t каждого слагаемого суммы (38). В этом случае μ_t равно (11), а (38) эквивалентно (15).

Допустим, что $\gamma(T) \to 0$ при $T \to \infty$, а игра является невырожденной, т.е. $v_T \to \infty$ при $T \to \infty$. Имеем $\sum_{t=1}^T \Delta v_t = v_T$ для всех T. По лемме Теплица [1]

$$\frac{1}{v_T}\left(2\sqrt{8(1 + \ln N)} \sum_{t=1}^T (\gamma(t))^{1/2} \Delta v_t\right) \to 0$$

при $T \to \infty$. Этот предел и неравенство (15) влекут асимптотическую состоятельность алгоритма FPL в среднем (16). Теорема доказана. $\triangle$

В работах [2] и [11] построены асимптотически состоятельные алгоритмы при полиномиальных верхних границах для пошаговых выигрышей. Наш алгоритм FPL также асимптотически состоятелен при полиномиальной верхней границе для пошаговых выигрышей.

Следствие 1. *Допустим, что* $\max_i |s_t^i| \le t^a$ *для всех* t *и*

$$\liminf_{t\to\infty} \frac{v_t}{t^{a+\delta}} > 0,$$

где $i = 1, \ldots, N$, a *и* δ *– положительные вещественные числа. Тогда*

$$E(s_{1:T}) \ge \max_i s_{1:T}^i - O(\sqrt{(1+\ln N)}T^{1-\frac{1}{2}\delta+a})$$

при $T \to \infty$, *где* $\gamma(t) = t^{-\delta}$ *и* μ_t *определены по* (11).

4. ПРИЛОЖЕНИЯ К ИГРЕ С НУЛЕВОЙ СУММОЙ

Для приложения нам не нужен случайный выбор эксперта в алгоритме FPL. Для построения смеси двух игровых стратегий используются вероятности выбора одной из двух стратегий. Поэтому мы проведем "дерандомизацию" алгоритма FPL. Мы интерпретируем математическое ожидание выигрышей $E(s_t)$ как взвешенную смесь пошаговых выигрышей экспертных стратегий. Более точно, на каждом шаге t, *Статистик* разделяет свои инвестиции пропорционально вероятностям экспертных стратегий (5) и (6), вычисленных алгоритмом FPL, и получает выигрыш

$$G_t = 2C(S_t - S_0)(P\{I_t = 1\} - P\{I_t = 2\})\Delta S_t$$

на шаге t, где C – произвольная положительная константа; $G_{1:T} = \sum_{t=1}^{T} G_t = E(s_{1:T})$ – суммарный выигрыш *Статистика*.

Если игра удовлетворяет условию $|s_t^1| / \sum_{i=1}^{t} |s_i^1| \le \gamma(t)$ для всех t, то по (15) мы получаем нижнюю оценку суммарного выигрыша

$$G_{1:T} \ge |s_{1:T}^1| - 8\sum_{t=1}^{T} (\gamma(t))^{1/2} |s_t^1|$$

для всех T.

Допустим, что $\gamma(t) = \mu$ для всех t. Тогда

$$G_{1:T} \ge |s_{1:T}^1| - 8\mu^{1/2} v_T$$

для всех T. Для получения этой оценки нам не нужно предполагать существование вычислимой верхней оценки $\gamma(t) = o(t)$ для флуктуаций игры.

Алгоритм FPL легко реализовать. На рис. 1–3 приведены результаты численных экспериментов игры на временном ряде, состоящем из почасовых значений цены акции компании ГАЗПРОМ 2009 г. (всего 100 значений), выгруженных с сайта FINAM.[6] На рис. 1 приведен график эволюции цены, на рис. 2 приведен график флуктуации игры.

[6] http://www.finam.ru/analysis/export/default.asp.

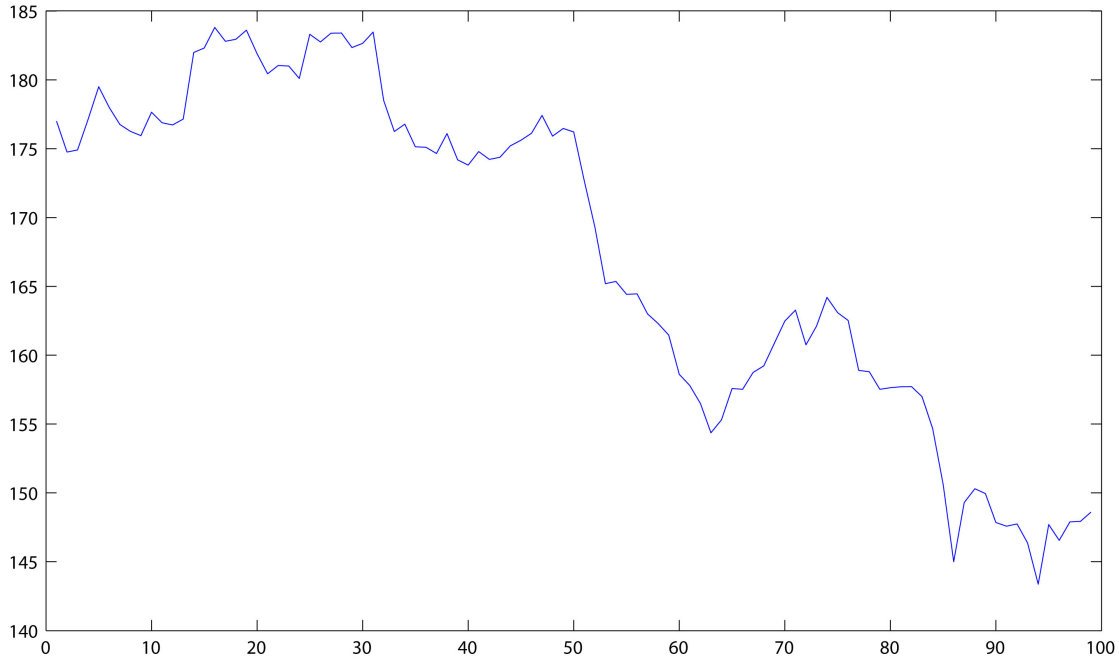

Рис. 1. Эволюция цены акции ГАЗПРОМа за 100 часов.

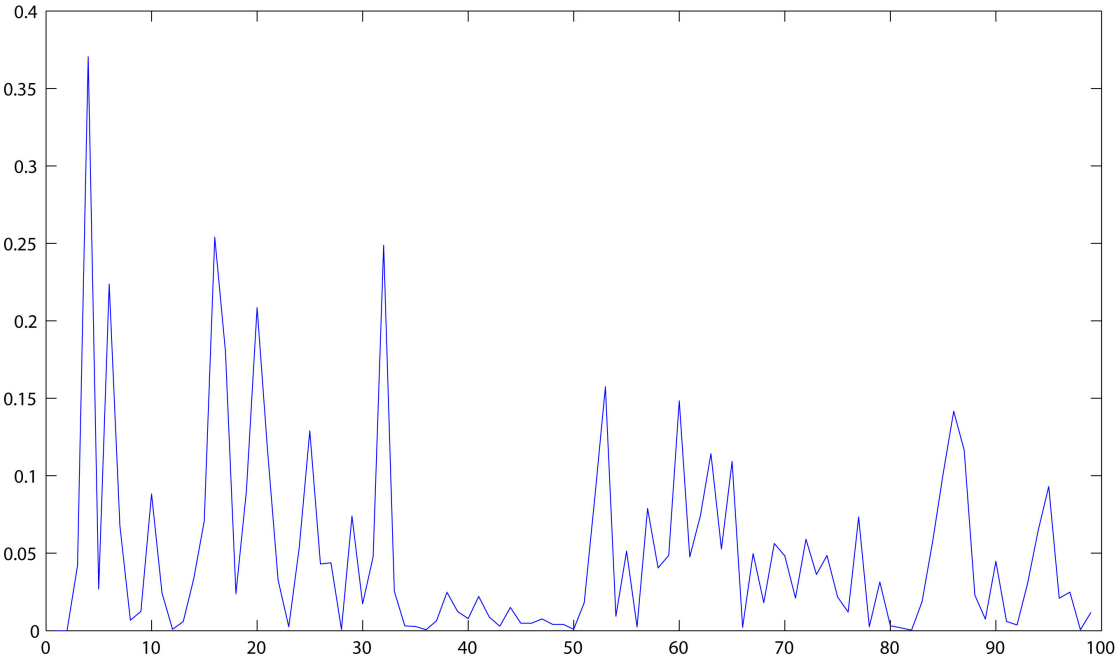

Рис. 2. Флуктуация игры.

На рис. 3 приведен график изменения выигрыша дерандомизированного алгоритма FPL за 100 раундов игры. Для вычисления весов алгоритма использовалась функция $\gamma(t) = t^{-1/2}$. По горизонтальной оси отложено время, по вертикальной оси – значения выигрышей. Две симметричные кривые – выигрыши экспертных стратегий. На протяжении игры выигрыши каждой из этих стратегий несколько раз были положительными или отрицательными. Пунктирная кривая между ними – эволюция среднего выигрыша алгоритма FPL, который тем не менее оказался положительным в конце периода. На рис. 3 верхняя кривая изображает эволюцию объема игры.

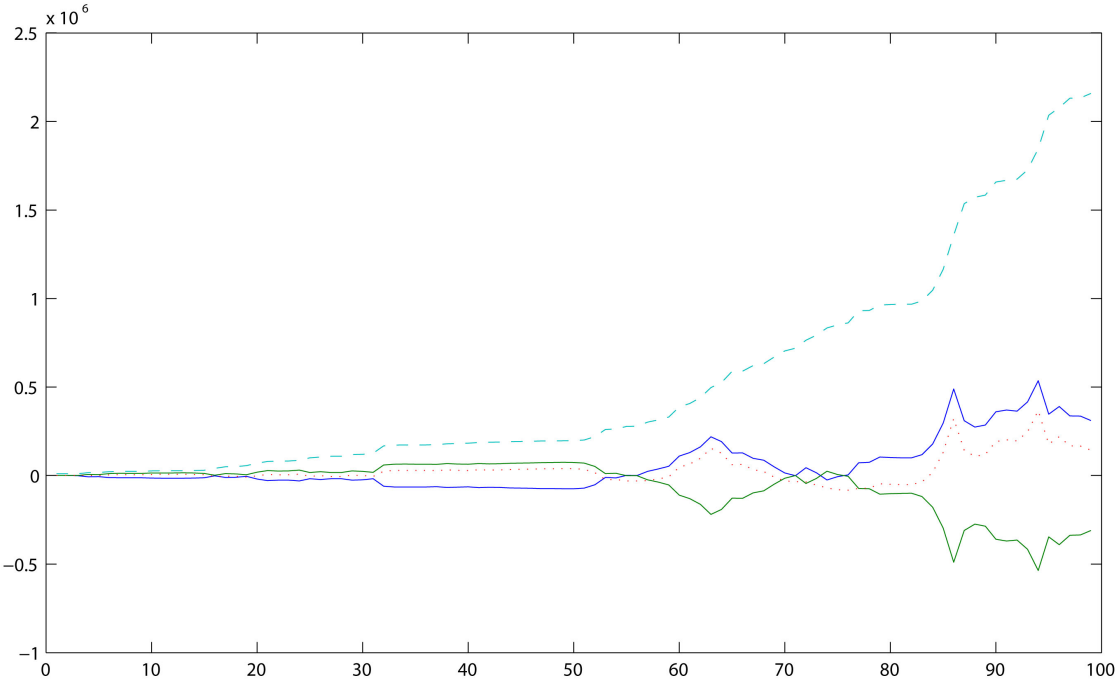

Рис. 3. Две симметричные кривые – выигрыши экспертных стратегий. Пунктирная кривая между ними – это средний выигрыш алгоритма FPL. Верхняя кривая – объем игры.

Приведенный алгоритм и его программная реализация могут использоваться для технического анализа временных рядов цен финансовых инструментов (акций, фьючерсов и др.).

5. ЗАКЛЮЧЕНИЕ

В данной работе мы рассматривали две различные задачи: первая из них заключается в том как использовать фрактальность броуновского движения цен финансового инструмента для получения выигрыша в некоторой естественной финансовой игре; вторая задача заключается в расширении методов теории предсказания с учетом экспертных стратегий на случай априори неограниченных выигрышей и потерь на каждом шаге игры. Несмотря на то, что эти проблемы выглядят далекими друг от друга, первая из них послужила мотивирующим примером для решения второй задачи, а вторая задача использовалась для решения первой задачи.

Некоторые задачи остаются нерешенными. Можно ли построить безопасные игровые стратегии в смысле книги [13]?[7]

В рамках теории предсказаний с использованием экспертных стратегий было бы полезно провести аналогичный анализ известных смешивающих алгоритмов (типа алгоритма взвешенного большинства) в случае неограниченных выигрышей и потерь.

Также имеется пробел между предложением 1 и теоремой 2, так как мы предполагаем в этой теореме, что имеется вычислимая верхняя оценка для сходимости флуктуаций к нулю: $\mathrm{fluc}(t) \leq \gamma(t) \to 0$, где $\gamma(t)$ – вычислимая функция заданная априори. Кроме этого, алгоритм использует эту функцию в своей работе для настройки параметров. Возникает вопрос, существуют ли смешивающие алгоритмы асимптотически состоятельные для случая когда $\mathrm{fluc}(t) \to 0$ при $t \to \infty$, которые не используют подобные верхние оценки.

СПИСОК ЛИТЕРАТУРЫ

1. Ширяев, А.Н. *Вероятность*. М.: Наука, 1980.
2. Allenberg, C., Auer, P., Gyorfi, L., Ottucsak, G. Hannan consistency in on-Line learning in case of unbounded losses under partial monitoring. In *Proceedings of the 17th International Conference ALT 2006* (Jose L. Balcazar, Philip M. Long, Frank Stephan (Eds.)), Lecture Notes in Computer Science. V. 4264. P. 229–243, Springer-Verlag, Berlin, Heidelberg, 2006.
3. Cheredito, P. Arbitrage in fractional Brownian motion. *Finance and Statistics*. 2003. V. 7. P. 533–553
4. Delbaen, F., Schachermayer, W. A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300 (1994), 463–520.
5. Freund, Y, Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*. 1997. V. 55. P.119–139
6. Hannan, J. Approximation to Bayes risk in repeated plays. In M. Dresher, A.W. Tucker, and P. Wolfe, editors, *Contributions to the Theory of Games*. V. 3. P. 97–139. Princeton University Press, 1957.
7. Hutter, M., Poland, J. Prediction with expert advice by following the perturbed leader for general weights. In *Proceedings of the 15th International Conference ALT 2004* (S.Ben-Dawid, J.Case, A.Maruoka (Eds.)), LNAI, V. 3244. P. 279–293. Springer-Verlag, Berlin, Heidelberg, 2004.
8. Kalai, A., Vempala, S. Efficient algorithms for online decisions. In *Proceedings of the 16th Annual Conference on Learning Theory COLT 2003* (Bernhard Scholkopf, Manfred K. Warmuth (Eds.)), LNCS, Volume 2777, 506-521, Springer-Verlag, Berlin, 2003. Extended version in Journal of Computer and System Sciences. 2005. V. 71. P. 291–307.
9. Littlestone, N., Warmuth, M.K. The weighted majority algorithm // *Information and Computation*. 1994. V. 108. P. 212–261.
10. Lugosi, G., Cesa-Bianchi, N. *Prediction, Learning and Games*. Cambridge University Press, New York, 2006.
11. Poland, J., Hutter, M. Defensive universal learning with experts. for general weight. In *Proceedings of the 16th International Conference ALT 2005* (S.Jain, H.U.Simon and E.Tomita (Eds.)), LNAI, Volume 3734, 356–370. Springer-Verlag, Berlin, Heidelberg, 2005.
12. Rogers, C. Arbitrage with fractional Brownian motion. *Mathematical Finance*. 1997. V. 7. P. 95–105
13. Shafer, G., Vovk, V. *Probability and Finance. It's Only a Game!* New York: Wiley. 2001.
14. Vovk, V. *A game-theoretic explanation of the* $\sqrt{dt}$ *effect*, Working paper #5. 2003, http://www.probabilityandfinance.com.

[7] Это означает, что *Статистик* может начать игру с некоторого капитала, при этом его стратегия никогда не приведет его к долгу, а в случае, когда цены следуют фрактальному броуновскому движению, *Статистик* даже получит выигрыш.

15. V'yugin, V.V. Learning Volatility of Discrete Time Series Using Prediction with Expert Advice. O. Watanabe and T. Zeugmann (Eds.), *SAGA 2009, Lecture Notes in Computer Science* 5792. P. 16–30, Springer-Verlag Berlin Heidelberg 2009.

FOLLOW THE PERTURBED LEADER ALGORITHM AND ITS APPLICATION FOR CONSTRUCTING GAME STRATEGIES

V.V. V'yugin

Institute for Information Transmission Problems,
Russian Academy of Sciences, Moscow, Russia
e-mail: vyugin@iitp.ru

Abstract: A modification of Follow the Perturbed Leader Algorithm is considered for the case where gains or losses of experts received or suffered at each step cannot be bounded in advance. We present the bounds for regret of this algorithm in terms of new notions of volume and fluctuation of the game. We prove that this algorithm has optimal performance in the case where fluctuations of one-step gains or losses of experts of the pool tend to zero. An application of the algorithm for algorithmic trading is presented. We define two expert strategies for buying and selling shares of a stock which receive arbitrage if there is a disagreement between "macro" and "micro" volatility of the stock price time series. We combine these strategies in one strategy using the derandomizing version of our master algorithm.

Keywords: prediction with expert advice, follow the perturbed leader, unbounded losses, adaptive learning rate, Hannan consistency, algorithmic trading strategy.