

Онлайн агрегирование вероятностных прогнозов на основе скорингового правила непрерывной ранжированной вероятности

В.В. Вьюгин, В.Г. Трунов

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

Поступила в редколлегию 11.05.2019

Аннотация—Рассматриваются методы формирования прогнозов в режиме онлайн и в форме вероятностных распределений будущих исходов. Различие между вероятностным прогнозом (распределением вероятностей) и числовым исходом измеряется с помощью функции потерь (скорингового правила). В практической статистике для оценки расхождения между вероятностными прогнозами и (количественными) исходами часто используется скоринговое правило непрерывной ранжированной вероятности (Continuous Ranked Probability Score – CRPS). В работе рассматривается случай, когда несколько конкурирующих методов (экспертов) дают свои прогнозы в виде функций распределения. Предложен алгоритм для агрегирования этих функций распределения в режиме онлайн. Получены оценки ошибки предложенного алгоритма в форме сравнения кумулятивных потерь алгоритма и потерь экспертных гипотез. В отличие от ранее известных оценок полученные оценки не зависят от времени. Представлены результаты численных экспериментов, иллюстрирующих предложенные методы.

КЛЮЧЕВЫЕ СЛОВА: предсказания с использованием экспертных стратегий, принятие решений в режиме онлайн, регрет, агрегирующий алгоритм, корректные скоринговые правила, правило непрерывной ранжированной вероятности.

1. ВВЕДЕНИЕ

Вероятностные прогнозы в форме распределения вероятностей будущих событий нашли широкое применение в таких областях, как метеорология, гидрология, экономика и демография (см. обсуждение в [1]). Вероятностные предсказания также используются в теории конформных предсказаний, где вероятностное распределение без допущений относительно его специфических параметров может быть присвоено любому алгоритму прогнозирования (см. [2]).

Различие между прогнозом и исходом измеряется функцией потерь (скоринговым правилом). Известными примерами скоринговых правил для непрерывных исходов являются логарифмическое правило и правило непрерывной ранжированной вероятности.

Логарифмическое скоринговое правило ([3]) определяется как $\text{LogS}(F, y) = -\log(F(y))$, где F – функция распределения вероятностей.

Правило непрерывной ранжированной вероятности (CRPS) определяется как

$$\text{CRPS}(F, y) = \int (F(u) - 1_{u \geq y})^2 du,$$

где $F(u)$ – функция распределения вероятностей, а y -исход (вещественное значение).¹

¹ Здесь $1_{u \geq y} = 1$, если $u \geq y$, в противном случае это 0.

Мы рассматриваем случай, когда несколько конкурирующих методов дают онлайн прогнозы в виде функций распределения вероятностей. Эти прогнозы могут привести к большим или меньшим потерям. Наша задача – объединить эти прогнозы в один агрегированный прогноз таким образом, чтобы минимизировать потери, исходя из имеющейся прошлой информации.

Эта проблема решается в рамках теории предсказания с использованием экспертных стратегий (Prediction with Expert Advice – PEA). Каждая экспертная стратегия определяется некоторой предсказательной моделью процесса, порождающего наблюдаемые данные. С течением времени модели могут становиться более или менее релевантными реальности. Факторы, влияющие на релевантность модели, скрыты от предсказателя, поэтому задача предсказания решается в рамках теоретико-игрового подхода, при котором не используется никаких предположений о том какая из возможных моделей релевантна в данный момент времени. При таком подходе агрегирующий алгоритм в режиме онлайн объединяет прогнозы экспертных моделей в зависимости от их потерь в прошлом. Разность между кумулятивными потерями агрегирующего алгоритма и кумулятивными потерями экспертного алгоритма за весь период прогнозирования называется регретом. Цель агрегирующего алгоритма заключается в том, чтобы минимизировать регрет относительно любой экспертной стратегии (см., например, [4], [5], [6], [7], [8]). В отличие от стандартной постановки задачи предсказания с использованием экспертных стратегий, в данной работе рассматривается случай, когда каждый эксперт представляет функцию вероятностного распределения, а не точечный прогноз. Агрегирующий алгоритм также представляет свой прогноз в форме функции распределения вероятностей, рассчитанной с использованием вероятностных прогнозов экспертов.

Качество экспертов и их прогнозов будем в данной работе измерять с помощью скорингового правила непрерывной ранжированной вероятности CRPS как функции потерь. На каждом временном шаге каждый эксперт выдает свой прогноз в форме распределения вероятностей. Агрегирующий алгоритм объединяет эти прогнозы в один консолидированный прогноз, который также является функцией распределения вероятностей. Эффективность агрегирующего алгоритма на любом временном интервале $[1, T]$ измеряется регретом, который представляет собой разницу между кумулятивными потерями агрегирующего алгоритма и кумулятивными потерями наилучшего эксперта.

Опубликовано много статей о вероятностных предсказаниях с использованием скорингового правила CRPS (напр., [1], [10], [11], [12], [13], [14], [15]). В подавляющем большинстве работ используются ансамблевые прогнозы и квантильные оценки функций распределения. Под ансамблями обычно понимают совокупность траекторий модели, вычисленных при измененных начальных условиях или возмущенных уравнениях модели. Ансамбль прогнозов может быть преобразован в кусочно-постоянную функцию распределения. Так, в работе [12] показано, что для вычисления CRPS можно интерпретировать ансамбль как квантильные оценки с неоднородными уровнями.

В данной работе рассматриваются также и другие формы представления вероятностных распределений. В некоторых случаях эксперты строят прогнозы на таких моделях данных (распределениях вероятностей), которые эксплицитно определены в аналитической форме. В настоящей статье предлагаются правила агрегирования таких моделей. Представлены формулы для прямого расчета агрегированного распределения вероятностей на основе распределений вероятностей, представленных экспертами в качестве их прогнозов.

Предлагаемые правила работают как в случае аналитических моделей, так и в случае использования детальных численных описаний функций распределения на основе ансамблевых или квантильных прогнозов.

Авторы [15] использовали метод экспоненциального онлайн смешивания для агрегирования ансамбля вероятностных прогнозов на основе CRPS в качестве функции потерь. Они показали,

что в этом случае теоретическая верхняя граница регрета составляет $O(\sqrt{T \ln N})$, где N – количество экспертов, а T – продолжительность временного интервала.

В этой статье мы получаем более точную верхнюю границу регрета для специального случая, когда результаты измерения и распределения вероятностей находятся в конечном интервале $[a, b]$ вещественной прямой. Мы показываем, что функция потерь $\text{CRPS}(F, y)$ является смешиваемой в смысле [8] и применяем агрегирующий алгоритм для получения не зависящей от времени верхней границы $\frac{b-a}{2} \ln N$ для регрета.²

В рамках теории предсказания с использованием экспертных стратегий процесс обучения представлен в виде игры. Эксперты и агрегирующий алгоритм наблюдают прошлые исходы, которые генерируются некоторым неизвестным им механизмом (называемым природой), и представляют свои прогнозы для неизвестного следующего исхода. После этого исход раскрывается природой и эксперты и агрегирующий алгоритм исчисляют свои потери на данном шаге.

В разделе 2 представлены некоторые детали агрегирующего алгоритма Вовка [8], а также его расширение на случай векторных прогнозов. В разделе 3 представлен метод вычисления агрегированного распределения вероятностей на основе распределений, представленных экспертами. Эффективность предложенных методов продемонстрирована результатами численных экспериментов, приведенных в разделе 4.

2. ПРЕДВАРИТЕЛЬНЫЕ СВЕДЕНИЯ

В этом разделе приведены основные определения и вспомогательные результаты теории предсказания с использованием экспертных стратегий, в частности, обучения со смешиваемыми функциями потерь.

Агрегирующий алгоритм. Рассмотрим обучение с функциями потерь $\lambda(\gamma, \omega)$, где $\omega \in \{0, 1\}$ – исход, а $\gamma \in [0, 1]$ – прогноз. Задан набор экспертов $E = \{1, \dots, N\}$.

Рассматривается следующая игра. В любом раунде $t = 1, 2, \dots$ каждый эксперт $i \in E$ выдает свой прогноз $f_{i,t}$, затем агрегирующий алгоритм формирует свое решение f_t , после чего становится известен исход ω_t . Каждый эксперт i несет потери $\lambda(f_{i,t}, \omega_t)$ и агрегатор несет потери $\lambda(f_t, \omega_t)$. Представим эту игру в виде протокола:

Определим $w_{i,1} = \frac{1}{N}$ для $1 \leq i \leq N$.

FOR $t = 1, \dots, T$

1. Получаем прогнозы экспертов $\mathbf{e}_t = (f_{i,t}^1, \dots, f_{i,t}^N)$, где $1 \leq i \leq N$.
2. Вычисляем прогноз алгоритма согласно правилу подстановки $f_t = \text{Subst}(\mathbf{e}_t, \mathbf{w}_t^*)$.³
3. Наблюдаем исход ω_t и вычисляем потери $\lambda(\omega_t, f_{i,t})$ экспертов $1 \leq i \leq N$, а также потери алгоритма $\lambda(\omega_t, f_t)$.
4. Обновляем веса экспертов $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\eta \lambda(\omega_t, f_{i,t})}.$$

ENDFOR

Агрегирующий алгоритм Вовка ([6], [8]) использует свой метод для вычисления прогноза агрегатора. Этот алгоритм присваивает экспертам веса $w_{i,t}$ и использует следующее правило

² Точные определения приведены в разделе 2.

³ Например, это могут быть правила (5) или (6), приведенные ниже.

обновления веса: $w_{i,1} = \frac{1}{N}$ и

$$w_{i,t+1} = w_{i,t} e^{-\eta \lambda(f_{i,t}, \omega_t)} \text{ for } t = 1, 2, \dots \quad (1)$$

Нормализованные веса определяются

$$w_{i,t}^* = \frac{w_{i,t}}{\sum_{j=1}^N w_{j,t}}. \quad (2)$$

Основным инструментом алгоритма является функция суперпредсказания

$$g_t(\omega) = -\frac{1}{\eta} \ln \sum_{i=1}^N e^{-\eta \lambda(f_{i,t}, \omega)} w_{i,t}^*. \quad (3)$$

Согласно [8], функция потерь $\lambda(\gamma, \omega)$ называется η -смешиваемой, если для произвольного распределения вероятностей $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ на множестве всех экспертов и для произвольного набора их предсказаний $\mathbf{e}_t = (f_{1,t}, \dots, f_{N,t})$ существует такое предсказание f_t , что выполнено неравенство

$$\lambda(f_t, \omega) \leq g_t(\omega) \text{ для всех } \omega. \quad (4)$$

Фиксируем некоторое правило вычисления такого прогноза f_t и будем писать $f_t = \text{Subst}(\mathbf{e}_t, \mathbf{w}_t^*)$. Соответствующая функция называется функцией или правилом подстановки.

Согласно [16], для η -смешиваемой функции потерь $\lambda(\gamma, \omega)$, где $\omega \in \{0, 1\}$ – исход, а $\gamma \in [0, 1]$ прогноз, соответствующее правило подстановки можно определить как

$$f_t = \text{Subst}(g_t) = \frac{1}{2} - \frac{1}{2\eta} \ln \frac{\sum_{i=1}^N w_{i,t}^* e^{-\eta \lambda(f_{i,t}, 0)}}{\sum_{i=1}^N w_{i,t}^* e^{-\eta \lambda(f_{i,t}, 1)}}. \quad (5)$$

Квадратичная функция потерь $\lambda(\gamma, \omega) = (\omega - \gamma)^2$ является η -смешиваемой для любого $0 < \eta \leq 2$. Подробнее см. в приложении А.

Отметим, что если функция потерь $\lambda(\gamma, \omega)$ является η -смешиваемой, то функция потерь $\Delta \lambda(\gamma, \omega)$ является $\frac{\eta}{\Delta}$ -смешиваемой для любого $\Delta > 0$. Детали см. в [6], [8] и [16].

Смешиваемость является обобщением понятия экспоненциальной вогнутости [4]. Функция потерь $\lambda(\gamma, \omega)$ называется η -экспоненциально вогнутой, если для любого ω функция $\exp(-\eta \lambda(\gamma, \omega))$ является вогнутой по γ . Для такой функции по определению вогнутости неравенство $\lambda(f_t, \omega) \leq g_t(\omega)$ также выполняется для всех ω , где

$$f_t = \sum_{i=1}^N w_{i,t}^* f_{i,t}. \quad (6)$$

Квадратичная функция потерь является η -экспоненциально вогнутой для $0 < \eta < \frac{1}{2}$.

Анализ регрета. Пусть $H_T = \sum_{t=1}^T \lambda(f_t, \omega_t)$ – кумулятивные потери агрегирующего алгоритма, а $L_T^i = \sum_{t=1}^T \lambda(f_{i,t}, \omega_t)$ – кумулятивные потери эксперта i за T шагов игры. Разница $R_T^i = H_T - L_T^i$ называется регретом относительно эксперта i , а $R_T = H_T - \min_i L_T^i$ – регрет относительно наилучшего эксперта.

Пусть $W_t = \sum_{i=1}^N w_{i,t}$. По определению, $g_t(\omega_t) = -\frac{1}{\eta} \ln \frac{W_{t+1}}{W_t}$, где $W_1 = 1$. По правилу обновления весов (1) мы получаем $w_{i,t+1} = \frac{1}{N} e^{-\eta L_t^i}$.

Тогда, согласно (4) и, сокращая одинаковые множители, мы получаем

$$H_T \leq \sum_{t=1}^T g_t(\omega_t) = -\frac{1}{\eta} \ln W_{T+1} \leq L_T^i + \frac{\ln N}{\eta} \quad (7)$$

для любого эксперта i . Таким образом, на произвольном интервале времени $[1, T]$ кумулятивные потери алгоритма не превосходят кумулятивные потери любого из экспертов с точностью до $\frac{\ln N}{\eta}$. Иными словами, $R_T \leq \frac{\ln N}{\eta}$.

Векторно-значные предсказания. Рассмотрим случай, когда эксперты и агрегатор дают прогнозы в виде векторов размерности d : на каждом раунде $t = 1, 2, \dots$ эксперт $i \in \{1, \dots, N\}$ представляет вектор прогнозов $\mathbf{f}_{i,t} = (f_{i,t}^1, \dots, f_{i,t}^d)$, а агрегатор дает вектор прогнозов $\mathbf{f}_t = (f_t^1, \dots, f_t^d)$. После этого будет получен вектор $\omega_t = (\omega_t^1, \dots, \omega_t^d)$ исходов и вычислены потери экспертов и агрегирующего алгоритма.

В работе [17] представлен следующий метод вычисления векторно-значных прогнозов агрегирующего алгоритма. Правило агрегирования применяется к каждой координате отдельно: определим $f_t^s = \text{Subst}(\mathbf{e}_t^s, \mathbf{w}_t^*)$ for $1 \leq s \leq d$, где $\mathbf{e}_t^s = (f_{1,t}^s, \dots, f_{N,t}^s)$ и $\mathbf{w}_t^* = (w_{i,1}^*, \dots, w_{i,N}^*)$.

Перепишем неравенство $\lambda(f_t^s, \omega) \leq g_t(\omega)$ как

$$e^{\eta \lambda(f_t^s, \omega)} \geq \sum_{i=1}^N e^{-\eta \lambda(f_{i,t}^s, \omega)} w_{i,t}^* \quad (8)$$

для $1 \leq s \leq d$. Данное неравенство выполнено для всех ω .

После того, как последовательность результатов $\omega_t = (\omega_t^1, \dots, \omega_t^d)$ будет объявлена, эксперты и агрегатор несут потери $\lambda(\omega_t, \mathbf{f}_{i,t}) = \sum_{s=1}^d \lambda(f_{i,t}^s, \omega_t^s)$ и $\lambda(\mathbf{f}_t, \omega_t) = \sum_{s=1}^d \lambda(f_t^s, \omega_t^s)$, где $\mathbf{f}_t = (f_t^1, \dots, f_t^d)$ and $\mathbf{f}_{i,t} = (f_{i,t}^1, \dots, f_{i,t}^d)$. Назовем функцию λ с векторно-значными аргументами обобщенной функцией потерь.

Перемножая неравенства (8) for $s = 1, \dots, d$, где $\omega = \omega_t^s$, мы получаем

$$e^{-\eta \sum_{s=1}^d \lambda(f_t^s, \omega_t^s)} \geq \prod_{s=1}^d \sum_{i=1}^N e^{-\eta \lambda(f_{i,t}^s, \omega_t^s)} w_{i,t}^*. \quad (9)$$

Согласно обобщенному неравенству Гёльдера,

$$\|F_1 F_2 \cdots F_d\|_r \leq \|F_1\|_{q_1} \|F_2\|_{q_2} \cdots \|F_d\|_{q_d},$$

где $\frac{1}{q_1} + \cdots + \frac{1}{q_d} = \frac{1}{r}$, $q_s \in (0, +\infty)$ и $F_s \in L^{q_s}$ для $1 \leq s \leq d$. Пусть $q_s = 1$ для всех $1 \leq s \leq d$, тогда $r = 1/d$.

Пусть $F_{i,s} = e^{-\eta \lambda(f_{i,t}^s, \omega_t^s)}$ для $s = 1, \dots, d$ и $\|F_s\|_1 = E_{i \sim \mathbf{w}^*} [F_{i,s}] = \sum_{i=1}^N F_{i,s} w_{i,t}^*$. Тогда

$$e^{-\eta \frac{1}{d} \sum_{s=1}^d \lambda(f_t^s, \omega_t^s)} \geq \sum_{i=1}^N e^{-\eta \frac{1}{d} \sum_{s=1}^d \lambda(f_{i,t}^s, \omega_t^s)} w_{i,t}^*.$$

или, эквивалентно,

$$e^{-\frac{\eta}{d} \lambda(\mathbf{f}_t, \omega_t)} \geq \sum_{i=1}^N e^{-\frac{\eta}{d} \lambda(\mathbf{f}_{i,t}, \omega_t)} w_{i,t}^*. \quad (10)$$

Неравенство (10) означает, что обобщенная функция потерь $\lambda(\mathbf{f}_t, \omega_t)$ является $\frac{\eta}{d}$ -смешиваемой.

3. АГРЕГИРОВАНИЕ ВЕРОЯТНОСТНЫХ ПРОГНОЗОВ

Допустим, что исходы y принадлежат интервалу $[a, b]$ вещественных чисел, где $a < b$, а прогнозы для y будут функциями распределения вероятностей $F : [a, b] \rightarrow [0, 1]$.⁴ Расхождение между прогнозом F и исходом y измеряется с помощью скорингового правила непрерывной ранжированной вероятности CRPS.

$$\text{CRPS}(F, y) = \int_a^b (F(u) - 1_{u \geq y})^2 du, \quad (11)$$

где $1_{u \geq y}$ является индикаторной функцией ([18]). Наименьшее возможное значение 0 этой функции потерь достигается, когда F концентрируется на y , а во всех остальных случаях значение $\text{CRPS}(F, y)$ будет положительным. Мы также вводим вспомогательное представление исхода y в виде двоичной переменной $\omega_{u,y} = 1_{u \geq y}$.

Мы рассматриваем игру на предсказания с использованием экспертных стратегий, где прогнозы экспертов и агрегирующего алгоритма представлены в виде функций вероятностных распределений. На каждом шаге t игры каждый эксперт $i \in \{1, \dots, N\}$ представляет свой прогноз – функцию распределения вероятностей $F_t^i(u)$, а агрегатор представляет свой прогноз – функцию распределения вероятностей $F_t(u)$. После этого предъясвляется исход $y_t \in [a, b]$, а эксперты и агрегатор несут потери $\text{CRPS}(F_t^i, y_t)$ и $\text{CRPS}(F_t, y_t)$. Цель агрегатора состоит в том, чтобы выдавать такие прогнозы, суммарные потери вследствие которых на произвольном начальном временном интервале будут не больше суммарных потерь наилучшего на данном интервале эксперта с точностью до некоторого определенного регрета.

Мы рассматриваем такую игру как предел последовательности игр с векторными прогнозами. Для этого аппроксимируем любые функции вероятностного распределения $F(y)$ кусочно-постоянными функциями $L(y)$. Любая такая функция L определяется точками $z_0, z_1, z_2, \dots, z_d$ и значениями $f_0, f_1, f_2, \dots, f_d$, где $a = z_0 < z_1 < z_2 < \dots < z_d = b$ и $0 = f_0 < f_1 < f_2 < \dots < f_d = 1$. По определению $L(y) = f_1$ для $z_0 \leq y < z_1$, $L(y) = f_2$ для $z_1 \leq y < z_2$, ..., $L(y) = f_d$ для $z_{d-1} \leq y < z_d$. Также предположим, что $z_{i+1} - z_i = \Delta$ для всех $0 \leq i < d$. По определению $\Delta = \frac{b-a}{d}$. В этом случае скоринговое правило непрерывной ранжированной вероятности принимает значение

$$\text{CRPS}(L, y) = \Delta \sum_{i=1}^d (f_i - \omega_{y,i})^2,$$

где $\omega_{y,i} = 1_{z_i \geq y} \in \{0, 1\}$ для $1 \leq i \leq d$.

Отождествим функцию L с вектором $\mathbf{f} = (f_1, f_2, \dots, f_d)$.

Так как квадратичная функция $\lambda(\gamma, \omega) = (\gamma - \omega)^2$ является 2-смешиваемой, где $\omega \in \{0, 1\}$, по результатам раздела 2, то соответствующая обобщенная функция потерь $\sum_{i=1}^d (f_i - \omega_i)^2$ является $\frac{2}{d}$ -смешиваемой, а тогда функция

$$\lambda(\mathbf{f}, \omega) = \Delta \sum_{i=1}^d (f_i - \omega_i)^2$$

является $\frac{2}{d\Delta} = \frac{2}{b-a}$ -смешиваемой, где $\omega = (\omega_1, \dots, \omega_d)$, $\omega_i \in \{0, 1\}$, и $\mathbf{f} = (f_1, \dots, f_d)$. Также $\text{CRPS}(L, y) = \lambda(\mathbf{f}, \omega_y)$, где $\omega_y = (\omega_{y,1}, \dots, \omega_{y,d})$.

Временно фиксируем размер сетки Δ (и точки z_1, \dots, z_d). Рассмотрим следующий протокол для вычисления кусочно-постоянной аппроксимации L_t :

⁴ Функция распределения – это неубывающая функция $F(y)$, определенная на интервале $[a, b]$ так, что $F(a) = 0$ и $F(b) = 1$.

Протокол 1

Определим $w_{i,1} = \frac{1}{N}$ для $1 \leq i \leq N$.

FOR $t = 1, \dots, T$

1. Получаем предсказания экспертов $\mathbf{f}_{i,t} = (f_{i,t}^1, \dots, f_{i,t}^d)$. (значения кусочно-постоянной функции $L_{i,t}$ на соответствующих интервалах), где $1 \leq i \leq N$.
2. Вычисляем кусочно-постоянную функцию L_t , основанную на прогнозах $\mathbf{f}_t = (f_t^1, \dots, f_t^d)$ агрегирующего алгоритма, где f_t^s определяется (5), а именно,

$$f_t^s = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t} e^{-2(f_{i,t}^s)^2}}{\sum_{i=1}^N w_{i,t} e^{-2(1-f_{i,t}^s)^2}} \quad (12)$$

для $1 \leq s \leq d$.⁵

3. Наблюдаем исход y_t и вычисляем потери $\text{CRPS}(L_t^i, y_t) = \Delta \sum_{s=1}^d (f_t^s - \omega_t^s)^2$ экспертов $1 \leq i \leq N$ и потери $\text{CRPS}(L_t, y_t) = \Delta \sum_{s=1}^d (f_t^s - \omega_t^s)^2$ агрегирующего алгоритма, где $\omega_t^s = 1_{z_s \geq y_t}$.
4. Обновляем веса экспертов $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\frac{2}{b-a} \text{CRPS}(L_t^i, y_t)}.$$

ENDFOR

На основе анализа Раздела 2 с использованием (7) получаем

$$\sum_{t=1}^T \text{CRPS}(L_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(L_t^i, y_t) + \frac{b-a}{2} \ln N \quad (13)$$

для любого i .

Для произвольного t (и a, b), полагая $\Delta \rightarrow 0$ (или, эквивалентно, $d \rightarrow \infty$) в (12), получим выражение для вычисления прогноза агрегирующего алгоритма $F_t(u)$ при заданных прогнозах $F_t^i(u)$ экспертов $1 \leq i \leq N$.

$$F_t(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t} e^{-2(F_t^i(u))^2}}{\sum_{i=1}^N w_{i,t} e^{-2(1-F_t^i(u))^2}}, \quad (14)$$

где $w_{i,t+1} = w_{i,t} e^{-\frac{2}{b-a} \text{CRPS}(F_t^i, y_t)}$ вычисляется рекурсивно. Несложно убедиться в том, что $F_t(u)$ представляет собой функцию вероятностного распределения.

Теорема 1. Для любого i и T ,

$$\sum_{t=1}^T \text{CRPS}(F_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(F_t^i, y_t) + \frac{b-a}{2} \ln N. \quad (15)$$

Доказательство. Поскольку для данных i и T неравенство (13) выполняется при любом размере сетки и оценка регрета не зависит от Δ , полагая $\Delta \rightarrow 0$, получим аналогичное неравенство (15) для предельных величин. \triangle

⁵ Аналогичным образом можно применить правило (6).

Квадратичная функция потерь также является η -экспоненциально вогнутой для $0 < \eta < \frac{1}{2}$. В этом случае правило (14) может быть заменено

$$F_t(u) = \sum_{i=1}^N w_{i,t}^* F_t^i(u), \tag{16}$$

где $w_{i,t}^* = \frac{w_{i,t}}{\sum_{j=1}^N w_{j,t}}$ – нормализованные веса. Соответствующие веса вычисляются рекурсивно

$$w_{i,t+1} = w_{i,t} e^{-\frac{1}{2(b-a)} \text{CRPS}(F_t^i, y_t)}. \tag{17}$$

В этом случае верхняя оценка (15) изменяется на

$$\sum_{t=1}^T \text{CRPS}(F_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(F_t^i, y_t) + 2(b-a) \ln N.$$

Представим протокол для расчета агрегирования вероятностного прогноза, с учетом вероятности прогнозов экспертов.

Протокол 2

Определяем $w_{i,1} = \frac{1}{N}$ при $1 \leq i \leq N$.

FOR $t = 1, \dots, T$

1. Получаем прогнозы экспертов – функции вероятностного распределения $F_t^i(u)$, где $1 \leq i \leq N$.
2. Вычисляем прогноз агрегирующего алгоритма – функцию вероятностного распределения $F_t(u)$:

$$F_t(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t} e^{-2(F_t^i(u))^2}}{\sum_{i=1}^N w_{i,t} e^{-2(1-F_t^i(u))^2}}. \tag{18}$$

3. Получаем исход y_t и вычисляем потери $\text{CRPS}(F_t^i, y_t) = \int_a^b (F_t^i(u) - 1_{u \geq y_t})^2 du$ экспертов $1 \leq i \leq N$ и потери $\text{CRPS}(F_t, y_t) = \int_a^b (F_t(u) - 1_{u \geq y_t})^2 du$ агрегирующего алгоритма.
4. Обновляем веса экспертов $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\frac{2}{b-a} \text{CRPS}(F_t^i, y_t)} \tag{19}$$

ENDFOR

Для экспоненциально вогнутых функций правила (18) и (19) заменяются на (16) и (17).

4. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

В этом разделе приведены результаты экспериментов, выполненных на синтетических данных, использование которых обеспечивает полную информацию об источниках неопределенности и позволяет количественно оценить и сравнить результаты применения методов агрегирования. Для многих распределений вероятностей моментные и усредненные значения CRPS могут быть представлены и вычислены в виде точных формул. Список таких распределений и пакеты программ для точных вычислений можно найти в работах ([1], [13], [15]).

Описание генерирующей модели. Исходные данные для проведения экспериментов были получены в результате выборки из смеси трех распределений вероятностей (генерирующих

моделей), где веса компонент смеси (априорные вероятности выбора каждого из трех элементов смеси) изменяются во времени по некоторому известному закону. В качестве генерирующих моделей для каждой из компонент смеси использовали усеченное нормальное распределение $\mathcal{N}(\mu_k, \sigma_k)$, где параметры заданы заранее и не изменяются в процессе эксперимента. Параметры усечения распределений $a < b$ задают ограниченный интервал изменения всех компонент исследуемого ряда. Временной интервал состоит из нескольких сегментов одинаковой длины, а веса компонент смеси зависят от времени. Производилось сравнение двух правил формирования агрегированного вероятностного распределения (14) и (16) на материале трех вариантов смешанной вероятностной модели, отличающихся характером изменения во времени относительных весов генерирующих моделей. В первом, наиболее простом варианте смешивания, на каждом из временных сегментов лидирует только одна из генерирующих моделей, которой приписывают априорную вероятность 1. При переходе от одного временного сегмента к другому лидер меняется. Во втором варианте в каждый момент времени смешиваются только две из трех генерирующих моделей и, наконец, в третьем эксперименте веса всех трех компонент непрерывно и гладко меняются во времени.

Описание прогнозирующей модели. Для прогнозирования числовых данных, получаемых в режиме онлайн от соответствующих смесей генерирующих моделей, использовался алгоритм, схема которого приведена в виде протокола 2. При этом использовались два правила вычисления прогнозов (14) и (16) и два соответствующих им правила адаптации весов (19) и (17).

При адаптации весов агрегирующего алгоритма использовали модификацию Fixed Share (см. [19]) протокола 2, где правило (19) или правило (17) адаптации весов заменяется на двухэтапную схему

$$w_{i,t}^\mu = \frac{w_{i,t} e^{-\frac{2}{b-a} \text{CRPS}(F_t^i, y_t)}}{\sum_{j=1}^N w_{j,t} e^{-\frac{2}{b-a} \text{CRPS}(F_t^j, y_t)}},$$

$$w_{i,t+1} = \frac{\alpha}{N} + (1 - \alpha)w_{i,t}^\mu,$$

где $0 < \alpha < 1$. В представленных экспериментах $\alpha = 0.001$.⁶

Перефразируя описание численной модели эксперимента, можно сказать, что есть три эксперта $i = 1, 2, 3$, каждый из которых предполагает, что исследуемый временной ряд получен в результате выборки из одного из заданных фиксированных распределений вероятностей. А именно, нормальных распределений с параметрами μ_i и σ_i усеченных на отрезке $[a, b]$.

На шаге t эксперимента каждый из трех экспертов оценивает балл очередной точки сигнала y_t на основании скорингового правила CRPS применительно к его распределению F_i , а агрегирующий алгоритм в онлайн стиле корректирует веса на основе анализа кумулятивных потерь на всей предыстории.

Результаты численных экспериментов 1–3. На рисунке 2 показаны графики плотности трех генерирующих распределений на отрезке усечения $[a, b] = [-10, 10]$ и графики скоринговой функции CRPS для этих распределений. Области определения всех трех генерирующих моделей пересекаются, но все они ограничены отрезком $[a, b]$, чтобы выполнить условие применимости метода агрегирования (14). На рисунке 3 слева показаны фрагменты случайных траекторий трех генерирующих моделей, а на графике справа приведены соответствующие этим моделям плотности вероятности.

Было проведено три вычислительных эксперимента, отличающихся характером формирования смеси генерирующих моделей. На рисунках 4, 9, 14 показаны основные этапы фор-

⁶ Используя переменный параметр α , мы можем получить границу регрета $O(\ln(TN))$ для соответствующего алгоритма.

мирования искусственных данных для проведения трех численных экспериментов. На этих рисунках в разделе А для наглядности приведены реализации случайных траекторий генерирующих моделей. В разделе В для тех же временных интервалов отображены априорные вероятности (фактические относительные веса), которые были использованы при агрегировании распределений вероятностей. В разделе С показаны результаты смешивания реализаций генерирующих моделей из раздела А с априорными вероятностями из раздела В.

Перейдем теперь к рассмотрению результатов вычислительных экспериментов. На Рисунке 5 приведены результаты онлайн-адаптации весов для эксперимента 1 с одним выраженным лидером на каждом временном сегменте. В первых трех разделах А, В и С, соответствующих генерирующим моделям (экспертам), показаны графики изменения адаптируемых весов по методу (14) (штриховая линия) и (16) (сплошная линия) в сравнении с истинным весом (пунктирная линия). Как видим, оба метода смешивания следуют за изменениями модели. При этом алгоритм, использующий правило (14), заметно быстрее адаптируется к новому лидеру, что можно объяснить более высокой скоростью обучения, которая в методе (14) оказывается в 4 раза выше, чем в методе (16).⁷ Это различие приводит к тому, что в эксперименте 1 кумулятивные потери алгоритма, использующего метод агрегирования (14), оказываются заметно меньше кумулятивных потерь алгоритма, использующего метод (16).

Поскольку в качестве критерия точности адаптации используется скоринговое правило CRPS, то естественно сравнивать методы агрегирования по величине накопленных потерь с потерями “идеального” алгоритма смешивания, который в каждый момент времени “знает” истинные значения компонент смеси, представленных в анализируемом временном ряду.

На рисунке 6 приведены графики суммарных потерь трех экспертов (непрерывные кривые), а также суммарные потери методов агрегирования (14) и (16) (штриховая и пунктирная кривые) в сравнении с накопленными потерями идеальной модели (штрих-пунктирная кривая). Видно некоторое преимущество метода (14) перед методом (16).

На следующих двух группах рисунков (9)–(16) показаны аналогичные результаты для экспериментов 2 и 3. В этих экспериментах используются более сложные способы смешивания исходных генерирующих моделей: в эксперименте 2 в каждый момент времени используется динамически изменяющаяся смесь двух из трех генерирующих моделей, а в эксперименте 3 используется динамически изменяющаяся смесь всех трех генерирующих моделей.

Во всех этих случаях агрегирующий алгоритм имеет существенно меньшие кумулятивные потери по сравнению с кумулятивными потерями исходных генерирующих моделей. При этом кумулятивные потери агрегирующего алгоритма близки к кумулятивным потерям истинной модели смешивания генерирующих моделей. При использовании более сложных вариантов динамического смешивания генерирующих моделей преимущество метода агрегирования (14) перед методом (16) уменьшается.

5. ЗАКЛЮЧЕНИЕ

В данной статье рассматривается проблема агрегирования вероятностных прогнозов. В этом случае популярным примером корректного скорингового правила для непрерывных исходов является непрерывный ранжированный показатель вероятности CRPS.

Представлен теоретический анализ возможностей использования функции непрерывной ранжированной вероятности CRPS для получения агрегированных прогнозов с использованием экспертных стратегий.

⁷ Отметим, что задача агрегирующего алгоритма как можно быстрее адаптироваться к изменившейся ситуации и повышать вес лидирующей модели.

Доказано, что функция потерь CRPS является смешиваемой и экспоненциально вогнутой. На основе этих свойств представлено обобщение агрегирующего алгоритма Вовка [8] на случай вероятностных прогнозов.

Представлены два правила вычисления прогнозов (14) и (16) на основе алгоритма Вовка [8]. Получены не зависящие от времени верхние границы регрета агрегирующего алгоритма. Получены точные формулы для вычисления агрегированного прогноза (распределения вероятностей) на основе вероятностных прогнозов, представленных экспертными моделями. Эти формулы могут быть использованы как в случае квантильных вероятностных прогнозов экспертов, полученных на основе экспериментальных данных, так и в тех случаях, когда эксперты используют аналитические модели для вычисления своих прогнозов.

Очевидным ограничением этих результатов является то, что они получены только для функций распределения сигнала, локализованных в конечных интервалах вещественной кривой.

Теоретические выводы проиллюстрированы результатами компьютерных экспериментов. Представлены результаты численных экспериментов на основе предложенных методов. Эти результаты показывают, что использование правил агрегирования прогнозов экспертов (14) и (16) приводят к существенно меньшим потерям, чем потери самих экспертов, и близки к потерям истинной модели генерации данных.

Эксперименты показали, что оба правила вычисления вероятностных прогнозов (14) и (16) приводят к аналогичным эмпирическим совокупным потерям в экспериментах 2 и 3, где использовались более сложные способы комбинирования исходных генерирующих моделей.

Теоретическая оценка регрета, полученного при пользовании правила (14), в четыре раза меньше оценки регрета, полученного при использовании правила (16). На практике это различие проявилось в первом эксперименте, в котором происходили резкие изменения генерирующей модели. В этом эксперименте преимущество метода (14) проявилось особенно заметно.

СПИСОК ЛИТЕРАТУРЫ

1. A. Jordan, F. Krüger, S. Lerch. Evaluating Probabilistic Forecasts with scoringRules, arXiv:1709.04743.
2. Vladimir Vovk, Jieli Shen, Valery Manokhin, Min-ge Xie. Nonparametric predictive distributions based on conformal prediction *Machine Learning*. 60, 82–102, 2017. ISSN: 0885-6125 (Print) 1573-0565 (Online)
3. I.J. Good. Rational Decisions. *Journal of the Royal Statistical Society B*, 14(1), 107–114, 1952. <https://www.jstor.org/stable/2984087>.
4. N. Cesa-Bianchi, G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
5. Y. Freund, R.E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*. 55:119–139, 1997.
6. V. Vovk, Aggregating strategies. In M. Fulk and J. Case, editors, *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, 371–383. San Mateo, CA, Morgan Kaufmann, 1990.
7. N. Littlestone, M. Warmuth. The weighted majority algorithm. *Information and Computation*. 108:212–261, 1994.
8. V. Vovk, A game of prediction with expert advice. *Journal of Computer and System Sciences*. 56(2), 153–173, 1998.
9. G.W. Brier. Verification of forecasts expressed in terms of probabilities. *Mon. Weather Rev.* 78. 1–3, 1950.
10. J. Bröcker, L.A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting* 22. 382–388, 2007.

11. J. Bröcker, L.A. Smith. From ensemble forecasts to predictive distribution functions. *Tellus A* 60. 663–678, 2008.
12. J. Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Q. J. R. Meteorol. Soc.* 138, 1611–1617, July 2012 B
13. A.E. Raftery, T. Gneiting, F. Balabdaoui, M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133, 1155–1174, 2005.
14. K. Bogner, K. Liechti, M.Zappa. Technical note: Combining quantile forecasts and predictive distributions of streamflows. *Hydrol. Earth Syst. Sci.*, 21, 5493–5502, 2017.
15. J. Thorey, V. Mallet and P. Baudin. Online learning with the Continuous Ranked Probability Score for ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society*, 143, 521–529, January 2017 A DOI:10.1002/qj.2940
16. V. Vovk. Competitive on-line statistics. *International Statistical Review* 69, 213–248, 2001.
17. D. Adamskiy, T. Bellotti, R. Dzhamtyrova, Y. Kalnishkan. Aggregating Algorithm for Prediction of Paks. arXiv:1710.08114 [cs.LG], 2017.
18. J.E. Matheson, R.L. Winkler. Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10), 1087–1096, 1976. doi:10.1287/mnsc.22.10.1087.
19. M. Herbster, M. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998.

ПРИЛОЖЕНИЕ

А. ФУНКЦИЯ ПОДСТАНОВКИ

Для произвольной функции потерь $\lambda(\gamma, \omega)$, где $\gamma \in [0, 1]$ и $\omega \in \{0, 1\}$, рассмотрим параметрическую кривую на плоскости

$$\left(e^{-\eta\lambda(\gamma,0)}, e^{-\eta\lambda(\gamma,1)} \right). \quad (\text{A.1})$$

Нас интересует случай, когда эта кривая является вогнутой. Условие вогнутости имеет вид

$$x'(\gamma)y''(\gamma) - x''(\gamma)y'(\gamma) \geq 0 \quad (\text{A.2})$$

для всех γ , где $x(\gamma) = e^{-\eta\lambda(\gamma,0)}$ и $y(\gamma) = e^{-\eta\lambda(\gamma,1)}$.

В частности, для квадратичной функции потерь $\lambda(\gamma, \omega) = (\gamma - \omega)^2$ будет $x(\gamma) = e^{-\eta\gamma^2}$ и $y(\gamma) = e^{-\eta(\gamma-1)^2}$. После элементарных преобразований неравенство (A.2) эквивалентно неравенству

$$\eta\gamma(1 - \gamma) \leq \frac{1}{2}.$$

Легко видеть, что так как $\gamma(1 - \gamma)$ принимает максимальное значение $\frac{1}{4}$ при $0 \leq \gamma \leq 1$, условие вогнутости выполнено для всех γ при $0 < \eta \leq 2$.

Условие η -смешиваемости означает, что для любого распределения $\mathbf{w} = (w_1, \dots, w_N)$ на множестве N экспертов и любых их прогнозов $\mathbf{f} = (f_1, \dots, f_N)$ найдется γ^* такое, что выполнены два неравенства

$$e^{-\eta(\lambda(\gamma^*, \omega))} \geq \sum_{i=1}^N w_i e^{-\eta\lambda(f_i, \omega)} \quad (\text{A.3})$$

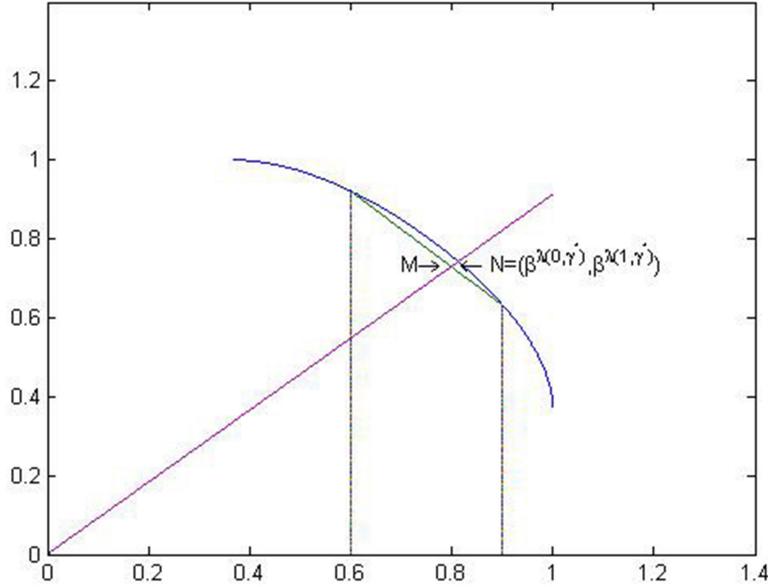


Рис. 1. Прямая, проходящая через точку M , отмечает точку $(\beta^{\lambda(0, \gamma^*)}, \beta^{\lambda(1, \gamma^*)})$ на кривой, где $\beta = e^{-\eta}$, по которой вычисляется предсказание $\gamma^* = \text{Subst}(\mathbf{f}, \mathbf{w})$.

при $\omega = 0, 1$. Точки $(e^{-\eta\lambda(f_i, 0)}, e^{-\eta\lambda(f_i, 1)})$, $i = 1, \dots, N$ лежат на кривой (A.1), а их выпуклая комбинация – точка M , лежит внутри выпуклой области, ограниченной этой кривой. Условие (A.3) означает, что абсцисса и ордината точки $N = (e^{-\eta\lambda(\gamma^*, 0)}, e^{-\eta\lambda(\gamma^*, 1)})$ не меньше чем абсцисса и ордината точки M . Найдем такую точку N . Прямая, проходящая через точку M , отмечает точку $N = (e^{-\eta\lambda(\gamma^*, 0)}, e^{-\eta\lambda(\gamma^*, 1)})$ на кривой (A.1) (см. рис. 1). Предсказание γ^* вычисляется из условия

$$\frac{e^{-\eta\lambda(\gamma^*, 1)}}{e^{-\eta\lambda(\gamma^*, 0)}} = \frac{\sum_{i=1}^N w_i e^{-\eta\lambda(f_i, 1)}}{\sum_{i=1}^N w_i e^{-\eta\lambda(f_i, 0)}}. \tag{A.4}$$

В частности, для квадратичной функции потерь из равенства (A.4) получаем выражение для γ^* :

$$\gamma^* = \text{Subst}(\mathbf{f}, \mathbf{w}) = \frac{1}{2} - \frac{1}{2\eta} \ln \frac{\sum_{i=1}^N w_i e^{-\eta f_i^2}}{\sum_{i=1}^N w_i e^{-\eta (f_i - 1)^2}}.$$

Условие смешиваемости квадратичной функции потерь позволяет взять $\eta = 2$.

Нетрудно проверить, что для любого $\omega \in [0, 1]$ функция $f(\gamma) = e^{-\eta(\gamma - \omega)^2}$ является вогнутой по $\gamma \in [0, 1]$ при $0 < \eta < \frac{1}{2}$. В этом случае величина $\gamma^* = \sum_{i=1}^N w_i f_i$ удовлетворяет неравенству (A.3) при любом $0 < \eta < \frac{1}{2}$ по определению вогнутости.

А. ГРАФИКИ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ

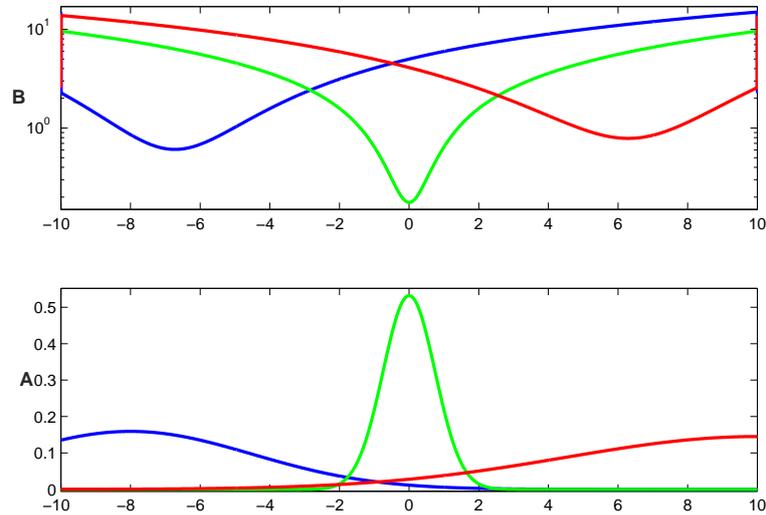


Рис. 2. Генерирующие модели с усеченным нормальным распределением. Плотности распределения трех генерирующих моделей (А) и графики CRPS для них в логарифмической шкале (В).

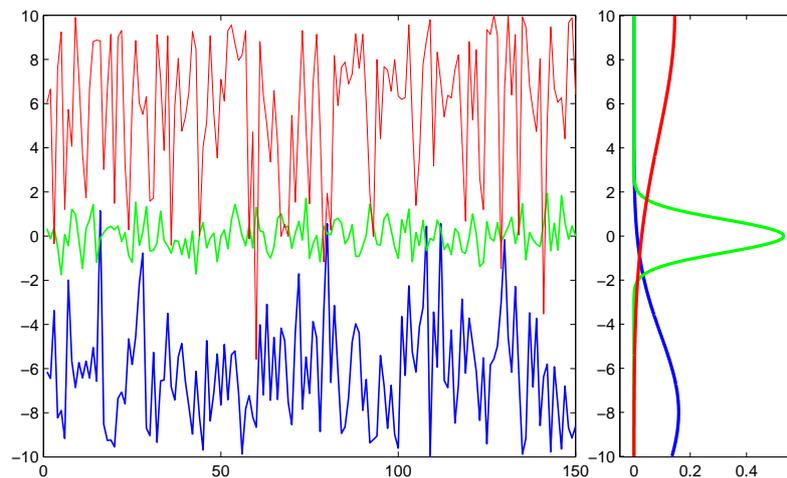


Рис. 3. Исходные данные: слева приведены реализации случайных траекторий трех генерирующих моделей; в правой части рисунка отображены соответствующие им плотности распределения.

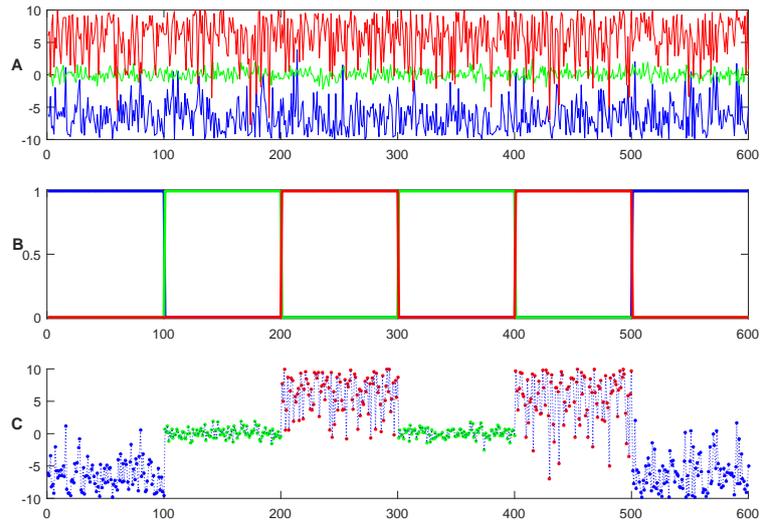


Рис. 4. Эксперимент 1. Основные этапы формирования модели генерации данных. В разделе **A** приведены реализации случайных траекторий генерирующих моделей. В разделе **B** для тех же временных интервалов отображены априорные вероятности (фактические относительные веса), которые были использованы при агрегировании распределений вероятностей. В разделе **C** показаны результаты смешивания реализаций трех моделей из раздела **A** с априорными вероятностями из раздела **B**.

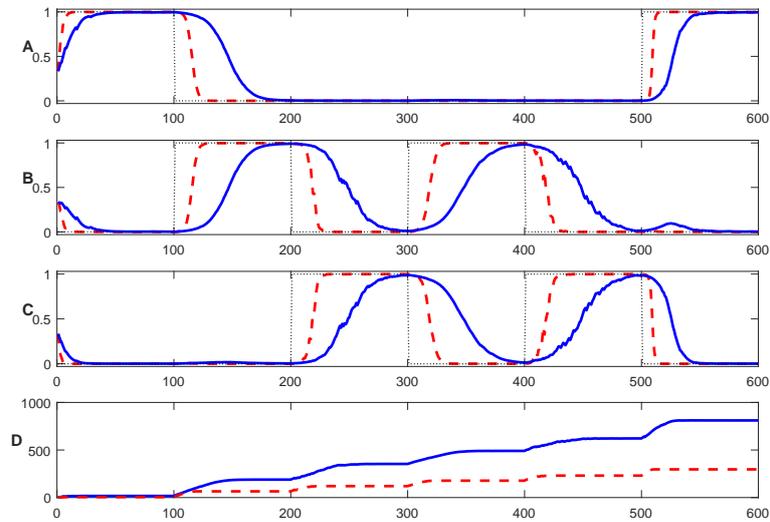


Рис. 5. Эксперимент 1. Результаты онлайн-адаптации весов экспертов. На каждом из трех разделов **A**, **B** и **C**, соответствующих генерирующим моделям (экспертам), показаны графики изменения адаптируемых весов по методу (14) (штриховая линия) и по методу (16) (сплошная линия) в сравнении с истинным весом (пунктирная линия). В разделе **D** приведены траектории кумулятивных потерь агрегирующего алгоритма для методов смешивания (14) (штриховая линия) и (16) (сплошная линия) за вычетом кумулятивных потерь истинной модели.

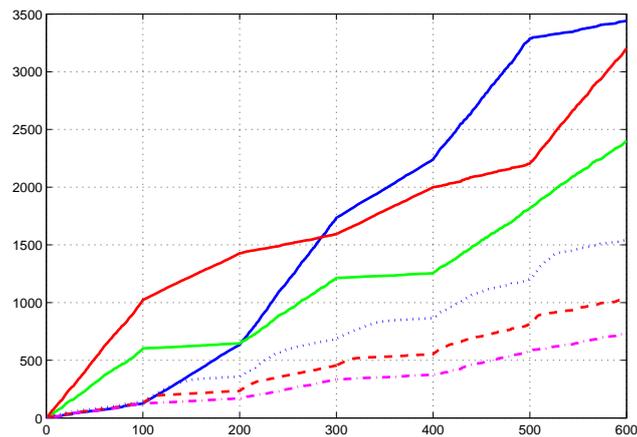


Рис. 6. Эксперимент 1. Кумулятивные потери экспертов (сплошные линии) и методов агрегирования (14) (штриховая линия) и (16) (пунктирная линия) в сравнении с суммарными потерями истинной (генерирующей) модели (штрих-пунктирная линия). Видно преимущество метода (14) перед методом (16).

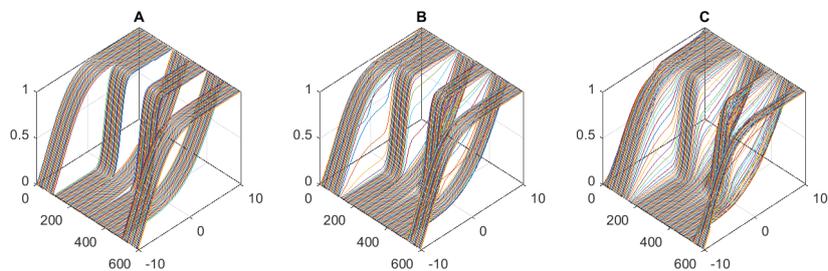


Рис. 7. Эксперимент 1. 3D отображение временной динамики функций распределения истинной модели **A**, а также результаты агрегирования функций распределения (14) **B** и (16) **C**.

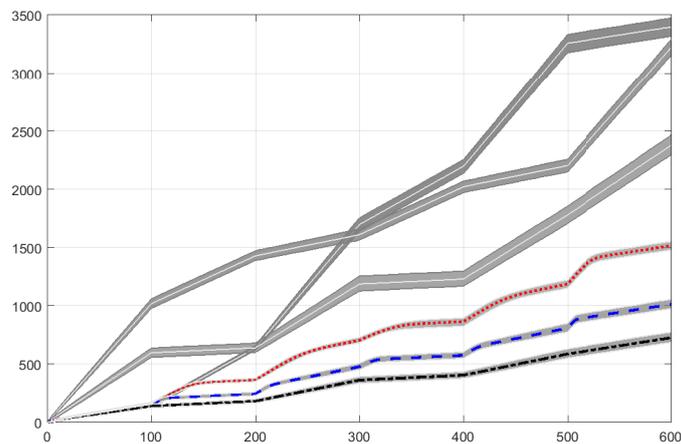


Рис. 8. Эксперимент 1. Результаты статистического анализа 1000 последовательных реализаций эксперимента. Полутонами отображены 0.05-0.95 квантильные полосы кумулятивных потерь. Сплошными линиями показаны медианы для кумулятивных потерь экспертов. Штриховая линия – медиана траекторий кумулятивных потерь агрегирующего алгоритма (14), пунктирная линия – медиана для метода агрегирования (16), и штрих-пунктирная линия – медиана кумулятивных потерь истинной модели.

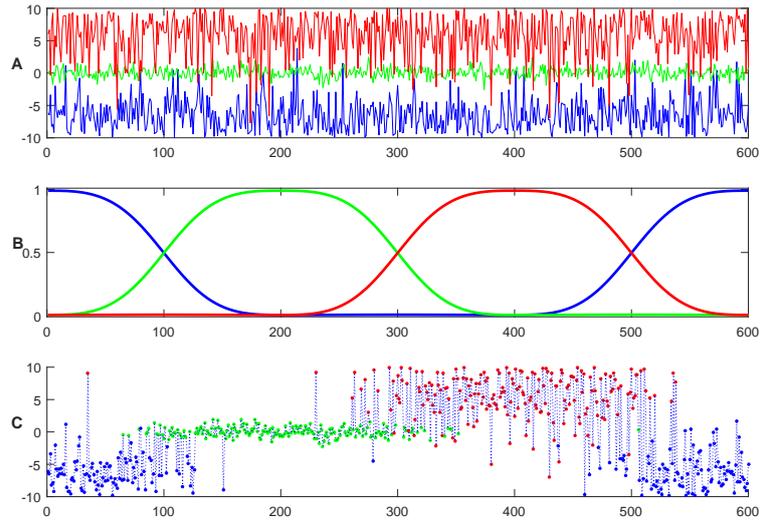


Рис. 9. Эксперимент 2. Основные этапы формирования численной модели. В разделе **A** приведены реализации случайных траекторий генерирующих моделей. В разделе **B** для тех же временных интервалов отображены априорные вероятности (фактические относительные веса), которые были использованы при смешивании распределений вероятностей. В разделе **C** показаны результаты смешивания реализаций трех моделей из раздела **A** с априорными вероятностями из раздела **B**.

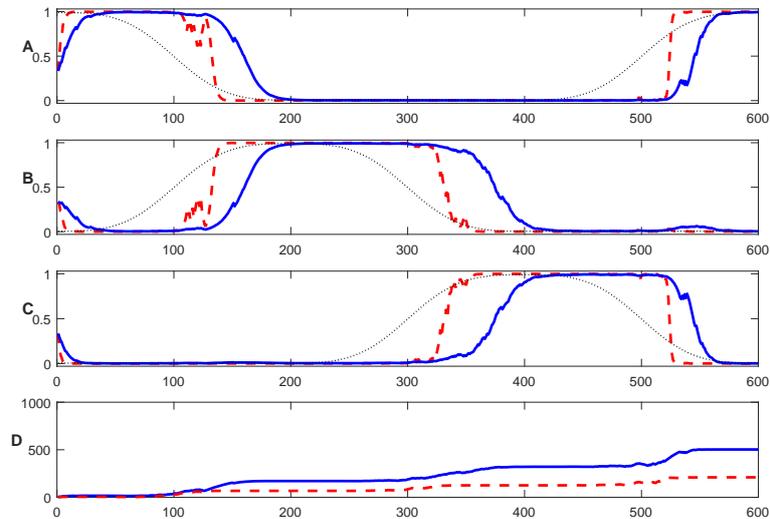


Рис. 10. Эксперимент 2. Результаты онлайн-адаптации весов экспертов в ситуации, когда веса компонент смеси плавно изменяются во времени, причем в каждый момент времени активны только две из трех генерирующих моделей. На каждом из трех разделов **A**, **B** и **C**, соответствующих генерирующим моделям (экспертам), показаны графики изменения адаптируемых весов по методу (14) (штриховая линия) и по методу (16) (сплошная линия) в сравнении с истинным весом (пунктирная линия). В разделе **D** приведены траектории кумулятивных потерь агрегирующего алгоритма для методов смешивания (14) (штриховая линия) и (16) (сплошная линия) за вычетом кумулятивных потерь истинной модели.

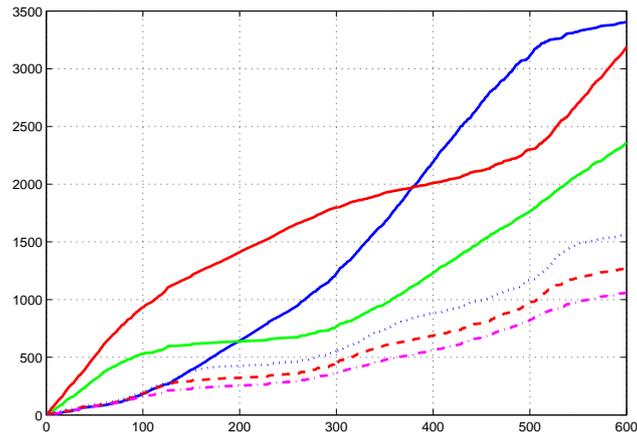


Рис. 11. Эксперимент 2. Кумулятивные потери экспертов (сплошные линии) и методов агрегирования (14) (штриховая линия) и (16) (пунктирная линия) в сравнении с суммарными потерями истинной (генерирующей) модели (штрих-пунктирная линия). Видно некоторое преимущество метода (14) перед методом (16).

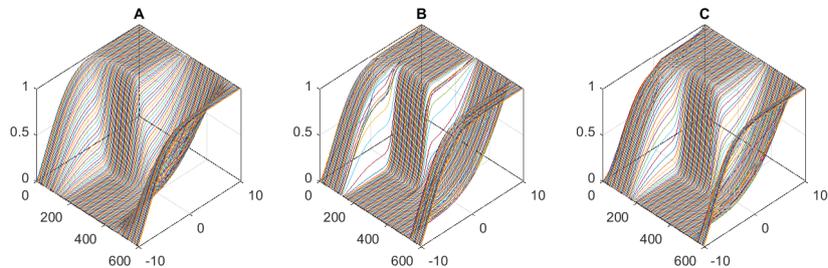


Рис. 12. Эксперимент 2. 3D отображение временной динамики функций распределения истинной модели *A*, а также результаты агрегирования функций распределения (14) *B* и (16) *C*.

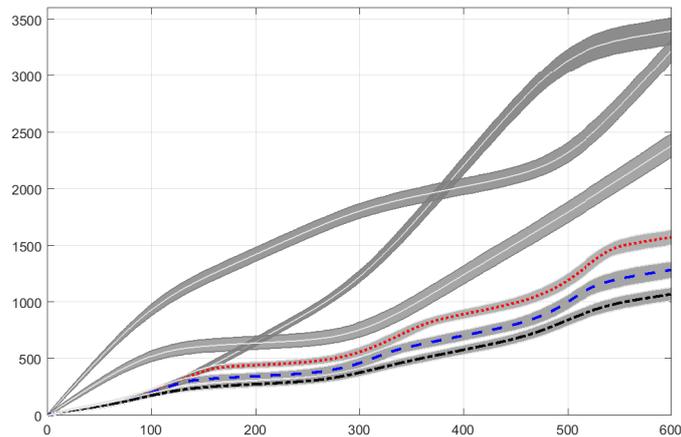


Рис. 13. Эксперимент 2. Результаты статистического анализа 1000 последовательных реализаций эксперимента. Полутонами отображены 0.05-0.95 квантильные полосы кумулятивных потерь. Сплошными линиями показаны медианы для кумулятивных потерь экспертов. Штриховая линия – медиана траекторий кумулятивных потерь агрегирующего алгоритма (14), пунктирная линия – медиана для метода агрегирования (16), и штрих-пунктирная линия – медиана кумулятивных потерь истинной модели.

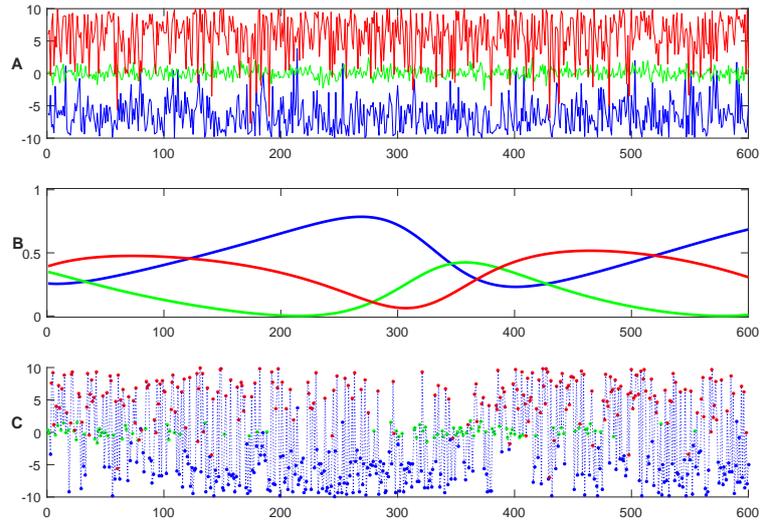


Рис. 14. Эксперимент 3. Основные этапы формирования численной модели. В разделе **A** приведены реализации случайных траекторий генерирующих моделей. В разделе **B** для тех же временных интервалов отображены априорные вероятности (фактически относительные веса), которые были использованы при агрегировании распределений вероятностей. В разделе **C** показаны результаты смешивания реализаций трех моделей из раздела **A** с априорными вероятностями из раздела **B**.

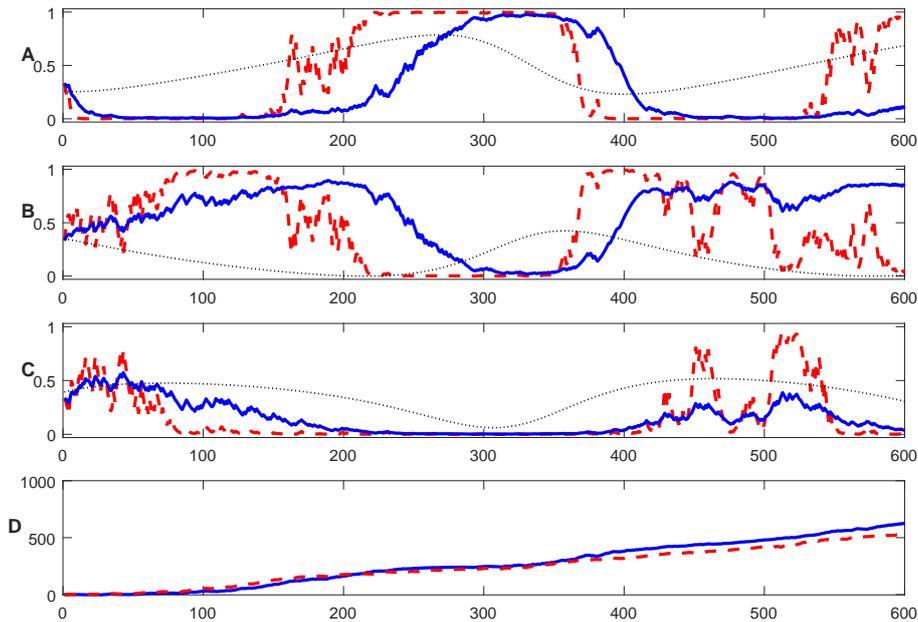


Рис. 15. Эксперимент 3. Результаты онлайн-адаптации весов экспертов. В каждом из трех разделов **A, B** и **C** соответствующих генерирующим моделям (экспертам), показаны графики изменения адаптируемых весов по методу (14) (штриховая линия) и по методу (16) (сплошная линия) в сравнении с истинным весом (пунктирная линия). В разделе **D** приведены траектории кумулятивных потерь агрегирующего алгоритма для методов смешивания (14) (штриховая линия) и (16) (сплошная линия) за вычетом кумулятивных потерь истинной модели.

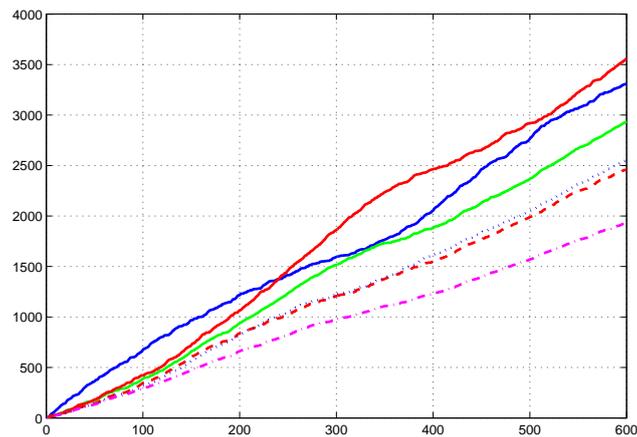


Рис. 16. Эксперимент 3. Кумулятивные потери экспертов (сплошные линии) и методов агрегирования (14) (штриховая линия) и (16) (пунктирная линия) в сравнении с суммарными потерями истинной (генерирующей) модели (штрих-пунктирная линия). Потери методов агрегирования (14) и (16) ведут себя сходным образом.

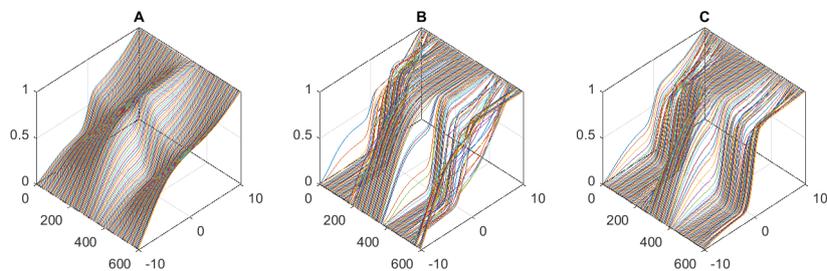


Рис. 17. Эксперимент 3. 3D отображение временной динамики функций распределения истинной модели **A**, а также результаты агрегирования функций распределения (14) **B** и (16) **C**.

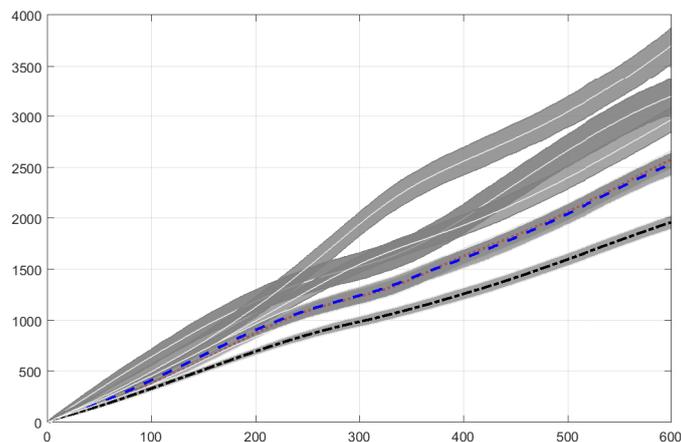


Рис. 18. Эксперимент 3. Результаты статистического анализа 1000 последовательных реализаций эксперимента. Полутонами отображены 0.05-0.95 квантильные полосы кумулятивных потерь. Сплошными линиями показаны медианы для кумулятивных потерь экспертов. Штриховая линия – медиана траекторий кумулятивных потерь агрегирующего алгоритма (14), пунктирная линия – медиана для метода агрегирования (16), и штрих-пунктирная линия – медиана кумулятивных потерь истинной модели.

Online Aggregation of Probabilistic Forecasts Based on the Continuous Ranked Probability Score

V.V. V'yugin, V.G. Trunov

Methods for generating predictions online and in the form of probability distributions of future outcomes are considered. The difference between the probabilistic forecast (probability distribution) and the numerical outcome is measured using the loss function (scoring rule). In practical statistics, the Continuous Ranked Probability Score (CRPS) is often used to estimate the discrepancy between probabilistic forecasts and (quantitative) outcomes. The paper considers the case when several competing methods (experts) give their online predictions as distribution functions. An algorithm is proposed for aggregating these distribution functions online. The performance bounds of the proposed algorithm are obtained in the form of a comparison of the cumulative loss of the algorithm and the loss of expert hypotheses. Unlike previously known estimates, the estimates obtained do not depend on time. The results of numerical experiments illustrating the proposed methods are presented.

KEYWORDS: decision theoretic online learning, prediction with expert advice, regret, aggregating algorithm, continuous ranked probability score – CRPS.