DOI <https://doi.org/10.18551/rjoas.2017-08.11>

ЭФФЕКТИВНЫЙ АЛГОРИТМ МУЛЬТИПАРАМЕТРИЧЕСКОГО ФУНКЦИОНАЛЬНО-СТАТИСТИЧЕСКОГО АНАЛИЗА И ЕГО КОМПЬЮТЕРНАЯ ВЕРСИЯ
EFFECTIVE ALGORITHM OF MULTIPARAMETRIC FUNCTIONAL-STATISTICAL ANALYSIS AND ITS COMPUTER VERSION

Прилуков А.Н., кандидат социологических наук
 Prilukov A.N., Candidate of Sociological Sciences

Институт горного дела Дальневосточного отделения РАН, Хабаровск, Россия
 Mining Institute of Russian Academy of Sciences, Khabarovsk, Russia
 E-mail: a_prilukov@mail.ru

АННОТАЦИЯ

Рассматриваются два метода математической статистики, разработанные более тридцати лет назад: *расширенная параметрическая регрессия* (РПР) и *анализ функционально-статистической динамики* (АФСД). Несмотря на принадлежность к одному крупному разделу математики, данные методы принципиально различаются характером решаемых с их помощью задач. РПР, первоначально обозначавшаяся как *толерантный регрессионный анализ*, за счёт расширенного состава её задействуемых параметров позволяет объединить и оптимизировать традиционные и вновь появляющиеся многочисленные разновидности конкретных моделей, неоправданно часто выдаваемых за самостоятельные методы и использующихся при восстановлении осложнённых стохастичностью зависимостей. АФСД, также не вполне удачно наречённый вначале *анализом коллинеарности*, предназначен для сравнения между собой как функциональных, так и статистических зависимостей по их параметрам, являющимся ключевыми для такого сопоставления, – нормированным амплитудам, соотношению параметризованных масштабов и взаимного сдвига. Помимо принадлежности к одному разделу математики рассматриваемые методы объединяют во многом совпадающие составы использующихся в них расчётных параметров, включая варьируемую метрику Минковского, позволяющую унифицировать и оптимизировать механизмы агрегирования стохастических ошибок (остатков). В статье демонстрируется, что близкое совпадение составов параметров ведёт к схожести алгоритмов решения задач обоих видов анализа, что в свою очередь даёт возможность обойтись одной компьютерной программой, позволяющей без каких-либо ограничений реализовать особенности каждого из методов. Этим же обстоятельством оправдывается объединение двух рассматриваемых методов в заголовке статьи под общим названием *мультипараметрического функционально-статистического анализа* (МФСА). Кроме сокращённого описания алгоритма и его программной реализации в статье приведены практические примеры, иллюстрирующие возможности и особенности рассматриваемых методов, а также намечающиеся направления их совершенствования.

ABSTRACT

In the article two mathematical statistics methods are discussed that have been developed more than three decades ago: the Augmented Parametric Regression (APR) and the Analysis of Functional-Statistical Dynamics (AFSD). Despite affiliation to the same major branch of mathematical statistics these methods profoundly diverge with respect to the tasks being solved by them. APR, initially labeled as Tolerant Regression Analysis, due to its augmented set of parameters enables it to consolidate and optimize traditional and numerous currently emerging regression models, which are used for stochastically contaminated dependences recovery and often are unjustifiably being rendered as self-contained methods. AFSD, also not quite aptly named at first as the Collinearity Analysis, is earmarked for functional and/or statistical dependences' mutual comparison according to their parameters, which play key roles in such juxtaposition, – their normalized amplitudes,

parameterized scaling factors, and mutual shifts (lags). Alongside with belonging to the same mathematical branch both discussed methods share closely coinciding calculating parameters, including Minkowski metric that makes it possible to unify and optimize the stochastic residuals aggregation mechanisms. Parameters' sets close coincidence leads to substantial similarity of algorithms that are used in tasks solving within both analyses types. In its turn this procures possibility to dispense with a single computer program that without any limitations realize both methods peculiarities. The same circumstance justifies both methods embracing in the article title under generalized heading of Multiparametric Functional-Statistical Analysis (MFSA). In addition to algorithm and its computer program realization shortcut descriptions the article contains examples that illustrate performance capabilities and particularities of discussed methods, as well as their outlined directions of further development.

КЛЮЧЕВЫЕ СЛОВА

Расширенная параметрическая регрессия, анализ функционально-статистической динамики, мультипараметрический функционально-статистический анализ, сеточно-градиентная оптимизация, алгоритм сеточной конвергенции.

KEY WORDS

Augmented parametric regression, analysis of functional-statistic dynamics, multiparametric functional-statistical analysis, gridding-gradient optimization, grid-converging algorithm.

Согласно одному из популярных обобщённых описаний, регрессионный анализ представляется выражением:

$$y_i = f(X_i; \Theta) + \varepsilon_i, \quad (1)$$

где y_i – i -е значение отклика (объясняемой переменной); f – функция, определяющая собой конкретный вид регрессионной модели; $X_i = x_{1i}, x_{2i}, \dots, x_{ki}$ – набор i -х значений векторов аргументов (объясняющих переменных или факторов); ε_i – i -е значение случайной компоненты уравнения (регрессионной ошибки); $i = 1, 2, \dots, n$ – номер наблюдения; $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ – параметры, в качестве которых обычно выступают расчётные коэффициенты регрессионной модели.

Демонстрируемый в (1) способ введения в уравнение регрессионных ошибок ε_i является аддитивным. При соблюдении определённых ограничений на характер совокупности $\{\varepsilon_i\}$ – их несмещённости, взаимной некоррелированности, независимости от величин X_i и y_i , распределённости по нормальному статистическому закону – этот способ обеспечивает возможность использования в ходе регрессионного анализа популярных методов наименьших модулей [1–2] и наименьших квадратов – см., например, публикации в журнале RJOAS авторов Acquah De-Graft H., Acquah S., Akpan S.B., Jafrizal, Shinta A., Suharno P., Wongnaa C.A. и других. Особую привлекательность методу наименьших квадратов, предложенному в 1805 г. Лежандром [3], а также появившемуся позднее методу наименьших модулей придаёт то обстоятельство, что в простых случаях они позволяют обходиться без громоздких вычислений, требующих обращения к компьютерам.

Следующей по популярности, пожалуй, является модель, строящаяся на предположении о мультипликативном воздействии регрессионных ошибок [4, с.74]. Её можно представить в виде:

$$y_i = f(X_i; \Theta) \times (1 + \varepsilon_i) = f(X_i; \Theta) \times \gamma_i. \quad (2)$$

Здесь γ_i – измеряемая относительно единицы мультипликативная регрессионная ошибка; ε_i – обычная, измеряемая относительно нуля, регрессионная ошибка аддитивного типа.

В различных разделах математической статистики находят применение и другие способы учёта влияния, оказываемого случайными величинами, подобными регрессионным ошибкам. В последнее время появляются также примеры обобщения разных способов учёта случайных отклонений, основывающиеся на представлениях о средних степенных значениях, называемых обобщёнными средними, средними по Гёльдеру или Минковскому [5].

Обобщённая средняя по Гёльдеру, например, описывается выражением [6]:

$$\bar{x}(m) = \left(\frac{1}{n} \sum_{i=1}^n x_i^m \right)^{1/m}. \quad (3)$$

Здесь $m \in (-\infty; \infty)$ – любое действительное число; x_i , $i = 1, 2, \dots, n$ – действительные числа из диапазона $x_i \in [0; \infty)$.

Стоит отметить, что некоторым дискретным значениям m в выражении (3) соответствуют разновидности средних величин, имеющие собственные имена:

$m_{\infty} \rightarrow \infty$ – $\max\{x_i\}$ – “среднее” в метрике доминирования;

$m_2 = 2$ – среднеквадратичное $\{x_i\}$;

$m_1 = 1$ – среднеарифметическое $\{x_i\}$;

$m_0 = 0$ – среднегеометрическое $\{x_i\}$;

$m_{-1} = -1$ – среднегармоническое $\{x_i\}$;

$m_{-\infty} \rightarrow -\infty$ – $\min\{x_i\}$.

Особой значимостью обладает то обстоятельство, что во всём числовом диапазоне $m \in (-\infty; \infty)$ средние значения (3) отвечают требованиям, предъявляемым к подобного рода мерам [7]: а) непрерывности; б) монотонности по каждому x_i ; в) симметричности, выражающейся в неизменности среднего при перестановке аргументов; г) равенстве средней от набора равных чисел их общему значению; д) неизменности общего среднего значения при разбивке набора $\{x_i\}$ на подгруппы и использовании вместо них средних по подгруппам.

Кроме того, из известного неравенства о средних, гласящего, что при любых $m_b > m_a$ имеет место:

$$\bar{x}(m_b) \geq \bar{x}(m_a), \quad (4)$$

для любых наборов $\{x_i\}$ выводятся соотношения [5]:

$$m_{\infty} = \max\{x_i\} \geq m_2\{x_i\} \geq m_1\{x_i\} \geq m_0\{x_i\} \geq m_{-1}\{x_i\} \geq m_{-\infty}\{x_i\} = \min\{x_i\} \quad (5)$$

Перечисленные свойства обобщённых средних величин свидетельствуют о том, что все они являются легитимными инструментами (мерами) при оценке качества аппроксимации и сообразно условиям конкретной задачи могут использоваться в регрессионном анализе и других разделах математической статистики. Методической трудностью, возникающей при этом, является обоснование выбора конкретной разновидности средней величины, адекватной решаемой задаче.

Преобладающий в настоящее время выбор среднеквадратичных величин объясняется рядом очевидных обстоятельств: 1) основательной разработанностью старейшего по времени создания метода наименьших квадратов, на основе которого вычисляются популярные статистические оценки, проверяются и обосновываются рабочие гипотезы исследования; 2) хорошей согласованностью среднеквадратичных величин с предполагаемым во многих случаях, хотя зачастую не вполне обоснованно, гауссовским (нормальным) распределением случайных величин, присутствующих в

вычислениях; 3) относительной лёгкостью вычислений, которые в простых случаях удаётся выполнять алгебраическими методами, в том числе без обращения к компьютерам.

В предложенных и обосновываемых автором методах математической статистики, реализованных с помощью описываемого в статье компьютерного алгоритма, используется обобщённый вид функционала (1), в котором изменяющиеся в ходе вычислений случайные величины ε_i включены в качестве аргументов, определяющих наряду с другими аргументами итоговый вид функции f :

$$y_i = f(X_i; \varepsilon_i; \Theta) \quad (6)$$

Расположением случайных величин ε_i в формуле (6) подчёркивается их двоякая природа. С одной стороны, как изначально независимые от исследователя величины ε_i родственны переменным $X_i = x_{1i}, x_{2i}, \dots, x_{ki}$. С другой стороны, с параметрами Θ ошибки $E = \{\varepsilon_i\}$ связывает способ, выбираемый исследователем для их влияющей на результаты вычислений обобщённой оценки.

С учётом последнего обстоятельства, а также того, что фигурирующие в уравнении (6) ошибки ε_i приобретают свои конкретные значения только в ходе вычислений, вид данного уравнения может быть подвергнут дальнейшему упрощению за счёт удаления из него компоненты ε_i и её включения в “сублимированный” виде в состав параметров Θ :

$$y_i = f(X_i; \Theta). \quad (7)$$

Здесь $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$ – набор параметров, отличающийся от общепринятого своим расширенным составом. В него фактически входят все переменные и постоянные величины, определяющие собой и так или иначе влияющие на ход и результаты вычислений. В расширенном составе параметров присутствуют: указание на тип применяемой функции f либо её формализованное описание, пригодное для автоматической расшифровки и программной реализации; начальные значения, диапазоны и шаги изменения функциональных коэффициентов и аргументов; условия выполнения избранных процедур и критерии переходов между ними; способы выдачи результатов и взаимодействия с исследователем и т.д.

На значения параметров Θ не накладывается априорных ограничений за исключением тех, которые диктуются их функциональным назначением. В частности, любой из параметров θ_i может быть представлен абсолютной либо относительной числовой величиной, ранговым коэффициентом, категориальной (номинальной) переменной, функциональным либо логическим оператором, простым либо составным критерием [8].

Одна из ключевых ролей в составе параметров Θ отводится варьируемой метрике Минковского $r \in (-\infty; \infty)$, определяющей собой способ усреднённого агрегирования ошибок, осуществляющегося по формуле:

$$D_r = \left(\frac{1}{n} \sum_{i=1}^n |\varepsilon_i|^r \right)^{1/r}. \quad (8)$$

D_r в выражении (8) представляет собой обобщённую среднюю ошибок (регрессионных остатков), подсчитанную при значении метрики Минковского равном r . От обобщённой средней по Гёльдеру (3) данное выражение отличается лишь переводом всех ε_i в область положительных значений $|\varepsilon_i| \in [0; \infty)$. Наряду с сохранением у D_r свойств обобщённого среднего такая операция существенно

упрощает вычисления за счёт исключения из них действий с комплексными величинами, появляющимися при оперировании дробными степенями отрицательных аргументов.

При принятых условиях функционал, описывающий решение задачи регрессионного анализа, может быть компактно изображён в виде:

$$D(Y; X; \Theta) = \left[\sum_{i=1}^n w_i |(y_i - \hat{y}_i) / z_i|^r \middle/ \sum_{i=1}^n w_i \right]^{1/r} \Rightarrow \min. \quad (9)$$

Здесь $Y=\{y_i\}$ – набор заданных дискретных значений регрессанда; $X = \{X_i\}$ – набор векторов значений независимых переменных; $\Theta=\{\theta_j\}$ – набор параметров, определяющих расчёты; $W=\{w_i\}$ – набор коэффициентов, используемых для взвешивания и цензурирования выборочных совокупностей. z_i в данном выражении представляют собой расчётные величины, с помощью которых производится нормирование регрессионных остатков, устраняющее по возможности их корреляцию с величинами y_i и $\hat{y}_i$.

Анализ функционально-статистической динамики (АФСД). Целью сопоставительного анализа динамики функциональных зависимостей, в том числе представленных статистическими выборками, является отыскание ключевых параметров, в концентрированном виде отражающих расхождение либо, наоборот, сходство между зависимостями – соотношения их амплитуд и скоростей протекания, а также опережения (отставания) одной зависимости относительно другой.

В реализованном варианте АФСД в качестве ключевых расчётных параметров выбраны смещение (лаг) между условными началами процессов, описываемых зависимостями, и разницы в скоростях протекания этих процессов. Соотношение амплитуд сравниваемых зависимостей отсутствовало в составе расчётных параметров, поскольку во всех выполненных расчётах сопоставлялись однотипные зависимости, изначально нормированные по амплитудам. В то же время возможность программного нормирования либо выравнивания амплитуд была предусмотрена за счёт выбора весовых коэффициентов, применяемых по отношению к ординатам сравниваемых зависимостей.

Для количественной оценки степени расхождения между зависимостями в АФСД, также как и в расширенной параметрической регрессии, используются расстояния в метрике Минковского. В исходном виде, абстрагирующемся от операций взаимного смещения и масштабирования, расстояние между анализируемыми непрерывными зависимостями описывается выражением [9, с.107]:

$$D_r(x, y) = \left[\int_v w(t) |x(t) - y(t)|^r dt \right]^{1/r}. \quad (10)$$

Здесь $x(t)$, $y(t)$ – сравниваемые зависимости; t – их общий аргумент; $w(t)$ – весовая функция; v – область определения функций x , y , w ; r – показатель (метрика) Минковского.

При дискретном задании функций x , y , w выражение (10) в статичном исходном состоянии и при условии совпадения абсцисс точек определения обеих зависимостей принимает вид:

$$D_r(x, y) = \left[\sum_{i=1}^n w_i |x_i - y_i|^r \right]^{1/r}. \quad (11)$$

где x_i, y_i, w_i – заданные дискретные значения функций x, y, w ; $i=1, 2, \dots, n$ – номер очередной дискретной точки; n – общее количество точек, для которых известны конкретные значения функций x, y, w .

Выбранные параметры, обозначаемые как s (смещение) и m (масштабирование), соотносятся с осью абсцисс, являющейся общей для сравниваемых зависимостей. В выражении (10) ось абсцисс присутствует явно в виде аргумента t , в (11) она играет роль виртуального континуума $t = (-\infty, \infty)$, вмещающего в себя наборы дискретных точек $(\{t_i\} \cup \{t_j\}) \subset t$.

Принципиально важно, что ось абсцисс в данном случае не обязана символизировать собой физическую, календарную, историческую или какую-либо другую разновидность времени. Более того, ассоциируемые с аргументом t переменные величины могут быть разными по своей субстанциальной природе, но при этом они должны играть ключевые роли в динамике сопоставляемых зависимостей и воздействовать на них в одном направлении.

Выражения (10), (11) являются заведомо упрощёнными, поскольку в них игнорируется разница в областях определения функций $x(t), y(t)$ либо соответствующих дискретных значений $\{x_i\}$ и $\{y_j\}$. В алгоритмах и компьютерных программах это преднамеренное упрощение устраняется за счёт внесения соответствующих корректировок в формулы, в том числе введением в них операций экстраполяции и интерполяции.

Для обеспечения несмещённости, повышения точности и унификации результатов расчётов функционалы (10) или (11) целесообразно пронормировать, как это иллюстрируется выражением (9). В результате этого, а также введения необходимых операций над переменными величинами выражение (11), например, принимает вид:

$$D(X; Y; \Theta) = \left\{ \frac{\sum_{i=1}^n p_i \hat{q}_i |(x_i - \hat{y}_i)/z_i|^r + \sum_{j=1}^n \hat{p}_j q_j |(\hat{x}_j - y_j)/z_j|^r}{\sum_{i=1}^n p_i \hat{q}_i + \sum_{j=1}^n \hat{p}_j q_j} \right\}^{1/r} \Rightarrow \min. \quad (12)$$

Усложнение функционала (12) по сравнению с предыдущим объясняется появлением семейства новых дискретных точек $J=\{j\}$ на оси абсцисс, к которым привязывается “ведомая” зависимость в результате её вариативного смещения и масштабирования относительно “ведущей”. А поскольку в результате названных двух операций семейства точек $I=\{i\}$ и $J=\{j\}$ оказываются взаимно смещёнными, требуется рассчитывать новые значения функций x и y на обновлённых участках их интерполяции и экстраполяции. Аналогичным метаморфозам подвергаются весовые коэффициенты w , которые к тому же нуждаются в разбивке на два подмножества $P = \{p_i\} \cup \{\hat{p}_i\}$ и $Q = \{q_j\} \cup \{\hat{q}_j\}$, относящиеся соответственно к функциям x и y .

В том важном и, следует думать, нередком случае, когда в качестве ведущей либо эталонной зависимости выбирается некая формализованная непрерывная функция $x(t)$, за счёт исключения необходимости применения операций интерполяции и экстраполяции по отношению к обеим функциям функционал (12) существенно упрощается:

$$D(X; Y; \Theta) = \left\{ \frac{\sum_{j=1}^n p(\hat{t}_j) q_j |[x(\hat{t}_j) - y_j]/z_j|^r}{\sum_{j=1}^n p(\hat{t}_j) q_j} \right\}^{1/r} \Rightarrow \min. \quad (13)$$

Здесь $\hat{t}_j$ – дискретное значение аргумента функции $x(t)$, рассчитанное с учётом сдвига абсциссы j -го значения функции y в результате применения к ней операций смещения и масштабирования.

Обращает на себя внимание значительная схожесть функционалов (9) и (13), различающихся между собой лишь характером фигурирующих в них переменных величин. Неявное различие состоит также в том, что в состав Θ в (13) входят варьируемые параметры смещения s и масштабирования m , отсутствующие в (9). Однако если придать этим параметрам фиксированные начальные значения $s_0=0$ и $m_0=1$, функционал (13) становится фактически тождественным функционалу (9) и, следовательно, вполне пригодным для выполнения расчётов методом расширенной параметрической регрессии. Тем самым находит очередное наглядное подтверждение родственная природа регрессионного и корреляционного анализов.

Применительно к тем нередко возникающим ситуациям, когда требуется выполнить обобщённое сопоставление не одной, а множества однотипных пар зависимостей X_l и Y_l , где $X_l=\{x_l\}$, $Y_l=\{y_l\}$, $l=1, 2, \dots, L$, L – количество пар сопоставляемых зависимостей, функционалы (12) и (13) должны быть откорректированы за счёт введения в состав их варьируемых параметров дополнительного элемента l . Естественно, это сопровождается добавлением цикла верхнего уровня в алгоритмы, реализующие функционалы (12) и (13), а также L -кратным увеличением продолжительности работы программы.

Следует однако учесть, что в описанном варианте расчётов обеспечивается отыскание параметров s_l и m_l для той пары зависимостей X_l и Y_l из их L -мерного набора, которые изначально наиболее близки друг к другу. Для отыскания усреднённых $\bar{s}$ и $\bar{m}$, характеризующих статистические наборы зависимостей X_l и Y_l в целом, требуется либо выполнить их предварительное усреднение, сопровождающееся последующим сопоставлением полученных $\bar{X}$ и $\bar{Y}$, либо произвести тем или иным способом усреднение множества значений $\{s_l, m_l\}$, полученных по отдельности на основе функционалов (12) и (13).

Алгоритм вычислений. Описываемая работа основывалась на интенсивном использовании нескольких поколений отечественных ЭВМ, начиная от Минск-22 и заканчивая ЕС-1052, входившей в серию Ряд 1 и обладавшей параметрами [10–11]: производительность – 700 тыс. операций/с; объём оперативной памяти 1÷8 МБ; ёмкость одного накопителя на магнитных дисках – 27 МБ; вместимость бобины накопителя на магнитной ленте ~100 МБ. При среднем времени доступа к пользовательской информации, хранившейся в основном на магнитных лентах, измеряемом десятками секунд, ЭВМ ЕС-1052 работала в многозадачном режиме, обслуживая одновременно до 15 клиентов аппаратно-программной системы коллективного доступа.

Для написания программ использовался универсальный язык программирования ПЛ/1 (Programming Language one), совмещающий в себе и существенно дополняющий возможности популярных специализированных языков ФОРТРАН, АЛГОЛ и КОБОЛ. Несмотря на чрезмерную усложнённость, которая, по мнению некоторых специалистов присуща ПЛ/1 [12, с.198], этот язык отличается компактностью, наглядностью и легко прослеживаемой логичностью своих программных листингов. Немаловажно также, что он позволял и, нужно думать, до сих пор позволяет получать эффективно работающие программы, в максимальной степени реализующие вычислительные возможности компьютеров.

Работа по созданию алгоритмов и программ, перемежавшаяся с выполнением многочисленных пробных и реальных расчётов, иллюстрируются в упрощённом виде рисунком 1.

Рисунок 1 – Укрупнённая структура алгоритма мультипараметрического функционально-статистического анализа, совмещённая со схемой его разработки и реализации

Использование в обоих описываемых методах функционалов (7) – (13), а также варьируемой метрики Минковского исключило из арсенала способов решения соответствующих задач многие известные методы. Широко используемые в линейном регрессионном анализе матричные методы, например, несовместимы с общим видом названных функционалов, в которых на функции f не накладываются необходимые в таких случаях ограничения. Градиентные методы поиска экстремумов

функций, нашедшие широкое распространение при решении различных прикладных задач, также в данных случаях оказываются малоприспособленными; прежде всего ввиду их невысокой надежности, а зачастую и непредсказуемости результатов, зависящих от выбора начальной точки и от конфигурации поверхности, отображающей целевую функцию пошагового поиска. Низкая надёжность градиентных способов объясняется в данном случае тем обстоятельством, что гиперповерхности, соответствующие в многомерных пространствах Θ минимизируемым функционалам D , нередко имеют сложный рельеф. Помимо мультимодальности, выражающейся в наличии многих локальных минимумов, в зависимости от применяемых способов интерполяции и экстраполяции на данных гиперповерхностях могут присутствовать резкие “изломы”, значительно затрудняющие использование градиентных способов.

Рисунок 2 – Характерные разновидности расчётных сеток: (а)–(б) ортогональные регулярные с координатной привязкой (а) узлов сетки, (б) центров ячеек [15,с.82]; (в) ортогональная с линейно изменяющимися координатными шагами [16,с.36]; (г) ортогональная регулярная неуниформная [16,с.34]; (д) композиционная [17,с.41]; (е) шестигранная мозаичная с наложенной на неё мозаикой из равносторонних треугольников [18,с.16]; (ж) мозаика, образованная суперпозицией регулярных одномерных сеток [19,с.5]; (и) “проволочное” обрамление моделируемой объёмной фигуры [20,с.4]; (к) контуры разрезов объёмной фигуры и их композиционная криволинейная сеточная модель [20,с.217]; (л) прямолинейная композиционная сетка, адаптированная к криволинейной поверхности моделируемого объекта [21,с.7].

Исходя из перечисленного, в качестве основного инструмента решения задач функционально-статистической оптимизации был выбран способ, известный как “метод точек и сеток” [13,с.281; 14,с.209]. Поскольку экспериментальная проверка одной из наиболее популярных разновидностей последнего – метода случайного поиска (Монте-Карло), – проведённая на первых этапах исследования, продемонстрировала его недостаточную надёжность, слабую предсказуемость и высокую затратность в

отношении используемых ресурсов ЭВМ, выбор был остановлен на ускоренно развивавшихся в то время операциях над регулярными сетками.

Рисунками 2 и 3 иллюстрируются в обобщённом виде произошедшие за последние несколько десятилетий расширение областей применения и диверсификация исследовательских процедур, осуществляющихся сеточными методами.

Рисунок 3 – Распространённые алгоритмические преобразования, выполняемые над расчётными сетками: (а)–(б) управление кривизной линий сетки [22,с.30]; (в)–(г) введение дополнительных наборов сеток [17,с.39-40]; (д) создание сетки, имеющей структуру направленного ациклического графа [23,с.5]; (е) создание композитной сетки наложением на исходную основу оптимально ориентированных мелкоячеечных сеточных фрагментов [24,с.7]; (ж) многосеточные преобразования по отношению к трёхмерным объектам, представляемым в виде «октодеревьев» [25,с.4; 26–27]; (и) схема решения разнообразных систем математических уравнений многосеточными рекурсивно-итерационными методами [28–30]; (к) схема реализации «конвергентного сеточного алгоритма» [31,с.189].

Наиболее интенсивно в последнее время развиваются сеточные методы, связанные с моделированием объёмных объектов и процессов в физике, химии, аэро- и гидродинамике, конструировании, архитектуре, визуализации и распознавании образов. Примеры соответствующих разновидностей расчётных сеток и преобразований над ними представлены на рисунках 2 (и–л) и 3 (ж–к).

В реализованном автором комплексном сеточно-градиентном алгоритме на первых его этапах (рисунок 1, блоки 4–8) используется многомерная квазирегулярная расчётная сетка вида рис. 2 (а). С регулярными сетками её объединяет ортогональность многомерного признакового пространства, предполагающая взаимную некоррелированность используемых признаков – параметров Θ

функционалов (9), (12), (13). От строго регулярных выбранную разновидность сеток отделяет использование в составе параметров Θ разнородных величин, в том числе ранговых и категориальных признаков, в отношении которых является бессмысленным утверждение о регулярности, выражающейся в равенстве расстояний между соседними грациями сеточных координат. Противоречит определению используемой сетки как регулярной также наличие статистической либо иной неявно выраженной взаимосвязанности между некоторыми параметрами, что, как показывают многочисленные выполненные вычислительные эксперименты, в определённой степени компенсируется выбором оптимального значения метрики Минковского r в названных функционалах.

Способ сеточных преобразований, применённый в разработанном автором алгоритме, имеет своего терминологического двойника, схематично изображённого на рисунке 3 (к). Содержание описываемого в [31, с.393-398] “алгоритма сеточной конвергенции” (Convergent gridding algorithm) заключается в итеративно повторяющейся процедуре дробления ячеек сетки, являющейся первоначально грубой регулярной топологической моделью участка земной поверхности, и присвоения её вновь появляющимся многочисленным узлам значений, рассчитываемых на основе усреднения значений близлежащих опорных триангуляционных точек, выбираемых из их достаточного представительного исходного массива. Процесс описанной “сеточной конвергенции” заканчивается после выполнения заданного количества его итераций или при достижении ячейками координатной сетки заданной минимальной величины.

Идея разработанного автором сеточного метода, схематично изображённого на рисунке 4, состоит в методичном уменьшении геометрических размеров квазирегулярной сетки при соответствующем увеличении её разрешающей способности, дополняемом её смещениям по осям координат, вплоть до как можно более полной конвергенции (слияния) сетки с отыскиваемой областью глобального минимума M многомерной целевой функции $D(\Theta)$. Отыскание точного значения координат Θ_M глобального минимума функции осуществляется более адекватным в таком случае градиентным методом, описываемым далее.

Реализация процедур сеточной конвергенции, отображаемой на рисунке 1 блоками 4–8, начинается с построения грубой исходной расчётной сетки, охватывающей целиком область определения функции $D(\Theta)$ – рисунок 4, итерация I. Первоначальные границы этой области совпадают с диапазонами изменения всех фигурирующих в расчётах переменных величин. При проведении регрессионного анализа, как это иллюстрируется рисунком 4, текущая область определения функции $D(\Theta)$ в натуральном выражении постепенно уменьшается за счёт повторяющегося K -кратного сокращения шагов расчётной сетки по всем её координатам. При этом сама уменьшенная в размерах сетка смещается так, чтобы равномерно охватывать найденную на одной из предыдущих итераций точку, которой соответствует минимальное из найденных к данному моменту значений функции $D(\Theta)$. Границы, в пределах которых локализируются формируемые более мелкие сетки, остаются при этом неизменными, свойственными первой итерации конвергенционного цикла.

В отличие от изложенного, при анализе функциональной динамики изначально предусматривается возможность выхода текущих областей определения функции $D(\Theta)$ за границы её первоначально заданной области, сопровождающегося её соответствующим расширением. Этим обеспечивается выполнение операций экстраполяции по отношению к участвующей в анализе “ведомой” функции.

Критериями завершения процедуры сеточной конвергенции (рисунок 1, блок 8) могут быть выбраны: а) выполнение заданного количества итераций или, что эквивалентно, достижение шагом приращения выбранного параметра θ_i заданной минимальной величины; б) прекращение уменьшения очередных расчётных значений $D(\Theta)$ на протяжении заданного количества итераций рассматриваемого конвергенционного процесса. В зависимости от особенностей решаемой задачи могут быть использованы и другие критерии завершения конвергенционного цикла, в том числе его полное исключение из алгоритма.

Рисунок 4 – Зависимость надёжности поиска глобального минимума целевой функции $D(\Theta)$ от значений коэффициента сеточной конвергенции K

Как это иллюстрируется рисунком 4, эффективность сеточной оптимизации в существенной степени зависит от выбора величины коэффициента конвергенции K . В частности, при сохранении неизменным правила $0 < K < 1$ и при приближении этого коэффициента к "1" ($1 - K \rightarrow \min$) возрастает плотность покрытия признакового пространства $\{\Theta\}$ узлами последовательно сменяющих друг друга сеток, а вместе с этим увеличиваются надёжность и точность расчетов. Возрастающие при этом количество итераций и линейно зависящего от них времени работы программы в определенных пределах оправдывают себя. Общее же правило состоит в том, что как коэффициент конвергенции K , так и условие окончания соответствующего цикла должны определяться особенностями решаемой задачи, а при необходимости – результатами пробных расчётов.

Поскольку, как это следует из вышесказанного, конвергентный алгоритм не обеспечивает абсолютной надёжности и точности получаемых результатов, возникает вопрос о преимуществах, обеспечивающих его конкурентоспособность по сравнению с другими сеточными методами. По отношению к популярному методу Монте-Карло это уже было сделано выше, поэтому есть смысл сопоставить предлагаемый метод с

наиболее простым в реализации использованием регулярных сеток. Труднооспоримое преимущество конвергентного алгоритма в этом случае состоит в многократном снижении размерности задачи и соответственно времени её решения на ЭВМ. Если при использовании “канонического” метода сеток и точек для получения заданных надёжности и точности расчётов требуется прибегать к построению чрезвычайно многомерных сеток, имеющих шаги приращения координат сопоставимые по величине с требующейся точностью расчётов, то при обращении к конвергентному алгоритму достаточно ограничиться первоначальным выбором достаточно грубых сеток, разрешающая способность которых автоматически возрастает по мере продолжения итерационного процесса. Нарастающее при этом в арифметической прогрессии время, затрачиваемое на повторяющиеся итерации циклического процесса, не идёт в сравнение с тем растущим в геометрической прогрессии временем, которое расходуется на обшёт многомерных сеток, наделённых требуемой разрешающей способностью.

С другой стороны, обращение к типам сеток и строящимся на них методам, приведённым в качестве примеров на рисунках 2 (в–ж) и 3 (а–е), во-первых, неизбежно сопряжено со значительным усложнением соответствующих вычислительных процедур и, во-вторых, требует знания, как правило, заранее неизвестных особенностей используемых в расчётах целевых функций $D(\Theta)$.

Отличаясь своими высокими экономичностью и скоростью решения, недостижимыми при использовании стандартного метода сеток и точек, сеточная конвергенция уступает последнему в отношении гарантируемых надёжности и точности получаемых результатов. Связано это, в частности, с неуклонным сокращением покрытия исходной области определения функции $D(\Theta)$ расчётной сеткой, уменьшающейся в процессе итераций по своим геометрическим размерам. При сложной конфигурации гиперповерхности, отображающей функцию $D(\Theta)$, это чревато возникновением ситуации, когда из-за своих малых размеров на всех последующих итерациях сетка перестаёт достигать области глобального минимума, находящейся, например, на дне вытянутого и извилистого жёлоба. В таких достаточно редких случаях в качестве отыскиваемого решения программой выдаётся некоторый его промежуточный результат, расчётные параметры $\{\hat{\theta}_i\}$ которого могут существенно отличаться от истинных.

Для избавления расчётов от подобных ошибок или, по крайней мере, снижения их величины и вероятности появления, сеточная конвергенция нуждается в дополнении градиентными методами. Поскольку экспериментальное опробование в этих целях стандартного метода наискорейшего спуска показало его недостаточную надёжность по отношению к решаемым задачам, ниже описывается один из реализованных автором подобных приёмов, проявивших себя с более выгодной стороны.

В основе приёма, названного интерполяционной оптимизацией, лежат представления об относительной гладкости функции $D(\Theta)$, характерной в первую очередь для её локальных участков. Из этих представлений вытекает, что на любом своём относительно малом участке данная многомерная функция может быть представлена в виде поверхности второго порядка в пространстве Θ , описываемой уравнением:

$$D(\Theta) = a_1\theta_1^2 + a_2\theta_2^2 + \dots + a_m\theta_m^2 + a_{m+1}\theta_1 + \dots + a_{m+2}\theta_2 + \dots + a_{2m}\theta_m + a_{2m+1}\theta_1\theta_2 + a_{2m+2}\theta_1\theta_3 + \dots + a_{M-1}\theta_{m-1}\theta_m + a_M, \quad (14)$$

где параметры $\theta_1, \theta_2, \dots, \theta_m$ выступают в качестве независимых переменных величин; $a_1, a_2, \dots, a_M$ – коэффициенты уравнения, значениями которых определяется конфигурация многомерной поверхности в выбранной её точке; M – максимальный номер коэффициента. Нетрудно убедиться, что M численно равен

количеству сочетаний из m по 2: $M = C_m^2$.

Очевидно, что в области минимума функции $D(\Theta)$ уравнением (14) описывается геометрическая фигура, являющаяся гиперпараболлоидом в пространстве размерности m . Следовательно, найдя с помощью описанного выше итерационного процесса приближённое положение минимума функции $D(\Theta)$ и произведя её аппроксимацию в найденной области, мы получаем возможность вычислить положение вершины гиперпараболлоида, с большей точностью индицирующей отыскиваемые координаты минимума названной функции, а тем самым дающей более точное решение задачи функционально-статистического анализа.

Вытекающая отсюда последовательность действий такова.

1. В многомерном признаковом пространстве Θ отыскиваются координаты некоторого количества точек $\hat{\Theta} = \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$, лежащих в окрестности найденной приближённой точки минимума, которые в совокупности с найденной точкой дают возможность построить систему из M уравнений. Очевидно, таких дополнительных точек должно быть $M - 1$.

2. В дополнительных точках вычисляются значения функции $D(\Theta)$.

3. Строится система из M уравнений, в которой коэффициенты $a_1, a_2, \dots, a_M$ выступают в качестве неизвестных величин. Одним из имеющихся стандартных способов отыскивается решение этой системы. Условием перехода к следующему шагу является неотрицательность первых m коэффициентов уравнения (14). Несоблюдение этого условия говорит о невозможности аппроксимации функции в данной её области с помощью гиперпараболлоида.

4. Считая найденные коэффициенты $a_1, a_2, \dots, a_M$ постоянными величинами, отыскиваются m частных производных уравнения (14), приравняваемых к нулю. Например, частная производная по θ_1 описывается уравнением:

$$2a_1\theta_1 + a_{m+1} + a_{2m+1}\theta_2 + a_{2m+2}\theta_3 + \dots + a_{3m-1}\theta_m = 0. \quad (15)$$

Выражения для производных komponуются в систему из m уравнений; отыскивается решение этой системы.

5. Найденное решение, состоящее из m значений параметров $\tilde{\Theta} = \tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_m$, интерпретируется как уточнённое положение минимума функции $D(\Theta)$. В качестве окончательного решения уравнения выбирается тот набор значений параметров $\hat{\Theta}$ или $\tilde{\Theta}$, при котором функция $D(\Theta)$ принимает минимальное значение.

Практическое использование описанной схемы уточнения результатов подтверждает её достаточно высокую эффективность. Так, при значениях метрики Минковского r близких к 2 интерполяционное уточнение в среднем на порядок повышает точность расчетов. В то же время логические рассуждения и практика решения задач показывают, что эта схема в некоторых случаях бывает неэффективной. Это, во-первых, обнаруживается тогда, когда в результате итерационного процесса положение минимума функции $D(\Theta)$, имеющей сложный рельеф, определено со значительной погрешностью. Отыскиваемое с помощью гиперпараболлической аппроксимации расчётное положение минимума, будучи в таких условиях слишком грубым, не даёт ожидаемого уточнения результатов.

Во-вторых, снижение эффективности гиперпараболлической аппроксимации происходит при использовании значений метрики r , существенно отличающихся от 2.

Эти обстоятельства говорят о необходимости продолжения поиска эффективных способов уточнения и повышения надёжности результатов. Одна из идей, ждущих своей реализации, заключается в том, что вид применяемой аппроксимации должен соответствовать используемому в расчётах значению метрики Минковского. Например,

при $r=1$ функция $D(\Theta)$ могла бы с большей эффективностью аппроксимироваться в виде гиперконуса в m -мерном пространстве Θ .

Другая идея в общем виде заключается в следующем. Через точки $\hat{\Theta}$, $\tilde{\Theta}$ и третью точку в пространстве Θ , лежащую между двумя первыми, проводится линия второго порядка. Если эта линия оказывается выпуклой вниз параболой, её вершина может быть принята за искомое решение.

Ещё одна идея состоит в усложнении конвергенционного процесса за счёт включения в каждую его итерационную фазу этапа интерполяционного уточнения промежуточного результата. Это может ускорить сходимость процесса, особенно в тех нередких случаях, когда область минимума функции $D(\Theta)$ имеет сложную конфигурацию.

Примеры решения прикладных задач. Описанный выше алгоритм был реализован в серии последовательно совершенствовавшихся программ, которые параллельно с доработкой интенсивно использовались в многочисленных практических расчётах. Краткие изложения и результаты последних представлены в публикациях [9; 32–38]. Ниже описываются две конкретные задачи, при решении одной из которых был использован метод расширенной параметрической регрессии [34–38]. Решение другой сопровождалось множеством взаимосвязанных расчётов, строившихся на использовании аппарата анализа функционально-статистической динамики [9; 32; 33; 36; 39].

Пример 1. Расширенная параметрическая регрессия.

Отыскивается конкретный вид некоторого эмпирического статистического распределения, подчиняющегося обобщённому закону распределения случайных величин [40, с.43]:

$$F(x) = Ae^{-ax^p} x^{n-t}, \quad (16)$$

где a , p , n – параметры, соответствующие индивидуальным особенностям кривых распределения; A – постоянная, вводимая с целью нормирования функции (16). Таким законом описывается, например, гранулометрический состав дисперсных смесей, образующихся при добыче и переработке твёрдых полезных ископаемых. В соответствии с имевшимися исходными данными решалась задача нахождения коэффициентов несколько видоизменённого по сравнению с (16) регрессионного уравнения:

$$\hat{y} = a_0 \exp(-a_1 d^{a_2}) d^{a_3}, \quad (17)$$

где a_0 , a_1 , a_2 , a_3 – отыскиваемые параметры уравнения; независимая переменная d – классификационный диаметр частиц в их дисперсной смеси; $\hat{y}$ – расчётная плотность объёма частиц данного классификационного диаметра. Исходные данные расчёта представлены в таблице 1, а его результаты, полученные при различных значениях метрики Минковского – в таблице 2 и на рисунке 5.

По виду поля рассеяния точек d на рисунке 5, соответствующих исходным данным (таблица 1), можно априорно прийти к заключению об одномодальном характере распределения объёмов частиц, что делает оправданным применение формулы (17). По рисунку, кроме того, наглядно прослеживается зависимость дисперсии объёмов частиц различных классов d от ожидаемой амплитуды функции плотности их распределения.

В программе, детальное описание которой из-за большого объёма не приводится, такому закону наложения ошибок с разной точностью отвечают пять задействованных способов их нормирования.

Таблица 1 – Экспериментальное гранулометрическое распределение состава дисперсных смесей*

i	d	y	i	d	y	i	d	y
1	0,07	0,007	21	0,58	0,28	41	4,46	0,055
2	0,214	0,08	22	2,08	0,107	42	10,7	0,008
3	0,5	0,58	23	5,0	0,06	43	0,357	0,023
4	1,79	0,35	24	10,0	0,044	44	1,25	0,385
5	5,7	0,05	25	0,166	0,02	45	4,46	0,103
6	10,0	0,013	26	0,583	0,206	46	10,7	0,002
7	0,179	0,12	27	2,08	0,215	47	0,25	0,047
8	0,893	0,57	28	5,0	0,086	48	0,875	0,5
9	2,14	0,22	29	10,0	0,009	49	1,875	0,414
10	5,0	0,05	30	0,167	0,06	50	3,75	0,031
11	10,9	0,08	31	0,583	0,411	51	8,75	0,001
12	0,179	0,1	32	2,083	0,143	52	0,438	0,13
13	0,893	0,39	33	5,0	0,092	53	1,56	0,42
14	2,14	0,148	34	10,0	0,015	54	3,75	0,023
15	5,0	0,065	35	0,357	0,231	55	7,5	0,009
16	10,7	0,007	36	1,25	0,43	56	0,217	0,225
17	0,416	0,129	37	4,46	0,062	57	0,76	0,88
18	2,08	0,232	38	10,7	0,006	58	2,72	0,086
19	5,0	0,07	39	0,357	0,294	59	10,9	0,003
20	10,0	0,005	40	1,25	0,46			

* i – порядковый номер образца, d – классификационный диаметр гранул дисперсной смеси, y – эмпирическое значение функции плотности распределения.

Таблица 2 – Результаты решения регрессионного уравнения (17) при различных значениях метрики Минковского r

r	a_0	a_1	a_2	a_3	$D_r(\theta)$
0,3	0,6539	0,9976	0,8703	1,066	0,30418
0,5	6,777	3,16	0,5087	1,565	0,35839
0,75	2,985	2,003	0,9649	1,947	0,40076
1,0	3,175	2,193	0,7748	1,757	0,43165
2,0	3,332	2,35	0,7591	1,609	0,51888
3,0	9,286	3,667	0,5811	1,935	0,57804

Рисунок 5 – Диаграмма рассеяния классификационного диаметра гранул d и результаты решения уравнения (17) при различных значениях метрики Минковского r

С учётом имеющих место резко выраженных колебаний средних величин ошибок, коррелирующих с ожидаемой амплитудой функции плотности распределения, в качестве наиболее подходящего выбран вариант z_i в функционале (9), описываемый выражением:

$$z_i = \bar{y}_i (1 + \hat{\sigma}_i / \bar{y}_i), \quad (18)$$

где z_i – нормировочный коэффициент; $\hat{\sigma}_i / \bar{y}_i$ представляют собой обобщённые коэффициенты вариации откликов y , подсчитываемые на выделяемых особо участках этой зависимости.

Негативное влияние на точность расчётов неуставленного закона распределения ошибок и предполагаемой автокорреляции откликов устраняется, насколько это возможно, подбором наилучшего способа агрегирования регрессионных ошибок, определяемого значением метрики Минковского r .

На первый взгляд данные таблицы 2 малоинформативны в отношении выбора этого показателя. Действительно, значения суммы регрессионных остатков функции $D_r(\theta)$, монотонно уменьшающиеся с уменьшением r , – явление закономерное, не связанное исключительно с улучшением качества аппроксимации. Среди других расчётных величин наглядная закономерность прослеживается только у a_2 и a_3 , достигающих максимумов при значении $r = 0,75$. Аналогичная картина прослеживается и в отношении коэффициентов a_0 и a_1 , хотя наблюдающиеся у них минимумы при $r = 0,75$ являются лишь локальными.

Один из объективных критериев для выбора оптимального значения r вытекает из сопоставления кривых, построенных по расчётным данным. Из рисунка 5 видно, что вместе с ростом значения r в диапазоне от 0,3 до 3 форма кривых, описываемых уравнением (17), претерпевает плавную эволюцию. В частности, их модальное значение вначале растёт вместе с ростом r , а при $r > 0,75$, столь же монотонно снижается. Вместе с этим визуально изменяется и степень соответствия получаемых кривых их ожидаемым очертаниям: кривая, соответствующая $r = 0,75$, лучше остальных аппроксимирует исследуемую зависимость в её наиболее критичной модальной области.

Ещё один объективный критерий для выбора r заключается в стабильности решения. Варьируя задаваемые границы диапазонов изменения $a_0, \dots, a_3$, число циклов итерационного процесса, коэффициент конвергенции K и другие параметры расчётов θ_i , можно приходить к существенно различающимся результатам. Решение считается стабильным, если его результаты оказываются близкими друг к другу при достаточно широком варьировании описанных начальных условий. В рассматриваемом примере это соответствует значениям r близким к 0,75.

Таким образом, одновременно несколько критериев – особенности в эволюции коэффициентов регрессионного уравнения, близость отображающей зависимость кривой к её ожидаемому виду, стабильность решения – приводят к одному значению метрики Минковского $r = 0,75$, которое в данном случае с достаточными основаниями можно считать оптимальным. Соответствующее этому значению r регрессионное уравнение имеет вид:

$$\hat{y} = 2,985 \times \exp\left(-2,003 \times d^{0,9649}\right) \times d^{1,947}. \quad (19)$$

Было бы опрометчивым выдавать полученное решение за заведомо безупречное, поскольку в иных условиях, – например, после тщательной выверки исходных данных и устранения возможных “загрязнений” выборки – сведённые в таблицу 2 расчётные параметры могут заметно измениться. Нельзя исключить вероятность того, что и само уравнение (17) не в полной мере соответствует исследуемому статистическому распределению. Замена же регрессионной модели

другой, более адекватной, может иметь своим результатом выбор иного значения метрики Минковского.

В качестве, по крайней мере, любопытного факта следует добавить, что значение метрики Минковского $r \approx 0,75$ оказалось близким к оптимальному и в ряде других выполненных автором расчётов с использованием реальных исходных данных.

Пример 2. Анализ функционально-статистической динамики.

Исходными данными для решения данной задачи явились анкеты, заполненные респондентами обширного социологического исследования, которым были охвачены 14 городов в различных регионах бывшего СССР [41]. Большинство из 257 вопросов анкеты были объединены в тематические группы, отражающие личностные и социальные характеристики респондентов, условия их жизни и трудовой деятельности, отношение к различным сторонам окружающей действительности. В частности фиксировались [39, с.114-117]: возраст, пол, уровень образования респондента; характер его трудовой и досуговой деятельности; интенсивность различных внепроизводственных занятий респондента; наличие в его квартире или доме бытовых удобств и предметов длительного пользования; субъективная оценка респондентом требований, предъявляемых к работе, и другие характеристики.

Из более чем трёх тысяч анкет, заполненных взрослыми жителями Хабаровска, полторы тысячи не имеющих нареканий к полноте и качеству заполнения были подвергнуты интенсивной компьютерной обработке с помощью разработанных автором формализованных методов, в том числе описываемых в статье. Результаты обработки вместе с исходными анкетами были переданы в единый информационный центр межрегионального исследования и использованы для совместно выполнявшихся анализов [32; 33].

Один из видов компьютерной обработки состоял в построении зависимостей, отражающих взаимосвязи различных видов внепроизводственной деятельности респондентов с их квалификационными характеристиками. Пример такого сопоставления представлен на рисунке 6.

Сравнивались между собой частоты участия рабочих и интеллигенции в различных досуговых видах деятельности, представляемые как функции их квалификационных уровней. При этом сами названные уровни выражались в разных показателях для рабочих и служащих. Более того, не совпадали даже соответствующие шкалы: в отношении рабочих использовалась пятизначная ранговая шкала, в отношении служащих – также ранговая, но представленная только четырьмя дискретными значениями.

Приведённые на рисунке результаты сопоставительного функционально-статистического анализа свидетельствуют о существенном сходстве культурно-досуговой активности, характерной для рабочих и служащих России в последней четверти XX в. Подтверждением этого являются значения $s \approx 0$ и $m \approx 1$ на рисунках 1 (а–е). С другой стороны, значительно отличающимися от названных значениями s и m , приведёнными на рисунках 1 (д, е), подтверждается наличие существенной разницы между двумя сравниваемыми социальными группами в отношении двух конкретных видов деятельности.

Следует заметить, что рисунком 6 иллюстрируются сугубо промежуточные результаты исследования, использовавшиеся для отладки разрабатываемых программ и оценки влияния, оказываемого на результаты расчётов используемыми в них варьируемыми параметрами. В частности, демонстрируемые данным рисунком результаты получены при значении метрики Минковского $r = 0,75$ и выбранном способе экстраполяции зависимостей, соответствующем их продолжению из конечных точек в направлении общего линейного тренда каждой экстраполируемой зависимости. Такое сочетание устанавливаемых исследователем параметров оказалось предпочтительнее других опробованных вариантов и было использовано в последующих массовых расчётах, проводившихся с тем же массивом исходных данных.

Рисунок 6 – Сопоставление интенсивностей посещения рабочими и интеллигенцией различных культурных мероприятий в зависимости от их квалификационных уровней: а) кино, б) драмтеатра, в) концертов, г) оперы и балета, д) спортивных состязаний, е) публичных лекций.

Определившие содержание рисунка 6 зависимости, характеризующие отдельные виды деятельности, в сгруппированном виде представлены на рисунке 7 (в). Особенность всех иллюстрируемых рисунком 7 обобщённых зависимостей состоит в их предельно упрощённом одномерном представлении – в виде точек на виртуальной оси, обозначенной как “интегральный уровень развития”. В качестве “эталонных” (опорных) при этом выбраны зависимости, относящиеся к работникам преимущественно физического труда, ранжированным в соответствии с характером их трудовой деятельности по пяти категориям – начиная от неквалифицированных

рабочих и заканчивая имеющими наивысшую квалификацию рабочими-интеллигентами. Сопоставляемая с рабочими социальная группа служащих и интеллигенции классифицировалась по четырём категориям – от исполнителей без специального образования до специалистов, руководителей и творческой интеллигенции, чьи трудовые обязанности предполагают как минимум наличие высшего образования.

Рисунок 7 – Различия в социально-экономическом статусе, достигнутом рабочими и служащими разных квалификационных категорий, по отношению к уровням их семейного материального благосостояния (а) и отдельным сторонам производственной и внепроизводственной деятельности (б–г) [32,с.26; 33,с.29; 39,с.138].

В качестве примечания к рисункам 6 и 7 необходимо отметить, что фигурирующие на них зависимости являются результатами статистического обобщения большого объёма первичной анкетной информации, выполнявшегося кратко описанными выше методами. При этом чаще всего названные операции были многоступенчатыми: 1) вначале обобщались ответы на один и тот же одиночный вопрос анкеты множества респондентов, рассортированных по изучаемым группам; 2) при наличии “батарейных” (однотипных, взаимосвязанных) вопросов обобщались ответы на них; 3) предыдущие результаты обобщались по отдельности для каждой из выделенных на первом этапе групп респондентов; 4) при необходимости статистически суммировались ответы респондентов, разделённых на группы на предыдущих этапах. По возможности все статистические усреднения осуществлялись с помощью встроенных в компьютерные программы процедур и выполнялись по правилам обобщённых средних величин, описываемых выражением (3).

Обсуждение результатов. Как отмечено выше, описанный алгоритм и реализующие его программы разрабатывались на базе отечественных ЭВМ, достаточно мощных для своего времени и, тем не менее, значительно уступавших по своим техническим параметрам современным аналогам. В первую очередь это касается оперативной и особенно долговременной памяти, на несколько порядков

уступавших по своим объёмам и быстродействию даже современным ноутбукам. Это наложило существенный отпечаток на структуру алгоритма и особенности его программной реализации. Положительное влияние данного обстоятельства выразилось в концентрации усилий на поиске и реализации программных решений, максимально экономичных в отношении используемых ресурсов ЭВМ, но в то же время приводящих к получению надёжных результатов. Отрицательный эффект ограниченности технических параметров ЭВМ проявился, например, в отказе от эффективных в вычислительном отношении алгоритмических и программных приёмов, требующих для своей реализации ёмкой и быстродействующей машинной памяти.

Трудновосполнимый удар по идущим полным ходом и нередко весьма успешным исследованиям и разработкам, в том числе авторским, был нанесён произошедшим в конце 1980-х годов переводом отечественной науки на самофинансирование и самоокупаемость [42]. В отношении автора это выразилось в выставлении “задним числом” счёта за использованное машинное время, сумма которого равнялась десятилетней зарплате автора. Результатами этого стали вынужденная смена места работы и полный отказ от основывавшихся на интенсивном использовании ЭВМ теоретико-прикладных инициатив, внезапно ставших экстремально дорогостоящими и поэтому практически неосуществимыми.

Из уже обдумывавшихся и частично начавших осуществляться идей, разноуровневых по своей предполагаемой значимости и оставшихся нереализованными, можно назвать следующие.

- Теоретический поиск, обоснование и разработка вариантов программной реализации средств, аналогичных существующим в общепринятых разновидностях регрессионного и корреляционного анализов и позволяющих оценивать качество выполненных расчётов, проверять связанные с ними статистические гипотезы, устанавливать доверительные интервалы расчётных величин.

- Обоснование и программная реализация способов интерполяции и экстраполяции анализируемых зависимостей, существенно расширяющих возможности единственного реализованного кусочно-линейного способа интерполяции и трёх реализованных аналогичных способов экстраполяции.

- Экспериментальная проверка и практическая реализация нескольких вариантов совершенствования сеточно-градиентного метода функциональной оптимизации, рассчитанных на повышение его надёжности, существующих в настоящее время на уровне в разной степени обоснованных идей и пробных экспериментов.

- Переработка алгоритма и его программной реализации, рассчитанная на максимально полное использование технических и программных возможностей современных компьютеров – многократно возросших объёмов их памяти, расширившегося и обновлённого программного обеспечения, в том числе за счёт существующих пакетов математико-статистической обработки информации.

Несмотря на скоростное прекращение поисково-исследовательских и прикладных работ, достигнутыми, по крайней мере, на экспериментально-демонстрационном, а частично и внедренческом уровнях можно считать следующие результаты.

1. Обоснована перспективность направления развития регрессионного анализа, связанного с расширением состава варьируемых величин, используемых в качестве параметров расчётов. Принципиальной новизной обладает, в частности, включение в состав параметров варьируемой метрики Минковского, позволяющей совместить в одной программе различные способы агрегирования регрессионных ошибок, в том числе далеко выходящие за рамки существующих, и тем самым сделать эту процедуру управляемой, обеспечивающей достижение оптимальных результатов.

2. Продемонстрировано, что многие считающиеся отдельными видами регрессионного анализа его по существу однотипные разновидности, различающиеся лишь конкретными особенностями используемых регрессионных моделей, характером переменных величин, способами интерполяции и экстраполяции зависимостей и т.д., могут быть унифицированы и объединены в одном алгоритме и его программной

реализации. Это способствует упрощению прикладных исследовательских поисков и повышению их эффективности.

3. Подтверждена возможность дополнения существующих методов математической статистики новым её направлением, обеспечивающим сопоставление сравниваемых функциональных и статистических зависимостей по их амплитуде, скорости протекания и взаимному смещению относительно выбранного ключевого параметра. Необходимость развития данного направления подтверждается наличием множества перспективных областей его приложений, относящихся к разнообразным отраслям производственной и научной деятельности.

4. Разработан и реализован в эффективных компьютерных программах комбинированный сеточно-градиентный алгоритм, позволяющий при достаточных уровнях точности и надёжности получать искомые результаты при минимальных расходах времени работы ЭВМ и других её технических ресурсов. Существенным достоинством алгоритма является его универсальность, позволяющая сочетать в одной программе решение задач расширенной параметрической регрессии с реализацией прикладных направлений анализа функционально-статистической динамики.

БИБЛИОГРАФИЯ

1. Мудров В.И., Кушко В.Л. Метод наименьших модулей. – М.: Знание, 1971. 61 с.
2. Рогачёв А.Ф. Параметризация эконометрических зависимостей методом наименьших модулей. – URL: <http://uecs.ru/instrumentalnii-metody-ekonomiki/item/388-2011-04-04-09-10-38>
3. Regression analysis. – URL: https://en.wikipedia.org/wiki/Regression_analysis
4. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. 755 с.
5. Среднее степенное. – URL: https://ru.wikipedia.org/wiki/Среднее_степенное
6. Mean. – URL: <https://en.wikipedia.org/wiki/Mean>
7. Среднее Колмогорова. – URL: https://ru.wikipedia.org/wiki/Среднее_Колмогорова
8. Прилуков А.Н. Структура критериев как инструментов социально-экономических исследований и повседневных коммуникаций. – URL: http://modernsciencejournal.org/release/2017/USN_2017_3_tom.pdf
9. Прилуков А.Н. Анализ коллинеарности как инструмент параметрического сопоставления социально-экономических зависимостей. – URL: https://rjoas.com/issue-2017-03/article_13.pdf
10. Электронная вычислительная машина EC-1052. – URL: <http://www.computer-museum.ru/histussr/2-52.htm>
11. Справочник по ЭВМ. – Киев: Наукова думка, 1989. 544 с. – URL: http://www.pseudology.org/science/Spravochnik_EVM.pdf
12. Казакова И.А. История вычислительной техники. – URL: <http://elib.ict.nsc.ru/jspui/bitstream/ICT/1262/1/histor.pdf>
13. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Статистика, 1973. 392 с.
14. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Кн.2. – М.: Финансы и статистика, 1987. 351 с.
15. Wesseling P. Introduction to multigrid methods. – URL: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19950019975.pdf>
16. Janka A. Multigrid for finite volumes: structured 2D case. – URL: http://perso.unifr.ch/ales.janka/papers/volume_aglomer.pdf
17. Мареев В.В., Станкова Е.Н. Многосеточные методы. – URL: <http://www.apmath.spbu.ru/ru/staff/stankova/publ/publ.pdf>
18. Effinger-Dean L., Bailey D. The Empire Problem in Penrose Tilings. – URL: <http://www.cs.williams.edu/~bailey/06le.pdf>

19. Bragg D. Tiling Spaces: Quasicrystals & Geometry. – URL: <https://www2.le.ac.uk/departments/mathematics/research/topological/documents/bragg-presentation-of-thesis>
20. GridPro GUI Manual Version 2.2. – URL: http://sp.gridpro.com/docs/WS_GUI_Manual_v6.6.pdf
21. Sterz O., Hauser A., Wittum G. Adaptive Local Multigrid Methods for the Solution of Time Harmonic Eddy Current Problems. – URL: <http://archiv.ub.uni-heidelberg.de/volltextserver/5610/1/preprint.pdf>
22. Вальгер С.А., Фёдорова Н.Н. Применение алгоритма адаптации расчётной сетки к решению уравнений Эйлера. – URL: www.ict.nsc.ru/jct/getfile.php?id=1495
23. Mitra S., Parashar M., Browne J.C. DAGH: User's Guide. – URL: <http://mgnet.org/mgnet/Codes/dagh/tutorial.pdf>
24. Flaherty J. E. Finite element analysis. – URL: <http://www.cs.rpi.edu/~flaherty/feaframe.html>
25. Virtualization-aware application framework for hierarchical multiscale simulations on a Grid. – URL: https://www.researchgate.net/profile/Aiichiro_Nakano/publication/228612748_Virtualization-aware_application_framework_for_hierarchical_multiscale_simulations_on_a_Grid/
26. Октодерево. – URL: <https://ru.wikipedia.org/wiki/Октодерево>
27. Sundar H., et al. Low-constant Parallel Algorithms for Finite Element Simulations using Linear Octrees. – URL: <http://sc07.supercomputing.org/schedule/pdf/pap117.pdf>
28. Speh M., Ruede U., Heilmann M. What are Multigrid Methods? – URL: <http://mgnet.org/mgnet/tutorials/MG/MG/MGwhat.html>
29. The Multigrid Workbench: Linear Iterations. – URL: <http://mgnet.org/mgnet/tutorials/xwb/liniter.html>
30. Многосеточный метод. – URL: https://ru.wikipedia.org/wiki/Многосеточный_метод
31. Schlumberger. Introduction to CPS-3. – URL: http://read.pudn.com/downloads130/doc/556189/train/train/GF4_IntroCPS3_manual.pdf
32. Прилуков А.Н., Кирх А.В. Многомерность социальной структуры и интегральные показатели // Социологические исследования в прибалтийских советских республиках. Ч.2. – Вильнюс: 1986. С. 14-27.
33. Prilukov A., Kirch A., Kukk T. Multidimensionality of social structure and integral indicators // Sociological research in the Baltic Soviet Republics. Part 2. – Vilnius: 1986. P. 14-30.
34. Прилуков А.Н. Толерантный регрессионный анализ и пример его программной реализации. – Хабаровск, 1988. 40 с. Деп. в ВИНТИ. № 8527-B88.
35. Прилуков А.Н. Толерантный регрессионный анализ: предпосылки, концепции, программная реализация. Препринт / ИГД ДВО АН СССР. – Владивосток, 1989. 48 с.
36. Прилуков А.Н. Новые формализованные методы анализа объектов и процессов минералопользования // Проблемы комплексного освоения георесурсов. – Хабаровск, ИГД ДВО РАН, 2010. С. 197-202.
37. Прилуков А.Н. Идеи толерантного регрессионного анализа и их реализация. – URL: http://modernsciencejournal.org/release/2017/USNO_2017_1_tom.pdf
38. Прилуков А.Н. Расширенная параметрическая регрессия: основные идеи и их реализация. – URL: https://rjoas.com/issue-2017-07/article_03.pdf
39. Прилуков А.Н. Развитие социально-классовой структуры: показатели и методы измерения. – URL: <http://search.rsl.ru/ru/record/01008035910>
40. Авдеев Н.А. Об аналитическом методе расчёта седиментометрического дисперсионного анализа. – Ростов-на-Дону: Изд-во Рост.ун-та, 1964. 202 с.
41. Советский город: социальная структура. – М.: Мысль, 1988. 286 с. – URL: <http://urss.ru/cgi-bin/db.pl?lang=Ru&blang=ru&page=Book&id=133870>
42. Шкабардня М.С. – URL: http://www.warheroes.ru/hero/hero.asp?Hero_id=20060