

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра «Прикладная математика и информатика»

(наименование кафедры)

01.04.02 ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА

(код и наименование направления подготовки)

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

(направленность (профиль))

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

на тему Алгоритм стохастической аппроксимации в задаче адаптации регрессионной модели

Студент

А.М. Войтович

(И.О. Фамилия)

(личная подпись)

Научный
руководитель

Г.А. Тырыгина

(И.О. Фамилия)

(личная подпись)

Руководитель программы

(ученая степень, звание, И.О. Фамилия)

(личная подпись)

« _____ » _____ 20 _____ Г.

Допустить к защите

Заведующий кафедрой к.т.н., доцент, А.В. Очеповский

(ученая степень, звание, И.О. Фамилия)

(личная подпись)

« _____ » _____ 20 _____ Г.

Тольятти 2018

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
Глава 1 Теоретические основы построения регрессионной модели.....	6
1.1 Основные предпосылки построения регрессионной модели.....	6
1.2 Построение регрессионной модели	9
Глава 2 Адаптация регрессионной модели методом Брауна	21
2.1 Применение метода Брауна для адаптации регрессионной модели	21
2.2 Реализация метода Брауна для адаптации регрессионной модели	51
Глава 3 Адаптация регрессионной модели методом стохастической аппроксимации.....	55
3.1 Теоретические вопросы одномерной стохастической аппроксимации ..	55
3.2 Практические вопросы использования одномерной стохастической аппроксимации.....	66
3.3 Применение стохастической аппроксимации для адаптации регрессионной модели.....	72
3.4 Реализация алгоритма стохастической аппроксимации для адаптации регрессионной модели.....	80
3.5 Сравнительный анализ эффективности адаптационных алгоритмов.....	83
ЗАКЛЮЧЕНИЕ	87
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ.....	89
ПРИЛОЖЕНИЕ А	94

ВВЕДЕНИЕ

Регрессионный анализ находит широкое применение при прогнозировании явлений в разных областях знаний. Практика показала, что регрессионные уравнения – хорошие измерители связей между явлениями.

Данные, полученные в результате исследования, постоянно меняются и обновляются. Регрессионные модели, построенные по этим данным, устаревают, что в свою очередь снижает точность прогнозов, полученных с помощью этих моделей. Для решения этой проблемы прибегают к методам, позволяющим адаптировать регрессионные модели с учётом поступления новой информации.

Одним из адаптационных методов является метод Брауна. Он позволяет адаптировать исходную модель, но при этом приходится учитывать исходные и новые данные, что в свою очередь является недостатком.

Предлагается другой алгоритм адаптации – метод стохастической аппроксимации. Он позволяет адаптировать регрессионную модель для новых данных, при этом не требуется заново строить модель с использованием исходных и полученных данных. Этим определяется актуальность темы.

Стохастическая аппроксимация – сравнительно молодой метод для решения задач в условиях неполной информации и широко используется в естественных и технических науках.

Профессор Х. Хотеллинг в статье, опубликованной, в 1941 г., обсудил многие идеи метода стохастической аппроксимации, затем в работах Фридмана и Севеджа, а также и других авторов появились родственные результаты. Бурхгольдер рассмотрел обобщённый метод стохастической аппроксимации. Итерационный метод стохастической аппроксимации был предложен Дворецким. В свою очередь, метод Дворецкого укладывается в обобщённые методы, предложенные Браверманом и Розеноэром и Волконским и Иванковым.

Но лишь Роббинс и Монро в своей основополагающей статье дали

формальную математическую трактовку этого вопроса. С тех пор во многих работах по стохастической аппроксимации, появившихся в теоретических и прикладных журналах, отмечалась необходимость использования этого метода в различных областях.

Цель работы:

Применить алгоритм стохастической аппроксимации для адаптации регрессионной модели.

Объект исследования:

Регрессионная модель.

Предмет исследования:

Адаптация регрессионной модели.

Для достижения цели сформулируем следующие задачи:

1. проанализировать алгоритм построения регрессионной модели;
2. использовать метод Брауна для адаптации регрессионной модели;
3. использовать метод стохастической аппроксимации для адаптации регрессионной модели и осуществить сравнительный анализ адаптационных алгоритмов.

Апробация результатов диссертации. Основные результаты докладывались на следующих конференциях:

- III Международная научно-практическая конференция (школе-семинаре) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук», 2017 г.
- IV Международная научно-практическая конференция (школе-семинаре) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук», 2018 г.

Научная новизна состоит в использовании алгоритма стохастической аппроксимации для адаптации регрессионной модели.

Теоретические основы исследования включают использование научных

трудов зарубежных и отечественных авторов по стохастической аппроксимации и регрессионным моделям.

Практическая значимость состоит в том, что предлагаемый подход может использоваться для адаптации регрессионных моделей.

Достоверность обусловлена тем, что полученные результаты вычислений не противоречат статистическим данным.

Основные положения, выносимые на защиту:

- разработка алгоритма стохастической аппроксимации для адаптации регрессионной модели;
- анализ результатов вычислительных экспериментов адаптации регрессионной модели методом Брауна и методом стохастической аппроксимации.

Текст диссертационной работы состоит из введения, трех глав, вывода по работе, списка литературы и приложения. Объем диссертации составляет 92 страниц, содержит 20 рисунков, список литературы включает 34 наименования.

В первой главе диссертации проводится анализ теоретических основ регрессионных моделей и методов их построения.

Во второй главе проводится анализ адаптационного метода Брауна.

В третьей главе проводится анализ алгоритма стохастической аппроксимации и его применение для адаптации регрессионных моделей. Описывается построение математической модели и алгоритма для ее решения. Программный продукт состоит из головного модуля, модулей содержащего математическую модель, модуля вычислений, а также модулей адаптации коэффициентов модели к изменяющимся данным и вывода результатов расчета.

В заключении описаны выводы результатов работы программного продукта.

В приложение находится листинг кода программного продукта.

Глава 1 Теоретические основы построения регрессионной модели

1.1 Основные предпосылки построения регрессионной модели

Для решения поставленных задач, требуется изучить основы построения регрессионных моделей. Один из видов регрессионных моделей описывает линейную зависимость: зависимое значение y с учётом погрешности ε представляет линейную функцию от k других независимых переменных x_1, x_2, \dots, x_k .

Случайная выборочная совокупность объёма n , полученная из генеральной совокупности $(y, x_1, x_2, \dots, x_k)$, на i -м наблюдение имеет вид $(y_i, x_{i1}, x_{i2}, \dots, x_{ik})$, при $i = 1, 2, \dots, n$.

Классической линейной моделью множественной регрессии является модель:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \text{ при } i = 1, 2, \dots, n \quad (1.1.1)$$

где $\beta_0, \beta_1, \dots, \beta_k$ — параметры модели, требующие оценки по выборке. [7]

Требования, выдвигаемые для независимых переменных и погрешности модели:

- независимые переменные x_1, x_2, \dots, x_k должны быть неслучайными, т. е. они получены без погрешности;
- переменные x_1, x_2, \dots, x_k не должны быть связаны линейной функциональной зависимостью;
- погрешность модели ε_i , являются независимыми величинами с математическим ожиданием равным нулю $M\varepsilon_i = 0$ и дисперсией равной: $D\varepsilon_i = \sigma^2$ при $i = 1, 2, \dots, n$. Коэффициент ковариации равен

$$\text{cov } \varepsilon_i, \varepsilon_l = M \varepsilon_i - M\varepsilon_i \varepsilon_l - M\varepsilon_l = M\varepsilon_i \varepsilon_l = \begin{cases} \sigma^2 & \text{при } i = l \\ 0 & \text{при } i \neq l \end{cases}, \quad (1.1.2)$$

где $i, l = 1, 2, \dots, n$;

- вектор значений погрешности модели $\varepsilon = \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n^T$ подчиняется n -мерному нормальному закону распределения. Вектор математических ожиданий погрешности равен нулю $M\varepsilon = 0$ и матрица ковариаций $\varepsilon \sigma^2 E_n$, т. е. $\varepsilon \in N_n(0; \sigma^2 E_n)$, где E_n — единичная матрица размером $n \times n$. [1]

Математическое ожидание зависимого значения y , заданное на векторе независимых переменных $X_i = x_{i1}, x_{i2}, \dots, x_{ik}^T$, при учёте (1.1.1) имеет вид:

$$y_i = M y_i / X_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}. \quad (1.1.3)$$

Полученное регрессионное уравнение, описывает функциональную зависимость среднего значения y от независимых переменных x_1, x_2, \dots, x_k .

В уравнение (1.1.3) свободным членом является значение β_0 . Свободный член обычно не рассматривается, так как ситуация, в которой все независимые переменные x_1, x_2, \dots, x_k равны нулю, бессмысленна. Примером может быть регрессионная модель, описывающая производительность труда, существование которой не имеет смысла без учёта рабочих мест, количества работников и т.д. [7,8]

Коэффициентами регрессии являются параметры $\beta_1, \beta_2, \dots, \beta_k$. Параметр регрессионной модели β_j в формуле (1.1.3) определяет, среднее значение изменения y , при увеличении переменной x_j на единицу, с учётом того, что остальные независимые переменные, входящие в регрессионную модель, не изменятся. Если прибавить единицу к независимой переменной x_{ik} в формуле (1.1.3), то получим выражение:

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + 1 = y_i + \beta_k. \quad (1.1.4)$$

Линейная регрессионная модель, записанная в матричном виде:

$$Y = X\beta + \varepsilon, \quad (1.1.5)$$

где $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ – вектор-столбец значений зависимых переменных (размерности n);

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \quad \text{— матрица значений независимых}$$

переменных (размерности $n \times k + 1$);

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad \text{— вектор-столбец параметров регрессионной модели,}$$

требующих оценки по выборке, (размерности $k + 1$);

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \text{— вектор-столбец значений погрешности регрессионной модели}$$

(размерности n).

С учётом того, что

$$M\varepsilon = 0, \quad (1.1.6)$$

где 0 — вектор-столбец размера n , значения которого равны нулю, а матрица ковариаций равна

$$\varepsilon = M\varepsilon\varepsilon^T = M \begin{pmatrix} \varepsilon_1 & & & & \\ \varepsilon_2 & & & & \\ \vdots & & & & \\ \vdots & & & & \\ \varepsilon_n & & & & \end{pmatrix} \varepsilon_1 \varepsilon_2 \dots \varepsilon_n = M \begin{pmatrix} \varepsilon_1^2 & \varepsilon_1 \varepsilon_2 & \dots & \varepsilon_1 \varepsilon_n \\ \varepsilon_2 \varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2 \varepsilon_n \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \varepsilon_n \varepsilon_1 & \varepsilon_n \varepsilon_2 & \dots & \varepsilon_n^2 \end{pmatrix}. \quad (1.1.7)$$

Из равенства (1.1.2) следует, что при $i = 1, 2, \dots, n$, $M\varepsilon_i^2 = \sigma^2$ и $M\varepsilon_{i_1} \varepsilon_{i_2} = 0$ когда $i_1 \neq i_2$, следовательно

$$\varepsilon = M \varepsilon \varepsilon^T = \sigma^2 E_n. \quad (1.1.8)$$

где $E_n = \begin{matrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & 1 \end{matrix}$ — единичная матрица размерности $(n \times n)$.

1.2 Построение регрессионной модели

Для оценки вектора параметров β используем метод наименьших квадратов. На основании условия минимизации скалярной суммы квадратов Q , находится оценка элементов вектора β ,

$$Q = Y - X\beta \quad Y - X\beta. \quad (1.2.1)$$

Действительно

$$Y - X\beta = \begin{matrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{matrix} - \begin{matrix} \beta_0 + \sum_{j=1}^k x_{1j}\beta_j \\ \dots \\ \beta_0 + \sum_{j=1}^k x_{nj}\beta_j \end{matrix} = \begin{matrix} y_1 - \beta_0 - \sum_{j=1}^k x_{1j}\beta_j \\ \dots \\ y_n - \beta_0 - \sum_{j=1}^k x_{nj}\beta_j \end{matrix}, \quad (1.2.2)$$

используя полученное выражение в формуле (1.2.1), при учёте условия (1.1.3) найдем скалярную сумму квадратов:

$$Q = \sum_{i=1}^n y_i - \beta_0 - \sum_{j=1}^k x_{ij}\beta_j \quad ^2 = \sum_{i=1}^n y_i - y_i \quad ^2. \quad (1.2.3)$$

Система уравнений $\frac{\partial Q}{\partial \beta_j} = 0$ является условием обращения скалярной суммы квадратов Q в минимум, при $j = 0, 1, 2, \dots, k$. Продифференцировав данную систему, получим:

$$-2X^T Y - X\beta = 0, \quad (1.2.4)$$

где X^T — транспонированная матрица, полученная на основе матрицы X .

Подставляя оценку b , полученную методом наименьших квадратов, вместо параметров регрессионной модели β , получаем

$$X^T Y = X^T X b, \quad (1.2.5)$$

где $b = b_0, b_1, \dots, b_k \quad ^T$ — вектор-столбец оценок (размерности $(k + 1)$).

Обе части уравнения перемножаются слева на обратную матрицу $X^T X^{-1}$. Полученное выражение имеет вид:

$$X^T X^{-1} X^T Y = X^T X^{-1} X^T X b, \quad (1.2.6)$$

B_{k+1}

оценка вектора β , полученная методом наименьших квадратов имеет вид:

$$b = \beta = X^T X^{-1} X^T Y. \quad (1.2.7)$$

Следует доказать несмещенность оценки методом наименьших квадратов.

Основываясь на том, что операции проводимые над матрицами, имеют свойства ассоциативности и дистрибутивности, с учётом формул (1.2.7) и (1.1.5) получим выражение

$$b = X^T X^{-1} X^T Y = X^T X^{-1} X^T X \beta + \varepsilon = X^T X^{-1} X^T X \beta + X^T X^{-1} X^T \varepsilon, \quad (1.2.8)$$

после преобразований получим

$$b = \beta + X^T X^{-1} X^T \varepsilon. \quad (1.2.9)$$

На основании (1.2.9) получим, что вектор оценок b подчиняется $(k + 1)$ -мерному нормальному закону распределения при условии нормального распределения вектора отклонений ε . Математическое ожидание Mb и матрица ковариаций b определяют закон распределения оценок b . Исходя из того, что матрица независимых значений X постоянна, то при учете формулы (1.1.6) получаем:

$$Mb = \beta + X^T X^{-1} X^T \varepsilon = \beta. \quad (1.2.10)$$

Выражение (1.2.10) показало, что полученные значения b являются несмещенными оценками.

Для линейной регрессионной модели вектор оценок b является несмещенной оценкой и дисперсия вектора β минимальна.

Матрица ковариаций вектора оценок b равна:

$$b = M (b - \beta) (b - \beta)^T. \quad (1.2.11)$$

После замены переменных по формуле (1.2.9) получим:

$$b = M (X^T X^{-1} X^T \varepsilon) (X^T X^{-1} X^T \varepsilon)^T. \quad (1.2.12)$$

С учётом свойств матриц, получим:

$$X^T X^{-1} X^T \varepsilon^T = \varepsilon^T X X^T X^{-1}, \quad (1.2.13)$$

на основании чего, получим

$$b = X^T X^{-1} X^T M \varepsilon \varepsilon^T X X^T X^{-1}. \quad (1.2.14)$$

Выполняя условие (1.1.7) и равенство $X = EX$, в котором E — единичная матрица, получаем:

$$b = \sigma^2 X^T X^{-1} X^T X X^T X^{-1}, \quad (1.2.15)$$

и матрица ковариации примет финальный вид

$$b = \sigma^2 X^T X^{-1}. \quad (1.2.16)$$

Определим смысл элементов матрицы ковариаций. Значения дисперсии элементов вектора оценок b находятся на главной диагонали матрицы (1.2.6). Остальные значения матрицы (1.2.6) являются значениями коэффициентов ковариации. Коэффициент ковариации, находящийся на пересечении j -го столбца и l -й строки матрицы (1.2.6) имеет вид

$$\text{cov } b_l b_j = M (b_l - \beta_l) (b_j - \beta_j), \quad (1.2.17)$$

где $l, j = 0, 1, 2, \dots, k$.

Оценка b_j коэффициента линейной регрессионной модели β_j $j = 0, 1, 2, \dots, k$ является линейной функцией (1.2.9) от погрешности ε . Оценка имеет нормальное распределение, значение математического ожидания равно β_j , а дисперсия имеет вид на основании формулы (1.2.6).

$$D b_{j-1} = \sigma^2 X^T X^{-1}_{jj}, \quad (1.2.18)$$

где $X^T X^{-1}_{jj}$ — элемент обратной матрицы $X^T X^{-1}$, находящийся на пересечении j -го столбца и j -й строки, где $j = 1, 2, \dots, k + 1$.

С целью нахождения оценки несмещенности s^2 остаточной дисперсии σ^2 , используем вектор остатков $e = Y - \hat{Y} = Y - Xb$.

На основании того, что $b = X^T X^{-1} X^T Y$ и $Y = X\beta + \varepsilon$, получаем вектор остатков равный:

$$e = X\beta + \varepsilon - X X^T X^{-1} X^T X\beta + \varepsilon = X\beta + \varepsilon - X X^T X^{-1} X^T X\beta - X X^T X^{-1} X^T \varepsilon = X\beta + \varepsilon - X\beta - X X^T X^{-1} X^T \varepsilon, \quad (1.2.19)$$

после преобразований, получаем

$$e = \varepsilon - X X^T X^{-1} X^T \varepsilon \quad (1.2.20)$$

Транспонированный вектор остатков имеет вид

$$e^T = \varepsilon^T - \varepsilon^T X X^T X^{-1} X^T. \quad (1.2.21)$$

Тогда

$$\begin{aligned} M y - X b^T y - X b &= M e^T e = M \varepsilon^T - \varepsilon^T X X^T X^{-1} X^T \times \varepsilon - \\ X X^T X^{-1} X^T \varepsilon &= M \varepsilon^T \varepsilon - M \varepsilon^T X X^T X^{-1} X^T \varepsilon - M \varepsilon^T X X^T X^{-1} X^T \varepsilon + \\ M \varepsilon^T X X^T X^{-1} X^T X X^T X^{-1} X^T \varepsilon. & \end{aligned} \quad (1.2.22)$$

Основываясь на том, что результатом произведения матрицы $X^T X$ на обратную матрицу $X^T X^{-1}$, является единичная матрица, заключительное слагаемое преобразуется в $M \varepsilon^T X X^T X^{-1} X^T \varepsilon$.

Учитывая данное преобразование, получаем

$$M e^T e = M \varepsilon^T \varepsilon - M \varepsilon^T X X^T X^{-1} X^T \varepsilon. \quad (1.2.22)$$

Принимая во внимание то, что скалярное произведение векторов $\varepsilon^T \varepsilon = \sum_{i=1}^n \varepsilon_i^2$ и $M \varepsilon_i^2 = \sigma^2$ при $i = 1, 2, \dots, n$, получим:

$$M \varepsilon^T \varepsilon = \sum_{i=1}^n M \varepsilon_i^2 = n \sigma^2. \quad (1.2.23)$$

В симметричности матрицы $C = X X^T X^{-1} X^T$ легко убедиться, т. е. в том, что $C^T = C$. Достаточно учитывать, что

$$X^T X^{-1 T} = X^T X^T^{-1} = X^T X^{-1}. \quad (1.2.24)$$

На основании симметричности данной матрицы, получаем:

$$M \varepsilon^T C \varepsilon = M \sum_{i,j} c_{ij} \varepsilon_i \varepsilon_j + M \sum_{i=1}^n c_{ii} \varepsilon_i^2, \quad (1.2.25)$$

где $i, j = 1, 2, \dots, n$.

Учитывая то, что элементы симметричной матрицы C не являются случайными и сумма математических ожиданий равно математическому ожиданию суммы, получаем:

$$M \varepsilon^T C \varepsilon = \sum_{i=1}^n c_{ii} M \varepsilon_i^2 + \sum_{i \neq j} c_{ij} M \varepsilon_i \varepsilon_j = \sigma^2 \sum_{i=1}^n c_{ii} = \sigma^2 \text{tr} C. \quad (1.2.26)$$

Следствием независимости значений ε_i , и следа матрицы C , равного сумме диагональных элементов этой матрицы $\sum_{i=1}^n c_{ii} = \text{tr} C$, является выражение $M \varepsilon_i^2 = \sigma^2$ при условии $M \varepsilon_i \varepsilon_j = 0$ где $i \neq j$.

При подстановке выражения вместо C , получаем

$$M \varepsilon^T C \varepsilon = M \varepsilon^T X X^T X^{-1} X^T \varepsilon = \sigma^2 \operatorname{tr} X X^T X^{-1} X^T = \sigma^2 \operatorname{tr} X^T X X^T X^{-1} = \sigma^2 \operatorname{tr} E_{k+1} = \sigma^2 (k+1). \quad (1.2.27)$$

В этом выражении используется особенность следа матрицы $\operatorname{tr} AC = \operatorname{tr} CA$; E_{k+1} — является единичной матрицей размера $(k+1) \times (k+1)$. Сумма диагональных элементов единичной матрицы равна $(k+1)$, т. е. $\operatorname{tr} E_{k+1} = k+1$.

Используя выражения (1.2.23) и (1.2.24), преобразуем формулу (1.2.22), которая будет иметь вид:

$$M(Y - Xb)^T(Y - Xb) = (n - k - 1) \sigma^2, \quad (1.2.28)$$

а оценка несмещенности остаточной дисперсии σ^2 будет иметь вид:

$$s^2 = \frac{1}{n - k - 1} (Y - Xb)^T(Y - Xb), \quad (1.2.29)$$

поскольку $M s^2 = \sigma^2$.

Используя критерий дисперсионного анализа, проведем проверку уравнения регрессии на значимость.

Сделаем предположение, о нормальном распределении вектора

$$\varepsilon \in N_n(\mu; \sigma^2 E_n). \quad (1.2.30)$$

В первую очередь покажем, что

$$Y^T Y = (Y - Xb)^T(Y - Xb) + Xb^T Xb. \quad (1.2.31)$$

Подставляя выражение (1.2.29) в преобразованную правую часть выражения (1.2.31), получаем

$$Y^T Y - b^T X^T(Y - Xb) + b^T X^T Xb = Y^T Y - Y^T Xb - b^T X^T Y + 2b^T X^T Xb = Y^T Y - Y^T Xb - Y^T Xb^T + 2b^T X^T X X^T X^{-1} X^T Y = Y^T Y.$$

Поскольку выражение $Y^T Xb = b_0 \sum_{i=1}^n y_i + \sum_{j=1}^k b_j \sum_{i=1}^n y_i x_{ij}$ является скалярной величиной, то $Y^T Xb = Y^T Xb^T = b^T X^T Y$.

Справедливость выражения (1.2.31) доказана. Рассмотрим каждое слагаемое данного выражения по отдельности.

$$Q_{\text{общ}} = Y^T Y = \sum_{i=1}^n y_i^2 \text{ является суммой квадратов отклонения } y_i \text{ от нуля.}$$

Первое рассматриваемое слагаемое выражения (1.2.31) имеет вид $Q_{\text{ост}} = Y - Xb^T Y - Xb = e^T e = \sum_{i=1}^n e_i^2$ и является сумма квадратов ошибок фактических значений от значений регрессионной модели $Y = Xb$.

Второе рассматриваемое слагаемое $Q_R = Xb^T Xb = \sum_{i=1}^n y^2$ является суммой квадратов отличия от нуля значений, обусловленных регрессионной моделью.

Таким образом получаем, $Q_{\text{общ}} = Q_R + Q_{\text{ост}}$.

Полученное выражение есть ни что иное, как общая вариация разложения на составляющие.

Ранг $Q_{\text{общ}}$ равняется сумме рангов Q_R и $Q_{\text{ост}}$. На основании теоремы Кохрана, значения Q_R и $Q_{\text{ост}}$ являются независимыми. Используя выражение (1.2.29) получаем $M \frac{Q_{\text{ост}}}{n-k-1} \sigma^2$.

Для значения Q_R , математическое ожидание имеет вид:

$$MQ_R = M Xb^T Xb = M b^T X^T Xb. \quad (1.2.32)$$

Используя формулы (1.1.5) и (1.2.7) и принимая в расчет, что $b^T = Y^T X X^T X^{-1}$, получим:

$$MQ_R = M \varepsilon^T - \beta^T X^T X X^T X^{-1} X^T X \beta + \varepsilon = M \varepsilon^T X X^T X^{-1} X^T \varepsilon + \beta^T X^T X \beta = (k+1) \sigma^2 + \beta^T X^T X \beta. \quad (1.2.33)$$

При условии, что $\beta = 0$, где 0 — вектор, состоящий из нулей, то $MQ_R = (k+1)\sigma^2$, и в результате получаем $M \frac{Q_R}{k+1} = \sigma^2$.

Доказано, что для $\beta = 0$, т. е. $\beta_0 = \beta_1 = \dots = \beta_k = 0$, несмещенные и независимые оценки $\frac{Q_{\text{ост}}}{n-k-1}$ и $\frac{Q_R}{k+1}$ являются оценками одной дисперсии σ^2 .

Гипотеза $H_0: \beta = 0$ проверяется выражением

$$F_{\text{набл}} = \frac{\frac{1}{k+1} Q_R}{\frac{1}{n-k-1} Q_{\text{ост}}}. \quad (1.2.33)$$

Выражение (1.2.33) имеет F -распределение с $(k+1)$ и $(n-k-1)$ степенями свободы, если выполняется гипотеза H_0 .

Анализ регрессионного уравнения завершается, когда все коэффициенты регрессионного уравнения равны нулю, т.е. регрессионное уравнение незначимо.[4]

В случае, когда гипотеза $H_0: \beta = 0$ отвергается, интерес представляет исследование значимости отдельных коэффициентов регрессии с последующим вычислением интервальных оценок для этих коэффициентов.

С помощью t -критерия, можно проверить гипотезу $H_0: \beta_j = 0$

$$t_j = \frac{b_j}{s \sqrt{X^T X^{-1}}_{jj}^{1/2}} \quad (1.2.34)$$

где $l = j + 1$, полученное выражение при выполнении гипотезы $H_0: \beta_j = 0$ имеет t -распределение с числом степеней свободы $n - k - 1$.

На рисунке 1.2.1 показано изменение значений y_i , по отношению к y .

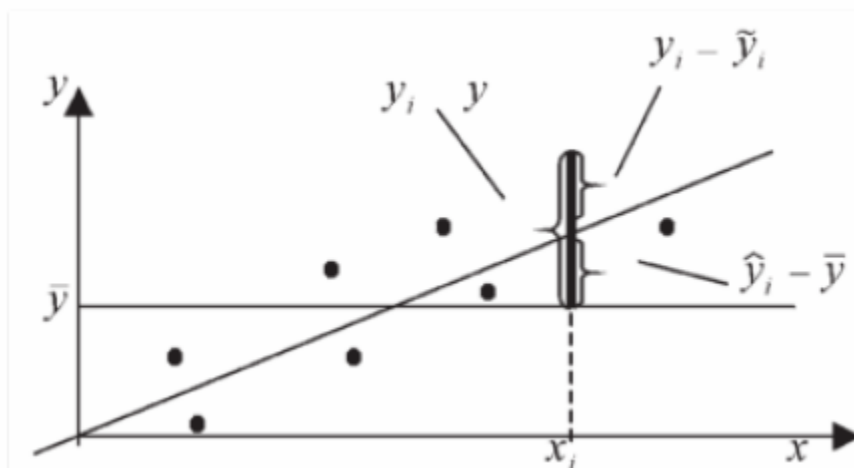


Рисунок 1.2.1 – Анализ изменения значений y_i относительно y

Покажем справедливость разложения

$$\sum_{i=1}^n y_i - y^2 = \sum_{i=1}^n y_i - y^2 + \sum_{i=1}^n y_i - y^2, \quad (1.2.35)$$

где $Q_{\text{общ}} = \sum_{i=1}^n y_i - y^2$ – полное изменение y_i относительно среднего значения y ;

$Q_R = \sum_{i=1}^n y_i - y^2$ – изменение относительно y , описываемая регрессионной моделью;

$Q_{\text{ост}} = \sum_{i=1}^n y_i - y_i^2 = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik}^2$ – изменение остатков регрессионной модели.

Преобразуем выражение полных изменений с целью доказательства справедливости выражения (1.2.35)

$$\begin{aligned} \sum_{i=1}^n y_i - y^2 &= \sum_{i=1}^n y_i - y + y_i - y^2 = \sum_{i=1}^n y_i - y_i^2 + \\ &+ \sum_{i=1}^n y_i - y^2 + 2 \sum_{i=1}^n y_i - y_i y_i - y. \end{aligned} \quad (1.2.36)$$

Введение понятие погрешности $e_i = y_i - y_i$, позволит получить значение удвоенного произведения которое будет равно нулю:

$$\begin{aligned} \sum_{i=1}^n e_i y_i - y &= \sum_{i=1}^n e_i b_0 + b_1 x_{i1} + \dots + b_k x_{ik} - y = b_0 \sum_{i=1}^n e_i + \\ &+ b_1 \sum_{i=1}^n e_i x_{i1} + \dots + b_k \sum_{i=1}^n e_i x_{ik} - y \sum_{i=1}^n e_i = 0, \end{aligned} \quad (1.2.37)$$

поскольку сумма погрешностей $\sum_{i=1}^n e_i$ и $\sum_{i=1}^n e_i x_{ij}$ при $j = 1, 2, \dots, k$ равна нулю. Вектор погрешности будет иметь вид

$$\begin{aligned} \sum_{i=1}^n e_i & \quad 1 \quad 1 \quad \dots \quad 1 \quad e_1 \\ \sum_{i=1}^n e_i x_{i1} &= \begin{matrix} x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \end{matrix} \quad e_2 \\ \dots & \quad \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n e_i x_{ik} & \quad x_{1k} \quad x_{2k} \quad \dots \quad x_{nk} \quad e_n \end{aligned} = X^T e = X^T Y - Xb =$$

$$X^T Y - X^T X X^T X^{-1} X^T Y = X^T Y - X^T Y = 0 = \begin{matrix} 0 \\ 0 \\ \dots \\ 0 \end{matrix}. \quad (1.2.38)$$

Значение удвоенного произведения равняется нулю: $2 \sum_{i=1}^n y_i - y_i y_i - y = 0$, поэтому разложение квадратичной формы справедливо и имеет вид

$$\sum_{i=1}^n y_i - y^2 = \sum_{i=1}^n y_i - y^2 + \sum_{i=1}^n y_i - y_i^2. \quad (1.2.39)$$

$Q_{\text{общ}} \qquad \qquad \qquad Q_R \qquad \qquad \qquad Q_{\text{ост}}$

Количество степеней свободы значения $Q'_{\text{общ}}$ равно сумме степеней значений Q'_R и $Q_{\text{ост}}$: $n - 1 = k + n - k - 1$, поэтому значения Q'_R и $Q_{\text{ост}}$ являются независимыми.[13]

Поделив правую и левую части выражения (1.2.35) на значение $\sum_{i=1}^n y_i - y^2$, получим:

$$1 = \frac{\sum_{i=1}^n y_i - y^2}{\sum_{i=1}^n y_i - y^2} + \frac{\sum_{i=1}^n y_i - y_i^2}{\sum_{i=1}^n y_i - y^2}. \quad (1.2.40)$$

В полученном выражение, первое слагаемое является оценкой множественного коэффициента детерминации R_y^2 , поскольку он описывает

долю изменения y , под воздействием независимых переменных x_1, x_2, \dots, x_k , используемых в регрессионной модели.

Тогда, на основании выражения (1.2.40), получаем:

$$R_y^2 = 1 - \frac{\sum_{i=1}^n y_i - y_i^2}{\sum_{i=1}^n y_i - y^2} \quad (1.2.41)$$

Результатом выражения (1.2.41) является то, что $0 < R_y^2 < 1$. Значение коэффициента детерминации $R_y^2 = 1$ означает, что прогнозируемые значения вычисляются с максимальной точностью, т.е. когда по независимым значениям x_1, x_2, \dots, x_k можно точно определить y , поскольку из условия $\sum_{i=1}^n y_i - y_i^2 = 0$ следует, что $y_i = y_i$, при $i = 1, 2, \dots, n$.

Если $R_y^2 = 0$, то изменение остатков равно полному изменению значения y . Это означает, что зависимость между y и независимыми переменными x_1, x_2, \dots, x_k отсутствует.

Когда между значением y и независимыми переменными x_1, x_2, \dots, x_k существует линейная зависимость, критерии статистики проверки гипотез $H_0: R_y^2 = 0$ и $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ совпадают.

Статистика для проверки гипотезы $H_0: R_y^2 = 0$, преобразованная с учетом выражений (1.2.40) и (1.2.41) будет иметь вид

$$F_{\text{набл}} = \frac{\frac{1}{k} R_y^2}{\frac{1}{n-k-1} (1-R_y^2)} = \frac{\frac{1}{k} \sum_{i=1}^n y_i - y_i^2}{\frac{1}{n-k-1} \sum_{i=1}^n y_i - y_i^2} = \frac{\frac{1}{k} Q'_R}{\frac{1}{n-k-1} Q_{\text{ост}}} \quad (1.2.42)$$

Полученная статистика может быть использована для проверки гипотезы $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$. В случае выполнения гипотезы H_0 , данная статистика будет иметь F -распределение со значениями степеней свободы равными $v_1 = k$ для числителя и $v_2 = n - k - 1$ для знаменателя.

Если полученная статистика является больше критического значения $F_{\text{кр}} \alpha; v_1; v_2$, с вероятностью погрешности α отвергается гипотеза H_0 . Следовательно $\beta' = \beta_1, \beta_2, \dots, \beta_k^T \neq 0$, что означает неравенство нулю хотя бы одного коэффициента регрессии.

В таком случае проводится оценка значимости отдельных регрессионных коэффициентов.

Предположим, что вектор отклонения ε обладает n -мерным нормальным распределением. Тогда вектор значений зависимых переменных Y в выражение (1.1.5) также обладает нормальным распределением. Следовательно элементы y_i являются независимыми, на основании их взаимной некоррелированности.

Оценка b_j $j = 0, 1, 2, \dots, k$ с математическим ожиданием равным β_j и дисперсией $D b_j = \sigma^2 X^T X^{-1} u$, где $l = j + 1$, на основании выражения (1.2.7) обладает нормальным законом распределения.

Исходя из этого, выражение $z = \frac{b_j - \beta_j}{\sigma \sqrt{X^T X^{-1} u}}^{1/2}$ обладает нормированным, нормальный закон распределения.

Статистика, опираясь на формулу (1.2.29), имеет вид

$$u^2 = \frac{ns^2}{\sigma^2} \quad (1.2.43)$$

обладает распределение χ^2 с количеством степеней свободы равным $(n - k - 1)$.

Рассмотрим случайная величина, имеющую вид

$$t = \frac{z}{u} \sqrt{v} \quad (1.2.44)$$

Если значение z стоящее в числителе, обладает нормальным нормированным распределением $z \in N(0, 1)$, и значение u^2 стоящее в знаменателе обладает распределением χ^2 с v степенями свободы $u^2 \in \chi^2(v)$, учитывая независимость значений z и u , то случайная величина t обладает распределением Стьюдента с v степенями свободы.

Заменяя значения z и u в выражение (1.2.44), получим выборочную характеристику, которая имеет вид

$$t_j = \frac{b_j - \beta_j}{s \sqrt{X^T X^{-1} u}}^{1/2} \quad (1.2.45)$$

Полученная характеристика обладает распределением Стьюдента с количеством степеней свободы равным $(n - k - 1)$. На основе выражения (1.2.34), вычислим с надежностью γ интервальную оценку для

$$\beta_j \in \varepsilon b_j \pm t_\gamma s \sqrt{X^T X^{-1}}_{jj}^{1/2} \quad (1.2.46)$$

В точке, которая определяется вектором X^0 размера $(k + 1)$, вычислим интервальную оценку y :

$$X^0 = [1, x_1^0, x_2^0, \dots, x_k^0]^T. \quad (1.2.47)$$

Откуда оценка несмещенности равна $y = (X^0)^T b$.

Действительно,

$$My = (X^0)^T Mb = (X^0)^T \beta = y, \quad (1.2.48)$$

$$Dy = (X^0)^T b \text{ cov} b (X^0)^T = \sigma^2 (X^0)^T X^T X^{-1} X^0, \quad (1.2.49)$$

где b — матрица ковариаций оценок b .

Статистика для нахождения доверительного интервала y , имеет вид

$$t = \frac{y - \hat{y}}{s \sqrt{(X^0)^T X^T X^{-1} X^0}}, \quad (1.2.50)$$

Полученная статистика обладает распределением Стьюдента со степенью свободы $(n - k - 1)$. Преобразуя неравенство $P \{ |T| < t_\gamma \} = \gamma$, вычислим интервальную оценку для y :

$$y_{X^0} \in \hat{y}_{X^0} \pm t_\gamma s \sqrt{(X^0)^T X^T X^{-1} X^0} \quad (1.2.51)$$

В полученном выражении, значение t_γ находится по таблице распределения Стьюдента.[17]

Результатом проведенных вычислений, является оценка интервала предсказания y_{n+1} с надежностью γ , имеет вид:

$$y_{n+1} \in (X^0)^T b \pm t_\gamma s \sqrt{(X^0)^T X^T X^{-1} X^0 + 1}. \quad (1.2.52)$$

В первой главе проанализированы предпосылки, и особенности построения регрессионной модели. Выявлены необходимые требования и условия, выдвигаемые перед составляющими элементами регрессионной модели, выполнение которых делаем модель корректной, а получаемые ей значения точными.

Проанализировано применение метода наименьших квадратов для оценки параметров регрессионной модели. Рассмотрены гипотезы, выдвигаемые для параметра β регрессионной модели. Вводится понятие погрешности вычислений. Используются разные статистики, позволяющие оценить особенности регрессионных моделей. Такие модели достаточно точно описывают разнообразные явления, но возникает необходимость адаптации для новых данных.

Глава 2 Адаптация регрессионной модели методом Брауна

2.1 Применение метода Брауна для адаптации регрессионной модели

Для адаптации регрессионной модели будем использовать метод краткосрочного прогнозирования Брауна. Используя временной ряд Y_T , без особо выраженной тенденции, требуется получить прогнозируемое значение Y_{T+1} . Для решения этой задачи удобно использовать среднюю арифметическую:

$$Y_{T+1} = Y = \frac{1}{T} \sum_{t=1}^T Y_t \quad (2.1.1)$$

Представленная средняя арифметическая характеризует средний уровень ряда. Отклонение значений от среднего уровня вызвано рядом причин.

Для стационарного процесса, который имеет нормальное распределение вероятностей, данная процедура не вызывает никаких сомнений и возражений. Получение наилучшей оценки математического ожидания процесса методом средней арифметической, можно только в том случае, когда он обладает нормальным распределением и стационарен. Прогнозирование методом средней арифметической будет неэффективно, если хотя бы одно из условий не выполняется.

Полученное отклонение, с целью понимания происходящих процессов, для эволюционных процессов играет большую роль, чем прошлые отклонения. Тем более, для прогноза более важными является текущие значения, чем старые данные. Например, с целью определить стоимость акции компании на завтра, куда важнее знать их текущую цену, чем значение годичной давности.

Выражение (2.1.1) в расширенной форме, имеет вид:

$$Y_{T+1} = \frac{1}{T} Y_T + \frac{1}{T} Y_{T-1} + \dots + \frac{1}{T} Y_t + \dots + \frac{1}{T} Y_1 \quad (2.1.2)$$

Или:

$$Y_{T+1} = v_T Y_T + v_{T-1} Y_{T-1} + \dots + v_t Y_t + \dots + v_1 Y_1 \quad (2.1.3)$$

Значение v_t является весом t -го наблюдения, причём легко удостовериться в том, что:

$$\sum_{t=1}^T v_t = 1 \quad (2.1.4)$$

Естественно, текущую информацию требуется учитывать в большей степени, чем прошлую. В математическом виде, данное требование выглядит следующим образом:

$$v_T > v_{T-1} > \dots > v_t > \dots > v_1 \quad (2.1.5)$$

Для получения формулы взвешенной средней, требуется выполнять условие (2.1.4) и использовать вычисленные весовые коэффициенты в выражение (2.1.3). Рядов, у которого весовые коэффициенты уменьшаются, а сумма элементов равна единице, существует огромное множество. Один из таких рядов имеет вид:

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \dots + \frac{1}{n \cdot n+1} + \dots = \sum_{n=1}^{\infty} \frac{1}{n \cdot n+1} \quad (2.1.6)$$

Преобразовать ряд таким образом, чтобы его элементы в сумме давали единицу, является возможным, если он сходится. Ряд вида

$$\frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{n!} + \dots \quad (2.1.7)$$

сходится к значению $e - 1$. После преобразования, сумма элементов будет равна единице, а сам ряд будет иметь вид:

$$\frac{1}{1! \cdot e-1} + \frac{1}{2! \cdot e-2} + \dots + \frac{1}{n! \cdot e-1} + \dots \quad (2.1.8)$$

Все прогнозируемые процессы различны. Для каждой модели, описывающей процесс, требуется индивидуальный подход при нахождении весовых коэффициентов взвешенной средней. Нахождение лучшего ряда для каждого процесса, является невозможным на практике.

Для решения проблемы нахождения взвешенной средней используемого ряда, требуется какая-либо процедура, которая по одной или нескольким характеристикам вычисляла требуемое значение.

Это является возможным, если весовые коэффициенты α имеют показательный характер:

$$S = \alpha + \alpha(1 - \alpha) + \alpha(1 - \alpha)^2 + \dots \quad (2.1.9)$$

Выражение (2.1.9) состоит из одного элемента α . Изменение элемента α влечет изменение модели прогнозирования. Модели, полученные таким образом, могут быть использованы для процессов, с разным характером изменения. Такой ряд называется экспоненциальным.[5]

С помощью показательного взвешенного ряда весовых коэффициентов легко посчитать среднее взвешенное для значения Y в момент времени T . Для момента времени $(T + 1)$, полученное значение будет являться прогнозной моделью. Обозначим полученное прогнозируемое значение Y_{T+1} .

Используем весовые коэффициенты выражения (2.1.9) с целью подстановки в (2.1.3). В результате такого преобразования, получим выражение:

$$Y_{T+1} = \alpha Y_T + \alpha(1 - \alpha) Y_{T-1} + \alpha(1 - \alpha)^2 Y_{T-2} + \dots \quad (2.1.10)$$

Если в выражение (2.1.10) вынести за скобки общий множитель $(1 - \alpha)$ то, получим выражение вида:

$$Y_{T+1} = \alpha Y_T + (1 - \alpha) [\alpha Y_{T-1} + \alpha(1 - \alpha) Y_{T-2} + \dots] \quad (2.1.11)$$

Значение суммы в квадратных скобках является значением взвешенной средней, полученном на основе предыдущих значений ряда. Учитывая это, получаем следующее выражение:

$$Y_{T+1} = \alpha Y_T + (1 - \alpha) Y_T \quad (2.1.12)$$

Р. Браун предложил формулу (2.1.12), которая оказалась удобной для вычисления и получила своё название по имени автора. Он называется методом экспоненциальной сглаживания.

Этот метод применим для рядов весовых коэффициентов, сумма элементов которого равна единице, и которые сходятся. При невыполнение поставленных требований, метод Брауна теряет смысл.

Р Брауна предложил использовать ряд весовых коэффициентов (2.1.9), который имеет смысл бесконечной геометрической прогрессией. Такой ряд

сходится к единице, когда все его члены выполняют условие: значение модуля члена ряда должно быть меньше единицы.

Это условие для используемого ряда, будет иметь вид:

$$1 - \alpha < 1 \quad (2.1.13)$$

Изменение постоянной сглаживания, на основании условия (2.1.13), происходит на интервале:

$$0 < \alpha < 2 \quad (2.1.14)$$

На основании теоремы Лейбница, если значение постоянной сглаживания больше единицы, то рассматриваемый ряд весовых коэффициентов сходится, и сумма элементов равна единице. Сама теорем говорит о том, что ряд

$$q_1 - q_2 + q_3 - q_4 + \dots + (-1)^{n-1} q_n + \dots, \quad (2.1.15)$$

сходится, если все элементы больше нуля и последовательность элементов q_n невозрастающая и истинно выражение

$$\lim_{n \rightarrow \infty} q_n = 0 \quad (2.1.16)$$

Используемый ряд на интервале $1 < \alpha < 2$, будет иметь следующий вид. Ряд значений

$$\alpha - \alpha 1 - \alpha + \alpha 1 - \alpha^2 - \alpha 1 - \alpha^3 + \dots \quad (2.1.17)$$

будет состоять только из положительных членов. Для того чтобы этот ряд сходился, необходимо выполнять условие (1.2.11). Это условие для используемого ряда выглядит следующим образом:

$$\lim_{n \rightarrow \infty} \alpha 1 - \alpha^n = 0 \quad (2.1.18)$$

Условие (2.1.18) всегда будет выполняться, потому что на интервале $0 < \alpha < 2$, значение модуля меньше единицы.

Вычисления постоянной сглаживания методом Брауна возможно, как при ограничение:

$$0 < \alpha < 1, \quad (2.1.19)$$

которое называется «классическим», так и при ограничение:

$$1 \leq \alpha < 2 \quad (2.1.20)$$

которое называется «запредельным множеством».

Значение α имеет название постоянная сглаживания. Данное название значение α получила, потому что, как и любая другая взвешенная средняя, данная модель усредняет прошлые значения. Это достигается за счёт сглаживания «провалов» и «пиков», как показано на рисунке 2.1.1.

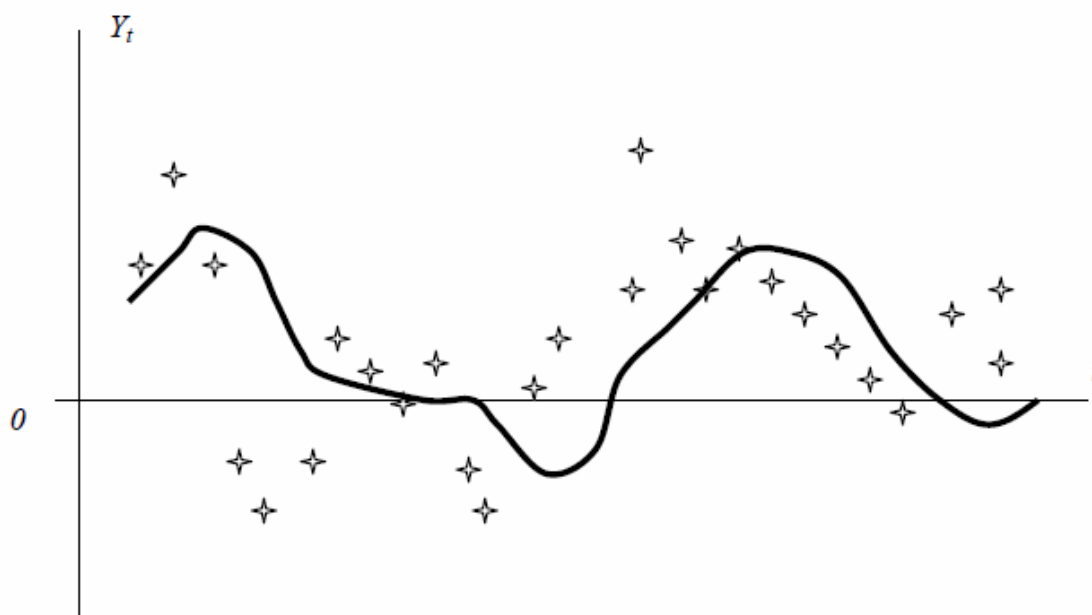


Рисунок 2.1.1 – Сглаживание тенденции ряда методом Брауна

Результат аппроксимации методом Брауна описывается значением постоянной сглаживания. Пусть используемый параметр α , лежащий на интервале от нуля до единицы (2.1.19), станет равным нулю.

Подставляя этот параметр в (2.1.11), получаем:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T = Y_T. \quad (2.1.21)$$

В результате, текущая информация не влияет на результаты модели. Такая модель не будет являться адаптивной.

Теперь выберем значение параметра α равным другому граничному значению, а именно – единице, и подставим в модель Брауна. Получим:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T = Y_T. \quad (2.1.22)$$

Результат показывает, что при таком значении α , прошлые значения не оказывают влияние на модель Брауна, что позволяет ей полностью адаптироваться к текущей информации.

То, как сильно происходит адаптация методом Брауна определяется параметром α . На рисунке 2.1.2 показана адаптация методом Брауна с двумя разными значениями параметра α . Один из графиков отображает адаптацию при $\alpha = 0,3$, а другой при $\alpha = 0,7$.

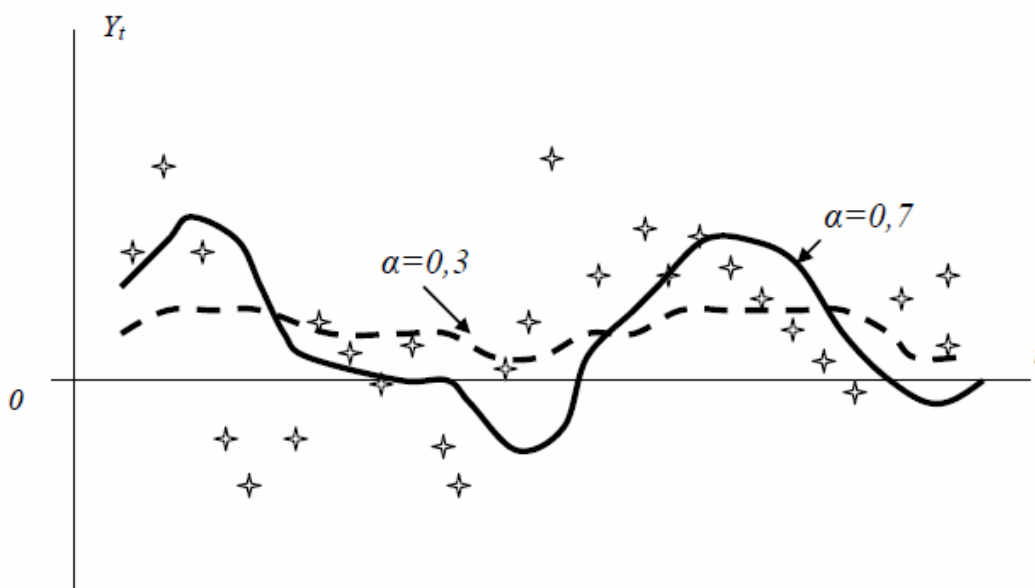


Рисунок 2.1.2 – Адаптация методом Брауна с разным значением α

Граничные случаи метода Брауна, определенные краевыми условиями (2.1.15) заключают в себе математический смысл. Принимая во внимание то, что граничные случаи метода Брауна удовлетворяют требованию, при котором значение параметра α больше либо равно единице, обозначим новый параметр, который будет иметь вид:

$$\beta = \alpha - 1, \quad 0 \leq \beta < 1. \quad (2.1.23)$$

Метод Брауна при использовании нового параметра β , примет следующий вид:

$$Y_{T+1} = Y_T + \beta (Y_T - Y_{T-1}). \quad (2.1.24)$$

Значение, полученное в скобках, является ошибкой аппроксимации

$$\varepsilon_T = Y_T - \hat{Y}_T, \quad (2.1.25)$$

Используя данное обозначение, модель (2.1.24) принимает вид:

$$Y_{T+1} = Y_T + \beta \varepsilon_T. \quad (2.1.26)$$

Полученное выражение (2.1.26) позволяет адаптировать методом Брауна, когда параметр постоянной сглаживания принимает граничные значения.

Модель Брауна становится полностью адаптивной к новой информации на основании формулы (2.1.26), поскольку в ней эта информация оказывает полное влияние на модель, а первое слагаемое в формуле представляет собой текущее наблюдение Y_T .

Метод Брауна учитывает отклонение вычисляемых значений от фактических значений ε_T , т.е. метод учитывает ошибку аппроксимации. Когда значение параметра β равно нулю, метод Брауна перестает учитывать ошибку аппроксимации, если значение параметра β равно единице, метод Брауна, на основании формулы (2.1.26) учитывает ошибку аппроксимации полностью. Это означает, что метод Брауна является невосприимчивым к ошибке аппроксимации. Нахождение значения параметра β на интервале от нуля до единицы, говорит о том, что метод Брауна учитывает ошибку аппроксимации с разной силой.[6]

Способ описания метода Брауна и точность вычисляемых значений, полученных этим методом, определяется значением параметра α . Появляется необходимость нахождения такого значения параметра α , чтобы точность получаемых прогнозов была наилучшей.

Для нахождения оптимального значения α используем метод ретропрогноза. Используя метод Брауна, описываем ряд значений Y_T . Для этого, задаем некоторое значение параметра α_1 и находим погрешность ретропрогноза:

$$\varepsilon_{t1} = Y_t - Y_{t1} \quad (2.1.27)$$

Полученная таким образом ошибка ретропрогноза плохо описывает поведение модели, построенной методом Брауна.

Точность получаемых значений при помощи метода Брауна, с учетом заданного параметра α , может быть описана некоторой величиной, такой как дисперсия, среднее отклонение или схожая характеристика.

Определение используемой характеристики обосновывается поставленными целями исследования. Одним из возможных выборов критерия является минимум дисперсии:

$$\sigma_1^2 = \frac{1}{T} \sum_t \varepsilon_{t1}^2 = \sum_t Y_t - Y_{t1}^2 \quad (2.1.28)$$

Вычислив для значения параметра α_1 дисперсию метода Брауна касательно начального ряда, выбирается другое значение параметра $\alpha_2 \neq \alpha_1$, находящееся на промежутке (2.1.14). Используя новый параметр α_2 находим ошибки ретропрогноза:

$$\varepsilon_{t2} = Y_t - Y_{t2} \quad (2.1.29)$$

Затем для параметра α_2 вычисляем дисперсию:

$$\sigma_2^2 = \frac{1}{T} \sum_t \varepsilon_{t2}^2 = \sum_t Y_t - Y_{t2}^2 \quad (2.1.30)$$

Изменяя значение параметра α , и для вычисляя для каждого дисперсию, получаем α_i, σ_i^2 . α_i, σ_i^2 является рядом различных значений параметров α и соответствующих им дисперсий. На рисунке 2.1.3 показано то, как находится наилучшее значение параметра α по значению дисперсии. Наилучшим параметр α будет тогда, когда дисперсия минимальна.

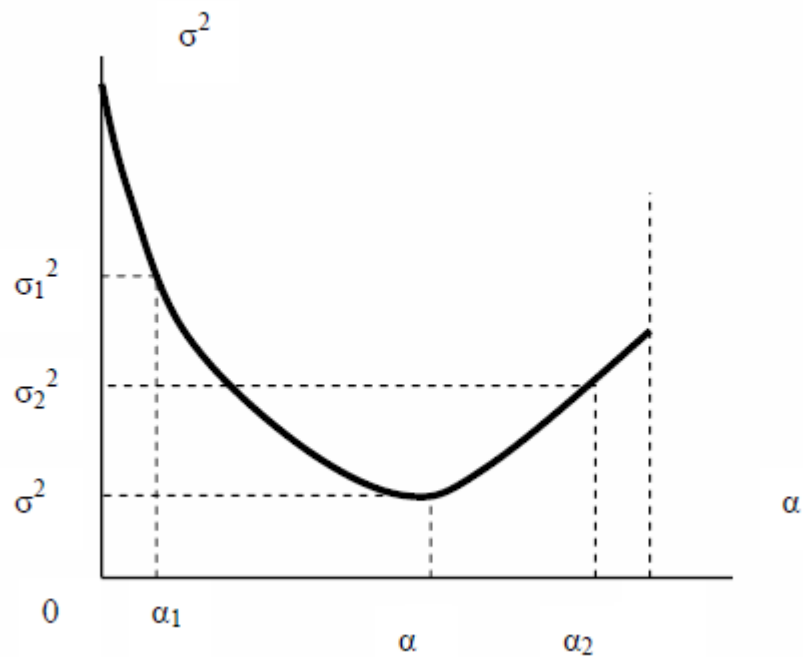


Рисунок 2.1.3 – Зависимость дисперсия ошибки от значения постоянного сглаживания

Поэтому, задача отыскания оптимального значения постоянной сглаживания α сводится к элементарному нахождению минимума данной функции. Для решения этой задачи можно воспользоваться методом перебора, поскольку параметр постоянного сглаживания имеет ограничения, но лучшим способом нахождения минимума неявно заданной функции является использование численных методов. Характер зависимости значения постоянной сглаживания и дисперсии ошибки имеет вид, показанный на рисунке 2.1.3. Однако бывают ситуации, когда представленная зависимость имеет не один, а несколько локальных минимумов. Ситуация, представленная на рисунке 2.1.4 является крайне редкой, но может встретиться на практике.

При использовании численных методов, когда функция имеет несколько минимумов, исследователь оказывается в сложной ситуации. Если значение параметра $\alpha = 0.01$, то с его увеличением, находится значение параметра α_1 , который является локальным минимумом. Данный параметр изображен на рисунке 2.1.4. Увеличение значения параметра α в точке α_1 , повлечет рост

погрешности дисперсии. Значение параметра α в точке α_1 можно принять наилучшее значение параметра α , но оно не будет являться единственным.

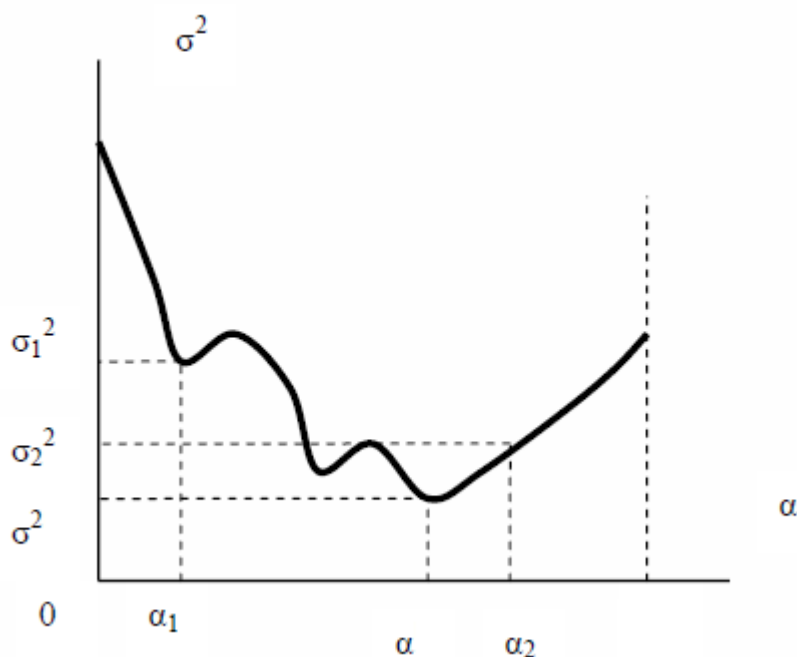


Рисунок 2.1.4 – Зависимость дисперсия ошибки от значения постоянной сглаживания с несколькими локальными минимумами

Для избегания подобной ситуации, следует увеличивать значение величины постоянного сглаживания на величину равную 0.1, что позволит получить соответствующие значения дисперсии ошибки. На основании вычисленных дисперсий, находится интервал оптимального значения, на котором происходит вычисление значения параметра α при помощи численных методов.

Ситуации, когда при значние параметра $\alpha < 0$, находится минимум дисперсии, что в свою очередь противоречит выражению (2.1.14). Это происходит при случайном или хаотичном поведение модели, построенной методом Брауна. Для предотвращения подобного исхода, требуется воспользоваться иным начальным значением Y_0 или методом прогнозирования.

Различные критерии выбора оценки оказывают разное воздействие на вычисление лучшего значения параметра α . Найдем наилучшее значение постоянной сглаживания по следующим двум критериям:

1) минимум дисперсии ошибки аппроксимации:

$$\frac{1}{T} \sum_{t=1}^T Y_t - \hat{Y}_t^2 \rightarrow \min \quad (2.1.31)$$

2) минимум суммы абсолютных ошибок аппроксимации:

$$\sum_{t=1}^T |Y_t - \hat{Y}_t| \rightarrow \min \quad (2.1.32)$$

Вычисления показали, что большие ряды в обоих случаях отбора дают схожие результаты при нахождении наилучшего значения постоянной сглаживания. Величина этого значения в зависимости от того, какой из двух критериев был выбран – (2.1.31) или (2.1.32), изменяется незначительно. В таком случае проблема выбора критерия нахождения значения постоянного сглаживания не стоит – выбирается любой критерий, в зависимости от удобства их использования для поиска оптимального значения α в том или ином случае. Оценка значения параметра α полученная формулами (2.1.32) и (2.1.31), не является одной и той же оценкой. Различие оценок, вычисленных по формулам (2.1.32) и (2.1.31) в большинстве случаев незначительно.

Нахождение лучшего значения параметра α , при использовании малой выборки, значительно влияет на результат прогнозирования методом Брауна. На малой выборке, значение параметра α , получаемое разными критериями, существенно различается. Нахождение значения параметра α на условном примере показало то, что первый критерий (2.1.31) даёт наилучшую оценку α , а второй критерий (2.1.32) даёт менее точную оценку. Параметр α , используемый в методе Брауна, при прогнозировании оказывает сильное влияние на результат.

На основании этого можно сделать вывод о том, что при значении параметра α находящегося на классическом интервале, то метод Брауна адаптивна. Если значение параметра α лежит не на интервале, то метод Брауна не только адаптивен, но еще и самообучаем. Начальный ряд определяет лучшее значение параметра α . Различие рядов, лучшее значение параметра α первого

ряда, находится на классическом интервале $0 < \alpha < 1$, и второго ряда, лучшее значение параметра α находится в запредельном множестве $1 < \alpha < 2$ отобразится проведением модельного эксперимента на условных примерах.

Результатом проведения эксперимента было выявлено, что значения параметра α , полученное разными критериями, различаются незначительно, но логарифмический критерий показал значительное отличие между полученными значениями параметра α .

Значение параметра α , находящееся в запредельном множестве $1 < \alpha < 2$, чаще является лучшим параметром. Исключением являются ряды, построенные синусоидой, параболой и экспонентой. Сумма элементов таких рядов, в графическом виде представляют неубывающую и невозрастающую последовательность значений. Значение параметра α является лучшим на классическом интервале.

Подведем итог проведенного эксперимента, относительно метода Брауна в запредельного множества. Если во время нахождения наилучшего значения параметр α находится на классическом интервале $0 < \alpha < 1$, то адаптация методом Брауна является достаточно результативной. Если лучшее значение параметра α лежит на запредельном множестве $1 < \alpha < 2$, то использование метода Брауна для адаптации модели получит неудовлетворительный результат. Существуют две причины появления данного исхода моделирования.

Первой причиной является выход процесса за пределы обычной динамики. У исследуемого процесса возникает направленное его развития. С целью описания такого процесса, следует использовать одну из регрессионных моделей.

Если исследуемый процесс находится между эволюционной динамикой и хаотичной динамикой, то это является второй причиной. Описание подобного процесса математической моделью невозможно. Адаптация такого процесса возможна методом Брауна, с значением параметра α находящегося на запредельном интервале $1 < \alpha < 2$.

При возникновении первой причины, требуется использовать модель, описывающую изменение регрессионного процесса оптимальным образом и использовать подходящий метод Брауна, который позволит получить точный результат. В случаях, когда регрессионной моделью невозможно описать исследуемый процесс, то не возможно использовать альтернативные адаптационные методы вместо метода Брауна с таким предельным значением параметра α .

Метод Брауна, или, как его принято называть, «метод экспоненциального сглаживания», является очень удобным в практическом применении с целью краткосрочного прогнозирования нестационарных процессов, как и необратимых процессов. Функциональность этого метода определяется и разнообразием интерпретаций его свойств.

Метод Брауна имеет вид:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T \quad (2.1.33)$$

На основании формулы (2.1.33) очевидны свойства метода Брауна. Когда значение параметра α равно нулю, метод Брауна является неадаптивным. Если значение параметра α равно единице, то метод Брауна адаптируется к новым данным в полном объеме.

Именно такой вид метода Брауна является наиболее используемым, и именно в этом виде отображает следующую интерпретацию значения постоянной сглаживания: α представляет собой некоторую среднюю взвешенную, используемую для построения прогнозного значения. Вычисляемое значение методом Брауна получается на основе двух составляющих. Первая составляющая представляет собой фактическое значение в момент времени t . Второй составляющей является полученное методом Брауна значение в тот же момент времени t . При использовании такого метода Брауна, значение параметра α находится на интервале $0 < \alpha < 1$. Чаще всего используется именно этот метод.

Применение метода Брауна для адаптации и вычисления прогнозируемого значения показано на рисунке 2.1.5. Результатом вычисления методом Брауна на основании значения I и II, является значение III, которое станет прогнозом на следующий момент времени, и будет обозначаться значением IV. Используя метод Брауна для значений IV и V, получаем значение VI, которое будет прогнозным значением VII. Вычисления производятся до тех пор, пока не получится искомое значение. При этом α в данном виде отображает влияние весов между фактом и прогнозом.

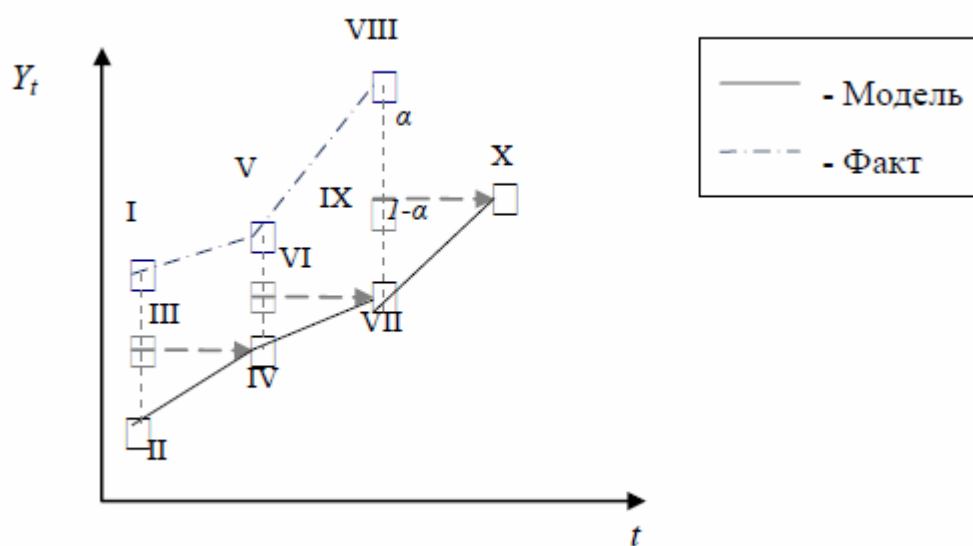


Рисунок 2.1.5 – Графический вид построения прогноза на основе модели (2.1.33)

Однако в данной ситуации такой вид модели накладывает ряд ограничений.

При раскрытии скобок второго произведения правой части модели (2.1.33) и после преобразования, метод Брауна будет иметь следующий вид:

$$Y_{T+1} = Y_T + \alpha Y_T - Y_T \quad (2.1.34)$$

Такой вид более наглядно показывает адаптивность метода Брауна: вычисляемое значение Y_{T+1} получается на основании прошлого вычисленного значения, а параметр α является коэффициентом адаптации для новых данных. Качество адаптации методом Брауна изменяется в зависимости от значения

параметра α : если значение параметра α равно $\alpha = 0.3$, то адаптация методом Брауна незначительна, если значение параметра α большое, и допустим равно $\alpha = 1.7$, то в этом случае модель, адаптируемая методом Брауна, быстро адаптируется под новые данные.

В правой части выражения (2.1.34), значение получаемое в скобках представляет собой текущую ошибку аппроксимации, которая имеет вид:

$$Y_T - \hat{Y}_T = \varepsilon_T, \quad (2.1.35)$$

тогда модель Брауна будет иметь вид:

$$Y_{T+1} = Y_T + \alpha \varepsilon_T. \quad (2.1.36)$$

Первое слагаемое метода Брауна является значением полученным методом в предыдущий момент времени, оно несет в себе данные о прошлых полученных элементах используемого ряда. Второе слагаемое, является результатом произведения значения параметра α и значения текущей ошибки аппроксимации, следовательно метод Брауна учитывает значение ошибки.

Адаптация методом Брауна описывается следующим образом. Когда значение, полученное методом Брауна, меньше фактического значения, то значение ошибки отрицательное. Следовательно, следующее вычисляемое методом значение будет уменьшаться на величину ошибки, умноженную на значение параметра α . Если значение ошибки положительное, то значение получаемое методом Брауна увеличивается на величину ошибки, умноженную на значение параметра α .

Адаптация методом Брауна заключается в приближение полученных значений к фактическим. На рисунке 2.1.6 показана работа этого метода.

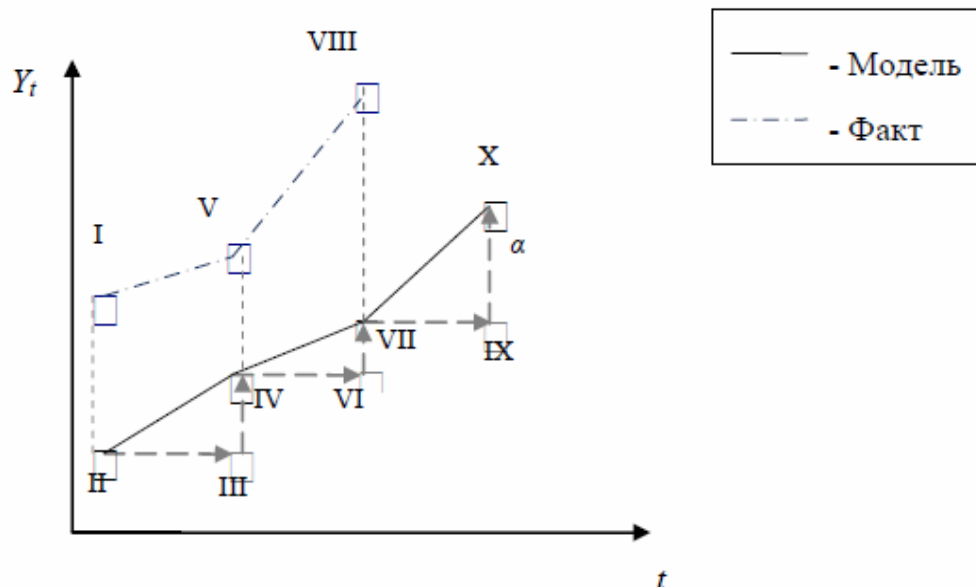


Рисунок 2.1.6 – Адаптация методом Брауна по формуле (2.1.34)

Полученное методом Брауна значение II, становится прогнозируемым значением в следующий момент времени и обозначается значением III, которое изменяется на величину ошибки аппроксимации между значениями I и II. Значение, полученное методом Брауна на основе значения III, становится прогнозируемым значением в следующий момент времени и обозначается IV. Это значение станет основой для вычисления следующего прогнозного значения VI и так далее.

При раскрытии скобок в формуле (2.1.33), модель Брауна принимает другой вид:

$$Y_{T+1} = \alpha Y_T + (1 - \alpha) Y_T = \alpha Y_T + Y_T - \alpha Y_T, \quad (2.1.37)$$

после математических преобразований в правой части в (2.1.37), получаем следующее выражение:

$$Y_{T+1} = \alpha Y_T + Y_T - \alpha Y_T + Y_T - Y_T. \quad (2.1.38)$$

При выносе из скобок выражение $\alpha - 1$ в (2.1.38), то получается новая форма, которая математически тождественна формам (2.1.33) и (2.1.34):

$$Y_{T+1} = Y_T + (\alpha - 1) (Y_T - Y_T). \quad (2.1.39)$$

Полученная модель получила немного другой смысл, благодаря такому представлению. Для удобства использования, требуется вместо $\alpha - 1$ ввести коэффициент $\beta = \alpha - 1$ и использовать формулу (2.1.35), тогда формула (2.1.37) имеет следующий вид:

$$Y_{T+1} = Y_T + \beta \varepsilon_T. \quad (2.1.40)$$

Графический вид метода прогнозирования в соответствии с (2.1.40) представлен на рисунке 2.1.7.

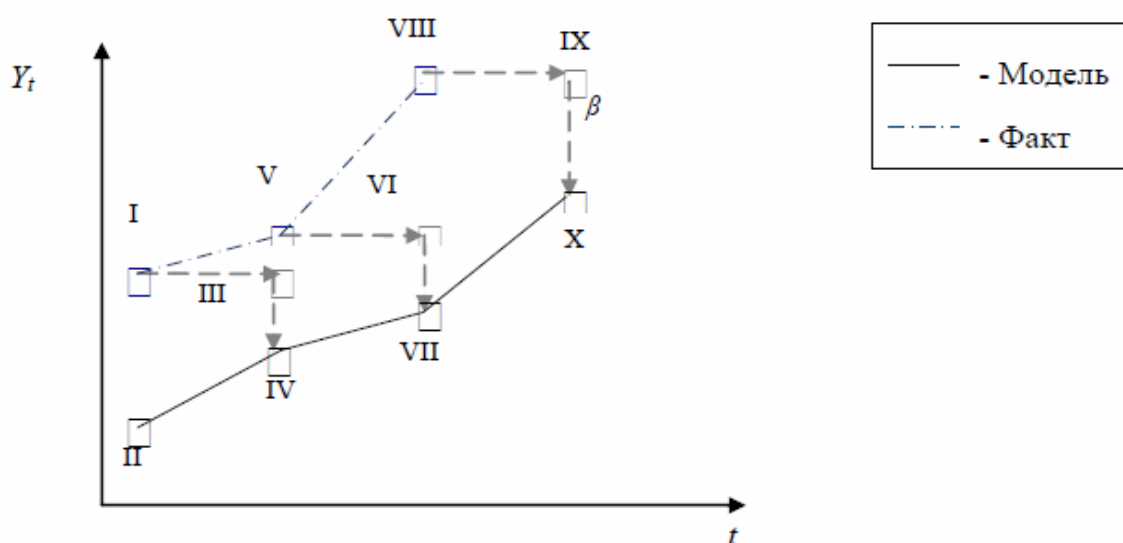


Рисунок 2.1.7 – Адаптация модели методом Брауна по формуле (2.1.40)

Данный метод адаптации напоминает метод представленный на рисунке 2.1.6, однако у него присутствуют некоторые отличия. Такая модель изначально строится на основе предыдущего фактического значения, а не на основе предыдущего вычисленного, т.е. значение в точке I переходит на следующее наблюдение в точку III. Полученное значение изменяется на значение ошибки фактической величины значения I от вычисляемого значения II, полученном методом Брауна в предыдущем моменте времени, которое соответствует значению параметра β .

При нахождении значения параметра α на классическом интервале от нуля до единицы, значение параметра $\beta = \alpha - 1$ находится на интервале $(-1; 0)$. При таком значении параметра β , значение текущей ошибки

положительно. Прогнозное значение, получаемое методом Брауна при положительной ошибке, увеличивается в сравнение с полученным значением в предыдущий момент времени, на значение ε_T . Использование метода Брауна для получения прогнозного значения, при значении параметра α находящегося на классическом интервале $0 < \alpha < 1$, приводит к инерционности. Это означает, что значение, получаемое методом Брауна, никогда не станет равным текущему значению.

При нахождение значения параметра α на запредельном интервале (2.1.20), значение параметра β находится в интервале $(0; 1)$. Прогнозное значение, получаемое методом Брауна при найденной ошибке, уменьшается в сравнение с полученным значением в предыдущий момент времени, на значение ε_T . Следствием этого, является превышение требуемой точности вычисления. Для предотвращения подобной ситуации, значение получаемое методом Брауна изменяется на величину ошибки, которая подкорректирована значением параметра β .

Метод Брауна, при значении параметра α , находящегося на классическом интервале, является инерционным, но когда параметр α находится на запредельном интервале, метод Брауна не обладает этим свойством. Необратимые процессы были поделены на два вида: первый вид – это хаотичные процессы, а второй вид – это эволюционные процессы. На основании этих видов, можно определить особенности метода Брауна.

На классическом интервале, когда значение параметра α находится на интервале $0 < \alpha < 1$ (2.1.19), используемый метод Брауна описывает изменяющиеся процессы, обладающие свойством инерционности. Такие процессы называются эволюционными.

На запредельном интервале (2.1.20), когда значение параметра α находится на интервале $1 < \alpha < 2$, используемый метод Брауна описывает процессы, не имеющие свойства инерционности. Такие процессы называются хаотическими.

Метод Брауна обладает ещё одной особенностью, которая не упоминалась ранее, для того чтобы не нарушать целостность изложения. Эта особенность заключается в том, что необходимо задавать начальное значение метода. Действительно, метод Брауна требует для последующий вычисления в первую очередь, на основании начального значения ряда Y_1 , получить значение метода во второй момент времени:

$$Y_2 = \alpha Y_1 + 1 - \alpha Y_1. \quad (2.1.41)$$

Первое значение данного выражения при заданном значении параметра α вычисляется достаточно просто, так как имеется фактическое значение Y_1 , а для получения второго значения выражения требуется значение, полученное методом Брауна в предыдущий момента времени, то есть Y_1 , а значение данной величины неизвестно. Очевидно, что без наличия первого вычисленного значения показателя модель бесполезна. Поэтому, метод Брауна необходимо изменить. Данным изменением будет добавление условия указания начального значения для метода Брауна. С учетом данного условия, метод Брауна примет вид:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T \text{ при заданном } Y_1 \quad (2.1.42)$$

Если первоначальное значение задано таким образом, что оно сильно отличается от фактического значения, то адаптация метод Брауна с таким начальным значением будет неэффективна. Поэтому требуется вычислять первоначальное значение на основании прогнозируемых значений.

В первую очередь требуется определить то, с какой силой первоначальное значение метода Брауна оказывает на точность вычисляемых методом прогнозных значений.

Весовой коэффициент первого значения, полученный методом Брауна на k шаге вычислений, основываясь на формуле (2.1.10), будет иметь вид $\alpha 1 - \alpha^k$. Вычисленный методом Брауна весовой коэффициент, на малом числе наблюдений, может получиться большим, следствием чего, начальное значение окажет значительное влияние на получаемое методом Брауна прогнозное

значение. На весовой коэффициент так же оказывает влияние значение параметра α , заданное для начального значения метода. Возьмем значение параметра α , находящееся на классическом интервале, близким единице, равное 0,8. Весовой коэффициент, полученный методом Брауна на пятом шаге вычислений, будет равен значению $0,8(1 - 0,8)^5 = 0,000256$. Влияние полученного значения на точность прогнозируемой методом Брауна величины, незначительно. Если же, значение параметра α , близко к нулю, и предположим равняется 0,2. Весовой коэффициент, полученный методом Брауна на пятом шаге вычислений, будет равен значению $0,2(1 - 0,2)^5 = 0,065536$.

Для значения параметра α , находящегося на запредельном интервале, ситуация аналогична. Возьмем значения параметра α равным 1,8. Весовой коэффициент, полученный методом Брауна на пятом шаге вычислений, будет равен значению $1,8(1 - 1,8)^5 = -0,58982$. Для получения точного прогнозируемого значения методом Брауна, стоит учитывать все возможные аспекты, потому что даже небольшое различие а значениях весовых коэффициентов, могут повлечь значительный рост погрешности вычислений, что приведет к неэффективности адаптации данного метода.

С ростом числа наблюдений весовой коэффициент, полученный методом Брауна для первого вычисляемого значения, перестает оказывать влияние на точность, и он принимает значение близкое к нулю, например, при значение постоянной сглаживания, равной 0,2, весовой коэффициент первого вычисляемого расчётного значения на $k = 30$ шаге наблюдения равно значению $0,2(1 - 0,2)^{30} = 0,000248$, и отклонения при вычислении Y стремится к нулю.

Для малых выборок ряда и малых значений постоянной сглаживания, оценка первоначального вычисляемого значения Y_1 требуется уделять особое внимание.

Выбор первого значения метода Брауна может происходить следующими способами:

- 1) оценка эксперта;
- 2) первое значение метода приравнивается фактическому значению;
- 3) берется средняя арифметическая первых фактических значений.

Получение первого значения Y_1 при помощи оценки эксперта, изначально имеет значительную погрешность. Но на большой выборке, влияние погрешности первого значения Y_1 незначительно, а скорость и простота получения первого значения метода Брауна, при помощи экспертной оценки, является основным преимуществом этого способа перед другими.

В ситуации, когда используется ряд с малым количеством наблюдений ($T \ll 40$), отклонение первого вычисляемого значения может оказать большое влияние на результат прогнозирования, особенно когда значение постоянной сглаживания мало. Поэтому в этом случае значимость правильной оценки первого вычисляемого значения модели Брауна очень высока. Оценка эксперта неточна для малых выборок, поэтому следует использовать другие способы оценки.

Второй способ выбора первого значения метода Брауна заключается в присвоение ему значения фактической величины. Этот способ используется чаще остальных, потому что прост и исключает субъективизм. Частым явлением бывает, что именно первое значение метода Брауна подвергается воздействию случайного отклонения и находится далеко от среднего уровня ряда. Поэтому в метод Брауна при этом способе оценивания величины первого вычисляемого наблюдения, учитывается возможность появления отклонения, которая в случае малой выборки оказывает существенное воздействие на результаты прогнозирования. Эффективное использование этого способа оценивания значения Y_1 осуществляется только на больших выборках, так как на малой выборке возникает ошибка аппроксимации, что в свою очередь сказывается на точности прогнозируемых значений.

На малых выборках эффективнее использовать третий способ оценки первого значения Y_1 , который требует вычислять среднюю арифметическую.

Средняя арифметическая считается на основе первых трех – пяти фактических значениях. С учетом условия, метод Брауна принимает вид:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_1 \text{ при } Y_1 = \frac{1}{T} \sum_{t=1}^T Y_t, T \leq 5 \quad (2.1.43)$$

Значение, полученное средней арифметической, берется как первое значение метода Брауна и используется для вычисления следующих значений. Этот способ нахождения первого значения метода Брауна, поскольку отклонение первых пяти фактических значений усредняется. Данный способ формализован и является лучшим, поскольку метод Брауна вычисляет значения по схожему принципу. Но как и любой способ нахождения Y_1 , он имеет недостатки: средняя арифметическая, является точным способом получения первого значения Y_1 , когда исследуемый процесс стационарен и имеет нормальное распределение, но метод Брауна используется для нестационарных и необратимых процессов. Остается неясным, какое количество первых членов ряда требуются для вычисления средней арифметической – два значения наблюдений маловато. Точного ответа не существует, поэтому приходится использовать субъективные способы выбора числа фактических значений, используемых для нахождения средней арифметической. [18]

Недостатком этого способа оценивания первого вычисляемого значения прогнозируемого показателя является существование другого способа, который лишен этого недостатка – этим способом является вычисление средней взвешенной первых фактических значений. Средняя взвешенная находится схожим способом с вычислением значений методом Брауна, то есть используются весовые коэффициенты, полученные по формуле (2.1.9). Данный способ предполагает, что вычисления методом Брауна начинается не с нахождения первого значения Y_1 , а с вычисления третьего значения Y_3 . При этом, третье значение Y_3 находится не методом Брауна, а вычисляется как средняя взвешенная по формуле:

$$Y_3 = \alpha Y_2 + 1 - \alpha Y_1. \quad (2.1.44)$$

Используя методом Брауна полученное третье значение Y_3 , вычисляются четвертое, пятое и последующие прогнозные значения.

Метод Брауна, с учетом полученных начальных принимает следующий вид:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T, T \geq 3, \text{ при } Y_3 = \alpha Y_2 + 1 - \alpha Y_1. \quad (2.1.45)$$

В этом методе отсутствует субъективизм, отклонение первых двух вычисленных значений усредняется благодаря значению параметра α , которое используется при нахождение всех прогнозных значений.

Недостатком этого способа является неравномерный и необоснованный подбор весовых коэффициентов для первого и второго вычисляемого значения. Если значение параметра α близко к нулю, то весовой коэффициент второго вычисляемого значения меньше первого. Если значение параметра α близко к единице, то весовой коэффициент у второго вычисляемого значения значительно больше, чем у первого.

Если значение параметра α находится на запредельном интервале, то весовой коэффициент первого вычисляемого значения становится отрицательным. Данный способ нахождения первого значения метода Брауна, не имеет конкретного обоснования, но при этом он лишен недостатков остальных способов.

Кроме рассмотренных способов, имеется еще один способ нахождения первого значения метода Брауна, который вытекает из смысла метода. Рассмотрим использование этого способа.

На практике сумма весовых коэффициентов не может равняться единице, а только стремиться к ней. Исключением является ситуация, когда значение параметра $\alpha = 1$. На основании этого, значение остаточного члена r_t в момент времени t находится по следующей формуле:

$$r_t = S - S_t = 1 - S_t, \quad (2.1.46)$$

где S – сумма элементов бесконечного ряда, равная единице, S_t – сумма первых t элементов конечного ряда.

Значение остаточного члена (2.1.46) на основе двух наблюдений вычисляется как:

$$r_2 = 1 - S_2 = 1 - \alpha + \alpha 1 - \alpha = 1 - \alpha - \alpha 1 - \alpha = 1 - \alpha^2. \quad (2.1.47)$$

Для трёх наблюдений остаточный член имеет вид:

$$r_3 = 1 - S_3 = 1 - \alpha + \alpha 1 - \alpha + \alpha 1 - \alpha^2 = 1 - \alpha^2 - \alpha 1 - \alpha^2 = 1 - \alpha^3. \quad (2.1.48)$$

Значение остаточного члена, на основе t наблюдений вычисляется следующим образом:

$$r_t = 1 - \alpha^t. \quad (2.1.49)$$

Но значение остаточного члена может быть равным нулю, когда выполняются условия:

- 1) значение параметра $\alpha = 1$,
- 2) количество наблюдений стремится к бесконечности $t \rightarrow \infty$.

Значение параметра α равное единице и бесконечные ряды на практике встречаются очень редко, поэтому значение экспоненциальной взвешенной средней в большинстве случаев не будет являться таковым, согласно условию (2.1.9).

Следовательно, весовые коэффициенты для рядов с малым количеством элементов, требуется подкорректировать с целью корректности работы метода Брауна.

На основании формулы (2.1.46), сумма весовых коэффициентов выглядит следующим образом:

$$S_t = 1 - r_t, \quad (2.1.50)$$

заменяя значение r_t в соответствии с формулой (2.1.49) получаем:

$$S_t = 1 - 1 - \alpha^t. \quad (2.1.51)$$

Следовательно, если перемножить все весовые коэффициенты в исходном ряде (2.1.9) на поправочный коэффициент:

$$\frac{1}{1 - 1 - \alpha^t}, \quad (2.1.52)$$

то сумма весовых коэффициентов S будет равна единице.

При использовании ряда состоящего из двух наблюдений, значение среднего экспоненциального взвешенного, которое является прогнозным значением третьего наблюдения, с учетом коэффициента (2.1.52) будет вычисляться как:

$$Y_3 = \frac{\alpha Y_2 + \alpha(1-\alpha) Y_1}{1 - (1-\alpha)^2} \quad (2.1.53)$$

Для ряда состоящего из трёх наблюдений:

$$Y_4 = \frac{\alpha Y_3 + \alpha(1-\alpha) Y_2 + \alpha(1-\alpha)^2 Y_1}{1 - (1-\alpha)^3} \quad (2.1.54)$$

Конечный вид для ряда состоящего из t наблюдений будет:

$$Y_{T+1} = \frac{\alpha Y_T + \alpha(1-\alpha) Y_{T-1} + \alpha(1-\alpha)^2 Y_{T-2} + \dots + \alpha(1-\alpha)^{T-1} Y_1}{1 - (1-\alpha)^T} = \frac{\sum_{t=1}^T \alpha(1-\alpha)^{T-t} Y_t}{1 - (1-\alpha)^T} \quad (2.1.55)$$

Исходя из формулы (2.1.55), появляется возможность выбора начального значения для метода Брауна. Тогда первое значения для метода вычисляется по формуле (2.1.55), а все следующие значения вычислять по формуле (2.1.11). Если необходимо начать вычисления с третьего наблюдения, то используется формула (2.1.53) для получения требуемого начального значения.[10,11]

Используемый метод обладает несколькими преимуществами:

- значение полученное по формуле (2.1.55) является экспоненциально взвешенной первых T значений исходного ряда. При этом все остальные значения ряда также вычисляются по формуле экспоненциально взвешенной т.е. самой моделью Брауна. Результатом таких вычислений является однородный ряд вычисленных значений, что позволяет предотвратить появления в наблюдаемом процессе посторонних факторов;
- при нахождение наилучшего значения постоянной сглаживания два первых наблюдения Y_1 и Y_2 не создают начального отклонения, так как используются в вычисление третьего значения через оптимальную постоянную сглаживания, которая участвует в вычисление всех элементов ряда;
- позволяет полностью избавиться от субъективизма.

Следовательно, получаем другой вариант модели Брауна, который учитывает начальные значения в полной:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T, T \geq 3, \text{ при } Y_3 = \frac{\alpha Y_2 + \alpha (1 - \alpha) Y_1}{1 - 1 - \alpha^2}. \quad (2.1.56)$$

Проведем сравнение работы полученных методов (2.1.43), (2.1.45) и (2.1.56) на условном примере. С целью нахождения отличий в нахождение прогнозных значений, полученных этими методами, был использован ряд, имеющий изменяющуюся тенденцию. Сравнение методов представлено на рисунке 2.1.8.

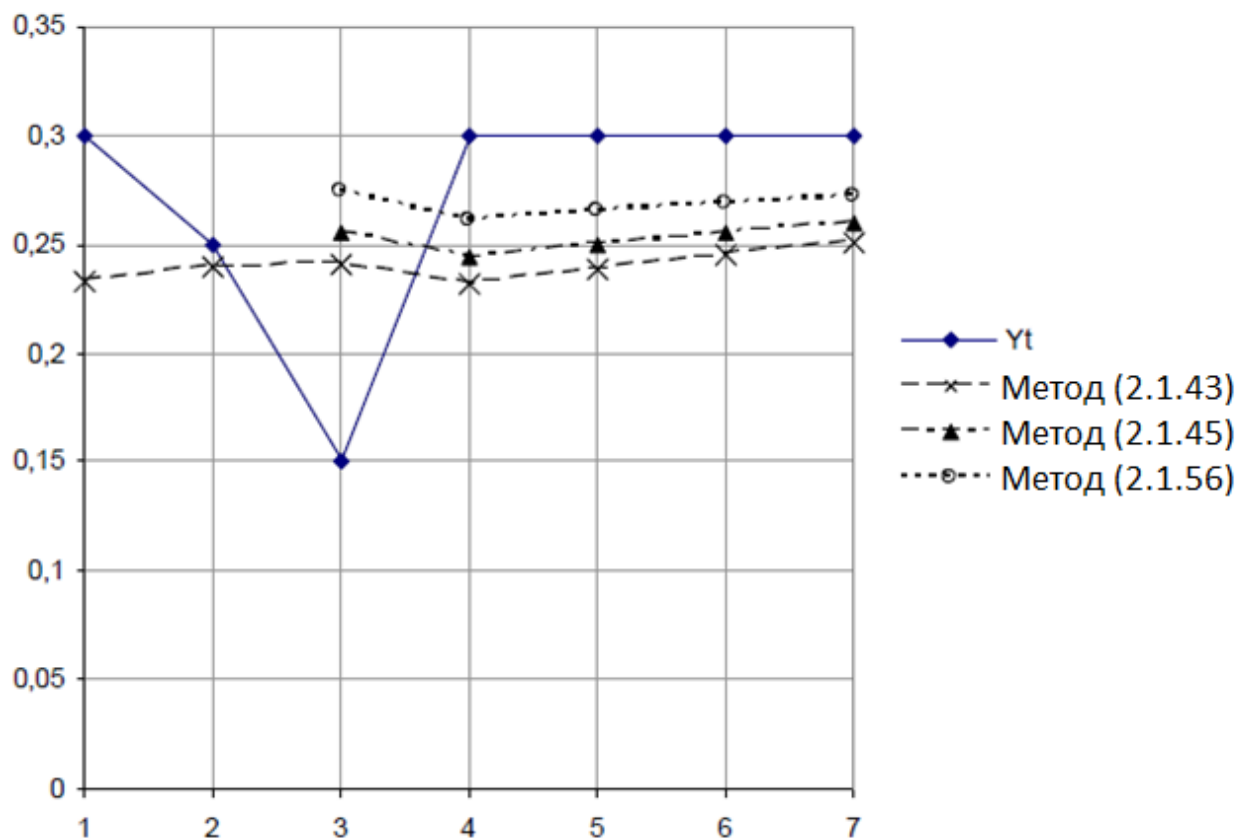


Рисунок 2.1.8 – Метод Брауна с разными начальными значениями при $\alpha = 0,1$

Значение параметра α для всех методов было равно 0,1. На рисунке 2.1.8 видно, что малом значение параметра α , способ получения первого значения играет большую роль при прогнозирование: значения полученные методом (2.1.56) оказываются ближе всех к фактическим значениям на 4 – 7 наблюдениях. Наилучшее значение параметра α является разным для каждого из используемых методов. Для метода (2.1.43) наилучшее значение параметра α составляет 0,21, для метода (2.1.45) значение параметра α равно 0, а для

метода (2.1.56) – значение параметра α находится за пределами условия (2.1.14) и равна – 0,17.

С увеличением значения параметра α , различие между получаемыми значениями методами (2.1.43), (2.1.45) и (2.1.56) уменьшается. Если значение параметра $\alpha = 1$, то значения, получаемые всеми методами, начиная с четвертого наблюдения совпадают. Сравнение методов представлено на рисунке 2.1.9.

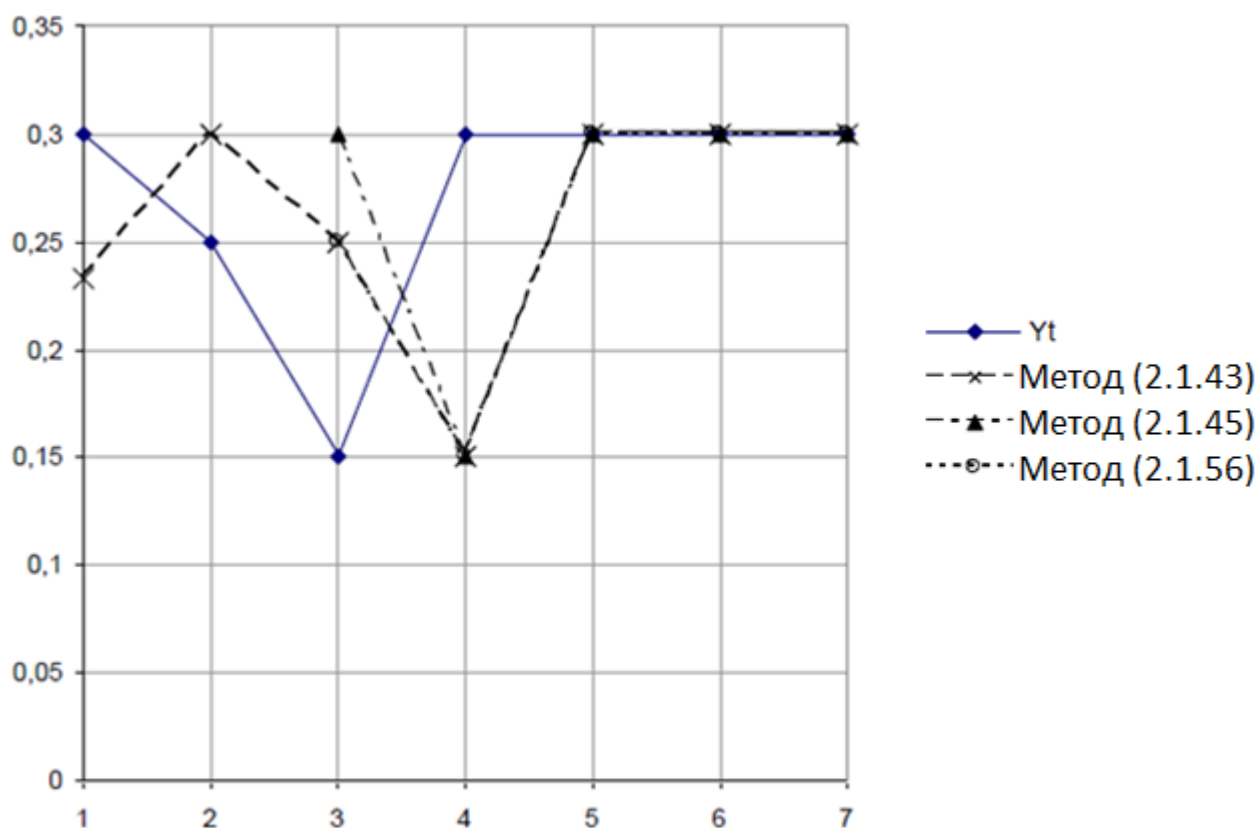


Рисунок 2.1.9 – Метод Брауна с разными начальными значениями при $\alpha = 1$

С увеличением номера наблюдения, различие получаемых методами значений увеличивается. Причиной расхождения является используемые способ задания первого значения для методов. На рисунке 2.1.10 показано, как ведут себя используемые методы при значении постоянной сглаживания $\alpha = 1,8$.

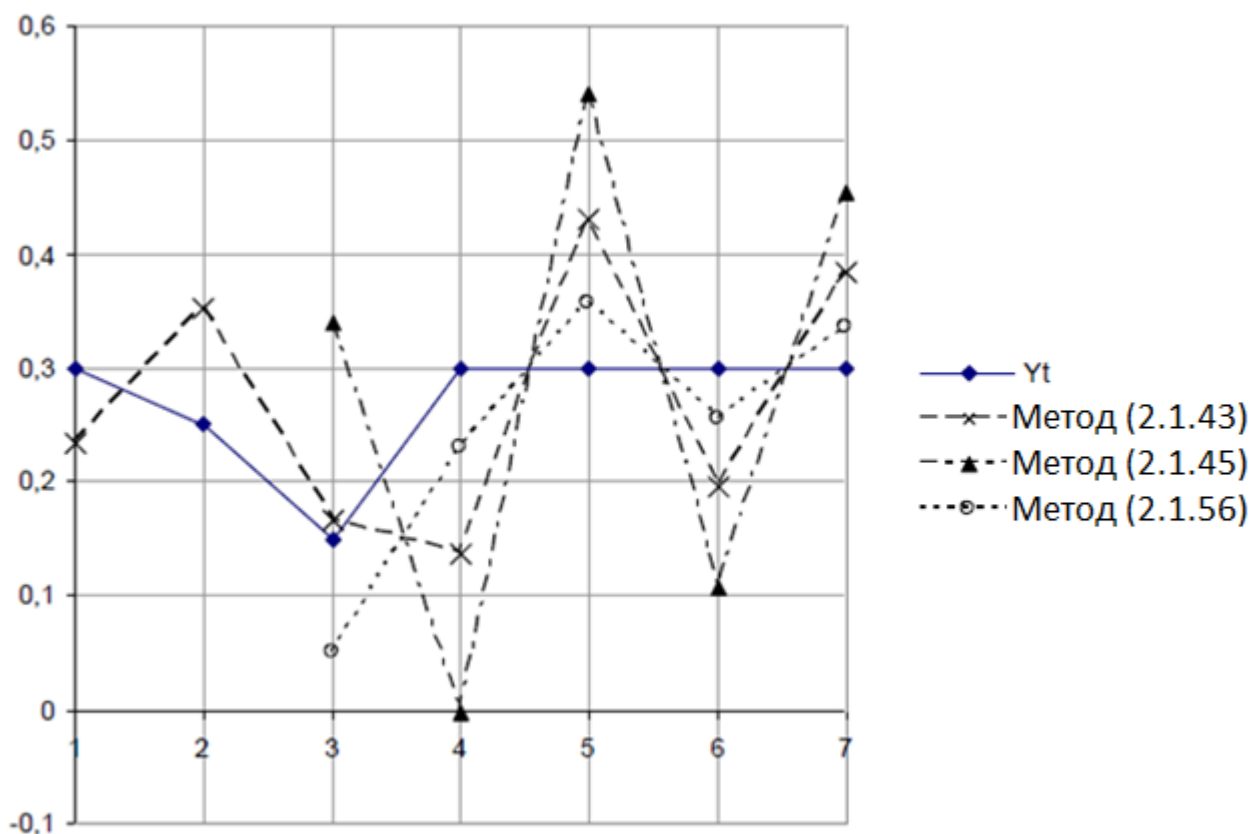


Рисунок 2.1.10 – Метод Брауна с разными начальными значениями при $\alpha = 1,8$

Различие между вычисляемыми значениями методами отчетливо видно. Причиной различия служат разные способы задания первого значения методов. По рисунку 2.1.10 видно, что значения, полученные методом (2.1.56) гораздо точнее значений, полученных другими методами.

Следовательно, способ вычисления первого значения метода Брауна оказывает большое влияние на точность получаемых значений. Если при использовании метода Брауна, значение параметра α получается близким к 0, или 2, то требуется использовать разные способы задания первого значения метода: возможно, другой способ позволит получить более точные прогнозные значения. При использовании малых выборок, эффективно использовать метод (2.1.45), или метод (2.1.56), использование которых сохраняет свойства метода Брауна.

При значении параметра α , значение которого приближается к 0 или 2, ряд весовых коэффициентов не сходится к единице. Существует множество методов, учитывающих данное свойство. Одним из таких методов является метод Р. Вельда, который предполагает использование поправочного коэффициента:

$$k_T = \frac{1}{\sum_{t=1}^T \alpha^{1-t}}. \quad (2.1.57)$$

Полученное методом Брауна значение Y_T перемножается на поправочный коэффициент. Полученное таким способом значение является более точным и имеет вид:

$$Y_T = k_T Y_T. \quad (2.1.58)$$

Использование метода Брауна для получения прогнозных значений становится эффективным только в том случае, если значение поправочного коэффициента приближается к значению 0,995.

Эта проблема может быть решена другим образом, без вычисления поправочного коэффициента.

При использовании малых выборок, вычисление значений методом Брауна имеет вид:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T, T \geq 3, \text{ при } Y_3 = \alpha Y_2 + 1 - \alpha Y_1, \quad (2.1.59)$$

или:

$$Y_{T+1} = \alpha Y_T + 1 - \alpha Y_T, T \geq 3, \text{ при } Y_3 = \frac{\alpha Y_2 + \alpha^{1-\alpha} Y_1}{1 - \alpha^2}. \quad (2.1.60)$$

При вычислении средней взвешенной для нахождения третьего прогнозного значения, ряд весовых коэффициентов в обоих случаях равняется единице:

$$\alpha + 1 - \alpha = 1 \text{ и } \frac{\alpha + \alpha^{1-\alpha}}{1 - (1-\alpha)^2} = \frac{2\alpha - \alpha^2}{1 - 1 + 2\alpha - \alpha^2} = \frac{2\alpha - \alpha^2}{2\alpha - \alpha^2} = 1, \quad (2.1.61)$$

Поэтому изменение метода Брауна при помощи поправочного коэффициента не требуется. Используя данный способ, в методе (2.1.45) вычислим четвертое прогнозное значение:

$$Y_4 = \alpha Y_3 + 1 - \alpha \quad Y_3 = \alpha Y_3 + 1 - \alpha \quad \alpha Y_2 + 1 - \alpha \quad Y_1 = \alpha Y_3 + 1 - \alpha \quad Y_2 + 1 - \alpha \quad {}^2 Y_1 \quad (2.1.62)$$

Сумма весовых коэффициентов выражения (2.1.62) равна единице:

$$s_3 = \alpha + \alpha 1 - \alpha + 1 - \alpha^2 = \alpha + \alpha - \alpha^2 + 1 - 2\alpha + \alpha^2 = 1 \quad (2.1.63)$$

Значение, полученное методом Брауна на пятом наблюдение имеет вид:

$$Y_5 = \alpha Y_4 + 1 - \alpha \quad Y_4 = \alpha Y_4 + 1 - \alpha \quad \alpha Y_3 + 1 - \alpha \quad Y_2 + 1 - \alpha \quad {}^2 Y_1 = \alpha Y_4 + 1 - \alpha \quad Y_3 + 1 - \alpha \quad {}^2 Y_2 + 1 - \alpha \quad {}^3 Y_1 \quad (2.1.64)$$

Вычислим сумму весовых коэффициентов выражения (2.1.64), которая является суммой ряда:

$$s_4 = \alpha + \alpha 1 - \alpha + 1 - \alpha^2 + 1 - \alpha^3 \quad (2.1.65)$$

Выделим сумму первых трёх слагаемых выражения (2.1.65) и представим в виде:

$$s = \alpha + \alpha 1 - \alpha + 1 - \alpha^2 = \alpha + \alpha 1 - \alpha + 1 - \alpha^2 - 1 - \alpha^2 + \alpha 1 - \alpha^2 \quad (2.1.66)$$

Откуда, при учете выражения (2.1.63), получим:

$$s = 1 - 1 - \alpha^2 + \alpha 1 - \alpha^2 = 1 - 1 - \alpha^3 \quad (2.1.67)$$

Используя значение полученное в выражение (2.1.52), получим:

$$s_4 = 1 - 1 - \alpha^3 + 1 - \alpha^3 = 1 \quad (2.1.68)$$

Сумма весовых коэффициентов снова равняется единице. Данная сумма будет равно единице для любого значения $T > 2$. Отсюда следует, что ряд весовых коэффициентов при использовании метода (2.1.56) будет сходиться к единице при любом значении $T > 2$.

Следовательно, использование метода Брауна с учетом условия задания начального коэффициента по методам (2.1.45) или (2.1.56), приводит к тому, что сумма весовых коэффициентов выражения (2.1.9) будет равняться единице, и использование поправочных коэффициентов в методе Брауна лишено всякого смысла. Поэтому, при использовании малых выборок, эффективное нахождение прогнозных значений методом Брауна, осуществляется по формулам (2.1.45) или (2.1.56). Данные методы, разработанные для малых рядов, но могут

использоваться и больших рядов. Это означает, что данные методы являются универсальными.

2.2 Реализация метода Брауна для адаптации регрессионной модели

Первым используемым алгоритмом адаптации является метод Брауна. Метод реализован на языке программирования Java. Исходные данные, используемые при построении метода Брауна были получены регрессионной моделью. Значение средней взвешенной α для метода Брауна при адаптации используемой регрессионной модели, вычисляется по следующей формуле:

$$\alpha = 2/(n + 1) \quad (2.2.1)$$

Значение α , вычисленное по формуле (2.2.1), описывает влияние исходных и новых данных оптимальным образом для используемой регрессионной модели. Первое значение было вычислено на основе среднего арифметического первых трех значений регрессионной модели.

Функция, реализующая метод Брауна на языке программирования Java, представлена на рисунке 2.2.1

```
double a = 2/(n+1);
yb[0]=(y[0]+y[1]+y[2])/3;
for (int i=1; i<n; i++)
{
    yb[i]=(1-a)*y[i-1]+a*yb[i-1];
    dif[i]=Math.pow((y[i]-yb[i]),2);
}
for (int i=0; i<n; i++)
{
    dif1=dif1+dif[i];
}
Ep=Math.sqrt(dif1/(n-1));
```

Рисунок 2.2.1 – Метод Брауна

Результаты адаптации регрессионной модели методом Брауна представлены на рисунке 2.2.2.

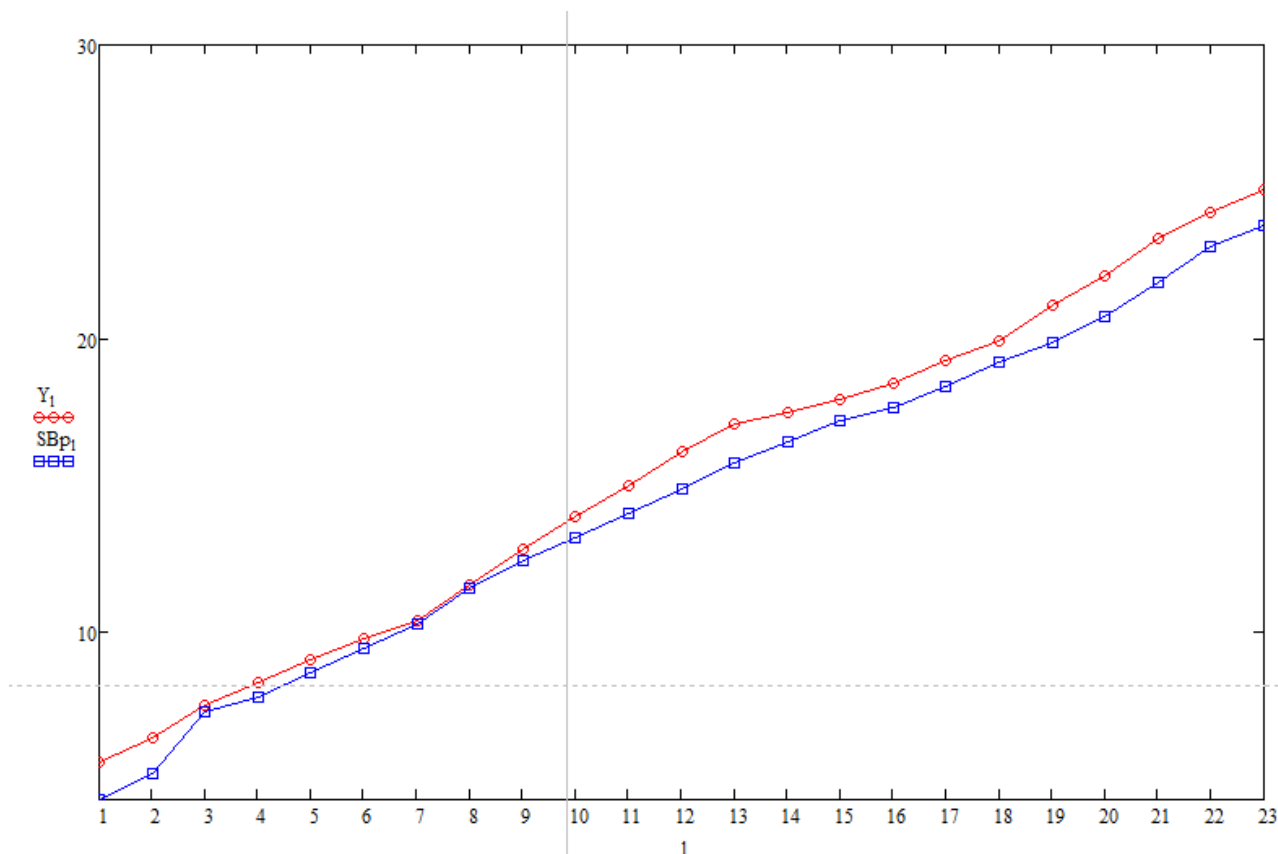


Рисунок 2.2.2 – Сравнение модели адаптированной методом Брауна с исходными данными

На рисунке 2.2.2 график Y_t – построен по исходным данным, график SBp_t – построен моделью, адаптированной методом Брауна, ось абсцисс – номер наблюдения, ось ординат – значение модели. Модель, полученная методом Брауна, имеет стандартную погрешность равную 0.923604. Сравнение исходной регрессионной модели и модели полученной при помощи метода Брауна представлено на рисунке 2.2.3.

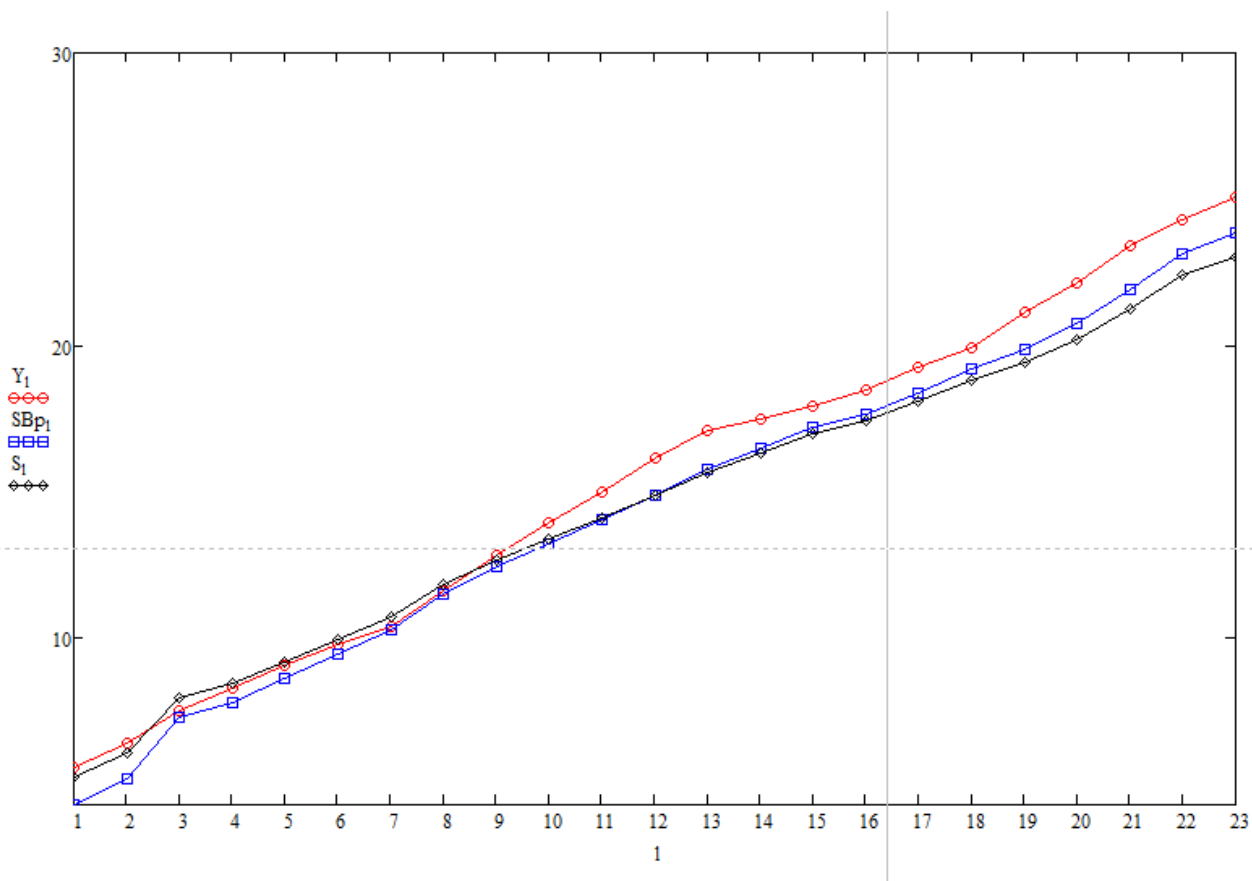


Рисунок 2.2.3 – Сравнение исходной регрессионной модели с моделью адаптированной методом Брауна

На рисунке 2.2.3 график Y_t – построен по исходным данным, график SBp_t – построен моделью, адаптированной методом Брауна, график S_t – исходная модель, ось абсцисс – номер наблюдения, ось ординат – значение модели.

Полученный результат показал, что полученная в результате адаптации методом Брауна модель, является точнее, но с целью дальнейшего прогнозирования нужно использовать методы, которые дадут более точный результат.

В данной главе был проанализирован адаптационный метод Брауна. Метод Брауна вычисляет прогнозное значение на основе ряда фактических значений. Было рассмотрено влияние значения средней взвешенной и способа задания первого значения на точность прогнозного значения методом Брауна.

Выявлены существенные недостатки метода Брауна, влияющие на точность прогнозируемого значения. Первым недостатком является

необходимость учета исходных и новых данных. Вторым недостатком является требование построения новой регрессионной модели. Третий недостаток – необходимость задания начального значения для вычислений модели Брауна и выбора подходящего способа задания этого значения. Невыполнение данных требования делает использование метода Брауна для адаптации неэффективным.

Реализована и проанализирована регрессионная модель, адаптированная методом Брауна. Сравнительный анализ показал, что адаптированная методом Брауна модель получает более точные прогнозы, чем исходная регрессионная модель. Но с учетом недостатков метода Брауна, эффективность этого метода недостаточна, поэтому требуется использовать другой метод адаптации.

Глава 3 Адаптация регрессионной модели методом стохастической аппроксимации

3.1 Теоретические вопросы одномерной стохастической аппроксимации

Общий вид итеративной процедуры стохастической аппроксимации представляется соотношением

$$x_{n+1} = x_n + a_n z x_n + \zeta \omega_n, x_n, \quad n = 1, 2, \dots, \quad (3.1.1)$$

где $z x$ – векторное поле, указывающее направление движения к точке, дающей решение задачи, в частности, экстремум математического ожидания случайной функции $\varphi \omega, x$;

$\zeta \omega, x$ – случайный вектор с нулевым математическим ожиданием $M\zeta = 0$;

a_n – последовательность значений, свободно выбираемая в достаточно широких пределах, характеризует длины шагов. Скаляры a_n могут интерпретироваться так же, как «параметры демпфирования колебаний» итеративного процесса. Чем меньше a_n , тем медленнее меняются значения x_n .

Классические схемы стохастической аппроксимации представляют собой, таким образом, итеративные методы градиентного типа. На каждом шаге процесса происходит смещение в произвольном направлении, которое определяется значениями исходных данных и схемой отдельной итерации. На одних шагах приближаемся к искомому оптимуму, на других удаляемся от него.

При использовании метода стохастической аппроксимации, сдвиг в нужном направлении происходит с большей вероятностью. В среднем за каждую итерацию происходит сдвиг в требуемом направлении. При большом количестве итераций в соответствии с законом больших чисел обеспечивается приближение к искомой оптимальной точке, если она единственна. Для того

чтобы добраться до оптимума из сколь угодно удаленной от него начальной точки и при этом обеспечить достаточное количество итераций, требуется проявление закона больших чисел, следовательно необходимо чтобы $\sum_{n=1}^{\infty} a_n$ расходился.

Чтобы после некоторого количества итераций, проведенных в окрестности искомой точки, не выйти, из нее, требуется, чтобы значение шага a_n при $n \rightarrow \infty$ стремилось к нулю. Таким образом, на классические методы стохастической аппроксимации можно смотреть как на итеративные методы градиентного типа решения безусловной экстремальной стохастической задачи

$$\inf_x M_{\omega} \varphi \omega, x . \quad (3.1.2)$$

Условия, при которых классические методы стохастической аппроксимации сходятся к искомому экстремуму, требуют выпуклости или по крайней мере одно экстремальности функции $M_{\omega} \varphi \omega, x$ по x . Это значит, что для использования стохастической аппроксимации в качестве итеративного метода решения задач стохастического программирования необходимо модифицировать классические схемы применительно к задачам условной-оптимизации и отказаться от требования выпуклости или одно экстремальности $M_{\omega} \varphi \omega, x$.

Рассматривая рекуррентное соотношение (3.1.1), как разностное уравнение, отвечающее дискретным моментам $n = 1, 2, \dots$, можно перейти к непрерывному случаю, заменив соотношение (3.1.1) дифференциальным уравнением.

В 1929 г. для отыскания корня уравнения

$$f(x) = \alpha \quad (3.1.3)$$

где $f(x)$ – заданная действительная функция;

α – действительное число.

Была предложена рекуррентная процедура

$$x_{n+1} = x_n + a_n (\alpha - f(x_n)), \quad n = 1, 2, \dots, \quad (3.1.4)$$

где x_1 – произвольное действительное число.

При некоторых ограничениях, наложенных на последовательность $\{a_n\}$, была доказана сходимость процедуры (3.1.4) к решению уравнения (3.1.3), ближайшему к x_1 .

В 1933 г. для вычисления значения переменной x , в котором достигается максимум $f(x)$, была предложена рекуррентная схема

$$x_{n+1} = x_n + \frac{a_n}{c_n} (f(x_n) + c_n - f(x_n - c_n)), n = 1, 2, \dots, \quad (3.1.5)$$

где x_1 – произвольное число.

При некоторых допущениях относительно функции $f(x)$ и последовательностей $\{a_n\}$ и $\{c_n\}$ была доказана сходимость процедуры к точке x^* в которой достигается максимум $f(x)$.

В прикладных задачах вычисление корня уравнения (3.1.3) и определение максимума функции $f(x)$ происходят обычно в условиях неполной информации относительно значений $f(x)$. Во многих задачах функция $f(x)$ задается в виде последовательности значений в точках x_1, x_2, \dots . Пусть в результате ошибок наблюдения зафиксированы значения $y_1 = y(\omega, x_1), y_2 = y(\omega, x_2), \dots$ случайной величины $y(\omega, x)$:

$$y(\omega, x_1) = f(x) + \omega(x) \quad (3.1.6)$$

Где ошибка наблюдения $\omega_i = \omega(x_i)$ – независимые между собой одинаково распределённые случайные величины с нулевым математическим ожиданием

$$M\omega(x) = \int_{-\infty}^{\infty} \omega dF = 0. \quad (3.1.7)$$

Обозначим условную функцию распределения $y(\omega, x)$ через $F_{y|x}$. Для каждого x имеем

$$My(\omega, x) = \int_{-\infty}^{\infty} y dF_{y|x} = f(x), \quad (3.1.8)$$

где $f(x)$ – функция регрессии, отвечающей распределению $F_{y|x}$.

Естественно попытаться построить итерационные процессы вида и (3.1.5), в которые вместо $f(x)$ входили бы случайные реализации наблюдений $y(\omega, x)$ и которые сходились бы в некотором смысле к корню уравнения (3.1.3)

или к экстремуму функции $f(x)$. Впервые такой процесс для отыскания корня уравнения построили Х. Роббинс и С. Монро. Используя их метод, Кифер и Вольфовиц предложили метод вычисления максимума математического ожидания случайной функции.

Целенаправленные рекуррентные процессы определения корня уравнения $M_{\omega} y_{\omega, x} = \alpha$ и отыскания экстремума функции $f(x) = M_{\omega} y_{\omega, x}$ по последовательным реализациям функции $y_{\omega, x}$ или некоторых ее характеристик получили название методов стохастической аппроксимации. Название подчеркивает, что сходимость процессов носит стохастический характер

В дальнейшем будут важны, главным образом, методы типа Кифера – Вольфовица, поскольку от них естественно перейти к итеративным схемам решения условных экстремальных задач, в которых параметры целевой функции и ограничений случайны.

Под классической схемой стохастической аппроксимации мы будем понимать итеративные схемы типа методов Роббинса – Монро и Кифера – Вольфовица и различные их модификации.

Рассмотрим итеративный метод Кифера — Вольфовица для отыскания точки единственного максимума функции регрессии $f(x)$.

Пусть $\{a_n\}$ и $\{c_n\}$ – последовательности положительных чисел, удовлетворяющих условиям

$$\sum_{n=1}^{\infty} a_n = \infty \quad (3.1.9)$$

$$\sum_{n=1}^{\infty} a_n c_n < \infty \quad (3.1.10)$$

$$\lim_{n \rightarrow \infty} c_n = 0 \quad (3.1.11)$$

$$\sum_{n=1}^{\infty} \frac{a_n^2}{c_n} < \infty \quad (3.1.12)$$

Метод Кифера – Вольфовица имеет вид:

$$x_{n+1} = x_n + \frac{a_n}{c_n} (y_{2n} - y_{2n-1}), n = 1, 2, \dots, \quad (3.1.13)$$

где x_1 – случайное значение;

y_{2n} и y_{2n-1} – независимые случайные величины с функциями распределения $F_{y|x_n+c_n}$ и $F_{y|x_n-c_n}$.

Чтобы метод (3.1.13) сходил к точке максимума функции $f(x)$, требуется наложить на функцию регрессии некоторые ограничения. В методе Кифера – Вольфовица не требуется дифференцируемость $f(x)$, именно поэтому метод нахождения экстремума $f(x)$ не является следствием метода нахождения корня уравнения $f'(x) = 0$. В случае существования производной $f'(x)$, это значение должно равняться нулю в точке $x = \Theta$, в которой достигается максимум $f(x)$. Поэтому естественно в общем случае требовать от функции $f(x)$, чтобы ее производная не превышала определенное значение в окрестности $x = \Theta$. Данное условие записывается следующим образом. [12]

Существуют такие постоянные значения $k_1 > 0$ и $k_2 > 0$, что

$$f(x_1) - f(x_2) < k_2(x_1 - x_2), \text{ как только } x_1 - \Theta + x_2 - \Theta < k_1. \quad (3.1.14)$$

Если функция $f(x)$ изменяется незначительно на достаточно большом отдалении от $x = \Theta$, то нахождение экстремума при помощи итеративной метода (3.1.13) будет происходить медленно. Поэтому естественно потребовать, чтобы за пределами некоторой окрестности точки $x = \Theta$ скорость изменения функции $f(x)$ была ограничена снизу. Данное условие звучит так.

Для любого значения $\delta > 0$ существует $\varepsilon(\delta) > 0$, такое, что

$$\inf_{0 < h < \frac{\delta}{2}} \frac{f(x+h) - f(x)}{h} > \varepsilon(\delta), \text{ как только } x - \Theta > \delta. \quad (3.1.15)$$

Чтобы ограничить рост функции $f(x)$, требуется выполнять следующее условие. Существуют постоянные значения $k_3 > 0$ и $k_4 > 0$, такие, что

$$f(x_1) - f(x_2) < k_4, \text{ как только } x_1 - x_2 < k_3. \quad (3.1.16)$$

Последнее требование, поставленное методом Кифера – Вольфовица, – одноэкстремальность функции $f(x)$. Данное требование выглядит следующим образом,

$$f(x) \text{ возрастает при } x < \Theta \text{ и убывает при } x > \Theta. \quad (3.1.17)$$

При выполнении условий (3.1.9) – (3.1.12) и (3.1.14) – (3.1.17) последовательность (3.1.13) сходится по вероятности $x = \Theta$.

Блум, в своих исследованиях не учитывает требование (3.1.14) и доказывает сходимость формулы (3.1.13) с вероятностью единица, а Дворецкий – показал, что в условиях теоремы Блума имеет место сходимость и в среднеквадратическом.

Рассмотрим особенности последовательностей случайных значений.

Пусть $x_1, x_2, \dots, x_n, \dots$ – случайная последовательность, а Θ – действительное число.

1. Последовательность x_n сходится к значению Θ с вероятностью единица, если $P \lim_{n \rightarrow \infty} x_n = \Theta = 1$.

2. Последовательность x_n сходится к значению Θ по вероятности, если $\lim_{n \rightarrow \infty} P |x_n - \Theta| > \varepsilon = 0, \forall \varepsilon > 0$.

3. Последовательность x_n сходится к значению Θ в среднеквадратическом, если $\lim_{n \rightarrow \infty} M |x_n - \Theta|^2 = \lim_{n \rightarrow \infty} x_n - \Theta^2 = 0$.

Следствием сходимости последовательности с вероятностью единица является сходимость по вероятности. На основании сходимости в среднеквадратическом также следует сходимость по вероятности.

Бурхгольдер рассмотрел обобщенный метод стохастической аппроксимации, из которого получаются методы Роббинса–Монро и Кифера – Вольфовица.

Пусть $f_n(x)$ – некоторая функция, а $z_n(x)$ – случайная величина с условной функцией распределения $F_{z|x}^n$ такая, что $M z_n | x = f_n(x)$; a_n – последовательность положительных чисел; x_1 – случайная величина.

Рассмотрим метод стохастической аппроксимации

$$x_{n+1} = x_n - a_n z_n, \quad n = 1, 2, \dots, \quad (3.1.18)$$

где z_n – случайная величина с условной функцией распределения $F_{z|x_n}^n$ на последовательности заданных значений $x_1, \dots, x_n, z_1, \dots, z_n$.

Пусть $F_{y|x}^k$ – функция распределения среднеарифметического k независимых случайных величин, каждая из которых имеет функцию распределения $F_{y|x}$. Если в выражение (3.1.18), значение $z_n = \alpha - y_n$ где y_n – случайная величина с функцией распределения $F_{y|x_n}^n$ на последовательности заданных значений $x_1, \dots, x_n, y_1, \dots, y_n$, то метод (3.1.18) является обобщением метода (3.1.4). Метод Роббинса – Монро соответствует значению функции распределения $F_{y|x}^k = F_{y|x}$, когда $k = 1$.

При подстановке в выражение (3.1.18) значения $z_n = \frac{1}{c_n} y_{2n} - y_{2n-1}$, где y_{2n} и y_{2n-1} случайные величины, условные функции распределения которых равны $F_{y|x_n+c_n}^n$ и $F_{y|x_n-c_n}^n$ соответственно на последовательности заданных значений $x_1, \dots, x_n, y_1, \dots, y_{n-1}$, в таком случае метод Бурхгольдера (3.1.18) является обобщением метода (3.1.5).

Метод Кифера – Вольфовица соответствует значению функции распределения $F_{y|x_k+c_k}^k = F_{y|x_k+c_k}$ и $F_{y|x_k-c_k}^k = F_{y|x_k-c_k}$, когда $k = 1$.

Главным различием метода (2.16) и методов (2.2) и (2.3), в которых функция $f(x)$ заменена на $y \omega, x$, является то, что условная функция распределения ошибок наблюдения $\omega(x)$ зависит не только от параметра x , но и от времени.

В своих исследованиях Бурхгольдер доказал справедливость следующего утверждения: пусть x_n – метод стохастической аппроксимации вида (3.1.16), Θ – действительное число, для которого можно подобрать функцию $\eta \varepsilon$, определенную на множестве положительных чисел, такую, что если $\varepsilon > 0$, $x - \Theta > \varepsilon$ и $n > \eta \varepsilon$ то

$$x - \Theta f_n x > 0, \quad (3.1.19)$$

$$\sup_{n,x} \frac{f_n x}{1+x} < \infty, \quad \sup \text{var} z_n x < \infty, \quad (3.1.20)$$

для $0 < \delta_1 < \delta_2 < \infty$

$$\lim_{n \rightarrow \infty} a_n \inf_{\delta_1 \leq x - \Theta \leq \delta_2} f_n x = \infty, \quad (3.1.21)$$

$$\sum_{n=1}^{\infty} a_n = \infty, \quad \sum_{n=1}^{\infty} a_n^2 < \infty. \quad (3.1.22)$$

Тогда $P \lim_{n \rightarrow \infty} x_n = \Theta = 1$.

Дворецкий предложил итерационный метод стохастической аппроксимации. Данный метод выглядит следующим образом

$$x_{n+1} = T_n x_1, \dots, x_n + z_n, \quad n = 1, 2, \dots, \quad (3.1.23)$$

где $T_n x_1, \dots, x_n$ – вычисляемое изменение, подверженное некоторым условиям;

z_n – случайная величина с законом распределения $F_{z|x}$ и с начальным значением равным нулю;

T_n – изменение, без отклонения;

z_n – случайное отклонение.

Метод Дворецкого сходится, когда случайное отклонение полученное в результате адаптации, является незначительным. Поэтому появляется возможность изучения детерминированной и стохастической составляющей метода (3.1.23).

Значимым условием, положенным в основу классических методов стохастической аппроксимации, является допущение об отсутствии систематических отклонений. Невыполнение данного условия приводит к тому, что метод не будет сходиться к оптимуму.[9]

Не менее значимым является требование сходимости ряда из дисперсий ошибок наблюдения $\sum_{n=1}^{\infty} Mz_n^2 < \infty$. При невыполнение данного условия, метод может не сходиться.

Для анализа особенностей метода Дворецкого, рассмотрим условия, которым должно удовлетворять детерминированное изменение $T_n x_1, \dots, x_n$.

Первое условие заключается в том, что при достаточно близком нахождение к искомой точке $x = \Theta$, сильные изменения нежелательны, поскольку существует вероятность удалиться от цели. Следовательно если $x_n - \Theta < \varepsilon$, то должно выполняться условие $x_{n+1} - x_n < \delta \varepsilon$, при достаточно больших значениях n

$$T_n x_1, \dots, x_n - \Theta < a_n \quad (3.1.24)$$

На последовательность значений a_n накладывается следующее требование

$$\lim_{n \rightarrow \infty} a_n = 0 \quad (3.1.25)$$

Второе условие заключается в том, что при отдаленном нахождение от $x = \Theta$, необходимо смещаться на такое значение, чтобы расстояние до оптимума сокращалось на каждом шаге.[34] Следовательно необходимо выполнять следующее условие

$$x_{n+1} - \Theta = T_n x_1, \dots, x_n - \Theta \leq x_n - \Theta - \gamma_n \quad (3.1.26)$$

где последовательность γ_n неотрицательных чисел, которая удовлетворяет условию

$$\sum_{n=1}^{\infty} \gamma_n = \infty \quad (3.1.27)$$

Дворецкий упрощает требование, для выражения (3.1.26):

$$T_n x_1, \dots, x_n - \Theta \leq 1 + \beta_n x_n - \Theta - \gamma_n \quad (3.1.28)$$

где β_n – последовательность неотрицательных чисел, которая удовлетворяет условию

$$\sum_{n=1}^{\infty} \beta_n < \infty \quad (3.1.29)$$

На основании условий (3.1.24) и (3.1.25), условие Дворецкого преобразуется и имеет вид

$$T_n x_1, \dots, x_n - \Theta \leq \max a_n, 1 + \beta_n x_n - \Theta - \gamma_n. \quad (3.1.30)$$

На начальное условие и погрешность z_n накладываются следующие требования:

$$M x_1^2 < \infty \quad (3.1.31)$$

$$M z_n^2 < \infty \quad (3.1.32)$$

$$M z_n | x_1, \dots, x_n = 0, \forall n \quad (3.1.33)$$

Теорема Дворецкого формулируется следующим образом: пусть a_n , β_n , γ_n – последовательности неотрицательных значений, которые удовлетворяют условиям (3.1.25), (3.1.27), (3.1.29), Θ – действительное число; $T_n x_1, \dots, x_n$ – вычисляемое изменение, удовлетворяющее условию (3.1.30) на

последовательности заданных значений x_1, \dots, x_n и z_n – случайные величины, которые удовлетворяю условиям (3.1.31) – (3.1.33). Следовательно, метод Дворецкого (3.1.23) сходится в среднеквадратическом и с вероятностью единица

$$\lim_{n \rightarrow \infty} M |x_n - \Theta|^2 = 0 \quad (3.1.34)$$

$$P \lim_{n \rightarrow \infty} x_n = \Theta = 1 \quad (3.1.35)$$

Теорема Дворецкого остается справедливой, если $\alpha_n, \beta_n, \gamma_n$ заменить функциями $\alpha_n(x_1, \dots, x_n), \beta_n(x_1, \dots, x_n), \gamma_n(x_1, \dots, x_n)$, такими, что

а) $\alpha_n(x_1, \dots, x_n)$ равномерно ограничены и $\lim_{n \rightarrow \infty} \alpha_n(x_1, \dots, x_n) = 0$ равномерно по всем последовательностям $x_1, x_2, \dots, x_n, \dots$;

б) $\beta_n(x_1, \dots, x_n)$ измеримы и ряд $\sum_{n=1}^{\infty} \beta_n(x_1, \dots, x_n)$ равномерно ограничен и равномерно сходится для всех последовательностей x_1, x_2, \dots ;

в) $\gamma_n(x_1, \dots, x_n)$ удовлетворяет условию $\sum_{n=1}^{\infty} \gamma_n(x_1, \dots, x_n) = \infty$ для всех последовательностей $x_1, x_2, \dots, x_n, \dots$, для которых $\sup_n x_n < \alpha$ для значения $\alpha > 0$.

Выполнение теоремы Дворецкого не прекращается, при замене условие (3.1.27) на условие $\sum_{n=1}^{\infty} \sup_{x_1, \dots, x_n} M z_n |x_1, \dots, x_n < \infty$. Так же, возможна замена условия (3.1.27) на требование, чтобы ряд $\sum_{n=1}^{\infty} M z_n |x_1, \dots, x_n$ был равномерно ограничен и равномерно сходиллся на всевозможных последовательностях значений $x_1, x_2, \dots, x_n, \dots$.

Выполнение общей теорема Дворецкого не нарушается, при замене требование к $\alpha_n(x_1, \dots, x_n)$ следующим условием:

$$P \alpha_1(x_1) < \infty = 1; P \alpha_n(x_1, \dots, x_n) \geq \alpha_{n+1}(x_1, \dots, x_{n+1}) = 1; \quad (3.1.36)$$

$$P \lim_{n \rightarrow \infty} \alpha_n(x_1, \dots, x_n) = 0 = 1. \quad (2.1.37)$$

Метод Роббинса – Монро и Кифера – Вольфовица являются особыми случаями метода Дворецкого. Для метода (3.1.2) – $T_n(x_n) = x_n + \alpha_n(\alpha - f(x_n))$, для метода (3.1.3) – $T_n(x_n) = x_n + \frac{\alpha_n}{c_n} (f(x_n) + c_n - f(x_n) - c_n)$.

Метод Дворецкого, является частным случаем обобщённого метода, предложенного Браверманном, Резонером, Волконским и Иванковым.

Дупач предложил обобщение метода Кифера – Вольфовица, которое позволяет учитывать изменение оптимального значения Θ в процессе аппроксимации.

Предположим, что $f_n x = f x - \Theta_n + \Theta_1$, $n = 2, 3, \dots$, где значение Θ_n является максимумом функции $f_n x$. Полученный метод стохастической аппроксимации выглядит следующим образом:

$$x_{n+1} = x_n + \frac{\alpha_n}{c_n} (y(x_n + c_n) - y(x_n - c_n)), \quad n = 1, 2, \dots, \quad (3.1.38)$$

где $\alpha_n = 1 + \frac{1}{n} \alpha$, а $y(x_n + c_n)$ и $y(x_n - c_n)$ – случайные значения, математическое ожидание которых соответственно равно $M y(x_n + c_n) | x_1, \dots, x_n = f_{n+1}(x_n + c_n)$ и $M y(x_n - c_n) | x_1, \dots, x_n = f_{n+1}(x_n - c_n)$.

Кроме того, предполагается, что $y(x_n + c_n)$ и $y(x_n - c_n)$ являются независимыми и значения их дисперсий ограничено значением постоянной σ^2 . В методе (3.1.30) x_1 – произвольная начальная оценка с ограниченной дисперсией, а последовательности α_n и c_n определяются соотношениями:

$$\alpha_n = \frac{a}{n^\alpha}, \quad a > 0; \quad \frac{3}{5} < \alpha < 1; \quad c_n = \frac{c}{n^\gamma}, \quad c > 0; \quad \frac{a}{6} \leq \gamma \leq a - \frac{1}{2}, \quad (3.1.39)$$

значение Θ_n изменяется так, что

$$\Theta_{n+1} - \left(1 + \frac{1}{n}\right) \Theta_n = O\left(\frac{1}{n^k}\right), \quad k > a. \quad (3.1.40)$$

Потребуем, чтобы функция $f(x)$ удовлетворяла следующим требованиям. Существуют положительные значения k_1, k_2, k_3 , такие, что

$$k_1(x - \Theta_1) \leq f'(x) \leq k_2(x - \Theta_1), \quad (3.1.41)$$

$$f'''(x) \leq k_3 \quad \text{при} \quad -\infty \leq x \leq \infty, \quad (3.1.42)$$

$f(x)$ возрастает при значении $x < \Theta_1$ и убывает при значении $x > \Theta_1$. Поэтому следующее утверждение является действительным: при выполнении поставленных условий, динамическая метод стохастической аппроксимации сходится в среднеквадратическом $x_n - \Theta_n$ к нулю. [15,16,33]

При этом

$$M |x_n - \Theta_n|^2 = \begin{cases} 0 \frac{1}{n^{\alpha-2\gamma}} & \text{при } x \geq \frac{3}{2}\alpha - \gamma \\ 0 \frac{1}{n^{2(x-\alpha)}} & \text{при } x < \frac{3}{2}\alpha - \gamma \end{cases} \quad (3.1.43)$$

На основании исследований Бунке, можно получить множество разнообразных обобщений метода стохастической аппроксимации. Используя особые виды данного метода, удастся вычислить последовательности значений, которые аппроксимирует медиану функции распределения, корень уравнения $f(x) = a$, максимум функции $f(x)$, где $f(x)$ – четко определяемая медиана случайного значения $y = \omega, x$.

3.2 Практические вопросы использования одномерной стохастической аппроксимации

Рассмотрим первую и самую простую процедуру стохастической аппроксимации, метод Роббинса – Монро.

Будем исследовать функцию $y(x)$, зависящую от одной независимой переменной и обращающуюся в нуль в единственной точке $x = x$, т. е. $y(x) = 0$. Предположим далее, что $y(x) < 0$ при $x < x$ и $y(x) > 0$ при $x > x$. Пусть помехи отсутствуют, α придано некоторое значение. Тогда по величине $y(x)$, соответствующей этому x , можно судить о том, где находится выбранное значение x — слева или справа от корня x , и определить направление движения к x . Если интервал неопределенности ограничен, то наиболее эффективным методом поиска является метод дихотомии. В применении к поиску корня эта процедура последовательного деления пополам получила название метода Больцано.

Предположим теперь, что измерения $y(x)$ производятся с ошибками. В этом случае необходимо различать истинное, но не известное нам значение функции y и известную, но, вероятно, не точную оценку этого значения, вычисленную по результатам измерений с помехами. [31,32] Обозначим эту оценку через z . Она является случайной величиной, ибо даже при постоянном

значении x наблюдения z флуктуируют случайным образом от одного эксперимента к другому. Пронумеруем все эксперименты, начиная с первого, и будем снабжать индексом n величины, относящиеся к n -му испытанию. С учетом этого запись $z(x_n)$ обозначает результат n -го эксперимента при некотором заданном значении x .

Индекс n указывает на то, что результат опыта зависит не только от величины x , но в значительной мере и от случая. Проиллюстрируем это примером. Пусть x_i и x_j – два равных между собой значения x , относящиеся к разным опытам (т. е. $x_i = x_j$, но $j \neq i$). Очевидно, им соответствуют и равные значения функции, т. е. $y(x_i) = y(x_j)$; однако благодаря случайному характеру помех результаты опытов, вообще говоря, будут различны, т. е. $z(x_i) \neq z(x_j)$.

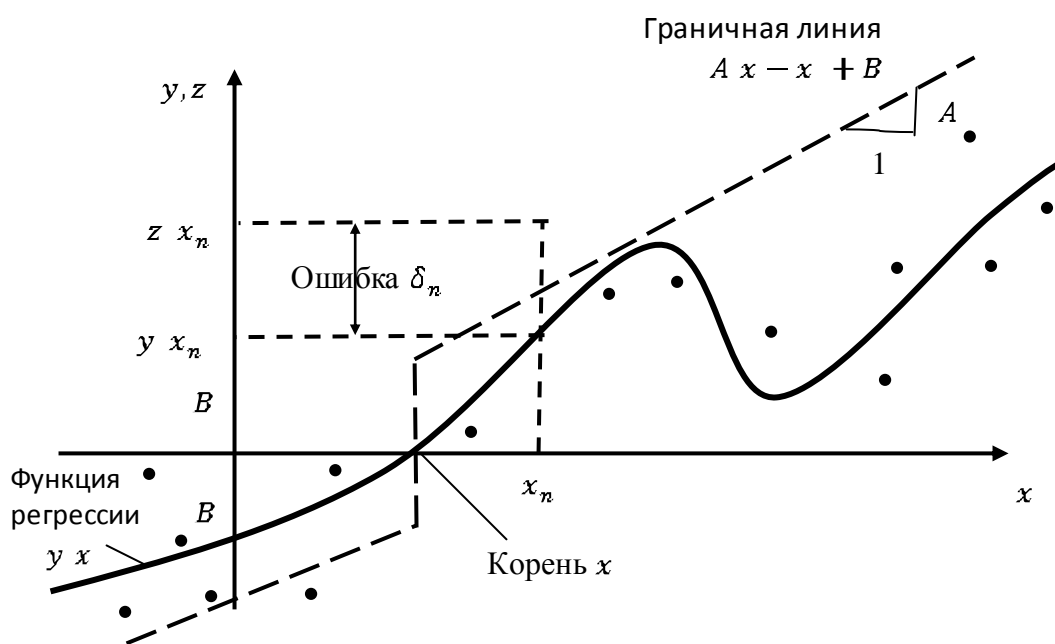


Рисунок 3.2.1 – Поиск корня в присутствии помех

Рисунок 3.2.1 иллюстрирует поведение функции $y(x)$, случайный характер наблюдений z и трудности, с которыми сталкивается экспериментатор при наличии помех. Сплошная линия обозначает исследуемую функцию $y(x)$, которую в статистике принято называть функцией регрессии. Результаты отдельных наблюдений изображены в виде точек, разбросанных вокруг этой линии.[27,29] Кроме того, на рисунке имеются точки, обведенные кружками.

Это – результаты наиболее неблагоприятных экспериментов, которые могут привести к поиску в неправильном направлении. Например, четыре такие точки, расположенные справа от x , лежат в области отрицательных значений и создают иллюзию того, что текущее значение x лежит левее x . В одном из экспериментов x фактически совпадает с величиной x . Однако его результат, искаженный помехами, создает впечатление того, что корень расположен левее. Впрочем, надо отметить, что большинство наблюдений, хотя и искажается помехами, но тем не менее указывает правильное направление поиска. [2]

В методе Роббинса – Монро независимая переменная x изменяется по закону

$$x_{n+1} = x_n - a_n z x_n , \quad (3.2.1)$$

где x_{n+1}, x_n – значения x соответственно в $(n + 1)$ -м и n -м экспериментах;

$z x_n$ – результат n -го эксперимента;

a_n – некоторый член последовательности положительных чисел, удовлетворяющей условию

$$\lim_{n \rightarrow \infty} a_n = 0. \quad (3.2.2)$$

Последовательность чисел a_n должна отвечать еще двум условиям. Сумма ее членов должна расходиться:

$$\sum_{n=1}^{\infty} a_n = \infty \quad (3.2.3)$$

а сумма квадратов членов – сходиться:

$$\sum_{n=1}^{\infty} a_n^2 < \infty \quad (3.2.4)$$

Этим условиям удовлетворяет, например, гармоническая последовательность чисел $\frac{1}{n}: 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$. Из всех последовательностей вида n^{-p} , отвечающих требованию расхождения суммы, эта последовательность обеспечивает наиболее быстрое сокращение длины шага.

Так как изменения на каждом шаге пропорциональны результату последнего наблюдения, то надо быть уверенным в том, что функция регрессии ограничена при всех конечных значениях x . Во избежание чрезмерно больших выбросов потребуем, чтобы функция $y(x)$ по обе стороны от корня была

ограничена прямыми линиями

$$y < Ax - x + B < \infty \quad (3.2.5)$$

где A и B – некоторые постоянные. Эти границы показаны на рисунке 3.2.1. Вообще говоря, данное требование не является чрезмерно жестким, так как на практике обычно известен некоторый конечный интервал, внутри которого лежит корень, и можно предположить, что в этом интервале данное условие удовлетворяется. Для доказательства сходимости процедуры Роббинса – Монро знание величин A и B необязательно. Однако если угловой коэффициент A известен хотя бы приближенно, то процесс поиска зачастую может быть ускорен.[19,38]

Роббинс и Монро показали, что их метод сходится в среднеквадратическом смысле. Это означает, что с увеличением числа наблюдений n математическое ожидание квадрата разности $(x_n - x)$ стремится к нулю, т. е.

$$\lim_{n \rightarrow \infty} E (x_n - x)^2 = 0. \quad (3.2.6)$$

Блему удалось усилить этот результат. Он показал, что при $n \rightarrow \infty$ последовательность x_n сходится к x с вероятностью единица, т. е.

$$p \lim x_n = x = 1 \quad (3.2.7)$$

где символ $p \{ \}$ означает вероятность события, стоящего в скобках.

Далее будут рассмотрены некоторые модификации алгоритмов стохастической аппроксимации.

Двухступенчатый метод. Кохран и Дэвис, обрабатывая статистически результаты биологических экспериментов, показали, что можно уменьшить среднеквадратичную ошибку оценки, используя двухступенчатую процедуру стохастической аппроксимации. Математически это сформулировано Вентером, доказавшим сходимость последовательности оценок к искомому значению с вероятностью единица.[3,27]

Ускоренная стохастическая аппроксимация. Если оценка величины θ находится в окрестности θ , то, применяя метод Роббинса – Монро, можно

ожидать, что в большинстве случаев она будет флуктуировать близ θ . Если же оценка не находится в окрестности θ , то нет оснований ожидать, что это будет так. Используя эту идею, Кестен предложил метод ускоренной стохастической аппроксимации.

Блум и Буркхольдер разработали свои итерационные методы и доказали сходимость с вероятностью единица. Одель предложил выбирать коэффициенты a_n на основании наблюдаемых значений Y_n .

Стохастическая аппроксимация при наличии тренда. Дупач модифицировал метод Роббинса – Монро для случая, когда оптимальное значение θ_n меняется в процессе аппроксимации. Этот результат может быть доказан обычным методом.

Модификация Фабиана. Фабиан предложил следующим образом изменить этот метод. Рекуррентное соотношение для процедуры Роббинса – Монро может быть переписано так:

$$x_{n+1} = x_n - a_n Y_n - \alpha \operatorname{sign} Y_n - \alpha .$$

Из него видно, что направление n -го шага процесса аппроксимации определяется в соответствии с $\operatorname{sign} Y_n - \alpha$, а длина шага равна $a_n Y_n - \alpha$. Такой выбор приемлем, если для больших значений $Y_n - \alpha$ можно ожидать большие значения $Y_n - \alpha$, однако это не гарантируется предположениями, в которых применяется метод Роббинса – Монро. Таким образом, можно изменить рекуррентное соотношение следующим образом:

$$x_{n+1} = x_n - a_n (\operatorname{sign} Y_n - \alpha) .$$

Вместо умножения a_n на $Y_n - \alpha$ предлагается умножать его на $\operatorname{sign} Y_n - \alpha$. Эта процедура оказывается полезной при условии, что оценка x_n находится в окрестности θ , а $Y_n - \alpha$ больше 1.

Спустя год после публикации статьи Роббинса и Монро появилась стохастический метод Кифера – Вольфовица, предназначенная специально для поиска максимума. Вместо нахождения производной $y'(x_n)$ в точке x_n они предложили вычислять средний угловой коэффициент

$$\frac{z(x_n + c_n) - z(x_n - c_n)}{2c_n} \quad (3.2.8)$$

где $z(x_n + c_n)$ и $z(x_n - c_n)$ – два наблюдения, соответствующих значениям $x_n + c_n, x_n - c_n$;

c_n – некоторая постоянная константа.

Эту идею иллюстрирует рисунок 3.2.2, на котором изображены типичная унимодальная функция и результаты измерений при наличии помех. Знак этого углового коэффициента определяет перспективное направление дальнейшего поиска.

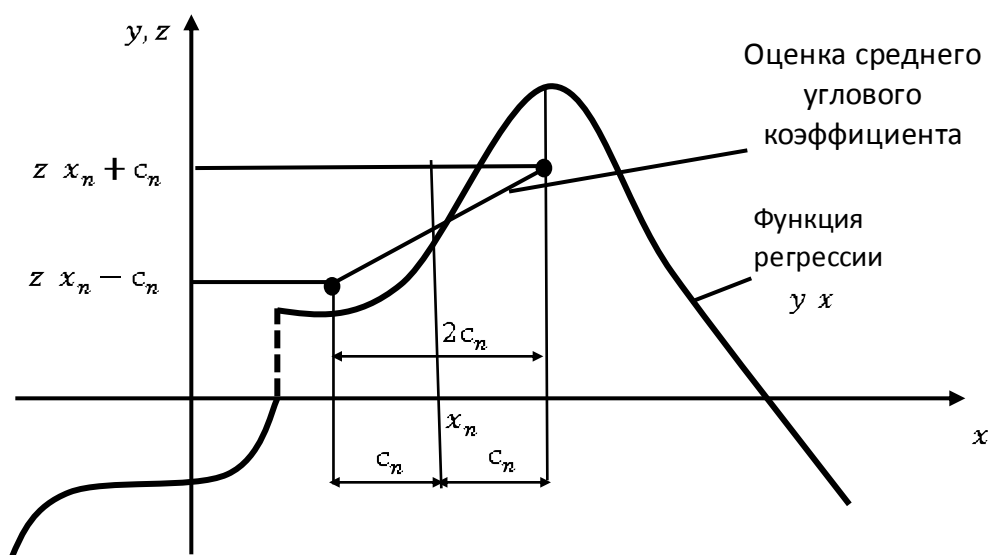


Рисунок 3.2.2 Оценка среднего углового коэффициента по методу Кифера – Вольфовица

В соответствии с методом Кифера – Вольфовица координата центра x_{n+1} следующей пары замеров вычисляется так:

$$x_{n+1} = x_n + \frac{a_n [z(x_n + c_n) - z(x_n - c_n)]}{c_n} \quad (3.2.9)$$

где a_n – некоторый член последовательности положительных чисел, определяющих длину шага;

c_n – расстояние между двумя предыдущими наблюдениями. В пределе и длина шага $2a_n$, и расстояние $2c_n$ стремятся к нулю, т. е.

$$\lim_{n \rightarrow \infty} a_n = 0, \quad (3.2.10)$$

$$\lim_{n \rightarrow \infty} c_n = 0, \quad (3.2.11)$$

так что процесс асимптотически сходится. Чтобы величина суммарного корректирующего воздействия была достаточна для достижения точки максимума, последовательность шагов должна удовлетворять условию

$$\sum_{n=1}^{\infty} a_n = \infty. \quad (3.2.12)$$

С другой стороны, для подавления помехи требуется, чтобы

$$\sum_{n=1}^{\infty} \left(\frac{a_n}{c_n}\right)^2 < \infty. \quad (3.2.13)$$

В первоначальной статье Кифер и Вольфовиц наложили на a_n и c_n еще ряд условий, как оказалось впоследствии, излишних.

Если функция регрессии y и помехи удовлетворяют некоторым разумным ограничениям, то процесс сходится к точке максимума унимодальной функции как в среднем квадратическом, так и с вероятностью единица. Как и в процедуре Роббинса – Монро, ошибка эксперимента должна быть равномерно ограниченной и иметь нулевое математическое ожидание. От функции регрессии y требуется, чтобы значения среднего углового коэффициента для любой пары наблюдений лежали в пределах некоторого сектора, ограниченного прямыми линиями, т. е. при всех $x_1 \neq x_2$ должно выполняться условие

$$y(x_2) - y(x_1) < A(x_2 - x_1) + B < \infty, \quad (3.2.14)$$

где x^* – истинная координата вершины;

A и B – некоторые постоянные.

Это ограничение аналогично уравнению в методе Роббинса – Монро; его назначение – исключить колебательность при поиске. Приведенные условия являются весьма слабыми и не вызывают на практике затруднений; им удовлетворяют даже функции вида $y = -x^2$ и $y = \exp -x^2$.

3.3 Применение стохастической аппроксимации для адаптации регрессионной модели

Существует множество разнообразных процессов, каждый из которых описывает соответствующая модель. Модель, описывающая объект, меняется, адаптируясь к изменениям в тенденциях. В качестве предмета адаптации выступают коэффициенты данной модели.

Адаптация коэффициентов модели будет продолжаться до тех пор, пока вычисленное значение показателя Y не приблизится к оптимальному значению U . Величина U является фактическим значением, которое подвергается воздействию различных факторов. Описание качественного состояния изменившейся системы является сутью адаптации регрессионной модели методом стохастической аппроксимации.

Для нахождения ожидаемого результата адаптации, следует учитывать, что показатель Y формируется при помощи трех составляющих:

1. детерминированной Y ;
2. случайной ε ;
3. неопределенной μ .

Учитывать все составляющие, при построение модели не представляется возможным. Следовательно исследуемый процесс двух составляющих: постоянной составляющей, представляющей из себя саму модель, и погрешности аппроксимации, описывающей воздействие случайных процессов, и воздействие неизвестных факторов и процессов:

$$Y = Y + \varepsilon \quad (3.3.1)$$

Погрешность, или ошибка аппроксимации ε определяет точность получаемых значений методом стохастической аппроксимации. Требование сведения погрешности вычислений к нулю не является необходимым, достаточно чтобы значение погрешности не должно превышать заданного значения η :

$$\varepsilon > \eta \quad (3.3.2)$$

Сутью адаптации методом стохастической аппроксимации является изменение коэффициентов модели таким образом, чтобы вычисленные

значения вновь соответствовали реальным значениям в заданных границах, обусловленных действием случайных факторов. [3,25,26]

Для адаптации регрессионных моделей воспользуемся методом Роббинса – Монро. Для вычисления точного управляющего значения x^* , выбираем убывающую с ростом числа испытаний n последовательность положительных значений y_n . Требуется за конечное число итераций, отыскать такое значение x^* , которое принадлежит множеству X , чтобы:

$$Y x^* = U \quad (3.3.3)$$

Для нахождения значения x используется метод стохастической аппроксимации Роббинса – Монро:

$$x_n = x_{n-1} + y_n (U - Y(x_{n-1})) \quad (3.3.4)$$

Положительное значение y_n является «параметром демпфирования колебаний». Способ задачи параметров демпфирования колебаний описывает характеристики алгоритма стохастической аппроксимации, одним из которых является скорость сходимости алгоритма к оптимальному значению.

Исследованию процессов адаптации посвящено множество работ, проведённых значительным числом специалистов в области математики и технической кибернетики, построенных на основе алгоритма стохастической аппроксимации Роббинса – Монро. Доказано, что если

$$y_n = \infty, \quad y_n^2 < \infty \quad (3.3.5)$$

то x стремится к x^* .

При использовании разных условий задания величины демпфирования колебаний, получаем разные методы адаптации:

- 1) метод с фиксированным шагом:

$$y_n = y < 1 \quad (3.3.6)$$

- 2) метод с переменным шагом, в котором количество испытаний n определяет значение демпфирования колебаний:

$$y_n = f(n) \quad (3.3.7)$$

3) метод с нелинейным шагом, когда значение демпфирования колебаний вычисляется на основе значений $Y[n]$ и $x[n]$, которые являются приближенными к оптимуму значениями:

$$y_n = F(Y, x, n) \quad (3.3.8)$$

Впервые на практике был использован алгоритм адаптации с переменным шагом. В этом случае задание параметров демпфирования колебаний может происходить различными способами. [14]

Преимуществом данного способа является простота и формализм. Изменение параметров зависит от шага аппроксимации, а не от результата вычисления.

Рассмотрим пример регрессионной однофакторной модели вида:

$$Y_t = f(x_t, a_i), i = 0, 1, 2, \dots, m - 1 \quad (3.3.9)$$

где a_i – коэффициенты модели;

m – количество коэффициентов модели;

x_t – фактор, оказывающий влияние на показатель Y_t .

Выражаем каждый коэффициент модели через значения x_t, Y_t и остальные коэффициенты. В результате получаем:

$$a_i = F(Y_t, x_t, a_0, a_1, \dots, a_{m-1}) \quad (3.3.10)$$

Теперь в некоторый момент времени наблюдения t , фактическое значение Y_t подставляется в выражение (3.3.10), заменяя тем самым полученное значения Y_t . Результатом замены элемента Y_t , является нахождение коэффициент a_{it} , который отличается от вычисленного коэффициента a_i . При подстановке вычисленного коэффициента a_{it} в исходную модель, полученное выражение описывает фактическое наблюдение:

$$a_{it} = F(Y_t, x_t, a_0, a_1, \dots, a_{m-1}) \quad (3.3.11)$$

Коэффициенты модели a_{it} , полученные таким способом, будут отличны от вычисленных ранее значений a_i . Коэффициенты a_{it} , вычисленные при помощи выражения (3.3.11) будут называться фактическими. Если построенная регрессионная модель описывает фактические данные с малой точностью, то

значения вычисленные коэффициенты будут значительно отличаться от фактических коэффициентов модели.

В таком случае требуется подкорректировать коэффициенты модели таким образом, чтобы их вычисленные значения стремились к фактическим. При адаптации регрессионной модели, полученные значения коэффициентов изменяется во времени, следовательно необходимо определить индекс t для обозначения этих коэффициентов, то есть будем их обозначать a_{it} .

Адаптация модели во времени происходит при помощи метода стохастической аппроксимации Роббинса – Монро:

$$a_{it} = a_{it-1} + \gamma (y_{it} - a_{it-1}) \quad (3.3.12)$$

где $a_{it} = a_{it-1}$;

N – номер коэффициента адаптации на предыдущем наблюдения.

С целью адаптации регрессионных моделей методом стохастической аппроксимации, может быть использован любой метод нахождения значения демпфирования колебаний.[24] Оптимальным методом для аддитивных регрессионных моделей является метод нахождения значения демпфирования колебаний с постоянным шагом, который имеет вид:

$$y_{it} = k_i \frac{\varepsilon_t - \eta}{\varepsilon_t} \quad (3.3.13)$$

где k_i – весовой коэффициент, определяющий степень адаптации вычисляемого коэффициента регрессионной модели в сравнение с другими коэффициентами. На значение весового коэффициента накладывается ряд ограничений. Одним из ограничения, является требование того, чтобы сумма всех этих коэффициентов равнялась единице:

$$\sum_i k_i = 1 \quad (3.3.14)$$

Поскольку коэффициенты регрессионной модели должны адаптироваться равномерно, поэтому необходимо, чтобы весовой коэффициент k_i был одинаковым для всех коэффициентов модели. Поэтому значение параметра демпфирования колебаний для всех коэффициентов рассчитывается по следующей формуле:

$$y_{it} = y_t = \frac{1}{m} \frac{\varepsilon_t - \eta}{\varepsilon_t} \quad (3.3.15)$$

Полученные по формуле (3.3.15) значения параметров демпфирования колебаний являются оптимальными, так как адаптация коэффициентов происходит за один шаг.

На основании логики адаптации регрессионных прогнозных моделей, адаптация коэффициентов модели происходит в том случае, если в какой то момент времени t , фактические значения и вычисленные значения коэффициентов a_{0t-1} и a_{1t-1} превышают допустимое значение:

$$\varepsilon_t > \eta \quad (3.3.16)$$

где $\varepsilon_t = Y_t - (a_{0t-1} + a_{1t-1}x_t)$.

Посчитаем для данного наблюдения параметр демпфирования колебаний:

$$y_t = \frac{1}{2} \frac{\varepsilon_t - \eta}{\varepsilon_t} \quad (3.3.17)$$

Выразим каждый коэффициент линейной однофакторной регрессионной модели через Y_t, x_t и оставшийся коэффициент. Вычисление коэффициента a_0 данной модели в момент времени t имеет вид:

$$a_{0t} = Y_t - a_{1t}x_t \quad (3.3.18)$$

Вычисление коэффициента a_1 :

$$a_{1t} = \frac{Y_t - a_{0t}}{x_t} \quad (3.3.19)$$

Следующим шагом подставим формулы вычисления коэффициентов в формулы алгоритма Роббинса-Монро:

$$a_{0t} n = a_{0t} n - 1 + y_t (Y_t - a_{1t-1}x_t - a_{0t} n - 1) \quad (3.3.20)$$

$$a_{1t} n = a_{1t} n - 1 + y_t \left(\frac{Y_t - a_{0t-1}}{x_t} - a_{1t} n - 1 \right) \quad (3.3.21)$$

В данном случае используем коэффициенты a_{0t-1} и a_{1t-1} как коэффициенты на начальном шаге адаптации $a_{0t} 0 = a_{0t-1}$ и $a_{1t} 0 = a_{1t-1}$. Прделав данную замену получаем следующие формулы:

$$a_{0t} n = a_{0t} n - 1 + y_t (Y_t - a_{1t-1}x_t - a_{0t-1}) \quad (3.3.22)$$

$$a_{1t} n = a_{1t} n - 1 + y_t \left(\frac{Y_t - a_{0t-1} - a_{1t-1}x_t}{x_t} \right) \quad (3.3.23)$$

Выражение, полученное в скобках, является ошибкой аппроксимации ε_t на текущем шаге. Поскольку адаптация модели происходит за один шаг, получаем несложную формулу для нахождения адаптированных коэффициентов:

$$a_{0t} = a_{0t-1} + y_t \varepsilon_t \quad (3.3.24)$$

$$a_{1t} = a_{1t-1} + y_t \left(\frac{\varepsilon_t}{x_t} \right) \quad (3.3.25)$$

На основании полученных выражений, становится ясным смысл метода адаптации регрессионной модели с помощью метода стохастической аппроксимации. Происходит подтягивание вычисленных значений к фактическим значениям на расстояние допустимой погрешности, равное η .

Исходя из произведенных вычислений была построена схема алгоритма адаптации модели представленная на рисунке 3.3.1. На начальном этапе происходит оценка коэффициентов модели на всём имеющемся множестве наблюдений.

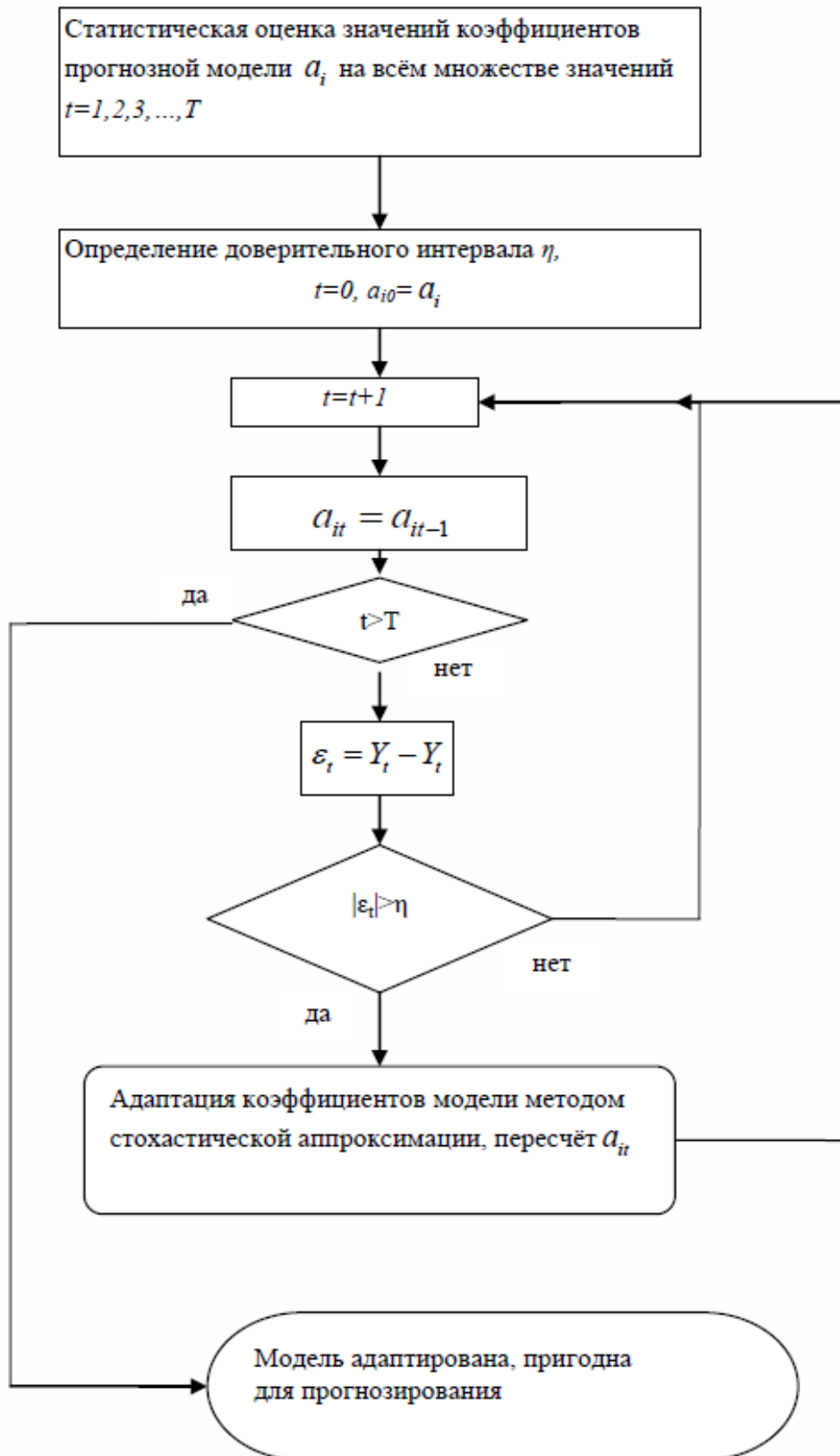


Рисунок 3.3.1. – Схема алгоритма адаптации модели

Модель должна хорошо описывать исходные данные. Для этого задается величина доверительной границы модели η , в рамках которой отклонения модели от фактических значений объясняются действием случайных величин, а выход за эти рамки служит основанием для адаптации.[20-23]

Последние адаптированные коэффициенты модели представляют ту модель, которая адаптируется в случаях изменения тенденций, если они были.

В том случае если тенденции не изменялись, то коэффициенты модели остаются в изначальном виде.

3.4 Реализация алгоритма стохастической аппроксимации для адаптации регрессионной модели

Вторым используемым алгоритмом адаптации является алгоритм стохастической аппроксимации. Алгоритм реализован на языке программирования Java. Для адаптации будет использована исходная регрессионная модель.

Перед началом адаптации, требуется задать допустимую погрешность адаптации. Для используемой регрессионной модели данное значение составляет $\eta = 0.28$. Адаптация алгоритмом стохастической аппроксимации означает последовательную корректировку коэффициентов адаптируемой модели. Функция адаптации данным алгоритмом представлена на рисунке 3.4.1. Полный листинг программы представлен в приложение А.


```

for (int i=0; i<n;i++)
{
    kofa[j]=kof[0];
    kofb[j]=kof[1];
    j++;
    sum=Math.abs(y[i]-(x[i]*kof[0]+kof[1]));
    sum=Math.round(sum*100)/100.0d;

    while(sum>Eps)
    {
        kofa[j]=kofa[j-1]
            +0.5*Math.abs((Math.abs(y[i]-(x[i]*kofa[j-1]+kofb[j-1]))-Eps)
                /(y[i]-(x[i]*kofa[j-1]+kofb[j-1])))
            *((y[i]-(x[i]*kofa[j-1]+kofb[j-1]))/x[i]);
        kofb[j]=kofb[j-1]
            +0.5*Math.abs((Math.abs(y[i]-(x[i]*kofa[j-1]+kofb[j-1]))-Eps)
                /(y[i]-(x[i]*kofa[j-1]+kofb[j-1])))
            *((y[i]-(x[i]*kofa[j-1]+kofb[j-1])));
        kof[0]=kofa[j];
        kof[1]=kofb[j];
        j++;
        sum=Math.abs(y[i]-(x[i]*kof[0]+kof[1]));
        sum=Math.round(sum*100)/100.0d;
    }
    j=0;
}
kof[0]=kofa[1];
kof[1]=kofb[1];

```

Рисунок 3.4.1 – Функция адаптации методом стохастической аппроксимации

Последовательно вычисляя коэффициенты таким образом, чтобы погрешность получаемых значений модели были меньше заданной точности. Результатом вычислений является аппроксимированная модель готовая для дальнейшего использования в прогнозирование. Вычислив коэффициенты мы получаем аппроксимированную модель, которая представлена на рисунке 3.4.2.

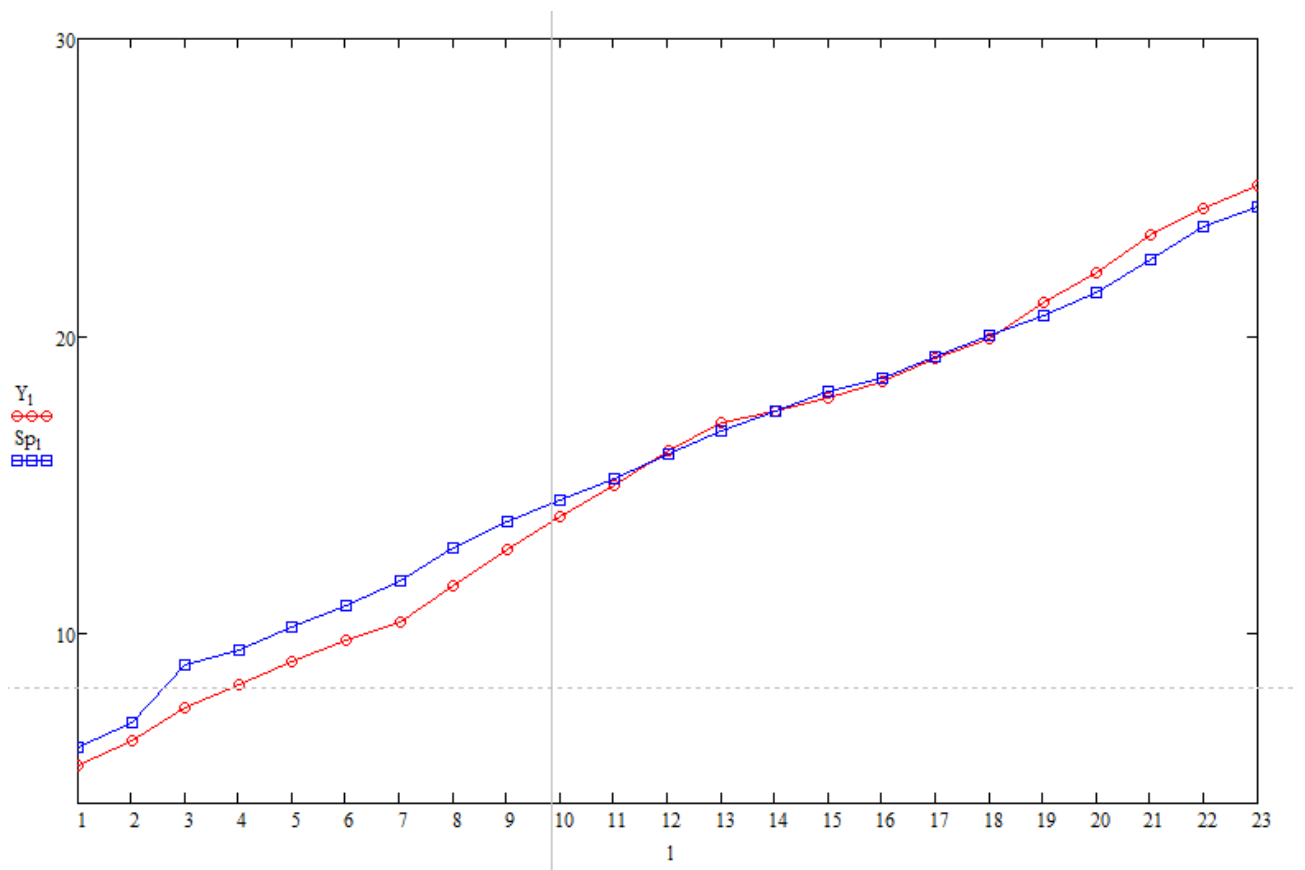


Рисунок 3.4.2 – Сравнение адаптированной модели с исходными данными

На рисунке 3.4.2 график Y_t – построен по исходным данным, график Sp_t – построен моделью, адаптированной алгоритмом стохастической аппроксимации, ось абсцисс – номер наблюдения, ось ординат – значение модели. Модель, полученная алгоритмом стохастической аппроксимации, сохраняет суть исходной модели, при этом «подгоняя» адаптированную модель под новые данные. Сравнение исходной регрессионной модели и модели полученной при помощи алгоритма стохастической аппроксимации представлено на рисунке 3.4.3.

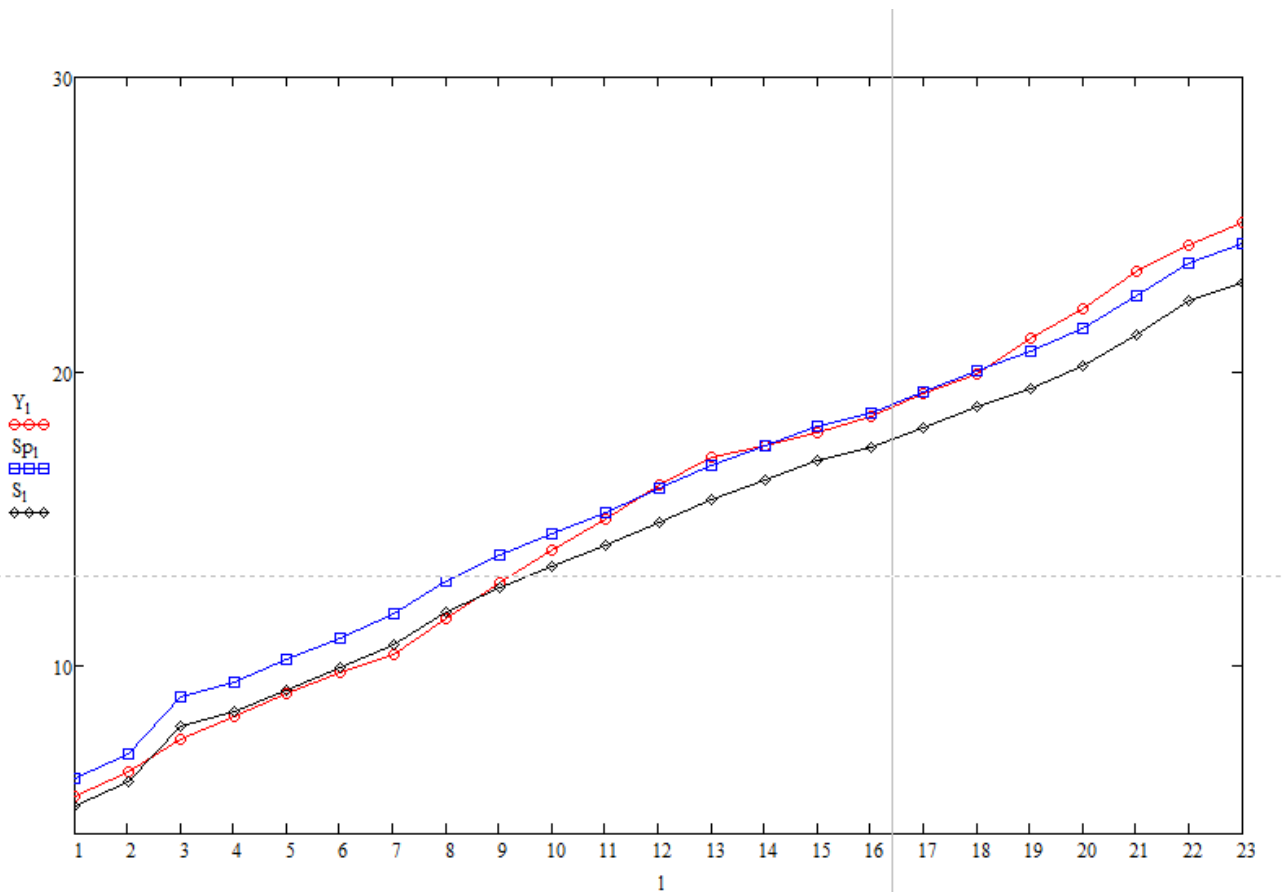


Рисунок 3.4.3 – Сравнение исходной модели с аппроксимированной

На рисунке 3.4.3 график Y_l – построен по исходным данным, график S_l – построен по исходной математической модели, график Sp_l – построен по аппроксимированной математической модели, ось абсцисс – номер наблюдения, ось ординат – значение модели. Исходя из полученных результатов, модель, адаптированная алгоритмом стохастической аппроксимации, целесообразнее использовать для дальнейшего использования с целью прогнозирования. Поскольку адаптированная регрессионная модель корректируется с учётом данных, которые были получены после построения исходной модели.

3.5 Сравнительный анализ эффективности адаптационных алгоритмов

С целью определения наиболее эффективного алгоритма адаптации, проведем сравнительный анализ моделей, полученных при помощи метода Брауна и алгоритма стохастической аппроксимации.

Проведем сравнительный анализ полученных моделей с исходной моделью, графики которых показаны на рисунке 3.5.1.

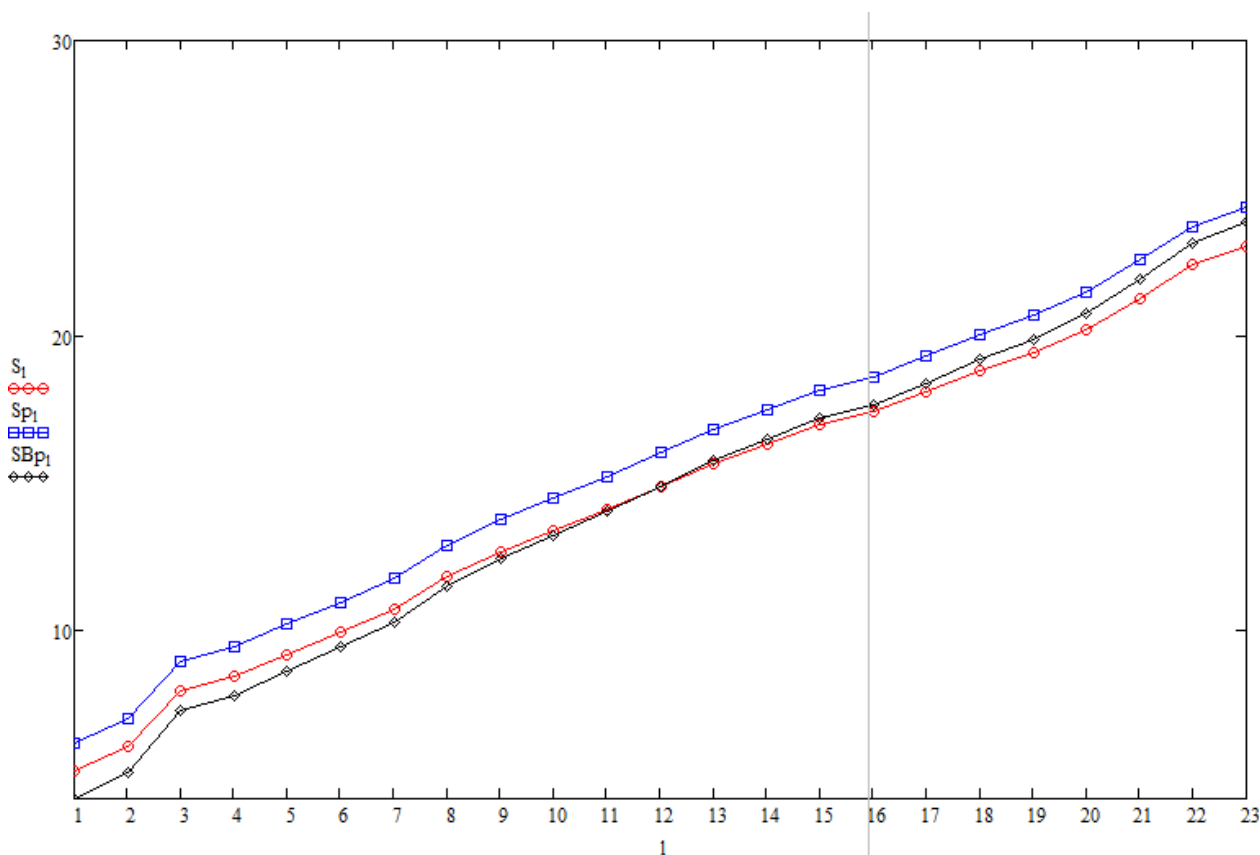


Рисунок 3.5.1 – Сравнение адаптированных моделей с исходной моделью

На рисунке 3.5.1 график S_t – исходная регрессионная модель, график Sp_t – построен моделью, адаптированной алгоритмом стохастической аппроксимации, график SBp_t – построен моделью, адаптированной методом Брауна, ось абсцисс – номер наблюдения, ось ординат – значение модели.

Модель, адаптированная стохастической аппроксимацией, является исходной моделью с изменёнными коэффициентами, что сохраняет исходную суть модели. Это является достоинством алгоритма стохастической аппроксимации.

В свою очередь метод Брауна требует построения новой модели, что требует больше времени и не всегда возможно, так как исходные данные могут быть недоступны – что в свою очередь является большим недостатком метода Брауна.

Проведем сравнительный анализ адаптированных моделей с точными исходными данными, графики которых представлены на рисунке 3.5.2.

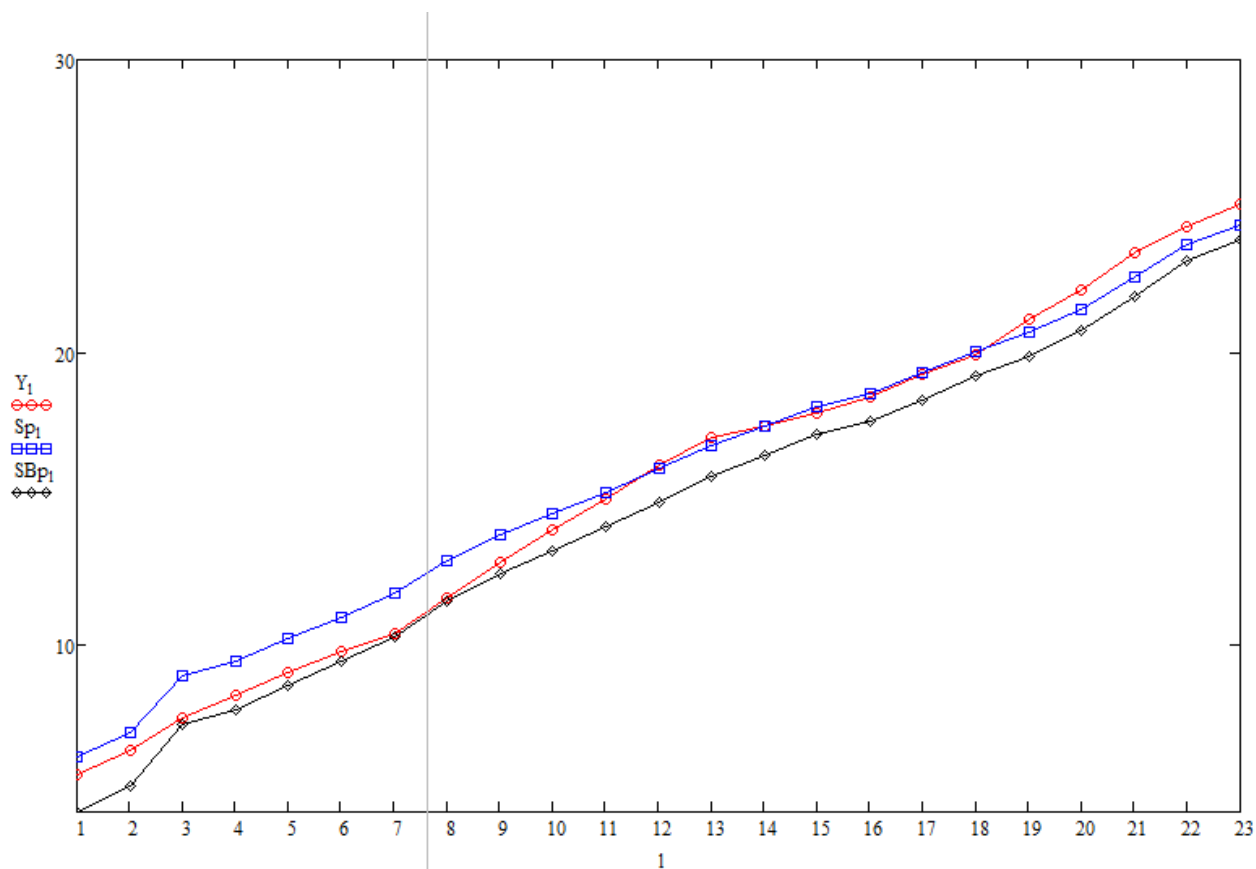


Рисунок 3.5.2 – Сравнение адаптированных моделей с исходными данными

На рисунке 3.5.2 график Y_l – построен по исходным данным, график Sp_l – построен моделью, адаптированной алгоритмом стохастической аппроксимации, график SBp_l – построен моделью, адаптированной методом Брауна, ось абсцисс – номер наблюдения, ось ординат – значение модели.

Модель, адаптированная методом Брауна, при построение требует задавать коэффициент α , который отображает влияние исходных и новых данных на новую модель. Нахождение этого коэффициента, с учетом того, что для большей точности он должен быть оптимальным, является ещё одним недостатком метода Брауна.

В свою очередь, модель, адаптированная алгоритмом стохастической аппроксимации, изменяется только под влиянием новых данных. Исходные данные не оказывают влияние на алгоритм стохастической аппроксимации, что является достоинством данного алгоритма. Ещё одним достоинством, по

сравнению с методом Брауна является то, что точность полученной модели зависит от задаваемой ошибки адаптации, с которой происходит сравнение адаптируемой модели с фактическими значениями.

Сравнительный анализ адаптационных алгоритмов, модели которых показаны на рисунке 3.5.2 показал, что прогнозы, полученные моделью, адаптированной алгоритмом стохастической аппроксимации, являются более точными, чем полученные моделью, адаптированной методом Брауна. При учете всех достоинств и недостатков используемых алгоритмов адаптации, рекомендуется использовать алгоритм стохастической аппроксимации, с целью адаптации регрессионных моделей. Такие регрессионные модели будут давать более эффективный результат при меньших затратах в сравнение с моделями, адаптированными методом Брауна.

В данной главе был рассмотрен метод стохастической аппроксимации, его теоретические основы и практическое применение для адаптации регрессионных моделей. Показаны модификации метода стохастической аппроксимации. Описывается программная реализация данного метода и результаты полученной адаптированной регрессионной модели. Проведен сравнительный анализ метода Брауна и метода стохастической аппроксимации, результатом которого является то, что значения, получаемые моделью, адаптированной методом стохастической аппроксимации, являются значительно точнее значений, полученных методом Брауна. Что доказало эффективность метода стохастической аппроксимации для адаптации регрессионной модели.

ЗАКЛЮЧЕНИЕ

Регрессионные модели широко применяются для прогнозирования разных явлений. Данная работа посвящена регрессионным моделям. В частности, их адаптации.

Поставленная задача, состоящая в анализе алгоритма построения регрессионной модели, достигнута.

В первой главе проведен анализ классической линейной модели множественной регрессии и предпосылки, описывающие требования к её построению. Проведен обзор метода построения регрессионной модели. Результатом проведённого анализа стали выводы о том, что с течением времени и появлением новой информации, регрессионная модель описывающая процесс, становится непригодной для использования. Использование такой регрессионной модели дает неточные результаты. Для предотвращения такого результата, регрессионную модель требуется адаптировать.

Проблема адаптации моделей является одной из основных в современных исследованиях, т.к. она характеризует точность получения результатов моделируемого процесса.

Поставленная задача, состоящая в использовании метода Брауна для адаптации регрессионной модели, достигнута.

Во второй главе описывается первый из использованных алгоритмов адаптации – метод Брауна. Этот метод позволяет адаптировать исходную модель и осуществлять прогнозирование. Адаптация методом Брауна строится на значение постоянной сглаживания, которое отображает уровень влияния данных на адаптируемую модель. Этот фактор является недостатком, потому что необходимо подбирать значение постоянной сглаживания таким образом, чтобы получить оптимальный результат. Ещё одним значимым недостатком метода Брауна является необходимость построения новой модели при поступлении новых данных.

Поставленная задача, состоящая в использовании алгоритма стохастической аппроксимации для адаптации регрессионной модели и осуществления сравнительного анализа адаптационных алгоритмов, достигнута.

В третьей главе описывается второй метод адаптации – алгоритм стохастической аппроксимации. Использование этого алгоритма возможно в условиях неполноты информации. Адаптация методом стохастической аппроксимации сводится к адаптации коэффициентов регрессионной модели, при поступлении новых данных. Таким образом исходная суть модели не подвергается изменениям, что в свою очередь является достоинством метода. Вывод, полученный из анализа работы метода стохастической аппроксимации является то, что он не требует построения новой регрессионной модели. Отклонение прогнозируемых значений, полученных адаптированной моделью, от фактических данных зависит от задаваемой точности метода.

Метод стохастической аппроксимации решает проблему получения более точных результатов с помощью регрессионной модели. Поэтому используемый метод является предпочтительным для адаптации регрессионных моделей. Это позволяет сократить затраты ресурсов на построение новых моделей и получить оптимальный результат адаптации.

Проведен сравнительный анализ метода Брауна и метода стохастической аппроксимации. Регрессионная модель, используемая для адаптации строилась на основе условных данных. Результатом сравнения является то, что прогнозируемые значения, получаемые моделью, адаптированной методом стохастической аппроксимации, находятся ближе к фактическим значениям, чем прогнозные значения, полученные методом Брауна. Это доказывает эффективность метода стохастической аппроксимации для адаптации регрессионной модели. Цель, поставленная в данной работе достигнута.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

Научная и методическая литература

1. Афанасьев В.Н. Анализ временных рядов и прогнозирование: учебник / В.Н. Афанасьев, Юзбашев М.М. – М.: Финансы и статистика, 2011. – 228 с.
2. Граничин О.Н. Рандомизированные алгоритмы оптимизации и оценивания при почти произвольных помехах / О.Н. Граничин, Б.Т. Поляк; Отв. Ред. А.В. Назин. – М.: Наука, 2013. – 291 с.
3. Дуброва Т.А. Статистические методы прогнозирования в экономике / Т.А. Дуброва – М.: Московская финансово-промышленная академия, 2014. – 60с.
4. Кремер Н.Ш. Эконометрика: учебник / Н.Ш. Кремер, Б.А. Путко – М.: ЮНИТИ-ДАНА, 2012. – 311 с.
5. Левицкий Е.М. Адаптивные эконометрические модели: учебник / Е.М. Левицкий. –Новосибирск: Наука, 2012. – 224с.
6. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособие / Ю.П. Лукашин. – М.:Финансы и статистика, 2013. – 416 с.
7. Мхитарян В.С. Эконометрика: учебник / В.С. Мхитарян, М.Ю. Архипова, В.А. Балаш, О.С. Балаш, Т.А. Дуброва, В.П. Сиротин. – М.: Проспект, 2011. – 384 с.
8. Мхитарян В.С. Эконометрика: учебно-методический комплекс / В.С. Мхитарян, М.Ю. Архипова, В.П. Сиротин. – М.: Изд. Центр ЕАОИ. 2010. - 144 с.
9. Семенов А.Д. Идентификация объектов управления: учеб. пособие / А.Д. Семенов, Д.В. Артамонов, А.В. Брюхачев. – Пенза: Изд-во Пенз. гос. ун-та. 2013. – 211 с.

10. Светуныхов С.Г. Методы социально-экономического прогнозирования: учебник / С.Г. Светуныхов, И.С. Светуныхов. – СПб.: Изд-во СПбГУЭФ, 2010. – 103с.

11. Светуныхов С.Г. Запредельные случаи метода Брауна в экономическом прогнозировании / С.Г. Светуныхов, А.В. Бутуханов, И.С. Светуныхов. – СПб.: Изд-во СПбГУЭФ. 2016. – 71с.

12. Уайлд Д.Дж. Методы поиска экстремума: монография / Д.Дж. Уайлд. – М.: Наука. Главная редакция физико-математической литературы, 2011. – 268с.

13. Фёрстер Э. Методы корреляционного и регрессионного анализа: учебник / Э. Фёрстер, Б. Рёнц. – М.: Изд-во «Финансы и статистика», 2012. – 302 с.

14. Фукунага К. Введение в статистическую теорию распознавания образов: монография / К. Фукунага. – М.: Наука. Главная редакция физико-математической литературы, 2013. -368с.

Электронные ресурсы

15. Балдин А.В. Стохастическая аппроксимация в моделях управления транспортными машинами /А.В. Балдин, В.Б. Борисевич, В.Ю. Строганов. 2010. [Электронный ресурс]: <http://vestniken.ru/articles/354/html/files/assets/basic-html/page1.html>

16. Граничин О.Н. Поисковые алгоритмы стохастической аппроксимации с рандомизацией на входе/ О.Н. Граничин. 2015. [Электронный ресурс]: <http://www.math.spbu.ru/user/gran/papers/GranPol2015.pdf>

17. Домбровский В.В. Эконометрика / В.В. Домбровский. 2016. [Электронный ресурс]: <http://sun.tsu.ru/mminfo/2016/Dombrovski/start.htm>

18. Светуныхов С.Г. Самообучающаяся модель краткосрочного прогнозирования социально-экономической динамики / С.Г. Светуныхов. 2010. [Электронный ресурс]:<https://www.hse.ru/sci/publications/27286573.html>

19. Хуан В. Стохастическая оптимизация сопряженных градиентов на основе метода наименьших квадратов для задач сейсмической инверсии /В.

Хуан, Х. В. Чжоу. 2014. [Электронный ресурс]:
http://www.iongeo.ru/media/img/index/pdffiles/143/TP_SEG_Stochastic_Conjugate_Gradient_Method_rus.pdf

Литература на иностранном языке

20. Behaim M. Stochastic Approximations And Differential Inclusions. / M. Behaim, J. Hofbauer, S. Sorin. 2010. [Электронный ресурс]:
<http://members.unine.ch/michel.benaim/perso/bhs.pdf>

21. Broadie M. General Bounds and Finite-Time Improvement for the Kiefer-Wolfowitz Stochastic Approximation Algorithm. / M. Broadie, D. Cicek, A. Zeevi. 2009. [Электронный ресурс]:
http://www.columbia.edu/~mnb2/broadie/Assets/GeneralBoundsAndFiniteTimeImprovementOnSAA_OR_20100730.pdf

22. Combes R. An introduction to stochastic approximation. 2013. [Электронный ресурс]:
https://people.kth.se/~alepro/DistriOptCourse/lecture_stoch_approx.pdf

23. Galtier, N M. Relative entropy minimizing noisy non-linear neural network to approximate stochastic processes / M. N Galtier, C. Marini, G. Wainrib, H. Jaeger. 2014. [Электронный ресурс]: <https://hal.archives-ouvertes.fr/hal-00994652/document>

24. Granichin O. N. Randomized Algorithms for Stochastic Approximation under Arbitrary Disturbances. 2011. [Электронный ресурс]:
<http://www.jhuapl.edu/spsa/PDF-SPSA/Granichin%20Article.pdf>

25. Karagiannis G. Parallel and Interacting Stochastic Approximation Annealing algorithms for global optimization. / G. Karagiannis, B. A. Konomi, G. Lin, F. Liang. 2015. [Электронный ресурс]:
https://www.math.purdue.edu/~gkaragia/papers/Supplamentary_material.pdf

26. Khong S.Z. On sampled-data extremum seeking control via stochastic approximation methods. / S.Z. Khong, Y. Tan, D. Netic, C. Manzie. 2013. [Электронный ресурс]:

http://people.eng.unimelb.edu.au/manzniec/Publications%20pdfs/13_Conf_Khong_b.pdf

27. Liang F. A Resampling-Based Stochastic Approximation Method for Analysis of Large Geostatistical Data / F. Liang, Y. Cheng, Q. Song, J. Park, P. Yang. 2013. [Электронный ресурс]:

<http://homepages.math.uic.edu/~minyang/Big%20Data%20Discussion%20Group/Resampling-based%20stochastic%20for%20analysis%20of%20large%20geostatistical%20data.pdf>

28. Mokkadem A. Revisiting Revesz's stochastic approximation method for the estimation of a regression function. / A. Mokkadem, M. Pelletier, Y. Slaoul. 2009. [Электронный ресурс]: <http://alea.impa.br/articles/v6/06-03.pdf>

29. Nemirovski A. Robust Stochastic Approximation Approach To Stochastic Programming. / A. Nemirovski, A. Juditsky, G. Lan, A. Shapiro. 2009. [Электронный ресурс]: http://www.optimization-online.org/DB_FILE/2007/09/1787.pdf

30. Odense S. Universal Approximation Results for the Temporal Restricted Boltzmann Machine and the Recurrent Temporal Restricted Boltzmann Machine / S. Odense, R. Edwards. 2016. [Электронный ресурс]: <http://jmlr.org/papers/volume17/15-478/15-478.pdf>

31. Ollivier Y. Training recurrent networks online without backtracking / Y. Ollivier, C. Tallec, G. Charpian. 2015. [Электронный ресурс]: <http://www.yann-ollivier.org/rech/pubs/nobacktrack.pdf>

32. Sun Y. Consensus-Type Stochastic Approximation Algorithms. 2012. [Электронный ресурс]: http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1556&context=oa_dissertations

33. Xu Z. A robust recurrent simultaneous perturbation stochastic approximation training algorithm for recurrent neural networks. / Z. Xu, Q. Song, D.

Wang. 2014. [Электронный ресурс]:
<https://dr.ntu.edu.sg/bitstream/handle/10220/20401/A%20Robust%20Recurrent%20Simultaneous%20Perturbation%20Stochastic%20Approximation%20Training%20of%20Recurrent%20Neural%20Networks.pdf?sequence=1&isAllowed=y>

34. Yu M. Stochastic Approximation And A Nonlocally Weighted Soft-Constrained Recursive Algorithm For Blind Separation Of Reverberant Speech Mixtures. / M. Yu, J. Xin. 2010. [Электронный ресурс]:
http://www.math.uci.edu/~jxin/dcds_mengxin.pdf

ПРИЛОЖЕНИЕ А

Листинг программного кода:

```
// Исходные данные
    public static double [] y = {5.57, 6.36, 7.50, 8.28, 9.06, 9.74, 10.36, 11.60,
12.79, 13.92, 14.95, 16.12, 17.10, 17.49, 17.90, 18.48, 19.22, 19.91, 21.10, 22.10,
23.40, 24.30, 25.05};
    public static double [] x = {1.953, 2.055, 2.291, 2.350, 2.443, 2.535, 2.634,
2.773, 2.878, 2.965, 3.056, 3.153, 3.252, 3.334, 3.415, 3.469, 3.551, 3.644, 3.721,
3.819, 3.950, 4.090, 4.170};
    // Точность вычисления
    public static double Eps = 0.28;
    public static void main(String[] args) {
    // Массивы значений
    double[] kof;
    double[] kofBr;
    double[] ybr;
    // Вызов метода вычисления коэффициентов модели
    kof=MNK(x,y,10);
    // Вызов метода стохастической аппроксимации
    kof=Stocha(x,y,kof,15);
    // Вывод результатов
    System.out.println(kof[0]);
    System.out.println(kof[1]);

    // Вызов метода Брауна
    ybr=Braun(x,y,23);
    kofBr=MNK(x,ybr,20);
    // Вывод результатов
    System.out.println(kofBr[0]);
    System.out.println(kofBr[1]);
    }
    // Метода вычисления коэффициентов модели
public static double[] MNK (double[] x, double [] y, double n)
    {
    double a, b, bufx, bufx2, bufy;
    double [] kof = new double[2];
    a=0;
    b=0;
    bufx=0;
```

```

bufx2=0;
bufy=0;
for (int i=0; i<n; i++)
{
    //Вычисление промежуточных значений
    a=a+x[i]*y[i];
    bufx=bufx+x[i];
    bufy=bufy+y[i];
    bufx2=bufx2 + x[i]*x[i];
}
//Вычисление коэффициентов
a=(a*n-bufx*bufy)/(n*bufx2-bufx*bufx);
b=(bufy-a*bufx)/n;
kof[0]=a;
kof[1]=b;
return kof;
}
//Метод стохастической аппроксимации
public static double[] Stocha (double[] x, double [] y, double [] kof, double n)
{
    double[] kofa = new double[10]; double[] kofb = new double[10];
    double sum=0;
    kofa[0]=kof[0];
    kofb[0]=kof[1];
    int j=0;
    for (int i=0; i<n;i++)
    {
        // Задание начальной точки аппроксимации
        kofa[j]=kof[0];
        kofb[j]=kof[1];
        j++;
        sum=Math.abs(y[i]-(x[i]*kof[0]+kof[1]));
        sum=Math.round(sum*100)/100.0d;
        // Условие прекращения вычисления новых коэффициентов
        while(sum>Eps)
        {
            // Изменение коэффициентов модели
            kofa[j]=kofa[j-1]
                +0.5*Math.abs((Math.abs(y[i]-(x[i]*kofa[j-1]+kofb[j-1]))-Eps)
                /(y[i]-(x[i]*kofa[j-1]+kofb[j-1]))))

```

```

        *((y[i]-(x[i]*kofa[j-1]+kofb[j-1]))/x[i]);
        kofb[j]=kofb[j-1]
        +0.5*Math.abs((Math.abs(y[i]-(x[i]*kofa[j-1]+kofb[j-1]))-Eps)
        /((y[i]-(x[i]*kofa[j-1]+kofb[j-1]))))
        *((y[i]-(x[i]*kofa[j-1]+kofb[j-1]))));
        kof[0]=kofa[j];
        kof[1]=kofb[j];
        j++;
        sum=Math.abs(y[i]-(x[i]*kof[0]+kof[1]));
        sum=Math.round(sum*100)/100.0d;
    }
    j=0;
}
kof[0]=kofa[1];
kof[1]=kofb[1];
return kof;
}
//Метод Брауна
public static double [] Braun(double[] x, double [] y, int n)
{
    double[] yb = new double[n];
    double[] dif = new double[n];
    double dif1=0;
    double Ep = 0;
    double a = 2/(n+1);
    // Вычисление начального значения
    yb[0]=(y[0]+y[1]+y[2])/3;
    for (int i=1; i<n; i++)
    {
        // Вычисление прогнозных значений
        yb[i]=(1-a)*y[i-1]+a*yb[i-1];
        dif[i]=Math.pow((y[i]-yb[i]),2);
    }
    // Вычисление погрешности вычислений
    for (int i=0; i<n; i++)
    {
        dif1=dif1+dif[i];
    }
    Ep=Math.sqrt(dif1/(n-1));
    System.out.println(Ep);
    return yb;
}
}

```