

ДИФФЕРЕНЦИАЛЬНЫЕ
УРАВНЕНИЯ
И
ПРОЦЕССЫ УПРАВЛЕНИЯ
№ 3, 2022
Электронный журнал,
рег. Эл № ФС77-39410 от 15.04.2010
ISSN 1817-2172
<http://diffjournal.spbu.ru/>
e-mail: jodiff@mail.ru

Моделирование динамических систем

О вероятностных цепочках с полиномиальным ростом

Логинова Н.В

Eram systems Montenegro DOO Podgorica

natalia.loginowa@gmail.com

Аннотация. В статье представлен один из методов математического моделирования процесса распределения ресурсов – метод вероятностных цепочек. Эта модель представляет собой дискретную динамическую систему, заданную на симплексе вероятностных векторов. Наиболее детально изучены вероятностные цепочки с логистическим и линейно-логарифмическим ростом. В данной работе рассмотрены цепочки с полиномиальным ростом. Полиномы системы могут быть заданы не единственным способом. В статье рассматриваются произвольный и оптимизированный методы определения полиномов. Оптимизация достигается за счет учета взаимосвязи влияния доли ресурса на одной территории на долю ресурса на другой территории. В качестве примера рассмотрены показатели торговли товарами в % от ВВП Германии, Франции, Италии, Испании. В работе также были построены цепочки с логистическим и линейно-логарифмическим ростом. Полученные результаты моделирования проанализированы с помощью коэффициента корреляции. Показано, что наибольшее значение указанный коэффициент принимает для цепочек с полиномиальным ростом, построенных оптимизированным способом.

Ключевые слова: вероятностные цепочки, математическое моделирование, полиномиальный рост, экстраполяция.

1. Введение

Для задач прогнозирования динамики различных экономических показателей, таких как производство/потребление участниками процесса какого-либо ресурса, или расходы на осуществление определенной деятельности, разработано достаточно много математических моделей. Наиболее часто используются методы машинного обучения [3], модели эконометрики [8].

В данной статье рассматриваются модели динамики распределения ресурсов с помощью так называемых вероятностных цепочек. Исходные данные о моделируемом процессе преобразуются к виду, описываемому в относительных долях ресурса, что приводит к использованию вероятностного распределения. Динамика изменения начального распределения описывается дискретной динамической системой, определенной на симплексе вероятностных векторов. Вероятностные цепочки были впервые рассмотрены в работах М. Сониса и Д. Хьюинса [9-12]. Наиболее часто используются модели с логистическим и линейно-логарифмическим ростом.

Опытные данные проверяются на корректность и на соответствие выбранной модели с помощью статистических критериев. В случае приемлемого соответствия строится прогноз на заданный промежуток времени. Логистические и логарифмически-линейные цепочки были успешно использованы в работах [1, 2] и [4, 5, 6] для построения прогноза для данных различной природы. Достоверность прогноза оценивается сравнением с эмпирическими данными. Важную роль играет оценка поведения динамической системы, используемой в качестве модели, в частности ее устойчивости. В работе [6] для цепочек с логарифмически-линейным ростом был проведен анализ устойчивости неподвижных точек модельной системы и рассмотрены примеры построения прогноза для разных типов устойчивости.

В данной работе мы рассматриваем способ построения прогноза на основе модели цепочек с полиномиальным ростом. Системы полиномов могут быть заданы не единственным образом, поэтому возникает вопрос о выборе наиболее подходящего представления. В работе рассмотрены два способа задания полиномов – случайным выбором и с помощью предварительного анализа взаимосвязи эмпирических данных разных участников. Сравнительный анализ результатов показывает, что второй способ дает более точный прогноз, но результаты отличаются незначительно. Показано, что цепочки с полиномиальным ростом дают лучший прогноз, чем логистические и линейно-логарифмические.

2. Основные определения

Вероятностный вектор — это совокупность величин p_i , таких, что

$$\begin{aligned} 0 \leq p_i \leq 1 \\ \sum_{i=1}^n p_i = 1, \end{aligned} \tag{1}$$

где $i = 1, \dots, n$, n — число групп или территорий.

Таким образом, совокупность эмпирических данных, представленных в виде вероятностных векторов (1) в рассматриваемые моменты времени, образует дискретную вероятностную цепочку.

Дискретная вероятностная $(1, n)$ -цепочка — это последовательность вероятностных векторов вида:

$$\begin{pmatrix} p_{1t} \\ \dots \\ p_{nt} \end{pmatrix}, t = 0, 1, \dots; 0 \leq p_{kt} \leq 1, \sum_{k=1}^n p_{kt} = 1. \tag{2}$$

Аналогично определяется вероятностная (m, n) -цепочка — это m последовательностей вероятностных векторов вида:

$$\begin{pmatrix} p_{1t}^s \\ \dots \\ p_{nt}^s \end{pmatrix}, t = 0, 1, \dots; s = 1, 2, \dots, m; \tag{3}$$

$$0 \leq p_{kt}^s \leq 1, \sum_{k=1}^n p_{kt}^s = 1.$$

Вероятностные цепочки вида (2) и (3) можно представить набором строго положительных порождающих функций — преобразований над вероятностными векторами.

Так, k -й элемент вероятностного вектора в момент времени t будет иметь вид

$$p_{k,t+1} = \frac{P_k(p_0, \dots, p_t)}{\sum_{i=1}^n P_i(p_0, \dots, p_t)}, \quad (4)$$

где $P_k(p_0, \dots, p_t)$ — строго положительная порождающая функция, и $k = 1, \dots, n, t = 0, 1, \dots$

Коэффициент корреляции, который позволяет оценить точность найденных данных, вычисляется по формуле

$$R = \frac{\sum_{t=0}^T \sum_{k=1}^n \hat{p}_{kt} \tilde{p}_{kt}}{\left(\left(\sum_{t=0}^T \sum_{k=1}^n \hat{p}_{kt}^2 \right)^{1/2} \left(\sum_{t=0}^T \sum_{k=1}^n \tilde{p}_{kt}^2 \right)^{1/2} \right)} \quad (5)$$

где значения \hat{p}_{kt} соответствуют эмпирической динамике, значения \tilde{p}_{kt} соответствуют найденным приближенным значениям. [11]

В работах [1, 2, 4, 5, 6] подробно рассмотрено построение вероятностных цепочек с логистическим ростом и с линейно-логарифмическим ростом. Здесь мы приведем основные данные, которые потребуются в работе.

Вероятностные цепочки с логистическим ростом определяются набором порождающих функций

$$P_k(p) = \gamma_k p_k, \quad (6)$$

где $\gamma_k > 0, k = 1, 2, \dots, n$ — скорость прироста или снижения доли одного вида ресурса на территории k или k -ого вида одного ресурса на одной территории.

Вероятностные цепочки с линейно-логарифмическим ростом задаются порождающими функциями Кобба-Дугласса:

$$P_k(p) = A_k p_1^{a_{k1}} p_2^{a_{k2}} \dots p_n^{a_{kn}}, \quad -\infty \leq a_{kj} \leq +\infty, A_k > 0, k = 1, \dots, n, \quad (7)$$

где A_1, \dots, A_n и элементы матрицы $[a_{ij}]_{i,j=1}^n$ параметры, определяемые в контексте задачи.

Для реализации построения цепочек с логистическим и линейно-логарифмическим ростом была написана и зарегистрирована [7] программа на C#. Сложность алгоритма построения цепочек с логистическим ростом составляет $O(n \cdot (T + T_{prog}))$, сложность алгоритма построения цепочек с линейно-логарифмическим ростом составляет $O(n^2 \cdot (T + T_{prog}))$, где T — количество лет, по которым имеются эмпирические данные, T_{prog} — количество лет, на которое строится прогноз, n — количество участников.

3. Алгоритм построения вероятностных цепочек с полиномиальным ростом

Цепочки с полиномиальным ростом являются частным случаем цепочек с рационально-полиномиальным ростом. Исходными данными являются вероятностные вектора вида $(x_{1t}, x_{2t}, \dots, x_{nt})$, вычисленные в определенный момент времени t . Функции, описывающие преобразование исходных данных, задаются полиномами X_i , так что

$$\sum_{i=1}^n X_i(x_{1t}, x_{2t}, \dots, x_{nt}) \equiv 1 \quad (8)$$

и $\sum_{i=1}^n x_{it} = 1$, для любого t .

В статье М. Сониса [11] для определения вида полиномов предлагается использовать равенство:

$$1 = (x_1 + x_2 + \dots + x_n)^m \quad (9)$$

означающее, что сумма представлена как полином некоторой степени m от суммы переменных [11]. Искомые полиномы строятся (не единственным образом) с учетом приведенного тождества (8), а именно сумма таких полиномов должна представлять собой полином некоторой степени m от суммы аргументов, т.е.

$$\begin{aligned} 1 &= (x_{1t} + x_{2t} + \dots + x_{nt})^m \\ &= X_1(x_{1t} + x_{2t} + \dots + x_{nt}) + X_2(x_{1t} + x_{2t} + \dots + x_{nt}) + \dots \\ &\quad + X_k(x_{1t} + x_{2t} + \dots + x_{nt}) \end{aligned} \quad (10)$$

Например, для $m, n = 2$ получаем

$$(x_{1t} + x_{2t})^2 = 1, \quad (11)$$

Откуда

$$X_{1t} + X_{2t} = x_{1t}^2 + x_{2t}^2 + 2x_{1t}, \quad (12)$$

и полиномы можно выбрать следующим образом

$$X_{1t}(x_{1t}, x_{2t}) = x_{1t}^2 \text{ и } X_{2t}(x_{1t}, x_{2t}) = x_{2t}^2 + 2x_{1t}x_{2t}. \quad (13)$$

Вероятностная цепочка с полиномиальным ростом имеет вид:

$$x_{kt+1} = X_k(x_{1t}, x_{2t}, \dots, x_{nt}), 1 \leq k \leq n. \quad (14)$$

По свойству построенных полиномов из тождества (8) следует, что

$$x_{1t+1} + x_{2t+1} + \dots + x_{nt+1} = 1. \quad (15)$$

Метод моделирования экономических данных на основе вероятностных цепочек с полиномиальным ростом для полиномов второй степени был применен в работе [5]. Полиномы были выбраны произвольным образом. Найденный коэффициент корреляции показал, что результат моделирования с помощью цепочек с полиномиальным ростом получился наиболее близким к эмпирическим данным по сравнению с результатами моделирования с помощью цепочек с логистическим и линейно-логарифмическим ростом.

В данной работе мы рассматриваем вариант выбора полиномов, который учитывает возможные зависимости между исходными данными и на конкретном примере сравниваем результаты прогноза. Полиномиальные цепочки строятся для $m = 2$, где m – это максимальная степень полиномов, а n обозначает количество видов единиц ресурсов в рамках одной совокупности.

Для определения коэффициентов полиномов нужно найти $n(n-1)/2$ (16)

коэффициентов перед произведениями различных переменных и n коэффициентов перед степенями переменных.

Таким образом, нужно определить $n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2}$ коэффициентов.

Тогда $n \cdot \frac{n(n+1)}{2}$ — это число переменных во всей модели.

Ищем многочлены в следующем виде:

$$X_1(x_1, x_2, \dots, x_n) = a_{111}x_1^2 + a_{122}x_2^2 + \dots + a_{1nn}x_n^2 + a_{112}x_1x_2 + \dots + a_{1(n-1)n}x_{n-1}x_n$$

⋮

$$X_i(x_1, x_2, \dots, x_n) = a_{i11}x_1^2 + a_{i22}x_2^2 + \dots + a_{inn}x_n^2 + a_{i12}x_1x_2 + \dots + a_{i(n-1)n}x_{n-1}x_n \quad (17)$$

$$X_n(x_1, x_2, \dots, x_n) = a_{n11}x_1^2 + a_{n22}x_2^2 + \dots + a_{nnn}x_n^2 + a_{n12}x_1x_2 + \dots + a_{n(n-1)n}x_{n-1}x_n,$$

где для упрощения обозначений опускаем момент времени t . [5]

Чтобы найти оптимальные значения для этих коэффициентов, мы используем эмпирические данные за некоторый период T . Для каждого периода t мы подставляем $x_k = p_{k,t}$ для всех $1 \leq k \leq n$ в правые части равенств, а в левые части подставляем значения вероятностного вектора за следующий период:

$$X_k(x_1, x_2, \dots, x_n) = p_{i,t+1} \quad (18)$$

Для нахождения коэффициентов получим систему из $T \cdot n$ уравнений вида:

$$p_{i,t+1} = a_{i11}p_{1,t}^2 + a_{i22}p_{2,t}^2 + \dots + a_{inn}p_{n,t}^2 + a_{i12}p_{1,t}p_{2,t} + \dots + a_{i(n-1)n}p_{n-1,t}p_{n,t}, \quad (19)$$

где $1 \leq i \leq n$, $1 \leq t \leq T$, (здесь $p_{i,t}$ — это уже конкретные числа, а не переменные).

Таким образом, мы получили большую неоднородную линейную систему из $n \cdot T$ уравнений с $n \cdot \frac{n(n+1)}{2}$ переменными.

$$\text{При } n \cdot T > n \cdot \frac{n(n+1)}{2} \quad (20)$$

получим переопределенную систему, у которой можно искать приближенное значение методом наименьших квадратов.

При решении системы учитываем ограничения на коэффициенты многочленов:

$$a_{ijk} \geq 0$$

$$\sum_{i=1}^n a_{ikk} = 1, \quad 1 \leq k \leq n \quad (21)$$

$$\sum_{i=1}^n a_{ikl} = 2, \quad 1 \leq j < k \leq n$$

Данные условия нужно учитывать дополнительно к основным $n \cdot T$ уравнениям.

4. Способы определения вида полиномов системы при построении цепочек с полиномиальным ростом

Нетрудно заметить, что при различных выборах полиномов мы будем получать разные модели. Рассмотрим влияние способа выбора вида полиномов системы на точность получаемого прогноза на примере модельной задачи распределения ресурсов между 2 участниками. Тогда

$$P_1 + P_2 = (x_1 + x_2)^2 = x_1^2 + x_2^2 + 2x_1x_2 \quad (22)$$

Зададим систему полиномов в общем виде с ограничениями: $a_{ijk} \geq 0$, $\sum_{i=1}^2 a_{ikk} = 1$, $\sum_{i=1}^2 a_{ikl}, k \neq l$.

$$\begin{aligned} P_1 &= a_{111}x_{11}^2 + a_{122}x_{21}^2 + a_{112}x_{11}x_{21} \\ P_2 &= a_{211}x_{11}^2 + a_{222}x_{22}^2 + a_{212}x_{12}x_{22} \end{aligned} \quad (23)$$

Таким образом, мы получаем систему из двух уравнений с 6 коэффициентами. Число коэффициентов можно уменьшить путем проведения предварительного анализа эмпирических данных. Мы рассмотрим влияние доли ресурсов на одной территории на долю ресурса на другой территории. Возможны три варианта взаимного влияния:

- 1) Коэффициент a_{kij} большой. Это свидетельствует о сильной взаимосвязи между x_k и $x_i x_j$.
- 2) Коэффициент $a_{kij} = \frac{2}{n}$, где n – количество стран. Сумма коэффициентов при попарных произведениях равна 2, $a_{1ij} + a_{2ij} + \dots + a_{nij} = 2$. Если все коэффициенты равны между собой, ($a_{kij} = \frac{2}{n}$), то слагаемое $x_i x_j$ вносит одинаковый вклад во все многочлены.
- 3) Коэффициент a_{kij} при слагаемом $x_i x_j$ в полиноме X_k близок к 0. Это показывает, что корреляция между x_k и $x_i x_j$ может быть отрицательной (увеличение $x_i x_j$ приводит к уменьшению x_k).

Таким образом, можно сделать следующие ограничения и допущения при определении вида полиномов системы:

- а) В многочлене X_k должны присутствовать все слагаемые с переменной x_k .
- б) Если в полиноме (P_2) произведение двух ($x_1 x_2$) не значимо и будет приравнено к 0, то это справедливо и для полинома (P_1).
- в) Коэффициенты при членах, в которые не входит x_i , (например $x_j x_k$) равны нулю. Это означает, что произведение $x_j x_k$, не влияет на x_i .

Таким образом, выбор вида полиномов зависит от взаимосвязи слагаемых $x_i x_j$ или x_i^2 . Часть коэффициентов системы может быть приравнена к 0, на часть коэффициентов наложено ограничение о том, что они не могут быть приравнены к 0. В результате в полиномах будут отсутствовать некоторые слагаемые и, таким образом, могут быть получены различные представления.

Нужно отметить, что независимо от числа коэффициентов, полученная система решается методом наименьших квадратов с ограничениями.

5. Пример применения цепочек с полиномиальным ростом к моделированию данных торговли продукцией

Построим полиномы вида:

$$X_i(x_1, x_2, \dots, x_n) = a_{i11}x_1^2 + a_{i22}x_2^2 + \dots + a_{inn}x_n^2 + a_{i12}x_1x_2 + \dots + a_{i(n-1)n}x_{n-1}x_n, \quad (24)$$

где зададим количеством видов единиц ресурсов равным 4 в рамках одной совокупности.

Для примера были взяты данные сбыта товаров (в % от ВВП) Германии, Франции, Италии, Испании [13]. Торговля продукцией как доля ВВП представляет собой сумму экспорта и импорта товаров, деленную на стоимость ВВП.

Таблица 1.

	Германия	Франция	Италия	Испания
1990	43.84	35.54	29.82	26.72
1991	42.42	35.37	28.26	26.59
1992	39.34	33.93	27.77	26.01
1993	34.89	33.18	29.80	27.07

1994	36.66	35.62	32.80	31.20
1995	38.18	36.88	37.44	34.39
1996	39.39	37.37	35.09	35.55
1997	43.34	40.41	36.28	36.68
1998	45.34	41.81	36.55	40.15
1999	46.38	42.96	36.44	37.78
2000	53.98	48.93	41.90	45.47
2001	54.41	47.37	41.19	43.25
2002	53.47	44.23	39.46	41.22
2003	54.33	42.97	37.92	40.28
2004	57.87	43.63	39.32	41.32
2005	61.42	44.06	40.80	41.74
2006	67.33	44.76	44.12	43.07
2007	69.45	44.80	45.77	43.65
2008	70.54	45.68	46.05	43.21
2009	60.23	38.87	37.51	35.04
2010	68.12	42.94	43.78	40.93
2011	72.88	46.01	47.21	46.20
2012	72.46	46.32	47.43	47.75
2013	70.36	44.91	46.59	48.60
2014	69.55	44.11	46.51	49.90
2015	70.83	44.17	47.27	49.71

Ищем 40 коэффициентов (10 коэффициентов в каждом из 4 уравнений системы), где 4 коэффициента – коэффициенты при переменных во второй степени и 6 коэффициентов – это коэффициенты при произведениях. Коэффициенты должны удовлетворять следующим условиям:

$$a_{ijk} \geq 0, \sum_{i=1}^4 a_{ikk} = 1, \sum_{i=1}^6 a_{ikl} = 2, k \neq l \quad (25)$$

$X_i \geq 0$ всегда остаются неотрицательными

$$X_1 + X_2 + X_3 + X_4$$

$$= x_{1t}^2 + x_{2t}^2 + x_{3t}^2 + x_{4t}^2 + 2x_{1t}x_{2t} + 2x_{1t}x_{3t} + 2x_{1t}x_{4t} + 2x_{2t}x_{3t} + 2x_{2t}x_{4t} + 2x_{3t}x_{4t} = (x_{1t} + x_{2t} + x_{3t} + x_{4t})^2 = 1 \quad (26)$$

Это преобразование переводит вероятностный вектор $x = (x_1, x_2, x_3, x_4)$ в вероятностный вектор $(X_1(x), X_2(x), X_3(x), X_4(x))$.

Нужно найти коэффициенты a_{ijk} .

Используем метод наименьших квадратов.

$$\min \sum_i^n e_i^2 = \min \sum_i^t (x_{it} - \overline{x_{it}})^2 \quad (27)$$

x_{it} - фактическое значение (из эмпирических данных)

$\overline{x_{it}} = X_i(x_{1t-1}, x_{2t-1}, \dots, x_{nt-1})$ — смоделированное значение

Эмпирические данные:

x_{11}	x_{21}	x_{31}	x_{41}
x_{12}	x_{22}	x_{32}	x_{42}
x_{13}	x_{23}	x_{33}	x_{43}
...

Функция X_i зависит от 40 коэффициентов.

$$\begin{aligned} \overline{x_{it}} &= X_i(x_{1(t-1)}, x_{2(t-1)}, x_{3(t-1)}, x_{4(t-1)}) \\ &= a_{i11}x_{1(t-1)}^2 \\ &\quad + a_{i22}x_{2(t-1)}^2 + a_{i33}x_{3(t-1)}^2 + a_{i44}x_{4(t-1)}^2 + a_{i12}x_{1(t-1)}x_{2(t-1)} \\ &\quad + a_{i13}x_{1(t-1)}x_{3(t-1)} + a_{i14}x_{1(t-1)}x_{4(t-1)} + a_{i23}x_{2(t-1)}x_{3(t-1)} \\ &\quad + a_{i24}x_{2(t-1)}x_{4(t-1)} + a_{i34}x_{3(t-1)}x_{4(t-1)} \end{aligned} \quad (28)$$

$X_i(x_{1(t-1)}, x_{2(t-1)}, x_{3(t-1)}, x_{4(t-1)}) = L_{i(t-1)}(a_{i11}, a_{i22}, a_{i33}, a_{i44}, a_{i12}, a_{i13}, a_{i14}, a_{i23}, a_{i24}, a_{i34})$ — линейная функция от весовых коэффициентов.

$(x_{it} - \overline{x_{it}})^2$ — квадрат отклонения предсказанного значения от эмпирического.

6. Реализация вычислений

Вычисления выполнены в пакете MATLAB. Была произведена оценка сложности программы снизу, то есть алгоритма построения цепочек при использовании метода наименьших квадратов без ограничений, сложность составила $O(n^9)$, где n — количество стран. В алгоритме использовался метод наименьших квадратов с линейными ограничениями, сложность которого оценить затруднительно, в связи с тем, что в программе была использована встроенная в MATLAB функция *lsqlin* [14]. Время вычисления результатов моделирования составило 1.17 секунд на компьютере с параметрами: процессор — AMD Ryzen 5 5600H 3.3 ГГц, RAM — 16 ГБ, объем SSD — 1 ТБ.

Функция находит решение, минимизируя сумму квадратов ошибок (SSE). С помощью этой функции находим коэффициенты системы. Далее находим смоделированные данные для промежутка времени, по которому известны эмпирические данные и, используя их, вычисляем значения для периода, на который строим прогноз. В рассматриваемом примере построение прогноза выполняется на 5 лет.

6.1. Моделирование данных с использованием неоптимизированной матрицы коэффициентов системы

Результат моделирования представлен в таблице 2.

Таблица. 2.

	Германия	Франция	Италия	Испания
1991	0.31195	0.25909	0.22456	0.20440
1992	0.31102	0.26376	0.21966	0.20556
1993	0.30204	0.26444	0.22354	0.20998
1994	0.27487	0.26409	0.23705	0.22400
1995	0.26914	0.25995	0.23704	0.23387
1996	0.26138	0.25069	0.24569	0.24224
1997	0.27228	0.25191	0.23337	0.24244
1998	0.27933	0.25584	0.22981	0.23502
1999	0.28428	0.25253	0.22266	0.24053
2000	0.28567	0.26004	0.22420	0.23009
2001	0.28919	0.25437	0.22129	0.23515
2002	0.29527	0.25179	0.22237	0.23057
2003	0.30274	0.24549	0.22205	0.22972
2004	0.31264	0.24214	0.21824	0.22698
2005	0.32014	0.23696	0.21797	0.22494
2006	0.32745	0.23197	0.21891	0.22167
2007	0.33656	0.22275	0.22220	0.21849
2008	0.33893	0.21841	0.22463	0.21803
2009	0.33958	0.22072	0.22476	0.21494
2010	0.34510	0.22451	0.22141	0.20898
2011	0.34401	0.21779	0.22438	0.21382
2012	0.34299	0.21506	0.22229	0.21966
2013	0.34061	0.21473	0.22113	0.22353
2014	0.33950	0.21160	0.21983	0.22907
2015	0.33904	0.20820	0.21886	0.23390
2016	0.34088	0.20675	0.22026	0.23211
2017	0.34718	0.20503	0.21841	0.22938
2018	0.35281	0.20326	0.21725	0.22668
2019	0.35771	0.20150	0.21657	0.22423
2020	0.36189	0.19975	0.21622	0.22214

На основе полученных данных построим графики данных торговли товарами (в % от ВВП) Германии, Франции, Италии, Испании за 1991-2020 гг..

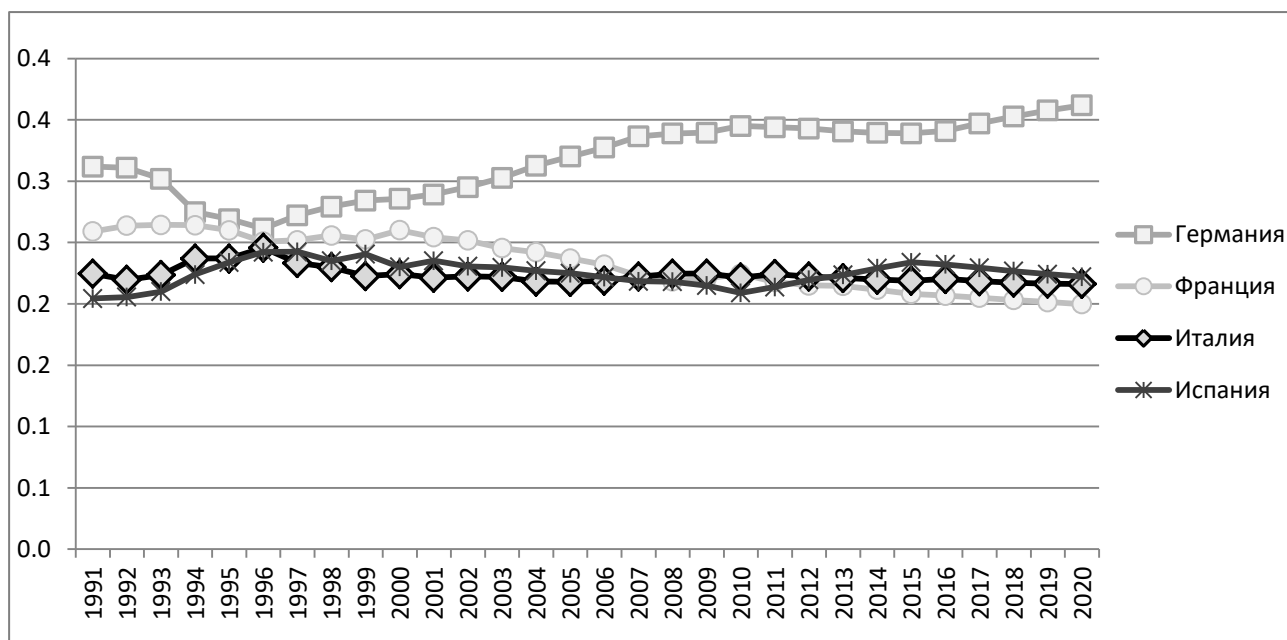


Рис. 1. Модель с полиномиальным ростом, построенная с использованием неоптимизированной матрицы

6.2. Моделирование данных с использованием оптимизированной матрицы коэффициентов системы

Теперь построим новый вид полиномов, где часть коэффициентов приравняем к 0 с учётом ограничений, описанных в разделе 4. Построим новый прогноз в MATLAB с использованием оптимизированной матрицы, содержащей коэффициенты системы. К нулю будут приравнены те коэффициенты, которые стоят при членах, в которые не входит x_i , (например x_j x_l). Результат приведен в таблице 3.

Таблица 3.

	Германия	Франция	Италия	Испания
1991	0.32048	0.25921	0.22414	0.19617
1992	0.31836	0.26482	0.21740	0.19943
1993	0.30826	0.26480	0.22217	0.20477
1994	0.27788	0.26187	0.23957	0.22068
1995	0.26852	0.25778	0.23955	0.23416
1996	0.25931	0.24676	0.25196	0.24197
1997	0.26788	0.25050	0.23494	0.24667
1998	0.27697	0.25515	0.22995	0.23793
1999	0.27841	0.25331	0.22018	0.24810
2000	0.28412	0.26047	0.22218	0.23323
2001	0.28508	0.25539	0.21842	0.24111
2002	0.29310	0.25254	0.22036	0.23401
2003	0.30073	0.24625	0.22048	0.23253

2004	0.31081	0.24358	0.21576	0.22984
2005	0.31890	0.23838	0.21592	0.22680
2006	0.32751	0.23316	0.21766	0.22167
2007	0.33816	0.22327	0.22268	0.21589
2008	0.34102	0.21848	0.22612	0.21437
2009	0.34294	0.22078	0.22618	0.21010
2010	0.35023	0.22515	0.22175	0.20287
2011	0.34759	0.21790	0.22587	0.20864
2012	0.34384	0.21549	0.22331	0.21736
2013	0.33975	0.21536	0.22178	0.22311
2014	0.33616	0.21241	0.22026	0.23116
2015	0.33356	0.20912	0.21924	0.23808
2016	0.33631	0.20741	0.22120	0.23508
2017	0.33868	0.20660	0.21939	0.23533
2018	0.34123	0.20591	0.21762	0.23524
2019	0.34391	0.20534	0.21595	0.23480
2020	0.34669	0.20488	0.21441	0.23401

На основе полученных данных построим графики данных торговли товарами (в % от ВВП) Германии, Франции, Италии, Испании за 1991-2020 гг..

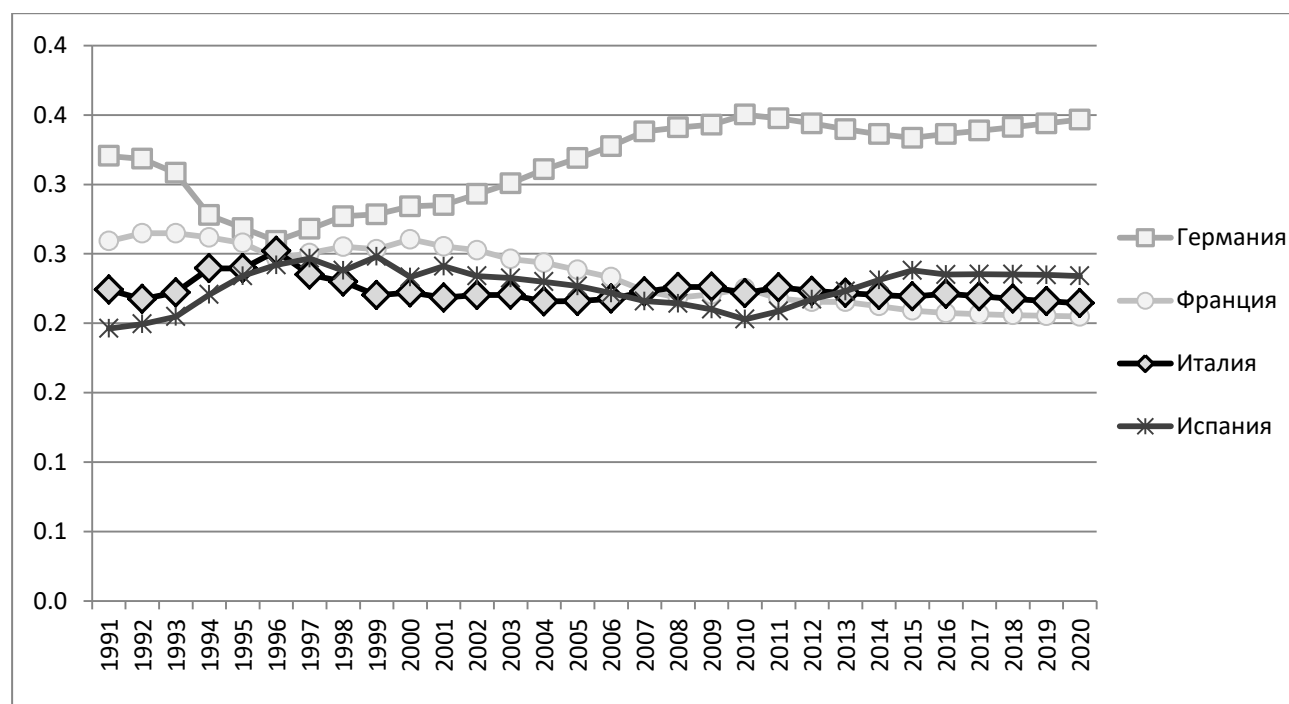


Рис. 2. Модель с полиномиальным ростом, построенная с использованием оптимизированной матрицы

Чтобы оценить полученные результаты, вычислим коэффициент корреляции между эмпирическими данными и смоделированными двумя способами данными.

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с полиномиальным ростом с использованием неоптимизированной матрицей, равен 0.999421293.

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с полиномиальным ростом с использованием оптимизированной матрицы, равен 0.999543693. Таким образом, второй способ точнее, но разница в результатах не очень значительна.

7. Сравнение результатов моделирования с помощью цепочек с полиномиальным ростом, логистическим и линейно-логарифмическим с эмпирическими данными

Чтобы оценить точность построенной модели, вычислим коэффициент корреляции между эмпирической динамикой и ее приближением с помощью формулы (5), где значениям \hat{p}_{kt} соответствуют эмпирические значения x_{it} , значениям \tilde{p}_{kt} соответствуют смоделированные значения \bar{x}_{it} .

Таблица. 4. Нормализованные эмпирические данные

	Германия	Франция	Италия	Испания
1990	0.32257	0.26145	0.21941	0.19657
1991	0.31981	0.26664	0.21305	0.20050
1992	0.30967	0.26705	0.21857	0.20471
1993	0.27926	0.26560	0.23849	0.21665
1994	0.26898	0.26139	0.24067	0.22897
1995	0.25993	0.25109	0.25485	0.23413
1996	0.26724	0.25352	0.23806	0.24119
1997	0.27655	0.25789	0.23151	0.23405
1998	0.27672	0.25514	0.22309	0.24505
1999	0.28355	0.26268	0.22278	0.23099
2000	0.28371	0.25714	0.22020	0.23895
2001	0.29218	0.25436	0.22120	0.23226
2002	0.29974	0.24796	0.22121	0.23109
2003	0.30958	0.24485	0.21605	0.22953
2004	0.31771	0.23953	0.21589	0.22687
2005	0.32667	0.23431	0.21701	0.22201
2006	0.33788	0.22460	0.22139	0.21613
2007	0.34101	0.21997	0.22470	0.21432
2008	0.34330	0.22229	0.22411	0.21030

2009	0.35087	0.22644	0.21855	0.20414
2010	0.34797	0.21935	0.22364	0.20904
2011	0.34329	0.21672	0.22238	0.21761
2012	0.33867	0.21648	0.22168	0.22317
2013	0.33430	0.21338	0.22139	0.23093
2014	0.33109	0.20996	0.22139	0.23756
2015	0.33413	0.20837	0.22300	0.23450
2016	0.33254	0.20870	0.22342	0.23534
2017	0.32850	0.20659	0.22735	0.23756
2018	0.32685	0.20608	0.22996	0.23710
2019	0.32413	0.20784	0.23324	0.23479
2020	0.32351	0.19860	0.23734	0.24056

Для того, чтобы сравнить точность смоделированных данных с точностью моделей с логистическим и линейно-логарифмическим ростом, построим цепочки с логистическим и линейно-логарифмическим ростом для того же набора эмпирических данных с интерполяцией на 1990-2015 гг. и экстраполяцией на 2015-2020 гг.

Таблица. 5. Вероятностные цепочки с логистическим ростом

	Германия	Франция	Италия	Испания
1990	0.30003	0.26424	0.22569	0.21004
1991	0.30073	0.26247	0.22567	0.21112
1992	0.30144	0.26071	0.22564	0.21221
1993	0.30214	0.25896	0.22561	0.21329
1994	0.30283	0.25721	0.22557	0.21438
1995	0.30353	0.25547	0.22553	0.21547
1996	0.30421	0.25374	0.22549	0.21656
1997	0.30490	0.25201	0.22544	0.21766
1998	0.30557	0.25029	0.22538	0.21875
1999	0.30625	0.24858	0.22533	0.21985
2000	0.30692	0.24687	0.22526	0.22094
2001	0.30759	0.24517	0.22520	0.22204
2002	0.30825	0.24348	0.22513	0.22314
2003	0.30891	0.24180	0.22505	0.22424
2004	0.30956	0.24012	0.22497	0.22534
2005	0.31021	0.23845	0.22489	0.22645

2006	0.31085	0.23679	0.22480	0.22755
2007	0.31149	0.23514	0.22471	0.22866
2008	0.31213	0.23349	0.22462	0.22976
2009	0.31276	0.23185	0.22452	0.23087
2010	0.31339	0.23022	0.22442	0.23198
2011	0.31401	0.22859	0.22431	0.23309
2012	0.31463	0.22697	0.22420	0.23420
2013	0.31524	0.22536	0.22408	0.23532
2014	0.31585	0.22376	0.22396	0.23643
2015	0.31645	0.22216	0.22384	0.23754
2016	0.31705	0.22057	0.22371	0.23866
2017	0.31765	0.21899	0.22358	0.23978
2018	0.31824	0.21742	0.22345	0.24089
2019	0.31883	0.21585	0.22331	0.24201
2020	0.31941	0.21430	0.22317	0.24313

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с логистическим ростом равен 0.998226081.

Таблица. 6. Вероятностные цепочки с линейно-логарифмическим ростом

	Германия	Франция	Италия	Испания
1991	0.30949	0.26080	0.22522	0.20449
1992	0.30299	0.25651	0.22832	0.21218
1993	0.30166	0.25023	0.22934	0.21877
1994	0.30447	0.24287	0.22889	0.22377
1995	0.31040	0.23509	0.22748	0.22703
1996	0.31829	0.22758	0.22554	0.22859
1997	0.32673	0.22112	0.22345	0.22869
1998	0.33440	0.21635	0.22155	0.22769
1999	0.34038	0.21347	0.22007	0.22608
2000	0.34442	0.21221	0.21910	0.22427
2001	0.34676	0.21203	0.21860	0.22260
2002	0.34787	0.21241	0.21846	0.22125
2003	0.34821	0.21296	0.21855	0.22027
2004	0.34813	0.21347	0.21876	0.21964

2005	0.34788	0.21383	0.21900	0.21929
2006	0.34761	0.21403	0.21923	0.21913
2007	0.34738	0.21410	0.21941	0.21911
2008	0.34723	0.21407	0.21954	0.21915
2009	0.34715	0.21400	0.21962	0.21923
2010	0.34713	0.21390	0.21967	0.21930
2011	0.34714	0.21381	0.21969	0.21937
2012	0.34717	0.21373	0.21969	0.21941
2013	0.34721	0.21367	0.21968	0.21944
2014	0.34725	0.21363	0.21967	0.21945
2015	0.34728	0.21361	0.21966	0.21946
2016	0.34730	0.21360	0.21965	0.21946
2017	0.34731	0.21359	0.21964	0.21945
2018	0.34732	0.21360	0.21964	0.21945
2019	0.34732	0.21360	0.21964	0.21944
2020	0.34732	0.21360	0.21963	0.21944

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с линейно-логарифмическим ростом равен 0.996673954.

Таким образом, в рассмотренном примере данные смоделированные с помощью цепочек с линейно-логарифмическим ростом, получились менее точными по отношению к эмпирическим данным. Данные, полученные в результате построения цепочек с полиномиальным ростом, получились ближе к эмпирическим данным.

8. Заключение

В представленной статье показан алгоритм построения вероятностных цепочек с полиномиальным ростом, предложенный М. Сонисом [11]. Приведены общие формулы для вычисления цепочек с полиномиальным ростом.

В исследовании продемонстрировано применение вероятностных цепочек с полиномиальным ростом на примере данных торговли товарами (в % от ВВП) Германии, Франции, Италии, Испании за 1991-2020 гг. Также в статье приведено описание алгоритма для построения вероятностных цепочек с помощью программы MATLAB для вышеуказанного примера. Интерполяция данных проведена на период с 1991 по 2015 год, и экстраполяция на период с 2015 по 2020 год.

Полученные результаты были оценены с помощью коэффициента корреляции. Найденные значения коэффициента корреляции для цепочек с полиномиальным, логистическим и линейно-логарифмическим ростом показывают, что модель с полиномиальным ростом ближе к эмпирическим данным и позволяет вычислить более точный прогноз по сравнению с цепочками с логистическим и линейно-логарифмическим ростом. При этом изменение метода выбора полиномов не оказало существенного влияния на коэффициент корреляции. Оптимизация матрицы коэффициентов позволяет ускорить вычисления.

Литература

- [1] Афанасьева Е.В. Моделирование процессов потребления экономических ресурсов с помощью вероятностных цепочек (на примере стран Западной Европы)//Научно - технические ведомости СПбГПУ: Информатика. Телекоммуникации. Управление. — СПб.: Политехн. ун-та, 2011. — № 3. — С. 93-97.
- [2] Афанасьева Е.В. Моделирование процессов распределения ресурсов с помощью вероятностных цепочек // Дифференциальные уравнения и процессы управления. — 2011. — № 3.
- [3] Бринк Х., Ричардс Д. Машинное обучение./ Бринк Х., Ричардс Д — СПб.: Питер, 2017. — 336 с.
- [4] Логинова Н.В. Об одном методе моделирования динамики социально-экономических процессов. Компьютерные инструменты в образовании, [S.l.], п. 2, р. 14-24, апр. 2018.
- [5] Логинова Н. В. (2020). Вероятностные цепочки с полиномиальным ростом как модель распределения ресурсов. Компьютерные инструменты в образовании, (3), 56-69.
- [6] Логинова Н. В., Ампилова Н.Б. Применение линейного бифуркационного анализа для оценки достоверности вероятностных цепочек, "Некоторые актуальные проблемы современной математики и математического образования. Герценовские чтения — 2019" Материалы научной конференции, 8-12 апреля 2019 г. — СПб.: Изд. РГПУ им. А.И. Герцена, 2019, с. 209-218.
- [7] Свидетельство о государственной регистрации программы для ЭВМ №2018666417 «Программа для построения прогноза изменения социальных и экономических данных на основе дискретных вероятностных цепочек» [Online]. Available: https://www1.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2018666417&TypeFile=html (дата обращения: 25.05.2022).
- [8] Эконометрика / под ред. И. И. Елисейевой. — М.: Финансы и статистика, 2005. — 576 с.
- [9] Hewings G. J. D. Regional industrial analysis and development / Geoffrey J. D. Hewings. — London: Methuen & Co, 1977. 180 с.
- [10] Social and demographic accounting / Ed. by Geoffrey J.D. Hewings a. Moss Madden. — Cambridge etc.: Cambridge univ. press, 1995. — IX, 242 с.
- [11] Sonis M. Discrete Non-Linear Probabilistic Chains (M. Drachlin and E. Litsyn eds) // Functional-Differential Equations, Ariel, Israel, 2003, 10:445-487.
- [12] Sonis M., Hewings G. Regional Competition and Complementarity: Comparative Advantages/Disadvantages and Increasing/Diminishing Returns in Discrete Relative Spatial Dynamics // Regional Competition Advances in Spatial Science / P. Batey, P. Friedrich. — Berlin: SpringerVerlag, 2001. — P. 139-157.
- [13] Группа Всемирного банка [Online]. Available: <https://data.worldbank.org> (дата обращения: 8.02.2022).
- [14] Archived documentation for previous versions of Matlab [Online]. Available: <https://www.mathworks.com/help/documentation-center> (дата обращения: 23.01.2022).

On probability chains with polynomial growth

Loginova N.V.

Ерам systems Montenegro DOO Podgorica

natalia.loginowa@gmail.com

Abstract. The article is devoted to method of mathematical modeling of the resource allocation process — the method of probabilistic chains. This model is a discrete dynamical system defined on a simplex of probability vectors. Probabilistic chains with logistic and linear-logarithmic growth have been studied in the most detail. In this paper, we consider chains with polynomial growth. The polynomials of the system can be specified by different ways. We compare arbitrary and optimized methods for choice of polynomials. The optimization is achieved by taking into account the relationship between the given resources. As an example, the data on trade of Germany, France, Italy, Spain are considered. The chains with logistic and linear-logarithmic growth are constructed as well. The obtained results of modeling were analyzed by using the correlation coefficient, which takes the greatest value for chains with polynomial growth constructed by the optimized way.

Keywords: probability chains, mathematical modeling, polynomial growth, extrapolation