

ВЕРОЯТНОСТНЫЕ ЦЕПОЧКИ С ПОЛИНОМИАЛЬНЫМ РОСТОМ КАК МОДЕЛЬ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ

Логинова Н. В.¹, инженер-программист, natalia.loginowa@gmail.com

¹ООО «ЭПАМ Систэмз», наб. Черной Речки, д. 41, 197342, Санкт-Петербург, Россия

Аннотация

В статье рассмотрен метод математического моделирования для построения прогноза изменения социально-экономических данных, основанный на использовании дискретных вероятностных цепочек. Исходные данные о распределении некоторого ресурса между несколькими участниками представлены в виде вероятностного вектора, а изменение этого распределения с течением времени описывается с помощью дискретной динамической системы, задаваемой определенной функцией. Достаточно хорошо изучены цепочки с линейным и логарифмически-линейным ростом. В данной работе мы рассматриваем вероятностные цепочки, в которых правая часть задана полиномами определенного вида. Алгоритм построения цепочек применен для исследования динамики распределения национального дохода Канады, Великобритании, США. Проведена оценка точности полученных результатов с помощью коэффициента корреляции, а также проведена оценка построенных дискретных динамических систем с помощью энтропии Шеннона.

Ключевые слова: динамические системы, вероятностные цепочки, экстраполяция, полиномиальный рост, энтропия Шеннона.

Цитирование: Логинова Н. В. Вероятностные цепочки с полиномиальным ростом как модель распределения ресурсов // Компьютерные инструменты в образовании. 2020. № 3. С. 56–69. doi:10.32603/2071-2340-2020-3-56-69

1. ВВЕДЕНИЕ

Моделирование процесса изменения социально-экономических данных и построение достоверного прогноза является актуальной и важной задачей. В данной работе мы рассматриваем модели, описываемые дискретными динамическими системами определенного вида, а именно так называемыми дискретными вероятностными цепочками. В таких моделях, предложенных и разработанных М. Сонисом и Д. Хьюинсом [8–10], распределение ресурса между участниками процесса выражается в относительных долях, и определение закона развития динамики такого процесса означает преобразование некоторого вероятностного вектора. Как дискретные динамические системы эти модели могут демонстрировать достаточно сложное поведение — от устойчивых режимов до хаотических, поэтому их применение позволяет исследовать различные классы процессов распределения ресурсов. Наиболее часто используются цепочки с логистическим и линейно-логарифмическим ростом [6, 7]. Применение вышеуказанных методов для

исследования таких социально-экономических показателей, как потребление различных видов ресурсов странами западной Европы, распределение исследователей по областям науки в РФ, распределение доходов между слоями населения РФ, было представлено в работах Афанасьевой Е. В. [1, 2]. Результаты моделирования с помощью вероятностных цепочек были использованы в [3] при анализе так называемого глобального перманентного спада в экономике развитых стран. Исследования вероятностных цепочек с логистическим и с линейно-логарифмическим ростом в зависимости от выбора статистических критериев были проведены в [4]. В работе [5] проведено моделирование экономических процессов и построение прогноза по эмпирическим данным на основе цепочек с линейно-логарифмическим ростом с предварительным анализом полученных моделей на устойчивость с помощью процедуры линейного бифуркационного анализа. Нужно отметить, что для некоторых видов данных полученная модель не имела устойчивых неподвижных точек и не наблюдалось тенденции к стабилизации процесса.

В настоящей работе рассматривается алгоритм построения цепочек с полиномиальным ростом и применение данного алгоритма к эмпирическим данным. В качестве инструмента для выполнения вычислений применяется пакет программ MATLAB [12]. Кроме того, в данном исследовании производится оценка полученных результатов, что позволяет сделать вывод о возможности применения данного метода для построения надежного прогноза изменения данных.

2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Вероятностный вектор — это совокупность величин p_i , таких, что: $0 \leq p_i \leq 1$, где $i = 1, \dots, n$, n — число групп или территорий,

$$\sum_{i=1}^n p_i = 1. \quad (1)$$

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 векторов вида

$$\begin{pmatrix} p_{1t} \\ p_{2t} \\ \vdots \\ p_{nt} \end{pmatrix}, \quad t = 0, 1, 2, \dots, \quad 0 \leq p_{kt} \leq 1, \quad \sum_{k=1}^n p_{kt} = 1. \quad (2)$$

Таким же способом, как и цепочка вида (2), определяется вероятностная (m, n) -цепочка — это m последовательностей вероятностных векторов вида:

$$\begin{pmatrix} p_{1t}^s \\ p_{2t}^s \\ \vdots \\ p_{nt}^s \end{pmatrix}, \quad t = 0, 1, 2, \dots, \quad s = 1, 2, \dots, m; \quad 0 \leq p_{kt}^s \leq 1, \quad \sum_{k=1}^n p_{kt}^s = 1. \quad (3)$$

Вероятностную цепочку 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

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

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

2.1. Коэффициент корреляции

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

$$R = \frac{\sum_{t=0}^T \sum_{k=1}^n \hat{p}_{kt} \tilde{p}_{kt}}{\left(\sum_{t=0}^T \sum_{k=1}^n \hat{p}_{kt}^2 \right)^{(1/2)} \cdot \left(\sum_{t=0}^T \sum_{k=1}^n \tilde{p}_{kt}^2 \right)^{(1/2)}}, \quad (4)$$

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

2.2. Энтропия Шеннона

Для каждого вероятностного вектора $p = (p_1, p_2, \dots, p_n)$ из симплекса S всех вероятностных векторов можно определить функцию энтропии Шеннона:

$$H(p) = - \sum_{k=1}^n p_k \log p_k. \quad (5)$$

Важное свойство этой функции — это неравенство Кульбака-Лейблера: для каждой пары вероятностных векторов $p = (p_1, p_2, \dots, p_n)$, $q = (q_1, q_2, \dots, q_n) \in S$ выполняется неравенство

$$- \sum_{k=1}^n p_k \log p_k \leq - \sum_{k=1}^n p_k \log q_k,$$

причем равенство выполняется тогда и только тогда, когда $p_k = q_k$ при всех $k = 1, 2, \dots, n$. Используем свойства логарифмов и получим:

$$\prod_{k=1}^n \left(\frac{q_k}{p_k} \right)^{p_k} \leq 1$$

Энтропия Шеннона центральной точки $\begin{pmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{pmatrix}$ симплекса S имеет максимальное зна-

чение $H_{max} = \log n$. На вершинах симплекса S (векторах вида $\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$) энтропия получает

минимальное значение $H_{min} = 0$. Поверхности в симплексе S с энтропией $H(p) = Const$ —

это симметричные концентрические овалы с центром в точке $\begin{pmatrix} 1/n \\ 1/n \\ \vdots \\ 1/n \end{pmatrix}$.

Энтропия $H(p)$ каждой точки $p = (p_1, p_2, \dots, p_n)$ из симплекса S увеличивается при приближении p к центральной точке [5].

3. ПОСТРОЕНИЕ ЦЕПОЧЕК С ПОЛИНОМИАЛЬНЫМ РОСТОМ

Цепочки с полиномиальным ростом являются частным случаем цепочек с рационально-полиномиальным ростом [9]. Исходными данными являются вероятностные вектора $(x_{1t}, x_{2t}, \dots, x_{nt})$, вычисленные в определенный момент времени t . Функции, описывающие преобразование исходных данных, задаются полиномами X_i , так что

$$\sum_{i=1}^n X_i(x_{1t}, x_{2t}, \dots, x_{nt}) \equiv 1 \text{ и } \sum_{i=1}^n x_{it} = 1, \text{ для любого } t. \quad (6)$$

Для определения вида полиномов мы можем использовать равенство:

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

означающее, что сумма представлена как полином некоторой степени m от суммы переменных. Искомые полиномы строятся с учетом приведенного тождества (6), а именно сумма таких полиномов должна представлять собой полином некоторой степени m от суммы аргументов, то есть

$$1 = (x_{1t} + x_{2t} + \dots + x_{nt})^m = X_1(x_{1t} + x_{2t} + \dots + x_{nt}) + X_2(x_{1t} + x_{2t} + \dots + x_{nt}) + \dots + X_k(x_{1t} + x_{2t} + \dots + x_{nt}).$$

Вероятностная цепочка с полиномиальным ростом имеет вид:

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

Заметим, что выбор полиномов в правой части может быть выполнен не единственным образом, поэтому в зависимости от выбора мы получаем различные системы.

Например, для $m, n = 2$ получаем равенство $(x_{1t} + x_{2t})^2 = 1$, которое может быть представлено в виде суммы двух полиномов

$$X_{1t} + X_{2t} = (x_{1t}^2 + x_{2t}^2 + 2x_{1t}x_{2t}) = (x_{1t} + x_{2t})^2 = 1.$$

Построим полиномы следующим образом

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

где для каждого момента времени t будет выполняться условие

$$x_{1t+1} = X_1(x_{1t}, x_{2t}).$$

По свойству построенных полиномов из тождества (6) $x_{1t} + x_{2t} + \dots + x_{nt} = 1$, следовательно и $x_{1t+1} + x_{2t+1} + \dots + x_{nt+1} = 1$.

В данной работе мы будем строить полиномиальные цепочки для $m = 2$, где m — это максимальная степень полиномов, а n обозначает количество видов единиц ресурсов

в рамках одной совокупности. Выбрать упорядоченную пару из n различных элементов можно $n \cdot (n - 1)$ способами. Число неупорядоченных пар равно $n \cdot (n - 1)/2$, то есть C_n^2 — биномиальный коэффициент.

Записываем задачу в виде линейной системы

$$\begin{aligned} X_1(x_1, x_2, \dots, x_n) &= a_{111}x_1^2 + a_{122}x_2^2 + \dots + a_{1nn}x_n^2 + a_{112}x_1x_2 + \dots + a_{1(n-1)n}x_{n-1}x_n \\ &\vdots \\ X_n(x_1, x_2, \dots, x_n) &= a_{n11}x_1^2 + a_{n22}x_2^2 + \dots + a_{nnn}x_n^2 + a_{n12}x_1x_2 + \dots + a_{n(n-1)n}x_{n-1}x_n, \end{aligned} \quad (7)$$

где каждый многочлен системы (7) имеет вид:

$$X_i(x_1, x_2, \dots, x_n) = a_{ni1}x_1^2 + a_{ni2}x_2^2 + \dots + a_{ninn}x_n^2 + a_{ni12}x_1x_2 + \dots + a_{ni(n-1)n}x_{n-1}x_n \quad (8)$$

Таким образом, получим

$$n + \frac{n(n-1)}{2} = \frac{2n + n^2 - n}{2} = \frac{n^2 + n}{2} = \frac{n(n+1)}{2}$$

неопределенных коэффициентов в каждом из n многочленов, где n — число коэффициентов перед переменными, возведёнными в степень, а $n \cdot (n - 1)/2$ — число коэффициентов перед произведениями различных переменных.

Всего в задаче будет $n \cdot n \cdot (n + 1)/2$ неизвестных весовых коэффициентов в многочленах вида (8), то есть $n \cdot n \cdot (n + 1)/2$ — это число переменных во всей модели.

4. ПРИМЕР

Построим полиномы вида (8) степени 2 с количеством видов единиц ресурсов, равным 3, в рамках одной совокупности. Тогда система (7) примет вид:

$$\begin{aligned} X_1(x_1, x_2, x_3) &= a_{111}x_1^2 + a_{122}x_2^2 + a_{133}x_3^2 + a_{112}x_1x_2 + a_{113}x_1x_3 + a_{123}x_2x_3, \\ X_2(x_1, x_2, x_3) &= a_{211}x_1^2 + a_{222}x_2^2 + a_{233}x_3^2 + a_{212}x_1x_2 + a_{213}x_1x_3 + a_{223}x_2x_3, \\ X_3(x_1, x_2, x_3) &= a_{311}x_1^2 + a_{322}x_2^2 + a_{333}x_3^2 + a_{312}x_1x_2 + a_{313}x_1x_3 + a_{323}x_2x_3. \end{aligned} \quad (9)$$

Количество коэффициентов в каждом многочлене равно $n + \frac{n(n-1)}{2}$, где n — количество коэффициентов при квадратах (их по 3 в каждом многочлене), $\frac{n(n-1)}{2}$ — количество коэффициентов при попарных произведениях. В итоге каждый многочлен содержит 6 коэффициентов, вся система содержит 18 коэффициентов, удовлетворяющих условиям

$$a_{ijk} \geq 0, \quad \sum_{i=1}^3 a_{ikk} = 1, \quad \sum_{i=1}^3 a_{ikl} = 2, \quad k \neq l, \quad x_i \geq 0.$$

Преобразование

$$X_1 + X_2 + X_3 = x_{1t}^2 + x_{2t}^2 + x_{3t}^2 + 2x_{1t}x_{2t} + 2x_{1t}x_{3t} + 2x_{2t}x_{3t} = (x_{1t} + x_{2t} + x_{3t})^2 = 1$$

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

Рассмотрим данные ВНД Канады, США, Великобритании [11] (табл. 1а). Преобразуем в нормированный вид x_{it} , где $\sum_{i=1}^n x_{it} = 1$ (табл. 1б).

Таблица 1. ВНД Канады, США, Великобритании¹

(а) Исходные данные				(б) Нормированные данные			
Год	Канада	США	Велико-британия	Год	Канада	США	Велико-британия
1985	13830	17510	9000	1985	0,34284	0,43406	0,22310
1986	14450	19160	10070	1986	0,33082	0,43864	0,23054
1987	16200	21460	12480	1987	0,32310	0,42800	0,24890
1988	18480	23580	15890	1988	0,31890	0,40690	0,27420
1989	19710	23860	17090	1989	0,32493	0,39334	0,28173
1990	20480	24150	18250	1990	0,32570	0,38406	0,29024
1991	20800	24370	19050	1991	0,32389	0,37948	0,29664
1992	21220	25780	21110	1992	0,31155	0,37851	0,30994
1993	21080	26480	20990	1993	0,30751	0,38629	0,30620
1994	20740	27750	21520	1994	0,29624	0,39637	0,30738
1995	20740	29150	20990	1995	0,29261	0,41126	0,29613
1996	20700	30380	22580	1996	0,28102	0,41244	0,30654
1997	21200	31390	24760	1997	0,27408	0,40582	0,32010
1998	20760	32150	26300	1998	0,26209	0,40588	0,33203
1999	21300	33780	27710	1999	0,25728	0,40802	0,33470
2000	22610	36070	28880	2000	0,25822	0,41195	0,32983
2001	22980	36840	28620	2001	0,25984	0,41655	0,32361
2002	23490	37470	28900	2002	0,26141	0,41698	0,32161
2003	25360	39950	32000	2003	0,26061	0,41054	0,32885
2004	29370	43680	37710	2004	0,26517	0,39437	0,34047
2005	34120	46340	42770	2005	0,27688	0,37604	0,34707
2006	37780	48080	45020	2006	0,28866	0,36736	0,34398
2007	41420	48640	47790	2007	0,30047	0,35285	0,34668
2008	44810	49330	48420	2008	0,31432	0,34603	0,33965
2009	43110	48050	43940	2009	0,31910	0,35566	0,32524
2010	44370	48950	41390	2010	0,32937	0,36337	0,30725
2011	47060	50460	40620	2011	0,34067	0,36528	0,29405
2012	50900	52540	41210	2012	0,35188	0,36322	0,28489
2013	52620	53660	42350	2013	0,35403	0,36103	0,28494
2014	51750	55380	43760	2014	0,34297	0,36702	0,29001
2015	47250	56070	43700	2015	0,32138	0,38138	0,29724

Интерполяция выполняется для периода 1985–2000 гг., экстраполяция на 2001–2015 гг. Используем метод наименьших квадратов.

$$\min \sum_i^n e_i^2 = \min \sum_i^t (x_{it} - \bar{x}_{it})^2,$$

где x_{it} — фактическое значение (из эмпирических данных), $\bar{x}_{it} = X_i(x_{1t-1}, x_{2t-1}, \dots, x_{nt-1})$ — смоделированное значение.

¹Биллионы долларов США.

Эмпирические данные:

x_{11}	x_{21}	x_{31}
x_{12}	x_{22}	x_{32}
x_{13}	x_{23}	x_{33}
\dots	\dots	\dots

Чтобы предсказать значения в следующий момент времени, мы подставляем значения из предыдущего момента времени:

$$\bar{x}_{it} = X_i(x_{11}, x_{21}, x_{31}) = a_{111}x_{11}^2 + a_{122}x_{21}^2 + a_{133}x_{31}^2 + a_{112}x_{11}^2x_{21} + a_{123}x_{21}x_{31},$$

$a_{111}x_{11}^2 + a_{122}x_{21}^2 + a_{133}x_{31}^2 + a_{112}x_{11}^2x_{21} + a_{123}x_{21}x_{31} = L_{11}(a_{111}, a_{122}, a_{122}, a_{133}, a_{112}, a_{113}, a_{123})$ — линейная функция от весовых коэффициентов. Квадрат отклонения предсказанного значения от эмпирического $(x_{12} - \bar{x}_{12})^2$ позволяет оценить точность прогноза.

5. ВЫЧИСЛЕНИЕ СМОДЕЛИРОВАННЫХ ЗНАЧЕНИЙ ДЛЯ ДАННЫХ ВНД КАНАДЫ, ВЕЛИКОБРИТАНИИ, США В ПРОГРАММЕ MATLAB

Задаём $N = 3$ — число стран и $M = 16$ — число лет. Находим оптимальную матрицу коэффициентов W на $M + 1$ годах.

1. Записываем в матрицу $PData$ все значения за период NY . Период состоит из количества лет M , для которого производится интерполяция, и прогнозируемого периода, для которого производится интерполяция. В рассматриваемом примере прогноз строится на 15 лет, а $M = 16$, то есть $NY = 31$ году по N странам. $PData$ — матрица размерности 31×3 .

2. Разбиваем матрицу $PData$ на две подматрицы PX (с первого года до M) и PY (со второго года до $M + 1$) размерностями 16×3 .

3. Записываем матрицу PXX , которая будет содержать элементы вида p_i^2, p_i, p_j и состоять из M строк, первые N столбцов матрицы соответствуют p_i^2 , затем $N \cdot (N - 1) / 2$ столбцов попарных произведений $p_i p_j$. Также создаем матрицу $Titles$ для хранения индексов переменных. Строка названий столбцов матрицы содержит индексы переменных.

4. Записываем задачу в виде линейной системы. Задаём переменные: $Kw = N \cdot N \cdot (N + 1) / 2$ — количество коэффициентов, $Pw = N \cdot (N + 1) / 2$ — количество коэффициентов в одном многочлене.

Изначально есть задача в виде $PXX \cdot Wmat = PY$. Мы записываем эту задачу в виде системы линейных уравнений для вектора из Kw коэффициентов в виде $C \cdot w = d$. Матрица C — это матрица PXX , записанная блоками на диагонали N раз. PXX имеет размеры $M \times N \cdot (N + 1) / 2$. Матрица C имеет размеры $N \cdot M$ (строк) $\times N \cdot N \cdot (N + 1) / 2$ (столбцов). Каждый блок относится к своей стране. Вектор d — это матрица PY , столбцы которой записаны в один длинный столбец. Матрица PY имеет размеры $M \times N$. Вектор d имеет размеры $M \times N \times 1$

5. С помощью встроенных функций MATLAB вычисляем матрицы A и b . $A = -eye(Kw)$ и $b = zeros(Kw, 1)$. Задано ограничение типа неравенства $A * w < b$. Здесь A — это матрица $-E = (-1) \cdot E$, а вектор b — это нулевой вектор. $-E \cdot w < 0 \Leftrightarrow E \cdot w > 0$. Задано условие неотрицательности всех коэффициентов во всех многочленах.

6. С помощью встроенных функций MATLAB вычисляем матрицы Aeq и beq . $Aeq = repmat(eye(Pw), 1, N)$, $beq = [ones(N, 1); 2 * ones(Pw - N, 1)]$. Ограничение типа равенства $Aeq * w = beq$.

7. Таким образом, все искомые матрицы для решения системы с ограничениями с помощью встроенной функции `lsqlin` найдены. $w = lsqlin(C, d, A, b, Aeq, beq)$ — решаем систему с ограничениями. Функция находит решение системы $C * w = d$ при ограничениях $A * w < b$, $Aeq * w = beq$, минимизируя сумму квадратов ошибок (SSE).

8. Полученный вектор w мы записываем по столбцам в матрицу $Wmat$. В каждом столбце i записаны коэффициенты многочлена для страны i . Вычисляем с помощью найденной матрицы PY_pred . $PY_pred = PXX \cdot Wmat$. В отличие от PY , в матрицу PY_pred записываются значения, полученные с помощью найденных многочленов.

9. $PY_prog(1:M, :) = PY_pred.PY_prog$ — это матрица прогнозных значений. Первые M строк берем из тренировочной задачи PY_pred , остальные $NY - M - 1$ лет рассчитываем значения с помощью дополнительно написанной функции `pred_next` по предыдущему году, подставляя в уже посчитанные многочлены. Функция `pred_next` находит значения N многочленов с известными коэффициентами по вероятностному вектору PX . $f = PXX * Wmat$ — вектор из квадратов и попарных произведений умноженный на матрицу коэффициентов, дает N новых прогнозируемых значений для следующего года [10].

Таким образом, получаем смоделированные значения в представленной ниже таблице 2a. Полученные значения можно сравнить с эмпирическими данными, представленными в таблице 2b.

Таблица 2. ВНД Канады, США, Великобритании¹

(a) Смоделированные данные

Год	Канада	США	Великобритания
1986	0,33548	0,42298	0,24154
1987	0,32389	0,42717	0,24894
1988	0,31667	0,41870	0,26463
1989	0,31294	0,40257	0,28449
1990	0,31889	0,39185	0,28926
1991	0,31973	0,38505	0,29521
1992	0,31804	0,38207	0,29989
1993	0,30621	0,38327	0,31052
1994	0,30226	0,38918	0,30855
1995	0,29137	0,39778	0,31086
1996	0,28773	0,40871	0,30357
1997	0,27665	0,41104	0,31231
1998	0,27009	0,40754	0,32237
1999	0,25864	0,40955	0,33181
2000	0,25404	0,41176	0,33421
2001	0,25490	0,41414	0,33095
2002	0,25172	0,41613	0,33215
2003	0,24867	0,41795	0,33338
2004	0,24575	0,41964	0,33461
2005	0,24296	0,42122	0,33582
2006	0,24029	0,42271	0,33700
2007	0,23774	0,42412	0,33814
2008	0,23530	0,42546	0,33924
2009	0,23297	0,42673	0,34030
2010	0,23074	0,42794	0,34132
2011	0,22862	0,42910	0,34229
2012	0,22659	0,43020	0,34321
2013	0,22465	0,43125	0,34410
2014	0,22281	0,43225	0,34495
2015	0,22105	0,43320	0,34575

(b) Эмпирические данные

Год	Канада	США	Великобритания
1986	0,33082	0,43864	0,23054
1987	0,32310	0,42800	0,24890
1988	0,31890	0,40690	0,27420
1989	0,32493	0,39334	0,28173
1990	0,32570	0,38406	0,29024
1991	0,32389	0,37948	0,29664
1992	0,31155	0,37851	0,30994
1993	0,30751	0,38629	0,30620
1994	0,29624	0,39637	0,30738
1995	0,29261	0,41126	0,29613
1996	0,28102	0,41244	0,30654
1997	0,27408	0,40582	0,32010
1998	0,26209	0,40588	0,33203
1999	0,25728	0,40802	0,33470
2000	0,25822	0,41195	0,32983
2001	0,25984	0,41655	0,32361
2002	0,26141	0,41698	0,32161
2003	0,26061	0,41054	0,32885
2004	0,26517	0,39437	0,34047
2005	0,27688	0,37604	0,34707
2006	0,28866	0,36736	0,34398
2007	0,30047	0,35285	0,34668
2008	0,31432	0,34603	0,33965
2009	0,31910	0,35566	0,32524
2010	0,32937	0,36337	0,30725
2011	0,34067	0,36528	0,29405
2012	0,35188	0,36322	0,28489
2013	0,35403	0,36103	0,28494
2014	0,34297	0,36702	0,29001
2015	0,32138	0,38138	0,29724

На основе полученных данных построим графики изменения ВНД Канады, США и Великобритании за 1986–2015 гг. (рис. 1).



Рис. 1. Модель с полиномиальным ростом

Далее построим график на основе эмпирических данных (рис. 2).

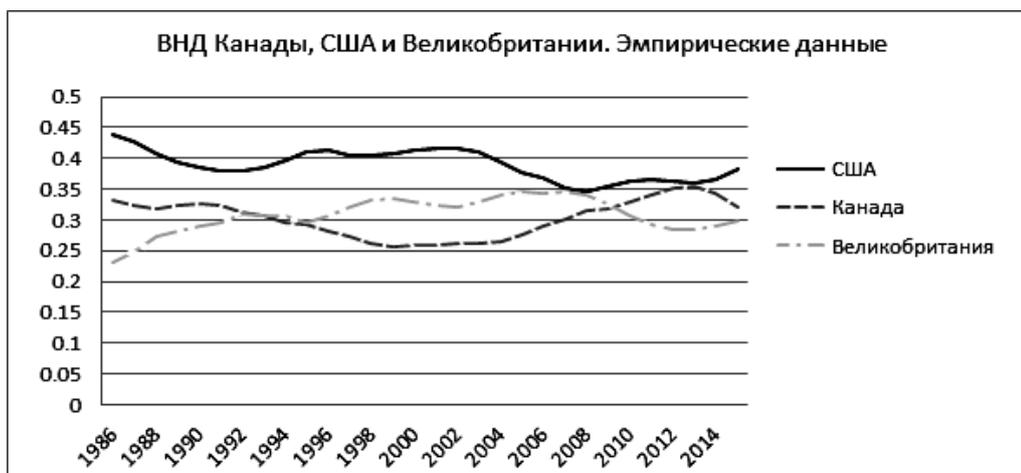


Рис. 2. Эмпирические данные

6. ОЦЕНКА ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Чтобы оценить точность построенной модели, вычислим коэффициент корреляции между эмпирической динамикой и ее приближением с помощью формулы (4), где значениям \hat{p}_{kt} соответствуют эмпирические значения x_{it} , значениям \tilde{p}_{kt} соответствуют смоделированные значения \bar{x}_{it} .

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с полиномиальным ростом равен 0,991974351.

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 1985–2000 гг. и экстраполяцией на 2001–2015 гг.

Таблица 3. ВНД Канады, Великобритании, США¹

(а) Вероятностные цепочки с логистическим ростом				(б) Вероятностные цепочки с линейно-логарифмическим ростом			
Год	Канада	США	Великобритания	Год	Канада	США	Великобритания
1986	0,33704	0,40833	0,25462	1986	0,33473	0,42910	0,23617
1987	0,33142	0,40824	0,26034	1987	0,32802	0,42062	0,25135
1988	0,32582	0,40806	0,26612	1988	0,32224	0,41084	0,26692
1989	0,32024	0,40779	0,27198	1989	0,31692	0,40207	0,28100
1990	0,31468	0,40742	0,27789	1990	0,31179	0,39584	0,29237
1991	0,30916	0,40697	0,28388	1991	0,30672	0,39241	0,30087
1992	0,30366	0,40642	0,28993	1992	0,30168	0,39123	0,30709
1993	0,29819	0,40578	0,29603	1993	0,29663	0,39157	0,31180
1994	0,29275	0,40505	0,30220	1994	0,29157	0,39287	0,31556
1995	0,28734	0,40423	0,30843	1995	0,28647	0,39482	0,31871
1996	0,28197	0,40331	0,31471	1996	0,28132	0,39725	0,32143
1997	0,27664	0,40231	0,32105	1997	0,27609	0,40010	0,32381
1998	0,27135	0,40122	0,32744	1998	0,27076	0,40337	0,32587
1999	0,26609	0,40003	0,33387	1999	0,26531	0,40708	0,32762
2000	0,26088	0,39876	0,34036	2000	0,25969	0,41129	0,32902
2001	0,25571	0,39740	0,34689	2001	0,25388	0,41606	0,33005
2002	0,25058	0,39595	0,35347	2002	0,4785	0,42151	0,33064
2003	0,24550	0,39442	0,36008	2003	0,24153	0,42777	0,33070
2004	0,24046	0,39280	0,36673	2004	0,23489	0,43502	0,33010
2005	0,23548	0,39110	0,37342	2005	0,22783	0,44351	0,32866
2006	0,23054	0,38931	0,38015	2006	0,22027	0,45361	0,32611
2007	0,22566	0,38744	0,38690	2007	0,21208	0,46586	0,32206
2008	0,22083	0,38549	0,39368	2008	0,20304	0,48108	0,31588
2009	0,21605	0,38346	0,40049	2009	0,19285	0,50061	0,30653
2010	0,21132	0,38135	0,40733	2010	0,18097	0,52680	0,29224
2011	0,20665	0,37917	0,41418	2011	0,16634	0,56400	0,26966
2012	0,20204	0,37691	0,42105	2012	0,14671	0,62104	0,23225
2013	0,19748	0,37458	0,42793	2013	0,11652	0,71585	0,16762
2014	0,19299	0,37218	0,43483	2014	0,06454	0,86828	0,06718
2015	0,18855	0,36971	0,44174	2015	0,00744	0,99033	0,00223

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с логистическим ростом равен 0,697891802.

Коэффициент корреляции между эмпирическими данными и данными, полученными в результате построения цепочек с линейно-логарифмическим ростом, равен 0,936117972.

Таким образом, в рассмотренном примере данные, смоделированные с помощью цепочек с полиномиальным ростом, получились ближе к эмпирическим данным. Данные, полученные в результате построения цепочек с логистическим ростом, получились менее точными по отношению к эмпирическим данным. Построим график, демонстрирующий изменения значений энтропии Шеннона в каждый момент времени за 1986–2015 года. Вычислим энтропию Шеннона по формуле (5) для цепочек с полиномиальным ростом, где значениям p_k соответствуют смоделированные значения x_i (рис. 3).

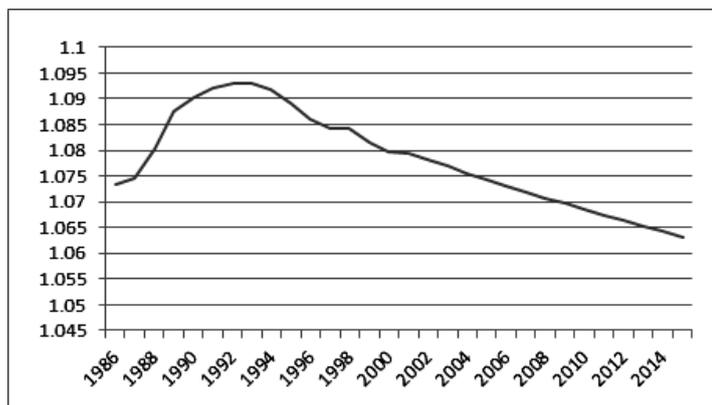


Рис. 3. Энтропия Шеннона для цепочек с полиномиальным ростом

График изменения значения вычисленной энтропии Шеннона показывает существенный рост в начале рассматриваемого периода, затем плавное снижение, небольшое увеличение в 1998 году, затем снова плавное снижение. Вычислим энтропию Шеннона по формуле (5) для цепочек с логистическим ростом (рис. 4).

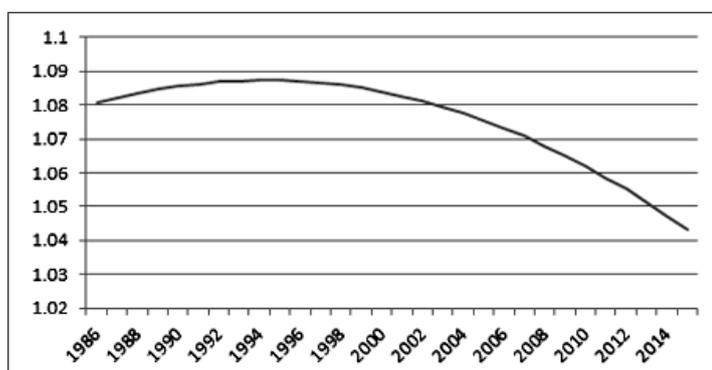


Рис. 4. Энтропия Шеннона для цепочек с логистическим ростом

График изменения значения вычисленной энтропии Шеннона, показывает небольшой, плавный рост в начале рассматриваемого периода, затем менее плавное снижение к концу рассматриваемого периода.

Вычислим энтропию Шеннона по формуле (5) для цепочек с линейно-логарифмическим ростом (рис. 5).

График изменения значений вычисленной энтропии Шеннона показывает примерно равные значения до 2010 года, затем, начиная с 2010 года, график демонстрирует резкое снижение к концу рассматриваемого периода.

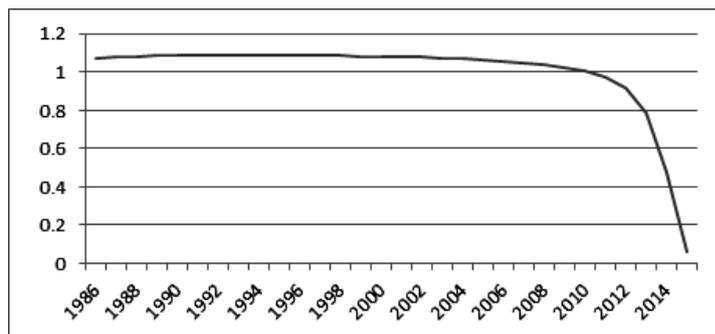


Рис. 5. Энтропия Шеннона для цепочек с линейно-логарифмическим ростом

Энтропию Шеннона можно рассматривать как меру недостатка информации о системе. Чем меньше энтропия, тем больше информация. В этом случае говорят о том, что система упорядочена. Чем больше энтропия, тем меньше информации о системе. Часто говорят о том, что система приближается к состоянию хаоса. Таким образом, видно, что во всех построенных моделях мера неопределённости постепенно уменьшается, принимая наименьшее значение в конце рассматриваемого периода.

7. ЗАКЛЮЧЕНИЕ

В представленной статье показан алгоритм построения вероятностных цепочек с полиномиальным ростом, предложенный М. Сонисом [9]. Приведены общие формулы для вычисления цепочек с полиномиальным ростом.

В исследовании продемонстрировано применение вероятностных цепочек с полиномиальным ростом на примере данных ВНД Канады, Великобритании и США. Также в статье приведено описание алгоритма для построения вероятностных цепочек с помощью программы MATLAB для вышеуказанного примера. Интерполяция данных проведена на период с 1985 по 2000 год, и экстраполяция на период с 2001 по 2015 год.

Полученные результаты были оценены с помощью коэффициента корреляции, а также энтропии Шеннона. Найденные значения коэффициента корреляции для цепочек с полиномиальным, логистическим и линейно-логарифмическим ростом дают основание сделать вывод о том, что модель с полиномиальным ростом ближе к эмпирическим данным и позволяет дать более точный прогноз по сравнению с цепочками с логистическим и линейно-логарифмическим ростом. Относительно результатов найденных значений энтропии Шеннона можно заметить, что во всех видах построенных вероятностных цепочек для спрогнозированного периода с 2001 по 2015 год показатель энтропии уменьшается и принимает наименьшее значение в 2015 году, что означает постепенное уменьшение меры недостатка информации.

Список литературы

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. 2011. № 3. С. 93–97.
2. *Афанасьева Е. В.* Моделирование процессов распределения ресурсов с помощью вероятностных цепочек // Дифференциальные уравнения и процессы управления. 2011. № 3. С. 84–137.

3. Гуревич Е. Квантовые принципы человеческих отношений. СПб.: Изд. дом «Сад искусств», 2017.
4. Логина Н. В. Об одном методе моделирования динамики социально-экономических процессов // Компьютерные инструменты в образовании. 2018. № 2. С. 14–24. doi: 10.32603/2071-2340-2018-2-14-24
5. Логина Н. В., Ампилова Н. Б. Применение линейного бифуркационного анализа для оценки достоверности вероятностных цепочек // Материалы научной конференции «Некоторые актуальные проблемы современной математики и математического образования. Герценовские чтения – 2019», 8–12 апреля 2019 г. СПб.: Изд. РГПУ им. А.И. Герцена, 2019. С. 209–218.
6. Hewings G. J. D. Regional industrial analysis and development. London: Methuen & Co, 1977. 180 p.
7. Shannon C. E. A. Mathematical Theory of Communication // The Bell System Technical Journal. 1948. Vol. 27, № 3. P. 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x
8. Hewings G. J. D., Madden M. Social and demographic accounting. Cambridge: Cambridge University Press, 1995. 242 p. doi: 10.1017/CBO9780511559860
9. Sonis M. Discrete Non-Linear Probabilistic Chains (M. Drachlin and E. Litsyn eds) // Functional-Differential Equations, Ariel, Israel, 2004. № 10. P. 593–639.
10. 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.
11. Группа Всемирного банка [Online]. Available: <https://data.worldbank.org> (дата обращения: 06.03.2020).
12. Archived documentation for previous versions of Matlab [Online]. Available: <https://www.mathworks.com/help/documentation-center> (дата обращения: 23.05.2020).

Поступила в редакцию 19.03.2020, окончательный вариант — 22.04.2020.

Логина Наталья Владиславовна, инженер-программист ООО «ЭПАМ Систэмз», natalia.loginowa@gmail.com

Computer tools in education, 2020

№ 3: 56–69

<http://cte.eltech.ru>

doi:10.32603/2071-2340-2020-3-56-69

Probability Chains with Polynomial Growth as a Model of Resource Distribution

Loginova N. V.¹, Software Engineer, natalia.loginowa@gmail.com

¹EPAM Systems, Ltd, 41, Chyornoy rechki embankment, 197342, Saint Petersburg, Russia

Abstract

The article is devoted to the method of discrete probability chains for constructing the forecast of changes in socio-economic data. Initial data about the distribution of the resource among several participants are presented in the form of the probabilistic vector, and its changing over time is described by a discrete dynamical system which is specified by a certain function. Chains with linear and logarithmic-linear growth have been well studied. In this paper, we consider the probabilistic chains in which the right-hand side is given by polynomials of a certain type. The results of the construction are applied to

the research of the dynamics of the distribution of the national income of Canada, Great Britain, and the United States. The accuracy of the results obtained is estimated by using the correlation coefficient, and the dynamics of the process modeled is estimated by using Shannon's entropy.

Keywords: *dynamic systems, probability chains, extrapolation, polynomial growth, Shannon entropy.*

Citation: N. V. Loginova, "Probability Chains with Polynomial Growth as a Resource Distribution Model," *Computer tools in education*, no. 3, pp. 56-69, 2020 (in Russian); doi: 10.32603/2071-2340-2020-2-56-69

References

1. E. V. Afanasyeva, "Nonlinear probabilistic chain-based modelling of the distribution of economic resources consumption (by the example of western european countries)," *St. Petersburg Polytechnical University Journal. Computer Science. Telecommunication and Control Systems*, no. 3, pp. 93–97, 2011 (in Russian).
2. E. V. Afanasyeva, "Probabilistic Chains Theory-Based Modelling of Socio-Economic Resources Distribution Processes," *Differential Equations and Control Processes*, no. 3, pp. 84–137, 2011 (in Russian).
3. E. Hourowitz, *Quantum Principles of Human Relations*, St. Petersburg, Russia: Sad iskusstv, 2017 (in Russian).
4. N. V. Loginova, "On a Method of Modeling of Socio-Economic Processes Dynamics," *Computer tools in education*, no. 2, pp. 14–24, 2018 (in Russian); doi: 10.32603/2071-2340-2018-2-14-24
5. N. V. Loginova and N. B. Ampilova, "On the application of linear bifurcation analysis to estimating the reliability of probabilistic chain models," in *Proc. of the scientific conference «Some actual problems of modern mathematics and mathematical education. Herzen Readings–2019», April 8–12, 2019*, St. Petersburg, Russia: RGPU, 2019, pp. 209–218 (in Russian).
6. G. J. D. Hewings, *Regional industrial analysis and development*, London: Methuen & Co, 1977.
7. C. E. A. Shannon, "Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948; doi: 10.1002/j.1538-7305.1948.tb01338.x
8. G. J. D. Hewings and M. Madden, eds., *Social and Demographic Accounting*, Cambridge: Cambridge University Press, 1995; doi: 10.1017/CBO9780511559860
9. M. Sonis, "Discrete Non-Linear Probabilistic Chains," *Functional-Differential Equations*, vol. 10, no. 3–4, pp. 593–639, 2004.
10. M. Sonis and G. J. D. Hewings, "Regional Competition and Complementarity: Comparative Advantage/Disadvantages and Increasing/Diminishing Returns in Discrete Relative Spatial Dynamics," in *Regional Competition*, Berlin: Springer, 2000, pp. 139–158; doi: 10.1007/978-3-662-04234-2_7
11. World Bank, *World Bank Group — International Development, Poverty, & Sustainability*, 2020. [Online]. Available: <https://data.worldbank.org>
12. MathWorks, "Archived documentation for previous versions of Matlab," *MathWorks*. [Online]. Available: <https://www.mathworks.com/help/documentation-center>

Received 19.03.2020, the final version — 22.04.2020.

Loginova Natalia, Software Engineer, EPAM Systems, Ltd, natalia.loginowa@gmail.com