EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 21. KÖIDE
ÜHISKONNATEADUSED. 1972, NR. 3

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 21
ОБЩЕСТВЕННЫЕ НАУКИ. 1972, № 3

https://doi.org/10.3176/hum.soc.sci.1972.3.04

Г. ФЕЛИЦИУС

АППРОКСИМАЦИЯ СТАТИСТИЧЕСКИХ РЯДОВ С ИСПОЛЬЗОВАНИЕМ ОРТОНОРМИРОВАННЫХ ФУНКЦИЙ

При экономико-статистических исследованиях производственных процессов часто возникает потребность в аппроксимации статистических рядов различными степенными полиномами. Обычно такая аппроксимация выполняется по способу наименьших квадратов с использованием системы нормальных уравнений, но в последнее время все большее внимание привлекают и другие методы, в том числе метод аппроксимации с использованием ортогональных функций Чебышева, выгодно отличающийся от других своей экономичностью. Однако использование ортогональных полиномов ограничено применимостью метода в основном для аппроксимации полными параболами различных степеней. Кроме того, ему присущи определенные сложности вычислительных алгоритмов и рекуррентных соотношений.

Эти в принципе преодолимые трудности особенно сказываются при реализации метода на малых вычислительных машинах, что побуждает исследователей обращаться в ряде случаев к классической схеме составления системы нормальных уравнений, предпочитая ее другим рациональным методам. В связи с этим возникает задача создать такой метод аппроксимации полиномами различных степеней, который сочетал бы в себе простоту алгоритмов с достаточной экономичностью вычислительных процедур.

В своем методе П. Л. Чебышев использовал только свойство ортогональности функций заданному полиному, получив при этом значительное упрощение вычислительного процесса. Оказывается, использование как свойства ортогональности, так и свойства нормированности подобных функций позволяет выявить новые, весьма ценные свойства способа наименьших квадратов и создать метод аппроксимации, выгодно отличающийся от применяемых в настоящее время.

Предлагаемый ниже метод аппроксимации, основанный на использовании ортонормированных функций, в определенной мере удовлетворяет этим требованиям как при использовании малых вычислительных машин, так и ЭВМ.

Пусть требуется определить по способу наименьших квадратов заданное таблицей значений аналитическое выражение полинома вида

$$y=a_0+a_1x+a_2x^2+\ldots+a_mx^m. \tag{1}$$

Решение данной задачи заключается в определении коэффициентов уравнения регрессии $a_0, a_1, a_2, \ldots, a_m$ из системы нормальных уравнений:

$$\left.\begin{aligned}
&a_0n+a_1\sum x+a_2\sum x^2+\ldots+a_m\sum x^m=\sum y\\
&a_0\sum x+a_1\sum x^2+a_2\sum x^3+\ldots+a_m\sum x^{m+1}=\sum yx\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_0\sum x^m+a_1\sum x^{m+1}+\ldots+a_m\sum x^{2m}=\sum yx^m,
\end{aligned}\right\} \tag{2}$$

где n — число членов статистического ряда. Применение ортонормированных функций позволяет при нахождении коэффициентов уравнения регрессии построить более простую систему уравнений, чем (2).

Введем в рассмотрение систему ортогональных и нормированных на множестве точек $x_1, x_2, \ldots, x_n$ к полиному (1) функций вида $\varphi_k(x)$, для которых выполняются условия

$$\sum\varphi_k(x)x^r=0, \text{ при } k\neq r \quad \text{и} \quad \sum\varphi_k(x)x^r=1, \text{ при } k=r.$$

Умножим полином (1) последовательно на систему ортонормированных функций $\varphi_0(x); \varphi_1(x); \ldots; \varphi_m(x)$. Суммируя полученные равенства по системе точек $x_1; x_2; x_3, \ldots, x_m$, составим из них следующую систему уравнений:

$$\left.\begin{array}{l}
a_0\sum\varphi_0(x)+a_1\sum x\varphi_0(x)+a_2\sum x^2\varphi_0(x)+\ldots+a_m\sum x^m\varphi_0(x)=\sum y\varphi_0(x)\\
a_0\sum\varphi_1(x)+a_1\sum x\varphi_1(x)+a_2\sum x^2\varphi_1(x)+\ldots+a_m\sum x^m\varphi_1(x)=\sum y\varphi_1(x)\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
a_0\sum\varphi_m(x)+a_1\sum x\varphi_m(x)+a_2\sum x^2\varphi_m(x)+\ldots+a_m\sum x^m\varphi_m(x)=\sum y\varphi_m(x).
\end{array}\right\} \quad (3)$$

Исходя из условий ортогональности и нормированности функций $\varphi_0(x); \varphi_1(x)\ldots$ $\ldots\varphi_m(x)$ сумма $\sum\varphi_0(x)=1$, а суммы $\sum\varphi_1(x)=\sum\varphi_2(x)=\ldots=\sum\varphi_m(x)=0$, в результате чего система уравнений (3) распадается на уравнение, связывающее значение коэффициента a_0 с другими коэффициентами уравнения регрессии вида

$$a_0=\sum y\varphi_0(x)-a_1\sum x\varphi_0(x)-a_2\sum x^2\varphi_0(x)-\ldots-a_m\sum x^m\varphi_0(x)=$$

$$=\frac{1}{n}\left[\sum y-a_1\sum x-a_2\sum x^2-\ldots-a_m\sum x^m\right]; \quad (4)$$

и систему уравнений, из которой могут быть определены $a_1, a_2\ldots a_m$:

$$\left.\begin{array}{l}
a_1\sum x\varphi_1(x)+a_2\sum x^2\varphi_1(x)+\ldots+a_m\sum x^m\varphi_1(x)=\sum y\varphi_1(x)\\
a_1\sum x\varphi_2(x)+a_2\sum x^2\varphi_2(x)+\ldots+a_m\sum x^m\varphi_2(x)=\sum y\varphi_2(x)\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
a_1\sum x\varphi_m(x)+a_2\sum x^2\varphi_m(x)+\ldots+a_m\sum x^m\varphi_m(x)=\sum y\varphi_m(x).
\end{array}\right\} \quad (5)$$

Для решения уравнения (4) и системы уравнений (5) необходимо также определить значения ортонормированных функций $\varphi_0(x); \varphi_1(x), \ldots, \varphi_m(x)$ для n точек независимой переменной x в равноотстоящих узлах $(x=1, 2, 3, \ldots, n)$, которые могут быть найдены из следующей системы уравнений:

$$\left.\begin{array}{l}
\varphi_0(x)n+\varphi_1(x)\sum x+\varphi_2(x)\sum x^2+\ldots+\varphi_m\sum x^m=1\\
\varphi_0(x)\sum x+\varphi_1(x)\sum x^2+\varphi_2(x)\sum x^3+\ldots+\varphi_m\sum x^{m+1}=x\\
\varphi_0(x)\sum x^2+\varphi_1(x)\sum x^2+\ldots+\varphi_m\sum x^{m+2}=x^2\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\varphi_0(x)\sum x^m+\varphi_1(x)\sum x^{m+1}+\ldots+\varphi_m\sum x^{2m}=x^m.
\end{array}\right\} \quad (6)$$

При этом возможен двоякий подход. Значения ортонормированных функций в равноотстоящих узлах аргумента x могут быть вычислены непосредственно путем подстановки в систему (6) значений $(x=1, 2, 3, \ldots, n)$ с последующим решением системы уравнений n раз. Или же в систему уравнений (6) вместо сумм вида $\sum x$; $\sum x^2$; $\sum x^m$ можно подставить их значения, выраженные через n и имеющие вид

$$\sum x=\frac{n(n+1)}{2}; \quad \sum x^2=\frac{n(n+1)(2n+1)}{6}; \quad \sum x^3=\frac{n^2(n+1)^2}{4} \quad \text{и т. д.}$$

Тогда в результате решения системы (6) можно получить аналитические выражения ортонормированных функций, по которым путем подстановки значений аргумента ($x=1, 2, 3, \ldots, n$) вычисляются постоянные коэффициенты, применяемые при последующих расчетах.

Такой подход, позволяющий избежать громоздких вычислений значений сумм аргумента x при показателе степени m выше третьей, был использован автором для параболической аппроксимации и экстраполяции таблично заданных функций в случае равноотстоящих значений аргумента [1]. По этому же методу были рассчитаны значения ортонормированных функций при различных m в интервале от $n=3$ до $n=25$, таблицы которых приведены в [2].

Если значения указанных функций подлежат определению на ЭВМ, можно рекомендовать первый подход, позволяющий оперировать системами уравнений более высоких порядков.

Сопоставление (2) с (4) и (5) показывает, что классической системе нормальных уравнений для определения коэффициентов уравнения регрессии параболического вида может соответствовать система уравнений, основанная на ортонормированном преобразовании функции от независимой переменной x.

Далее оказалось, что отличие системы ортонормированных функций от системы нормальных уравнений способа наименьших квадратов не только формально.

Прежде всего система уравнений (5) имеет преимущество перед классической системой нормальных уравнений, заключающееся в том, что порядок системы, основанной на применении ортонормированных функций, на единицу меньше, что значительно снижает объем вычислений и вероятность ошибок.

Кроме того, при выравнивании статистических рядов по параболам и гиперболам различных степеней значения ортонормированных функций в равноотстоящих узлах могут быть представлены в виде целых чисел или простых дробей, что облегчает осуществление попарных произведений под знаком сумм. Существенно также, что большинство вычислений по данному методу имеет простой алгоритм, единый для всех элементов системы уравнений.

Достоинство предлагаемой системы уравнений в ее универсальности. При подстановке в нее вместо независимой переменной x значений аргументов многих переменных, например $x_1, x_2, x_3, \ldots, x_m$, одна и та же система уравнений может быть использована при определении коэффициентов уравнений как полиноминальной регрессии, так и множественной.

В силу ортонормированности функций система уравнений (4) и (5) для случая равноотстоящих значений независимой переменной x распадается на отдельные тождества вида

$$a_0=\sum y\varphi_0(x);\ \ a_1=\sum y\varphi_1(x);\ \ \ldots;\ \ a_m=\sum y\varphi_m(x),$$

позволяющие исключительно простым способом определять все коэффициенты уравнения регрессии, а также находить точки экстраполяции сглаживаемых функций [3].

При реализации предлагаемого метода на ЭВМ появляется возможность использовать одну и ту же систему уравнений для аппроксимации различными полиномами при линейной и нелинейной множественной корреляции, а также применения их различных сочетаний для отдельных факториальных признаков.

При использовании ортонормированных функций облегчается создание на базе одной системы уравнений универсальной вычислительной программы для аппроксимации полиномами различного вида, в том числе и тригонометрическими, в различных сочетаниях при парной и множественной корреляции.

Указанное следует из того, что все элементы системы уравнения (5) линейно независимы, а это позволяет, не нарушая целостности системы, производить в нем различные подстановки и преобразования как аргумента, так и функции.

Так, если в системе (5) умножить значение функции правых частей всех уравнений на x^m без изменения других элементов системы, то она станет пригодной для аппроксимации гиперболами степени m.

Кроме того, система (5) позволяет проводить аппроксимацию таблично заданных функций как в обычной системе отсчета, так и при переносе начала отсчета либо в середину статистического ряда, либо в любую точку оси абсцисс.

Нет сомнения, что в процессе дальнейших исследований выявятся и другие положительные свойства ортонормированных функций, однако перечисленных выше вполне достаточно для того, чтобы по достоинству оценить предлагаемый метод.

Подтвердим сказанное примером аппроксимации статистического ряда полиномом сложного вида с использованием ортонормированных функций. Исходные данные — табличные значения аргумента — x и функции — y — приведены в графах 1 и 2 таблицы; в качестве аппроксимирующей выбираем зависимость вида:

$$y=\ln\left(a_0+a_1\sqrt{x}+a_2\ln x^2+a_3\sin\frac{2\pi x_i}{x_n}\right).$$

Достаточно сложный вид аппроксимирующей зависимости выбран для того, чтобы проиллюстрировать разрешающую способность предлагаемого метода. Заметим в этой связи, что решение подобной задачи известными методами сопряжено со значительными вычислительными трудностями, которые сложно преодолеть при ручном способе расчета.

В графах 3—6 таблицы приведены преобразованные значения исходных табличных данных функции и аргумента, остальные графы таблицы являются расчетными.

Ортонормированные коэффициенты функций $\varphi_3(x)$, $\varphi_2(x)$ и $\varphi_1(x)$ в графах 7, 12 и 17 взяты из таблиц, табулированных в [2]. В нижней строке таблицы приведены суммы преобразованных табличных значений функции и аргумента, а также суммы их попарных произведений на ортонормированные коэффициенты.

Из итоговых данных таблицы составим следующую систему уравнений:

$$\left.\begin{aligned}2{,}833\,a_1+7{,}38150\,a_2+1{,}648\,a_3&=16{,}0150\\ a_1+2{,}1478\,a_2+0{,}3550\,a_3&=\ \ 4{,}7334\\ 0{,}1667\,a_1+0{,}2621\,a_2+0{,}0021\,a_3&=\ \ 0{,}6154.\end{aligned}\right\}$$

Решая полученную систему третьего порядка по схеме Сарруса, имеем: $a_1=1$; $a_2=3$; $a_3=-2$; откуда

$$a_0=\frac{1}{4}\,[46{,}0968+11{,}0-3\cdot 13{,}6048+2\cdot 1{,}8588]=$$

$$=\frac{1}{4}\,[60{,}8144-40{,}8144]=\frac{20}{4}=5.$$

Рассмотренная выше система уравнений с использованием ортонормированных функций по праву может называться канонической, поскольку обладает целым рядом замечательных свойств, присущих системам нормальных уравнений различного вида, построенных для решения локальных задач способа наименьших квадратов. Единственным недостатком является то, что для случая неравноотстоящих значений независимой переменной решение будет приближенным.

Для аппроксимации по зависимостям практически любого вида и множественной корреляции в случае если аргумент задан в неравноотстоящих узлах могут быть получены функции, подобные ортонормированным, использование которых позволяет определять точные значения коэффициентов уравнения регрессии.

Если через $\psi_1(x_i)$; $\psi_2(x_i)$, ..., $\psi_m(x_i)$ обозначить табличные значения подобных ортонормированных функций, то они могут быть вычислены по следующему алгоритму:

$$\psi_1(x_i)=x_i-\frac{\sum x_i}{n};\quad \psi_2=x_i{}^2-\frac{\sum x_i{}^2}{n};\ \ldots\ \psi_m=x_i{}^m-\frac{\sum x_i{}^m}{n}.$$

x	y	$\sqrt{x}$	$\ln x^2$	$\sin \frac{2\pi x_i}{x_n}$	e^y	$\varphi_3(x)$	$\sqrt{x}\varphi_3(x)$	$\ln x^2\varphi_3(x)$	$\sin \frac{2\pi x_i}{x_n}\varphi_3(x)$	$e^y\varphi_3(x)$	$\varphi_2(x)$	$\sqrt{x}\varphi_2(x)$
1	2	3	4	5	6	7	8	9	10	11	12	13
1	1,2528	1	0	0,2504	3,4992	—1	—1	0	—0,2504	—3,4992	3	3
4	2,2641	2	2,7726	0,8480	9,6218	3	6	8,3178	2,5440	28,8654	—8	—16
9	2,6144	3	4,3944	0,7604	13,6624	—3	—9	—13,1832	—2,2812	—40,9872	7	21
25	2,9607	5	6,4378	0	19,3134	1	5	6,4378	0	19,3134	—2	—10
						6	1	1,5724	0,0124	3,6924	2	—2
		11,0	13,6048	1,8588	46,0968		0,1667	0,2621	0,0021	0,6154		—1

$\ln x^2\varphi_2(x)$	$\sin \frac{2\pi x_i}{x_n}\varphi_2(x)$	$e^y\varphi_2(x)$	$\varphi_1(x)$	$\sqrt{x}\varphi_1(x)$	$\ln x^2\varphi_1(x)$	$\sin \frac{2\pi x_i}{x_n}\varphi_1(x)$	$e^y\varphi_1(x)$
14	15	16	17	18	19	20	21
0	0,7512	10,4976	—26	—26	0	—6,5104	—90,9792
—22,1808	—6,7840	—76,9744	57	114	158,0382	48,3360	548,4426
30,7608	5,3228	95,6368	—42	—126	—184,5648	—31,9368	—573,8208
—12,8756	0	—38,6268	11	55	70,8158	0	212,4474
—4,2956	—0,7100	—9,4668	6	17	44,2892	9,8888	96,0900
—2,1478	—0,3550	—4,7334		2,8333	7,3815	1,6481	16,0150

Указанные функции могут быть подставлены в систему уравнений (5) вместо ортонормированных функций $\varphi_1(x)$; $\varphi_2(x)$; ... $\varphi_m(x)$, порядок решения системы остается прежним.

В заключение следует отметить, что в Институте экономики АН ЭССР разработана на этой основе универсальная программа нахождения коэффициентов парной и множественной регрессий для аппроксимации по широкому классу зависимостей.

ЛИТЕРАТУРА

1. Г. Фелициус, Выравнивание и экстраполяция эмпирических распределений с применением постоянных коэффициентов. Известия АН ЭССР. Общественные науки, 1970, **19**, № 1, 28—33.
2. Методика статистической обработки эмпирических данных. Руководящий технический материал РТМ44-70 (проект), изд. Комитета стандартов, мер и измерительных приборов при СМ СССР. М., 1970 (ротапринт).
3. Г. Фелициус, Упрощенный метод аппроксимации, интерполяции и экстраполяции таблично заданных функций. Экономика и математические методы, 1971. **VII**, 6, стр. 933—938.

Институт экономики
Академии наук Эстонской ССР

Поступила в редакцию
31/VIII 1971

G. FELICIUS

STATISTILISTE RIDADE APROKSIMEERIMINE ORTONORMEERITUD FUNKTSIOONIDEGA

Resümee

Artiklis käsitletakse paaris ja mitmese, lineaarse ja mittelineaarse regressioonivõrrandi kordajate leidmist vähimruutude meetodil, kasutades ortonormeeritud funktsioone, kusjuures sõltumatu muutuja muutub mittekonstantse sammuga. Teiste tuntud meetoditega võrreldes võimaldab kõnesolev alandada normaalvõrrandite süsteemi järku ning lihtsustab arvutusi. Seega võimaldab ta laiendada vähimruutude meetodit ning paaris ja mitmese korrelatsiooni kasutusala. Kirjeldatud meetod on realiseeritud elektronarvutil lihtsa universaalse programmina.

Eesti NSV Teaduste Akadeemia
Majanduse Instituut

Toimetusse saabunud
31. VIII 1971

G. FELICIUS

APPROXIMATION OF STATISTICAL SERIES BY ORTHONORMED FUNCTIONS

Summary

The author discusses the method of determining the coefficients of the equation of double and multiple, linear and nonlinear regressions by means of the least squares method for distribution values of the independent variable with the utilization of orthonormed functions.

The method, in comparison with the known ones, allows to decrease the order of the system of normal equations and simplifies the computations carried out on small computers.

For the realization of the method on the computer, simple universal programs may be created, permitting to widen the sphere of application of the method of the least squares in the case of a double and multiple correlation.

Academy of Sciences of the Estonian SSR,
Institute of Economics

Received
Aug. 31, 1971