EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 19. KÕIDE
ÜHISKONNATEADUSED. 1970, NR. 1

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 19
ОБЩЕСТВЕННЫЕ НАУКИ. 1970, № 1

https://doi.org/10.3176/hum.soc.sci.1970.1.03

Г. ФЕЛИЦИУС

ВЫРАВНИВАНИЕ И ЭКСТРАПОЛИРОВАНИЕ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ С ПРИМЕНЕНИЕМ ПОСТОЯННЫХ КОЭФФИЦИЕНТОВ

В мировой практике обработки результатов наблюдений выравнивание эмпирических рядов распределений осуществляется в основном по способу наименьших квадратов. Среди различных модификаций этого способа наибольшее распространение находят выравнивания на основе составления и решения системы нормальных уравнений и по методу П. Л. Чебышева с использованием таблиц постоянных коэффициентов, вычисленных на базе специальных полиномов.

Классический метод выравнивания с решением системы уравнений весьма громоздок и сопряжен с некоторыми вычислительными трудностями. Так, при выравнивании по параболе третьей степени необходимо решить систему из пяти определителей четвертого порядка.

Метод выравнивания при помощи полиномов Чебышева менее трудоемок и рекомендован для параболического выравнивания эмпирических распределений [1]. Использование полиноминальных коэффициентов несколько сокращает объем вычислений, но не сводит их к минимуму, не говоря уже о сложности вычислительных процедур.

Указанные трудности позволяет преодолеть ЭЦВМ, однако она не всегда доступна, а в ряде случаев применение ее может оказаться и нецелесообразным.

Между тем в последнее время наука и практика испытывают потребность в общедоступном методе выравнивания эмпирических распределений с минимально возможным объемом вычислений. Кроме того, получившие широкое распространение методы прогнозирования различных процессов, особенно производственных и экономических, требуют создания простых вычислительных методов экстраполяции эмпирических рядов распределений по способу наименьших квадратов.

Этим требованиям наиболее полно отвечает предлагаемая методика выравнивания и экстраполяции эмпирических распределений для равноотстоящих узлов по способу наименьших квадратов, основанная на предварительном исчислении и использовании постоянных числовых коэффициентов.

Для определения коэффициентов получены рекуррентные формулы, вывод которых рассмотрим для случая линейной аппроксимации функции, заданной таблично.

Пусть $y_1, y_2, y_3, \ldots, y_n$ — значения функции некоторого процесса, независимая переменная которого принимает целочисленные значения $x=1, 2, 3, \ldots, n$, и требуется найти линейную корреляционную зависимость между ними вида $y=a_1x+a_0$ по способу наименьших квадратов.

Система нормальных уравнений для определения коэффициентов a_1 и a_0 имеет следующий вид:

$$\left.\begin{aligned} na_0+a_1\sum x &= \sum y \\ a_0\sum x+a_1\sum x^2 &= \sum xy \end{aligned}\right\}.$$

Решая эту систему, получим формулы для определения коэффициентов a_1 и a_0 в общем виде:

$$a_1=[n\sum xy-\sum x\sum y]/[n\sum x^2-(\sum x)^2];$$

$$a_0=[\sum x^2\sum y-\sum x\sum xy]/[n\sum x^2-(\sum x)^2].$$

Заметим, что входящие в формулы для коэффициентов суммы $\sum x$ и $\sum x^2$ для случая равноотстоящих узлов представляют собой сумму первых n чисел натурального ряда в первой и второй степени, а, следовательно, могут быть представлены как

$$\sum x=\frac{n(n+1)}{2};\quad \sum x^2=\frac{n(n+1)(2n+1)}{6},$$

где n — число суммируемых членов ряда.

Подставим значения сумм в формулы коэффициентов, для чего сначала определим знаменатель выражений

$$n\sum x^2-(\sum x)^2=n\,n(n+1)(2n+1)/6-n^2(n+1)^2/4=$$
$$=n^2(n+1)[(2n+1)/3-(n+1)/2]/2=n^2(n+1)(n-1)/12.$$

Числитель коэффициента a_1 можно представить в виде $n\sum xy-\sum x\sum y=$ $=(2n\sum xy-n(n+1)\sum y)/2$,

откуда получим преобразованное значение

$$a_1=\sum_{i=1}^{i=n} y_i\,6[2x_i-(n+1)]/n(n+1)(n-1). \tag{1}$$

Аналогично представим числитель коэффициента a_0:

$$\sum x^2\sum y-\sum x\sum xy=n(n+1)(2n+1)\sum y/[6-n(n+1)\sum xy/2].$$

Окончательно найдем выражение для a_0 в виде

$$a_0=\sum_{i=1}^{i=n} y_i\,2(2n+1-3x_i)/n(n-1). \tag{2}$$

Выражения (1) и (2) являются функциями двух переменных n и x_i, где x_i принимает значения натурального ряда чисел от единицы до n. Значения этих функций, вычисленные при $i=1, 2, 3, \ldots, (n-1), n$, представляют собой коэффициенты, по которым осуществляется аппроксимация эмпирического ряда распределения, и могут быть найдены из следующих выражений:

$$K_{a_1}=6(2x_i-n-1)/n(n+1)(n-1); \tag{3}$$

$$K_{a_0}=2(2n+1-3x_i)/n(n-1). \tag{4}$$

Возможность использования (3) и (4) в качестве постоянных коэффициентов аппроксимации вытекает из (1) и (2), поскольку в них производится суммирование попарных произведений y_i на значения функции $f(x_i)$ при одинаковых индексах суммирования ($i=1, 2, 3, \ldots, n$ и $i=x_i$).

Вычисление постоянных коэффициентов и их использование для определения параметров аппроксимирующей прямой рассмотрим на примере.

Пусть известны следующие значения ряда распределения при равноотстоящих значениях независимой переменной:

$y_i=10, 8, 9, 6, 4, 2$ и 1; при $x_i=1, 2, 3, 4, 5, 6$ и 7.

Необходимо найти коэффициенты a_1 и a_0 прямолинейного уравнения регрессии $y=a_1x+a_0$.

Подставляя в (3) и (4) количество равноотстоящих узлов интерполяции $n=7$, получим следующее выражение для определения постоянных коэффициентов:

$$K_{a_1}=(x_i-4)/28;\quad K_{a_0}=(5-x)/7.$$

Придавая x_i последовательно значения $x_i=1, 2, 3, \ldots, 7$, получим следующие коэффициенты:

$$28\ K_{a_1}=+3,\ -2,\ -1,\ 0,\ +1,\ +2,\ +3;$$

$$7\ K_{a_0}=+4,\ +3,\ +2,\ +1,\ 0,\ -1,\ -2.$$

Перемножив y_i и K_{a_1}, а также y_i и K_{a_0} и сложив полученные произведения, окончательно получим:

$$a_1=(-3\times10-2\times8-1\times9+0\times6+1\times4+2\times2+3\times1)/28=-11/7;$$

$$a_0=(4\times10+3\times8+2\times9+1\times6+0\times4-1\times2-2\times1)/7=12.$$

Следовательно, уравнение регрессии будет иметь вид

$$y=-11/7x+12.$$

Коэффициенты уравнения регрессии могут быть получены и из системы нормальных уравнений, если предварительно найти, что $\sum x_i^2=140$; $\sum x_i=28$; $\sum y_i=40$; $\sum y_i x_i=116$,

тогда

$$a_1=7(116-160)/49\times4=-44/28=-11/7;$$

аналогично

$$a_0=(140\times40-28\times116)/(7\times140-784)=2352/196=12.$$

Из приведенного примера видно, что определение коэффициентов уравнения регрессий классическим способом (из системы нормальных уравнений) и по предлагаемой методике приводит к одним и тем же результатам, несмотря на различие в вычислительных процедурах. Однако при предлагаемом методе постоянные коэффициенты K_{a_1} и K_{a_0} могут быть вычислены заранее и сведены в таблицы, которые используются по мере надобности, что не только упрощает вычисления, но сводит до минимума возможность появления ошибок.

Рассмотренный способ определения коэффициентов уравнения прямолинейной регрессии может быть распространен и на обратную пропорциональную зависимость между величинами.

Если правую часть уравнения вида $y=a_1x+a_0$ разделить на независимую переменную x, то можно получить уравнение равносторонней гиперболы $y_1=a_1+a_0/x$. Очевидно, что от уравнения прямой можно перейти к гиперболической зависимости, если независимую переменную разделить на x, а коэффициенты K_{a_1} и K_{a_0} умножить на x_i. В этом случае постоянные коэффициенты для гиперболической зависимости вида $y=a_1/x+a_0$ можно вычислить путем умножения соответствующих уравнений (3) и (4) на x_i и замены их индексов, т. е. из выражений вида

$$K_{a_1}=2[x(2n+1)-3x^2]/n(n-1); \tag{5}$$

$$K_{a_0}=6[2x^2-x(n+1)]/n(n+1)(n-1) \tag{6}$$

В случае экспоненциальной зависимости $y=a_0a_1^x$ постоянные коэффициенты для определения уравнения регрессии могут быть вычислены из (3) и (4), поскольку эта зависимость путем логарифмирования может быть сведена к линейной вида

$$\lg y=x\lg a_1+\lg a_0.$$

При использовании указанных постоянных коэффициентов следует лишь несколько изменить порядок вычислений в формуле для определения уравнения регрессии, а именно

$$\lg a_1 = \sum_{i=1}^{i=n} \lg y_i K_{a_1}; \quad \lg a_0 = \sum_{i=1}^{i=n} \lg y_i K_{a_0}, \tag{7}$$

где K_{a_1} и K_{a_0} берутся из (3) и (4).

Коэффициенты уравнения регрессии a_1 и a_0 могут быть найдены из (7) по таблицам антилогарифмов.

Переходя к определению постоянных коэффициентов уравнений регрессии параболического типа, заметим, что ограниченные размеры данной статьи и некоторая сложность полученных выражений не позволяют нам полностью привести формулы и их вывод. Ограничимся формулами для определения постоянных коэффициентов уравнения регрессии второго порядка $y = a_2x^2 + a_1x + a_0$, которые приводятся без доказательства.

$$K_{a_2} = 30[(n+1)(n+2) - 6x_i(n+1) + 6x_i^2] / n(n+1)(n+2)(n-1)(n-2); \tag{8}$$

$$K_{a_1} = 6[-3(n+2)(2n+1) + 2x_i(2n+1)(8n+11)/(n+1) - 30x_i^2] :$$
$$: n(n+2)(n-1)(n-2); \tag{9}$$

$$K_{a_0} = 3[(3n^2+3n+2) - 6x_i(2n+1) + 10x_i^2] / n(n-1)(n-2). \tag{10}$$

Подчеркнем, что некоторая громоздкость выражений для вычисления постоянных коэффициентов уравнений регрессии второй и высших степеней еще не означает, что сложны и вычисленные по ним коэффициенты. Наоборот, полученные по выражениям (3)—(10) коэффициенты просты и не затрудняют последующих вычислений.

Анализ показывает, что при параболическом выравнивании постоянные коэффициенты, вычисленные по предлагаемой методике, для коэффициента уравнения регрессии при члене с высшей степенью совпадают с полиноминальными коэффициентами метода Чебышева, приведенными, например, в [1].

Указанное свидетельствует о существовании тесной связи между этими методами и можно полагать, что предлагаемый метод станет существенным дополнением полиноминального метода Чебышева. Наряду с этим предлагаемый метод имеет и важное самостоятельное значение, поскольку он позволяет расширить границы применения способа наименьших квадратов для обработки эмпирических рядов распределений.

В последнее время широко распространены методы прогнозирования различных процессов, использующие в качестве математического аппарата способы экстраполяции зависимостей, полученных для предыстории процесса. Обычно для этого применяется либо метод линейной экстраполяции, либо метод экспоненциального сглаживания. Метод экстраполяции по способу наименьших квадратов, несмотря на его преимущества, почти не используется из-за сложности вычислительных процедур.

Предлагаемый метод для экстраполяции эмпирических рядов распределений по способу наименьших квадратов существенно отличается от известных и позволяет простыми средствами получать значения экстраполируемой функции на один, два и более шагов вперед. Он основан на предварительном вычислении постоянных коэффициентов, по которым определяется значение аппроксимирующей функции для узлов $n+1$, $n+2$ и т. д.

Формулы для определения постоянных коэффициентов могут быть получены из выражений (4), (6), (7), (10) для a_0, если изменить в них порядок подстановки независимого переменного x_i, положив $x_i = n$, $(n-1)$, ... , 3, 2, 1. В этом случае мы получим формулы для определения постоянных коэффициентов экстраполяции на один шаг вперед.

Указанное следует из того, что свободный член уравнения регрессии a_0 характеризует значение экстраполируемой функции на один шаг назад и при изменении направления экстраполяции соответствует изменению порядка суммирования одного из сомножителей под знаком сумм.

Аналогично могут быть получены формулы для вычисления постоянных коэффициентов экстраполяции на два и более шагов вперед. Ниже приведены выражения для вычисления коэффициентов экстраполяции на один и два шага вперед.

Для уравнения регрессии вида $y=a_1x+a_0$:

$$K_{n+1}=2(3x_i-n-2)/n(n-1);$$

$$K_{n+2}=2(n+3)[3x_i-(n^2+6n+5)/(n+3)]/n(n+1)(n-1). \quad (11)$$

Для уравнения регрессии $y=a_1/x+a_0$:

$$K_{n+1}=6[(n+1)^2-3x_i(n+1)+2x_i^2]/n(n+1)(n-1);$$

$$K_{n+2}=2[3x_i^2(n+3)-x_i(n^2+6n+5)]/n(n+1)(n+2)(n-1). \quad (12)$$

Для уравнения регрессии $y=a_2x^2+a_1x+a_0$:

$$K_{n+1}=3[(n^2+5n+6)-2x_i(4n+7)+10x_i^2]/\ldots/n(n-1)(n-2);$$

$$K_{n+2}=3[(n^3+14n^2+43n+30)-2x_i(4n^2+39n+41)+$$

$$+10x_i^2(n+7)]/n(n+1)(n-1)(n-2). \quad (13)$$

Коэффициенты экстраполяции вычисляются по приведенным формулам путем подстановки в них значений независимого переменного $x_i=1, 2, 3, \ldots, n-1, n$, причем индексы при K означают экстраполяцию вправо на один и два шага соответственно.

Применение предлагаемого метода не ограничивается рамками выравнивания эмпирических распределений и экстраполяцией и позволяет расширить существующие возможности способа наименьших квадратов.

Прежде всего необходимо указать на возможность не только экстраполяции, но и интерполяции функции регрессии без нахождения самого уравнения регрессии, причем объем вычислений не возрастает, а даже уменьшается, так как при нахождении значения одной экстраполируемой или интерполируемой точки производятся такие же вычисления, что и при определении каждого коэффициента уравнения a_0, a_1, a_2, a_3 и т. д.

Упрощенным способом, без нахождения коэффициентов уравнения регрессии, могут быть определены по значениям ряда распределения производная и интеграл исследуемой функции, найдены точки перегиба и экстремальные точки кривой аппроксимации, корни уравнения, параметры касательных и т. д.

Если эмпирическими рядами распределений заданы две или более линии регрессии, то, не вычисляя их параметры, можно, производя соответствующие действия над постоянными коэффициентами, осуществлять умножение и деление описывающих их функций и другие действия над ними.

В тех случаях, когда не предъявляются повышенные требования к точности вычислений, выравнивание и экстраполяцию эмпирических распределений можно еще более упростить, для чего по вышеприведенным уравнениям вместо таблиц постоянных коэффициентов составить соответствующие номограммы. Графическое выравнивание и экстраполяция могут быть сведены к выполнению лишь операций сложения над значениями эмпирической функции распределения.

В дальнейшем, возможно, выявятся и другие области применения предлагаемого метода.

Выводы

1. Предлагаемый метод выравнивания и экстраполяции на основе постоянных коэффициентов упростит вычисления за счет использования предварительно составленных таблиц или номограмм.

2. Использование метода расширяет возможности математической обработки результатов экспериментов по способу наименьших квадратов, особенно по их экстраполяции, что может быть использовано при прогнозировании различных производственных и экономических процессов.

ЛИТЕРАТУРА

1. Методика статистической обработки эмпирических данных. Руководящий технический материал РТМ 44-76. Государственное издательство стандартов. М., 1963.

Институт экономики
Академии наук Эстонской ССР

Поступила в редакцию
3/VI 1969

G. FELICIUS

PÜSIVATE KOEFITSIENTIDE ABIL SAADUD EMPIIRILISTE JAOTUSRIDADE ÕGVENDAMISEST JA EKSTRAPOLEERIMISEST

Resümee

Esitatakse uus meetod püsivate arvuliste koefitsientide abil saadud empiiriliste jaotusridade õgvendamiseks ja ekstrapoleerimiseks. Tuuakse ära valemid koefitsientide arvutamiseks vähemruutude meetodil, kui punktide arv muutub.

Käesolev meetod on laialdaselt rakendatav, lihtne ja kasutatav tootmisprotsesside prognoosimiseks.

Eesti NSV Teaduste Akadeemia
Majanduse Instituut

Saabus toimetusse
3. VI 1969

G. FELICIUS

EVENING OUT AND EXTRAPOLATION OF EMPIRIC DISTRIBUTIONS WITH AN APPLICATION OF CONSTANT COEFFICIENTS

Summary

The author presents a new method of evening out and extrapolation of empiric rows of distribution, with the help of constant numeric coefficients.

Formulas are presented for computing coefficients in case of different values of the sum of points to be evened out, with the method of smallest squares.

The author points out the wide applicability of the method, its simplicity and the possibility of extrapolating the evened-out values of the function for forecasting industrial processes.

Academy of Sciences of the Estonian SSR,
Institute of Economics

Received
June 3, 1969