8. Инютин С. А. Вычислительные задачи большой алгоритмической сложности и модулярная арифметика // Вестник Тюменского госуниверситета. 2002. № 3. С. 3-9.

Сергей Александрович БАРДАСОВ — доцент кафедры экономики и управления собственностью, кандидат физико-математических наук

УДК 519.21(075.8)

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОГО ЧИСЛА ГРУПП ИНТЕРВАЛЬНОГО ВАРИАЦИОННОГО РЯДА

АННОТАЦИЯ. Функция максимального правдоподобия используется для определения оптимального числа групп вариационного ряда.

The function of the maximal plausibility is used for definition of optimum number of groups of variational series.

Пусть имеется n объектов, которые характеризуются ранжированными в порядке возрастания численными значениями $x_1 \leq x_2 \leq \ldots \leq x_n$ некоторого признака. Построим интервальный вариационный ряд, то есть образуем m групп. Определим для полученного ряда функцию максимального правдоподобия, зависящую от числа групп m и от способа их получения, по формуле:

$$L = L(x_1, x_2, \ldots, x_n; m) = \left(\frac{f_1}{n\lambda_1}\right)^{f_1} \cdot \left(\frac{f_2}{n\lambda_2}\right)^{f_2} \cdot \ldots \cdot \left(\frac{f_m}{n\lambda_m}\right)^{f_m}, \qquad (1)$$

где f_i – количество значений признака в i – й группе $\left(\sum_{i=1}^{m} f_i = n,\ f_i \geq 1\right)$,

λ_i – длина i – го группового интервала $\left(\lambda_i > 0,\ \sum_{i=1}^{m} \lambda_i = \Lambda = x_n - x_1\right)$.

Таким образом, если какое-либо значение признака x оказалось в i – й группе, то плотность вероятности в точке x полагается равной $\left(\frac{f_i}{n\lambda_i}\right)$. При определении функции (1) полагаем, что при построении ряда не должно быть пустых групп и групповых интервалов, равных нулю.

Рассмотрим возможность получения критерия для оптимального числа групп m, при котором функция правдоподобия примет максимальное значение. В случае равных интервалов, имеем:

$$L_1 = \left(\frac{f_1}{n\lambda}\right)^{f_1} \cdot \left(\frac{f_2}{n\lambda}\right)^{f_2} \cdot \ldots \cdot \left(\frac{f_m}{n\lambda}\right)^{f_m} = \frac{f_1^{f_1} \cdot f_2^{f_2} \cdot \ldots \cdot f_m^{f_m}}{n^n \lambda^n} =$$

$$= \frac{f_1^{f_1} \cdot f_2^{f_2} \cdot \ldots \cdot f_m^{f_m} \cdot m^n}{n^n \Lambda^n}, \quad \lambda = \frac{\Lambda}{m}.$$

Выражение $n^n \Lambda^n$ в знаменателе не зависит от числа групп m, число объектов n фиксировано и $f_1^{f_1} \cdot f_2^{f_2} \cdot \ldots \cdot f_m^{f_m} \cdot m^n \geq 1$. Тогда, избавляясь от знаменателя и извлекая корень n-й степени, получим, что функция правдоподобия примет максимальное значение, когда

$$m \sqrt[n]{f_1^{f_1} \cdot f_2^{f_2} \cdot \ldots \cdot f_m^{f_m}} \to \max \cdot \qquad (2)$$

Таким образом, в случае равных интервалов, произведение средней геометрической величины из групповых частот и числа групп должно принимать максимальное значение. В случае равных частот, имеем:

$$L_2 = \left(\frac{f}{n\lambda_1}\right)^f \cdot \left(\frac{f}{n\lambda_2}\right)^f \cdot \ldots \cdot \left(\frac{f}{n\lambda_m}\right)^f = \frac{f^n}{n^n(\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_m)^f} =$$

$$= \frac{1}{m^n(\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_m)^{\frac{n}{m}}} = \frac{1}{\left(m \sqrt[m]{\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_m}\right)^n}, \quad f = \frac{n}{m}.$$

Тогда получим, что функция правдоподобия примет максимальное значение, когда

$$m \sqrt[m]{\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_m} \to \min. \qquad (3)$$

Таким образом, в случае равных частот, произведение средней геометрической величины из групповых интервалов и числа групп должно принимать минимальное значение.

Формулы (2) и (3) позволяют сравнивать между собой результаты группировок с различным числом групп и выбирать из них группировку, которой соответствует большее значение функции правдоподобия. Данное утверждение относится ко всем распределениям за исключением равномерного, так как для него функция правдоподобия не зависит от числа групп. Действительно, в этом случае все частоты равны $f = n/m$ и все интервалы равны $\lambda = \Lambda/m$. Функция правдоподобия равна

$$L = \left(\frac{f_1}{n\lambda_1}\right)^{f_1} \cdot \left(\frac{f_2}{n\lambda_2}\right)^{f_2} \cdot \ldots \cdot \left(\frac{f_m}{n\lambda_m}\right)^{f_m} = \frac{f^{fm}}{n^{fm}(\Lambda/m)^{fm}} = \frac{(n/m)^{fm}}{n^{fm}(\Lambda/m)^{fm}} = \frac{1}{\Lambda^n},$$

следовательно, не зависит от числа групп.

Для оценки числа групп обычно применяют формулу Стерджесса

$$m = 1 + \log_2 n, \qquad (4)$$

которая дает хороший результат, когда распределение близко к нормальному закону. Ранее автором было предложено оценивать число групп по формуле

$$m = \frac{x_n - x_1}{6\sigma}(1 + \log_2 n), \qquad (5)$$

где x_1 и x_n – минимальное и максимальное значения признака,

σ – стандартное отклонение.

Для распределений близких к нормальному закону формулы (4) и (5) рекомендуют одинаковое число групп.

Возьмем данные (табл. 1), распределение которых отличается от нормального закона, оценим число групп по формулам (4) и (5), а затем определим оптимальное число групп с использованием функции правдоподобия. Для представленных данных x_n-x_1=2,5 $\sigma \approx 0{,}58$ Тогда по формуле Стерджесса число групп $m = 7$, а по формуле (5) $m = 5$. Затем было оценено выражение (2) при числе групп равном $m = 1,2,3,4,5,6,7$. Величина $m \sqrt[n]{f_1^{f_1} \cdot f_2^{f_2} \cdot \ldots \cdot f_m^{f_m}}$ оказалась максимальной при $m = 5$.

Таблица 1

Значения признака X

1,0	1,3	1,5	1,6	1,7	1,9	2,05	2,1
2,15	2,2	2,25	2,3	2,35	2,4	2,5	2,52
2,55	2,58	2,6	2,63	2,65	2,71	2,75	2,78
2,81	2,83	2,88	2,9	2,95	3,0	3,01	3,03
3,04	3,05	3,07	3,09	3,1	3,11	3,12	3,15
3,16	3,18	3,19	3,2	3,22	3,24	3,25	3,26
3,28	3,3	3,31	3,33	3,25	3,36	3,37	3,39
3,4	3,41	3,42	3,43	3,46	3,48	3,49	3,5

Можно рекомендовать следующий порядок действий для выбора оптимального числа групп. Сначала необходимо оценить число групп m_1 и m_2 по формулам (4) и (5). Пусть, например, $m_1 < m_2$. Затем по формулам (2) или (3) определить оптимальное число групп из отрезка $[m_1 - k, m_2 + k$, где k – небольшое целое число, например $k = 2$.

Владимир Александрович ДАВЫДЕНКО –
доктор социологических наук, заведующий кафедрой экономической социологии

Гульнара Фатыховна РОМАШКИНА –
доктор социологических наук, профессор кафедры экономической социологии

Сергей Николаевич ЧУКАНОВ –
доктор технических наук, профессор Института математики СО РАН, Омский филиал, заведующий лабораторией «Моделирование сложных систем»

УДК 519.6

*МОДЕЛИРОВАНИЕ СОЦИАЛЬНЫХ СЕТЕЙ**

АННОТАЦИЯ Рассмотрены общие свойствах различных видов сетей, описывающих процессы в сложных социальных системах; статистические свойства сетей; модели социальных сетей; поведение социальных систем с сетевой структурой.

The general properties of various kinds of networks, corresponding to processes in complex social systems, statistical properties of the networks, models of social networks, as well as behavior of social systems with a network structure are considered in this paper.

* Работа выполнена при поддержке программы «Университеты России» в рамках гранта УР.10.01.422 «Социолого-математическое моделирование социальных сетей».