

теплообменников // Применение ЭВМ в оптимальном планировании и проектировании. - Новосибирск, 1982. - С.81-90.

3. Schatzoff M., Tsao R., Feinberg S. *Efficient Calculation of All Possible Regressions*. *Technometrics*, 1968, - V. 10. - N 4. - p. 769-779.

4. Себер Дж. Регрессионный линейный анализ. - М.: Мир, 1980. - 456 с.

5. Численные методы условной оптимизации. - М.: Мир, 1977. - 290 с.

УДК 519.25

Б. Ю. Лемешко, А. Г. Панфилов,
Л. В. Панфилова
(Новосиб. электротехн. ин-т)

АСИМПТОТИЧЕСКИ ОПТИМАЛЬНОЕ ГРУППИРОВАНИЕ ДАННЫХ ПРИ АНАЛИЗЕ НАБЛЮДЕНИЙ, РАСПРЕДЕЛЕННЫХ ПО ЗАКОНАМ ЛАПЛАСА И ДВОЙНОГО ПОКАЗАТЕЛЬНОГО.

Оценки параметров распределений по группированным данным находят обычно методом максимального правдоподобия, методом минимума χ^2 или одним из родственных ему методов [1, 2]. При некоторых условиях [1, 3-6] эти оценки существуют, единственны и обладают свойствами состоятельности и асимптотической эффективности. При этом асимптотическая дисперсионная матрица оценки вектора параметров распределения θ определяется соотношением

$$D[\theta] = N^{-1} J_r^{-1}(\theta).$$

Здесь N - объем выборки; $J_r(\theta)$ - информационная матрица Фишера по группированным данным,

$$J_r(\theta) = \sum_{i=1}^k \frac{\nabla P_i(\theta) \nabla P_i^T(\theta)}{P_i(\theta)},$$

где $P_i(\theta)$ - вероятность попадания наблюдения в i -й интервал. Так как $J_r(\theta)$ - положительно определенная матрица, то с ростом потерь информации, вызванных группированием данных, ухудшается дисперсионная матрица оценок параметров.

С другой стороны, как показано в работе [7], мощность критерия тем выше, чем меньше потери информации от группирования.

Используя зависимость $J_r(\theta)$ от граничных точек интервалов и задаваясь определенным критерием, можно минимизировать потери информации от группирования, разбивая соответствующим образом область определения случайной величины на интервалы группирования.

В данной работе решается задача асимптотически оптимального группирования для распределений Лапласа и двойного показательного.

Функция плотности закона распределения Лапласа имеет вид

$$f(x) = \frac{\theta}{2} e^{-\theta|x|},$$

где $\theta > 0$, $x \in (-\infty, +\infty)$.

При решении задачи асимптотически оптимального группирования для параметра θ максимизировалась относительная асимптотическая информация

$$A = J_r(\theta) / J_H(\theta) = \sum_{i=1}^k A_i,$$

где

$$A_i = \begin{cases} \frac{1}{2} \frac{(t_i e^{t_i} - t_{i-1} e^{t_{i-1}})^2}{e^{t_i} - e^{t_{i-1}}}, & t_i < 0 \\ \frac{1}{2} \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_{i-1}} - e^{-t_i}}, & t_{i-1} > 0 \\ \frac{1}{4} \frac{(t_i e^{-t_i} - t_{i-1} e^{t_{i-1}})^2}{1 - \frac{1}{2}(e^{-t_i} + e^{t_{i-1}})}, & t_{i-1} < 0, t_i > 0 \end{cases}$$

и $t_i = \theta x_i$. Решение задачи для числа интервалов группирования от 2 до 10 приведено в табл. I. В случае четного k имеется два эквивалентных решения. Групповые пределы второго решения зеркально относительно нуля соответствуют значениям, приведенным в табл. I.

Функция плотности двойного показательного распределения

$$f(x) = \theta_1 \theta_2 e^{-\theta_1 x - \theta_2 e^{-\theta_1 x}},$$

где $\theta_1, \theta_2 > 0$, $x \in (-\infty, +\infty)$, определяется двумя параметрами.

Информационное количество Фишера по группированным данным о параметре θ_2

$$J_r(\theta_2) = \frac{1}{\theta_2^2} \sum_{i=1}^k \frac{(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})^2}{e^{-t_i} - e^{-t_{i-1}}},$$

где $t_i = \theta_1 x_i$,

а по негруппированным

$$J_H(\theta_2) = \frac{1}{\theta_2^2}$$

При решении задачи асимптотически оптимального группирования для параметра θ_2 максимизировалась относительная асимптотическая информация о параметре θ_2 :

$$A = J_r(\theta_2) / J_H(\theta_2).$$

Решение задачи приведено в табл.2.

Информационное количество Фишера о параметре θ_1 по группированным данным

$$J_r(\theta_1) = \frac{1}{\theta_1^2} \sum_{i=1}^k \frac{(t_i \ln(t_i/\theta_2) \cdot e^{-t_i} - t_{i-1} \ln(t_{i-1}/\theta_2) \cdot e^{-t_{i-1}})^2}{e^{-t_i} - e^{-t_{i-1}}},$$

где $t_i = \theta_2 e^{-\theta_1 x_i}$

а по негруппированным

$$J_H(\theta_1) = [\ln^2 \theta_2 + 2(C-1) \ln \theta_2 + C^2 - 2C + 1 + \frac{\pi^2}{6}] / \theta_1^2,$$

где $C = 0,5772$ - постоянная Эйлера.

Относительная асимптотическая информация о параметре $A = J_r(\theta_1) / J_H(\theta_1)$ зависит от θ_2 .

Решение задачи асимптотически оптимального группирования получено для различных значений θ_2 . В табл.3 приведены групповые пределы в виде $t_i = \theta_2 e^{-\theta_1 x_i}$ для значения θ_2 , равного 7.

Задача асимптотически оптимального группирования при одновременном оценивании двух параметров двойного показательного распределения связана с максимизацией определителя информационной матрицы Фишера по группированным данным или, точнее, с максимизацией отношения определителя по группированным данным к определителю информационной матрицы по негруппированным.

Информационная матрица по группированной выборке определяется выражением

$$J_r = \begin{bmatrix} J_r(\theta_1) & J_r(\theta_1, \theta_2) \\ J_r(\theta_1, \theta_2) & J_r(\theta_2) \end{bmatrix},$$

где

$$J_r(\theta_1, \theta_2) = \sum_{i=1}^k \frac{\partial \ln P_i}{\partial \theta_1} \cdot \frac{\partial \ln P_i}{\partial \theta_2} \cdot P_i.$$

Здесь P_i - вероятность попадания наблюдения в i -й интервал. Непосредственно .

Таблица I
Асимптотически оптимальные групповые пределы в виде $t_i = \hat{\Sigma} x_i$
для оценивания параметра θ распределения Лапласа

K	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	1.8416									0.2920
3	-1.5936	1.5936								0.6476
4	-2.7044	-1.1121	1.5183							0.7303
5	-2.6112	-1.0176	1.0176	2.6112						0.8203
6	-2.5670	-0.9745	0.8045	1.8233	3.4173					0.8548
7	-3.3652	-1.7716	-0.7544	0.7541	1.7716	3.3652				0.8910
8	-3.3329	-1.7388	-0.7266	0.6311	1.3816	2.3972	3.9879			0.9087
9	-3.9657	-2.3720	-1.3545	-0.6004	0.6004	1.3545	2.3720	3.9657		0.9269
10	-3.9468	-2.3487	-1.3335	-0.5817	0.5217	1.1246	1.8811	2.9054	4.5228	0.9371

Таблица 2

Асимптотически оптимальные групповые пределы для оценивания параметра θ_1 двойного показательного распределения в виде

$$t_i = \theta_1 e^{-\theta_1 x_i}$$

K	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	1.5936									0.6476
3	2.6112	1.0176								0.8203
4	3.3652	1.7716	0.7541							0.8910
5	3.9657	2.3720	1.3545	0.6004						0.9269
6	4.4650	2.8714	1.8538	1.0997	0.4993					0.9476
7	4.8925	3.2989	2.2813	1.5273	0.9269	0.4276				0.9606
8	5.2665	3.6729	2.6553	1.9012	1.3008	0.8015	3.3739			0.9693
9	5.5988	4.0052	2.9876	2.2336	1.6331	1.1338	0.7063	0.3323		0.9754
10	5.8979	4.3042	3.2866	2.5326	1.9322	1.4329	1.0053	0.6314	0.2990	0.9798

Таблица 3

Асимптотически оптимальные граничные точки в виде $t_i - e^{-\theta_1 x_i}$ и соответствующие значения асимптотической относительной информации при оценивании параметра θ_1 двойного показательного закона распределения ($\theta_2 = 7$)

K	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	A
2	0.0681									0.6774
3	0.1177	0.0279								0.8397
4	0.1532	0.0587	0.0143							0.9036
5	0.1797	0.0848	0.0344	0.0082						0.9353
6	0.2016	0.1077	0.0548	0.0226	0.0053					0.9533
7	0.2170	0.1257	0.0718	0.0369	0.0149	0.0034				0.9644
8	0.2303	0.1410	0.0869	0.0502	0.0260	0.0106	0.0023			0.9718
9	0.2437	0.1559	0.1027	0.0652	0.0885	0.0198	0.0079	0.0017		0.9770
10	0.2541	0.1687	0.1143	0.0765	0.0495	0.0295	0.0155	0.0061	0.0014	0.9807

Таблица 4

Асимптотически оптимальные групповые пределы в виде
 $t_i = \theta_2 e^{-\theta_1 x_i}$ для оценивания двух параметров двойного
 показательного распределения или при проверке гипотез
 о согласии по критерию χ^2 Пирсона

k	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
3	2.6067	0.2731						
4	3.4137	1.3979	0.2109					
5	3.8606	1.9590	0.5123	0.1044				
6	4.4096	2.5726	1.2269	0.3649	0.0772			
7	4.7959	2.9922	1.7192	0.6758	0.2318	0.0501		
8	5.2049	3.4285	2.2041	1.1904	0.4837	0.1740	0.0377	
9	5.5273	3.7667	2.5713	1.6027	0.7829	0.3431	0.1269	0.0275
10	5.8478	4.1024	2.9269	1.9932	1.1805	0.5770	0.2638	0.0988
11	6.1270	4.3930	3.2319	2.3192	1.5344	0.8560	0.4359	0.2046
12	6.3853	4.6589	3.5103	2.6163	1.8570	1.1789	0.6517	0.3434
13	6.6208	4.9016	3.7623	2.8810	2.1401	1.4807	0.9030	0.5106
14	6.8444	5.1314	3.9997	3.1286	2.4019	1.7608	1.1798	0.7116
15	7.0506	5.3425	4.2169	3.3538	2.6381	2.0116	1.4426	0.9387

k	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	A
3							0.4079
4							0.5272
5							0.6836
6							0.7571
7							0.8109
8							0.8490
9							0.8756
10	0.0213						0.8963
11	0.0771	0.0165					0.9123
12	0.1638	0.0618	0.0123				0.9248
13	0.2754	0.1326	0.0500	0.0106			0.9349
14	0.3126	0.2261	0.1094	0.0412	0.0087		0.9431
15	0.5734	0.3394	0.1881	0.0913	0.0344	0.0072	0.9498

$$\gamma_r(\theta_1, \theta_2) = \frac{1}{\theta_1 \theta_2} \sum_{i=1}^n \frac{(t_i \ln \frac{t_i}{\theta_2} e^{-t_i - t_{i-1}} \ln \frac{t_{i-1}}{\theta_2} e^{-t_{i-1}})(t_i e^{-t_i} - t_{i-1} e^{-t_{i-1}})}{e^{-t_i} - e^{-t_{i-1}}}.$$

Информационная матрица Фишера по негруппированным данным задается соотношением

$$J_H = \begin{bmatrix} J_H(\theta_1) & J_H(\theta_1, \theta_2) \\ J_H(\theta_1, \theta_2) & J_H(\theta_2) \end{bmatrix},$$

где

$$J_H(\theta_1, \theta_2) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(x)}{\partial \theta_1} \cdot \frac{\partial \ln f(x)}{\partial \theta_2} \right) f(x) dx.$$

В данном случае

$$J_H(\theta_1, \theta_2) = 1 - c - \ln \theta_2.$$

При решении задачи асимптотически оптимального группирования максимизировалась величина

$$A = \det J_r / \det J_H.$$

Причем

$$\det J_H = \frac{1.5449}{\theta_1^2 \theta_2^2}.$$

Результаты решения этой задачи приведены в табл.4.

Список литературы

1. Рао С.Р. Линейные статистические методы и их применения. - М.: Наука, 1968. - 548 с.
2. Кендал М., Стьюарт А. Статистические выводы и связи. - М.: Наука, 1973. - 900 с.
3. Кулльдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. - М.: Наука, 1966. - 176 с.
4. Бодин Н.А. Оценка параметров распределения по группированным наблюдениям. - Труды ордена Ленина Математического института им. В.А.Стеклова. - Л., 1970. - Т.III. - С.110-154.
5. Лемешко Б.Ю. Оценивание параметров распределений по группированным наблюдениям. - В кн.: Вопросы кибернетики. М.: 1977, вып. 30. - С.80-96.

6. Денисов В.И. Математическое обеспечение системы ЭВМ - экспериментатор. - М.: Наука, 1977. - 252 с.

7. Денисов В.И., Лемешко Б.Ю. Оптимальное группирование при обработке экспериментальных данных // Измерительные информационные системы / Новосиб. электротехн. ин-т. - Новосибирск, 1979. - С. 5-14.

УДК 519.24

Шляпников Д.А.
(Новосиб. электротехн. ин-т)

О НЕКОТОРЫХ СИТУАЦИЯХ, ПОРОЖДАЮЩИХ ПРЕОБРАЗОВАНИЯ

Опишем основные ситуации, возникающие в задачах идентификации при обработке экспериментальных данных, и некоторые возможные пути использования преобразований для реализации этих ситуаций. Будем рассматривать конкретно регрессионные модели в случае объемов наблюдений, достаточных для статистических выводов. Кроме того, ограничимся тем фактом, что исследователя не интересует структура модели и степень влияния отдельных факторов на отклик. Цель обработки экспериментальных данных - получение значений функции отклика при различных условиях. Аналогичные же цели можно поставить в случае дисперсионной модели.

Рассмотрим процедуру преобразования в случае, когда наблюдения представляют собой однородную выборку из некоторой генеральной совокупности с неизвестным математическим ожиданием и дисперсией. При нарушении основных предпосылок, связанных с нормальностью закона распределения и однородностью дисперсии, получение несмещенных оценок с минимальной дисперсией весьма затруднительно. Улучшить свойства выборки можно, подобрав подходящее нормализующее преобразование, используя при этом либо алгоритм Бокса-Кокса [1,5], основанный на методе максимального правдоподобия, либо алгоритм Кэскея [2], предполагающего, что исходная выборка имеет закон распределения пирсоновского типа. Преобразованные выборочные данные распределены квазинормально и допускают построение оценок с наилучшими свойствами. Полученные оценки для математического ожидания и дисперсии нормализованной выборки составляют достаточную систему статистик и позволяют найти единственные с минимальной дисперсией несмещенные оценки для математического ожидания и дисперсии исходных данных.