응용통계연구 제 6권 2호
1993년 9월 pp.247-252.

On L_1 Regression Coefficients[1)]

C. S. Hong, H. J. Choi[2)]

Abstract

Consider minimizing the sum of absolute deviations for multiple regression models. If a regression line is assumed to pass a given point, then we can find that the L_1 regression coefficient can be defined in terms of the weighted medians of the slopes from each data point to the given point. Therefore L_1 method could be regarded to find the optimal point which regression line passes over.

1. Introduction

Let's consider the general regression model. The vector tf sample observations $\underset{\sim}{Y}$ is expressed as a linear combination of k explanatory vectors $\underset{\sim}{X}_j$ plus an disturbance vector $\underset{\sim}{u}$ possessing some distributions :

$$\begin{aligned}\underset{\sim}{Y} &= \underset{\sim}{1}\alpha + \beta_1 \underset{\sim}{X}_1 + \beta_2 \underset{\sim}{X}_2 + \cdots + \beta_k \underset{\sim}{X}_k + \underset{\sim}{u} \\ &= \underset{\sim}{1}\alpha + X\underset{\sim}{\beta} + \underset{\sim}{u}\ ,\end{aligned}$$

where each vector possesses n elements, $\underset{\sim}{1}$ is a unit vector, $(\underset{\sim}{1}, X) = (\underset{\sim}{1}, \underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_k)$ is the $(n\times(k+1))$ matrix and $(\alpha, \underset{\sim}{\beta}') = (\alpha, \beta_1, \beta_2, \cdots, \beta_k)$ is a vector of unknown parameters.

The method of minimizing the sum of absolute deviations from a predicted regression model, which is called a L_1 regression model, has been studied for a long period of time. The L_1 estimator $(a, \underset{\sim}{b}') = (a, b_1, b_2, \cdots, b_k)$ of $(\alpha, \underset{\sim}{\beta}')$ is a solution to the problem

$$\min S(a, \underset{\sim}{b}) = \sum_{i=1}^{n} | Y_i - a - \sum_{j=1}^{k} X_{ij} b_j | . \tag{1}$$

Kennedy and Gentle(1980) reported that there are two types of algorithms to compute L_1 estimators. The first type is concerned as Simplex method. Wagner(1957), Davies(1967), Appa and Smith(1973), Barrodale and Roberts(1973) and Gentle, Kennedy and Sposito(1977a, 1977b) etc. considered it as linear programming problems (See Sposito(1975) for more details and a general discussion). The other type of algorithm uses iterative computing methods. Schlossmacher(1973) gave an iterative method for L_1 estimator using iterative reweighted least square. He obtained L_1 estimators using least square method with weights

---

1) This paper was supported by Non Directed Research Fund, Korea Research Foundation, 1991.

2) C. S. Hong is Associate Professor and H. J. Choi is Graduate student, Department of Statisics, Sung Kyun Kwan University, Seoul, 110-745, Korea. Authors thank to the Refrees for many useful comments.

of absolute residuals, and mentioned that his method performed better than linear programming. However, Armstrong and Frome(1976) discussed that this method is rather inefficient. Bloomfield and Steiger(1980) proposed another efficient iterative method which searches for a set of data points to make a corresponding deviation vanish.

In this paper, we suggest an alternative method for L_1 estimators. Let $(\underline{X_0}', Y_0) \equiv (X_{01}, X_{02}, \cdots, X_{0k}, Y_0)$ be a given point (optimal point) which the L_1 regression line passes over. Then a L_1 problem in (1) might be replaced with

$$\min_{\underline{b}} \quad S(\underline{b} \mid (\underline{X_0}', Y_0)) = \min_{\underline{b}} \sum_{i=1}^{n} |(Y_i - Y_0) - \sum_{j=1}^{k} b_j (X_{ij} - X_{0j})| . \qquad (2)$$

When the point (X_0, Y_0) is given in a simple L_1 regression model, Karst(1958) developed algorithms to compute L_1 estimator (L_1 regression coefficient) using order statistics of the slopes $\{(Y_i - Y_0)/(X_i - X_0)\}$. Hereby we will extend Karst's algorithm and explain how to obtain the L_1 estimators alternatively. Moreover, for multiple regression models, we will show that L_1 regression coefficients are the convergent weighted medians of the slopes. In this paper, it will be discussed an alternative method to find L_1 estimators for general regression models.

2. L_1 Regression coefficients

For a given point $(\underline{X_0}', Y_0) \in R^{k+1}$, L_1 regression coefficients $\underline{b}$ so as to minimize $S(\underline{b} \mid (\underline{X_0}', Y_0))$ in (2) could be obtained by certain iteration method. With initial values of $\{b_h ; h \neq j\}$, the estimate of the coefficient b_j is the weighted median of the slopes

$$\{((Y_i - Y_0) - \sum_{h \neq j} b_h (X_{ih} - X_{0h}))/(X_{ij} - X_{0j}) ; j = 1, 2, \cdots, k\} .$$

Since the coefficients $\{b_j ; j = 1, 2, \cdots, k\}$ must be estimated simultaneously, we could use an iteration method in order to obtain these with certain initial values. And these estimates are convergent since the function of $S(\underline{b} \mid (\underline{X_0}', Y_0))$ in (2) is well known to be piecewise convex. Therefore the estimates of L_1 regression coefficient might be the values that the weighted medians converge. And we could define such medians as the convergent weighted medians.

Theorem. For a given point $(\underline{X_0}', Y_0)$ which a regression line passes on, the estimates of the L_1 regression coefficients $\{b_j ; j = 1, 2, \cdots, k\}$ are the convergent weighted medians of the slopes

$$\{((Y_i - Y_0) - \sum_{h \neq j} b_h (X_{ih} - X_{0h}))/(X_{ij} - X_{0j}) ; j = 1, 2, \cdots, k\} .$$

with some weights $\{|X_{ij} - X_{0j}| ; j = 1, 2, \cdots, k\}$.

Also we can make a note about the well-known simple regression models as the following:

Corollary. L_1 estimate of the simple regression coefficient, b, is the weighted median of the slopes, $\{(Y_i - Y_0)/(X_i - X_0)\}$, with weights $\{|X_i - X_0|\}$.

If we knew the location of one point which L_1 regression line passed on, the estimators could be defined and obtained by the above theorem. Hence L_1 method to estimate regression parameters could be regarded as to find an optimal point on $(k+1)$th euclidian space:

$$\min_{(\underline{X}',Y)\in R^{k+1}} S(\underline{b} \mid (\underline{X}',Y)) . \qquad (3)$$

The optimal points might be said all points on the regression line. Since the response variable Y is the functions of k-independent variables, we can reduce to the kth dimensionality with a given value which is one of the response variable Y, for example the first observation Y_1 or the response variable mean, $\overline{Y}$. Then this method (3) would be substitute for finding an optimal point of the kth dimensional space rather than $(k+1)$th dimensional space as the following :

$$\min_{\underline{X}\in R^k} S(\underline{b} \mid (\underline{X}',Y_1)) .$$

The above method could be programmed with a golden section search technique(See Kennedy and Gentle(1980)) to obtain the minimum value of $S(\underline{b} \mid (\underline{X}',Y_1))$.

In order to demonstrate the alternative method, we extracted a multiple regression data from Draper and Smith (1981, pp. 629-630). Here we list below the data in (Table 1).

(Table 1) data list

Y	X_1	X_2	X_3	X_4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
95.9	7	52	6	33
109.2	11	55	9	22
102.7	3	71	17	6
72.5	1	31	22	44
93.1	2	54	18	22
115.9	21	47	4	26
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12

First of all, we may set the first observation Y_1 = 78.5 as a given value. Now, by using some search method, we can find an optimal point on 4th dimensional space rather than 5th dimension such as (X_{01}=7.0, X_{02}=26.0, X_{03}=6.0, X_{04}=60.0), which fortunately coincides the first observation of X's. With this optimal point, we could find the estimates of L_1 regression coefficients through the iteration steps mentioned above. (Table 2) shows the convergent step.

(Table 2) Convergent Step

b_1	b_2	b_3	b_4
0.000004	0.735714	-0.499997	-0.118214
0.898431	0.681863	-0.094999	-0.048510
1.370500	0.677267	0.207422	0.007738
1.703953	0.689240	0.318420	0.019268
1.729851	0.691543	0.323840	0.021485
1.732556	0.691783	0.324406	0.021710
1.732839	0.691808	0.324465	0.021741
1.732860	0.691811	0.234471	0.021743
1.732871	0.691811	0.324472	0.021743
1.732869	0.691811	0.324470	0.021743

Finally, we obtained the fitted L_1 regression line as

$$\hat{Y} = 45.1314 + 1.7329\ X_1 + 0.6918\ X_2 + 0.3245\ X_3 + 0.0217\ X_4\ .$$

3. Conclusion

The main ideas of this paper are followed by quite simple facts. Suppose the simple L_1 regression line intersects a point (X_0, Y_0). Then regression coefficient is found to be the weighted median of slopes from data points to a certain optimal point over which the regression line passes. The weights could be defined as the distance of X-axis from each explanatory value to a given optimal point, $|X_i - X_0|$. And for the given first value of the response variable, choose an optimal point with respect to the independent variable such that the sum of absolute residuals should be minimized by iterative methods.

Moreover, we extend the above method to multiple regression models. In these cases, each regression coefficient cannot be mentioned as the median of some slopes when a multiple regression line has intersected a certain given point. Another fact we find for general regression models is that L_1 regression coefficients could be defined in terms of the convergent weighted medians of some slopes. With these situations, we did the analogous argument to find an optimal point on the kth dimensional space so that the sum of absolute residuals would be minimized by iterative computations. But it is hard to think that this alternative method is better than other methods (for example, Simplex method, Schlossmacher's method or Bloomfield and Steiger's method) to find L_1 estimators, because it does matter of computing work with some search method.

References

[1] Amstrong, R. D. and E. L. Frome (1976), "A Comparison of Two Algorithms for Absolute Deviation Curve Fitting.", *Journal of the American Statistical Association*, **71**, 328-330.

[2] Appa, G. and Smith, C. (1973), "On L_1 and Chebyshev Estimation.", *Journal of Mathematical Programming*, **8**, 73-87.

[3] Barrodale, I. and Roberts, F. D. K. (1973), "An improved Algorithm for Discrete L_1 Linear Approximation.", *SIAM, Journal of Numerical Analysis*, Vol. **10**, No.5, 839-848.

[4] Bloomfield, G. and Steiger, W. (1980), "Least Absolute Deviations Curve-Fitting.", *SIAM, Journal of Science Statistical Computation*, Vol. 1, No. 2, 290-301.

[5] Davies, M. (1967), "Linear Approximations Using the Criterion of Least Total Deviations", *Journal of the Royal Statistical Society*, (**B**), 101-109.

[6] Gentle, J. E., Kennedy, W. J. and Sposito, V. A. (1977a), "On Least Absolute Deviations Estimation.", *Communications in Statistics*, (**A**) **6**, 839-845.

[7] Gentle J. E., Kennedy, W. J. and Sposito, V. A. (1977b), "On Properties of L_1 Estimators.", *Mathematical Programming*, **12**, 138-140.

[8] Karst, O. (1958), "Linear Curve Fitting Using Least Deviations.", *Journal of the American Statistical Association*, **53**, 118-132.

[9] Kennedy, W. J. and Gentle, J. E. (1980), *Statistical Computing*, Marcel Dekker Inc.

[10] Schlossmacher, E. J. (1973), "An Iterative Technique for Absolute Deviations Curve Fitting.", *Journal of the American Statistical Association*, **68**, 857-859.

[11] Sposito, V. A. (1975), *Linear and Non-Linear Programming*, The Iowa State University Press, Ames Iowa.

[12] Wagner, H. M. (1957), "Linear Porgramming Techniques for Regression Analysis., *Journal of the American Statistical Association*, **54**, 206-212.

L_1 회귀 계수에 관한 연구[1)]

홍종선 , 최현집[2)]

요 약

회귀계수 추정을 위해 잔차의 절대값의 합을 최소화하는 과정에서 회귀선이 어떤 점을 지난다고 가정하였을 경우에는 L_1 회귀계수는 표본 관측점과 주어진 점과의 기울기들의 가중 중앙값으로 정의할 수 있음을 보였다. 그러므로 L_1 방법은 회귀선이 통과하는 하나의 최적점을 발견하는 것으로 간주될 수 있다.

1) 이 논문은 1991년 문교부 한국학술진흥재단의 자유공모과제 학술연구 조성비에 의해 연구되었음.
2) (110-745) 서울특별시 종로구 명륜동 3-53 성균관대학교 통계학과.