行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※
※ ※
※ **分位估計量**(2/2) ※
※ ※
※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：□個別型計畫　□整合型計畫

計畫編號：NSC 89-2118-M-009-020-

執行期間：89 年 8 月 1 日至 90 年 7 月 31 日

計畫主持人：陳鄰安教授

共同主持人：

本成果報告包括以下應繳交之附件：

□赴國外出差或研習心得報告一份

□赴大陸地區出差或研習心得報告一份

□出席國際學術會議心得報告及發表之論文各一份

□國際合作研究計畫國外研究報告書一份

執行單位：國立交通大學統計研究所

中 華 民 國 90 年 10 月 25 日

行政院國家科學委員會專題研究計畫成果報告
分位估計量(2/2)

計畫編號：NSC 89-2118-M-009-020-
執行期限：89 年 8 月 1 日至 90 年 7 月 31 日
主持人：陳鄰安　交大統計所
共同主持人：
計畫參與人員：

一、中文摘要

我們利用二維分配函數定義了兩種二維迴歸分位向量。此迴歸分位向量及截斷平均數的大樣本分配均予推導出來。再者，我們也進行了資料分析。

關鍵詞：分位向量、迴歸、截斷平均數

Abstract

We introduce two different types of bivariate regression quantile which are defined through the bivariate distribution function. The large sample representations of sample type bivariate regression quantiles are both developed. For parameter estimation, affine equivariant bivariate trimmed mean and bivariate trimmed covariance matrix have also been introduced where large sample property of the bivariate trimmed mean has been investigated. Results of simulation studies for the bivariate quantiles and trimmed mean have been presented. On the other hand, results of data analysis and regression quantiles of higher dimensions are also displayed.

KEYWORDS: Quantile; Regression; Trimmed Mean.

二、Introduction

Order statistics or quantiles are the basis for a variety of useful exploratory and robust procedures for univariate data. It is desirable to extend these procedures to multivariate data, but the lack of a natural ordering for multivariate data (Kendall, 1966, Bell and Haller, 1969) has hindered the definition of quantiles and hence the definition of procedures based on them in multivariate problems.

Much of the work in generalizing quantiles to multivariate models has concentrated on the particular case of the median of the multivariate location model. Weber (1909) defined the multivariate l_1 median by minimizing the multivariate version of the absolute residuals. More recently, Oja (1983) defined the multivariate simplex median by minimizing the sum of volumes of simplices with vertices on the observations, and Liu (1988 and 1990) introduced the simplicial depth median maximizing an empirical simplicial depth function. In these approaches, the median by Weber is not affine equivariant but those by Oja and Liu are. An excellent review of the work on the multivariate median is given by Small (1990). General multivariate quantiles (which of course include the multivariate median and extremes as special cases) are more difficult to define. The approach of taking a minimization problem whose solution is the univariate quantile has been generalized, called the M-quantile, by Breckling and Chambers (1988) and Koltchinski (1997) to the multivariate model. Unfortunately these two generalizations are all not affine equivariant. The approach of Einmahl and Mason (1992) considered the properties of the volume of the smalles Borel set that contains at least fraction of the multivariate data points. This variable, the volume, is no longer a multivariate quantity and then is not natural as a quantile point. As noted by Chaudhuri (1996), most authors, as the above ones, try to introduce descriptive statistics that generalize the concept of univariate quantiles to the multivariate setup without discussing what they are trying to estimate. That is, almost no attention is paid to the underlying population quantile. As the fact that the univariate quantile is an inverse function of a distribution function, recently, Chen and Welsh (1999) introduced a multivariate quantile through the inveses of joint and marginal distribution functions of variables or their transforms for the multivariate location model. introduced a bivariate quantile for bivariate location. With using inverses of joint distributions, this quantile then satisfies some properties of space partition with specified probabilities on sets.

Relatively to the location model, the study for quantile in the multivariate regression model seems receive even lesser attention. As an exemption, Koenker and Portnoy (1987) introduced the multivariate M-estimation for this model. This study by Koenker and Portnoy has the advantage of introducing a generalized robust type estimator for multivariate regression, however, the estimators proposed in their paper involved a weighted matrix that is assumed to be known and then make the estimators not affine equivariant. The interest in this paper is to introduce the multivariate quantile to the multivariate regression model through the conditional joint and marginal distributions of the dependent variables and their transforms. For simplicity, we first specify the population bivariate regression quantile in terms of the underlying error cumulative distribution and then construct estimators of these population quantiles through the regression quantile by Koenker and Bassett (1978). We show that this bivariate regression quantile satisfies a property of maximizing variance in terms of all directions of linear transformations. For estimation, affine equivariant sample bivariate regression quantiles are introduced and studied. Some properties as being a quantile of data portioning for the sample bivariate regression quantile have been stated. The setting of bivariate regression quantile leads naturally to the development of affine equivariant bivariate regression trimmed mean and covariance matrix. Results of simulation for quantile and trimmed mean and also a data analysis have all been provided.

We define two different types of bivariate regression quantile points in Section 2. We present sample estimators of these bivariate regression quantile points and establish their large sample properties in Sections 3 and 4. We introduce a bivariate trimmed mean in Section 5. We apply the bivariate regression quantiles and trimmed mean in Section 5 and briefly discuss extensions to higher dimensions in Section 6.

三、 References

[1]. Bai, Z.-D. and He, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. Annals of Mathematical Statistics. \bf 27, 1616-1637.

[2]. Bell, C. B. and Haller, H. S. (1969). Bivariate symmetry tests: parametric and nonparametric. Annals of Mathematical Statistics. 40, 259-269.

[3]. Breckling, J. and Chambers, R. (1988). M-quantiles. Biometrika. 75, 761-771.

[4]. Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. Journal of the American Statistical Association. 91, 862-872.

[5]. Chen, L. A. and Chiang, Y. C. (1996), Symmetric type quantile and trimmed means for location and linear regression model. Journal of Nonparametric Statistics. 7, 171-185.

[6]. Chen, L-A and Welsh, A. H. (2001), Bivariate quantiles. Journal of Multivariate Analysis. To appaer. \smallskip \ref Chen,

[7] .L-A, Welsh, A. H. and Chan, W. (1999) Linear winsorized means for the linear regression model. Statistica Sinica. 11, 147-172.

[8] .Einhmahl, J. H. J. and Mason, D. (1992). Generalized quantile process. The Annals of Statistics. 20, 1062-1078.

[9] .Gnanadesikan, R. (1997). Methods for statistical data analysis of multivariate observations. Wiley, New York.

[10] .Jureckova, J. (1984). Regression quantiles and trimmed least squares estimator under general design. Kybernetika. 20, 345-357.

[11] .Kendall, M. G. (1966). Discrimination

and classification. Krishnaiah I. 165-184.
[12] .Koenker, R. and Bassett, G. J. (1978). Regression quantiles. Econometrica. 46, 33-50.
[13] .Koenker, R. and Portnoy, S. (1990). M estimation of multivariate regression. Journal of the American Statistical Association. 85, 1060-1068. [14] .Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. The Annals of Statistics. 25, 435-477.
[15] .Liu, R. Y. (1988). On a notion of simplicial depth. Proceeding of National academy Science, USA, 18, 1732-1734.
[16] .Liu, R. Y. (1990). On a notion of data depth based on random simplices. The Annals of Statistics. 18, 405-414.
[17] .Maddala, G. S. (1988), Introduction to Econometrics, New York: Macmillan Publishing Company.
[18] .Oja, H. (1983). Descriptive statistics for multivariate distributions. Statistics and Probability Letters. 1, 327-332.
[19] .Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. Journal of the American Statistical Association 75, 828-838.
[20] .Small, C. G. (1990). A survey of multidimensional medians. International Statistical Review, 58, 273-277.
[21] .Weber, A. (1909). Uber den standort der industrien, Tubingen, Alfred Weber's Theory of Location of Industries, University of Chicago Press.
[22] .Welsh, A. H. (1987). The trimmed maen in the linear model. Annals of Statistics. 15, 20-36.