行政院國家科學委員會專題研究計畫成果報告

多維平行體及其應用

計畫編號：NSC90-2118-M009-008
執行期限：90 年 8 月 1 日至 91 年 7 月 31 日

主持人：陳鄰安　　交通大學統計所

一、中文摘要

我們提出一個多維平行體的觀念。此一多維平行體宛如單維 quantile 可用於建造多維度的截斷平均數。我們也計算分析了此一多維度截斷平均數的近似變異數並且做了蒙地卡羅模擬。

關鍵詞：多維平行體；分位向量；截斷平均數

Abstract

We propose a multivariate parallelogram that can play the role of the univariate quantile in location model. We then use it to introduce a multivariate trimmed mean. The asymptotic efficiency of the multivariate trimmed mean is studied by its asymptotic variance and Monte Carlo simulation.

Keywords: Multivariate parallelogram, Quantile, Trimmed mean.

1. Introduction

Let y1,...,yn be a random sample from a univariate population with a distribution function and an empirical distribution function obtained from this sample. The quantile interval with two quantile as ends plays a very important role in statistical inference. Basically this interval acts in two aspects. First, as a region with a particular coverage probability, the interval is a natural estimator for the scale parameters like range and interquartile range. With this property, it can be used to define a process capability index for process capability assessment, especially for non-normal processes. Second, this interval is routinely used in classifying the observations of a sample into good or bad observations in robust estimation such as the trimmed mean and Winsorized mean.

There are some attempts of proposing analogues for quantiles or order statistics in high dimensions. The approach of taking a minimization problem whose solution is the univariate quantile, generalizing the minimization problem to the multivariate case and then defining multivariate quantiles to be solutions of this minimization problem has been taken by Breckling and Chambers (1988) and Koltchinski (1997). Chaudhuri (1996) considered a geometric quantile that uses the geometry of multivariate data clouds. Chakraborty (2001) used a transformation retransformation technique to introduce a multivariate quantile. But none is satisfactory in defining multivariate regions for constructing descriptive statistics, due to the lacking of a natural ordering in multidimensional data. Unlike the above approaches, Chen and Welsh (2002) proposed a bivariate quantile that satisfy an analogous property to that of the univariate quantiles in that they partition R^{2} into sets with a specified probability content. This method determines the sets sequentially such that the second set is determined with the first set that makes the distribution of the second quantile element involves it of the first element. When we extend this to multiple dimensional case, the large sample theory will be too complicated to study and then not easy for application, based on quantile's asymptotic distribution, such as statistical inference. An attempt in the literature has been done for multivariate median estimation. For example, 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. An excellent review of this work is given by Small (1990).

Chaudhuri (1996) criticized that in the literature there is little efforts in developing multivariate descriptive statistics that are relevant about some population parameters. He further explained that most authors only introduced certain descriptive statistics that merely generalize the concept of univariate statistics to multivariate setup and there are no clear population analogues for these descriptive statistics. Although the approach by Chen and Welsh (2002) does serve an estimator for estimating a multivariate population parameter, however, alternatives easier in theoretical study and practical applications are worth to develop. The major purpose of this paper is to define a multivariate parallelogram region as a counterpart of the univariate quantile interval and propose a statistic to estimate it. Not only proving an estimation of population parameter, this sample multivariate parallelogram does easy in establishing many multivariate descriptive statistics, for example, multivariate versions of scale estimators, process capability index and trimmed mean. However, we will study the multivariate trimmed mean in this paper only.

With a similar idea that Huber (1973, 1981) used in constructing a location-scale equivariant studentized M-estimator for location, the multivariate quantile points are introduced that naturally are used to construct the multivariate parallelogram. Large-sample properties of the multivariate quantile points and trimmed means constructing by this parallelogram are studied. Asymptotic generalized variances of the multivariate studentized trimmed mean and Cramer-Rao lower bounds under various multivariate contaminated normal distributions are computed. The study reveals that the studentized trimmed mean is quite efficient.

Reference

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

[2] Barnett, V. (1976), The ordering of multivariate data. Journal of Royal Statistical Society, A. 139, 318-344.

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

[4] Cacoullos, T. and DeCicco, H. (1967). On the distribution of the bivariate range. Technometrics. 476-480.

[5] Chakraborty, B. (2001), On affine equivariant multivariate quantiles. Annals of the Institute of Statistical Mathematics. 53, 380-403.

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

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

[8] Chen, L.-A. and Welsh, A. H. (2002). Distribution-function-based bivariate quantiles. Journal of Multivariate Analysis, to appear.

[9] Chen, L-A, Welsh, A. H. and Chan, W. (2001). Estimators for the linear regression model based on Winsorized observations. Statistica Sinica. 11, 147-172.

[10] Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. New York: Academic Press.

[11] Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multi-response data. Biometrics. 28, 81-124.

[12] Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. The Annals of Statistics. 1, 799-821.

[13] Huber, P. J. (1981). Robust statistics. New York: Wiley.

[14] Jureckova, J. (1977). Asymptotic relation

of M-Estimates and R-Estimates in linear regression model. The Annals of Mathematical Statistics. 5, 464-472.

[15] Jureckova, J. and Sen, P. K. (1996). Robust statistical procedures. Wiley, New York.

[16] Liu, R. Y. (1988). On a notion of simplicial depth. Proceeding of National academy Science, USA. 18, 1732-1734.

[17] Liu, R. Y. (1990). On a notion of data depth based on random simplices. The Annals of Statistics. 18, 405-414.

[18] Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. The Annals of Statistics. 4, 51-67.

[19] Oja, H. (1983). Descriptive statistics for multivariate distributions. Statistics and Probability Letters. 1, 327-332.

[20] Ruppert, D. and Carroll, R. J. (1980), Trimmed least squares estimation in the linear model. Journal of American Statistical Association. 75, 828-838.

[21] Small, C. G. (1990). A survey of multidimensional medians. International Statistical Review} 58, 273-277.

[22] Taam, W. Subbaiah, P. and Liddy, J. W. (1993), A note on multivariate capability indices. Journal of Applied Statistics. 20, 339-351.

附件：封面格式

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※

※　　　　　　　　　　　　　　　　　　　　　　　　※

※　　　　　　　　多維平行體及其應用　　　　　　　　※

※　　　　　　　　　　　　　　　　　　　　　　　　※

※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：□個別型計畫　　□整合型計畫

計畫編號：NSC 90-2118-M-009-008

執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：陳鄰安

共同主持人：

計畫參與人員：沈欣怡、林旻靜、林玫芩

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：國立交通大學統計所

中　華　民　國　91 年 10　月 30　日