行政院國家科學委員會專題研究計畫成果報告

HELLINGER 距離下統計模型的幾何結構與估計法

一、中文摘要

在傳統上的大樣本理論，概似比的卡方漸近是一非常重要的性質，所以對於概似比的分佈也變成統計上的一個主要問題。在過去對此問題的理論結果只可適用於一般規則的模型，在這篇文章中我們發現若用貝氏的方法來看此問題，則我們可以把傳統的結果推廣到一般上我們所認為的不規則的模型上。這對於統計上許多不規則模型及有斷點的模型有非常重要的貢獻。相信這一個結果對於把參數空間賦予一個幾何距離，然後再來研究大　樣本時估計的收斂性與幾何距離的收斂性會有很大的影響與貢獻。

關鍵詞：大樣本理論、概似比、規則的模型、有斷點的模型、幾何距離、估計的收斂性

Abstract

The results of Wilks (1938) had three drawbacks. First of all, the likelihood function has to be smoothed enough in order to admit a Taylor's expansion. Secondly, the MLE has to be asymptotically normal and this itself relies on Taylor's expansions and the central limit theorem. Thirdly, assumptions on the independence of observations are typically made. We contend that much simpler fundamental insight to the Wilks theorem is available: if the contour set of the likelihood function around the MLE are of fan shape, then the Wilks type of results hold. The classical Wilks' theorem corresponds to the situations where the contour set is ellipsoid. In general, the asymptotic normality of the MLE is not required, neither does the asymptotic distribution of the MLE have to exist. We constructed an example where the MLE is not asymptotically normal, but Wilks' type of results hold.

We obtain more general results by using Bayesian approach and considering the likelihood contour sets being fan-shaped. This provides an insightful geometric understanding and a useful extension of the likelihood ratio theory. As a result, even if the MLEs are not asymptotically normal, the likelihood ratio statistics can still be asymptotically Chi-squared-distributed. In this sense, the traditional wilks theorem is a special case of our generalized results. Further, we demonstrate that the limiting distributions of the log-likelihood ratio statistics are, in general, gamma distributions. We believe that this important result will be widely used by statisticians.

Keywords: likelihood function, Taylor's expansion, MLE, asymptotically normal, Bayesian approach, Chi-squared-distributed, gamma distributions.

二、緣由與目的

One of the most celebrated folk theorems in statistics is that twice the logarithm of a maximum likelihood ratio statistic is asymptotically chi-distributed. This result is due to Wilks (1938) and is proved via Taylor's expansions of likelihood functions and by assuming that the maximum likelihood estimator (MLE) is asymptotically normal. See also Wald (1941), Wilks (1962) and heuristics given in the popular textbooks such as Cox and Hinkley (1974), Kendall and

Stuart (1979), among others. While this understanding is insightful, it has three drawbacks. First of all, the likelihood function has to be smoothed enough in order to admit a Taylor's expansion. Secondly, the MLE has to be asymptotically normal and this itself relies on Taylor's expansions and the central limit theorem. Thirdly, assumptions on the independence of observations are typically made. Technical proofs of the first two steps above are by no mean simple. This is probably why rigorous statements and heuristic proofs are suppressed in many popular graduate textbooks. See for example page 229 of Bickel and Doksum (1977), page 486 of Lehmann (1986) and page 381 of Casella and Berger (1990).

We contend that much simpler fundamental insight to the Wilks theorem is available: if the contour set of the likelihood function around the MLE are of fan shape, then the Wilks type of results hold. The classical Wilks' theorem corresponds to the situations where the contour set is ellipsoid. In general, the asymptotic normality of the MLE is not required, neither does the asymptotic distribution of the MLE have to exist. One can easily construct an example where the MLE is not asymptotically normal, but Wilks' type of results hold. An additional benefit is that our technical proof is simple and can be understood without much probability background.

三、結果與討論

In the project we derived the asympototic posterior distribution of the log-likelihood ratio statistic which is of gamma type distribution. By using the asymptotic posterior distribution of the log-likelihood ratio statistic, we obtained the asymptotic frequentist distribution of log-likelihood ratio statistic. Therefore, all the results we obtained up to now are based on large sample size. We apply the same method to the cases of small sample size. That is, we use the posterior distribution of log-likelihood ratio statistic to estimate the frequeist distribution of the log-likelihood ratio statistic.

In the derivation of the asymptotic distribution of log-likelihood ratio statistic, we only require the prior to be a nonnegative function, and all the nonnegative priors will result the same asymptotic distribution of log-likelihood ratio statistic. However, in the finite sample cases, the choice prior distribution becomes very important, the posterior distribution of log-likelihood ratio statistic are effected by the choice of the prior distribution. From Welch (1963), we know that, in regular models, if we use the square root of Fish information as prior, then the confidence interval obtained from the posterior distribution has the right converage probability up to the second order. Therefore, we adapt the prior proposed by Hung and Wang (1996) which is an extension of the square root of Fisher information to the non-regular models.

In the finite sample problem, first, we try to see how the gamma distributions approximate the true distribution. Secondly, from we see that, even in the irregular models, the asymptotic distribution of log-likelihood ratio statistic is of gamma type distributions with scale parameter equal to 1. Therefore, we try to approximate the posterior distribution by a gamma type distribution with scale parameter equal to 1. For finding this approximation, when parameter is of one-dimension, we observe the behavior of estimator of parameter around some neighborhood of the true parameter. The

condition about the likelihood contour set being of fan-shaped can be reduced to that the value of error is proportional to $(2w)^r$. Therefore, for fixed estimate of parameter, the slope of the curve of log 2w agains t log error is r. The detail procedure is in , my paper. In the first example, the maximum likelihood estimator is the smallest order statistic and its asymptotic distribution is an exponential distribution which is not a normal distribution. For six different sample sizes, we use the method mentioned above to estimate the parameter r. We observed that the slope of log 2W against log error is always equal 1 for any sample sizes. Then draw three curves on the same plot. The nonsmooth one is the empirical distribution of log 2W. The solid one is the distribution function of gamma. And the non-solid one is the gamma approximation function for the true distribution. We see that these three curves are very close even when the sample size is small. It demonstrates that even for small sample sizes, the distribution of 2W is a true gamma type distribution and our technique used to estimate the parameter r works well.

For another example, the first derivative of the likelihood function relatived to parameter does not exist. For twelve different sample sizes, we estimate the slope of log 2W against log error. While the sample size is one, the Estimator is 1. Also, we observed that The slope increases as the sample size increases. When the sample size is large enough, slope is close to 2. Then draw three curves, that is, the empirical distribution, the distribution of gamma, and the gamma approximation function. The plot also

demonstrates, for any kind of sample sizes, the distribution of 2W is close to a gamma type distribution and our technique works well. We noticed that in this example, the distribution of log-likelihood ratio statistic changes from gamma distribution of degree 1 to degree 2 as sample size changes from one to infinity.

Finally, consider the last example, the maximum likelihood estimator statistic is again not asymptotically normal-distributed. For twelve sample sizes, we estimate the slope of log 2W against log error. While the sample size is equal 1, the estimator slope equal 2. We observed that slope decreases as the sample size increases. If the sample size is large enough slope is close to 1. Then draw three curves on the same plot as which we done before. In the n=2 case, the result is not as well as other cases, but does not go much far wrong. We noticed that in this example, the distribution of log-likelihood ratio statistic changes from gamma distribution of degree 2 to degree 1 as sample size changes from one to infinity.

In this project, even the Taylor's expansion of the likelihood function do not exist or the asymptotic distribution of the maximum likelihood estimator is not a normal distribution, our method still work well. The key point is that the likelihood contour sets are of fan-shaped. And, therefore, the log-likelihood ratio statistic is asymptotically gamma-distributed.

四、計畫成果自評

　　相信本計畫的研究結果對於把參數空間賦予一個幾何距離，然後再來研究大樣本時估計的收斂性與幾何距離的收斂性會有很大的影響與貢獻。同時，對於小樣

本時的概似比函數的分配估計也會有很有
意義的逼近，一定可以廣泛的被統計學家
所使用。

五、參考文獻

(1) Bickel, P.J. and Doksum, K.A. (1977). Mathematical Statistics: Basic ideas and selected topics. Prentice Hall, Englewood Cliffs.

(2) Bickel, P.J. and Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction --- A Bayesian argument, The Annals of Statistics, 18, 1070-1090.

(3) Casella, G. and Berger, R.L. (1990). Statistical Inference. Duxbury, Belmont.

(4) Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics, Chapman and Hall, London.

(5) David, A.P. (1991). Fisherian inference in likelihood and frequential frames of reference (with discussion), Journal of Royal Statistics Society B},53, 79-109.

(6) Kendall, M. and Stuart, A. (1979). The advanced Theory of Statistics, Volume II: Inference and Relationship}, 4th edition. Macmillan, New York.

(7) Le Cam, L. and Yang, G.L. (1990).em Asymptotic in Statistics: Some Basic Concepts. Springer-Verlag, New York.

(8) Lehmann, E. L. (1986). Testing Statistical Hypotheses}, 2nd edition. Wiley, New York.

(9) Scheffe, H. (1947). A useful convergence theorem for probability distributions. Ann. Math. Statist., 18, 434-438.

(10) Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. J. Amer. Statist. Assoc, 81, 82-86.

(11) Wald, A. (1941). Asymptotically most powerful tests of statistical hypotheses. Ann. Math. Statist, 12, 1--19. Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9, 60-62.

(12) Wilks, S.S. (1962). Mathematical Statistics. Wiley, New York.

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
※※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　※
※　HELLINGER 距離統計模型的幾何結構與估計法(2/2)　※
※　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：□個別型計畫　　□整合型計畫

計畫編號：NSC90 – 2118 - M009 - 013

執行期間：2001 年 8 月 1 日至　2002 年 7 月 31 日

計畫主持人：洪慧念

共同主持人：

計畫參與人員：

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：

中　華　民　國　91　年　10　月　28　日