行政院國家科學委員會專題研究計畫成果報告
-橢球形區域之估計-

Preparation of NSC Project Reports
-Estimation of an elliptically shaped domain-

計畫編號：NSC 90-2118-M-004-009
執行期限：90 年 8 月 1 日至 91 年 7 月 31 日
主持人：蔡紋琦 執行機構及單位名稱:國立政治大學統計學系
共同主持人：無 執行機構及單位名稱：無
計畫參與人員：周奇勳 執行機構及單位名稱：國立政治大學統計學系
賴信宏 國立政治大學統計學系

一、中文摘要

設法去找出一個未知區域的形狀、大小、或位置在很多科學領域中常常是一件基本且關鍵的工作。例如對環境研究者而言，他們可能需要畫定出被某污染物所污染的區域；對地質學者而言，則他們可能需要找出某一特定礦物的分布範圍。而其統計語言可寫成：藉由區域中隨機觀察到的位置來推測整個可能散佈的範圍。

假設 S 是一個有界我們想要估計的區域，Ś 則是所有可能之區域所收集起來的集合，譬如 S 是一個橢圓球，Ś 則是所有橢圓球形狀區域所形成的集合。則對任何一般情況，我們可以證明出最大擬似估計和貝氏估計會強收斂到真正的區域(在差集合測度的距離之下）。不過對於其極限分佈，目前則只能做到用二維的 Cramer-von Mises 型檢定來做檢查，此一方法有其缺失，即必須先給定一個可能的極限分佈，然後再做檢定，得到的並不是百分之百確定正確的分佈答案。

關鍵詞：形狀，橢圓球，不連續點，最大擬似估計，貝氏估計，強收斂，二維 Cramer-von Mises 型檢定

Abstract

Estimating the location, shape, and size of an unknown region of interest is usually an important task in many science disciplines. In environmental studies, the geographical spread of a pollutant is frequently crucial. In Geology, it is often required to find the covering area of a specific mineral substance. One general formulation of this kind of problem in Statistical language would be: estimating an unknown domain S of interest based on n points randomly selected from it.

Let S be the bounded domain that we wish to estimate and Ś be the collection of the domains with certain features that we believe S owns. For example, S is an ellipse and Ś is the ellipse family. Under some weak conditions on Ś, we can show that the maximum likelihood estimate and the Bayes estimate are strongly consistent with respect to the set-difference distance. As to the limiting distribution, the exact formulation is still not available. So far, we are able to use the bivariate Cramer-von Mises type of test to check for any possible limiting distribution of the estimates. However, it does not provide a 100% sure conclusions.

Keywords: shape, ellipse, discontinuities, maximum likelihood estimate, Bayes estimate, strong consistency, bivariate Cramer-von Mises type of test

二、Introduction

In biology, the size and the shape of home range within a community of a species of animal are often a starting point for the analysis of a social system. In forestry, estimating the geographical edge of a rare species of plant based on sighting of individuals is an important issue as well. This project examines a mathematical problem motivated by the applications above. The statistical model can be stated as follows: in an Euclidean space R^k, n independent observations (X_1,x_2,…,X_n) uniformly distributed in an unknown compact domain of interest (S) are available. We want to estimate the region S based on information of X_1,…,X_n.

When one does not restrict any requirement on the shape of the region S besides convexity, the convex hull of X_1,…X_n turns out to be the maximum likelihood estimate of S. One can find some related results about the limiting behavior of the convex hull in the literatures. If one believes that S is likely spherical and assumes the center of symmetry is known, for instance, S is an L_p ball with center at the origin of the k-dimensional Euclidean space, the strong consistency of the maximum likelihood estimate and the Bayes estimate of S has been proved as well. However, no limiting distribution has been yet obtained. Therefore one focus of this project is on the limiting distribution of the estimates. Another focus of this project is to liberate some assumptions on the shape and the location of S. For example, S can be an ellipse with an unknown center or a linear combination of several ellipses. The summary of the results for these two aspects of consideration is stated separately in the following two sections.

三、Strong Consistency of the Estimates

Denote Ś the family of domains from which the true region S comes. Under some regularity conditions on Ś, the maximum likelihood estimate (Ŝ_n_mle, the set in Ś which contains all the observations X_1,…,X_n with smallest Lebesgue volume) is strong consistent to the true set under the set-difference metric; namely,

P(λ (Ŝ_n_mle s S)h 0 as n h ∞)=1.

This comes from the theorem below:

THEOREM: Let S_0 be the true region of interest and Ŝ_n be a function of X_1,…,X_n satisfying

(f(X_1, Ŝ_n) f(X_2,Ŝ_n) … f(X_n, Ŝ_n))/
(f(X_1,S_0) f(X_2,S_0) … f(X_n,S_0)) >=c

for some positive constant c and for any n, X_1,…X_n, where f(X,S)=(1/ λ (S)) 1_{X∈S}. If for any neighborhood U of S_0 (with respect to the set-difference metric), we have

P_{S_0}(lim_ {n h ∞ }
(sup_{S∈ Ś\U}f(X_1,S) f(X_2,S)…f(X_n,S))/
(f(X_1,S_0) f(X_2,S_0) … f(X_n),S_0)=0)=1.

Then P_{S_0}(λ (Ŝ_n s S)h 0 as n h ∞)=1.

A similar result for the Bayes estimate can be derived as well. We skip the statement here.

四、Limiting Distributions of the Estimates

Since the model assumes that X_1,…X_n are uniformly selected from the region S of interest. The support S itself is or related to the parameter indeed. In other words, the boundary points of S are exact the points of discontinuity of the density of the random variable observed. Therefore, the first thought for obtaining the exact form of the limiting distributions of the estimates is to try to connect the estimates to a stochastic process that may be easier to deal with. It turns out that we are, unfortunately, unable to do it still. Therefore another approach is considered.

Suppose S is an L_p ball with center at the origin of the Euclidean space. Namely, we can use a two-dimensional parameter (the shape and the size) to characterize the region S. Consider the modified bivariate Cramer-von Mises statistic that is used to test whether the distribution of one population differs from the distribution of another

population. Suppose we are able to "guess" the limiting distribution of the estimates. Then we can "test" if the conjecture is correct or not by first generating a bunch of two-dimentional points of the estimates and then use the modified bivariate Cramer-von Mises type of test to compare the simulated points of the estimates with the limiting distribution believed. However, as one can see that this is not a really good approach as we need to specify the conjectured limiting distribution and we know that any test does not guarantee a one hundred percent correct conclusion.

五、參考文獻

[1] H. Braker, T. Hsing, and N.H. Bingham (1998), On the Hausdorff distance between a convex set and an interior random convex hull, adv. Appl. Probab.

[2] Herman Rubin (1961), The Estimation of Discontinuities in Multivariate Densities, and Related Problems in Stochastic Processes, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability

[3] Dale L. Zimmerman (1993), A Bivariate Cramer-von Mises Type of Test for Spatial Randomness, Applied Statistics

[4] Stephen E. Syrjala (1996), A Statistical Test for a Difference Between the Spatial Distributions of two Populations, Ecology

[5] Derek S. Cotterill, Miklos Csorgo (1982), On the Limiting Distribution of and Critical Values for the Multivariate Cramer-Von Mises Statistic, Annals of Statistics

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

※※※※※※※※※※※※※※※※※※※※※※※※※※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※　　　　　　　　（計畫名稱）　　　　　　　　　　※
※　　　　　　**橢球形區域之估計**　　　　　　　　※
※　　　　　　　　　　　　　　　　　　　　　　　　※
※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：x□個別型計畫　　□整合型計畫

計畫編號：NSC　90 － 2118 － M － 004 － 009 －

執行期間：　90　年　8　月　1　日至　91　年　7　月　31　日

計畫主持人：蔡紋琦

共同主持人：無

計畫參與人員：周奇勳、賴信宏

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：國立政治大學統計學系

中　華　民　國　91　年　10　月　30　日