行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※
※ ※
※ 在幾何學上的污染下使用發散性測量局部敏感性 ※
※ ※
※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別： ☒個別型計畫 □整合型計畫
計畫編號：NSC 90－2118－M－032－010
執行期間：90 年 8 月 1 日至 91 年 7 月 31 日

計畫主持人：婁國仁 副教授

執行單位：淡江大學管理科學學系

中 華 民 國 91 年 12 月 18 日

行政院國家科學委員會專題研究計畫成果報告

在幾何學上的污染下使用發散性測量局部敏感性
Measuring Local Sensitivity using Divergence under Geometric Contamination

計畫編號：NSC 90-2118-M-032-010

執行期限：民國 90 年 8 月 1 日至 91 年 7 月 31 日

主持人：婁國仁 淡江大學管理科學學系 副教授

研究助理：陳怡雯 淡江大學管理科學學系碩士班研究生

臺北縣淡水鎮英專路 151 號
Tel: (02)2621-5656 ext. 2845 E-mail: 109880@mail.tku.edu.tw

一、中文摘要

對事前分配（Prior）的選擇，貝式程序的敏感性是很多貝式研究者關心的主題。傳統上，在貝式推論裡的敏感度分析或穩健議題上被分成兩大類一整體的與局部的敏感性。在整體性的分析裡，考慮在一組合理的事先分配中研究事後（Posterior）特徵的變化，然而在局部性分析裡是探討以某一被引出的（相信的）事前分配，在其附近做小干擾，觀察其影響性。不管怎樣，貝式分析強烈地依賴模型的假設，利用事前與概似（Likelihood）的運用。本計畫中，我們探討雙重干擾（事前且/或概似）的作用影響在事後推論上。尤其，我們開發局部敏感性測度，為了同時對兩者作干擾，觀察事後敏感性如何。然後，使用兩機率間之距離概念，作幾何學上的型態干擾，使用發散性測度，應用在加權分配問題上，取得局部敏感性的結果。

關鍵詞：貝式因子，貝式穩健性，機率種類，微敏感性，干擾，總變異量，發散性， 加權分配。

Abstract

The sensitivity of Bayes procedures to the choice of a prior distribution is a major concern for many Bayesians. Traditionally, the sensitivity analysis or the robustness issues in Bayesian inference can be classified into two broad categories, global and local sensitivity. In global analysis, one considers a class of reasonable priors and studies the variations of posterior features, whereas in local analysis, the effects of minor perturbations around some elicited priors are studied. However, a Bayesian analysis strongly depends on modeling assumptions which make use of both prior and likelihood. In this project we investigate the effects of dual perturbations (prior and/or likelihood) on the posterior inference. In particular, we develop local sensitivity measures to detect how sensitive the posterior is with respect to simultaneous perturbations in both prior and likelihood. We then study the effects on posterior distributions (or measures) using some notion of distance between probability measures. Local sensitivity measures are also obtained using the notion of divergence measures for geometric type of perturbations with weighted distribution problems.

Keywords: Bayes Factor, Bayesian Robustness, Classes of Probabilities, Infinitesimal Sensitivity, Perturbation, Total Variation, Divergence, Weighted Distribution

二、計畫緣由與目的

A Bayesian analysis depends strongly on the modeling assumptions, which make use of both prior and likelihood. Even after fitting a standard statistical model to a given set of data, one does not feel comfortable unless some sensitivity checks are made for model adequacy. One way to measure the sensitivity of the present model is to perturb the standard model to a larger (global) or smaller (local) amount in potentially conceivable directions to determine the effect of such alterations on the analysis. Although to a strict Bayesian the prior is subjective, it is often difficult to specify or elicit a method that would yield a convincing prior. The situation becomes more difficult for high dimensional parameters. Thus, to perform a complete Bayesian analysis, one must use some sensitivity measures to check model adequacy. Notable references are due to Berger (1984, 1985, 1986, 1990) and the references contained therein. Thus, the sensitivity analysis or the robustness issues in Bayesian inference can be classified into two broad categories, global and local sensitivity. In global analysis one considers a class of reasonable priors and studies the variations of such posterior features as posterior mean, variance or credible regions etc., which determines how changes in assumptions may change the answer. Alternatively, in local analysis the effects of minor perturbations around some elicited priors are studied along several conceivable directions.

Papers involving global sensitivity analysis are due to Berger (1990), Srinivasan and Truszczynska (1990), Basu and Dasgupta (1992), Sivaganesan (1993) and the references therein. In contrast, a small but quickly growing literature on Bayesian local sensitivity has developed ; see Basu, Jamalamadaka and Liu (1993), Gustafson and Wasserman (1993), Gustafson (1994), Ruggeri and Wasserman (1993) and Ghosh and Dey (1994).

The major advantage of local sensitivity analysis is realized particularly in multivariate problems, where the global analysis is too time consuming and often analytically intractable. In the multivariate scenario, several questions arise. For example, how sensitive is the posterior marginal density for one parameter when the prior of another parameter or the likelihood corresponding to another parameter changes? These problems can be handled in a reasonable manner through local sensitivity analysis.

To develop any reasonable sensitivity measures one needs to interpose certain basic concepts. For example, What class of perturbations are to be considered?, and How do we assess the discrepancies between the models generated through these perturbations? Although there are several ways to perturb a model, we will confine ourselves to geometric contamination. There have appeared a few papers concerning the effects of dual perturbation (prior and/or likelihood) on posterior inference. Such effects were considered by Lavine (1991) and studied by Basu (1994). Here, we develop local sensitivity measures to detect how sensitive the posterior is with respect to the simultaneous perturbations in both prior and likelihood. We observe that the local sensitivity measures are expressible in terms of certain Bayes factors.

三、研究模型的摘要

We need some common notations and definitions throughout the project.

Suppose X is a random variable with density $f(x|\theta)$ where θ is the parameter of interest. We allow both X and θ to be vector valued. Let Θ be the parameter space with A. Suppose further P denotes the class of prior probability measures on (Θ, A), where $P \in P$ is a particular member.

The marginal density of X is defined as

$$m_P(x,f) = \int_\Theta f(x|\theta)P(d\theta)$$

and the posterior probability measure corresponding to P and f is defined as

$$P^x(f) = f(x|\bullet)P / m_P(x.f)$$

In weighted distribution problem, a realization x of X under $f(x|\theta)$ enters the investigators record with probability proportional to a weight function $w(x)$. Clearly, the recorded x is not an observation on X, but on the r.v. X^w, say, having pdf

$$f^{w}(x|\theta) = \frac{w(x)f(x|\theta)}{E_f[w(x)]}$$

where $E_f[w(x)] = \int w(x)\, f(x|\theta)\mathrm{d}x$ is the normalizing constant. The r.v. X^w is called the weighted version of X and its distribution in relation to that of X is called the weighted distribution with weight function w.

Rao (1965) first unified the concept of weighted distribution. Patil and Rao (1977) discuss how truncation models and damaged observations can give rise to weighted distributions. Patil (1981) and Bayarri and DeGroot (1992) suggest several applications of weighted distribution. Larose and Dey (1994) studied weighted distributions in the context of model selection in a Bayesian framework whereas Bayarri and Berger (1993) consider robustness issues for the weight function.

We consider the geometric ε-contamination classes. The perturbed prior and the likelihood are respectively given as

$$dQ_{\varepsilon_1} = c(\varepsilon_1)(dP)^{1-\varepsilon_1}(dQ)^{\varepsilon_1}, \quad 0 \le \varepsilon_1 \le 1$$

and

$$g_{\varepsilon_2} = c(\varepsilon_2) f^{1-\varepsilon_2} g^{\varepsilon_2}, \quad 0 \le \varepsilon_2 \le 1$$

where P is the elicited prior, Q is the contamination which belongs to a certain class, for example, a class of priors with unimodal densities. Similarly, f is the elicited likelihood (based on sufficient statistic) and g is the contamination.

where $c(\varepsilon_1)$ and $c(\varepsilon_2)$ are appropriate normalizing constants. We further assume f and g are compatible in the sense that both may belong to the location-scale family. The class of geometric ε-contamination was first introduced in the context of Bayesian robustness by Gelfand and Dey (1991).

We use φ -divergence between two probability measures as introduced by Csiszar (1977). Formally the φ-divergence between two probability measures P and Q is defined as

$$D_{\varphi}(P,Q)=\int \varphi\left(\frac{dP}{dQ}\right)dQ \qquad (1)$$

where φ is a smooth convex function such that $\varphi(1)=0$. Several well-known divergence measures, e.g., Kullback-Liebler, Hellinger distance, Chi-squared distance, etc. can be obtained by the appropriate choice of the φ-functions. Goel (1983) used φ-divergence to measure information in hierarchical Bayesian models. Sensitivity diagnostics based on φ -divergence were studied by Delampady and Dey (1994), Dey and Birmiwal (1994), Gustafson and Wasserman (1993) and Ghosh and Dey (1994).

Now, to define several sensitivity measures based on the φ-divergence, we assume interchange of limit with integral. We define the local sensitivity measure for the prior perturbations as

$$s_{11}(x)=\lim_{\varepsilon_1\to 0^+}\frac{D_{\varphi}\left(Q_{\varepsilon_1}^{x}(f),P^{x}(f)\right)}{D_{\varphi}\left(Q_{\varepsilon_1},P\right)} = \frac{\frac{\partial^2}{\partial\varepsilon_1^2}D_{\varphi}\left(Q_{\varepsilon_1}^{x}(f),P^{x}(f)\right)\Big|_{\varepsilon_1=0}}{\frac{\partial^2}{\partial\varepsilon_1^2}D_{\varphi}\left(Q_{\varepsilon_1},P\right)\Big|_{\varepsilon_1=0_1}} \qquad (2)$$

Equation (2) follows from (1) by using Taylor's expansion. The quantity $s_{11}(x)$ gives the ratio of the local curvature of φ-divergences between posterior and prior.

Similarly, to define the sensitivity measure for the likelihood perturbation, we consider

$$s_{22}(x)=\lim_{\varepsilon_2\to 0^+}\frac{D_{\varphi}\left(P^{x}\left(g_{\varepsilon_2}\right),P^{x}(f)\right)}{D_{\varphi}\left(g_{\varepsilon_2},f\right)} = \frac{\frac{\partial^2}{\partial\varepsilon_2^2}D_{\varphi}\left(P^{x}\left(g_{\varepsilon_2}\right),P^{x}(f)\right)\Big|_{\varepsilon_2=0}}{\frac{\partial^2}{\partial\varepsilon_2^2}D_{\varphi}\left(g_{\varepsilon_2},f\right)\Big|_{\varepsilon_2=0_1}}$$

Finally, a local sensitivity measure which captures both perturbations can be defined as

$$s_{12}(x)=\lim_{\varepsilon_1\to 0^+,\varepsilon_2\to 0^+}\frac{D_{\varphi}\left(P^{x}\left(g_{\varepsilon_2}\right),P^{x}(f)\right)}{D_{\varphi}\left(Q_{\varepsilon_1},P\right)} = \frac{\frac{\partial^2}{\partial\varepsilon_2^2}D_{\varphi}\left(P^{x}\left(g_{\varepsilon_2}\right),P^{x}(f)\right)\Big|_{\varepsilon_2=0}}{\frac{\partial^2}{\partial\varepsilon_1^2}D_{\varphi}\left(Q_{\varepsilon_1},P\right)\Big|_{\varepsilon_1=0}}$$

四、結果與討論

Now, in the following results, we give the formula for the local sensitivity measures under geometric perturbations.

Result 1. *Under geometric perturbation, the local sensitivity measures based on the φ divergences are given as*

$$s_{11}(x)=\frac{Var_{P^x(f)}\left(\log\frac{dQ}{dP}\right)}{Var_{P}\left(\log\frac{dQ}{dP}\right)}$$

$$s_{22}(x)=\frac{Var_{P^x(f)}\left(\log\frac{g}{f}\right)}{Var_{f}\left(\log\frac{g}{f}\right)}$$

and

$$s_{12}(x)=\frac{Var_{P^x(f)}\left(\log\frac{g}{f}\right)}{Var_{P}\left(\log\frac{dP}{dQ}\right)}$$

Let us first assume that for a fixed weight function $w(x)$, the likelihood function is perturbed. Then, under geometric perturbation of the likelihood function i.e., $g_{\varepsilon_2}=c(\varepsilon_2)f^{1-\varepsilon_2}g^{\varepsilon_2}$, it follows that

$$E_{g_{\varepsilon_2}}[w(x)]=c(\varepsilon_2)E_f\left[w(x)\gamma^{\varepsilon_2}\right]$$

where $\gamma={}^{g}\!/\!_{f}$. Thus the posterior can be expressed as

$$P^{x}\left(g_{\varepsilon_2}^{w}\right)=\frac{\gamma^{\varepsilon_2}E_f[w(x)]}{E_f\left[w(x)\gamma^{\varepsilon_2}\right]}\frac{E_{P^x(f)}\left[{}^{1}\!/\!_{E_f[w(x)]}\right]}{E_{P^x(f)}\left[\gamma^{\varepsilon_2}/E_f\left[w(x)\gamma^{\varepsilon_2}\right]\right]}P^{x}\left(f^{w}\right)$$

Now we can define the local sensitivity measures in the context of weighted distributions as

$$s_{22}^{w}(x)=\lim_{\varepsilon_2\to 0^+}\frac{D_{\varphi}\left(P^{x}\left(g_{\varepsilon_2}^{w}\right),P^{x}(f)\right)}{D_{\varphi}\left(g_{\varepsilon_2}^{w},f\right)}$$

For fixed weight function, we can similarly define local sensitivity measures $s_{11}^{w}(x)$ and $s_{12}^{w}(x)$ for prior perturbation and joint effect of prior and likelihood perturbation. The following result gives a computational formula for $s_{22}^{w}(x)$ under geometric perturbation.

Result 2. *For fixed weight function, the local sensitivity measure for the likelihood perturbation is given as follows. Under geometric perturbation,* $s_{22}^{w}(x)$ *is given as*

$$s_{22}^{w}(x)=\frac{Var_{P^x(f^w)}\left(\log\frac{g}{f}-\frac{E_f\left[w(x)\log\frac{g}{f}\right]}{E_f[w(x)]}\right)}{Var_f\left(\log\frac{g}{f}\right)}$$

Similar results hold for prior perturbations. The following result gives the computing formula for the prior perturbation.

Result 3. *The local sensitivity measure under geometric perturbation of the prior in the case of weighted distributions is given as follows: under geometric perturbation*

$$s_{11}^{w}(x)=\frac{Var_{P^x(f^w)}\left(\log\frac{dQ}{dP}\right)}{Var_P\left(\log\frac{dQ}{dP}\right)}$$

Similarly, we can define the effect of the joint perturbation and define $s_{12}^{w}(x)$ under both perturba tions.

Let us now consider the perturbation of the weight function. If we consider linear perturbation of the form

$$w(x)=(1-\varepsilon_3)w_0(x)+\varepsilon_3 v(x) \quad , \quad 0\le\varepsilon_3\le 1$$

then one would expect that if v_1 and v_2 are two equivalent weight function in the sense that $f^{v_1}(x|\theta)$ is same as $f^{v_2}(x|\theta)$, then they should give rise to the same $w(x)$. Such, however, is not the case.

Consequently, we put further restrictions on the choice of the weight functions. We assume that the weight functions $w(x)$'s are nonnegative and

$\int w(x)dx < \infty$. Then we define

$w^{*}(x)=w(x)/\int w(x)dx$.

These assumptions are not, of course, always valid. Therefore, we can define a local sensitivity measure for the perturbation of the weight function as

$$s_{33}^{w}(x)=\lim_{\varepsilon_3\to 0^+}\frac{D_\varphi\left(P_w^x(f),P_{w_0}^x(f)\right)}{D_\varphi\left(w^{*}(x),w_0^{*}(x)\right)}$$

The following result gives the computing formula for $s_{33}^{w}(x)$.

Result 4. *Under perturbation of the weight function, the local sensitivity measure is given as follows: under geometric perturbation*

$$s_{33}^{w}(x)=\frac{Var_{P_{w_0}^x(f)}\left(\frac{E_f\left[w_0(x)\log\frac{v(x)}{w_0(x)}\right]}{E_f w_0(x)}\right)}{Var_{w_0^{*}(x)}\left(\log\frac{v(x)}{w_0(x)}\right)}$$

五、計畫成果自評

Suppose we have two different φ-functions, say φ_1 associated with the prior divergence and φ_2 associated with the posterior divergence. Then we can define the local sensitivity measure for the prior perturbation as

$$s_{11}=\lim_{\varepsilon_1\to 0^+}\frac{D_{\varphi_2}\left(Q_{\varepsilon_1}^x(f),P^x(f)\right)}{D_{\varphi_1}\left(Q_{\varepsilon_1},P\right)}$$

An overall local sensitivity measure can be given as the matrix

$$S(x)=\begin{pmatrix} s_{11}(x) & s_{12}(x)\\ s_{12}(x) & s_{22}(x)\end{pmatrix}$$

For a given problem, the matrix $S(x)$ can be computed and its eigenvalues can be used as its maximum and minimum direction of sensitivity. We can use this $S(x)$-matrix to investigate the influence of the observation x as follows. Suppose *g det(A)* denotes the generalized determinant of a matrix A , i.e. *g det(A)* = product of the non-zero eigenvalues of A . Now suppose we perturb the observation x by δ to obtain $x+\delta$ where δ is of the same dimension as that of x . Recompute the $S(x)$-matrix and call it $S(x+\delta)$, then find the relative change $|(g\det(S(x+\delta))-g\det(S(x)))/g\det(S(x))|$. If this relative change is large, then we may say that x is an influential observation. We can apply the same principle to study the robustness of prior and likelihood using $g\det(S(x))$, as in general the S-matrix depends on both prior and the likelihood.

In this research, we have reached the goals in my proposal, but it can be done more research in the future.

After the report is completed, the empirical results will be submitted to the related conferences and journals.

References

1. Basu, S. (1994). ``Posterior sensitivity to the sampling distribution and the prior: more than one observation." Research Report 66, Department of Mathematical Sciences, University of Arkansas.

2. Basu, S., Jammalamadaka, S.R. and Liu (1993). ``Local posterior robustness with parametric priors: maximum and average sensitivity." Technical Report 239,Department of Statistics and Applied Probability, University of California, Santa Barbara.

3. Basu, S. and Dasgupta, A. (1992). ``Bayesian analysis with distribution bands: the role of the loss function." Technical Report 208, Department of Statistics and Applied Probability, University of California, Santa Barbara.

4. Berger, J. (1984). ``The robust Bayesian Viewpont (with discussion)." In Robustness in Bayesian Statistics} (J. Kadane ed.): North -Holland, Amsterdam.

5. Berger, J. (1985). Statistical Decision Theory (2^{nd} Edition). Springer-Verlag: N.Y.

6. Berger, J. (1986). Comment on: ``On the consistency of Bayes estimates" by Diaconis, P. and Freedman, D. Ann. Statist., 14, 30-37.

7. Berger, J. (1990). ``Robust Bayesian analysis: sensitivity to the prior." J. Statist. Plan. Inference, 25, 303-328.

8. Csiszar, I. (1977). ``Information measures: a critical survey." Transactions of 7th Prague Conference on Information Theory, Statistical Decision Functions and Random Process 1974, Boston: Reidel Publications.

9. Gelfand, A.E. and Dey, D.K. (1991). ``On Bayesian robustness of contaminated classes of priors." Statistics and Decisions, 9, 63-80.

10. Ghosh, S.K. and Dey, D.K. (1994). Sensitivity diagnostics and robustness issues in Bayesian inference. Technical Report No. 94-30, Department of Statistics, University of Connecticut.

11. Gustafson, P. (1994). ``Local sensitivity in Bayesian statistics." Ph.D. dissertation, Department of Statistics, Carnegie Mellon University.

12. Gustafson, P. and Wasserman, L. (1993). ``Local sensitivity diagnostics for Bayesian inference." Technical Report 574, Department of Statistics, Carnegie Mellon University.

13. Lavine, M. (1991). ``Sensitivity in Bayesian Statistics: the prior and the likelihood." J. Amer. Statist. Assoc., 86, 396-399.

14. Ruggeri, F. and Wasserman, L. (1993). ``Infinitesimal sensitivity of posterior distributions." To appear in Canad. J.Statist.

15. Sivaganesan, S. (1993). ``Robust Bayesian diagnostics." J. Statist. Plan. Inf., 35, 171-188.

16. Sivaganesan,S. and Berger, J.O. (1989). ``Ranges of posterior measures for priors with unimodal contaminations." Ann. Statist., 17, 868-889.

17. Srinivasan, C. and Truszcynska, H. (1990). ``On the ranges of posterior quantities." Technial Report 294, Department of Statistics, University of Kentucky.