八十八年度行政院國科會專題研究計劃成果報告

計劃編號：NSC 88-2118-M-002-007

執行期限：87/08/01 ¼ 88/07/31

主持人：蕭朱杏　　台大流病所生物醫學統計組

利用事後分配的馬可夫鏈樣本來估計邊際密度函數與貝氏因子

Using Markov Chains to Estimate Marginal Distributions and Bayes factor

計畫中文摘要：

貝氏因子(Bayes factor) **是貝氏統計方法中對於推論假設檢定**(hypothesis testing)**所用的一個統計量。它的定義可以表成兩個資料的邊際機率**(marginal probability of data)**的比值。然而，貝氏因子的計算在實際應用上卻常常遇到困難。在過去要解決這個問題常使用** Laplace method**；現在，拜電腦(軟硬體)快速發展所賜，則常使用** Markov chain Monte Carlo (MCMC) **的方法。不過，直接運用**MCMC**的方法只能提供我們有興趣之隨機變數**(parameter of interest) **的事後樣本**(posterior sample)**，並不能直接得到貝氏因子的值。解決方法一是善加運用事後機率樣本，直接估計在每一個模式下資料的邊際機率。簡單的說，我希望能夠找到事後樣本中的某個值(如** mode **或** mean **)，以及在該點的機率密度函數值，然後再估計資料的邊際機率與貝氏因子。這個方法只是眾多貝氏因子估計值中的一個；但是，它提供另外一個簡單好算的選擇。**

關鍵詞：Bayes factor, Laplace-Metropolis, MCMC, Random effect

十七、計畫英文摘要：

Bayes factor is a commonly used quantity in Bayesian testing hypotheses. Its definition is the ratio of posterior odds to prior odds in favor of the null hypothesis. This quantity can be rewritten, using Bayes' theorem, as a ratio of marginal probabilities. Its computation is often complicated and difficult in practice, especially for complex hierarchical models such as random effect model. When the analytical integration is not feasible, one can use either the asymptotic Laplace

approximation, or the numerical method. The Markov chain Monte Carlo (MCMC) method is one efficient way which has been used widely among statisticians. The first goal of the current research plan is to develop a numerical method using Markov chains. I intend to use the posterior samples, assumed coming easily from established software like BUGS, to estimate the mean (or the mode) and the value of the probability density function at that point. This approach was first mentioned in Raftery (1996) but has not yet been investigated. It would require the technique from nonparametric density estimate. The second goal of this research plan is to increase the feasibility of the pseudo-prior for more general cases.

Introduction

The use of Bayes factors (BF) has been an important research topic in model selection and hypotheses testing. Its definition is the ratio of posterior odds to prior odds, and can be written as the ratio of marginal probabilities under each competing hypothesis respectively:

$$\text{BF} = \Pr(\text{data} \mid \text{model 1}) / \Pr(\text{data} \mid \text{model 2})\,. \qquad (1)$$

When the value of Bayes factor is large, it indicates positive evidence in favor of model 1. Jeffreys (1961) has tabulated various magnitudes for the strength of evidence in different situations. If the number of competing models is more than two, such as in the case for model selection, then we need to evaluate the posterior probability of each model conditioning on the observed data:

$$\Pr(\text{model } i \mid \text{data}) = \frac{\Pr(\text{data} \mid \text{model } i)\Pr(\text{model } i)}{\sum_j \Pr(\text{data} \mid \text{model } j)\Pr(\text{model } j)}, \quad i = 1, 2, \ldots, k\,. \qquad (2)$$

Either in equation (1) or (2), the parameters are not seen; in fact, all other parameters need to be integrated out. The issue now is the approximation to integration. In the past, Laplace (Tierney and Kadane 1986) approximation has been used widely; however, not all complex models are suitable for the analytically approximation. The regularity conditions may fail or may be difficult to verify for these cases. Recently, the Markov chain Monte Carlo methods have drawn great attention for this purpose. Various approximations based on the Markov chains of the posterior samples have been proposed and discussed (Gilks, Richardson, and Spiegelhalter 1996; Raftery 1996). Raftery (1996) discusses and summaries various Monte Carlo methods using different importance sampling functions (and thus different samples from either the prior or the posterior distribution) to approximate the integration or the marginal probability. Some approximations are useful when the samples are from the prior distribution; some are for samples from posterior distribution.

With the potential difficulty of generating samples, we focus on developing the approximation method using posterior samples only, in particular the posterior samples from Markov chains. Carlin and Chib (1995) propose the formulation of pseudo-prior distribution to connect the model i and parameters θ_j under model j ($j \div i$) to complete the specification of a full set of conditional probability distributions for Gibbs sampler. This approach also focuses on the use of and generations of posterior samples of the parameters and the discrete model index. Lewis and Raftery (1997) propose the Laplace-Metropolis estimator for approximation in the case of random effect logistic regression model. Basically it combines the Laplace and MCMC methods for each integration.

Nonparametric estimate:

Our approach, similar to Lewis and Raftery (1997) and diCiccio, Kass, Raftery, and Wasserman (1997), also focuses on the numerical methods, uses the Markov chains of posterior samples, but applies different Monte Carlo method. In fact, we will try to estimate the mean or mode and the posterior density function at that point assuming the posterior samples are already available. The value C to be estimated is defined as

$$C = \frac{P(y|\bar{\theta})\pi(\bar{\theta})}{f(\bar{\theta}|y)} = \frac{P(y|\hat{\theta})\pi(\hat{\theta})}{f(\hat{\theta}|y)},$$

and the value of C remains the same for every θ. A natural estimate for C is

$$C = \frac{P(y|\bar{\theta})\pi(\bar{\theta})}{f(\bar{\theta}|y)} = \frac{P(y|\hat{\theta})\pi(\hat{\theta})}{f(\hat{\theta}|y)}$$

where $\hat{f}$ is the non-parametric estimate for f.

Based on the results in Parzen (1962) and Eddy (1980), under regularity conditions the nonparametric estimate of f at the point x is asymptotically normal,

$$\frac{\hat{f}(x) - E(f(x))}{[\mathrm{var}(\hat{f}(x))]^{1/2}} \cong N(0,1)$$

where

$$\mathrm{var}(\hat{f}(x)) \approx \frac{1}{nh} f(x) \int K^2$$

and K(t) is a kernel density. Assuming that the integretion above is bounded, we then obtain $\hat{f}(x) - E(\hat{f}(x)) = o(nh)^{-1/2}$. Furthermore, because the expectation contains the bias, we obtain

$$\hat{C} = C[1 + o((nh)^{-1/2})]$$

where n is the sample size and h is the width chosen in the non-parametric density estimate.

Results:

Our simulation indicates that this approximation is fairly accurate. In addition, as mentioned above, this approximation is also easy to derive as long as the posterior samples are easy to derive. We are currently deriving the multivariate case and carrying out the multivariate simulation. The manuscript will be finished soon.

References:

1. Carlin , B. P. and Chib, S. (1995) `Bayesian Model Choice via Markov Chain Monte Carlo Methods', *Journal of the Royal Statistical Society, Series B*, 57, 473-484.
2. DiCiccio, T. J., Kass, R. E., Raftery, A., and Wasserman, L. (1997) `Computing Bayes Factors by Combining Simulation and Asymptotic Approximations', *Journal of the American Statistical Association*, 92, 903-915.
3. Eddy, W. F. (1980) `Optimum Kernel Estimators of the Mode', *The Annals of Statistics*, 8, 870-882.
4. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall.
5. Jeffreys, H. (1961) *Theory of Probability*, Appendix B. Third edition. Oxford University Press.
6. Lewis, S. M. and Raftery, A. D. (1997) `Estimating Bayes Factors via Posterior Simulation with the Laplace-Metropolis Estimator', *Journal of American Statistical Association,* 92, 648-655.
7. Newton, M. A. and Raftery, A. E. (1994) `Approximate Bayesian Inference by the Weighted Likelihood Bootstrap (with discussion)', *Journal of the Royal Statistical Society, Series B*, 56, 3-48.
8. Parzen, E. (1962) `On Estimation of A Probability Density Function And Mode', *Ann. Math. Statist.* 33, 1065-1076.
9. Raftery, A. E. (1996) `Hypothesis Testing and Model Selection', in *Markov Chain Monte Carlo in Practice*, pp.163-184, London: Chapman & Hall.
10. Tierney, L. and Kadane, J. B. (1986) `Accurate Approximations for Posterior Moments and Marginal Densities', *Journal of American Statistical Association,* 81, 82-86.
11. Zeger, S. L. and Karim, M. R. (1991) `Generalized Linear Models with Random Effects; A Gibbs Sampling Approach', *Journal of American Statistical Association*, 86, 79-85.