行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※

平滑直方圖的貝氏方法之研究

A Study on the Bayesian Methods for Histogram Smoothing

※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：☑個別型計畫　□整合型計畫

計畫編號：NSC　88－2118－M－004－001

NSC　89－2118－M－004－005

執行期間：87 年 8 月 1 日至 89 年 7 月 31 日

計畫主持人：姜志銘

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：國立政治大學應用數學系

中　華　民　國　八　十　九　年　十　月　十　八　日

行政院國家科學委員會專題研究計畫成果報告

平滑直方圖的貝氏方法之研究

A Study on the Bayesian Methods for Histogram Smoothing

計畫編號：NSC 88-2118-M-004-001
NSC 89-2118-M-004-005
執行期限：87 年 8 月 1 日至 89 年 7 月 31 日
主持人：姜志銘　政大應用數學系
jiangt@math.nccu.edu.tw
計畫參與人員：洪淑玲　政大統計系
彭志弘　政大應數系
劉猷銘　政大應數系

一、中文摘要

直方圖資料（histogram data）的貝氏推論一直苦於缺乏方便的先驗分配（prior distribution）以顯示所探討母體（underlying population）的臨近類別（neighboring categories）的頻率（frequency）平滑的先驗知識（prior knowledge），過濾變量的 Dirichlet 分配已經由 Dickey 及主持人（1988）發展出來，並且已經證實可解決前述的問題。本計畫把此種先驗分配的應用推廣到一般實務上經常發生的失去部分訊息的直方圖資料（即某些資料可能無法分辨它們來自那一個單一的類別），並且探討了準貝氏法在計算上的可行性。

關鍵詞：貝氏推論、直方圖資料、區域性平滑、準貝氏

Abstract

Bayesian inference for histogram data has suffered from the scarcity of convenient prior distributions permitting the expression of prior knowledge that the underlying population frequencies vary smoothly over neighboring categories. The filtered-variate Dirichlet distributions have been developed by Dickey and Jiang (1988) and have been shown to be able to deal with the above mentioned problems. In this research, we extend the methods to deal with censored histogram data (i.e. the data that can't be distinguished between categories). In addition, quasi-Bayes computational methods were also explored.

Keywords: Bayesian Inference, Histogram Data, Locally Smoothing, Quasi-Bayes

Introduction

In social science, we may want to estimate the proportions of households by the number of persons in the household. In medical science, patients are treated with a medication for a disease. We may want to estimate the rates of recurrence of the disease in several time intervals, say 0-1 year, 1-2 year, 2-3 year, and 3 years and above. From the Bayesian perspectives all of these examples have the property that adjacent categorical probabilities (proportions or rates) are very likely to be smooth or positively related. How do we assess the appropriate prior distribution that would take these properties into consideration? We shall assume that prior distribution family to be the filtered-variate Dirichlet distributions, which were first given by Dickey and Jiang (1998). Several prior assessment methods were also given there.

How do we, based on random sample data, estimate these categorical probabilities (proportions or rates) with the incorporation of the prior smoothness? The sample data that are reported may be fully classified, in which case the data are said to be non-censored. On the other hand, the data are said to be censored if the data are partially classified. The Bayesian treatments of the censored categorical data can be seen; for example in Karson and Wrobleski (1970), Antelman (1972), Kaufman and King (1973), Basu and Pereira (1982), Albert and Gupta (1983), Gunel (1984), Smith, Choi, and Gunel (1985), Albert (1885), Kadane (1985), Dickey, Jiang and Kadane (1987), and Paulino and Pereira (1992 and 1995). However, they did not consider the problem of the prior smoothness. Dickey and Jiang (1998) first give the solutions to these problems when non-censored data are observed.

In this research, we further solve the problems when censored data are observed. Notations and Bayesian inference with censored data and smooth prior are given next. Further results are given in Jiang (2000).

Notations and Bayesian Inferences

Assume that the population can be divided into k categories and let $\theta_1, \theta_2, \cdots, \theta_k$ be the categorical probabilities of these k categories. To estimate these θ_i's, a random sample of n subjects is taken. Let $n_1, n_2, \cdots, n_k$ be the number of subjects that truthfully report themselves in the first, second, ... , or k-th category, respectively. Here $n = n_1 + n_2 + \cdots + n_k$. The likelihood function is proportional to

$$l_1(\underline{\theta}) = \prod_{i=1}^{k} \theta_i^{n_i}, \qquad (2.1)$$

It is well known that the usual conjugate family of prior distributions for this likelihood function (2.1) is the Dirichlet, $\underline{\theta} \sim D(\underline{b})$, $\underline{b} = (b_1, b_2, \cdots, b_k)$, each $0 < b_i < \infty$. The p.d.f. (probability density function) of $\underline{\theta}$ is expressed as

$$f_1(\underline{\theta}) = \frac{1}{B(\underline{b})} \prod_{i=1}^{k} \theta_i^{b_i - 1}, \quad \underline{\theta} \in \Delta^k, \qquad (2.2)$$

where

$$B(\underline{b}) = \prod_{i=1}^{k} \Gamma(b_i) / \Gamma(b_+), \ b_+ = \sum_{i=1}^{k} b_i, \text{ and}$$

$$\Delta^k = \{\underline{\theta}: \text{ each } \theta_i \geq 0 \text{ and } \theta_+ = 1\}.$$

A high positive prior correlation of adjacent or near-neighbor category probabilities is desirable to express a prior belief in local smoothness. This can be seen from the expression for prior expected squared differences,

$$E[(\theta_i - \theta_j)^2] = [E(\theta_i) - E(\theta_j)]^2$$
$$+ [var(\theta_i) + var(\theta_j)] - 2\,cov(\theta_i, \theta_j). \qquad (2.3)$$

However, the Dirichlet random variables are negatively correlated. For example,

$$Cor(\theta_i, \theta_j)$$
$$= -\{b_i \cdot b_j / [(1 - b_i)(1 - b_j)]\}^{1/2}. \qquad (2.4)$$

Dickey and Jiang (1998) introduce the filtered-variate Dirichlet distributions for local smoothness and define $\underline{\theta}$ to be a filtered-variate Dirichlet distribution with a constant matrix $G(k \times m)$ and a $m \times 1$ parameter vector $\underline{b}$, denoted by $\underline{\theta} \sim F_G D(\underline{b})$, if $\underline{\theta} = G\underline{\alpha}$ and the $m \times 1$ vector $\underline{\alpha}$ has a Dirichlet distribution with parameter $\underline{b}$. Note that, from the fact that $\underline{\theta}$ and $\underline{\alpha}$ are probability vectors, all entries of G are nonnegative and each column sums to unity. A Dirichlet distribution can be generalized to a generalized Dirichlet distribution, also called a Dickey's distribution. This was first given by Dickey (1983). A $m \times 1$ random probability vector $\underline{\alpha}$ is said to have a Dickey's distribution with parameters $\underline{b}$, G^T, $\underline{n}$, denoted by $\underline{\alpha} \sim D(\underline{b}, G^T, \underline{n})$, if the p.d.f. of $\underline{\alpha}$ is

$$\frac{1}{B(\underline{b})} \left(\prod_{j=1}^{m} \alpha_j^{b_j - 1}\right)$$
$$\cdot \left[\prod_{i=1}^{k} \left(\sum_{j=1}^{m} \alpha_j g_{ij}\right)^{n_i}\right] / \Re(\underline{b}, G^T, -\underline{n}), \qquad (2.5)$$

where $\underline{n} = (n_1, \cdots, n_k)$ and the normalizing constant $\Re$ in (2.5) is a special case of Carlson's multiple hypergeometric function (Carlson, 1977). Similarly, a filtered-variate Dirichlet distribution can be generalized to a filtered-variate Dickey distribution. That is, $\underline{\theta}$ is said to have a filtered-variate Dickey distribution if $\underline{\theta} = G\underline{\alpha}$ and $\underline{\alpha}$ has a Dickey distribution. Hence, if $\underline{\alpha} \sim D(\underline{b}, G^T, \underline{n})$ and $\underline{\theta} = G\underline{\alpha}$, then $\underline{\theta} \sim F_G D(\underline{b},\ G^T,\ \underline{n})$. It can be shown that, with the filtered-variate Dirichlet prior distribution $\underline{\theta} \sim F_G D(\underline{b})$ and likelihood function (2.1), the posterior distribution is the filtered-variate Dickey distribution

$$\underline{\theta} \mid \underline{n} \sim F_G D(\underline{b}, G^T, \underline{n}). \tag{2.6}$$

Assume another random sample of m subjects is taken. Let $m_1, m_2, \cdots, m_J$ be the number of subjects that truthfully report themselves in the first, second, $\cdots$, or J-th set of categories, respectively. For example, if a certain set of categories contains the first and second categories, then when we say a subject truthfully reports him/herself in this set of categories, we mean that this subject is from either the first or the second category and we cannot distinguish between them. Therefore, the likelihood function is then proportional to

$$l_2(\underline{\theta}) = \prod_{j=1}^{J} (\sum_{i=1}^{k} \theta_i h_{ij})^{m_j}, \tag{2.7}$$

where h_{ij} is the conditional probability that a subject is reported in the j-th set of categories, given that this subject is from the i-th category. This type of report data with missing distinction between categories is called censored data. In addition, a report in the j-th categories is called noninformative if $h_{ij} = h_{i'j}$ whenever both i-th and i'-th categories are within j-th set of categories. For censored data, we shall only consider the noninformatively censored data here. Dickey, Jiang and Kadane (1987) have further discussions on the noninformatively censored data.

In practice, it is likely that a sample contains the combination of the above two different types of data. Hence, the likelihood function is proportional to

$$l(\underline{\theta}) = l_1(\underline{\theta}) \cdot l_2(\underline{\theta})$$

$$= (\prod_{i=1}^{k} \theta_i^{n_i}) \cdot [\prod_{j=1}^{J} (\sum_{i=1}^{k} \theta_i h_{ij})^{m_j}]. \tag{2.8}$$

The posterior distribution can then be expressed as

$$\underline{\theta} | \underline{n}, \underline{m} \sim F_G D(\underline{b},\ [G^T \vdots G^T H], (\underline{n}, \underline{m})), \tag{2.9}$$

where H is a $k \times J$ matrix with h_{ij} being its i, j-th entry. Note that this is again a filtered-variate Dickey distribution and the distribution (2.6) is a special case of the distribution (2.9).

The posterior mean minimizes expected squared error and is an attractive and natural estimate for $\underline{\theta}$. The posterior mean can be expressed as

$$E(\underline{\theta} | \underline{n}, \underline{m}) = G\underline{w}, \tag{2.10}$$

where the i-th entry of m-vector $\underline{w}$ is w_i and

$$w_i = \frac{b_i}{b_+} \cdot \frac{\Re(\underline{b} + \underline{\delta}_{(i)}, [G^T \vdots G^T H], -(\underline{n}, \underline{m}))}{\Re(\underline{b}, [G^T \vdots G^T H], -(\underline{n}, \underline{m}))}, \tag{2.11}$$

or as

$$E(\theta_i | \underline{n}, \underline{m}) = \frac{\Re(\underline{b},\ (G^T \vdots G^T H),\ -(\underline{n} + \underline{\delta}_{(i)}, \underline{m}))}{\Re(\underline{b},\ (G^T \vdots G^T H),\ -(\underline{n}, \underline{m}))}$$

$$i = 1, \cdots, m, \tag{2.12}$$

where $\underline{\delta}_{(i)} = (\delta_{i1}, \cdots, \delta_{ik})$ with $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for all $j \neq i$. Similarly, the posterior variance can be expressed as

$$\text{var}(\underline{\theta} \mid \underline{n}, \underline{m}) = GS, \tag{2.13}$$

where S is an $m \times m$ matrix and its i, j-th entry

$$s_{ij} = \frac{b_i \cdot (b_i + 1)}{b_+ \cdot (b_+ + 1)} \cdot \frac{\Re(\underline{b} + \underline{\delta}_{(i)} + \underline{\delta}_{(i)}, [G^T \vdots G^T H], -(\underline{n}, \underline{m}))}{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}, \underline{m}))} - w_i^2, \text{ if } i = j, \tag{2.14a}$$

or

$$s_{ij} = \frac{b_i \cdot b_j}{b_+ \cdot (b_+ + 1)} \cdot \frac{\Re(\underline{b} + \underline{\delta}_{(i)} + \underline{\delta}_{(j)}, (G^T \vdots G^T H), -(\underline{n}, \underline{m}))}{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}, \underline{m}))} - w_i w_j, \text{ if } i \neq j. \tag{2.14b}$$

Alternatively, the posterior variance can also be expressed as

$$\operatorname{cov}(\theta_i, \theta_j \mid \underline{n}, \underline{m})$$

$$= \frac{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}+\underline{\delta}_{(i)}+\underline{\delta}_{(j)}, \underline{m}))}{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}, \underline{m}))}$$

$$- \frac{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}+\underline{\delta}_{(i)}, \underline{m}))}{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}, \underline{m}))}$$

$$\cdot \frac{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}+\underline{\delta}_{(j)}, \underline{m}))}{\Re(\underline{b}, (G^T \vdots G^T H), -(\underline{n}, \underline{m}))}. \qquad (2.15)$$

The computations for the posterior moments (2.10-2.15) involve the computations of Carlson's functions $\Re$, which can be computed with the methods given by Jiang, Kadane and Dickey (1992). Alternatively, we may use the quasi-Bayes methods to compute the posterior moments. For further discussions, see Jiang and Dickey (2000), Peng (2000), and Hung (1998).

References

Albert, J. H., and Gupta, A. K. (1983), "Bayesian Estimation Methods for 2×2 Contingency Tables Using Mixtures of Dirichlet Distribution," *Journal of the American Statistical Association*, 78, 708-717.

Albert, J. H. (1985), "Bayesian Estimation Methods for Incomplete Two-Way Contingency Tables Using Prior Belief of Association," in *Bayesian Statistics 2*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Amsterdam: North-Holland, pp. 589-602.

Antelman, G. R. (1972), "Interrelated Bernoulli Process," *Journal of the American Statistical Association*, 67, 831-841.

Basu, D., and Pereira, C. A. de B. (1982), "On the Bayesian Analysis of Categorical Data: The Problem of Nonresponse," *Journal of Statistical Planning and Inference*, 6, 345-362.

Carlson. B. C. (1977). *Special Functions of Applied Mathematics*, New York: Academic Press.

Dickey, J. M. (1983), "Multiple Hypergeometric Functions: Probabilistic Interpretations and Statistical Uses," *Journal of the American Statistical Association*, 78, 328-637.

Dickey, J. M., Jiang, J. M., and Kadane, J. B. (1987), "Bayesian Methods for Censored Categorical Data," *Journal of the American Statistical Association*, 82, 773-781.

Dickey, J. M., Jiang, T. (1998), "Filtered-Variate Prior Distributions for Histograms Smoothing," *Journal of the American Statistical Association*, 93, 651-662.

Gunel, E. (1984), " A Bayesian Analysis of the Multinomial Model for a Dichotomous Response With Nonrespondents," *Communications in Statistics - Theory and Methods*, 13, 737-751.

Hung, S. (1998), "Bayesian Analysis for Censored Categorical Data."(In Chinese) Ph.D. dissertation, National Chengchi Univ., Dept. of Statistics.

Jiang, T. (2000), "Bayesian Analysis of the Censored Data with Smoothed Prior Distribution." Presented in the 2000 Joint Statistical Meetings, Aug. 13-17, 2000, Indianapolis, Indiana.

Jiang, T. and Dickey, J. (2000), "Bayesian Approaches to Categorical Data with informative Censoring." To be published.

Jiang, T. J., Kadane, J. B., and Dickey, J. M. (1992), " Computation of Carlson's Multiple Hypergeometric Function $\Re$ for Bayesian Applications," *Journal of Computational and Graphical Statistics*, 1, 231-251.

Kadane, J. B. (1985), " Is Victimization Chronic? A Bayesian Analysis of the Multinomial Missing Data," *Journal of Econometrics*, 29, 47-67.

Karson, M. J., and Wrobleski, W. J. (1970), "A Bayesian Analysis of Binomial Data With a Partially Informative Category," in *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 532-534.

Kaufman, G. M., and King, B. (1973), "A Bayesian Analysis of Nonresponse in Dichotomous Process," *Journal of the American Statistical Association*, 68, 670-678.

Paulino, C. D. M., and Pereira, C. A. de B. (1992), "Bayesian Analysis of Categorical Data Informatively Censored," *Communications in Statistics - Theory and Methods*, 21, 2689-1705.

Paulino, C. D. M., and Periera, C. A. de B. (1995), "Bayesian Methods for Categorical Data under Informative General Censoring," *Biometrika*, 82, 439-446

Peng, J. (2000), "A comparison on Bayesian and Quasi-Bayesian Methods for Histogram Smoothing."(In Chinese) Master thesis, National Chengchi Univ., Dept. of Math. Sciences.

Smith, P. J., Chio, S. C., and Gunel, E. (1985), "Bayesian Analysis of a 2×2 Contingency Tables With Both Completely and Partially Cross-Classified Data", *Journal of Educational Statistics*, 10, 31-4