行政院國家科學委員會專題研究計畫 成果報告

二階段抽樣法在非對稱損失函數下貝氏序列估計之研究 研究成果報告(精簡版)

- 計畫主持人: 黃連成
- 計畫參與人員: 碩士班研究生-兼任助理人員:曾仲宏 博士班研究生-兼任助理人員:劉士嘉 博士班研究生-兼任助理人員:陳孝先

公開 資訊 : 本計書涉及專利或其他智慧財產權, 2年後可公開查詢

中 華 民 國 101 年 10 月 08 日

- 中 文 摘 要 : 貝氏的架構下,考慮研究使用非對稱的 LINEX(linear exponential)損失函數來估計特殊一維指數族(oneparameter exponential family)分佈的平均值並且每個觀察 值有一固定成本的序列估計問題。本研究計畫對 LINEX 損失 函數的貝氏序列估計問題,在給定事先分佈(prior distribution)下,提出二階段法則(two-stage procedure) 並證明它具有漸近點最優(asymptotically pointwise optimal)和漸近最佳(asymptotically optimal)性質。除此 之外,將提出一個具有穩健性(robust)的二階段法則,此法 則與資料的分佈、事先分佈無關,並將證明在某些條件下的 事先分佈,它如同給定事先分佈下的漸近點最優法則所具有 的漸近性質。
- 中文關鍵詞: 漸近最佳性,漸近點最優,LINEX損失函數,序列估計,二 階段法則。

英文摘要:

英文關鍵詞:

A Robust Two-Stage Procedure in Bayes Sequential Estimation of a Particular Exponential Family Under LINEX Loss

Leng-Cheng Hwang[∗] Chia-Chen Yang Department of Statistics Department of Finance Tunghai University Lingtung University

Taichung, Taiwan, R.O.C. Taichung, Taiwan, R.O.C.

Abstract

The problem of Bayes sequential estimation of the unknown parameter in a particular exponential family of distributions is considered under LINEX loss function for estimation error and a fixed cost for each observation. Instead of fully sequential sampling, a twostage sampling technique is introduced to solve the problem in this paper. The proposed two-stage procedure is robust in the sense that it does not depend on the parameters of the conjugate prior. It is shown that the two-stage procedure is asymptotically pointwise optimal and asymptotically optimal for a large class of the conjugate priors.

Keywords: Asymptotically optimal; Asymptotically pointwise optimal; Bayes sequential estimation; LINEX loss function; Two-stage procedure. Mathematical Subject Classifications: 62L12.

* Corresponding author (lchwang@thu.edu.tw)

1. Introduction

The Bayes sequential estimation problem is to seek an optimal sequential procedure which includes an optimal stopping time and a Bayes estimate. The Bayes estimate is usually obtained in the problem. Hence the Bayes sequential estimation problem is reduced to finding an optimal stopping time.

It is well known that the optimal stopping time exists in Bayes sequential estimation problem in certain case under mild regularity conditions. However, the exact determination of optimal stopping time appears to be a formidable task, in practice; see, e.g., Chow, Robbins and Siegmund (1971). Due to this difficulty in finding explicit optimal rules, some procedures have been proposed with the goal of finding "asymptotically" optimal rules. For instance, Bickel and Yahav (1967) provided a simple but very attractive large sample approximation to optimal rules, namely "asymptotically pointwise optimal" (APO) rules. Since Bickel and Yahav's initiation of this idea, many papers have envolved on developing APO rules in various different contexts; see, e.g. Bickel and Yahav (1968), Gleser and Kunte (1976), Woodroofe (1981), Martinsek (1987), Ghosh and Hoekstra (1995), Hwang (2001) and Hwang and Karunamuni (2008), among others.

From the practical standpoint, purely sequential procedures suffer, especially when money and time are important design factors. Multistage methods of sampling techniques are used in statistical inference. The idea of group sampling done in two stages from a normal population is proposed by Stein (1945). Cox (1952) extended the double sampling techniques to cover a wider class of problems. Within the classical non-Bayesian framework, Mukhopadhyay (1980) and Ghosh and Mukhopadhyay (1981) proposed twostage procedures instead of purely sequential procedures in sequential interval and point estimation problems. The two-stage sampling techniques are applied to Bayes sequential estimation for the exponential distribution under the squared error loss by Hwang (1999).

In this paper, the problem of Bayes sequential estimation of the unknown parameter in a particular exponential family of distributions with LINEX loss function and fixed cost for each observation is considered. Given a conjugate prior, Jokiel-Rokita (2011) derived an APO procedure depending on the parameters of the prior distribution, and it is shown to be asymptotically optimal (AO). Instead of fully sequential sampling, a twostage sampling technique is introduced to solve the problem in Section 2. The proposed two-stage procedure is robust in the sense that it does not depend on the parameters of the prior distribution. It is shown that the two-stage procedure also shares the asymptotic properties with the APO procedure. The proofs of some auxiliary lemmas in order to obtain main theorems in Section 2 are given in Section 3.

2. Two-stage procedure and its asymptotic properties

Let X_1, X_2, \cdots be a sequence of independent and identically distributed random variables from a particular exponential family of distributions with a density function of the form

$$
f_{\theta}(x) = \theta^{\frac{k}{2}} e^{-s(x)\theta}, \ x \in R,
$$

with respect to some σ -finite measure, where $\theta > 0$ is an unknown parameter, k is a positive constant, and $s(x)$ is a nonnegative function. The particular exponential family of distributions was introduced by Rahman and Gupta (1993) . In the case k is a positive integer, the family is called transformed chi-square family. Some distributions belong to the particular exponential family, for example, the normal distribution with known mean, the gamma distribution with known shape parameter and the Pareto distribution with known scale parameter. More details can be referred to Table 1 of Jokiel-Rokita (2011).

Suppose that we are interested in estimating θ . Having recorded n observations $X_1, \dots,$ X_n , we assume that the loss incurred in estimating θ by $d_n = d_n(X_1, \dots, X_n)$ is $L(\theta, d_n)$ + cn, where

$$
L(\theta, d_n) = \exp(a(d_n - \theta)) - a(d_n - \theta) - 1, \ a \neq 0,
$$

is the LINEX loss and $c > 0$ is the cost for each observation. One notes that $bL(\theta, d_n)$ with $b > 0$ is a general form of the LINEX loss function, hence c can also be regarded as a relative weight with respect to the general form of the LINEX loss. The LINEX loss function was first introduced by Varian (1975). It is a very useful asymmetric loss function that increases approximately exponentially on one side of zero and approximately linearly on the other side. In the case $a < 0$, the loss function indicates that underestimation is more costly than overestimation. The opposite is true when $a > 0$.

Suppose that θ has a gamma prior distribution $\Gamma(\alpha, \lambda)$ with a density function of the

form

$$
\pi(\theta) = \frac{\lambda^{\alpha} \theta^{\alpha - 1}}{\Gamma(\alpha)} e^{-\lambda \theta}, \ \theta > 0,
$$

where $\alpha > 0$ and $\lambda > 0$. For convenience, we denote $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, $S_n = \sum_{i=1}^n s(X_i)$ and $\bar{S}_n = \frac{S_n}{n}$ $\frac{S_n}{n}$ for all $n \geq 1$. One notes that for given θ , $s(X_i)$ has a gamma distribution $\Gamma(\frac{k}{2},\theta)$ and then \bar{S}_n has a gamma distribution $\Gamma(\frac{nk}{2},n\theta)$. It is easy to see that the posterior distribution of θ given \mathcal{F}_n is the gamma distribution $\Gamma(\alpha_n,\lambda_n)$, where $\alpha_n = \alpha + \frac{nk}{2}$ $\frac{ik}{2}$ and $\lambda_n = \lambda + S_n$. By straightforward calculations and assume $a > -\lambda$, we can obtain that for a given stopping time T , the optimal estimate is the Bayes estimate

$$
\tilde{\theta}_T = \frac{\alpha_T}{a} \log \left(1 + \frac{a}{\lambda_T} \right).
$$

Then the Bayes risk of the Bayes sequential procedure $(T, \tilde{\theta}_T)$ is equal to

$$
E\left\{\frac{a\alpha_T}{\lambda_T} - \alpha_T \log\left(1 + \frac{a}{\lambda_T}\right) + cT\right\}.
$$

Hence , finding an optimal Bayes sequential procedure for the sequential problem is equivalent to constructing an optimal stopping time for the sequence $\{L_n(c); n \geq 1\}$, where $L_n(c) = Y_n + cn$ and

$$
Y_n = \frac{a\alpha_n}{\lambda_n} - \alpha_n \log \left(1 + \frac{a}{\lambda_n} \right).
$$

Bickel and Yahav (1967, 1968) described methods for finding a family of stopping times $\{t_c; c > 0\}$ which is APO with respect to $\{Y_n + cn; n \ge 1\}$, that is,

$$
\lim_{c \to 0} \frac{Y_{t_c} + ct_c}{\inf_n E(Y_n + cn)} = 1
$$
 a.s.

They also showed that this family of stopping times is AO, that is,

$$
\lim_{c \to 0} \frac{E(Y_{t_c} + ct_c)}{\inf_T E(Y_T + cT)} = 1,
$$

where the infimum extends over all \mathcal{F}_n -stopping times T.

To this case, Jokiel-Rokita (2011) showed that the sequence $\{L_n(c); n \geq 1\}$ satisfies the conditions of Theorem 2.1 in Bickel and Yahav (1967) and Theorem 3.1 in Bickel and Yahav (1968). Hence Jokiel-Rokita (2011) obtained that under the condition $a > -\lambda$, the family of stopping times $\{N_c; c > 0\}$ defined by

$$
N_c = \inf \left\{ n \ge \left[\frac{4}{k} \right] + 1 : Y_n \le cn \right\}, \ c > 0,
$$

with $[x]$ denoting the integer part of x, is APO and AO with respect to the sequence ${L_n(c); n \geq 1}.$

The sequential procedure $(N_c, \hat{\theta}_{N_c})$ depends on the known parameters of the gamma prior. When the parameters of the prior distribution are misspecified or unknown, the procedure is not appropriate. Hence, we would like to propose a procedure, which is independent of prior parameters, but at the same time it still possesses some asymptotic properties. Here we concentrate on a two-stage procedure instead of purely sequential procedure for the sake of simplicity and economy.

Notice that the result $nY_n \to \frac{a^2\theta^2}{k}$ $\frac{\partial^2 \theta^2}{\partial k}$ a.s. obtained by Jokiel-Rokita (2011) and $\bar{S}_n \to \frac{k}{2\theta}$ a.s. assured by the strong law of large numbers. Now we describe the two-stage procedure. The procedure, by means of the definition of N_c , takes an initial sample of size $n_0 = n_0(c)$ $[\delta c^{-\gamma}] + 1$ for some $\delta > 0$ and for some $0 < \gamma < \frac{1}{2}$, a second sample to bring the sample size to √

$$
T_c = \max\left\{n_0, \left[\frac{|a|\sqrt{k}}{2\sqrt{c}\bar{S}_{n_0}}\right] + 1\right\}.
$$

Then θ can be estimated by $\hat{\theta}_{T_c} = k/(2(\bar{S}_{T_c} + b_{T_c}))$, where $\bar{S}_{T_c} = \frac{1}{T_c}$ $\frac{1}{T_c} \sum_{i=1}^{T_c} s(X_i)$ and $b_{T_c} = \frac{b}{T_c}$ T_c with a fixed constant $b > 0$. Here the proposed two-stage procedure $(T_c, \hat{\theta}_{T_c})$ is robust in the sense that it does not depend on the parameters of the gamma prior.

Let the posterior risk of the estimator $\hat{\theta}_n$ be

$$
Y_n^* = E(L(\theta, \hat{\theta}_n) | \mathcal{F}_n)
$$

= $\left(\frac{\lambda_n}{a + \lambda_n}\right)^{\alpha_n} e^{ak/(2(\bar{S}_n + b_n))} - \frac{ak}{2(\bar{S}_n + b_n)} + \frac{a\alpha_n}{\lambda_n} - 1,$

where $b_n = \frac{b}{n}$ $\frac{b}{n}$ for all $n \geq 1$. Then the performance of the two-stage procedure $(T_c, \hat{\theta}_{T_c})$ will be measured by its Bayes risk

$$
R(T_c, \hat{\theta}_{T_c}) = E(L(\theta, \hat{\theta}_{T_c}) + cT_c) = E(Y_{T_c}^* + cT_c).
$$

The family of the two-stage stopping times $\{T_c; c > 0\}$ and the two-stage procedure $(T_c, \hat{\theta}_{T_c})$ are APO and AO for a large class of gamma prior distributions in the following Theorem 2.1 and Theorem 2.2, respectively. The proofs for the two main theorems will be given in Section 3.

Theorem 2.1. (i) ${T_c: c > 0}$ is APO with respect to ${Y_n + cn; n \ge 1}$ and ${Y_n^* + cn; n \ge 1}$. (ii) $\frac{Y_{T_c}^* + cT_c}{Y_{N_c} + cN_c} \to 1$ a.s. as $c \to 0$.

Theorem 2.2. If either $-\lambda < a < 0$ or $0 < a < b, \alpha \geq 1, \lambda > a$, then the Bayes risk of the two-stage procedure $(T_c, \hat{\theta}_{T_c})$ is

$$
R(T_c, \hat{\theta}_{T_c}) = \inf_{T} E\{L(\theta, \tilde{\theta}_T) + cT\} + o(\sqrt{c})
$$

=
$$
\frac{2|a|}{\sqrt{k}} \frac{\alpha}{\lambda} \sqrt{c} + o(\sqrt{c}) \text{ as } c \to 0,
$$

where the infimum extends over all \mathcal{F}_n -stopping times T.

3. Proof

In order to prove Theorem 2.1 and Theorem 2.2, we will develop some auxiliary results, whereas the proofs of the lemmas will be omitted in here.

Lemma 3.1. We have $\sqrt{c}T_c \rightarrow \frac{|a|\theta}{\sqrt{k}}$ a.s. as $c \rightarrow 0$.

Lemma 3.2. We have $nY_n^* \to \frac{a^2\theta^2}{k}$ $\frac{2\theta^2}{k}$ a.s. as $n \to \infty$.

Lemma 3.3. For any given $p > 1$, there exists an integrable random variable that dominates $(\sqrt{c}T_c)^p$ for all sufficiently small c.

Lemma 3.4. For any given $p > 1$, there exists an integrable random variable that dominates $(\theta \bar{S}_{T_c})^{-p}$ for all sufficiently small c.

Lemma 3.5. For any $p > 1$, $\left\{ \left(\frac{\sqrt{cT_c}}{a} \right)$ $\frac{\overline{c}T_c}{\theta}\Big)^{-p}$; $c>0$ \mathcal{L} is uniformly integrable.

Lemma 3.6. If $p > 0$ and either the case $a < 0$ and $\lambda + ap > 0$ or the other case $a > 0, \alpha \geq 1$ and $\lambda > ap \cdot \max\{1, \frac{\lambda}{b}\}\$ $\frac{\lambda}{b}$, then $\{e^{p\eta_{T_c}}; c > 0\}$ is uniformly integrable, where η_{T_c} is between 0 and $a(k/(2(\bar{S}_{T_c}+b_{T_c}))-\theta)$.

Proof of Theorem 2.1.

It follows from the definition of the APO rule N_c and the result $nY_n \to \frac{a^2\theta^2}{k}$ $\frac{2\theta^2}{k}$ a.s. in Jokiel-Rokita (2011) that $cN_c^2 \rightarrow \frac{a^2\theta^2}{k}$ $\frac{\mu_{\theta}^2}{k}$ a.s. Hence, by Lemma 3.1, we have $\frac{T_c}{N_c} \to 1$ a.s. Then, by the Remark of Theorem 2.1 in Bickel and Yahav (1967), we obtain $\{T_c; c > 0\}$ is APO with respect to $\{Y_n + cn; n \geq 1\}$.

Using the fact that $Y_n^* \geq Y_n$ a.s., we obtain the following inequalities

$$
1 \le \frac{Y_{T_c}^* + cT_c}{\inf_n(Y_n^* + cn)} \le \frac{Y_{N_c} + cN_c}{\inf_n(Y_n + cn)} \cdot \frac{Y_{T_c}^* + cT_c}{Y_{N_c} + cN_c}.
$$

Hence, by Lemmas 3.1 and 3.2, we have

$$
\frac{Y_{T_c}^* + cT_c}{Y_{N_c} + cN_c} = \frac{T_c Y_{T_c}^* + cT_c^2}{N_c Y_{N_c} + cN_c^2} \cdot \frac{N_c}{T_c} \to 1 \text{ a.s.}
$$

Then, by the results of the APO rule N_c ,

$$
\frac{Y_{T_c}^* + cT_c}{\inf_n(Y_n^* + cn)} \to 1 \text{ a.s.},
$$

that is, ${T_c; c > 0}$ is APO with respect to ${Y_n^* + cn; n \ge 1}$. The part (i) thus follows, and the proof of the part (ii) is also complete. \Box

Proof of Theorem 2.2.

It follows from Lemmas 3.1 and 3.3 that

$$
E(cT_c) = \sqrt{c} \frac{|a|}{\sqrt{k}} E\theta + o(\sqrt{c})
$$

$$
= \frac{|a|}{\sqrt{k}} \frac{\alpha}{\lambda} \sqrt{c} + o(\sqrt{c}).
$$

Using Taylor's theorem, we obtain

$$
\frac{1}{\sqrt{c}}L(\theta, \hat{\theta}_{T_c}) = \frac{1}{\sqrt{c}}e^{\eta_{T_c}} \frac{a^2}{2} \left(\frac{k}{2(\bar{S}_{T_c} + b_{T_c})} - \theta\right)^2 \n= e^{\eta_{T_c}} \frac{a^2 k}{4\sqrt{c}T_c(\bar{S}_{T_c} + b_{T_c})^2} \left(\frac{\sqrt{T_c}(\bar{S}_{T_c} + b_{T_c} - \frac{k}{2\theta})}{\sqrt{\frac{k}{2\theta^2}}}\right)^2,
$$

where η_{T_c} is between 0 and $a(k/(2(\bar{S}_{T_c}+b_{T_c}))-\theta)$. It follows from the fact $\eta_{T_c} \to 0$ a.s., Anscombe's theorem and Slutsky's theorem that

$$
\frac{1}{\sqrt{c}}L(\theta, \hat{\theta}_{T_c}) \stackrel{D}{\longrightarrow} \frac{|a|}{\sqrt{k}}G,
$$

where G is defined by $G(y) = EF_{\chi_1^2}(\frac{y}{\theta})$ $\frac{y}{\theta}$ for all $y \in R$, and $F_{\chi_1^2}$ denotes the chi-squared distribution function with one degree of freedom.

On the other hand, we can rewrite

$$
\frac{1}{\sqrt{c}}L(\theta, \hat{\theta}_{T_c}) = e^{\eta_{T_c}} \frac{a^2 \theta^2}{2\sqrt{c}(\bar{S}_{T_c} + b_{T_c})^2} \left(\bar{S}_{T_c} + b_{T_c} - \frac{k}{2\theta}\right)^2
$$
\n
$$
\leq O(1)e^{\eta_{T_c}} \left(\frac{1}{\theta \bar{S}_{T_c}}\right)^2 \left(\frac{\theta}{\sqrt{c}T_c}\right)^2 \left\{\sqrt{c}\theta^2 + \left(c^{\frac{1}{4}} \sum_{i=1}^{T_c} \frac{s(X_i) - \frac{k}{2\theta}}{\sqrt{\frac{k}{2\theta^2}}}\right)^2\right\}.
$$

It follows from Lemma 2.3 of Hwang (1999) and Lemma 3.3 that for all sufficiently small c, $\sqrt{ }$ $c^{\frac{1}{4}}\sum_{i=1}^{T_c}$ $\frac{s(X_i) - \frac{k}{2\theta}}{\sqrt{\frac{k}{2\theta^2}}}$ \setminus^p is uniformly integrable for any given $p \geq 2$. Together with Lemmas 3.4, 3.5 and 3.6, we obtain for all sufficiently small c, $\frac{1}{\sqrt{2}}$ $\bar{c}_c L(\theta, \hat{\theta}_{T_c})$ is uniformly integrable. The conditions are needed here. Hence we have

$$
EL(\theta, \hat{\theta}_{T_c}) = \sqrt{c} \frac{|a|}{\sqrt{k}} E\theta + o(\sqrt{c})
$$

$$
= \frac{|a|}{\sqrt{k}} \frac{\alpha}{\lambda} \sqrt{c} + o(\sqrt{c}).
$$

The proof is thus complete. \Box

Acknowledgements

Part of this work was supported by the National Science Council of R.O.C.

References

- Bickel, P.J. and Yahav, J.A. (1967) Asymptotically pointwise optimal procedures in sequential analysis. Proceedings of Fifth Berkeley Symposium in Mathematical Statistics and Probability 1, 401-413. University of California Press
- Bickel, P.J. and Yahav, J.A. (1968) Asymptotically optimal Bayes and minimax procedures in sequential estimation. Ann. Math. Statist. **39**, 442-456
- Chow, Y.S., Robbins, H. and Siegmund, D. (1971) Great Expectations: The Theory of Optimal Stopping. Houghton Mifflin, Boston
- Cox, D.R. (1952) Estimation by double sampling. Biometrika 39, 217-227
- Ghosh, M. and Mukhopadhyay, N. (1981) Consistency and asymptotic efficiency of two stage and sequential estimation procedures. Sankhya 43 Ser. A, 220-227
- Ghosh, M. and Hoekstra, R.M. (1995) A.P.O. rules in hierarchical Bayes regression models. Sequential Analysis 14, 99-115
- Gleser, L.J., Kunte, S. (1976) On asymptotically optimal sequential Bayes interval estimation procedures. Ann. Statist. 4, 685-711
- Hwang, L.-C. (1999) Two-stage approach to Bayes sequential estimation in the exponential distribution. Statistics and Probablity Letters 45, 277-282
- Hwang, L.-C. (1999) A robust asymptotically optimal procedure in Bayes sequential estimation. Statistica Sinica 9, 893-904
- Hwang, L.-C. (2001) Asymptotically pointwise optimal rules in the Poisson process. Sequential Analysis 20, 13-23
- Hwang, L.-C. and Karunamuni, R.J. (2008) Asymptotically pointwise optimal allocation rules in Bayes sequential estimation. Statistics and Probability Letters 78, 2490-2495
- Jokiel-Rokita, A. (2011) Bayes sequential estimation for a particular exponential family of distributions under LINEX loss. Metrika 74, 211-2194
- Martinsek, A. T. (1987) Empirical Bayes methods in sequential estimation. Sequential Analysis 6, 119-137
- Mukhopadhyay, N. (1980) A consistent and asymptotically efficient two-stage procedure to construct fixed- width confidence interval for the mean. Metrika 27, 281-284
- Rahman, M.S. and Gupta, R.P. (1993) Family of transformed chi-square distributions. Commun. Stat. Theory Methods 22, 135-146
- Stein, C. (1945) A two sample test for a linear hypothesis whose power is independent of the variance. Ann. Math. Stat. 16, 243-258
- Varian, H.R. (1975) A Bayesian approach to real estate assessment. in Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage, S. E. Fienberg and A. Zellner, eds., 195-208. Amsterdam: North Holland
- Woodroofe, M. (1981) A.P.O. rules are asymptotically non-deficient for estimation with squared error loss. Z. Wahrsch. verw. Gebiete 58, 331-341

國科會補助計畫衍生研發成果推廣資料表

日期:2012/10/05

100 年度專題研究計畫研究成果彙整表

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價 值(簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性)、是否適 合在學術期刊發表或申請專利、主要發現或其他有關價值等,作一綜合評估。

