RRPA88010192 (12P)

行政院國家科學委員會專題研究計畫成果報告

二階段抽樣法的多維自己序列估計

計畫類別：☑個別型計畫　□整合型計畫

計畫編號：NSC 88－2118－M－032－001

執行期間：87 年 8 月 1 日至 88 年 7 月 31 日

個別型計畫：計畫主持人：黃連成
共同主持人：

整合型計畫：總計畫主持人：
子計畫主持人：

註：整合型計畫總報告與子計畫成果報告請分開編印各成一冊，彙整一起繳送國科會。

處理方式：□可立即對外提供參考
（請打 v）☑一年後可對外提供參考
□兩年後可對外提供參考
（必要時，本會得展延發表時限）

執行單位：淡江大學數學系

中華民國 88 年 10 月 12 日

Two-Stage Approach to Bayes Sequential Estimation in the Exponential Distribution

Leng-Cheng Hwang
Department of Mathematics
Tamkang University
Tamsui, Taipei, Taiwan, R.O.C.

Abstract: For squared error loss plus linear cost, the problem of Bayes sequential estimation of the mean is considered. A two-stage procedure, not depending on the prior distribution, for the scale exponential family is proposed in this paper. It is shown that the proposed two-stage procedure shares the first order efficiency properties with the fully sequential procedures for a large class of prior distributions.

AMS 1991 subject classification: 62L12.

Key words and phrases: Asymptotically Bayes, asymptotically pointwise optimal, Bayes sequential estimation, two-stage procedure.

1. Introduction

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

It follows from Theorems 4.4 and 4.5 of Chow, Robbins and Siegmund (1971) that an optimal stopping rule exists, but the exact determination of the optimal stopping rule appears to be a formidable task, in practice. Bickel and Yahav (1967, 1968) describe methods for finding a family of stopping rules which is asymptotically pointwise optimal (A.P.O.) and they have shown that the A.P.O. rule is asymptotically Bayes, that is, the ratio of the Bayes risk of the A.P.O. rule and the Bayes risk of the optimal stopping rule goes to one as the cost per unit sample approaches zero. Later, the second order efficiency of the A.P.O. rule is discussed in Woodroofe (1981) and Rehaïlia (1984), etc.

When the prior distribution is a conjugate prior with unknown parameters and when some previously observed auxiliary data are available, parametric empirical Bayes procedures have been proposed by Martinsek (1987) for the exponential and normal cases, and similar parametric empirical Bayes procedures are studied by Hwang (1992) for the Bernoulli and Poisson cases. Ghosh and Hoekstra (1989) consider the estimation of the multivariate normal mean using two-stage priors and squared error loss. A.P.O. rules are developed under certain hierarchical Bayes (H.B.) models. Ghosh and Hoekstra (1995) extend the results of Ghosh and Hoekstra (1989) to a multivariate regression setting. In one-parameter exponential family, Karunamuni (1996) constructs a parametric empirical Bayes procedure by means of the A.P.O. procedure and shows that the parametric empirical Bayes procedure achieves the most string envelope risk–the Bayes risk of the optimal sequential procedure as the number of components increases to infinity for each the cost per unit sample, under some restrictive conditions.

When the prior is completely unknown, Bickel and Yahav (1968) propose a sequential procedure without using any auxiliary data. The procedure is asymptotically Bayes for a large class of prior distributions. Two sequential procedures of the type of Bickel and Yahav (1968), for the scale exponential and location normal families respectively, are A.P.O. and the second order approximation of the Bayes risk is established for a large class of prior distributions in Hwang (1997).

The procedures proposed in all the references are fully sequential procedures. In this paper, a two-stage procedure instead of fully sequential procedures for the scale exponential family is proposed. The proposed two-stage procedure does not depend on the prior and does not use any auxiliary data, and it is shown to be A.P.O. and asymptotically Bayes for a large class of prior distributions.

2. Formulation and main results

Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed (i.i.d.) random variables with density function

$$f_\theta(x) = \frac{1}{\theta} exp(-\frac{x}{\theta}), \qquad x > 0, \theta > 0.$$

Suppose that θ has a continuous bounded density with respect to Lebesgue measure such that $E(\theta^2) < \infty$. It is desired to estimate the conditional mean of X_1, $E_\theta(X_1) = \theta$, subject to the loss function

$$(\delta_n(X_1, \cdots, X_n) - \theta)^2 + cn, \qquad c > 0,$$

if one stops with first n observations and estimates θ by $\delta_n(X_1, \cdots, X_n)$. The goal is to minimize the Bayes risk over all stopping rules and over all estimators.

It is well known that for any stopping rule t, the Bayes risk is minimized by the posterior mean of θ, that is $\delta_t = E(\theta|\mathcal{F}_t)$, where $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$ for each $n \geq 1$. Therefore the Bayes risk of a sequential procedure $(t, E(\theta|\mathcal{F}_t))$ is equal to $E(Var(\theta|\mathcal{F}_t) + ct)$. Hence, finding an optimal sequential procedure for this problem is equivalent to constructing an optimal stopping rule for the sequence $\{Z_n, n \geq 1\}$, where

$$Z_n = Var(\theta|\mathcal{F}_n) + cn.$$

It follows from Theorems 2.1 and 3.1 of Bickel and Yahav (1967) that the A.P.O. rule is $U_c = \inf\{n \geq 1 : Var(\theta|\mathcal{F}_n) < nc\}$, $c > 0$. It means that, for any stopping rules $\{s_c, c > 0\}$, we have

$$\overline{\lim_{c\to 0}} \frac{Z_{U_c}}{Z_{s_c}} \leq 1 \quad \text{a.s.}$$

In view of $E(Var(\theta|\mathcal{F}_n)) \leq E(\theta^2)/n$, we know from Theorem 3.1 of Bickel and Yahav (1968) that the A.P.O. rule U_c is asymptotically Bayes, that is,

$$\begin{aligned} E(Var(\theta|\mathcal{F}_{U_c}) + cU_c) &= \inf_s E(Var(\theta|\mathcal{F}_s) + cs) + o(\sqrt{c}) \\ &= 2\sqrt{c}E(\theta) + o(\sqrt{c}) \quad \text{as } c \to 0, \end{aligned}$$

where the infimum extends over all $\mathcal{F}_n$-stopping rules s.

The A.P.O. procedure $(U_c, E(\theta|\mathcal{F}_{U_c}))$ depends on the prior distribution of θ, which is sometimes unknown or misspecified. Hwang (1997) suggests the stopping rule

$$T_c = \inf\{n \geq 3 : \bar{X}_n^2 < cn^2\}, \quad c > 0,$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, and estimate θ by $\bar{X}_{T_c}$. The stopping rule T_c and the sequential procedure $(T_c, \bar{X}_{T_c})$ are respectively shown to be A.P.O. with respect to $\{Z_n, n \geq 1\}$ and

$$\begin{aligned} E\{(\bar{X}_{T_c} - \theta)^2 + cT_c\} &= 2\sqrt{c}E(\theta) + 3c + o(c) \\ &= \inf_s E(Var(\theta|\mathcal{F}_s) + cs) + o(\sqrt{c}) \quad \text{as } c \to 0. \end{aligned}$$

Here it is shown that the first order efficiency may be obtained by procedures that take observations in two stages. The procedure studied, by means of the definition of T_c, takes an initial sample of size $n_0 = n_0(c) = [\delta c^{-\alpha}] + 1$ for some $\delta > 0$ and for some $0 < \alpha < \frac{1}{2}$, then a second sample to bring the total sample size to

$$N_c = \max\{n_0, [c^{-\frac{1}{2}}\bar{X}_{n_0}] + 1\},$$

where $[x]$ denotes the integer part of x. So we propose the two-stage procedure $(N_c, \bar{X}_{N_c})$ instead of the sequential procedure $(T_c, \bar{X}_{T_c})$. The proposed two-stage procedure is similar to those introduced by Ghosh and Mukhopadhyay (1981) and Hall (1981) for sequential point and interval estimation and sequential interval estimation, respectively, within the classical non-Bayesian framework.

Since $\sqrt{c}N_c \to \theta$ a.s. and $\sqrt{c}T_c \to \theta$ a.s., we have $\frac{T_c}{N_c} \to 1$ a.s.. From this fact, N_c is also A.P.O. with respect to $\{Z_n, n \geq 1\}$. On the other hand, the following theorem tells us that the two-stage procedure $(N_c, \bar{X}_{N_c})$ is also asymptotically Bayes for a large class of prior distributions.

Theorem If $E(X_1^3) < \infty$, then $E\{(\bar{X}_{N_c} - \theta)^2 + cN_c\} = 2\sqrt{c}E(\theta) + o(\sqrt{c})$ as $c \to 0$.

3. Proofs

We will now develop some auxiliary results on uniform integrability and apply these results to prove the main theorem.

Lemma 1 If $E(X_1^p) < \infty$ for some $p > 1$, then $\{(\sqrt{c}N_c)^p, c > 0\}$ is dominated by an integrable random variable.

Proof: Using the definition of N_c, we have

$$\begin{aligned}
(\sqrt{c}N_c)^p &= (\sqrt{c}N_c)^p \mathbf{1}_{\{N_c = [c^{-\frac{1}{2}}\bar{X}_{n_0}]+1\}} + (\sqrt{c}N_c)^p \mathbf{1}_{\{N_c = n_0\}} \\
&\leq (\bar{X}_{n_0} + \sqrt{c}n_0)^p + (\sqrt{c}n_0)^p \\
&\leq M\{\sup_{n\geq 1}(\bar{X}_n)^p + 1\}
\end{aligned}$$

for some finite constant $M > 0$, where $\mathbf{1}_A$ denotes the indicator function of A.

By noting that $\{\bar{X}_n, \sigma(\bar{X}_k, k \geq n); n \geq 1\}$ is a reverse martingale and using Doob's inequality,

$$E\{\sup_{n\geq 1}(\bar{X}_n)^p\} \leq (\frac{p}{p-1})^p E(X_1^p) < \infty.$$

This completes the proof of the lemma.

Lemma 2 If $E(X_1^{\frac{p}{2}}) < \infty$ for some $p > 2$, then $\{|c^{\frac{1}{4}} \sum_{i=1}^{N_c}(\frac{X_i}{\theta} - 1)|^p, c > 0\}$ is uniformly integrable.

Proof: We first note that Lemma 5 of Chow and Yu (1981) also holds if the filtration $\mathcal{F}_n = \sigma(Y_1, \cdots, Y_n)$ is replaced by a filtration $\mathcal{G}_n$ such that $\mathcal{F}_n \subset \mathcal{G}_n$ for each $n \geq 0$, and $\mathcal{G}_n$ and $\sigma(Y_{n+1})$ are independent. (The setting in Lemma 5 of Chow and Yu (1981) is used here.)

Observe that $\theta, \frac{X_1}{\theta}, \frac{X_2}{\theta}, \cdots$ are independent, and the $\frac{X_i}{\theta}$ are exponentially distributed with mean 1. The assertion thus follows from $\sigma(X_1, \cdots, X_n) \subset \sigma(\theta, \frac{X_1}{\theta}, \cdots, \frac{X_n}{\theta})$ and the fact that N_c are $\sigma(X_1, \cdots, X_n)$-stopping times.

Lemma 3 Assume that $E(X_1^p) < \infty$ for some $p > 0$, then $\{(\frac{\sqrt{c}N_c}{\theta})^{-p}, c > 0\}$ is uniformly integrable.

Proof: Let random variables θ and $X_1, X_2, \cdots$ be defined on a probability space $(\Omega, \mathcal{F}, P)$. Let $P_{\underset{\sim}{X}} : \mathcal{B}^\infty \times \Omega \to [0,1]$ be a regular conditional distribution for $\underset{\sim}{X} = (X_1, X_2, \cdots)$ given $\sigma(\theta)$ such that for each $w \in \Omega$ the coordinate random variables $\{\xi_n, n \geq 1\}$ of probability space $(R^\infty, \mathcal{B}^\infty, P_{\underset{\sim}{X}}(\cdot, w))$ are i.i.d. and for almost all $w \in \Omega$ the ξ_n are exponentially distributed with mean $\theta(w)$.

Let $x = (x_1, x_2, \cdots)$ and we define $s_c = \max\{n_0, [c^{-\frac{1}{2}}\bar{\xi}_{n_0}] + 1\}$, $c > 0$, where $\bar{\xi}_{n_0} = \frac{1}{n_0}\sum_{i=1}^{n_0} \xi_i$. Note that

$$E(X_1|\theta)(w) = \int_{R^\infty} x_1 P_{\underset{\sim}{X}}(dx, w) \equiv E^w \xi_1 = \theta(w) \quad \text{a.s.}$$

$$Var(X_1|\theta)(w) = \int_{R^\infty} x_1^2 P_{\underset{\sim}{X}}(dx, w) - (\int_{R^\infty} x_1 P_{\underset{\sim}{X}}(dx, w))^2 \equiv Var^w \xi_1 = \theta^2(w) \text{ a.s.,}$$

where E^w and Var^w denote respectively the expectation and variance with respect to $P_{\underset{\sim}{X}}(\cdot, w)$.

Let $0 < \gamma < 1$. Using Markov's inequality, for almost all $w \in \Omega$

$$\begin{aligned}
& P_{\underset{\sim}{X}}(\{\sqrt{c}s_c < \gamma E^w \xi_1\}, w) \\
\le\ & P_{\underset{\sim}{X}}(\{\sqrt{c}([c^{-\frac{1}{2}}\bar{\xi}_{n_0}] + 1) < \gamma E^w \xi_1\}, w) \\
\le\ & P_{\underset{\sim}{X}}(\{exp(-t\bar{\xi}_{n_0}) > exp(-t\gamma E^w \xi_1)\}, w) \\
\le\ & exp(t\gamma E^w \xi_1) E^w(exp(-t\bar{\xi}_{n_0})) \\
=\ & exp\{-n_0(ln(1 + \frac{t}{n_0} E^w \xi_1) - \frac{t}{n_0}\gamma E^w \xi_1)\}
\end{aligned}$$

for all $t > 0$. Then, by minimizing $exp\{-n_0(ln(1+\frac{t}{n_0}E^w\xi_1) - \frac{t}{n_0}\gamma E^w \xi_1)\}$ with respect to $t > 0$,

$$P_{\underset{\sim}{X}}(\{\sqrt{c}s_c < \gamma E^w \xi_1\}, w) \le exp(-n_0(\gamma - 1 - ln\gamma)) \le exp(-kc^{-\alpha}), \tag{1}$$

where $k = \delta(\gamma - 1 - ln\gamma) > 0$.

It follows from the properties of regular conditional distributions and (1) that

$$\begin{aligned}
& E\{(\frac{\sqrt{c}N_c}{\theta})^{-p} \mathbf{1}_{\{\sqrt{c}N_c < \gamma\theta\}}\} \le EE^w\{(E^w\xi_1)^p c^{-\frac{p}{2}} \mathbf{1}_{\{\sqrt{c}s_c < \gamma E^w \xi_1\}}\} \\
\le\ & c^{-\frac{p}{2}} exp(-kc^{-\alpha}) E(\theta^p) = o(1),
\end{aligned} \tag{2}$$

and

$$\begin{aligned}
& P\{(\frac{\sqrt{c}N_c}{\theta})^{-p} \mathbf{1}_{\{\sqrt{c}N_c < \gamma\theta\}} > \epsilon\} \le EP_{\underset{\sim}{X}}(\{\sqrt{c}s_c < \gamma E^w \xi_1\}, w) \\
\le\ & exp(-kc^{-\alpha}) = o(1)
\end{aligned} \tag{3}$$

for all $\epsilon > 0$. Combining (2), (3) and $E\{(\frac{\sqrt{c}N_c}{\theta})^{-(p+1)} \mathbf{1}_{\{\sqrt{c}N_c \ge \gamma\theta\}}\} \le \gamma^{-(p+1)}$, we obtain that $\{(\frac{\sqrt{c}N_c}{\theta})^{-p}, c > 0\}$ is uniformly integrable. The lemma thus follows.

Proof of theorem:

The random variables ξ_i, the probability measures $P_{\mathfrak{X}}(\cdot,\omega)$ and the stopping times s_c are the same as those described in the proof of Lemma 3. Using the properties of regular conditional distributions, Anscombe's theorem and the bounded convergence theorem, for any $y \in R$

$$\begin{aligned}
& lim_{c\to 0} P\{c^{-\frac{1}{2}}(\bar{X}_{N_c}-\theta)^2 \le y\} \\
= \; & lim_{c\to 0} EP_{\mathfrak{X}}(\{c^{-\frac{1}{2}}(\bar{\xi}_{s_c}-E^w\xi_1)^2 \le y\}, w) \\
= \; & lim_{c\to 0} EP_{\mathfrak{X}}(\{(\frac{\bar{\xi}_{s_c}-E^w\xi_1}{\sqrt{\frac{Var^w\xi_1}{s_c}}})^2 \frac{Var^w\xi_1}{\sqrt{c}s_c} \le y\}, w) \\
= \; & EF_{\chi_1^2}(\frac{y}{\theta}),
\end{aligned}$$

where $F_{\chi_1^2}$ denotes the chi-squared distribution function with one degree of freedom. Therefore

$$c^{-\frac{1}{2}}(\bar{X}_{N_c}-\theta)^2 \xrightarrow{\mathcal{D}} F, \tag{4}$$

where $\xrightarrow{\mathcal{D}}$ means weak convergence and F is the limiting distribution defined by $F(y) = EF_{\chi_1^2}(\frac{y}{\theta})$ for all $y \in R$.

In view of

$$c^{-\frac{1}{2}}(\bar{X}_{N_c}-\theta)^2 = (c^{\frac{1}{4}}\sum_{i=1}^{N_c}(\frac{X_i}{\theta}-1))^2(\frac{\sqrt{c}N_c}{\theta})^{-2}, \tag{5}$$

one obtains the uniform integrability of $\{c^{-\frac{1}{2}}(\bar{X}_{N_c}-\theta)^2, c>0\}$ by Lemmas 2 and 3.

Combining (4) and the uniform integrability of (5), we have

$$E(\bar{X}_{N_c}-\theta)^2 = \sqrt{c}E(\theta) + o(\sqrt{c}). \tag{6}$$

It is easy to see that $\sqrt{c}N_c \to \theta$ a.s.. Together with the uniform integrability of $\{\sqrt{c}N_c, c>0\}$ assured by Lemma 1, one obtains

$$E(cN_c) = \sqrt{c}E(\theta) + o(\sqrt{c}). \tag{7}$$

Combining (6) and (7), we have

$$E\{(\bar{X}_{N_c}-\theta)^2+cN_c\}=2\sqrt{c}E(\theta)+o(\sqrt{c}).$$

The proof is thus complete.

Acknowledgements

The author would like to express his sincerest gratitude to Professor I-Shou Chang and Professor Chao A. Hsiung for introducing related multistage methods to him. The author is also grateful to the referee for his valuable comments.

References

Bickel, P. and J. Yahav (1967), Asymptotically pointwise optimal procedures in sequential analysis, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** (University of California) pp. 401-413.

Bickel, P. and J. Yahav (1968), Asymptotically optimal Bayes and minimax procedures in sequential estimation, *Ann. Math. Statist.* **39**, 442-456.

Chow, Y.S., H. Robbins and D. Siegmund (1971), *Great Expectations: The Theory of Optimal Stopping* (Houghton Mifflin, Boston).

Chow, Y.S. and H. Teicher (1988), *Probability Theory: Independence, Interchangeability, Martingales* (Springer-Verlag, New York, 2nd ed.).

Chow, Y.S. and K.F. Yu (1981), On the performance of a sequential procedure for the estimation of the mean, *Ann. Statist.* **9**, 184-189.

Ghosh, M. and R.M. Hoekstra (1989), A.P.O. rules in hierarchical and empirical Bayes models, *Sequential Analysis* **8**, 79-100.

Ghosh, M. and R.M. Hoekstra (1995), A.P.O. rules in hierarchical Bayes regression models, *Sequential Analysis* **14**, 99-115.

Ghosh, M. and N. Mukhopadhyay (1981), Consistency and asymptotic efficiency of two stage and sequential estimation procedures, *Sankhyā A* **43**, 220-227.

Hall, P. (1981), Asymptotic theory of triple sampling for sequential estimation of a mean, *Ann. Statist.* **9**, 1229-1238.

Hwang, L.-C. (1992), Empirical Bayes approach to Bayes sequential estimation: the Poisson and Bernoulli cases, *Commun. Statist. Theor. Meth.* **21**, 2293-2308.

Hwang, L.-C. (1997), A Bayesian approach to sequential estimation without using the prior information, *Sequential Analysis* **16**, 319-343.

Karunamuni, R.J. (1996), Empirical Bayes sequential estimation for exponential families: the untruncated component, *Ann. Inst. Statist. Math.* **48**, 711-730.

Martinsek, A.T. (1987), Empirical Bayes methods in sequential estimation, *Sequential Analysis* **6**, 119-137.

Rehaïlia, M.E.H. (1984), Asymptotic sequential analysis in the A.P.O. rule performance, *Sequential Analysis* **3**, 155-174.

Woodroofe, M. (1981), A.P.O. rules are asymptotically non-deficient for estimation with squared error loss, *Z. Wahrsch. verw. Gebiete* **58**, 331-341.