A ROBUST ASYMPTOTICALLY OPTIMAL PROCEDURE IN BAYES SEQUENTIAL ESTIMATION

Leng-Cheng Hwang

Tamkang University

Abstract: The problem of sequential estimation of the mean, subject to the loss defined as the sum of squared error loss and sampling costs, is considered within the Bayesian framework. It is shown that the sequential procedure, as proposed by Chow and Yu (1981) in classical non-Bayesian sequential estimation, is, in fact, asymptotically Bayes for a large class of prior distributions. The proposed procedure, without using any auxiliary data, is robust in the sense that it does not depend on the distribution of outcome variables and the prior.

Key words and phrases: Asymptotically Bayes, Bayes sequential estimation, Bayes risk, optimal sequential procedure, prior distributions.

1. Introduction

Let $X_1, X_2, \ldots$ be independent observations from some population with a parameter θ. Suppose that θ is itself a random variable having a distribution. It is desired to estimate a real-valued function $m(\theta)$, subject to the loss measured by the sum of squared error and the sampling costs.

The Bayes sequential estimation problem is to seek an optimal sequential procedure which includes an optimal stopping rule and a Bayes estimate (i.e. the posterior mean of $m(\theta)$ for squared error loss). There are many papers that discussed this problem. For example, Chow, Robbins and Siegmund (1971) proved the existence of optimal stopping rules for a Bayes sequential estimation problem; Alvo (1977) obtained a lower bound for the Bayes risk of optimal stopping rules for the one-parameter exponential family; Rasmussen (1980) provided an optimal stopping time for the problem of estimating the normal mean with unknown variance and conjugate priors for the mean and variance; Woodroofe (1981) proved that, in the case of the one-parameter exponential family and conjugate priors, the asymptotically pointwise optimal rules are asymptotically non-deficient, that is, the difference between their Bayes risks and the Bayes risk of optimal stopping rules is $o(c)$, where c is the sampling cost per observation; Rehaïlia (1984) extended the result of Woodroofe (1981) to the case of non-conjugate priors. All these discussions are for the situation that the prior is known.

When the prior distribution is a conjugate prior with unknown parameters and when some previously observed auxiliary data are available, parametric empirical Bayes procedures were proposed by Martinsek (1987) for the exponential and the normal cases, and similar parametric empirical Bayes procedures were studied by Hwang (1992) for the Bernoulli and Poisson cases. These procedures were shown to be asymptotically non-deficient. When the prior is completely unknown, Bickel and Yahav (1968), without using any auxiliary data, proposed, instead, a sequential procedure which depends on the distribution of X. They showed that under some undesirably stringent conditions (see pp. 455-456 of Bickel and Yahav (1968)), the procedure is asymptotically Bayes, that is, the Bayes risk of the proposed sequential procedure and the Bayes risk of the optimal sequential procedure are asymptotically equivalent.

Suppose that we are in the situation that the unknown mean $m(\theta) = E_\theta(X)$ is to be estimated, the prior distribution of θ is unknown, and previous auxiliary data are not available. It is desirable to establish a more general sequential procedure which depends on the present data, but not on the distribution of X, such that it is asymptotically Bayes whatever be the true prior. The plan of this paper is as follows. In Section 2, for arbitrary distributions, a naive sequential procedure is proposed and its Bayes risk is shown to be asymptotically not greater than the Bayes risk of the optimal fixed-sample-size procedure. In Section 3, a one-parameter exponential family of distributions is studied and the procedure proposed in Section 2, without using any auxiliary data, is applied to estimate the mean. The procedure is shown to be asymptotically Bayes for a large class of prior distributions. The conditions needed here are much simpler than the ones given in Bickel and Yahav (1968).

We note that the proposed sequential procedure is, in fact, the one commonly used in classical non-Bayesian sequential estimation problems (see eg. Chow and Yu (1981)). Therefore, in this paper we have linked the two types of sequential estimation problems by showing that the naive sequential procedure is in fact asymptotically Bayes, whatever be the true prior, in the problem of sequential estimation of the mean subject to the loss defined by the sum of squared error loss and sampling costs.

2. Bayes Sequential Estimation of the Mean

We consider $X, X_1, X_2, \ldots$ to be a sequence of independent, identically distributed (i.i.d.) random variables with density f_θ with respect to a σ-finite measure μ, where the parameter θ may be multi-dimensional. Let E_θ denote the expectation with respect to f_θ. Here we treat θ as a realization of a random variable. We take Θ to be the parameter space and are given a prior distribution G on Θ. Suppose that we are interested in estimating $E_\theta(X)$. Having recorded n

observations $X_1, \ldots, X_n$, we assume that the loss incurred in estimating $E_\theta(X)$ by $\delta_n(X_1, \ldots, X_n)$ is

$$(\delta_n(X_1, \ldots, X_n) - E_\theta(X))^2 + cn, \qquad (2.1)$$

where c is the cost per unit sample.

Suppose that we adopt the sample mean $\bar{X}_n$ as our estimate. For a fixed sample size n, the corresponding Bayes risk is given by

$$R_n = E\{(\bar{X}_n - E(X \mid \theta))^2 + cn\} = E\Big(\frac{\text{Var}\,(X \mid \theta)}{n} + cn\Big),$$

where E denotes the expectation with respect to the overall probability measure. The risk R_n is minimized by taking the sample size n_0, which satisfies

$$\Big[\big(\frac{E(\text{Var}\,(X \mid \theta))}{c}\big)^{\frac{1}{2}}\Big] \leq n_0 \leq \Big[\big(\frac{E(\text{Var}\,(X \mid \theta))}{c}\big)^{\frac{1}{2}}\Big] + 1 \qquad (2.2)$$

with $[x]$ denoting the integer part of x. We thus have $R_{n_0} \simeq 2\sqrt{c}(E(\text{Var}\,(X \mid \theta)))^{\frac{1}{2}}$.

The value of θ under study represents only a single observation from the distribution G, and a check of validity of G is not possible if no previous auxiliary experiments are available. In case the prior distribution G is misspecified, unknown, or irrelevant to the present θ, the optimal fixed sample size n_0 can not be obtained, and there is no fixed-sample-size procedure that will attain R_{n_0}. We then wish to find a sequential procedure such that its Bayes risk will not be greater than R_{n_0} asymptotically.

Let

$$t_c = \inf\{n \geq n_c : \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 < n^2 c\}, \qquad (2.3)$$

where n_c is a positive integer, which may depend on c. The inequalities in (2.2) suggest that the stopping rule t_c may be a good candidate and we shall estimate $E_\theta(X)$ by $\bar{X}_{t_c}$. The performance of the sequential procedure $(t_c, \bar{X}_{t_c})$ will be measured by its Bayes risk $R(t_c, \bar{X}_{t_c})$. In Theorem 2.1, we show that under some regularity conditions,

$$R(t_c, \bar{X}_{t_c}) = 2\sqrt{c}E(\text{Var}\,(X \mid \theta))^{\frac{1}{2}} + o(\sqrt{c}), \quad \text{as} \ \ c \to 0. \qquad (2.4)$$

It is easy to see that $R_{n_0} \geq 2\sqrt{c}E(\text{Var}\,(X \mid \theta))^{\frac{1}{2}}$, indicating that the Bayes risk $R(t_c, \bar{X}_{t_c})$ of the sequential procedure $(t_c, \bar{X}_{t_c})$ is asymptotically not greater than the minimum Bayes risk R_{n_0} of the fixed-sample-size procedure.

The following conditions (A) and (B) will be needed in Lemma 2.2 and Theorem 2.1.

Condition (A). There exists a constant $\alpha > 0$ and $\theta_1, \theta_2 \in \Theta$ such that

$$\sup_{\theta \in \Theta} f_\theta(x) \leq O(1) f_{\theta_1}(x) \quad \text{for all } x > \alpha,$$

$$\sup_{\theta \in \Theta} f_\theta(x) \leq O(1) f_{\theta_2}(x) \quad \text{for all } x < -\alpha.$$

Condition (B). There exist constants $0 < a < b < \infty$ such that $|E(X|\theta)| < b$ and $\text{Var}(X|\theta) \in (a, b)$ a.s..

Remark. Conditions (A) and (B) hold in many cases including, for example, the one-parameter exponential family, the standard Weibull distribution, uniform distributions whose parameter spaces are one-dimensional bounded and closed intervals, and normal distributions with unknown means and variances whose parameter spaces are two-dimensional bounded and closed intervals.

Theorem 2.1. *Under conditions* (A) *and* (B), *if* $E|X|^{3p} < \infty$ *for some* $p > 1$, *and* $\delta c^{-1/4} \leq n_c = o(c^{-1/2})$ *for some* $\delta > 0$, *then* (2.4) *holds.*

Remark. This theorem tells us that the procedure $(t_c, \bar{X}_{t_c})$ proposed in Chow-Yu (1981) is not only risk efficient from the frequentist point of view, but also risk efficient from the Bayesian point of view for general prior distributions.

In order to prove Theorem 2.1, we first develop some lemmas concerning uniform integrability.

Lemma 2.1. *If* $E|X|^p < \infty$ *for some* $p > 2$ *and* $n_c \leq O(c^{-1/2})$, *then* $\{(\sqrt{c}t_c)^p, c > 0\}$ *is dominated by an integrable random variable.*

Lemma 2.2. *Under conditions* (A) *and* (B), *if* $\delta c^{-1/4} \leq n_c$ *for some* $\delta > 0$, *then for any* $p > 0$, $\{(\sqrt{c}t_c)^{-p}, c > 0\}$ *is uniformly integrable.*

The following lemma is a Bayesian version of Lemma 5 of Chow and Yu (1981).

Lemma 2.3. *Let* $\{s_c, c > 0\}$ *be a family of* $\sigma(\theta, X_1, \ldots, X_n)$-*stopping times such that for some* $p \geq 2$, $b > 1$, $\{(\sqrt{c}s_c)^{pb/2}, c > 0\}$ *is uniformly integrable. Let* h *be a measurable function of* X *and* θ. *Assume that* $E|h(X, \theta)|^{pa} < \infty$ *for* $a > 1$ *such that* $\frac{1}{a} + \frac{1}{b} = 1$. *Then* $\{|c^{1/4} \sum_{i=1}^{s_c} (h(X_i, \theta) - E(h(X, \theta)|\theta))|^p, c > 0\}$ *is uniformly integrable.*

Remark. Lemma 2.3 also holds if the filtration $\sigma(\theta, X_1, \ldots, X_n)$ is replaced by a filtration $\mathcal{G}_n$ such that $\sigma(\theta, X_1, \ldots, X_n) \subset \mathcal{G}_n$ for each $n \geq 0$, and $\mathcal{G}_n$ and $\sigma(X_{n+1})$ are conditionally independent given $\sigma(\theta)$.

With Lemma 2.1-2.3 we can now prove the main theorem. The proofs for the three lemmas will be given in the Appendix.

Proof of Theorem 2.1. Let θ and the $X's$ 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 $\underset{\sim}{X} = (X_1, X_2, \ldots)$ given θ such that for each $w \in \Omega$, the coordinate random variables $\{\xi_n, n \geq 1\}$ in the induced probability space $(R^\infty, \mathcal{B}^\infty, P_{\underset{\sim}{X}}(\cdot, w))$ are i.i.d..

Let $x = (x_1, x_2, \ldots)$ and define $s_c = \inf\{n \geq n_c : \frac{1}{n}\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2 < n^2 c\}, c > 0$, where $\bar{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$. Note that

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

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

Using the properties of regular conditional distributions, Anscombe's Theorem and the Bounded Convergence Theorem, for any $y \in R$,

$$\lim_{c \to 0} P\{c^{-1/2}(\bar{X}_{t_c} - E(X|\theta))^2 \leq y\}$$

$$= \lim_{c \to 0} EP_{\underset{\sim}{X}}(\{c^{-1/2}(\bar{\xi}_{s_c} - E^w \xi_1)^2 \leq y\}, w)$$

$$= \lim_{c \to 0} EP_{\underset{\sim}{X}}(\{(\frac{\bar{\xi}_{s_c} - E^w \xi_1}{\sqrt{\frac{\text{Var}^w \xi_1}{s_c}}})^2 \frac{\text{Var}^w \xi_1}{\sqrt{c}s_c} \leq y\}, w)$$

$$= EF_{\chi_1^2}(\frac{y}{\sqrt{\text{Var}^w \xi_1}}),$$

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

$$c^{-\frac{1}{2}}(\bar{X}_{t_c} - E(X|\theta))^2 \xrightarrow{\mathcal{D}} F, \tag{2.5}$$

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

In view of

$$c^{-\frac{1}{2}}(\bar{X}_{t_c} - E(X|\theta))^2 = (c^{\frac{1}{4}}\sum_{i=1}^{t_c}(X_i - E(X|\theta)))^2(\sqrt{c}t_c)^{-2}, \tag{2.6}$$

one obtains the uniform integrability of $\{c^{-\frac{1}{2}}(\bar{X}_{t_c} - E(X|\theta))^2, c > 0\}$ by Lemmas 2.1, 2.2 and 2.3. The moment conditions are needed here.

Combining (2.5) and the uniform integrability of (2.6), we have

$$E(\bar{X}_{t_c} - E(X|\theta))^2 = \sqrt{c}E(\text{Var}(X|\theta))^{\frac{1}{2}} + o(\sqrt{c}). \tag{2.7}$$

Note that the expectation of the distribution function F is $E(\text{Var}(X|\theta))^{\frac{1}{2}}$.

It is easy to see that $\sqrt{c}t_c \longrightarrow (\text{Var}\,(X|\theta))^{\frac{1}{2}}$ $a.s..$ Together with the uniform integrability of $\{\sqrt{c}t_c, c > 0\}$ assured by Lemma 2.1, we have

$$E(ct_c) = \sqrt{c}E(\text{Var}\,(X|\theta))^{\frac{1}{2}} + o(\sqrt{c}). \qquad (2.8)$$

Combining (2.7) and (2.8), we have

$$R(t_c, \bar{X}_{t_c}) = 2\sqrt{c}E(\text{Var}\,(X|\theta))^{\frac{1}{2}} + o(\sqrt{c}).$$

The proof is thus complete.

3. Application to the One-Parameter Exponential Family

Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with a density function of the form

$$f_\theta(x) = \exp(\theta x - A(\theta)), \quad x \in R, \quad \theta \in \Theta,$$

with respect to some σ-finite measure. Here Θ denotes the natural parameter space. It follows that Θ is an interval, finite or infinite, and we assume that $A(\cdot)$ has continuous second derivatives in the interior of Θ and $A'(\cdot) \neq 0$.

Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for each $n \geq 1$. The Bayes risk for a sequential procedure consisting of a stopping rule t and an $\mathcal{F}_t$-measurable function δ_t is $R(t, \delta_t) = E((\delta_t - A'(\theta))^2 + ct)$. We know that for any stopping rule t, the Bayes estimator of $A'(\theta)$ is given by $\delta_t = E(A'(\theta)|\mathcal{F}_t)$. Then the Bayes risk of the sequential procedure $(t, E(A'(\theta)|\mathcal{F}_t))$ is $R_t = E(\text{Var}\,(A'(\theta)|\mathcal{F}_t) + ct)$.

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

$$Z_n = \text{Var}\,(A'(\theta)|\mathcal{F}_n) + cn. \qquad (3.1)$$

Here we are interested in finding a family of stopping rules $\{t(c), c > 0\}$ such that $t(\cdot)$ is asymptotically pointwise optimal (A.P.O.), that is, for any stopping rules $\{s(c), c > 0\}$, $\overline{\lim}_{c\to 0} Z_{t(c)}/Z_{s(c)} \leq 1$ $a.s..$

Let

$$U(c) = \inf\{n \geq 1 : \text{Var}\,(A'(\theta)|\mathcal{F}_n) < nc\}, \quad c > 0.$$

In view of $E_\zeta(\frac{\partial \ln f_\zeta(X)}{\partial \zeta})^2 = 1/A''(\theta)$ where $\zeta = A'(\theta)$, and $E(\text{Var}\,(A'(\theta)|\mathcal{F}_n)) \leq E(A''(\theta))/n$, we know from the results in Bickel and Yahav (1967, 1968) that if $A'(\theta)$ has a continuous bounded density with respect to the Lebesgue measure and $E(X^2)$ is finite, then the stopping rules $\{U(c), c > 0\}$ is A.P.O. with respect to (3.1); it is also asymptotically Bayes, that is,

$$R_{U(c)} = \inf_s R_s + o(\sqrt{c}) = 2\sqrt{c}E(A''(\theta))^{\frac{1}{2}} + o(\sqrt{c}), \qquad (3.2)$$

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

Note that $\text{Var}\,(A'(\theta)|\mathcal{F}_n)$ and $E(A'(\theta)|\mathcal{F}_n)$ depend on the prior distribution of θ, which is sometimes unknown or misspecified. We would thus like to find an alternative procedure, which does not depend on the prior distribution and still possesses the good properties of being A.P.O. and asymptotically Bayes with respect to a large class of priors.

From Bickel and Yahav (1967), $n\text{Var}\,(A'(\theta)|\mathcal{F}_n) \longrightarrow \text{Var}\,_\theta(X) = A''(\theta)$ $a.s.$. In view of the definition of $U(c)$ and the properties of one-parameter exponential family, we propose the sequential procedure $(t_c, \bar{X}_{t_c})$, where t_c is defined in (2.3). It follows directly from Theorem 2.1 and (3.2) that $(t_c, \bar{X}_{t_c})$ is asymptotically Bayes with respect to a large class of prior distributions. More precisely, we have

Theorem 3.1. *Assume that Θ is bounded and closed, and $A'(\theta)$ has a continuous bounded density with respect to the Lebesgue measure. If $E|X|^{3p} < \infty$ for some $p > 1$, and $\delta c^{-1/4} \leq n_c = o(c^{-1/2})$ as $c \to 0$, for some $\delta > 0$, then the Bayes risk of $(t_c, \bar{X}_{t_c})$ for estimating $E_\theta(X) = A'(\theta)$ subject to the loss function (2.1) is*

$$
\begin{aligned}
R(t_c, \bar{X}_{t_c}) &= 2\sqrt{c}E(A''(\theta))^{\frac{1}{2}} + o(\sqrt{c}) \\
&= \inf_s R(s, E(A'(\theta)|\mathcal{F}_s)) + o(\sqrt{c}) \quad as \quad c \to 0,
\end{aligned}
$$

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

Acknowledgements

Appendix

Proof of Lemma 2.1. On $\{t_c \geq n_c+1\}$, we have $\sqrt{c}t_c \leq 2(\sup_{n\geq 2} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2)^{\frac{1}{2}}$, $c > 0$. Hence

$$
(\sqrt{c}t_c)^p \leq O(1) + 2^p \Big(\sup_{n\geq 2} \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2\Big)^{\frac{p}{2}}, \quad c > 0.
$$

The assertion now follows from Doob's inequality and the fact that $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ is a backward martingale.

Proof of Lemma 2.2. Let random variables $\theta, X, X_1, X_2, \ldots$ 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 $\underset{\sim}{X} = (X_1, X_2, \ldots)$ given θ, such that for every $w \in \Omega$ the coordinate random variables $\{\xi_n, n \geq 1\}$ in the induced probability space $(R^\infty, \mathcal{B}^\infty, P_{\underset{\sim}{X}}(\cdot, w))$ are i.i.d..

Let $s_c = \inf\{n \geq n_c : \frac{1}{n}\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2 < n^2 c\}, c > 0$, where $\bar{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$. Note that $E(X|\theta)(w) = E^w \xi_1$ $a.s.$ and $\mathrm{Var}\,(X|\theta)(w) = \mathrm{Var}^w(\xi_1)$ $a.s.$. For almost all $w \in \Omega$, there exists a constant M such that the truncated random variables $\eta_i = (\xi_i \wedge M) \vee (-M)$ satisfy $\mathrm{Var}^w(\eta_1) \in (a,b)$ for some $0 < a < b < \infty$. Conditions (A) and (B) are needed here. Note that $\sum_{i=1}^n (\eta_i - \bar{\eta}_n)^2 \leq \sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2$, where $\bar{\eta}_n = \frac{1}{n}\sum_{i=1}^n \eta_i$.

Define $r_c = \inf\{n \geq n_c : \frac{1}{n}\sum_{i=1}^n (\eta_i - \bar{\eta}_n)^2 < n^2 c\}, c > 0$. For $0 < \gamma < 1$,

$$P_{\underset{\sim}{X}}(\{\sqrt{c}r_c < \gamma\}, w) \geq P_{\underset{\sim}{X}}(\{\sqrt{c}s_c < \gamma\}, w),$$

and we have

$$P_{\underset{\sim}{X}}(\{\sqrt{c}r_c < \gamma\sqrt{\mathrm{Var}^w \eta_1}\}, w)$$

$$\leq P_{\underset{\sim}{X}}(\{\frac{1}{j}\sum_{i=1}^j (\eta_i - \bar{\eta}_j)^2 < j^2 c, \quad \text{for} \quad \text{some} \quad n_c \leq j \leq n_c'\}, w)$$

$$\leq \sum_{j=n_c}^{n_c'} P_{\underset{\sim}{X}}(\{\frac{1}{j}\sum_{i=1}^j (\eta_i - \bar{\eta}_j)^2 < j^2 c\}, w)$$

$$= \sum_{j=n_c}^{n_c'} P_{\underset{\sim}{X}}(\{\frac{1}{j}\sum_{i=1}^j (\mathrm{Var}^w \eta_1 - (\eta_i - E^w \eta_1)^2) + (\frac{1}{j}\sum_{i=1}^j (\eta_i - E^w \eta_1))^2 > \mathrm{Var}^w \eta_1 - j^2 c\}, w)$$

$$\leq \sum_{j=n_c}^{n_c'} P_{\underset{\sim}{X}}(\{\frac{1}{j}\sum_{i=1}^j (\mathrm{Var}^w \eta_1 - (\eta_i - E^w \eta_1)^2) + (\frac{1}{j}\sum_{i=1}^j (\eta_i - E^w \eta_1))^2 > 2\epsilon \mathrm{Var}^w \eta_1\}, w)$$

$$\leq \sum_{j=n_c}^{n_c'} P_{\underset{\sim}{X}}(\{\frac{1}{j}\sum_{i=1}^j (\mathrm{Var}^w \eta_1 - (\eta_i - E^w \eta_1)^2) > \epsilon \mathrm{Var}^w \eta_1\}, w)$$

$$+ \sum_{j=n_c}^{n_c'} P_{\underset{\sim}{X}}(\{(\frac{1}{j}\sum_{i=1}^j (\eta_i - E^w \eta_1))^2 > \epsilon \mathrm{Var}^w \eta_1\}, w), \tag{A.1}$$

where $n_c' = [\gamma\sqrt{\frac{\mathrm{Var}^w \eta_1}{c}}]$ and the second-to-last inequality is due to the fact that for $n_c \leq j \leq n_c'$, $\mathrm{Var}^w \eta_1 - j^2 c \geq \mathrm{Var}^w \eta_1 (1 - \gamma^2) \equiv (\mathrm{Var}^w \eta_1) 2\epsilon$. For convenience, let the first term in (A.1) be denoted by $\sum_{j=n_c}^{n_c'} I_j$ and the second term by $\sum_{j=n_c}^{n_c'} II_j$. Then

$$I_j = P_{\underset{\sim}{X}}\Big(\Big\{\frac{\sum_{i=1}^j (\mathrm{Var}^w \eta_1 - (\eta_i - E^w \eta_1)^2)}{\sqrt{j\mathrm{Var}^w(\eta_1 - E^w \eta_1)^2}} > \frac{j\epsilon \mathrm{Var}^w \eta_1}{\sqrt{j\mathrm{Var}^w(\eta_1 - E^w \eta_1)^2}}\Big\}, w\Big).$$

Let $d_j = ess \, \sup \max_{1 \le i \le j} \left| \frac{\mathrm{Var}^{\,w}\eta_1 - (\eta_i - E^w\eta_1)^2}{\sqrt{j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}} \right|$, which is bounded by

$\frac{b+4M^2}{\sqrt{j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}}$ due to truncation. If $\frac{j\epsilon\mathrm{Var}^{\,w}\eta_1}{\sqrt{j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}}d_j \le 1$, then
by an inequality on p.266 of Loeve (1977),

$$I_j \le \exp\left(\frac{-j^2\epsilon^2(\mathrm{Var}^{\,w}\eta_1)^2}{4j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}\right) \le \exp\left(-\frac{j\epsilon^2 a^2}{64M^4}\right). \qquad (A.2)$$

If $\frac{j\epsilon\mathrm{Var}^{\,w}\eta_1}{\sqrt{j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}}d_j \ge 1$, then

$$I_j \le \exp\left(-\frac{j\epsilon\mathrm{Var}^{\,w}\eta_1}{4\sqrt{j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}}\frac{\sqrt{j\mathrm{Var}^{\,w}(\eta_1 - E^w\eta_1)^2}}{b+4M^2}\right)$$
$$\le \exp\left(\frac{-j\epsilon a}{4(b+4M^2)}\right). \qquad (A.3)$$

Combining (A.2) and (A.3), we have

$$I_j \le \exp(-k_1 j), \qquad (A.4)$$

where $k_1 = \min\{\frac{\epsilon^2 a^2}{64M^4}, \frac{\epsilon a}{4(b+4M^2)}\}$. As for the second term,

$$II_j = P_{\underline{x}}\left(\left\{\sum_{i=1}^{j}(\eta_i - E^w\eta_1) > j\sqrt{\epsilon\mathrm{Var}^{\,w}\eta_1}\right\}, w\right)$$
$$+P_{\underline{x}}\left(\left\{\sum_{i=1}^{j}(-\eta_i + E^w\eta_1) > j\sqrt{\epsilon\mathrm{Var}^{\,w}\eta_1}\right\}, w\right)$$
$$\le \exp(-k_2 j) + \exp(-k_2 j) \qquad (A.5)$$

where $k_2 = \min\{\frac{\epsilon}{4}, \frac{\sqrt{\epsilon a}}{8M}\}$. The inequality in (A.5) follows from a similar argument
as for (A.4).

Let $k = \min\{k_1, k_2\}$. Then combining (A.1), (A.4) and (A.5),

$$P_{\underline{x}}\left(\{\sqrt{c}r_c < \gamma\sqrt{\mathrm{Var}^{\,w}\eta_1}\}, w\right) \le \sum_{j=n_c}^{n'_c} 3\exp(-kj)$$
$$\le \sum_{j=n_c}^{n'_c} 3\exp(-k\delta c^{-\frac{1}{4}}) \le 3\gamma\sqrt{\frac{\mathrm{Var}^{\,w}\eta_1}{c}}\exp(-k\delta c^{-\frac{1}{4}})$$
$$\le 3\gamma\sqrt{b}c^{-\frac{1}{2}}\exp(-k\delta c^{-\frac{1}{4}}). \qquad (A.6)$$

The second and the third inequalities are from the definition of n_c and n'_c respectively.

In order to obtain the uniform integrability of $\{(\sqrt{c}t_c)^{-p}, c > 0\}$, we need to show that for $0 < \gamma < 1$, $P\{\sqrt{c}t_c < \gamma\} = o(c^{\frac{p}{2}})$ (see Lemma 1 of Chow and Yu (1981)).

$$P\{\sqrt{c}t_c < \gamma\sqrt{a}\} = \int_\Omega P_{\underset{\sim}{x}}(\{\sqrt{c}s_c < \gamma\sqrt{a}\}, w)dP(w)$$

$$\leq \int_\Omega P_{\underset{\sim}{x}}(\{\sqrt{c}s_c < \gamma\sqrt{\operatorname{Var}{}^w\eta_1}\}, w)dP(w)$$

$$\leq \int_\Omega P_{\underset{\sim}{x}}(\{\sqrt{c}r_c < \gamma\sqrt{\operatorname{Var}{}^w\eta_1}\}, w)dP(w)$$

$$\leq O(1)\int_\Omega c^{-\frac{1}{2}}\exp(-k\delta c^{-\frac{1}{4}})dP(w) = o(c^{\frac{p}{2}}). \tag{A.7}$$

The last inequality in (A.7) follows from (A.6). The proof is thus complete.

Proof of Lemma 2.3. For convenience, let $W_n(\theta) = h(X_n, \theta) - E(h(X, \theta)|\theta)$. Let $s'_c = s_c \wedge N$, where $N = [Kc^{-\frac{1}{2}}]$ for $K \geq 1$. For any $\delta > 0$, there exists a positive constant K_1 such that $E|Y_n|^p < \delta$ for all $n \geq 1$, where $Y_n = W_n(\theta)1_{\{|W_n(\theta)|\geq K_1\}} - E(W_n(\theta)1_{\{|W_n(\theta)|\geq K_1\}}|\theta)$. Put $\mathcal{F}_o = \sigma(\theta)$, $\mathcal{F}_n = \sigma(\theta, X_1, \ldots, X_n)$, $n \geq 1$, and

$$Z_n = W_n(\theta) - Y_n = W_n(\theta)1_{\{|W_n(\theta)|<K_1\}} - E(W_n(\theta)1_{\{|W_n(\theta)|<K_1\}}|\theta).$$

Note that $E(Y_{n+1}|\mathcal{F}_n) = E(Y_{n+1}|\theta) = 0$ for all $n \geq 0$, and s_c and s'_c are $\sigma(\theta, X_1, \ldots, X_n)$ - stopping rules. Then $\{\sum_{i=1}^n Y_i 1_{\{s'_c \geq i\}}, n \geq 1\}$, $\{\sum_{i=1}^n Z_i 1_{\{s'_c \geq i\}}, n \geq 1\}$ and $\{\sum_{i=N+1}^n W_i(\theta)1_{\{s_c \geq i\}}, n \geq N+1\}$ are $\mathcal{F}_n$-martingales. Using Burkholder, Davis and Gundy's inequality (1972) (see p.409 of Chow and Teicher (1988)) and Jensen's inequality, we have for some constant $A > 0$,

$$E|\sum_{i=1}^{s'_c} Y_i|^p \leq AE(\sum_{i=1}^\infty E(Y_i^2 1_{\{s'_c \geq i\}}|\mathcal{F}_{i-1}))^{p/2} + AE(\sup_{i\geq 1}|Y_i|1_{\{s'_c \geq i\}})^p$$

$$\leq AE(\sum_{i=1}^\infty 1_{\{s'_c \geq i\}}E(Y_i^2|\theta))^{p/2} + AE(\sum_{i=1}^N |Y_i|^p 1_{\{s'_c \geq i\}})$$

$$= AE(s'_c E(Y_1^2|\theta))^{p/2} + AE(\sum_{i=1}^N 1_{\{s'_c \geq i\}}E(|Y_i|^p|\mathcal{F}_{i-1}))$$

$$\leq AE((s'_c)^{p/2}E(|Y_1|^p|\theta)) + AE(s'_c E(|Y_1|^p|\theta))$$

$$\leq 2AN^{p/2}E|Y_1|^p \leq 2AN^{p/2}\delta.$$

Hence

$$\sup_{c>0} E|c^{1/4}\sum_{i=1}^{s'_c} Y_i|^p \leq 2AK^{p/2}\delta = o(1) \quad \text{as} \quad \delta \to 0. \tag{A.8}$$

Similarly,

$$E|\sum_{i=1}^{s_c'} Z_i|^{p+1} \le AE(\sum_{i=1}^{\infty} E(Z_i^2 1_{\{s_c'\ge i\}}|\mathcal{F}_{i-1}))^{\frac{p+1}{2}} + AE(\sup_{i\ge 1}|Z_i|1_{\{s_c'\ge i\}})^{p+1}$$

$$\le 2AE((s_c')^{\frac{p+1}{2}}E(|Z_1|^{p+1}|\theta)) \le 2AN^{\frac{p+1}{2}}E|Z_1|^{p+1}.$$

Then

$$\sup_{c>0} E|c^{1/4}\sum_{i=1}^{s_c'} Z_i|^{p+1} \le 2A(2K_1)^{p+1}K^{\frac{p+1}{2}} < \infty. \qquad (A.9)$$

From (A.8) and (A.9), $|c^{1/4}\sum_{i=1}^{s_c'} W_i(\theta)|^p$ is uniformly integrable. Similar to (A.8), by Burkholder, Davis and Gundy's (1972), Jensen's and Hölder's inequalities, we have

$$E|\sum_{i=N+1}^{\infty} W_i(\theta)1_{\{s_c\ge i\}}|^p \le 2AE((\sum_{i=N+1}^{\infty} 1_{\{s_c\ge i\}})^{p/2}E(|W_1(\theta)|^p|\theta))$$

$$\le 2AE^{1/a}|W_1(\theta)|^{pa}E^{1/b}|\sum_{i=N+1}^{\infty} 1_{\{s_c\ge i\}}|^{pb/2}.$$

Theorefore,

$$\sup_{c>0} E|c^{1/4}\sum_{i=N+1}^{\infty} W_i(\theta)1_{\{s_c\ge i\}}|^p$$

$$\le 2AE^{1/a}|W_1(\theta)|^{pa}(\sup_{c>0}\int_{\{\sqrt{c}s_c>K\}}(\sqrt{c}s_c)^{pb/2}dP)^{1/b}$$

$$= o(1) \quad \text{as} \quad K\to\infty. \qquad (A.10)$$

The last equality follows from the assumptions. Write

$$\sum_{i=1}^{s_c} W_i(\theta) = \sum_{i=1}^{s_c'} W_i(\theta) + \sum_{i=N+1}^{\infty} W_i(\theta)1_{\{s_c\ge i\}} \qquad (A.11)$$

The lemma now follows from (A.10) and (A.11).

References

Alvo, M. (1977). Bayesian sequential estimation. *Ann. Statist.* **5**, 955-968.

Bickel, P. J. and Yahav, J. A. (1967). Asymptotically pointwise optimal procedures in sequential analysis. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **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.

Burkholder, D. L., Davis, B. J. and Gundy, R. F. (1972). Integral inequalities for convex
    functions of operators on martingales. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*
    2, 223-240. University of California Press.

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

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

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

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

Loeve, M. (1977). *Probability Theory I.* 4th edition. Springer-Verlag, Berlin.

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

Rasmussen, S. L. (1980). A Bayesian approach to a problem in sequential estimation. *Ann.
    Statist.* **8**, 1229-1234.

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.

Department of Mathematics, Tamkang University, Tamsui, Taipei, Taiwan

E-mail: lchwang@math.tku.edu.tw