

## ON THE PROPERTIES OF STATISTICAL SEQUENTIAL DECISION PROCEDURES

MASAFUMI AKAHIRA AND KEN-ICHI KOIKE

### 1. INTRODUCTION

In statistical inference, we consider estimation problems of parameters based on the sample of previously fixed size in most cases. However, it seems to be more practical to introduce an appropriate stopping rule without fixing the size of a sample in advance and to treat sequential estimation procedures according to a sequential sampling plan based on the rule. In this paper we discuss various properties including the optimality of sequential decision procedures in statistical inference. First, in the nonsequential case, the concepts of sufficiency, completeness, etc. are known to be important to obtain the best estimator, and they are also useful for deriving the (uniformly) minimum variance unbiased sequential estimation procedure, etc. in the sequential sampling plan. For the case of Bernoulli trials in the sequential sampling plan, the geometrical necessary and sufficient conditions for completeness are given by Girshick et al. [GMS46] and Lehmann and Stein [LS50]. For the case of multinomial trials, such conditions for completeness as an unsolved problem are also adopted by Linnik and Romanovsky [LR72], and a much stronger sufficient condition is given by Kremers [Kr90]. In this paper we provide a comparatively weak sufficient condition for completeness which becomes a necessary condition in the case of Bernoulli trials. Next, in the case of sequential binomial sampling, a sufficient condition for a stopping rule to be closed, i.e., to stop at finite steps with probability 1, is given in [GMS46] and Wolfowitz [Wo46], and here, in the case of sequential multinomial sampling, the condition is extended to that of [Wo46] up to the equivalent order. In a sequential unbiased estimation problem on a function of probability  $p$  of success in the Bernoulli trials, there is a contention that an unbiasedly estimable parameter must be continuous in  $p$ , but it is shown that it cannot be valid. We also discuss a sequential unbiased estimation in the case of multinomial trials.

In the nonsequential case, it is known that the Cramér-Rao inequality plays an important part in obtaining the efficient estimator. The inequality is extended by Wolfowitz [Wo47] to the sequential case, and a sequential estimation procedure attaining the lower bound by the inequality is said to be efficient. However, it is rare that such an efficient sequential estimation procedure exists, and it is seen that the lower bound by the Wolfowitz inequality is not attainable in most cases ([Gh87], [St90]). This seems to be quite different from the fact that there exist efficient estimators for an exponential family of distributions in the nonsequential

---

This article originally appeared in Japanese in *Sūgaku* **48** (2) (1996), 184–195.  
1991 *Mathematics Subject Classification*. Primary 62L12; Secondary 62F10, 62F12.

case. In this paper, for the multinomial case in the sequential sampling plan we discuss which sequential estimation procedure consisting of a stopping rule and an unbiased estimator attains the lower bound. Wasan [Wa64] also considered some sequential estimation problems in consideration of the size of a sample in the case of Bernoulli trials, and we here extend the case to that of multinomial trials. Further we show a sequential estimation procedure to be admissible without imposing unbiasedness on it.

Finally, we discuss asymptotically optimum sequential estimation procedures since it is generally difficult to get the optimal fixed size of a sample in the sequential case.

## 2. SEQUENTIAL DECISION PROCEDURES IN THE MULTINOMIAL SAMPLING

First, suppose that  $X^{(1)}, X^{(2)}, \dots$  is a sequence of independent and identically distributed  $k$ -dimensional multinomial trials, that is, for each  $i = 1, 2, \dots$ ,  $X^{(i)} := (X_1^{(i)}, \dots, X_k^{(i)})$  is a random vector with  $X_j^{(i)} \in \{0, 1\}$  ( $j = 1, \dots, k$ ),  $\sum_{j=1}^k X_j^{(i)} = 1$ , and, for  $\theta = (\theta_1, \dots, \theta_{k-1})$  with  $0 < \theta_j < 1$  ( $j = 1, \dots, k-1$ ) and  $\sum_{j=1}^{k-1} \theta_j < 1$ ,

$$P_\theta \left\{ X_j^{(i)} = 1 \right\} = \theta_j \quad (j = 1, \dots, k),$$

where  $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$ .

For the above sequence  $\{X^{(n)}\}$  in the sequential sampling, a decision whether or not to sample  $X^{(n+1)}$  is based upon  $X^{(1)}, \dots, X^{(n)}$  for each positive integer  $n$ . Then, the size of the sample may be a random variable specified by a sampling plan under consideration. It is often denoted by  $N$ . Since the random vector  $Y^{(N)} := \sum_{i=1}^N X^{(i)}$  is a sufficient statistic for  $\theta$ , it is enough to consider only estimators based upon  $Y^{(N)}$  ([F67]). Now we define the stopping rule  $\varphi$  as

$$\varphi(z) = (\varphi_0, \varphi_1(x^{(1)}), \varphi_2(x^{(1)}, x^{(2)}), \dots),$$

where  $z = (x^{(1)}, x^{(2)}, \dots)$  and, for each  $j = 1, 2, \dots$ ,  $\varphi_j$  is defined on the sample space of  $(X^{(1)}, \dots, X^{(j)})$  and  $0 \leq \varphi_j \leq 1$  for all  $j = 0, 1, \dots$ . For each  $j = 1, 2, \dots$ , the function  $\varphi_j(x^{(1)}, \dots, x^{(j)})$  represents the conditional probability that a statistician stops sampling, given that he has taken  $X^{(1)} = x^{(1)}, \dots, X^{(j)} = x^{(j)}$ , and also  $\varphi_0$  is a constant representing the probability of taking no observations at all. Henceforth we often denote the stopping rule by the size  $N$  of the sample instead of  $\varphi$  and call the pair  $(N, Y^{(N)})$  a sequential decision procedure.

Here, to avoid the case when the sampling continues forever, we assume that the sequential decision procedure is closed, that is,  $P_\theta(N < \infty) = 1$  for all  $\theta$ . In the above model, the outcome of such  $Y^{(N)}$  can be represented as a random walk that starts from the origin in the  $k$ -fold direct products  $\mathbf{N}_0^k$  of a set of all the nonnegative integers. For a given stopping rule  $\varphi$ , the probability mass function of  $Y^{(N)}$  is given by

$$P_\theta \left\{ Y^{(N)} = y \right\} = c(y) \prod_{j=1}^k \theta_j^{y_j},$$

where  $0 \leq c(y) \leq \left( \sum_{j=1}^k y_j \right)! / \prod_{j=1}^k y_j!$  with  $y = (y_1, \dots, y_k)$ . In the non-sequential case, it is well known how to obtain the uniformly minimum variance

unbiased estimator based on the complete sufficient statistic, and it can be extended to the sequential case. So, if  $Y^{(N)}$  is complete as a statistic, that is,  $E_\theta[g(Y^{(N)})] = 0$  implies  $g(Y^{(N)}) = 0$  a.e.  $P_\theta$  for all  $\theta \in \Theta$ , the sequential decision procedure  $(N, Y^{(N)})$  is said to be *complete*. In the next section we shall discuss (necessary and) sufficient conditions for the completeness.

### 3. COMPLETENESS OF SEQUENTIAL DECISION PROCEDURES

First we consider a non-randomized stopping rule based only on the sequence  $\{Y^{(N)}\}$  of the sufficient statistics, that is, in the above  $\varphi$ , each  $\varphi_n$  takes on only the values 0 or 1 depending only on  $y^{(n)}$ . In this section we restrict attention to the space  $\mathbf{N}_0^k$  and use the following terminologies. (i) For a point  $y = (y_1, \dots, y_k) \in \mathbf{N}_0^k$ ,  $\sum_{j=1}^k y_j$  is called *an index of  $y$* . (ii) A point  $y$  is said to be *accessible* if  $P_\theta\{Y^{(m)} = y, N \geq m\} > 0$ . (iii) A point  $y$  is called a *continuation point* if it is accessible and  $\varphi_m(y) = 0$ , and a set of all the points is called a *continuation region* (see Figure 3.1). A sequential decision procedure is said to be *bounded* if there exists some positive constant  $c$  such that, for any accessible point  $y = (y_1, \dots, y_k)$ ,  $\sum_{j=1}^k y_j \leq c$ , and also *simple* if the convex hull of the continuation region on each index contains no points except for continuation points (see Figures 3.2 and 3.3 on p. 200).

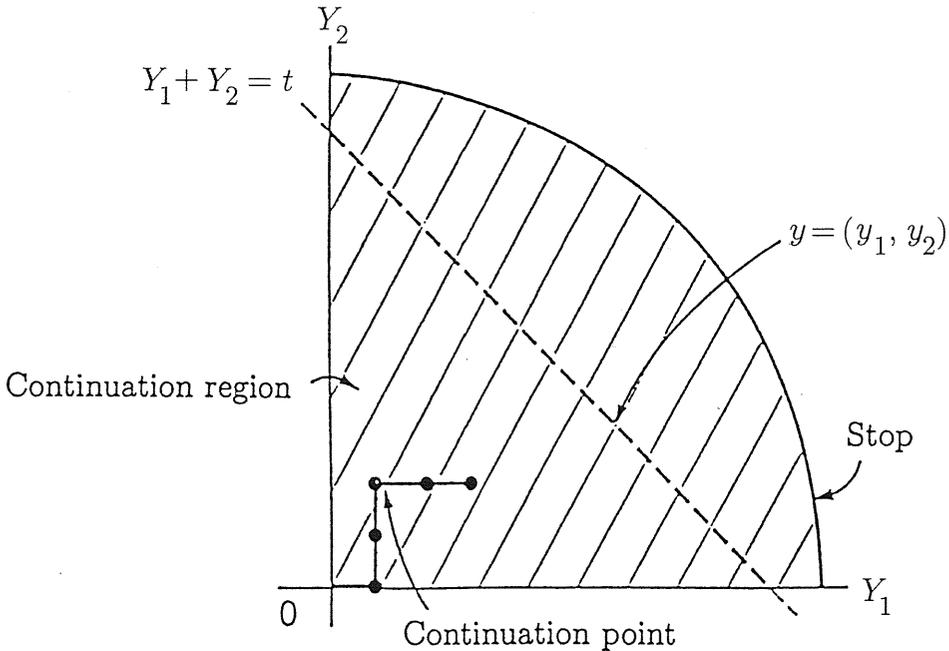


FIGURE 3.1. The case when  $k = 2$  and  $t$  is the index of a point  $y = (y_1, y_2)$

Then we have the following result on the completeness of a sequential decision procedure ([KoA93c]).

**Theorem 3.1.** *A bounded and simple sequential decision procedure is complete.*

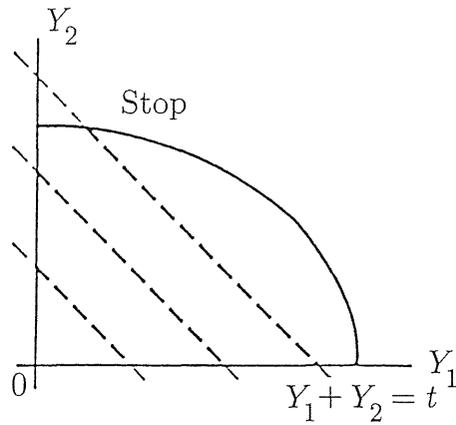


FIGURE 3.2. The case when the sequential decision procedure is simple

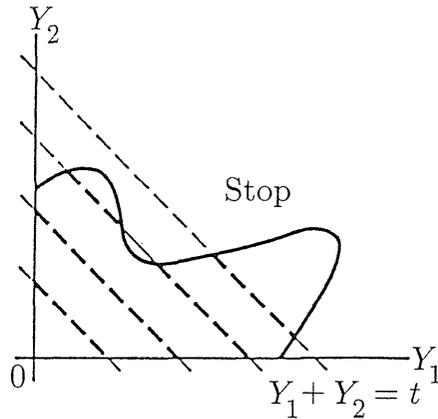


FIGURE 3.3. The case when the sequential decision procedure is not simple

In relation to the above theorem, it is known that, in the sequential binomial case, the sequential decision procedure is complete if and only if it is simple ([GMS46], [LS50]). As the other sufficient condition for completeness in the sequential multinomial case, Kremers [Kr90] states that the sequential decision procedure is bounded and the convex hull of all the continuation regions contains no points except for continuation points. The above theorem also becomes an answer to one of the unsolved problems proposed by Linnik and Romanovsky [LR72]. Some necessary conditions for the sequential decision procedure to be complete are given in [LR72] and Koike and Akahira [KoA93b].

Next we consider a randomized stopping rule. Let  $\{q_n\}$  be a sequence of constants such that  $q_n \geq 0$  ( $n = 0, 1, 2, \dots$ ) and  $\sum_{n=0}^{\infty} q_n = 1$ . For a stopping rule  $\varphi$  in the previous section we assume that  $\varphi_n = q_n$  for each nonnegative integer  $n$ . Then the stopping rule is randomized. We also have for any estimator  $f(Y^{(N)})$  based on

the sufficient statistic  $Y^{(N)}$

$$E_{\theta} \left[ f(Y^{(N)}) \right] = \sum_{n=0}^{\infty} \sum_{y_1 + \dots + y_k = n} q_n n! f(y_1, \dots, y_k) \prod_{j=1}^k \theta_j^{y_j} / y_j!,$$

where  $\sum_{y_1 + \dots + y_k = n}$  means that we take the sum with respect to all possible combinations with  $\sum_{j=1}^k y_j = n$ . Then we have the following ([KoA93c]).

**Theorem 3.2.** *Under the above randomized stopping rule, a necessary and sufficient condition for the sequential decision procedure to be complete is that there exists a unique nonnegative integer  $n$  satisfying  $q_n = 1$ .*

For example, the above result may be applied to the following estimation problem. Let  $\delta = (\delta_0, \delta_1, \dots, \delta_n, \dots)$  be an unbiased estimator of a function  $g(\theta)$  of  $\theta$ . Under the condition that  $\sum_{n=0}^{\infty} E_{\theta}[q_n \delta_n] = g(\theta)$  for all  $\theta$ , it is seen to be desirable to minimize  $\sum_{n=0}^{\infty} E_{\theta}[q_n \{\delta_n - g(\theta)\}^2]$ . Now, since, for each  $n$ ,  $\varphi_n$  is independent of  $Y^{(N)}$ , it is enough to obtain  $\delta_n$  minimizing  $E_{\theta}[\{\delta_n - g(\theta)\}^2]$  for each  $n$  under the above condition. Hence this problem can be treated in a way similar to the nonsequential case. Also, an advantage of the necessary and sufficient condition of Theorem 3.2 is to be able to check it easily.

#### 4. CLOSEDNESS FOR STOPPING RULES AND UNBIASED ESTIMATION

First we consider the non-randomized stopping rule based only on the sufficient statistic  $Y^{(N)}$ . Then it follows from the definition that any bounded sequential decision procedure is closed. Now, as a sufficient condition for an unbounded sequential decision procedure to be closed, we have the following ([Ko93]).

**Theorem 4.1.** *A sufficient condition for a sequential decision procedure to be closed is that*

$$\liminf_{n \rightarrow \infty} A(n) / n^{(k-1)/2} < \infty,$$

where  $A(n)$  is the number of accessible points of the index  $n$ .

It is noted that the condition of Theorem 4.1 for the case  $k = 2$ , i.e., the binomial case, coincides with that of Wolfowitz [Wo46]. It is also shown by Sato [Sa95] that the condition is the best in some sense but not necessary. Applying the Rao-Blackwell method we have the following ([Ko93]).

**Theorem 4.2.** *For any closed sequential decision procedure,*

$$\delta_j(y) = \frac{c_j(y)}{c(y)}$$

is an unbiased estimator of  $\theta_j$  for  $j = 1, \dots, k$ , where, for each  $j$ ,  $\delta_j(y)$  is defined on the sample space of  $Y^{(N)}$  and  $c_j(y)$  denotes the number of paths from the point  $(0, \dots, 0, \underset{j}{1}, 0, \dots, 0)$  with 1 in the only  $j$ -th component to  $y$ .

**Corollary 4.1.** For any closed sequential decision procedure and its accessible point  $(a_1, \dots, a_k)$ ,

$$\hat{\delta}(y) = \frac{c_0(y)}{c(y)}$$

is an unbiased estimator of  $\prod_{j=1}^k \theta_j^{a_j}$ , where  $\hat{\delta}$  is defined on the sample space of  $Y^{(N)}$  and  $c_0(y)$  denotes the number of paths that pass through  $(a_1, \dots, a_k)$  and terminate in  $y$ .

We apply Theorem 4.2 to some stopping rules and can practically obtain the unbiased estimator ([Ko93]).

**Example 4.1.** We extend the stopping rule of Kremers [Kr87] to the  $k$ -dimensional case. For positive integers  $n_1, n_2$  ( $n_1 < n_2$ ),  $m_1, \dots, m_k$ , let the stopping rule  $N$  be

$$N = \min_r \left\{ r \geq n_1 : \left( Y_1^{(r)} \geq m_1 \text{ or } \dots \text{ or } Y_k^{(r)} \geq m_k \right) \text{ or } r \geq n_2 \right\}$$

(see Figure 4.1).

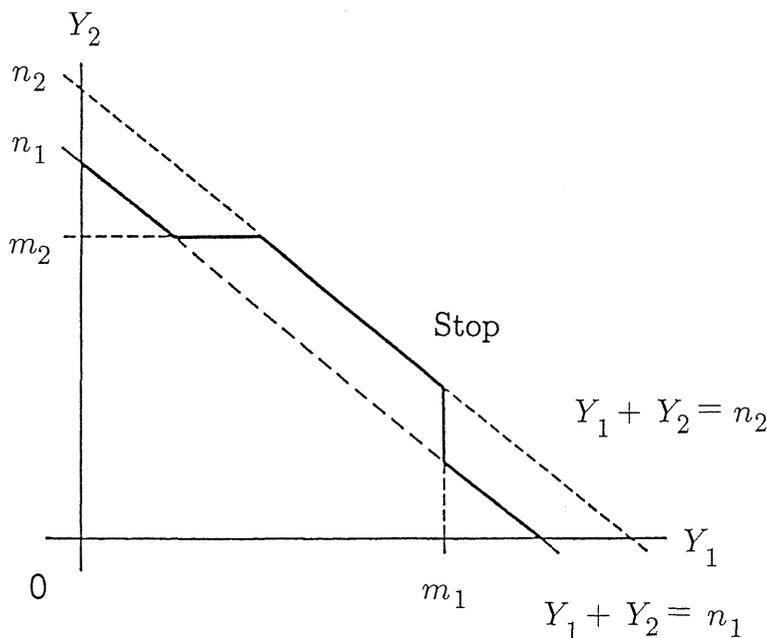


FIGURE 4.1. The stopping rule  $N$  in the case when  $k = 2$  ([Kr87])

Then it is easily seen that, for each  $j = 1, \dots, k$ , the  $\delta_j$  of Theorem 4.2 is given by

$$\delta_j(Y^{(N)}) = \begin{cases} \frac{Y_j^{(N)}}{N}, & \text{if } N = n_1 \text{ or} \\ & N = n_2 \text{ and } (Y_1^{(N)} < m_1 \text{ and } \dots \text{ and } Y_k^{(N)} < m_k), \\ \frac{Y_j^{(N)} - X_j^{(N)}}{N-1}, & \text{otherwise.} \end{cases}$$

Since, by Theorem 3.1, the sequential decision procedure is complete, it is seen that, for each  $j$ , the above  $\delta_j$  is the uniformly minimum variance unbiased estimator of  $\theta_j$ .

5. UNBIASED ESTIMATION FOR THE SEQUENTIAL BINOMIAL SAMPLING

We consider the randomized stopping rule in §3. That is, let  $\{q_n\}$  be a sequence of constants satisfying  $q_n \geq 0$  ( $n = 0, 1, 2, \dots$ ) and  $\sum_{n=0}^{\infty} q_n = 1$ , and for a stopping rule  $\varphi$  in §2 assume that  $\varphi_n = q_n$  for each nonnegative integer  $n$ . Then we consider the unbiased estimability of a parametric function in the sequential binomial sampling. In the nonsequential case, one can treat this with the problem of estimating a function  $g(p)$  of  $p$  based on  $n$  independent Bernoulli trials with a probability  $p$  of success. As a necessary and sufficient condition for  $g(p)$  to be unbiasedly estimable, it is known that  $g(p)$  is a polynomial of degree equal to or less than  $n$  ([L83]). Recently there was a contention that the estimable parameter must be continuous in  $p$  ([BhBo90]), but in this section it is shown that the claim cannot be valid.

In the setup of §2, we consider the case when  $k = 2$  and  $\theta_1 = p$ . Define an estimator  $e$  of a function  $g(p)$  of  $p$  by a real-valued function defined on the sample space of  $Y^{(N)}$ . Since  $\theta_2 = 1 - p$ , it follows that

$$E_p \left[ e(Y^{(N)}) \right] = \sum_{n=0}^{\infty} q_n \sum_{x+y=n} e(x, y) \binom{n}{x} p^x (1-p)^y.$$

If there exists an estimator  $e$  such that for any  $p$

$$\begin{aligned} \sum_{x,y} |e(x, y)| P_p \left\{ Y^{(N)} = (x, y) \right\} < \infty, \\ \sum_{x,y} e(x, y) P_p \left\{ Y^{(N)} = (x, y) \right\} = g(p), \end{aligned}$$

then the parametric function  $g(p)$  is said to be *unbiasedly estimable*. Now we assume that there is a sequence  $\{\hat{g}_n(x, y)\}$  of functions such that for any  $p$  with  $0 < p < 1$

$$\lim_{n \rightarrow \infty} \sum_{x+y=n} \hat{g}_n(x, y) \binom{n}{x} p^x (1-p)^y = g(p),$$

where  $\hat{g}_n(x, y)$  is defined over  $\{(x, y) | x + y = n, 0 \leq x \leq n\}$  for  $n = 0, 1, 2, \dots$ . Then, letting  $\hat{g}_0(0, 0) = \hat{g}_{n-1}(-1, n) = \hat{g}_{n-1}(n, -1) = 0$  for  $n = 1, 2, \dots$ , we define an estimator  $e^*$  of  $g(p)$  as

$$\begin{aligned} e^*(x, y) &= \left\{ \hat{g}_n(x, y) - \frac{x}{n} \hat{g}_{n-1}(x-1, y) - \frac{y}{n} \hat{g}_{n-1}(x, y-1) \right\} / q_n \\ &\qquad\qquad\qquad (n = x + y = 1, 2, \dots), \\ e^*(0, 0) &= 0. \end{aligned}$$

Then we have the following ([ATK92]).

**Theorem 5.1.** *If*

$$\sum_{n=0}^{\infty} q_n \sum_{x+y=n} |e^*(x, y)| \binom{n}{x} p^x (1-p)^y < \infty$$

for all  $p$  ( $0 < p < 1$ ), then  $g(p)$  is unbiasedly estimable.

Using Theorem 5.1, we can construct a discontinuous function  $g(p)$  that is unbiasedly estimable.

**Example 5.1.** Define a function  $g(p)$  by

$$g(p) = \begin{cases} 1 & \text{for } p > 1/2, \\ 0 & \text{for } p = 1/2, \\ -1 & \text{for } p < 1/2, \end{cases}$$

and also a function  $\hat{g}_n(x, y)$  by

$$\hat{g}_n(x, y) = \begin{cases} (1/a_n)(x - y) & \text{for } |x - y| \leq a_n, \\ 1 & \text{for } x - y > a_n, \\ -1 & \text{for } x - y < -a_n, \end{cases}$$

where  $\{a_n\}$  is an increasing sequence of positive numbers in  $n$  such that

$$\lim_{n \rightarrow \infty} (a_n/n^\gamma) = c (> 0) \quad \text{for } 1/2 < \gamma < 1.$$

Constructing the estimator  $e^*(x, y)$  based on  $\hat{g}_n(x, y)$  as is stated in the above, we can verify that the condition of Theorem 5.1 is satisfied; hence the function  $g(p)$  is unbiasedly estimable.

Further, using an estimator depending on the path, we can get a similar result to the above in the case of the nonrandomized stopping rule ([ATK92]).

## 6. INFORMATION INEQUALITIES AND EFFICIENCIES OF SEQUENTIAL ESTIMATION PROCEDURES BASED ON THE MULTINOMIAL TRIALS

In this section we consider an extension of the Cramér-Rao type lower bound for sequential estimation procedures in the multiparameter case and also discuss efficiency in the sense of attaining the lower bound in the estimation problem of  $\theta$  for the multinomial trials in §2 ([KoA94]). We assume the following regularity conditions (A1)–(A7).

- (A1) A parameter space  $\Omega$  is an open subset of the Euclidean  $r$ -space  $\mathbf{R}^r$  and an element of  $\Omega$  is denoted by  $\theta := (\theta_1, \dots, \theta_r)'$ . For each  $i = 1, \dots, k$ , let  $g_i(\theta)$  be a real-valued partially differentiable function with respect to  $\theta_j$  for  $j = 1, \dots, r$  and  $g(\theta) := (g_1(\theta), \dots, g_k(\theta))'$ . Let  $h_{ij}(\theta) = \partial g_i(\theta) / \partial \theta_j$  ( $i = 1, \dots, k; j = 1, \dots, r$ ) and  $H(\theta) = \{h_{ij}(\theta)\}$ .
- (A2) A family of probability measures  $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$  defined on a sample space  $(\mathcal{X}, \mathcal{A})$  is dominated by a  $\sigma$ -finite measure  $\mu$  and for each  $\theta \in \Omega$  its density function  $dP_\theta/d\mu$  is denoted by  $p(\cdot, \theta)$ . Let  $X_1, X_2, \dots$  be a sequence

of independent and identically distributed random variables according to a distribution  $P_\theta$ , where  $\theta \in \Omega$ .

- (A3) The support  $\{x : p(x, \theta) > 0\}$  of  $p(\cdot, \theta)$  is independent of  $\theta$ .
- (A4) For a. a.  $x[\mu]$ ,  $p(x, \theta)$  is partially differentiable with respect to  $\theta_j$  for  $j = 1, \dots, r$ .
- (A5) The size  $N$  of the sample is a random variable satisfying  $0 < E_\theta(N) < \infty$  for all  $\theta \in \Omega$ .
- (A6) Let  $\varphi^{(n)}(X_1, \dots, X_n) = (\varphi_1^{(n)}(X_1, \dots, X_n), \dots, \varphi_k^{(n)}(X_1, \dots, X_n))'$  for  $n = 1, 2, \dots$ . Then, for an unbiased estimator  $\varphi = (\varphi^{(1)}, \varphi^{(2)}, \dots)$  of  $g(\theta)$ , the partial derivatives with respect to  $\theta_j$  of the left-hand sides of

$$\sum_{n=1}^{\infty} \int_{\{N=n\}} \varphi^{(n)}(x_1, \dots, x_n) \prod_{l=1}^n p(x_l, \theta) \prod_{l=1}^n \mu(dx_l) = g(\theta),$$

$$\sum_{n=1}^{\infty} \int_{\{N=n\}} \prod_{l=1}^n p(x_l, \theta) \prod_{l=1}^n \mu(dx_l) = 1,$$

can be obtained by differentiating under the integral sign for  $j = 1, \dots, r$ .

- (A7) Let

$$I_{ij}(\theta) = E_\theta \left[ \frac{\partial \log p(X_1, \theta)}{\partial \theta_i} \frac{\partial \log p(X_1, \theta)}{\partial \theta_j} \right]$$

for  $i, j = 1, \dots, r$ . Then  $I_{ii}$  is finite for  $i = 1, \dots, r$  and the Fisher information matrix  $I(\theta) = \{I_{ij}(\theta)\}$  is a positive definite symmetric matrix for all  $\theta \in \Omega$ .

Under the regularity conditions we have the following ([Ko96]).

**Theorem 6.1.** *Assume that the conditions (A1)–(A7) hold. Then the matrix*

$$\text{Cov}_\theta(\varphi) - H(\theta) \{E_\theta(N)I(\theta)\}^{-1} H(\theta)'$$

is positive semidefinite for all  $\theta \in \Omega$ , where  $\text{Cov}_\theta(\varphi)$  denotes the covariance matrix of  $\varphi$ . In addition, if  $r = k$  and  $H(\theta)$  is nonsingular, then for the generalized variance  $|\text{Cov}_\theta(\varphi)|$ ,

$$|\text{Cov}_\theta(\varphi)| \geq \frac{|H(\theta)|^2}{|E_\theta(N)I(\theta)|}.$$

Equality holds in the above if and only if

$$\text{Cov}_\theta(\varphi) = H(\theta) \{E_\theta(N)I(\theta)\}^{-1} H(\theta)'.$$

In this case, for any  $\theta \in \Omega$  and  $i = 1, \dots, r$ ,

$$\varphi_i^{(N)}(X_1, \dots, X_N) = \sum_{j=1}^k a_j^{(i)}(\theta) \left\{ \sum_{l=1}^N \frac{\partial}{\partial \theta_j} \log p(X_l, \theta) + b_j^{(i)}(\theta) \right\} \quad P_\theta\text{-a.s.},$$

where  $a_j^{(i)}(\theta)$  and  $b_j^{(i)}(\theta)$  are the functions depending only on  $\theta$  for  $i = 1, \dots, r$  and  $j = 1, \dots, k$ .

In particular, in the case of  $k = 1$ , the above inequality is called the Wolfowitz inequality.

**Corollary 6.1.** *Assume that the conditions (A1)–(A7) hold. Let*

$$g(\theta) = \theta + b(\theta), \quad b(\theta) = (b_1(\theta), \dots, b_k(\theta))',$$

$$B(\theta) = \{B_{ij}(\theta)\}, \quad B_{ij}(\theta) = \frac{\partial b_i(\theta)}{\partial \theta_j} \quad (i, j = 1, \dots, k).$$

*If the loss function is given by  $W(\theta, d) = (d - \theta)'(d - \theta)$ , then the risk function  $R(\theta, \varphi) := E_\theta[W(\theta, \varphi)]$  of  $\varphi$  satisfies the following inequality:*

$$R(\theta, \varphi) \geq \text{tr} \left[ b(\theta)b(\theta)' + (\mathbf{E}_k + B(\theta)) \{E_\theta(N)I(\theta)\}^{-1} (\mathbf{E}_k + B(\theta)') \right],$$

*where  $\text{tr}[A]$  and  $\mathbf{E}_k$  represent the trace of  $A$  and the identity matrix with  $k$  degrees, respectively.*

Henceforth we consider the estimation problem for a sequential multinomial sampling. A stopping rule is called a *single sampling plan of size  $n$*  (or SSP( $n$ ) for short) if there exists a positive integer  $n$  such that  $N = n$  with probability 1. A stopping rule is also called an *inverse sampling plan with index  $(i, n)$*  (or ISP( $i, n$ ) for short) if there exist a positive integer  $n$  and  $i$  with  $1 \leq i \leq k$  such that

$$\sum_{j=0}^i Y_j^{(N)} = n,$$

where the permutation of components of  $Y^{(N)}$  is allowed. Here, without loss of generality, we take the sum of the first  $i$  coordinates.

DeGroot [D59] obtained the efficient sequential estimation procedure in the case of  $k = 2$ . Bhat and Kulkarni [BK66] showed that for any  $k$  the efficient sequential estimation procedure in the sense of attaining the lower bound for the variance exists in the only cases of SSP and ISP and the estimators are of only linear form in the cases. The fact is also shown to be true for the generalized variance ([Ko96]).

We extend the case of Bernoulli trials in Wasan [Wa64] to that of multinomial ones and consider the following problems ([Ko96]).

- (I) minimizing  $\sum_{j=1}^{k-1} \text{Var}_\theta(\delta_j^{(N)})$  under  $E_\theta(N) \leq n$  for positive integer  $n$ ,
- (II) minimizing  $E_\theta(N)$  under  $\sum_{j=1}^{k-1} \text{Var}_\theta(\delta_j^{(N)}) \leq a$  for some positive number  $a$ ,
- (III) minimizing  $\sum_{j=1}^{k-1} \text{Var}_\theta(\delta_j^{(N)}) + cE_\theta(N)$  for some positive constant  $c$  as a cost per observation.

*Case (I).* Let  $\Delta$  be the totality of pairs of a stopping rule satisfying suitable regularity conditions and an unbiased estimator  $\delta^{(N)} = (\delta_1^{(N)}, \dots, \delta_{k-1}^{(N)})$  of  $\theta$ . Let

$$\Delta_n^{(1)} := \left\{ (N, \delta^{(N)}) \in \Delta : \sup_\theta E_\theta(N) \leq n \right\}$$

for  $n = 1, 2, \dots$ . Taking SSP( $n$ ) as a stopping rule and

$$\hat{\delta}^{(N)} = \frac{1}{N} \left( \sum_{i=1}^N X_1^{(i)}, \dots, \sum_{i=1}^N X_{k-1}^{(i)} \right)$$

as an estimator, we see that the sequential decision procedure  $(SSP(n), \hat{\delta}^{(N)})$  uniformly minimizes the sum of variances.

Case (II). Put  $a := (k - 1)^2 / (nk^2)$  and let

$$\Delta_a^{(2)} := \left\{ (N, \delta^{(N)}) \in \Delta : \sup_{\theta} \sum_{j=1}^{k-1} \text{Var}_{\theta}(\delta_j^{(N)}) \leq a \right\}.$$

Then  $(SSP(n), \hat{\delta}^{(N)})$  is admissible in  $\Delta_a^{(2)}$ .

Case (III). We can get reasonable solutions in the only special situation. That is,  $n = 1/(\sqrt{c}k)$  is an integer, and  $(SSP(n), \hat{\delta}^{(N)})$  is admissible in  $\Delta$ .

Next we consider the case where the condition of unbiasedness is not necessarily assumed. This corresponds to the problem of whether something like the Stein type estimator in the case of the multivariate normal distribution exists or not in the sequential multinomial sampling plan. First we extend the parameter space of  $\theta$  to

$$\left\{ \theta = (\theta_1, \dots, \theta_{k-1}) \in \mathbf{R}^{k-1} : 0 \leq \theta_j \leq 1 \quad (j = 1, \dots, k-1), \quad \sum_{j=1}^{k-1} \theta_j \leq 1 \right\}.$$

We take

$$L(\theta, a) = \sum_{j=1}^{k-1} \frac{(a_j - \theta_j)^2}{\theta_j}$$

as a loss function, where  $\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j$ ,  $a = (a_1, \dots, a_{k-1})$ ,  $a_k = 1 - \sum_{j=1}^{k-1} a_j$  and  $0/0 = 0$ . Let  $\Sigma = \{\sigma_{ij}\}$  be the covariance matrix of  $(X_1^{(1)}, \dots, X_{k-1}^{(1)})$ , that is,

$$\sigma_{ij} = \begin{cases} \theta_i(1 - \theta_i) & (i = j), \\ -\theta_i\theta_j & (i \neq j) \end{cases}$$

for  $i, j = 1, \dots, k - 1$ . Then a straightforward calculation yields that the loss function becomes

$$L(\theta, a) = (a - \theta)\Sigma^{-1}(a - \theta)'$$

if  $\Sigma$  is nonsingular ([OS79]). Let  $\Delta^*$  be a class of all the pairs of a stopping rule  $N$  satisfying suitable regularity conditions and an estimator  $\delta^{(N)} = (\delta_1^{(N)}, \dots, \delta_{k-1}^{(N)})$  (which is not necessarily unbiased) for  $\theta$ . Let

$$\Delta_n^{(1)*} = \left\{ (N, \delta^{(N)}) \in \Delta^* : \sup_{\theta} E_{\theta}(N) \leq n \right\} \quad (n = 1, 2, \dots).$$

Then we have the following ([Ko96], [KoA94]).

**Theorem 6.2.** *For  $n \geq 1$ , the sequential decision procedure  $(SSP(n), \hat{\delta}^{(N)})$  is admissible and minimax in  $\Delta_n^{(1)*}$ .*

Finally, suppose that a prior distribution of  $(\theta_1, \dots, \theta_{k-1})$  is the Dirichlet one with a density

$$\Gamma(\nu_1 + \dots + \nu_k) \prod_{j=1}^k \frac{\theta_j^{\nu_j - 1}}{\Gamma(\nu_j)} \quad (\nu_j > 0, j = 1, \dots, k)$$

with  $\sum_{j=1}^k \theta_j = 1$  and  $\theta_j > 0$  ( $j = 1, \dots, k$ ). Assume that a loss function  $L(\theta, a)$  is given by  $M(\theta - a)^2$  and a cost per observation is equal to  $c$ , where  $M$  and  $c$  are positive constants. Then we consider the Bayesian sequential estimation of  $\theta_j$  for  $j = 1, \dots, k$ . For any stopping rule the Bayes terminal decision procedure is given by a nonrandomized decision procedure  $d(X^{(1)}, X^{(2)}, \dots) = (d_1(X^{(1)}), d_2(X^{(1)}, X^{(2)}), \dots)$ , where

$$d_m(X^{(1)}, \dots, X^{(m)}) = \frac{\nu_j + \sum_{i=1}^m X_j^{(i)}}{m + \sum_{j=1}^k \nu_j} \quad (j = 1, \dots, k; m = 1, 2, \dots).$$

We can also obtain the optimal stopping rule in the sense of minimizing the Bayes risk using a backward induction ([KoA93a]).

### 7. ASYMPTOTIC THEORY OF SEQUENTIAL ESTIMATION

In sequential estimation, it is not generally easy to obtain an optimal fixed size for the sample. So, it is usual to consider it asymptotically. It is ordinarily discussed as follows ([M83]). Suppose that  $X_1, X_2, \dots$  is a sequence of independent and identically distributed random variables with mean  $\mu$  and variance  $\sigma^2$ . Let  $n$  be a sample size. In the case when we estimate  $\mu$  by the sample mean  $\bar{X}_n := (1/n) \sum_{i=1}^n X_i$ , we use the loss function  $L_n := A(\bar{X}_n - \mu)^2 + n$ , where  $A$  is a positive constant. If  $n$  is a fixed sample size and  $\sigma$  is known, then  $n = n_0$  minimizing the risk  $R_n := E(L_n) = A\sigma^2 n^{-1} + n$  is given by  $n_0 \approx A^{1/2}\sigma$  and the value of its risk becomes  $R_{n_0} \approx 2A^{1/2}\sigma$ . If  $\sigma$  is unknown, we cannot use  $n_0$ , and there is not a procedure for getting a fixed sample size attaining the risk  $R_{n_0}$ . In this case one can estimate  $\mu$  by  $\bar{X}_T$  using the stopping rule

$$\begin{aligned} T = T_A &= \inf \left\{ n \geq n_A : n \geq A^{1/2} \left[ n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]^{1/2} \right\} \\ &= \inf \left\{ n \geq n_A : n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \leq A^{-1} n^2 \right\}, \end{aligned}$$

where  $n_A$  is a natural number depending on  $A$ . Indeed, Robbins [R59] considered the above sequential estimation procedure for the normal distribution. In the normal case, Starr [S66] showed its asymptotic risk efficiency, that is,  $R_T/R_{n_0} \rightarrow 1$  as  $A \rightarrow \infty$ , Starr and Woodroffe [SW69] verified that, on the regret,  $R_T - R_{n_0} = O(1)$  as  $A \rightarrow \infty$ , and Woodroffe [W77] obtained the second-order approximation of the

risk of the sequential estimation procedure, i.e.,  $R_T - R_{n_0} = 1/2 + o(1)$  as  $A \rightarrow \infty$  (see [M83] for other distributions). Takada [T92] also pointed out that there exists a sequential estimation procedure with asymptotically negative regret. On the other hand, there was a trial to make the risk asymptotically smaller using a bias-adjustment of the sequential procedure ([UI94]).

Next, from the viewpoint of higher-order asymptotic theory, we obtain the Bhattacharyya type lower bound for the risk of asymptotically unbiased estimation procedures and show that the sequential maximum likelihood estimation procedure attains the bound. Let  $X_1, X_2, \dots, X_n, \dots$  be a sequence of independent and identically distributed random variables with a density function  $f(x, \theta)$  (w.r.t. a  $\sigma$ -finite measure  $\mu$ ), where  $\theta$  is a real-valued parameter. We assume that  $E_\theta(n) = \nu(\theta) + o(1)$ ,  $\text{Var}_\theta(n)/\nu(\theta) = O(1)$ ,  $E_\theta(n^k)/\{\nu(\theta)\}^k = O(1)$  ( $k = 2, 3, 4$ ),  $\{(\partial^k/\partial\theta^k)\nu(\theta)\}/\nu(\theta) = O(1)$  ( $k = 1, 2$ ). We also define a risk of the estimator  $\hat{\theta}_n := \hat{\theta}_n(X_1, \dots, X_n)$  with a stopping rule by

$$r_\nu := E_\theta \left[ (\hat{\theta}_n - \theta)^2 \right] + c\nu(\theta).$$

Here  $c$  is some positive constant and  $c\nu(\theta)$  represents the average cost and we denote  $\nu(\theta)$  by  $\nu$ . We also impose appropriate regularity conditions on  $f(x, \theta)$ . Further we define  $I(\theta) := E_\theta[\{\ell^{(1)}(\theta, X)\}^2]$ ,  $J(\theta) := E_\theta[\ell^{(1)}(\theta, X)\ell^{(2)}(\theta, X)]$ ,  $M(\theta) := E_\theta[\{\ell^{(2)}(\theta, X)\}^2] - I^2(\theta)$  and  $N(\theta) := E_\theta[\{\ell^{(1)}(\theta, X)\}^2\ell^{(2)}(\theta, X)] + I^2(\theta)$ , where  $\ell^{(k)}(\theta, x) = (\partial^k/\partial\theta^k)\ell(\theta, x)$  ( $k = 1, 2$ ) with  $\ell(\theta, x) = \log f(x, \theta)$ . Then we have the following ([A94]).

**Theorem 7.1 (Bhattacharyya type lower bound).** *Under suitable regularity conditions, for any asymptotically unbiased estimator  $\hat{\theta}_n$  with a stopping rule, i.e.,  $E_\theta(\hat{\theta}_n) = \theta + o(1/\nu)$  it follows that*

$$r_\nu \geq 2\sqrt{\frac{c}{I(\theta)}} + \frac{cJ^2(\theta)}{2I^3(\theta)} + o(c)$$

as  $c \rightarrow 0$ .

We also consider a sequential estimation procedure attaining the above Bhattacharyya type lower bound. Let  $\hat{\theta}_{ML}$  be a maximum likelihood (ML) estimator based on  $X_1, \dots, X_n$ . We denote  $\hat{\theta}_{ML}^*$  to be a bias-adjusted ML estimator so that  $E_\theta(\hat{\theta}_{ML}^*) = \theta + o(1/\nu)$ . Then we have the following ([A94]).

**Theorem 7.2.** *Assume that suitable regularity conditions hold. Suppose that the stopping rule  $S_0$  is so determined that the observation is stopped at  $n$  satisfying*

$$-\sum_{i=1}^n \ell^{(2)}(\hat{\theta}_{ML}^*, X_i) = \nu(\hat{\theta}_{ML}^*) I(\hat{\theta}_{ML}^*) + C(\hat{\theta}_{ML}^*) + \varepsilon,$$

where  $\nu(\theta) = 1/\sqrt{cI(\theta)} + O(1)$  with

$$C(\theta) = \frac{J(\theta)\nu'(\theta)}{I(\theta)\nu(\theta)} - \frac{\nu''(\theta)}{2\nu(\theta)} + \frac{1}{2I(\theta)}\{M(\theta) + N(\theta)\}$$

and some random variable  $\varepsilon$  with  $E_\theta(\varepsilon) = o(1)$ . Then the risk of the bias-adjusted ML estimator  $\hat{\theta}_{ML}^*$  with the stopping rule  $S_0$  is given by

$$r_\nu^{(0)} := E_\theta \left[ (\hat{\theta}_{ML}^* - \theta)^2 \right] + c\nu(\theta) = 2\sqrt{\frac{c}{I(\theta)}} + \frac{cJ^2(\theta)}{2I^3(\theta)} + o(c)$$

as  $c \rightarrow 0$ , that is, the risk of the bias-adjusted ML estimation procedure  $(S_0, \hat{\theta}_{ML}^*)$  attains the Bhattacharyya type lower bound given in Theorem 7.1 up to the order  $o(c)$ .

One intuitively obvious stopping rule  $S_1$  could be to determine by the equality  $n = 1/\sqrt{cI(\hat{\theta}_{ML})}$ . Then the bias-adjusted ML estimation procedure  $(S_1, \hat{\theta}_{ML}^*)$  has the risk

$$r_\nu^{(1)} := 2\sqrt{\frac{c}{I(\theta)}} + \frac{cJ^2(\theta)}{2I^3(\theta)} + \frac{c}{I^3(\theta)} \{I(\theta)M(\theta) - J^2(\theta)\} + o(c)$$

as  $c \rightarrow 0$ , which implies that the risk of the sequential estimation procedure  $(S_1, \hat{\theta}_{ML}^*)$  is generally larger than that of  $(S_0, \hat{\theta}_{ML}^*)$  since  $I(\theta)M(\theta) \geq J^2(\theta)$  by the Schwarz inequality.

In the following example we obtain the risks of bias-adjusted sequential ML estimation procedures and show that they attain the above Bhattacharyya type lower bound in the case of Gamma distribution ([IU95]).

**Example 7.1.** Suppose that  $X_1, X_2, \dots, X_n, \dots$  is a sequence of independent and identically distributed random variables with a density function

$$f(x, \theta) = \begin{cases} \frac{1}{\Gamma(\lambda)} \left(\frac{\lambda}{\theta}\right)^\lambda x^{\lambda-1} e^{-\lambda x/\theta} & \text{for } x > 0, \\ 0 & \text{for } x \leq 0, \end{cases}$$

where  $\lambda$  is a positive constant and  $\theta$  is an unknown positive-valued parameter. Then the ML estimator  $\hat{\theta}_{ML}$  of  $\theta$  is given by  $\bar{X}_n := \sum_{i=1}^n X_i/n$ . Now we consider a stopping rule so that the observation is stopped at

$$N = N_c := \inf \left\{ n \geq m : \sum_{i=1}^n X_i \leq (\lambda c)^{1/2} n^2 \ell_n \right\},$$

where  $m$  is predetermined and  $\ell_n = 1 + (\ell^*/n) + o(1/n)$  with a constant  $\ell^*$  independent of  $n$ . For  $m > 2\lambda^{-1}$  the risk of the sequential ML estimation procedure  $\delta := (N, \bar{X}_N)$  is given by

$$r_\nu(\delta) := E_\theta \left[ (\bar{X}_N - \theta)^2 \right] + c\nu(\theta) = 2\theta\sqrt{\frac{c}{\lambda}} + \frac{3c}{\lambda} + o(c)$$

as  $c \rightarrow 0$ . If we consider two sequential bias-adjusted ML estimators

$$\bar{X}_N^{(1)} := \bar{X}_N + \sqrt{\frac{c}{\lambda}} \quad \text{and} \quad \bar{X}_N^{(2)} := \left(1 + \frac{1}{\lambda N}\right) \bar{X}_N,$$

then for  $m > \lambda^{-1}$

$$E_{\theta} \left[ \bar{X}_N^{(j)} \right] = \theta + o(\sqrt{c})$$

as  $c \rightarrow 0$  for  $j = 1, 2$ , which implies that for  $m > 2\lambda^{-1}$  the risks of the sequential bias-adjusted ML estimation procedures  $\delta^{(j)} := (N, \bar{X}_N^{(j)})$  ( $j = 1, 2$ ) are given by

$$r_{\nu} \left( \delta^{(j)} \right) = 2\theta \sqrt{\frac{c}{\lambda}} + \frac{2c}{\lambda} + o(c)$$

as  $c \rightarrow 0$  for  $j = 1, 2$ . On the other hand, since, in the Gamma case,  $I(\theta) = \lambda/\theta^2$  and  $J(\theta) = -2\lambda/\theta^3$ , it follows from Theorem 7.1 that the Bhattacharyya type lower bound is equal to

$$2\sqrt{\frac{c}{I(\theta)}} + \frac{cJ^2(\theta)}{2I^3(\theta)} + o(c) = 2\theta \sqrt{\frac{c}{\lambda}} + \frac{2c}{\lambda} + o(c).$$

Hence it is seen that both of the sequential bias-adjusted ML estimation procedures  $\delta^{(1)}$  and  $\delta^{(2)}$  coincide with the Bhattacharyya type lower bound for the risk up to the order  $o(c)$ .

Without taking the cost into consideration, we already obtained the Bhattacharyya type lower bound for the asymptotic variance of asymptotically unbiased estimation procedures and showed that an appropriate sequential bias-adjusted ML estimation procedure attained the bound and was uniformly third-order asymptotically efficient in the sense that it attained the bound for the asymptotic distribution of sequential estimation procedures up to the third order, i.e., the order  $o(1/\nu)$  ([TA88], [AT89], [AT91]). We also showed that the sequential discretized likelihood estimation procedure was asymptotically equivalent to the sequential bias-adjusted ML estimation procedure up to the order  $o(1/\nu)$  ([A95]). These mean that, in the sequential case, the use of an appropriate stopping rule provides us with more information, which gives us a stronger result than in the nonsequential case. Indeed, in the nonsequential case, the bias-adjusted ML estimator has the loss  $\{I(\theta)M(\theta) - J^2(\theta)\}/I^2(\theta)$  of information, but, in the sequential case the bias-adjusted ML estimation procedure has no loss of information. This is also shown as the fact that the conformal embedding curvature vanishes for a curved exponential family of distributions from the viewpoint of differential geometry ([OAT91]).

For the sequential interval estimation, Hall [Ha81] discussed the regret using a triple sampling method in the construction of confidence intervals of the mean of the normal distribution. The method is very useful for various problems of sequential estimation. For example, using the method, Honda [Ho92] considered the construction of a mean parameter of a one-parameter exponential family of distributions.

#### REFERENCES

- [A94] Akahira, M., *Second order asymptotic efficiency in terms of the risk in sequential estimation*, *Selecta Statistica Canadiana* **9** (1994), 25–38.
- [A95] Akahira, M., *The Bhattacharyya type bound for the asymptotic variance and the sequential discretized likelihood estimation procedure*, *Sequential Analysis* **14** (1995), 193–204.
- [AT89] Akahira, M. and Takeuchi, K., *Third order asymptotic efficiency of the sequential maximum likelihood estimation procedure*, *Sequential Analysis* **8** (1989), 333–359.

- [AT91] Akahira, M. and Takeuchi, K., *Second order asymptotic efficiency in terms of the asymptotic variance of sequential estimation procedures in the presence of nuisance parameters*, Sequential Analysis **10** (1991), 27–43.
- [ATK92] Akahira, M., Takeuchi, K., and Koike, K., *Unbiased estimation in sequential binomial sampling*, Rep. Statist. Appl. Res. Un. Japan Sci. Engrs. **39** (1992), 1–13.
- [BhBo90] Bhandari, B. K. and Bose, P., *Existence of unbiased estimates in sequential binomial experiments*, Sankhyā, Ser. A **52** (1990), 127–130, Corrigenda *ibid.* Ser. A **55** (1993), 327.
- [BK66] Bhat, B. R. and Kulkarni, N. V., *On efficient multinomial estimation*, J. Roy. Statist. Soc. Ser. B **28** (1966), 45–52.
- [D59] DeGroot, M. H., *Unbiased binomial sequential estimation*, Ann. Math. Statist. **30** (1959), 80–101.
- [F67] Ferguson, T. S., *Mathematical Statistics*, Academic Press, New York, 1967.
- [Gh87] Ghosh, B. K., *On the attainment of the Cramér-Rao bound in the sequential case*, Sequential Analysis **6** (1987), 267–288.
- [GMS46] Girshick, M. A., Mosteller, F., and Savage, L. J., *Unbiased estimates for certain binomial sampling problems with applications*, Ann. Math. Statist. **17** (1946), 13–23.
- [Ha81] Hall, P., *Asymptotic theory of triple sampling for sequential estimation of a mean*, Ann. Statist. **9** (1981), 1229–1238.
- [Ho92] Honda, T., *Construction of a confidence interval by triple sampling*, Sequential Analysis **11** (1992), 273–287.
- [IU95] Isogai, E. and Uno, C., *On the sequential point estimation of the mean of a gamma distribution*, Statist. Probab. Lett. **22** (1995), 287–293.
- [Ko93] Koike, K., *Unbiased estimation for sequential multinomial sampling plans*, Sequential Analysis **12** (1993), 253–260.
- [Ko94] Koike, K., *On the admissibility of estimation procedures in sequential multinomial sampling plans*, Proc. Sympos., Res. Inst. Math. Sci., Kyoto Univ. **879** (1994), 1–13. (in Japanese)
- [Ko96] Koike, K., *On the optimum properties of sequential estimation procedures in the multinomial sampling plans*, Sequential Analysis **15** (1996), 285–298.
- [KoA93a] Koike, K. and Akahira, M., *Bayes sequential decision rules in the multinomial sampling*, Proc. Sympos., Res. Inst. Math. Sci., Kyoto Univ. **842** (1993), 1–11. (in Japanese)
- [KoA93b] Koike, K. and Akahira, M., *Optimal estimation procedures for sequential sampling*, In: Proceedings of the Meeting on Statistical Inference and Sequential Analysis, 1993, pp. 32–39. (in Japanese)
- [KoA93c] Koike, K. and Akahira, M., *Completeness for sequential sampling plans*, Sequential Analysis **12** (1993), 211–218.
- [KoA94] Koike, K. and Akahira, M., *Optimal sequential estimation procedures in multinomial sampling plans*, International Conference on Stochastic Models and Optimal Stopping, Nanzan Univ., Nagoya, 1994, pp. 176–192.
- [KoA95] Koike, K. and Akahira, M., *On the Bhattacharyya type inequality in the sequential estimation*, Proc. Sympos., Res. Inst. Math. Sci., Kyoto Univ. **916** (1995), 180–188. (in Japanese)
- [Kr87] Kremers, W. K., *An improved estimator of the mean for a sequential binomial sampling plan*, Technometrics **29** (1987), 109–112.
- [Kr90] Kremers, W. K., *Completeness and unbiased estimation in sequential multinomial sampling*, Sequential Analysis **9** (1990), 43–58.
- [L83] Lehmann, E. L., *Theory of Point Estimation*, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, New York, 1983.
- [LS50] Lehmann, E. L. and Stein, C., *Completeness in the sequential case*, Ann. Math. Statist. **21** (1950), 376–385.
- [LR72] Linnik, Yu. V. and Romanovsky, I. V., *Some new results in sequential estimation theory*, Proc. Sixth Berkeley Sympos. Math. Statist. Prob., vol. 1, 1972, pp. 85–96.
- [M83] Martinsek, A. T., *Second order approximation to the risk of a sequential procedure*, Ann. Statist. **11** (1983), 827–836.
- [OAT91] Okamoto, I., Amari, S. and Takeuchi, K., *Asymptotic theory of sequential estimation: Differential geometrical approach*, Ann. Statist. **19** (1991), 961–981.

- [OS79] Olkin, I. and Sobel, M., *Admissible and minimax estimation for the multinomial distribution and for  $k$  independent binomial distributions*, Ann. Statist. **7** (1979), 284–290.
- [R59] Robbins, H., *Sequential estimation of the mean of a normal population*, Probability and Statistics (H. Cramér Volume), Almqvist and Wiksell, Uppsala, Sweden, 1959, pp. 235–245.
- [Sa95] Sato, M., *A note on “Unbiased estimation for sequential sampling plans”*, Sequential Analysis **14** (1995), 335–340.
- [S66] Starr, N., *On the asymptotic efficiency of a sequential procedure for estimating the mean*, Ann. Math. Statist. **37** (1966), 1173–1185.
- [SW69] Starr, N. and Woodroffe, M., *Remarks on sequential point estimation*, Proc. Nat. Acad. Sci. U.S.A. **63** (1969), 285–288.
- [St90] Stefanof, V. T., *A note on the attainment of the Cramér-Rao bound in the sequential case*, Sequential Analysis **9** (1990), 327–334.
- [T92] Takada, Y., *A sequential procedure with asymptotically negative regret for estimating a normal mean*, Ann. Statist. **20** (1992), 562–569.
- [TA88] Takeuchi, K. and Akahira, M., *Second order asymptotic efficiency in terms of asymptotic variances of sequential maximum likelihood estimation procedures*, In: Statistical Theory and Data Analysis II (K. Matusita, ed.), North Holland, Amsterdam, 1988, pp. 191–196.
- [UI94] Uno, C. and Isogai, E., *A note on sequential estimation of the normal variance*, Sequential Analysis **13** (1994), 201–206.
- [Wa64] Wasan, M. T., *Sequential optimum procedures for unbiased estimation of a binomial parameter*, Technometrics **6** (1964), 259–271.
- [Wo46] Wolfowitz, J., *On sequential binomial estimation*, Ann. Math. Statist. **17** (1946), 489–493.
- [Wo47] Wolfowitz, J., *The efficiency of sequential estimates and Wald’s equation for sequential processes*, Ann. Math. Statist. **18** (1947), 215–230.
- [W77] Woodroffe, M., *Second order approximations for sequential point and interval estimation*, Ann. Statist. **5** (1977), 984–995.

Translated by M. AKAHIRA

INSTITUTE OF MATHEMATICS, UNIVERSITY OF TSUKUBA, 1-1-1 TENNODAI, TSUKUBA-SHI, IBARAKI-KEN 305-8571, JAPAN