Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 64(2) 2010 137-144

# Sequential Cramér-Rao and Bhattacharyya Bounds: Work of G.R. Seth and Afterwards

**J.K. Ghosh[1,2] and Sumitra Purkayastha[2*]**
[1]*Purdue University, West Lafayette, USA*
[2]*Indian Statistical Institute, Kolkata, India*

Received 09 March 2010; Revised 24 May 2010; Accepted 24 May 2010

**SUMMARY**

A sequential version of Cramér-Rao inequality was obtained by Wolfowitz (1947). Bhattacharyya inequality (1946, 1947) can be seen as a refinement of Cramér-Rao inequality. We discuss at length its sequential version as obtained by Seth (1949). We also discuss results of Seth and others, notably Ghosh (1987), on the impossibility of attainment of equality in sequential Cramér-Rao inequality, where attainment will mean here and thereafter attainment for all values of the underlying parameter $\theta$ except where it is stated otherwise. We also discuss briefly why sequential estimation remains important, notwithstanding the above non-existence results.

*Keywords*: Bhattacharyya inequality, Cramér-Rao inequality, Sequential estimation.

## 1. INTRODUCTION

Motivated by the importance of Cramér-Rao inequality, researchers in the 1940's began asking if an analogous fact holds in the context of sequential estimation as well. Wolfowitz (1947) showed that there is indeed a sequential version of Cramér-Rao inequality. Professor G.R. Seth, in whose memory this paper is being written, wrote his doctoral dissertation under the supervision of Professor J. Wolfowitz at Columbia University. His work revolved around a study of the Bhattacharyya inequality (1946, 1947), which can be seen as a refinement of Cramér-Rao inequality, in the context of sequential estimation. Seth was also interested in the attainment of Cramér-Rao inequality in the sequential case. Part of the work done in his dissertation was reported in a paper which appeared in *Annals of Mathematical Statistics* in 1949 (Seth, 1949). This paper deals with both the points mentioned above as well as interesting related points like sharpness of Bhattacharyya bounds, orthogonal polynomials and characterization of certain exponential families through orthogonal families arising as Bhattacharyya's covariate functions.

The organization of the paper is as follows. In Section 2, we provide a brief review of Seth (1949) and some related work. In Section 3, we provide a more detailed review of work related to the sequential Cramér-Rao inequality. In Section 4, we describe some work on attainment of sequential Bhattacharyya bounds and state a conjecture. The paper ends with some concluding remarks which appear in Section 5.

## 2. WORK OF PROFESSOR G.R. SETH: BASED ON SETH (1949)

### 2.1 Wolfowitz (1947): Forerunner of Seth (1949)

The Cramér-Rao inequality, for samples of fixed size, was discovered independently by Rao (1945) and Cramér (1946a, 1946b), and several other distinguished authors, e.g., Fréchet (1943) and Darmois (1945), whose papers were published later, and Fábian and Hannan (1977), Pitman (1979) and Müller-Funk *et al.*

*Corresponding author : Sumitra Purkayastha
*E-mail address* : sumitra@isical.ac.in, s.purkayastha@gmail.com

(1989), the last three giving much better suficient conditions.

The Cramér-Rao inequality led researchers to the following question: is there a similar result which is relevant in the context of sequential estimation? Wolfowitz (1947) showed that the answer to this question is affirmative. His main result, proved for independent and identically distributed (i.i.d.) observations, is stated below in Theorem 1.

**Theorem 1.** (Sequential Cramér-Rao inequality in an i.i.d. set-up: Wolfowitz (1947))

Let $X_1$, $X_2$, ... be a sequence of i.i.d. random variables with common probability density function given by $f_\theta(x)$ with respect to some $\sigma$-finite measure $\mu$, where $\mu$ is either the Lebesgue measure $\lambda$ or the counting measure, and also where $\theta \in \Theta$, an open interval of $\mathbb{R}$. Let $N$ be a stopping time with $P_\theta(N = 0) = 0$ and $E_\theta(N) < \infty$ for all $\theta \in \Theta$. Also let $T_N = \delta(X_1 ..., X_N)$ be an estimator of a parametric function $\theta$ such that $E_\theta(T_N) = \theta$. We make the following assumptions.

(1) The derivative $\partial f_\theta(x)/\partial\theta$ exists for all $\theta$ and almost all $x$ $[\mu]$. We define $\partial \log f_\theta(x)/\partial\theta$ as zero whenever $f_\theta(x) = 0$; thus $\partial \log f_\theta(x)/\partial\theta$ is defined for all $\theta \in \Theta$ and almost all $x$ $[\mu]$. We also assume that $E_\theta[\partial \log f_\theta(X_1)/\partial\theta] = 0$ and that $E_\theta[\partial \log f_\theta(X_1)/\partial\theta]^2 > 0$ for all $\theta \in \Theta$.

(2) $E_\theta\left(\sum_{i=1}^N |\partial \log f_\theta(X_i)/\partial\theta|\right)^2$ exists for all $\theta \in \Theta$.

(3) For any integral $j$, there exists a non-negative $\mu$-measurable function $T_j(x_1, ..., x_j)$ such that

(a) $\left|\delta(x_1,...,x_j)\dfrac{\partial}{\partial\theta}\prod_{i=1}^{j} f_\theta(x_i)\right| < T_j(x_1, ...., x_j)$

for all $\theta \in \Theta$ and almost all $(x_1, ..., x_j) \in \{N = j\}$ $[\mu]$,

(b) $\int_{N=j} T_j(x_1, ..., x_j)\, d\mu(x_1) ... d\mu(x_j)$

is finite.

(4) Let

$$t_j(\theta) := \int_{N=j} \delta(x_1, ..., x_j) \prod_{i=1}^{j} f_\theta(x_i)\, d\mu(x_1) ... d\mu(x_j).$$

We postulate the uniform convergence of the series

$$\sum_{j=1}^{\infty} \frac{dt_j(\theta)}{d\theta}$$

for all $\theta \in \Theta$. (The existence of $\dfrac{dt_j(\theta)}{d\theta}$ is implied by (3))

Then

$$\mathrm{Var}_\theta(T_N) \geq \frac{1}{E_\theta(N)\,E_\theta([\partial \log p_\theta(X_1)/\partial\theta]^2)} \tag{1}$$

**Remark 1.** It can be see that the inequality in (1) formally looks the same as its fixed-sample size counterpart, except that one has to replace the sample size $n$ by the expectation of the associated stopping time N.

**Remark 2.** Suppose $X_1$, $X_2$, ... are i.i.d. random variables having common distribution $P_\theta$, $\theta \in \Theta$, where $\Theta \subseteq \mathbb{R}$ is open, with pdf/pmf with respect to *a* $\sigma$-finite measure $\mu$ given by

$$f_\theta(x) = \exp(\theta x - \psi(\theta)),\ \theta \in \Theta \tag{2}$$

Also, let $T_N = \delta(X_1, ..., X_N)$ be an unbiased estimator of a parametric function $\gamma(\theta)$ that is differentiable with respect to $\theta$. Then (1) has the following form:

$$\mathrm{Var}_\theta(T_N) \geq \frac{[\gamma'(\theta)]^2}{E_\theta(N)\,\psi''(\theta)} \tag{3}$$

In particular, if $X_i$'s are $N(\theta, 1)$, then $\psi(\theta) = \theta^2/2$ so that the right-hand side of (3) becomes $[\gamma'(\theta)]^2/E_\theta(N)$. Also, if $X_i$'s are Bin($\theta$, 1), then the right-hand side of (3) becomes $[\{\gamma'(\theta)\}^2\theta(1 - \theta)]/E_\theta(N)$.

**Remark 3.** Ghosh *et al*. (1997, p. 96) contains a version of sequential Cramér-Rao inequality where (1) the $X_i$'s are neither independent nor have identical distributions, (2) the statistic $T_N = \delta(X_1, ..., X_N)$ is an estimator of a parametric function $\gamma(\theta)$ which is possibly biased, and (3) the conditions which allow differentiation under the summation and the integral signs are not separately stated. This result is stated below.

**Theorem 2**. (Sequential Cramér-Rao inequality: vide Ghosh *et al*. (1997), p. 96)

Let $X_1, X_2, \ldots$ be a sequence of random variables with joint probability density function of $(X_1, \ldots, X_j)$ being given by $f_\theta(x_1, \ldots, x_j)$ with respect to some $\sigma$-finite measure $\mu^j$ for every $j \geq 1$. Let $N$ be a stopping time with $P_\theta(N = 0) = 0$ and $P_\theta(N < \infty) = 1$ for all $\theta \in \Theta$. Also let $T_N = \delta(X_1, \ldots, X_N)$ be an estimator of a parametric function $\gamma(\theta)$ such that $E_\theta(T_N) = \gamma(\theta) + b(\theta)$. It is assumed that $\gamma(\theta)$, $b(\theta)$, and $\log f_\theta(x_1, \ldots, x_j)$ (for every $j \geq 1$) are all differentiable with respect to $\theta$ and $E_\theta[\partial \log f_\theta(x_1, \ldots, x_j)/\partial\theta]^2 > 0$ for every $j \geq 1$. Further, assume that differentiation under the summation and the integral signs are valid in $E_\theta(1) = 1$ and $E_\theta(T_N) = \gamma(\theta) + b(\theta)$. Then

$$\mathrm{Var}_\theta(T_N) \geq \frac{[\gamma'(\theta) + b'(\theta)]^2}{E_\theta([\partial \log f_\theta(X_1, \ldots, X_N)/\partial\theta]^2)}$$

**Remark 4.** Ghosh *et al*. (1997, p. 97) contains a statement of Theorem 2 in the special case when the $X_i$'s are i.i.d.

## 2.2 Seth (1949)

Bhattacharyya (1946, 1947) extended the Cramér-Rao inequality, providing a lower bound to the variance of an unbiased estimator based not only on the first derivative of the likelihood function but higher derivatives as well. This result is known as the *Bhattacharyya inequality*.

The results mentioned in the preceding paragraph were discovered at around the same time. It was therefore natural that a sequential analogue of the Bhattacharyya inequality will also be sought by researchers. In a major achievement, Professor G.R. Seth obtained a sequential version of the Bhattacharyya inequality in his doctoral dissertation written under the supervision of Professor J. Wolfowitz at Columbia University. Seth's inequality (Seth, 1949) is stated below.

**Theorem 3.** (Sequential Bhattacharyya inequality: Seth (1949))

Let $X_1, X_2, \ldots$ be a sequence of random variables with joint probability density function of $(X_1, \ldots, X_j)$ being given by $f_\theta(x_1, \ldots, x_j)$ with respect to some $\sigma$-finite measure $\mu^j$ for every $j \geq 1$, where $\mu^j$ is either the Lebesgue measure on $\mathbb{R}^j$ or the counting measure, and also where $\theta \in \Theta$, an open interval of $\mathbb{R}$. Let $N$ be a stopping time with $P_\theta(N = 0) = 0$ and $P_\theta(N < \infty) = 1$ for all $\theta \in \Theta$. Also, let $T_N = \delta(X_1, \ldots, X_N)$ be an unbiased estimator of a parametric function $\gamma(\theta)$ that is differentiable $k$ times with respect to $\theta$. Let $\alpha := ((\mathrm{d}\gamma(\theta)/\mathrm{d}\theta), \ldots, (\mathrm{d}^k\gamma(\theta)/\mathrm{d}\theta^k))^T$. We make the following assumptions.

(1) For every $j \geq 1$, the derivatives $\partial^i f_\theta(x_1, \ldots, x_j)/\partial\theta^i$, $i = 1, \ldots, k$, exist for all $\theta \in \Theta$ and almost all $(x_1, \ldots, x_j) \in \{N = j\}$ $[\mu^j]$. We define

$$\frac{1}{f_\theta(x_1, \ldots, x_j)} \frac{\partial^i f_\theta(x_1, \ldots, x_j)}{\partial\theta^i} = 0$$

whenever $f_\theta(x_1, \ldots, x_j) = 0$; thus

$$\phi_i(j) \equiv \frac{1}{f_\theta(x_1, \ldots, x_j)} \frac{\partial^i f_\theta(x_1, \ldots, x_j)}{\partial\theta^i}$$

is defined for all $\theta \in \Theta$ and almost all $(x_1, \ldots, x_j) \in \{N = j\}$ $[\mu^j]$.

(2) For any integral $j$, there exists non-negative $\mu^j$-measurable functions $T_{i,j}(x_1, \ldots, x_j)$, $i = 1, \ldots, k$, such that

(a) $$\left| \delta(x_1, \ldots, x_j) \frac{\partial^i f_\theta(x_1, \ldots, x_j)}{\partial\theta^i} \right| < T_{i,j}(x_1, \ldots, x_j)$$

for all $\theta \in \Theta$ and almost all $(x_1, \ldots, x_j) \in \{N = j\}$ $[\mu^j]$,

(b) $\int_{N=j} T_{i,j}(x_1, \ldots, x_j)\, \mathrm{d}\mu^j$, $i = 1, \ldots, k$, are finite.

(3) Let $t_j(\theta) := \int_{N=j} \delta(x_1, \ldots, x_j) f_\theta(x_1, \ldots, x_j)\, \mathrm{d}\mu^j$.

We postulate the uniform convergence of the series

$$\sum_{j=1}^{\infty} \frac{\mathrm{d}^i}{\mathrm{d}\theta^i} t_j(\theta), \; i = 1, \ldots, k,$$

for all $\theta \in \Theta$. (The existence of $\frac{\mathrm{d}^i}{\mathrm{d}\theta^i} t_j(\theta)$ is implied by (3))

(4) For every $j \geq 1$, there exists functions $S_i(x_1, \ldots, x_j)$ for $i = 1, \ldots, k$, such that when $\delta(x_1, \ldots, x_j)$ and $T_{i,j}(x_1, \ldots, x_j)$ are replaced by unity and $S_i(x_1, \ldots, x_j)$ respectively, conditions (2) and (3) still hold good.

(5) The covariance matrix of $(\phi_1(N), ..., \phi_k(N))^T$ exists and is non-singular for almost all $\theta \in \Theta$. We denote this matrix by $\Lambda \equiv ((\lambda_{st}))$.

Then

$$\text{Var}_\theta(T_N) \geq \alpha^T \Lambda^{-1} \alpha \quad (4)$$

**Remark 5.** Suppose $X_i$'s are as in Remark 2 (vide (2)) and $N$ is a stopping time with $P_\theta(N = 0) = 0$ and $E_\theta(N^4) < \infty$ for all $\theta \in \Theta$. Suppose $k = 2$ in Theorem 3. Let $\gamma(\theta)$ be twice differentiable. Defining, for $j \geq 1$, $S_j = \Sigma_{s=1}^{j} X_s$, it is easy to see that $\phi_1(N) = S_N - N\psi'(\theta)$, $\phi_2(N) = [S_N - N\psi'(\theta)]^2 - N\psi''(\theta)$. Note that $\mu := E_\theta(X_1) = \psi'(\theta)$, $\sigma^2 := \text{Var}_\theta(X_1) = \psi''(\theta)$, $\mu_3 := E_\theta(X_1 - \mu)^3 = \psi^{(3)}(\theta)$. Let us also note that $E_\theta(\phi_1(N)) = E_\theta(\phi_2(N)) = 0$ for every $\theta \in \Theta$ (cf. (2.3.9) and (2.3.10) of Seth (1949)). Finding the lower bound in (3) requires computing now the entries of the matrix $\Lambda = ((\lambda_{st}))$. Notice now that $\lambda_{st} = E_\theta[\phi_s(N)\phi_t(N)]$, $s, t = 1, 2$. Proceeding as in Example 1 in § 4.1 of Seth (1949), we see that $\lambda_{11} = \psi''(\theta)E_\theta(N)$, $\lambda_{12} = E_\theta(N)\psi^{(3)}(\theta) + 2E_\theta[N(S_N - N\mu)]$, and $\lambda_{22} = E_\theta[\{S_N - N\psi'(\theta)\}^4] - 2\psi''(\theta)E_\theta[N\{S_N - N\psi'(\theta)\}^2] + \{\psi''(\theta)\}^2 E_\theta(N^2)$.

In addition to the result stated above, Seth (1949) also derives several other important results. We mention some of them below.

- He obtained a condition under which the (sequential) Bhattacharyya lower bound obtained by him was sharper than the (sequential) Cramér-Rao lower bound obtained by Wolfowitz (1947). More generally, he obtained a condition under which the (sequential) Bhattacharyya lower bound, based on a set of derivatives of the likelihood function upto and including a fixed order will improve if the derivative of the next order is considered.
- Attainment of the lower bound in sequential Cramér-Rao inequality and fixed-sample size Bhattacharyya inequality received considerable attention in Seth (1949).
  - He studied two examples in each of which the sequential Bhattacharyya lower bound is greater than the sequential Cramér-Rao lower bound given byWolfowitz (1947). These examples demonstrate both unattainability of sequential Cramér-Rao lower bound and desirability of finding a higher lower bound than the one given by Wolfowitz (1947).
  - With a view to understanding when the Wolfowitz lower bound (1947) will be attained, he considered an i.i.d. set-up, a stopping time $N$ having finite expectation $E_\theta(N)$, admitting first two derivatives for all $\theta$ with the first derivative of $E_\theta(N)$ being either zero for all $\theta$ or never zero. Finally, in presence of an *efficient* estimate (Cramér 1946b), he showed that the Wolfowitz lower bound (1947) for variance of unbiased estimates will be attained only with a fixed-sample size procedure. This generalizes a similar result obtained earlier by Blackwell and Girshick (1947).
  - He obtained necessary and sufficient conditions for the attainment of the Bhattacharyya lower bound in the fixed-sample size case.
  - Finally, while studying equality in the inequality obtained by him, Seth (1949) showed that successive derivatives of likelihood functions, divided by the corresponding likelihood, constitute classes of orthogonal polynomials with respect to weight functions given by (1) normal probability density function (pdf) with unknown mean, (2) normal pdf with unknown variance, (3) binomial probability mass function (pmf) with unknown probability of success, and (4) Poisson pmf with unknown mean. It is indeed both remarkable and surprising that such orthogonal polynomials were obtained from purely statistical consideration like obtaining lower bound to unbiased estimators.

**Remark 6.** Characterization of distributions based on Bhattacharyya bounds have been addressed in the literature. Shanbhag (1972) showed that for the $3 \times 3$ Bhattacharyya matrix of a distribution having an exponential-type density function to be diagonal, it is necessary and sufficient that it be one of the following: normal, poisson, gamma, binomial, and negative binomial. He seems to have missed one of Morris's

(1982) families. Shanbhag (1972) also showed that for these distributions an $s \times s$ Bhattacharyya matrix is well-defined for all $s$ and is diagonal. It may be noted that the Bhattacharyya bounds are easy to write down in the diagonal case.

**Remark 7.** Ghosh and Sathe (1987) proved that for all unbiasedly estimable functions and all multiparameter exponential families, Bhattacharyya bounds converge to the variance of the minimum variance unbiased estimate as the order of the Bhattacharyya matrix goes to infinity. They also point out that for the exponential families characterized by Shanbag (1972), the variance $\mathrm{Var}_\theta(X)$ is a quadratic function of the mean $E_\theta(X)$ (where $X$ is what Morris (1982) calls the natural random variable for exponential families). Morris (1982) has characterized six exponential families for which $\mathrm{Var}_\theta(X)$ is a quadratic function of $\mathrm{Var}_\theta(X)$. His six families are the following well-known examples: normal, poisson, gamma, binomial, and negative binomial, and one new one. Morris (1982) gives many reasons as to why these families are of special interest within the class of exponential families.

## 3. ATTAINMENT OF SEQUENTIAL CRAMÉR-RAO LOWER BOUND

Research on attainment of Cramér-Rao lower bound (for all $\theta$) has interested many researchers. Wijsman (1973) obtained the first technically complete proof of a result giving necessary and sufficient conditions which will ensure attainment of this bound. This result is stated below.

**Theorem 4.** (Attainment of the Cramér-Rao lower bound: Wijsman (1973))

Suppose the sample space is an arbitrary measure space $(\chi, \mathcal{A}, \mu)$, with $\mu$ $\sigma$-finite. The parameter space is the measure space $(\Theta, \mathcal{B}, \nu)$ with $\Theta$ a Borel subset of the real line, $\mathcal{B}$ the Borel subsets of $\Theta$ and $\nu$ Lebesgue measure. There is given a random variable $X$ with values in $\chi$ and distribution $P_\theta$ given by $P_\theta(A) = \int_A p_\theta(x)\, d\mu(x)$, $\theta \in \Theta$. For convenience differentiation with respect to $\theta$ will be denoted by $D$. Any integration with respect to $\mu$ will always be understood to be over the whole of $\chi$. Let $m$ be a real-valued function on $\Theta$, not identically constant; let $t(X)$ be an unbiased estimator of $m(\theta)$. We shall make the following regularity conditions.

(a) $\Theta$ is an open interval (possibly infinite or semi-infinite);

(b) $p_\theta(x) > 0$ for every $\theta \in \Theta$, $x \in \chi$, $p_\theta(\cdot)$ is $\mathcal{A}$-measurable for every $\theta \in \Theta$, and $p_\theta(x)$ is a continuously differentiable function of $\theta$ for every $x \in \chi$;

(c) $0 < \mathrm{Var}_\theta D\log p_\theta(X) < \infty$ for every $\theta \in \Theta$;

(d) $\int p_\theta(x)\, d\mu(x)$ can be differentiated under the integral sign with respect to $\theta$;

(e) $\int t(x)\, p_\theta(x)\, d\mu(x)$ is finite and can be differentiated under the integral sign with respect to $\theta$.

Then the inequality

$$\mathrm{Var}_\theta(t(X)) \geq \frac{[m'(\theta)^2]}{\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log p_\theta(X)\right)}$$

is an equality for all $\theta$ if and only if there exists $K \in \mathcal{A}$ with $\mu(K) = 0$ such that for $x \in \chi - K$, $\theta \in \Theta$,

$$p_\theta(x) = c(\theta)h(x)\exp(q(\theta)t(x))$$

in which $c(\theta)$ and $h(x)$ are $> 0$, $q$ is strictly monotonic, and both $c$ and $q$ are continuously differentiable.

In other words, Wijsman (1973) proved that the variance of an unbiased estimator of a function of a real parameter attains the Cramér-Rao lower bound if and only if the family of distributions is a one-parameter exponential family. Later Joshi (1976) also addressed similar issues.

The question of attainment of the sequential Cramér-Rao lower bound (for all $\theta$) in the sequential case was taken up and answered in the negative by Seth (1949) for $E_\theta(X)$. Ghosh (1987) contains a detailed study of equality in sequential Cramér-Rao inequality. Trybula (1968) and Linnik and Romanovsky (1972) also addressed this issue. More explicitly, Ghosh (1987) showed that when the observations follow a one-parameter exponential family of distributions the bound can be attained for one or all values of the parameter under strictly sequential rules only in a very special case, namely, for the Bernoulli distribution. In what follows, we briefly discuss the work by Ghosh (1987), including a *less formal* presentation of its main theorem (Theorem 5 below) and some heuristic explanation as to why the result is expected. A good presentation of the arguments of Ghosh (1987) appears in Section 4.3 of Ghosh *et al*. (1997).

**Theorem 5.** (Attainment of sequential Cramér-Rao lower bound: Ghosh (1987))

Suppose $X_1, X_2, \ldots$ are as in Remark 2. Consider a strictly sequential procedure with stopping time $N$ and a statistic $T_N \equiv T_N(X_1, \ldots, X_N)$ with $E_\theta(T_N) = \gamma(\theta)$, where $\gamma(\theta)$ is a differentiable function. For this one-parameter family, the inequality in (3) becomes an equality if and only if $X_1$ has as its support a certain set containing two points and the stopping time is either a constant or one particular $N$.

We present now a partly heuristic argument explaining why in presence of attainment of sequential Cramér-Rao inequality, $X_1$ must be as in Theorem 5. Sequential Cramér-Rao inequality is established by employing Cauchy-Schwartz inequality on the covariance of $T_N(X_1, \ldots, X_N)$ and $\partial \log f_\theta(X_1, \ldots, X_N)/\partial\theta = S_N - N\psi'(\theta)$. Let $S_n := X_1 + \cdots + X_n$. Thus, for understanding the equality in sequential Cramér-Rao inequality, if we tried to argue as for the fixed-sample size case, we would consider estimates of the form

$$T_N = \gamma(\theta) + B(\theta)(S_N - N\psi'(\theta)) \;\; \forall \;\; \theta \tag{5}$$

Since right-hand side of (5) involves $\theta$, the left-hand side cannot be a statistic (i.e., free of $\theta$) except for the following two cases, namely,

(1) the fixed-sample size case when $S_N - N\psi'(\theta)$ reduces to $S_m - m\psi'(\theta)$ and hence we may take $T_N = T_m = A + BS_m$,

(2) the sequential case when $S_N$ is a linear function of $N$, say $A + BN$. Then $T_N - \gamma(\theta) = B(\theta)\,(S_N - N\psi'(\theta)) = B_1N$ +some function of $\theta = B_1(N - E_\theta(N))$ as $T_N$ is unbiased. Hence, $T_N = B_2N$. This argument is formalized in Lemma 1 of Ghosh (1987). He then shows that this violates the assumption $E_\theta(N) < \infty \;\; \forall \;\; \theta$, if the $X_i$'s have distribution belonging to an exponential family other than the Bernoulli.

**Remark 8.** In passing, we mention that Ghosh (1987) also characterized the rules for which equality in (1) is attained. These are essentially negative binomial sampling schemes. De-Groot (1959) called such sequential sampling plans efficient. We also note from a careful scrutiny of Theorem 1 of Ghosh (1987), that under the assumption of sampling Bernoulli variables and stopping time obtained by Ghosh (1987), the estimator of $\gamma(\theta) = \theta$, has a negative binomial distribution which is a one-parameter exponential family. This explains why attainment of Cramér-Rao lower bound is expected in this situation.

**Remark 9.** A particular version of the inequality, obtained byWolfowitz (1947), was obtained in the same year by Blackwell and Girshick (1947) who restricted their attention to a situation where the underlying parameter itself is to be estimated and the sum of observations is a sufficient statistic. They also proved that the corresponding inequality is attainable only with a fixed sample size procedure. Also, Simons (1980) gave elegant construction of unbiased sequential estimators of the normal mean which attain smaller variance than the corresponding Cramér-Rao bound for a single $\theta$. Simons (1980) conjectures that such $\theta$'s will have Lebesgue measure zero. Finally, let us also mention that Stefanov (1985) studied efficient sequential estimation, in the Cramér-Rao sense, in stochastic processes whose corresponding sufficient statistics are processes with stationary independent increments. Ghosh's (1987) proof of Theorem 5, which is the basic result on non-existence, rests on Stefanov's (1985) condition that for attainment of the lower bound for all $\theta$, $S_N$ must be a linear function of $N$.

## 4. SEQUENTIAL CRAMÉR-RAO AND BHATTACHARYYA LOWER BOUNDS: COMPARISON AND A CONJECTURE ON ATTAINMENT OF THE LATTER

The (sequential) Bhattacharyya bounds are greater than or equal to the (sequential) Cramér-Rao bounds. Some more details are given below.

(1) For the family of normal distributions with mean $\theta$ and with Wald's stopping time corresponding to testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, Seth (1949) has considered the sequential Bhattacharyya lower bound with $k = 2$, where $k$ is the maximum order of partial derivative of the joint probability density function with respect to the parameter, assumed to exist (vide Theorem 3). He has proved that it, i.e., the sequential Bhattacharyya lower bound, is greater than the corresponding Cramér-Rao lower bound. The algebraic steps proceed along the following direction.

Recall from Remark 5 that $\phi_1(N) = \Sigma_{i=1}^{N}(X_i - \theta) = S_N - N\theta$ and $\phi_2(N) = (S_N - N\theta)^2 - N$. The fact that sequential Bhattacharyya lower bound is greater than the corresponding Cramér-Rao lower bound is a consequence of the fact that $E_\theta[\phi_1(N)\phi_2(N)]$ is not identically zero in $\theta$. This last fact is implied by the fact that $E_\theta(N)$ is not identically zero in $\theta$.

(2) For the family of Bernoulli distributions with mean $p$ and with Wald's stopping time corresponding to testing $H_0 : p = p_0$ against $H_1 : p = p_1$, he has proved that with $k = 2$, the Bhattacharyya inequality for estimating $\gamma(p) = p$ ceases to be an equality for every $\theta$. The algebraic steps are essentially same as those in the preceding example.

In the light of our heuristic argument to explain why the non-existence result of Ghosh (1987) is expected to be true, it is clear that one should be able to prove a similar general result on attainment of Bhattacharyya bound in the sequential case for exponential families.

## 5. CONCLUDING REMARKS

We have surveyed the contributions of Professor Seth relating to sequential lower bounds and orthogonal polynomials and previous or further work on these topics by distinguished contributors. As far as attainment of sequential lower bounds is concerned, the results are theoretically interesting but negative from a practical point if view. This often leads to a pessimism about all sequential estimation, see, e.g. Ghosh (1987). That such pessimism is unwarranted can be seen from many fruitful applications of sequential estimation in Ghosh *et al*. (1997), of which two- or three-stage bounded length confidence intervals are the most popular. They have been applied in survey sampling (with a pilot sample as the first-stage sample) and entomology. Point estimation can also benefit from sequential methods when the estimated variance is a function of observations, and one wishes to control its value at a pre-assigned level by sampling sequentially. See also the interesting remarks along similar lines in Simons (1980).

### ACKNOWLEDGEMENT

The authors are grateful to the referee for offering valuable suggestions.

### REFERENCES

Bhattacharyya, A. (1946, 1947). On some analogues of the amount of information and their use in statistical estimation. *Sankhyā*, **8**, 1-14, 201-208.

Blackwell, D. and Girshick, M.A. (1947). A lower bound for the variance of some unbiased sequential estimates. *Ann. Math. Statist*., **18**, 277-280.

Cramér, H. (1946a). *Mathematical Methods of Statistics*. Princeton University Press, NJ.

Cramér, H. (1946b). A contribution to the theory of statistical estimation. *Skand. Act. Tidskr*., **29**, 85-94.

Darmois, G. (1945). Sur les limites de la dispersion de certaines estimations. *Rev. Inst. Int. Statist*., **13**, 9-15.

DeGroot, M.H. (1959). Unbiased sequential estimation for binomial populations. *Ann. Math. Statist*., **30**, 80-101.

Fábian, V. and Hannan, J. (1977). On the Cramér-Rao inequality. *Ann. Statist*., **5**, 197-205.

Fréchet, M. (1943). Sur l'extension de certaines evaluations statistique de petits echantillons. *Rev. Inst. Int. Statist*., **11**, 182-205.

Ghosh, B.K. (1987). On the attainment of the Cramér-Rao bound in the sequential case. *Sequential Analysis*, **6**, 267-288.

Ghosh, J.K. and Sathe, Y.S. (1987). Convergence of Bhattacharya bounds-revisited. *Sankhyā*, **A49**, 37-42.

Ghosh, M., Mukhopadhyay, N., and Sen, P.K. (1997). *Sequential Estimation*. Wiley, New York.

Joshi, V.M. (1976). On the attainment of the Cramér-Rao lower bound. *Ann. Statist*., **4**, 998-1002.

Linnik, Yu.V. and Romanovsky, I.V. (1972). Some new results in sequential estimation theory. *Proc. Sixth Berkeley Symp. on Math. Statist. Probab*., 85-96.

Morris, C.N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist*., **10**, 65-80.

Müller-Funk, U., Pukelsheim, F., and Witting, H. (1989). On the attainment of the Cramér-Rao bound in $L_r$ differentiable families of distributions. *Ann. Statist*., **17**, 1742-1748.

Pitman, E.J.G. (1979). *Some Basic Theory for Statistical Inference*. Chapman & Hall, London.

Rao, C.R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc*., **37**, 81-91.

Seth, G.R. (1949). On the variance of estimates. *Ann. Math. Statist*., **20**, 1-27.

Shanbhag, D.N. (1972). Some characterizations based on Bhattacharyya matrix. *J. Appl. Probab*., **9**, 580-587.

Simons, G. (1980). Sequential estimators and the Cramér-Rao lower bound. *J. Statist. Plann. Inf*., **4**, 67-74.

Stefanov, V. (1985). On efficient stopping times. *Stoch. Proc. Appl*., **2**, 305-314.

Trybula, S. (1968). Sequential estimation in processes with independent increments. *Dissertationes Mathematicae*, **60**, 1-50.

Wijsman, R.A. (1973). On the attainment of the Cramér-Rao lower bound. *Ann. Statist*., **1**, 538-542.

Wolfowitz, J. (1947). The efficiency of sequential estimates and Wald's equation for sequential processes. *Ann. Math. Statist*., **18**, 215-230.