RAO-BLACKWELL ESTIMATOR OF DISTRIBUTION FUNCTION AND ITS USE IN DETERMINING BUNDLE STRENGTH OF FILAMENTS

by

B. B. Bhattacharyya

Department of Statistics
North Carolina State University, Raleigh, N. C.

Institute of Statistics Mimeo Series No. 1196

December 1978

RAO-BLACKWELL ESTIMATOR OF DISTRIBUTION FUNCTION AND ITS USE IN DETERMINING BUNDLE STRENGTH OF FILAMENTS

ABSTRACT

Small sample and large sample properties of the estimate of the bundle strength of filaments based on Rao-Blackwell empirical process have been studied. The sequence of statistics has been shown to be a reverse submartingale process. Martingal convergence theorems have been used to establish the properties of the statistics.

AMS Classification: Primary 60B10
Secondary 62B99

Key Words & Phrases: Gaussian process, Rao-Blackwell estimator, reverse martingale, transitive sufficiency, weak convergence, bundle strength

1. INTRODUCTION

Let $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$ be the ordered values of $X_1, X_2, \cdots X_n$ representing the strengths (nonnegative random variables) of individual filaments in a bundle of n parallel filaments of equal length. If we assume that the force of a free load on the bundle is distributed equally on each filament and the strength of an individual filament is independent of the number of filaments in a bundle, then the mimimum load B_n beyond which all filaments of the bundle give way is defined to be the strength of the bundle.

Now if a bundle breaks under load L, then the inequalities $nX_{n,1} \leq L$, $(n-1)X_{n,2} \leq L, \cdots, X_{n,n} \leq L$ are simultaneously satisfied. Hence the bundle strength B_n can be represented as

$$B_n = \max\{nX_{n,1}, (n-1)X_{n,2}, \cdots, X_{n,n}\}. \tag{1.1}$$

The small sample and large sample statistical properties of $B_n^* = B_n/n$ have been investigated by Daniels (1945), Suh-Bhattacharyya-Grandage (1970), Sen-Bhattacharyya-Suh (1973), Phoenix and Taylor (1973) and recently by Sen-Bhattacharyya (1976).

We observed that if $F_n(x)$ is the empirical distribution function of $X_1, X_2, \cdots, X_n$, then $n^{-1}B_n$ can be represented as $\sup_{x\geq 0}\{x(1-F_n(x))\}$. In our previous papers this representation was used to establish the large sample properties of the distribution of mean bundle strength. In the present paper we assume that $\{X_i, i \geq 1\}$ is a sequence of iid random variables having an absolutely continous distribution function $F(x, \underset{\sim}{\theta})$ where $\underset{\sim}{\theta}$ belongs to a subset of a finite dimensional Euclidean space and that there exists a complete sufficient statistic (vector) $\underset{\sim}{T}_n$ where

$$\underset{\sim}{T}_n' = (T_{n1}, T_{n2}, \cdots, T_{nq}) \tag{1.2}$$

for some $q \geq 1$, q finite and independent on n .

Therefore the Rao-Blackwell empirical distribution function

$$\psi_n(x) = E\{F_n(x) | T_n\} \tag{1.3}$$

is the unique minimum variance unbiased estimator of $F(x, \underset{\sim}{\theta})$.

The object of the present paper is to study the small sample and large sample properties of the statistic $\sup_{x \geq 0}\{x(1 - \psi_n(x))\}$. Section 2 deals with the basic assumptions and preliminary notions. In Section 3 a few basic lemmas and convergence and moment convergence of $\sup_{x \geq 0}\{x(1 - \psi_n(x))\}$ have been established. The asymptotic normality has been proved in Section 4 and a few concluding remarks have been made in the final section.

2. ASSUMPTIONS AND NOTATIONS

Let us define

$$h_o(\underset{\sim}{\theta}) = \sup\{x(1 - F(x, \underset{\sim}{\theta}))\}. \tag{2.1}$$

(I) We assume that there exists a unique $x_o(\underset{\sim}{\theta})$ such that

$$h_o(\underset{\sim}{\theta}) = x_o(\underset{\sim}{\theta})\{1 - F(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta})\} \tag{2.2}$$

and for every $0 < \epsilon < h_o(\underset{\sim}{\theta})$, there esists a $\delta(\underset{\sim}{\theta}) > 0$ such that

$$\sup\{x(1 - F(x, \underset{\sim}{\theta})) ;\ |x - x_o(\underset{\sim}{\theta})| > \delta(\underset{\sim}{\theta})\} < h_o(\underset{\sim}{\theta}) - \epsilon . \tag{2.3}$$

(II) Moreover for $|x - x_o(\underset{\sim}{\theta})| \leq \delta\ (\underset{\sim}{\theta})$,

$$c_2|x - x_o(\underset{\sim}{\theta})|^{k_2} \leq h_o(\underset{\sim}{\theta}) - x(1 - F(x, \underset{\sim}{\theta})) \leq c_1|x - x_o(\underset{\sim}{\theta})|^{k_1} \tag{2.4}$$

where c_1, c_2 ($< \infty$ but may depend on $\underset{\sim}{\theta}$) and k_1, k_2, $1 \leq k_1$, $k_2 < \infty$ are postive constants.

(III) We also assume that

$$0 < F(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta}) < 1 \tag{2.5}$$

and $F(x, \underset{\sim}{\theta})$ has a continuous and finite density function $f(x, \underset{\sim}{\theta})$ for all $|x - x_o(\underset{\sim}{\theta})| < \delta(\underset{\sim}{\theta})$, where

$$0 < f(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta}) < \infty \tag{2.6}$$

(IV) In Bhattacharyya and Sen (1977) sufficient conditions for asymptotic normality of $n^{\frac{1}{2}}\{\psi_n(x) - F(x, \underset{\sim}{\theta})\}$ have been given. In the present paper we shall assume $n^{\frac{1}{2}}\{\psi_n(x) - F(x, \underset{\sim}{\theta})\}$ converges in distribution to a Gaussian random process with an appropriate topology and the Gaussian random process is continuous with probability one. Moreever there exists a subset $\chi \subset R^1$ such that for every $x_i \in \chi$, $i = 1, 2, \cdots, m$,

$$\lim_{n\to\infty} n\, E\{\psi_n(x_i) - F(x_i, \underset{\sim}{\theta})\}\{\psi_n(x_j) - F(x_j, \underset{\sim}{\theta})\} = h(x_i, x_j, \underset{\sim}{\theta}) \tag{2.7}$$

and the matrix $h(x_i, x_j, \underset{\sim}{\theta})$ is positive definite for every $m \geq 1$.

(V) $x_o(\underset{\sim}{\theta}) \in \chi$ for every $\underset{\sim}{\theta}$.

(VI) $E_{\underset{\sim}{\theta}}(X_i^2) < \infty$.

3. SOME BASIC LEMANS

Lemma 3.1. Let $\{X_i, i \geq 1\}$ be an arbitrary sequence of non-negative random variables (not necessarily stationary or independent), such that for some $p > 0$,

$$\sup_i \int_t^\infty x^p \, d\, P(X_i \leq x) = \beta(t) \to 0 \quad \text{as } t\to\infty. \tag{3.1}$$

Then for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\max_{1\le i\le n} X_i \ge \epsilon\, n^{\frac{1}{p}}\} = 0 . \tag{3.2}$$

Proof: For every $\epsilon > 0$, $\beta(\epsilon\, n^{\frac{1}{p}})\to 0$ as $n\to\infty$. Consequently there exists an an $n_o(\epsilon, \delta)$ such that

$$\beta(\epsilon\, n^{\frac{1}{p}}) \le \delta\, \epsilon^p \ \text{ for } n \ge n_o(\epsilon, \delta) \tag{3.3}$$

Also from (3.1),

$$\epsilon^p n\, P\{X_i \ge \epsilon\, n^{\frac{1}{p}}\} \le \int_{\epsilon\, n^{\frac{1}{p}}}^{\infty} x^p \, d\,P(X_i \le x) = \beta(\epsilon\, n^{\frac{1}{p}}) . \tag{3.4}$$

Hence by (3.3) and (3.4),

$$\begin{aligned} P\{\max_{1\le i\le n} X_i \ge \epsilon\, n^{\frac{1}{p}}\} &\le \sum_{i=1}^{n} P\{X_i \ge \epsilon\, n^{\frac{1}{p}}\} \\ &\le \sum_{i=1}^{n} \beta(\epsilon\, n^{\frac{1}{p}})/\epsilon^p\, n = \epsilon^{-p}\beta(\epsilon\, n^{\frac{1}{p}}) \le \delta \end{aligned} \tag{3.5}$$

for $n \ge n_o(\epsilon, \delta)$.

Lemma 3.2 For every $\epsilon > 0$, there exists a positive k_ϵ which is finite and independent of $\underset{\sim}{\theta}$ such that for some n_o,

$$P\{\mathrm{Sup}(n^{\frac{1}{2}}|\psi_n(x) - F(x, \underset{\sim}{\theta})|) \ge k_\epsilon\} < \epsilon\, ,\ n \ge n_o \tag{3.6}$$

uniformly in $\underset{\sim}{\theta}$.

Proof: By Markov inequality, the left hand side of (3.6) is bounded above $k_\epsilon^{-1} E\{\sup_x n^{\frac{1}{2}}|\psi_n(x) - F(x, \underset{\sim}{\theta})|\}$. Nothing that $\psi_n(x) = E\{F_n(x)|\underset{\sim}{T}_n\}$ and using Jensen's inequality on conditional expectation

$$E\{\sup_x n^{\frac{1}{2}}|\psi_n(x) - F(x, \underset{\sim}{\theta})|\} \le E\{\sup_x n^{\frac{1}{2}}|F_n(x) - F(x, \underset{\sim}{\theta})|\}. \qquad (3.7)$$

From Dvoretsky, Kiefer and Wolfowitz (1956, p646) we have

$$P\{\sup_x n^{\frac{1}{2}}(F(x, \underset{\sim}{\theta}) - F_n(x)) > r\} \le c_1 e^{-2r^2}$$

for all $r \ge 0$ where c_1 does not depend on n or θ. From this result the lemma follows.

Lemma 3.3 For every sequence $\{\epsilon_n\}$ of positive $\epsilon_n > 0$ with $\lim \epsilon_n = 0$,

$$\sup\{n^{\frac{1}{2}}|\psi_n(x) - \psi_n(x_o(\underset{\sim}{\theta})) - F(x, \underset{\sim}{\theta}) + F(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta})|;\ |x - x_o(\underset{\sim}{\theta})| \le \epsilon_n\} \xrightarrow{p} 0. \qquad (3.8)$$

Proof: The proof follows easily from the tightness of the Rao-Blackwell empirical process. (Bhattacharyya-Sen (1977)). Let $\sigma_n(X) = \sigma(X_1, X_2, \cdots, X_n)$ be the σ-field generated by $X_1, X_2, \cdots, X_n$ and $\sigma(\underset{\sim}{T}_n)$ be the σ-field generated by $\underset{\sim}{T}_n$. Also let $\mathcal{T}_n = \sigma(\underset{\sim}{T}_n, \underset{\sim}{T}_{n+1}, \cdots)$ and therefore $\mathcal{T}_n$ is a monotone non-increasing σ-field.

Lemma 3.4 $\{\sup x(1 - \psi_n(x)), \mathcal{T}_n\}$ is a reverse submartingale process.

Proof: Since $\{X_n, n \ge 1\}$ are iid random variables and $\{\underset{\sim}{T}_n\}$ is complete and sufficient, it follows from Bahadur (1954, p443) that it is transitive. Hence by Wijsman Theorem (Zacks 1971, p84) $\sigma_n(\underset{\sim}{X})$ and $\sigma(\underset{\sim}{T}_{n+1})$ are conditionally independent given $\sigma(\underset{\sim}{T}_n)$.

$$\begin{aligned}\psi_{n+1}(x) &= E\{F_n(x)|\sigma(\underset{\sim}{T}_{n+1})\} = E\{E\{F_n(x)|\sigma(\underset{\sim}{T}_n, \underset{\sim}{T}_{n+1})\}|\sigma(\underset{\sim}{T}_{n+1})\\ &= E\{\psi_n(x)|\sigma(\underset{\sim}{T}_{n+1})\} = E\{\psi_n(x)|\mathcal{T}_{n+1}\} \text{ a.s.}\end{aligned} \qquad (3.9)$$

Hence $\{\psi_n(x), \mathcal{T}_n\}$ is a reverse martingale process. From this the result follows immediately.

<u>Theorem 3.1</u> Under assumptions (VI) for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\sup_x |x(1-\psi_n(x)) - x(1-F(x, \underset{\sim}{\theta}))| > \epsilon\} = 0 \qquad (3.10)$$

<u>Proof</u>:

$$\begin{aligned} &\sup_{x\geq 0} |x(1-\psi_n(x) - x(1-F(x, \underset{\sim}{\theta}))| \\ &\leq \max\{ \sup_{0\leq x\leq X_{n,n}} X_{n,n}|\psi_n(x) - F(x, \underset{\sim}{\theta}|, \qquad (3.11) \\ &\qquad \sup_{x>X_{n,n}} x(1-F(x, \underset{\sim}{\theta}))\} \end{aligned}$$

By Lemma 3.1 $X_{n,n} = o_p(n^{\frac{1}{2}})$ and $n^{\frac{1}{2}} \sup_{0\leq x<\infty} |\psi_n(x) - F(x, \underset{\sim}{\theta})|$ is bound in probability by Lemma 3.2. Moreover $E_{\underset{\sim}{\theta}}(X^2) < \infty$ implies $\lim_{x\to\infty} x(1-F(x, \underset{\sim}{\theta})) = 0$ and hence $\sup_{x\geq X_{n,n}} x(1-F(x, \underset{\sim}{\theta})) \xrightarrow{p} 0$ as $X_{n,n}\to\infty$ as $n\to\infty$.

<u>Corollary 1</u> Under the conditions of Theorem 3.1, $\sup_x x(1-\psi_n(x))$ converges almost surely to $\sup_x x(1 - F(x, \underset{\sim}{\theta}))$.

<u>Proof</u>: Since

$$\begin{aligned} &|\sup_x x(1-\psi_n(x) - \sup_x x(1-F(x, \underset{\sim}{\theta}))| \\ &\leq \sup_x |x(1-\psi_n(x) - x(1-F(x, \underset{\sim}{\theta}))| \xrightarrow{p} 0, \end{aligned} \qquad (3.12)$$

and by Lemma 3.4, $\sup_x x(1-\psi_n(x))$ is a nonnegative reverse martingale which converges almost surely to a random variable, the Corollary follows.

<u>Corollary 2</u> If $x_{n\,o}(1-\psi_n(x_{n\,o})) = \sup_x x(1-\psi_n(x))$, then $x_{n\,o}$ converges almost surely to $x_o(\underset{\sim}{\theta})$.

<u>Proof</u>: Since $x(1-\psi_n(x))$ converges a.s. to the continous function $x(1-F(x, \underset{\sim}{\theta}))$ uniformly in x and $x_o(\underset{\sim}{\theta})$ is unique by (2.2), hence $x_{n\,o}$ converges a.s. to $x_o(\underset{\sim}{\theta})$.

Corollary 3 Under the conditions of Theorem 3.1 and Corollary 2, $\psi_n(x_{n\,o}) - F(x_o(\underset{\sim}{\theta})) \xrightarrow{as} 0$.

Proof: $\psi(x_{n\,o}) - F(x_o(\underset{\sim}{\theta})) = F(x_{n\,o}) - F(x_o(\underset{\sim}{\theta})) + \psi_n(x_{n\,o}) - F(x_{n\,o})$.

By Corollary 2 and continuity of F, $F(x_{n\,o}) - F(x_o(\underset{\sim}{\theta})) \xrightarrow{as} 0$ and also $\psi_n(x_{n\,o}) - F(x_{n\,o}) \xrightarrow{as} 0$. Hence the result.

Theorem 3.2 Under the same conditions as in Theorem 3.1,

$$\lim_{n\to\infty} E\{\sup_x x(1-\psi_n(x))\}^k \downarrow (x_o(\underset{\sim}{\theta}))^k \, (1-F(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta})^k \quad \text{for } k=1, 2, \cdots$$

Proof: This follows directly from the fact $\{\sup_x x(1-\psi_n(x))\}^k$, $k=1, 2, \cdots$ are reverse submartingale (Chung 1974, p238) sequences.

4. ASYMPTOTIC NORMALITY

Theorem 4.1 Under the conditions stated in Section 2 $n^{\frac{1}{2}}(\{\sup x(1-\psi_n(x)\} - h_o(\underset{\sim}{\theta}))$ converges in law to a normal distribution with mean 0 and variance $(x_o(\underset{\sim}{\theta}))^2 \, h(x_o(\underset{\sim}{\theta}), x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta})$.

Proof: Let us choose a $\delta(\underset{\sim}{\theta})$ satisfying (2.3) and (2.4). Then using (2.4) and (2.9), it is easy to show

$$\max\{\sup_x x(1-\psi_n(x);\ |x-x_o(\underset{\sim}{\theta})| > \delta(\underset{\sim}{\theta})\} \leq h_o(\underset{\sim}{\theta}) - \eta \tag{4.1}$$

where $\eta > 0$, in probability as $n\to\infty$.

With the definition k_2 (≥ 1) in (2.4) we let $\beta = 1/2k_2$ and put

$$I_n(\beta) = \{x:\ |x-x_o(\underset{\sim}{\theta})| < n^{-\beta}\log n\} \tag{4.2}$$

$$I_n^*(\beta) = \{x:\ |x-x_o(\underset{\sim}{\theta})| < \delta(\underset{\sim}{\theta}) \text{ but } x \notin I_n(\beta)\}. \tag{4.3}$$

Following the same lines of argument as in Sen-Bhattacharyya (1976), we can show

$$\sup_x\{x(1-\psi_n(x));\ x \in I_n^*(\beta)\} \leq h_o(\underset{\sim}{\theta}) - d\,n^{-\frac{1}{2}}(\log n)^{k_2} \qquad (4.4)$$

in probability as $n\to\infty$ where $0 < d < c_2$.

Finally, for $x \in I_n(\beta)$,

$$|\text{Sup } x(1-\psi_n(x)) - h_o(\underset{\sim}{\theta}) - x_o(\underset{\sim}{\theta})\{\psi_n(x_o(\underset{\sim}{\theta})) - F(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta})\}| = o_p(n^{-\frac{1}{2}}) \qquad (4.5)$$

Since $x_o(\underset{\sim}{\theta})\{\psi_n(x_o(\underset{\sim}{\theta})) - F(x_o(\underset{\sim}{\theta}), \underset{\sim}{\theta})\}$ is $o_p(n^{-\frac{1}{2}})$, it follows from (4.1), (4.4) and (4.5)

$$\text{Sup } x(1-\psi_n(x)) = \{\text{Sup } x(1-\psi_n(x));\ x \in I_n(\beta)\} \qquad (4.6)$$

in probability as $n\to\infty$.

Combining (4.5), 4.6) and assumptions (IV), (V) of Section 2 Theorem 4.1 is established.

5. CONCLUDING REMARKS

An alternative method of proving Theorem 4.1 may be given using Continous Mapping Theorem (Billingsley 1968) as was done by Phoenix and Taylor (1973). We can also use the method used in this paper to establish asymptotic properties of a general class of nonnegative statistics expressible as an extremum of a functional of Rao-Blackwell empirical process.

Acknowledgement:

I am indebted to Dr. P. K. Sen of the University of North Carolina at Chapel Hill for helpful discussions.

REFERENCES

[1] Bahadur, R. R. (1954). Sufficiency and statistical decision function. Ann. Math. Stat. 25:423-462.

[2] Bhattacharyya, B. B. and Sen, P. K. (1977). Weak convergence of the Rao-Blackwell estimator of a distribution function. Annals of Probability 5:500-510.

[3] Billingsley, P. (1968). Convergence of Probability Measures. John Wiley, New York.

[4] Chung, Kai Lai (1974). A Course in Probability Theory. Academic Press, New York.

[5] Daniels, H. E. (1945). The statistical theory of the strength of a bundle of threads. Proc. Roy. Soc. A. 183:405-425.

[6] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of classical multinormal estimator. Ann. Math. Stat. 27:642-669.

[7] Phoenix, S. I. and Taylor, H. M. (1973). The asymptotic strength distribution of a fiber bundle. Advances in Applied Probability. 5:200-216.

[8] Sen, P. K. Bhattacharyya, B. B. and Suh, M. W. (1973). Limiting behaviour of the extremum of a certain sample function. Ann. Statistics. 1:297-311.

[9] Sen, P. K. and Bhattacharyya, B. B. (1976). Asymptotic normality of the extremum of a sample function. Z. Wahrscheinlichkeitstheorie Verw Gebiete 14:113-118.

[10] Suh, M. W. Bhattacharyya, B. B. and Grandage, A. (1970). On the distribution and moments of the strength of a bundle of filaments. Journal of Applied Probability 7(3):712-720.

[11] Zacks, S. (1971). The Theory of Statistical Inference. John Wiley. New York.