مجله علوم دانشگاه تهران جلد ۱۰ شماره‌های ۳و۴ سال ۱۳۶۰ ازصفحه ۲۷ تا ۳٤

چند خاصیت بهینگی بر آوردهای برداری بیز تجربی

دکتر محمد رضا مشکانی

گروه آمار دانشکدهٔ جامع انفورماتیك و مدیریت ، دانشگاه شهید بهشتی ، اوین ۱۹۸۳۲ تهران

چکیده

دراین مقاله به‌بررسی بعضی خواص بهین برآوردکننده‌های بیزتجربی مربوط به‌پارامترهای برداری توزیعهای آماری می‌پردازیم. هدف آن است که نتایجی شبیه به‌نتایج مربوط به‌پارامترهای اسکالری بیابیم. برای دستیابی به این هدف باالهام‌گرفتن ازنتایج مربوط به‌برآوردکننده‌های یک‌بعدی، مفاهیم ذیربط رادر مورد برآوردکننده‌های برداری تعریف کرده، شرایط بهینگی آنها را تعیین میکنیم. این مفاهیم عبارتند از سازگاری، بهینگی دراحتمال وبهینگی مجانبی. بااستفاده از روابط موجود بین وجوه مختلف‌گرایشهای‌متغیرهای تصادفی ثابت میکنیم که بهینگی مجانبی مولفه‌های مربوط به بردار برآورد کننده بیزتجربی شرط کافی برای بهینگی مجانبی بردار برآورد کننده است. بدین‌ترتیب، راه آسانتری برای امتحان کردن بهینگی مجانبی فراهم میکنیم که درکارهای آماری سودمند تواند بود.

Some optimal properties of multivariate empirical Bayes procedures

Dr. M. R. Meshkani

Department of Statistics, Shahid Beheshti University, Eveen 19832, *Tehran, Iran*

Abstract

This work addresses itself to the study of some potimal properties of Empirical Bayes estimators of the vector value parameters of the distributions. The goal is to provide some tools similar to those found for scalar parmeters. Relevant notions are defined and conditions of optimality for such estimators are stated. Utilizing the relations between - different modes of convergence, it has been established that componentwise asymptotic optimality is sufficient for vector asymptotic optimality. Thus, providing an easier way to check for asymptotic optimality.

1. Introduction

With a view to future use, we shall extend some aspects of the Empirical Bayes (EB) approach for univariate problems to provide some tools for dealing with multivariate problem. These results can then be applied to some specific problems in our later works. One application is seen in Mohammad-zadeh (1986).

The EB approach to statistical decision problems is applicable when we encounter the same decision problem in a sequence repeatedly and independently with a fixed but unknown prior distribution for the parameter. we do not expect all decision problems in practice to be embedded in such a sequence. However, when they are, the EB approach offers certain advantages over any approach which ignores the fact that the parameter is itself a random variable. This approach also has advantages over approaches which assume a personal prior not changing with experience, Robbins (1964).

The statistical decision problem with which we shall be concerned is of the following character. Let R^m denote the m-dimensional Euclidean space. Suppose we have :

(1) a parameter space $\Omega \subset R^m$,

(2) a non-observable random vector Λ on Ω with a generic distribution function (d. f.) G,

(3) an observable random vector $\mathbf{X}$ on a space $\mathcal{X}$ with conditional d. f. $F(\cdot \mid \Lambda)$, $\Lambda \in \Omega$,

(4) an action space $\mathcal{A}$ with elements $\mathbf{A}$, and

(5) a loss function $L(\mathbf{A}, \Lambda) \geq 0$, $\mathbf{A} \in \mathcal{A}$, $\Lambda \in \Omega$.

The problem is to choose a decision function $\mathbf{d}: \mathcal{X} \rightarrow \mathcal{A}$ such that upon observing $\mathbf{x}$ we shall take the action $\mathbf{d}(\mathbf{x})$ and incur the loss $L(\mathbf{d}(\mathbf{x}), \Lambda)$. The Bayes risk relative to G is defined as

$$W(\mathbf{d}, G) = \int_\Omega \int_{\mathcal{X}} L[\mathbf{d}(\mathbf{x}), \Lambda]\, dF(\mathbf{x} \mid \Lambda)\, dG(\Lambda) \quad (1.1)$$

Any decision function $\mathbf{d}_G$ which minimizes (1.1) is called a Bayes rule relative to G. When G is known, the Bayes risk of $\mathbf{d}_G$ is

$$W(G) = W(\mathbf{d}_G, G) = \inf_d W(\mathbf{d}, G) \quad (1.2)$$

There remains the question of what to do when G is not known. It is assumed that G exists. Thus, $W(\mathbf{d}, G)$ is an appropriate criterion for the performance of any $\mathbf{d}$.

The standard EB assumption is that there are independent pairs of random vectors $(\Lambda, \mathbf{X})$ with outcomes

$$(\Lambda_1, \mathbf{x}_1), \ldots, (\Lambda_n, \mathbf{x}_n); (\Lambda_{n+1} \equiv \Lambda, \mathbf{x}_{n+1} \equiv \mathbf{x}) \quad (1.3)$$

The values of Λ_i, $i \in N_1 = \{1, \ldots, n+1\}$, always remain unknown. At the $(n+1)$ **st** stage, a decision is to be made about Λ. we have the «past data» $\mathbf{x}_i$, $i \in N = \{1, \ldots, n\}$, which can be used to construct a decision function about Λ of the form

$$\mathbf{d}_n(\mathbf{x}) = \mathbf{d}(\mathbf{x}_1, \ldots, \mathbf{x}_n; \mathbf{x}) = \{\ldots, d_{nj}(\mathbf{x}), \ldots\} \quad (1.4)$$

The function (1.4) which maps $\mathcal{X}$ into $\mathcal{A}$ is called an EB rule. In the estimation problem, the rule is an estimator. Then, the EB estimator, being a function of the «past data», has a Bayes risk

$$E_n[W(\mathbf{d}_n, G)] = E_n\{\int_\Omega \int_{\mathcal{X}} L[\mathbf{d}_n(\mathbf{x}), \Lambda]\, dF(\mathbf{x} \mid \Lambda)\, dG(\Lambda)\} \quad (1.5)$$

where E_n is the expectation with respect to the joint distribution of $\{\mathbf{x}_i, i \in N\}$. By definition,

$$E_n[W(d_n, G)] \geq W(G) \quad (1.6)$$

A decision is called asymptotically optimal (a. o.) relative to G, if

$$\lim_{n \rightarrow \infty} E_n[W(\mathbf{d}_n, G)] = W(G)$$

Therefore, it is reasonable to believe that a. o. EB - estimators will share optimal properies of the Bayes estimators, if there is enough past information.

We can represent a $k \times l$ matrix as a vector of kl components. This representation makes possible to consider the decision problems for matrices in the framework of the above formulation. In fact, the above formulation is suitable for three levels of complication. When $k = l = 1$, we have one parameter $K > 1$, $l = 1$ give a k - vector of parameters; and k, $l > 1$ lead to a $k \times l$ matrix of parameters.

2. Review of the univariate case

In this section, we shall review the relevant results from the univariate theory of the EB approach to statistical decision problem. The problem of special interest to us is the estimation of a parameter for a probability distribution.

The idea of EB estimation began with Robbins (1955). He derived a. o. estimates for parameters of certain common univariate distributions. Neyman (1962) called it a breakthrough in the theory of statis - tical decision making. To avoid repeating ourselves, consider the equations (1. 1) - (1. 3) written for scalars d, X, and Λ; denote them by W (d, G), W(G), and (Λ_i, x_i), $i \varepsilon N_1$, respectively. Accordingly, we consider (1. 4) when we have a scalar $d_n(x)$. Since x is a realization of $X \equiv X_{n+1}$, it will be considered fixed in the spirit of Bayesian theory. Thus, in the sequel the expectation and limit operators refer to the joint distribution of $\{X_i,\ i \varepsilon N\}$. Some desirable properties have been defined for the estimators.

Definition 2. 1.

An EB estimator, $d_n(x)$, is said to be «*consistent*» relative to G, if it converges in probability to the respective Bayes estimator, $d_G(x)$, as the size of «past data» increases indefinitely. This is denoted by

$$d_n(x) \xrightarrow{p} d_G(\mathbf{x})$$

Definition 2. 2. (Robbins, 1964)

An EB estimator, $d_n(x)$, is said to be *asymptotically optimal* (a. o.), relative to G, if

$$\lim_{n \to \infty} E_n \{ W [d_n(x), G] \} = W(G)$$

where E_n is the expectation with respect to the joint distribution of $\{X_i,\ i \varepsilon N\}$. Robbins (1964) gave a theorem to verify the a. o. of an EB estimator. Rutherford and Krutchkoff (1969) noted that for an unbounded loss function, however, including the popular squared errors loss, Robbins' condition

$$\int_\Omega \{\sup_d L(d, \Lambda)\, dG(\Lambda)\} < \infty$$

may not hold. To amend, they proposed the weaker notion of ε - a. o. which means for any arbitrary $\varepsilon > 0$

$$\lim_{n \to \infty} E_n \{ W [d_n(x), G] \} \leqslant W(G) + \varepsilon$$

This holds for prior distribution with $E\{|\Lambda|^{2+\gamma}\} < \infty$, for some $\gamma > O$. We note that this restriction on the class of prior distributions for Λ holds for most priors wich are useful in practice.

To relax the assumptions on the prior, Maritz (1970) proposed the truncated EB estimators and asymptotic optimality in probability a. o. (p).

Definition 2. 3.

An EB estimator, $d_n(x)$, is said to be *asymptotically optimal in probability* a. o. (p), relative to G, if

$$W[d_n(\mathbf{x}), G] \xrightarrow{p} W(G), \text{ as } n \to \infty$$

We shall show that a. o. (p) is equivalent to a. o. under some conditions.

Let A and B be finite, although possibly large negative or positive numbers. If Λ is known to be bounded in either direction, A and B are taken as those appropriate bounds.

Theorem 2. 4. (Maritz, 1970)

Let $d_n(x)$ be an EB estimator of $\wedge$, truncated at A and B. Let $d_G(x)$ be the corresponding Bayes estimator of $\wedge$. Assume

$$\int_\Omega \int_{\mathfrak{X}} d^2 G(x)\, dF(x \mid \wedge)\, dG(\wedge) < \infty \tag{2. 1}$$

If $d_n(x) \xrightarrow{p} d_G(x)$ as $n \to \infty$ for all $x \in \mathfrak{X}$ such that $d_G(x) \in (A, B)$, then $d_n(x)$ is a. o. (p). Squared error loss is assumed. □

We observe that (2. 1) is not a restrictive condition. If (2. 1) does not hold, then the Bayes estimator has an infinite Bayes risk and consideration of an EB estimator does not make sense. We shall see later that theorem 2. 4 in conjunction with an observation, which we shall make, will prove helpful in establishing a. o. of our estimators.

Recently, Deely and Zimmer (1976) returned to Robbins' idea of a. o. and proved it under a weaker condition.

Theorem 2. 5. (Deely and Zimmer, 1976)

Let $d_n(x)$ be a consistent EB estimator of $\wedge$. Let the loss function be such that,

$$L[d_n(\mathbf{x}), \wedge] \xrightarrow{p} L[d_G(\mathbf{x}), \wedge],\ \wedge \in \Omega \text{ as } n \to \infty$$

Suppose there exists a sequence of functions

$h_n(\mathbf{x}, \wedge) = h_n(\mathbf{x}_1, \dots, \mathbf{x}_n; \mathbf{x}, \wedge)$ such that for all $x \in \mathfrak{X}$, $\wedge \in \Omega$

(i) $h_n(\mathbf{x}, \wedge) \xrightarrow{p} h(\mathbf{x}, \wedge)$,

(ii) $L[d_n(\mathbf{x}), \wedge] \le h_n(\mathbf{x}, \wedge)$ for $n = 1, 2, \dots,$

and

(iii) $\lim_{n \to \infty} E\, h_n[(\mathbf{x}, \wedge)] = E[\lim_{n \to \infty} h_n(\mathbf{x}, \wedge)] < \infty$

Then, $d_n(x)$ is a. o.

The proof is based on a result known as the extended dominated convergence theorem. We shall use it to prove multivariate version of Theorem (2. 5) later. Therefore, it is presented here as a lemma.

Lemma 2. 6. (pratt, 1960)

Let $(\mathfrak{X}, \mathcal{B}, P)$ be a probability space and let $\{f_n\}$ and $\{g_n\}$ be two sequences of measurable functions such that

(i) $f_n \xrightarrow{p} f$, $g_n \xrightarrow{p} g$,

(ii) $o \le f_n \le g_n$ for $n = 1, 2 \dots,$

and

(iii) $\lim_{n \to \infty} \int g_n\, dp = \int g\, dp < \infty$.

Then,

$$\lim_{n \to \infty} \int f_n\, dp = \int f\, dp < \infty$$

In theorem 2. 5, $d_n(x) \xrightarrow{p} d_G(x)$ will imply

$$L[d_n(x), \wedge] \xrightarrow{p} L[d_G(x), \wedge]$$

if $L(.,.)$ is continuous in the first argument. Convergence in probability with respect to X_i, $i = 1, 2, \dots$ at each $(X, \wedge)$ implies convergence in probability on the product pace $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \dots \times \mathfrak{X}_n \times \Omega$.
Now set $f_n = L[d_n(x), \wedge]$ and $g_n = h_n(x, \wedge)$. The result follows from Lemma - (2. 6.).
It is observed that $W[d_n(x), G]$ is a r. v. depending on $\{X_i, i \in N\}$. In fact, definition (2. 2) states its convergence in the first mean and definition (2. 3.) states its convergence in probability to W (G). by virtue of the following lemma, we show that these two definitions are equivalent.

Lemma 2. 7. (Serfling, 1976)

Let X_n and X be generic random variables. Suppose $X_n \xrightarrow{P} X$ as $n \to \infty$ and there is a random variable Y such that $|X_n| \leq |Y|$ with probability one (w. p. 1) for all n. If $E(|Y|^r) < \infty$ then X_n converges to X in **r th** mean, $r \geq 1$.

Theorem 2. 8

Let $d_n(x)$ be an EB estimator of $\wedge$ truncated at A, B. Let $\wedge$ have a prior d. f. G such that $E(\wedge^2) < \infty$. If the loss function is a squared error loss, $L(d, \wedge) = (d - \wedge)2$, then $W[d_n(x), G] < \infty$ w. P. 1.

Proof.

From definition (1. 1)

$$O \leq W[d_n(x), G] = E[d_n(x) - \wedge]^2 \leq$$
$$2\{E[d_n(x)]^2 + E(\wedge^2)\}$$

by C_r - inequality [Loève (1977, p. 197)]. Since $d_n(x)$ belongs to (A, B),

$$E[d_n(x)]^2 \leq (B-A)^2 < \infty \text{ W.p.1.}$$

Now, if we define Y in Lemma 2. 7 by

$$p\{Y = 2E[d_n(x)]^2 + 2E(\wedge^2) = 1,$$

then $E(|Y|) < \infty$.

We shall be dealing with estimation of parameters. Therefore, it is always true that $E(\wedge^2) < \infty$. In this case we see that the conditions of theorem 2. 4 are more easily checked than those of theorem 2. 5. Whenever we are dealing with squared error loss , establishment of a. o. (p)implies . a. o This is a consequence of lemma 2. 7 and theorem 2. 8.

All the previous results concern the large sample properties of the EB estimators. The distribution of the EB estimators for small sizes of the past data are extremely hard to derive. Therefore, their small sample properties have generally only been sudied by simulation, Maritz (1970).

3. Extention to multivariate case

The results known for the univariate case can be extended to include the multivariate case, i. e., to include the vector valued quantities. Therefore, we develop here the appropriate results for the multivariate theory.

In the sequel, we shall frequently use the notion of convergence of a vector to another vector either in probability or in distribution. Convergence in prob - ability can be defined in different ways, depending on the metric used. we give the following definition and a necessary and sufficient condtition for convergence of a vector in terms of convergence of its components.

Definition 3. 1.

Let $\mathbf{X}$ and $\mathbf{X}_n$, n=1, 2, ... be s - dimensional random vectors on a probability space $(\theta, \mathcal{F}, p)$. The random vector X_n *converges in probability* to X if for every $\varepsilon > 0$,

$$p(\theta: \|\mathbf{X}_n - \mathbf{X}\| > \varepsilon) \to O, \text{ as } n \to \infty$$

where $\|\mathbf{X}\| = \left[\sum_{j\varepsilon S} x_j^2\right]^{1/2}$.

This is denoted by $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$.

Lemma 3. 2.

Let $\mathbf{X}$ and $\{\mathbf{X}_n, n = 1, 2, \ldots\}$ be s - dimensional random vectors on a probability space $(\theta, \mathcal{F}, p)$. The random vector $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ if and only if $X_{nj} \xrightarrow{p} X_j$ for all j.

Proof.

That this is a necessary condition is obvious. Let us prove that it is also a sufficient condition. For every $\varepsilon > 0$, let

$$A = \{\theta: \left[\sum_{j\varepsilon S} (X_{nj} - X_j)\right]^{1/2} > \varepsilon\}$$

and

$$A_j = \{\theta: |X_{nj}| > [\varepsilon^2/_s]^{1/2}\}, \qquad j\varepsilon S.$$

Thus, we need to show $A \subset \bigcup_{j\varepsilon S} A_j$. If $\theta \notin A_j$, for all j, then $\theta \notin \bigcup_{j\varepsilon S} A_j$ That is,

$$(X_{nj}-X_j)^2 \leq \varepsilon^2/s,\ j\ \varepsilon\ S.$$

Thus,

$$\sum_{j\varepsilon S} (X_{nj}-X_j)^2 = \| \mathbf{X}_n - \mathbf{X} \|^2 \leq \varepsilon^2$$

and $\theta \notin A$. Therefore, $A \subset \bigcup_{j\varepsilon S} A_j$.

Definition 3. 3.

A vector valued EB estimator $\mathbf{d}_n(\mathbf{x})$, is said to be «*consistent*» relative to G, if it converges in probability to the respective Bayes estimate, $\mathbf{d}_G(\mathbf{x})$ as the size of the «past data» increases indefinitely. This is denoted by

$$\mathbf{d}_n(\mathbf{x}) \xrightarrow{p} \mathbf{d}_G(\mathbf{x}).$$

Definition 3. 4.

A vector valued EB estimator, $\mathbf{d}_n(\mathbf{x})$ is said to be *asymptotically optimal* (a. o.) relative to G, if

$$\lim_{n\to\infty} E_n \{W[\mathbf{d}_n(\mathbf{x}),G]\} = W(G),$$

where E_n is the expectation with respect to the joint distribution of $\{\mathbf{X}_i,\ i\ \varepsilon\ N\}$.

Definition 3. 5.

A vector valued EB estimator, $\mathbf{d}_n(\mathbf{x})$ is said to be *asymptotically optimal in probability* a. o. (p), relative to G, if

$$W[\mathbf{d}_n(\mathbf{x}), G] \xrightarrow{p} W(G), \text{ as } n\to\infty$$

where p refers to the joint distribution of $\{\mathbf{X}_i,\ i\ \varepsilon\ N\}$.

Theorem 2. 5 can now be stated for vector valued estimator .

Theorem 3. 6.

Let $\mathbf{d}_n(\mathbf{x})$ be a «consistent» EB estimator of the parameter vector Λ. Let the loss function be such that

$$L[\mathbf{d}_n(\mathbf{x}), \Lambda] \xrightarrow{p} L[\mathbf{d}_G(\mathbf{x}), \Lambda]$$

for each Λ. Suppose there exists a sequence of functions $h_n(\mathbf{x}, \Lambda) = h_n(\mathbf{x}_1, \ldots, \mathbf{x}_n; \mathbf{x}, \Lambda)$ such that for all n and Λ

(i) $h_n(\mathbf{x}, \Lambda) \xrightarrow{p} h(\mathbf{x}, \Lambda)$,

(ii) $L[\mathbf{d}_n(\mathbf{x}), \Lambda] \leq h_n(\mathbf{x}, \Lambda)$, for $n = 1,2,..$

and

(iii) $\lim_{n\to\infty} E[h_n(\mathbf{x}), \Lambda)] = E\{\lim_{n\to\infty} h_n(\mathbf{x}, \Lambda)\} < \infty$.

Then, $\mathbf{d}_n(\mathbf{x})$ is a. o.

Proof.

The proof is similar to that of Theorem 2. 5 where we now et

$$f_n = L[\mathbf{d}_n(\mathbf{x}), \Lambda] \text{ and } g_n = h_n(\mathbf{x}, \Lambda).$$

The above theorem is not very helpful in establishing a. o. of an EB estimator $\mathbf{d}_n(\mathbf{x})$ because finding a sequence $h_n(\mathbf{X}, \Lambda)$ may not be easy. Thus, utillzing the continuous nature of the squared error loss function, we shall relegate a.o. of a vector to a. o. of its components.

From now on, we specialize $L(\mathbf{d}, \Lambda)$ to

$$L(\mathbf{d}, \Lambda) = \sum_{j\varepsilon S} L_j(d_j, \Lambda_j) = \sum_{j\varepsilon S} (d_j - \Lambda_j)^2 \tag{3. 1}$$

[see De Greoot (1970)]. Consequently,

$$W(\mathbf{d}, G) = \sum_{j\varepsilon S} W_j(d_j, G_j) \tag{3. 2}$$

where

$$W_j(d_j, G_j) = \int_\Omega \int_{\mathscr{X}} (d_j - \Lambda_j)^2\, dF(\mathbf{x} \mid \Lambda)$$

Lemma 3. 7.

For $L(\mathbf{d}, \Lambda)$ given in (3. 1), the minimum Bayes risk for $\mathbf{d}$ is achieved by having minimum Bayes risk for all d_j, $j\ \varepsilon\ S$.

Proof.

From (1. 1),

$$W(\mathbf{d},G)=\sum_{j\varepsilon S}\int_{\Omega}\int_{\mathfrak{X}}(d_j-\Lambda_j)^2 dF(\mathbf{x}\mid\Lambda)dG(\Lambda)$$

$$=\sum_{j\varepsilon S} W_j(d_j, G_j).$$

Now, it is clear that if d_j has the minimum Bayes risk $W_j(\mathbf{d}_j, G_j)$ for all $j \varepsilon S$, then $W(\mathbf{d}, \mathbf{G})$ achieves its minimum.

To obtain the Bayes estimate $\mathbf{d}_G$ relative to G, we need to obtain Bayes estimate of d_{Gj} relative to G_j for each j.

Theorem 3. 8.

For a squared error loss function, a vector valued EB estimator, $\mathbf{d}_n(\mathbf{x})$, is a. o. (p) if and only if $\mathbf{d}_{nj}(\mathbf{x})$ is a. o. (p) for all $j \varepsilon S$.

Poorf.

From (3. 2), it is obvious that if $\mathbf{d}_{nj}(\mathbf{x})$ is not a. o. (p) for all j, then there is one j such that

$$W_j[d_{nj}(\mathbf{x}), G_j] \overset{p}{\not\Rightarrow} W_j(G_j).$$

That is,

$$W[\mathbf{d}_n(\mathbf{x}), G] = \sum_{j\varepsilon S} W_j[d_{nj}(\mathbf{x}), G_j]$$

$$\overset{p}{\not\Rightarrow} \sum_{j\varepsilon S} W_j(G_j) = W(G),$$

and $\mathbf{d}_n(\mathbf{x})$ is not a. o. (P). on the other hand, if $\mathbf{d}_{nj}(\mathbf{x})$ is a. o. (p) for all $j \varepsilon S$, then

$$W_j[d_{nj}(\mathbf{x}), G_j] \xrightarrow{p} W_j(G_j).$$

Hence,

$$W[\mathbf{d}_n(\mathbf{x}), G] = \sum_{j\varepsilon S} W_j[\mathbf{d}_{nj}(\mathbf{x}), G_j]$$

$$\xrightarrow{p} \sum_{j\varepsilon S} W_j(G_j) = W(G).$$

Corollary 3. 9.

Let $\boldsymbol{\Lambda}$ have $E(\boldsymbol{\Lambda}_j^{\,2}) < \infty$ for all $j \varepsilon S$. Let $\mathbf{d}_n(\mathbf{x})$ be the EB estimator relative to the squared error loss function. Then definitions 3. 4 and 3. 5 are equivalent.

Proof.

That a. o. implies a. o. (p) follows from the fact that convergence in the rth mean implies convergence in probability. To show the other direction, we note that by Lemma 2. 7 and Theorem 2. 8 for all $j \varepsilon S$,

$$E_n\{|W_j[d_{nj}(\mathbf{x}), G_j]-W_j(G_j)|\}\to O, \text{ as } n\to\infty.$$

Thus,

$$E_n\{|W[\mathbf{d}_n(\mathbf{x}), G]-W(G)|\}=$$

$$E_n\{|\sum_{j\varepsilon S} W_j[\mathbf{d}_{nj}(\mathbf{x}), G_j]-\sum_{j\varepsilon S} W_j(G)|\}\leqslant$$

$$\sum_{j\varepsilon S} E_n\{|W_j[\mathbf{d}_{nj}(\mathbf{x}), G_j]-W_j(G_j)|\}\longrightarrow 0$$

i. e., $\mathbf{d}_n(\mathbf{x})$ is a. o.

Therefore, to establish a. o. of a vector valued EB estimator, we need to show that each component is a. o. (p).

References

Deely, J. J. and Zimmer, W. J. (1976) Asymptotic optimality of empirical Baye procedure. *Annal. of Statis.* **4.** 576 - 580.

DeGroot, M. H. (1970) *Optimal Statistical Decisions.* McGraw - Hill New York.

Loève M, (1977) *Probability Theory I,* 4th edition. Springer Verlag, New York.

Maritz, J. S. (1970) *Empirical Bayes Methods.* Methuen Monographs, Methuen, London.

Mohammadzadeh, M. (1986) *Empirical Bayes estima - tion in exponential family.* M. S. Thesis, University of Tarbeyat Modarress, Tehran (in Persian).

Neyman, J. (1962) Two breakthroughs in the theory of Statistical decision making. *Rev. Inter. Statis. Inst.* **30**, 11 - 27.

Pratt, J. W. (1960) On interchanging limits and integrals, *Annal. of Math. Statis.* **31**, 74 - 77.

Robbins. H. (1955) An empirical Bayes approach to Statistics. *Proc. Third Berkeley Symp. Math. Statis. Prob.* **1**, 157 - 164.

Robbins, H. (1964) The empirical Bayes approach to Statistical decision problems. *Annal. of Statis.* **35**, 1 - 20.

Rutherford, J. R. and Krutchkoff, R. G. (1969) - ε - Asymptotic optimality of empircal Bayes estimator. *Biometrika* **56**, 220.

Serfling, R. J. (1976) *Approximation Theorems of Mathematical Statistics* Dept. of Statistics, FSU, - Tallahassee, Florida.

Susarla, V. (1982) Empirical Bayes Theory in *Ency - clopedia of Statistical Sciences* **vol. 2** Kotz, s. and Johnson, S. (eds.) John Wiley, New York.