The Annals of Statistics
1974, Vol. 2, No. 1, 118-133

RATE OF CONVERGENCE IN THE SEQUENCE-COMPOUND SQUARED-DISTANCE LOSS ESTIMATION PROBLEM FOR A FAMILY OF *m*-VARIATE NORMAL DISTRIBUTIONS[1]

BY V. SUSARLA

University of Wisconsin-Milwaukee

This paper is concerned with rates of convergence in the sequence-compound decision problem when the component problem is the squared-distance loss estimation of the mean of m-variate normal distribution with covariance matrix I. Section 1 introduces some notation and discusses the earlier work related to this problem. In Section 2, we prove two lemmas which are required in later sections. Sections 3, 4 and 5 exhibit sequence-compound decision procedures whose modified regrets are $O(n^{-1/m+4})$, near $O(n^{-\frac{1}{4}})$ and near $O(n^{-\frac{1}{2}})$ respectively, all the orders being uniform in parameter sequences concerned. In Section 6, comparisons have been made between the procedures given in Sections 3, 4 and 5. We conclude the paper with a few remarks.

1. Introduction. In order to describe the sequence-compound decision problem, we have to describe the setup of a statistical decision problem to which we refer, later on, as the component problem. We follow closely the notation of [3].

In the component problem there is a family of probability measures $\mathscr{P} = \{P_\theta \mid \theta \in \Theta\}$ defined on a σ-field $\mathscr{B}$ of subsets of $\mathscr{X}$, an action space A and a loss function L defined on $\Theta \times A$. The component problem is to decide about θ based on a realization x of a random variable X distributed as P_θ belonging to $\mathscr{P}$. For any randomized procedure η, let $R(\theta, \eta)$ denote the risk of using η to decide about θ and for any distribution G on Θ, let $R(G)$ denote the Bayes risk against G in the component problem.

Suppose we have a sequence of such component problems. Specifically let $\mathbf{X} = \{\underline{X}_n\}$ be a sequence of independent random variables with, for each n, $\underline{X}_n$ distributed as P_{θ_n} in $\mathscr{P}$. A sequence-compound decision problem is one in which, for any n, the decision about θ_n is allowed to depend on $\underline{X}_1, \cdots, \underline{X}_n$ and the loss at any particular stage n is taken to be the Cesaro sum of the losses in the first n component problems. For any sequence-compound decision procedure (for definition, among other references, see [3]) $\boldsymbol{\eta} = \{\eta_n\}$, let $R_n(\boldsymbol{\theta}, \boldsymbol{\eta})$ denote the Cesaro sum of the risks of using the first n components of $\boldsymbol{\eta}$ to decide about

Received April 1971; revised April 1973.

[1] Research supported by National Science Foundation, Grant No. GP-23480. This paper is based on the author's dissertation at Michigan State University.

AMS classifications. Primary 62C25; Secondary 62C99, 62C25.

Key words and phrases. Decision theory, squared-distance loss estimation, sequence-compound estimators.

the corresponding n components of $\boldsymbol{\theta} = \{\theta_n\}$ and let

$$D_n(\boldsymbol{\theta}, \boldsymbol{\eta}) = R_n(\boldsymbol{\theta}, \boldsymbol{\eta}) - R(G_n) \tag{1.1}$$

where G_n is the empiric distribution of $\theta_1, \cdots, \theta_n$. D_n is called the modified regret of $\boldsymbol{\eta}$.

For any simple symmetric sequence-compound decision procedure (again, for definition, see [3]) $\boldsymbol{\eta}$, it is known that $D_n(\boldsymbol{\theta}, \boldsymbol{\eta}) \geqq 0$ for all $\boldsymbol{\theta}$ and $D_n(\boldsymbol{\theta}, \boldsymbol{\eta}) \geqq \varepsilon > 0$ (ε is independent of n) for some $\boldsymbol{\theta}$ whenever $\inf_\eta \sup_\theta \{R(\theta, \eta) - \inf_a R(\theta, a)\} > 0$. However, some non-simple procedures have been exhibited whose D_n functions are bounded by a null sequence (often, this null sequence is independent of $\boldsymbol{\theta}$ and converges to zero with certain rate) for special choices of Θ, A, $\mathscr{P}$, and L. (see references [2], [3], [5], [6], [7], [10], [11] and [12].)

Commonly, the idea is to define the procedure $\boldsymbol{\eta}$ is such a way that its nth component is a natural estimate of $\psi_{n-1}(\underline{X}_n)$, where ψ_n is a Bayes response versus G_n, and show that $\mathbf{P}_n[|\eta_n - \psi_{n-1}(\underline{X}_n)|] \to 0$ (possibly with certain rate) when Θ, A, $\mathscr{P}$, and L satisfy certain conditions. This idea is related to the inequalities (8.8) and (8.11) of [6] which, in our notation, can be stated as

$$n^{-1} \textstyle\sum_{j=1}^n R(\theta_j, \psi_j) \leqq R(G_n) \leqq n^{-1} \sum_{j=1}^n R(\theta_j, \psi_{j-1}) \tag{1.2}$$

where ψ_0 is any arbitrary decision rule. The sequence-compound procedures exhibited in this paper use this idea.

The result most comparable to the results of this paper appears in [3], where a sequence-compound decision procedure with modified regret $O(n^{-\frac{1}{4}})$ (uniform in parameter sequences) is exhibited for the case when the component problem is the squared-distance loss estimation of the mean (lying in a compact set) of a univariate normal distribution. A generalization of this result is given in Section 3. For other rate results, see [3].

Hereinafter, we specialize the notation to the case of our interest. Let $\Theta = [-\alpha, \alpha]^m$, P_θ be the m-variate normal distribution with mean vector θ and the identity matrix as the covariance matrix and L be the squared-distance function. With p_θ denoting the multivariate normal density corresponding to P_θ, we have

$$\psi_n(x) = E[\theta \mid x] = \frac{\int \theta p_\theta(x)\, dG_n(\theta)}{\int p_\theta(x)\, dG_n(\theta)} = x + \bar{q}(x)/\bar{p}(x) \tag{1.3}$$
$$= x + (\log \bar{p})'(x)$$

with $\bar{p}$ denoting the mixed density $\int p_\theta\, dG_n$, $\bar{q}$ and $(\log \bar{p})'$ denoting the vectors of partial derivatives of $\bar{p}$ and $\log \bar{p}$ respectively. The two alternative forms given in (1.3) for ψ_n, the discussion around (1.2) and Lemma 2.2 of Section 2 of this paper motivated the definitions of the $(n+1)$th components of the sequence-compound procedures exhibited in Sections 3, 4 and 5. Specifically, the form $x + (\log \bar{p})'(x)$ for ψ_n has been exploited in Section 3 while the other form has been exploited in Sections 4 and 5.

The following notation, abbreviations, and conventions will be used throughout the paper:

Let μ denote the Lebesgue measure on (E^m, B^m). For any two points $u = (u_1, \cdots, u_m)$ and $v = (v_1, \cdots, v_m)$ in E^m, let $||u||^2 = \sum_{i=1}^m u_i^2$, $|u| = \sum_{i=1}^m |u_i|$ and $(u, v) = \sum_{i=1}^m u_i v_i$. The inequalities $||u|| \leqq |u| \leqq m^{\frac{1}{2}}||u||$ will be used without further comment. A vector in E^m will be denoted by $\langle\ \rangle$ with the general coordinate of the vector exhibited inside the brackets. $c_1, c_2, \cdots$ denote constants depending on m and α.

Throughout the paper, Φ and ϕ denote the standard normal distribution function and its density respectively. We suppress the arguments of functions whenever it is convenient to do so and use sets to denote their own indicator functions. Ratios of the form 0/0 are taken to be equal to unity. For any measure ν, the notation $\nu[f]$ or νf will often be used for $\int f\, d\nu$. P_n, p_n and $\mathbf{P}_n$ abbreviate P_{θ_n}, p_{θ_n} and the product measure $P_1 \times p_2 \times \cdots \times P_n$ respectively. $\sum$ stands for summation over j from 1 through n. In Sections 3 through 6 $\underline{X}_{n+1}$ has been abbreviated to $X = (X_1, \cdots, X_m)'$. ψ_n has been abbreviated to $\psi = \langle\psi_l\rangle$ in Sections 3, 4, and 5.

The orders stated are uniform in the parameter sequences $\boldsymbol{\theta}$ in $\times_n[-\alpha, \alpha]^m$, and as such, the range of the parameter sequence $\boldsymbol{\theta}$ is to be taken as $\times_n[-\alpha, \alpha]^m$ in the results of the paper.

The main results of Sections 3, 5, and 6 use the following lemma which is a consequence of the Berry–Esséen Theorem (see [9], page 288.)

LEMMA 1.1. *If $\tau_1, \cdots, \tau_n$ are independent random variables with $|\tau_i| \leqq d$ for all i, then*

$$|P[\sum \tau_i \geqq 0] - \Phi(E(\sum \tau_i)/(\mathrm{Var}(\sum \tau_i))^{\frac{1}{2}})| \leqq cd/(\mathrm{Var}(\sum \tau_i))^{\frac{1}{2}}$$

where c is the Berry–Esséen constant.

2. A bound for the modified regret $D_n(\boldsymbol{\theta}, \boldsymbol{\eta})$. We state and prove two lemmas which are higher dimensional generalizations of Proposition 1 and Corollary 1 of Chapter I of [2] for the case of the family of normal distributions $\mathscr{P} = \{P_\theta \mid \theta \in [-\alpha, \alpha]^m\}$.

LEMMA 2.1. $\mathbf{P}_n[|\psi_n - \psi_{n-1}|] \leqq n^{-1}2m\alpha \exp 4m\alpha^2$ *for $n > 1$.*

PROOF. From $\psi_n = G_n[\theta p_\theta]/G_n[p_\theta]$, the triangle inequality and Jensen's inequality, it follows that

$$|\psi_n - \psi_{n-1}| = p_n(\textstyle\sum_{i=1}^{n-1} p_i)^{-1}(\sum p_j)^{-1}|\sum_{i=1}^{n-1}(\theta_i - \theta_n)p_i| \tag{2.1}$$
$$\leqq 2m\alpha p_n(\textstyle\sum p_j)^{-1} \leqq n^{-2}2m\alpha p_n(\sum p_j^{-1})\,.$$

Since $P_n[p_n p_j^{-1}] = \exp(||\theta_n - \theta_j||^2) \leqq \exp 4m\alpha^2$, the proof of the lemma is complete in view of (2.1). □

LEMMA 2.2. *If $\boldsymbol{\eta}$ is in $\times_n[-\alpha, \alpha]^m$, then*

$$|D_n(\boldsymbol{\theta}, \boldsymbol{\eta})| \leqq 4\alpha n^{-1} \textstyle\sum \mathbf{P}_j[|\eta_j - \psi_{j-1}(\underline{X}_j)|] + O(n^{-1}\log n)$$

where ψ_0 is an arbitrary decision rule taking values in $[-\alpha, \alpha]^m$.

PROOF. From (1.2) we have

$$n^{-1} \sum P_j[\|\psi_j - \theta_j\|^2] \leqq R(G_n) \leqq n^{-1} \sum P_j[\|\psi_{j-1} - \theta_j\|^2] . \tag{2.2}$$

For $i = j-1$ and j, we have $\|\eta_j - \theta_j\|^2 - \|\psi_i - \theta_j\|^2 = (\eta_j + \psi_i - 2\theta_j, \eta_j - \psi_i) \leqq 4\alpha|\eta_j - \psi_i|$, since η_j, θ_j, the supports of G_j and G_{j-1}, and hence ψ_j and ψ_{j-1} are all in $[-\alpha, \alpha]^m$; hence (2.2) and (1.1) yield

$$\begin{aligned} -4\alpha n^{-1} \sum \mathbf{P}_j[|\eta_j - \psi_{j-1}(\underline{X}_j)|] \\ \leqq D_n(\boldsymbol{\theta}, \boldsymbol{\eta}) \leqq 4\alpha n^{-1} \sum \mathbf{P}_j[|\eta_j - \psi_j(\underline{X}_j)|] . \end{aligned} \tag{2.3}$$

The triangle inequality applied to the rhs of (2.3) and Lemma 2.1 complete the proof of the lemma. □

Lemma 2.2 motivated all the sequence-compound procedures of this paper, since it shows that by taking η_i to be an estimate of Bayes estimate versus G_{i-1} in the component problem, one might be able to obtain a rate for $D_n(\boldsymbol{\theta}, \boldsymbol{\eta}) \to 0$ by first obtaining a rate for $\mathbf{P}_i[|\eta_i - \psi_{i-1}(\underline{X}_i)|] \to 0$ as $i \uparrow \infty$. The next three sections deal with the rate of convergence of $\mathbf{P}_i[|\eta_i - \psi_{i-1}(\underline{X}_i)|] \to 0$ for three choices of η_i and use Lemma 2.2 to obtain a rate of convergence for the modified regrets of the corresponding sequence-compound procedures.

3. A rate of convergence for $D_n(\boldsymbol{\theta}, \boldsymbol{\psi}^{**})$ with ψ^{**} based on a divided difference estimator for the log of a density. In this section, we define the sequence-compound decision procedure ψ^{**} whose $(n+1)$th component is a divided difference estimator for $\underline{X}_{n+1} + (\log \bar{p})'(\underline{X}_{n+1})$, and show that its modified regret is $O(n^{-1/m+4})$. This result generalizes the result stated in Chapter III of [2] for the $m = 1$ case. Some notation, which is similar to that of [2] for the $m = 1$ case, is required to define ψ^{**}. The functions, introduced below for each n, are abbreviated by omission of their dependency on n.

Let $\bar{F}$ denote the average of the distributions of $\underline{X}_1, \cdots, \underline{X}_n$. For each x in E^m with coordinates $x_1, \cdots, x_m$, let $R = \times_{j=1}^m I_j$ where $I_j = [x_j, x_j + h]$ for $j = 1, \cdots, m$ with $h > 0$ and for $l = 1, \cdots, m$, $R_l = \times_{j=1}^m I_j'$ with $I_j' = I_j$ for $j \neq l$, $I_l' = I_l + k = [x_j + k, x_j + k + h]$ and $0 < h \leqq k$.

For any distribution F on E^m, let $t(F)$ denote the vector-valued function $\langle k^{-1} \log (FR_l/FR)\rangle$ from E^m to E^m where FR and FR_l represent measures of R and R_l under F.

Using a generalization of a particular case of Cauchy's mean value theorem (Lemma 3.3), we can show that $t(\bar{F}) - (\log \bar{p})' \to 0$ as $n \uparrow \infty$ and $k \downarrow 0$ (Lemma 3.4). Hence, in view of (1.3) and Lemma 2.2, it suffices to define ψ_{n+1}^{**}, the $(n+1)$th component of $\boldsymbol{\psi}^{**}$, such that $\mathbf{P}_{n+1}[|\psi_{n+1}^{**}(\underline{X}_1, \cdots, \underline{X}_{n+1}) - (\underline{X}_{n+1} + t(\bar{F})(\underline{X}_{n+1}))|] \to 0$ as $n \uparrow \infty$ and $k \downarrow 0$. This last convergence is achievable by taking ψ_{n+1}^{**} to be a natural estimate of $\underline{X}_{n+1} + t(\bar{F})(\underline{X}_{n+1})$. Therefore, let

$$\begin{aligned} \psi_{n+1}^{**} &= \mathrm{tr}\,(\underline{X}_{n+1} + t(F^*)(\underline{X}_{n+1})) \qquad \text{and} \\ \psi_{n+1}^{*} &= \mathrm{tr}'\,(\underline{X}_{n+1} + t(F^*)(\underline{X}_{n+1})) \end{aligned} \tag{3.1}$$

where F^* is the empirical distribution of $\underline{X}_1, \cdots, \underline{X}_n$ and tr and tr′ stand for coordinatewise retraction to $[-\alpha, \alpha]$ and $[-\alpha', \alpha']$, with $\alpha' = \alpha + k + h$, respectively.

In this section, ψ_n, ψ^{**}_{n+1}, ψ^*_{n+1}, $t(\bar{F})$, and $t(F^*)$ are abbreviated to ψ, ψ^{**}, ψ^*, t and t^* respectively.

Since ψ is in $[-\alpha, \alpha]^m$ and $\psi^{**} = \operatorname{tr}(\psi^*)$, $|\psi^{**} - \psi| \leqq |\psi^* - \psi|$ and, therefore, by the triangle inequality,

$$(3.2) \qquad \mathbf{P}_{n+1}[|\psi^{**} - \psi(X)|] \leqq \mathbf{P}_{n+1}[|\psi^* - (X + t(X))|] + \mathbf{P}_{n+1}[|X + t(X) - \psi(X)|] .$$

Now we prove some lemmas which are required to prove the main result of this section.

LEMMA 3.1. *For x in E^m,*

(1) $x + t(x) \in [-\alpha - h - k/2, \alpha]^m$,

(2) $\bar{F}R_l \geqq h^m \bar{p} \exp - [\{(k + h)(|x| + m^{\frac{1}{2}}\alpha + mh)/2\}]$

where $\bar{p}$ is the density of $\bar{F}$ at x and

(3) $h\bar{F}R_l' \leqq (k + h)\bar{F}R_l \exp (k + h)(|x_l| + \alpha' + k + h)$

for $l = 1, \cdots, m$ where $R_l' = \times_{j=1}^m I_j''$ with $I_j'' = I_j$ for $j \neq l$ and $I_l'' =$ $[x_l, x_l + k + h]$.

PROOF. Since the proofs of the three results are similar, we prove only the first result. For the proofs of other results, see Lemma 3, Section 1.2 of [12]. In the proof, let F_j denote the distribution of $\underline{X}_j$ and $\theta_{j1}, \cdots, \theta_{jm}$ denote the coordinates of θ_j. Let $l \in \{1, \cdots, m\}$.

PROOF OF (1). Since the coordinates of $\underline{X}_j$ are independent,

$$\frac{F_j R_l}{F_j R} = \frac{\Phi(x_l - \theta_{jl} + k + h) - \Phi(x_l - \theta_{jl} + k)}{\Phi(x_l - \theta_{jl} + h) - \Phi(x_l - \theta_{jl})} .$$

An application of Cauchy's mean value theorem (see [4], page 81) to the rhs gives the existence of ω in (0, 1) such that

$$\frac{F_j R_l}{F_j R} = \exp - k(x_l - \theta_{jl} + \omega h + k/2) .$$

Hence, since $|\theta_{jl}| \leqq \alpha$, $\exp - k(x_l + \alpha + \omega h + k/2) \leqq F_j R_l/F_j R \leqq \exp k(\alpha - x_l)$. These bounds are independent of j and hence equivalent to the result in (1) since $kt = \langle \log (\bar{F}R_l/\bar{F}R) \rangle$. □

Now we bound the integrals appearing on the rhs of (3.2). The method of bounding the first integral is essentially a generalization of that given in Chapter III of [2] while a different method leads to a simpler bound on the second integral.

LEMMA 3.2. *For $0 < 2k \leqq (\alpha + 3)^{-1}$,*

$$(3.3) \qquad \mathbf{P}_{n+1}[|\psi^* - (X + t(X))|] \leqq c_1 \left(\frac{k + h}{nk^2 h^{m+1}}\right)^{\frac{1}{2}} + c_2 \left(\frac{1}{nh^m}\right)^{\frac{1}{2}} .$$

PROOF. Let ψ_l^*, t_l^* and t_l denote the lth coordinates of ψ^*, t^* and t respectively. We will prove that $\mathbf{P}_{n+1}[|\psi_l^* - X_l - t_l|] \leqq m^{-1}(\text{rhs}\,(3.3))$. Throughout the proof, $t = t(X)$.

From (1) of Lemma 3.1 and since $\psi_l^* \in [-\alpha', \alpha']$, we obtain that

$$\begin{aligned}\mathbf{P}_n[|\psi_l^* - X_l - t_l|] &= \textstyle\int_0^{2\alpha'} \mathbf{P}_n[|\psi_l^* - X_l - t_l| > u]\, du \\ &\leqq \textstyle\int_0^{2\alpha'} \mathbf{P}_n[|t_l^* - t_l| > u]\, du \\ &= \textstyle\int_0^{2\alpha'} \mathbf{P}_n[t_l^* - t_l > u]\, du \\ &\quad + \textstyle\int_{-2\alpha'}^{0} \mathbf{P}_n[t_l^* - t_l < u]\, du\,. \end{aligned} \tag{3.4}$$

We now bound $\mathbf{P}_n[t_l^* - t_l > u]$. Fix X and let

$$\begin{aligned} Y_j(u) &= 1 & \underline{X}_j &\in R_l \\ &= -\exp k(t_l + u) & \underline{X}_j &\in R \\ &= 0 & &\text{otherwise.}\end{aligned} \tag{3.5}$$

With $\beta^2 = \operatorname{Var}(\sum Y_j)$, we first prove the following required sublemma.

SUBLEMMA. *For* $|u| \leqq 2\alpha'$,

$$|\mathbf{P}_n[\textstyle\sum Y_j \geqq 0] - \Phi(\beta^{-1} \sum P_j Y_j)| \leqq c_4 \dfrac{\exp k(\alpha + 2\alpha' + |X_l|)}{n^{\frac12}(\bar{F}R_l)^{\frac12}}\,. \tag{3.6}$$

PROOF OF THE SUBLEMMA. Since $|u| \leqq 2\alpha'$ and $X_l + t_l \leqq \alpha$ by (1) of Lemma 3.1,

$$\max\{|Y_j| \mid 1 \leqq j \leqq n\} \leqq \exp k(\alpha + 2\alpha' + |X_l|)\,. \tag{3.7}$$

Therefore, by Lemma 1.1, the lhs of (3.6) $\leqq c_2\beta^{-1}\exp k(\alpha + 2\alpha' + |X_l|)$. The proof is completed by showing that $\beta^2 \geqq c_3^2 n\bar{F}R_l$ for some c_3^2. In view of (3.5), $\operatorname{Var}(Y_j) \geqq \operatorname{Var}(Y_j^+) = F_jR_l(1 - F_jR_l) \geqq c_3^2F_jR_l$ since $F_jR_l \leqq (\Phi(h/2) - \Phi(-h/2))^m$ and $h \leqq k \leqq \{2(\alpha + 3)\}^{-1}$. Hence

$$\beta^2 = \operatorname{Var}(\textstyle\sum Y_j) = \sum \operatorname{Var}(Y_j) \geqq c_3^2 n\bar{F}R_l\,. \tag{3.8}$$

Returning to the lemma, we have, using (3.5) and (3.7),

$$\beta^2 = \operatorname{Var}(\textstyle\sum Y_j) \leqq \sum P_j Y_j^2 \leqq n\bar{F}R_l' \exp 2k(\alpha + 2\alpha' + |X_l|)\,. \tag{3.9}$$

From $\bar{F}R\exp kt_l = \bar{F}R_l$ we have (a) $\sum P_jY_j = n\bar{F}R_l(1 - \exp ku) \leqq -nk\bar{F}R_lu$, so that $\Phi(\beta^{-1}\sum P_jY_j) \leqq \Phi(-(nk^2h^m)^{\frac12}uf)$ where

$$f = \bar{F}R_l(h^m\bar{F}R_l')^{-\frac12}\exp -k(\alpha + 2\alpha' + |X_l|)\,; \tag{3.10}$$

and we have also, using the definitions of t_l^* and Y_j,

(b) $[t_l^* - t_l > u] \leqq [\sum Y_j \geqq 0]$, so that, by (3.6), for $0 < u < 2\alpha'$,

$$\mathbf{P}_n[t_l^* - t_l > u] \leqq \Phi(\beta^{-1}\textstyle\sum P_jY_j) + \text{bound in the sublemma.} \tag{3.11}$$

Hence, for $0 < u < 2\alpha'$,

$$\mathbf{P}_n[t_l^* - t_l > u] \leqq \Phi(-(nk^2h^m)^{\frac12}uf) + \text{bound in the sublemma.} \tag{3.12}$$

A similar argument gives that, for $-2\alpha' < u < 0$, $\mathbf{P}_n[t_l{}^* - t_l < u] \leqq \Phi(2^{-1}(nk^2h^m)^{\frac{1}{2}}uf) +$ bound in the sublemma. Since $\int_0^{2\alpha'} \Phi(-au)\,du \leqq a^{-1}$ for $a > 0$, we obtain from (3.4) that

$$\mathbf{P}_n[|\psi_l{}^* - X_l - t_l] \leqq \frac{3}{(nk^2h^m)^{\frac{1}{2}}f} + 4\alpha' \text{ (bound in the sublemma)}\,.$$

Hence, the proof is completed by showing that the P_{n+1}-integrals of $m(h(k+h)^{-1})^{\frac{1}{2}}f^{-1}$ and $m(nh^m)^{\frac{1}{2}}$ (bound in the sublemma) are uniformly bounded in n with the help of Lemma 3.1.

By (2) and (3) of Lemma 3.1, the assumption $0 < h \leqq k \leqq \frac{1}{6}$ and (3.11), we obtain that $[h(k+h)^{-1}]^{\frac{1}{2}}f^{-1} \leqq c_5[\bar{p}(X)]^{-\frac{1}{2}} \exp\{(3|X_l| + |X|/2\}$ for some c_5. Since $\exp - (||u|| + m^{\frac{1}{2}}\alpha)^2 \leqq (2\pi)^m p_i{}^2(u) \leqq \exp - (||u||^2 - 2||u||m^{\frac{1}{2}}\alpha)$, the P_{n+1}-integral of the above upper bound for $(h(k+h)^{-1})^{\frac{1}{2}}f^{-1}$ is uniformly bounded in n. Similar treatment for the other integral completes the proof of the lemma. □

Lemma 3.4, which bounds the second integral of the rhs of (3.2), uses the following lemma, which is proved in Chapter I of [12].

LEMMA 3.3. *For each $j = 1, \cdots, n$, $i = 1, \cdots, m$, let the functions f_{ji}, g_{ji} be real-valued, continuous on $[a_i, b_i]$ and differentiable on (a_i, b_i) and let the derivatives of g_{ji} be in $(0, \infty)$. Then there exist $(c_1, \cdots, c_m)$ in $\times_{i=1}^m (a_i, b_i)$ such that*

$$\frac{\sum \pi f_{ji}|_{a_i}^{b_i}}{\sum \pi g_{ji}|_{a_i}^{b_i}} = \frac{\sum \pi f'_{ji}(c_i)}{\sum \pi g'_{ji}(c_i)}$$

where π stands for the product over i from 1 through m and f'_{ji}, g'_{ji} are the derivatives of f_{ji}, g_{ji}.

LEMMA 3.4. $|X_l + t_l(X) - \psi_l(X)| \leqq k(1+\alpha^2) + h(1+m\alpha^2)$ *for* $l = 1, \cdots, m$.

PROOF. In the proof, let $t_l = t_l(X)$, $\psi_l = \psi_l(X)$; H abbreviate $h^{-m}\bar{F}R$ and e_l denote the unit vector in the lth direction. Since $t_l = k^{-1}[\log H(X + ke_l) - \log H(X)]$, by the mean value theorem, $t_l = \partial \log H/\partial X_l(X + k\varepsilon e_l)$ for some ε in $(0, 1)$. Since $\psi_l - X_l = \partial \log \bar{p}/\partial X_l$ by (1.3), the triangle inequality yields

$$|X_l + t_l - \psi_l \leqq |I_1| + |I_2| \tag{3.13}$$

where

$$I_1 = \partial \log \bar{p}/\partial X_l|_X^{X+\varepsilon k e_l} \tag{3.14}$$

and

$$I_2 = \partial \log H/\partial X_l(X + \varepsilon k e_l) - \partial \log \bar{p}/\partial X_l(X + \varepsilon k e_l)\,. \tag{3.15}$$

By the mean value theorem, $I_1 = \varepsilon k\, \partial^2 \log \bar{p}/\partial X_l{}^2(X + \varepsilon^* k e_l)$ for some ε^* in $(0, \varepsilon)$. With $\theta_{j1}, \cdots, \theta_{jm}$ denoting the coordinates of θ_j for any j, we have

$$1 + \frac{\partial^2 \log \bar{p}}{\partial X_l{}^2} = \frac{\sum (X_l - \theta_{jl})^2 p_j}{\sum p_j} - \left(\frac{\sum (X_l - \theta_{jl}) p_j}{\sum p_j}\right)^2.$$

The rhs of this equality can be recognized as the conditional variance of the lth coordinate of $Y - \theta$ given Y when the joint distribution of the pair (Y, θ) has the distribution resulting from G_n on θ and P_θ on Y for given θ. Hence, since the support of G_n lies in $[-\alpha, \alpha]^m$, $|\partial^2 \log \bar{p}/\partial X_l^2| \leqq 1 + \alpha^2$. Hence, since we have seen above that I_1 is $\varepsilon k(\partial^2 \log \bar{p}/\partial X_l^2)$ evaluated at $X + \varepsilon^* k e_l$,

$$(3.16) \qquad |I_1| \leqq k(1 + \alpha^2) .$$

We complete the proof of the lemma by showing that $|I_2| \leqq h(1 + m\alpha^2)$ with the help of Lemma 3.3. The definition of H and R give that

$$nh^m \frac{\partial H}{\partial X_l} = \sum [\phi(X_l - \theta_{jl} + h) - \phi(X_l - \theta_{jl})] \times \prod_{i \neq l} [\Phi(X_i - \theta_{ji} + h) - \Phi(X_i - \theta_{ji})] .$$

Now we apply Lemma 3.3 to the ratio $\partial H/\partial X_l/H$ with the following identification. For $j = 1, \cdots, n$, $f_{ji} = g_{ji} = \Phi(\cdot - \theta_{ji})$ for $i \neq l$, $f_{jl} = \phi(\cdot - \theta_{jl})$, $g_{jl} = \Phi(\cdot - \theta_{jl})$ and $(a_i, b_i) = (X_i, X_i + h)$ for $i = 1, \cdots, m$. Then, by Lemma 3.3, $\partial \log H/\partial X_l = \partial \log \bar{p}/\partial X_l(X + h\delta)$ for some δ in $(0, 1)^m$. Subtracting $\partial \log \bar{p}/\partial X_l$ from both sides of the last equality followed by an application of the mean value theorem to the rhs of the resulting equality as a function of h gives

$$(3.17) \qquad \frac{\partial \log H}{\partial X_l} - \frac{\partial \log \bar{p}}{\partial X_l} = h \sum_{i=1}^m \delta_i \frac{\partial^2 \log \bar{p}}{\partial X_i \, \partial X_l} (X + h'\delta)$$

for some h' in $(0, h)$. Since, it can be shown that, for $i \neq l$, $\partial^2 \log \bar{p}/\partial X_i \, \partial X_l$ is the i, lth element in the covariance matrix of $\theta - Y$ conditional on Y where Y and θ are as described earlier and since it is already shown that $|\partial^2 \log \bar{p}/\partial X_l^2| \leqq 1 + \alpha^2$, the fact that the support of G_n lies in $[-\alpha, \alpha]^m$, (3.15) and (3.17) imply that $|I_2| \leqq h(1 + m\alpha^2)$. This inequality together with (3.13) through (3.16) gives the result. □

The choices of h and k given in the following main result of the section are optimal for the rate of convergence to zero of the expression obtained by adding the right-hand sides of the results of Lemmas 3.2 and 3.4.

THEOREM 3.1. *If* $h^{m+4} = n^{-1}$, $k^{m+4} = a\, n^{-1}$ *for* a *in* $[1, \infty)$ *and* $\boldsymbol{\psi}^{**} = \{\psi_n^{**}\}$ *where*

$$(3.18) \qquad \psi_1^{**} = 0 \textit{ and } \psi_{n+1}^{**} = \psi^{**} \; (\psi^{**} \textit{ is defined by } (3.2)) \textit{ for } n \geqq 2, \textit{ then}$$

$$\mathbf{P}_{n+1}[|\psi_{n+1}^{**} - \psi_n(\underline{X}_{n+1})|] = O(n^{-1/m+4}) \qquad \textit{and} \qquad D_n(\boldsymbol{\theta}, \boldsymbol{\psi}^{**}) = O(n^{-1/m+4}) .$$

PROOF. The first result of the theorem is a direct consequence of (3.2), Lemmas 3.2 and 3.3 and the definitions of k and h. Since, every component of $\boldsymbol{\psi}^{**}$ lies in $[-\alpha, \alpha]^m$ due to (3.18), the second result follows from the first result and Lemma 2.2. □

4. Rates near $O(n^{-\frac{1}{2}})$ for $D_n(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}})$ with $\hat{\boldsymbol{\psi}}$ based on kernel estimators for a density and its derivative. In this section, we exhibit a sequence-compound procedure $\hat{\boldsymbol{\psi}} = \{\hat{\psi}_n\}$ belonging to a class of procedures whose modified regret is near $O(n^{-\frac{1}{2}})$. To motivate the definition of $\hat{\psi}_{n+1}$, we recall that a rate of convergence for $D_n(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}) \to 0$ would be achieved by first obtaining a rate for $\mathbf{P}_{n+1}[|\hat{\psi}_{n+1} - \psi_n(X)|] \to 0$. Thus we require $\hat{\psi}_{n+1}$ to be an estimate of $\psi_n(X) = X + \bar{q}(X)/\bar{p}(X)$ where $\bar{p}$ and $\bar{q}$ are as in (1.3). Hence, ψ_n can be estimated by first estimating $\bar{p}$ and $\bar{q}$ by the estimates $\bar{\hat{p}}$ and $\bar{\hat{q}}$ (defined below by (4.4) and (4.5)) and then plugging in these estimates in the form for ψ_n given above. The resulting estimator, $X + \bar{\hat{q}}(X)/\bar{\hat{p}}(X)$, after proper truncation, is the estimator $\hat{\psi}_{n+1}$ defined in (4.10). The estimates $\bar{\hat{q}}$ and $\bar{\hat{p}}$ depend on real-valued functions $K_0, \cdots, K_m$ on E^m which are similar to those introduced by Johns and Van Ryzin [8] in the context of an empirical Bayes two-action problem in univariate exponential families. In order to achieve certain rates of convergence for $\mathbf{P}_n[|\bar{\hat{q}} - \bar{q}|] \to 0$ and $\mathbf{P}_n[|\bar{\hat{p}} - p|] \to 0$ (and hence for $\mathbf{P}_{n+1}[|\hat{\psi}_{n+1} - \psi_n(X)|] \to 0$), we impose certain orthogonality conditions on $K_0, \cdots, K_m$, again similar to those given in Theorems 3 and 4 of Johns and Van Ryzin [8].

For $l = 0, 1, \cdots, m$, let K_l be bounded on E^m with $\mu[||z||^s K_l] = s!\, c_{ls} < \infty$ and, for all nonnegative integers $t_1, \cdots, t_m$,

$$(4.1) \qquad \mu[K_0 \textstyle\prod_{i=1}^m z_i^{t_i}] = 1 \quad \text{or} \quad 0 \quad \text{according as} \quad \sum_{i=1}^m t_i = 0$$

or in $\{1, \cdots, s-1\}$ and, for $1 \leqq l \leqq m$, $z_l K_l$ satisfies (4.1) with s replaced by $s - 1$. Functions $K_0, \cdots, K_m$ satisfying these conditions and an additional condition (5.3) have been given at the end of Section 5.

The intent and result of these conditions on $K_0, \cdots, K_m$ is that, if f is a function on E^m with partials of order s uniformly bounded by M, then the substitution of the sth order Taylor expansion with Lagrange's form of remainder shows

$$(4.2) \qquad |\mu[fK_0] - f(0)| \leqq Mc_{0s}$$

and if, in addition, f and all its partials not involving the lth variable vanish at zero,

$$(4.3) \qquad |\mu[fK_l] - f_l(0)| \leqq Mc_{ls}$$

where f_l stands for the first partial of f wrt the lth variable.

For $\varepsilon, \delta > 0$, define

$$(4.4) \qquad \hat{p}_j(X) = \varepsilon^{-m} K_0(\varepsilon^{-1}(\underline{X}_j - X)),\ n\bar{\hat{p}} = \sum \hat{p}_j\,, \qquad \text{and} \quad \bar{\hat{q}} = \langle \bar{\hat{q}}_l \rangle$$

where

$$(4.5) \qquad n\bar{\hat{q}}_l = \sum \hat{q}_{jl} \qquad \text{with} \quad \delta^{m+1}\hat{q}_{jl} = \tfrac{1}{2}K_l(I_l\delta^{-1}(\underline{X}_j - X)) - K_l(\delta^{-1}(\underline{X}_j - X))$$

where I_l is the $m \times m$ matrix reduced by $\frac{1}{2}$ at the lth diagonal element.

Now we state and prove some lemmas concerning the estimates $\bar{\hat{p}}$ and $\bar{\hat{q}}$ of $\bar{p}$ and $\bar{q}$. We do not require the condition that $\theta_j \in [-\alpha, \alpha]^m$ for all j to prove the following lemma.

LEMMA 4.1. (a) $\mathbf{P}_n[|\bar{\hat{p}} - \bar{p}(X)|] \leqq c_1(\varepsilon^s + (n\varepsilon^m)^{-\frac{1}{2}})$
(b) $\mathbf{P}_n[|\bar{\hat{q}} - \bar{q}(X)|] \leqq c_2(\delta^{s-1} + (n\delta^{m+2})^{-\frac{1}{2}})$.

PROOF. Since the proofs of these two results are similar, except that proof of (b) uses (4.3) while that of (a) uses (4.2), we prove (b) only. For details of proof of (a), see Lemma 7 of [12]. Let $\bar{\hat{q}}_l$ denote the lth coordinate of $\bar{\hat{q}}$. We obtain by (4.5) and the change of variables that

$$\mu[p_j\hat{q}_{jl}] = \delta^{-(m+1)} \int p_j(u)[\tfrac{1}{2}K_l(I_l\delta^{-1}(u - X)) - K_l(\delta^{-1}(u - X))]\, d\mu(u)$$
$$= \delta^{-1} \int K_l(v)(p_j(X + I_l^{-1}\delta v) - p_j(X + \delta v))\, d\mu(v)\,.$$

Since $p_j(X + I_l^{-1}\delta v) - p_j(X + \delta v)$ and its partials not involving the lth variable vanish at $v = 0$, (4.3) implies that the rhs of the above equality differs from the lth partial of p_j by at most $c_3\delta^{s-1}$ for some c_3. Hence,

$$|\mathbf{P}_n[\bar{\hat{q}}] - \bar{q}| \leqq mc_3\delta^{s-1}\,. \tag{4.6}$$

Let $V_X(\bar{\hat{q}}_l)$ denote the conditional variance of $\bar{\hat{q}}_l$ given X. By the inequality $(a + b)^2 \leqq 2(a^2 + b^2)$ and change of variables, we have

$$\mu[p_j\hat{q}_{jl}^2] \leqq (\delta^{m+2})^{-1} \int K_l^2(u)\{2^{-1}p_j(X + I_l^{-1}\delta u) + 2p_j(X + \delta u)\}\, d\mu(u)\,.$$

Since $p_j \leqq 1$ and $\mu[K_l^2] < \infty$ by assumptions on K_l,

$$V_X(\bar{\hat{q}}_l) \leqq n^{-2} \sum \mu[p_j\hat{q}_{jl}^2] \leqq c_4(n\delta^{m+2})^{-1}\,. \tag{4.7}$$

Since $\mathbf{P}_n[|\bar{\hat{q}} - \bar{q}|] \leqq |\mathbf{P}_n[\bar{\hat{q}}] - \bar{q}| + \sum_{l=1}^{m} V_X^{\frac{1}{2}}(\bar{\hat{q}}_l)$, (b) follows from (4.6) and (4.7). □

Since $X + \bar{q}/\bar{p}$ is in $[-\alpha, \alpha]^m$ and since θ_{n+1} in $[-\alpha, \alpha]^m$ implies that $P_{n+1}[|X|]$ is uniformly bounded, the following inequality is a consequence of (a) of Lemma 4.1.

$$\mathbf{P}_{n+1}[|\bar{q}||(\bar{\hat{p}}/\bar{p}) - 1|] \leqq c_9\{\varepsilon^s + (n\varepsilon^m)^{-\frac{1}{2}}\}\,. \tag{4.8}$$

LEMMA 4.2. *For any a in* (0, 1) *and* $\beta > 0$, $P_{n+1}[\bar{p} < \beta] \leqq c_5\beta^a$.

PROOF. With $Z = X - \theta_{n+1}$ and therefore $|X - \theta_j| \leqq |Z| + 2m\alpha$

$$p_j(X) = (2\pi)^{-m/2} \exp - (||X - \theta_j||^2/2)$$
$$\geqq (2\pi)^{-m/2} \exp\{-(m^{\frac{1}{2}}||Z|| + 2m\alpha)^2/2\}\,. \tag{4.9}$$

Let M be the minimum value of $||Z||$ for which the rhs of (4.9) $\leqq \beta$. Since

$$P[||Z||^2 > 2t] \leqq e^{-bt} \int_0^\infty e^{-(1-b)u/2}u^{(m-2)/2}\, du/2^{m/2}\Gamma(m/2) \leqq e^{-bt}(1 - b)^{-m/2}$$

for all $t > 0$ and b in (0, 1), we get from (4.9) that

$$\beta^{-a}P_{n+1}[\bar{p} < \beta] \leqq \beta^{-a}P_{n+1}[||Z|| > M]$$
$$\leqq c_8^{-a}(1 - b)^{-m/2} \exp[\{a(M + 2m\alpha)^2 - bM^2\}/2]$$

where the rhs expression is bounded in M for $0 < a < b < 1$. □

COROLLARY 4.2. *For any a in* (0, 1) *and* $\beta > 0$,

$$P_{n+1}[(|\bar{q}|/\bar{p})[\bar{p} < \beta]] \leqq c_6\beta^a\,.$$

PROOF. Since $|\bar{q}|/\bar{p} \leqq m\alpha + |X|$ and, therefore, has uniformly bounded moments of any order, Hölder's inequality yields for $r > 1$ the bound

$$(P_{n+1}[(|\bar{q}|/\bar{p})^{r/(r-1)}])^{(r-1)/r}(P_{n+1}[\bar{p} < \beta])^{1/r}$$

for the lhs of the result of the corollary. By Lemma 4.2, $P_{n+1}[\bar{p} < \beta] \leqq c_5\beta^b$ for b in (0, 1). Choosing r such that $ar = b$, we get the result of the corollary. □

Now we define the procedure $\hat{\boldsymbol{\psi}}$ as follows. Let

$$(4.10) \qquad \hat{\psi}_1 = 0 \quad \text{and} \quad \hat{\psi}_{n+1} = \mathrm{tr}\,(X + (\bar{q}/\bar{p}')) \qquad \text{for } n \geqq 1$$

where tr, as in Section 3, stands for coordinatewise retraction to $[-\alpha, \alpha]$ and for y in R^1, let $y' = y \vee \beta$ where $\beta > 0$.

In the following lemma, δ is chosen to minimize the bound in (b) of Lemma 4.1, ε is chosen in such a way that the bound in (a) of Lemma 4.1 is smaller than the optimal bound in (b) of Lemma 4.1 and β is chosen to balance the bound on $|\hat{\psi} - \psi|$($\hat{\psi}$ is an abbreviation for $\hat{\psi}_{n+1}$) occurring in the proof of the following lemma.

LEMMA 4.3. *For each positive integer s and a in* (0, 1), *there exists c_7 such that if $\delta^{2s+m} = n^{-1}$, $\delta^{(m+2)/m} \leqq \varepsilon \leqq \delta^{(s-1)/s}$ and $\beta^{1+a} = \delta^{s-1}$, then*

$$\mathbf{P}_{n+1}[|\hat{\psi} - \psi(X)|] \leqq c_7 n^{-(s-1)a/(2s+m)(1+a)} .$$

PROOF. Since ψ lies in $[-\alpha, \alpha]^m$ and since $\hat{\psi}$ is the retraction of $X + \bar{q}/\bar{p}$ to $[-\alpha, \alpha]^m$, we have, by the inequality $\bar{p}' \geqq \beta$, that

$$|\hat{\psi} - \psi| \leqq |(\bar{q}/\bar{p}') - (\bar{q}/\bar{p})| \leqq \beta^{-1}\{|\bar{q} - \bar{q}| + (|\bar{q}|/\bar{p})|\bar{p} - \bar{p}'|\} .$$

Since $|\bar{p} - \bar{p}'| \leqq |\bar{p} - \bar{p}| + \beta[\bar{p} < \beta]$, by (b) of Lemma 4.1, (4.8) and Corollary 4.2, it follows that

$$\mathbf{P}_{n+1}[|\hat{\psi} - \psi] \leqq c_7\{\beta^{-1}[\delta^{s-1} + (n\delta^{m+2})^{-\frac{1}{2}} + \varepsilon^s + (n\varepsilon^m)^{-\frac{1}{2}}] + \beta^a\}$$

for some constant c_7. This inequality together with assumptions on ε, δ and β completes the proof of the lemma. □

THEOREM 4.1. *If the hypothesis of Lemma* 4.3 *is satisfied and $\hat{\boldsymbol{\psi}}$ is defined by* (4.10), *then*

$$D_n(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}}) = O(n^{-(s-1)a/(2s+m)(1+a)}) .$$

PROOF. Since, by definition, every component of $\hat{\boldsymbol{\psi}}$ lies in $[-\alpha, \alpha]^m$, the theorem is a consequence of the above lemma and Lemma 2.2 of Section 2. □

5. Rates near $O(n^{-\frac{1}{2}})$ for $D_n(\boldsymbol{\theta}, {}_0\hat{\boldsymbol{\psi}})$ with ${}_0\hat{\boldsymbol{\psi}}$, a particular $\hat{\psi}$. For a fixed positive integer $s(>1)$, letting ${}_0\hat{\boldsymbol{\psi}}$ denote a specialization, less a retraction to $[\beta, \infty)$, of $\hat{\boldsymbol{\psi}}$ defined by (4.10), with an additional assumption (5.3) on $K_0, \cdots, K_m$, we show that $D_n(\boldsymbol{\theta}, {}_0\hat{\boldsymbol{\psi}}) = O(n^{-(s-1)/2(s+m+1)})$. The proof of this result, to some extent, goes along the lines of proof of Lemma 3.2; but additionally uses the rates at which $|\mathbf{P}_n[\bar{p}] - \bar{p}(X)| \to 0$ and $|\mathbf{P}_n[\bar{q}] - \bar{q}(X)| \to 0$ as $n \uparrow \infty$.

After specializing $\bar{p}$ and $\bar{q}$ defined by (4.4) and (4.5) respectively by setting

$\varepsilon = \delta = h(>0)$ in their definitions, let

$$ {}_0\hat{\psi} = \operatorname{tr}(X + (\bar{\hat{q}}/\bar{\hat{p}}))[\bar{\hat{p}} > 0] \tag{5.1}$$

where tr stands for coordinatewise retraction to $[-\alpha, \alpha]$. With $Z_j = h^{-1}(\underline{X}_j - X)$ for $j = 1, \cdots, n$, and $V_l(u) = h[u + \{\bar{q}_l(X)/\bar{p}(X)\}]$, let $Y_j(u) = \langle Y_{jl}(u)\rangle$ with

$$ Y_{jl}(u) = h^{m+1}\hat{q}_{jl} - h^m V_l(u)\hat{p}_j = (\tfrac{1}{2}K_l \circ I_l - K_l - V_l(u)K_0) \circ Z_j \tag{5.2}$$

for $j = 1, \cdots, n$ where $\circ$ denotes composition of functions.

LEMMA 5.1. *If $K_0, \cdots, K_m$ satisfy conditions involving* (4.1) *and, additionally, are such that for $0 < u \leqq 2\alpha$, $h \leqq 1$,*

$$ c_1 e^{-c_2\|X\|} \leqq \frac{\operatorname{Var}(Y_{jl}(\pm u)), \operatorname{Var}(K_0 \circ Z_j)}{h^m \phi(\|X\|)} \leqq c_3 e^{c_4\|X\|} \tag{5.3}$$

for $l = 1, \cdots, m$, then $\mathbf{P}_{n+1}[|{}_0\hat{\psi} - \psi(X)|] \leqq c_5\{(nh^{m+2})^{-\frac{1}{2}} + (nh^{2s+m})^{\frac{1}{2}}\}$.

PROOF. (Outline only. For details, see Section 4 of [12].) Let ${}_0\hat{\psi}_l$, ψ_l, $\hat{q}_l$ and $\bar{q}_l$ denote the lth coordinates of ${}_0\hat{\psi}$, ψ, $\bar{\hat{q}}$ and $\bar{q}$ respectively.

By definition of ${}_0\hat{\psi}_l$ and the fact that ψ_l lies in $[-\alpha, \alpha]$,

$$ |{}_0\hat{\psi}_l - \psi_l(X)| \leqq (|(\bar{\hat{q}}_l/\bar{\hat{p}}) - (\bar{q}_l(X)/\bar{p}(X))| \wedge 2\alpha)[\bar{\hat{p}} > 0] + \alpha[\bar{\hat{p}} \leqq 0]\,. $$

Therefore, the definition of Y_{jl} in (5.2) and the equalities $\sum \bar{\hat{q}}_{jl} = n\bar{q}_l$ and $\sum \hat{p}_j = n\bar{\hat{p}}$ imply that

$$ \begin{aligned} \mathbf{P}_n[|{}_0\hat{\psi}_l - \psi_l(X)|] &\leqq \textstyle\int_0^{2\alpha} \mathbf{P}_n[\bar{\hat{p}} > 0\,;\ \sum Y_{jl}(u) > 0 \quad \text{or} \\ &\qquad -\textstyle\sum Y_{jl}(-u) > 0]\,du + \alpha\mathbf{P}_n[\bar{\hat{p}} \leqq 0] \\ &\leqq \textstyle\int_0^{2\alpha} (\mathbf{P}_n[\sum Y_{jl}(u) > 0] \\ &\qquad + \mathbf{P}_n[-\textstyle\sum Y_{jl}(-u) > 0])\,du + \alpha\mathbf{P}_n[\bar{\hat{p}} \leqq 0]\,. \end{aligned} \tag{5.4}$$

We first bound the integrand of the first term. Since $\psi_l(X) = X_l + \{\bar{q}_l(X)/\bar{p}(X)\} \leqq \alpha$, we have for $0 < u \leqq 2\alpha$

$$ |V_l(\pm u)| \leqq (|X_l| + 3\alpha) \qquad \text{and} \qquad |Y_{jl}(\pm u)| \leqq c_6(1 + |X_l|)\,. \tag{5.5}$$

Also by (5.3),

$$ \begin{aligned} \underline{\beta}^2 = c_1 nh^m e^{-c_2\|X\|}\phi(\|X\|) &\leqq \beta^2(\pm u) = \operatorname{Var}(\textstyle\sum Y_{jl}(\pm u)) \\ &\leqq c_3 nh^m e^{c_4\|X\|}\phi(\|X\|) = \bar{\beta}^2\,. \end{aligned} \tag{5.6}$$

(5.5) and (5.6) together with an application of Lemma 1.1 to the two sets of random variables $Y_{1l}(u), \cdots, Y_{nl}(u)$ and $-Y_{1l}(-u), \cdots, -Y_{nl}(-u)$ followed by a few manipulations involving a triangle inequality, (4.4), and (3.6) of [12] give the inequality

$$ \begin{aligned} &\max\{\mathbf{P}_n[\textstyle\sum Y_{jl}(u) > 0], \mathbf{P}_n[-\textstyle\sum Y_{jl}(-u) > 0]\} \\ &\qquad \leqq \Phi(-nh^{m+1}u\bar{p}(X)/\bar{\beta}) + c_6\underline{\beta}^{-1}(1 + nh^{s+m})(1 + |X_l|)\,. \end{aligned} \tag{5.7}$$

Similarly, one can show that

$$\mathbf{P}_n[\bar{p} \leqq 0] \leqq \bar{\beta}/nh^m\bar{p}(X) + c_7\underline{\beta}^{-1}(1 + nh^{s+m}) \tag{5.8}$$

by using normal tail bound (see [1], page 166). Therefore, by (5.4), (5.7) and (5.8),

$$\begin{aligned}\mathbf{P}_n[|{}_0\hat{\psi}_l - \psi_l(X)|] &\leqq 2 \textstyle\int_0^{2\alpha} \Phi(-nh^{m+1}u\bar{p}(X)/\bar{\beta})\, du \\ &\quad + \frac{\bar{\beta}}{nh^{m+1}\bar{p}(X)} + \frac{c_7(1 + nh^{m+s})(1 + |X_l|)}{\underline{\beta}} \\ &\leqq 3\bar{\beta}(nh^{m+1}\bar{p}(X))^{-1} + \frac{c_7(1 + nh^{m+s})(1 + |X_l|)}{\underline{\beta}}\,.\end{aligned} \tag{5.9}$$

Now we show that the P_{n+1} integral of the rhs of the above inequality leads to the rhs of the result. Since $(nh^m)^{-1}\bar{\beta}^2 = c_3\phi(||X||)\exp(c_4||X||)$ by (5.6) and since $\exp - (||u|| + m^{\frac{1}{2}}\alpha)^2 \leqq (2\pi)^m p_i^2(u) \leqq \exp - (||u||^2 - 2m^{\frac{1}{2}}\alpha||u||)$, the P_{n+1}-integral of the first term of the rhs of (5.9) is bounded by $c_5(nh^{m+2})^{-\frac{1}{2}}$ for some c_5. Similar treatment of the second term of the rhs of (5.9) results in $c_5(nh^{2s+m})^{\frac{1}{2}}$. □

Now we define the procedure ${}_0\hat{\boldsymbol{\psi}}$ as follows. Let

$${}_0\hat{\psi}_1 = 0 \quad \text{and} \quad {}_0\hat{\psi}_{n+1} = {}_0\hat{\psi} \qquad \text{for } n \geqq 1 \tag{5.10}$$

where ${}_0\hat{\psi}$ is defined by (5.1).

THEOREM 5.1. *If $K_0, \cdots, K_m$ satisfy the conditions of Lemma* 5.1, $h = an^{-1/(s+m+1)}$ *for $a \in (0, \infty)$ and ${}_0\hat{\boldsymbol{\psi}}$ is defined by* (5.10), *then*

$$D_n(\boldsymbol{\theta}, {}_0\hat{\boldsymbol{\psi}}) = O(n^{-(s-1)/2(s+m+1)})\,.$$

PROOF. Since, by definition, every component of ${}_0\hat{\boldsymbol{\psi}}$ lies in $[-\alpha, \alpha]^m$, the result of the theorem is a direct consequence of the hypothesis on h, Lemma 5.1 and Lemma 2.2. □

Now we exhibit kernel functions $K_0, \cdots, K_m$ satisfying the conditions of Lemma 5.1 (and hence of Theorem 5.1). We develop these kernels in $m = 2$ case for the sake of the simplicity of the notation.

Let $[c_{ij}]$ be an $\infty \times \infty$ matrix whose ijth element is c_{ij}. For each pair of positive integers i, j, let $R^{i,j}$ denote the indicator function of the southwest quadrant of (i, j) intersected with the northeast quadrant of $(0, 0)$. We will determine $[a_{ij}]$, $[b_{ij1}]$ and $[b_{ij2}]$ with finitely many coordinates different from zero such that

$$K_0 = \textstyle\sum_{i,j} a_{ij}R^{i,j}\,, \qquad K_1 = \sum_{i,j} b_{ij1}R^{i,j} \qquad \text{and} \qquad K_2 = \sum_{i,j} b_{ij2}R^{i,j} \tag{5.11}$$

satisfy the conditions of Lemma 5.1.

For any two positive integers S, T, let $[a_{ij}]_{S,T}$ denote the modification of $[a_{ij}]$ obtained by replacing a_{ij} by zero if $i > S$ or $j > T$. We now note that for any two sets of distinct positive integers $k_1, \cdots, k_S$ and $l_1, \cdots, l_T$, the vectors

$$[i^{k_1}j^{l_1}]_{S,T}, \cdots, [i^{k_S}j^{l_T}]_{S,T} \quad \text{are a basis for} \quad E^{ST}\,. \tag{5.12}$$

(For $\sum_{r,t} c_{rt}[i^{k_r} j^{l_t}]_{S,T} = [0]$ iff $\sum_{r,t} c_{rt} i^{k_r} j^{l_t} = 0$ has the roots $\{1, \cdots, S\} \times \{1, \cdots, T\}$, which by iterative application of Descarte's rule of signs (see [13], page 121) requires c_{rt} to vanish.) We use this fact to show that certain norms are different from zero and to show that a certain coefficient is different from zero. The kernel conditions (4.1) on K_0 and K_1 specialize to the following requirements on the inner products,

$$\begin{aligned} ([a_{ij}], [i^{l_1} j^{l_2}]) &= 1 \quad \text{for} \quad l_1 = l_2 = 1 \\ &= 0 \quad \text{for} \quad 1 \leqq l_1, l_2, 3 \leqq l_1 + l_2 \leqq s + 1 \end{aligned}$$

and

$$\begin{aligned} ([b_{ij1}], [i^{l_1} j^{l_2}]) &= 1 \quad \text{for} \quad l_1 = 2, l_2 = 1 \\ &= 0 \quad \text{for} \quad 4 \leqq l_1 + l_2 \leqq s + 1 \,. \end{aligned}$$

We choose $[a_{ij}]$ for simplicity to be the projection of

$$(5.13) \qquad [ij]_{s,s} \quad \text{on} \quad \perp \{[i^{l_1} j^{l_2}]_{s,s} \mid 1 \leqq l_1, l_2, 3 \leqq l_1 + l_2 \leqq s + 1\}$$
divided by its squared norm,

and in order to satisfy the variance requirements (5.3), we take $[b_{ij1}]$ to be the projection of

$$(5.14) \qquad [i^2 j]_{s,s} \quad \text{on} \quad \perp \{[i^{l_1} j^{l_2}]_{s,s} \mid (l_1, l_2) \neq (2, 1), 1 \leqq l_1, l_2 \leqq s\}$$
divided by its squared norm.

The squared norms are different from zero by the aforenoted linear independence for $(S, T) = (s, s)$. Moreover $b_{sj1} \neq 0$ for some j in $\{1, \cdots, s\}$ for otherwise $[b_{ij1}]$ defined by (5.14) will lie in $R^{(s-1)s}$ and is orthogonal to a basis in $R^{(s-1)s}$, hence is zero. Let $M = \max \{j \mid b_{sj1} \neq 0\}$.

With A denoting the bound on K_0, K_1 and K_2, we have by the definition of Y_{j1} in (5.2) and the definitions of K_0 and K_1 in (5.11) that

$$(5.15) \qquad \begin{aligned} \operatorname{Var}(Y_{j1}) &\leqq A^2(1 \cdot 5 + V)^2 P_j[Z_j \in (0, s) \times (0, s)] \\ &= A^2(1 \cdot 5 + V)^2 P_j[\underline{X}_j \in (X, X + sh) \times (X, X + sh)] \end{aligned}$$

for fixed X. By the mean value theorem, the probability appearing on the rhs of (5.15) is $s^2 h^2 p_j(X + \xi sh)$ for some ξ in the unit square. Hence, factoring out $h^2 \phi(||X||)$ from $s^2 h^2 p_j(X + \xi sh)$ we obtain from the restriction $h \leqq 1$ and the inequality (5.5) that the rhs of (5.15) (and hence $\operatorname{Var}(Y_{j1})$) is bounded by the rhs of (5.3) specialized to $m = 2$ case for suitable c_3 and c_4.

Now we show that when the kernel functions involved in the definition of Y_{j1} are given by (5.11), the lower bound requirement for $\operatorname{Var}(Y_{j1})$ is also satisfied when $m = 2$. We observe that Y_{j1} defined by (5.2) and (5.11) takes a finite number of values including zero and $2^{-1} b_{sM1}$. The probability that it takes the value zero is $P_j[Z_j \notin (0, 2s) \times (0, s)]$ and that it takes the value $2^{-1} b_{sM1}$ is $P_j[Z_j \in (2(s - 1), 2s) \times (M - 1, M)]$. Since the variance of any random variable is bounded below by the expectation of the conditional (conditioned on $Z_j \notin (0, 2s) \times (0, s)$ or $Z_j \in (2(s - 1), 2s) \times (M - 1, M)$) variance, we obtain that $\mathrm{Var}(Y_{j1})$ is bounded below by

$$(5.16)\quad 4^{-1}b_{sM1}^2 P_j[Z_j \in (2(s-1), 2s) \times (M-1, M)] \times \left\{\frac{P_j[Z_j \notin (0, 2s) \times (0, s)]}{P_j[Z_j \notin (0, 2s) \times (0, s)] + P_j[Z_j \in (2(s-1), 2s) \times (M-1), M)]}\right\}.$$

Since the expression in the curly brackets can be bounded by a constant c_{13} and since $P_j[Z_j \in (2(s-1), 2s) \times (M-1, M)] = h^2 p_j(X + \xi h)$ for some ξ (in $(2(s-1), 2s) \times (M-1, M)$), the lower bound in (5.3) specialized to $m = 2$ case for $\mathrm{Var}(Y_{j1})$ follows from (5.16) since $b_{sM1} \neq 0$ by our previous observation. Similarly, by following the arguments given above, we can show that $\mathrm{Var}(Y_{j2})$ and $\mathrm{Var}(\sum K_0 \circ Z_j)$ also satisfy (5.3).

6. A lower bound for $D_n(\mathbf{0}, \boldsymbol{\phi}^{**})$. In this section we compare the rate of convergence of the regret of $\boldsymbol{\phi}^{**}$ to those of $\hat{\boldsymbol{\phi}}$ and ${}_0\hat{\boldsymbol{\phi}}$ using the following result. With k and h ($\in (0, k]) \to 0$ as $n \to \infty$ and

$$\beta^2 = nk^2h^m,$$

by using Lemmas 1.1 and 3.1, we prove

THEOREM 6.1. *If $\beta(h + k/2) \to a$ ($\in (0, \infty)$), then*

$$D_n(\mathbf{0}, \boldsymbol{\phi}^{**}) \geqq c_1^2\beta^{-2} \quad \textit{for large} \quad n.$$

PROOF. Recall that $\phi^{**}(=\langle\phi_l^{**}\rangle)$, the $(n+1)$th component of ϕ^{**}, is given by (3.21). Since $nD_n(\mathbf{0}, \boldsymbol{\phi}^{**}) = \sum \mathbf{P}_j[||\phi_j^{**}||^2]$ by (1.1), $R(G_n) = 0$ for degenerate G_n, and since $\mathbf{P}_{n+1}[||\phi^{**}||^2] \geqq \mathbf{P}_{n+1}[[X_l > \alpha]\phi_l^{**}]$, the proof of the theorem is completed by showing that $\mathbf{P}_{n+1}[[X_l > \alpha]\phi_l^{**}] \geqq c_1\beta^{-1}$. (For details, see Section 5 of [12].) □

REMARK. For the choice of h and k given in Theorem 3.1, namely, $h = n^{-1/m+4}$ and $k = an^{-1/m+4}$ with a in $[1, \infty)$, we obtain by the theorem proved above that

$$D_n(\mathbf{0}, \boldsymbol{\phi}^{**}) \geqq c_1^2 n^{-2/m+4}.$$

For any $a > 0$, Theorem 4.1 shows that we can define sequence-compound decision procedures $\hat{\boldsymbol{\phi}}$ such that $D_n(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}) = O(n^{-(1/4-a)})$. Hence, for $a > 1/36$, $m \geqq 5$ and sufficiently large n,

$$\sup D_n(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}) \leqq n^{-(1/4-a)} \leqq n^{-2/m+4} \leqq \sup D_n(\boldsymbol{\theta}, \boldsymbol{\phi}^{**})$$

where the sup is taken over all parameter sequences $\boldsymbol{\theta}$ in $\times_n[-\alpha, \alpha]^m$.

For any positive integer s, Theorem 5.1 shows that we can define procedures ${}_0\hat{\boldsymbol{\phi}}$ such that $D_n(\boldsymbol{\theta}, {}_0\hat{\boldsymbol{\phi}}) = O(n^{-(s-1)/2(s+m+1)})$. Hence, if $s > 5 + 8/m$, the procedure ${}_0\hat{\boldsymbol{\phi}}$ is better than $\boldsymbol{\phi}^{**}$ in the sense in which $\hat{\boldsymbol{\phi}}$ is better than $\boldsymbol{\phi}^{**}$.

7. Concluding remarks. The result of Theorem 5.1 is of most interest since it establishes a rate near $O(n^{-\frac{1}{2}})$ which is in line with the results obtained by Gilliland [2, 3], Hannan [5, 6], and Johns [7].

The main reason for our not being able to exhibit sequence-compound procedures whose modified regrets are $O(n^{-\frac{1}{2}})$ is that unbiased estimates do not seem to exist for $\bar{p}$, the average of the densities of $\underline{X}_1, \cdots, \underline{X}_n$ and for the vector of partial derivatives of $\bar{p}$.

The author has already obtained results comparable to the results of Section 5 when the component problem is a linear loss two-action problem concerning the scale parameter of a negative exponential distribution.

Acknowledgment. The author wishes to thank Professor James F. Hannan for suggesting the problem and for his guidance and encouragement. Thanks are also due to the referee for suggestions which shortened the paper considerably.

REFERENCES

[1] FELLER, WILLIAM (1957). *An Introduction to Probability Theory and its Applications* **1** (2nd ed.). Wiley, New York.

[2] GILLILAND, DENNIS C. (1966). Approximation to Bayes risk in sequences of non-finite decision problems. RM-162, Department of Statistics and Probability, Michigan State Univ.

[3] GILLILAND, DENNIS C. (1968). Sequential compound estimation. *Ann. Math. Statist.* **39** 1890–1904.

[4] GRAVES, LAWRENCE M. (1946). *The Theory of Functions of Real Variables*, (2nd ed). Macmillan, New York.

[5] HANNAN, JAMES F. (1956). The dynamic statistical decision problem when the component problem involves a finite number, *m*, of distributions (abstract). *Ann. Math. Statist.* **27** 212.

[6] HANNAN, JAMES F. (1957). Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games* **3** 97–139.

[7] JOHNS, M. V. JR. (1967). Two-action compound decision problems. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 463–478. Univ. of California Press.

[8] JOHNS, M. V. JR. and VAN RYZIN, J. (1967). Convergence rates for empirical Bayes two-action problems II. Continuous case. Technical Report No. 132, Department of Statistics, Stanford Univ.

[9] LOÈVE, MICHEL (1963). *Probability Theory*, (3rd ed.). Van Nostrand, Princeton.

[10] SAMUEL, ESTER (1965). Sequential compound estimators. *Ann. Math. Statist.* **36** 879–889.

[11] SAMUEL, ESTER (1966). Sequential compound rules for the finite decision problem. *J. Roy. Statist. Soc. Ser. B* **28** 63–72.

[12] SUSARLA, V. (1970). Rates of convergence in sequence-compound squared-distance loss estimation and two-action problems. RM-262, Department of Statistics and Probability, Michigan State Univ.

[13] USPENSKY, J. V. (1948). *Theory of Equations* (1st ed.). McGraw-Hill, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN
MILWAUKEE, WISCONSIN 53201