REPETITIVE PLAY IN FINITE STATISTICAL GAMES WITH UNKNOWN DISTRIBUTIONS[1]

By J. Van Ryzin

Argonne National Laboratory

1. Introduction and summary. This paper is concerned with repetitive sequential play in finite statistical games (decision problems) from the statistician's point of view. We shall assume that the statistician's move at stage k may depend on the previous $k - 1$ moves of Nature as well as the random variable $\mathbf{X}_k = (X_1, \cdots, X_k)$, where the X_i are independent observations (r.v.'s) (possibly vector-valued) from the sequence of statistical games, $k = 1, 2, \cdots$. The play is repetitive in the sense that each component game is identical in structure, with only the moves of the statistician and Nature changing. Furthermore, we impose no assumptions regarding the behavior of the parameter sequence of Nature's moves. The statistician does have the added disadvantage that the finite class of distributions in the component game is not fully specified. However, he does know that class in question has: either (i) all members with discrete distributions or (ii) all members with q-dimensional a.e. continuous Lebesgue densities.

This same problem when the distributions are fully known has been treated in [6] for statistical as well as more general games in which Nature's space is finite. In the case where the distributions are completely specified but the history of the past moves is unknown to the statistician, see [20], [22], [27], and [28]. The development in this paper is closely connected to and motivated by these results, particularly those of the preceding paper [27].

If for fixed N, the empirical distribution p_N of Nature's moves is known, then the statistician could use as a rule for each of the N component games a strategy Bayes against p_N having risk $\phi(p_N)$. In all the papers cited in the previous paragraph, the aim was to construct for the statistician, when p_N is unknown and N not specified, a sequence of randomized decision functions whose Nth average loss minus $\phi(p_N)$ approaches zero (or has an upper bound approaching zero) in a suitable sense as the number of repetitions of play, N, increases. However, in the case of statistical games, all of the above results require that the finite class of distributions be fully specified. In this paper we remove that assumption by estimating the distributions sequentially based on past moves and observations. Then in the present play of the component game the statistician substitutes these estimators into a procedure which is Bayes against the empirical distribution of Nature's previous moves. The resulting sequence of procedures is shown to be "asymptotically good" in the sense that the average

Received 10 May 1965; revised 4 March 1966.

[1] Work performed under the auspices of the United States Atomic Energy Commission.

loss over the N games W_N minus the Bayes risk $\phi(p_N)$ approaches zero (in an appropriate sense) as N, the number of games played, increases.

In Section 2 we introduce notation and preliminaries. Section 3 discusses play in repetitive games and defines the proposed sequential procedures $\mathbf{t} = \{\mathbf{t}_k\}$. In Section 4 we prove preliminary results upon which all proofs are founded.

Section 5 considers the discrete case giving uniform (in sequences of Nature's moves) convergence theorems (as $N \to \infty$) for the quantity $W_N - \phi(p_N)$. Theorem 5.1 is a uniform convergence theorem of $O(N^{-\frac{1}{2}})$ of the expected value of $W_N - \phi(p_N)$ for finite discrete classes, each member of which is non-degenerate and satisfies a certain tail probability condition. Under the same conditions, Theorem 5.2 gives uniform convergence to zero in probability for the quantity $N^{\frac{1}{2}} (\log N)^{-1} \{W_N - \phi(p_N)\}$ as $N \to \infty$. Uniform convergence of $W_N - \phi(p_N) \to 0$ in probability for general non-degenerate finite discrete class is presented in Theorem 5.3.

Section 6 treats the estimation problem for densities needed to form the randomized strategy sequences $\mathbf{t}$ in the continuous case. The results stated are based on a paper by Cacoullos [3] generalizing the univariate results of Parzen [15].

In Section 7, we present results for the continuous case. Theorem 7.1 and its corollary give uniform convergence of $W_N - \phi(p_N)$ to zero in probability and of its expectation to zero, respectively. The finite continuous classes of Theorem 7.1 are very general in the sense that each member is a continuous a.e. density.

Finally, in Section 8 we draw certain conclusions and relate our results to similar results obtained elsewhere.

The novelty of the paper rests in the fact that through the past history of Nature's moves and the observations connected with past play, one can construct a sequential strategy, $\mathbf{t} = \{\mathbf{t}_k\}$, with very little knowledge about the finite class of distributions, which approaches asymptotic "optimal" play. The lack of knowledge on the finite class of distributions distinguishes this work from the related "repetitive type" problems in games and/or decision theory treated in [1], [2], [4], [6], [7], [8], [9], [10], [12], [17], [18], [19], [20], [21], [22], [24], [25], [26], [27], [28], and [29].

For possible applications of this work see Neyman [14], especially his Example 3 and his discussion relating to the work of Blackwell [2].

2. Preliminaries. See Section 2 of the preceding paper [27]. Reference to Section 2 and equations therein for the remainder of *this* paper will be taken as a reference to the corresponding Section 2 of [27].

3. Sequential strategies in repetitive statistical games. Consider again the problem stated in Section 2. If in such a problem the statistician knows at stage k, the first k observations $\mathbf{X}_k = (X_1, \cdots, X_k)$ and the previous $k - 1$ moves of Nature represented by the $k - 1$ parameter values, $\boldsymbol{\theta}_{k-1} = (\theta_1, \cdots, \theta_{k-1})$, $(\boldsymbol{\theta}_0 = 0)$, then the procedure in (2.1) at stage k will be a function of $\mathbf{X}_k$ and $\boldsymbol{\theta}_{k-1}$, i.e., in (2.1) we have

$$\mathbf{t}_k(\boldsymbol{\theta}, \mathbf{X}_k) = \mathbf{t}_k(\boldsymbol{\theta}_{k-1}, \mathbf{X}_k), \qquad k = 1, 2, \cdots. \tag{3.1}$$

To further specify the strategies $\mathbf{t} = \{\mathbf{t}_k\}$ in (3.1) consider what else is known to the statistician. Recall that in Section 1, we specified that the finite class $\mathcal{P} = \{P_\theta ; \theta \varepsilon \Omega\}$ is unknown to the statistician. However, he does know that for all problems either $\mathcal{P} = \mathcal{P}_1$ or $\mathcal{P} = \mathcal{P}_2$, where

$\mathcal{P}_1 = \{P_\theta \mid$ For each $\theta \varepsilon \Omega$, P_θ is a discrete distribution on the countable set $\mathfrak{X}$ with $\mathfrak{F}$ the σ-field generated by points and μ is counting measure$\}$

$\mathcal{P}_2 = \{P_\theta \mid$ For each $\theta \varepsilon \Omega$, P_θ is a q-dimensional distribution on $\mathfrak{X}$, Euclidean q-space, having an a.e.(μ) continuous density with μ as Lebesgue measure and $\mathfrak{F}$ the σ-field of Lebesgue measurable sets$\}$

We shall refer to the situation in which $\mathcal{P}$ is $\mathcal{P}_1$ as the *discrete case* and in which $\mathcal{P}$ is $\mathcal{P}_2$ as the *continuous case*.

The problem is now to select a sequence of strategies $\mathbf{t}$ approaching "optimal" play in an appropriate sense. This can be done in both the discrete and continuous cases cited above. The sense in which a procedure $\mathbf{t}$ is deemed good is to show the regret function $R_N(\boldsymbol{\theta}, \mathbf{t}) - \phi(p_N(\boldsymbol{\theta}))$ in (2.12) approaches zero at some rate uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ as $N \to \infty$. See Theorems 5.1 and Corollaries 5.3.1 and 7.1.1. We also examine the more delicate problem of convergence of the average loss of N games. That is, we examine conditions under which $W_N(\boldsymbol{\theta}, \mathbf{t}) - \phi(p_N(\boldsymbol{\theta}))$ converges in probability uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ as $N \to \infty$, where $W_N(\boldsymbol{\theta}, \mathbf{t})$ is given by (2.5). Studying this last type of convergence, which although technically more difficult, provides more insight and enables this work to be more closely related to the "experience theory" approach of [4], [10], [24], and [29]. Also as was pointed out by Samuel in [22] one actually incurs losses rather than risks.

Returning to the problem of proposing sequences of strategies $\mathbf{t}$, note that the procedure $t(\xi, x)$ with components in (2.9) depends on the finite family $f = (f_1(x), \cdots, f_m(x))$ of densities (2.2) as well as on ξ and x. Hence, we write t in (2.9) as $t(f, \xi) = t(f, \xi, x)$ with components $d = 1, \cdots, n$ given by,

$$\begin{aligned} t_d(f, \xi, x) &= 0 && \text{if} \quad (\xi, L^d f(x)) > \min_j \ (\xi, L^j f(x)) \\ &= 1 && \text{if} \quad (\xi, L^d f(x)) < \min_{j \neq d} \ (\xi, L^j f(x)) \\ &= \text{arbitrary} && \text{if} \quad (\xi, L^d f(x)) = \min_j \ (\xi, L^j f(x)). \end{aligned} \tag{3.2}$$

Similarly, (2.8) and (2.11) rewritten displaying this dependence become

$$r(f, \xi, t) = \int \textstyle\sum_{d=1}^{n} (\xi, L^d f(x)) t_d(x) \, d\mu(x) \tag{3.3}$$

$$\phi(f, \xi) = r(\xi, t(f, \xi)). \tag{3.4}$$

Note also that from the definition of $\phi(f, \xi)$ as $\inf_t r(f, \xi, t)$, we have

$$r(f, \xi, t) \geqq \phi(f, \xi) \quad \text{for all} \quad t, f \text{ and } \xi. \tag{3.5}$$

The sequential strategies we propose can now be simply defined. At stage k, $k \geqq 2$, estimate for each $\theta \varepsilon \Omega$ the density $f_\theta(x)$ by $\hat{f}_{k\theta}(x)$ where this estimate is based on the past moves of Nature $\boldsymbol{\theta}_{k-1}$ and the past observations $\mathbf{X}_{k-1}$. Substitute these estimates into a current Bayes solution (represented by (3.2)) on X_k with respect to the $(k - 1)$st stage empirical distribution on Ω given by $p_{k-1}(\boldsymbol{\theta})$. The resulting sequence $\mathbf{t}$ is the one we shall study.

Specifically, let $\mathbf{g} = \{g_k(u, x)\}$ be a sequence of real-valued measurable functions on $\mathfrak{X} \times \mathfrak{X}$ each of which is an estimator of $f_\theta(x)$ for each $\theta \varepsilon \Omega$, in the sense that $g_k(u, x) \geqq 0$ on $\mathfrak{X} \times \mathfrak{X}$, $\int g_k(u, x)\, d\mu(x) = 1$ a.e. μ and $g_{k\theta}(x) = E_\theta g_k(U, x) = \int g_k(u, x) f_\theta(u)\, d\mu(u) < \infty$ a.e. μ. In the discrete case, we give a sequence of unbiased estimators in the sense that $g_{k\theta}(x) = f_\theta(x)$ a.e. μ, while in the continuous case, the sequence $\mathbf{g}$ will be asymptotically unbiased in the sense that $g_{k\theta}(x) \to f_\theta(x)$ a.e. μ as $k \to \infty$. Precise discussion of estimator sequences $\mathbf{g}$ are deferred until later.

Given the sequence $\mathbf{g}$ of estimators, define for each $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, the vector sequence of estimators $\hat{\mathbf{f}} = \{\hat{f}_k\}$ with $\hat{f}_k = (\hat{f}_{k1}(x), \cdots, \hat{f}_{km}(x))$ where

$$(3.6) \qquad \begin{aligned} \hat{f}_{k\theta}(x) &= \{\textstyle\sum_{\nu=1}^{k} \delta_{\theta_\nu \theta}\}^{-1} \textstyle\sum_{\nu=1}^{k} \delta_{\theta_\nu \theta}\, g_\nu(X_\nu, x) && \text{if} \quad p_{k,\theta}(\boldsymbol{\theta}) > 0. \\ &= 0 && \text{if} \quad p_{k,\theta}(\boldsymbol{\theta}) = 0 \end{aligned}$$

Note that the sequence $\hat{\mathbf{f}}$ of estimators depends on $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ even though we have not displayed this dependence in (3.6).

The rule we propose substitutes $\hat{f}_{k-1} = (\hat{f}_{k-1,1}, \cdots, \hat{f}_{k-1,m})$ for $f = (f_1, \cdots, f_m)$ and $p_{k-1}(\boldsymbol{\theta})$ for ξ in (3.2) in the kth component game, $k \geqq 1$, (taking $\xi_0 = \hat{f}_0 = \mathbf{0}$, the zero vector in E^m). Thus, we define $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ in (3.1) as

$$(3.7) \qquad \begin{aligned} &\mathbf{t}_k^*(\boldsymbol{\theta}_{k-1}, \mathbf{X}_k) = t(\hat{f}_{k-1}, p_{k-1}(\boldsymbol{\theta}), X_k) \text{ with } d\text{th component} \\ &t_d(\hat{f}_{k-1}, p_{k-1}(\boldsymbol{\theta}), X_k),\ d = 1, \cdots, n, \text{ where } t_d(f, \xi, x) \text{ is given by (3.2)} \\ &\text{and } \hat{f}_{k-1} \text{ for } k \geqq 2 \text{ has its } \theta\text{th component defined by (3.6).} \end{aligned}$$

We assume that the arbitrary values in (3.2) (which may be functions of f and ξ) are defined to include the range of $(\hat{f}_{k-1}, p_{k-1})$, $k \geqq 1$. We note that this can always be done. For example, one such selection is $t(f, \xi, x)$ as in (2.10).

A sequence $\mathbf{t}^*$ so defined satisfies two lemmas which are vital to the asymptotic results obtained in the discrete case (Section 5) and the continuous case (Section 7) and are given in the next section.

4. Some useful lemmas.

LEMMA 4.1. *Let* $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ *be a strategy sequence defined by* (3.7) *and let* $w_{k-1}(\theta, x) = (L_\theta, t(\hat{f}_{k-1}, p_{k-1}, x))$. *Then*

$$(4.1) \qquad N^{-1} \textstyle\sum_{k=1}^{N} \mathbf{E}[w_{k-1}(\theta_k, X_k) \mid \mathbf{X}_{k-1}] \leqq A_N + B_N + \phi(\hat{f}_N, p_N)$$

$$(4.2) \qquad N^{-1} \textstyle\sum_{k=1}^{N} \mathbf{E}[w_{k-1}(\theta_k, X_k) \mid \mathbf{X}_{k-1}] \geqq A_N + \phi(\hat{f}_N, p_N),$$

where $\mathbf{E}[\quad \mid \mathbf{X}_{k-1}]$ *denotes conditional expectation given* $\mathbf{X}_{k-1}$ *(unconditional expectation for* $k = 1$), $\phi(f, \xi)$ *is as in* (3.4) *and*

$$A_N = N^{-1} \textstyle\sum_{k=1}^{N} \int w_{k-1}(\theta_k, x)\{f_{\theta_k}(x) - g_k(X_k, x)\}\, d\mu(x)$$

$$B_N = N^{-1} \textstyle\sum_{k=1}^{N} \int \{w_{k-1}(\theta_k, x) - w_k(\theta_k, x)\} g_k(X_k, x)\, d\mu(x).$$

PROOF. To obtain Inequality (4.1) observe that

$$w_{k-1}(\theta_k, x)f_{\theta_k}(x) = w_{k-1}(\theta_k, x)\{f_{\theta_k}(x) - g_k(X_k, x)\} + \{w_{k-1}(\theta_k, x) - w_k(\theta_k, x)\}g_k(X_k, x) + w_k(\theta_k, x)g_k(X_k, x).$$

This equality when averaged on k, together with noting that

$$\mathbf{E}[w_{k-1}(\theta_k, X_k) \mid \mathbf{X}_{k-1}] = \int w_{k-1}(\theta_k, x)f_{\theta_k}(x)\, d\mu(x) \tag{4.3}$$

yields (4.1) with inequality replaced by equality and $\phi(\hat{f}_n, \xi_n)$ replaced by

$$C_N = N^{-1} \textstyle\sum_{k=1}^{N} \int w_k(\theta_k, x)g_k(X_k, x)\, d\mu(x). \tag{4.4}$$

To complete the proof of (4.1) it suffices to show that $C_N \leqq \phi(\hat{f}_N, p_N)$. Let $g_k' = g_k(X_k, x)\mathbf{1}$, where $\mathbf{1}$ is the vector of 1's in E^m and let $w_k'(x)$ denote the m-vector $(w_k(1, x), \cdots, w_k(m, x))$. Then, with $\psi_k = \sum_{\nu=1}^{k} \epsilon_{\theta_\nu} (\psi_0 = \mathbf{0}$ and $\epsilon_i = (\delta_{i1}, \cdots, \delta_{im}))$, we can express using (3.6) the integrand of the kth summand of (4.4) as

$$w_k(\theta_k, x)g_k(X_k, x) = (\epsilon_{\theta_k} g_k', w_k'(x)) = (\psi_k \hat{f}_k - \psi_{k-1}\hat{f}_{k-1}, w_k'(x)).$$

By rearranging the order of summation in (4.4) we obtain from this equality and (4.4),

$$C_N = N^{-1} \textstyle\sum_{k=1}^{N-1} \int (\psi_k \hat{f}_k(x), w_k'(x) - w_{k+1}'(x))\, d\mu(x) + \int (p_N \hat{f}_N(x), w_N'(x))\, d\mu(x). \tag{4.5}$$

But the kth summand of the first term on the right-hand side of (4.5) can be written as (see (3.2), (3.3), and (3.4)),

$$k \int \textstyle\sum_{d=1}^{n} \{(p_k \hat{f}_k(x), L^d)\{t_d(\hat{f}_k, p_k, x) - t_d(\hat{f}_{k+1}, p_{k+1}, x)\}\}\, d\mu(x) = k\{\phi(\hat{f}_k, p_k) - r(\hat{f}_k, p_k, t(\hat{f}_{k+1}, p_{k+1}))\},$$

which by (3.5) is non-positive. Hence, by (4.5) we have

$$C_N \leqq \int (p_N \hat{f}_N(x), w_N'(x))\, d\mu(x) = \int \textstyle\sum_{d=1}^{n} (p_N \hat{f}_N(x), L^d) t_d(\hat{f}_N, p_N, x)\, d\mu(x) = \phi(\hat{f}_N, p_N),$$

where the last equality follows by (3.3) and (3.4). This completes the proof of (4.1).

To obtain Inequality (4.2) observe that

$$w_{k-1}(\theta_k, x)f_{\theta_k}(x) = w_{k-1}(\theta_k, x)\{f_{\theta_k}(x) - g_k(X_k, x)\} + w_{k-1}(\theta_k, x)g_k(X_k, x).$$

This equality, when averaged on k, yields by (4.3) the equation (4.2) with inequality replaced by equality and $\phi(\hat{f}_N, p_N)$ replaced by

$$D_N = N^{-1} \sum_{k=1}^{N} \int w_{k-1}(\theta_k, x) g_k(X_k, x)\, d\mu(x). \tag{4.6}$$

But the integrand of the kth summand of (4.6) can be written as

$$\begin{aligned} w_{k-1}(\theta_k, x) g_k(X_k, x) &= (\epsilon_{\theta_k} g_k', w_{k-1}'(x)) \\ &= (\psi_k \hat{f}_k - \psi_{k-1}\hat{f}_{k-1}, w_{k-1}'(x)), \end{aligned}$$

which by a rearrangement in order of summation and (4.6) yields

$$\begin{aligned} D_N &= N^{-1} \sum_{k=1}^{N} \int (\psi_k \hat{f}_k(x), w_{k-1}'(x) - w_k'(x)) d\mu(x) \\ &\quad + \int (p_N \hat{f}_N(x), w_N'(x)) d\mu(x) \\ &= N^{-1} \sum_{k=1}^{N} k\{r(\hat{f}_k, p_k, t(\hat{f}_{k-1}, p_{k-1})) - \phi(\hat{f}_k, p_k)\} \\ &\quad + \phi(\hat{f}_N, p_N). \end{aligned}$$

Thus, by (3.5) applied for each k in the first term of the above we see $D_N \geqq \phi(\hat{f}_N, p_N)$ and (4.2) is proved.

The following lemma is a generalization of Lemma 2 of [22] to uniformity on an index set Θ.

LEMMA 4.2. *Let $(S, \mathcal{A})$ be a measurable space upon which is defined a class $\{P_\theta \mid \theta \varepsilon \Theta\}$ of probability measures. Let $\{Y_k(\theta)\} = \{Y_k(\theta, s)\}$ be a martingale sequence of real-valued random variables for every $\theta \varepsilon \Theta$ and define $\sigma_k^2(\theta) =$ Var $\{Y_k(\theta) - Y_{k-1}(\theta)\}$, $k \geqq 1$, $Y_0(\theta) = 0$. For every $\theta \varepsilon \Theta$, let $\{b_k(\theta)\}$ be a sequence such that $b_k \leqq b_k(\theta) \leqq \beta_k$, where $b_k \uparrow \infty$. If there exists a finite positive constant C independent of $\theta \varepsilon \Theta$ such that*

$$\sum_{k=1}^{\infty} \{\sigma_k^2(\theta)/b_k^2(\theta)\} \leqq C \text{ for } \theta \varepsilon \Theta \tag{4.7}$$

then,

(4.8) $b_N^{-1}(\theta) Y_N(\theta) \rightarrow_{\text{a.s.}} 0$ *and uniformly in* $\theta \varepsilon \Theta$. *(p-uniformly in the sense of Parzen* [16], *that is,* $P_\theta[b_N^{-1}(\theta) Y_N(\theta) \geqq \epsilon$ *for some* $N \geqq N_0]$ $\rightarrow 0$ *uniformly in* $\theta \varepsilon \Theta$ *as* $N \rightarrow \infty$).

PROOF. Define $X_k(\theta) = Y_k(\theta) - Y_{k-1}(\theta)$, $k \geqq 1$. By the martingale property $E[X_k(\theta) \mid X_1(\theta), \cdots, X_{k-1}(\theta)] = 0$ a.s. for every $\theta \varepsilon \Theta$. Use the extended Kolmogorov inequality (Loève, [13], C, p. 386) in the proof of Theorem 16A in Parzen [16] in place of the Kolmogorov inequality, take $EX_k(\theta) = 0$ therein and obtain $b_N^{-1}(\theta) \sum_{k=1}^{N} X_k(\theta) = b_N^{-1}(\theta) Y_N(\theta) \rightarrow 0$ a.s. and uniformly in $\theta \varepsilon \Theta$.

Hereafter, when talking about a.s. (or in probability) uniform convergence of sequences of r.v.'s $\{Y_k(\theta)\}$ we shall always mean p-uniform strong (or weak) convergence in the sense of Parzen [16] as in (4.8) above.

Lemma 4.2 yields the following result to be used in connection with Lemma 4.1 for proving later results.

LEMMA 4.3. *Let* $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ *be a strategy sequence defined by* (3.7). *Then,* (4.9) $N^{\frac{1}{2}}(\log N)^{-1}Z_N(\boldsymbol{\theta}_N, \mathbf{X}_N) \rightarrow_{\text{a.s.}} 0$ (**P**) *and uniformly in* $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$,

where $Z_N(\boldsymbol{\theta}_N, \mathbf{X}_N) = W_N(\boldsymbol{\theta}, \mathbf{t}^*) - N^{-1}\sum_{k=1}^N \mathbf{E}[w_{k-1}(\theta_k, X_k) \mid \mathbf{X}_{k-1}]$, *where* $W_N(\boldsymbol{\theta}, \mathbf{t}^*)$ *is defined by* (2.5).

Proof. Fix $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. Define $Y_N = NZ_N(\boldsymbol{\theta}_N, \mathbf{X}_N)$. By the definition of $w_{k-1}(\theta_k X_k)$ (see Lemma 4.1), we have $Y_N = Y_{N-1} + w_{N-1}(\theta_N, X_N) - \mathbf{E}[w_{N-1}(\theta_N, X_N) \mid \mathbf{X}_{N-1}]$. But by the smoothing property of conditional expectations $\mathbf{E}[w_{N-1}(\theta_N, X_N) \mid Y_1, \cdots, Y_{N-1}] = \mathbf{E}[\mathbf{E}[w_{N-1}(\theta_N, X_N)|\mathbf{X}_{N-1}]|Y_1, \cdots, Y_{N-1}]$. Hence, $\mathbf{E}[Y_N \mid Y_1, \cdots, Y_{N-1}] = Y_{N-1}$, $N \geqq 1$, and $Y_N = Y_N(\boldsymbol{\theta})$ is a martingale for each $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. Noting that $|Y_N(\boldsymbol{\theta}) - Y_{N-1}(\boldsymbol{\theta})| \leqq 2L^*$, we can apply Lemma 4.2 with $b_k(\boldsymbol{\theta}) \equiv k^{\frac{1}{2}} \log k$, $k \geqq 2$, $b_1(\boldsymbol{\theta}) = 1$, and $C = (2L^*)^2\{\sum_{k=2}^{\infty}(k \log^2 k)^{-1} + 1\}$ to obtain $N^{-\frac{1}{2}}(\log N)^{-1}Y_N(\boldsymbol{\theta}) \rightarrow 0$ a.s. (**P**) and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, from whence (4.9) follows.

In addition to Lemmas 4.1 and 4.3 the following inequalities are utilized in later proofs. As a direct consequence of (3.3), (3.4), and (3.5) we have a.e. **P**,

$$
\begin{aligned}
&\int \textstyle\sum_{d=1}^n (L^d, p_N(\hat{f}_N(x) - f(x)))t_d(\hat{f}_N, p_N, x)\,d\mu(x) \\
(4.10) \qquad &\leqq \phi(\hat{f}_N, p_N) - \phi(f, p_N) \\
&\leqq \int \textstyle\sum_{d=1}^n (L^d, p_N(\hat{f}_N(x) - f(x)))t_d(f, p_N, x)\,d\mu(x).
\end{aligned}
$$

Since the t_d's are probabilities on $\mathfrak{D}$ and thus bounded by 1, we see that (4.10) implies a.e. **P**,

$$
\begin{aligned}
(4.11) \quad &|\phi(\hat{f}_N, p_N) - \phi(f, p_N)| \\
&\leqq \textstyle\sum_{d=1}^n \sum_{\theta=1}^m |L(\theta, d)| \int |p_{N\theta}\{\hat{f}_{N\theta}(x) - f_\theta(x)\}|\,d\mu(x).
\end{aligned}
$$

With the aid of results given in this section we now pass to consideration of the discrete case.

5. The discrete case. In the discrete case the class $\mathcal{P}$ is taken as $\mathcal{P}_1$ of Section 3. Thus $f_\theta(x) = \Pr\{X = x\}$ and $0 \leqq f_\theta(x) \leqq 1$, $\sum_{x \varepsilon \mathfrak{X}} f_\theta(x) = \int f_\theta(x)\,d\mu(x) = 1$.

In order to arrive at the strategy sequence $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ defined by (3.7) we need only specify the sequence $\mathbf{g} = \{g_k(u, x)\}$ of estimators. In the discrete case we take

$$(5.1) \qquad g_k(u, x) = g(u, x), k = 1, 2, \cdots,$$

where $g(u, x) = 1$ or 0 as $u = x$ or $u \neq x$. Then, the sequence $\mathbf{g}$ satisfies the necessary conditions of Section 3 since $g(u, x) \geqq 0$, is measurable on $\mathfrak{X} \times \mathfrak{X}$, $\sum_{x \varepsilon \mathfrak{X}} g(u, x) = 1(\int g_k(u, x)\,d\mu(x) = 1)$ for $u \varepsilon \mathfrak{X}$, and $g_{k\theta}(x) = E_\theta g(X, x) = f_\theta(x) < \infty$ a.e.μ. Furthermore, from this last equality we have by (3.6) $\hat{f}_{k\theta}(x) = \{kp_{k\theta}\}^{-1}\sum_{\nu=1}^k \delta_{\theta_\nu\theta}\, g(X_\nu, x)$ the kth stage empirical distribution of those random variable X_ν which are distributed as P_θ, satisfies.

$$(5.2) \qquad \mathbf{E}\hat{f}_{k\theta}(x) = f_\theta(x), \qquad x \varepsilon \mathfrak{X}, \theta = 1, \cdots, m.$$

That is, $\hat{f}_k(x) = \{\hat{f}_{k1}(x), \cdots, \hat{f}_{km}(x)\}$, is the kth stage unbiased estimate of the vector of probabilities $f(x) = \{f_1(x), \cdots, f_m(x)\}$ for all $x \varepsilon \mathfrak{X}$.

Henceforth, in referring to the discrete case we shall always take $g_k(u, x)$ as in (5.1) to define the sequence strategy $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ of (3.7). The following theorem holds.

THEOREM 5.1. *In the discrete case, if P_θ is non-degenerate and $\sum_{x \varepsilon \mathfrak{X}} \{f_\theta(x)\}^{\frac{1}{2}} < \infty$ for each $\theta \varepsilon \Omega$, then the sequence $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ of (3.7) satisfies $|R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta}))| \leqq c' N^{-\frac{1}{2}}$ where c' is independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.*

PROOF. Observe that by definition of $w_{k-1}(\theta_k, X_k)$ in Lemma 4.1, we have

$$R_N(\boldsymbol{\theta}, \mathbf{t}^*) = N^{-1} \textstyle\sum_{k=1}^N \mathbf{E}\{\mathbf{E}[w_{k-1}(\theta_k, X_k)|\mathbf{X}_{k-1}]\}. \tag{5.3}$$

Hence, by Lemma 4.1 the result will be proved if we appropriately uniformly bound the terms $\mathbf{E}A_N$ and $\mathbf{E}\{\phi(\hat{f}_N, p_N) - \phi(f, p_N)\}$ from above and below and $\mathbf{E}B_N$ from above.

By unbiasedness of $g(X_k, x)$ and independence of X_k and $w_{k-1}(\theta_k, x)$, $k \geqq 1$, we have

$$\mathbf{E}A_N = 0 \quad \text{uniformly in} \quad \boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}. \tag{5.4}$$

Similarly, (5.2) and the upper inequality of (4.10) imply

$$\mathbf{E}\{\phi(\hat{f}_N, p_N) - \phi(f, p_N)\} \leqq 0 \quad \text{uniformly in} \quad \boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}. \tag{5.5}$$

Thus, (5.3), (5.4), (5.5), and Lemma 4.1 complete the proof if we find positive constants α_1 and α_2 independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ such that (i) $\mathbf{E}\{\phi(\hat{f}_N, p_n) - \phi(f, p_N)\} \geqq -\alpha_1 N^{-\frac{1}{2}}$ and (ii) $\mathbf{E}B_N \leqq \alpha_2 N^{-\frac{1}{2}}$.

(i) Note that for $\theta \varepsilon \Omega$, $x \varepsilon \mathfrak{X}$, we have

$$\begin{aligned} \mathbf{E}|\hat{f}_{N\theta}(x) - f_\theta(x)|^2 &= \{Np_{N\theta}\}^{-1} E_\theta\{g(X, x) - f_\theta(x)\}^2 \\ &= \{Np_{N\theta}\}^{-1} f_\theta(x)(1 - f_\theta(x)). \end{aligned} \tag{5.6}$$

In Inequality (4.11), interchange of the order of integration and the Schwarz inequality combine with (5.6) to yield

$$\mathbf{E}|\phi(\hat{f}_N, p_n) - \phi(f, p_N)| \leqq N^{-\frac{1}{2}} \textstyle\sum_{d=1}^n \sum_{\theta=1}^m |L(\theta, d)| p_{N\theta}^{\frac{1}{2}} q_\theta$$

where $q_\theta = \sum_{x \varepsilon \mathfrak{X}} \{f_\theta(x)(1 - f_\theta(x))\}^{\frac{1}{2}}$. The finiteness of q_θ for $\theta \varepsilon \Omega$ follows from the summability assumption on $f_\theta^{\frac{1}{2}}(x)$. Uniformity in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ follows from the Schwarz m-space inequality applied to the above to yield

$$\mathbf{E}|\phi(\hat{f}_N, p_N) - \phi(f, p_N)| \leqq N^{-\frac{1}{2}} \alpha_1, \tag{5.7}$$

with $\alpha_1 = \sum_{d=1}^n \|L^d q\|$, $q = (q_1, \cdots, q_m)$. Thus, (i) is proved.

(ii) To bound $\mathbf{E}B_N$ observe that (3.7), (3.2) the definition of $w_k(\theta_k, x)$ (see Lemma 4.1), and our use of bracket notation for characteristic functions yields,

$$\begin{aligned} &|\{w_{k-1}(\theta_k, x) - w_k(\theta_k, x)\} g_k(X_k, x)| \\ &= |\textstyle\sum_{d \neq d'} L_{\theta_k}^{dd'} g_k(X_k, x) t_d(\hat{f}_{k-1}, p_{k-1}, x) t_{d'}(\hat{f}_k, p_k, x)| \\ &\leqq \textstyle\sum_{d \neq d'} |L_{\theta_k}^{dd'}| g_k(X_k, x) [-L_{\theta_k}^{dd'} g_k(X_k, x) \leqq \sum_{\nu=1}^{k-1} L_{\theta_\nu}^{dd'} g_\nu(X_\nu, x) \leqq 0]. \end{aligned} \tag{5.8}$$

Fix θ, d, d', $(d \neq d')$, k such that $\theta_k = \theta$ and x. Apply the Berry-Esseen normal approximation in the form of Lemma 2.1 to the sum of the $(k - 1)p_{k-1,\theta}$ variables $S_{k-1,\theta} = \sum_{\nu=1}^{k-1} L_{\theta_\nu}^{dd'} \delta_{\theta_\nu\theta} g(X_\nu, x)$ in the $\mathbf{E}_{k-1}$ integral of the characteristic function on the right-hand side of (5.8) for fixed X_ν, $\nu < k$, $\theta_\nu \neq \theta$. Since the sum $S_{k-1,\theta}$ of independent and identically distributed random variables falls into an interval of length $|L_\theta^{dd'}| g(X_k, x) \leqq L$ when this characteristic function is one and has variance $s_{k-1,\theta}^2(x) = |L_\theta^{dd'}|^2 (k-1) p_{k-1,\theta}\{f_\theta(x)(1 - f_\theta(x))\}$, we have the Berry-Esseen bound if $p_{k-1,\theta} > 0$, $|L_\theta^{dd'}| > 0$, and $0 < f_\theta(x) < 1$ given by

$$\mathbf{E}_{k-1}[-L_\theta^{dd'} g(X_k, x) \leqq \textstyle\sum_{\nu=1}^{k-1} L_{\theta_\nu}^{dd'} g(X_\nu, x) \leqq 0] \leqq \{s_{k-1,\theta}(x)\}^{-1}\gamma, \tag{5.9}$$

where $\gamma = L(2\pi)^{-\frac{1}{2}} + 2\beta$, β the Berry-Esseen constant. Hence, with the convention that $y/0 = \infty$ for y positive real, (5.8) and (5.9) imply

$$\begin{aligned} &\mathbf{E}|w_{k-1}(\theta_k, x) - w_k(\theta_k, x)| g(X_k, x) \\ &\quad\leqq \textstyle\sum_{d \neq d'} |L_\theta^{dd'}| \min\{f_\theta(x), \gamma s_{k-1,\theta}^{-1}(x) f_\theta(x)\} \\ &\quad\leqq n(n-1)\min\{L f_\theta(x), \gamma\{(k-1)p_{k-1,\theta}\}^{-\frac{1}{2}}\{f_\theta(x)/(1 - f_\theta(x))\}^{\frac{1}{2}}\}. \end{aligned} \tag{5.10}$$

Summing (5.10) on all $\theta_k = \theta$, followed by summing on $\theta = 1, \cdots, m$ yields after interchanging integration on $\mathbf{E}$ and summation on x,

$$\begin{aligned} \mathbf{E}|B_N| &\leqq \textstyle\sum_{x \varepsilon \mathfrak{X}} N^{-1} \sum_{k=1}^{N} \mathbf{E}|w_{k-1}(\theta_k, x) - w_k(\theta_k, x)| g(X_k, x) \\ &\leqq n(n-1) \textstyle\sum_x \sum_\theta (N^{-1} \sum_{j=1}^{N p_{N\theta}} \min\{L f_\theta(x), \gamma (j-1)^{-\frac{1}{2}} \\ &\qquad \{f_\theta(x)/1 - f_\theta(x)\}^{\frac{1}{2}}\}) \\ &\leqq n(n-1) \textstyle\sum_x \{mL \max_\theta f_\theta(x) N^{-1} + \min\{L \max_\theta f_\theta(x), \\ &\qquad \gamma a_N q(x)\}\} \end{aligned} \tag{5.11}$$

where $q(x) = \sum_{\theta=1}^{m} \{f_\theta(x)/1 - f_\theta(x)\}^{\frac{1}{2}}$ and $a_N = N^{-1} \sum_{j=2}^{N} (j-1)^{-\frac{1}{2}} \leqq 2N^{-\frac{1}{2}}$ by (4.10) of [27].

Finally, observe that $\sum_{x \varepsilon \mathfrak{X}} \max_\theta f_\theta(x) \leqq m$ and $q_0 = \sum_{x \varepsilon \mathfrak{X}} q(x)$ is a finite constant by non-degeneracy of P_θ and summability of $f_\theta^{\frac{1}{2}}(x)$ for all $\theta \varepsilon \Omega$. Hence, by (5.11) we have

$$\mathbf{E}|B_N| \leqq n(n-1)\{\min\{mL, 2\gamma q_0 N^{-\frac{1}{2}}\} + m^2 L N^{-1}\} \tag{5.12}$$

from whence (ii) follows. This completes the proof.

THEOREM 5.2. *In the discrete case, if P_θ is non-degenerate and $\sum_{x \varepsilon \mathfrak{X}} \{f_\theta(x)\}^{\frac{1}{2}} < \infty$ for each $\theta \varepsilon \Omega$, then the sequence $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ of (3.7) satisfies $N^{\frac{1}{2}} (\log N)^{-1} \{W_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta}))\} \to_{\mathbf{P}} 0$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.*

PROOF. By Lemmas 4.1 and 4.3 it suffices to show that (i) $N^{\frac{1}{2}} (\log N)^{-1} A_N \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, (ii) $N^{\frac{1}{2}} (\log N)^{-1} B_N \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, and (iii) $N^{\frac{1}{2}} (\log N)^{-1}\{\phi(\hat{f}_N, p_N) - \phi(f, p_N)\} \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. Markov's inequality (Loève, p. 158) together with (5.12) implies (ii), and with (5.7) implies (iii). To obtain (i), define $Y_N = Y_N(\boldsymbol{\theta}) = N A_N$, and observe that for $N \geqq 1$, $Y_0 = 0$.

$$\begin{aligned} \mathbf{E}[Y_N | Y_1, \cdots, Y_{N-1}] &= Y_{N-1} \\ &\quad + \mathbf{E}\{\textstyle\sum_{x \varepsilon \mathfrak{X}} w_{N-1}(\theta_N, x)\{f_{\theta_N}(x) - g(X_N, x)\} | Y_1, \cdots, Y_{N-1}\}. \end{aligned}$$

Use Fubini's theorem for conditional expectations to interchange $\mathbf{E}[\ \mid Y_1, \cdots, Y_{N-1}]$ and summation on $\mathfrak{X}$ in the above to obtain by independence of X_N and $Y_1, \cdots, Y_{N-1}$ and unbiasedness of $g(X_N, x)$ that $Y_N(\boldsymbol{\theta})$ for each $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ is a martingale sequence with $|Y_k(\boldsymbol{\theta}) - Y_{k-1}(\boldsymbol{\theta})| \leqq 2L^*$. Hence, by Lemma 4.2 with $b_N(\boldsymbol{\theta}) \equiv N^{\frac{1}{2}} \log N$, $N \geqq 2$, we obtain $\{N^{\frac{1}{2}} \log N\}^{-1} Y_N(\boldsymbol{\theta}) = N^{\frac{1}{2}} (\log N)^{-1} A_N \to 0$ a.s. and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, from whence (i) follows. This completes the proof.

We remark that in the above proof the convergence in (i) was strong (with probability one) convergence and that with care and (4.11) the same can be shown for the convergence in (iii). Hence, Theorem 5.2 would hold strongly if in (ii) we could show that $N^{\frac{1}{2}} (\log N)^{-1} B_N \to 0$ with probability one and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. However, this we were unable to do, thus requiring weak convergence in Theorem 5.2.

It is easy to see that the condition $\sum_x f_\theta^{\frac{1}{2}}(x) < \infty$ of Theorems 5.1 and 5.2 is satisfied when $\mathfrak{X}$ is finite or when P_θ is a lattice distribution on the real line having finite second moment.

The following theorem shows that the assumption of summability of $f_\theta^{\frac{1}{2}}(x)$ on $\mathfrak{X}$ may be dropped to obtain a result of lower order convergence.

THEOREM 5.3. *In the discrete case, if P_θ is non-degenerate for each $\theta \varepsilon \Omega$, then the sequence $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ in (3.7) satisfies $W_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta})) \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.*

PROOF. In view of Lemmas 4.1 and 4.3 and the proof of (i) of Theorem 5.2 it suffices to show (i) $B_N \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ and (ii) $\phi(\hat{f}_N, p_N) - \phi(f, p_N) \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

To verify (i), note that in (5.11) non-degeneracy of P_θ for $\theta \varepsilon \Omega$ furnishes $\lim_N\{a_N q(x)\} = 0$ for all $x \varepsilon \mathfrak{X}$. Thus, since $\min \{L \max_\theta f_\theta(x), \gamma a_N q(x)\}$ is independent of $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ and bounded by $L \sum_{\theta=1}^m f_\theta(x)$, the bounded convergence theorem applied in (5.11) yields $\mathbf{E}|B_N| \to 0$ as $N \to \infty$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. Markov's inequality now completes the proof of (i).

To verify (ii), observe that for each $x \varepsilon \mathfrak{X}$, $\theta \varepsilon \Omega$, the quantity $p_{N\theta}\{\hat{f}_{N\theta}(x) - f_\theta(x)\}$ is the average of N independent random variables $Y_{k\theta}(\boldsymbol{\theta}, x) = \delta_{\theta_k \theta}\{g(X_k, x) - f_\theta(x)\}$ where $Y_{k\theta}(\boldsymbol{\theta}, x)$ is uniformly bounded in $\boldsymbol{\theta}$ and k and hence by Theorem 15A of Parzen [16] (or Lemma 4.1), $p_{N\theta}\{\hat{f}_{N\theta}(x) - f_\theta(x)\} \to_{\text{a.s.}} 0$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ for each $x \varepsilon \mathfrak{X}$. Thus, $p_{N\theta}|\hat{f}_{N\theta}(x) - f_\theta(x)| \leqq 2$ implies $\sup_{\boldsymbol{\theta}} \mathbf{E} p_{N\theta}|\hat{f}_{N\theta}(x) - f_\theta(x)| \to 0$ as $N \to \infty$ for $x \varepsilon \mathfrak{X}$ and is, for all N, bounded (by unbiasedness of $\hat{f}_{N\theta}(x)$ in (5.2)) by $2f_\theta(x)$, which is summable on $\mathfrak{X}$. Hence the bounded convergence theorem implies

$$\sum_{x \varepsilon \mathfrak{X}} \mathbf{E} p_{N\theta}|\hat{f}_{N\theta}(x) - f_\theta(x)| \to 0 \tag{5.13}$$

uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$ as $N \to \infty$. Using Markov's inequality, (4.11) and Fubini's theorem to obtain

$$\begin{aligned} \epsilon \mathbf{E}[|\phi(\hat{f}_N, p_N) - \phi(f, p_N)| \geqq \epsilon] &\leqq \mathbf{E}|\phi(\hat{f}_N, p_N) - \phi(f, p_N)| \\ &\leqq nL^* \textstyle\sum_{\theta=1}^m \sum_{x \varepsilon \mathfrak{X}} \mathbf{E} p_{N\theta}|\hat{f}_{N\theta}(x) - f_\theta(x)|, \end{aligned}$$

we see that (ii) now follows from (5.13). This completes the proof.

By boundedness of $W_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta}))$, we have

COROLLARY 5.3.1. *In the discrete case, if P_θ is non-degenerate for each $\theta \varepsilon \Omega$, then the sequence $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ satisfies $R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta})) \to 0$ uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.*

We now turn our consideration to the continuous case.

6. Estimation in the continuous case. Estimation of Lebesgue densities has been treated for $q = 1$ by Parzen [15] and for $q > 1$ by Cacoullos [3]. The following is due to them and is stated here as Lemma 6.1. Let $u = (u_1, \cdots, u_q)$ denote a point in E^q and $\int \zeta(u)\, du$ indicate integration with respect to q-dimensional Lebesgue measure for integrable ζ.

LEMMA 6.1. *Let $K(u)$ be a Borel function on E^q such that*

$$\sup_u |K(u)| < \infty, \tag{6.1}$$

$$\int |K(u)|\, du < \infty, \tag{6.2}$$

$$\lim_{\|u\| \to \infty} \|u\|^q |K(u)| = 0. \tag{6.3}$$

Let $\zeta(u)$ be a Lebesgue-integrable real-valued function on E^q and define

$$\zeta_k(x) = h_k^{-q} \int K(h_k^{-1}u)\zeta(x - u)du \tag{6.4}$$

where $\{h_k\}$ is a sequence of numbers with

$$\lim_k h_k = 0, \qquad h_k > 0,\ h_k \downarrow. \tag{6.5}$$

Then for every $x \varepsilon C(\zeta)$, the continuity set of ζ,

$$\lim_k \zeta_k(x) = \zeta(x) \int K(u)\, du. \tag{6.6}$$

In particular, if

$$K(u) \geqq 0 \quad \text{and} \quad \int K(u)\, du = 1 \tag{6.7}$$

then for $x \varepsilon C(\zeta)$,

$$\lim_k \zeta_k(x) = \zeta(x). \tag{6.8}$$

We now define the sequence, $\mathbf{g} = \{g_k(u, x)\}$, of estimators of the densities $f_\theta(x)$ for the continuous case by taking,

$$g_k(u, x) = h_k^{-q} K(h_k^{-1}(x - u)), \tag{6.9}$$

where $K(u)$ is a Borel function on E^q satisfying (6.1), (6.3), and (6.7) and $\{h_k\}$ is a sequence for which (6.5) holds. Then the sequence $\mathbf{g}$ satisfies the necessary conditions of Section 3 since $g_k(u, x) \geqq 0$, is measurable on $E^q \times E^q$, and

$$\int g_k(u, x)\, dx = \int h_k^{-q} K(h_k^{-1}(x - u))\, dx = 1 \text{ for } u \varepsilon E^q. \tag{6.10}$$

Furthermore, Lemma 6.1 guarantees by (6.8) that the estimators $g_k(u, x)$ are asymptotically unbiased in the sense that as $k \to \infty$

$$\begin{aligned} g_{k\theta}(x) &= E_\theta g_k(X, x) = \int h_k^{-q} K(h_k^{-1}(x - u)) f_\theta(u)\, du \\ &= \int h_k^{-q} K(h_k^{-1}u) f_\theta(x - u)\, du \to f_\theta(x) \text{ a.e.}\mu, \end{aligned} \tag{6.11}$$

since $C(f_\theta)$, $\theta \varepsilon \Omega$, has its complement of μ(q-dimensional Lebesgue) measure zero by the assumptions imposed on the class $\mathcal{P}_2$.

Having specified the sequence $\mathbf{g} = \{g_k(u, x)\}$ of estimators (via the kernel function $K(u)$ and (6.9), we see that (3.6) specifies the sequence $\hat{\mathbf{f}} = \{\hat{f}_k\}$ of estimators of $f = (f_1, \cdots f_m)$ used in definining, via (3.7), the strategy sequence $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ for the continuous case. Although the sequence $\mathbf{t}^*$ depends on the kernel function $K(u)$, we suppress this dependence assuming that for the remainder of the paper a fixed kernel $K(u)$ satisfying (6.1), (6.3), and (6.7) has been chosen. For possible choices of $K(u)$ see Table 1 of [15] for $q = 1$ or Table A of [3] for $q > 1$. With the sequence $\mathbf{t}^*$ so specified we examine its convergence properties in the next section. Before doing so however, we state the following consequence of Lemma 6.1 which is given for general densities in [15] (Theorem 2A) and in [3] (Lemma 2.1 and Corollary 2.1).

LEMMA 6.2. *Let* $\alpha \geqq 1$ *and let* (6.5) *hold. Then,* $g_k(X, x)$ *where* X *is distributed as* P_θ, $\theta \varepsilon \Omega$, *satisfies*

$$\lim_k h_k^{q(\alpha-1)} E_\theta g_k^\alpha(X, x) = f_\theta(x) \int K^\alpha(u) du. \tag{6.12}$$

for $x \varepsilon C(f_\theta)$. *In particular, for* $x \varepsilon C(f_\theta)$, *we have,*

$$\lim_k\{h_k^q V_\theta\{g_k(X, x)\}\} = f_\theta(x) \int K^2(u)\, du, \tag{6.13}$$

where $V_\theta\{h(X)\}$ *represents the variance of the* r.v. $h(X)$ *under* P_θ.

Having specified the form of the estimating sequence $\mathbf{g}$ in the continuous case and given certain properties thereof, we now examine the convergence properties of the resulting procedure sequence $\mathbf{t}^*$.

7. The continuous case. The following measure theoretic lemma will be found useful in the sequel.

LEMMA 7.1. *Let* $(\mathfrak{X}', \mathfrak{F}', \mu')$ *be a measure space upon which are defined sequences of integrable functions* $\{\eta_k\}$ *and* $\{\zeta_k\}$ *such that* $|\eta_k(x)| \leqq |\zeta_k(x)|$ a.e.μ'. *If* $\eta_k \to \eta$ *in measure and* $\zeta_k \to \zeta$ *in mean, then* $\eta_k \to \eta$ *in mean.*

PROOF. Noting that for $F \varepsilon \mathfrak{F}'$, $\nu_k'(F) = \int_F |\eta_k|\, d\mu' \leqq \int_F |\zeta_k|\, d\mu' = \nu_k(F)$, the result follows from Theorem C, p. 108 of Halmos [8].

Let $\{h_k\}$ be a sequence satisfying

$$\lim_k k h_k^{4q} = \infty. \tag{7.1}$$

THEOREM 7.1. *In the continuous case, the strategy sequence* $\mathbf{t}^* = \{\mathbf{t}_k^*\}$, *defined by* (3.6), (3.7), *and* (6.9), *with* $\{h_n\}$ *satisfying* (6.5) *and* (7.1) *and* $K(u)$ *satisfying* (6.1), (6.3), *and* (6.7), *satisfies* $W_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta})) \to 0$ *in probability uniformly in* $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

PROOF. By Lemmas 4.1 and 4.3 it suffices to prove that A_N, B_N and $\{\phi(\hat{f}_N, p_N) - \phi(f, p_N)\} \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$. We show more by verifying (i) $A_N \to 0$ with probability one and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, (ii) $\{\phi(\hat{f}_N, p_N) - \phi(f, p_N)\} \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, and (iii) $B_N \to 0$ in probability and uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

(i) Observe that since $\int f_\theta(x)\, dx = \int g_{k\theta}(x)\, dx = 1$, $g_{k\theta}(x)$ given by (6.11), Scheffé's theorem (Scheffé [23]) and (6.11) imply $\max_\theta \int |f_\theta(x) - g_{k\theta}(x)|\, dx \to 0$ as $k \to \infty$. Hence, boundedness of $w_{k-1}(\theta_k, x)$ by L^* and Toeplitz's lemma (Loève, [13], p. 238) yield.

$$\begin{aligned} N^{-1} \textstyle\sum_{k=1}^N \int w_{k-1}(\theta_k, x)\{f_{\theta_k}(x) - g_{k\theta_k}(x)\} dx \\ \leqq L^* N^{-1} \textstyle\sum_{k=1}^N \max_\theta \int |f_\theta(x) - g_{k\theta}(x)|\, dx \\ \to 0 \text{ as } N \to \infty \text{ uniformly in } \boldsymbol{\theta} \in \boldsymbol{\Omega}. \end{aligned} \tag{7.2}$$

Note that the left-hand side of (7.2) differs from A_N by the term

$$A_N' = N^{-1} \textstyle\sum_{k=1}^N \int w_{k-1}(\theta_k, x)\{E_{\theta_k} g_k(X, x) - g_k(X_k, x)\}\, dx, \tag{7.3}$$

where $Y_N' = NA_N'$ is seen to be a martingale sequence by arguments similar to those used for $Y_N = NA_N$ in the proof of (i) of Theorem 5.2. Hence, since $|Y_k'(\boldsymbol{\theta}) - Y_{k-1}'(\boldsymbol{\theta})| \leqq 2L^*$, $(Y_0' = 0)$, we have by Lemma 4.2, $(\log N)^{-1} N^{\frac{1}{2}} A_N' \to_{\text{a.s.}} 0$ uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, from whence (7.2) and (7.3) imply (i).

(ii) By definition of $g_{k\theta}(x)$ in (6.11) we have for $\theta \in \Omega$,

$$\begin{aligned} \sup_{\boldsymbol{\theta}} \int |p_{N\theta}\{\mathbf{E}\hat{f}_{N\theta}(x) - f_\theta(x)\}| dx \\ = \sup_{\boldsymbol{\theta}} \int |N^{-1} \textstyle\sum_{k=1}^N \delta_{\theta_k \theta}\{g_{k\theta}(x) - f_\theta(x)\}|\, dx \\ \leqq N^{-1} \textstyle\sum_{k=1}^N \int |g_{k\theta}(x) - f_\theta(x)|\, dx \\ \to 0 \text{ as } n \to \infty \text{ uniformly in } \boldsymbol{\theta} \in \boldsymbol{\Omega} \text{ as in (7.2)}. \end{aligned} \tag{7.4}$$

Next, observe that by Fubini's theorem, we have

$$\sup_{\boldsymbol{\theta}} \mathbf{E} \int |p_{N\theta}\{\hat{f}_{N\theta}(x) - \mathbf{E}\hat{f}_{N\theta}(x)\}|\, dx \leqq \int \zeta_{N\theta}(x)\, dx \tag{7.5}$$

where,

$$\zeta_{N\theta}(x) = \sup_{\boldsymbol{\theta}} \mathbf{E}|p_{N\theta}\{\hat{f}_{N\theta}(x) - \mathbf{E}\hat{f}_{N\theta}(x)\}|. \tag{7.6}$$

But by the Schwarz inequality, independence of the X_k, and monotonicity of the $\{h_k\}$, we have

$$\begin{aligned} \zeta_{N\theta}(x) &\leqq \sup_{\boldsymbol{\theta}} \{N^{-2} \textstyle\sum_{k=1}^N \delta_{\theta_k \theta} V_\theta\{g_k(X, x)\}\}^{\frac{1}{2}} \\ &\leqq (N h_N^q)^{-\frac{1}{2}} \{N^{-1} \textstyle\sum_{k=1}^N h_k^q V_\theta\{g_k(X, x)\}\}^{\frac{1}{2}}. \end{aligned}$$

Hence, by (7.1) and noting that (6.13) and Toeplitz's lemma imply

$$N^{-1} \textstyle\sum_{k=1}^N h_k^q V_\theta\{g_k(X, x)\} \to f_\theta(x) \int K^2(u)\, du,$$

for $x \in C(f_\theta)$, we have

$$\zeta_{N\theta}(x) \to 0 \text{ for all } x \in C(f_\theta) \text{ as } N \to \infty. \tag{7.7}$$

We combine (7.7) and Lemma 7.1 in the following manner: note that $g_{N\theta}^*(x) = N^{-1} \sum_{k=1}^N g_{k\theta}(x) \to f_\theta(x)$ a.e.μ (for all $x \in C(f_\theta)$ whose complement has Lebesgue measure zero); and that by Scheffé's theorem we obtain $g_{N\theta}^*(x) \to f_\theta(x)$ in mean from whence Lemma 7.1, $\zeta_{N\theta}(x) \leqq 2g_{N\theta}^*(x)$, and (7.7) implying $\zeta_{N\theta}(x) \to 0$ in measure (actually a.e.) yield $\zeta_{N\theta}(x) \to 0$ in mean. Hence, by (7.5) and (7.6), we have as $N \to \infty$

(7.8) $$\sup_{\boldsymbol{\theta}} \mathbf{E} \int p_{N\theta}|\hat{f}_{N\theta}(x) - \mathbf{E}\hat{f}_{N\theta}(x)|\, dx \to 0.$$

Finally, by Markov's inequality we have

$$\epsilon \mathbf{E}[|\phi(\hat{f}_N, p_N) - \phi(f, p_N)| \geqq \epsilon] \leqq \mathbf{E}|\phi(\hat{f}_N, p_N) - \phi(f, p_N)|,$$

from which (4.11), (7.4) and (7.8) yield the desired result (ii) after bounding the integrand on the right-hand side of (4.11) by

$$p_{N\theta}|\hat{f}_{N\theta}(x) - \mathbf{E}\hat{f}_{N\theta}(x)| + p_{N\theta}|\mathbf{E}\hat{f}_{N\theta}(x) - f_\theta(x)|.$$

(iii) By Markov's inequality to prove (iii) it suffices to show that uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Omega}$,

(7.9) $$\mathbf{E}|B_N| \to 0 \text{ as } N \to \infty.$$

To do this we shall make use of Inequality (5.8) which holds for the continuous case also. In (5.8) fix θ, d, d', $(d \neq d')$, k such that $\theta_k = \theta$ and x. Apply the Berry-Esseen normal approximation of Lemma 2.1 to the sum of the $(k-1)p_{k-1,\theta}$ variables $S_{k-1,\theta} = \sum_{\nu=1}^{k-1} \delta_{\theta_\nu\theta} L_\theta^{dd'} g_\nu(X_\nu, x)$ in the $\mathbf{E}_{k-1}$ integral of the characteristic function on the right-hand side of (5.8) for fixed X_ν, $\nu < k$, $\theta_\nu \neq \theta$. Since the sum $S_{k-1,\theta}$ falls into an interval of length $|L_\theta^{dd'}| g_k(X_k, x)$ when this characteristic function is unity and has variance $|L_\theta^{dd'}|^2 s_{k-1,\theta}^2(\boldsymbol{\theta}, x)$ with

(7.10) $$s_{k-1,\theta}^2(\boldsymbol{\theta}, x) = \sum_{\nu=1}^{k-1} \delta_{\theta_\nu\theta} V_\theta(g_\nu(X, x)),$$

we have by Lemma 2.1, if $p_{k-1,\theta}(\boldsymbol{\theta}) > 0$, and $|L_\theta^{dd'}| > 0$,

(7.11) $$\begin{aligned} \mathbf{E}_{k-1}[-L_\theta^{dd'} g_k(X_k, x) \leqq \textstyle\sum_{\nu=1}^{k-1} L_{\theta_\nu}^{dd'} g_\nu(X_\nu, x) \leqq 0] \\ \leqq (2\pi)^{-\frac{1}{2}} g_k(X_k, x) s_{k-1,\theta}^{-k}(\boldsymbol{\theta}, x) + 2\beta\zeta_{k-1,\theta}(\boldsymbol{\theta}, x) \end{aligned}$$

where,

(7.12) $$\zeta_{k-1,\theta}(\boldsymbol{\theta}, x) = s_{k-1,\theta}^{-3}(\boldsymbol{\theta}, x)\zeta'_{k-1,\theta}(\boldsymbol{\theta}, x),$$

with

(7.13) $$\zeta'_{k-1,\theta}(\boldsymbol{\theta}, x) = \sum_{\nu=1}^{k-1} \delta_{\theta_\nu\theta} E_\theta |g_\nu(X, x) - g_{\nu\theta}(x)|^3.$$

Noting that (7.11) is always bounded by unity and summing first on k such that $\theta_k = \theta$, and then on d, d', $d \neq d'$, we have by (7.11), (5.8), and the definition of B_N,

(7.14) $$\mathbf{E}|B_N| \leqq n(n-1)L^{\cdot} \sum_{\theta=1}^{m} \int \{\alpha_{N\theta}(\boldsymbol{\theta}, x) + \alpha'_{N\theta}(\boldsymbol{\theta}, x)\}\, dx,$$

where

(7.15) $$\alpha_{N\theta}(\boldsymbol{\theta}, x) = N^{-1} \sum_{k=1}^{N} \delta_{\theta_k\theta} \min\{g_{k\theta}(x), (2\pi)^{-\frac{1}{2}} s_{k-1,\theta}^{-1}(\boldsymbol{\theta}, x) E_\theta g_k^2(X, x)\}$$

and,

$$(7.16)\qquad \alpha'_{N\theta}(\boldsymbol{\theta}, x) = N^{-1} \textstyle\sum_{k=1}^{N} \delta_{\theta_k\theta} g_{k\theta}(x) \min\{1, 2\beta\zeta_{k-1,\theta}(\boldsymbol{\theta}, x)\}$$

with $\zeta_{k-1,\theta}(\boldsymbol{\theta}, x)$ given by (7.12) and $y/0 = \infty$ for y positive and real.

In finding the limit for the right-hand side of (7.14), fix θ and let $\epsilon > 0$ be given. Fix $x \in C(f_\theta)$. By Lemma 6.2 and (6.11) there exists a $k_\theta'(x) \geqq 1$ such that for $k \geqq k_\theta'(x)$,

$$(7.17)\qquad |h_k^q V_\theta\{g_k(X, x)\} - f_\theta(x) \textstyle\int K^2(u)\, du| \leqq \epsilon,$$

and,

$$(7.18)\qquad k^{-1} \textstyle\sum_{\nu=1}^{k} |g_{\nu\theta}(x) - f_\theta(x)| \leqq \epsilon.$$

Furthermore, we see that (6.11) and Lemma 6.2, imply that as $N \to \infty$,

$$(7.19)\qquad \gamma_{N\theta}(x) = \max_{k \leqq N} g_{k\theta}(x) \to \gamma_\theta(x) < \infty,$$

$$(7.20)\qquad \gamma'_{N\theta}(x) = \max_{k \leqq N} \{h_k^q E_\theta g_k^2(X, x)\} \to \gamma_\theta'(x) < \infty,$$

$$(7.21)\qquad \gamma''_{N\theta}(x) = \max_{k \leqq N} \{h_k^{2q} E_\theta |g_k(X, x) - g_{k\theta}(x)|^3\} \to \gamma_\theta''(x) < \infty,$$

since, for fixed $x \in C(f_\theta)$, $\{\gamma_{N\theta}(x)\}$, $\{\gamma'_{N\theta}(x)\}$, and $\{\gamma''_{N\theta}(x)\}$ are monotone bounded sequences.

Let $F_\theta(\epsilon) = \{x \mid f_\theta(x) \int K^2(u)\, du \geqq 2\epsilon\}$. Assume $x \in F_\theta(\epsilon)$. Then for all $k \geqq k_\theta'(x)$, (7.10), (7.17), and monotonicity of the h_k imply,

$$(7.22)\qquad \begin{aligned} s_{k,\theta}^2(\boldsymbol{\theta}, x) &\geqq \textstyle\sum_{\nu=k_\theta'(x)}^{k} \delta_{\theta_\nu\theta} h_1^{-q}\{f_\theta(x) \int K^2(u)\, du - \epsilon\} \\ &\geqq \textstyle\sum_{\nu=k_\theta'(x)}^{k} \delta_{\theta_\nu\theta} h_1^{-q}\epsilon. \end{aligned}$$

Let $k_\theta^*(\theta)$ be the first subscript in the sequence $\boldsymbol{\theta} = \{\theta_k\}$ such that $\theta_k = \theta$. Define $k_\theta = \max\{k_\theta'(x) + 1, k_\theta^*(\boldsymbol{\theta}) + 1\}$. Then, as $N \to \infty$, we have by (7.22), (7.20) and monotonicity of the h_k,

$$(7.23)\qquad \begin{aligned} &N^{-1} \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_k\theta} s_{k-1,\theta}^{-1}(\boldsymbol{\theta}, x) E_\theta g_k^2(X, x) \\ &\quad\leqq (N h_N^q)^{-1} (\epsilon^{-1} h_1^q)^{\frac{1}{2}} \gamma'_{N\theta}(x) \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_\nu\theta} \{\sum_{\nu=k_\theta'(x)}^{k-1} \delta_{\theta_\nu\theta}\}^{-\frac{1}{2}} \\ &\quad\leqq (N^{\frac{1}{2}} h_N^q)^{-1} (\epsilon^{-1} h_1^q)^{\frac{1}{2}} 2\gamma'_{N\theta}(x) \to 0, \end{aligned}$$

where the last inequality follows from (4.10) of [27] implying that

$$(7.24)\qquad \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_k\theta} \{\sum_{\nu=k_\theta'(x)}^{k-1} \delta_{\theta_\nu\theta}\}^{-\frac{1}{2}} \leqq \sum_{j=2}^{N-1} (j-1)^{-\frac{1}{2}} \leqq 2N^{\frac{1}{2}}$$

and convergence follows from (7.1) and (7.20). Also, note that as $N \to \infty$, (7.19) implies

$$(7.25)\qquad N^{-1} \textstyle\sum_{k=1}^{k_\theta - 1} \delta_{\theta_k\theta} g_{k\theta}(x) \leqq k_\theta'(x) (N^{-1} \gamma_{N\theta}(x)) \to 0.$$

Since (7.15) is bounded by the sum of (7.25) and $(2\pi)^{-\frac{1}{2}}$ times (7.23), and since the bounds on the extreme left-hand sides in these two equations are independent of $\boldsymbol{\theta} \in \boldsymbol{\Omega}$, we have proved that,

$$(7.26)\qquad \lim_N \{\sup_{\boldsymbol{\theta}} \alpha_{N\theta}(\boldsymbol{\theta}, x)\} = 0 \text{ for } x \in F_\theta(\epsilon) \cap C(f_\theta).$$

Now assume $x \,\varepsilon\, F_\theta'(\epsilon)$, the complement of $F_\theta(\epsilon)$. Then, (7.15) and (7.18) imply for $N \geqq k_\theta'(x)$,

$$\sup_{\boldsymbol{\theta}} \alpha_{N\theta}(\boldsymbol{\theta}, x) \leqq N^{-1} \sum_{k=1}^{N} g_{k\theta}(x) \leqq \epsilon + f_\theta(x) < \epsilon(1 + 2\{\int K^2(u)\, du\}^{-1}), \tag{7.27}$$

from whence it follows by arbitrariness of ϵ and (7.26) that

$$\lim_N \{\sup_{\boldsymbol{\theta}} \alpha_{N\theta}(\boldsymbol{\theta}, x)\} = 0 \text{ for } x \,\varepsilon\, C(f_\theta). \tag{7.28}$$

Similar to (7.23), we have (7.12), (7.13), (7.22), and (7.21) combining to yield, for $x \,\varepsilon\, F_\theta(\epsilon)$,

$$\begin{aligned} &N^{-1} \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_k\theta}\zeta_{k-1,\theta}(\boldsymbol{\theta}, x) \\ &\quad\leqq (h_1^{\,q}\epsilon^{-1})^{\frac{3}{2}}(N^{-1} \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_k\theta}\{\sum_{\nu=k_\theta'(x)}^{k-1} \delta_{\theta_\nu\theta}\}^{-\frac{3}{2}}\zeta_{k-1,\theta}'(\boldsymbol{\theta}, x)) \\ &\quad\leqq (h_1^{\,q}\epsilon^{-1})^{\frac{3}{2}}\gamma_{N\theta}''(x)(Nh_N^{\,2q})^{-1} \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_k\theta}\{\sum_{\nu=k_\theta'(x)}^{k-1} \delta_{\theta_\nu\theta}\}^{-\frac{3}{2}} \sum_{\nu=1}^{k-1} \delta_{\theta_\nu\theta} \\ &\quad\leqq (h_1^{\,q}\epsilon^{-1})^{\frac{3}{2}}\gamma_{N\theta}''(x)\{2(Nh_N^{\,4q})^{-\frac{1}{2}} \\ &\qquad + (Nh_N^{\,2q})^{-1}(\textstyle\sum_{k=1}^{k_\theta'(x)-1} \delta_{\theta_k\theta}) \sum_{k=k_\theta}^{N} \delta_{\theta_k\theta}\{\sum_{\nu=k_\theta'(x)}^{k-1} \delta_{\theta_\nu\theta}\}^{-\frac{3}{2}}\} \end{aligned} \tag{7.29}$$

where the last inequality is via (7.24). Observe that

$$(\textstyle\sum_{k=1}^{k_\theta'(x)-1} \delta_{\theta_k\theta}) \sum_{k=k_\theta}^{N} \delta_{\theta_k\theta}\{\sum_{\nu=k_\theta'(x)}^{k-1} \delta_{\theta_\nu\theta}\}^{-\frac{3}{2}} < k_\theta'(x) \sum_{j=1}^{N-1} j^{-\frac{3}{2}} < 3k_\theta'(x).$$

Thus, (7.29) yields

$$\begin{aligned} N^{-1} \textstyle\sum_{k=k_\theta}^{N} \delta_{\theta_k\theta}\zeta_{k-1,\theta}(\boldsymbol{\theta}, x) &< 3(h_1^{\,q}\epsilon^{-1})^{\frac{3}{2}}\gamma_{N\theta}''(x)(Nh_N^{\,4q})^{-\frac{1}{2}}\{1 + k_\theta'(x)N^{-\frac{1}{2}}\} \\ &\to 0 \text{ as } N \to \infty, \text{ for } x \,\varepsilon\, F_\theta(\epsilon) \cap C(f_\theta), \end{aligned} \tag{7.30}$$

where the convergence follows from (7.1) and (7.21). Since (7.16) is bounded by the sum of (7.25) and 2β times (7.30) with the upper bounds in these two equations being independent of $\boldsymbol{\theta} \,\varepsilon\, \boldsymbol{\Omega}$, we have

$$\lim_N \{\sup_{\boldsymbol{\theta}} \alpha_{N\theta}'(\boldsymbol{\theta}, x)\} = 0 \text{ for } x \,\varepsilon\, F_\theta(\epsilon) \cap C(f_\theta). \tag{7.31}$$

On $F_\theta'(\epsilon)$ use (7.27) with $\alpha_{N\theta}'(\boldsymbol{\theta}, x)$ replacing $\alpha_{N\theta}(\boldsymbol{\theta}, x)$ to obtain by an argument similar to that yielding (7.28),

$$\lim_N \{\sup_{\boldsymbol{\theta}} \alpha_{N\theta}'(\boldsymbol{\theta}, x)\} = 0 \text{ for } x \,\varepsilon\, C(f_\theta). \tag{7.32}$$

Next, observe that

$$\sup_{\boldsymbol{\theta}} \{\alpha_{N\theta}(\boldsymbol{\theta}, x) + \alpha_{N\theta}'(\boldsymbol{\theta}, x)\} \leqq 2g_{N\theta}^{*}(x), \tag{7.33}$$

where $g_{N\theta}^{*}(x) = N^{-1} \sum_{k=1}^{N} g_{k\theta}(x)$ converges a.e.μ to $f_\theta(x)$ by (6.11) and Toeplitz's lemma. But by Scheffé's theorem since $g_{N\theta}^{*}(x)$ is a density this implies $g_{N\theta}^{*}(x) \to f_\theta(x)$ in mean. Hence, Lemma 7.1, (7.28), (7.32), and (7.33) and the fact that by assumptions on the class $\mathcal{P}_2$ the Lebesgue measure of $C'(f_\theta)$, the complement of $C(f_\theta)$, is zero we have

$$\sup_{\boldsymbol{\theta}}\{\alpha_{N\theta}(\boldsymbol{\theta}, x) + \alpha_{N\theta}'(\boldsymbol{\theta}, x)\} \to 0 \text{ in mean.} \tag{7.34}$$

Equations (7.34) and (7.14) yield (7.9) uniformly in $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$, which by Markov's inequality completes the proof of (iii).

The theorem is thus proved by (i), (ii), and (iii).

Boundedness of $W_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta}))$ yields,

COROLLARY 7.1. *In the continuous case, if the assumptions of Theorem* 7.1 *are satisfied, then the strategy sequence* $\mathbf{t}^* = \{\mathbf{t}_k^*\}$ *satisfies* $R_N(\boldsymbol{\theta}, \mathbf{t}^*) - \phi(f, p_N(\boldsymbol{\theta})) \to 0$ *uniformly in* $\boldsymbol{\theta} \varepsilon \boldsymbol{\Omega}$.

8. Final remarks. We make the following remarks.

If we let the θ_k be independent identically distributed random variables with $\Pr\{\theta_k = \theta\} = \xi_\theta'$, $\xi_\theta' \geqq 0$, $\sum_{\theta=1}^m \xi_\theta' = 1$, then the results of this paper are closely related to the "experience theory" work of [4], [10], [24], and [29]. Specifically, replace $\phi(f, p_N(\boldsymbol{\theta}))$ by $\phi(f, \xi')$, $\xi' = (\xi_1', \cdots, \xi_m')$, in all theorems and interpret in probability statements of Theorems 5.2, 5.3, and 7.1 with regard to the infinite product measure $\mathbf{P} = \times_{k=1}^\infty P_k'$, where P_k' is the joint distribution of (θ_k, X_k) which for each k is i.i.d. with density $\xi_\theta' f_\theta(x)$. Then since $E\{\phi(f, p_N(\boldsymbol{\theta})) - \phi(f, \xi')\}$ is of $O(N^{-\frac{1}{2}})$ (see proof of Theorem 6.1 of [27]) the thus modified Theorems 5.2, 5.3, and 7.1 yield "weak" experience theory results with unit delay time. Compare, for example, Spacek's theorem in [24] and its improved version Theorem 2 in [10] as well as the more general results given by Theorem 9.2 of [29]. What distinguishes our work from these results is that we have *not* required full knowledge of the distributions. Under the assumption of i.i.d. θ_k's this paper (for unit delay time) can thus be viewed as a generalization of experience theory from known to unknown distributions. However, we are able only to obtain weak convergence results rather than strong convergence.

For a more detailed discussion of experience theory see [4], [10], [24], and [29].

Under the assumption of randomness of the θ_k's, Theorems 5.1, 5.3.1, and 7.1.1 are interesting to compare with the "empirical Bayes" approach in [11], [18], [19], and [21], which study convergence to Bayes risk of the kth component risk rather than average risk. In particular, the "non-parametric" results of Johns [11] examine empirical Bayes procedures for unknown distributions (discrete and continuous case).

In a repetitive statistical game against an opponent, the sequential strategies $\mathbf{t}^*$ of this paper permit the statistician to take advantage of player I (Nature) if he does not play a minimax strategy, even when player II (statistician) does *not* know I's class of pure strategies (in this case the class $\mathcal{P}$). However, player I may "control" his average loss about $\phi(f, \xi') = v$, the value of the game, where ξ' is the maximin strategy for I, by choosing the θ_k's independently according to ξ'. Yet player II is protected in this case since $\mathbf{t}^*$ asymptotically "controls" in probability the average loss (Theorems 5.2, 5.3, and 7.1) about v also. Blackwell in [2] and [3] and Katz in [12] have considered this phenomenon of controlling play in general finite repetitive games.

Again from the game theory point of view Theorem 5.1 (or Corollary 7.1 for lower order) has its analog for more general games in Theorem 4 of [6]. However, in the case of statistical games this theorem would require that the densities $f = (f_1, \cdots, f_m)$ be completely specified. In order to obtain Theorem 4 in [6], Hannan introduces the idea of playing, at stage k, Bayes against a random perturbation of the $(k - 1)$st stage empirical distribution of player I's moves. This randomization is necessary since merely playing Bayes against $p_{k-1}(\boldsymbol{\theta})$ at stage k will not guarantee that the Nth average payoff risk) is asymptotically equivalent to the Nth stage empirical Bayes risk $\phi(p_N(\boldsymbol{\theta}))$. In our case, this randomization is furnished by the estimator sequence $\hat{\mathbf{f}}$.

9. Acknowledgment. The author wishes to thank the referee for pointing out changes which considerably improved an earlier version of the paper.

REFERENCES

[1] Blackwell, David (1956). An analogue of the minimax theorem for vector payoffs. *Pacific J. Math.* **6** 1–8.

[2] Blackwell, David (1954). Controlled random walks. *Proc. Internat. Congress Math.* Amsterdam **3** 336–338.

[3] Cacoullos, T. (1964). Estimation of a multivariate density. Technical Report No. 40, Department of Statistics, University of Minnesota.

[4] Driml, M. and Hans, O. (1960). On experience theory problems. *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes.* Czechoslovak Academy, Prague. 93–111.

[5] Halmos, Paul R. (1950). *Measure Theory.* Van Nostrand, Princeton.

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

[7] 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.

[8] Hannan, James F. and Robbins, Herbert (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37–51.

[9] Hannan, J. F. and Van Ryzin, J. R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743–1752.

[10] Hans, O. (1962). On conditional Bayes loss. *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes.* Czeckoslovak Academy, Prague. 207–221.

[11] Johns, M. V., Jr. (1957). Non-parametric empirical Bayes procedures. *Ann. Math. Statist.* **28** 649–669.

[12] Katz, Melvin (1960). Infinitely repeatable games. *Pacific J. Math.* **10** 879–885.

[13] Loève, Michel (1960). *Probability Theory* (2nd ed.). Van Nostrand, Princeton.

[14] Neyman, J. (1962). Two breakthroughs in the theory of statistical decision making. *Review of the International Statistical Institute.* **30** 11–27.

[15] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

[16] Parzen, E. (1954). On uniform convergence of families of sequences of random variables. *Univ. of California Publ. in Statist.* **2** 23–54. Univ. of California Press.

[17] Robbins, Herbert (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131–148. Univ. of California Press.

[18] ROBBINS, HERBERT (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 157–163. Univ. of California Press.

[19] ROBBINS, HERBERT (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[20] SAMUEL, E. (1963). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079–1094.

[21] SAMUEL, E. (1963). An empirical Bayes approach to the testing of certain parametric hypothesis. *Ann. Math. Statist.* **34** 1370–1385.

[22] SAMUEL, E. (1964). Convergence of the losses of certain decision rules for the sequential compound decision problem. *Ann. Math. Statist.* **35** 1606–1621.

[23] SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–438.

[24] SPACEK, A. (1956). An elementary experience problem. *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes.* Czechoslovak Acadeny, Prague. 253–258.

[25] VAN RYZIN, J. R. (1964). Asymptotic solutions to compound decision problems. Ph.D. Thesis, Dept. of Statistics, Michigan State University. Also appears as an Argonne National Laboratory Report, ANL-6820.

[26] VAN RYZIN, J. R. (1966). The compound decision problems with $m \times n$ finite loss matrix. *Ann. Math. Statist.* **37** 412–424.

[27] VAN RYZIN, J. (1966). The sequential compound decision problem with $m \times n$ finite loss matrix. *Ann. Math. Statist.* **37** 954–975.

[28] VAN RYZIN, J. (1965). Uniform convergence of losses in the fixed and sequential $m \times n$ compound decision problem. (Abstract). *Ann. Math. Statist.* **36** 1323–1324.

[29] WINKELBAUER, K. (1956). Experience in games of strategy and in statistical decision. *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes.* Czechoslovak Academy, Prague. 297–354