The Annals of Statistics
1979, Vol. 7, No. 4, 847–859

A REDUCTION THEOREM FOR CERTAIN SEQUENTIAL EXPERIMENTS. II[1]

BY L. LE CAM

University of California, Berkeley

The paper studies experiments which, for nonrandom stopping rules, resemble Koopman-Darmois families. It is shown that asymptotically one can limit oneself to sequential stopping rules which depend only on the terms entering in the Koopman-Darmois approximations, whether or not these terms are sums of independent variables. One can also obtain asymptotic results by studying similar problems on suitable processes with independent increments.

1. Introduction. This is an addendum, and a correction, to the paper [5] with the same title.

One main purpose of the paper in question was to show that for experiments which have suitable asymptotic resemblance with Koopman-Darmois families, asymptotic results on sequential decision problems may be obtained by studying similar problems on an appropriate process with independent increments.

Part of the reduction is to replace stopping times of fairly nondescript sequences of σ-fields by stopping times adapted to the statistics which occur in the Koopman-Darmois approximations. This was the object of Proposition 2 of [5].

Unfortunately, as pointed out by Dr. Albrecht Irle, the proof of Proposition 2 in [5] is incorrect unless the σ-fields involved satisfy a transitivity relation of the Bahadur type [1]. Here the transitivity becomes available only in the limit.

In the present paper we give an alternate argument showing first that reduction to the limit process is possible. This gives an essential improvement of Proposition 3 of [5] and allows us to deduce that Proposition 2 of [5] is valid after all. The method of proof corresponds to what was called in [5] "an embryo of a proof". It is rather lengthy.

Section 2 recalls the necessary notation and assumptions as well as some minor results of [5].

Section 3 elaborates several lemmas concerning independence of variables involved in limiting distributions or continuity properties for stopping times.

Section 4 gives the main results.

Some general features of the technique of proof used here are related to the methods used by P. Bickel and J. Yahav in [2]. However, the assumptions and results are not comparable.

Received June 1977; revised May 1978.

[1]This research was prepared with the support of National Science Foundation Grant No. MPS 75-10376 and U.S. Army Research Office Grant No. ARO-DA-AG 29-76-G-0167.

AMS 1970 *subject classification.* G2B15, 62L12.

Key words and phrases. Koopman-Darmois families, stopping times, processes with independent increments, approximations of experiments.

2. Notations and assumptions. The structure considered in [5] involves a set Θ and a function ξ from Θ to a finite dimensional vector space called H. There is also a particular point θ_0 for which one can assume that $\xi(\theta_0)$ is the origin of H.

In the present paper we shall dispense with the set Θ and work directly on the range $\Xi = \xi(\Theta)$ of ξ. This does not change anything essential and simplifies notation.

It will be assumed throughout that Ξ satisfies the following requirement:

ASSUMPTION A. The closure $\bar{\Xi}$ of Ξ contains an open neighborhood of the origin of H.

On the set Ξ there is a certain real valued function β. The pair (Ξ, β) will remain fixed throughout.

One considers also a sequence $\{\mathcal{S}_\nu: \nu = 1, 2, \cdots\}$ of systems $\mathcal{S}_\nu$ each composed of the following objects.

(1) A set $\mathcal{X}_\nu$, with an increasing sequence $\{\mathcal{A}_\nu(n), n = 0, 1, 2, \cdots\}$ of σ-fields. It is assumed that $\mathcal{A}_\nu(0)$ is always the trivial σ-field of $\mathcal{X}_\nu$.

(2) For each $\mathcal{A}_\nu(n)$ and each $\xi \in \Xi$, there is a probability measure $P_\nu[\xi, n]$ defined on $\mathcal{A}_\nu(n)$. These measures are coherent in the sense that if $m < n$ then $P_\nu[\xi, m]$ is the restriction of $P_\nu[\xi, n]$ to $\mathcal{A}_\nu(m)$.

(3) For each pair (n, ν) there is an $\mathcal{A}_\nu(n)$ measurable map $S_{n,\nu}$ of $\mathcal{X}_\nu$ into the dual H' of the finite dimensional space H.

(4) To each ν is assigned an integer n_ν in such a way that $n_\nu \to \infty$ if $\nu \to \infty$.

The coherence property required in (2) above means that, at the cost of performing a suitable completion operation on $\mathcal{X}_\nu$, one could regard the $P_\nu[\xi, n]$ as restrictions to $\mathcal{A}_\nu(n)$ of probability measures defined on the σ-field generated by the union of the $\mathcal{A}_\nu(n)$. We shall proceed to use language as if this was the case. However, the property in question is irrelevant in the present context.

Let $f_\nu(\xi, n)$ be the Radon-Nikodym density with respect to $P_\nu[0, n]$ of the part of $P_\nu[\xi, n]$ which is dominated by $P_\nu[0, n]$. This is defined on $\{\mathcal{X}_\nu, \mathcal{A}_\nu(n)\}$ and assumed to be $\mathcal{A}_\nu(n)$-measurable.

On the systems $\{\mathcal{S}_\nu: \nu = 1, 2, \cdots\}$ we shall impose an assumption as follows.

ASSUMPTION (K.D.). The sequence $\{\mathcal{S}_\nu\}$ satisfies the asymptotic Koopman-Darmois assumption for the function β and the maps $S_{n,\nu}$ in the sense that for every integer b and every $\xi \in \Xi$:

(i) the sequences $\{P_\nu[\xi, bn_\nu]\}$ and $\{P_\nu[0, bn_\nu]\}$ are contiguous, and,

(ii) the expressions

$$\sup_{n < bn_\nu} \left| f_\nu(\xi, n) - \exp\left\{\xi S_{n,\nu} - \frac{n}{n_\nu}\beta(\xi)\right\}\right|$$

tend to zero in probability.

For each system $\mathcal{S}_\nu$ we shall also consider integer valued stopping variables N_ν which are adapted to the sequence $\{\mathcal{A}_\nu(n), n = 0, 1, \cdots\}$ in the sense that the set

$A_{n,\nu} = \{N_\nu = n\}$ belong to $\mathcal{A}_\nu(n)$. Each such stopping variable defines a σ-field which will be denoted $\mathcal{A}_\nu(N_\nu)$. It is the σ-field defined on the sum $\cup_n A_{n,\nu}$ by taking on $A_{n,\nu}$ the trace of $\mathcal{A}_\nu(n)$.

The restriction to $\mathcal{A}_\nu(N_\nu)$ of the coherent sequence $\{P_\nu[\xi, n]; n = 0, 1, \cdots\}$ will be noted $P_\nu[\xi, N_\nu]$.

Note that the corresponding likelihood ratios may be written

$$\frac{dP_\nu[\xi, N_\nu]}{dP_\nu[0, N_\nu]} = \Sigma_n I_{A_{n,\nu}} f_\nu[\xi, n]$$
$$= f_\nu[\xi, N_\nu].$$

We shall consider only, and exclusively, sequences $\{N_\nu\}$ of stopping variables subject to the following requirement.

Assumption B. For each $\xi \in \Xi$ the sequence of distributions $\mathcal{L}(N_\nu/n_\nu|\xi)$ is relatively compact on $[0, \infty)$.

It will be convenient to introduce a variety of stochastic processes as follows. Let t denote a "time variable", $t \in [0, \infty)$. The process S_ν is defined by letting $S_\nu(t) = S_{n,\nu}$ whenever the integer part $[n_\nu t]$ of $n_\nu t$ is equal to n.

These processes are adapted to the σ-fields $\{\mathcal{B}_\nu(t)\}$ for $\mathcal{B}_\nu(t) = \mathcal{A}_\nu(n)$ if $[n_\nu t] = n$.

Note that the processes in question are right continuous with left limits. Also the σ-fields $\mathcal{B}_\nu(t)$ are right continuous.

In this framework the stopping variables N_ν become stopping times T_ν adapted to the $\{\mathcal{B}_\nu(t); t \in [0, \infty)\}$ if one writes $T_\nu = N_\nu/n_\nu$.

Conversely, if T_ν is a stopping time adapted to the increasing family $\{\mathcal{B}_\nu(t); t \in [0, \infty)\}$ then the variable $N_\nu = [n_\nu T_\nu]$ is a stopping variable for the initial family $\{\mathcal{A}_\nu(n); n = 0, 1, \cdots\}$.

The probability measures corresponding to this system will be denoted $Q_\nu(\xi, T_\nu)$ so that $Q_\nu[\xi, T_\nu] = P_\nu[\xi, N_\nu]$.

The experiments $\mathcal{E}_\nu(T_\nu) = \{Q_\nu(\xi, T_\nu); \xi \in \Xi\}$ will be called the experiments attached to T_ν.

It is shown in [5] that for systems $\mathcal{S}_\nu$ which satisfy both (A) and (K.D.) and for $\xi = 0$ the finite dimensional distributions of the processes S_ν converge to that of a process S which has independent increments and satisfies the relation

$$E \exp\{\xi S(t)\} = \exp\{t\beta(\xi)\}$$

for all $t \in [0, \infty)$ and all $\xi \in \Xi$.

We shall need further on a suitable version of this process, adapted to an increasing family of σ-fields $\{\mathcal{F}_t; t \in [0, \infty)\}$ as follows.

Let Ω be the set of functions from $[0, \infty)$ to H' which have left limits and which are right continuous. Let Ω_1 be the product $\Omega_1 = \Omega \times [0, 1]$ of Ω by the interval $[0, 1]$. The σ-field $\mathcal{F}_t$, on Ω_1, is generated by the evaluation maps $\{(\omega(s), \gamma); s \leq t\}$ with $(\omega(s), \gamma) \in H' \times [0, 1]$.

On Ω one takes for measure P_0' the distribution of the process S described above. On [0, 1] one places the Lebesgue measure λ and one takes on Ω_1 the direct product $P_0' \otimes \lambda$.

Besides the measure P_0 we shall also need measures $P_{\xi, t}$ defined on the σ-fields $\mathcal{F}_t$ and $\mathcal{F}_{t+} = \cap_s\{\mathcal{F}_s; s > t\}$. The measure $P_{\xi, t}$ has density

$$\exp\{\xi\omega(t) - t\beta(\xi)\}$$

with respect to P_0.

The main purpose of [5] could be described as follows. Consider (subject to (A), (K.D.) and (B)) sequential decision problems for the systems $\{\mathbb{S}_\nu\}$. Then, asymptotically, one may replace the σ-fields $\mathcal{B}_\nu(t)$ by the σ-fields $\mathcal{F}_{t+}$. Also, one may replace the families $Q_\nu(\xi, t)$ by the limit family $P_{\xi, t}$.

3. Basic lemmas concerning the system $\{\mathbb{S}_\nu\}$ *It will be assumed throughout, even if this is not explicitly mentioned, that the sequence* $\{\mathbb{S}_\nu\}$ *satisfies the assumptions* (A) *and* (K.D.).

It is also assumed that all the stopping variables under consideration satisfy Assumption B.

Let τ_ν be a stopping time adapted to the σ-fields $\{\mathcal{B}_\nu(t); t \in [0, \infty)\}$. Let X_ν be the random variable $X_\nu = S_\nu(\tau_\nu)$. Take a number $\alpha > 0$ and let $\tau_\nu' = \tau_\nu + \alpha$. Let Z_ν be the difference $Z_\nu = S_\nu(\tau_\nu') - S_\nu(\tau_\nu)$.

Finally let Y_ν be a $\mathcal{B}_\nu(\tau_\nu)$ measurable map from $\mathcal{X}_\nu$ to a certain fixed Euclidean space.

LEMMA 1. *Assume that for the measures corresponding to* $\xi = 0$ *the joint distributions* $\mathcal{L}[\tau_\nu, X_\nu, Y_\nu, Z_\nu|0]$ *converge to a limit* $L_0 = \mathcal{L}[t, X, Y, Z|0]$. *Then, for each* $\xi \in \Xi$, *the distributions* $\mathcal{L}[\tau_\nu, X_\nu, Y_\nu, Z_\nu|\xi]$ *converge to a limit* L_ξ *which has, for density with respect to* L_0, *the expression* $\exp\{\xi(x + z) - (t + \alpha)\beta(\xi)\}$.

Furthermore, the experiment $\{L_\xi; \xi \in \Xi\}$ *so obtained is the weak limit of the experiments* $\mathcal{E}_\nu(\tau_\nu')$.

PROOF. This is essentially the same statement as that of Lemma 2 of [5]. According to the Koopman-Darmois assumption and the continuity assertion of Lemma 1 of [5], the differences

$$\frac{dQ_\nu(\xi, \tau_\nu')}{dQ_\nu(0, \tau_\nu')} - \exp\{\xi S_\nu(\tau_\nu') - \tau_\nu'\beta(\xi)\}$$

tend to zero in probability as $\nu \to \infty$.

Fixing a value of ξ, the usual contiguity argument shows that the joint limiting distribution of the quadruplet $(\tau_\nu', X_\nu, Y_\nu, Z_\nu)$ has, with respect to L_0, a density $\exp\{\xi s - t'\beta(\xi)\}$ with $s = x + z$ and $t' = t + \alpha$. The first assertion follows.

The assertion concerning the limit of the experiments $\mathcal{E}_\nu(\tau_\nu')$ is then a consequence of the fact that the pairs $[S_\nu(\tau_\nu'), \tau_\nu']$ and, therefore, also the pairs $[S_\nu(\tau_\nu'), \tau_\nu]$, are distinguished sequences of asymptotically sufficient statistics for the sequence $\{\mathcal{E}_\nu(\tau_\nu'); \nu = 1, 2, \cdots\}$. (See [3], page 50).

REMARK. The preceding lemma shows in particular that the limit of the experiments $\mathcal{E}_\nu(\tau'_\nu)$ is entirely determined by the limit of the joint distributions $\mathcal{L}[\tau'_\nu, S_\nu(\tau'_\nu)|0]$. This will be used repeatedly in the sequel. Note however that if $\mathcal{E}_\nu(\tau_\nu)$ tends to a limit $\mathcal{E}$, it does not necessarily follow that $\mathcal{L}[\tau_\nu, X_\nu|0]$ tends to a limit, since the Laplace transform involved in Lemma 1 may well fail to determine the underlying measures.

The following proposition will play a very essential role in all our arguments.

PROPOSITION 1. *With the same assumptions and notations as in Lemma* 1, *the limiting distributions L_ξ of the quadruplets (T, X, Y, Z) are such that Z is independent of (T, X, Y). It has a Laplace transform*

$$E_0 \exp\{\xi Z\} = \exp\{\alpha\beta(\xi)\}$$

for all ξ interior to the closure of Ξ.

PROOF. Let g be an arbitrary bounded measurable function defined on the space of triplets (t, x, y). Write the distribution L_0 in the form

$$L_0(dt, dx, dy, dz) = M(dt, dx, dy)K(dz|t, x, y)$$

for a marginal distribution M and a conditional distribution K. Consider the expectation $E_\xi g(T, X, Y)$ for the distribution L_ξ. This may be written

$$\int g(t, x, y) \exp\{\xi(x + z) - (t + \alpha)\beta(\xi)\}\, dKdM$$
$$= \int g(t, x, y) \exp\{\xi x - t\beta(\xi)\}\phi(\xi, t, x, y)\, dM,$$

with

$$\phi(\xi, t, x, y) = \int \exp\{\xi z - \alpha\beta(\xi)\}K(dz|t, x, y).$$

However, Lemma 1 can also be applied to the limiting distribution of the triplets (τ_ν, X_ν, Y_ν) themselves, yielding

$$E_\xi g(T, X, Y) = \int g(t, x, y) \exp\{\xi x - t\beta(\xi)\}M(dt, dx, dy).$$

Since g can be selected arbitrarily, it follows that, for each ξ and almost all triplets (t, x, y), one must have $\phi(\xi, t, x, y) = 1$.

Applying this to a dense countable set of values of ξ one concludes that for almost all (t, x, y) the Laplace transform $\int \exp\{\xi z\}K[dz|t, x, y]$ equals $\exp\{\alpha\beta(\xi)\}$ for all ξ in an open neighborhood of the origin. This concludes the proof of the proposition.

Instead of one number $\alpha > 0$ one could consider a finite sequence, say $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_m$ and a vector $\vec{Z}_\nu = \{Z_\nu^i, i = 1, 2, \cdots, m\}$ with $Z_\nu^i = S_\nu(\tau_\nu + \alpha_i) - S_\nu(\tau_\nu + \alpha_{i-1})$.

This gives the following corollary.

COROLLARY. *In the joint limiting distribution*

$$\mathcal{L}[T, X, Y, \vec{Z}|0] = \lim_\nu \mathcal{L}[\tau_\nu, X_\nu, Y_\nu, \vec{Z}_\nu|0]$$

the triplet (T, X, Y) is independent of $\vec{Z}$.

PROOF. First consider the pairs $\{(\tau_\nu, X_\nu, Y_\nu, Z_\nu^i, i = 1, \cdots, m-1), (Z_\nu^m)\}$. The first vector is measurable for the σ-field $\mathfrak{B}_\nu(\tau_\nu + \alpha_{m-1})$. It may take the place of Y_ν in an application of Proposition 1. This gives the independence of $(T, X, Y, Z^i, i \leqslant m-1)$ and Z^m. Then we can proceed downward, with $(T, X, Y, Z^i, i \leqslant m-2)$ and Z^{m-1} and so forth. The result follows. To state the next result we shall introduce a definition as follows.

Let $\{T_\nu\}$ and $\{T_\nu^*\}$ be two sequences of stopping times, with both T_ν and T_ν^* adapted to the increasing family $\mathfrak{B}_\nu(t)$; $t \in [0, \infty)$.

DEFINITION 1. The sequences $\{T_\nu\}$ and $\{T_\nu^*\}$ will be called asymptotically equivalent if

(i) for any weak cluster point $(\mathcal{E}, \mathcal{E}^*)$ of the sequence of pairs of experiments $(\mathcal{E}_\nu(T_\nu), \mathcal{E}_\nu(T_\nu^*))$ the two experiments $\mathcal{E}$ and $\mathcal{E}^*$ are equivalent;

(ii) for every $\xi \in \Xi$, the differences $\mathcal{L}[T_\nu|\xi] - \mathcal{L}[T_\nu^*|\xi]$ tend to zero, for the dual Lipschitz norm of measures on $[0, \infty)$.

This says essentially that T_ν and T_ν^* yield about the same information and that their distributions differ little.

PROPOSITION 2. *Let T_ν and T_ν^* be stopping times of the family $\mathfrak{B}_\nu(t)$; $t \in [0, \infty)$. Assume that the conditions* (A), (K.D.) *and* (B) *are all satisfied. Assume also that, as $\nu \to \infty$, the difference $T_\nu - T_\nu^*$ tends to zero in probability. Then*

(i) *$\{T_\nu\}$ and $\{T_\nu^*\}$ are asymptotically equivalent; and*

(ii) *the difference $S_\nu(T_\nu^*) - S_\nu(T_\nu)$ tends to zero in probability.*

PROOF. Let F be a finite subset of Ξ which contains the origin and a basis of the linear space H spanned by Ξ. Take a number $\alpha > 0$ and consider the stopping times $T_\nu^1 = T_\nu \wedge T_\nu^*$, $T_\nu^2 = T_\nu \vee T_\nu^*$, and $T_\nu^3 = T_\nu^1 + \alpha$.

Any one of these stopping times yields an experiment $\mathcal{E}_\nu(T_\nu^i)$. We shall consider that the experiments in question are indexed by the set F only.

For the informational ordering of experiments one may write $\mathcal{E}_\nu(T_\nu^1) \leqslant \mathcal{E}_\nu(T_\nu) \leqslant \mathcal{E}_\nu(T_\nu^2)$, and similarly for T_ν^*.

Let $(\mathcal{E}^1, \mathcal{E}^2, \mathcal{E}^3)$ be a cluster point of the sequence of triplets $(\mathcal{E}_\nu(T_\nu^1), \mathcal{E}_\nu(T_\nu^2), \mathcal{E}_\nu(T_\nu^3))$. It is clear that $\mathcal{E}^1 \leqslant \mathcal{E}^2$, but one has also $\mathcal{E}^2 \leqslant \mathcal{E}^3$. To show this consider the set B_ν where $T_\nu^2 \leqslant T_\nu^3$. This set belongs to both σ-fields $\mathfrak{B}_\nu(T_\nu^2)$ and $\mathfrak{B}_\nu(T_\nu^3)$. By assumption the probability of B_ν tends to unity for $\xi = 0$ and therefore also for all $\xi \in \Xi$ by contiguity. On B_ν the measures $Q_\nu(\xi, T_\nu^2)$ and $Q_\nu(\xi, T_\nu^3)$ coincide. Thus the deficiency of $\mathcal{E}_\nu(T_\nu^3)$ with respect to $\mathcal{E}_\nu(T_\nu^2)$ does not exceed $\sup_{\xi \in F}\{1 - Q_\nu[\xi, T_\nu^2; B_\nu]\}$. This gives the desired inequality $\mathcal{E}^2 \leqslant \mathcal{E}^3$.

Now we shall show that for each given $\varepsilon > 0$ it is possible to select α so small that the deficiency $\delta(\mathcal{E}^1, \mathcal{E}^3)$ does not exceed ε.

For this purpose let $Z_\nu = S_\nu(T_\nu^3) - S_\nu(T_\nu^1)$ and let L_0 be a cluster point of the sequence $\mathcal{L}\{T_\nu^1, S_\nu(T_\nu^1), Z_\nu|0\}$ corresponding to the cluster point $(\mathcal{E}^1, \mathcal{E}^2, \mathcal{E}^3)$. Then, according to Lemma 1, the experiment $\mathcal{E}^3$ is representable by the measures

L_ξ, $\xi \in F$ such that L_ξ has density

$$g^3(t, x, z) = \exp\{\xi(x + z) - (t + \alpha)\beta(\xi)\}$$

with respect to L_0. The experiment $\mathcal{E}^1$ can be represented similarly, but with densities

$$g^1(t, x, z) = \exp\{\xi x - t\beta(\xi)\}$$

with respect to the same L_0.

The Hellinger affinity between g^1 and g^3 is equal to $\exp\{\alpha[\beta(\xi/2) - \frac{1}{2}\beta(\xi)]\}$. This can obviously be made as close to one as desired by taking α sufficiently close to zero. Using the inequalities between Hellinger distances and L_1-norms (see for instance [3]), one concludes that statement (i) of the proposition holds.

To show that (ii) holds, we shall show that $S_\nu(T_\nu) - S_\nu(T_\nu^1)$ tends to zero. Since the same argument will apply to T_ν^* also, the result will follow. For this purpose, take two elements, 0 and ξ of the finite set F and let $\mu_\nu = Q_\nu(0, T_\nu) + Q_\nu(\xi, T_\nu)$ on the σ-field $\mathcal{B}_\nu(T_\nu) \supset \mathcal{B}_\nu(T_\nu^1)$. Then $Q_\nu(\xi, T_\nu)$ has, with respect to μ_ν, a Radon-Nikodym density V_ν on $\mathcal{B}_\nu(T_\nu)$ and a density V_ν^1 on $\mathcal{B}_\nu(T_\nu^1)$.

Let δ_ν be the deficiency $\delta[\mathcal{E}_\nu(T_\nu^1), \mathcal{E}_\nu(T_\nu)]$. According to the inequalities of [4] one may write $\int |V_\nu^1 - V_\nu|^2 d\mu_\nu \leq 2\delta_\nu$. Thus, by part (i) these integrals tend to zero. According to the contiguity assumption included in (K.D.) this implies that $\log(V_\nu/1 - V_\nu) - \log(V_\nu^1/1 - V_\nu^1)$ also tends to zero in probability. However, by assumption, this latter difference is asymptotically equivalent to

$$\xi[S_\nu(T_\nu) - S_\nu(T_\nu^1)] - (T_\nu - T_\nu^1)\beta(\xi).$$

One concludes that $\xi[S_\nu(T_\nu) - S_\nu(T_\nu^1)]$ tends to zero in probability for all $\xi \in F$. Since F contains a basis of H, this concludes the proof.

REMARK. An alternate proof can be carried out without having recourse to the deficiencies and to the inequalities of [4] by using martingale inequalities on the likelihood ratios.

In the next section we shall use a consequence of Proposition 2 applicable to dyadic approximations of the times T_ν. These are defined as follows.

For each integer $m > 1$, the mth dyadic approximation $T_{m,\nu}$ of T_ν is defined by first letting $T_{m,\nu}^* = k2^{-m}$ whenever $(k - 1)2^{-m} < T_\nu \leq k2^{-m}$ and then letting $T_{m,\nu} = (2^m) \wedge T_{m,\nu}^*$.

COROLLARY. *Let $\{T_\nu\}$ be a sequence of stopping times such that $\mathfrak{L}[T_\nu, S_\nu(T_\nu)|0]$ converges to a limit on $[0, \infty) \times H'$. Then, for every $\xi \in \Xi$ one has*

$$\lim_{m,\nu} E_\xi\{1 \wedge \|S_\nu(T_{m,\nu}) - S_\nu(T_\nu)\|\} = 0$$

if m and ν both tend to infinity in any arbitrary manner.

In particular, for every $\xi \in \Xi$ the limit of $\mathfrak{L}[T_\nu, S_\nu(T_\nu)|\xi]$ is equal to the iterated limit

$$\lim_m \lim_\nu \mathfrak{L}[T_{m,\nu}, S_\nu(T_{m,\nu})|\xi].$$

NOTE. In this iterated limit, the first limit over ν may not exist. What the statement is supposed to mean is that for any arbitrary selected cluster point $M_m(\xi)$ of $\mathcal{L}[T_{m,\nu}, S_\nu(T_{m,\nu})|\theta]$ the limit $\lim_m M_m(\xi)$ is $\lim_\nu \mathcal{L}[T_\nu, S_\nu(T_\nu)|\xi]$.

PROOF. It is sufficient to show that for every ε one can find a number $r(\varepsilon)$ so large that $\nu \geqslant r(\varepsilon)$ and $m \geqslant r(\varepsilon)$ will imply $E_0\{1 \wedge \|S_\nu(T_\nu) - S_\nu(T_{m,\nu})\|\} < \varepsilon$.

Assuming the contrary, one could readily construct sequences which would violate Proposition 2. Hence the assertion.

4. Reduction theorems. In the present section we shall first show that the limits of the sequences of experiments $\{\mathcal{E}_\nu(T_\nu)\}$ can be realized on the process with independent increments described at the end of Section 2. Then it will be argued that, asymptotically, it is enough to consider stopping times which are adapted to the increasing family of σ-fields $\{\mathcal{F}_{t+}, t \in [0, \infty)\}$ instead of the possibly more general $\mathcal{B}_\nu(t)$.

The proof of the possibility of reduction to the process with independent increments has been separated into two parts for greater clarity. It should be clear that the main part of the argument is the first part referring only to dyadic times.

Recall that the space Ω_1 is a product $\Omega_1 = \Omega \times [0, 1]$ where Ω is the space of functions from $[0, \infty)$ to H' which have left limits and are right continuous. For $\xi = 0$, the measure P_0 used on Ω_1 is the product $P_0' \otimes \lambda$ where λ is the Lebesgue measure and where P_0' is the distribution of a process ω with independent increments such that $E_0 \exp\{\xi\omega(t)\} = \exp\{t\beta(\xi)\}$.

Let $D_m = \{t_k;\ k = 0, 1, 2, \cdots, 2^{2m}\}$, with $t_k = k2^{-m}$. Then the set $\mathcal{W}(m)$ of functions from D_m to H' is a factor of Ω. The measures induced on $\mathcal{W}_1(m) = \mathcal{W}(m) \times [0, 1]$ by P_0 or P_ξ will still be noted P_0 and P_ξ, by abuse of notation.

In the following statement the integer m will be kept fixed and omitted from the notation.

PROPOSITION 3. *Let the assumptions* (A) *and* (K.D.) *be satisfied and let* T_ν *be a stopping time adapted to the* σ*-fields* $\mathcal{B}_\nu(t)$, $t \in [0, \infty)$.

Assume that T_ν *takes its values in* D *and that* $\mathcal{L}\{T_\nu, S_\nu(T_\nu)|0\}$ *converges to a limit. Then, there is on* $\mathcal{W}_1$ *a stopping time* τ *adapted to the* σ*-fields* $\mathcal{F}_t$, $t \in D$ *such that*

(i) *the distribution* L_0 *of* $[\tau, \omega(\tau)]$ *for* P_0 *is the limit of the distributions* $\mathcal{L}[T_\nu, S_\nu(T_\nu)|0]$;

(ii) *the experiment given by the measures*

$$dL_\xi = \exp\{\xi\omega(\tau) - \tau\beta(\xi)\}\, dL_0$$

is the limit of the experiments $\mathcal{E}_\nu(T_\nu)$.

PROOF. Let $\mathcal{V}$ be the space of functions from D to the product $\{0, 1\} \times H'$ of the two point set $\{0, 1\}$ by the finite dimensional space H'. Let $J_\nu(k)$ be the indicator of the event $T_\nu \leqslant t_k$, for $t_k \in D$ and let $Z_{\nu,k}$ be the difference $Z_{\nu,k} = S_\nu(t_k) - S_\nu(t_{k-1})$, $k = 1, 2, \cdots, 2^{2m}$. Put $Z_{\nu,0} = 0$. Then the vector $\{J_\nu(k), Z_{\nu,k};\ k = 0, 1, \cdots, 2^{2m}\}$ is a random element which takes values in the space $\mathcal{V}$. The corresponding joint distributions, for $Q_\nu[0, 2^{2m}]$, have at least one cluster point, say M.

To this M one may apply Proposition 1 repeatedly (as indicated in the remark which follows Proposition 1) using instead of the stopping time τ_ν of that proposition, times which are deterministic, with values t_j, $t_j \in D$. The vectors Y_ν which occur in Proposition 1 will then be taken equal to the vectors $Y_\nu = \{J_\nu(t_k), Z_{\nu,k}; k \leqslant j\}$ and the difference called Z_ν previously will be $Z_{\nu,j+1} = S_\nu(t_{j+1}) - S_\nu(t_j)$, in the present case.

Let $V = \{(U(t), Z(t)); t \in D\}$ be a random element which has distribution M on $\mathcal{V}$. According to Proposition 1, for each $j = 0, 1, \cdots, 2^{2m}$ the sets of variables

$$\{U(s), Z(s); s \leqslant t_j\} \quad \text{and} \quad \{Z(t); t \geqslant t_{j+1}\}$$

are independent.

Let $\mathcal{G}_t$ be the σ-field generated on $\mathcal{V}$ by the coordinate evaluations $\{z(s); s \leqslant t\}$. For each $t \in D$, let ϕ_t be the conditional expectation $E_0[U(t)|\mathcal{G}]$ where $\mathcal{G}$ is the σ-field $\mathcal{G} = \cup_{t \in D}\mathcal{G}_t$.

We claim that one can select versions of the ϕ_t so that

(i) ϕ_t is $\mathcal{G}_t$ measurable;

(ii) $0 \leqslant \phi_0 \leqslant \phi_s \leqslant \phi_t \leqslant \phi_b = 1$ (with $b = t_{2^{2m}}$ for simplicity).

To prove this, let f be any bounded $\mathcal{G}_t$ measurable function and let g be any bounded function which is measurable with respect to the σ-field $\mathcal{G}_t'$ generated by $\{Z(s); s > t, s \in D\}$.

According to Proposition 1, the pair $(f, U(t))$ is independent of g. Thus ϕ_t must satisfy the relation

$$E_0 fU(t)g = E_0[fU(t)]E_0(g) = E_0[f\phi_t g].$$

However, if $\psi_t = E_0[fU(t)|\mathcal{G}_t]$ one has also $E_0 fU(t) = E_0 f\psi_t$ and $E_0 f\psi_t g = E_0 f\psi_t E_0 g$. Since this is true for all pairs (f, g) it follows that $\phi_t = \psi_t$ almost everywhere. Thus one may select for ϕ_t a $\mathcal{G}_t$ measurable function such that $0 \leqslant \phi_t \leqslant 1$. By construction $J_\nu(s) \leqslant J_\nu(t)$ if $s \leqslant t$. It follows that $U(s) \leqslant U(t)$ almost surely and therefore that $\phi_s \leqslant \phi_t$ almost surely.

One can also let $\phi_b = 1$ for the last element b of D. Then, replacing ϕ_t by the equivalent $\bar{\phi}_t = \sup\{\phi_s; s \leqslant t\}$ gives the desired functions.

Now let $\mathcal{W}$ be the space of H' valued functions defined on D. Map the second components called z of the vectors $v \in \mathcal{V}$ into $\mathcal{W}$ by the operation $w(t) = \Sigma_{s<t} z(s)$. The selected cluster point M induces in this manner a certain distribution on $\mathcal{W}$. It is no other than the distribution P_0' induced by our process with independent increments.

Let $\mathcal{W}_1 = \mathcal{W} \times [0, 1]$ be given the product measure $P_0 = P_0' \otimes \lambda$. For each pair $(w, \gamma) \in \mathcal{W}_1$ define a stopping time τ as the first $t \in D$ for which $\phi_t(w) \geqslant \gamma$. Similarly define T on $\mathcal{V}$ by $T \leqslant t$ if $U(t) = 1$.

We claim that the distributions $\mathcal{L}[T, w(T)]$ and $\mathcal{L}[\tau, w(\tau)]$ are the same. To see this consider a particular $t \in D$ and a bounded measurable function f defined on

$D \times H'$. Assume that $f(u, x) = 0$ unless $u = t$ and consider the expectations $E_0 f[T, w(T)]$. Let s be the element of D which immediately precedes t. By assumption

$$E_0 f[T, \mathcal{W}(T)] = E_0[U(t) - U(s)]f[t, w(t)]$$
$$= E_0(\phi_t - \phi_s)f[t, w(t)].$$

However, for the stopping time τ the conditional probability $P_0[\tau \leqslant t|w]$ is equal to ϕ_t. Thus $E_0(\phi_t - \phi_s)f[t, w(t)]$ is also equal to $E_0 f[\tau, w(\tau)]$.

The identity of distributions follows. The second statement is a consequence of Lemma 1, hence the result as claimed.

This leads to the following result.

THEOREM 1. *Let the assumptions* (A) *and* (K.D.) *be satisfied by the systems* $\mathcal{S}_\nu$ *and let* T_ν *be a sequence of stopping times which satisfies condition B. The time* T_ν *is assumed to be adapted to the family* $\mathcal{B}_\nu(t)$, $t \in [0, \infty)$.

Assume, in addition, that the sequence of distributions $\mathcal{L}\{T_\nu, S_\nu(T_\nu)|0\}$ *converges to a limit.*

Then there is on Ω_1 *a stopping time* τ *adapted to the family* $\mathcal{F}_{t+}$; $t \in [0, \infty)$ *such that*

(i) *the distribution* L_ξ *of* $(\tau, \omega(\tau))$ *for* P_ξ *is the limit of the distributions* $\mathcal{L}[T_\nu, S_\nu(T_\nu)|\xi]$

(ii) *the experiment given by the measures*

$$dL_\xi = \exp\{\xi\omega(\tau) - \tau\beta(\xi)\}dL_0$$

is the limit of the experiments $\mathcal{E}_\nu(T_\nu)$.

PROOF. Consider the dyadic sets $D_m = \{t_k;\ k = 0, 1, 2, \cdots, 2^{2m}\}$, $t_k = k2^{-m}$ and form the corresponding spaces $\mathcal{V}_m$, $\mathcal{W}(m)$ and $\mathcal{W}_1(m)$. Let $T_{m,\nu}$ be the mth dyadic approximation to T_ν. This has values in D_m. Proceeding as in the proof of Proposition 3, the $(T_{m,\nu}, S_\nu(T_{m,\nu}))$ induce a certain distribution $M_{m,\nu}$ on $\mathcal{V}_m$. Taking an appropriate subsequence if necessary, one may assume that, for all integers m, $M_{m,\nu}$ have limits on their space $\mathcal{V}_m$.

Thus the construction used in Proposition 3 with σ-fields $\mathcal{G}_t$ becomes available. It can also be carried out directly on Ω_1 itself, using the σ-fields $\mathcal{F}_t$ instead of the $\mathcal{G}_t$.

This yields certain stopping times τ_m.

The successive τ_m are compatible. In fact the possible values of τ_m are the subset $\{k2^{-(m+1)}\}$ of D_{m+1}, where k is even and $k \leqslant 2^{2m+1}$. If τ_{m+1} takes a value $(2n-1)2^{-(m+1)}$ or $(2n)2^{-(m+1)}$ with $2n \leqslant 2^{2m+1}$ then τ_m must almost surely take the value $2n2^{-(m+1)}$.

It follows that for every integer b, as soon as $b < 2^{2m}$, the $\tau_m \wedge b$ form a sequence which decreases almost surely to a limit $\tau \wedge b$ with $\tau = \lim_m \tau_m$.

This stopping time τ is a stopping time of the σ-fields $\mathcal{F}_{t+}$; $t \in [0, \infty)$. The decreasing property implies that $\mathcal{L}[\tau_m, \omega(\tau_m)]$ converges to $\mathcal{L}[\tau, \omega(\tau)]$, since the ω are right continuous.

However, by Proposition 3 and the corollary of Proposition 2, we have

$$\mathcal{L}[\tau, \omega(\tau)] = \lim_m \lim_\nu \mathcal{L}[T_{m,\nu}, S_\nu(T_{m,\nu})]$$
$$= \lim_\nu \mathcal{L}[T_\nu, S_\nu, (T_\nu)].$$

This concludes the proof of statement (i) of the theorem. The second statement is then immediate and the result is completely proved.

Theorem 1 effectively reduces the study of the possible limits of the experiments $\mathcal{E}_\nu(T_\nu)$ and associated distributions $\mathcal{L}[T_\nu, S_\nu(T_\nu)|\xi]$ to the study of analogous objects defined on the limit process with values in Ω_1.

However, this need not be the most convenient way to proceed. For many purposes it is more satisfactory to note first that one can replace the general σ-fields $\mathfrak{B}_\nu(t)$ by the σ-fields $\mathfrak{F}_t$. Because of this we shall now restate Proposition 2 of [5] and elaborate a proof of it.

Take pairs (S_ν, γ) which consist of the process S_ν itself and of a variable γ uniformly distributed on [0, 1]. These can be considered random elements in the space Ω_1. They are adapted to the σ-fields $\mathfrak{F}_t$; $t \in [0, \infty)$ themselves.

THEOREM 2. *Let the assumptions* (A) *and* (K.D.) *be satisfied and let T_ν be a sequence of stopping times such that T_ν is adapted to the family $\mathfrak{B}_\nu(t)$; $t \in [0, \infty)$ and satisfies assumption* (B).

Then there are stopping times τ_ν defined on Ω_1 and adapted to the σ-fields $\mathfrak{F}_t$; $t \in [0, \infty)$ such that as $\nu \to \infty$ the Prohorov distance between $\mathcal{L}[T_\nu, S_\nu(T_\nu)|\xi]$ and $\mathcal{L}[\tau_\nu, S_\nu(\tau_\nu)|\xi]$ tends to zero for all $\xi \in \Xi$.

PROOF. Let us first show that it is sufficient to prove the result for the dyadic approximations $T_{m,\nu}$ instead of the T_ν themselves. Indeed, suppose constructed times $\tau_{m,\nu}$ such that, for all m, the Prohorov distance $\pi_\nu(m)$ between $\mathcal{L}[T_{m,\nu}, S_\nu(T_{m,\nu})|0]$ and $\mathcal{L}[\tau_{m,\nu}, S_\nu(\tau_{m,\nu})|0]$ tends to zero. Then one can find an increasing sequence $\{m_\nu\}$ such that $\pi_\nu(m_\nu)$ still tends to zero. The corresponding distances $\pi_\nu(m_\nu, \xi)$ computed for distributions induced for the parameter ξ will also tend to zero according to the usual contiguity arguments. The result for the T_ν follows then from the corollary of Proposition 2.

Thus it is sufficient to proceed assuming that T_ν itself takes values in the dyadic set $D = \{t_k; k = 0, 1, \cdots, 2^{2m}\}$, $t_k = k2^{-m}$.

Introduce the spaces $\mathfrak{W}$ and $\mathfrak{W}_1$ used in the proof of Proposition 3. Let $\overline{\mathfrak{W}}$ be the space of functions from D to the product $\{0, 1\} \times H'$ of the two point set $\{0, 1\}$ by the space H' and let $\overline{\mathfrak{W}}_1 = \overline{\mathfrak{W}} \times [0, 1]$.

The elements of $\overline{\mathfrak{W}}$ are vectors $\{(u(s), w(s)); s \in D\}$. Let $\mathfrak{F}_t$ be the σ-field generated on $\overline{\mathfrak{W}}_1$ by the evaluations $(w(s), \gamma)$, $s \leq t$, $\gamma \in [0, 1]$ and let $\overline{\mathfrak{F}}_t$ be the σ-field generated by the evaluations $(u(s), w(s))$, $s \leq t$.

Then T_ν may be redefined as a stopping time adapted to the family $\overline{\mathfrak{F}}_t$; $t \in D$, while the τ_ν of interest are adapted to $\mathfrak{F}_t$; $t \in D$.

Select such a τ_ν so that the Prohorov distance, say a_ν, between $\mathcal{L}[\tau_\nu, S_\nu(\tau_\nu)|0]$ and $\mathcal{L}[T_\nu, S_\nu(T_\nu)|0]$ does not exceed the infimum of all such possible distances by more than $2^{-\nu}$. Let M_ν be the distribution induced by T_ν on $\overline{\mathcal{W}}_1$.

Suppose that a_ν does not tend to zero. Then there is an $\varepsilon_0 > 0$ and a subsequence, say $\nu(r)$, such that $a_{\nu(r)} \geqslant \varepsilon_0$ for all r. One may assume that the subsequence has been chosen so that along it the distributions $\mathcal{L}[\tau_\nu, S_\nu(\tau_\nu)|0]$, $\mathcal{L}[T_\nu, S_\nu(T_\nu)|0]$ and M_ν will converge. If M is the limit of the M_ν, Proposition 3 shows that there is on $\overline{\mathcal{W}}_1$ a stopping time of the σ-fields $\mathcal{F}_t$; $t \in D$ such that $\mathcal{L}[\tau, S(\tau)]$ is the limit of the $\mathcal{L}[T_\nu, S_\nu(T_\nu)|0]$.

Let A_t be the subset of $\overline{\mathcal{W}}_1$ where $\tau = t$. Consider it as a subset of the product $\mathcal{W}_t \times [0, 1]$ where $\mathcal{W}_t$ is a product of $H_s' \equiv H'$, for $s \leqslant t$.

Then, for each $\varepsilon \in (0, \varepsilon_0)$, there is a compact set K_t and an open set G_t, such that in $\mathcal{W}_t \times [0, 1]$ one has $K_t \subset A_t \subset G_t$ and $M(G_t - K_t) < \varepsilon 2^{-4m}$.

For each $\alpha > 0$ let K_t^α be the set of points in $\mathcal{W}_t \times [0, 1]$ whose distance to K_t is at most α. Then, for α small enough, $K_t^\alpha \subset G_t$. Also, at least one of these $K_t^\alpha \subset G_t$ must have a boundary which has M measure zero. Select such a set K_t^α and call it B_t. Let $\overline{B}_t = \cup_s\{B_s;\ s \leqslant t\}$ and let τ' be the stopping time such that $\tau' \leqslant t$ if (w, γ) belongs to $\overline{B}_t$.

The sets $\overline{B}_t$ can also be used in the same way to define stopping times, say τ_ν', for the processes $\{S_\nu(t);\ t \in D\}$.

Since all the B_s have negligible boundaries for M, so does $\overline{B}_t$. Thus $M_\nu(\overline{B}_t) \to M(\overline{B}_t)$ for all $t \in D$. In particular $\mathcal{L}[\tau_\nu', S_\nu(\tau_\nu')|0]$ converges to $\mathcal{L}[\tau', w(\tau')|0]$. However, $\mathcal{L}[\tau, w(\tau)|0]$ and $\mathcal{L}[\tau', w(\tau')|0]$ differ by at most $\varepsilon < \varepsilon_0$.

This yields a contradiction to the assertion that $a_\nu \geqslant \varepsilon_0$. The desired result follows.

5. Concluding remarks. Our previous paper [5] described how the foregoing theorems could be applied to obtain lower bounds on asymptotic variances of estimates or upper bounds on the power of sequential tests.

Because at that time we did not have available the equivalent of Theorem 1, the descriptions had to be limited to cases where the sequences of stopping times T_ν were "regular" as explained there. Theorem 1 says that in fact all sequences (subject to (B)) are regular.

As mentioned earlier, the technique of passing through dyadic times is reminiscent of the "block " technique used by Bickel and Yahav in [2]. However, here it is mainly used to allow application of the transitivity conditions on the limit experiments. For the problems considered in [2], transitivity is available all the way along.

Of course, under the assumptions made here we cannot conclude (as in [2]) that the Bayes risks for the systems $\mathcal{S}_\nu$ will converge to the Bayes risks on the limiting processes. (One can assert that the limit of the Bayes risk is not smaller than the Bayes risk computed on the limit process.) This is partly because we have not controlled in any way the expectations $E(T_\nu)$ but only the distributions $\mathcal{L}(T_\nu)$, and partly because we have dealt only with pointwise convergence on Ξ. Convergences which are uniform on compacts could be induced at the cost of assuming that the distributions $P_\nu[\xi, bn_\nu]$ satisfy additional conditions such as the condition that if $\xi_\nu \to \xi$ then $\|P_\nu[\xi_\nu, bn_\nu] - P_\nu[\xi, bn_\nu]\|$ tends to zero.

REFERENCES

[1] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

[2] BICKEL, P. J. and YAHAV, J. A. (1972). On the Wiener process approximation to Bayesian sequential testing problems. *Sixth Berkeley Symp. Math. Statist. and Probability*. 57–83. Univ. California Press, Berkeley.

[3] LE CAM, L. (1974a). *Notes on Asymptotic Methods in Statistical Decision Theory*. Centre de Recherches Mathématiques, Univ. Montréal.

[4] LE CAM, L. (1974b). On the information contained in additional observations. *Ann. Statist.* **2** 630–649.

[5] LE CAM, L. (1977). A reduction theorem for certain sequential experiments. In *Statistical Decision Theory and Related Topics*. Vol. II. (S. Gupta and D. Moore, eds.) 223–244. Academic Press, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720