The Annals of Statistics
1996, Vol. 24, No. 1, 436–448

COMPARISON OF SEQUENTIAL EXPERIMENTS

BY EITAN GREENSHTEIN

Ben-Gurion University

A generalization for the theory of comparison of experiments is given to the case of sequential experiments. We investigate only the case of "0" deficiency. Applications are given to the case of exponential experiments.

Introduction. The theory of comparison of experiments deals with the following problem. Suppose two kinds of observations are available to a statistician. The observations are of two random variables having two different laws of distribution depending on the same parameter set. Some inference should be made with a resulting loss depending on the true parameter which is unknown. The statistician should choose which observation to take before making the inference. Usually one observation is better than the other depending on the type of loss and on the prior information. In some cases, one observation is better than the other regardless of the loss or prior information. The last case is of special interest.

Previous research on comparison of experiments has been confined to nonsequential experiments. In this work we will examine the problem in the case of comparison of two sequential experiments. Recent works on comparison of experiments in the context of sequential analysis are Greenshtein [6] and Greenshtein and Torgersen [8] and [9].

The concept of experiment is defined by a sample space $\mathscr{X}$, σ-algebra $\mathscr{B}^X$ and a collection of measures F_θ, $\theta \in \Theta$. Let $(\mathscr{X}, \mathscr{B}^X, F_\theta)$ and $(\mathscr{Y}, \mathscr{B}^Y, G_\theta)$, $\theta \in \Theta$, be two experiments, to be referred as X and Y. A criterion to determine whether one experiment is more informative than (sufficient for) the other was suggested by Bohnenblust, Shapley and Sherman [3] and is the following: X is a sufficient experiment for Y, if for every action space A, loss function $L(\theta, a)$, $\theta \in \Theta$, $a \in A$, and procedure δ depending on Y, there exists a procedure δ' depending on X such that the associated risk functions satisfy $R(\theta, \delta') \le R(\theta, \delta)$ for every θ.

Blackwell [2] considered the same problem and suggested the following criterion: X is a sufficient experiment for Y if there exists a Markov kernel δ such that $\forall\ A \in \mathscr{B}^Y$, $G_\theta(A) = \int \delta(A|x)\, dF_\theta(x)$ or, equivalently, $G_\theta(dy) = \int \delta(dy|x) F_\theta(dx)$.

The last criterion in words: the distribution of Y under θ can be achieved by a randomization after observing X without knowing θ. Blackwell [2] and later Le Cam [12] showed the equivalence of those two criteria. Two important references for the work done on the subject are Torgersen [14] and Strasser [13].

Received September 1993; revised June 1995.
AMS 1991 *subject classifications*. 62B15, 62L10.
Key words and phrases. Sufficient experiment, sequential experiment, Markov kernel.

A generalization of this concept to sequential experiments is the following: let $\{X_i\}$ and $\{Y_i\}$ be two sequences of r.v.'s $(X_1, \dots, X_n) \sim F_\theta^n$ and $(Y_1, \dots, Y_n) \sim G_\theta^n$, $\theta \in \Theta$, $1 \le n \le \infty$; $\{X_i\}$ is sequentially sufficient for $\{Y_i\}$ if there exists a sequence of Markov kernels $\delta_1(dy_1|x_1)$, $\delta_2(dy_2|x_1, x_2, y_1)$, $\delta_3(dy_3|x_1, x_2, x_3, y_1, y_2), \dots$ satisfying:

$$G_\theta^1(dy_1) = \int \delta_1(dy_1|x_1)\, dF_\theta(x_1),$$

$$G_\theta^2(dy_1, dy_2) = \int\int \delta_1(dy_1|x_1)\, \delta_2(dy_2|x_1, x_2, y_1)\, dF_\theta^1(x_1)\, dF^2(x_2|x_1),$$

$$\vdots$$

In words, the process $\{Y_i\}$ may be sequentially randomized from the process $\{X_i\}$ without knowing θ.

In Section 2 we give two motivating examples. Example 2.1 is simple and surprising; it shows that $(X_1, \dots, X_n)$ sufficient for $(Y_1, \dots, Y_n)$ for every fixed n does not imply in general that $\{X_i\}$ is sequentially sufficient for $\{Y_i\}$.

In Section 3 we prove two main results. We give a condition under which $(X_1, \dots, X_n)$ sufficient for $(Y_1, \dots, Y_n)$ for every n implies $\{X_i\}$ is sequentially sufficient for $\{Y_i\}$. The second result is that in some cases X is sufficient for Y if and only if observing X is equivalent to observing Y and an independent additional experiment D.

1. Formulation of a sequential procedure and preliminary results. In this formulation we will follow closely Brown [4]. Let $X_1, \dots, X_m$, $m \le \infty$, be a sequence of random variables distributed according to the law $F_\theta(dx_1, \dots, dx_m)$, $\theta \in \Theta$. Suppose a statistician, while observing the process, may choose at each stage $n \le m$ an action a_n. Finally, there is a loss $L(\theta, a_1, \dots, a_m)$ incurred from taking the actions $a_1, \dots, a_m$ when θ is the true parameter.

We will now state this more formally. Let $X_1, \dots, X_m$, $m \le \infty$, be a sequence of r.v.'s. Denote by $\mathscr{B}_n^X$ the σ-algebra generated by X_n, $\mathscr{B}_0^X$ a trivial σ-field, $\mathscr{B}_{(n)}^X$ the σ-algebra generated by $X_1, \dots, X_n$ and $\mathscr{B}^X$ the σ-algebra generated by $X_1, \dots, X_m$. Let $F_\theta(dx_1, \dots, dx_m)$ be a parametrized family of distributions on the product space $\times \mathscr{X}_i$, with the σ-algebra $\mathscr{B}^X$.

Assume there exists a set $A \subseteq K \subseteq \times_{n=0}^m K_n$ of possible sequences of actions. K_0 consists of actions that are taken without observations like start sampling or do not start sampling. Give K the Tychonoff topology. Let $\mathscr{A}_n$ be the Borel field on K_n, $\mathscr{A}_{(n)}$ the Borel field on $\times_{i=0}^n K_i$ and $\mathscr{A}$ the Borel field on K.

Definition 1.1. A sequential decision procedure is a set of conditional measures $\{\delta_n\colon n = 0, \dots, m\}$ satisfying, for $n \ge 1$,

(i) $\delta_n(\cdot|x, a)$ is a probability measure on K_n. Here $x = (x_1, \dots, x_m)$ and $a = (a_0, \dots, a_m)$, and δ_0 is a probability measure on K_0.

(ii) $\delta_n(C|\cdot,\cdot)$ is $\mathscr{B}^X_{(n)} \times \mathscr{A}_{(n-1)}$-measurable for each $C \in \mathscr{A}_n$.
(iii) $\delta_n(C|\cdot, a)$ is $\mathscr{B}_{(n)}$-measurable for each $a \in A$, $C \in \mathscr{A}_n$.

Let $L(\theta, a_1, \dots, a_m)$ be a loss function which for every θ is $\mathscr{A}$-measurable.

A procedure $\{\delta_n\} = \Delta$ determines a stochastic process on the space $(\times_{n=1}^m X_n) \times (\times_{n=0}^m K_n)$, with the σ-algebra $\mathscr{B}^X \times \mathscr{A}$ and measure $H_{\theta\Delta}(dx_1, \dots, dx_m, da_0, \dots, da_m)$. The description of this process in words is: choose an action a_0 with distribution determined by δ_0. Observe X_1 with distribution as the marginal of $F_\theta(dx_1, \dots)$ on X_1. Then choose a_1 with distribution $\delta_n(da_1|x_1, a_0)$ and so on. It is shown in [4] that this process is well defined. Denote the marginal of $H_{\theta,\Delta}$ on $\mathscr{A}$ as $\mu_{\theta\Delta}(da_0, \dots, da_m)$.

DEFINITION 1.2. A sequential decision procedure such that $\mu_{\theta\Delta}(A) = 1$ for each θ will be called an available sequential decision procedure. Here $A \subseteq K \subseteq \times_{n=0}^m K_n$ is assumed to be compact and hence measurable.

DEFINITION 1.3. The risk function is defined by

$$R(\theta, \Delta) = \int L(\theta, a_1, \dots, a_m)\, d\mu_{\theta\Delta}(a_1, \dots, a_m). \tag{1.1}$$

DEFINITION 1.4. A triple $(\mathscr{X}, \mathscr{B}^X, F_\theta, \theta \in \Theta)$ is called an experiment.

DEFINITION 1.5. A sequential experiment is defined by $(\times_{i=1}^m \mathscr{X}_i, \mathscr{B}^X, F_\theta, \theta \in \Theta)$.

When there is no ambiguity we will refer to these experiments as the experiment X and the sequential experiment $\{X_i\}$.

DEFINITION 1.6. An experiment $(\mathscr{X}, \mathscr{B}^X, F_\theta, \theta \in \Theta)$ is sufficient for $(\mathscr{Y}, \mathscr{B}^Y, G_\theta, \theta \in \Theta)$, denoted $X \supseteq Y$, if and only if for every action space A, loss function $L(\theta, a)$, $\theta \in \Theta$, $a \in A$, and decision procedure δ depending on Y, there exists a procedure δ' depending on X such that $R(\theta, \delta') \le R(\theta, \delta)$ for every θ.

THEOREM 1.1 (Le Cam [12]). *Suppose Y is Borelian and $F_\theta \ll V$ for some dominating measure V. Then $X \supseteq Y$ if and only if there is a function $\delta(B|x)$, $B \in \mathscr{B}^Y$, $x \in \mathscr{X}$, such that*:

(i) *For each $x \in \mathscr{X}$, $\delta(\cdot|x)$ is a probability measure on $\mathscr{Y}$*;
(ii) *For each $B \in \mathscr{B}^Y$, $\delta(B|\cdot)$ is $\mathscr{B}^X$-measurable*;
(iii) *For each $B \in \mathscr{B}^Y$, $G_\theta(B) = \int \delta(B|x)\, dF_\theta(x)$.*

Conditions (i) and (ii) define $\delta(\cdot|\cdot)$ to be a Markov kernel. When all of these conditions are satisfied, we say that the experiment Y is a randomization of X.

DEFINITION 1.6a. $(\times\mathscr{X}_i, \mathscr{B}^X, F_\theta, \theta \in \Theta)$ is sequentially sufficient for $(\times\mathscr{Y}_i, \mathscr{B}^Y, G_\theta, \theta \in \Theta)$, denoted $\{X_i\} \supseteq_{\text{seq}} \{Y_i\}$, if and only if for any action space $A \subseteq K = \times K_n$, loss function $L(\theta, a_1, \dots, a_m)$ and available sequential decision procedure $\Delta = \{\delta_n\}$ depending on $\{Y_i\}$, there exists an available $\Delta' = \{\delta'_n\}$ depending on $\{X_i\}$ such that the associated risk functions satisfy $R(\theta, \Delta') \leq R(\theta, \Delta)$ for each θ.

Denote by F_θ^n the restriction of F_θ to $B_{(n)}$.

THEOREM 1.1a. *Suppose $F_\theta^n \ll \nu^n$ for every n, where $\{\nu^n\}$ is a sequence of dominating measures, and $Y_1, \dots, Y_m$ is Borelian. Then $\{X_i\} \supseteq_{\text{seq}} \{Y_i\}$ if and only if there exists $\{\delta_n\} = \Delta$ satisfying the three conditions in Definition* 1.1 *where $\mathscr{Y}_1 \times \cdots \times \mathscr{Y}_m, \mathscr{B}_n^Y, \mathscr{B}_{(n)}^Y, \mathscr{B}^Y$ play the role of $K, \mathscr{A}_n, \mathscr{A}_{(n)}, \mathscr{A}$ such that*

$$\mu_{\theta\Delta}(dy_1, \dots, dy_m) = G_\theta(dy_1, \dots, dy_m) \quad \textit{for each } \theta.$$

PROOF. Follow the ideas of Le Cam; see Greenshtein [6]. □

Theorem 1.1a is the analogue of Theorem 1 of Le Cam to the sequential case. It states that $\{X_i\}$ is sequentially sufficient for $\{Y_i\}$ if and only if $\{Y_i\}$ can be sequentially randomized from $\{X_i\}$ without knowing θ, as informally described in the Introduction. Δ may be viewed as a Markov kernel between the nonsequential experiments $(X_1, \dots, X_m)$ and $(Y_1, \dots, Y_m)$; it is a special kind of Markov kernel and we call it a "sequential Markov kernel."

2. Motivating examples. In the two examples given in this section, we will investigate the following: for two sequential experiments we consider the relation where $(\mathscr{X}_1 \times \cdots \times \mathscr{X}_n, \mathscr{B}_{(n)}^X, F_\theta) \supseteq (\mathscr{Y}_1 \times \cdots \times \mathscr{Y}_n, \mathscr{B}_{(n)}^Y, G_\theta)$ for every n. The first example will show that this relation does not imply $\{X_i\}$ is sequentially sufficient for $\{Y_i\}$. The second example will indicate that sequential sufficiency might be implied by sufficiency for every fixed n under some additional conditions. These examples will motivate Theorems 3.1 and 3.2.

The following example demonstrates that in a sequential testing problem, a statistician having two observations $X \supseteq Y$ might prefer to observe first the less informative one.

EXAMPLE 2.1. Let Y_1, Y_2 be independent r.v.'s with distribution:

Under θ_0, $Y_1 \sim \text{Bernoulli}(\frac{1}{3})$, $Y_2 \sim \text{Bernoulli}(\frac{1}{4})$.

Under θ_1, $Y_1 \sim \text{Bernoulli}(\frac{2}{3})$, $Y_2 \sim \text{Bernoulli}(\frac{3}{4})$.

Let $X_1 = Y_2$ and $X_2 = Y_1$. Here $X_1 \supseteq Y_1$; the Markov kernel from the experiment X_1 to Y_1 is $\delta(1/1) = \frac{5}{6}$, $\delta(1/0) = \frac{1}{6}$, $\delta(0/1) = \frac{1}{6}$, $\delta(0/0) = \frac{5}{6}$. Obviously, $(X_1, X_2) \supseteq (Y_1, Y_2)$.

We will describe a sequential decision problem which shows that $\{X_1, X_2\}$ is not sequentially sufficient for $\{Y_1, Y_2\}$. Suppose that the cost of a first observation is 0, and the cost of a second observation is $c > 0$. The terminal actions are "θ_1" and "θ_0." Define the loss function $L(\theta_0, \theta_1) = 1$, $L(\theta_1, \theta_0) = a$, $L(\cdot, \cdot) = 0$ otherwise. Let δ^0 be the following procedure depending on $\{Y_i\}$. Observe Y_1; if $Y_1 = 0$ decide θ_0; if $Y_1 = 1$ take another observation. Then decide θ_1 if $Y_2 = 1$; decide θ_0 if $Y_2 = 0$. The risk associated with δ^0 is

$$R(\theta_0, \delta^0) = \tfrac{1}{3} \cdot c + \tfrac{1}{3} \cdot \tfrac{1}{4}, \qquad R(\theta_1, \delta^0) = \tfrac{2}{3} \cdot c + \tfrac{2}{3} \cdot \tfrac{1}{4} \cdot a + \tfrac{1}{3} \cdot a = \tfrac{2}{3} \cdot c + \tfrac{1}{2} \cdot a.$$

In the following we will show that for a suitable choice of a and c, there is no δ^1 depending on $\{X_i\}$ that improves upon δ^0 for every θ.

It can be shown [6] that there are only five admissible nonrandomized sequential procedures δ^i, $i = 1, \dots, 5$, based on $\{X_1, X_2\}$. Their associated risks $r_i = (R(\theta_0, \delta^i), R(\theta_1, \delta^i))$, when we take $a = \frac{1}{10}$ and $c = \frac{1}{100}$, are: $r_1 = (1, 0)$, $r_2 = (0, \frac{1}{10})$, $r_3 = (\frac{1}{4}, \frac{1}{40})$, $r_4 = (\frac{1}{400} + \frac{1}{12}, \frac{3}{400} + \frac{1}{20})$, $r_5 = (\frac{3}{400} + \frac{1}{2}, \frac{1}{400} + \frac{1}{120})$; we also have $r_0 = (R(\theta_0, \delta^0), R(\theta_1, \delta^0)) = (\frac{1}{300} + \frac{1}{12}, \frac{2}{300} + \frac{1}{20})$.

It may be verified that r_0 cannot be dominated by a convex combination of $r_1, \dots, r_5$; hence it is not achievable by a sequential procedure based on $\{X_1, X_2\}$.

Before starting the next example the following definitions are needed.

Definition 2.1. Let $(\mathscr{D}, \mathscr{B}^D, F_\theta)$ and $(\mathscr{Y}, \mathscr{B}^Y, G_\theta)$, $\theta \in \Theta$, be two experiments. The experiment consisting of two independent experiments Y and D is the following: the sample space $\mathscr{D} \times \mathscr{Y}$, the σ-algebra generated by $\mathscr{B}^D \times \mathscr{B}^Y$ and the product measure $F_\theta \times G_\theta$.

Definition 2.2. For two experiments X and Y, $X \approx Y$ iff $X \supseteq Y$ and $Y \supseteq X$.

Example 2.2. Let $X_1 \sim N(\theta, 1)$, $X_2 \sim N(\theta, 2)$, $Y_1 = X_2$, $Y_2 = X_1$. X_i are independent. Here we have $X_1 \supseteq Y_1$, because Y_1 can be randomized in the following way from X_1. Let $Z \sim N(0, 1)$, then $X_1 + Z \sim N(\theta, 2)$, that is no matter what θ is, $X_1 + Z$ has the same distribution as Y_1. Obviously $X_1, X_2 \approx Y_1, Y_2$.

We will explain now why $X_1, X_2 \supseteq_{\text{seq}} Y_1, Y_2$. Observe that $X_1 \approx (Y_1', D)$, where (Y_1', D) is the experiment consisting of two independent experiments $Y_1' \sim N(\theta, 2)$ and $D \sim N(\theta, 2)$. Thus, $X_1, X_2 \approx_{\text{seq}} (Y_1', D), X_2 \approx_{\text{seq}} (Y_1', D), Y_1$. Similarly, $Y_1 Y_2 \approx_{\text{seq}} Y_1, (Y_1', D)$. Hence it is enough to show $(Y_1', D), Y_1 \supseteq_{\text{seq}} Y_1, (Y_1', D)$. This is easy because an experimenter observing at the first stage (Y_1', D) may ignore D and act as if only Y_1' was observed, doing as well as an experimenter observing Y_1. In the second stage, the first experimenter observes D and Y_1, and can do as well as the second experimenter who observes (Y_1', D) at that stage. In Theorem 3.2 we will see that the factorization of X_1 to (Y_1', D) is also a necessary condition for sequential sufficiency in this setting.

In Example 2.2 we have shown a case where X_i are independent, Y_i are independent and $\{X_i\} \supseteq_{\text{seq}} \{Y_i\}$, other than the obvious case where $X_i \supseteq Y_i$ for each i.

3. Main results.

CONVENTION. In this section, for a given measure $H(dx_1, \dots, dx_m)$, $H(dx_{i_1}, \dots, dx_{i_k})$ will be understood as the marginal of $H(dx_1, \dots, dx_m)$ on the σ-algebra generated by $(X_{i_1}, \dots, X_{i_k})$.

Let $(\{\mathscr{X}_i\}, \mathscr{B}^X, \tilde{F}_\theta, \theta \in \Theta)$ and $(\{\mathscr{Y}_i\}, B^Y, G_\theta, \theta \in \Theta)$ be two sequential experiments. Let S_n be a sufficient statistic for θ based on $X_1, \dots, X_n$, $n = 1, 2, \dots$. Let $Y_{(n)} = (Y_1, \dots, Y_n)$. Let $F_\theta(ds_1, ds_2, \dots)$ be the induced measure on $S_1 \times S_2 \times \cdots$.

THEOREM 3.1. *Suppose S_n is boundedly complete, $n = 1, 2, \dots$. Then $\{X_n\} \supseteq_{\text{seq}} \{Y_n\}$ if and only if $S_n \supseteq Y_{(n)}$ for every n.*

Before proving the theorem we need the following lemmas and definitions.

DEFINITION 3.1. Let $\{X_i\}$ and $\{Y_i\}$ be two sequential experiments. Let $\{S_n\}$ be a sequence of sufficient statistics for θ based on $X_1, \dots, X_n$. Let $\{T_n\}$ be a sequence of Markov kernels from $(S_n, \mathscr{B}^{S_n})$ to $(Y_{(n)}, \mathscr{B}^Y_{(n)})$. The sequence will be called compatible if and only if

$$T_n(A_k | S_n) = E(T_k(A_k | S_k) | S_n), \qquad A_k \in \mathscr{B}^Y_{(k)}, \tag{3.1}$$

for every n, k, $1 \le k \le n$.

LEMMA 3.1. *Suppose $S_n \supseteq Y_{(n)}$ for every n, and suppose S_n is boundedly complete. Let T_n be the Markov kernel from the experiment S_n to $Y_{(n)}$; that is, T_n satisfies $\int T_n(A_k | s_n) F_\theta(ds_n) = G_\theta(A_k)$, $A_k \in \mathscr{B}^Y_{(k)}$, $1 \le k \le n$, $n = 1, 2, \dots$ (by completeness T_n is unique). Then the sequence T_n is compatible.*

PROOF. Observe that

$$\int \left[\int T_k(A_k | s_k) F(ds_k | s_n) \right] F_\theta(ds_n) = \int T_k(A_k | s_k) F_\theta(ds_k) = G_\theta(A_k).$$

Here $F(ds_k | s_n)$ is independent of θ by sufficiency. Now (3.1) follows from bounded completeness. □

LEMMA 3.2. *Let $\{X_i\}$ and $\{Y_i\}$ be two sequential experiments. Let S_n be sufficient for $X_1, \dots, X_n$, $n = 1, 2, \dots$. Assume there exists a sequence of Markov kernels T_n satisfying:*

(i) *$\{T_n\}$ is a compatible sequence;*
(ii) $\int T_n(A_n | s_n) F_\theta(ds_n) = G_\theta(A_n)$, $A_n \in B^Y_{(n)}$.

Then $\{X_n\} \supseteq_{\text{seq}} \{Y_n\}$.

PROOF. Define $\delta_1(dy_1 | s_1) = T_1(dy_1 | s_1)$. Define $\delta_n(dy_n | s_n, y_{(n-1)})$ to be the conditional distribution formed from $T_n(dy_{(n)} | s_n)$ by conditioning on $Y_{(n-1)}$. A proof that $\delta_n(\cdot|\cdot,\cdot)$ satisfy the conditions in Definition 1.1 may be found in [6]. Here $y_n, \mathscr{B}_n^Y, \mathscr{B}_{(n)}^Y$ play the role of $a_n, \mathscr{A}_n, \mathscr{A}_{(n)}$ in Definition 1.1.

Consider the process described in Section 1 induced by $\{S_n\}$ and $\Delta = \{\delta_n\}$. Denote the measure on this process $H_{\theta,\Delta}(ds_1, dy_1, ds_2, dy_2, \dots)$. The marginal $H_{\theta,\Delta}(dy_1, \dots, dy_n)$ was denoted $\mu_{\theta,\Delta}$ in Section 1. By Theorem 1.1a in order to establish the proof of this lemma, it is enough to show that $\mu_{\theta,\Delta} = G_\theta(dy_1, dy_2, \dots)$. By Kolmogorov consistency it is enough to show for every n that $G_\theta(dy_1, \dots, dy_n) = \mu_{\theta,\Delta}(dy_1, \dots, dy_n)$. Suppose we have shown for every $k < n$ that

$$H_{\theta,\Delta}(dy_{(k)}) = \int T_k(dy_{(k)}|s_k) F_\theta(ds_k).$$

We will show it for n. This will imply $H_{\theta,\Delta}(dy_{(n)}) = G_\theta(dy_{(n)})$, because by construction $H_{\theta,\Delta}(ds_n) = F_\theta(ds_n)$, and now $H_{\theta,\Delta}(dy_{(n)}) = G_\theta(dy_{(n)})$ follows from (ii):

$$\begin{aligned} &H_{\theta,\Delta}(ds_{n-1}, ds_n, dy_{(n-1)}, dy_n) \\ &\quad = F_\theta(ds_{n-1}) T_{n-1}(dy_{(n-1)}|s_{n-1}) F_\theta(ds_n|s_{n-1})\, \delta_n(dy_n|s_n, y_{(n-1)}) \\ &\quad = F(ds_{n-1}|s_n) T_{n-1}(dy_{(n-1)}|s_{n-1}) F_\theta(ds_n)\, \delta_n(dy_n|s_n, y_{(n-1)}). \end{aligned}$$

The first equality follows from the induction hypothesis upon realizing that $F_\theta(ds_{n-1}) = H_{\theta,\Delta}(ds_{n-1})$. The second equality follows because

$$F_\theta(ds_{n-1}) F_\theta(ds_n|s_{n-1}) = F(ds_{n-1}|s_n) F_\theta(ds_n).$$

From the compatibility assumption it follows that

$$\int T_{n-1}(dy_{(n-1)}|s_{n-1}) F(ds_{(n-1)}|s_n) = T_n(dy_{(n-1)}|s_n).$$

Thus,

$$\begin{aligned} &\int_{S_{n-1}} H_{\theta,\Delta}(ds_{n-1}, ds_n, dy_{(n-1)}, dy_n) \\ &\quad = T_n(dy_{(n-1)}|s_n)\, \delta_n(dy_n|s_n, y_{(n-1)}) F_\theta(ds_n) \\ &\quad = T_n(dy_{(n)}|s_n) F_\theta(ds_n). \end{aligned}$$

Finally,

$$\begin{aligned} &\int_{S_n} \int_{S_{n-1}} H_{\theta,\Delta}(ds_{n-1}, ds_n, dy_{(n-1)}, dy_n) \\ &\quad = \int T_n(dy_{(n)}|s_n) F_\theta(ds_n) = G_\theta(dy_{(n)}). \end{aligned}$$

This completes the proof. □

PROOF OF THEOREM 3.1. From Lemmas 3.1 and 3.2, it is easy to conclude the proof. □

Consider Example 2.2. We have shown that if $X \sim N(\theta, \sigma_1^2)$ and $Y \sim N(\theta, \sigma_2^2)$, $\sigma_1^2 \le \sigma_2^2$ are independent, then $(X, Y) \supseteq_{\text{seq}} (Y, X)$. Now it can also be deduced from Theorem 3.1. Originally it was shown by a factorization of the experiment X, that is, showing that X is equivalent to two independent experiments $Y' \sim N(\theta, \sigma_2^2)$ and $D \sim N(\theta, \sigma_3^2)$. In the following theorem we will show that this factorization criterion is necessary for a sequence X_1, X_2 to be sequentially sufficient for Y_1, Y_2, where $Y_1 = X_2$ and $Y_2 = X_1$. This result will imply a general factorization theorem for exponential experiments.

First we introduce some notation and other preliminaries. Suppose $(X_1, X_2) \supseteq_{\text{seq}} (Y_1, Y_2)$. Consider the experiment $((\mathscr{X}_1, \mathscr{X}_2, \mathscr{Y}_1, \mathscr{Y}_2), \mathscr{B}^{X_1 X_2 Y_1 Y_2}, H_{\theta\Delta}(dx_1, dx_2\, dy_1\, dy_2))$, $\theta \in \Theta$, where $H_{\theta\Delta}$ is the measure induced by $\delta_1(dy_1|x_1)$ and $\delta_2(dy_2|x_1, x_2, y_1)$. We will refer in the sequel to experiments that are induced from the experiment (X_1, X_2, Y_1, Y_2) in the following way: for each θ there is a conditional distribution $H_{\theta,\Delta}(dx_1, dx_2|Y_1 = y_1)$ and an experiment $((\mathscr{X}_1, \mathscr{X}_2), \mathscr{B}^{X_1, X_2}, H_{\theta\Delta}(dx_1, dx_2|Y_1 = y_1), \ \theta \in \Theta)$. Denote such an experiment as $(X_1, X_2|Y_1 = y_1)$. Similarly define $(Y_2|Y_1 = y_1)$ and $(X_1|Y_1 = y_1)$.

REMARK 3.1. In the experiment (X_1, X_2, Y_1, Y_2), (X_1, X_2) is a sufficient statistic. The reason is that the distribution of Y_1, Y_2 conditional on $X_1 = x_1$ and $X_2 = x_2$ is independent of θ.

THEOREM 3.2. *Let X_1, X_2 and Y_1, Y_2 be two sequential experiments. Assume*:

(i) *X_i, $i = 1, 2$, are independent, and Y_i, $i = 1, 2$, are independent.*
(ii) *$X_1 \approx Y_2$ and $X_2 \approx Y_1$.*

Then $(X_1, X_2) \supseteq_{\text{seq}} (Y_1, Y_2)$ if and only if $X_1 \approx (Y_1, D)$, where (Y_1, D) is the experiment consisting of two independent experiments Y_1 and D.

Before proving Theorem 3.2 we need the following:

Let $(\mathscr{X}, \mathscr{B}^X, F_\theta)$ be an experiment and let $\{\lambda\}$ be the set of all probability distributions on the parameter set $\{\theta\}$ with a finite support. Here the value of $\lambda(\cdot)$ at a point θ is the point mass of the distribution λ.

DEFINITION 3.2. The functional

$$H_X(\lambda) = \int \prod_{\theta \in \Theta} f_\theta^{\lambda(\theta)}(x)\, d\eta(x), \qquad \lambda \in \{\lambda\},$$

where $f_\theta(x) = dF_\theta(x)/d\eta(x)$ for some dominating measure η, is the Hellinger transform of the experiment X.

The following can be shown (Strasser [13]):

(a) $X \supseteq Y$ implies $H_X(\lambda) \le H_Y(\lambda)$ for every $\lambda \in \{\lambda\}$.
(b) $X \approx Y$ if and only if $H_X(\lambda) = H_Y(\lambda)$ for every $\lambda \in \{\lambda\}$.
(c) X and Y are independent implies $H_{(X,Y)}(\lambda) = H_X(\lambda) \cdot H_Y(\lambda)$.

Another general fact we will use (Strasser [13]) is: for dominated families F_θ and G_θ, $\theta \in \Theta$.

(d) $(\mathscr{X}, \mathscr{B}^X, F_\theta, \theta \in \Theta) \supseteq (\mathscr{Y}, \mathscr{B}^Y, G_\theta, \theta \in \Theta)$ if and only if for every finite subset $\tilde{\Theta}$, $(\mathscr{X}, \mathscr{B}^X, F_\theta, \theta \in \tilde{\Theta}) \supseteq (\mathscr{Y}, \mathscr{B}^Y, G_\theta, \theta \in \tilde{\Theta})$.

LEMMA 3.3. *Suppose* $(X_1, X_2) \supseteq_{\text{seq}} (Y_1, Y_2)$. *Then* $(X_1, X_2 | Y_1 = y_1) \supseteq (Y_2 | Y_1 = y_1)$ *for almost every* y_1.

PROOF. Let $A \in \mathscr{B}^{Y_2}$; then

$$H_{\theta, \Delta}(A | Y_1 = y_1) = \int H_{\theta, \Delta}(A | x_1, x_2, y_1) \, dH_{\theta, \Delta}(x_1, x_2 | Y_1 = y_1).$$

By sufficiency (Remark 3.1) $H_{\theta, \Delta}(A | x_1, x_2, y_1) = H_\Delta(A | x_1, x_2, y_1)$ is independent of θ and can be viewed as the desired Markov kernel between the experiments. □

PROOF OF THEOREM 3.2. "If" is obvious; let us prove "only if." By (d) it is enough to prove $X_1 \approx (Y_1, D)$ for every experiment with finite parameter space $\tilde{\Theta} \subseteq \Theta$. Applying Lemma 3.3 we get: $(X_1, X_2 | Y_1 = y_1) \supseteq (Y_2 | Y_1 = y_1)$ a.e. $H_{\theta, \Delta}(dy_1)$. Our first step is to show that $(X_1, X_2 | Y_1 = y_1) \approx (Y_2 | Y_1 = y_1)$ a.e. Observe that (X_1, X_2) is sufficient for (X_1, X_2, Y_1) (Remark 3.1); hence by (b) $H_{(X_1, X_2)}(\lambda) = H_{(X_1, X_2, Y_1)}(\lambda)$ for every $\lambda \in \{\lambda\}$. Thus

$$H_{(X_1 X_2)}(\lambda) = \int \prod_{\theta \in \Theta} h_\theta(x_1, x_2, y_1)^{\lambda(\theta)} \, d\eta(x_1, x_2, y_1),$$

where η is any measure dominating $H_{\theta, \Delta}(dx_1, dx_2, dy_1, dy_2)$ and $h_\theta = dH_{\theta, \Delta} / d\eta$. Now

$$\frac{dH_{\theta, \Delta}(x_1, x_2 | Y_1 = y_1)}{d\eta(x_1, x_2 | Y_1 = y_1)} = \frac{h_\theta(x_1, x_2, y_1)}{\Psi_\theta(y_1)},$$

where $\Psi_\theta(y_1) = \int h_\theta(x_1, x_2, y_1) \, d\eta(x_1, x_2 | Y_1 = y_1)$. Notice that $\Psi_\theta(y_1) = dH_{\theta, \Delta}(y_1) / d\eta(y_1)$. $H_{(X_1, X_2)}(\lambda)$ can be written now as

$$\text{(i)} \quad H_{(X_1, X_2)}(\lambda) = \int \int \prod_{\theta \in \tilde{\Theta}} \Psi_\theta^{\lambda(\theta)}(y_1) \times \frac{h_\theta^{\lambda(\theta)}(x_1, x_2, y_1)}{\Psi_\theta^{\lambda(\theta)}(y_1)} \, d\eta(x_1, x_2 | Y_1 = y_1) \, d\eta(y_1)$$
$$= \int H_{(X_1, X_2 | Y_1 = y_1)}(\lambda) \prod_{\theta \in \tilde{\Theta}} \Psi_\theta^{\lambda(\theta)}(y_1) \, d\eta(y_1).$$

Suppose $(X_1, X_2 | Y_1 = y_1) \supseteq X_1$ and $X_1 \not\supseteq (X_1, X_2 | Y_1 = y_1)$ on a set of positive measure $\eta(dy_1)$. We will show this implies that there exists λ_0 such that

$H_{(X_1, X_2|Y_1=y_1)}(\lambda_0) < H_{X_1}(\lambda_0)$ on a set with positive measure $\eta(dy_1)$ which will imply

$$\text{(ii)} \quad \int H_{(X_1, X_2|Y_1=y_1)}(\lambda_0) \prod_{\theta \in \tilde{\Theta}} \Psi_\theta^{\lambda_0(\theta)}(y_1)\, d\eta(y_1) < H_{X_1}(\lambda_0) H_{Y_1}(\lambda_0)$$
$$= H_{X_1}(\lambda_0) H_{X_2}(\lambda_0).$$

Results (i) and (ii) lead to the contradiction $H_{(X_1, X_2)}(\lambda_0) < H_{(X_1, X_2)}(\lambda_0)$. Now we will show the existence of such λ_0. Let $\tilde{\Theta} = (\theta_1, \dots, \theta_n)$, and consider the following measure space: sample space $R^n \times Y_1$, where R^n is the n dimension Euclidean space, with the obvious σ-algebra and measure which is the product of $\eta(dy_1)$ and Lebesgue. Let $A = \{(\lambda, y_1) | \lambda \in R^n,\ y_1 \in Y_1, H_{(X_1, X_2|Y_1=y_1)}(\lambda) < H_{X_1}(\lambda)\}$. By assumption and using (a) and (b) there exists a set of positive measure $\eta(dy_1)$ satisfying: $H_{(X_1, X_2|Y_1=y_1)}(\lambda) < H_{X_1}(\lambda)$ for some $\lambda \in \{\lambda\}$. Since the Hellinger transform is a continuous function if $H_{(X_1, X_2|Y_1=y_1)}(\cdot) < H_{X_1}(\cdot)$ for some λ, the strict inequality holds for a set of positive Lebesgue measure. Then by Fubini's theorem A has a positive measure, and, using Fubini's theorem again, we deduce that there exists λ_0 such that $H_{(X_1, X_2|Y_1=y_1)}(\lambda_0) < H_{X_1}(\lambda_0)$ on a set of positive measure $\eta(dy_1)$. As noted, this leads to a contradiction. Hence $(X_1, X_2 | Y_1 = y_1) \approx X_1$ almost everywhere $\eta(dy_1)$.

Since by assumption X_2 is independent of X_1, and by construction Y_1 is independent of X_2, we may conclude: $(X_1, X_2 | Y_1 = y_1) \approx ((X_1 | Y_1 = y_1), X_2)$, where the last experiment consists of two independent experiments $(X_1 | Y_1 = y_1)$ and X_2. By (b) and (c) we get

$$H_{X_1}(\lambda) = H_{(X_1, X_2|Y_1=y_1)}(\lambda) = H_{(X_1|Y_1=y_1)}(\lambda) \cdot H_{X_2}(\lambda)$$

for every λ and almost every y_1. Hence there exists y_1^0 such that $H_{(X_1|Y_1=y_1^0)}(\lambda) = H_{(X_1, X_2|Y_1=y_1)}(\lambda)$ for every λ and almost every y_1. Denote the experiment $(X_1 | Y_1 = y_1^0)$ as D. Then

$$\begin{aligned} H_{(X_1)}(\lambda) &= H_{(X_1, Y_1)}(\lambda) \\ &= \iint \prod_{\theta \in \tilde{\Theta}} h(x_1, y_1)^{\lambda(\theta)}\, d\eta(x_1|y_1)\, d\eta(y_1) \\ &= \iint \prod_{\theta \in \tilde{\Theta}} \Psi_\theta^{\lambda(\theta)}(y_1) \frac{h(x_1, y_1)^{\lambda(\theta)}}{\Psi_\theta^{\lambda(\theta)}(y_1)}\, d\eta\,(x_1/y_1)\, d\eta\,(y_1) \\ &= H_D(\lambda) \cdot H_{X_2}(\lambda). \end{aligned}$$

By (b) we conclude $X_1 \approx (X_2, D)$. □

Applications. In the remaining part of this section we will show how the theory is applied for experiments $(\mathscr{X}, \mathscr{B}^X, F_\theta, \theta \in \Theta)$ when $\{F_\theta\}$ is an exponential family. The following is an immediate corollary of Theorem 3.1.

THEOREM 3.3. *Let* $\{X_i\}_{i=1}^m$ *and* $\{Y_i\}_{i=1}^m$, $m \le \infty$, *be two sequential experiments with parameter set* $\Theta \subseteq R^k$. *Suppose*:

(i) *There exists a sequence of sufficient statistics* $S_n = S_n(X_1, \dots, X_n)$ *such that*

$$dF_\theta^n(x_{(n)}) = \exp(\theta \cdot S_n - \Psi(\theta))\, d\mu_n(x_{(n)}).$$

(ii) Θ *has a nonvoid interior*.

Then $\{X_i\} \supseteq_{\text{seq}} \{Y_i\}$ *if and only if* $X_{(n)} \supseteq Y_{(n)}$ *for every* n.

PROOF. This is true because S_n is complete and sufficient when Θ has a nonvoid interior. □

EXAMPLE 3.1. Consider the linear experiments:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = M_1 \cdot \beta + \varepsilon_1 \quad \text{and} \quad \begin{pmatrix} Z_1 \\ \vdots \\ Z_m \end{pmatrix} = M_2 \cdot \beta + \varepsilon_2, \qquad \varepsilon_i \sim N(0, \sigma^2 I),$$

where M_1 and M_2 are two $m \times k$-dimensional matrices and ε_i are m-dimensional random vectors. Here the unknown parameter θ is the k-dimensional vector β. Hansen and Torgersen [10] showed that $(Y_1, \dots, Y_m) \supseteq (Z_1, \dots, Z_m)$ if and only if $((M_1' M_1) - (M_2' M_2))$ is positive semidefinite. If we consider $\{Y_i\}$ and $\{Z_i\}$ as sequential experiments, the conditions of Theorem 3.3 are satisfied and we can deduce the following: $(Y_1, \dots, Y_m) \supseteq_{\text{seq}} (Z_1, \dots, Z_m)$ if and only if for every n, $((M_1^{(n)\prime} M_1^{(n)}) - (M_2^{(n)\prime} M_2^{(n)}))$ is positive semidefinite. Here $M^{(n)}$ is the matrix consisting of the first n rows of X.

Another application gives a slight improvement of the following theorem, which was proved independently by Ehm and Müller [5] and Janssen [11].

THEOREM 3.4. *Let* $(\mathscr{X}, \mathscr{B}^X, F_\theta)$ *and* $(\mathscr{Y}, \mathscr{B}^Y, G_\theta)$, $\theta \in \Theta$, *be two experiments. Suppose* $\{F_\theta\}$ *and* $\{G_\theta\}$ *are exponential families and* Θ *has nonvoid interior. Assume* X *and* Y *are the canonical observations and* Θ *is the canonical parameter set. Then* $X \supseteq Y$ *implies*:

(i) $X \approx (Y, D)$, *where* (Y, D) *is an experiment consisting of two independent experiments* $(\mathscr{Y}, \mathscr{B}^Y, G_\theta)$ *and* $(\mathscr{D}, \mathscr{B}^D, K_\theta)$.
(ii) $\{K_\theta\}$ *is an exponential family*.

THEOREM 3.4a. *The conclusion of Theorem* 3.4 *remains valid if we replace the condition that* Θ *has nonvoid interior by the* (*weaker*) *condition that* $X + Y$ *is boundedly complete when* X, Y *are independent*.

PROOF. (i) Consider the two-stage sequential experiments X, Y and Y, X, where X and Y are independent. By Theorem 3.1, $(X, Y) \supseteq_{\text{seq}} (Y, X)$; hence, by Theorem 3.2, $X \approx (Y, D)$.

(ii) Consider the measure $H_{\theta,\delta}(dx, dy)$ induced by $\delta_1(dy|x)$. Then

$$H_{\theta,\delta}(dx, dy) = \exp(\theta \cdot x - \Psi(\theta))\ \delta_1(dy|x)\ d\mu(x),$$

where $F_\theta(dx) = \exp(\theta \cdot x - \Psi(\theta))\, d\mu(x)$. From the proof of Theorem 3.2, $D \approx (X|Y = y_0)$. Denote $\omega(dx, dy) = \delta_1(dy|x)\mu(dx)$. Then

$$\frac{dK_\theta}{d\mu} = \frac{dH_{\theta,\delta}(X|Y = y_0)}{d\mu} = \frac{\exp(\theta \cdot x - \Psi(\theta))\,\omega(dx|Y = y_0)}{\int \exp(\theta \cdot x - \Psi(\theta))\,\omega(dx|Y = y_0)}. \qquad \square$$

Further application of the results in this section to sequential testing may be found in Greenshtein [7].

Concluding remarks.

REMARK 1. Consider the two nonsequential experiments $((\mathscr{X}_1, \dots, \mathscr{X}_m), \mathscr{B}^X, F_\theta)$ and $((\mathscr{Y}_1, \dots, \mathscr{Y}_m), \mathscr{B}^Y, G_\theta)$, $\theta \in \Theta$. Let $\mathscr{D}$ be the set of all sequential Markov kernels Δ (see the end of Section 1) between the experiments. A natural generalization of Le Cam's [12] concept of deficiency to sequential experiments is to define: $\varepsilon = \inf_{\Delta \in \mathscr{D}} \sup_{\theta \in \Theta} \|F_\theta \Delta - G_\theta\|$. Here the norm is of total variation, and ε is defined as the deficiency between the induced sequential experiments.

The concept was shown to be fruitful in asymptotic theory in defining limits of sequences of experiments. The analogue is to consider limits of sequences of sequential experiments.

REMARK 2. The first generalization of sufficiency concepts to sequential analysis was done by Bahadur [1]. He introduced the idea of transitivity, extending the concept of sufficient statistics to sequential analysis. His motivation was to summarize given data.

Let $\{X_i\}$ be a sequential experiment, and let $\{S_i\}$ be a sequence of sufficient statistics. A natural question in the context of our paper is whether the two sequential experiments are equivalent. This is plausible since for every fixed n the nonsequential experiments $(X_1, \dots, X_n)$ and $(S_1, \dots, S_n)$ are equivalent. In our setting a decision at stage n may be a function of $(S_1, \dots, S_n) = V_n$, since, unlike in [1], we are not motivated by summarizing the data. Still, examples may be given where the sequence $\{V_i\}$ is not transitive and whence the experiments $\{X_i\}$ and $\{S_i\}$ are not equivalent. Such examples were first pointed out by Bahadur, and they demonstrate a phenomenon similar to our Example 2.1. Theorem 3.1 implies that if a sequence of complete sufficient statistics exists, then the experiments $\{S_n\}$ and $\{X_n\}$ are sequentially equivalent.

Acknowledgment. This work is part of the author's Ph.D. thesis at Cornell University. The author would like to thank his thesis advisor, Professor L. D. Brown, for his most valuable help and advice.

REFERENCES

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

[2] BLACKWELL, D. (1953). Equivalent comparison of experiments. *Ann. Math. Statist.* **24** 265–272.

[3] BOHNENBLUST, H. F., SHAPLEY, L. S. and SHERMAN, S. (1949). Reconnaissance in game theory. Unpublished Rand research memorandum.

[4] BROWN, L. D. (1977). Closure theorems for sequential design processes. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and D. S. Moore, eds.) **2** 57–91. Academic Press, New York.

[5] EHM, W. and MÜLLER, D. W. (1983). Factorizing the information contained in an experiment, conditionally on the observed value of a statistic. *Z. Wahrsch. Verw. Gebiete* **65** 121–134.

[6] GREENSHTEIN, E. (1990). On some decision theory problems in sequential analysis. Ph.D. dissertation, Statistics Center, Cornell Univ.

[7] GREENSHTEIN, E. (1993). An independent sequentially equivalent presentation of dependent experiments. *Sequential Analysis*. To appear.

[8] GREENSHTEIN, E. and TORGERSEN, E. N. (1993). Statistical information and expected number of observation for sequential experiments. Unpublished manuscript.

[9] GREENSHTEIN, E. and TORGERSEN, E. N. (1994). Ordering of sequentially sampled exponential experiments. Unpublished manuscript.

[10] HANSEN, O. H. and TORGERSEN, E. N. (1974). Comparison of linear normal experiments. *Ann. Statist.* **2** 367–373.

[11] JANSSEN, A. (1988). A convolution theorem for comparison of exponential experiments. Technical Report, Dept. Mathematics, Univ. Siegen.

[12] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.

[13] STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

[14] TORGERSEN, E. N. (1991). *Comparison of Experiments*. Cambridge Univ. Press.

DEPARTMENT OF INDUSTRIAL ENGINEERING
AND MANAGEMENT
BEN-GURION UNIVERSITY
BEER-SHEVA, 84105
ISRAEL