Comparison of Sequential Experiments

by

Eitan Greenshtein

Technical Report # 90–19

Department of Statistics
Purdue University

April, 1990

Δ

COMPARISON OF SEQUENTIAL EXPERIMENTS

by

Eitan Greenshtein

ABSTRACT

A generalization for the theory of comparison of experiments is given to the case of sequential 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 what the loss or the prior information are. The last case is of a 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.

The concept of experiment is defined by a sample space X, σ algebra $\mathcal{B}^X$, and a collection of measures F_θ $\theta \in \Theta$. Let $(X, \mathcal{B}^X, F_\theta)$ and $(Y, \mathcal{B}^Y, G_\theta)$ $\theta \in \Theta$ be two experiments. A criterion to determine whether one experiment is more informative than (sufficient for) the other was suggested by Bohnenblust, Shapley and Sherman [2] 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 $R(\theta, \delta') \leq R(\theta, \delta)$ for every θ.

Blackwell [1] considered the same problem and suggested the following criterion: X is a sufficient experiment for Y if there exist a markov kernel δ such that $\forall\, A \in \mathcal{B}^Y, G_\theta(A) = \int \delta(A|x) dF_\theta(x)$.

The last criterion in words: The distribution of Y under θ can be achieved by a randomization after observing X without knowing θ. Blackwell [1] and later LeCam [9] showed the equivalence of those two criteria. Two important references for the work done on the subject are Torgerson [12] and Strasser [11].

In Section 1 we will formulate a general sequential decision problem. Section 2 analogues to Blackwell and B.S.S. criteria are defined for sequential experiments. Equivalence between the criteria is proved. In Section 3 two motivating examples are given. In Section 4 two main theorems are given. In the first theorem, for two sequential experiments $\{X_i\}$ and $\{Y_i\}$ it is proved that: If for every fixed size experiment $(X_1, \ldots, X_n)$ is sufficient for $(Y_1, \ldots, Y_n)$ and for every n $(X_1, \ldots, X_n)$ has a complete sufficient statistics, then $\{X_i\}$ is sequentially sufficient for $\{Y_i\}$. The second theorem gives a necessary and sufficient condition for a sequence X_1, X_2 to be sequentially sufficient for Y_1, Y_2, where $Y_1 = X_2$ and $Y_2 = X_1$. In Section 5 applications are given to the case of comparison of sequential exponential experiments, that is when the distributions involved belong to an exponential

family.

Section 1: Formulation of a Sequential Procedure

In this formulation we will follow closely Brown [3]. Let $X_1, \ldots, X_m$ $m \leq \infty$ be a sequence of random variables distributed according to the law $F_\theta(dx_1, \ldots, dx_m)$ $\theta \in \Theta$. Suppose a statistician, while observing the process, may choose at each stage $n \leq m$ an action a_n. Finally there is a loss $L(\theta, a_1, \ldots, a_m)$ incurred from taking the action $a_1, \ldots, a_m$ when θ is the true parameter.

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

Assume there exists a set $A \subseteq K \subseteq \underset{n=0}{\overset{m}{X}} K_n$ 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 $\mathcal{A}_n$ be the Borel field on K_n, $\mathcal{A}_{(n)}$ the Borel field on $\underset{i=0}{\overset{n}{X}} K_i$, $\mathcal{A}$ the Borel field on K.

Definition 1: A sequential decision procedure is a set of conditional measures $\{\delta_n : n = 0, \ldots, m\}$ satisfying for $n \geq 1$

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

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

(iii) $\delta_n(C|\cdot, a)$ is $\mathcal{B}_{(n)}$ measurable for each $a \in A, C \in \mathcal{A}_n$.

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

A set $\{\delta_n\} = \Delta$, determines a stochastic process on the space $\left(\overset{m}{\underset{n=1}{X}} X_n\right) \times \left(\overset{m}{\underset{n=0}{X}} K_n\right)$, with the σ algebra $\mathcal{B}^X \times \mathcal{A}$ and measure $H_{\theta\Delta}(dx_1, \ldots, dx_m, da_0, \ldots, da_m)$. 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 $F_\theta(dx_1, \ldots)$ on X_1. Then choose a_1 with distribution $\delta_n(da_1|x_1, a_0)$ and so on. It is shown in [3] that this process is well defined. Denote the marginal of $H_{\theta,\Delta}$ on $\mathcal{A}$ as $\mu_{\theta\Delta}(da_0, \ldots, da_m)$.

Definition 2: A sequential decision procedure such that $\mu_{\theta\Delta}(A) = 1$ for each θ will be called an available sequential decision procedure. Here $A \subseteq K = \overset{m}{\underset{n=0}{X}} K_n$ assumed to be compact and hence measurable.

Definition 3: The risk function is defined:

$$R(\theta, \Delta) = \int L(\theta, a_1, \ldots, a_m) d\mu_{\theta\Delta}(a_1, \ldots, a_m). \tag{2}$$

Assumptions:

(i) $L(\theta_i, \cdot)$ is continuous for each θ, and $L(\cdot, \cdot)$ is bounded.

(ii) A is compact.

(iii) $m < \infty$

(iv) $F_\theta \ll \nu$ (F_θ is dominated by a measure ν).

Under the above assumptions it is proved in [3] that:

Theorem 1: There exists a minimax procedure, and a least favorable sequence of priors supported on finite subsets of Θ.

In the second part of the next section, we will explain how ordinary or common types of sequential procedures are included in this formulation.

Section 2

In this section the notion of sufficient experiment initiated by Blackwell [1] will be introduced. A generalization of a fundamental theorem by LeCam in comparison of experiments is then given, for the case of sequential experiments.

Definition 1: A triple $(X, \mathcal{B}^X, F_\theta\ \theta \in \Theta)$ is called an experiment.

Definition 2: A sequential experiment is defined by $(\overset{m}{\underset{i=1}{X}} X_i, \mathcal{B}^X, F_\theta\ \theta \in \Theta)$.

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

Definition 3: An experiment $(X, \mathcal{B}^X, F_\theta\ \theta \in \Theta)$ is sufficient for $(Y, \mathcal{B}^Y, G_\theta\ \theta \in \Theta)$ 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: $R(\theta, \delta') \leq R(\theta, \delta)$ for every θ.

Theorem 1: LeCam [7].

Suppose Y is Borelian and $F_\theta \ll V$. Then $X \supseteq Y$ if and only if there is a function $\delta(B|x)\ B \in \mathcal{B}^Y, x \in X$ such that:

(i) For each $x \in X, \delta(\cdot|x)$ is a probability measure on Y.

(ii) For each $B \in \mathcal{B}^Y, \delta(B|\cdot)$ is B^X measurable.

(iii) For each $B \in \mathcal{B}^Y, G_\theta(B) = \int \delta(B|x) dF_\theta(x)$.

Conditions i, ii define $\delta(\cdot|\cdot)$ to be a Markov kernel. When i, ii and iii are satisfied we shall say: Experiment Y is a randomization of X.

Definition 3a: $(\times X_i, \mathcal{B}^X, F_\theta)$ is sequentially sufficient for $(\times Y_i, \mathcal{B}^Y, G_\theta)$ denoted

$\{X_i\} \underset{\text{seq}}{\subseteq} \{Y_i\}$ if and only if: For any action space $A \subseteq K = \times K_n$, loss function $L(\theta, a_1, \ldots, a_m)$ 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 F_θ^n the restriction of F_θ to $B_{(n)}$.

Theorem 1a: Suppose $F_\theta^n \ll \nu^n$ for every n, and $Y_1, \ldots, Y_m$ is Borelian. Then $\{X_i\} \underset{\text{seq}}{\subseteq} \{Y_i\}$ if and only if:

(i) There exists $\{\delta_n\} = \Delta$ satisfying the three conditions in Definition 1.1 where $Y_n \times \ldots \times Y_m, B_n^Y, B_{(n)}^Y, B^Y$ play the role of $K, \mathcal{A}_n, \mathcal{A}_{(n)}, \mathcal{A}$.

(ii) $\mu_{\theta\Delta}(dy_1, \ldots, dy_m) = G_\theta(dy_1, \ldots, dy_m)$ for each θ.

Proof: We will follow the ideas in LeCam's proof. The proof of the if part is easy, we will prove the only if part.

First assume $m < \infty$. We may assume w.l.o.g. that $Y_n \times \ldots \times Y_m$ is a compact polish space, since $Y_1 \times \ldots \times Y_m$ is Borelian so that there is an one to one bimeasurable map of $\times Y_i$ into [0, 1], and we may consider the closure of this map by defining $\nu(c) = 0$ for C the set of all points added by the closure. Take the action space A to be identical to the sample space $\times Y_i$. Consider the decision rule based on $\{Y_i\}\Delta = \{\delta_n\}$ where:

$$\delta_n(E|y_1, \ldots, y_m, y_1, \ldots, y_m) = \begin{cases} 1 & y_n \in E \\ 0 & \text{otherwise} \end{cases}$$

Define an auxiliary sequential experiment $(\times X_i, \mathcal{B}^X, F_{(\theta,L)})$; in the auxiliary experiment the parameter set consists of all pairs (θ, L) where $\theta \in \Theta$ and $L(\theta, y_1, \ldots, y_m)$ is a loss function such that $L(\theta, \cdot)$ is continuous bounded by 1 for each θ. Here $F_{(\theta,L)}(A) = F_\theta(A)$

for every $A \in \mathcal{B}^Y$. Define the loss in the auxiliary problem as:

$$\tilde{L}((\theta, L), y_1, \ldots, y_m) = L(\theta, y) - \int L(\theta, y) dG_\theta(y) \quad y = y_1, \ldots, y_m.$$

Let P be a finitely supported prior on the auxiliary parameter set. We index the parameters with positive prior as (θ_k, L_k) $k = 1, \ldots, K$. Now re-index them as $\theta_i, \times L_j, i = 1, \ldots, I$ $j = 1, \ldots, J_i$ in the obvious way so that $\theta_j \neq \theta_k$ for $j \neq k$.

The Bayes risk $B(p)$, and the Bayes procedure $\{\delta_n^p\} = \Delta^p$ in the auxiliary problem with prior P satisfy:

$$\begin{aligned} B(P) &= \int R((\theta_i, L_j), \Delta^p) dP(\theta_i, L_j) \\ &= \int \left[\int L_j(\theta_i, y) d\mu_{\theta_i \Delta^p}(y) - \int L(\theta_i, y) dG_{\theta_i}(y) \right] dP(\theta_i, L_j) \\ &= \sum_i \left[\int \sum_j L_j(\theta_i, y) d\mu_{\theta_i \Delta^p}(y) - L_j(\theta_i, y) dG_{\theta_i}(y)) P(L_j | \theta_i) \right] P(\theta_i) \\ &= \sum_i \left[\int L^*(\theta_i, y) \left[d\mu_{\theta_i \Delta^p}(y) - dG_{\theta_i}(y) \right] \right] P(\theta_i) \end{aligned}$$

Here L^* is implicitly defined. Now the last expression can be written as:

$$B(p) = \sum_i R(\theta_i \Delta^p) - R(\theta_i, \Delta)$$

here the risk $R(\theta, \cdot)$ is with respect to L^*. By hypothesis there exists Δ' such that $R(\theta_i \Delta') - R(\theta_i, \Delta) \leq 0$ for each θ_i, hence $B(p) < 0$.

By Theorem 1.1 there exists a least favorable sequence of priors $\{P_n\}$ such that the minimax value m satisfy $m = \sup_{\{P_n\}} B(P_n)$. $m \leq 0$ since $B(P_n) \leq 0$ for every p_n.

We conclude: There exists $\Delta^m = \{\delta_n^m\}$ minimax procedure such that:

$$\int L(\theta, y) d\mu_{\theta \Delta^m}(y) \leq \int L(\theta, y) dG_\theta(y)$$

for every θ and continuous bounded L. Thus $\mu_{\theta, \Delta^m}(dy) = G_\theta(dy)$.

In the case $m = \infty$. We conclude by previous considerations that there exists an infinite sequence $\{\Delta^n\}$ where $\Delta^1 = \{\delta_1^1, \ldots\}, \Delta^2 = \{\delta_1^2, \delta_2^2, \ldots\}, \ldots$ such that for every n $\mu_{\theta\Delta^n}(dy) = G_\theta^n(dy)$. By diagonalization argument we can construct a subsequence Δ^{n_k} that approaches a limit Δ^m in the same sense as [3]. It follows now that: $\mu_{\theta\Delta^m} = G_\theta$. $\square$

In this work we consider somewhat artificially wide and general class of sequential decision problems, as formulated in Section 1. The reason is that we feel the proofs of the main theorems are easier and more natural using the general definitions; also in view of Theorem 1 Section 4, it seems that we should not expect additional interesting examples of "Sequential–Sufficiency" if we consider a narrower class of decision problems. In the remaining of this section we will examine what happens when we narrow ourselves to special types of sequential decision problems. First some definitions.

Let $X_1, \ldots, X_m$ $m \leq \infty$ be a sequence of random variables.

Definition 4: A stopping rule N is a random variable whose range is $(0, 1, \ldots)$, satisfying the conditions: The event $(N \leq k)$ belongs to $\mathcal{B}_{(k)}$. $\mathcal{B}_{(k)} = \sigma(X_1, \ldots, X_k)$.

Definition 5: The σ algebra related to a stopping rule N, denoted $B_{(N)}$, is the set:

$$\{A | A \in \sigma(X_1, \ldots, X_m), A \cap (N = k) \in \mathcal{B}_{(k)} \quad \forall\, k = 0, 1, 2, \ldots\}$$

where $\mathcal{B}_{(0)}$ is a trivial σ algebra.

Commonly the theory deals with procedures in the following setting. A set $\tilde{A}$ of terminal actions is given, with loss function $\tilde{L}(\theta, \tilde{a})$ $\theta \in \Theta, \tilde{a} \in \tilde{A}$. The available procedures are all the pairs $<N, \delta>$ where N is a stopping rule and δ is a Markov kernel from $\left(\underset{i=1}{\overset{m}{X}} X_i, B_{(N)}, F_\theta\right)$ to $\tilde{A}$. The loss incurred by an $<N, \delta>$ procedure which takes n observations and a terminal action $\tilde{a}$ is:

1 $$L(\theta, n, \tilde{a}) = c \cdot n + \tilde{L}(\theta, \tilde{a})\ c > 0.$$

We will now point out why this setting is a special case of the general setting described in Section 1. Define K_n the available actions at stage n to be $\tilde{A} \cup \{\{S\} \cup \{C\}\}$, i.e. we are adding two points to $\tilde{A}$. The set $A \subseteq \overset{m}{\underset{n=0}{X}} K_n$ of the available sequences consists of all sequences of the form $(C_1, C_2, \ldots, C_{n-1}, \tilde{a}_n, S_{n+1}, S_{n+2}, \ldots)$, here $C_i = C, S_{n+i} = S$ and $\tilde{a}_n \in \tilde{A}$. The meaning of the action C_i is to continue sampling and S_{n+i} are some trivial actions allowed after a terminal action $\tilde{a}_n$. Finally the loss:

$$L(\theta, C_1, \ldots, C_{n-1}, a_n, S_{n+1}, \ldots) = c \cdot n + \tilde{L}(\theta, \tilde{a}_n).$$

Definition 6: We will say that $(\times X_i, \mathcal{B}^X, F_\theta)$ is "sequentially sufficient for $(\times Y_i, \mathcal{B}^Y, G_\theta)$ for $<N, \delta>$ problems" if and only if for every action space $\tilde{A}$, loss function $L(\theta, n, a)$ of the form 1, and procedure $<N, \delta>$ depending on $\{Y_i\}$, there exists a procedure $<N', \delta'>$ depending on X_i such that:

$$E_\theta L(\theta, N', \delta') \leq E_\theta L(\theta, N, \delta)$$

for every θ.

Theorem 2: Assume $F_\theta^n \ll \nu^n \ \forall\, n$. Then $\{X_i\}$ is sufficient for $\{Y_i\}$ for $<N, \delta>$ problems, denoted $\{X_i\} \underset{<N,\delta>}{\supseteq} \{Y_i\}$ if and only if for any stopping rule N depending on $\{Y_i\}$, there exists a stopping rule N' depending on $\{X_i\}$ such that:

(i) $E_\theta(N') \leq E_\theta(N)$

(ii) $(\times X_i, B_{(N')}, F_\theta) \supseteq (\times X_i, B_{(N)}, G_\theta)$. (The sufficiency here is in the nonsequential sense.)

Proof: The idea of the proof: For a given stopping rule N depending on Y_i, take $\tilde{A}$ to be identical with $B_{(N)}$. Then consider the auxiliary experiment $(\times X_i, \mathcal{B}^X, F_{\theta \times L \times c})$ $\theta \in \Theta, L$ loss function, $c > 0$. Now similarly to Theorem 1, one can get the result (see [6]).

The following example will be of two sequential experiments, in which $\{X_i\}$ is not sequentially sufficient for $\{Y_i\}$, but it is sequentially sufficient for $<N, \delta>$ problems.

Example 1: The example involves a two stage sequential experiment and a parameter set consisting of two parameters θ_0 and θ_1.

The distribution of X_1, X_2:

Under θ_0 $X_1 \equiv 0,\ X_2 \equiv 0$

Under θ_1 $X_1 = 0,\ X_2 \equiv 1$

The distribution of Y_1, Y_2:

Under θ_0 $Y_1 \sim$ Bernoulli $(\frac{3}{8}),\ Y_2 \equiv 0$

Under θ_0 $Y_0 \sim$ Bernoulli $(\frac{5}{8}),\ Y_2 \equiv 0$

Obviously X_1, X_2 is not sequentially sufficient for Y_1, Y_2. In order to show X_1, X_2 is sequentially sufficient for Y_1, Y_2 for $< N, \delta >$ problems we have to show: For every stopping rule N depending on Y_1, Y_2, there exists, N' depending on X_1, X_2 such that: $(X_1 \times X_2, B_{(N')}, F_\theta) \supseteq (Y_1 \times Y_2, B_{(N)}, G_\theta)$ and such that $E_\theta(N) \supseteq E_\theta(N')$ for $\theta = \theta_0$ and $\theta = \theta_1$. It is easy to see that we need only to consider stopping rules N of the form: with probability $(1 - p)$ take no observations, with probability p take one observation. The corresponding N' will be: Take no observations with probability $1 - \frac{P}{2}$, take two observations with probability $\frac{P}{2}$.

The resulting experiment induced by N, has the following distribution under θ_0 and θ_1. (The sufficient statistics z for (Y_1, Y_N) is defined to be 2 when $N = 0$)

$$z \overset{\theta_0}{\sim} \begin{cases} 2 & 1-P \\ 0 & P \cdot \frac{5}{8} \\ 1 & P \cdot \frac{3}{8} \end{cases} \qquad z \overset{\theta_1}{\sim} \begin{cases} 2 & 1-P \\ 0 & P \cdot \frac{3}{8} \\ 1 & P \cdot \frac{5}{8} \end{cases}$$

The resulting experiment induced by N' (define z similarly).

$$z \overset{\theta_0}{\sim} \begin{cases} 2 & 1-\frac{p}{2} \\ 0 & \frac{p}{2} \\ 1 & 0 \end{cases} \qquad z \overset{\theta_1}{\sim} \begin{cases} 2 & 1-\frac{p}{2} \\ 0 & 0 \\ 1 & \frac{p}{2} \end{cases}$$

One can check the following:

(i) $E_\theta(N') = E_\theta(N)$

(ii) The experiment induced by N can be randomized from the experiment induced by N' with the Markov kernel: $\delta(2/2) = \frac{1-P}{1-\frac{P}{2}}, \delta(0/2) = \frac{P}{2(2-P)}, \delta(1/2) = \frac{P}{2(2-P)}$, $\delta(1/1) = \frac{6}{8}$, $\delta(0/1) = \frac{2}{8}$, $\delta(1/0) = \frac{2}{8}$, $\delta(0/0) = \frac{6}{8}$. i.e. $(X_1 \times X_2, B_{(N')}, F_\theta) \supseteq (Y_1 \times Y_2, B_{(N)}, G_\theta)$.

Section 3

In the two examples given in this section we will investigate the following: For two sequential experiments we consider the relation where $(X_1 \times \ldots \times X_n, \mathcal{B}^X_{(n)}, F_\theta) \supseteq (Y_1 \times \ldots \times Y_n, \mathcal{B}^Y_{(n)}, G_\theta)$ for every n. The first example will show 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 us for Theorem 1 and Theorem 2 in the next section.

Example 1: Let Y_1, Y_2 be independent r.v. with distribution:

Under $\theta_0 Y_1 \sim$ Bernoulli $(\frac{1}{3}), Y_2 \sim$ Bernoulli $(\frac{1}{4})$

Under $\theta_1 Y_1 \sim$ Bernoulli $(\frac{2}{3}), Y_2 \sim$ 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 where (X_1, X_2) is not sequentially sufficient for Y_1, Y_2. Suppose the cost of first observation is 0, and the cost of the 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:

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

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 the following are the only admissible nonrandomized procedures and their associated risks $r = (R(\theta_0, \delta), R(\theta_1, \delta))$.

δ^1: Decide θ_1 with probability one, $r_1 = (1, 0)$.

δ^2: Decide θ_0 with probability one, $r_2 = (0, a)$

δ^3: Decide θ_0 if $X_1 = 0, \theta_1$ if $X_1 = 1, r_3 = (\frac{1}{4}, \frac{1}{4}a)$

δ^4: Decide θ_0 if $X_1 = 0$, if $X_1 = 1$ observe X_2. Decide θ_i if $X_2 = i$. $r_4 = (\frac{1}{4}c + \frac{1}{12}, \frac{3}{4}c + \frac{1}{2}a)$

δ^5: Decide θ_1 if $X_1 = 0$, if $X_1 = 1$ observe X_2. Decide θ_i if $X_2 = i$. $r_5 = (\frac{3}{4}c + \frac{1}{2}, \frac{1}{4}c + \frac{1}{12}a)$.

Take $a = \frac{1}{10}$ and $c = \frac{1}{100}$. Then $r_1 = (1, 0), r_2 = (0, \frac{1}{10}), r_3 = (\frac{1}{4}, \frac{1}{40}), r_4 = (\frac{1}{400} + \frac{1}{120}, \frac{3}{400} + \frac{1}{12}), r_5 = (\frac{3}{400} + \frac{1}{2}, \frac{1}{400} + \frac{1}{120})$ and $r_0 = (R(\theta_0, \delta^0), R(\theta_1, \delta^0) = (\frac{1}{300} + \frac{1}{12}, \frac{2}{300} + \frac{1}{20})$

Figure 1: Risk points $r_0, \ldots, r_5$

In order to show the risk point r_0 cannot be achieved by a randomized procedure, it is enough to check that no $p, q \geq 0$ $p + q = 1$, exists such that:

$$p \cdot \frac{1}{4} + q\left(\frac{1}{400} + \frac{1}{12}\right) \leq \frac{1}{300} + \frac{1}{12}$$

$$p \cdot \frac{1}{40} + q\left(\frac{3}{400} + \frac{1}{20}\right) \leq \frac{2}{300} + \frac{1}{20}$$

which is indeed the case.

Remark 1: Giving a cost $c = \frac{1}{100}$ for the first observation does not change the result, it only adds a few more admissible procedures and corresponding computations. Hence we can conclude: $X_1 \supseteq Y_1$ and $(X_1, X_2) \supseteq (Y_1, Y_2)$ does not imply that $\{X_i\} \underset{<N,\delta>}{\supseteq} \{Y_i\}$.

Before starting the next example the following definitions are needed.

Definition 1: Let $(D, \mathcal{B}^D, F_\theta)$ and $(Y, \mathcal{B}^Y, G_\theta)$ be two experiments. The experiment consisting of two independent experiments Y and D is the following: Smaple space $D \times Y, \sigma$ algebra generated by $\mathcal{B}^D \times \mathcal{B}^Y$, and the product measure $F_\theta \times G_\theta$.

Definition 2: For two experiments X and $Y, X \approx Y$ iff $X \supseteq Y$ and $Y \supseteq X$.

Example 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 $X_1 + Z \sim N(\theta,2)$, i.e. 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 $X_1X_2 \underset{\text{seq}}{\supseteq} Y_1, Y_2$. $X_1 \approx (Y_1', D)$ 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: $X_1, X_2 \underset{\text{seq}}{\approx} (Y_1'D), X_2 \underset{\text{seq}}{\approx} (Y_1'D), Y_1$. Similarly $Y_1Y_2 \underset{\text{seq}}{\approx} Y_1, (Y_1'D)$. Hence it is enough to show $(Y_1'D), Y_1 \underset{\text{seq}}{\supseteq} Y_1, (Y_1'D)$. The latest 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, thus can do as good as an experimenter observing Y_1. In the second stage, the first experimenter observes D and Y_1, and can do as good as the second experimenter who observes (Y_1', D) at that stage.

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

Section 4

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

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

Theorem 1: Suppose S_n is boundly complete $n = 1, 2, \ldots$. Then $\{S_n\} \underset{\text{seq}}{\supseteq} \{Y_n\}$ if and only if $S_n \supseteq Y_{(n)}$ for every n.

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

Definition 1: Let $\{X_i\}$ and $\{Y_i\}$ be two sequential experiments. Let T_n be a sequence of Markov kernels from $(S_n, \mathcal{B}^{S_n})$ to $(Y_{(n)}, \mathcal{B}^Y_{(n)})$. The sequence will be called compatible if and only if:

$$(1) \qquad T_n(A_k|S_n) = E(T_k(A_k|S_k)|S_n)\ A_k \in \mathcal{B}^Y_k,$$

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

Lemma 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_m to $Y_{(m)}$, i.e. T_n satisfies $\int T_n(A_k|s_n)F_\theta(ds_n) = G_\theta(A_k), A_k \in \mathcal{B}'_{(k)}, n = 1, 2, \ldots$ (by completeness T_n is unique). Then the sequence T_n is compatible.

Proof: By assumption $\int T_n(A_k|s_n)F_\theta(ds_n) = G_\theta(A_k)$. Also:

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

Here $F(ds_k|s_m)$ is independent of θ by sufficiency. Now (1) follows from bounded completenss.

Lemma 2: Let $\{X_i\}$ and $\{Y_i\}$ be two sequential experiments. Let $\{S_n\}$ be a sequence of sufficient statistics for $X_1, \ldots, X_n, n = 1, 2, \ldots$. 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 $\{S_n\} \underset{\text{seq}}{\supseteq} \{Y_n\}$.

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 $Y_{(n-1)}$. We will postpone to

Lemma 3 the proof that $\delta_n(\cdot|\cdot,\cdot)$ satisfy the conditions in Definition 1.1. Here $y_n, \mathcal{B}_n^Y, \mathcal{B}_{(n)}^Y$ play the role of $a_n, \mathcal{A}_n, \mathcal{A}_{(n)}$ 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 $H_{\theta,\Delta}(ds_1, dy_1, ds_2, dy_2, \ldots)$. The marginal $H_{\theta,\Delta}(dy_1, \ldots, dy_n)$ was denoted $\mu_{\theta,\Delta}$ in Seciton 1. By Theorem 2.1a in order to establish the proof of this lemma, it is enough to show that $\mu_{\theta,\Delta} = G_\theta(dy_1, dy_2, \ldots)$. By Kolmogorov consistency it is enough to show for every n $G_\theta(dy_1, \ldots, dy_n) = \mu_{\theta,\Delta}(dy_1, \ldots, dy_n)$. Suppose we have shown for every $k < n$ that:

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

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

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

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:

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

Finally:

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

This completes the proof, except for Lemma 3 below.

Lemma 3: $\delta_n(dy_n|s_n, y_{(n-1)})$ satisfies the conditions in Definition 1.1, i.e.:

(i) $\Delta_n(\cdot|s_n, y_{(n-1)})$ is a probability measure on $\mathcal{B}_n^Y$.

(ii) $\delta_n(A|\cdot,\cdot)$ is $\mathcal{B}_{(n)}^S \times B_{(n-1)}^Y$ measurable, $A \in \mathcal{B}_n^Y$.

(iii) $\delta_n(A|s_n,\cdot)$ is $B_{(n-1)}^Y$ measurable.

Proof:

(i) is immediate from the existence of the conditional distribution for the distribution $T_n(dy_{(n)}|s_n)$ conditioning on $Y_{(n-1)}$.

(ii) Consider the following probability space: Sample space $S_n \times Y_{(n)}$, Borel field generated by $\mathcal{B}_n^{S_n} \times \mathcal{B}_n^Y$, and measure H_{θ,T_n} where:

$$H_{\theta,T_n}(A_1 \times A_2) = \int_{A_1} T_n(A_2|s_n)F_\theta(ds_n), A_1 \in \mathcal{B}_n^{S_n}, A_2 \in \mathcal{B}_{(n)}^Y.$$

By construction $T_n(A|s_n) = H_{\theta,T_n}(A|s_n)$ a.e. Hence there exists a version of the conditional distribution such that $H_{\theta,T_n}(\cdot|s_n) = T_n(A|s_n)$. Thus:

$$H_{\theta,T_n}(A|s_n, y_{(n-1)}) = T_n(A|s_n, y_{(n-1)}) = \delta_n(A|s_n, y_{(n-1)}) \text{ a.e.}$$

Now $\delta_n(A|\cdot,\cdot)$ is $B_{(n)}^{S_n} \times B_{(n-1)}^Y$ measurable, because $H_{\theta,T_n}(A|\cdot,\cdot)$ is.

(iii) This part follows because if $f(x,y)$ is measurable with respect to the Borel field $\mathcal{B}^X \times \mathcal{B}^Y$ then $f(x,\cdot)$ is measurable $\mathcal{B}^Y$.

Proof of Theorem 1: From Lemma 1 and Lemma 2 it is easy to conclude the proof. □

In the remaining of this section we will prove a factorization theorem.

Consider Example 2 in Section 3. We have shown that if $X \sim N(\theta, \sigma_1^2)$ and $Y \sim N(\theta, \sigma_2^2)$ $\sigma_1^2 \leq \sigma_2^2$ then $(X, Y) \underset{\text{seq}}{\supseteq} (Y, X)$. Now it can also be deduced from Theorem 1. Originally it was shown by a factorization of the experiment X, i.e. 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 next theorem of this section 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 $Y_2 = X_1$.

First some notations and other preliminaries. Suppose $(X_1, X_2) \underset{\text{seq}}{\supseteq} (Y_1, Y_2)$. Consider the experiment $\left((X_1, X_2, Y_1, Y_2), \mathcal{B}^{X_1 X_2 Y_1 Y_2}, H_{\theta\Delta}(dx_1 dx_2 dy_1 dy_2)\right)$ where $H_{\theta\Delta}$ is the measure induced by the relevant $\delta_1(dy_1|x_1)$ 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 $H_{\theta,\Delta}(dx_1, dx_2|Y_1 = y_2)$ and an experiment $\left((X_1, X_2), \mathcal{B}^{X_1, X_2}, H_{\theta\Delta}(dx_1, dx_2|y_1 = y_1)\right)$. Denote such an experiment as $(X_1, X_2|Y_1 = y_1)$.

Remark 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.

The Hellinger Transform

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

Definition 2: The measure valued functional

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

where $f_\theta(x) = \frac{dF_\theta(x)}{d\eta(x)}$ is the Hellinger transform of the experiment X.

The following can be shown (Strasser [11]).

(a) $X \supseteq Y$ implies $H_X(\lambda) \leq H_Y(\lambda)$ 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 [11]) is: For dominated families F_θ and G_θ $\theta \in \Theta$.

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

Lemma 4: Suppose $(X_1, X_2) \underset{\text{seq}}{\supseteq} (Y_1, Y_2)$. Then $(X_1, X_2 | Y_1 = y) \supseteq (Y_2 | Y_1 = y)$ for almost every y.

Proof: Let $A \in \mathcal{B}^{Y_2}$ then:

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

By sufficiency (Remark 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.

Theorem 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 $(X_1, X_2) \underset{\text{seq}}{\supseteq} (Y_1, Y_2)$ 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.

Proof: 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 4 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. Since (X_1, X_2) is sufficient for (X_1, X_2, Y_1) (Remark 1), by (b) $H_{(X_1,X_2)}(\lambda) = H_{(X_1,X_2,Y_1)}(\lambda)$ for every $\lambda \in \{\lambda\}$. Thus

$$H_{(X_1X_2)}(\lambda) = \int \underset{\theta\in\Theta}{\pi} h_\theta(x_1, x_2, y_1)^{\lambda(\theta)} d\eta(x_1, x_2, y_1)$$

where η is any measure dominating $H_{\theta,\Delta}(dx_1, dx_2, dy_1, dy_2)$ and $h_\theta = \frac{dH_{\theta,\Delta}}{d\eta}$. 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 $\Psi_\theta(y_1) = \int h_\theta 9x_1, x_2, y_1) d\eta(x_1, x_2|Y_1 = y_1)$. Notice that $\Psi_\theta(y_1) = \frac{dH_{\theta,\Delta}(y_1)}{d\eta(y_1)}$. $H_{(X_1,X_2)}(\lambda)$ can be written now as:

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

$$\int H_{(X_1,X_2|Y_1=y_1)(\lambda)} \underset{\theta\in\tilde{\Theta}}{\pi} \Psi_\theta^{\lambda(\theta)}(y_1) d\eta(y_1).$$

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 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:

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

(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 $\tilde{\Theta} = (\theta_1, \ldots, \theta_n)$, 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 Fubbini's theorem A has a positive measure, and using Fubbini's theorem again we deduce 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.

$$\begin{aligned} H_{(X_1)}(\lambda) &= H_{(X_1,Y_1)}(\lambda) \\ &= \int \underset{\theta \in \tilde{\Theta}}{\pi} h(x_1, y_1)^{\lambda(\theta)} d\eta(x_1|y_1) d\eta(y_1) \\ &= H_D(\lambda) \cdot H_{X_2}(\lambda) \end{aligned}$$

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

Section 5

In this section we will show how the theory is applied for experiments $(X, \mathcal{B}^X, F_\theta)$ when F_θ is an exponential family. The following is an immediate corollary of Theorem 4.1.

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

(i) There exists a sequence of sufficient statistics $X_{(n)}$ and $Y_{(n)}$ such that

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

(ii) Θ has non void interior. Then $\{X_i\} \underset{\text{seq}}{\supseteq} \{Y_i\}$ if and only if $X_{(n)} \supseteq Y_{(n)}$ for every n.

Proof: This is true because $X_{(n)}$ is complete and sufficient when Θ has non-void interior.

Example 1: Consider the linear experiments:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = X_1 \cdot \beta + \varepsilon_1 \text{ and } \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = X_2 \cdot \beta + \varepsilon_2 \quad \varepsilon_i \sim N(0, \sigma^2 I).$$

Here the unknown parameter $\theta = \beta$. Hansen and Torgerson [7], showed that $(Y_1, \ldots, Y_m) \supseteq (Z_1, \ldots, z_m)$ if and only if $((X_1' X_1) - (X_2' X_2))$ is positive semi-definite. If we consider $\{Y_i\}$ and $\{Z_i\}$ as sequential experiments, the conditions of corollary 1 are satisfied and we can deduce the following: $(Y_1, \ldots, Y_m) \underset{\text{seq}}{\supseteq} (Z_1, \ldots, Z_m)$ if and only if for every n $((X_1^{(n)'} X_1^{(n)}) - (X_2^{(n)'} X_2^{(n)}))$ is positive semi-definite. Here $X^{(n)}$ is the matrix consisting of the first n rows of X.

Another application gives a slight improvement of the following Theorem 1. Theorem 1 was proved independently by W. Ehnn and P.W. Müller [5] and by A. Janssen [8].

Theorem 1: Let $(X, \mathcal{B}^X, F_\theta)$ and $(Y, \mathcal{B}^Y, G_\theta)\theta \in \Theta$ be two experiments. Suppose F_θ and G_θ

are exponential families and Θ has non-void 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 $(Y, \mathcal{B}^Y, G_\theta)$ and $(D, \mathcal{B}^D, K_\theta)$.

(ii) K_θ is an exponential family.

Theorem 1a: The conclusion of Theorem 1 remains valid if we replace the condition that Θ has non-void interior by the (weaker) condition that $X+Y$ is boundedly complete when X, Y are independent.

Proof: (i) Consider the two stages sequential experiments X, Y and Y, X where X and Y are independent. By Theorem 4.1 $(X, Y) \underset{\text{seq}}{\supseteq} (Y, X)$, hence by Theorem 4.2 $X \approx (Y, D)$.

(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 4.2 $D \approx (X|Y = y_0)$. Denote $\omega(dx, dy) = \delta(dy|x)\mu(dx)$. Then:

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

Remark 1: Theorem 1a is true also if $Y = (Y_1, \ldots, Y_n)$ each $Y_i \in R^k$ is a canonical observation from an exponential family.

Monotonicity of Bayes Sequential Tests. Another application of the theory involves generalizing Sobel's result [10] about monotonicity of Bayes sequential tests as stated in the following Theorem 2.

Let $\{X_i\}$ be a sequence of r.v. distributed $F_\theta(dx_1, dx_2, \ldots)$. Let S_n be a sequence of sufficient statistics distributed $F_\theta^n(ds_n), \theta \in R$. Let $H_0 : \theta \leq \theta_0, H_1 : \theta > \theta_0$. Consider an

$<N, \delta>$ setting in which the space of terminal actions consists of two actions H_0 and H_1. Assume loss function $L(\theta, H_i, n) = c \cdot n + \tilde{L}(\theta, H_i)$, where $\tilde{L}(\theta, H_0)$ is non decreasing and $\tilde{L}(\theta, H_1)$ is non increasing as a function of θ.

Definition 1: An $<N, \delta>$ procedure is called monotone, if at the n'th stage there exists a_n and b_n such that the procedure decides H_0 if $S_n < a_n, H_1$ if $S_n > b_n$, takes at least one more observation if $S_n \in (a_n, b_n)$. The procedure either takes one more observation or decides H_0 if $S_n = a_n$. The procedure either takes one more observation or decides H_1 if $S_n = b_n$.

Theorem 2: (Sobel [10]) Let $Y_n, n = 1, \ldots, m \leq \infty$ be a sequence of independent real valued random variables with distribution $F_\theta^n(dy_n), F_\theta^n$ is an exponential family. $S_n = \sum_{i=1}^{n} Y_i$ a sufficient statistics. Then for the testing problem described above every Bayes procedure is monotone.

Theorem 2a: Replace the condition Y_i are independent by the condition: There exists a sequence of sufficient statistics $S_1, S_2, \ldots$ such that the distributions $F_\theta^n(ds_n)$ are an exponential family with S_n being the canonical observations. Assume further that the canonical parameter set Θ has a non-void interior. Then the conclusion of Theorem 2 remains valid.

Example 2: Let $(Y_1, \ldots, Y_m)$ be a multivariate normal distributed vector with $EY_i = \theta, i = 1, \ldots, m$, and covariance matrix $\Sigma \neq I.\theta$ unknown, Σ known.

Example 3: Let $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ be a multivariate normal distributed vector with $E\underset{\sim}{Y} = \underset{\sim}{\theta}$ and covariance matrix $\sigma^2 A, \sigma^2$ is an unknown real valued parameter, $\underset{\sim}{\theta}$ and A are known.

Examples 2 and 3 are not covered by Sobel's theorem but are covered by Theorem 2a.

Comment: It can be shown that the cases covered by Theorem 2a are already covered by

Brown Cohen and Strawderman [4]. The consideration involve arguments similar to the following Lemmas 1 and 2, and a further argument about transitivity.

Lemma 1: Under the conditions of Theorem 2a:

(i) There exists a sequence of independent random variables $X_1, \ldots, X_m$ such that the distribution of X_i under θ $F_\theta(dx_i)$, is an exponential family with X_i canonical observation, and $(X_1, \ldots, X_m) \underset{\text{seq}}{\approx} (Y_1, \ldots, Y_m)$

(ii) $(X_1, \ldots, X_m) \underset{\text{seq}}{\approx} (S_1, \ldots, S_m)$.

Proof: We will construct the sequence $\{X_i\}$.$X_1 = S_1$. Suppose $X_1, \ldots, X_{n-1}$ are already defined and $(X_1, \ldots, X_{n-1}) \approx S_{n-1} \subseteq S_n$. By Remark 1 on Theorem 1a there exists an experiment D such that D is independent of $(X_1, \ldots, X_{n-1})$ and $((X_1, \ldots, X_{n-1}), D) \approx S_n$. Denote $X_n = D$. By Theorem 1a D may be taken as a canonical observation from an exponential family with canonical parameter set Θ.

Now since $(X_1, \ldots, X_n) \approx (S_1, \ldots, S_n)$ for every n, we conclude by Theorem 4.1 $\{X_i\} \underset{\text{seq}}{\approx} \{S_i\} \underset{\text{seq}}{\approx} \{Y_i\}$.

Lemma 2: Suppose $(X, B^X, F_\theta) \approx (Y, B^Y, G_\theta)$ $\theta \in \Theta$. Assume

$$F_\theta(dx) = \exp(\theta \cdot x - \Psi(\theta))\mu(dx) \text{ and } G_\theta(dy) = \exp(\theta \cdot y - \rho(\theta))\nu(dy),$$

F_θ and G_θ are minimal exponential families. Then there exist a constant C such that:

$$\exp(\theta \cdot x - \Psi(\theta) = \exp(\theta \cdot (x + C) - \rho(\theta)).$$

Proof: It can be shown [11], that $X \approx Y$ implies for $\theta_1, \ldots, \theta_n \{\exp(\theta_i \cdot x - \Psi(\theta_i)\}_{i=1}^n$ has the same distribution as $\{\exp(\theta_i \cdot y - \Psi(\theta_i)\}_{i=1}^n$ when $X \sim \mu$ and $Y \sim \nu$. Hence both distributions have the same support and for x in the support there exists $y(x)$ such that

$\theta_i \cdot x - \Psi(\theta_i) = \theta_i y(x) - \rho(\theta_i), i = 1, \ldots, n$ or $\Psi(\theta_i) = \rho(\theta_i) + \theta_i(x - y(x))$. We can conclude now $x - y(x) = C$.

Proof of Theorem 2a: Let $H_{\theta,\Delta}(s_1, \ldots, s_m, x_1, \ldots, x_m)$ be the measure induced by $F^i_\theta(ds_i)$ and the relevant $\delta_1(x_1|s_1), \delta_2(dx_2|x_1, s_1, s_2) \ldots$ for sequences S_i and X_i as in Lemma 1. Let $d\tilde{H}_{\theta,\Delta}(s_k, \sum_{i=1}^{k} x_i)$ be the joint distribution of S_k and $\sum_{i=1}^{k} X_i$ under $H_{\theta,\Delta}$. By Lemma 2 $S_k + C_k = \sum_{i=1}^{k} X_i$ for some constant C_k, because $S_k \approx \sum_{i=1}^{k} X_i$.

The risk of a Bayes procedure based on $\{X_i\}$ is the same as the Bayes risk of a procedure based on $\{S_i\}$. By Sobel's result at stage n, a Bayes procedure based on $\{X_i\}$ keeps sampling if $a_n < \sum_{i=1}^{n} X_i < b_n$, which is equivalent to $a_n - C_n < S_n < b_n - C_n$. Similarly for other actions.

Acknowledgement

This work is part of the author's Ph.D. thesis at Cornell University, 1990. The author would like to thank his thesis advisor, Professor L.D. Brown, for his most valuable help and advice.

References

[1] Blackwell, D. (1953). Equivalent comparison of experiments. *Ann. Math. Stat.*, 24, 265–272.

[2] Bohnenblust, H.F., Shapley, L.S.; and S. Sherman. Unpublished.

[3] Brown, L.D. (1977). Closure theorems for sequential design processes. In *Statistical Decision Theory and Related Topics* II. 57–91, Academic Press.

[4] Brown, L.D., Cohen, A., and W.E. Strawderman (1979). Monotonicity of Bayes sequential tests. *Ann. Stat.*, 1222–1230.

[5] Ehnn, W. and D.W. Müller (1983). Factorizing the information contained in an experiment. *Wahrsch*, **65**, 121–134.

[6] Greenshtein, E., Ph.D. Thesis. Cornell University, 1990.

[7] Hansen, O.H. and E.N. Torgerson (1974). Comparison of linear normal experiments. *Ann. Stat.*, **2**, 367–373.

[8] Janssen, Arnold (1988). Unpublished.

[9] LeCam, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Stat.*, **35**, 1419–1455.

[10] Sobel, M. (1952). An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Stat.*, **23**, 319–337.

[11] Strasser, H. (1985). *Mathematical Theory of Statistics.* Berlin, New York, W. de Gruyter.

[12] Torgerson, E.N. (1976). Comparison of statistical experiments. *Scan. J. Stat.*, 182–208.