SEQUENTIAL METHODS IN STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 16
PWN–POLISH SCIENTIFIC PUBLISHERS
WARSAW 1985

SUFFICIENCY IN SEQUENTIAL EXPERIMENTAL DESIGN PROBLEMS

H. HECKENDORFF

Section of Mathematics, Technological University, Karl-Marx-Stadt, G.D.R.

In the paper the principles of sufficiency and transitivity will be applied to sequentially designed experiments. By generalization of the usual sufficiency concept to the present case there are conditions that allow the choice of design rules, stopping rules and terminal action rules on the basis of sufficient and transitive sequences of statistics or σ-fields. These results are generalizations of corresponding statements of Gray ([5]), who has studied only the dominated case. Moreover, the approach of Gray to the definition of transitivity is changed in some sense.

1. Introduction

Sequential statistical procedures are characterized by the parallel performance of observation and statistical inference, so that the observations up to time n will influence further sampling. In the simplest case this influence consists in the decision of stopping or continuation at time n. In the sequential design case from the continuation decision it follows the necessity to design the observations for the next step in dependence on the past observation results.

In view of numerical realization of such a sequential procedure it is useful to have the possibility to store not all past data, but only a set of current values. This can be achieved by application of simplified processing algorithms (such as stochastic approximation procedures) or by using sufficient sequences of statistics or σ-fields, allowing the data reduction without loss of information.

2. Experimental design rules

The mathematical model of a real experiment is the statistical experiment. That is a triplet $(\mathscr{X}, \mathfrak{A}, \mathscr{P})$, consisting of the sample space $\mathscr{X}$, of a chosen σ-

field $\mathfrak{A}$ of events observable on principle, and of a family $\mathscr{P}$ of probability measures. Without loss of generality $\mathscr{P}$ can be represented in the form $\mathscr{P} = \{P_\theta, \theta \in \Theta\}$, Θ being a parameter set, not necessary in a Euclidean space. Experimental design arises if, in a statistical problem, there is the possibility to have different statistical experiments. It will be assumed that the results of these experiments are in the same measurable space, and the experiments differ only in the family $\mathscr{P}$, that is, instead of $\mathscr{P}$ there will be a family $\mathscr{P}^e$, $\mathscr{P}^e = \{P_\theta^e, \theta \in \Theta\}$, where e is a parameter, chosen in a measurable space $(\mathscr{E}, \mathfrak{E})$. The parameter e, $e \in \mathscr{E}$, is chosen before the experiment will be started, and the experiment consists in a realization of a random element X in $(\mathscr{X}, \mathfrak{A}, \mathscr{P}^e)$. This scheme includes the design situation in problems of regression analysis, analysis of variance, control problems, etc.

The element e, $e \in \mathscr{E}$, is called the *experimental design*, and the choice of e results by a design procedure.

In the following, randomized designs will also be taken into consideration. A randomized design π will be a probability measure on $(\mathscr{E}, \mathfrak{E})$, and in application the design π is realized by realization of a random element V with distribution π on $(\mathscr{E}, \mathfrak{E})$, giving an element e, $e \in \mathscr{E}$. With this e the element X in $(\mathscr{X}, \mathfrak{A}, \mathscr{P}^e)$ will be realized.

For sequential design it must be assumed that the whole experiment can be performed by fractions, so that the whole result consists of a sequence of fractional results. In this sense a sequentially performed experiment is the realization of a sequence $V_1, X_1, V_2, X_2, \ldots$, in which X_n at stage n is a random element with realizations in $(\mathscr{X}_n, \mathfrak{A}_n)$, V_n is a random element with realizations in $(\mathscr{E}_n, \mathfrak{E}_n)$, $n \in \mathbf{N}$.

For all further consideration the following assumption should be fulfilled.

General assumption. $(\mathscr{X}_n, \mathfrak{A}_n)$, $(E_n \mathfrak{F}_n)$ $n \in \mathbf{N}$, are Borelian spaces, that is, measure-isomorphic spaces to a complete separable metric space with the σ-field of the Borel sets. Let

$$H_n = (V_1, X_1, \ldots, V_n, X_n).$$

$$\mathscr{G}_n = \mathscr{E}_1 \times \mathscr{X}_1 \times \ldots \times \mathscr{E}_n \times \mathscr{X}_n, \quad n \in \mathbf{N}.$$

$$\mathfrak{F}_n \sim \mathfrak{E}_1 \otimes \mathfrak{A}_1 \otimes \ldots \otimes \mathfrak{E}_n \otimes \mathfrak{A}_n.$$

Then the following definition can be given.

DEFINITION 1. A *sequential design rule* ϱ is a sequence $\{\pi_n\}_{n \in \mathbf{N}}$ in which $\pi_1(\cdot)$ is a probability measure on $(\mathscr{E}_1, \mathfrak{E}_1)$, $\pi_n(\cdot; h_{n-1})$, $n > 1$, are transition probabilities from $(\mathscr{G}_{n-1}, \mathfrak{F}_{n-1})$ into $(\mathscr{E}_n, \mathfrak{E}_n)$.

That is, $\pi_1(\cdot)$ is the distribution of V_1, $\pi_n(\cdot; \cdot)$ is the distribution of V_n depending on h_{n-1}, the realization of H_{n-1}. (The general assumption of Borelian spaces to be used ensures in further the existence of conditional

probability distributions.) For $X_1, X_2, \ldots$ a distribution model will be assumed in the form that X_{n+1} has a distribution $P_\theta^{(n+1)}(\cdot; h_n, v_{n+1})$, $\theta \in \Theta$, depending on the realization v_{n+1} of the design V_{n+1} and, perhaps, on the realization h_n of H_n. This distribution may be thought as a transition probability from $(\mathcal{G}_n \times \mathcal{E}_{n+1}, \mathfrak{F}_n \times \mathfrak{E}_{n+1})$ into $(\mathcal{X}_{n+1}, \mathfrak{A}_{n+1})$.

With a design rule ϱ and a distribution model the whole statistical experiment is fixed. It may be set:

$$\mathcal{X} = \mathcal{E}_1 \times \mathcal{X}_1 \times \ldots,$$
$$\mathfrak{A} = \mathfrak{E}_1 \otimes \mathfrak{A}_1 \otimes \ldots,$$
$$\mathcal{P}^\varrho = \{P_\theta^\varrho, \theta \in \Theta\},$$

where the P_θ^ϱ are given by the finite-dimensional distributions $P_{\theta,n}^\varrho$ on $(\mathcal{G}_n, \mathfrak{F}_n)$ with

$$P_{\theta,n}^\varrho = \pi_1 \times P_\theta^{(1)} \times \ldots \times \pi_n \times P_\theta^{(n)}$$

in the sense of the Ionescu-Tulcea theorem.

It will be noted that the statistical experiment includes the design results, the realizations of $V_1, V_2, \ldots$.

Remark 1. A design rule ϱ, for which the π_n, $n > 1$, are transition probabilities only from $(\mathcal{E}_1 \times \ldots \times \mathcal{E}_{n-1}, \mathfrak{E}_1 \otimes \ldots \otimes \mathfrak{E}_{n-1})$ into $(\mathcal{E}_n, \mathfrak{E}_n)$, independently of $x_1, \ldots, x_{n-1}$, is called an *iterative design rule*. There are papers in which sequential design is limited by iterative design rules, for instance [3].

Remark 2. A *nonrandomized design rule* $\varkappa$ is a sequence $\{\varkappa_n\}_{n \in N}$, in which $\varkappa_1$ is a fixed element e_1 of E_1, $\varkappa_n$, $n > 1$, is a measurable function from $(\mathcal{G}_{n-1}, \mathfrak{F}_{n-1})$ into $(\mathcal{E}_n, \mathfrak{E}_n)$. By the assumption of Borelian spaces it is ensured that nonrandomized design rules may be considered as special sequential rules with distributions π_n degenerate at points $\varkappa_n(h_{n-1})$. (On Borelian spaces the single points are Borel sets.)

Remark 3. If the $P_\theta^{(n)}(\cdot; h_{n-1}, v_n)$, $n \geqslant 1$, for any $v_1, x_1, \ldots$ are dominated by σ-finite measures $\mu^{(n)}$ on $(\mathcal{X}_n, \mathfrak{A}_n)$ with densities $f_\theta^{(n)}(\cdot; h_{n-1}, v_n)$, $n \geqslant 1$, then $P_{\theta,n}^\varrho$ for any design rule ϱ is dominated by the measure

$$\mu_n^\varrho = \pi_1 \times \mu^{(1)} \times \ldots \times \pi_n \times \mu^{(n)}, \tag{1}$$

and the density (likelihood function) is

$$f_{\theta,n}(h_n) = f_\theta^{(1)}(x_1; v_1) \prod_{j=2}^{n} f_\theta^{(j)}(x_j; h_{j-1}, v_j). \tag{2}$$

This follows from the general Fubini theorem.

If the π_n, $n \geqslant 1$, are also dominated with densities $p_1(\cdot)$, $p_n(\cdot; h_{n-1})$, then

$$f_{\theta,n}^\varrho(h_n) = f_\theta^{(1)}(x_1; v_1) \prod_{j=2}^{n} f_\theta^{(j)}(x_j; h_{j-1}, v_j) p_1(v_1) \prod_{j=2}^{n} p_j(v_j; h_{j-1}). \tag{3}$$

In the latter case the rule ϱ could be called a *dominated design rule.* Gray ([5]) deals with this very case.

3. Sample space

For reasons of realization, a statistical experiment must be specified by a sample size.

To simplify further denotations all introduced σ-fields will be assumed to be cylindric sub-σ-fields of $\mathfrak{A}$ on $\mathfrak{X}$. Then $\mathfrak{F}_n$, $n \geqslant 1$, with

$$\mathfrak{F}_n = \mathfrak{E}_1 \vee \mathfrak{A}_1 \vee \ldots \vee \mathfrak{E}_n \vee \mathfrak{A}_n \tag{4}$$

being the smallest σ-field containing $\mathfrak{E}_1$, $\mathfrak{A}_1$, ..., $\mathfrak{E}_n$, $\mathfrak{A}_n$.

The sample size may be a fixed or a random number.

(a) Fixed sample size n. In this case the sequentially designed experiment is the realization of the sequence $(V_1, X_1, \ldots, V_n, X_n)$ of fixed size n, and the sample space becomes $(\mathfrak{X}, \mathfrak{F}_n, \mathscr{P}_n^{\varrho})$.

(b) Random sample size τ. The sequence $\{\mathfrak{F}_n\}_{n \in N}$ is a sequence of sub-σ-fields of $\mathfrak{A}$, which is nondecreasing. Further, let $\mathfrak{F}_\infty = \sigma(\mathfrak{F}_n, n \in N)$. Then a random sample size may be introduced in relation to the sequence $\{\mathfrak{F}_n\}_{n \in N}$ by the following definition.

DEFINITION 2. 1. A *Markov sample size* τ is a Markov time in relation to $\{\mathfrak{F}_n\}_{n \in N}$, that is a random variable on $(\mathfrak{X}, \mathfrak{A})$ with realizations in $\bar{N}$ and with $\{\tau = n\} \in \mathfrak{F}_n$ for every $n \in \bar{N}$.

2. A *randomized sample size* τ is a random variable on $(\mathfrak{X}, \mathfrak{A})$ with realizations in $\bar{N}$ with

$$\begin{aligned} &\text{(a)} \quad P_\theta^{\varrho}(\{\tau = n\} \mid \mathfrak{F}_n) = P_\theta^{\varrho}(\{\tau = n\} \mid \mathfrak{F}_\infty) \qquad \text{a.e.,} \quad \theta \in \Theta, \\ &\text{(b)} \quad P_\theta^{\varrho}(\{\tau = n\} \mid \mathfrak{F}_n) \text{ has a version independent of } \theta, \ \theta \in \Theta. \end{aligned} \tag{5}$$

The requirement (a) ensures that $\{\tau = n\}$ has fixed probability under $\mathfrak{F}_n$, which is not changed by further observation, in other words, $\{\tau = n\}$ and $\mathfrak{F}_\infty$ are conditionally independent under $\mathfrak{F}_n$. The requirement (b) allows the statistical realization of τ without knowledge about the parameter θ.

A randomized sample size τ can be realized by a stopping rule $\{\varphi_n\}_{n \in N}$ in the sense of ([1]), where φ_n, $n \in N$, are $\mathfrak{F}_n$-measurable functions on $\mathfrak{X}$ with values in $[0, 1]$. For a given τ it may be set

$$\varphi_n(x) = \begin{cases} 0, & \text{if} \quad P_\theta^{\varrho}(\{\tau = n\} \mid \mathfrak{F}_n) = 0, \\ \dfrac{P_\theta^{\varrho}(\{\tau = n\} \mid \mathfrak{F}_n)}{P_\theta^{\varrho}(\{\tau \leqslant n\} \mid \mathfrak{F}_n)}(x), & \text{otherwise.} \end{cases} \tag{6}$$

(By definition 2 the conditional probabilities in (6) may be chosen independent of θ, $\theta \in \Theta$.)

On the other hand, every stopping rule $\{\varphi_n\}_{n\in N}$ determines a randomized sample size τ.

With the sample size τ the sample space becomes $(\mathcal{X}, \mathfrak{F}_\tau, \mathcal{P}_\tau^\varrho)$, where

$$\mathfrak{F}_\tau = \sigma(\mathfrak{F}_n \cap \{\tau = n\}, n\in \bar{N})$$

is the σ-field of observable events and $\mathcal{P}_\tau^\varrho$ is the family of measures $P_{\theta,\tau}^\varrho$ restricted from $\mathfrak{A}$ to $\mathfrak{F}_\tau$.

4. The concept of sufficiency

For reasons of sequential design the usual definition of sufficiency must be slightly generalized.

DEFINITION 3. A sub-sequence $\{\mathfrak{G}_n\}_{n\in N}$ of σ-fields on $(\mathcal{X}, \mathfrak{A}, \mathcal{P}^\varrho)$ is *sufficient* for $\{\mathfrak{F}_n\}_{n\in N}$, if

(1) $\mathfrak{G}_n \subseteq \mathfrak{F}_n$, $n\in N$,

(2) for every F_n, $F_n \in \mathfrak{F}_n$, and for any design rule ϱ for the conditional probability $P_\theta^\varrho(F_n \mid \mathfrak{G}_n)$ there is a version $P_\cdot^\varrho(F_n \mid \mathfrak{G}_n)$ independent of θ, $\theta\in\Theta$, $n\in N$.

A sequence $\{T_n\}_{n\in N}$ of statistics on $(\mathcal{X}, \mathfrak{A}, \mathcal{P}^\varrho)$ is *sufficient* for $\{(V_1, X_1, \ldots, V_n, X_n)\}_{n\in N}$, if T_n is a $\mathfrak{F}_n$-measurable function and $\{\sigma(T_n)\}_{n\in N}$ satisfies (2), $n\in N$. From this definition the following question arises. If the sequence $\{\mathfrak{G}_n\}_{n\in\bar{N}}$ is sufficient for $\{\mathfrak{F}_n\}_{n\in N}$, τ is a randomized sample size, will $\mathfrak{G}_\tau$, $\mathfrak{G}_\tau = \sigma(\{\tau = n\} \cap \mathfrak{G}_n, n\in\bar{N})$, then be sufficient for $\mathfrak{F}_\tau$? The answer will be positive (see [1], [2], [6]). But the analogous statement is not true, if "sufficient" is replaced by "minimal-sufficient".

In accordance with the nondesigned experiment case the definition of transitivity may be given in the following way.

DEFINITION 4. A sequence $\{\mathfrak{G}_n\}_{n\in N}$ of σ-fields on $(\mathcal{X}, \mathfrak{A}, \mathcal{P}^\varrho)$ is *transitive* in relation to $\{\mathfrak{F}_n\}_{n\in N}$, if

(1) $\mathfrak{G}_n \subseteq \mathfrak{F}_n$, $n\in N$,

(2) for any design rule ϱ the σ-fields $\mathfrak{G}_{n+1}$ and $\mathfrak{F}_n \vee \mathfrak{E}_{n+1}$ are conditionally independent under $\mathfrak{G}_n \vee \mathfrak{E}_{n+1}$, that is

$$P_\theta^\varrho(G_{n+1} \mid \mathfrak{F}_n \vee \mathfrak{E}_{n+1}) = P_\theta^\varrho(G_{n+1} \mid \mathfrak{G}_n \vee \mathfrak{E}_{n+1}) \quad \text{a.e.,}$$

$$G_{n+1}\in\mathfrak{G}_{n+1},\ \theta\in\Theta,\ n\in N. \tag{7}$$

A sequence $\{T_n\}_{n\in N}$ of statistics on $(\mathcal{X}, \mathfrak{A}, \mathcal{P}^\varrho)$ is *transitive* in relation to $\{\mathfrak{F}_n\}_{n\in N}$, if $\{\mathfrak{G}_n\}_{n\in N}$ with $\mathfrak{G}_n = \sigma(T_n)$, $n\in N$, is transitive in relation to $\{\mathfrak{F}_n\}_{n\in N}$.

Remark. Definition 4 substantially differs from the corresponding POLS-definition (policy sufficiency) of [5]. His POLS-definition, different in the fixed-sample and random-sample cases, is difficult to understand.

Now the following theorem may be proved.

THEOREM. *Let* $(\mathcal{X}, \mathfrak{F}_\tau, \mathscr{P}_\tau^\varrho)$ *be the sample space to the design rule* ϱ *and the randomized sample size* τ, *given by a stopping rule* $\{\varphi_n\}_{n\in\mathbb{N}}$ *in relation to* $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$, *with* $\tau < \infty$. *If the sequence* $\{\mathfrak{G}_n\}_{n\in\mathbb{N}}$ *of* σ*-fields on* $(\mathcal{X}, \mathfrak{A}, \mathscr{P}^\varrho)$ *is sufficient for* $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ *and transitive in relation to* $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$, *then the design rule* ϱ *and the stopping rule* $\{\varphi_n\}_{n\in\mathbb{N}}$ *may be replaced by rules* $\tilde{\varrho}$ *and* $\{\tilde{\varphi}_n\}_{n\in\mathbb{N}}$ *on the basis of the sequence* $\{\mathfrak{G}_n\}_{n\in\mathbb{N}}$, *so that* $P_{\theta,\tau}^{\varrho}$ *and* $P_{\theta,\tilde{\tau}}^{\tilde{\varrho}}$ *will coincide on* $\mathfrak{G}_\tau$.

($\tilde{\varrho} = \{\tilde{\pi}_n\}_{n\in\mathbb{N}}$, $\{\tilde{\varphi}_n\}_{n\in\mathbb{N}}$ based on $\{\mathfrak{G}_n\}_{n\in\mathbb{N}}$ means that the $\tilde{\pi}_n$ are transition probabilities from $\mathfrak{G}_{n-1}$ to $\mathfrak{E}_n$ and the $\tilde{\varphi}_n$ are $\mathfrak{G}_n$-measurable functions.)

Proof. The proof uses some aspects associated with designs in addition to the usual conditional expectation technique.

1° If $\varrho = \{\pi_n\}_{n\in\mathbb{N}}$ is any design rule in relation to $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$, then by the sufficiency of $\{\mathfrak{G}_n\}_{n\in\mathbb{N}}$ for $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ the $\mathfrak{F}_n$-measurable functions $\pi_{n+1}(B_{n+1}; \cdot)$, $B_{n+1}\in\mathfrak{E}_{n+1}$, $n \geqslant 1$, may be replaced by $\mathfrak{G}_n$-measurable functions $\tilde{\pi}_{n+1}(B_{n+1}; \cdot)$ with

$$\tilde{\pi}_{n+1}(B_{n+1}; h_n) = E_{P_\theta^\varrho}(\pi_{n+1}(B_{n+1}; H_n)\,|\,\mathfrak{G}_n)(h_n). \tag{8}$$

From the assumption of Borelian spaces it follows that the sequence $\tilde{\pi}_1 = \pi_1$, $\tilde{\pi}_2, \ldots$ defines a design rule $\tilde{\varrho}$. The $\tilde{\pi}_{n+1}$ in equation (8) may depend on π_1, $\pi_2, \ldots$, but they are independent of θ, $\theta\in\Theta$, and $\mathfrak{G}_n$-measurable functions of h_n, $h_n\in\mathscr{G}_n$.

Now it will be chosen that for any $(\mathfrak{G}_n \vee \mathfrak{E}_{n+1})$-measurable function $g(h_n, v_{n+1})$ on $\mathscr{H}_n \times \mathcal{E}_{n+1}$ the equation

$$\int_{\mathcal{X}} g(h_n, v_{n+1})\,\tilde{\pi}_{n+1}(dv_{n+1}; h_n) = E_{P_{\theta,n}^\varrho}\Big[\int_{\mathcal{X}} g(H_n, v_{n+1})\,\pi_{n+1}(dv_{n+1}; H_n)\,|\,\mathfrak{G}_n\Big](h_n) \tag{9}$$

holds. It suffices to prove (9) for functions of the form

$$g(h_n, v_{n+1}) = \chi_{G_n}(h_n)\,\chi_{B_{n+1}}(v_{n+1})$$

with $G_n\in\mathfrak{G}_n$, $B_{n+1}\in\mathfrak{E}_{n+1}$. By definition of $\tilde{\pi}_{n+1}$ it is

$$\begin{aligned} E_{P_{\theta,n}^\varrho}\Big[\int_{\mathcal{X}} \chi_{G_n}(H_n)\,\chi_{B_{n+1}}(v_{n+1})\,\pi_{n+1}(dv_{n+1}; H_n)\,|\,\mathfrak{G}_n\Big](h_n) \\ &= E_{P_{\theta,n}^\varrho}[\chi_{G_n}(H_n)\,\pi_{n+1}(B_{n+1}; H_n)\,|\,\mathfrak{G}_n](h_n) \\ &= \chi_{G_n}(h_n)\,E_{P_{\theta,n}^\varrho}[\pi_{n+1}(B_{n+1}; H_n)\,|\,\mathfrak{G}_n](h_n) \\ &= \chi_{G_n}(h_n)\,\tilde{\pi}_{n+1}(B_{n+1}; h_n) \\ &= \int_{\mathcal{X}} \chi_{G_n}(h_n)\,\chi_{B_{n+1}}(v_{n+1})\,\tilde{\pi}_{n+1}(dv_{n+1}; h_n), \end{aligned}$$

so that (9) holds.

2° Using equation (9) it will be proved that the measures P_θ^ϱ and $P_\theta^{\tilde\varrho}$ coincide on $\mathfrak{G}_n$, $n\in N$, that is

$$P_{\theta,n}^\varrho(G_n) = P_{\theta,n}^{\tilde\varrho}(G_n), \qquad G_n\in\mathfrak{G}_n. \tag{10}$$

This will be achieved by induction. For $n = 1$ it is $\pi_1 = \tilde\pi_1$, and (10) trivially holds.

If (10) holds for a n, $n \geqslant 1$, then using properties of the conditional expectation and equation (9)

$$\begin{aligned}
P_{\theta,n+1}^\varrho(G_{n+1}) &= \int_{\mathcal{G}_{n+1}} \chi_{G_{n+1}}(h_{n+1})\, P_{\theta,n+1}^\varrho(dh_{n+1}) \\
&= \int_{\mathcal{G}_n}\int_{\mathcal{E}_{n+1}}\int_{\mathcal{X}_{n+1}} \chi_{G_{n+1}}(h_{n+1})\, P_\theta^{(n+1)}(dx_{n+1};\, h_n,\, v_{n+1})\times \\
&\qquad\qquad \times \pi_{n+1}(dv_{n+1};\, h_n)\, P_{\theta,n}^\varrho(dh_n) \\
&= \int_{\mathcal{G}_{n+1}} E_{P_{\theta,n}^\varrho}\Big[\int_{\mathcal{E}_{n+1}}\int_{\mathcal{X}_{n+1}} \chi_{G_{n+1}}(H_n;\, v_{n+1},\, x_{n+1}) \\
&\qquad \times P_\theta^{(n+1)}(dx_{n+1};\, H_n,\, v_{n+1})\,\pi_{n+1}(dv_{n+1};\, H_n)\,\Big|\,\mathfrak{G}_n\Big]\, P_{\theta,n}^\varrho(dh_n) \\
&= \int_{\mathcal{G}_n}\int_{\mathcal{E}_{n+1}}\int_{\mathcal{X}_{n+1}} \chi_{G_{n+1}}(h_{n+1})\, P_\theta^{(n+1)}(dx_{n+1};\, h_n,\, v_{n+1})\times \\
&\qquad\qquad \times \tilde\pi_{n+1}(dv_{n+1};\, h_n)\, P_{\theta,n}^\varrho(dh_n).
\end{aligned}$$

In the latter integral the integration relative to $P_{\theta,n}^\varrho$ may be replaced by integration relative to $P_{\theta,n}^{\tilde\varrho}$. (The integrand is $\mathfrak{G}_n$-measurable.) Thus, it follows

$$P_{\theta,n+1}^\varrho(G_{n+1}) = P_{\theta,n+1}^{\tilde\varrho}(G_{n+1}), \qquad G_{n+1}\in\mathfrak{G}_{n+1}.$$

3° To finish the proof it suffices to show that

$$P_\theta^{\tilde\varrho}(G_{n+1}\mid\tilde{\mathfrak{F}}_n) = P_\theta^{\tilde\varrho}(G_{n+1}\mid\mathfrak{G}_n),\ G_{n+1}\in\mathfrak{G}_{n+1}. \tag{11}$$

From Definition 4

$$\begin{aligned}
P_\theta^{\tilde\varrho}(G_{n+1}\mid\tilde{\mathfrak{F}}_n) &= \int_{E_{n+1}} P_\theta^{\tilde\gamma}(G_{n+1}\mid\tilde{\mathfrak{F}}_n\vee\mathfrak{E}_{n+1})(h_n,\, v_{n+1})\,\tilde\pi_{n+1}(dv_{n+1};\, h_n) \\
&= \int_{E_{n+1}} P_\theta^{\tilde\varrho}(G_{n+1}\mid\mathfrak{G}_n\vee\mathfrak{E}_{n+1})(h_n,\, v_{n+1})\,\tilde\pi_{n+1}(dv_{n+1};\, h_n) \\
&= P_\theta^{\tilde\gamma}(G_{n+1}\mid\mathfrak{G}_n),
\end{aligned}$$

so that (11) holds.

Thus, the sufficiency and transitivity assumptions of Bahadur's reduction theorem ([1]) for the nondesigned experiment case are given on $(\mathcal{X},\, \mathfrak{A},\, \mathcal{P}^{\tilde\varrho})$. If τ is a randomized sample size in relation to $\{\tilde{\mathfrak{F}}_n\}_{n\in N}$ on $(\mathcal{X},\, \mathfrak{A},\, \mathcal{P}^\varrho)$, then it will also be on $(\mathcal{X},\, \mathfrak{A},\, \mathcal{P}^{\tilde\varrho})$. By the theorem of Bahadur ([1]) the stopping rule $\{\varphi_n\}_{n\in N}$ may be replaced by a rule $\{\tilde\varphi_n\}_{n\in N}$ with $\mathfrak{G}_n$-measurable functions $\tilde\varphi_n$, $n\in N$, so that $P_{\theta,\tau}^{\tilde\varrho}$ and $P_{\theta,\tilde\tau}^{\tilde\gamma}$ $\theta\in\Theta$, will coincide on $\mathfrak{G}_\tau$. ∎

5. Corollaries

First conditions for verification of the sufficiency and transitivity of a sequence of σ-fields or statistics will be given.

In the dominated case a factorization theorem may be applied to verify the sufficiency.

LEMMA 1. *If the $P^{\varrho}_{\theta,n}$ are dominated by measures μ^{ϱ}_n with densities $f_{\theta,n}(h_n)$ in accordance with equation (2), then a sequence $\{T_n\}_{n\in N}$ of statistics will be sufficient for $\{\mathfrak{F}_n\}_{n\in N}$, if and only if there are measurable functions $g_{n,\theta}(t_n)$ and $\mathfrak{F}_n$-measurable functions $l_n(h_n)$ such that*

$$f_{\theta,n}(h_n) = g_{n,\theta}(t_n)\, l_n(h_n), \qquad n\in N. \tag{12}$$

If ϱ is also dominated, then, in accordance with equation (3), the factorization becomes

$$f^{\varrho}_{\theta,n}(h_n) = g_{n,\theta}(t_n)\, l^{\varrho}_n(h_n), \qquad n\in N. \tag{13}$$

Proof. The proof is a repetition of the usual consideration in the nondesign case.

It may be noted that the functions $g_{n,\theta}$ and l_n in equation (12) and $g_{n,\theta}$ in equation (13) are independent of ϱ. ∎

For the transitivity the following lemma is useful.

LEMMA 2. *A sequence $\{T_n\}_{n\in N}$ of $(R^m, \mathfrak{B}^m)$-valued statistics is transitive, if*

(1) *there are $(\mathfrak{B}^m \otimes \mathfrak{E}_{n+1} \otimes \mathfrak{A}_{n+1})$-measurable functions g_n with*

$$T_{n+1} = g_n(T_n, V_{n+1}, X_{n+1}), \qquad n\in N, \tag{14}$$

(2) *the σ-fields $\mathfrak{A}_{n+1}$ and $\mathfrak{F}_n \vee \mathfrak{E}_{n+1}$ are conditionally independent under $\mathfrak{G}_n \vee \mathfrak{E}_{n+1}$, $\mathfrak{G}_n = \sigma(T_n)$, that is*

$$P^{\varrho}_{\theta}(A_{n+1} \mid \mathfrak{F}_n \vee \mathfrak{E}_{n+1}) = P^{\varrho}_{\theta}(A_{n+1} \mid \mathfrak{G}_n \vee \mathfrak{E}_{n+1}) \qquad a.e., \tag{15}$$

$A_{n+1}\in \mathfrak{A}_{n+1}$, $\theta\in\Theta$, $n\in N$.

Condition (2) holds, if

(3) *the distribution $P^{(n+1)}_{\theta}(\cdot; h_n, v_{n+1})$ does not depend on h_n, $h_n\in \mathcal{G}_n$.*

Proof. (a) Condition (1) of Definition 4 follows from condition (1) of the Lemma 2.

If $\mathfrak{A}_{n+1}$ and $\mathfrak{F}_n \vee \mathfrak{E}_{n+1}$ are conditionally independent under $\mathfrak{G}_n \vee \mathfrak{E}_{n+1}$, then also $\mathfrak{G}_n \vee \mathfrak{E}_{n+1} \vee \mathfrak{A}_{n+1}$ and $\mathfrak{F}_n \vee \mathfrak{E}_{n+1}$ are conditionally independent under $\mathfrak{G}_n \vee \mathfrak{E}_{n+1}$. But from equation (14)

$$\mathfrak{G}_{n+1} \subseteq \mathfrak{G}_n \vee \mathfrak{E}_{n+1} \vee \mathfrak{A}_{n+1},$$

thus, condition (2) of Definition 4 is valid.

(b) It will be stated the equation

$$P_\theta^\varrho(A_{n+1} \mid \mathfrak{F}_n \vee \mathfrak{E}_{n+1})(h_n, v_{n+1}) = P_\theta^{(n+1)}(A_{n+1}; h_n, v_{n+1}) \tag{16}$$

under condition (1) and (3) of the lemma. From (16) and the condition (3) of the Lemma 2 it follows that $P_\theta^\varrho(A_{n+1} \mid \mathfrak{F}_n \vee \mathfrak{E}_{n+1})$, $A_{n+1} \in \mathfrak{A}_{n+1}$, trivially is a $(\mathfrak{G}_n \vee \mathfrak{E}_{n+1})$-measurable function, and condition (2) of Definition 4 is valid.

To prove (16) it may be noted that by definition

$$\int_C P_\theta^\varrho(A_{n+1} \mid \mathfrak{F}_n \vee \mathfrak{E}_{n+1}) dP_\theta^\varrho = P_\theta^\varrho(A_{n+1} \cap C)$$

for all $C \in \mathfrak{F}_n \vee \mathfrak{E}_{n+1}$. For

$$C = B_1 \times A_1 \times \ldots \times A_n \times B_{n+1}, \qquad A_i \in \mathfrak{A}_i,\ B_i \in \mathfrak{E}_i,$$

the equation

$$\int_C P_\theta^\varrho(A_{n+1} \mid \mathfrak{F}_n \vee \mathfrak{E}_{n+1})(h_n, v_{n+1}) dP_\theta^\varrho = \int_C P_\theta^{(n+1)}(A_{n+1}; h_n, v_{n+1}) dP_\theta^\varrho$$

may be verified using the general Fubini theorem, and for all $C \in \mathfrak{F}_n \vee \mathfrak{E}_{n+1}$ this equation follows from here in the usual way. Both integrands are $(\mathfrak{F}_n \vee \mathfrak{E}_{n+1})$-measurable functions, and the validity of (16) a.e. follows.

The general reduction theorem allows the following practicable corollaries.

COROLLARY 1. *Let be $\lambda(\cdot, \cdot)$ a (randomized) decision rule from $(\mathscr{X}, \mathfrak{F}_\tau, \mathscr{P}_\tau^\varrho)$ into a Borelian decision space $(D, \mathfrak{D})$. Then in the situation of the Theorem λ may be replaced by a decision rule $\tilde{\lambda}$ from $(\mathscr{X}, \mathfrak{G}_\tau, \mathscr{P}_\tau^\varrho)$ into $(D, \mathfrak{D})$ equivalent to λ in the sense*

$$\int_{\mathscr{X}} \lambda(x, C) dP_{\theta,\tau}^\varrho = \int_{\mathscr{X}} \tilde{\lambda}(x, C) dP_{\theta,\tilde{\tau}}^\varrho,$$

$C \in \mathfrak{D}$, $\theta \in \Theta$.

COROLLARY 2. *Let $L(d, \theta)$ on $D \times \Theta$ be a decision loss function and $c_n(h_n, \theta)$ the costs of observation h_n. Then the loss function associated with a statistical procedure (ϱ, τ, λ) is*

$$W_{\varrho,\tau,\lambda}(x, \theta) = \int_D L(d, \theta) d\lambda(x, \cdot) + c_{\tau(x)}(x, \theta),$$

and under suitable conditions the risk function

$$R(\varrho, \tau, \lambda; \theta) = \int_{\mathscr{X}} W_{\varrho,\tau,\lambda}(x, \theta) dP_{\theta,\tau}^\varrho$$

exists.

If in the situation of Lemma 2 $c_{\tau(\cdot)}(\cdot, \theta)$ is a $\mathfrak{G}_\tau$-measurable function of x, $x \in \mathcal{X}$, then for any statistical procedure (ϱ, τ, λ) connected with $\{\mathfrak{F}_n\}_{n\in N}$ there is a procedure $(\tilde{\varrho}, \tilde{\tau}, \tilde{\lambda})$ based on $\{\mathfrak{G}_n\}_{n\in N}$ with

$$R(\varrho, \tau, \lambda; \theta) = R(\tilde{\varrho}, \tilde{\tau}, \tilde{\lambda}; \theta), \qquad \theta \in \Theta.$$

Remark 1. The theorem and its corollaries cannot be formulated without using randomized rules. Even in the case where ϱ, τ, λ are non-randomized rules the equivalent rules $\tilde{\varrho}$, $\tilde{\tau}$, $\tilde{\lambda}$ may be randomized.

Remark 2. With a concrete risk function $R(\varrho, \tau, \lambda; \theta)$ an optimization problem arises to choose a statistical procedure. Unfortunately, only in a few special cases global-optimal strategies exist. For the regression problem some aspects of construction of statistical procedures including sequential design rules are discussed by Heckendorff and Martin ([7]), and Gebhardt and Heckendorff ([4]).

6. Examples

1° *Two-armed bandit problem.* At each stage n one of two random variables Y_i, $i = 1, 2$, with

$$Y_i = \begin{cases} 1 & \text{with probability } p_i, \\ 0 & \text{with probability } 1 - p_i, \end{cases} \quad i = 1, 2,$$

may be observed. The p_i, $i = 1, 2$, are unknown probabilities. Here

$$\mathcal{X}_n = \{0, 1\}, \qquad E_n = \{0, 1\},$$

where $e = 0$ indicates to observe Y_1 and $e = 1$ to observe Y_2. X_n, $n \in N$, is the random variable observed at time n. The parameter θ is $\theta = (p_1, p_2)$ with $\Theta = [0, 1] \times [0, 1]$.

Let the distribution model be the following:

$$P_\theta^{(n+1)}(X_{n+1} = x_{n+1}; h_n, v_{n+1}) = P_\theta^{(n+1)}(X_{n+1} = x_{n+1}; v_{n+1}), \qquad n \in N,$$

and

$$P_\theta^{(n)}(X_n = x_n; v_n) = p_1^{x_n(1-v_n)}(1-p_1)^{(1-x_n)(1-v_n)} p_2^{x_n v_n}(1-p_2)^{(1-x_n)v_n}$$
$$x_n = 0, 1, \; v_n = 0, 1.$$

A design rule ϱ may be given by the probabilities

$$\pi_1(V_1 = v_1), \qquad v_1 = 0, 1,$$

$$\pi_{n+1}(V_{n+1} = v_{n+1}; h_n), \qquad v_{n+1} = 0, 1, \; n \geqslant 1.$$

Then

$$P^{\varrho}_{\theta,n}(H_n = h_n) = P^{\varrho}_{\theta,n}(V_1 = v_1, X_1 = x_1, \ldots, V_n = v_n, X_n = x_n)$$
$$= \prod_{j=1}^{n} P^{(j)}_{\theta}(X_j = x_j; v_j)\pi_1(v_1) \prod_{j=2}^{n} \pi_j(v_j; v_1, x_1, \ldots, v_{j-1}, x_{j-1})$$
$$= f^{\varrho}_{\theta,n}(h_n)$$

is the density in relation to the $2n$-dimensional counting measure ν^{2n}.

From here, the density relative to $\nu^1 \times \pi_1 \times \ldots \times \pi_n$ is

$$\prod_{j=1}^{n} P^{(j)}_{\theta}(X_j = x_j; v_j) = f_{\theta,n}(h_n)$$
$$= p_1^{t_{1n}-t_{3n}}(1-p_1)^{n-t_{1n}-t_{2n}-t_{3n}} p_2^{t_{3n}}(1-p_2)^{t_{2n}-t_{3n}}$$

with

$$t_n = (t_{1n}, t_{2n}, t_{3n}) = \Big(\sum_{j=1}^{n} x_j, \sum_{j=1}^{n} v_j, \sum_{j=1}^{n} x_j v_j\Big).$$

From Lemma 1 follows that $T_n = (T_{1n}, T_{2n}, T_{3n})$ is a sequence of three-dimensional sufficient statistics for the observation sequence (including the design variables!), and from Lemma 2 the transitivity of this sequence may easily be seen.

2° *Regression.* Let

$$(\mathscr{X}_n, \mathfrak{A}_n) = (R^1, \mathfrak{B}^1), \qquad (E_n, \mathfrak{E}_n) = ([-1, 1], \mathfrak{B}_{[-1,1]}), \qquad n \in \boldsymbol{N},$$

and

$$f^{(j)}_{\theta}(x_j; h_{j-1}, v_j) = f^{(j)}_{\theta}(x_j; v_j)$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left[-\frac{1}{2\sigma^2}(x_j - \alpha - \beta v_j)^2\right], \qquad j \in \boldsymbol{N}.$$

For any design rule ϱ the likelihood function in accordance with equation (2) has the form

$$\prod_{j=1}^{n} f^{(j)}_{\theta}(x_j; v_j) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[-\frac{n\alpha^2}{2\sigma^2}\right] \times$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}\left[\beta^2 \sum_{j=1}^{n} v_j^2 + 2\alpha\beta \sum_{j=1}^{n} v_j + \sum_{j=1}^{n} x_j^2 - 2\alpha \sum_{j=1}^{n} x_j - 2\beta \sum_{j=1}^{n} x_j v_j\right]\right\},$$

so that from Lemma 1 and Lemma 2 the sequence $T_n = \Big(\sum_{j=1}^{n} X_j, \sum_{j=1}^{n} V_j,$

$\left(\sum_{j=1}^{n} X_j^2, \sum_{j=1}^{n} V_j^2, \sum_{j=1}^{n} X_j V_j\right)$, $n \in N$, is a sequence of sufficient and transitive statistics, and the theorem allows, without loss of optimality, any statistical procedure (ϱ, τ, λ) to be based on this sequence.

References

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

[2] R. Döhler, *Dominierbarkeit und Suffizienz in der Sequentialanalyse*, Math. Operationsforsch. Statist. Ser. Statist. **12** (1981), 101–134.

[3] V. V. Fedorov, *Theory of optimal experiment*, Nauka, Moscow 1971, in Russian.

[4] R. Gebhardt and H. Heckendorff, *Zur sequentiellen Versuchsplanung für das Regressionsproblem*, Math. Operationsforsch. Statist. Ser. Statist., **14** (1983), 355–366.

[5] K. B. Gray, *Sequential selection of experiments*, Ann. Math. Statist. **39** (1968), 1953–1977.

[6] H. Heckendorff, *Sequential statistical structures*, in: R. Bartoszyński, J. Koronacki, and R. Zieliński (eds.), *Mathematical Statistics*, Banach Center Publ., vol. 6, PWN–Polish Scientific Publishers, Warsaw 1980, 151–157.

[7] H. Heckendorff and J. Martin, *Versuchsplanung für die zweistufige Konstruktion von Konfidenzbändern vorgegebener Breite für eine Regressionsfunktion*, Freiberger Forschungshefte, Band D 128, 1980, 57–81.

Presented to the semester
Sequential Methods in Statistics
September 7–December 11, 1981