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

SEQUENTIAL AND ITERATIVE DESIGNS OF EXPERIMENTS

ANDREJ PÁZMAN

*Mathematical Institute, Slovak Academy of Sciences,
Bratislawa, Czechoslovakia*

1. Introduction

We are given a set X of possible trials (of elementary experiments). At each trial $x \in X$ we can observe a random variable $y(x)$ which is distributed according to $p(\cdot|x, \theta)$ on R^1. Here θ stands for the unknown vector of parameters. We suppose that $\theta \in \Theta$ = a given subset of R^m.

The experimenter can choose a sequence of trials $x_1, \ldots, x_N$ which are to be performed in the experiment. We suppose that the random variables $y(x_1), \ldots, y(x_N)$ observed in the experiment are independent.

The choice of trials $x_1, \ldots, x_N$ can be done either a priori, before performing observations, or sequentially, according to the scheme

$$\ldots \text{design} \ldots \text{observation} \ldots \text{estimation of } \theta \ldots \text{design} \ldots, \text{etc.}$$

In the linear regression model, which is considered in Section 2, we can predict the variances of linear estimates without performing observations. As a consequence, instead of a sequential design, we use an iterative procedure for computing designs:

$$\ldots \text{design} \ldots \text{computation of variances} \ldots \text{design} \ldots, \text{etc.}$$

In Section 2 there is a brief survey on iterative design procedures in linear regression models. In Section 3 the nonlinear models are considered.

2. Iterative designs in the linear regression model

In this section we suppose that $\Theta = R^m$ and that the mean of $y(x)$ is

$$E_\theta y(x) = f'(x)\theta \qquad (\theta \in R^m, x \in X), \tag{1}$$

and the variance is

$$\operatorname{Var} y(x) = \sigma^2(x),$$

where $f_1(x),\dots,f_m(x)$ and $\sigma^2(x) > 0$ are known continuous functions on the set X, which is supposed to be compact, and $\theta := (\theta_1,\dots,\theta_m)'$ is the vector of unknown parameters. The functions $f_1(x),\dots,f_m(x)$ are linearly independent.

The matrix

$$M = \sum_{k=1}^{N} f(x_k) f'(x_k) \sigma^{-2}(x_k) \tag{2}$$

is the information matrix corresponding to the design $x_1,\dots,x_N$. It is well known that, if M is nonsingular, the vector of the BLUE-s of $\theta_1,\dots,\theta_m$ is given by

$$\hat{\theta} = \operatorname{Arg}\min_{\theta\in R^m} \sum_{i=1}^{N} \sigma^{-2}(x_i)[y(x_i) - f'(x_i)\theta]^2 = M^{-1} \sum_{i=1}^{N} f(x_i)\sigma^{-2}(x_i)\, y(x_i),$$

and its covariance matrix is

$$\operatorname{Var}(\hat{\theta}|M) = M^{-1}. \tag{3}$$

For a general M (singular or nonsingular), a function of the parameters $g'\theta$ is estimable without bias if and only if the vector g can be expressed as $g = Mu$ for some $u \in R^m$. Then the variance of the BLUE of $g'\theta$ is

$$\operatorname{Var}(g'\hat{\theta}|M) = g' M^- g,$$

where M^- is a g-inverse of M.

Designs of experiments are compared, roughly speaking, according to the value of $\operatorname{Var}(g'\hat{\theta}|M)$ for different vectors g. Hence the role of the information matrix in comparing designs is of crucial importance. It is emphasized still more if we consider robust estimates instead of the least squares estimates, since the covariance matrix of the robust estimates of $\theta_1,\dots,\theta_m$ tends to kM^{-1} with $N \to \infty$, where k depends on the method of estimation but not on the design (cf. [2]).

Let us denote by $c(x)$ the cost per one observation in a trial x, by $N(x)$ the number of repetitions of x in the design $x_1,\dots,x_N$, and by $c = \sum_{x\in X} N(x)c(x)$ the total cost of the experiment.

The *design measure* is defined by

$$\xi(x) := N(x)c(x)/c \qquad (x \in X).$$

Evidently

$$M = cM(\xi),$$

where $M(\xi)$ is the information matrix of the design measure ξ,

$$M(\xi) := \sum_{x\in X} \tilde{f}(x)\tilde{f}'(x)\xi(x), \tag{4}$$

and where $\tilde{f}(x) := f(x)\sigma^{-1}(x)c^{-1/2}(x)$. Without restriction on generality we shall write $f(x)$ instead of $\tilde{f}(x)$ in (4).

We identify the set of all possible design measures with the set Ξ of all probability measures on X which are supported by finite sets. The set

$$\mathfrak{M} := \{M(\xi) \colon \xi \in \Xi\} \tag{5}$$

is convex and compact.

An optimality criterion is given by a function $\phi\colon \mathfrak{M} \to R^1$, which expresses in a statistically meaningful way the aim of the experimenter. Using such a criterion, the experimenter tries to find a $\xi^* \in \Xi$ such that

$$\xi^* = \operatorname{Arg}\min_{\xi \in \Xi} \phi\,[M(\xi)]. \tag{6}$$

The design measure ξ^* is called *ϕ-optimal.* A variety of different functions ϕ have been proposed as optimality criteria. They can be divided into two main classes:

a) *Global (regular) optimality criteria.* They are defined by: $\phi(M) < \infty \Leftrightarrow M$ is nonsingular. This means that $\phi(M)$ is finite iff all parameters $\theta_1, \dots, \theta_m$ are estimable without bias.

b) *Partial (singular) optimality criteria.* They are defined by: There is an $M \in \mathfrak{M}$ such that M is singular and that $\phi(M) < \infty$. Partial criteria are used when there are nuisance (useless) parameters, or functions of parameters.

Examples of global optimality criteria.

$$\phi_{p,H}(M) = \left[\frac{1}{m}\operatorname{tr}(HM^{-1}H')^p\right]^{1/p} \qquad (M \in \mathfrak{M}, \det M \neq 0)$$

where $p \in (0, 1)$, and H is a nonsingular $m \times m$ matrix ([3]).

Special cases. *D-optimality* ($p \to 0$, $H = I$), *E-optimality* ($p \to \infty$, $H = I$), *A-optimality* ($p = 1$, $H = I$), *L-optimality* ($p = 1$, $H =$ arbitrary).

$$\phi_{\mathfrak{G}}(M) = \max\{g'(M)^{-1}g \colon g \in \mathfrak{G}\}$$

where $\mathfrak{G}$ is a set of vectors in R^m which span R^m.

Special cases. *G-optimality* ($\mathfrak{G} = \{f(x) \colon x \in X\}$), *E-optimality* ($\mathfrak{G} = \{g \colon g \in R^m, \|g\| = 1\}$).

Examples of partial optimality criteria.

$$\tilde{\phi}_{s,H}(M) = \begin{cases} [(1/s)\operatorname{tr}(H'M^{-}H)^p]^{1/p} & \text{if } H = MV \text{ for some } m \times s \text{ matrix } V, \\ \infty & \text{otherwise} \end{cases}$$

where $s < p$, H an $m \times s$ matrix of rank s ([3]).

$$\tilde{\phi}_{\mathfrak{G}}(M) = \begin{cases} \max\{g'M^{-}g\colon\ g \in \mathfrak{G}\} & \text{if } g = Mu,\ \forall g \in \mathfrak{G}, \\ \infty & \text{otherwise} \end{cases}$$

where the set $\mathfrak{G} \subset R^m$ does not span R^m.

$$\Phi^{(r)}(M) = \sum_{\substack{i_1,\ldots,i_r=1 \\ j_1,\ldots,j_r=1}}^{m} \gamma_{i_1,\ldots,i_r}\{M^{-}\}_{i_1 j_1}\cdots\{M^{-}\}_{i_r j_r}\gamma_{j_1,\ldots,j_r}$$

(the variance of the estimate for a polynomial

$$p(\theta) := \sum_{i_1,\ldots,i_r} \gamma_{i_1,\ldots,i_r}\theta_{i_1}\cdots\theta_{i_r},$$

as proved in [4], Chapter IV).

Global criterion functions are continuous on $\mathfrak{M}$. Partial criterion functions are continuous on $\mathfrak{M}_+ := \{M \in \mathfrak{M}\colon \det M \neq 0\}$ but are only lower semicontinuous on $\mathfrak{M}$. All given criterion functions are convex on $\mathfrak{M}$. Some of them are differentiable on $\mathfrak{M}_+$, i.e., there is a gradient $\nabla\phi(M)$,

$$\{\nabla\phi(M)\}_{ij} = \frac{\partial\phi(M)}{\partial\{M\}_{ij}} \qquad (i, j = 1, \ldots, m). \tag{7}$$

For the other (nondifferentiable) criteria convexity implies the existence of the directional derivative

$$\partial\phi(M, L) := \lim_{\beta \downarrow 0} \frac{\phi[(1-\beta)M + \beta L] - \phi(M)}{\beta} \qquad (L \in \mathfrak{M},\ M \in \mathfrak{M}_+).$$

In the differentiable case we have, evidently,

$$\partial\phi(M, L) = \operatorname{tr} \nabla\phi(M)(L - M). \tag{8}$$

There are many iterative procedures to compute optimal designs under different optimality criteria. We shall discuss some common features.

An iterative procedure to compute a ϕ-optimal design (eq. (6)) is started by an arbitrary design measure ξ_0 such that $\det M(\xi_0) \neq 0$. At the nth step of the procedure a design measure ξ_{n+1} is computed from ξ_n as

$$\xi_{n+1} = (1-\beta_n)\xi_n + \beta_n \varkappa_n, \tag{9}$$

where $\beta_n \in (0, 1)$, $\varkappa_n \in \Xi$. The choice of β_n and of $\varkappa_n$ has to ensure the convergence of the sequence $\{\phi[M(\xi_n)]\}_{n=0}^{\infty}$ to $\min_{M \in \mathfrak{M}} \phi(M)$. The procedure is stopped at some step n according to a stopping rule. Then ξ_n is considered as approximately ϕ-optimal.

The requirement $\det M(\xi_0) \neq 0$ implies that $\det M(\xi_n) \neq 0$, as follows from (9) and from the positive semidefiniteness of an information matrix (eg. (4)). Hence the whole iterative procedure does not exceed the frame of the set $\mathfrak{M}_+$. Nevertheless, $\mathfrak{M}_+$ is dense in $\mathfrak{M}$, since for every $M \in \mathfrak{M}$, $\bar{M} \in \mathfrak{M}_+$ we have $\lim_{n\to\infty}(1-1/n)M+1/n\bar{M} = M$. Hence there are no principal difficulties with iterative procedures for global criterion functions (which are continuous on $\mathfrak{M}$). The same cannot be said about partial optimality criteria.

Fortunately, there is a simple and effective universal stopping rule for convex optimality criteria ([3]). It is expressed by the implication

$$[\phi[M(\xi)] < \infty \quad \text{and} \quad (\forall M \in \mathfrak{M})(\partial\phi[M(\xi), M] \geqslant -\delta)]$$
$$\Rightarrow \phi[M(\xi)] \leqslant \min_{M\in\mathfrak{M}} \phi(M)+\delta, \tag{10}$$

which is a direct consequence of the convexity of ϕ. Since $\{f(x)f'(x)\colon x\in X\}$ is the set of extreme points of $\mathfrak{M}$, for differentiable optimality criterion functions ϕ we obtain from (10)

$$[\phi[M(\xi)] < \infty \text{ and } \min_{x\in X} f'(x)\nabla\phi[M(\xi)]f(x) \geqslant \operatorname{tr} M(\xi)\nabla\phi[M(\xi)]-\delta]$$
$$\Rightarrow \phi[M(\xi)] \leqslant \min_{M\in\mathfrak{M}} \phi(M)+\delta. \tag{11}$$

The most intricate task in proposing a new procedure for computing optimal designs is to prove its convergence. We shall sketch a general scheme of such a proof. To ensure this convergence it is required first that $\lim_{n\to\infty}\beta_n = 0$, and that $\sum_{n=0}^{\infty}\beta_n = \infty$. The choice of the correcting design $\varkappa_n$ in equation (9), which is the "essence" of the procedure, can be formalized as follows: with every $M\in\mathfrak{M}_+$ we associate a closed convex set $\mathfrak{K}(M)\subset\mathfrak{M}$ and take $\varkappa_n$ such that

$$\frac{M(\xi_{n+1})-M(\xi_n)}{\beta_n} = M(\varkappa_n)-M(\xi_n)\in\mathfrak{K}[M(\xi_n)]. \tag{12}$$

We suppose that the mapping $\mathfrak{K}$ is closed in the sense that

$$M_n\in\mathfrak{M},\ L_n\in\mathfrak{K}(M_n),\ M_n\to M,\ L_n\to L \ \Rightarrow\ L\in\mathfrak{K}(M),$$

and bounded by a number $C<\infty$,

$$\|L\| \leqslant C \qquad (L\in\mathfrak{K}(M),\ M\in\mathfrak{M}), \tag{13}$$

where $\|L\|^2 = \operatorname{tr} LL'$. The mapping

$$\tau\colon\ t\in\langle 0,\infty) \to M(\xi_n)+\Big(t-\sum_0^n \beta_i\Big)\frac{M(\xi_{n+1})-M(\xi_n)}{\beta_n} \tag{14}$$

is a "trajectory" in $\mathfrak{M}_+$ corresponding to the iterative procedure. From equation (13) it follows that

$$\frac{\|\tau(t_1)-\tau(t_2)\|}{|t_1-t_2|} < C \qquad (t_1, t_2\in\langle 0,\infty)), \tag{15}$$

and from (12) it follows that $d\tau/dt \in \mathfrak{K}[M(\xi_n)]$ if $t \in \langle \sum_0^{n-1} \beta_i, \sum_0^n \beta_i \rangle$. With n tending to infinity this interval becomes shorter, and we find that "approximately"

$$\frac{d\tau}{dt} \in \mathfrak{K}[\tau(t)]. \tag{16}$$

The "trajectories" in $\mathfrak{M}_+$ which satisfy (15), but also (16) for almost every $t \in \langle 0, \infty)$, are of special interest. Intuitively they correspond to "trajectories" of iterative procedures with "infinitely small $\beta_n - s$".

If we can show that

a) the global criterion function ϕ can be extended to a finite convex function on a set $\mathfrak{U} \subset \mathfrak{M}_+$ which is open in the span of $\mathfrak{M}_+$,

b) the function $t \in \langle 0, \infty) \to \phi[\tau(t)]$ is decreasing on "trajectories" which satisfy (15) and (16) unless $\tau(t)$ is the information matrix of a ϕ-optimal design,

c) $+\infty$ is not a limit point of the sequence $\{\phi[M(\xi_n)]\}_{n=0}^{\infty}$,

then the procedure is convergent, i.e.,

$$\lim_{n \to \infty} \phi[M(\xi_n)] = \min_{M \in \mathfrak{M}} \phi(M).$$

Further details and proofs concerning the content of this section can be found in [4].

3. Sequential designs in nonlinear models

In this section we shall suppose that the parameter set Θ is an open bounded set in R^m, that the set X of possible trials is finite and that $y(x)$ is distributed according to an arbitrary probability distribution $p(\cdot|x, \theta)$ on R^1. We shall suppose further that for any Borel set $A \subset R^1$ and any $x \in X$ the function $\theta \in \Theta \to p(A|x, \theta)$ is continuous.

A sequential design of an experiment is a successive choice of trials $x_1, x_2, \ldots, x_i, \ldots$ which is based on the results of previous observations. This means that at the ith step of the sequential procedure we have to decide whether to stop further observations or to choose another trial x_{i+1} in an optimal way, and to observe $y(x_{i+1})$.

Let us denote by $x_{(i)} = (x_1, \ldots, x_i)$ the sequence of the first i trials and by $y_{(i)} = (y(x_1), \ldots, y(x_i))$ the sequence of the observed random variables.

To be general, we shall suppose that the stopping time and the choice of x_{i+1} at the ith step are random. The stopping time is defined by a sequence

$$\varphi := \{\varphi_i(x_{(i)}, y_{(i)})\}_{i=0}^{\infty},$$

where $\varphi_i(x_{(i)}, y_{(i)})$ is the probability of stopping the experiment at the ith step if $y_{(i)}$ has been observed in the sequence of trials $x_{(i)}$. The choice of new trials is given by a sequence

$$\delta := \{\delta_i(\cdot|x_{(i)}, y_{(i)})\}_{i=0}^{\infty},$$

where $\delta_i(\cdot|x_{(i)}, y_{(i)})$ is a probability distribution on X and $\delta_i(x|x_{(i)}, y_{(i)})$ is the probability of $x_{i+1} = x$ if $y_{(i)}$ was observed and the procedure has not been stopped before the $(i+1)$st step.

The *sequential design of the experiment* is thus defined by the pair (φ, δ).

An optimality criterion is given by a sequence of functions $\{q(x_{(i)}, y_{(i)})\}_{i=0}^{\infty}$ such that $-q(x_{(i)}, y_{(i)})$ expresses the quality of the experiment if it is stopped at the ith step after observing $y_{(i)}$ in the sequence of trials $x_{(i)}$.

EXAMPLE 1. Let us suppose that there is a probability density $f(y|x, \theta) = dP(y|x, \theta)/d\lambda$ with respect to the Lebesgue measure λ which satisfies the usual conditions of regularity. Let $K \subset \Theta$ be a compact set and let us denote by $\hat{\theta}^{(i)} = \hat{\theta}(x_{(i)}, y_{(i)})$ the M.L. estimate

$$\hat{\theta}^{(i)} := \operatorname*{Arg\,min}_{\theta \in K} \sum_{j=1}^{i} \log f(y(x_j)|x_j, \theta). \tag{17}$$

The Fisher information matrix is

$$\{M_F^{(i)}(\theta)\}_{kl} := \sum_{j=1}^{i} E_\theta \left[\frac{\log f(y|x_j, \theta)}{\partial \theta_k} \frac{\log f(y|x_j, \theta)}{\partial \theta_l} \right] \qquad (k, l = 1, \ldots, m) \tag{18}$$

We may define

$$q(x_{(i)}, y_{(i)}) = \log \det M_K^{(i)}(\hat{\theta}^{(i)})$$

(the *local D-optimality*), or

$$q(x_{(i)}, y_{(i)}) = \begin{cases} \sum_{k=1}^{m} \{[M_F^{(i)}(\hat{\theta}^{(i)})]^{-1}\}_{kk}; & \text{if } M_F^{(i)}(\hat{\theta}^{(i)}) \text{ is non singular,} \\ \infty & \text{otherwise} \end{cases}$$

(the *local A-optimality*), etc.

EXAMPLE 2. Under the assumptions given in Example 1, take η_0 as an a priori p.d. on K. We may define

$$q(x_{(i)}, y_{(i)}) = \int_K \log \det M_F^{(i)}(\theta)\, \eta_0(d\theta)$$

(the *Bayesian D-optimality*), etc.

We return to the general case. Let us denote

$$\psi_i(x_{(i)}, y_{(i)}) := \prod_{j=0}^{i} [1 - \varphi_j(x_{(j)}, y_{(j)})],$$

$$\Delta_i(x_{(i)}, y_{(i)}) := \prod_{j=0}^{i} \delta_j(x_{j+1}|x_{(j)}, y_{(j)}),$$

$$P(\cdot|x_{(i)}, \theta) := \prod_{j=1}^{i} p(\cdot|x_j, \theta)$$

(the joint p.d. of $y_{(i)}$, if $x_{(i)}$ is given)

$$P^{(i)}(x_{(i)}, dy_{(i)}|\theta) = \Delta_{i-1}(x_{(i-1)}, y_{(i-1)})\, P(dy_{(i)}|x_{(i)}, \theta) \tag{19}$$

(the joint p.d. of $x_{(i)}$, $y_{(i)}$), and let $E_\theta^{(i)}$ denote the mean with respect to $P^{(i)}(\cdot|\theta)$.

The distribution function of the optimality criterion function q can be written as

$$F_\theta(b|\varphi, \delta) = \sum_{i=0}^{\gamma} E_\theta^{(i)}[\chi_{B(b)}\varphi_i\psi_{i-1}] \qquad (b \in R^1), \tag{20}$$

where

$$B(b) := \{(i, x_{(i)}, y_{(i)}) \colon q(x_{(i)}, y_{(i)}) < b\},$$

and $\varphi_i\psi_{i-1} = \varphi_i \prod_{j=0}^{i-1} (1 - \varphi_j)$ is the probability of stopping exactly at the ith step. Two sequential designs (φ, δ) and $(\bar{\varphi}, \bar{\delta})$ are considered as equivalent (with respect to q) iff

$$F_\theta(b|\varphi, \delta) = F_\theta(b|\bar{\varphi}, \bar{\delta}) \qquad (b \in R^1).$$

Sometimes, it is useful to consider an a priori p.d. η_0 on K with $\eta_0(K) = 1$. For simplicity let us suppose that K is finite. The a posteriori p.d. is

$$\eta(\theta|x_{(i)}, y_{(i)}) = h(y_{(i)}, x_{(i)}, \theta)\,\eta_0(\theta), \tag{21}$$

where $h(\cdot, x_{(i)}, \theta)$ is the density of $P(\cdot|x_{(i)}, \theta)$ with respect to

$$\mu(\cdot) := \sum_{\theta \in K} P(\cdot|x_{(i)}, \theta)\,\eta_0(\theta). \tag{22}$$

Evidently,

$$\eta(K|x_{(i)}, y_{(i)}) = 1.$$

THEOREM. *Let $q(x_{(i)}, y_{(i)})$ be a function of $\eta(\cdot|x_{(i)}, y_{(i)})$,*

$$q(x_{(i)}, y_{(i)}) = Q_i[\eta(\cdot|x_{(i)}, y_{(i)})] \qquad (i = 0, 1, \ldots).$$

Then to every sequential design φ, δ *there is an equivalent sequential design* $(\bar{\varphi}, \bar{\delta})$ *such that*

$$\varphi_i(x_{(i)}, y_{(i)}) = \phi_i[\eta(\cdot|x_{(i)}, y_{(i)})],$$
$$\delta_i(x_{i+1}|x_{(i)}, y_{(i)}) = D_i[x_{i+1}, \eta(\cdot|x_{(i)}, y_{(i)})]$$

for some functions ϕ_i, D_i $(i = 0, 1, \ldots)$.

Note. The criterion function in Example 2 satisfies the assumptions in the theorem.

COROLLARY. *If the criterion functions given in Example* 1 *are used, then it is sufficient to consider only those sequential designs which depend on* $x_{(i)}$, $y_{(i)}$ *only through the likelihood function*

$$\theta \in \Theta \to \sum_{j=1}^{i} \log f(y_j|x_j, \theta).$$

Proof. If η_0 is proportional to the counting measure on K then

$$\eta(\theta|x_{(i)}, y_{(i)}) = \frac{\prod_{j=1}^{i} f[y_j|x_j, \theta]}{\sum_{\theta \in K} \prod_{j=1}^{i} f[y_j|x_j, \theta]}.$$

Moreover, the M.L. estimate $\hat{\theta}^{(i)}$ can be expressed as

$$\hat{\theta}^{(i)} = \operatorname{Arg} \min_{\theta \in K} \log \frac{\eta(\theta|x_{(i)}, y_{(i)})}{\eta_0(\theta)}. \quad \blacksquare$$

LEMMA. *The sequence of statistics*

$$(x_{(i)}, y_{(i)}) \in X^i \times R^i \to \eta(\cdot|x_{(i)}, y_{(i)}) \qquad (i = 0, 1, \ldots)$$

is a transitive sequence of sufficient statistics (in the sense of [1]) *with respect to the measures* $\{P^{(i)}(\cdot|\theta); \theta \in K\}$ $(i = 1, 2, \ldots)$.

Proof. Let ν be the counting measure on the set X^i (i.e., $\nu(x_{(i)}) = 1$; $(x_{(i)} \in X^i)$) and let μ be given by equation (22). According to (19) and to (21)

$$\frac{dP^{(i)}(x_{(i)}, y_{(i)}|\theta)}{d\mu \times \nu} = \frac{dP(y_{(i)}|x_{(i)}, \theta)}{d\mu} = G[\eta(\cdot|x_{(i)}, y_{(i)}), \theta],$$

where the function G is defined by

$$G[\eta(\cdot), \theta] = \eta(\theta)/\eta_0(\theta).$$

Hence the sufficiency is proved. Further

$$\frac{\eta(\theta|x_{(i)}, y_{(i)})}{\eta(\theta|x_{(i-1)}, y_{(i-1)})} = \frac{dp(y_i|x_i, \theta)}{d\mu_i},$$

where $\mu_i(\cdot) = \sum_{\theta \in K} p(\cdot|x_i, \theta)\eta(\theta|x_{(i-1)}, y_{(i-1)})$. Hence $\eta(\cdot|x_{(i)}, y_{(i)})$ depends only on x_i, y_i and on $\eta(\cdot|x_{(i-1)}, y_{(i-1)})$. The transitivity follows from [1]. ∎

Proof of the theorem. Let us define

$$\bar{\varphi}_i(x_{(i)}, y_{(i)}) = \frac{E_\theta^{(i)}[\varphi_i \psi_{i-1}|\eta(\cdot|x_{(i)}, y_{(i)})]}{E_\theta^{(i)}[\psi_{i-1}|\eta(\cdot|x_{(i)}, y_{(i)})]}, \tag{23}$$

and

$$\bar{\delta}_i(x_{i+1}|x_{(i)}, y_{(i)}) = \frac{E_\theta^{(i)}[\delta_i(x_{i+1}|\cdot)\psi_{i-1}|\eta(\cdot|x_{(i)}, y_{(i)})]}{E_\theta^{(i)}[\psi_{i-1}|\eta(\cdot|x_{(i)}, y_{(i)})]}. \tag{24}$$

According to (20) it is enough to prove that

$$E_\theta^{(i)}[g\circ\eta_i\, \varphi_i \psi_{i-1}] = \bar{E}^{(i)}[g\circ\eta_i\, \bar{\varphi}_i \bar{\psi}_{i-1}] \tag{25}$$

for any bounded function g defined on the set of all probability measures on K, where $\eta_i := \eta(\cdot|x_{(i)}, y_{(i)})$. We proceed by induction on i. Since $\varphi_0 = \bar{\varphi}_0$, $\delta_0(x_1) = \bar{\delta}_0(x_1)$, equation (25) is evident for $i = 1$. Suppose that (25) is true for $i = k-1$. We have according to (23)

$$\begin{aligned} E_\theta^{(k)}[g\circ\eta_k\, \varphi_k \psi_k] &= E_\theta^{(k)}\{g\circ\eta_k\, E_\theta^{(k)}[\varphi_k \psi_{k-1}|\eta_k]\} \\ &= E_\theta^{(k)}\{g\circ\eta_k\, E_\theta^{(k)}[\bar{\varphi}_k \psi_{k-1}|\eta_k]\} \\ &= E_\theta^{(k)}\{g\circ\eta_k\, \bar{\varphi}_k \psi_{k-1}\}. \end{aligned} \tag{26}$$

We shall use the following notation: If l: $X^k \times R^k \to R^1$ then $l^{(x_k, y_k)}$: $X^{k-1} \times R^{k-1} \to R^1$ will be defined by

$$l^{(x_k, y_k)}(x_{(k-1)}, y_{(k-1)}) = l(x_{(k)}, y_{(k)}).$$

From the transitivity of $\{\eta(\cdot|x_{(i)}, y_{(i)})\}_{i=1}^\infty$ it follows that $[g\circ\eta_k\, \bar{\varphi}_k]^{(x_k, y_k)}$ is a function of $\eta(\cdot|x_{(k-1)}, y_{(k-1)})$. Hence, using (19), we obtain

$$\begin{aligned} &E_\theta^{(k)}[g\circ\eta_k\, \bar{\varphi}_k \psi_{k-1}] \\ &= \sum_{x_k \in X} \int_R E_\theta^{(k-1)}\{\delta_k(x_k|\cdot)[g\circ\eta_k\, \bar{\varphi}_k]^{(x_k, x_k)} \psi_{k-1}\}\, p(dy_k|x_k, \theta) \\ &= \sum_{x_k \in X} \int_R E_\theta^{(k-1)}\{[g\circ\eta_k\, \bar{\varphi}_k]^{(x_k, y_k)} E_\theta^{(k-1)}[\delta_k(x_k|\cdot)\psi_{k-1}]\}\, p(dy_k|x_k, \theta) \\ &= \sum_{x_k \in X} \int_R E_\theta^{(k-1)}\{[g\circ\eta_k\, \bar{\varphi}_k]\bar{\delta}_k(x_k|\cdot)\psi_{k-1}\}\, p(dy_k|x_k, \theta), \end{aligned} \tag{27}$$

the last equality being a consequence of (24). Since $\psi_{k-1} = (1-\varphi_{k-1})\psi_{k-2}$, using the validity of (25) for $i = k-1$, we obtain

$$E_\theta^{(k-1)}\{[g\circ\eta_k\, \bar{\varphi}_k]\bar{\delta}_k(x_k|\cdot)\psi_{k-1}\} = E_\theta^{(k-1)}\{[g\circ\eta_k\, \bar{\varphi}_k]\bar{\delta}_k(x_k|\cdot)\bar{\psi}_{k-1}\}. \tag{28}$$

Substituting (28) in (27) and comparing with (26), we obtain (25). ∎

References

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

[2] J. Jurečková, *Robust estimation in a linear regression model*, in: R. Bartoszyński, J. Koronacki and R. Zieliński (eds.), *Mathematical Statistics*, Banach Center Publ., vol. 6, PWN–Polish Scientific Publishers, Warszawa 1980, 169–174.

[3] J. Kiefer, *General equivalence theory for optimum designs (approximate theory)*, Ann. Statist. **2** (1974), 849–879.

[4] A. Pázman, *Foundations of Optimization of Experiments*, Veda, Bratislava 1980, in Slovak. (The English translation will appear in 1985 ed. by Reidler.)

[5] C. F. Wu, *Some iterative procedures for generating non-singular optimal designs*, Commun. Statist. A – Theory Methods, (1978), 1399–1412.

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