ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XVI, 4 (1980)

W. NÄTHER and J. PILZ (Freiberg, G.D.R.)

ESTIMATION AND EXPERIMENTAL DESIGN IN A LINEAR REGRESSION MODEL USING PRIOR INFORMATION

1. Introduction. The unknown parameters of a linear regression model are usually estimated by the method of least squares, and G- or D-optimum designs are the most frequently investigated experimental designs for the least squares estimator (LSE). Both the LSE and, e.g., G- and D-optimum designs in no manner make use of prior knowledge about the unknown regression parameters or about the domain of forecast. But prior information is given mostly in real situations. Therefore, it is of practical importance to combine prior information and actual information in a well-defined way to obtain more efficient methods of estimation and experimental design in a linear regression model.

The present paper deals with this problem in a Bayesian terminology. In Section 2 we give a decision theoretic formulation of the estimation and designing problem in a linear regression model. Then we are concerned with a special Bayes estimator (BE) for which we summarize some useful properties (Section 3). In Section 4 we deal with the designing problem for this BE. The main goal is to establish a certain robustness of admissibility and D-optimality of designs relative to the underlying estimator (LSE or BE). We show that under certain conditions designs being admissible for the BE are also admissible for the LSE (and conversely) and, for a simple linear regression model, D-optimum designs for the LSE are also D-optimum for the BE.

2. Definitions, notions and decision theoretic formulation of the problem. Let $\boldsymbol{R}^k$ be the k-dimensional Euclidean space, $\boldsymbol{R}^+$ the set of positive real numbers, P_Z the distribution of a random variable (or vector) Z, $\mathrm{Var}Z$ the variance (or covariance matrix) of Z, and E_Z the operator of expectation with respect to the distribution P_Z. If no confusion is possible, E_Z is denoted shortly by E. By A' we denote the transpose of a matrix (or vector) A.

Let $B \subset \boldsymbol{R}^k$. The random variables $Y(x)$, $x \in B$, are assumed to satisfy the linear regression model

$$\mathrm{E}Y(x) = \theta' f(x), \quad \theta \in T \subseteq \boldsymbol{R}^r,$$
$$\operatorname{Var} Y(x) = 1/\lambda, \quad \lambda \in K \subseteq \boldsymbol{R}^+,$$

where $f(\cdot) = (f_1(\cdot), \ldots, f_r(\cdot))'$ is a vector of known real-valued and linearly independent functions, $\theta = (\theta_1, \ldots, \theta_r)'$ and λ are unknown parameters (shortly, $\delta = (\theta', \lambda)'$).

Let $V \subset B$ be the experimental region. Any n-tuple $V_n = (x_1, \ldots, x_n)$ belonging to V^n is called an *exact experimental design* or *design* for short. At the points x_i of a design V_n we observe n realizations of $Y(x)$, i.e., a realization y of the so-called observation vector $Y(V_n) = (Y(x_1), \ldots, Y(x_n))'$. The observations are assumed to be uncorrelated, i.e., we have $\operatorname{Var} Y(V_n) = \lambda^{-1} I_n$ (I_n denotes the $(n \times n)$-identity matrix).

Based on a design V_n from a certain set $V^{(n)}$ of designs of size n and on the sample y, a linear function of the parameter θ, say $\psi(x, \theta) = A(x)\theta \in \boldsymbol{R}^m$, $m \leqslant r$, is to be estimated for the points x from a set $H \subset B$. $A(x)$ is a given $(m \times r)$-matrix and H is called the *domain of forecast*. The estimator for $\psi(x, \theta)$ is denoted by $\hat{\psi}(x, Y(V_n))$, and $\hat{\psi}$ is to be taken from some set D of measurable decision functions. The goodness of estimation is valued by a risk $R(\delta, x; \hat{\psi}, V_n)$ which is generated by a non-negative loss function $L(\cdot\,;\cdot)$ on $\boldsymbol{R}^m \times \boldsymbol{R}^m$ according to the formula

$$R(\delta, x; \hat{\psi}, V_n) = \mathrm{E}_{Y(V_n)/\delta} L\big(\psi(x, \theta); \hat{\psi}(x, Y(V_n))\big).$$

$L(\psi(x, \theta); \hat{\psi}(x, y))$ measures the loss incurred in estimating the true parameter $\psi(x, \theta)$ by $\hat{\psi}(x, y)$, $\mathrm{E}_{Y(V_n)/\delta}$ denotes the conditional expectation of $Y(V_n)$ for given δ, and the conditional distribution $P_{Y(V_n)/\delta}$ is assumed to be known.

The common problem of optimum experimental design in a linear regression model can be interpreted as a statistical decision problem

$$G = [T \times K \times H, D \times V^{(n)}, R], \tag{1}$$

where the elements $(\theta, \lambda, x) \in T \times K \times H$ are the states of "nature" and the elements $(\hat{\psi}, V_n) \in D \times V^{(n)}$ are the strategies of the "experimentalist".

Now we assume that the experimentalist, in addition to the sample information $Y(V_n)$, has a certain prior knowledge which he can express by a prior distribution $P_{(\Delta,X)}$ on $T \times K \times H$, i.e., $\delta = (\theta, \lambda) \in T \times K$ and $x \in H$ become random vectors $\Delta = (\Theta, \Lambda)$ and X, respectively. A solution of the decision problem (1), which regards both the actual and the prior information, is the Bayes solution $(\hat{\psi}^*, V_n^*)$ with respect to $P_{(\Delta,X)}$. If

$$\varrho(P_{(\Delta,X)}; \hat{\psi}, V_n) = \mathrm{E}_{(\Delta,X)} R(\Delta, X; \hat{\psi}, V_n)$$

denotes the Bayes risk of the strategies $(\hat{\psi}, V_n) \in D \times V^{(n)}$ with respect to $P_{(\Delta,X)}$, then $(\hat{\psi}^*, V_n^*)$ satisfies the equation

$$(2) \qquad \varrho(P_{(\Delta,X)}; \hat{\psi}^*, V_n^*) = \inf_{(\hat{\psi}, V_n) \in D \times V^{(n)}} \varrho(P_{(\Delta,X)}; \hat{\psi}, V_n).$$

Before we are going to search for a Bayes strategy $(\hat{\psi}^*, V_n^*)$, we briefly describe, for comparison, the classical approach to the experimental design problem.

Usually, $\psi(x, \theta)$ is estimated by the LSE. If

$$F(V_n) = \big(f(x_1), \ldots, f(x_n)\big)'$$

is of full rank, i.e., $\operatorname{rank} F(V_n) = r$, then the LSE $\tilde{\psi}$ is given by

$$\tilde{\psi}\big(x, Y(V_n)\big) = A(x)\tilde{\theta}\big(Y(V_n)\big) = A(x)\big(F(V_n)'F(V_n)\big)^{-1}F(V_n)'Y(V_n).$$

The covariance matrix

$$\operatorname{Var}\tilde{\psi} = \lambda^{-1}A(x)\big(F(V_n)'F(V_n)\big)^{-1}A(x)'$$

of $\tilde{\psi}$ depends on the design V_n only by the so-called *information matrix*

$$M(V_n) = n^{-1}F(V_n)'F(V_n).$$

Hence, classical criteria of optimality of designs for the LSE are suitable functionals of $M(V_n)$. For example, $\tilde{V}_n$ is called *D-optimum* in $V^{(n)}$ for LSE if

$$(3) \qquad \det M(\tilde{V}_n) = \sup_{V_n \in V^{(n)}} \det M(V_n),$$

and it is called *I-optimum* (relative to p) in $V^{(n)}$ for LSE if

$$(4) \qquad \int_H \operatorname{tr} A(x)M(\tilde{V}_n)^{-1}A(x)'p(x)\,dx = \inf_{V_n \in V^{(n)}} \int_H \operatorname{tr} A(x)M(V_n)^{-1}A(x)'p(x)\,dx$$

holds for some given weight function p on H. Here $\operatorname{tr} C$ denotes the trace of a matrix C.

Let $V_n^{(1)}$ and $V_n^{(2)}$ be two designs and let $\mathfrak{M}_r^{\geqslant}$ be the set of positive semidefinite matrices of order r. Then $V_n^{(1)}$ is called *better* than $V_n^{(2)}$ if

$$(5) \qquad M(V_n^{(1)}) - M(V_n^{(2)}) \in \mathfrak{M}_r^{\geqslant}, \quad M(V_n^{(1)}) \neq M(V_n^{(2)}).$$

This definition is appropriate to the usual notion of optimum design, for $V_n^{(1)}$ is preferred to $V_n^{(2)}$ by most of the known criteria if (5) holds. $\bar{V}_n$ is called *admissible* for LSE if there exists no design $V_n \in V^{(n)}$ which is better than $\bar{V}_n$. For example, D- and I-optimum designs are admissible in

$$V^{(n)} = \{V_n \in V^n \colon \det M(V_n) > 0\}$$

(see [1]). However, this classical conception of admissibility does not agree with an exact decision-theoretic definition of preference based on the risk function.

It is possible to decompose problem (2) and to obtain the components $\hat{\psi}^*$ and V_n^* of a Bayes strategy $(\hat{\psi}^*, V_n^*)$ separately as solutions of decision problems G_1 and G_2 with a less complicated structure. We first consider the *estimation problem*

$$G_1 = [T \times K, D, R]$$

for fixed $x \in H$ and $V_n \in V^{(n)}$. Let $P_{\Delta/x}$ be the conditional distribution of $\Delta = (\Theta, \Lambda)$ for given $x \in X$ and let $\hat{\psi}^* \in D$ satisfy

$$\forall x \in H\, \forall \hat{\psi} \in D\, \forall V_n \in V^{(n)}\colon\ \mathrm{E}_{\Delta/x} R(\Delta, x; \hat{\psi}^*, V_n) \leqslant \mathrm{E}_{\Delta/x} R(\Delta, x; \hat{\psi}, V_n). \tag{6}$$

This means that for any pair $(x, V_n) \in H \times V^{(n)}$ the estimator $\hat{\psi}^*$ is a Bayesian solution in G_1 with respect to $P_{\Delta/x}$. The decision problem

$$G_2 = [H, V^{(n)}, R^*]$$

with the risk

$$R^*(x, V_n) = \mathrm{E}_{\Delta/x} R(\Delta, x; \hat{\psi}^*, V_n) \tag{7}$$

is called the *designing problem* for $\hat{\psi}^*$. An optimum design $V_n^* \in V^{(n)}$ can be obtained as a Bayesian solution in G_2 with respect to the marginal prior distribution P_X , i.e.,

$$\mathrm{E}_X R^*(X; V_n^*) = \inf_{V_n \in V^{(n)}} \mathrm{E}_X R^*(X; V_n).$$

LEMMA 1. *Assume that the Bayes risk $\varrho(P_{(\Delta,X)}; \hat{\psi}, V_n)$ exists for every strategy $(\hat{\psi}, V_n) \in D \times V^{(n)}$. If $\hat{\psi}^*$ is an estimator satisfying* (6) *and V_n^* is Bayesian in G_2 with respect to P_X, then $(\hat{\psi}^*, V_n^*)$ is a Bayes strategy in G with respect to $P_{(\Delta,X)}$.*

Proof. By (6) and the Bayes optimality of V_n^*, for any $(\hat{\psi}, V_n) \in D \times V^{(n)}$ we have

$$\begin{aligned} \varrho(P_{(\Delta,X)}; \hat{\psi}, V_n) &= \mathrm{E}_X \mathrm{E}_{\Delta/X} R(\Delta, X; \hat{\psi}, V_n) \\ &= \int_H \mathrm{E}_{\Delta/x} R(\Delta, x; \hat{\psi}, V_n)\, dP_X(x) \geqslant \int_H \mathrm{E}_{\Delta/x} R(\Delta, x; \hat{\psi}^*, V_n)\, dP_X(x) \\ &= \mathrm{E}_X R^*(X; V_n) \geqslant \mathrm{E}_X R^*(X; V_n^*) = \varrho(P_{(\Delta,X)}; \hat{\psi}^*, V_n^*). \end{aligned}$$

Remark 1. Lemma 1 can be generalized for a separate determination of optimal components not only of Bayesian strategies but also of so-called Q-optimum strategies $(\hat{\psi}^*, V_n^*) \in D \times V^{(n)}$ minimizing the expression $QR(\cdot, \cdot; \hat{\psi}, V_n)$; Q is an operator acting on $T \times K \times H$ to make the risk independent of the parameters (θ, λ, x). For the Bayesian case,

$Q = \mathrm{E}_{(\Delta, X)}$ is to be chosen. The generalization, which includes, e.g., the minimax optimality ($Q = \sup\limits_{(\theta,\lambda,x)\in T\times K\times H}$), may be found in [2].

3. Estimation problem. Lemma 1 entitles us to consider at first the estimation problem G_1 with fixed, but arbitrary $x \in H$, $V_n \in V^{(n)}$. We will consider a Bayes solution in G_1 for a special but important case which meets many real situations.

ASSUMPTION 1. Let the random vectors Δ and X be independent, i.e., $P_{\Delta/x} = P_\Delta$.

ASSUMPTION 2. Let the loss function be quadratic, i.e.,

$$L\big(\psi(x, \theta); \hat{\psi}(x, y)\big) = \big(\psi(x, \theta) - \hat{\psi}(x, y)\big)' U \big(\psi(x, \theta) - \hat{\psi}(x, y)\big) =: \|\psi(x, \theta) - \hat{\psi}(x, y)\|_U^2,$$

where U is an arbitrary matrix from $\mathfrak{M}_m^{\geqslant}$.

ASSUMPTION 3. Let the observations be normally distributed, i.e.,

$$P_{Y(V_n)/\delta} = N\big(F(V_n)\theta, \lambda^{-1} I_n\big).$$

Assumption 3 implies, e.g., $T = \boldsymbol{R}^r$ and $K = \boldsymbol{R}^+$.

ASSUMPTION 4. Let P_Δ be a normal gamma-distribution with the density

$$p(\theta, \lambda | a, \nu, \mu, \Phi) = \begin{cases} c\lambda^{(\nu+r)/2-1} \exp\left\{-\dfrac{\lambda}{2a}(\nu + \|\theta - \mu\|_{\Phi^{-1}}^2)\right\} & \text{if } \theta \in \boldsymbol{R}^r,\ \lambda > 0, \\ 0 & \text{if } \lambda \leqslant 0 \end{cases}$$

($a > 0$, $\nu > 0$, $\mu \in \boldsymbol{R}^r$, $\Phi \in \mathfrak{M}_r^{>}$).

This is a conjugate prior distribution of

$$P_{Y(V_n)/\delta} = N\big(F(V_n)\theta, \lambda^{-1} I_n\big)$$

(see [8]). It will be abbreviated by

$$P_\Delta = NG(a, \nu, \mu, \Phi).$$

Under these four assumptions the estimator

$$(8) \quad \hat{\psi}^*\big(x, Y(V_n)\big) = A(x)\left[F(V_n)'F(V_n) + \frac{1}{a}\Phi^{-1}\right]^{-1}\left[F(V_n)'Y(V_n) + \frac{1}{a}\Phi^{-1}\mu\right]$$

is Bayesian in $G_1 = [T \times K, D, R]$ with respect to P_Δ (see [8]).

PROPOSITION 1. *The estimator $\hat{\psi}^*$ is admissible in G_1.*

Indeed, under assumptions 1-4, $\hat{\psi}^*$ is the unique BE. Moreover, $\hat{\psi}^*$ is robust, to a certain extent, under a change of the loss function and the prior distribution:

PROPOSITION 2. *Let assumptions* 1, 3 *and* 4 *be satisfied. Then* $\hat{\psi}^*$ *is also a Bayes estimator relative to loss functions*

$$L(\psi(x,\theta);\hat{\psi}(x,y)) = L(\|\psi(x,\theta)-\hat{\psi}(x,y)\|_U^2)$$

which are non-decreasing functions of the argument $\|\psi(x,\theta)-\hat{\psi}(x,y)\|_U^2$.

This result follows from Lemma 1 in [6], which is proved in a similar way as Theorem 4.6.1 in [3], by observing that the marginal prior $P_{\Theta/y}$ of Θ is a t-distribution with the mode

$$\left[F(V_n)'F(V_n)+\frac{1}{a}\Phi^{-1}\right]^{-1}\left[F(V_n)'y+\frac{1}{a}\Phi^{-1}\mu\right].$$

If the prior distribution is not known precisely but is a member of the set

$$\mathfrak{P} = \left\{P_\Delta\colon \mathrm{E}(\Theta|\lambda) = \mu, \mathrm{Var}(\Theta|\lambda) = \frac{a}{\lambda}\Phi, \mathrm{E}\Lambda = a\right\},$$

then the estimator $\hat{\psi}^*$ turns out to be restricted minimax:

PROPOSITION 3. *Under assumptions* 1, 2 *and* 3 *we have* (see [9])

$$\sup_{P_\Delta\in\mathfrak{P}}\mathrm{E}_\Delta R(\Delta,x;\hat{\psi}^*,V_n) = \inf_{\hat{\psi}\in D}\sup_{P_\Delta\in\mathfrak{P}}\mathrm{E}_\Delta R(\Delta,x;\hat{\psi},V_n).$$

It can be shown that the estimator $\hat{\psi}^*$ is also optimum in some sense relative to certain forms of non-Bayesian prior knowledge. This is, for example, the case if it is known a priori that

$$(\theta-\mu)'\Phi^{-1}(\theta-\mu)\leqslant 1$$

(i.e., the regression parameter belongs to an ellipsoid with the centre μ) and $\lambda\in[a,\infty)$ (i.e., a is the least precision of the observation or, equivalently, a^{-1} is an upper bound for the variance).

PROPOSITION 4. *Let* $T_0 = \{\theta\in\boldsymbol{R}^r\colon (\theta-\mu)'\Phi^{-1}(\theta-\mu)\leqslant 1\}$, $K_0 = [a,\infty)$ *and let* D_L *be the set of all linear estimators for* $\psi(x,\theta)$. *If*

$$\operatorname{rank} A(x)'UA(x) = 1,$$

then $\hat{\psi}^*(x, Y(V_n))$ *is minimax in* $G_1 = [T_0\times K_0, D_L, R]$.

Proof. Any linear estimator $\hat{\psi}\in D_L$ can be written as

$$\hat{\psi}(x, Y(V_n)) = A(x)[WY(V_n)+w]$$

with some $(r\times n)$-matrix W and $w\in\boldsymbol{R}^n$. Moreover, since $\operatorname{rank}A(x)'UA(x) = 1$, there exists a vector $u\in\boldsymbol{R}^r$ such that $A(x)'UA(x) = uu'$. Consequently,

$$\begin{aligned}R(\theta,\lambda,x;\hat{\psi},V_n) &= \mathrm{E}_{Y(V_n)/\theta,\lambda}\|\theta-WY(V_n)-w\|_{uu'}^2\\ &= \{[w+(WF(V_n)-I_r)\theta]'u\}^2+\frac{1}{\lambda}u'WW'u\\ &=: q(\theta,\lambda,W,w).\end{aligned}$$

Let

$$W_\lambda = [F(V_n)'F(V_n)+(\lambda\Phi)^{-1}]^{-1}F(V_n)' \quad \text{and} \quad w_\lambda = [I_r - W_\lambda F(V_n)]\mu.$$

Then we have

$$\sup_{\theta\in T_0} q(\theta, \lambda, W_\lambda, w_\lambda) = \lambda^{-1}u'W_\lambda W_\lambda' u + \sup_{\tilde\theta\in\tilde T_0}\{u'[W_\lambda F(V_n)-I_r]\tilde\theta\}^2,$$

where $\tilde T_0 = \{\tilde\theta \in \boldsymbol{R}^r: \tilde\theta'\Phi^{-1}\tilde\theta \leqslant 1\}$. Now, for any λ

$$\begin{aligned}\sup_{\theta\in T_0} q(\theta, \lambda, W_\lambda, w_\lambda) &= \lambda^{-1}u'[F(V_n)'F(V_n)+(\lambda\Phi)^{-1}]^{-1}u\\ &= \inf_{W,w}\sup_{\theta\in T_0} q(\theta, \lambda, W, w) = \sup_{\theta\in T_0}\inf_{W,w} q(\theta, \lambda, W, w)\end{aligned}$$

(see [5], Sections 2 and 3). Thus we have

$$\begin{aligned}\sup_{\lambda\in K_0}\sup_{\theta\in T_0} q(\theta, \lambda, W_\lambda, w_\lambda) &= \sup_{\lambda\in K_0}\sup_{\theta\in T_0}\inf_{\hat\psi\in D_L} R(\theta, \lambda, x; \hat\psi, V_n)\\ &\leqslant \inf_{\hat\psi\in D_L}\sup_{\lambda\in K_0}\sup_{\theta\in T_0} R(\theta, \lambda, x; \hat\psi, V_n)\end{aligned}$$

(the last inequality is known from game theory) and

$$\begin{aligned}\sup_{\lambda\in K_0}\sup_{\theta\in T_0} q(\theta, \lambda, W_\lambda, w_\lambda) &= a^{-1}u'[F(V_n)'F(V_n)+(a\Phi)^{-1}]^{-1}u\\ &= \sup_{\lambda\in K_0}\sup_{\theta\in T_0} q(\theta, \lambda, W_a, w_a).\end{aligned}$$

But

$$\hat\psi^*(x, Y(V_n)) = A(x)[F(V_n)'F(V_n)+(a\Phi)^{-1}]^{-1}[F(V_n)'Y(V_n)+(a\Phi)^{-1}\mu]$$

can be written as

$$\hat\psi^*(x, Y(V_n)) = A(x)[W_a Y(V_n)+w_a].$$

Hence

$$\begin{aligned}\sup_{\lambda\in K_0}\sup_{\theta\in T_0} q(\theta, \lambda, W_\lambda, w_\lambda) &= \sup_{\lambda\in K_0}\sup_{\theta\in T_0} R(\theta, \lambda, x; \hat\psi^*, V_n)\\ &\leqslant \inf_{\hat\psi\in D_L}\sup_{\lambda\in K_0}\sup_{\theta\in T_0} R(\theta, \lambda, x; \hat\psi, V_n).\end{aligned}$$

It should be noted that no distributional assumptions are necessary for this minimax property (with the exception that the covariance matrix of $Y(V_n)$ is $\lambda^{-1}I_n$ with $\lambda \geqslant a$).

A further and more detailed analysis of estimators involving certain non-Bayesian prior knowledge, but having the same structure as the BE (8), may be found in [2] and [7].

Under certain conditions $\hat\psi^*$ can be expanded in a series which makes it possible to compute the BE approximately, starting with the LSE $\tilde\psi$. To do so let $\|S\|$ denote the usual Euclidean norm of a matrix $S = (s_{ij})$, i.e.,

$$\|S\| = (\operatorname{tr} S'S)^{1/2} = \left(\sum s_{ij}^2\right)^{1/2}.$$

PROPOSITION 5. *Assume that* $\operatorname{rank} F(V_n) = r$. *Let*

$$J(V_n) = -\frac{1}{a}[\Phi F(V_n)'F(V_n)]^{-1}$$

and let $\tilde{\psi}(x, Y(V_n)) = A(x)\tilde{\theta}(Y(V_n))$ *be the LSE of* $\psi(x, \theta)$. *If* $\|J(V_n)\| < 1$, *then*

$$\hat{\psi}^*(x, Y(V_n)) = A(x)\Big\{\tilde{\theta}(Y(V_n)) - J(V_n)\mu + \sum_{k=1}^{\infty} J(V_n)^k [\tilde{\theta}(Y(V_n)) - J(V_n)\mu]\Big\}.$$

Proof. In functional analysis the series expansion

$$[J^0 - J]^{-1} = \sum_{k=0}^{\infty} J^k$$

is known to hold for any linear operator J in a Banach space $(E, \|\cdot\|_E)$ if

$$\sup_{e \in E_0} \|Je\|_E < 1, \qquad E_0 = \{e \in E \colon \|e\|_E \leqslant 1\},$$

where J^0 denotes the unity operator in E. Now, $\hat{\psi}^*(x, Y(V_n))$ can be written as

$$\begin{aligned}\hat{\psi}^*&(x, Y(V_n)) \\ &= A(x)[I_r + (a\Phi F(V_n)'F(V_n))^{-1}]^{-1}(F(V_n)'F(V_n))^{-1}\Big(F(V_n)'Y(V_n) + \\ &\quad + \frac{1}{a}\Phi^{-1}\mu\Big) \\ &= A(x)[I_r - J(V_n)]^{-1}[\tilde{\theta}(Y(V_n)) - J(V_n)\mu].\end{aligned}$$

The result follows with $E_0 = \{e \in \boldsymbol{R}^r \colon e'e \leqslant 1\}$ by observing that

$$\sup_{e \in E_0} \|J(V_n)e\|_{I_r}^2 = \sup_{e \in E_0} e'J(V_n)'J(V_n)e \leqslant \operatorname{tr} J(V_n)'J(V_n) = \|J(V_n)\|^2 < 1.$$

Remark 2. The condition $\|J(V_n)\| < 1$ is satisfied, for example, if the prior knowledge about the regression parameter is sufficiently vague (i.e., if a is sufficiently large). In the limit case $a \to \infty$ the BE $\hat{\psi}^*$ and the LSE $\tilde{\psi}$ coincide.

4. Designing problem. According to Lemma 1 we now consider the designing problem $G_2 = [H, V^{(n)}, R^*]$ for the BE $\hat{\psi}^*$. At first we compute the risk R^* defined by (7), i.e., the Bayes risk of $\hat{\psi}^*$ with respect to $P_\Delta = NG(a, \nu, \mu, \Phi)$.

Let $x \in H$ and $S(x) = A(x)' U A(x)$.

LEMMA 2. *If assumptions* 1-4 *are satisfied, then*

$$R^*(x, V_n) = \mathrm{E}_\Lambda R(\Lambda, x; \hat{\psi}^*, V_n) \tag{9}$$
$$= [\mathrm{E}(1/\Lambda)] \operatorname{tr}\{S(x)[F(V_n)'F(V_n) + (a\Phi)^{-1}]^{-1}\}.$$

Proof. Let $F = F(V_n)$ and $\Phi_1 = [F'F + (a\Phi)^{-1}]^{-1}$. With the assumed quadratic loss we get

$$R(\theta, \lambda, x; \hat{\psi}^*, V_n) = \operatorname{tr} S(x) \Phi_1 \left[(a\Phi)^{-1}(\theta - \mu)(\theta - \mu)'(a\Phi)^{-1} + \frac{1}{\lambda} F'F\right] \Phi_1.$$

From the definition of $P_\Delta = NG(a, \nu, \mu, \Phi)$ it follows that the conditional distribution $P_{\Theta/\lambda}$ is normal with mean μ and covariance matrix $(a/\lambda)\Phi$, so that

$$R^*(x, V_n) = \mathrm{E}_\Lambda \mathrm{E}_{\Theta/\Lambda} R(\Theta, \Lambda, x; \hat{\psi}^*, V_n)$$
$$= \mathrm{E}_\Lambda \left\{ \operatorname{tr} S(x) \Phi_1 \left[(a\Phi)^{-1} \frac{a}{\Lambda} \Phi (a\Phi)^{-1} + \frac{1}{\Lambda} F'F\right] \Phi_1 \right\}$$
$$= [\mathrm{E}(1/\Lambda)] \operatorname{tr} S(x) \Phi_1.$$

With the prior distribution P_X on the domain of forecast H an optimum design V_n^* is to be determined according to Lemma 1 and formula (9) by solving the problem

$$\mathrm{E}_X \operatorname{tr} S(x)[nM(V_n^*) + (a\Phi)^{-1}]^{-1} \tag{10}$$
$$= \inf_{V_n \in V^{(n)}} \mathrm{E}_X \operatorname{tr} S(x)[nM(V_n) + (a\Phi)^{-1}]^{-1}.$$

A design V_n^* satisfying (10) may be called *I-optimum* for the BE. Indeed, if P_X has a density $p(\cdot)$ with respect to Lebesgue measure and if $U = I_m$, then equation (10) takes the form

$$\int_H \operatorname{tr} A(x)[nM(V_n^*) + (a\Phi)^{-1}]^{-1} A(x)' p(x)\, dx$$
$$= \inf_{V_n \in V^{(n)}} \int_H \operatorname{tr} A(x)[nM(V_n) + (a\Phi)^{-1}]^{-1} A(x)' p(x)\, dx$$

and the analogy with (4) is obvious. Here the matrix

$$\bar{M}(V_n) = M(V_n) + \frac{1}{an} \Phi^{-1} \tag{11}$$

takes over the role of the information matrix for the estimator $\hat{\psi}^*$.

Remark 3. $(an\bar{M}(V_n))^{-1}$ is the covariance matrix of the preposterior marginal distribution of Θ, i.e. (see [2]),

$$\mathrm{E}\{\operatorname{Var}(\Theta \mid Y(V_n))\} = (an\bar{M}(V_n))^{-1}.$$

In the sense of a proper decision theoretic definition we say that a design $V_n^{(1)} \in V^{(n)}$ is *R^*-better* than a design $V_n^{(2)} \in V^{(n)}$ if

$$\forall x \in H\colon\ R^*(x;\ V_n^{(1)}) \leqslant R^*(x;\ V_n^{(2)})$$

and

$$\exists x_0 \in H\colon\ R^*(x_0;\ V_n^{(1)}) < R^*(x_0;\ V_n^{(2)}).$$

The design $\bar{V}_n$ is called *R^*-admissible* for $\hat{\psi}^*$ if there exists no design $V_n \in V^{(n)}$ which is R^*-better than $\bar{V}_n$. Let $\mathfrak{M}[Q]$ and $\mathfrak{L}(E)$ denote the column space of a matrix Q and the linear space generated by the linear combinations of vectors of a linear manifold E, respectively. Now the following interesting connection of usual admissibility with R^*-admissibility can be established.

THEOREM 1. *Let U be of full rank m and assume that*

$$\mathfrak{L}\Big(\bigcup_{x \in H} \mathfrak{M}[A(x)'\,U^{1/2}]\Big) = \boldsymbol{R}^r. \tag{12}$$

If assumptions 1-4 are satisfied, then every design being R^*-admissible for the BE $\hat{\psi}^*$ is admissible also for the LSE.

Proof. Assume that $\bar{V}_n \in V^{(n)}$ is R^*-admissible for $\hat{\psi}^*$ but not admissible for LSE. Consequently, there exists a design $V_n \in V^{(n)}$ such that $M(V_n) - M(\bar{V}_n)$ is positive semidefinite. This implies

$$\bar{M}(\bar{V}_n)^{-1} - \bar{M}(V_n)^{-1} \in \mathfrak{M}_r^{\geqslant},$$

so that

$$\begin{aligned}&\operatorname{tr} S(x)\,[\bar{M}(\bar{V}_n)^{-1} - \bar{M}(V_n)^{-1}]\\ &\qquad = \operatorname{tr} U^{1/2} A(x)\,[\bar{M}(\bar{V}_n)^{-1} - \bar{M}(V_n)^{-1}]\,A(x)'\,U^{1/2} =: \operatorname{tr} Q(x) \geqslant 0\end{aligned}$$

for every $x \in H$. Moreover, there exists at least one point in H for which $\operatorname{tr} Q(x)$ is positive. Otherwise, we would infer that for every $x \in H$ all eigenvalues of $Q(x)$ are zero, i.e., $Q(x)$ is the null matrix. This would imply

$$\forall x \in H\ \forall z \in \boldsymbol{R}^m\colon\ z'Q(x)z = 0$$

or, equivalently,

$$\forall x \in H\ \forall q \in \mathfrak{M}[A(x)'\,U^{1/2}]\colon\ q'[\bar{M}(\bar{V}_n)^{-1} - \bar{M}(V_n)^{-1}]q = 0$$

which is impossible because of condition (12) and $\bar{M}(\bar{V}_n) \neq \bar{M}(V_n)$. Thus

$$\begin{aligned}R^*(x;\ V_n) &= n^{-1}[\mathrm{E}(1/\Lambda)] \operatorname{tr} S(x) \bar{M}(V_n)^{-1}\\ &\leqslant n^{-1}[\mathrm{E}(1/\Lambda)] \operatorname{tr} S(x) \bar{M}(\bar{V}_n)^{-1} = R^*(x;\ \bar{V}_n)\end{aligned}$$

for every $x \in H$ and strict inequality holds for at least one point in H.

Remark 4. For a special case where $A(x) = f(x)'$, i.e., $m = 1$ and $\psi(x, \theta) = f(x)'\theta$, and U is a positive real number, (12) reduces to the condition that H contains at least r points $x_1, \ldots, x_r$ for which the vectors $f(x_1), \ldots, f(x_r)$ are linearly independent.

The inversion of the statement of Theorem 1 does not hold in general. But if we define "better than" analogously to (5), i.e., $V_n^{(1)}$ is called *better* for BE than $V_n^{(2)}$ if

$$\bar{M}(V_n^{(1)}) - \bar{M}(V_n^{(2)}) \in \mathfrak{M}_r^{\geqslant},$$

then admissibility of designs based on LSE and BE, respectively, coincides.

In analogy with (3) a design V_n^+ is called D-optimum in $V^{(n)}$ for BE if

$$\det \bar{M}(V_n^+) = \sup_{V_n \in V^{(n)}} \det \bar{M}(V_n).$$

This means that a D-optimum design minimizes the generalized variance $\det\big(\mathrm{E}\{\mathrm{Var}(\Theta \mid Y(V_n))\}\big)$ of the preposterior marginal distribution of the regression coefficients (see Remark 3).

Obviously, I- and D-optimum designs for BE are R^*-admissible. Because of the invariance of the admissibility of designs relative to the underlying BE and LSE it is possible that even the optimal designs for $\hat{\psi}^*$ and $\tilde{\psi}$, respectively, coincide.

For the case of D-optimality this is true if we consider the simple linear regression model

$$\mathrm{E}Y(x) = \theta_1 + \theta_2 x, \qquad V = [-1, +1]. \tag{13}$$

Then we have

$$M(V_n) = \begin{pmatrix} 1 & m_1 \\ m_1 & m_2 \end{pmatrix}, \quad m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

and the D-optimum design V_n^+ for LSE in model (13) is given by

$$V_n^+ = \begin{cases} (n/2)x_i = -1, \ (n/2)x_i = 1 & \text{if } n \text{ is even}, \\ [(n+1)/2]x_i = -1, \ [(n-1)/2]x_i = 1 & \text{if } n \text{ is odd} \end{cases} \tag{14}$$

(see [4]).

Let the matrix Φ of the prior distribution be diagonal,

$$\Phi = \operatorname{diag}(\varphi_1, \varphi_2), \tag{15}$$

i.e., the regression coefficients are uncorrelated a priori. The D-optimum design for $\hat{\varphi}^*$ maximizes

$$(16)\qquad \det\bar{M}(V_n) = \det\left(M(V_n)+\frac{1}{an}\Phi^{-1}\right) = m_2-m_1^2+\frac{m_2}{an\varphi_1}+\frac{1+1/an\varphi_1}{an\varphi_2},$$

i.e., it maximizes

$$m_2-m_1^2+\frac{m_2}{an\varphi_1} = \det M(V_n)+\frac{m_2}{an\varphi_1}.$$

The first summand is maximized by V_n^+ from (14) and the second summand attains its maximum value $1/an\varphi_1$ also at V_n^+. Thus V_n^+ maximizes (16) and we have proved

THEOREM 2. *Under condition* (15), *the D-optimum design for the LSE is D-optimum for the BE in model* (13).

Concluding, we remark that in the asymptotic case ($n\to\infty$) the designing problem based on $M(V_n)$ coincides with the designing problem based on $\bar{M}(V_n)$ (see (11)). This reflects the fact that in the asymptotic case the sample information dominates the prior information.

References

[1] H. Bandemer, A. Bellmann, W. Jung und K. Richter, *Optimale Versuchsplanung*, Berlin 1973.

[2] — *Theorie und Praxis der optimalen Versuchsplanung*, vol. I, Berlin 1977.

[3] T. S. Ferguson, *Mathematical statistics, A decision-theoretic approach*, New York 1967.

[4] G. Hohmann and W. Jung, *On sequential and nonsequential D-optimal design*, Biom. Zeitschr. 17 (1975), p. 329-336.

[5] Й. Кукс и В. Ольман, *Минимаксная линейная оценка коэффициентов регресии*, Изв. АН Эст. ССР, Физ. мат., 21 (1972), p. 66-72.

[6] W. Näther, *The choice of estimators and experimental designs in a linear regression model according to a joint criterion of optimality*, Math. Operat. u. Statist. 6 (1975), p. 677-686.

[7] J. Pilz, *Bayesian estimation and design in linear regression models*, in print in Teubner-Verlag Leipzig.

[8] H. Raiffa and R. Schlaifer, *Applied statistical decision theory*, Boston 1961.

[9] S. L. Wind, *An empirical Bayes approach to multiple linear regression*, Ann. Statist. 1 (1973), p. 93-103.

SEKTION MATHEMATIK
BERGAKADEMIE FREIBERG
DDR-92 FREIBERG

Received on 26. 6. 1975;
revised version on 29. 7. 1977

W. NÄTHER i J. PILZ (Freiberg, NRD)

ESTYMACJA I PLANOWANIE DOŚWIADCZEŃ W LINIOWYM MODELU REGRESJI PRZY UŻYCIU INFORMACJI APRIORYCZNEJ

STRESZCZENIE

W pracy przedstawia się zagadnienie planowania doświadczeń w liniowym modelu regresji przy użyciu informacji apriorycznej jako bayesowski problem decyzji statystycznej, który można rozłożyć na problem estymacji i problem planowania doświadczenia. Dla zwykłego estymatora bayesowskiego z kwadratową funkcją straty, obserwacji o rozkładach normalnych i sprzężonego rozkładu apriorycznego wykazuje się jego odporność na zmianę straty i rozkładu apriorycznego i dowodzi pewnej własności minimaksowej. Formułuje się problem wyboru optymalnego doświadczenia dla estymatora bayesowskiego i stwierdza pewną niezmienniczość zwykłego pojęcia doświadczeń dopuszczalnych. Ponadto w przypadku prostego modelu regresji liniowej dowodzi się, że doświadczenia D-optymalne dla estymatorów najmniejszych kwadratów i bayesowskiego są jednakowe, jeśli współczynniki regresji są a priori nieskorelowane.