EXACT D-OPTIMAL DESIGNS FOR THE REGRESSION LINE WITH LINEAR OR RELAY PERTURBED VARIANCE OF OBSERVATIONS

V. P. Kirlitsa
Belarusain State University
Minsk, Belarus
e-mail: `Kirlitsa@bsu.by`

Abstract

Conditions when the design for model of heteroscedastic observations subject to linear or relay perturbation turns to the design for homoscedastic observations are found in this article.

Let the experimental results at the design points $x_i, i = 1, 2, \ldots, n$, be described by equation:

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon(x_i), \quad i = 1, \ldots, n, \tag{1}$$

where y_i are observed variables; θ_0, θ_1 are unknown parameters; x_i are controllable variables from interval $[-1; 1]$; $\varepsilon(x_i)$ are random errors with mean zero.The variance $D\{\varepsilon(x_i)\} = d(x_i)$ accepts constant value $d_1, d_1 > 0$, when $x_i \in [-1, 1], -1 < c \leq 1$, i.e. on this interval $(c, 1]$ variance $d(x)$ has linear perturbation

$$d(x) = \frac{(d_1 - d_2)x + d_2 c - d_1}{c - 1}, \quad d_2 > 0, \tag{2}$$

or it has relay perturbation

$$d(x) = d_2, \quad d_2 \neq d_1, \tag{3}$$

where d_1, d_2, c are the fixed constants.

For homoscedastic observations, i.e. $d(x) = d_1, x \in [-1, 1]$, D-optimal design is

$$\varepsilon_n^0 = \left\{ \begin{array}{cc} -1, & 1 \\ n_1, & n - n_1 \end{array} \right\}, \tag{4}$$

where $n_1 = s$ for even $n = 2s$ and $n_1 = s$ or $n_1 = s + 1$ for odd $n = 2s + 1$.The estimations of unknown parameters which is constructed under the design (4) do not depend from d_1 and are

$$\hat{\theta}_0 = \frac{1}{2n_1(n - n_1)} \left\{ (n - n_1) \sum_{i=1}^{n_1} y_{1i} + n_1 \sum_{i=1}^{n-n_1} y_{2i} \right\}, \tag{5}$$

$$\hat{\theta}_1 = \frac{1}{2n_1(n - n_1)} \left\{ n_1 \sum_{i=1}^{n-n_1} y_{2i} - (n - n_1) \sum_{i=1}^{n_1} y_{1i} \right\}, \tag{6}$$

where y_{1i} are observations in the point -1 and y_{2i} are observations in the point 1.

In this paper it will be shown that the design (4) and corresponding to it estimations (5), (6) not to change for defined parameters d_1, d_2, c which determines perturbation of homoscedastic observations.

The basic results of this paper are based on the following theorem.

Theorem 1. *Support points of exact D-optimal design for model of observations (1) with linear perturbations of variance (2) (or relay perturbation of variance (3)) are -1, c, 1.*

Proof of this theorem is omitted.The full proof of this theorem is presented in [2].

From the theorem 1 follow that exact D-optimal design ε_n^0 is

$$\varepsilon_n^0 = \left\{ \begin{array}{ccc} -1, & c, & 1 \\ n_1^0, & n_2^0 & n - n_1^0 - n_2^0 \end{array} \right\}, \tag{7}$$

where n_1^0, n_2^0 are solutions of a problem of quadratic programming

$$f(n_1, n_2) = \frac{1}{kd_1^2} \left\{ -4n_1^2 - (c-1)^2 n_2^2 + b(k,c) n_1 n_2 + 4nn_2 + n(c-1)^2 n_2 \right\} \to \max, \tag{8}$$

where

$$b(k,c) = k(1+c)^2 - c^2 + 2c - 5, k = \frac{d_2}{d_1}$$

and maximization in (8) is taken on set

$$G = \{n_1, n_2 : 0 \le n_1 \le n-1, 0 \le n_2 \le n-1, 1 \le n_1 + n_2 \le n\}. \tag{9}$$

We investigate the following problem. What is the set of parameters d_1, c, d_2 which can guarantee that design (7) for heteroscedastic observations convert to the design (4) for homoscedastic observations? There are two approaches to resolve this problem.

The first approach consists in numerical solution of a problem of quadratic programming (8), (9) and choose those values d_1, c, d_2 which guarantee that the design (7) convert to the design (4), i.e. it is necessary to find some conditions which allow to conclude that $n_2^0 = 0$. One effective method of solution numerical the formulated problem (8), (9) is presented in paper [3]. But this method to determinate conditions when $n_2^0 = 0$ is a little effective.

The second approach to our problem consist in finding the analytical conditions which can help us to form optimal design of experiments (4) stable to perturbation of variance $d(x)$.

In article [4] it is proved that design (4) does not change if variance of observations $d(x)$ satisfies to an inequality

$$d(x) \ge \varphi(x), x \in [-1, 1], \tag{10}$$

where

$$\varphi(x) = \frac{d_1}{4} \left\{ (1+k)x^2 + 2(k-1)x + 1 + k \right\}, \tag{11}$$

and (11) is a parabola which is passing through points (-1,d_1) and (1,d_2).

In the case when the inequality (10) is not holds the spectrum of the optimal design (7) can have two or three points.

Our task is investigate when inequality (10) holds good, i.e. when perturbation of variance (2), or (3) not influence on optimal design (4).

At first we will consider a case of linear perturbation (2) then we will consider a case of relay perturbation (3).

1. Case of linear perturbation. In this section we will consider six subcases.

1.1. In the case when $c = -1$ the inequality (10) holds. In this case the variance of observations linearly increases or decreases on the interval [-1,1]. It follows from the fact that function (11) is convex. Thus in this case the design (4) and estimations of unknown parameters corresponding to it remain invariable. Estimations of unknown parameters do not depend on values d_1 and d_2. It follows from this that they are mutually eliminated in the course of their construction.

1.2. In the case when $0 < k \leq 1$, i.e. if $0 < d_2 \leq d_1$, inequality (10) holds and it guarantee that design (4) and corresponding to it estimations (5), (6) remains invariable. It follows from this that function $\varphi(x)$ is convex and function $d(x)$ is concave for $x \in [-1, 1]$.

1.3. In the case when $k > 1$, i.e. if $d_2 > d_1$, inequality (10) holds if $-1 < c \leq c_1$ where

$$c_1 = \frac{3-k}{1+k}$$

is solution of equation $\varphi(x) = d_1$. The substation of this fact is reduced in paper [4]. So, in the case $d_2 > d_1$ and $-1 < c \leq c_1$ D-optimal design for heteroscedastic observations look like (4) and estimations of unknown parameters corresponding to it are equal (5), (6) and not depend from $d_1, \quad d_2$. However in this subcase it is necessary to know the value exact of k, i.e. to know precisely in how much time d_2 is more d_1. It as a rule is not known precisely for the experimenter.This restruction will be eliminated in a following subcases.

1.4. Let $k > 1, k \in [k_-, \quad k_+], \quad k_- > 1$, where $k_-, \quad k_+$ are fixed and known to the experimenter. Function

$$f(k) = \frac{3-k}{1+k}$$

decreases monotonically on the interval $[k_-, \quad k_+]$ as its derivative is negative. Hence, if k varies from k_- to k_+ value c_1 belong to an interval $[c_{1-}, \quad c_{1+}]$ where

$$c_{1-} = \frac{3-k_+}{1+k_+}, \quad c_{1+} = \frac{3-k_-}{1+k_-}.$$

Using outcomes of subcase 1.3 we conclude that if $c \leq c_{1-}$ in this case the design (7) is converted in the design (4) for all $k \in [k_-, k_+]$.

1.5. In this subcase we will assume that values c and k are not known precisely and $k > 1$. But the experimenter know precisely that $c \in [c_-, c_+], k \in [k_-, k_+], k_- > 1$, where c_-, c_+, k_-, k_+ are fixed values. If $c_+ \leq c_{1-}$, it agree a subcase 1.4, we can conclude that the optimal design (7) is convert in the design (4) for homoscedastic observations.

1.6. In this subcase we will assume, in difference from the previous subcase, that c, k are random variables. In this subcase it is possible to indicate conditions when the design (7) convert to the design (4). However these conditions will be fulfilled with certain probability which we can calculate. The limited volume of this paper does not allow to investigate this subcase completely.

2. Case of relay perturbation. Case of relay perturbation is investigated in the same way as a case 1. The limited volume of given article also does not allow to investigate this case in details.

References

[1] Moyssiadis C., Kounias S., (1983). Exact D-optimal observations 2^k designs of resolutions III, when $N \equiv 1$ or 2 mod 4. *Math. Operationsforch. U. Statist.* Vol. **14**, pp. 367-379.

[2] Kirlitsa V.P. (2009). Exact D-optimal designs for line regression with piecewise linear changes variance of observations. *Probability theory, mathematical statistics and their applications. The collection of scientific papers.* Vol. **2**, pp. 56-61.(In russian).

[3] Kirlitsa V.P. (2010). Exact D-optimal designs of experiments for line regression which are robust respect to linear perturbation of variance of observations.*Probability theory, mathematical statistics and their applications. The collection of scientific papers.*Vol. **3**, pp. 127-131.(In russian).

[4] Kirlitsa V.P. (2008). D-optimal designs of experiments robust with respect to change variance of observations for line regression. *Vestnik BSU* ,Vol. **1**, pp. 165-167. (In russian).