Optimal Design for Experiments with Potentially Failing Trials

Hackl, Peter

Published: 01/01/1994

Document Version
Publisher's PDF, also known as Version of record

Optimal Design for Experiments with Potentially Failing Trials

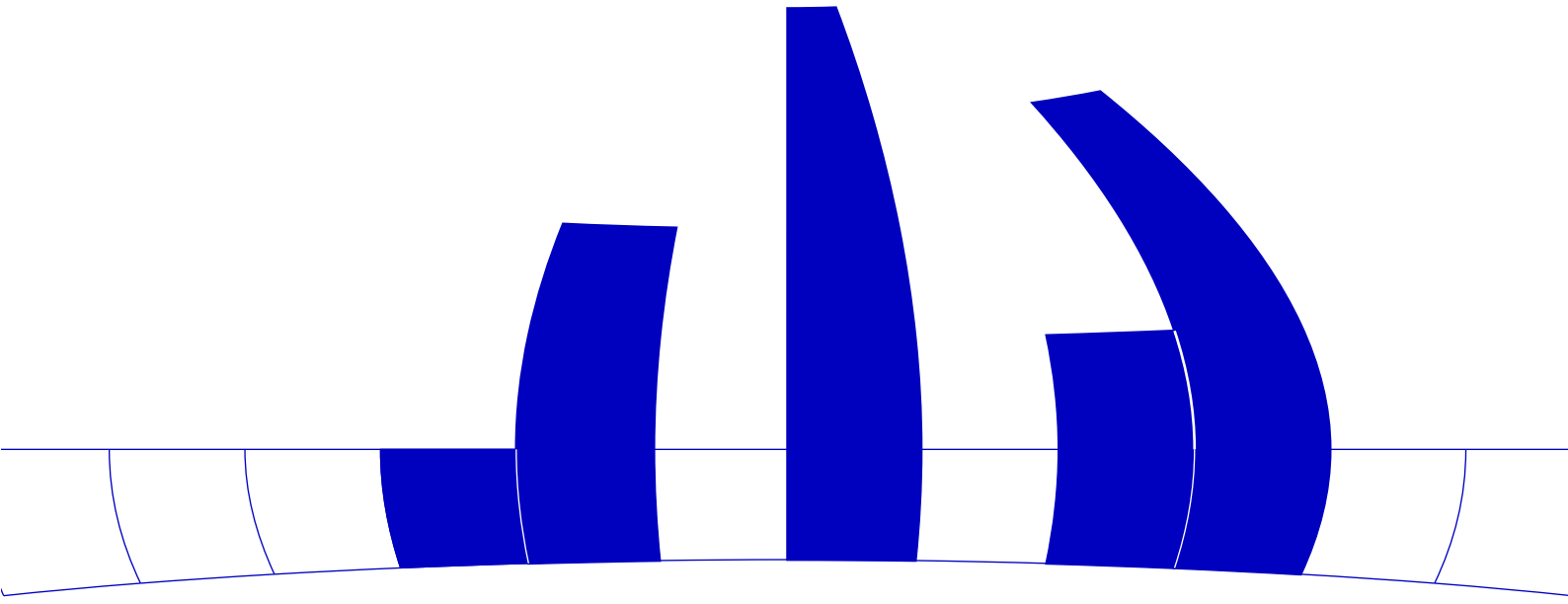
Peter Hackl

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 39
October 1994

http://statmath.wu-wien.ac.at/



Optimal Design for Experiments with Potentially Failing Trials

by

Peter Hackl[1]

Abstract: We discuss the problem of optimal allocation of the design points of an experiment for the case where the trials may fail with non-zero probability. Numerical results for D-optimal designs are given for estimating the coefficients of a polynomial regression. For small sample sizes these designs may deviate substantially from the corresponding designs in the case of certain response. They can be less efficient, but are less affected by failing trials.

Keywords: polynomial regression, failing trials, D-optimality.

[1]Department of Statistics, University of Economics and Business Administration, Vienna, Austria.

1 Introduction

Application of the standard theory of optimal design is based on the assumption that all trials of an experiment result in corresponding observations of the response variable. However, we are never sure that the responses of all trials will really be available when the experiment is actually performed. This leads us to consider designs for potentially failing trials, i.e., designs where the probability for getting a response is less than one for some or all sites in the design space.

The problem of potentially failing trials has been treated by Hedayat and Majumdar (1983) and later by Das and Sinha (1994) in the context of redesigning experiments due to shortened resources: Which observation should be dropped given the probabilities that the trials will be failing at the sites of the planned experiment. Our paper discusses a more general question: What is the optimal experiment in the case of possibly failing trials.

The paper is organized as follows. In Section 2 we state the problem and present criteria that allows us to assess the optimality of candidate designs. In Section 3 we show on the basis of a quadratic polynomial how D-optimal designs for estimating the model coefficients change due to failing trials.

2 Statement of the Problem

Let us assume that the response variable Y is determined by a regressor variable x according to the model

$$Y = \beta_1 + \beta_2 f_2 + \ldots + \beta_m f_m + u = f^T \beta + u \tag{1}$$

where the components of $f = [1, f_2, \ldots, f_m]^T$ are functions of x: $f_i = f_i(x)$, and β is an m-vector, and the error term u has expectation zero and variance σ^2 for all x; the error terms are assumed to be pairwise independent.

Our aim is to estimate the coefficient vector β. For that reason we plan to take a sample of Y's at a set of sites. The set $\mathcal{X}$ of all sites of interest is the design

space for our experiment. A design is represented by a measure ξ over $\mathcal{X}$; given that $\mathcal{X}$ comprises k points, the design is written as

$$\xi = \left\{ \begin{array}{ccc} x_1 & \dots & x_k \\ n_1 & \dots & n_k \end{array} \right\}$$

or shorter as $\xi = \{x_1, \dots, x_k; n_1, \dots, n_k\}$, where x_i, $i = 1, \dots, k$, are distinct elements from $\mathcal{X}$ and the integers n_i, $i = 1, \dots, k$, indicate the number of observations taken at x_i, and $\sum_{i=1}^{k} n_i = r$.

For such an r-trial design, the information matrix for the parameter vector β is $F^T F = \sum_{i=1}^{k} f_i f_i^T n_i$; the matrix F consists of k rows f_i^T. The information matrix per observation is

$$M(\xi) = \frac{1}{r} \sum_{i=1}^{k} f_i f_i^T n_i = \frac{1}{r} F^T F = \frac{\sigma^2}{r} [D_{\hat{\beta}}(\xi)]^{-1} \tag{2}$$

where $D_{\hat{\beta}}(\xi)$ is the corresponding covariance matrix of the LS-estimator $\hat{\beta}$. The covariance matrix is, given a quadratic loss function, a measure of the expected loss or risk in estimating β by $\hat{\beta}$. Consequently, the information matrix M and the covariance matrix D are important devices for the choice of an experimental design. An optimal design is a design ξ^* that minimizes a suitable measure of imprecision or risk: The most popular criterion is the D–optimality in which $\Psi[M(\xi)] = \log |M(\xi)|^{-1} = -\log |M(\xi)|$ is minimized. This implies the minimization of the determinant $|D_{\hat{\beta}}(\xi)|$ of the expected loss.

In the case of potentially failing trials, i.e., when the probability is less than one that a planned observations is actually observed, the covariance matrix or the expected loss has to take this fact into account: Let $1 - p_i$ be the probability that the trial at site x_i fails ($i = 1, \dots, k$); and let us assume that the events that trials are failing are independent. As a consequence, the number ρ of observations that are available after the execution of an r-trial experiment for estimating β fulfills the inequality $\rho \leq r$. As the LS-estimator $\hat{\beta}$ requires that ρ is at least m, the event that $\hat{\beta}$ cannot be calculated on the basis of the realized sample has a non-zero probability.

For an r-trial design ξ_u for an experiment with potentially failing trials, the expected loss in estimating the m-vector β of coefficients is the weighted sum of covariance matrices that correspond to the various sets of possibly available observations:

$$D_{\hat{\beta}}(\xi_u) = \sum_{\rho=1}^{r} \frac{1}{\rho} \sum_{j \in J_\rho} p_{\rho,j}^{(r)} \left[M(\xi_{\rho,j}^{(r)}) \right]^{-1} ; \tag{3}$$

here, for each ρ the set J_ρ corresponds to the set of all possible subsets of size ρ from the sites $\mathcal{X}(\xi_u)$ of ξ_u, i.e., j counts the ρ-trial designs $\xi_{\rho,j}^{(r)}$; the weight for the design $\xi_{\rho,j}^{(r)}$ is the probability

$$p_{\rho,j}^{(r)} = \prod_{i \in I_{\rho,j}^{(a)}(\xi_u)} p_i \prod_{i \in I_{\rho,j}^{(m)}(\xi_u)} (1 - p_i) \tag{4}$$

where the index sets $I_{\rho,j}^{(a)}(\xi_u)$ and $I_{\rho,j}^{(m)}(\xi_u)$ correspond to the subsets of sites where the responses are available and missing, respectively; the information matrices per observation are

$$M(\xi_{\rho,j}^{(r)}) = \frac{1}{\rho} \sum_{i \in I_{\rho,j}^{(a)}(\xi_u)} f_i f_i^T \tag{5}$$

The determinant $|D_{\hat{\beta}}(\xi)|$ of the expected loss is unbounded if the information matrix $M(\xi)$ for the design ξ is singular; of course, such a design is no candidate for being chosen as an optimal design. Given a realistic structure of the p_i, $i = 1, \ldots, k$, we have to expect that singular information matrices are found among those in (3) for any design ξ_u. This means that $D_{\hat{\beta}}(\xi_u)$ is with high probability unbounded. We have therefore to look for a appropriate policy with respect to designs with singular information matrix. Our suggestion is to add a suitably chosen penalty term which is a multiple of the probability that a design results in a singular information matrix:

$$|D_{\hat{\beta}}(\xi_u)| = \left| \sum_{\rho=1}^{r} \frac{1}{\rho} \sum_{j \in J_\rho^r} p_{\rho,j}^{(r)} \left[M(\xi_{\rho,j}^{(r)}) \right]^{-1} + \lambda P_{r,\rho}^s I_m \right| ; \tag{6}$$

here, J_ρ^r and J_ρ^s correspond to the set of designs with regular and singular information matrix, respectively; λ is the penalty and $P_{r,\rho}^s$ is the probability that a design results in a singular information matrix. The effect of the choice of λ is analyzed in the following section.

An optimal design ξ for estimating β minimizes the determinant $|D_{\hat{\beta}}(\xi)|$ of the covariance matrix of the estimate $\hat{\beta}$. In the case of certain response the basic idea of algorithms for obtaining D-optimal designs is the relation

$$|M_{r+1}| = |M_r| \left[\frac{r}{r+1} + \frac{d(x, \xi_r)}{r} \right]$$

where

$$d(x, \xi_r) = \frac{r \operatorname{Var}\{\hat{y}_x\}}{\sigma^2} = f^T(x) M_r^{-1} f(x)$$

is the standardized variance of the forecast $\hat{y}_x$ for the response at x, based on the observation from design ξ_r. This relation implies that the addition of a trial at the site where $d(x, \xi_r)$ is maximum results in the largest possible increase of the information matrix. This relation can be used to approach the optimal design, e.g., by iteratively adding and deleting sites. In the potentially failing trial case, however, the effect of adding a trial is a weighted mean of the effects of adding the trial to all the subdesigns that can occur due to failing trials; the weights are functions of the probabilities $p_{\rho,j}^{(r)}$ and the information matrices $M(\xi_{\rho,j}^{(r)})$. As a consequence, the numerical effort in assessing possible candidates to be added or deleted is considerable.

3 Examples

On the basis of three examples, we investigate the effect of potentially failing trials on optimal design. For the probabilities of failing trials we choose three structures:

(a) constant probabilities

(b) constant probabilities and independence of failing trials

(c) increasing probabilities

We assume that the response variable Y is generated according to the quadratic polynomial

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + u = f^T \beta + u\,; \tag{7}$$

x is a non-random regressor variable with design space

$$\mathcal{X}_k = \left\{0, \frac{1}{k-1}, \ldots, 1\right\};$$

the error term u has expectation zero and variance σ^2 for all x in $\mathcal{X}_k$ and the error terms for any two observations are independent.

For the case where potentially failing trials are ignored, the D-optimal continuous design with $\mathcal{X} = [0,1]$ is $\xi_c^* = \{0, 0.5, 1; \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$. Exact designs for the discrete design space $\mathcal{X}_k$ are similar in the sense that the observations are concentrated in sites that are close to the sites of ξ_c^*. *Table 1* gives corresponding designs for some values of k and sample sizes r together with the value of the optimality criterion in column "$p_0 = 1$"; the designs are indicated by "1" at sites where trials are planned and "." otherwise. The designs were found by comparing the optimality criteria of suitable candidates.

3.1 A simple probability structure

Let us assume that at all sites at most one observation is taken; thus, the design of our experiment has the form $\xi_u = \{x_1, \ldots, x_r; 1, \ldots, 1\}$, where the $x_i \in \mathcal{X}_k$ are the selected sites. In modelling the probability structure we choose probabilities p_i, $i = 1, \ldots, k$, for the events that a response is observed at all sites except at x_i: $p_1 = \ldots = p_k = p$. Then, $p_0 = 1 - \sum_{i=1}^{k} p_i$ is the probability for the event that the responses of all sites are available or the responses of at least two sites are missing. Given these assumptions, we find for the expected loss or the covariance matrix of the LS-estimates $\hat{\beta}$

$$\begin{aligned} D_p(\hat{\beta}) &= \sum_{j=i}^{r} p_j D(\hat{\beta}_{-j}) + \left(1 - \sum_{j=1}^{r} p_j\right) D(\hat{\beta}) \\ &= \sigma^2 \left\{ \frac{1}{r-1} \sum_{j=1}^{r} p_j [M(\xi_{-j})]^{-1} + \frac{1}{r} \left(1 - \sum_{j=1}^{r} p_j\right) [M(\xi)]^{-1} \right\}; \end{aligned} \quad (8)$$

here, ξ_{-j} and $\hat{\beta}_{-j}$ are the design and the LS-estimate, respectively, where the observation at site x_j is missing.

Table 1 shows the D-optimal designs for $p_0 = 1$, 0.8, and 0.2 for three values of k and some sample sizes r, together with the value of the optimality criterion. As it can be expected, for a given design the value of the optimality criterion increases with decreasing p_0; this corresponds to the fact that with decreasing p_0, the weight of non-optimal designs increases which become effective due to failing trials. Even more interesting is the fact that the optimal design changes for small values of r with decreasing p_0. E.g., for $k = 8$ and $r = 4$, the optimal designs are (A) '1 . . 1 1 . . 1' and (B) '1 . 1 . . 1 . 1' for the case of certain response and potentially failing trials, respectively. The reason is that the latter design is more robust against failing trials: A missing observation in $x = 0$ (or $x = 1$) increases the optimality criterion by a factor 100 (from 8.17 to 817.0) in case of design A but only by 7.2 (from 18.2 to 130.72) in case B. The effect of omitting one of the inner observations is about the same: increase from 8.17 to 16.67 in case A, from 18.2 to 24.01 in case B.

3.2 Independently failing trials

We assume again that at all sites at most one observation is taken. For all sites the probability is p that a trial is failing, and the events that trials are failing are assumed to be independent. Thus, the probabilities $p_{\rho,j}^{(r)}$ in (6) are

$$p_{\rho,j}^{(r)} = \binom{r}{\rho} p^{\rho}(1-p)^{r-\rho} .$$

For the numerical illustration, two values for p were chosen out of the interval (B) [.86, .92] and (C) [.57, .73] so that the probability $P\{S < k-1\}$ is about 0.8 and 0.2, respectively, that the number S of non-failing trials is less that $k-1$. The corresponding probabilities are shown in the following table.

Table 2 shows the D-optimal designs for the cases (A) $p = 1$ and (B) and (C) as stated above, again for the three values of k and the sample sizes r, that were used in Section 3.1. In all reported cases, the penalty was set to zero ($\lambda = 0$). However, a nonzero penalty increases the value of the optimality criterion but only for very large values of λ the ranking of the designs. Contrary to the cases

of Section 3.1, the value of the optimality criterion can be increased or decreased in the cases (B) and (C) as compared to the case $p = 1$. Particularly for small values of r, the optimality criterion of the optimal designs in cases (B) and (C) are considerably larger than that of the optimal design for $p = 1$. In agreement with the results in Section 3.1, the optimal designs tend to give weight to points that are away from the center with increasing p.

	k	p	$P\{S < k-1\}$	$P\{S = k-1\}$	$P\{S = k\}$	$P\{S \leq 2\}$
(B)	6	0.860	0.200	0.395	0.405	0.005
	8	0.895	0.202	0.386	0.412	0.000
	10	0.917	0.199	0.381	0.420	0.000
(C)	6	0.575	0.804	0.160	0.036	0.216
	8	0.670	0.799	0.160	0.041	0.019
	10	0.729	0.800	0.158	0.042	0.001

3.3 Increasing probability for failing trials

Here, the probability structure is the same as that in Section 3.1 except that we let the probabilities grow: The probability, that a response is observed at all sites except at x_i, is now $p_i = p(i-1)$, $i = 1, \ldots, k$, where p is chosen so that $p_0 = 1 - \sum_{i=1}^{k} p_i$ gets a prespecified value.

Table 3 shows the D-optimal designs for $p_0 = 1$, 0.8, and 0.2 for three values of k and some sample sizes r, together with the value of the optimality criterion. Here, in cases of unsymmetrical designs the optimality criterion of the mirrored designs is not the same. E.g., for $k = 8$, $r = 4$ and $p_0 = 0.2$, the optimality criteria for the optimal design '1 . 1 . . . 1 1' and the design '1 1 . . . 1 . 1' are 41.21 and 108.26, respectively. This is due to the fact that the loss of an observation is much more serious in the second case: A failing trial both at $x = 0$ in the former and at $x = 1$ in the latter case results in a design criterion of 294.12; the probabilities that this happens is 0 and 0.167 in the two cases; similarly, the design criterion is 66.69 for a failing trial in $x = 2/7$ and

$x = 5/7$, and the corresponding probabilities are 0.038 and 0.111, respectively. As a consequence, the effect of potentially failing trials is reduced as compared to the case of constant probability of failing trials as discussed in Section 3.1. So, the ratio of the optimality criterion of the optimal designs for potentially failing trials ($p_0 = 0.2$) and certain response is decreased from 5.0 to 3.9 when $k = 10$ and $r = 4$.

4 Concluding Remarks

When real-life experiments are performed it cannot be guaranteed that each trial will result in an observation. In the case where the data are to be used for estimating the model parameters and the experimental design is optimized, the loss of this uncertainty is twofold: A reduced number of observations (1) increases the standard error of the estimates, and (2) potentially decreases the efficiency of the design. The size of the effects depends on the probabilities that trials fail to give response at the sites of the design space. These effects must be expected to be serious for small and moderate sample sizes. This is demonstrated for simple cases in the examples of Section 3.

Acknowledgement

The author is indebted to Michael Maderbacher for his assistance in the computations.

References

Hedayat, A.S. and Majumdar, D. (1984). Redesigning Experiments, pp. 113-140 in A.M. Abouammoh, E. El-Neweihi, E. Aly, M.A. Alosh, (Eds.), *Developments in Statistics and Its Applications. Proceedings of the First Saudi Symposium on Statistics and Its Applications.* Riyadh: King Saud University Library.

Das, A. and Sinha, B.K. (1994). Optimal Allocation of Observations Under Experimental Constraints: Resources Scarcity and External Force's Intervention. *Research Report*, Indian Statistical Institute, Calcutta.

Table 1: Exact D-optimal designs for experiments with constant probability for the trials to fail and $p_0 = 1$, 0.8, and 0.2; three sizes k of the design space $\mathcal{X}_k$ and various sample sizes r. In cases of unsymmetrical designs, the corresponding mirrored design has the same value of the optimality criterion. The respective values of the optimality criterion is given for three values of p_0; the minimal values of the optimality criterion are indicated by '*'.

		$\lvert D(\xi)\rvert$			
k	r	$p_0 = 1$	$p_0 = 0.8$	$p_0 = 0.2$	ξ
6	4	8.35*	65.12	649.36	1 . 1 1 . 1
		14.36	23.71*	72.15*	1 1 . . 1 1
	5	5.75*	15.09	65.82	1 1 1 1 . 1
		6.34	9.46*	23.95*	1 1 1 . 1 1
	6	3.99	5.66	13.04	1 1 1 1 1 1
8	4	8.17*	103.72	1169.13	1 . . 1 1 . . 1
		10.14	18.20*	61.03*	1 . 1 . . 1 . 1
	5	5.45*	15.30	59.56	1 . . 1 1 . 1 1
		6.30	8.02*	14.99*	1 1 . 1 . . 1 1
	6	3.63	4.43	7.49	1 1 . 1 1 . 1 1
	7	2.71	3.26	5.32	1 1 . 1 1 1 1 1
	8	2.08	2.47	3.89	1 1 1 1 1 1 1 1
10	4	8.10*	152.99	1855.79	1 . . . 1 1 . . . 1
		10.32	17.63*	40.70*	1 . 1 . . . 1 . . 1
	5	5.20*	16.28	61.62	1 . . . 1 1 . . 1 1
		6.08	7.42*	12.54*	1 1 . . 1 . . . 1 1
	6	3.34	3.84	5.65	1 1 . . 1 1 . . 1 1
	7	2.41	2.75	3.94	1 1 . 1 1 1 . . 1 1
	8	1.83	2.08	2.93	1 1 . 1 1 1 1 . 1 1
	9	1.49	1.68	2.32	1 1 1 1 1 1 1 . 1 1
	10	1.22	1.36	1.84	1 1 1 1 1 1 1 1 1 1

Table 2: Exact D-optimal designs for experiments with uncertain response and probability for non-failing trial (A) $p = 1$, (B) $p \in [.86, .92]$, and (C) $p \in [.57, .73]$ and independent failing trials; three sizes k of the design space $\mathcal{X}_k$ and various sample sizes r. In cases of unsymmetrical designs, the corresponding mirrored design has the same value of the optimality criterion. The minimal values of the optimality criterion are indicated by '*'.

		$\lvert D(\xi)\rvert$			
k	r	(A)	(B)	(C)	ξ
6	4	8.35*	8.30	3.63	1 . 1 1 . 1
		14.36	1.02*	0.31*	1 1 . . 1 1
	5	5.75*	4.66	22.54	1 1 1 1 . 1
		6.34	3.47*	17.42*	1 1 1 . 1 1
	6	3.99	0.92	44.49	1 1 1 1 1 1
8	4	8.17*	19.21	24.45	1 . . 1 1 . . 1
		10.14	1.03*	0.90	1 . 1 . . 1 . 1
		22.08	1.17	0.82*	1 1 . . . . 1 1
	5	5.45*	3.39	41.79	1 1 . 1 1 . . 1
		6.30	1.38*	16.37*	1 1 . 1 . 1 . 1
	6	3.63	0.39	33.91	1 1 . 1 1 . 1 1
	7	2.71	0.10	31.56	1 1 . 1 1 1 1 1
	8	2.08	0.02	17.44	1 1 1 1 1 1 1 1
10	4	8.10*	34.59	79.79	1 . . . 1 1 . . . 1
		10.32	0.74*	0.90*	1 . 1 . . . . 1 . 1
	5	5.20*	3.45	76.81	1 . . . 1 1 . . 1 1
		6.08	0.69*	10.50*	1 . 1 . 1 . . 1 . 1
	6	3.34*	0.28	36.91	1 1 . . 1 1 . . 1 1
		7.17	0.19*	19.58	1 . 1 . 1 . 1 . 1 1
		7.39	0.23	19.05*	1 . 1 . 1 1 . 1 . 1
	7	2.41*	0.06	15.61	1 1 . 1 1 1 . . 1 1
		4.71	0.04*	14.37	1 1 . 1 1 . 1 . 1 1
		5.01	0.06	12.60*	1 1 . 1 1 1 . 1 . 1
	8	1.83*	0.01	4.22*	1 1 . 1 1 1 1 . 1 1
		3.66	0.01*	12.49	1 1 1 . 1 1 . 1 1 1
	9	1.49*	0.01	1.46*	1 1 1 1 1 1 1 . 1 1
		2.85	0.01*	2.21	1 1 1 1 1 1 . 1 1 1
	10	1.22	0.01	0.42	1 1 1 1 1 1 1 1 1 1

Table 3: Exact D-optimal designs for experiments with increasing probability for trials to fail and $p_0 = 1$, 0.8, and 0.2; three sizes k of the design space $\mathcal{X}_k$ and various sample sizes r. The respective values of the optimality criterion is given for three values of p_0; the minimal values of the optimality criterion are indicated by '*'.

		$\lvert D(\xi)\rvert$			
k	r	$p_0 = 1$	$p_0 = 0.8$	$p_0 = 0.2$	ξ
6	4	8.35*	51.97	397.93	1 . 1 1 . 1
		9.81	20.12*	83.19	1 . 1 . 1 1
		14.36	23.33	64.52*	1 1 . . 1 1
	5	5.75*	18.46	73.01	1 1 1 1 . 1
		6.34	9.11*	20.67*	1 1 1 . 1 1
	6	3.99	5.61	12.09	1 1 1 1 1 1
8	4	8.17*	71.90	585.93	1 . . 1 1 . . 1
		10.14	17.63*	50.35	1 . 1 . . 1 . 1
		12.56	17.75	41.21*	1 . 1 . . . 1 1
	5	5.45*	21.05	80.00	1 1 . 1 1 . . 1
		5.83	7.48*	14.01	1 . 1 . 1 . 1 1
		6.30	7.89	13.87*	1 1 . 1 . . 1 1
	6	3.63	4.42	7.24	1 1 . 1 1 . 1 1
	7	2.71	3.25	5.14	1 1 1 1 1 . 1 1
	8	2.08	2.46	3.80	1 1 1 1 1 1 1 1
10	4	8.10*	93.02	784.31	1 . . . 1 1 . . . 1
		9.11	14.48*	31.89*	1 . . 1 . . 1 . . 1
	5	5.20*	7.94	19.68	1 . . . 1 1 . . 1 1
		5.44	6.69*	11.49	1 . . 1 . 1 . . 1 1
		5.91	6.99	10.89*	1 . 1 . . 1 . . 1 1
	6	3.34	3.84	5.54	1 1 . . 1 1 . . 1 1
	7	2.41	2.73	3.84	1 1 . 1 1 1 . . 1 1
	8	1.83	2.07	2.85	1 1 . 1 1 1 1 . 1 1
	9	1.49	1.67	2.26	1 1 . 1 1 1 1 1 1 1
	10	1.22	1.36	1.82	1 1 1 1 1 1 1 1 1 1