The Korean Communications
in Statistics Vol. 4, No. 1, 1997
pp. 129-138

Choosing Optimal Design Points in Two Dimensional Space using Voronoi Tessellation

Dongryeon Park[1)]

Abstract

In this paper, the problem for choosing design points in the two dimensional case is considered. In the one dimensional case, given the design density function, we can choose design points using the quantile function. However, in the two dimensional case, there is no clear definition of the percentile. Therefore, the idea of choosing design points in the univariate case can not be applied directly to the two dimensional case. We convert this problem into an optimization problem using the Voronoi diagram.

1. Introduction

A sensitivity experiment is characterized by a response surface that relates the stimulus level applied to an experimental subject to the probability of response. The outcome of experiment is assumed dichotomous, response or nonresponse. The observed reaction Y_i of the ith subject at stimulus level $\boldsymbol{x}_i$, $i=1,\dots,n$ is encoded by $Y_i=0$ (if nonresponse) or $Y_i=1$ (if response) where $\boldsymbol{x}_i$ is a $k\times 1$ vector of independent variables. The probability of response is related to stimulus level $\boldsymbol{x}$ by a quantal response surface $p(\boldsymbol{x})$, i.e.,

$$p(\boldsymbol{x}_i) = \Pr(Y_i=1) \quad , i=1,\dots,n \tag{1}$$

The specification of the stimulus level $\boldsymbol{x}_i$ forms the design of the experiment. In this paper, we consider only the two dimensional case in which the design point $\boldsymbol{x}_i$ lie in $\boldsymbol{R}^2$.

As an estimator for $p(\boldsymbol{x})$, we define the kernel estimator

$$\hat{p}(\boldsymbol{x})=\frac{1}{b^2}\sum_{i=1}^{n}\int_{A_i}K\left(\frac{\boldsymbol{x}-s}{b}\right)ds\ Y_i \tag{2}$$

where b is a sequence of a positive bandwidths depending on n such that

1) Assistant Professor, Department of Statistics, Hanshin University, Osan, Kyunggi-do, 447-791, Korea

$$b \to 0, \quad nb^2 \to \infty \quad \text{as} \quad n \to \infty$$

and where K is a kernel function, and where A_i is a partition of Ω such that $\boldsymbol{x}_i \in A_i$ and $\bigcup_i A_i = \Omega$ and $A_i \cap A_j = \emptyset$, for *all* $i \neq j$ where $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ are the design points. Assume $\Omega = [0,1]^2$. In addition we assume $K(\boldsymbol{u})$ is continuous and has a compact support.

In the one dimensional case, Muller and Schmitt (1988) assumed that there is a strictly positive design density $f(x)$ which determines the design points uniquely, and they derived the optimal design density $f^*(x)$ which minimizes the asymptotic IMSE (Integrated MSE). Then the optimal design points $x_1^*, \cdots, x_n^*$ are chosen according to

$$\int_0^{x_i^*} f^*(t)dt = \frac{i-1}{n-1}. \tag{3}$$

The quantile function uniquely determines the design points in one dimension.

In the two dimensional case, however, there is no clear definition of the percentile, so the idea of choosing design points in univariate case can not be applied directly to the two dimensional case. Therefore, even though Park (1995) derived the optimal design density $f^*(\boldsymbol{x})$ which minimizes the asymptotic IMSE of our estimator $\hat{p}(\boldsymbol{x})$ of (2), there is no unique way we can choose the optimal design points $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*$ using $f^*(\boldsymbol{x})$.

In this paper, we will consider the problem for choosing design points in the two dimensional case using the design density.

2. Choosing Optimal Design Points

Suppose the optimal design density $f^*(\boldsymbol{x})$ is given and we want to choose optimal design points $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*$ from $f^*(\boldsymbol{x})$. Then we wish to choose optimal design points $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*$ such that

$$\int_{A_i} f^*(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{n}, \qquad \boldsymbol{x}_i^* \in A_i. \tag{4}$$

Therefore,

$$\Delta A_i = \frac{1}{nf(\boldsymbol{a}_i)}, \quad \text{for some} \quad \boldsymbol{a}_i \in A_i \tag{5}$$

where ΔA_i is the area of A_i, so $\Delta A_i = O(n^{-1})$.

However there is no unique way to select optimal design points such that (4) is satisfied, so we need to restrict A_i in a reasonable way.

Suppose we set

$$A_i = \{ \boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_i^*\| \leq \|\boldsymbol{x} - \boldsymbol{x}_j^*\| \quad \text{for} \quad j \in I_n \quad \text{but} \quad j \neq i \} \tag{6}$$

where $I_n = \{1, \cdots, n\}$ and $\|\cdot\|$ is the Euclidean distance. Then A_i is the Voronoi polygon associated with $\boldsymbol{x}_i^*$ and the set given by $A = \{A_1, \cdots, A_n\}$ is the Voronoi diagram generated by the optimal design points.

So we wish to choose the optimal design points by constructing the Voronoi diagram such that (4) is satisfied for all i. However, there is no obvious way to do this. Therefore we have to resort to an approximate approach. We convert our problem into an optimization problem in the following way :

1. Define the objective function F

$$F(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \int_{A_i} \|\boldsymbol{x} - \boldsymbol{x}_i\|^2 f^*(\boldsymbol{x})\, d\boldsymbol{x} \tag{7}$$

   where A_i is the Voronoi polygon associated with $\boldsymbol{x}_i$ and $\|\cdot\|$ is the Euclidean distance.

2. Choose $\{\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*\}$ such that $\{\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*\}$ minimizes F.

The above optimization problem is a geographical optimization problem and it has been considered by Iri *et al.* (1984). To be specific, let us consider n "facilities" placed, respectively, at points $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ of the two dimensional Euclidean space $\boldsymbol{R}^2$. The territories of these facilities are the Voronoi regions A_i defined by (6). We consider furthermore a distribution of "inhabitants" represented by $f^*(\boldsymbol{x})$, to be called the population density function. We assume that an individual inhabitant living in region A_i will use the ith facility and the cost for user to gain access to facility is a function of the Euclidean distance of $\boldsymbol{x}$ and the location $\boldsymbol{x}_i$ of the nearest facility.

In this optimization problem, we want to determine the optimum locations of a given number of facilities for a given population distribution, and this is the problem of minimizing the objective function F in (7) which is the total cost connected with the serviceability of the facilities. Therefore to minimize the total cost, we will put more facilities where the population is dense.

Thus, if we choose $\{\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*\}$ such that it minimizes the objective function F of (7), then we can expect the area of A_i which is the Voronoi polygon associated with $\boldsymbol{x}_i^*$ is inversely proportional to $f^*(\boldsymbol{x})$.

Note that from (4) and (5), ΔA_i is also inversely proportional to $f^*(\boldsymbol{x})$. Therefore we claim that the optimization problem is approximately equivalent to our original problem.

3. Optimal Design Algorithm

In this section, we consider how to implement the optimal design idea of the previous section. First of all, in terms of the objective function F, we can formally state the optimal design algorithm.

Optimal Design Algorithm

Given an optimal design density function $f^*(\boldsymbol{x})$, choose the optimal design points $\boldsymbol{x}_1^*, \cdots, \boldsymbol{x}_n^*$ such that it minimizes

$$\sum_{i=1}^{n} \int_{A_i} \| \boldsymbol{x} - \boldsymbol{x}_i^* \|^2 f^*(\boldsymbol{x})\, d\boldsymbol{x} \tag{8}$$

Since each $\boldsymbol{x}_i$ has 2 coordinates (x_{i1}, x_{i2}), the problem is essentially minimization with respect to $2n$ variables. Let $\boldsymbol{X}$ be the $2n$-dimensional unknown vector whose components are the components of n vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$:

$$\boldsymbol{X} = [\ \boldsymbol{x}_1^T, \cdots, \ \boldsymbol{x}_n^T\ ].$$

Our computational procedure for above problem is written algorithmically as follows.

Step 1. Choose an arbitrary $\boldsymbol{X}^{(0)} \in \boldsymbol{R}^{2n}$, and set $k = 0$ (denoted by $k \leftarrow 0$).

Step 2. Compute the gradient, $\nabla F(\boldsymbol{X}^{(k)})$, then determine the direction vector $\boldsymbol{d}^{(k)}$ at the point $\boldsymbol{X}^{(k)}$

Step 3. Determine the step size $\alpha^{(k)}$ at the point $\boldsymbol{X}^{(k)}$

Step 4. Get the new approximation by

$$\boldsymbol{X}^{(k+1)} \leftarrow \boldsymbol{X}^{(k)} + \alpha^{(k)} \boldsymbol{d}^{(k)}$$

Step 5. Repeat Step 2 - 4 until some stopping criterion is satisfied.

To compute the gradient $\nabla F(\boldsymbol{X}^{(k)})$, we need the partial derivatives of F. The partial derivatives of F in (7) with respect to x_{ij} consists of two components, one comming from the derivatives of the integrand function and the other from the variation of the region A_i itself as well as of the region A_j's adjacent to A_i due to the variation of $\boldsymbol{x}_i$. The first component takes the form:

$$\int_{A_i} 2(x_{ij}-x_j)f^*(\boldsymbol{x})d\boldsymbol{x}$$

and the second component is the sum of the terms due to the variation of A_i and the variation of adjacent region A_j . However, the terms constituting the second component cancel themselves, so that only the first component remain. Therefore, the first derivative is given by

$$\frac{\partial F}{\partial x_{ij}} = \int_{A_i} 2(x_{ij}-x_j)f^*(\boldsymbol{x})d\boldsymbol{x}, \quad j=1,2, \ i=1,\dots,n$$

where $\boldsymbol{x} = (x_1, x_2)$ and x_{ij}, $j=1,2$, is the element of $\boldsymbol{x}_i$, so $\boldsymbol{x}_i^T = (x_{i1}, x_{i2})$. The detailed derivation is given by Iri *et al.* (1984) and Okabe (1992).

Using the steepest descent method with this derivative, we can obtain a local optimal solution. Alternatively, we may use the Newton method to gain faster convergence. For this use, we need the second partial derivatives of F. The detailed derivation of the second partial derivatives of F is given by Iri *et al.* (1984). Then the direction vector is given by

$$\boldsymbol{d}^{(k)T} = -[\nabla^2 F(X^{(k)})]^{-1} \nabla F(X^{(k)}),$$

where $\nabla^2 F(X^{(k)})$ is the Hessian matrix of $F(X^{(k)})$. Although we may use the Newton method, the calculation of the Hessian matrix becomes time consuming for a large number of design points when the design density is not uniform. To shorten computing times, Iri *et al.* (1984) recommend, on the basis of their numerical experiment, to use the quasi-Newton method with the diagonal matrix H, whose diagonal elements are given by

$$\frac{\partial^2 F}{\partial x_{ij}^2} = \int_{A_i} 2f^*(\boldsymbol{x})d\boldsymbol{x}, \quad j=1,2$$

where $\boldsymbol{x}_i^T = (x_{i1}, x_{i2})$ and $\boldsymbol{x} = (x_1, x_2)$.

The line search algorithm is based on the Goldstein's rule to use $\widehat{\alpha}$ satisfying

$$\mu_2 \widehat{\alpha}\, \boldsymbol{d}^{(k)} \cdot \nabla F(X^{(k)}) \le F(X^{(k)} + \widehat{\alpha}\, \boldsymbol{d}^{(k)}) - F(X^{(k)}) \le \mu_1 \widehat{\alpha}\, \boldsymbol{d}^{(k)} \cdot \nabla F(X^{(k)})$$

with predetermined constants μ_1 and μ_2 $(0 < \mu_1 < \mu_2 < 1)$. We used $\mu_1 = 10^{-3}$ and $\mu_2 = 0.5$, which are recommended values in Suzuki *et al.* (1991). The stopping rule of the repetition is given by

$$\max_{i,j} |x_{ij}^{(k+1)} - x_{ij}^{(k)}| \le 10^{-5}.$$

The algorithm for Voronoi construction which we used is based on the grid method. We used 20×20 grid points throughout our numerical exercises, and each grid point is the center of pixel sized $\frac{1}{20}\times\frac{1}{20}$. All we have to do to construct the Voronoi diagram is for each generator which is design point, to identify the pixels which belong to the generator.

4. Examples

In this section, we choose design points using the optimal design algorithm of the previous section for several design densities. As a design density, we used the bivariate normal density with different parameters :

$$\text{Model 1 } (f_1(\boldsymbol{x})) : \ \mu_1=\mu_2=.5, \ \sigma_1=\sigma_2=.05, \ \rho=.5$$

$$\text{Model 2 } (f_2(\boldsymbol{x})) : \ \mu_1=\mu_2=.5, \ \sigma_1=\sigma_2=.1, \ \rho=.5$$

$$\text{Model 3 } (f_3(\boldsymbol{x})) : \ \mu_1=\mu_2=.5, \ \sigma_1=\sigma_2=.15, \ \rho=.5$$

In Figure 1-6, we have plotted the design density function and the optimal configuration of $n=25$ for each model in the unit square. We can see that the selected design points for each model represent the design density very well. Therefore, by the optimal design algorithm of (8), we can solve the problem for choosing the design points in the two dimensional case.

5. Conclusion

In one dimensional case, if we want to determine the design point from the design density function, we can use the quantile function which uniquely determines the design points. However, in two dimensional case, since there is no clear definition of the percentile, the idea of choosing design points in univariate case can not be applied directly. In this paper, we provide one possible solution for this problem. Even though our method is not exact approach, we can verify our algorithm works very well through our numerical exercises.

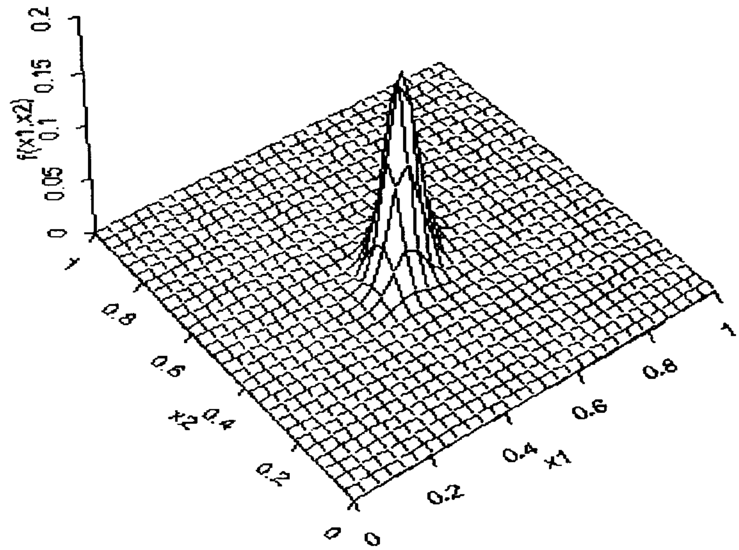

Figure 1: Design Density of $f_1(\boldsymbol{x})$

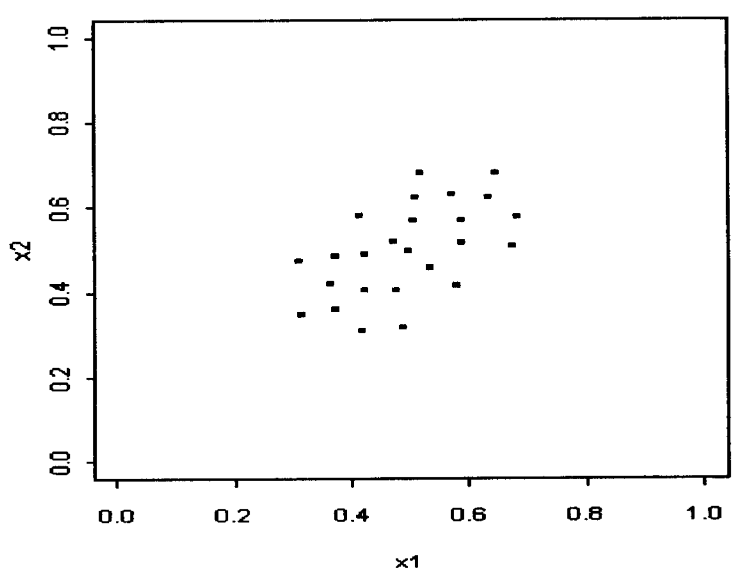

Figure 2: Location of Design Points for $f_1(\boldsymbol{x})$

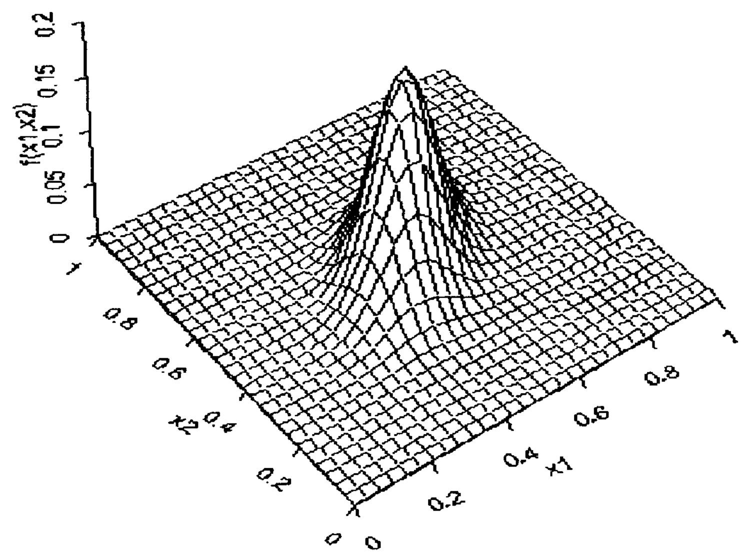

Figure 3: Design Density of $f_2(\boldsymbol{x})$

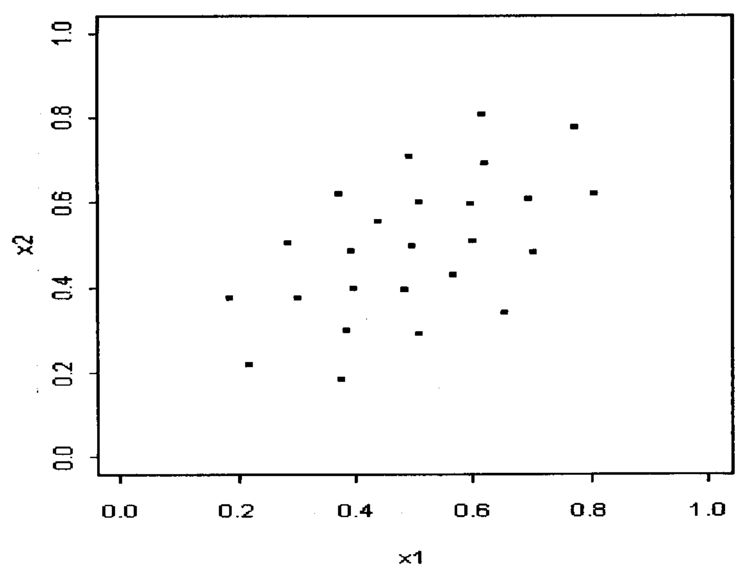

Figure 4: Location of Design Points for $f_2(\boldsymbol{x})$

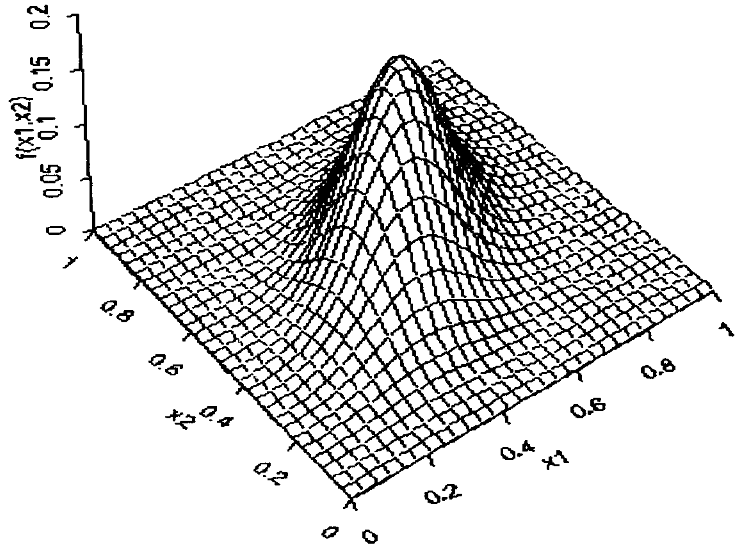

Figure 5: Design Density of $f_3(\boldsymbol{x})$

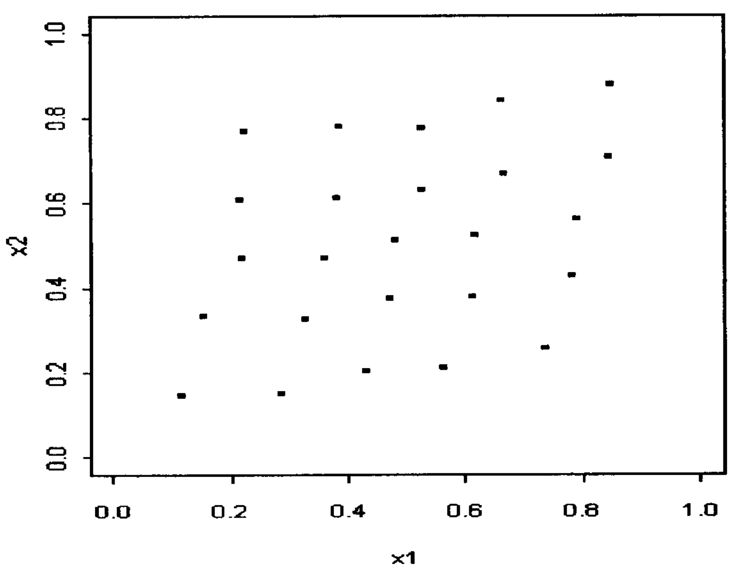

Figure 6: Location of Design Points for $f_3(\boldsymbol{x})$

References

[1] Iri, M., Murota, K. and Ohya, T. (1984), "A Fast Voronoi-diagram Algorithm with Applications to Geographical Optimization Problem", in *System Modelling and Optimization*, ed. P. Thoft-Christensen, Springer, Berlin, pp 273-288

[2] Muller, H. and Schmitt, T. (1988), "Kernel and Probit Estimates in Quantal Bioassay", *Journal of the American Statistical Association*, 83, 750-759

[3] Okabe, A. (1992), *Spatial Tessellations : Concepts and Applications of Voronoi Diagrams*, New York, John Wiley

[4] Park, D. R. (1995), "Optimal Designs for Multivariate Nonparametric Kernel Regression with Binary Data", *The Korean Communications in Statistics*, 2, 243-248

[5] Suzuki, T., Asami, Y. and Okabe, A. (1991), "Sequential Location - Allocation of Public Facilities in One - and Two - dimensional Space : Comparison of Several Policies", *Mathematical Programming*, 52, 125-146