SYNTHESIS OF OPTIMUM EXPERIMENTAL PLANS ENSURING COMPUTATION OF INTEGRAL PARAMETERS USING A REGRESSION EQUATION.

(SOME EXAMPLES)

Dr. Manuel Carames
Assistant Professors
Mathematics Department
Miami-Dade College, North Campus
11380 NW 27 Avenue, Miami, Florida 33167, USA
Emails: mcarames@mdc.edu

ABSTRACT

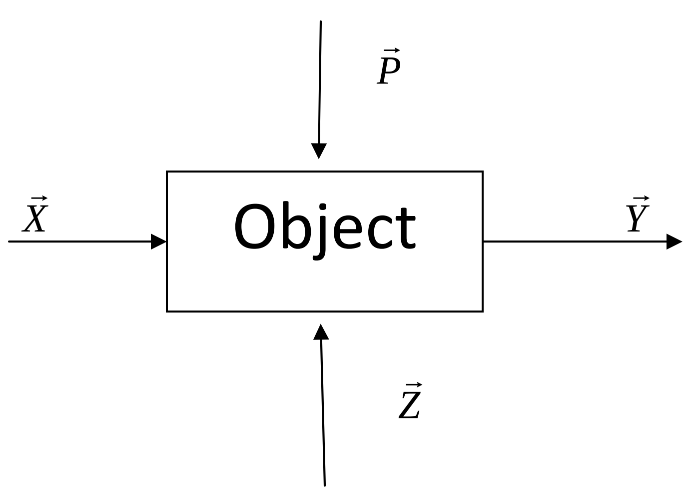
This is an attempt to share with my colleagues some experiences related to Optimal Experimental Plans and particularly to ones that were introduced by me and have the general name of I plans (I because they are plans to estimate integral indexes of the response function).

Theme: Optimal Experimental Plans
Key words: Experiment, estimation, statistic, system, model

Introduction: In different scenarios, engineers, researchers, etc need to estimate not the values of the response functions in certain points, they need to estimate some integral indexes of this function. To this idea is dedicated this paper

Body:

What is the researched object?



The set of parameters which define the state of the object is divided in the following groups:

- Input variables (vector of input variables) $\vec{x} = \left(x_1, x_2 \cdots x_m\right)$. In this group we have the controllable parameters of the object. The values of these parameters are inside given intervals and they are given by the schedule of the technological process or technical constrains and are of the form $x_{i\,\min} \le x_i \le x_{i\,\max}$; $i = 1,2,\cdots,m$;
- Output variables (vector of output variables) $\vec{y} = \left(y_1, y_2, \cdots y_r\right)$. In this group we have the variables that contain information about quantity and quality characteristics of the output product.
- In group $\vec{z}$ we include the variables that we can control but we cannot manipulate. $\vec{z} = \left(z_1, z_2, \cdots z_s\right)$.
- In group $\vec{p}$ we include variables that we cannot control neither manipulate. In this group we have noises and we do not know the points where they are applied, we neither know the time characteristics nor the power of then. $\vec{p} = \left(p_1 p_2 \cdots p_l\right)$.

It is clear from the way that we defined the universe of signals that only the once that belong to $\vec{x}$ could be manipulated.

In case of active experiment all the variables that belong to $\vec{p}$ will be represented by equivalent addition noise e applied to the output.

Variables from $\vec{z}$, whose characteristics during the experiment are known will be considered as variables that belong to group $\vec{x}$.

Let us assume that we have one output signal, then the response function can be represented as

$y = \eta\left(\vec{x}, \vec{\theta}\right) + e$ where

e is the noise with the following properties:

$E(e) = 0$ the mean of the noise is zero. E is the mean operator ;

$E\left(e^i . e^j\right) = 0 \forall i \ne j$ this means in different point of the factor space x_i, x_j where the output is measured e^i and e^j are not correlated ;

$D\left(e^i\right) = \sigma^2 \forall i$ this means that dispersion of the noise in all points of factor space $\vec{x}_i$ is the same.

The last two conditions can be written other way around $E\{e.e'\} = \sigma^2 W$ where

W is the unit matrix with dimension equals to $n \times n$;

n is the amount of experiments (amount of measures of the output;

$e' = (e_1, e_2, \cdots, e_n)$ is the vector of the values of the errors in the points where the measures were taken;

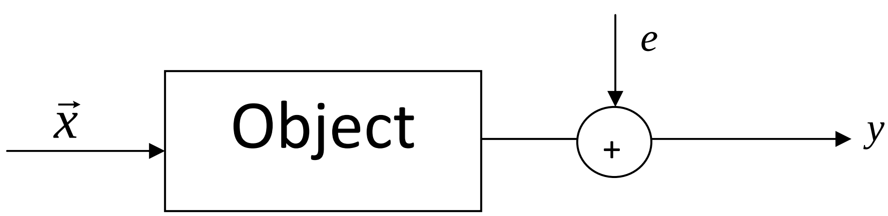
E is the mean operator;

$$\eta(\vec{x}, \vec{\theta}) = \sum_{i=1}^{k} \theta_i f_i(\vec{x});$$

θ_i - coefficients of the model;

$f_i(\vec{x})$- given functions of input variables.

The schema of the object can be given now by the following sketch:



What do we want to do?

Sometimes we want to find not the discrete value of the response function and instead we want to find some integral indexes of the function, it means

$$\overline{\mu}_j = \int_{\Omega} y\omega_j(\vec{x})d\vec{x} = \int_{\Omega}\left[\eta(\vec{x}, \vec{\theta}) + e\right]\omega_j(\vec{x})d\vec{x} = \int_{\Omega}\left[\sum_{i=1}^{k} \theta_i f_i(\vec{x}) + e\right]\omega_j(\vec{x})d\vec{x}$$

The estimation of the indexes we will find:

$$\hat{\overline{\mu}}_j = E(\overline{\mu}_j) = \int_{\Omega}\sum_{i=1}^{k} \hat{\theta}_i f_i(\vec{x})\omega(\vec{x})d\vec{x};$$

where $\overline{\mu}_j$ is the estimated index;

y is the output parameter (output variable);

$\hat{\theta}_i$ is the estimation of the coefficient θ_i of the model;

$\omega_j(\vec{x})$ given not random weight function that in general depends of the input factors;

Ω is the given region of factor space where we calculate the integral idexes;

$j = \overline{1,M}$; M is the amount of calculated integral indexes.

We will look for the estimators of the coefficients of the model inside linear class of estimators, what means:

$\hat{\theta} = T\vec{y}$;

Where $y' = (y_1, y_2, \cdots y_n)$ is the vector of the values of response function in different points $\vec{x}_i$ of the factor space;

n is the amount of experiments and

T is a matrix with dimension $k \times n$

We will look for estimators with the following characteristics

- $E\left(\hat{\vec{\bar{\mu}}}\right) = \vec{\bar{\mu}}_{real}$ what means that their expected value will be the same as the real value of the integral indexes;
- $\lim\limits_{n \to \infty} p\left[\left(\hat{\vec{\bar{\mu}}} - \vec{\bar{\mu}}_{real}\right)'\left(\hat{\vec{\bar{\mu}}}_n - \vec{\bar{\mu}}_{real}\right) \geq \xi\right] = 0$ what means that the estimation converge from the probabilistic point of view to the real values of the estimated parameters. Index n signify that estimation $\hat{\vec{\bar{\mu}}}_n$ was obtained after n measures and ξ is any in front given positive number and
- $D\left(\hat{\vec{\bar{\mu}}}\right) \leq D\left(\hat{\vec{\bar{\mu}}}_0\right)$ where $D\left(\hat{\vec{\bar{\mu}}}\right)$ covariance matrix and $D\left(\hat{\vec{\bar{\mu}}}_0\right)$ is the covariance matrix of any unbiased estimations of $\vec{\bar{\mu}}_0$.

These estimators are known as the best linear estimators and the estimations that we obtain are the best linear estimation of the integral index of the response function.

Optimal Experimental Plans

It is known that in order for us to improve the quality of the properties of the statistic estimations we need, somehow, to select points in the factor space to do our measures, it means, WE NEED TO PLAN THE EXPERIMENT!!

Today we can find in the literature different type of experimental plans.

When we plan an active experiment, the necessary statistic material for the estimation of the parameters is collected following a define research program. Research program is the experimental plan and it satisfies certain criterion of optimality.

About Experimental Plans

They are divided in exact and continuous plans.

Exact plans are optimal for a given number of observations N

The task of finding an optimal exact plan is done by finding where, in the plan region, measurements should be taken to satisfy the given criterion of optimality.

If the obtained plan is concentrated in $n \le N$ points them we can define the observation frequency in point l as $\xi_l = r_l / N$, where r_l is the number of observations done in point x_l. From what was said we have $\sum_{l=1}^{n} \xi_l = 1$ where ξ_l proportion of observations that were done on point x_l considering the total amount of done observations as the unit.

The main characteristic for exact plans is that $r_l = \xi_l N$ where r_l is a positive integer.

Continuous plans are not related to a specific number of observations. These plans are given by a positive probability metric $\xi(x)$, that is $\int_x d\xi(x) = 1$. A continuous norm plan ε is the following set of magnitudes :

$$\varepsilon = \begin{Bmatrix} x_1 & x_2 & . & . & . & x_n \\ \xi_1 & \xi_2 & . & . & . & \xi_n \end{Bmatrix}$$

where x_i are points of the spectra of the plan

and $x_i \in X$ where X is the planning region;

ξ_i is the frequency of observations in corresponding points of the plan.

We can find a correspondence between Norm Plans, Norm Information Matrix of the Plan and Norm Covariance Matrix.

Optimal Experimental Plans

Norm Information Matrix of the Plan is given by: $L(\varepsilon) = \int\limits_{x} f(x) f'(x) d\xi(x)$ where

$f'(x) = (f_1(x), f_2(x), \cdots, f_k(x))$ is the base in which the response function is decomposed.

In the case that the metric is contained in a finite number of points we have :

$$L(\varepsilon) = \sum_{i=1}^{n} \xi_i f(x_i) f'(x_i)$$

And the Norm Covariance Matrix is:

$$D(\varepsilon) = L^{-1}(\varepsilon)$$

Statistic properties of the estimation of integral index of response functions

Let us assume that $\eta(\vec{x}, \vec{\theta})$ is a function linear related to parameters, this is: $\eta(\vec{x}, \vec{\theta}) = \theta' f(\vec{x})$ where $f'(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \cdots f_k(\vec{x}))$.

In the points $x_1, x_2, \cdots x_n$ were done independently measures $y_1, y_2, \cdots, y_n$ with dispersions equal to $\sigma^2{}_1, \sigma^2{}_2, \cdots, \sigma^2{}_n$

We will analyze only linear estimation for θ what means that we are looking for such estimations that could be represented as $\hat{\theta} = Ty$ where $y' = (y_1, y_2, \cdots y_n)$ and T is a matrix with dimensions $k \times n$.

It is known from literature the following theorem:

The best linear estimation for the unknown parameters θ are $\hat{\theta} = B^{-1} y$ where matrix B is equal to $B = \sum_{i=1}^{n} v_i f(\vec{x}_i) f'(\vec{x}_i)$ and is not a degenerated matrix; $v_i = \sigma^{-2}{}_i$. Covariance matrix of estimation is equal to $D(\hat{\theta}) = B^{-1}$.

One corollary of the above theorem is that the estimation of any linear combination $t = C\theta$ will be $\hat{t} = C\hat{\theta}$. Covariance matrix of estimations $\hat{t}$ is equal to $D(\hat{t}) = CD(\hat{\theta})C'$

Optimal Experimental Plans

Let us demonstrate that our estimations of $\vec{\bar{\mu}}$ (integral indexes of response function) are linear combination of the coefficient of the model:

$$\eta(\vec{x},\vec{\theta}) = \sum_{j=1}^{k} \theta_j f_j(\vec{x})$$

$$E[\bar{\mu}_l] = \int_{\Omega} \omega_l(\vec{x})\eta(\vec{x},\vec{\theta})d\vec{x} = \int_{\Omega} \omega_l(\vec{x}) \sum_{j=1}^{x} \theta_j f_j(\vec{x})d\vec{x} =$$

$$= \theta_1 \int_{\Omega} \omega_l(\vec{x}) f_1(\vec{x})d\vec{x} + \cdots + \theta_k \int_{\Omega} \omega_l(\vec{x}) f_k(\vec{x})d\vec{x}$$

Using matrix notation we can represent the last expression as:

$$\vec{\bar{\mu}} = \begin{pmatrix} \int_{\Omega} \omega_1 f_1 dx & \cdots & \int_{\Omega} \omega_1 f_m dx \\ \vdots & \cdots & \vdots \\ \int_{\Omega} \omega_l f_1 dx & \cdots & \int_{\Omega} \omega_l f_m dx \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_m \end{pmatrix}$$

Where $\vec{\bar{\mu}} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_l \end{pmatrix}$

It is known too that the best linear estimation for $\hat{\theta}$ minimizes the sum of the weight squares of the difference between the real value and the one that is calculated by the model (Lease Square Method, LSM)

$$S(\theta) = \sum_{i=1}^{n} v_i \left[y_i - f'(\vec{x}_i)\theta\right]^2$$

Optimal Experimental Plans

If we put together what was said in the statistic properties of the estimation and what LSM says we can conclude that covariance matrix for our integral indexes is the following:

$D(\hat{\mu}) = \phi D(\hat{\theta})\phi'$ where $\phi = \begin{pmatrix} \int\limits_{\Omega} \omega_1(\vec{x}) f_1(\vec{x}) d\vec{x} & \cdots & \int\limits_{\Omega} \omega_1(\vec{x}) f_k(\vec{x}) d\vec{x} \\ \vdots & \vdots & \vdots \\ \int\limits_{\Omega} \omega_m(\vec{x}) f_1(\vec{x}) d\vec{x} & \cdots & \int\limits_{\Omega} \omega_m(\vec{x}) f_1(\vec{x}) d\vec{x} \end{pmatrix}$

Why is actual the problem?

In different situation scientists, engineers and researchers could find one or more of the following real problems:

- estimation of the mean value of the response function and not the function itself;
- estimation of the volume of raw material in a mine;
- estimation of the amplitude of harmonics in a complex signal;
- estimation of some ordinates of transformed function;
- estimation of the power of a signal for given frequencies.

What is the objective of the experiment?

Estimate $E\lfloor \bar{\mu}_l \rfloor$;

Where $E\lfloor \bar{\mu}_l \rfloor = \int\limits_{\Omega} \omega_l(\vec{x}) \eta(\vec{x}, \vec{\theta}) dx$

And $\eta(\vec{x}, \vec{\theta}) + e = \sum\limits_{i=1}^{k} \theta_i f_i(\vec{x}) + e$ is the response function

which is known except for the numerical values of the parameters.

Ω Is the given region of factor space;

$l = \overline{1, M}$; M number of integral indexes that need to be calculated and

$\omega_l(\vec{x})$ given non-random weight function of input factors.

What do we need to find?

We need to find: ε^* that belongs to the class of continuous and norm plans E such that produces the minimum value of variance of the estimator $\vec{\bar{\mu}}$.

Definition of Experimental I_D Optimal Plan

Plan ε^* is an I_D optimal plan if and only if it minimizes the determinant of the Covariance Matrix, that is,

$$\min_{\varepsilon\in E} \det D\left(\hat{\vec{\bar{\mu}}}_\varepsilon\right) = \det D\left(\hat{\vec{\bar{\mu}}}_{\varepsilon^*}\right) \quad \text{or} \quad \max_{\varepsilon\in E} \det L\left(\hat{\vec{\bar{\mu}}}_{\varepsilon*}\right) = \det L\left(\hat{\vec{\bar{\mu}}}_{\varepsilon*}\right)$$

Where L is the plan Information Matrix

I_D is the plan which minimizes the volume of the dispersion ellipsoid.

Some examples:

The response function mean value estimation. Application # 1

We will assume that we have a weight function $\omega_1 \equiv 1$ then:

$$E\left[\bar{\mu}_1\right] = \int_\Omega \eta\left(\vec{x}, \vec{\theta}\right) dx$$ **, is the mean value of the response function;**

Meaning: in different technological objects the inputs can change

$$\vec{x}_{i\,\min} \le \vec{x}_i \le \vec{x}_{i\,\max}$$

we need to estimate the mean value of the output index.

Estimation of the amplitude of given harmonic of the response function. Application # 2

The given information is the same as before with the exception of weight functions. In this problem they are the same as the Fourier series,

$$\vec{\omega}(x) = \left[\sin(x), \cos(x), \sin(2x), \cos(2x), \cdots, \sin(Mx), \cos(Mx)\right]$$

$$E\left[\overline{\mu}^{1}\right]=\frac{1}{L}\int_{-L}^{L}\cos\frac{k\pi}{L}x\,\eta\left(x,\overline{\theta}\right)dx \qquad E\left[\overline{\mu}^{2}\right]=\frac{1}{L}\int_{-L}^{L}\sin\frac{k\pi}{L}x\,\eta\left(x,\overline{\theta}\right)dx$$

Meaning: a parametric signal which describes complex periodic movement is given and we need to select some of its components

Estimation of some ordinates of a given transform function. Application # 3

Given: parametric signal $\eta\left(\vec{x},\vec{\theta}\right)$ in the region X

A transformation to region O_M is performed.

This means

$$\eta_{\vec{x}\in X}\left(\vec{x},\vec{\theta}\right)\rightarrow\tilde{\eta}_{\omega\in O_M}\left(\omega,\vec{\beta}\right) \quad \text{where} \quad \tilde{\eta}\left(\omega,\vec{\beta}\right)=\int_X h(\omega,\vec{x})\,\eta\left(\vec{x},\vec{\theta}\right)dx$$

and $h(\omega,\vec{x})$ weight functions that areused to go from region X to region O_M

Significance: if $h(\omega,x)=e^{i\omega x}$ then we are applying Fourier Transformation to the input signal and we want to estimate the power of this signal for given frequencies.

Analytic synthesis of a I_D plan

Given: $\eta\left(x,\vec{\theta}\right)=\theta_0+\theta_1 x \quad x\in X=\left[-1,1\right]$

$$\varepsilon^{*}\in E=\begin{Bmatrix}-1 & 0 & 1\\ l_1 & 1-2l_1 & l_1\end{Bmatrix}$$

We need to find a plan E which belongs to class norm and continuous plans and it will give the minimum dispersion for the estimation of mean value of the response function in the region X

$$D\left(\hat{S}\right)\underset{\varepsilon\in E}{\rightarrow}\min$$

That is,

Let us find the Information Matrix of the Plan. $L(\varepsilon)=\sum_{i=1}^{3} l_i f(x_i) f'(x_i)=$

$$=l_1\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}+(1-2l_1)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}+l_1\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 2l_1\end{pmatrix}$$

Vector ϕ in this case has the following form $\phi=(2,0)$

Them $D(\hat{S})=\phi D(\hat{\theta})\phi'=\phi[L(\varepsilon)]^{-1}\phi'=\ (2,0)\begin{pmatrix}1 & 0\\ 0 & \frac{1}{2}l_1\end{pmatrix}\begin{pmatrix}2\\ 0\end{pmatrix}=4$

This result tells us that estimation dispersion of the mean of the response function does not depend of the frequencies of the plan, meaning that we can distribute the resources the way we want.

We can use 50% in the extremes (any person will think that this is the correct distribution and different type of optimal experimental plans recommended so) or you can put all your resources in the center of the plan.

The physical meaning of this result is that for accurate estimation of the mean value of a straight line you can distribute the resources uniformly at the ends of the plan region or you can put the resources all on the plan center to accurately estimate the free term.

Some conclusions

- It is really convenient to have optimal experimental plans that specifically deal with integral indexes of the response function.

- Some results were totally unexpected.
- There are many applications for these plans.
- There is a numerical method to synthesize these type of plans based on nonlinear optimization and use the Rosembrok's algorithm on rotation of coordinates.

Optimal Experimental Plans

References

- Doctoral Thesis “Synthesis Optimal Experimental Plans for Calculation Integral Indexes of the Response Function” by Manuel Caramés, Moscow 1983 (in Russian);
- Letski E., Caramés M. Experimental Plans for Accelerated *Life Testing.* News for Machinery Industry. !979, No 2. (in Russian)
- Experiment Theory by Nalimov V.V. Nauka 1971 (in Russian);
- Experimental Optimal Theory by Fedorov V.V. Nauka 1971 (in Russian);
- Experimental Design for Researching Technological Processes. Hartman K., Letski E., Shefer V. Mir 1977 (in Russian);
- Krug G. K. Doctoral Thesis “Methods of Statistic Analysis for Objects that need to be Controlled” Moscow 1968.