ОБЪЕДИНЕННЫЙ
ИНСТИТУТ
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

Дубна

E5 - 3247

V.V. Fedorov, A. Pázman

DESIGN OF EXPERIMENTS BASED ON THE MEASURE OF INFORMATION

ЛАБОРАТОРИЯ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ
И АВТОМАТИЗАЦИИ

1967.

E5 - 3247

V.V. Fedorov, A. Pázman

DESIGN OF EXPERIMENTS BASED ON THE MEASURE OF INFORMATION

In the case of a nonlinear dependence of measured quantities on the studied parameters (for example, in the phase shift analysis of scattering of elementary particles) a situation offen occurs when the likelihood function[1] has two or more peaks with nearly equal levels.

The position of these peaks is considered to define the sets of the possible estimates of the parameters. The height of the peaks defines, roughly speaking, the reliability of these estimates and the width of each peak corresponds to the error matrix (dispersion matrix) of the proper set of estimates.

If the aim of the experimenter is only to discriminate the sets of the estimates and he is not interested how the error of each parameter is decreased then planning as has been described in ref.[2] is necessary. If the experimenter is going the maximal amount of information on some parameters he is to use a method which specifies and discriminates the estimates at the same time.

The earlier methods of design and comparison of experiments which are based on the error matrix are not applicable in our case. Therefore, a more general measure of our knowledge than the error matrix should be used, namely the measure of amount of information which is based on the notion of entropy. The investigation of the general properties of this measure applied to the comparison of our experiments is given in ref.[3]. In planning experiments this measure was used for the first time in refs.[4] and[5], but in the particular relatively simple case when the measured quantities are linearly dependent on the parameters studied.

This paper is aimed at using the ideas which have been developped in ref.[3] in planning experiments in the case of nonlinear parametrization and developing some computing methods.

Let us denote $\eta(\vec{\theta}, x)$ the measured quantity,

$$\vec{\theta} = \begin{pmatrix} \theta_1 \\ \cdot \\ \cdot \\ \theta_m \end{pmatrix} \tag{1a}$$

the parameters and let

$$\vec{y} = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \tag{1b}$$

be the results of experiments. Then

$$I(\vec{y}) = \int_{\Omega} p(\vec{\theta}/\vec{y}) \ln p(\vec{\theta}/\vec{y})\, d\vec{\theta} \tag{2}$$

is the entropy amount of information which is contained in the experiments with results $\vec{y}$. Here $p(\vec{\theta}/\vec{y})$ is the normalized likelihood function,

$$\int_{\Omega} p(\vec{\theta}/\vec{y})\, d\vec{\theta} = 1 \tag{3}$$

and Ω is the region of the possible values of $\vec{\theta}$.

If the planned experiment consists in the measurements of η at some point x during the time t, then the amount of information is increased by

$$\Delta I(x,t,y) = \int_{\Omega} p(\vec{\theta}/y) \ln p(\vec{\theta}/y)\, d\vec{\theta} - \int_{\Omega} p(\vec{\theta}) \ln p(\vec{\theta})\, d\vec{\theta} \tag{4}$$

where $p(\vec{\theta})$ is the value of the normalized likelihood function at $t=0$, y is the result of the measurement at the point x. The planning consists in finding such a point at which the increase of the amount of information will be maximal. The function $\Delta I(x,t,y)$ depends on the value of y which is unknown before the experiment. We shall consider, therefore, as optimal the point x_0 at which the expected value of the mean increase of the amount of information is maximal, i.e.

$$\Delta J(t, x_0) = \max_x \Delta J(t, x). \tag{5}$$

Here

$$\Delta J(t,x) = \int_Y p(y) \int_{\Omega} p(\vec{\theta}/y) \ln p(\vec{\theta}/y)\, d\vec{\theta}\, dy - \int_{\Omega} p(\vec{\theta}) \ln p(\vec{\theta})\, d\vec{\theta} =$$

$$= \int_Y \int_{\Omega} p(y, \vec{\theta}) \ln \frac{p(y, \vec{\theta})}{p(y)\, p(\vec{\theta})}\, dy\, d\vec{\theta} \tag{6}$$

$$p(y) = \int_{\Omega} p(y/\vec{\theta})\, p(\vec{\theta})\, d\vec{\theta} \tag{7}$$

$$p(\vec{\theta}/y) = \frac{p(y/\vec{\theta})\, p(\vec{\theta})}{p(y)} \tag{8}$$

$$p(y, \vec{\theta}) = p(y/\vec{\theta})\, p(\vec{\theta}) = p(\vec{\theta}/y)\, p(y) \tag{9}$$

(see for example ref.[6]).

$p(y/\vec{\theta})$ is the conditional probability density of y assumed to be known a priori.

In the case of scattering experiments for a "sufficiently rich statistics":

$$p(y/\vec{\theta}) = \frac{1}{\sqrt{2\pi(\lambda(x)\,t)^{-1}}}\, e^{-\frac{1}{2}(y-\eta(\vec{\theta},x))^2 \lambda(x)\, t} \tag{10}$$

where $\lambda(x)$ is the efficiency of the measurement of $\eta(\vec{\theta},x)$.

Further we shall assume that (10) is true. If the experimenter is interested in the maximal amount of information on k required parameters, the increase of information on these "useful" parameters is given as

$$\Delta J(t,x,\vec{\omega}) = \int_Y \int_\Omega p(y,\vec{\theta}) \ln \frac{p(y,\vec{\omega})}{p(y)\,p(\vec{\omega})}\, dy\, d\vec{\theta} \tag{11}$$

where

$$\vec{\omega} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} \text{ - are useful parameters,}$$

$$\vec{\phi} = \begin{pmatrix} \theta_{k+1} \\ \vdots \\ \theta_m \end{pmatrix} \text{ - are nuisance parameters,}$$

$$p(\vec{\omega}) = \int p(\vec{\theta})\, d\vec{\phi} \tag{12}$$

and
$$p(y,\vec{\omega}) = \int p(y,\vec{\theta})\, d\vec{\phi} \tag{13}$$

In a general case for arbitrary functions $p(\vec{\theta})$ and $\eta(x,\vec{\theta})$ the direct computing of the mean increase of information $\Delta J(t,x,\vec{\omega})$ is a cumbersome computing problem because one has to compute multidimensional integrals for each value of x. In this paper some approximate results are obtained.

Let us assume that the time required for the additional experiment is short, i.e.

$$\Delta J(t,x,\vec{\omega}) = \frac{\partial J(t,x,\vec{\omega})}{\partial t}\Big|_{t=0} t \tag{14}$$

In this case, as follows from (5), the optimal point corresponds to

$$\max_x \frac{\partial J(t,x,\vec{\omega})}{\partial t}\Big|_{t=0} \tag{15}$$

In Appendix I the following equality is proved

$$\frac{\partial J(t,x,\vec{\omega})}{\partial t}\Big|_{t=0} = \frac{\lambda(x)}{2} \left\{ \int \frac{\left[\int \eta(\vec{\theta},x)\, p(\vec{\theta})\, d\vec{\phi}\right]^2}{p(\vec{\omega})}\, d\vec{\omega} - \left[\int \eta(\vec{\theta},x)\, p(\vec{\theta})\, d\vec{\theta}\right]^2 \right\} \tag{16}$$

In many cases of practical interest we can approximate the likelihood function by a superposition of the normal probability densities:

$$p(\vec{\theta}) = \sum_{i=1}^{N} W_i\, p_i(\vec{\theta}) \tag{17}$$

Here

$$p_i(\vec{\theta}) = \frac{1}{(2\pi)^{\frac{m}{2}}\, |D_i|^{1/2}}\, e^{-\frac{1}{2}(\vec{\theta}-\hat{\vec{\theta}}_i)' D_i (\vec{\theta}-\hat{\vec{\theta}}_i)}, \tag{18}$$

$\hat{\vec{\theta}}_i$ is the i-th set of estimates,

D_i is the corresponding error matrix

and
$$W_i = p(\hat{\vec{\theta}}_i)\, |D_i|^{1/2}$$

The approximation of $p(\vec{\theta})$ given in (17) corresponds to the quasi-linearity of $\eta(\vec{\theta},x)$ in the neighbourhood of each estimate $\hat{\vec{\theta}}_i$:

$$\eta(\vec{\theta},x) = \eta(\hat{\vec{\theta}}_i, x) + \vec{f}_i'(x)(\vec{\theta}-\hat{\vec{\theta}}_i) \tag{20}$$

where

$$f_{i,\alpha} = \frac{\partial \eta(\vec{\theta},x)}{\partial \theta_\alpha}\Big|_{\vec{\theta}=\hat{\vec{\theta}}_i}$$

We shall compute the rate of the increase of the amount of information under the assumptions from (17) to (20) for the two most important cases: $k=m$ and $k=1$

A. $k = m$

Substituting (17) and (20) into (16) we obtain

$$\frac{\partial J}{\partial t}\Big|_{t=0} = \frac{\lambda(x)}{2}\left[\sum_{i=1}^{N} W_i \sigma_i^2(x) + \sum_{i=1}^{N} W_i\left(\eta_i(x) - \sum_{k=1}^{N} W_k \eta_k(x)\right)^2\right] \tag{21}$$

where $\eta_i(x) = \eta(\hat{\vec{\theta}}_i, x)$ and $\sigma_i^2(x) = \vec{f}_i'(x) D_i \vec{f}_i(x)$ is the squared corridor of errors of the curve $\eta_i(x)$.

If only one set of estimates exists (i.e. $p(\vec{\theta}) = p_1(\vec{\theta})$) then (21) coincides with (7) from /5/. If the corridors of errors are narrow compared to the distances between the curves $\eta_1(x), \dots, \eta_N(x)$, the measurement must be made at the point x which gives the maximal dispersion of these curves.

B. $k = 1$

If one parameter is specified ($\vec{\omega} = \theta_1$) then we obtain from (16), (17) and (20) (see Appendix 2)

$$\frac{\partial J(t, x, \theta_1)}{\partial t}\Big|_{t=0} = \frac{\lambda(x)}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} W_i W_j \{\eta_i(x)\eta_j(x) a_{ij} + + 2\eta_i(x) g_j(x) b_{ij} + g_i(x) g_j(x) c_{ij}\} \tag{22}$$

where

$$g_i(x) = f_{i,1}(x) + \sum_{\alpha=2}^{m} \frac{D_{i,1\alpha}}{D_{i,11}} f_{i,\alpha}(x)$$

and a_{ij}, b_{ij}, c_{ij} are constants (see Appendix 2: (A.23)-(A.25)).

APPENDIX I

Teorem

If $\vec{\theta} = \begin{pmatrix}\vec{\omega}\\ \vec{\phi}\end{pmatrix}$ and if

$$p(y|\vec{\theta}) = \sqrt{\frac{\lambda(x)t}{2\pi}}\, e^{-\lambda(x)t\frac{[y-\eta(\vec{\theta},x)]^2}{2}} \tag{A.1}$$

then

$$\frac{\partial J(t, x, \vec{\omega})}{\partial t}\Big|_{t=0} = \frac{\lambda(x)}{2}\{\int p(\vec{\omega})[E_{\vec{\phi}/\vec{\omega}}\eta(\vec{\theta}, x)]^2 d\vec{\omega} - [E_{\vec{\theta}}\eta(\vec{\theta}, x)]^2\} \tag{A.2}$$

and

$$\frac{\partial^2 J(t, x, \vec{\omega})}{\partial t^2}\Big| = \frac{\lambda^2(x)}{4}\{\int p(\vec{\omega})[E_{\vec{\phi}/\vec{\omega}}(\eta(\vec{\theta}, x) - E_{\vec{\phi}/\vec{\omega}}\eta(\vec{\theta}, x))^2]^2 d\vec{\omega} - - [E_{\vec{\theta}}(\eta(\vec{\theta}, x) - E_{\vec{\theta}}\eta(\vec{\theta}, x))^2]^2\} \tag{A.3}$$

where

$$E_{\vec{\theta}}\eta(\vec{\theta}, x) = \int p(\vec{\theta})\,\eta(\vec{\theta}, x)\, d\vec{\theta} \tag{A.4a}$$

and

$$E_{\vec{\phi}/\vec{\omega}}\eta(\vec{\theta}, x) = \int p(\vec{\phi}/\vec{\omega})\,\eta(\vec{\theta}, x)\, d\vec{\phi} \tag{A.4b}$$

Proof

By definition:

$$\frac{\partial J(t, x, \vec{\omega})}{\partial t}\Big|_{t=0} = \lim_{t\to 0} \frac{\Delta J(t, x, \vec{\omega})}{t} \tag{A.5}$$

From (11) and (A.1) we obtain

$$\Delta J(t, x, \vec{\omega}) = \int p(\vec{\theta})\, p(y/\vec{\theta}) \ln \frac{\int p(\vec{\theta})\, e^{\frac{t}{2}\eta(\vec{\theta},x)[2y-\eta(\vec{\theta},x)]} d\vec{\phi}}{p(\vec{\omega}) \int p(\vec{\theta})\, e^{\frac{t}{2}\eta(\vec{\theta},x)[2y-\eta(\vec{\theta},x)]} d\vec{\theta}}\, d\vec{\theta}\, dy \tag{A.6}$$

Let us denote

$$\int p(\vec{\theta})\, e^{\frac{t}{2}\eta(\vec{\theta},x)[2y-\eta(\vec{\theta},x)]}\, d\vec{\phi} = u(t) \qquad \text{(A.7a)}$$

$$\int p(\vec{\theta})\, e^{\frac{t}{2}\eta(\vec{\theta},x)[2y-\eta(\vec{\theta},x)]}\, d\vec{\theta} = v(t) \qquad \text{(A.7b)}$$

and

$$\ln \frac{u(t)}{p(\vec{\omega})\, v(t)} = h(t,y,\vec{\omega}) \qquad \text{(A.8)}$$

The Taylor expansion gives

$$h(t,y,\vec{\omega}) = h(0,y,\vec{\omega}) + t\,\dot{h}(0,y,\vec{\omega}) + \frac{t^2}{2}\ddot{h}(0,y,\vec{\omega}) + \frac{t^3}{6}\dddot{h}(r,y,\vec{\omega}) \qquad \text{(A.9)}$$

where

$$0 \le r \le t$$

It will be readily seen that

$$h(0,y,\vec{\omega}) = 0,\ \dot{h}(0,y,\vec{\omega}) = \frac{\dot{u}(0)}{u(0)} - \frac{\dot{v}(0)}{v(0)} \qquad \text{(A.10)}$$

$$\ddot{h}(0,y,\vec{\omega}) = \frac{\ddot{u}(0)}{u(0)} - \left[\frac{\dot{u}(0)}{u(0)}\right]^2 - \frac{\ddot{v}(0)}{v(0)} + \left[\frac{\dot{v}(0)}{v(0)}\right]^2 \qquad \text{(A.11)}$$

$$\dddot{h}(r,y,\vec{\omega}) = \frac{\dddot{u}(r)}{u(r)} - 3\frac{\ddot{u}(r)}{u(r)}\frac{\dot{u}(r)}{u(r)} + 2\left[\frac{\dot{u}(r)}{u(r)}\right]^3 - \frac{\dddot{v}(r)}{v(r)} + 3\frac{\ddot{v}(r)}{v(r)}\frac{\dot{v}(r)}{v(r)} - 2\left[\frac{\dot{v}(r)}{v(r)}\right]^3 \qquad \text{(A.12)}$$

The derivatives $\dot{u}(t), \dot{v}(t)$ and so on, can be obtained by derivating (A.7a) and (A.7b). For example

$$\dot{u}(t) = \frac{1}{2}\int p(\vec{\theta})\, e^{\frac{t}{2}\eta(\vec{\theta},x)[2y-\eta(\vec{\theta},x)]}\, \eta(\vec{\theta},x)[2y-\eta(\vec{\theta},x)]\, d\vec{\phi} \qquad \text{(A.13)}$$

We can interchange the order of derivation and integration in (A.13) since the integrated function in (A.13) is a monoton function of time (see ref. /7/).

From (A.5)-(A.9) we obtain

$$\left.\frac{\partial J(t,x,\vec{\omega})}{\partial t}\right|_{t=0} = \lim_{t\to 0} \Big\{ \int p(\vec{\theta})\, p(y|\vec{\theta})\, \dot{h}(0,y,\vec{\omega})\, d\vec{\theta}\, dy + \frac{t}{2}\int p(\vec{\theta})\, p(y|\vec{\theta})\, \ddot{h}(0,y,\vec{\omega})\, d\vec{\theta}\, dy + \frac{t^2}{6}\int p(\vec{\theta})\, p(y|\vec{\theta})\, \dddot{h}(r,\vec{\omega},y)\, d\vec{\theta}\, dy \Big\} \qquad \text{(A.14)}$$

Using the known relation between the central moments of the normal distribution $p(y|\vec{\theta})$ (ref. /6/)

$$m_{2s} = (2s-1)!!\,\sigma^{2s} \qquad \text{(A.15)}$$

and the equality $\lambda(x)\, t = \sigma^2$ it can be shown that in (A.14) the k-th term A_k of the decomposition with respect to t is proportional to

$$A_k \approx t^k m_k \approx t^k \sigma^k \approx t^{\frac{k}{2}} \quad \text{if } k \text{ is even}$$

and

$$A_k \approx t^k \sigma^{k-1} \quad \text{if } k \text{ is odd} \qquad \text{(A.16)}$$

It follows from (A.16) that for computing $\frac{\partial J}{\partial t}$ it is sufficient to take in (A.14) only the two first terms and to compute only those terms of (A.11) which contain y^2.

From (A.10), (A.11), and (A.13) we obtain

$$\dot{h}(0,y,\vec{\omega}) = \frac{\eta(\vec{\theta},x)\int p(\vec{\theta})\eta(\vec{\theta},x)\,d\vec{\phi} - \frac{1}{2}\int p(\vec{\theta})\eta^2(\vec{\theta},x)\,d\vec{\phi}}{p(\vec{\omega})} \qquad (A.17)$$

$$-\eta(\vec{\theta},x)\int p(\vec{\theta})\,\eta(\vec{\theta},x)\,d\vec{\theta} + \frac{1}{2}\int p(\vec{\theta})\,\eta(\vec{\theta},x)\,d\vec{\theta}$$

$$\ddot{h}(0,y,\vec{\omega}) = y^2\left\{\frac{\int p(\vec{\theta})\eta^2(\vec{\theta},x)\,d\vec{\phi} - [\int p(\vec{\theta})\eta(\vec{\theta},x)\,d\vec{\phi}]^2}{p(\vec{\omega})} - \int p(\vec{\theta})\,\eta^2(\vec{\theta},x)\,d\vec{\theta} + \right.$$

$$\left. + [\int p(\vec{\theta})\,\eta(\vec{\theta},x)\,d\vec{\theta}]^2\right\} \quad + \qquad (A.18)$$

terms linear with respect to y.

Substituting (A.17), (A.18) and (A.1) into (A.14) we obtain after simple computation

$$\left.\frac{\partial J(t,x,\vec{\omega})}{\partial t}\right|_{t=0} = \frac{\lambda(x)}{2}\left\{\int \frac{[\int p(\vec{\theta})\eta(\vec{\theta},x)\,d\vec{\phi}]^2}{p(\vec{\omega})}\,d\vec{\omega} - [\int p(\vec{\theta})\eta(\vec{\theta},x)\,d\vec{\theta}]^2\right\} \qquad (A.19)$$

what is equal to (A.2)

The second derivation of the information (A.3) can be computed analogously. The value of $\frac{\partial^2 J}{\partial t^2}$ can be used for the estimation of the accuracy of the relation (14).

APPENDIX 2

We shall compute $\frac{\partial J}{\partial t}$ for $k = 1$. We obtain from (17) and (20)

$$E_{\theta_2,\dots,\theta_m/\theta_1}\,\eta(\vec{\theta},x) =$$

$$= \sum_{i=1}^{N} W_i \int [\eta_i(x) + \vec{f}_i(x)(\vec{\theta} - \hat{\vec{\theta}}_i)]\, p_i(\theta_2,\dots,\theta_m/\theta_1)\,d\theta_2 \dots d\theta_m \frac{p_i(\theta_1)}{\sum_{k=1}^{N} W_k\, p_k(\theta_1)} \qquad (A.20)$$

Here

$$p_i(\theta_1) = \frac{1}{\sqrt{2\pi D_{i,11}}}\, e^{-\frac{1}{2}\frac{(\theta_1 - \hat{\theta}_{i,1})^2}{D_{i,11}}}$$

and $p_i(\theta_2,\dots,\theta_m/\theta_1)$ is a conditional normal probability density. Computing the integral in (A.20) we obtain (about the mean value of a conditional normal distribution see ref. /8/ p. 45):

$$E_{\theta_2,\dots,\theta_m/\theta_1}\eta(\vec{\theta},x) = \sum_{i=1}^{N} W_i\{\eta_i(x) + (\theta_1 - \hat{\theta}_{i,1})[f_{i,1}(x) + \sum_{\alpha=2}^{m}\frac{D_{i,1\alpha}}{D_{i,11}} f_{i,\alpha}(x)]\}\frac{p_i(\theta_1)}{\sum_{k=1}^{N} W_k\, p_k(\theta_1)} \qquad (A.21)$$

Similarly, we obtain

$$E_{\vec{\theta}}\,\eta(\vec{\theta},x) = \sum_{i=1}^{N} W_i\,\eta_i(x) \qquad (A.22)$$

Substituting (A.21) and (A.22) into (A.2) we obtain (22). The constants a_{ij}, b_{ij}, c_{ij} are equal to the following integrals

$$a_{ij} = \int \frac{p_i(\theta_1)\,p_j(\theta_1)}{\sum_{k=1}^{N} W_k\, p_k(\theta_1)}\,d\theta_1 - 1 \qquad (A.23)$$

$$b_{ij} = \int (\theta_1 - \hat{\theta}_{i,1})\frac{p_i(\theta_1)\,p_j(\theta_1)}{\sum_{k=1}^{N} W_k\, p_k(\theta_1)}\,d\theta_1 \qquad (A.24)$$

$$c_{ij} = \int (\theta_1 - \hat{\theta}_{i,1})(\theta_1 - \hat{\theta}_{j,1})\frac{p_i(\theta_1)\,p_j(\theta_1)}{\sum_{k=1}^{N} W_k\, p_k(\theta_1)}\,d\theta_1 \qquad (A.25)$$

References

1. Н.П. Клепиков, С.Н. Соколов. Анализ и планирование экспериментов методом максимума правдоподобия. "Наука", М-1964.
2. V.V.Fedorov, A.Pázman. Preprint E1-3066. Dubna, 1966.
3. D.Lindley. Ann.Math.Stat. 27, 986 (1956).

4. M.Stone. Ann.Math.Stat. 30, 55 (1959).
5. С.Н. Соколов. Теория вероятностей и ее применение. 8, 95 (1963), 8,318 (1963).
6. H.Cramér. Mathematical Methods of Statistics. Princeton, 1946.
7. Г.М.Фихтенгольц. Курс дифференциального и интегрального вычисления. Физматгиз 1959, т. II.
8. Т. Андерсон. Введение в многомерный статистический анализ. Физматгиз М-1963 стр.45.

Received by Publishing Department
on March 30, 1967.