ОБЪЕДИНЕННЫЙ ИНСТИТУТ ЯДЕРНЫХ ИССЛЕДОВАНИЙ

S.N. Sokolov

D-573

MEASUREMENT YIELDING MAXIMUM INFORMATION AND CONTINUOUS PLANNING OF EXPERIMENTS

Дубна 1961

S.N. Sokolov

D-573

MEASUREMENT YIELDING MAXIMUM INFORMATION AND CONTINUOUS PLANNING OF EXPERIMENTS

Summary

It is shown that the rate of accumulation of information for a given group of parameters during the measurement of some function is equal to the partial variance of that function, provided the method of least squares is valid. On this basis the problem of the continuous planning of indirect experiments is solved.

Introduction

The planning of experiments seeks to obtain the largest information for a given group of parameters at the expense of the least efforts. Efforts may imply time, money, and other undesirable factors and their combinations which can be measured numerically and should be reduced to minimum. Among all problems arising in planning experiments, only statistical ones will be discussed here.

A mention that the use of mathematical statistics is necessary for the rational organization of experiments, can be found even in the early works of the creators of the modern theory of statistical analysis of experiments. In particular, R.A.Fisher who first formulated the principle of maximum likelihood, considered planning as the main inverse problem of statistics. However, Fisher and his followers developed exclusively the theory of planning direct experiments under badly controlled conditions that is experiments in industry, trade, agriculture, and so on. At the same time, physics and related sciences deal, as a rule, with indirect experiments when a great number of different measurements are combined in order to calculate a small group of parameters having an immediate theoretical interest. Here, in passing from the immediate results of observation to the values of measured parameters there is a method of least squares which complicates the picture. The theory of planning indirect experiments scarcely developed, and in the latest literature one can find only papers having far too distant relation (if any) to the actual problems of planning physical experiments.

In 1958 N.P.Klepikov and the author of the present article investigated the planning of experiments in which extrapolation of the measured curve in the inaccessible regions was needed. In particular, we had in mind measurements of scattering amplitude at zero angle. The problem was solved, and the results were expounded in the book/1/, which contained a chapter devoted to the planning of experiments. In this chapter the problem just mentioned and some its generalizations are treated, the notion of difficulty function h is introduced, and many questions connected with planning are discussed.

The solution of the extrapolation problem showed that there exist, in fact, most advantageous points of measurement and most rational distribution of total experimental time among these points. It showed also that a comparatively small displacement of the points off their best positions or the violation of rational distribution of time diminishes sharply the precision of extrapolation what confirms the practical urgency of planning . It is worth noting that an obvious graphical form was found for the recipe for choosing a variant of experimental equipment and positions of measurements, as well as for estimating a precision attainable at extrapolation.

The main shortcoming of the planning procedure proposed in paper /1/ was always understood to be a static formulation of the problem when the whole experiment is planned at once, the difficulty function h being supposed known beforehand and no earlier, "unplanned" measurements to exist. Meanwhile, in reality the data necessary for planning are accumulating and becoming more accurate only little by little in the course of an experiment, so it is very important to be able to find most advantageous positions for measurements simultaneously with the measurements themselves that is to make a continuous planning. Indeed, no experiment is ever born as a whole and complete thing, and after some groping measurements and taking into account all the existing data one should decide what step to make next, that is, what measurement to choose next in order to reach the aim : the determination of theoretically important parameters with a sufficient precision, as fast as possible. The choice of the best next measurement depending on the results thus far obtained may be also called the dynamical planning. In the course of a continuous planning, neither the best measurement positions, nor the time distribution coincide with those obtained by the static planning (although they are tending to the latter ones when the experiment is infinitely continued), since the measurements already made change the importance of individual regions .

In the foregoing we assume that the results of the observations are normally distributed and linearly dependent upon the parameters a we are determining. The formulae obtained below will always be valid in practice when the method of least squares can be applied. This implies, on the one hand, that the dependence of the direct results ot observations on the parameters we are measuring may be non-linear, and on the other , that the distributions of probabilities of results of observations may be somewhat different from the Gaussian ones, but this non-linearity and the violation of a Gaussian form of the distributions must be reasonably small.

In fact, for applications it is essential only that the distribution of the parameters a would be alike the Gaussian distribution in the vicinity of two standard deviations and that the error matrix would keep its usual meaning.

1. The Problem

When we measure experimentally a certain curve $y(a,x)$ at some points x_i we do not usually take interest in all the parameters a which this curve depends upon. There are a lot of examples illustrating that in order to specify one-two quantities A important from the theoretical point of view an experiment had to be arranged on the measurement of the curve dependent upon 10-15 unknown parameters. As usual, this happens because the parameters of interest cannot be measured directly, so that there remains nothing but to measure them together with the ballast parameters which are containing in the dependence $y(a,x)$ but which we are not interested in. In planning an experiment, one wishes, naturally, to spend as much time as possible to specify the quantities which are of interest, and as less time as possible, to specify those balast parameters which have to be measured, unfortunately, jointly with the parameters of interest.

A continuous planning of an experiment is aimed at finding such a measurement among all those to be made at the next moment that would yield the largest additional information possible about the given group of quantities and would require the least labour consumption. As we shall see later this formulation of the problem implies that the accuracy of each of the parameters A or their linear combinations are uniformly important to us. It may sometimes happen that some of A-s are relatively more important than others. In this case a more general functional expressing the accuracy of A-s should be used instead of information. We shall not discuss this more general case in this paper.

It is clear intuitively, that those measurement points are most advantageous at which the dependence of the curve $y(a,x)$ on the quantities we are interested in, is expressed most strongly. On the contrary, when the behaviour of the curve $y(a,x)$ is mainly determined by the ballast parameters, the results of the measurements will bring almost nothing new about the quantities for the measurements of which the experiment is arranged. Therefore, information about A-s must be connected in some way with the sensitivity of $y(a,x)$ to the variations of A-s. To investigate this connection we have first to know what amount of information (and about what quantities) can be obtained from the measurement of the dependence $y(a,x)$ at each point x.

Further we will go on considering only the planning of such experiments the purpose of which is to specify the given group of the quantities $A = \{ a_1, ..., a_r\}$. Evidently, one may specify some quantities only when some primitive, groping measurements of these quantities have already been performed. Then, basing on such preliminary experiments, one may try to answer the question how to choose the next measurement in order to obtain, by spending minimum efforts, as much additional information as possible about the quantities $a_1, ..., a_r$.

Additional information about the quantities A may be, generally speaking, obtained by measuring any function $y(c)$ of the parameters c (known with errors), provided the parameters are intercorrelated in some way with the quantities A. For instance, the parameters c were measured earlier jointly with A or are their functions. In the latter case, not loosing in generality, one can consider some of the parameters c coinciding with some of the quantities A. The rest of the parameters (we shall denote them by B) will hereafter be referred to as ballast ones. (By calling them "ballast" we stress that it is not the aim of the present experiment to specify them *). A set of all the quantities under consideration - i.e., the quantities A and ballast parameters B will be denoted by $a = \{ a_1, ..., a_m\}$.

* In practice it may turn out useful to measure $y(c)$ even if all the parameters c are ballast ones.

Let us suppose that the information about the parameters a which is available at the beginning of the experiment, is given in the form of a probability distribution $p(a_1, \ldots, a_m)$. We restrict ourselves to the case when the distribution $p(a)$ is normal and may be fully characterized by setting the mean values $\bar{a}_1, \ldots, \bar{a}_m$ and the error matrix $\sigma^2_{kk'}$, $k, k' = 1, \ldots, m$. Then the distribution of the parameters $a_1, \ldots, a_r$ we take interest in, will be characterized by the corresponding submatrix $\sigma^2_{jj'}$, $j, j' = 1, \ldots, r$ which is obtained by cutting out from the matrix $\sigma^2_{kk'}$ the lines and columns referring to the ballast parameters $a_{r+1}, \ldots, a_m$. Let us express the available amount of information about the parameters $a_1, \ldots, a_r$ in terms of the matrix $\sigma^2_{jj'}$.

By definition, the amount of the information q in the "communication" $p(a_1, \ldots, a_r)$ is given by

$$q = \int p(a_1, \ldots, a_r) \log p(a_1, \ldots, a_r)\, da_1 \ldots da_r + c, \tag{1.2}$$

in which the normalization additive constant c and the base of the logarithm should be fixed. For a future analysis the most convenient normalization is

$$q = 2 \int p \ln p \, da_1 \ldots da_r + r(1 + \ln 2\pi), \tag{1.3}$$

so, for one normally distributed parameter with the variance $\sigma^2 = \frac{1}{w}$, we get

$$q = \ln w. \tag{1.4}$$

For the set of r quantities with the error matrix $\sigma^2_{jj'}$, expression (1.3) yields

$$q(a_1, \ldots, a_r) = -\ln |\sigma^2|, \tag{1.5}$$

where $|\sigma^2|$ stands for the determinant of the matrix $\sigma^2_{jj'}$.

In the following, when denoting by $y_\xi(a)$, a set of all functions which can be measured in the given experiment we shall mean by ξ (as well as by the "point" ξ) a set of all numbers characterizing the measurement uniquely, including those specifying a choice, if any, of experimental equipment. a will include all the quantities A and ballast parameters B which can be encounted in the dependencies y.

Let us make a certain additional measurement ξ. Then, according to the method of least squares, the specified values of the parameters a are determined by the condition

$$M = w_\xi [y_\xi(a) - y_\xi]^2 + \sum_{k,k'=1}^{m} (a_k - \bar{a}_k)\, z_{kk'}\, (a_{k'} - \bar{a}_{k'}) = \min \tag{1.6}$$

where $w_\xi = \frac{1}{\sigma_\xi^2}$ is the weight of the additional measurement equal to the inverse variance of the result of the measurement y_ξ , and $z_{kk'}$ is the matrix inverse to $\sigma^2_{kk'}$.

From (1.6), we get by the usual routine that the errors in the specified values of the parameters will be estimated by the matrix $(z + \Delta z)^{-1}$ inverse to the matrix $(z + \Delta z)$, the latter being equal to

$$z + \Delta z = z_{kk'} + w_\xi \frac{\partial y_\xi(a)}{\partial a_k} \frac{\partial y_\xi(a)}{\partial a_{k'}} . \qquad (1.7)$$

If one assumes that the values $\bar{a}$ had been obtained experimentally, i.e., by measuring certain dependences $y_\xi(a)$, then

$$z_{kk'} = \sum_\xi \frac{\partial y_\xi(a)}{\partial a_k} \frac{\partial y_\xi(a)}{\partial a_{k'}} w_\xi \qquad (1.8)$$

hence

$$\frac{\partial z_{kk'}}{\partial w_\xi} = \frac{\partial y_\xi(a)}{\partial a_k} \frac{\partial y_\xi(a)}{\partial a_{k'}} . \qquad (1.9)$$

Expression (1.7) is equivalent to (1.9) in the linear case. But when the dependence $y_\xi(a)$ is not quite linear over a equality (1.9) should be preferred .

Departing from (1.9), it is possible to find the rate of the accumulation of the information when the measurement ξ increases the weight w_ξ

$$q'_{w_\xi} = \frac{\partial q(A)}{\partial w_\xi} . \qquad (1.10)$$

However, if one starts just from formulae (1.5) and (1.9), then for each measurement we are interested in, the calculations should be performed which are as cumbersome as those of the complete analysis of the problem of m parameters by the method of the least squares .

A relative advantage of the measurements ξ depends not only upon the speed with which the information as a function of the weight w is accumulated, but also upon the efficiency of the measurement ξ. The efficiency oi the measurement $\lambda(\xi)$ is convenient to put equal to the increment of the weight of the measurement gained at a price of efforts Λ equal to unity*. Evidently, the efficiency of the measurement of some curve is zero outside the experimentally accessible interval. The function $\lambda(\xi)$ may be predicted according to the analysis of the experimental conditions or established experimentally if the measurements have been already started.

Evidently, the most advantageous measurement is the one for which the speed of accumulating the information (about the parameters Λ) as a function of labour consuming

$$V(\xi) = \frac{\partial\, q\,(A)}{\partial\, w_{\xi}}\;\lambda\,(\xi) \tag{1.11}$$

will be maximum. To provide for the condition

$$V\,(\xi) = max \tag{1.12}$$

it is only necessary to find a simple way of calculating the function $q\,\acute{w}_{\xi}$. In the next section a theorem is set forth which establishes the relationship between the rate of accumulating the information and the variance of the function $y_{\xi}(a)$, solving,thereby, the problem of finding the most advantageous measurement.

2. The Basic Theorem of Planning

Let there be given the error matrix z^{-1}_{kk}, of the parameters $a = a_1, \ldots, a_m$. We call the parameters $A = a_1, \ldots, a_r$ main group ; the remaining parameters $B = a_{r+1}, \ldots, a_n$ make up an additional group.

In the matrix z^{-1} we single out the submatrix R^{-1} which is referred only to the parameters of the main group , and the submatrix T^{-1} which is referred only to the additional group of the parameters

---

*The efficiency may be determined formally by the equality $\lambda(\xi) = \frac{dw_{\xi}}{d\Lambda(\xi)}$ since the proportionality tacitly admitted in a verbal formulation, does not matter. The relation $W = c\,\Lambda$ is important only for the statical planning. For that reason it is stressed in the book/1/when the difficulty function $h(x)$ is introduced. This difficulty function is connected with $\lambda\,(x)$ by the relation $\lambda\,(x) = \frac{1}{h^2(x)}$.

$$z^{-1} = \left(\begin{array}{c|c} R^{-1} & S^{-1} \\ \hline \tilde{S}^{-1} & T^{-1} \end{array} \right) \qquad (2.1)$$

(the rectangular matrices S^{-1} and $\tilde{S}^{-1}$ show the relationship between the parameters of different groups).

Let among the parameters a there be no linearly-dependent ones, so exists the matrix z inverse to the matrix z^{-1}. We divide the matrix z into submatrices in the same manner as it was done with the matrix z^{-1}:

$$z = \left(\begin{array}{c|c} N & P \\ \hline \bar{P} & Q \end{array} \right) . \qquad (2.2)$$

We call the matrix Q^{-1} inverse to the matrix Q the error matrix of the parameters of the additional group under the condition that parameters of thé main group A are fixed.

Let us call the quantity $q(A)$ given by the formula

$$q(A) = -\ln \det R^{-1} = -\ln |R^{-1}| \qquad (2.3)$$

the information concerning the group of the parameters **A** .

Let $y_\xi(a)$ be an arbitrary function of the index ξ (the variable ξ may be both descrete and continuous) and a linear function of the parameters a

$$y_\xi(a) = c(\xi) + \sum_{k=1}^{m} \phi_k(\xi)\, a_k . \qquad (2.4)$$

Let as a result of the measurement ξ, the elements of the matrix z acquire the increments

$$\Delta z_{kk'} = w \frac{\partial z_{kk'}}{\partial w_\xi} = w\, \phi_k(\xi)\, \phi_{k'}(\xi) \qquad (2.5)$$

where $w = \frac{1}{\sigma_y^2}$ is the weight of the measurement equal to the inverse variance of the result of the measurement σ_y.

Then the theorem takes place : the rate of the accumulation of the information about the given group A of the parameters a in the course of the measurement of the function $y_\xi(a)$ is equal to a decrease of the variance of the function $y_\xi(a)$ in fixing * this group of the parameters.

According to the notations introduced above, the theorem states that

$$\frac{\partial q(a_1,\dots,a_r)}{\partial w_\xi} = \sigma_m^2(\xi) - \sigma_{m-r}^2(\xi) \equiv$$

(2.6)

$$\equiv \sum_{k,k'=1}^{m} \phi_k(\xi)\, z_{kk'}^{-1}\, \phi_{k'}(\xi) - \sum_{k,k'=r+1}^{m} \phi_k(\xi)\, Q_{kk'}^{-1}\, \phi_{k'}(\xi) \quad .$$

The meaning of the theorem becomes more obvious in case when there are no ballast parameters and all m parameters enter the group A.Then

$$\sigma_{m-m}^2(x) = 0,$$

and

$$\frac{\partial q}{\partial w_x} = \sigma_m^2(x) \quad ,$$

that is, the derivative of information is equal to the square of the corridor of errors $\sigma(x)$.

Proof.

Let every submatrix into which matrices z and z^{-1} are split, and its inverse one, if any, be put in one-to-one correspondence with the matrix of the m-th order , having zeros on the place of the missing elements. For example,

$$p = \left(\begin{array}{c|c} 0 & P \\ \hline 0 & 0 \end{array}\right) \quad . \qquad (2.7)$$

*When the dependence $y_\xi(a)$ is linear by the parameters a only approximately, the parameters should be fixed equal to their mean values which they had before the fixing according to previous measurements .

Such matrices added by zeros will be denoted by small letters. Evidently

$$z = n + p + \bar{p} + q, \qquad z^{-1} = r^{-1} + s^{-1} + \bar{s}^{-1} + t^{-1} . \tag{2.8}$$

Then, the statement of the theorem assumes the form .

$$\frac{\partial q(A)}{\partial w_\xi} = - \frac{\partial \ln |R^{-1}|}{\partial w_\xi} = \phi_p z^{-1}_{pp} \phi_{p'} - \phi_p q^{-1}_{pp'} \phi_{p'} \tag{2.9}$$

(the summation $\sum_{1}^{m}$ over the dumb, repeating indices is implied here and in the foregoing).

We begin to prove (2.9) by varying the information. Making use of the well-known relation

$$\partial | X | = | X | \; X^{-1}_{kk'} \partial X_{k'k} , \tag{2.10}$$

we get that

$$- \frac{\partial \ln |R^{-1}|}{\partial w} = - R_{kk'} \frac{\partial R^{-1}_{k'k}}{\partial w} . \tag{2.11}$$

The expression (2.11) contains the derivatives of the matrix elements R^{-1} which can be found by varying the equality

$$z_{pk'} z^{-1}_{k'k} = \delta_{pk} . \tag{2.12}$$

Indeed, by varying (2.12), we have

$$(z_{pk'} + \partial z_{pk'}) (z^{-1}_{k'k} + \partial z^{-1}_{k'k}) = \delta_{pk}, \tag{2.13}$$

hence

$$\frac{\partial z^{-1}_{k'k}}{\partial w} = - z^{-1}_{k'p'} \frac{\partial z_{p'p}}{\partial w} z^{-1}_{pk} . \tag{2.14}$$

Substituting (2.14) into (2.11) and remembering, that by the definition

$$\frac{\partial z_{p'p}}{\partial w} = \phi_{p'} \phi_p , \tag{2.15}$$

we are led to an explicit expression of the derivative of information in terms of the matrices introduced above

$$\frac{\partial q(A)}{\partial w} = \phi_p z^{-1}_{pk} r_{kk'} z^{-1}_{k'p'} \phi_{p'} \quad . \tag{2.16}$$

Since ϕ_p are arbitrary functions of ξ, then to prove (2.9) it is necessary to prove the matrix equality

$$z^{-1}_{pk} r_{kk'} z^{-1}_{k'p'} = z^{-1}_{pp'} - q^{-1}_{pp'} = \left(\begin{array}{c|c} R^{-1} & S^{-1} \\ \hline \bar{S}^{-1} & T^{-1} - Q^{-1} \end{array}\right) \quad . \tag{2.17}$$

A direct calculation of the left-hand side of equality (2.17) yields

$$z^{-1}_{pk} r_{kk'} z^{-1}_{k'p'} = \left(\begin{array}{c|c} R^{-1} & S^{-1} \\ \hline \bar{S}^{-1} & \bar{S}^{-1} R S^{-1} \end{array}\right) \quad . \tag{2.18}$$

On the other hand, it follows from formulae (2.7), (2.8), and (2.12) that

$$\bar{p} r^{-1} + q \bar{s}^{-1} = 0 , \tag{2.19}$$

$$\bar{p} s^{-1} + q t^{-1} = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & \begin{matrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{matrix} \end{array}\right) \quad . \tag{2.20}$$

By multiplying (2.20) from the left hand side by the matrix q^{-1}, we have

$$- q^{-1} \bar{p} s^{-1} = t^{-1} - q^{-1}. \tag{2.21}$$

By multiplying, further, (2.19) from the left by the matrix q^{-1} and from the right by the matrix r, we obtain

$$q^{-1} \bar{p} + \bar{s}^{-1} r = 0, \tag{2.22}$$

whence

$$\bar{s}^{-1} r s^{-1} = t^{-1} - q^{-1}. \tag{2.23}$$

Thus, equality (2.17) is correct. The theorem is proved.

Higher Derivatives of Information

It easy to show by a direct calculation, that

$$\frac{\partial \sigma^2(\xi_1)}{\partial w_{\xi_2}} = \sigma^2(\xi_1, \xi_2)\, \sigma^2(\xi_2, \xi_1) \tag{2.24}$$

where

$$\sigma^2(\xi_1, \xi_2) = \phi_k\ (\xi_1)\ z^{-1}_{kk'} \phi_{k'}(\xi_2). \tag{2.25}$$

The formulae for the higher derivatives of information follow immediately e.g.,

$$\frac{\partial^2 q\,(A)}{\partial w_{\xi_1} \partial w_{\xi_2}} = [\sigma_m^2\,(\xi_1, \xi_2)]^2 - [\sigma^2_{m-r}\,(\xi_1, \xi_2)]^2 \tag{2.26}$$

Corollary 1.

The speed of the accumulation of information is a quantity invariant with respect, at least, to the linear substitution of the parameters inside the main and additional (ballast) groups

$$a_i \to b_i = \sum_{j=1}^{r} c_{ij}\, a_j; \qquad 1 < i < r;$$

$$a_i \to b_i = \sum_{j=r+1}^{m} c_{ij}\, a_j; \qquad r+1 < i < m; \tag{2.27}$$

Indeed, one can become sure by an elementary check, that both the variance $\sigma^2_m(\xi)$ and the variance $\sigma^2_{m-r}(\xi)$ are invariant with respect to the substitution (2.27). Therefore, the quantity $\frac{\partial q}{\partial w}$, being their difference, remains also invariant to such a substitution.

Corollary 1 shows that the increment of the accuracy of every linear combination of the parameters A is of the same importance for the growth of the information about A.

Corollary 1 proves to be useful in planning the measurements of the quantities which do not enter $y(a)$ explicitly. Because of this, they cannot be fixed simply by cutting out from the matrix z. For instance, we take interest in the value of the curve $y_\xi(a)$ at a certain point ξ_0, i.e., in the quantity $b = y_{\xi_0}(a)$. In practice it is possible to fix the parameter b (what is necessary for planning) by adding (with the help of (1.7)) at the point ξ_0 "the measurement" $y_0 = y_{\xi_0}(\bar{a})$ with the weight "heavy" enough and to use the variance

$$\sigma^2_m(\xi)\,\Big|_{y_{\xi_0}(a) = y_0} \tag{2.28}$$

instead of $\sigma^2_{m-1}(\xi)$ in the calculation of the function q'_{w_ξ}

$$\frac{\partial q}{\partial w_\xi} = \sigma^2_m(\xi) - \sigma^2_m(\xi)\Big|_{y_{\xi_0}(a) \simeq y_0} . \tag{2.29}$$

The invariance of the quantity q'_{w_ξ} provides for the validity and unambiguity of similar methods.

Corollary 2.

The quantity q'_{w_ξ} is restricted by the inequality

$$q'_{w_\xi} < \frac{1}{w_\xi} \tag{2.30}$$

which follows immediately from the meaning of the quantity $\sigma^2_m(\xi)$.

Corollary 3.

The quantity q'_{w_ξ} satisfies the inequality

$$\sum_\xi q'_{w_\xi} \cdot w_\xi < r , \tag{2.31}$$

where w_ξ are the weights of the measurements $y_\xi(a)$.

The inequality (2.31) turns into the equality

$$\sum_\xi q'_{w_\xi} \cdot w_\xi = r , \tag{2.32}$$

if all the information q is obtained by measuring y_ξ, since in this case the identity

$$\sum_\xi \sigma^2_m(\xi)\ w_\xi = m , \tag{2.33}$$

is valid what can be verified by a direct calculation.

Distribution of Information among Measurements

In the analysis of an experiment it is often important to know how much did every measurement help to reduce the errors of the parameters $\boldsymbol{A}$. According to (2.32), the rate of the accumulation of the information satisfies the relation

$$\sum_\xi q'_{w_\xi} \frac{w_\xi}{r} = 1 . \tag{2.34}$$

This relation shows that the separate terms $q'_{w_\xi} \frac{w_\xi}{r}$ may, in a certain sense, characterize the relative amount of the information (about the parameters A) obtained from each measurement ξ.

Strictly speaking, the information given by some measurement ξ_0 depends on the order in which the measurements ξ were made . In the analysis of a completed experiment the actual order of measure - ments is of no importance. If we assume that all the measurements were made simultaneously and the weights were increasing, being always proportional to the weights w_ξ finally obtained, we get just the quantity $q'_{w_{\xi_0}} \frac{w_{\xi_0}}{r}$ for the relative amount of information given by the measurement ξ_0.

3. Drift and Jumps of the Most Advantageous Measurement in the Course of an Experiment

It is always important to know in practice what measurement becomes most advantageous

$$v(\xi) = max$$

in future since the measuring apparatus must be adjusted before hand for this change . Let us first investigate qualitatively probable movement of ξ, assuming ξ being a continuous variable.

By formula (1.11)

$$v(\xi) = q'_{w_\xi} \lambda(\xi)$$

two co-factors enter the function $v(\xi)$. As for the function of the efficiency $\lambda(\xi)$, its changes have a character of unpredicted corrections, so that we shall consider it constant in time. On the contrary, the function q'_{w_ξ} is steadily decreasing after each measurement, what may cause large displacements of the most advantageous measurement point ξ. Really, from (2.30) one can obtain the inequality

$$v(\xi) < \frac{\lambda(\xi)}{w_\xi} \tag{2.1}$$

whence it follows that as the measurements are being made at the point ξ where $v(\xi)$ reaches the maximum, the magnitude of this maximum is rapidly falling. The maximum itself as well as the most advantageous measurement point ξ may drift aside .

Let us plot time t *vs* most advantageous point ξ where we are measuring at a moment t (Fig.1). Let the point 0 show the moment when the planning had been started after some preliminary (unplanned) measurements and the point ξ_0, $v(\xi_0) = max$ had been found.

Let each measurement require a unit time, and the next most advantageous measurement point $\xi_1, \xi_2, \ldots$ is calculated anew before each following measurement (in Fig . 1 at the moments 1,2 ...).

It should be expected that the best point ξ will be displacing only insignificantly until it happens at a certain moment (moment 8 in Fig. 1) that the maximum of the function $v(\xi)$ (maximum n.1) corresponding to this point would became lower than one of the earlier minor maxima (maximum n.2) of this function, and the most advantageous measurement point would jump there (to the point ξ_8 in Fig. 1.).

Since the measurements decrease most strongly just the maxima at which they are being made, then the most advantageous measurement point will further jump for a while between maxima n. 1 and n. 2. Then there will appear a third maximum (moment 19 in Fig.1) and so on , until their number reaches the total number of the parameters m (in Fig. 1, the case $m = 3$ is shown).

When the number of maxima becomes equal to that of the parameters, there will be no new maxima of the function $v(\xi)$, and the point ξ will begin, obeying a certain law, to "visit" all the m maxima, providing, thereby, some distribution of efforts between them. The function $v(\xi)$ will be getting smaller at the same rate for all ξ , so the maxima will cease to move and remain near some points $\xi_\infty(1),\ldots,\xi_\infty(m)$. Evidently, for the function $v(\xi)$ to remain further unchanged in shape (what is most advantageous from the point of view of the speed with which the information is being accumulated), it is necessary to spend the efforts for new measurements at the points $\xi_\infty(1),\ldots,\xi_\infty(m)$ proportional to those already spent in their vicinities. As for the calculation of the next most advantageous measurement point, it can be stopped.

If the function $\lambda(\xi)$ is not strongly dependent upon the measurements made, or this dependence is known, it enables one to calculate the movement of the most advantageous measurement point beforehand, since the second co-factor q'_{w_ξ} entering $v(\xi)$ depends only upon the weights w_ξ and not upon the results of the observations *.

Indeed, having found ξ_0 , we can add conditionally the weight $w_{\xi_0} = \lambda(\xi_0)$ at the this point , then to find the point ξ_1 , to add again conditionally the weight $w(\xi_1) = \lambda(\xi_1)$ and so forth, imitating, thereby, the whole process of a continuous planning of an experiment. When all the m maxima become apparent, the calculation can be stopped and the graph like that shown in Fig. 1 can be drawn. This graph will become now the plan of an experiment. If the function $\lambda(\xi)$ changes strongly in the course of the experiment, a further plan must be calculated anew.

4. Conclusion

Let us illustrate by a real example what advantage in the accuracy may be expected due to the planning of an experiment. In paper /2/, the angular distributions of neutron-proton scattering at different energies were used for the determination of meson-nucleon coupling constant. It is most convenient to be concerned with the determination of the constant f^2 from the experiment at 380-400 MeV.

---

* If the function $y_\xi(a)$ depends upon the parameters a not quite linearly, then the function q'_{w_ξ} will depend to a certain extent upon the results of the measurements y_ξ. But this, however, can be usually neglected.

As can be seen from the figure given in paper /2/ the efforts spent for measuring $\frac{d\sigma}{d\Omega}$ in the angular interval 0° - 180° were distributed approximately uniformly. The calculation of the quantities $q'_w(\Theta_i)\ w_i$ has shown that in such distribution of the efforts 7 extreme points at angles Θ close to 180° yielded 90 % of the information about f^2, whereas the rest 29 points added only 10%. If the experiment under discussion is considered as the one aimed at the determination of the constant f^2 only, then it would become possible, if the planning is applied, to determine the constant f^2 2-3 times more accurately without additional time and money consumption. A similar specification of the constant f^2 with the previous distribution of efforts would require a 4-9-fold increase of time spent by an experimental physicist during his work with the accelerator.

In universal (for example, preliminary) experiments where almost all of the parameters a entering the dependence $y_{\xi}(a)$ are of immediate interest (enter the group A), the application of planning will give much more modest advantage in accuracy. On the contrary, in experiments for which the ratio $\frac{m}{r}$ is great (for the experiment described above $\frac{m}{r}$ is equal to 10), one can expect, if the planning of experiments is applied, an essential specification of the results.

The author takes the opportunity to express his gratitude to Prof. J.A.Smorodinsky for fruitful discussions of the work and to Dr. N.P.Klepikov for critical remarks.

Received by Publishing Department
on July 20, 1960.

References

1. Н.П.Клепиков, С.Н.Соколов. Анализ экспериментальных данных методом максимума правдоподобия, гл.У, Preprint ОИЯИ, Р-235 /1958/.
2. N.S.Amaglobeli, Yu.M.Kazarinov, S.N.Sokolov., I.N.Silin. Determination of the coupling constant of pion-nucleon interaction by differential cross section for elastic (np)-scattering at 90, 380-400, 630 MEV.

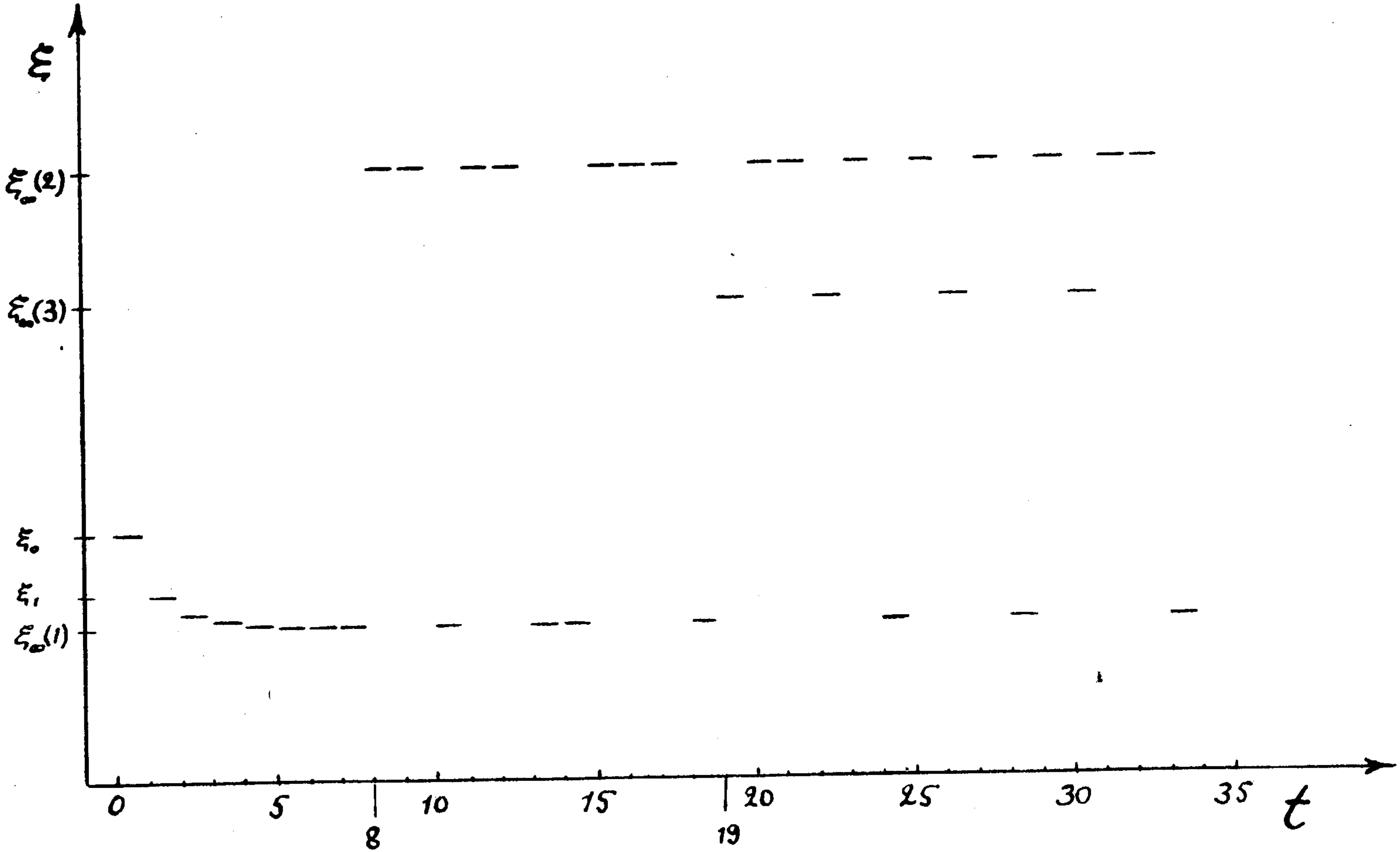

Fig. 1