ОБЪЕДИНЕННЫЙ ИНСТИТУТ
ЯДЕРНЫХ ИССЛЕДОВАНИЙ

С.Н. Соколов

Д-573

# ИЗМЕРЕНИЕ, ДАЮЩЕЕ НАИБОЛЬШУЮ ИНФОРМАЦИЮ, И ЕГО ОПРЕДЕЛЕНИЕ ПРИ НЕПРЕРЫВНОМ ПЛАНИРОВАНИИ ЭКСПЕРИМЕНТА

Дубна 1960 год

## Аннотация

На базе математической статистики полностью решается основная задача непрерывного планирования физических экспериментов: в процессе выполнения эксперимента и в зависимости от информации, которая уже накоплена, среди всех измерений, которые можно сделать в следующий момент, выбирается такое, которое дает о данной группе величин наибольшую информацию при наименьшей затрате усилий /времени, средств и т.п./. Развивается качественная картина перемещения наиболее выгодного измерения и обсуждается один пример.

S. N. Sokolov

Д - 573

С.Н.Соколов

# ИЗМЕРЕНИЕ, ДАЮЩЕЕ НАИБОЛЬШУЮ ИНФОРМАЦИЮ, И ЕГО ОПРЕДЕЛЕНИЕ ПРИ НЕПРЕРЫВНОМ ПЛАНИРОВАНИИ ЭКСПЕРИМЕНТА

## Введение

Планирование эксперимента преследует цель получения наибольшей информации о заданной группе величин при наименьшей затрате усилий. Под усилиями могут подразумеваться время, средства и другие нежелательные факторы и их комбинации, которые можно измерить численно и которые следует свести к минимуму. Так как извлечение информации из экспериментальных данных делается с помощью математической статистики, то естественно, что статистические проблемы возникают при планировании любого эксперимента.

Упоминание о том, что использование математической статистики необходимо для рациональной постановки экспериментов, можно встретить еще у создателей современной теории статистической обработки экспериментов, в частности у Р.А.Фишера, выдвинувшего принцип максимума правдоподобия. Фишер называет планирование экспериментов основной обратной задачей статистики. Однако он и его последователи развили исключительно теорию планирования прямых экспериментов, проводимых в плохо контролируемых условиях, то есть промышленных, торговых, сельско-хозяйственных и других подобных экспериментов. В то же время физика и смежные с ней науки имеют дело почти исключительно с косвенными измерениями, когда большое число разных измерений объединяется и из их совокупности вычисляются значения нескольких величин, имеющих прямой теоретический интерес. В этом случае на пути между непосредственными результатами наблюдений и значениями измеряемых величин обычно лежит метод наименьших квадратов. Планирование косвенных экспериментов почти не развивалось, и в позднейшей литературе можно найти лишь очень частные результаты, имеющие весьма косвенное отношение к реальным задачам планирования физических экспериментов.

Первой попыткой, имеющей дело с реальной экспериментальной ситуацией, явилось решение Н.П.Клепиковым совместно с автором настоящей статьи задачи о планировании измерений в случае, когда требуется экстраполяция или интерполяция измеренной кривой в области, недоступные или труднодоступные для непосредственного наблюдения. Результаты изложены в книге /1/, в которую включена глава, посвященная планированию экспериментов. Там разбирается упомянутая выше задача и некоторые ее обобщения, вводится понятие функции трудности $h$ и обсуждаются многие связанные с планированием вопросы.

Решение задачи на экстраполяцию показало, что действительно существуют наивыгоднейшие точки, в которых следует проводить измерения, и наиболее рациональное распределение суммарного времени измерения между этими точками. Выяснилось также, что сравнительно небольшое смещение точек с их наилучших мест или нарушение рационального распределения времен резко снижает точность экстраполяции, что подтверждает практическую актуальность планирования. Существенно, что рецепт выбора вариантов экспериментального оборудования и мест измерений, а также вычисление достижимой точности экстраполяции удалось представить в наглядной графической форме.

Основным недостатком предлагаемого в книге /1/ планирования является то, что задача рассматривается как статическая, то есть весь эксперимент планируется сразу, причем предполагается, что заранее известна трудность $h$ измерения отдельных участков кривой и никаких ранее выполненных "внеплановых" измерений этой кривой нет. Между тем, все необходимые для планирования сведения накапливаются и уточняются в процессе эксперимента постепенно, и крайне важно уметь находить наивыгоднейшие места измерений одновременно с выполнением самих измерений, то есть производить непрерывное планирование. Действительно, эксперимент никогда не рождается целиком, и после некоторых нащупывающих измерений и при учете уже накопленной наукой информации возникает вопрос, как сделать следующий шаг, то есть какое измерение выбрать следующим, чтобы наискорейшим путем двигаться к цели - к определению теоретически важных параметров с достаточной точностью. Выбор наилучшего шага с учетом уже полученных результатов можно назвать также динамическим планированием.

При непрерывном планировании ни наилучшие места измерений, ни распределение времен не совпадают с теми, которые дает статическое планирование /хотя при бесконечном продолжении эксперимента и стремятся к ним/, так как уже выполненные измерения изменяют важность отдельных участков.

В дальнейшем мы предположим, что результаты наблюдений распределены нормально и связаны линейно с определяемыми параметрами $a$. На практике получаемые ниже формулы будут верны всегда, когда только применим метод наименьших квадратов. Это означает, что связь непосредственных результатов наблюдений с измеряемыми параметрами может быть и нелинейной и что распределения вероятностей результатов наблюдений могут отклоняться от распределения Гаусса, но эта нелинейность и "негауссовость" должны быть в разумных пределах.

Фактически для приложений существенно только, чтобы распределение параметров $\alpha$ в окрестности двух стандартных отклонений было похоже на распределение Гаусса и чтобы матрица ошибок сохраняла свой обычный смысл.

## § 1. Постановка задачи

Когда на эксперименте в некоторых точках $x_i$ измеряется некоторая кривая $y(\alpha, x)$, то обычно не все параметры $\alpha$, от которых зависит эта кривая, представляют для нас непосредственный интерес. Можно привести немало примеров, когда ради уточнения одной-двух теоретически важных величин ставился эксперимент по измерению кривой, зависящей от 10-15 неизвестных параметров. Обычно это происходит из-за того, что либо интересующие нас параметры не поддаются прямому измерению, либо их прямое измерение очень трудно, так что нам не остается ничего лучшего, как измерять их совместно с не интересующими нас балластными параметрами, входящими в зависимость $y(\alpha, x)$. При этом возникает естественное желание, прежде чем делать очередное измерение, найти такую точку $x$, при измерении в которой возможно бо́льшая часть времени и средств тратилась бы на уточнение интересующих нас величин, и возможно меньшая - на уточнение тех балластных параметров, которые в силу необходимости измеряются с ними совместно.

Непрерывное планирование эксперимента ставит целью отыскание среди всех измерений, которые можно сделать в следующий момент, такого измерения, которое даст наибольшую дополнительную информацию о заданной группе величин при наименьшей затрате усилий. Основная трудность планирования экспериментов лежит в том, что непосредственные результаты наблюдений с увеличением желаемой информации связаны довольно сложным образом.

Интуитивно ясно, что наиболее выгодными для измерения будут те точки $x$, в которых зависимость кривой $y(\alpha, x)$ от интересующих нас величин выражена наиболее сильно, и, наоборот, там, где ход кривой $y(\alpha, x)$ определяется, в основном, балластными параметрами, результаты измерений почти ничего не скажут о тех величинах, ради измерения которых делается эксперимент. Однако для планирования, как показывает практика, интуитивные соображения дают недостаточно точные указания, так что нам потребуется в первую очередь знать количественно, сколько информации и о каких величинах дает измерение данной зависимости $y(\alpha, x)$ в каждой точке $x$

Мы будем далее рассматривать планирование только таких экспериментов, целью которых является уточнение заданной группы величин $A=\{\alpha_1,\dots,\alpha_r\}$. Естественно, что уточнять величины можно только тогда, когда они уже грубо известны, "нащупаны", то есть когда часть опытов по их измерению уже как-то сделана. Тогда можно уже пытаться дать обоснованный ответ на вопрос, как выбрать следующее измерение, чтобы при наименьшей затрате усилий получить как можно больше дополнительной информации о величинах $\alpha_1,\dots,\alpha_r$.

Дополнительная информация о величинах $A$ может быть, вообще говоря, получена измерением любой функции $y(C)$ /известных с ошибками/ параметров $C$, если только параметры $C$ как-то связаны с величинами А – например, ранее измерялись с ними совместно или являются их функциями. В последнем случае, не проигрывая в общности, можно считать, что часть из параметров $C$ совпадает с некоторыми из величин А. Остальные параметры /обозначим их через $B$ / будем и далее называть балластными, подчеркивая этим названием, что их уточнение не является целью данного эксперимента.[x/] Совокупность всех входящих в рассмотрение величин – т.е. величин А и балластных параметров $B$ – будем обозначать через $\alpha=\{\alpha_1,\dots,\alpha_m\}$.

Примем, что сведения, которыми мы располагаем перед началом эксперимента о параметрах $\alpha$, заключается в том, что нам дана оценка их совместного распределения вероятности р $/\alpha_1,\dots,\alpha_m/$. Ограничимся случаем, когда распределение $p(\alpha)$ нормально и, соответственно, может быть исчерпывающим образом охарактеризовано заданием средних значений $\bar{\alpha}_1,\dots,\bar{\alpha}_m$ и матрицей ошибок $\sigma^2_{kk'}$, $k,k'=1,\dots,m$. Тогда распределение интересующей нас группы параметров $\alpha_1,\dots,\alpha_r$ будет характеризоваться соответствующей подматрицей $\sigma^2_{jj'}$, $j,j'=1,\dots,r$, которая получится вычеркиванием из матрицы $\sigma^2_{kk'}$ строк и столбцов, относящихся к балластным параметрам $\alpha_{r+1},\dots,\alpha_m$. Выразим количество информации о параметрах $\alpha_1,\dots,\alpha_r$, которым мы располагаем, через матрицу $\sigma^2_{jj'}$.

---

[x/] На практике может оказаться, что полезно измерять $y(C)$, даже если все параметры $C$ – балластные.

По определению, количество информации $q$ в "сообщении" $p(a_1, \dots, a_r)$ дается формулой

$$q = \int p(a_1, \dots, a_r) \log p(a_1, \dots, a_r)\, da_1 \cdots da_r + C, \qquad /1.2/$$

в которой следует фиксировать нормировочную аддитивную постоянную $C$ и основание логарифма. Наиболее удобной для дальнейшего будет нормировка

$$q = 2 \int p \ln p \, da_1 \cdots da_r + r(1 + \ln 2\pi), \qquad /1.3/$$

при которой для одного нормально распределенного параметра с дисперсией $\sigma^2 = \frac{1}{W}$ получаем

$$q = \ln W. \qquad /1.4/$$

Для совокупности $r$ величин с матрицей ошибок $\sigma^2_{jj'}$ вычисления по формуле /1.3/ дают

$$q(a_1, \dots, a_r) = -\ln |\sigma^2|, \qquad /1.5/$$

где $|\sigma^2|$ – определитель матрицы $\sigma^2_{jj'}$.

Прежде чем переходить к дальнейшему, надо уточнить понятие измерения. На практике нередко случается, что одни и те же величины можно измерять разными, не равноценными по своей трудоемкости, способами. Например, имеются два взаимоисключающих варианта экспериментального оборудования – № 1 и № 2, – один из которых лучше измеряет "левый" конец некоторой зависимости $y(x)$, а другой –"правый" конец той же зависимости, и между этими вариантами нужно выбрать. Поэтому в дальнейшем, обозначая множество всех функций, доступных для измерения на данном эксперименте, через $y(\alpha, \xi)$, мы будем под $\xi$ /а также под "точкой" $\xi$ / подразумевать совокупность всех чисел, однозначно характеризующих измерение, в том числе, указывающих вариант оборудования, а под $\alpha$ – совокупность величин $A$ и всех балластных параметров $B$, которые только могут встретиться в измеряемой зависимости.

Сделаем некоторое дополнительное измерение $\xi$. Тогда согласно методу наименьших квадратов, уточненные значения параметров $\alpha$ определятся из условия

$$M = w\left[y(\alpha, \xi) - y\right]^2 + \sum_{k,k'=1}^{m} (\alpha_k - \bar{\alpha}_k) Z_{kk'} (\alpha_{k'} - \bar{\alpha}_{k'}) = \min, \qquad /1.6/$$

где $w = \frac{1}{\sigma^2}$ – вес дополнительного измерения, равный обратной дисперсии результата измерения $y$, и где $Z_{kk'}$ – матрица, обратная к матрице $\sigma^2_{kk'}$.

Из /1.6/ обычной техникой получаем, что ошибки уточненных значений параметров будут оцениваться матрицей $(Z+\Delta Z)^{-1}$, обратной к матрице $Z+\Delta Z$, равной

$$Z+\Delta Z = Z_{kk'} + w\,\frac{\partial y(a,\xi)}{\partial a_k}\,\frac{\partial y(a,\xi)}{\partial a_{k'}}\,. \qquad /1.7/$$

Если считать, что значения $\bar{a}$ были в свою очередь получены экспериментально, то есть измерением некоторых зависимостей $y(a,\xi)$, то

$$Z_{kk'} = \int \frac{\partial y(a,\xi)}{\partial a_k}\,\frac{\partial y(a,\xi)}{\partial a_{k'}}\, w(\xi)\, d\xi\,, \qquad /1.8/$$

откуда

$$\frac{\delta Z_{kk'}}{\delta w(\xi)} = \frac{\partial y(a,\xi)}{\partial a_k}\,\frac{\partial y(a,\xi)}{\partial a_{k'}}\,. \qquad /1.9/$$

Выражение /1.7/ эквивалентно выражению /1.9/ в линейном случае, но, когда зависимость $y(a,\xi)$ не вполне линейна по $a$, равенству /1.9/ следует отдавать предпочтение.

Отправляясь от выражения /1.9/, можно уже найти скорость накопления информации $q$ при выполнении измерения $\xi$

$$q'_w(\xi) = \frac{\delta q(A)}{\delta w(\xi)}\,. \qquad /1.10/$$

Однако, если исходить непосредственно из формул /1.5/ и /1.9/, то для каждого интересующего нас измерения $\xi$ придется провести вычисления, примерно равные по объему полному анализу методом наименьших квадратов задачи на $m$ параметров, что в целом, для прослеживания всей кривой $q'_w(\xi)$, составит довольно громоздкий расчет.

Относительная выгодность тех или иных измерений $\xi$ зависит не только от скорости накопления информации как функции веса $w$, но и от того насколько измерение $\xi$ эффективно. Эффективность измерений $\lambda(\xi)$ удобнее всего определить как приращение веса измерения $w$, если на это измерение затратить усилия $\Lambda$, равные единице[x]. Очевидно, эффективность $\lambda(\xi)$ равна нулю вне интервала, доступного для измерений, и велика там, где измерять удобно. Функция $\lambda(x)$ может быть предсказана на основании анализа условий эксперимента или установлена опытным путем, если измерения уже начаты.

Очевидно, наиболее выгодным измерением является то, для которого скорость накопления информации /о параметрах $A$ / как функция затрачиваемых усилий

$$v(\xi) = \frac{\delta q(A)}{\delta w(\xi)} \lambda(\xi) \qquad /1.11/$$

будет максимальной. Чтобы обеспечить условие

$$V(\xi) = max\,, \qquad /1.12/$$

надо только найти простой способ вычисления функции $q'_w(\xi)$. В следующем параграфе излагается теорема, которая устанавливает связь скорости накопления информации с дисперсией кривой $y(a,\xi)$ и тем самым дает практическое решение задачи отыскания наивыгоднейшего измерения.

## § 2. Основная теорема планирования

Пусть дана матрица ошибок $Z^{-1}_{kk'}$ параметров $a = a_1, \ldots, a_m$. Параметры $A = a_1, \ldots, a_r$, которые будут нас интересовать, мы назовем основной группой: оставшиеся параметры $B = a_{r+1}, \ldots, a_m$ составят дополнительную группу.

---

x/ Эффективность можно определить формально равенством $\lambda(\xi) = \dfrac{\partial w(\xi)}{\partial \Lambda(\xi)}$ так как для дальнейшего пропорциональная связь веса и усилий $w = c\Lambda$, неявно предполагаемая в данной в тексте словесной формулировке, несущественна. Соотношение $w = c\Lambda$ важно только для статического планирования, поэтому оно подчеркивается в книге /1/ при введении функции трудности $h(x)$, связанной с $\lambda(x)$ соотношением $\lambda(x) = \dfrac{1}{h^2(x)}$.

Выделим в матрице $Z^{-1}$ подматрицу $R^{-1}$, которая относится только к параметрам основной группы, и подматрицу $T^{-1}$, которая относится только к дополнительной группе параметров:

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

/прямоугольные матрицы $S^{-1}$ и $\bar{S}^{-1}$ дают связь параметров разных групп/.

Пусть среди параметров $a$ нет линейно-зависимых, и существует матрица $Z$, обратная к матрице $Z^{-1}$. Разобьем матрицу $Z$ на подматрицы таким же образом, как это было сделано с матрицей $Z^{-1}$:

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

Назовем матрицу $Q^{-1}$, обратную к матрице $Q$, матрицей ошибок параметров дополнительной группы при условии, что параметры основной группы $A$ фиксированы.

Назовем величину $q(A)$, заданную формулой

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

информацией относительно группы параметров $A$.

Пусть $y(a,\xi)$ является произвольной функцией переменной $\xi$ /переменная $\xi$ может быть как дискретной, так и непрерывной/ и линейной функцией параметров $a$:

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

Пусть в результате измерения функции $y(a,\xi)$ в точке $\xi$ элементы матрицы $Z$ получают приращения

$$\Delta Z_{kk'} = w \frac{\delta Z_{kk'}}{\delta w(\xi)} \equiv w\, \varphi_k(\xi)\, \varphi_{k'}(\xi), \qquad /2.5/$$

где $W = \frac{1}{\sigma^2}$ – вес проведенного измерения, равный обратной дисперсии результата измерения.

Тогда имеет место теорема: скорость накопления информации о данной группе $A$ параметров $a$ при проведении измерений функции $y(a,\xi)$ /в точке $\xi$ /равна уменьшению /в этой точке $\xi$ / дисперсии функции $y(a,\xi)$ при фиксации[x/] этой группы параметров.

Во введенных выше обозначениях утверждение теоремы имеет вид

$$\frac{\delta q(a_1,\dots,a_r)}{\delta\, w(\xi)} = \sigma^2_m(\xi) - \sigma^2_{m-r}(\xi) \equiv$$

$$\equiv \sum_{k,k'=1}^{m} \varphi_k(\xi)\, Z^{-1}_{kk'}\, \varphi_{k'}(\xi) - \sum_{k,k'=r}^{m} \varphi_k(\xi)\, Q^{-1}_{kk'}\, \varphi_{k'}(\xi). \qquad /2.6/$$

Содержание теоремы становится более наглядным в частном случае, когда нет балластных параметров и все $m$ параметров входят в группу $A$. Тогда

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

и

$$\frac{\delta q}{\delta w(x)} = \sigma^2_m(x),$$

то есть производная информации равна квадрату коридора ошибок $\sigma(x)$.

Доказательство.

Каждой из подматриц, на которые разбиты матрицы $Z$ и $Z^{-1}$, и тем обратным матрицам, которые имеются к этим подматрицам, поставим в соответствие матрицу порядка $m$, поставив на место недостающих элементов нули, например,

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

---

x/ В случаях, когда зависимость $y(a,\xi)$ линейна по параметрам $a$ только приближенно, параметры следует фиксировать равными их средним значениям, которые они имели до фиксации согласно прежним измерениям.

Такие, дополненные нулями, матрицы будем обозначать малыми буквами. Очевидно, что

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

Тогда утверждение теоремы примет вид

$$\frac{\delta q(A)}{\delta w(\xi)} = -\frac{\delta \ln|R^{-1}|}{\delta w(x)} = \varphi_p Z^{-1}_{pq}\varphi_q - \varphi_p q^{-1}_{pq}\varphi_q \qquad /2.9/$$

/суммирование $\sum\limits_{1}^{m}$ по дважды встречающимся индексам здесь и в дальнейшем подразумевается/.

Доказательство формулы /2.9/ начнем с варьирования информации. Воспользовавшись известным соотношением

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

мы получим, что

$$-\frac{\delta \ln|R^{-1}|}{\delta w} = -R_{kk'}\frac{\delta R^{-1}_{k'k}}{\delta w}\,. \qquad /2.11/$$

В /2.11/ входят вариационные производные элементов матрицы $R^{-1}$, которые можно найти, варьируя тождество

$$Z_{pk'}\, Z^{-1}_{k'k} = \delta_{pk}\,. \qquad /2.12/$$

Действительно, варьируя /2.12/, имеем

$$\left(Z_{pk'} + \delta Z_{pk'}\right)\left(Z^{-1}_{k'k} + \delta Z^{-1}_{k'k}\right) = \delta_{pk}, \qquad /2.13/$$

откуда

$$\frac{\delta Z^{-1}_{k'k}}{\delta w} = -Z^{-1}_{k'q}\frac{\partial Z_{qp}}{\delta w} Z^{-1}_{pk}\,. \qquad /2.14/$$

Подставляя /2.14/ в /2.11/ и вспоминая, что, по определению,

$$\frac{\delta Z_{qp}}{\delta w} = \varphi_q \varphi_p \ , \qquad /2.15/$$

приходим к явному выражению вариационной производной информации через введенные выше матрицы

$$\frac{\delta q}{\delta w} = \varphi_p Z^{-1}_{pk} \tau_{kk'} Z^{-1}_{k'q} \varphi_q \ . \qquad /2.16/$$

Так как $\varphi_p$ – произвольные функции, то для справедливости формулы /2.9/ нужно доказать матричное равенство

$$Z^{-1}_{pk} \tau_{kk'} Z^{-1}_{k'q} = Z^{-1}_{pq} - q^{-1}_{pq} = \left(\begin{array}{c|c} R^{-1} & S^{-1} \\ \hline \bar{S}^{-1} & T^{-1} - Q^{-1} \end{array}\right) . \qquad /2.17/$$

Непосредственное вычисление левой части равенства /2.17/ дает

$$Z^{-1}_{pk} \tau_{kk'} Z^{-1}_{k'q} = \left(\begin{array}{c|c} R^{-1} & S^{-1} \\ \hline \bar{S}^{-1} & \bar{S}^{-1} R S^{-1} \end{array}\right) . \qquad /2.18/$$

С другой стороны, из формул /2.7/ , /2.8/ и /2.12/ следует

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

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

Домножая /2.20/ слева на $q^{-1}$, имеем

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

Домножая, далее, 2.19 слева на $q^{-1}$ и справа на $r$, получим

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

откуда

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

Таким образом, равенство /2.17/ справедливо, и теорема доказана.

Высшие вариационные производные информации

Прямым вычислением легко показать, что

$$\frac{\delta\, \sigma^2(\xi_1)}{\delta w(\xi_2)} = \sigma^2(\xi_1, \xi_2)\, \sigma^2(\xi_2, \xi_1) \,, \qquad /2.24/$$

где

$$\sigma^2(\xi_1, \xi_2) = \varphi_k(\xi_1)\, Z^{-1}_{kk'}\, \varphi_{k'}(\xi_2) \,. \qquad /2.25/$$

Отсюда непосредственно следуют формулы для высших вариационных производных информации, например:

$$\frac{\delta^2 q}{\delta w(\xi_1)\, \delta w(\xi_2)} = \left[\sigma^2_m(\xi_1, \xi_2)\right]^2 - \left[\sigma^2_{m-r}(\xi_1, \xi_2)\right]^2. \qquad /2.26/$$

Следствие 1.

Скорость накопления информации является величиной, инвариантной по крайней мере к линейному переопределению параметров внутри основной и дополнительной /балластной/ групп в отдельности

$$\begin{cases} a_1, \dots, a_r \to b_1(a_1, \dots, a_r), \dots, b_r(a_1, \dots, a_r) \\ a_{r+1}, \dots, a_m \to b_{r+1}(a_{r+1}, \dots, a_m), \dots, b_m(a_{r+1}, \dots, a_m) \end{cases} \qquad /2.27/$$

Действительно, элементарной проверкой можно убедиться, что как дисперсия $\sigma^2_m(\xi)$, так и дисперсия $\sigma^2_{m-2}(\xi)$ инвариантны к замене /2.27/. Поэтому величина $\frac{\delta q}{\delta w}$, будучи их разностью, также остается инвариантной при такой замене.

Следствие 1 оказывается полезным при планировании измерения величин зависимость кривой $y(a,\xi)$ от которых не найдена в явном виде, благодаря чему их нельзя фиксировать простым вычеркиванием из матрицы $Z$. Например, пусть нас интересует значение кривой $y(a,\xi)$ в некоторой точке $\xi_0$, то есть величина $b = y(a,\xi_0)$. Практически нужное для планирования фиксирование параметра $b$ можно осуществить, добавив условно в точке $\xi_0$ "измерение" $y_0 = y(\bar{a},\xi_0)$ с достаточно большим весом, найдя с помощью формулы /1.7/ дисперсию $\sigma^2_m(\xi)$ при учете этого "измерения"

$$\sigma^2_m(\xi)\Big|_{y(a,\xi_0)\simeq y_0} \qquad /2.28/$$

и использовав функцию /2.28/ в качестве $\sigma^2_{m-1}(\xi)$ при подсчете функции $q'_w(\xi)$

$$\frac{\delta q}{\delta w(\xi)} = \sigma^2_m(\xi) - \sigma^2_m(\xi)\Big|_{y(a,\xi_0)\simeq y_0} \qquad /2.29/$$

Инвариантность величины $q'_w(\xi)$ гарантирует нам законность и однозначность подобных приемов.

<u>Следствие 2</u>.

Величина $q'_w(\xi)$ в точках, где имеются измерения, ограничена неравенством

$$q'_w(\xi) \leqslant \frac{1}{w(\xi)}, \qquad /2.30/$$

которое непосредственно вытекает из смысла величины $\sigma^2_m(\xi)$.

<u>Следствие 3</u>

Величина $q'_w(\xi)$ удовлетворяет неравенству

$$\sum_i q'_w(\xi_i)\,w(\xi_i) \leqslant \tau, \qquad /2.31/$$

где $W(\xi_i)$ - веса выполненных измерений кривой $y(a,\xi)$.

Неравенство /2.31/ превращается в равенство

$$\sum_i q'_w(\xi_i)\,W(\xi_i) = r\,, \qquad /2.32/$$

если вся информация $q$ получена измерением только кривой $y(a,\xi)$, так как в этом случае справедливо тождество

$$\sum_i \sigma^2_m(\xi_i)\,W(\xi_i) = m\,, \qquad /2.33/$$

которое можно проверить непосредственным вычислением.

## § 3. Перемещение наивыгоднейшей точки по мере выполнения эксперимента

На практике всегда бывает важно знать, куда переместится наивыгоднейшее измерение $\xi$

$$V(\xi) = max$$

в дальнейшем, так как измерительная аппаратура должна быть заранее приспособлена для соответствующей перестройки. Нарисуем сначала качественную картину перемещения $\xi$ со временем.

Согласно формуле /1.11/

$$V(\xi) = q'_w(\xi)\,\lambda(\xi)$$

в функцию $V(\xi)$ входят два сомножителя. Что касается функции эффективности $\lambda(\xi)$, то ее изменения обычно носят характер уточнений, не поддающихся предсказанию заранее, так что мы будем считать ее постоянной во времени. Напротив, функция $q'_w(\xi)$ неуклонно уменьшается после каждого измерения, что может вызвать большие перемещения наивыгоднейшей точки. Действительно, из неравенства /2.30/ вытекает неравенство

$$V(\xi) \leqslant \frac{\lambda(\xi)}{W(\xi)}\,, \qquad /3.1/$$

из которого с очевидностью следует, что по мере выполнения измерений в точке, где $V(\xi)$ достигает максимума, величина этого максимума быстро падает. Сам максимум и вместе с ним наивыгоднейшая точка $\xi$ при этом могут несколько смещаться в сторону.

Изобразим перемещение наивыгоднейшей точки $\xi$ графически, отложив по оси абсцисс время $t$, а по оси ординат - точку $\xi$, в которой производятся измерения и где в данный момент накапливается вес /рис. 1/. Пусть точка $O$ изображает момент, когда после некоторых предварительных /внеплановых/ измерений было начато планирование и найдена точка $\xi_0$, $V(\xi_0) = max$.

Пусть каждое измерение требует единичного времени и перед началом каждого следующего измерения /на рис. 1 в моменты 1,2.../ очередная наивыгоднейшая точка $\xi_1, \xi_2, \ldots$ подсчитывается заново.

Следует ожидать, что некоторое время наилучшая точка $\xi$ будет только незначительно перемещаться, пока в некоторый момент /момент 8 на рис. 1/ не обнаружится, что соответствующий этой точке максимум функции $V(\xi)$ /максимум № 1/ стал ниже одного из ранее второстепенных максимумов /максимума № 2/ этой функции, и наивыгоднейшая точка перескочила туда /в точку $\xi_8$ на рис. 1/.

Так как измерения сильнее всего уменьшают именно те максимумы, в которых они производятся, то наивыгоднейшая точка будет в дальнейшем некоторое время перескакивать между максимумами № 1 и № 2. Затем выявится третий максимум /момент 19 на рис. 1/ и так далее, пока их число не достигнет полного числа параметров $m$ / на рис. 1 изображен случай $m = 3$/.

Когда число максимумов сравняется с числом параметров, новых максимумов у функции $V(\xi)$ появляться уже не будет, и точка $\xi$ начнет по некоторому закону "посещать" все $m$ максимумов, обеспечивая таким способом некоторое распределение усилий между ними. Функция $V(\xi)$ станет при этом уменьшаться оставаясь подобной самой себе, так что практически прекратится и перемещение максимумов, которые будут в дальнейшем оставаться вблизи некоторых точек $\xi_\infty(1), \ldots, \xi_\infty(m)$ . Очевидно, для того, чтобы функция $V(\xi)$ оставалась в дальнейшем подобной самой себе /что наиболее выгодно с точки зрения скорости накопления информации/, на новые измерения в точках $\xi_\infty(1), \ldots, \xi_\infty(m)$ надо затрачивать усилия пропорционально уже затраченным в их окрестностях, и в дальнейшем расчет очередной наивыгоднейшей точки можно прекратить.

Если функция $\lambda(\xi)$ не сильно зависит от производимых измерений, или эта зависимость известна, всю картину перемещения наивыгоднейшей точки можно рассчитать заранее, воспользовавшись тем, что функция $q'_w(\xi)$ зависит только от весов $w(\xi_i)$ и не зависит от самих результатов наблюдений $y_i$ x/.

Действительно, найдя $\xi_0$, мы можем добавить условно в этой точке вес $w(\xi_0)=\lambda(\xi_0)$, найти точку $\xi_1$, добавить в ней опять-таки условно вес $w(\xi_1)=\lambda(\xi_1)$, и так далее, имитируя таким образом весь процесс непрерывного планирования эксперимента. Когда выявятся все $m$ максимумов, расчет можно закончить и нарисовать график, подобный рис. 1, который теперь станет планом эксперимента. В случае, если во время эксперимента фукция $\lambda(\xi)$ неожиданно сильно изменится, дальнейший план, естественно, придется рассчитать заново.

## § 4. Заключение

Посмотрим на реальном примере, какой приблизительно выигрыш в точности можно ожидать от применения планирования эксперимента. В работе /2/ из угловых распределений нейтрон-протонного рассеяния при различных энергиях искалась константа $\pi$-мезон-нуклонной связи $f^2$. В качестве примера удобнее всего остановиться на определении константы $f^2$ из опыта при энергии 380-400 Мэв.

Согласно следствию 3, для скорости накопления информации относительно величины $f^2$ существует соотношение

$$\sum_i q'_w(\theta_i)w_i = 1, \qquad /4.1/$$

---

x/ Если функция $y(\alpha,\xi)$ зависит от параметров $\alpha$ не вполне линейно, то функция $q'_w(\xi)$ будет в некоторой степени зависеть от результатов измерений $y_i$, но этим обычно можно пренебречь.

которое показывает, что отдельные слагаемые $q'_w(\theta_i)w_i$ могут в некотором смысле характеризовать процентный вклад $i$ - го измерения в информацию[x/] о величине $f^2$.

Как это видно из приводимого в работе /2/ рисунка, усилия по измерению $\frac{d\sigma}{d\Omega}$ в интервале углов $0^\circ - 180^\circ$ были распределены приблизительно равномерно. Подсчет величин $q'_w(\theta_i)w_i$ показал, что при этом распределении усилий 7 крайних точек при углах $\theta$ близких к $180^\circ$ внесли 90% информации, а на остальные 29 точек пришлось только 10% всей информации. Если обсуждаемый опыт рассматривать как опыт по определению только константы $f^2$, то, применив к нему планирование, без дополнительных затрат времени и средств константу $f^2$ можно было бы из этого эксперимента определить в 2-3 раза точнее. Такое же уточнение константы $f^2$ при прежнем распределении усилий потребовало бы 4-х - 9-и кратного увеличения времени работы на ускорителе.

В универсальных /нащупывающих/ экспериментах, где почти все входящие в зависимость $y(\alpha, \xi)$ параметры $\alpha$ представляют непосредственный интерес /входят в группу $A$ /, применение планирования даст, очевидно, значительно более скромный выигрыш в точности. Наоборот, в специализированных экспериментах, для которых отношение $\frac{m}{\tau}$ велико /для опыта, приведенного выше $\frac{m}{\tau}$ равняется 10/, можно ожидать от применения планирования существенного уточнения результатов.

---

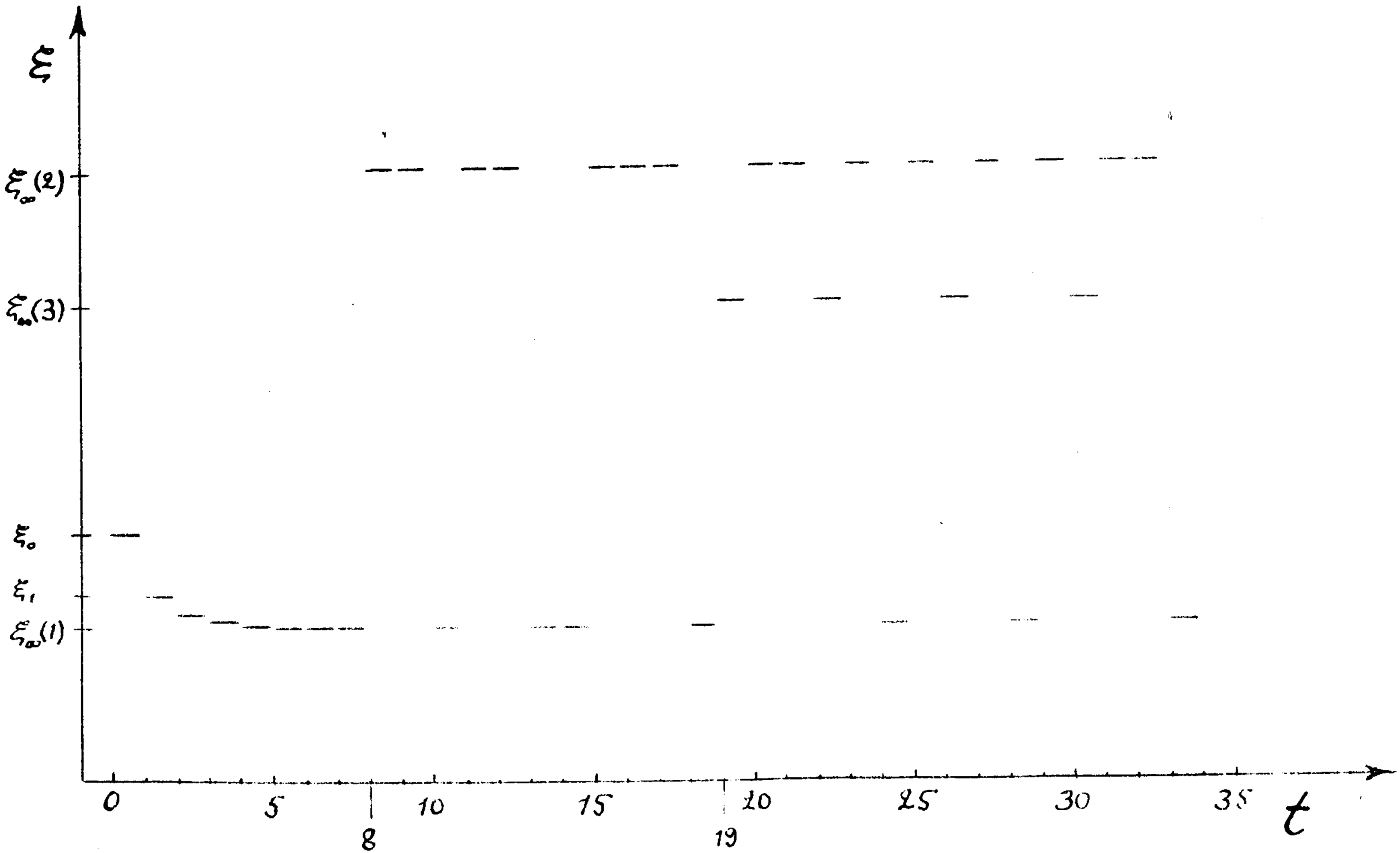
x/ Строго говоря, ответить на вопрос: "Сколько информации вносят такие-то измерения?"- однозначно нельзя, так как информация - нелинейная функция весов. Однако можно, например, подсчитать, сколько было бы потеряно информации, если веса таких-то измерений уменьшить в 10 раз, по сравнению с потерей информации при пропорциональном уменьшении весов всех измерений в те же 10 раз. Величины $q'_w W$ соответствуют бесконечно малому пропорциональному уменьшению весов.

## Литература


1. Н.П.Клепиков, С.Н.Соколов. Анализ экспериментальных данных методом максимума правдоподобия, гл. У. Препринт ОИЯИ, Р-235 /1958/.

2. Н.С.Амаглобели, Ю.М.Казаринов, С.Н.Соколов, И.Н.Силин. Определение константы $f^2$ из данных по дифференциальным сечениям упругого $n-p$ - рассеяния при энергиях 90, 380-400 и 630 Мэв. Препринт ОИЯИ, Д-5[illegible]5 /1960/.


Рукопись поступила в издательский отдел
20 июля 1960 года.



$\xi$
$\xi_\infty(2)$
$\xi_\infty(3)$
$\xi_0$
$\xi_1$
$\xi_\infty(1)$
0
5
8
10
15
19
20
25
30
35
$t$


Рис. 1.