

Основы математического планирования эксперимента

Историческая справка

До середины XVIII века вопросами организации эксперимента целиком занимались **экспериментаторы**.

Уделом математиков была обработка уже проведенного эксперимента.

Постепенно стало ясно, что речь должна идти не только об обработке экспериментальных данных, а об оптимальной процедуре математико-статистического анализа.

Такие процедуры и были разработаны усилиями многих математиков.

Основные этапы становления планирования эксперимента:

- метод наименьших квадратов – (А.Лежандр, К.Гаусс, конец 18-начало 19 века);
- основы регрессионного и корреляционного анализа (Ф. Гальтон, К.Пирсон, конец 19 - начало 20 века);
- концепция малых выборок (Госсет, более известный под псевдонимом «Стьюдент», начало 20 века);
- основы математического планирования эксперимента (Р.Фишер, середина 20 века);
- разработка последовательной стратегии экспериментирования, шаговая стратегия экспериментирования (Бокс и Уилсон)

Причем получается **определенная сбалансированность между стремлением к минимизации числа опытов и уровнем точности и надежности полученных результатов.**

Хорошо **спланированный эксперимент обеспечивает оптимальную обработку результатов**, и, следовательно, возможность четких статистических выводов.

Однако, в основе статистических методов обработки данных (дисперсионный и регрессионный анализ) лежат определенные **предпосылки о свойствах законов распределения случайных величин, их независимости, однородности дисперсий** и т.д., что в реальных задачах выполняется далеко не всегда.

Совокупность таких предпосылок принято называть **моделью ситуации.**

Возникает вопрос: зачем оптимально планировать эксперимент, если нет уверенности в том, выполняются ли предпосылки принятой модели ситуации?

В конце 70-х годов 20 века центр тяжести переместился на проблему принятия **решения при выборе модели ситуации и обработке данных.**

Так возникло новое направление, известное под названием **анализа данных.**

Здесь можно выделить такие основные этапы, как

- проверка выполнимости предпосылок модели ситуации;
- использование априорной информации (байесовские методы);
- применение устойчивых (робастных) процедур в случае нарушения тех или иных предпосылок или невозможности их проверки.

Все это стимулирует в последнее время развитие робастных и непараметрических методов анализа.

Таким образом, **экспериментатор должен наилучшим образом выбрать модель ситуации, план эксперимента и метод обработки.**

Основные понятия и определения

Под экспериментом будем понимать совокупность операций совершаемых над объектом исследования с целью получения информации об его свойствах

Важнейшей задачей методов обработки полученной в ходе эксперимента информации является **задача построения математической модели изучаемого явления, процесса, объекта.**

Ее можно использовать и при анализе процессов и при проектировании объектов.

Можно получить хорошо аппроксимирующую математическую модель, если целенаправленно применяется активный эксперимент.

Другой **задачей обработки полученной в ходе эксперимента информации является задача оптимизации**, т.е. нахождения такой комбинации влияющих независимых переменных, при которой выбранный показатель оптимальности принимает экстремальное значение.

Опыт – это отдельная экспериментальная часть.

План эксперимента – совокупность данных, определяющих число, условия и порядок проведения опытов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям, совокупность действий направленных на разработку стратегии экспериментирования (от получения априорной информации до получения работоспособной математической модели или определения оптимальных условий).

Это целенаправленное управление экспериментом, реализуемое в условиях неполного знания механизма изучаемого явления.

В процессе измерений, последующей обработки данных, а также формализации результатов в виде математической модели, **возникают погрешности и теряется часть информации, содержащейся в исходных данных.**

Применение методов планирования эксперимента позволяет определить погрешность математической модели и судить о ее адекватности.

Если точность модели оказывается недостаточной, то применение методов **планирования эксперимента позволяет модернизировать математическую модель с проведением дополнительных опытов без потери предыдущей информации и с минимальными затратами.**

Цель планирования эксперимента – нахождение таких условий и правил проведения опытов, при которых удастся получить надежную и достоверную информацию об объекте с наименьшей затратой труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности.

Пусть интересующее нас свойство (**Y**) объекта зависит от нескольких (**n**) независимых переменных (**X₁, X₂, ..., X_n**) и мы хотим выяснить характер этой зависимости - **Y=F(X₁, X₂, ..., X_n)**, о которой мы имеем лишь общее представление.

Величина **Y** – отклик, а сама зависимость **Y=F(X₁, X₂, ..., X_n)** – функция отклика.

Независимые переменные **X₁, X₂, ..., X_n** – факторы.

Диапазоны изменения факторов задают область определения **Y**.

Если принять, что каждому фактору соответствует координатная ось, то полученное **пространство называется факторным пространством.**

При **n=2** область определения **Y** представляется собой прямоугольник, при **n=3** – куб, при **n > 3** - гиперкуб.

При выборе диапазонов изменения факторов нужно учитывать их совместимость, т.е. контролировать, чтобы в этих диапазонах любые сочетания факторов были бы реализуемы в опытах и не приводили бы к абсурду.

Для каждого из факторов указывают граничные значения

$$X_{i \min} \leq X_i \leq X_{i \max}, i = 1, \dots, n.$$

Регрессионный анализ функции отклика предназначен для получения ее математической модели в виде уравнения регрессии

$$Y = F(X_1, X_2, \dots, X_n; B_1, B_2, \dots, B_m) + e$$

B_1, B_2, \dots, B_m – некоторые коэффициенты;

e – погрешность.

Среди основных методов планирования, применяемых на разных этапах исследования, используют:

- планирование **отсеивающего эксперимента**, основное значение которого выделение из всей совокупности факторов группы существенных факторов, подлежащих дальнейшему детальному изучению;
- планирование **эксперимента для дисперсионного анализа**, т.е. составление планов для объектов с качественными факторами;
- планирование *регрессионного эксперимента*, позволяющего получать регрессионные модели (полиномиальные и иные);
- планирование **экстремального эксперимента**, в котором главная задача — экспериментальная оптимизация объекта исследования;
- планирование при **изучении динамических процессов** и т.д.

Инициатором применения планирования эксперимента является Рональд А. Фишер, другой автор известных первых работ – Френк Йетс.

Далее идеи планирования эксперимента формировались в трудах Дж. Бокса, Дж. Кифера, Г.К. Круга, Е.В. Маркова и др.

В настоящее время методы планирования эксперимента заложены в специализированных пакетах программных продуктов, например: **StatGrapfics, Statistica, SPSS, SYSTAT** и др.

Представление результатов экспериментов

При использовании методов планирования эксперимента необходимо найти ответы на 4 вопроса:

1. • Какие сочетания факторов и сколько таких сочетаний необходимо взять для определения функции отклика?
2. • Как найти коэффициенты B_0, B_1, \dots, B_m ?
3. • Как оценить точность представления функции отклика?
4. • Как использовать полученное представление для поиска оптимальных значений Y ?

Геометрическое представление функции отклика в факторном пространстве X_1, X_2, \dots, X_n (рис.1).

При трех и более факторах задача становится неразрешимой. Если и будут найдены решения, то использовать совокупность выражений достаточно трудно, а часто и не реально

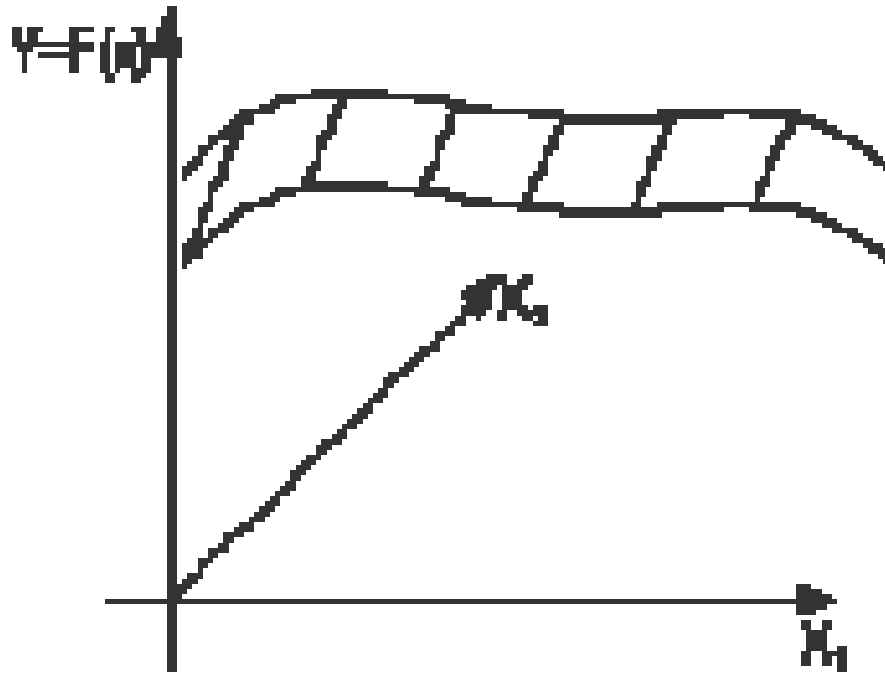


Рис. 1. Поверхность отклика

Например, пусть необходимо исследовать влияние движущейся жидкости и внутреннего диаметра на потери давления и расход жидкости проектировании гидравлической сети (рис. 2).



Если в диапазоне изменения каждого фактора взять хотя бы по шесть точек для того чтобы выполнить опыты при всех возможных сочетаниях значений факторов (их три) необходимо выполнить опытов и сформировать по $6^2=36$ кривых для каждой из двух функций отклика.

Если мы хотим хотя бы продублировать опыты чтобы снизить погрешность, то число опытов пропорционально возрастает, поэтому произвольное выполнение опытов при числе факторов более двух и использование их результатов – практически нереально.

v	м/с	0,5	0,6	0,7	0,8	0,9	1,0
ρ	кг/м ³	800	900	1000	1100	1200	1300
d	м	0,05	0,075	0,1	0,125	0,15	0,175

Разложение функции отклика в степенной ряд, кодирование факторов

Если заранее не известно аналитическое выражение функции отклика, то можно рассматривать не саму функцию, а ее разложение, например, в степенной ряд в виде полинома

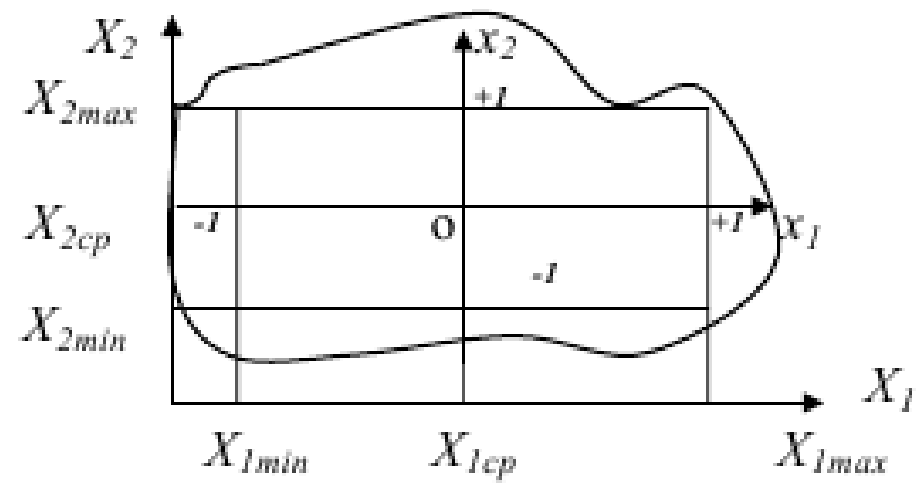
$$Y=B_0 + B_1X_1 + \dots + B_nX_n + B_{12}X_1X_2 + \dots B_{nn-1}X_nX_{n-1} + B_{11}X_{12}+ \dots + B_{nn}X_{n2} + \dots$$

Разложение в степенной ряд функции возможно в том случае, если сама функция является непрерывной. На практике обычно ограничиваются числом членов степенного ряда и аппроксимируют функцию полиномом некоторой степени.

Факторы могут иметь разные размерности (А, В, Вт, об/мин) и резко отличаться количественно.

В теории планирования эксперимента используют кодирование факторов.

Рис. 3. Пространство кодированных факторов



Эта операция заключается в выборе нового масштаба для кодированных факторов (рис. 3), причем такого, чтобы минимальное значение кодированных факторов соответствовало “-1”, а максимальное значение “+1”, а также в переносе начала координат в точку с координатами $X_{1cp}, X_{2cp}, \dots, X_{ncp}$

$$X_{icp} = \frac{x_{i\min} + x_{i\max}}{2}.$$

Текущее значение кодированного фактора

$$x_i = \frac{x_i - x_{icp}}{x_{icp} - x_{i \min}} = \frac{x_i - x_{icp}}{x_{i \max} - x_{icp}} = \frac{2x_i - x_{i \max} - x_{i \min}}{x_{i \max} - x_{i \min}} (4).$$

где X_i – именованное (абсолютное) значение фактора; x_i – кодированное значение фактора;

$X_{icp} - X_{i \min} = X_{i \max} - X_{icp}$ – интервал варьирования фактора.

Граница совместимости факторов указана на рис. 3 в виде кривой линии.

Если фактор изменяется дискретно, например он является качественным, то каждому уровню этого кодированного фактора присваиваются числа в диапазоне от +1 до -1.

Так при двух уровнях это +1 и -1, при трех уровнях +1, 0, -1 и т.д.

Функция отклика может быть выражена через кодированные факторы $Y=f(x_1,..., x_n)$ и записана в полиномиальном виде

$$Y=b_0+b_1x_1+b_2x_2+...+b_nx_n+b_{12}x_1x_2+...+b_{nn-1}x_{n-1}x_n+b_{11}x_{12}+...+b_{nn}x_{n2}+.... \tag{4}$$

Очевидно, что $B_i \neq b_i$

но

$$Y=F(X_i,..., X_b,..., X_n)= f(x_i,... x_i,..., X_n)$$

Для полинома, записанного в кодированных факторах, степень влияния факторов или их сочетаний на функцию отклика определяется величиной их коэффициента **b_i**.
 Для полинома в именованных факторах величина коэффициента **B_i** еще не говорит однозначно о степени влияния этого фактора или их сочетаний на функцию отклика.

Задача определения коэффициентов уравнения регрессии.

Для определения $m+1$ коэффициента полинома необходимо не менее $m+1$ уравнений (опытов).

Полученные коэффициенты B позволяют сформировать уравнение функции отклика при $m+1$ членах уравнения.

Если **точность** этого уравнения оказалась недостаточной, то требуется взять уравнение с большим числом членов и начать все заново так как все коэффициенты B оказываются зависимыми друг от друга.

Это возникает при использовании пассивного эксперимента. Однако если целенаправленно использовать активный эксперимент и особым образом построить матрицу сочетаний факторов в опытах X , использовать планирование эксперимента, то коэффициенты полинома определяются независимо друг от друга.

Стратегия применения планов заключается в принципе постепенного планирования – **постепенного усложнения модели.**

Начинают с простейшей модели, находят для нее коэффициенты, определяется ее точность. Если точность не удовлетворяет, то планирование и модель постепенно усложняются.