S. Akhnazarova
V. Kafarov

EXPERIMENT OPTIMIZATION IN CHEMISTRY AND CHEMICAL ENGINEERING

Translated from Russian
by
Vladimir M. Matskovsky
and
Alexander P. Repyev

MIR PUBLISHERS
MOSCOW

S. Akhnazarova
V. Kafarov

EXPERIMENT OPTIMIZATION IN CHEMISTRY AND CHEMICAL ENGINEERING

Translated from Russian
by
Vladimir M. Matskovsky
and
Alexander P. Repyev

1 9 9 2 1

MIR PUBLISHERS
MOSCOW

First published 1982
Revised from the 1978 Russian edition

The Greek Alphabet

$A\ \alpha$	Alpha	$I\ \iota$	Iota	$P\ \varrho$	Rho
$B\ \beta$	Beta	$K\ \varkappa$	Kappa	$\Sigma\ \sigma$	Sigma
$\Gamma\ \gamma$	Gamma	$\Lambda\ \lambda$	Lambda	$T\ \tau$	Tau
$\Delta\ \delta$	Delta	$M\ \mu$	Mu	$\Upsilon\ \upsilon$	Upsilon
$E\ \varepsilon$	Epsilon	$N\ \nu$	Nu	$\Phi\ \varphi$	Phi
$Z\ \zeta$	Zeta	$\Xi\ \xi$	Xi	$X\ \chi$	Chi
$H\ \eta$	Eta	$O\ o$	Omicron	$\Psi\ \psi$	Psi
$\Theta\ \theta$	Theta	$\Pi\ \pi$	Pi	$\Omega\ \omega$	Omega

На английском языке

Printed in Hungarian People's Republic

EXPERIMENT OPTIMIZATION IN CHEMISTRY AND CHEMICAL ENGINEERING

С. Л. Ахназарова,
В. В. Кафаров

ОПТИМИЗАЦИЯ ЭКСПЕРИМЕНТА В ХИМИИ И ХИМИЧЕСКОЙ ТЕХНОЛОГИИ

ИЗДАТЕЛЬСТВО
«ВЫСШАЯ ШКОЛА»
МОСКВА

CONTENTS

INTRODUCTION

Experiment has always been a major tool in tackling practical problems and testing theoretical hypotheses in chemistry and chemical engineering. Traditional experimentation, however, involves a good deal of effort and time, especially where complex processes are involved. A very efficient way to enhance the value of research and to cut down the process development time is through designed experiment, that is, by optimizing experiment at every stage from inception, through research and development, to engineering and production.

At present, process analysis, design, optimization and performance prediction in chemical engineering are above all based on mathematical modelling [1]. If the designer has at his disposal complete information (thermodynamic, kinetic and hydrodynamic data) about the process or plant of interest, he can develop a deterministic mathematical model as a set of ordinary or partial differential equations. In order to determine the coefficients appearing in the equations and also to verify the model, he runs an experiment.

If the available information is not complete, the designer undertakes a functional study of the process or plant; to this end, he runs an experiment and observes the input and output process or plant variables.

In Fig. I.1, the measurable or *controllable* input variables are designated $x_1, \ldots, x_k$; *uncontrollable* or *random* process variables (noise) are designated $w_1, \ldots, w_l$; and the output variables (responses) are designated $y_1, \ldots, y_m$.

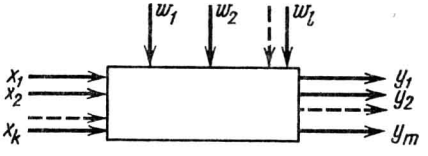

Fig. I. 1. Multivariate process with multiple inputs

The term "random" applies to variables that can be accounted for with difficulty, if at all. This is true, for example, of the fall in the activity of a catalyst, variations in the state of a heat transfer surface, variations in the ambient temperature, etc. The input variables $x_1, \ldots, x_k$ form the basis (or basic) set of variables, because they control the experiment. Of course,

this division of process variables into random and basic is arbitrary. The designer may class as random any variable not included in the basis, even though it may be known well. Depending on the objective sought or the experimental capabilities available, some measurable variables may be included in process (or plant) noise. This will of course impair the accuracy of the mathematical model. An output variable may be any technical or economic index of the process or plant. Applying regression or correlation analysis to experimental data, the designer can establish relationships between the various variables and determine the conditions of an optimum.

A typical mathematical model is

$$y = \varphi(x_1, x_2, \ldots, x_k) \tag{I.1}$$

where y is the dependent variable, or the *response*, and x_i's are the independent variables, or *factors*. They occupy what is known as the *factor space*, and the graphical representation of the response function is called the *response surface*.

In process analysis by statistical methods, the mathematical model is most frequently a polynomial—a truncated Taylor series into which the response function, Eq. (I.1), is expanded:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{\substack{u,j=1 \\ u \neq j}}^{k} \beta_{uj} x_u x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \ldots \tag{I.2}$$

where

$$\beta_j = \partial\varphi/\partial x_j \qquad \beta_{uj} = \partial^2\varphi/\partial x_u\,\partial x_j \qquad \beta_{jj} = \partial^2\varphi/2\,\partial x_j^2$$

In any real process, there always are uncontrollable and unmeasurable variables, and the response y usually varies at random. This is why the evaluation of experimental data yields *sample regression coefficients* b_0, b_j, b_{uj}, and b_{jj}, which are estimates of the theoretical coefficients β_0, β_j, β_{uj}, and β_{jj}. The estimated regression equation developed on the basis of an experiment will then take the form

$$y = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{\substack{u,j=1 \\ u \neq j}}^{k} b_{uj} x_u x_j + \sum_{j=1}^{k} b_{jj} x_j^2 \tag{I.3}$$

The coefficient b_0 is the *free term* of the regression equation, the coefficients b_j are the *linear terms*, the coefficients b_{jj} are the *quadratic terms*, and the coefficients b_{uj} are the *interaction terms*.

Any physical quantity associated with a process (temperature, pressure, flow rate, etc.) varies with time in a random manner and is therefore a *random process*. Over an observation interval, a random process takes on some specific form unknown in advance, which we shall refer to as a *realization* of the random process. By definition, a random process implies the existence of an infinite number of such realizations. By noting the realizations of the random process at regular intervals, we obtain a set of

random variables or sample functions. The time interval must be sufficiently large for the sample functions, or random variables, to be taken from independent experiments.

Random processes can be *stationary* (Fig. I.2) and *nonstationary* (Fig. I.3). Stationary random processes can be described as those processes which are independent of the choice of zero on the time axis. Another idea of stationarity is that a time translation of a sample function results in a similar sample function of the random process. Analysis of a stationary

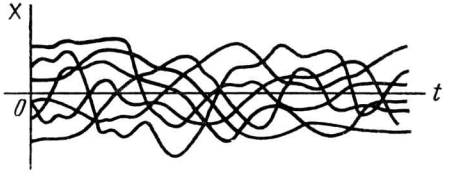

Fig. I. 2. Stationary random process

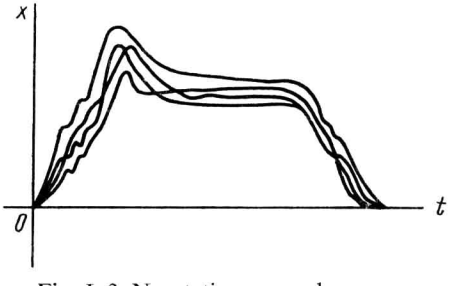

Fig. I. 3. Nonstationary random process

random process within any time interval will yield the same probability characteristics. In contrast, nonstationary random processes are those which depend on the choice of zero on the time axis, and a time translation of a sample function will not result in a similar sample function of the random process. Obviously, for nonstationary random processes the statistical properties of interest will be different in different observation intervals. If the physical quantities representative of some chemical-engineering process vary in the manner of a nonstationary random process, a model in the form of Eq. (I.3), which is an algebraic equation with constant coefficients, cannot be built in principle. This applies, for example, to catalytic reactors if the characteristics of the catalyst change abruptly in the course of service.

The statistical data necessary for analysis (model input specifications) are gathered by conducting an experiment directly on the process or plant of interest. This may be a passive or an active experiment. During a *passive experiment*, the more traditional one of the two, a large series of measurements is carried out, with the values of each of the independent variables

altered from measurement to measurement in turn. This procedure can also be used in the course of normal service of the process or plant. The data thus obtained are then processed by the methods of classical regression or correlation analysis [2-7]. An *active experiment* is conducted to a predetermined plan or design (this is known as experimental design). In an active experiment, the independent variables are altered all at the same time, so that their interaction is evaluated at once, and the amount of experimentation can substantially be cut down. The choice of an experimental design depends on the prior information available and the objective sought. The strategy of the experiment may be adjusted to suit a particular stage in the overall process or plant evaluation.

The use of experimental design (or design of experiment) was for the first time proposed by R. Fisher of Great Britain in the 1930s, but credit for the now widely used methods of experimental design goes to Box and Wilson of the United States [8]. Despite the limitations of passive experiment and classical regression analysis [2], they are still widely used under industrial conditions, because information about the behaviour of the process or plant can be obtained without upsetting the normal course of production. At present, the methods of experimental design widely used under laboratory and pilot-production conditions [9, 10, 11] are seldom employed under industrial conditions [12]. However, it may be expected that further advances in experimental design with reference to industrial conditions and progress in technology will enable experiment optimization to be applied at any stage of process development and evaluation.

Part One

ANALYSIS OF EXPERIMENT BY STATISTICAL METHODS

Chapter One

MAIN STATISTICAL CHARACTERISTICS OF RANDOM VARIABLES

1.1. Random Variables. Axioms of Probability Theory. Distribution Laws

A *random variable* is that which, in a trial, takes on a value unpredictable from the conditions of the experiment. In fact, a random variable possesses a set of allowable values, but it takes on only one at each particular trial. In contrast to non-random quantities which vary in value only when the conditions (parameters) of a trial are changed, a random variable can take on different values even though the parameters remain unchanged. Variations in a random variable from measurement to measurement are related to unobservable (random) factors.

It is convenient to characterize a random variable by specifying the set of values that it can take on. In this respect, random variables can be *discrete* and *continuous*. The values that a discrete random variable can take on can be counted in advance; in other words, a discrete variable can take on only distinct values in an interval. The values that a continuous random variable can take on cannot be counted in advance, because they fill the interval continuously, and the variable has the probability of taking any value.

It is not sufficient to specify the set of allowable values in order to characterize a random variable. For complete specification, it is necessary to state which values it can take on and how often. Suppose that a discrete random variable X can take on the values $x_1, x_2, \ldots, x_k$ as a result of an experiment. The ratio between the number m of observations or trials when the random variable X takes on the value x_i and the total number n of trials is called the *relative frequency of the event* $X = x_i$. The relative frequency m/n itself is a random variable and changes according to the number of trials made. With a very long series of trials, the relative frequency tends to stabilize about some value, p_i, called the *probability of the event* $X = x_i$, or statistically,

$$p_i = P(X = x_i) \approx m/n \tag{1.1}$$

According to Bernoulli's theorem, the probability P of the relative frequency of an event, m/n, differing from the probability p_i of the event by more

than ε tends to zero as n tends to infinity for any positive value of ε:

$$\lim_{n \to \infty} P\{|m/n - p| > \varepsilon\} = 0 \qquad (\varepsilon > 0) \tag{1.1a}$$

Kolmogorov of the Soviet Union has formulated the following axioms of probability theory.

1. The probability of a random event A is a nonnegative number:

$$P(A) \geq 0 \tag{1.2}$$

2. The probability of a certain event U is equal to unity:

$$P(U) = 1 \tag{1.3}$$

and the probability of an impossible event V is equal to zero:

$$P(V) = 0 \tag{1.4}$$

Thus,

$$0 \leq P \leq 1 \tag{1.5}$$

The sum of several events $(A_1 + A_2 + \ldots + A_n)$ is the event consisting in the occurrence of at least one of these events.

3. The probability that at least one of several disjoint events $A_1, A_2, \ldots A_n$ will occur (union is designated by the symbol $\cup$) is equal to the sum of the probabilities of these events:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + P(A_3) \tag{1.6}$$

This is the addition law of probability.

The product of several events (intersection is designated by the symbol $\cap$), $A_1 \cap A_2 \cap \ldots \cap A_n$, is the event consisting of all points common to $A_1, A_2, \ldots, A_n$.

Random events $A_1, A_2, \ldots, A_n$ are called *statistically independent*, if the probability of any one of them is independent of the probability of any other of these events. The probability of the product of several independent events (joint event) is equal to the product of the probabilities of these events:

$$P(A_1 \cap A_2 \ldots, \cap A_n) = P(A_1)\, P(A_2) \ldots P(A_n) \tag{1.7}$$

Event A is said to be dependent on (related to) event B, if the probability of event A varies according as event B has occurred. The revised probability of A when it is known that B has occurred is called the *conditional probability of A given B* and is denoted by $P(A|B)$.

For related (or mutually dependent) events, the probability of the product of two events is equal to the product of the probability of one by the conditional probability of the other given the first event occurs:

$$P(A \cap B) = P(A)\, P(B|A) \tag{1.8}$$

This is called the multiplication law of compound probabilities. Similarly, if event B precedes event A and affects it in some way, then

$$P(A \cap B) = P(B)\, P(A|B) \tag{1.9}$$

Example 1.1. The probability of trouble-free operation for a computer depends on three series-connected units each of which may fail independently of the other two. The probability of trouble-free operation for the first unit is $P(A_1) = 0.9$, for the second $P(A_2) = 0.8$, and for the third, $P(A_3) = 0.8$. Determine the probability of trouble-free operation for the computer as a whole.

Solution. According to the multiplication law of compound probabilities, Eq. (1.7),

$$P(A) = P(A_1)\,P(A_2)\,P(A_3) = 0.9 \times 0.8 \times 0.8 = 0.576$$

The sum of the probabilities of all likely values of a random variable is equal to unity

$$\sum_{i=1}^{n} p_i = 1 \tag{1.10}$$

because it is true that the random variable will take on one of its values during an experiment. This total probability is distributed among the individual values in a certain definite manner.

A discrete random variable can completely be specified by giving its values arranged in a probability sequence, that is, with the probability p_i stated for each value x_i:

x_i	x_1	x_2	x_3	...	x_n
p_i	p_1	p_2	p_3	...	p_n

Any relation connecting all likely values of a random variable to their respective probabilities is known as the *probability distribution* of that variable. The probability sequence given above is a form of representation of probability distribution.

The distribution of a continuous random variable cannot be specified by giving the probabilities of only a few values. A continuous variable can take so many values that for most of them the probability to assume the given values would be zero; that is, an event can occur, although its probability has been predicted to be zero. For continuous random variables, one is interested in the probability that, as an outcome of an experiment, the value of the variable will occur within a predetermined interval of numbers. It is convenient to use the probability that $X < x$, where x is an arbitrary real number and X is the random variable. This probability is a function of x

$$P(X < x) = F(x) \tag{1.11}$$

and is called the *distribution function* of the random variable.

A distribution function can be used to specify the distribution of a discrete as well as a continuous random variable. By definition, $F(x)$ is a nondiminishing function of x; if $x_1 \leqslant x_2$, then $F(x_1) \leqslant F(x_2)$ (Fig. 1.1*a*). The ordinate of the curve corresponding to point x_1 represents the probability that a trial will show the random variable value, X, less than x_1. The difference of the ordinates corresponding to points x_1 and x_2 gives

the probability that the values of the random variable will lie in the interval between x_1 and x_2:

$$P(x_1 \leqslant X \leqslant x_2) = F(x_2) - F(x_1) \tag{1.12}$$

At the extreme values of the argument, the distribution function takes on values 0 and 1:

$$F(-\infty) = 0 \qquad F(+\infty) = 1 \tag{1.13}$$

The distribution function of a discrete random variable is always a discontinuous step function which jumps at points corresponding to the likely values of the random variable and is equal to the probabilities of these values (Fig. 1.1*b*). The sum of all jumps is 1.

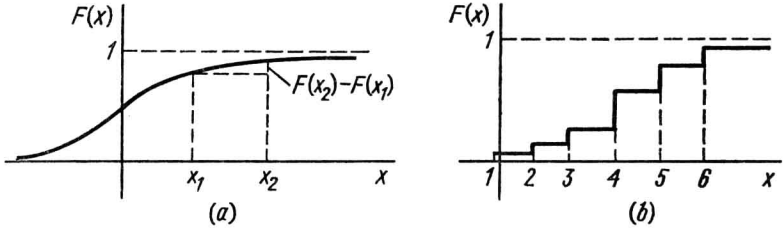

Fig. 1. 1. Distribution function of: (*a*) a continuous random variable, and (*b*) a discrete random variable

For a continuous random variable, use is more frequently made of a derivative of the distribution function, known as the *probability density function* of the random variable X. If $F(x)$ is continuous and differentiable, then

$$f(x) = F'(x) \tag{1.14}$$

The probability density function $f(x)$ also specifies the random variable completely.

The probability density function is a nonnegative function (Fig. 1.2.) The area bounded by the x-axis, lines x_1 and x_2, and the density curve represents the probability that the random veriable will take on values in the interval $[x_1, x_2]$:

$$P(x_1) \leqslant X \leqslant x_2) = \int_{x_1}^{x_2} f(x)\,dx = F(x_2) - F(x_1) \tag{1.15}$$

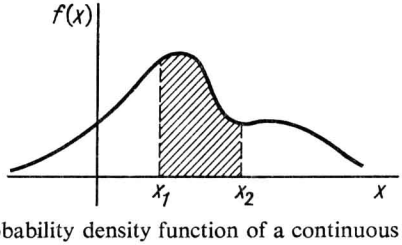

Fig. 1. 2. Probability density function of a continuous random variable

Notably,

$$F(x) = P(-\infty \leqslant X \leqslant x) = \int_{-\infty}^{x} f(x)\,dx \tag{1.16}$$

Hence comes another important property of the probability density function, namely

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{1.17}$$

because the occurrence of the random variable within the interval $-\infty < X < +\infty$ is a true event.

1.2. Numerical Characteristics

Instead of specifying a random variable by giving its probability distributions, in applied problems this is often done by giving numerical characteristics—real numbers that describe the salient features of the random variable. They are called the *moments* of a random variable. For a random variable X or the associated distribution function, the kth moment μ_k about a is the expected value of $(x-a)^k$ whenever this exists. For a discrete random variable with values $\{x_i\}$ and probabilities $\{p_i\}$

$$\mu_k = \sum (x_i - a)^k p_i$$

and for a continuous random variable with probability density function f,

$$\mu_k = \int_{-\infty}^{\infty} (x-a)^k f(x)\,dx.$$

The moments about zero ($a = 0$) are called raw moments. For a discrete random variable, the kth raw moment is given by

$$\alpha_k = \sum_{i=1}^{n} x_i^k p_i \qquad k = 1, 2, \ldots \tag{1.18}$$

and for a continuous random variable,

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x)\,dx \tag{1.19}$$

The first raw moment ($k = 1$) of a random variable is known as its arithmetic mean, designated variously as $E\{X\}$, μ_x, or μ. For discrete random variables, it is

$$\alpha_1 = \mu = E\{X\} = \sum_{i=1}^{n} x_i p_i \tag{1.20}$$

and for continuous random variables,

$$\alpha_1 = \mu = E\{X\} = \int_{-\infty}^{\infty} x f(x)\,dx \tag{1.21}$$

If a is the mean, then μ_k is the central moment of order k. The kth central moment of a discrete random variable is given by

$$\mu_k = \sum_{i=1}^{n} (x_i - \mu)^k p_i \tag{1.22}$$

and for a continuous random variable,

$$\mu_k = \int_{-\infty}^{\infty} (x-\mu)^k f(x)\, dx \tag{1.23}$$

The first central moment is zero always, $\mu_1 = 0$. The second central moment is called the *variance* of a random variable; it is a measure or variability or dispersion of the random variable. The variance of a random variable is the expected value of the square of the deviation of the random variable from its expected value or mean, and is designated as Var $\{X\}$, σ_x^2, or σ^2. Thus,

$$\text{Var}\{X\} = E\{(X-\mu)^2\} \tag{1.24}$$

For a discrete random variable, the variance is

$$\text{Var}\{X\} = \mu_2 = \sum_{i=1}^{n} (x_i-\mu)^2 p_i \tag{1.25}$$

and for a continuous random variable,

$$\text{Var}\{X\} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \tag{1.26}$$

The square root of the second central moment is called the *standard deviation* of a random variable:

$$\sigma_x = \sqrt{\text{Var}\{X\}} = \sqrt{\mu_2} \tag{1.27}$$

The third central moment divided by σ^3 is called the *coefficient of skewness:*

$$\gamma_1 = \mu_3/\sigma^3 \tag{1.28}$$

The third central moment is a measure of the symmetry of the distribution of a random variable with respect to the mean. In terms of the raw moments, it is expressed as

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3 \tag{1.29}$$

The fourth central moment is given by

$$\mu_4 = \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4 \tag{1.30}$$

It characterizes the sharpness of the peak about the mode, also called *peakedness* or *kurtosis*. The *coefficient* of *kurtosis* or *excess* is defined as

$$\gamma_2 = \mu_4/\sigma^4 - 3 \tag{1.31}$$

Plots of density functions with nonzero coefficients of skewness and excess are shown in Fig. 1.3. For comparison, it shows a dashed curve having the same mean μ and variance σ^2, but with the coefficients of skewness and excess equal to zero.

The moments exist if the respective integrals (for continuous random variables) exist or if the respective sequences (for discrete random variables) converge absolutely. For random variables with limited values, the moments exist always.