ESTIMATION OF PARAMETERS OF THE MIXTURE OF TWO NORMAL DISTRIBUTIONS BY NONLINEAR OPTIMIZATION

Josef VRBA, Antonín HAVLÍČEK and Jan ČERMÁK

Institute of Chemical Process Fundamentals,
Czechoslovak Academy of Sciences, 165 02 Prague 6 - Suchdol

Received November 15th, 1982

A method for two distributions mixture fitting is proposed. It has been proved by testing, that the method is for fitting of histograms more efficient than those using the moments of high orders.

Techniques based on measurement of random-character variables and subsequent evaluation of their statistical characteristics have been increasingly used in studies of typical two-phase chemical engineering systems (gas–liquid, gas–solid). Valuable informations in the amplitude region can be obtained from histograms. Due to the mechanism of processes in two-phase systems — alternative occurrence of individual phases in a given point of the equipment and stochastic nature of processes — resulting histogram corresponds to a mixture of two different frequency distributions. With an increasing number of variables influencing the process such a mixture can be viewed (according to the central limit theorem) as the mixture of two normal distributions.

As examples temperature histograms obtained during studies of boiling mechanism[1], histograms of porosity in bubble column reactors reflecting the presence of two different sets of bubbles formed by two different mechanisms[2], pressure histograms obtained by pressure transmitters in inhomogeneous fluidized beds containing dense phase and bubbles[3,4], *etc.*, may be given. To obtain data needed for mechanism evaluation and for individual system description, parameters of both distributions to which appropriate histograms are related have to be estimated.

A similar problem can be encountered during the evaluation of experiments of the input signal–response output type if a distribution function corresponding to normal distribution is used for partial responses approximation. As an example we can mention *e.g.* the evaluation of a response signal obtained from the chromatographic analysis of a binary mixture in cases when the complete components separation cannot be accomplished by the choice of a proper packing or when packing of limited length has to be used due to time limitation. Similarly, the described procedure can be applied for spectrophotometric analysis of two components with overlapping spectra.

The problem of resolution of the mixture of two normal distributions has been solved by numerous authors. Several methods for both general and specific cases can be found *e.g.* in work by Cohen[3]. These methods are based on the comparison between moments of experimental data and estimates of parameters of the two distributions. In the general case, when estimates of two mean values, two standard deviations and of the proportionality constant have to be obtained, it is necessary to determine the first five moments and to find a solution of a polynomial equation of the ninth order. The solution of the polynomial equation can be avoided using an alternative iterative procedure. According to this procedure initial estimates (based on the assumption of equal standard deviations) are used for the original system of moment equations.

Experiences with experimental data evaluation point out, however, the low estimation accuracy especially for the higher-order moments. An alternative approach was therefore chosen to obtain a procedure for two normal distributions resolution which was developed as a part of system software for on-line data processing on a desk calculator. The problem was formulated as a search for an extreme of the loss function following the preliminary procedure of numerical data filtration[4]. The method can be modified easily for another type of distribution this being impossible in the case of the procedure based upon moment equations. If the type of distribution cannot be assumed *a priori* the most advantageous model can be automatically chosen on the basis of minimal loss function values. The nonweighted sum of squares of deviations of experimental data and model outputs is considered as the loss function.

The aim of this work is to prove the feasibility of the direct estimation of parameters of a mixture of two normal distributions using the nonlinear optimization method.

THEORETICAL

Applying the least squares technique, the problem of estimation of parameters of the two normal distributions mixture can be transformed into the problem of nonlinear optimization in the space of parameters Ω; vector of parameters $\boldsymbol{\Theta} \in \Omega$ is co-ordinated to the regression equation (in general form)

$$\mathbf{Y} = \eta(\mathbf{x}, \boldsymbol{\Theta}) + \mathbf{e} = \eta(\mathbf{x}, \boldsymbol{\Theta}^{\wedge}) + \boldsymbol{\varepsilon} \tag{1}$$

according to the condition

$$\begin{aligned} \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon} &= [\mathbf{Y} - \eta(\mathbf{x}, \boldsymbol{\Theta}^{\wedge})]^{\mathrm{T}} [\mathbf{Y} - \eta(\mathbf{x}, \boldsymbol{\Theta}^{\wedge})] = \\ &= \min_{\boldsymbol{\beta}\in\Omega} \{[\mathbf{Y} - \eta(\mathbf{x}, \boldsymbol{\beta})]^{\mathrm{T}} [\mathbf{Y} - \eta(\mathbf{x}, \boldsymbol{\beta})]\}, \end{aligned} \tag{2}$$

where $\boldsymbol{\beta}$, $\boldsymbol{\Theta}^{\wedge}$ are estimates of the vector

$\boldsymbol{\Theta} = (C_1, C_2, \sigma_1, \sigma_2, \mu_1, \mu_2)^{\mathrm{T}}$ corresponding to the case of the mixture of two normal distributions

$$\begin{aligned} \eta(x_i, \boldsymbol{\Theta}) &= \frac{C_1}{\sqrt{(2\pi)}\,\sigma_1} \exp\left[-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}\right] + \\ &+ \frac{C_2}{\sqrt{(2\pi)}\,\sigma_2} \exp\left[-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}\right] \end{aligned} \tag{3}$$

where

$$C_1 + C_2 = 1\,. \tag{4}$$

To solve the optimization problem

$$F(\boldsymbol{\beta}) = \sum_{i=1}^{N} [Y_i - \eta(x_i, \boldsymbol{\beta})]^2 \stackrel{!}{=} \text{minimum} \tag{5}$$

an arbitrary effective technique of nonlinear programming can be applied, the derivative-free procedure UNCOM (ref.[5]) and the Gauss least squares method were used in our case. The UNCOM procedure fulfiles the demands both on the accuracy of $\boldsymbol{\Theta}$ estimation and on the calculation speed.*

Unambiguous resolution of the two distributions is generally possible in the case of existence of three local extremes of the function (*3*), more specifically of two maximums (existing in a vicinity of mean values μ_1, μ_2) and of a single minimum (lying between the two maximum values). A general relation for the extreme value x^+ can be derived from the condition $\mathrm{d}\eta(\boldsymbol{x}, \boldsymbol{\Theta})/\mathrm{d}x = 0$ in the form

$$\frac{C_1}{C_2}\left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{u}{v} = \exp\left[(u^2 - v^2)/2\right], \tag{6}$$

where

$$u = \frac{x^+ - \mu_1}{\sigma_1} > 0 \tag{7}$$

$$v = \frac{\mu_2 - x^+}{\sigma_2} > 0 \tag{8}$$

$$x^+ \in (\mu_1, \mu_2).$$

Equation (*6*) can be written as

$$z \equiv z_l = z_r \tag{9}$$

where

$$z_l = \ln\left[\frac{C_1}{C_2}\left(\frac{\sigma_2}{\sigma_1}\right)^3 \frac{x^+ - \mu_1}{\mu_2 - x^+}\right] \tag{10}$$

* The method UNCOM seeks for an unconstrained extreme of a general nonlinear function. This method belongs to the group of derivative-free methods determining the local extreme of a function. A single-dimension search procedure based upon a specific form of a quadratic interpolation is used to determine the optimal length of step in the given direction from the given point. Considering the solution of the problem of parameter estimation for the mixture of two normal distributions, the UNCOM method proved to be suitable for implementation on a desk calculator (*e.g.* HP 9821 A).

$$z_r = \frac{1}{2}\left[\left(\frac{x^+ - \mu_1}{\sigma_1}\right)^2 - \left(\frac{\mu_2 - x^+}{\sigma_2}\right)^2\right] \quad (11)$$

and the three extremes in question can be determined numerically or graphically as a common solution of functions $z_l(x)$ and $z_r(x)$. As an example, the determination of the roots of Eq. (9) for $C_1 = C_2 = 0{\cdot}5$, $\mu_1 = 4$, $\mu_2 = 6$ and $\sigma_1 = \sigma_2 = \sigma = 0{\cdot}75$ and $\sigma_1 = \sigma_2 = \sigma = 0{\cdot}90$, resp., is shown in Figs 1*a* and 1*b*. Apparently, a successive convergence of roots occurs with increasing value of σ till finally full identity of roots takes place for $\sigma = (\mu_2 - \mu_1)/2 = 1$ (triple root – maximum). In our symetrical case the condition

$$\Delta\mu = \mu_2 - \mu_1 > 2\sigma \quad (12)$$

has to be fulfilled for the existence of three different roots. These considerations are however valid only for the theoretical model described by Eq. (3) which is not affected by random disturbances (*i.e.* $\mathbf{e} = \mathbf{O}$).

The mean values μ_1 and μ_2 exhibit the best adaptability to the decrease of value of the objective function $F(\boldsymbol{\beta})$, to a less extent this is valid also for coefficients C_1 and C_2 (for C_2 condition (4) is valid). Initial approximation of standard deviations σ_1, σ_2 proved to be controlling for the rate of calculation. Whereas the initial approximations of parameter values μ_1 and μ_2 can be determined with relative accuracy (*e.g.* from a graphical record of pairs x_i, Y_i) it is very difficult to obtain appropriate estimations of σ_1 and σ_2. Both the initial estimation and the satisfactory solution

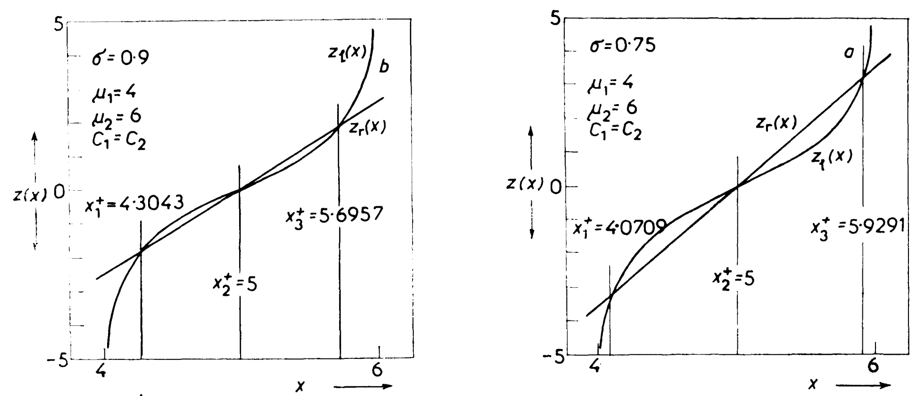

FIG. 1
Graphical solution of Eqs (9)—(11)

of the optimization problem (5) depend in principle on the feasibility of the two distributions resolution. The starting approximation of the values of calculated parameters can be obtained by different ways; in the paper we present a method which proved to be suitable in all cases solved.

Our theoretical model can be written formally as

$$y(x|\,\sigma_1, \mu_1, \sigma_2, \mu_2) = C_1 y_1(x|\,\sigma_1, \mu_1) + C_2 y_2(x|\,\sigma_2, \mu_2)\,. \tag{13}$$

Let us suppose that the coordinates of both peaks of the function (3) (assuming that they are at least visually distinguishable) can be with satisfactory accuracy considered as the coordinates of mean values μ_1, μ_2. The following approximative relations are valid for function values in both peaks (in agreement with Eq. (13))

$$y_{\max 1} \doteq y(\mu_1|\,\sigma_1, \mu_1, \sigma_2, \mu_2) = C_1(\sqrt{(2\pi)}\,\sigma_1)^{-1} + C_2 y_2(\mu_1|\,\sigma_2, \mu_2)\,, \tag{14}$$

$$y_{\max 2} \doteq y(\mu_2|\,\sigma_1, \mu_1, \sigma_2, \mu_2) = C_2(\sqrt{(2\pi)}\,\sigma_2)^{-1} + C_1 y_1(\mu_2|\,\sigma_1, \mu_1)\,. \tag{15}$$

The positions of peaks $y_{\max 1}, y_{\max 2}$ can be determined and used as the initial estimates of μ_1 and μ_2. Initial estimates of coefficients C_1 and C_2 can be obtained on the basis of their properties-weights of individual distributions (see condition (4)), the ratio C_1/C_2 corresponds to the ratio of subintegral areas of both distributions. Functional values of each of the two distributions are to a minimum extent influenced by the values of the other one in external halves of their interval. These external half-intervals can be therefore used for the estimation of the size of individual areas and the value of the coefficient C_1 can be determined from the relation

$$C_1 \doteq A/(A + B) \tag{16}$$

where

$$A = \int_{\mu_1 - m}^{\mu_1} y(x|\,\sigma_1, \mu_1, \sigma_1, \mu_2) \doteq \sum_{i = j_{\mu_1} - j_m}^{j_{\mu_1}} y(x_i|\,\sigma_1, \mu_1, \sigma_2, \mu_2) \tag{17}$$

$$B = \int_{\mu_2}^{\mu_2 + m} y(x|\,\sigma_1, \mu_1, \sigma_2, \mu_2) \doteq \sum_{i = j_{\mu_2}}^{j_{\mu_2} + j_m} y(x_i|\,\sigma_1, \mu_1, \sigma_2, \mu_2) \tag{18}$$

$$m = \min\,[\mu_1 - x_0, x_N - \mu_2] \tag{19}$$

$j\mu_1, j_m, j\mu_2$ – sequence indexes of frequency interval.

For the determination of C_2 the condition (*4*) is valid. The procedure based upon the use of complete half-intervals $(\mu_1 - x_0)$, $(x_N - \mu_2)$ yielded in all cases significantly worse initial estimation of C_1 than the procedure working with the cut-outs (segments) $(\mu_1 - m)$, $(\mu_2 + m)$ (see above).

The estimation of values of the standard deviations σ_1, σ_2 can be then based upon the estimated values of μ_1, μ_2, C_1 and C_2 (which consequently influences the estimation quality). The estimation procedure can be described by an iterative block.

$$\sigma_1 = \sigma_2 = 1$$

$$\sigma_1 = C_1/\{[y_{\max 1} - C_2 y_2(\mu_1 | \sigma_2, \mu_2)] \sqrt{(2\pi)}\}$$
$$\sigma_2 = C_2/\{[y_{\max 2} - C_1 y_1(\mu_2 | \sigma_1, \mu_1)] \sqrt{(2\pi)}\}$$

convergence? — no

yes

Functional values $y_{\max 1}$, $y_{\max 2}$ and corresponding coordinates can be determined easily for the theoretical model (*3*) unaffected by random disturbances. When such disturbances occur individual points (Y_i, x_i) are more or less dispersed on both sides of the assumed smoothed line. This can be observed in cases of histograms obtained from amplitude analysis of limited experimental data sets and from the time behaviour of output signals in cases of input–output experiments. Determination of peaks to be searched for is under such circumstances practically impossible both visually and numericaly. Smoothing of real frequency curves by a simple digital filter proved to be helpful in such cases. The digital filtering was aimed at the preliminary removal of the fluctuating component causing the data scattering while not transmitting any information in the modelled system.

The simplest type of the digital filter removing the higher frequencies is the summation low-frequency filter which can be described by the equation

$$y_{Fi} = h_{-n}y_{i-n} + \ldots + h_{-1}y_{i-1} + h_0y_i + h_1y_{i+1} + \ldots + h_ny_{i+n} \tag{20}$$

$$i = 1, 3, \ldots, N$$

$$j = 1, 2, \ldots, n$$

where y_{Fi} is a filter output value in a discrete interval i; y_{i+j} is an input filtered value in a discrete interval $i + j$; h_j is a weighting coefficient and n is a filter order.
In the simplest case, the filter with constant weighting coefficients

$$h_i = 1/(2n + 1) = h \tag{21}$$

can be used for filtering.

Such a summation-type filter lets only very slow signals (*i.e.* signals with frequency close to zero) pass through without an amplitude distortion. If therefore data containing both the slow signal component, *i.e.* functional dependency (in out case the sum of the two distribution functions) and the fast component (fluctuations) are filtered, the functional dependence obtained is distorted by fluctuations which are significantly reduced by the summation filter used. Estimation of parameters of our model described by Eqs (*13*)–(*19*) is then applied to such smoothed data. In cases when only the position of a single peak is apparent from the graphic representation (or from the analysis of numerical data sequence) coordinates of the second mean value can be estimated *e.g.* according to the position of a fracture on the smoothed line of supposed shape (inflection *etc.*). In such a case the position of estimated mean value coordinates is given explicitely.

The parameter estimates obtained as solutions of the optimization problems (*2*) or (*5*) have to be tested regarding their confidence region. On doing this it is assumed *a priori* that the data to be analyzed can be described by the model (*3*). The approximative confidence region of nonlinear model parameters can be obtained only from the linearized model; the linearization is performed in the vicinity of the vector of nonlinear model parameters[6]. For a linear or linearized model respectively expressed by the relation

$$Y_i = \sum_{j=1}^{n} \Theta_j\varphi_j(x_i) + e_i\,; \quad i = 1, 2, \ldots, N \tag{22}$$

where x_i are values of an independent variable; Y_i are values of a dependent variable *i.e.* experimental (measured) data; φ_j is a known function of x; Θ_j is an unknown parameter and e_i is a measurement error and model disturbance a matrix equation can be written

$$\mathbf{Y}(N, 1) = \boldsymbol{\Phi}(N, n)\,\boldsymbol{\Theta}(n, 1) + \mathbf{e}(N, 1) \tag{23}$$

where

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)^T ,$$

$$\boldsymbol{\Phi} = \begin{pmatrix} \varphi_1(x_1), & \varphi_2(x_1), & \ldots, & \varphi_n(x_1) \\ \varphi_1(x_2), & \varphi_2(x_2), & \ldots, & \varphi_n(x_2) \\ \cdots & \cdots & \cdots & \cdots \\ \varphi_1(x_N), & \varphi_2(x_N), & \ldots, & \varphi_n(x_N) \end{pmatrix} ,$$

$$\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_n)^T ,$$

$$\mathbf{e} = (e_1, e_2, \ldots, e_N)^T ,$$

and N is a number of data pairs and n is a number of model parameters.

Estimation of parameters $\boldsymbol{\Theta}^\wedge$ by the least squares method is based upon the minimization of the expression

$$\mathbf{e}^T\mathbf{e} = (\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge)^T (\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge) = \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge - \boldsymbol{\Theta}^{\wedge T}\boldsymbol{\Phi}^T\mathbf{Y} + \boldsymbol{\Theta}^{\wedge T}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge \quad (24)$$

so that

$$-2\boldsymbol{\Phi}^T(\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge) = \mathbf{O}$$

or

$$\boldsymbol{\Theta}^\wedge = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T\mathbf{Y} . \quad (25)$$

Supposing that $\boldsymbol{\Phi}$ and $\mathbf{e}$ are mutually independent then

$$E\{\boldsymbol{\Theta}^\wedge\} = E\{(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T\mathbf{Y}\} = E\{(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Theta} + (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T\mathbf{e}\} =$$
$$= E\{\boldsymbol{\Theta} + (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T\mathbf{e}\} = \boldsymbol{\Theta} + (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\, \boldsymbol{\Phi}^T E\{e\} = \boldsymbol{\Theta} . \quad (26)$$

If $\boldsymbol{\Phi}$ and $\mathbf{e}$ are independent and e_i are normally distributed with zero mean value and with a variance σ^2 then

$$E\{(\boldsymbol{\Theta}^\wedge - \boldsymbol{\Theta})(\boldsymbol{\Theta}^\wedge - \boldsymbol{\Theta})^T\} = \sigma^2(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1} . \quad (27)$$

When the value of variance σ^2 is unknown its estimate

$$s^2 = (\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge)^T(\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge)/(N - n) \quad (28)$$

can be used. From Eqs (27) and (28) it follows then for the variance of parameters estimate

$$\sigma^2_{\Theta^\wedge} = (\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge)^T (\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\Theta}^\wedge)(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}/(N - n) . \quad (29)$$

The estimation of the confidence region for the parameters vector $\boldsymbol{\Theta}^\wedge$ can be made using the so called Student test. The variable

$$\mathbf{t} = (\boldsymbol{\Theta}^\wedge - \boldsymbol{\Theta})/\sigma_{\boldsymbol{\Theta}^\wedge} \qquad (30)$$

is a random variable described by the Student distribution. For the given confidence level determined by the coefficient α, the inequality

$$\mathbf{t}_D \leqq \mathbf{t} \leqq \mathbf{t}_H \qquad (31)$$

can be written, where indices D and H denote lower and upper boundaries exceeded by the quantity $\mathbf{t}$ with $100\alpha\%$ probability (for given number of degrees of freedom $\gamma = N - n$). Linearization of our nonlinear model can be done in the vicinity of parameter values obtained by the optimization method for the nonlinear model. The first two terms of the Taylor expansion

$$y(x) = y(x|\,\boldsymbol{\Theta} = \boldsymbol{\Theta}^\wedge) + \sum_{i=1}^{n} \left(\frac{\partial y(x)}{\partial \Theta_i}\right)\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (\Theta_i - \hat{\Theta}_i) \qquad (32)$$

can be used for linearization. This model linearized in parameters is (as to its form) identical with the model (*22*), where $\boldsymbol{\Theta}$ is the vector of parameters of the linearized model from Eqs (*25*) and (*26*) and $\boldsymbol{\Theta}^\wedge$ is the vector of parameters estimates obtained from the original nonlinear model. Our model – the mixture of the two normal istributions (*3*) can be written in a linearized form as

$$\begin{aligned} y(x_i) \doteq{} & \frac{\hat{C}_1}{\sqrt{(2\pi)}\,\hat{\sigma}_1} \exp\left(-\frac{(x_i - \hat{\mu}_1)^2}{2\hat{\sigma}_1^2}\right) + \frac{\hat{C}_2}{\sqrt{(2\pi)}\,\hat{\sigma}_2} \exp\left(-\frac{(x_i - \hat{\mu}_2)^2}{2\hat{\sigma}_2^2}\right) + \\ & + \frac{\partial y(x_i)}{\partial C_1}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (C_1 - \hat{C}_1) + \frac{\partial y(x_i)}{\partial C_2}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (C_2 - \hat{C}_2) + \\ & + \frac{\partial y(x_i)}{\partial \sigma_1}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (\sigma_1 - \hat{\sigma}_1) + \frac{\partial y(x_i)}{\partial \sigma_2}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (\sigma_2 - \hat{\sigma}_2) + \\ & + \frac{\partial y(x_i)}{\partial \mu_1}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (\mu_1 - \hat{\mu}_1) + \frac{\partial y(x_i)}{\partial \mu_2}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} (\mu_2 - \hat{\mu}_2) \end{aligned} \qquad (33)$$

where

$$\frac{\partial y(x_i)}{\partial C_k}\bigg|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^\wedge} = B_{ki}/\hat{C}_k \qquad (34)$$

$$\left.\frac{\partial y(x_i)}{\sigma_k}\right|_{\Theta=\Theta^\wedge} = B_{ki}(A_{ki}^2 - 1)/\hat{\sigma}_k \tag{35}$$

$$\left.\frac{\partial y(x_i)}{\partial \mu_k}\right|_{\Theta=\Theta^\wedge} = A_{ki}B_{ki}/\hat{\sigma}_k \tag{36}$$

$$A_{ki} = (x_i - \hat{\mu}_k)/\hat{\sigma}_k \tag{37}$$

$$B_{ki} = \hat{C}_k \exp\left(-A_{ik}^2/2\right)/\sqrt{(2\pi)}\,\hat{\sigma}_k) \quad k = 1, 2\,; \quad i = 1, 2, \ldots, N\,. \tag{38}$$

The constant term of the Taylor expansion can be transferred to the left side of Eq. (*33*) and following relations can be written in agreement with the general relation (*22*)

$$Y_i' = Y_i + B_{1i}(A_{1i}^2 - 1) + B_{2i}(A_{2i}^2 - 1) + A_{1i}B_{1i}\hat{\mu}_i/\hat{\sigma}_1 + A_{2i}B_{2i}\hat{\mu}_2/\hat{\sigma}_2 \tag{39}$$

$$Y_i' = \sum_{j=1}^{6} \varphi_{ij}\Theta_j \tag{40}$$

where

$$\varphi_{i1} = B_{1i}/\hat{C}_1\,, \quad \Theta_1 = C_1\,; \tag{41}$$

$$\varphi_{i2} = B_{2i}/\hat{C}_2\,, \quad \Theta_2 = C_2\,; \tag{42}$$

$$\varphi_{i3} = B_{1i}(A_{1i}^2 - 1)/\hat{\sigma}_1\,, \quad \Theta_3 = \sigma_1\,; \tag{43}$$

$$\varphi_{i4} = B_{2i}(A_{2i}^2 - 1)/\hat{\sigma}_2\,, \quad \Theta_4 = \sigma_2\,; \tag{44}$$

$$\varphi_{i5} = A_{1i}B_{1i}/\hat{\sigma}_1\,, \quad \Theta_5 = \mu_1\,; \tag{45}$$

$$\varphi_{i6} = A_{2i}B_{2i}/\hat{\sigma}_2\,, \quad \Theta_6 = \mu_2\,. \tag{46}$$

RESULTS

This part of the work is devoted to practical problems of smoothing of relative frequency curves by probability density functions. As the first step pf preliminary appreciation of a curve shape it is necessary to estimate if the curve can be described by the single probability density function or by the mixture of two probability density functions, having in a general case parameters C_1, C_2, σ_1, σ_2, μ_1, μ_2. If the experimental data curve exhibits a distinctive saddle between two peaks (after data filtering, if needed) the model of two distributions mixture can be expected to be appropriate. Analogously this is true if an evident plateau can be observed on a single-peak curve. When such a saddle (plateau) is not apparent both models *i.e.* model of a single distribution or that of the mixture of two distributions, can be expected to fit the data. In such cases only the loss function values corresponding to data smoothing by the two models in question can be used as an exact criterion whereas the visual comparison of experimental curves with the model ones can be misleading in some cases. The theoretical curve of single normal distribution with parameters $N(5; 2)$ can be *e.g.* approximated with the accuracy corresponding to the loss function value 10^{-6} by the theoretical curve of the mixture of two normal distributions with parameters $N_1(4{\cdot}39; 1{\cdot}90)$ and $N_2(5{\cdot}61; 1{\cdot}90)$, $C_1 = C_2 = 0{\cdot}5$. The curves are practically undistinguishable even in the case of very fine graphical presentation. If however the order of calculation accuracy is increased gradually (which indeed results in a consequent decrease of the calculation speed) the solution of the problem approaches theoretical values and the curve can be identified as that of single distribution. Even in cases when the hypothesis of the presence of two different distributions in data set concerned is not unambiguously supported by physical mechanism of the process, the experience obtained from the simulated data treatment suggest that it is considerably safer to start the determination of the initial approximation of parameters estimates assuming the two distributions existence. For the alternate initial assumption it has been proved that even when a distinctive plateau of the frequency curve can be observed the local loss function extreme corresponding to a single distribution curve can be obtained as the solution. Selection between the two types of models has to be based therefore upon a rigorous analysis of the problem to be solved.

The distinguishability of the model of a two distributions mixture depends on relations between $\Delta\mu$ and the magnitude of σ_1, σ_2 on one hand and on relation between C_1 and C_2 on the other hand. It has however to be pointed out that the effect of individual factors is always displayed simultaneously in an increased extent. As a general rule, the ability of model recognition increases with increasing $\Delta\mu$ while decreasing steeply with increasing σ_1, σ_2. Considering the isolated effect of C_1 and C_2 only, the equality $C_1 = C_2 = 0{\cdot}5$ can be postulated as the best distinguishability condition. The condition (*12*) (in the symmetrical case $\sigma_1 = \sigma_2$ and $C_1 = C_2$) can be used as an approximate criterion of distinguishability, the controlling calculations proved however that the distinguishability demanded could be reached even when the condition (*12*) was not fulfilled. Apart from the effect of factors mentioned above, the influence of random disturbances was further studied, as caused by experimental errors and/or by the insufficient quality of a frequency curve (histogram) due to improperly chosen frequency class width. The existence of such random disturbances makes the distinguishability problem even more complicated.

Results of some typical problems of the estimation of parameters of a two normal distributions mixture are presented in Table I. Except for experiments *9* and *10*

TABLE I

Parameters estimation results

Experiment	Data type	C_1	C_2	μ_1	μ_2	σ_1	σ_2	$F(\beta)$	NFE[h]
	1[d]	0·50	0·50	3·00	7·00	1·00	1·00		
	2[e]	0·50	0·50	3·00	7·00	1·00	1·00		
1 A[a]	3[f]	0·50	0·50	3·00	7·00	1·00	1·00	0·5—07	1
	4[g]	0·00	0·00	0·00	0·00	0·00	0·00		
	1	0·50	0·50	3·00	7·00	1·00	1·00		
	2	0·50	0·50	3·01	7·03	0·99	0·00		
1 B[b]	3	0·50	0·50	3·01	7·03	0·99	0·98	0·4—02	80
	4	0·02	0·02	0·04	0·04	0·05	0·05		
	1	0·50	0·50	3·00	7·00	2·00	2·00		
	2	0·49	0·51	2·98	6·98	1·99	2·01		
2 A	3	0·50	0·50	3·00	7·00	2·00	2·00	0·2—06	116
	4	0·00	0·00	0·00	0·00	0·00	0·00		
	1	0·50	0·50	3·00	7·00	3·00	3·00		
	2	0·50	0·50	3·00	7·00	3·00	3·00		
3 A	3	0·50	0·50	3·01	7·01	3·00	3·00	0·6—07	22
	4	0·04	0·04	0·14	0·15	0·04	0·04		
	1	0·90	0·10	3·00	7·00	1·00	1·00		
	2	0·90	0·10	3·00	7·00	1·00	1·00		
4 A	3	0·90	0·10	3·00	7·00	1·00	1·00	0·7—07	1
	4	0·00	0·00	0·00	0·00	0·00	0·00		
	1	0·90	0·10	3·00	7·00	1·00	1·00		
	2	0·91	0·09	3·01	7·14	1·00	0·85		
4 B	3	0·90	0·09	3·01	7·13	1·00	0·86	0·4—02	82
	4	0·02	0·02	0·02	0·18	0·02	0·19		
	1	0·90	0·10	3·00	7·00	1·30	1·30		
	2	0·90	0·10	3·00	6·99	1·30	1·32		
5 A	3	0·90	0·10	3·00	7·00	1·30	1·30	0·3—05	36
	4	0·00	0·00	0·00	0·00	0·00	0·00		
	1	0·90	0·10	3·00	7·00	1·30	1·30		
	2	0·90	0·10	3·02	7·19	1·29	1·19		
5 B	3	0·90	0·09	3·02	7·18	1·29	1·17	0·4—02	238
	4	0·03	0·03	0·04	0·36	0·05	0·38		
	1	0·50	0·50	3·00	7·00	0·50	2·00		
	2	0·50	0·50	3·00	7·00	0·50	2·00		
6 A	3	0·50	0·50	3·00	7·00	0·50	2·00	0·2—06	141
	4	0·00	0·00	0·00	0·00	0·00	0·00		

TABLE I
(*Continued*)

Experiment	Data type	C_1	C_2	μ_1	μ_2	σ_1	σ_2	$F(\beta)$	NFE[h]
	1	0·50	0·50	3·00	7·00	0·50	2·00		
	2	0·51	0·49	3·00	7·07	0·51	0·97		
6 B	3	0·51	0·50	3·00	7·07	0·51	1·98	0·4—02	179
	4	0·02	0·04	0·02	0·14	0·02	0·19		
	1	0·50	0·50	3·00	4·00	0·50	0·50		
	2	0·51	0·49	3·00	4·00	0·51	0·50		
7 A	3	0·50	0·50	3·00	4·00	0·50	0·50	0·1—04	56
	4	0·00	0·00	0·00	0·00	0·00	0·00		
	1	0·50	0·50	3·00	4·00	0·50	0·50		
	2	0·50	0·50	2·99	4·00	0·51	0·50		
7 B	3	0·49	0·51	2·99	3·99	0·50	0·50	0·4—02	96
	4	0·13	0·13	0·12	0·11	0·06	0·06		
	1	0·10	0·90	3·00	4·00	0·40	2·00		
	2	0·10	0·90	3·00	4·00	0·04	2·00		
8 A	3	0·10	0·90	3·00	4·00	0·40	2·00	0·1—05	329
	4	0·00	0·00	0·00	0·00	0·00	0·00		
	1	0·10	0·90	3·00	4·00	0·40	2·00		
	2	0·11	0·89	2·99	4·04	0·45	2·00		
8 B	3	0·11	0·87	2·98	4·04	0·46	1·98	0·4—02	359
	4	0·02	0·04	0·06	0·09	0·07	0·08		
	1	—	—	—	—	—	—		
	2	0·44	0·56	4·78	5·59	0·56	0·22		
9 C[c]	3	0·52	0·54	4·80	5·60	0·63	0·21	0·3—01	219
	4	0·09	0·08	0·13	0·01	0·11	0·02		
	1	—	—	—	—	—	—		
	2	0·61	0·39	4·61	5·94	0·94	0·25		
10 C	3	0·66	0·37	4·65	5·94	1·00	0·25	0·3—01	165
	4	0·09	0·06	0·15	0·02	0·14	0·03		

[a] Theoretical values; [b] theoretical values with superposed noise 0·01; [c] real data (experiment 9—27 frequency classes, experiment 10—40 frequency classes); [d] theoretical value; [e] calculated value (from the original model); [f] calculated value (from the linearized model); [g] reliability interval (for the linearized model) at $\alpha = 0{\cdot}05$; [h] number of function evaluation in optimization.

dealing with real data from a fluidized-bed reactor, all other data were simulated on the calculator. The simulation was carried out for mixtures of two theoretical curves at two noise levels 0·00 and 0·01. Higher noise levels than the defined level 0·01 do not fulfil conditions of proper histogram formation (too low frequency in individual classes). Fifty values were used in all cases for simulated data, corresponding to 50 classes in a real histogram. It has to be remembered that the smoothness of frequency curves can be up to a certain level raised by the use of smaller class number *i.e.* by the frequency increase in individual classes. Parameters of the theoretical model of mixtures of two normal distributions were chosen with respect to the maximum distinguishability proof (*i.e.* even for the cases when condition (*12*) is not fulfilled) considering further all possible shapes of histograms from real data. It was observed that in problems with simulated data the noise in the cases of short realizations (50 values) could cause systematic deviations in the shape of probability curves. The loss function value was therefore in the cases of such curves with noise systematically lower for parameter estimates obtained than for curves with the theoretical parameters values.

Initial parameter approximations for the optimization procedure application were determined using the procedure described in the theoretical part (Eqs (*13*)–(*19*)). In Table I four values are given for each parameter, individual values corresponding to the theoretical value, to the values calculated for the original model and for the linearized model respectively and to the confidence interval for $\alpha = 0.05$ (calculated from the linearized relation). Despite the fact that the reliability intervals ($\pm$) were determined only indirectly they can be used with a sufficient accuracy for the characterization of reliability of the calculated nonlinear model parameters. Three types of problems can be distinguished in Table I. The first type of problems deals with symmetrical cases ($C_1 = C_2$ and $\sigma_1 = \sigma_2$), the second type includes asymmetrical cases and the problems of the third type are the real data problems. Figs 2 and 4 show cases when $\Delta\mu = 2\sigma$, Fig. 3 corresponds to the case $\Delta\mu < 2\sigma$. The asymmetrical cases shown in Fig 5 and 6 correspond approximately to the symmetrical cases for $\Delta\mu \leqq 2\sigma$. In all problems tested, the parameters were determined with accuracy of two decimal points. Such an accuracy represents a certain compromise between possible practical demands on the reliability of obtained results and a reasonable calculation time. Number of determinations of $F(\beta)$ values ranged for optimization procedures between 1 and 330 for the theoretical curves without any noise and between 80 and 440 for the curves with noise; lower values were obtained for curves with distinct peaks around both mean values, the higher values correspond to single-peak curves. The frequency curves and curves of probability densities with calculated parameters are presented in Figs 2–8. An illustrative example of a smoothed frequency curve after the numerical data filtration is further given in Figs 2*b*, 3*b*, 4*b*, 5*b* and 6*b*. The original histograms are shown in Figs 7 and 8 representing results obtained from experimental data. These experimental data were obtained from pressure fluctuation measurements carried out on the fluidized-bed reactor described elsewhere[4].

CONCLUSION

An extensive set of simulation calculations proved that the effectiveness of the procedure suggested depends on the choice of initial approximations as in all cases the existence of several local extremes of the minimized loss function can be expected

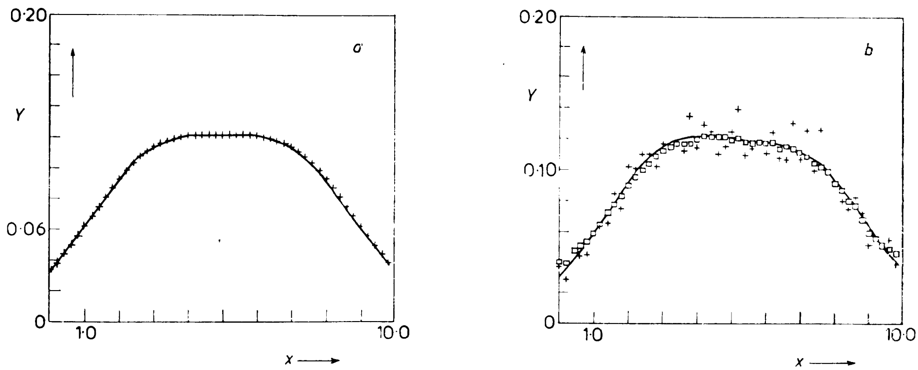

FIG. 2

Probability density function — parameters estimation. Symmetrical case; $\Delta\mu = 2\sigma$; + original data; □ filtered data; — fitted curve; *a* noise level 0·00 (Exp. 2A); *b* noise level 0·01 (Exp. 2B)

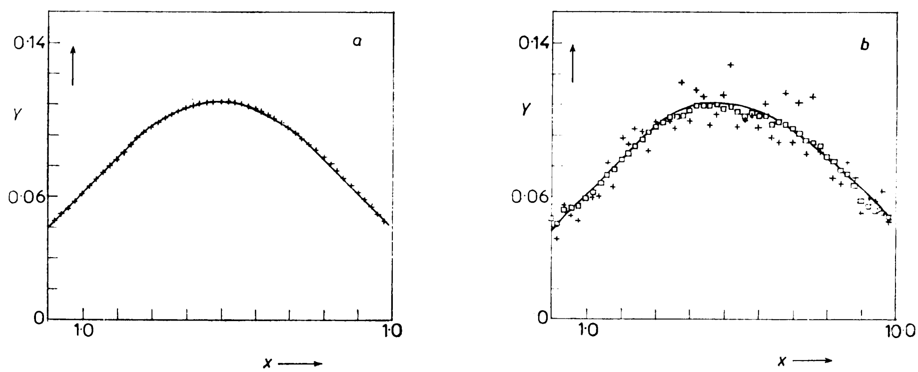

FIG. 3

Probability density function — parameters estimation. Symmetrical case; $\Delta\mu = 2\sigma$; + original data; □ filtered data; — fitted curve; *a* noise level 0·00 (Exp. 3A); *b* noise level 0·01 (Exp. 3B) due to the high nonlinearity of the problem solved. The strategy of initial approximations choice utilized the numerical pre-filtration of histogram values or response curves and a subsequent choice of initial approximations of mean values according to the position of apparent peaks of the distribution mixture and of proportionality constants derived from them on the basis of ratios of segments of subintegral areas. Using such strategy, the extreme to be looked for could be found in all cases after a reasonable number of iterations. In the case when original curve did not indicate the presence of two distributions in the mixture, it was always useful to expect it as the calculation algorithm ensured in such a case even a reliable determination

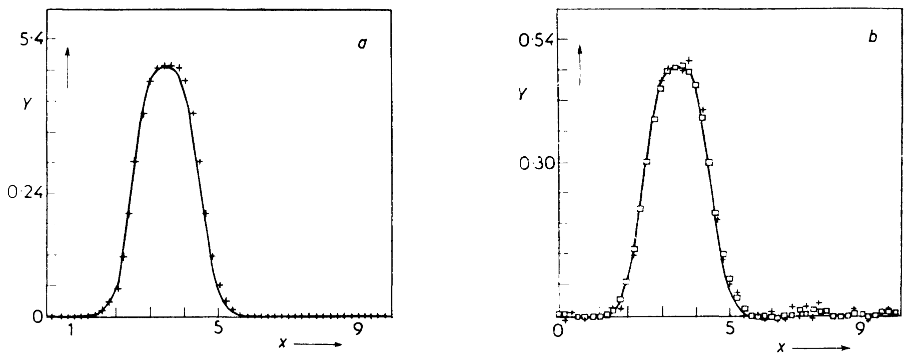

FIG. 4

Probability density function — parameters estimation. Asymmetrical case; $\Delta\mu = 2\sigma$; + original data; □ filtered data; — fitted curve; *a* noise level 0·00 (Exp. 5A); *b* noise level 0·01 (Exp. 5B)

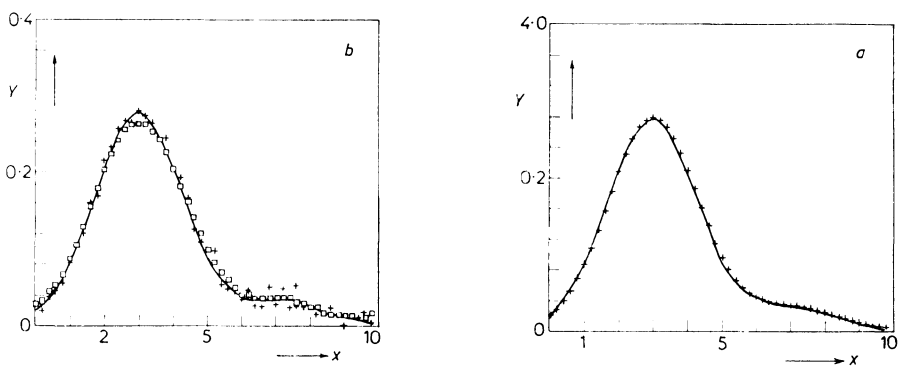

FIG. 5

Probability density function — parameters estimation. Asymmetrical case; + original data; □ filtered data; — fitted curve; *a* noise level 0·00 (Exp. 7A); *b* noise level 0·01 (Exp. 7B) of the estimates of a single distribution parameters. In the opposite case, however, *i.e.* when the single distribution was assumed at the beginning, the resolution desired could never be reached.

The procedure suggested in this work can be used even for the evaluation of pre--treated data accessible in the form of histograms (data from amplitude analysers) or response curves. In this respect it is superior to procedures described in literature which are based upon moments of higher orders. Such procedures require calcula-

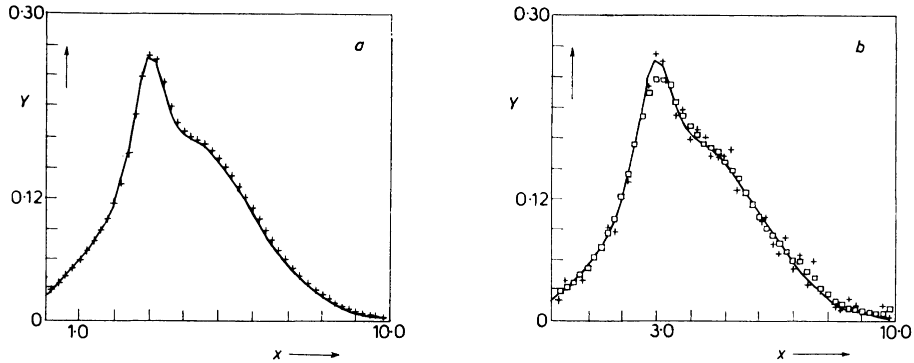

FIG. 6

Probability density function — parameters estimation. Asymmetrical case; + original data; □ filtered data; — fitted curve; *a* noise level 0·00 (Exp. 8A); *b* noise level 0·01 (Exp. 8B)

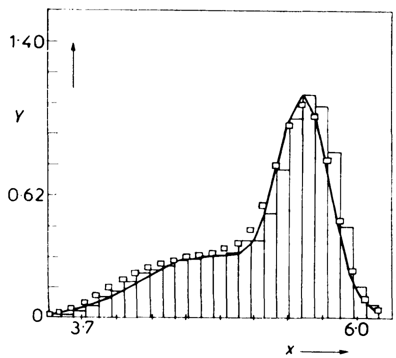

FIG. 7

Probability density function — parameters estimation. Experimental data from a fluidized bed reactor (Exp. 9); original data (histogram); □ filtered data; — fitted curve

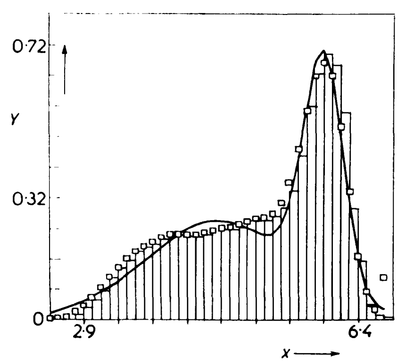

FIG. 8

Probability density function — parameters estimation. Experimental data from a fluidized bed reactor (Exp. 10); original data (histogram); □ filtered data; — fitted curve tion of the higher order moments from original data in order to provide a good accuracy of estimates. The procedure presented in this paper can be recommended for routine evaluation on desk calculators considering both the comparatively low demands on the memory capacity and the calculation speed requirre.

LIST OF SYMBOLS

x independent variable
η dependent model variable
y dependent regression line variable
Y dependent experimental data variable
e random disturbance
ε residue
C proportionality coefficient
μ mean value
σ standard deviation
F loss function

REFERENCES

1. Jones O. C., Delhaye M. J.: Int. J. Multiphase Flow *3*, 89 (1976),
2. Kölbel H., Beinheuer R., Langemann H.: Chem. Eng. Tech. *44*, 697 (1972).
3. Cohen A. C.: Technometrics *9*, 15 (1967).
4. Havlíček A., Čermák J., Gustavsson I.: Chem. Eng. J. *20*, 95 (1980).
5. Vrba J.: Chem. Prům. *29*, 449 (1979).
6. Himmelblau D. M.: *Process Analysis by Statistical Methods.* Wiley, New York 1970.