Chemical Engineering Journal 332 (2018) 303–311

Contents lists available at ScienceDirect

Chemical Engineering Journal

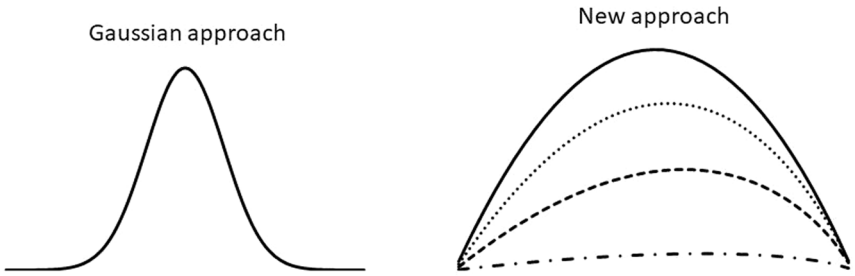
journal homepage: www.elsevier.com/locate/cej

A family of kinetic distributions for interpretation of experimental fluctuations in kinetic problems

Henrique Pacheco, Fabrício Thiengo, Martin Schmal, José Carlos Pinto*

Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, Cidade Universitária – CP: 68502, Rio de Janeiro 21941-972, RJ, Brazil

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Kinetic error distribution
Experimental error
Chemical kinetics
Confidence interval
Statistical analysis

ABSTRACT

The computation of confidence intervals frequently leads to arguable results due to lack of rigor when experimental errors are analyzed in kinetic experiments. Particularly, the usual Gaussian approach may not be adequate when the variable of interest is the reactant conversion, as this variable is constrained between very hard limits: 0 and 1. For this reason, the present work focuses on the development of analytical and numerical procedures for more accurate description of experimental errors in first-order reaction systems, which can be eventually extended to more complex reaction processes. Based on the proposed analytical and numerical schemes, new statistical distributions (named here as the kinetic distributions) can be derived to allow for more appropriate representation of conversion fluctuations and the respective statistical quantities, including the confidence intervals, which can be used more advantageously for analyses of kinetic data. In particular, it is shown that conversion errors are heteroscedastic, going through a point of maximum when conversion is allowed to increase from 0 to 1, and that confidence intervals are not symmetrical in respect to the averages, as assumed by Gaussian analyses.

1. Introduction

Mathematical models find valuable and widespread use in the field of catalysis. From the most fundamental theoretical aspects to the most complex reaction systems, they are used by researchers to enlighten reaction mechanisms, fit experimental data, and validate proposed hypotheses [1,2]. Besides, models are used in all sorts of kinetic studies, including very different experimental problems, such as chemical vapor deposition of carbon nanotubes, enzymatic assays, and sewage treatment [3–5].

Kinetic models depend on model parameters that are difficult (not to say impossible) to measure and must be inferred from available experimental data. Definition of model parameters is fundamental during model building because they describe the relative importance of distinct experimental effects on the analyzed process responses. Without proper determination of the model parameters, models become useless.

* Corresponding author.
E-mail address: pinto@peq.coppe.ufrj.br (J.C. Pinto).

http://dx.doi.org/10.1016/j.cej.2017.09.076
Received 16 June 2017; Received in revised form 9 September 2017; Accepted 11 September 2017
Available online 15 September 2017
1385-8947/

Nomenclature

a	Variability of catalytic activity
b	First parameter for the variability of concentration measurements
c	Second parameter for the variability of concentration measurements
C_i	Concentration of species i
C_i^e	Concentration of species i at equilibrium.
C_{A0}	Initial concentration of species A
C_A^m	Measured value of the concentration of species A
k_1	Specific reaction rate for the direction reaction
k_2	Specific reaction rate for the reverse reaction
k'	Simplified specific reaction rate
k'^m	Measured value of k'
N	Number of random numbers to be used in the numerical procedure
NE	Number of experiments
$P_{cum}(x_i)$	Cumulative probability of variable x at point i
S_x	Sample standard deviation of x
t	Time of reaction
x	Reaction conversion
x^m	Measured value of x
$\overline{x}$	Sample mean of x
ε_c	Experimental fluctuations of concentration
$\varepsilon_{k'}$	Experimental fluctuations of k'
$\varepsilon_C^{(1)}$	Error measurement at initial concentration
$\varepsilon_C^{(2)}$	Error measurement at equilibrium concentration
$\varepsilon_C^{(3)}$	Error measurement at sample concentration
ε_x	Conversion fluctuations
$\overline{\varepsilon_x}$	Measurement bias of x
$\wp_z(z)$	Probability distribution of variable z
σ_a^2	Variance of a
σ_a	Standard deviation of a
μ_a	Mean of a

When the proposed study requires the evaluation of model parameters, it is common to estimate the unknown parameters through minimization of an objective function (or metrics) that represents the distance between the values predicted by the model and the values obtained experimentally [2]. In order to formulate the objective function, it is usually necessary to postulate several simplifying assumptions, either to reduce the experimental load or simply because it is impossible to know beforehand all features of a particular experimental system. Most times, however, these assumptions are not validated or verified, which may lead to inappropriate conclusions and parameter values.

In order to illustrate this point, Schwaab et al. [6] showed that the confidence region of parameter values may present very peculiar shapes when the nonlinear behavior of model responses and experimental measurements are considered. Such region can be quite distinct from the hyper-ellipsoidal shape commonly assumed for confidence regions, as obtained for linear model responses and Gaussian distribution of measurement fluctuations. Similarly, Schwaab et al. [7–9] showed that improper model parameterization may lead to highly correlated model parameters, which can significantly degrade the final model performance. Finally, it is usually assumed that fluctuation errors are constant throughout the experimental region, even though this cannot be supported by independent error analyses, as discussed by Alberton et al. [10].

Certainly, one of the commonest assumptions is the Gaussian distribution for the experimental fluctuations. This assumption is so popular that many do not understand that, although the Gaussian distribution is a model that may find suitable applications in numerous situations, it may also be inadequate for the interpretation of several other practical problems [2]. For this reason, the proper characterization of experimental fluctuations can be of paramount importance for correct assessment of experimental analyses and interpretation of kinetic models. Despite that, the detailed characterization of experimental fluctuations in kinetic studies is often neglected, due to difficulty to investigate how these unavoidable fluctuations depend upon the reaction conditions, among other reasons. Particularly, the behavior of the experimental fluctuations can be linked to specific characteristics of the experimental system, including measuring techniques and operation procedures, naturally causing this type of investigation to be challenging [2].

As experimental fluctuations are related to uncontrolled random causes, the proper characterization of experimental variability requires replication of experimental trials a sufficiently high number of times [11], discouraging the fundamental investigation of fluctuations and explaining why certain statistical distribution models are needed and assumed to be valid *a priori* during an experimental investigation. Particularly, the Gaussian distribution is useful because it may be applied to a large array of physical problems, provides an asymptote for problems dominated by infinitely many sources of variability and requires the definition of only two parameters (mean and variance) for its use. Moreover, the Gaussian distribution is mathematically tractable, allowing for a number of important theoretical developments, which include the derivation of t-Student, F-Fisher, and chi-square tests for analyses of means and variances of experimental data samples [2,11].

When the Gaussian distribution is used to define boundaries for conversion and selectivity measurements in kinetic studies, however, anomalous results may be obtained. For instance, confidence intervals may lie outside the [0,1] interval (which makes no physical sense), because the Gaussian distribution is defined in the infinite domain, while conversions and selectivities lie in the much narrower finite interval [0,1]. An obvious conclusion is that conversion and selectivity measurements do not follow the Gaussian distribution, although it may be true that this distribution may provide useful fits for experimental fluctuations in certain experimental systems.

Based on the previous paragraphs, the present work focuses on the development of analytical and numerical procedures for more accurate description of experimental errors in first-order reaction systems. Based on the proposed analytical and numerical schemes, new statistical distributions (named here as the *kinetic distributions of fluctuation measurements*) are derived to allow for more appropriate representation of conversion fluctuations and the respective statistical quantities, including the confidence intervals, which can be used more advantageously for analyses of kinetic data. In order to do that, a first-order reaction is assumed to take place in a model reacting system, as several systems can be represented with good accuracy by first-order reaction models. Besides, it is well known that more complex nonlinear functions can be represented locally by simpler models, especially when experimental fluctuations are not too large [12]. Nevertheless, as shown in the proposed numerical development, this underlying assumption does not constitute a major drawback of the proposed analysis, for the first-order reaction rate assumption can be easily relaxed in more involving numerical analyses in order to represent more complex reaction systems.

Stochastic methods are largely applied in Chemical Engineering, both as an alternative for finding global maxima and as an elegant, efficient way to validate simulations [13–15]. What is proposed in this work is, though, to explore the error distributions themselves via a stochastic approach, something that is seldom seen in the scientific literature. This is of the utmost relevance since errors in the most fundamental variables of the problem must have their statistical