Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physica A 352 (2005) 379–396
PHYSICA A
www.elsevier.com/locate/physa

Application of the generalized mean value function to the statistical detection of water in decane by near-infrared spectroscopy

Raoul R. Nigmatullin[a], Adam Moroz[b], Geoff Smith[b,*]

[a]*Theoretical Physics Department, Kazan State University, Kremlevskaya Str.18, Kazan, Tatarstan, Russia*
[b]*School of Pharmacy, De Monfort University, The Gateway, Leicester LE1 9BH, UK*

Received 15 March 2004; received in revised form 29 November 2004
Available online 26 January 2005

Abstract

The generalized mean value (GMV) function, defined as $G_N^{(p)} = (\Delta_N^{(p)})^{1/p}$(where $\Delta_N^{(p)}$ is the absolute value of the moment of the pth order), was used here to differentiate between statistically close random sequences or those sequences containing large numbers of measured points ($N \gg 1$). The approach taken was to find (with the help of the eigen-coordinates method) the approximate analytical function for any $G_N^{(p)}$, and in so doing demonstrate that it is inherently possible to express *quantitatively* the reduced characteristics of *any* random sequence, in terms of the 'universal' set of fitting parameters defined by this function. The introduction of a 'universal' set of the reduced parameters in the moment space then provides the instrument for the comparison of different random sequences. Applications for this new method are evident for many branches of the analytical sciences, but especially in cases when visual 'labels' (e.g. resonance lines), which serve as an indication of the presence of an additive, are either absent or 'contaminated' strongly by *noise*. Those fitting parameters from the approximate analytical expression, which depend on the concentration of the small additive can then be used for the construction of the quasi-monotonic line, defined as the *calibration curve*. Real experiments based on the treatment of near-infrared (NIR) spectra obtained for decane (the initial matrix) with water (the additive) confirm the efficiency of this simple approach. In contrast, the more conventional statistical method, based on cluster analysis, failed to establish the desired calibration curve. This simple and universal approach, which is

*Corresponding author.
E-mail address: gsmith02@dmu.ac.uk (G. Smith).

doi:10.1016/j.physa.2004.12.055

free from model assumptions, can be used for any set of random sequences (e.g. spectrograms) if it is necessary to compare them *quantitatively* with each other.

PACS: 02.60.Ed, Pn; 06.20.Dk; 07.05.Kf, Rm

Keywords: A. Generalized mean value; A. Higher moments; B. Random sequences; C. Eigen-coordinates method

1. Introduction

Almost all statistical approaches used in spectroscopy, such as cluster analysis, principal component analysis (PCA), and discriminant analysis (DA), are based on the use of first- and second-order statistical moments, see, e.g. recent publications [1–5]. The statistics based on the usage of higher-order moments are to all intents and purposes not used in this field, though certain applications for improving the quality of image processing have been developed [6,7]. In particular, in paper [7], an algorithm for an object recognition, based on the invariants derived from the second- and third-order moments, has been developed.

Common use of lower moments includes the well-known average value and standard deviation (which are based on the calculation of the first and the second moments, respectively) and the third and the fourth moments (which are related to the calculation of the skewness and kurtosis). The skewness and kurtosis are defined as dimensionless moments, in contrast to the mean and standard deviation, which have the same dimensions as the measured quantities. The skewness characterizes the degree of asymmetry of the distribution around its mean value. A positive (or negative) value of the skewness implies a distribution with a high number of large (or smaller) values than would be expected from a Gaussian distribution. The kurtosis measures the relative sharpness or flatness of a distribution, compared to a normal distribution with the same mean and standard deviation; the larger the kurtosis, the sharper the distribution. As far as we know, the systematic use of the statistics based on higher moments, with order $p>4$, is *not* known.

Attentive analysis shows that for the characterization of random sequences, and their comparison with each other, one can use the space of higher moments, whereby the number of moments is determined by the plateau region of the generalized mean value (GMV) function, as defined below. Moreover, by working within the space of moments, one can transform *any* random sequence to a smooth function. This smooth function can be approximated with very high accuracy by an analytical function, which contains a finite number of fitting parameters. This 'universal' set of fitting (i.e. reduced) parameters, identified with the help of the eigen-coordinates (ECs) method, represents as a specific 'fingerprint' (or replica) of the random sequence considered. This identification enables one to find differences between random sequences of *any* nature, and does not require any prior supposition, e.g. that the noise conforms to Gaussian statistics.