MATHEMATICAL AND STATISTICAL MODELS IN THE BIOSCIENCES RESEARCH GROUP

Albert Sorribas
Departament de Ciències Mèdiques Bàsiques
(Bioestadística)
Universitat de Lleida

A complex system, such as a metabolic pathway, a physiological network or a population, requires developing specific tools for analyzing its properties. Mathematical models deal with such complexity by building-up approximate descriptions that can be used to critically inspect system's properties. Among other possibilities, ordinary differential equations models (ODE) provide an appropriate framework from elementary chemical reactions to population dynamics. A critical step in defining appropriate ODE models is to choose a suitable formalism for representing the underlying processes. If the goal is to analyze the whole system, a mathematical description based on a too detailed representation of individual processes may be inappropriate and may lead to unpractical models. Our group, established since 1993 in the Department of Basic Medical Sciences of the University of Lleida, has specialized in a specific approach known as *power-law formalism*. This formalism provides an approximate representation for non-linear functions and leads to mathematical models that can be used for characterizing large-system's properties. Basically, this formalism is based on a Taylor series approximation in logarithmic coordinates, which provides a power-law approximation in Cartesian coordinates. As a result, the models based in this formalism provide a good representation of the non-linear properties of the underlying processes and yet they are easily analyzable both algebraically and numerically. Modeling issues that have been addressed using this formalism encompases from regulatory networks to modeling statistical distributions. An up-to-date information on published papers and last results can be found at the web page: www.udl.es/usuaris/q3695988/WebPL/main.htm.

Development of the power-law formalism for system modeling and analysis: One of the main concerns in using mathematical models is parameter estimation. In metabolic systems experimental data come from different sources and design of optimal experiments for parameter estimation is seldom possible. In such cases, alternative strategies must be used, including the possibility of mixing both experimental and qualitative data based on published observations and on expertise. Issues concerning parameter estimation either from steady-state measurements or from dynamic observations and a method for identifying the regulatory estructure of a pathway can be found in: Sorribas *et al*, 1993; Cascante, Sorribas and Canela, 1994; Sorribas and Cascante, 1994. Formally, the power-law formalism can be related to other existing formalisms such as Generalized Lotka Volterra models. Within metabolic studies, a common alternati-

ve is to use the Metabolic Control Analysis approach. Comparison between different formalisms can be found in: Curto, Sorribas and Cascante, 1995; Soribas, Curto and Cascante, 1995; Cascante, Curto and Sorribas, 1995. More recently, we have turned to develop a new approach based on least-squares criteria as a definition for the power-law formalism. This new approach may be more suitable for model definition from steady-state measurements (Hernández-Bermejo, Fairén and Sorribas, 1999, 2000; Sorribas and Hernández-Bermejo, 2002).

Modeling and analyzing large systems: One of the critical steps in using a mathematical model is validation. This is specially important in models that are build using different sources of information. The power-law formalism provides tools for a systematic validation in large models. These tools are related to sensitivity analysis and a sound knowledge of the biological framework (Curto *et al*, 1997, 1998). Tools for evaluating the effect of parameter uncertainties on large models can be found in: de Atauri, Sorribas and Cascante, 2000. Current work on modeling and analyzing large systems through the use of these techniques in our group includes: (*i*) A model of cell cycle in yeast, with inclusion of stochastic effects related to cellular compartimentation, (*ii*) Bridging the gap between Genomics and Physiology through the development of a model based on yeast response to oxidative stress and data from DNA-microarrays, and (*iii*) A model of the spread of HIV infection in our community.

S-distributions: The power-law formalism can be used to define approximate models in many instances. One of the most interesting applications is the modeling of statistical distributions by means of a family of distributions known as S-distribution (Voit, 1992). We have studied some important characteristics of this family, including its potential use as general method for generating a random sample of a given distribution (Hernández-Bermejo and Sorribas, 2001) and the representation of dynamic changes in distribution associated to a growth process (Voit and Sorribas, 2000). From a practical point of view, we have developed a method for estimating conditional S-distributions (Sorribas, March and Voit, 2000; March *et al.*, 2002) and for Receiver Operating Characteristic curves in medical diagnostic problems (Sorribas, March and Trujillano, 2002). A numerical method for maximum-likelihood estimation for S-distributions has also been defined (March *et al.*, 2002). Our current research aims to generalize the S-distribution family, to further develop numerical tools for random number generation based on this family, and to define hypothesis tests based on S-distributions.

References

1. Cascante, M., Curto, R. and Sorribas, A. (1995). «Comparative characterization pathway of Saccharomyces cerevisiae using Biochemical Systems Theory and Metabolic Control Analysis: Steady-state analysis». *Mathematical Bioscience*, 130, 51-69.

2. Cascante, M., Sorribas, A. and Canela, E. I. (1994). «Enzyme-enzyme interactions and metabolite channelling: alternative mechanisms and their evolutive significance». *Biochemical Journal*, 298, 313-320.
3. Curto, R., Sorribas, A. and Cascante, M. (1995). «Comparative characterization pathway of Saccharomyces cerevisiae using Biochemical Systems Theory and Metabolic Control Analysis: Model definition and nomenclature». *Mathematical Biosciences*, 130, 25-50.
4. Curto, R., Voit, E. O., Sorribas, A. and Cascante, M. (1997). «Validation and steady-state analysis of a power-law model of purine metabolism». *Biochemical Journal*, 324, 761-775.
5. Curto, R., Voit, E. O., Sorribas, A. and Cascante, M. (1998). «A power-law model of purine metabolism: comparison of alternative modelling strategies». *Mathematical Biosciences*, 151, 1-49.
6. de Atauri, P., Sorribas, A. and Cascante, M. (2000). «Evaluation of the effect of boundary model assumptions». *Biotechnology Bioengineering*, 68, 18-30.
7. Hernández-Bermejo, B. and Sorribas, A. (2001). «Analytical quantile solution for the S-distribution, random number generation and statistical data modelling». *Biomedical Journal*, 43, 1017-1025.
8. Hernández-Bermejo, B., Fairén, V. and Sorribas, A. (1999). «Power-law modelling based on least-squares minimization criteria». *Mathematical Biosciences*, 161, 83-97.
9. Hernández-Bermejo, B., Fairén, V. and Sorribas, A. (2000). «Power-law modelling based on least-squares criteria: consequences for systems analysis and simulation». *Mathematical Biosciences*, 167, 87-107.
10. March, J., Trujillano, J., Tort, M. and Sorribas, A. (2002). «Estimating conditional distributions using a method based on S-distributions: Reference percentile curves for body mass index in Spanish Children». *Growth Development and Aging* (submitted for publication).
11. Sorribas, A. and Cascante, M. (1994). «Structure identifiability in metabolic pathways: parameter estimation in models based on the power-law formalism». *Biochemical Journal*, 298, 303-311.
12. Sorribas, A. and Hernández-Bermejo, B. (2002). «Modeling a metabolic pathway from steady-state measurements on the intact system: on the usefulness of least-squares derived power-law models». *Mathematical Biosciences* (submitted for publication).
13. Sorribas, A., Curto, R. and Cascante, M. (1995). «Comparative characterization pathway of *Saccharomyces cerevisiæ* using Biochemical Systems Theory and Metabolic Control Analysis: Model validation and dynamic behaviour». *Mathematical Biosciences*, 130, 71-84.
14. Sorribas, A., March, J. and Trujillano, J. (2002). «A new parametric mtehod based on S-distributions for computing Receiver Operating Characteristic curves for continuous diagnostic tests». *Statistics in Medecine* (in press).

15. Sorribas, A., March, J. and Voit, E. O. (2000). «Data modelling using S-distributions: Utility in estimating age-related trends in cross-sectional studies». *Statistics in Medecine*, 19, 697-713.
16. Sorribas, A., Samitier, S., Canela, E. I. and Cascante, M. (1993). «Metabolic pathway characterization from transient response data obtained in situ: parameter estimation in S-system models». *Journal of Theoretical Biology*, 162, 81-102.
17. Voit, E. O. (1992). «The S-distribution: A tool for approximation and classification of univariate, unimodal probability distributions». *Biomedical Journal*, 7, 855-878.
18. Voit, E. O. and Sorribas, A. (2000). «Computer modeling of dynamically changing distributions of random variables». *Mathematical and Computer Modelling*, 31, 217-225.