Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 132 (2005) 33–51
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Modeling disease dynamics and survivor functions by sanogenesis curves

A.G. Bart[a,*], V.A. Bart[b], A. Steland[c], M.L. Zaslavskiy[a]

[a] *Department of Mathematics and Mechanics, St. Petersburg University, Universitetskaya av.26 Petrodvoretz, St. Petersburg 198904, Russia*
[b] *Research Institute of Cardiology, St. Petersburg, Russia*
[c] *Ruhr-Universitat Bochum, Germany*

Available online 11 September 2004

Abstract

We propose to analyse the development of a disease by sanogenesis curves which model the result of interacting exitatory and inhibitory factors on a disease. Assuming that high-dimensional data describing the disease course are driven by a latent complex-valued Gaussian process with Markovian structure, we can identify the sanogenesis curve as the real part of the covariance function of the latent process. By applying techniques of stochastic process theory and partially inverse functions theory this finding allows to estimate the model parameters. In addition, the sanogensis curve also suggests a new model for survival times, where failures (deads) are only observed during critical time periods (crises) defined by the sanogenesis curve. We illustrate our approach by analyzing two real data sets from medicine.

Keywords: Sanogenesis curve; Survivor function; Partially inverse function; Factor analysis; Correlation function; Critical and latent time points; Confidence regions

1. Introduction

This paper is devoted to a new approach to analyze the development of a disease and related survivor functions. Our research is primarily motivated by medical applications where often high-dimensional data about the development of a disease is given and should be used in the analysis. Given a generic survival time T, i.e., a non-negative random variable

* Corresponding author
E-mail address: agb@ab2119.spb.edu (A.G. Bart).

0378-3758/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.jspi.2004.06.014

(r.v.), the most important statistical quantities are the survivor function defined by $H(t) = P(T > t), t \in \mathbb{R}$ and the hazard function $h(t) = -\mathrm{d}\ \ln\ H(t)/\mathrm{d}t$. There exists a rich literature discussing approaches to model these quantities. We refer to Barlow and Proschan (1975), Cox and Oakes (1984), Gavrilov and Gavrilova (1991), Klein and Goel (1992), and the references given there. These classic approaches share the following features, which may be unpleasant for certain applications.

First, most classic approaches assume smooth and rather simple survivor functions. Medical data of disease processes are often characterized by a discrete-continuous character reflecting the dynamics of decay periods and latent periods. In particular, these latent periods play an essential role in the dynamics of many diseases sometimes determining their course, see Bart et al. (1980). One may consider non-parametric models, e.g., the Kaplan–Meier estimates of the survivor function, cf. Cox and Oakes (1984) and Glantz (1994), but in many cases an underlying parametric model taking account of discontinuities and latent periods is more appealing.

Second, survivor functions describe only the result of a complex regulatory interaction process of the involved (biological) system. In the sequel we call these underlying processes basic processes. The general problem, however, has an integral character. Our aim is to provide a relatively detailed framework for the system behavior which drives the distribution of the observed survival times and additional variables which describe the disease dynamics. In real applications the survivor functions are usually accompanied by rich information about the basic processes. In that case, the medical data usually consists of high-dimensional basic information about the patient's disease development, e.g., blood measurements, heart rate, CD4 cell counts playing an important role in the immune system, etc. Consequently, we have a large number of additional observed variables for each patient. Traditionally, that kind of information is considered as additional data and is taken into account in the form of explanatory variables as in the proportional hazard model, see Cox and Oakes (1984). In that model the shape of the hazard function remains the same, but the initial function $h_0(t)$ is multiplied by a factor, $\psi(\beta, \mathbf{z})$, depending on a parameter vector β (so that $\psi(0, \mathbf{z}) = 1$) and on a vector $\mathbf{z}$ of explanatory variables (covariates), which alone contains all the additional information, yielding the model $h(t) = \psi(\beta, \mathbf{z})h_0(t)$.

The present paper proposes a completely different approach to deal with these problems. Following Bart et al. (1980), Bart and Alekseyeff (1999), and Bart (2002) a parametric model of the survivor function is formulated, which is based on an analysis of the dynamics due to the main regulatory interactions in the disease course. In a first stage we propose to perform a factor analysis to reduce the complexity of the variables describing the disease course to two main factors. Factor analyses are quite common in medicine, see e.g. Lloyd and Ledermann (1984). The resulting two-dimensional process is used to estimate a parametric model for the sanogensis curve. We regard that function as the main subject of an analysis of a disease, and as an important starting point to derive survivor functions. In medical–biological systems the sanogenetic (inhibitory) factors retard a pathological process driven by exitatory factors. When confronted with a disease, an organism's immune systems initiates regulatory mechanisms. The sanogensis functions describes the dynamics of these interactions and has therefore a medical interpretation.

Our approach is to derive a survivor function directly from an (estimated) sanogenesis curve by a transformation which essentially puts an upper bound on those oscillation of