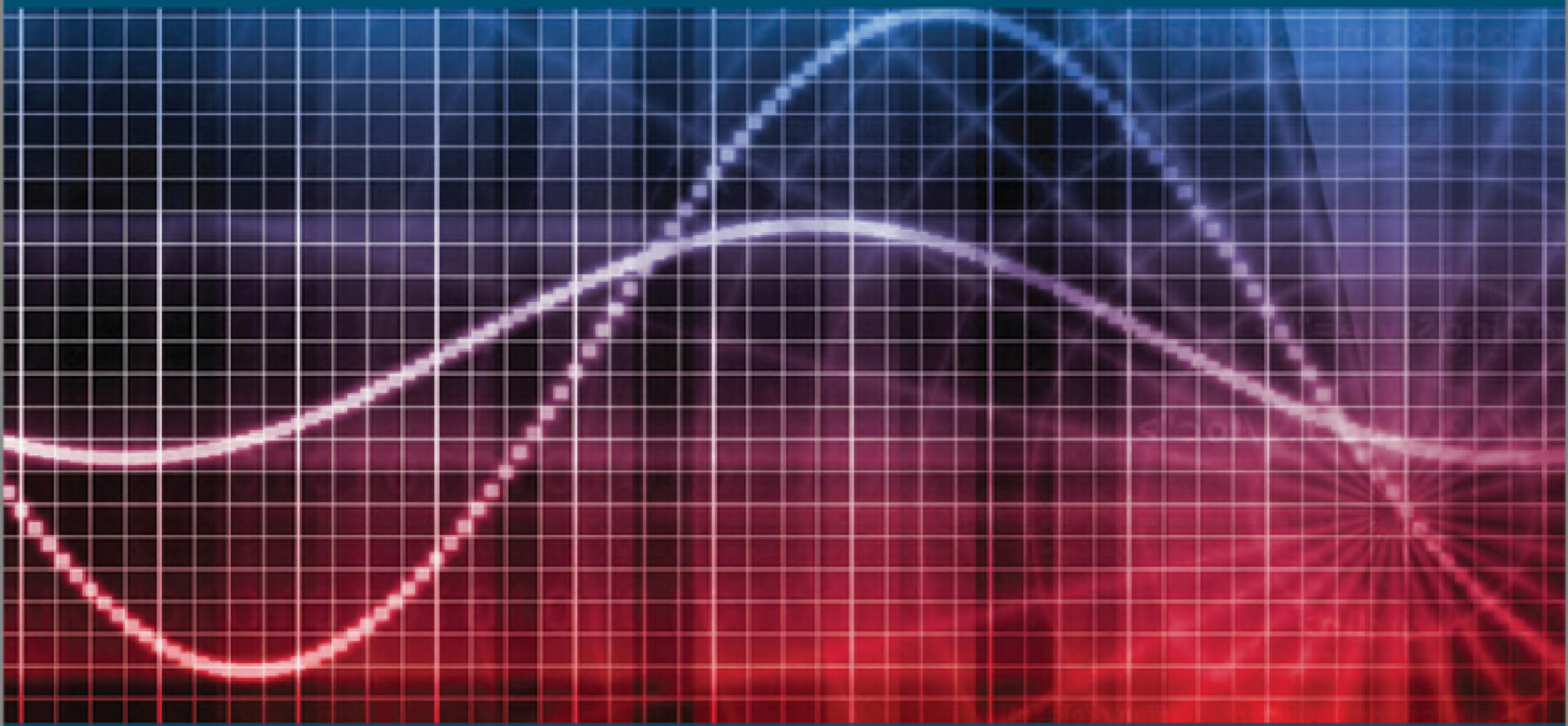
MATHEMATICS AND STATISTICS SERIES



Statistical Models and Methods for Reliability and Survival Analysis

Edited by
Vincent Couallier, Léo Gerville-Réache
Catherine Huber-Carol, Nikolaos Limnios
and Mounir Mesbah

iSTE

WILEY

Statistical Models and Methods
for Reliability and Survival Analysis

Statistical Models and Methods for Reliability and Survival Analysis

Edited by
Vincent Couallier
Léo Gerville-Réache
Catherine Huber-Carol
Nikolaos Limnios
Mounir Mesbah

Series Editor
Nikolaos Limnios

iSTE

WILEY

First published 2014 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Control Number: 2013951426

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-619-8

MIX
Paper from responsible sources
FSC® C013604
FSC www.fsc.org

Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY

Table of Contents

Preface

Statistical Models and Methods for Reliability and Survival Analysis is a volume of contributions by specialists in statistical theory and their applications that provide up-to-date developments in methods used in survival analysis, statistical goodness of fit, stochastic processes for system reliability among others. Many of them are related to the work of Professor M. Nikulin in statistics which has spanned 30 years. The contributors accepted this challenging project to gather various contributions with a wide range of techniques and results, all of them on the topics of the past S2MRSA conference, 4–6th July 2012, dedicated to M. Nikulin for his 20th anniversary as a professor in the Bordeaux Segalen University (http://www.sm.u-bordeaux2.fr/Stat.Bordeaux.2012). The book is intended for researchers interested in statistical methodology and models useful in survival analysis, system reliability and statistical testing for censored and non-censored data.

Vincent COUALLIER
Léo GERVILLE-RÉACHE
Catherine HUBER-CAROL
Nikolaos LIMNIOS
Mounir MESBAH
October 2013

Biography of Mikhail Stepanovitch Nikouline

From his native city of St. Petersburg, whose name was Leningrad when he was born there on April 29, 1944, Mikhail Stepanovitch Nikulin (Nikouline by French passport) visited many foreign countries due to his interest in several different domains in his favorite disciplines: mathematics and statistics. It was Professor Login N. Bolshev who allowed him to be aware of his talent in mathematics and directed his thesis in the field of "Probability Theory and Mathematical Statistics".

He obtained his Master's degree in mathematics from the State University in St. Petersburg in 1966. In the meanwhile, he had met Valia, a young lady who had many admirers, and he was proud and happy to be the one she chose to be her husband. They married in early 1963 when they were both under 20. Valia was a scientist like her husband and they had a daughter named Hélène at the end of the same year.

Then they decided to learn French intensively in order to go to Brazzaville, capital of the "République Populaire du Congo", where they spent three years. There, both were professors of mathematics in the *Collège d'Enseignement Général* from 1966 to 1969. Just after leaving Congo, their second child Alexis was born; both children were very promising. Then they went back to their native country. There, M.S. Nikouline became a PhD student at the Steklov Mathematical Institute in Moscow, Academy of Science of Russia, while Valia was doing research in mathematics at the Department of Applied Mathematics at the State University in St. Petersburg. Under the supervision of Professor L.N. Bolshev, for whom Nikouline always had a great admiration for both his mathematical skill and his personal human qualities, and whose early death at the age of 56 saddened Nikouline greatly, he obtained his PhD thesis entitled "A generalization of chi square tests".

Chapter written by Vincent COUALLIER, Léo GERVILLE-RÉACHE, Catherine HUBER-CAROL, Nikolaos LIMNIOS and Mounir MESBAH.

From 1974 to 2006, he was a member of the famous Ildar Ibragimov's Laboratory of Statistical Methods at the Mathematical Institute of V. Steklov (St. Petersburg).

From 1988 to 1992, M.S. Nikouline was a docent in the Department of Statistics and Probability at the State University in St. Petersburg. In 1992, he was hired as an Associate Professor at the University Victor Segalen-Bordeaux 2, France, where he spent his career serving in various positions: Professor in 1996, Full Professor in 2006 and Distinguished Professor in 2008 until 2012 when he became Emeritus. During the same time, from 1996 to 2001, he was the head of the Department "SCIMS- Science et Modélisation", and, from 1999 to 2006, he was the head of the Laboratory "Equipe d'Accueil 2961: Mathematical Statistics and its Applications".

He published a large number of papers (more than 250), alone for most of them, or jointly, with several colleagues and friends from the USSR, Europe and Canada, such as V. Solev, V. Bagdonavicius, V.G. Voinov, N.N. Lyashenko, K. Dhzapridze, P. Greenwood, L. Gerville-Réache, J. Kruopis and many others.

His first field of interest, following the theme of his PhD thesis, was goodness-of-fit tests and best unbiased estimators for parametric models. Between 1973 and 1992, he published more than 30 papers on this subject.

From 1993 to 2013, his subjects of interest became more diversified. We can see from the references of Nikulin's papers that although he was interested in parametric models, goodness of fit for them and unbiased estimators, he also tackled semi-parametric and non-parametric models that are more general. In particular, he was interested in human survival data analysis and he was one of the few people who made the link with reliability whose traditional field is industry that seemed for a long time to be very much apart from the medical field. From a mathematical point of view, many advances in any of these two domains could be provided by the other domain. With this perspective in mind, M.S. Nikouline organized with three French colleagues M. Mesbah of the University of Vannes (Britanny) and then University of Paris VI, N. Limnios of the University of Technology of Compiègne and C. Huber-Carol of the University of Paris Descartes, a European Seminar devoted to the "Survival Analysis and Reliability Theory".

During his career in Bordeaux, Nikouline has directed 11 PhD theses and most of his students have obtained good positions either in universities or in industry.

Moreover, Nikouline was the author and editor of a number of volumes on mathematical statistics. He has also written several articles for the *Encyclopaedia* and was part of several editorial boards and scientific societies such as the International Statistical Institute (ISI) and the Mathematical Society of St. Petersburg. Throughout his long career, he has met a large number of colleagues and as he is a friendly and generous person, many of them became his friends and were always very faithful to him.

PART 1

Statistical Models and Methods

Chapter 1

Unidimensionality, Agreement and Concordance Probability

The evaluation and comparison of various methods often arise in medical research. For example, the evaluation of reproducibility of a new measurement technique often needs a comparison with the established technique, and image interpretation is often read by two or more observers. In this chapter, we provide a review of the measures of agreement and association, describe the statistical models underlying the Cronbach's alpha coefficient (CAC) and the backward reliability curve (BRC), the kappa coefficient, and present a general approach based on the concept of concordance probability. In particular, we illustrate the relationship between the concordance probability and various existing measures of agreement and association, namely Kendall's τ, Somer's D, area under receiver operating characteristic (ROC) curve and Harrell's c-index. In addition, we review the estimation of concordance probability and present its large sample properties. Recent developments in the analysis of right censored data are also presented.

1.1. Introduction

The evaluation and comparison of various methods often arise in medical research. For example, the evaluation of reproducibility of a measurement technique often needs a comparison with the established technique, and the interpretation of a computerized tomography (CT) or magnetic resonance imaging (MRI) scan is often read by two or more observers. There is considerable literature on the measure of

---

Chapter written by Zhezhen JIN and Mounir MESBAH.

agreement (see [CHO 04], [BAR 07], [WAT 10], [SHO 04] and [LIN 10]). The methods vary with different types of measurement, i.e. continuous or categorical measurements. When the response variable is continuous, there are several intuitive approaches, namely comparison of means, Cronbach's coefficient alpha (CAC), various correlation coefficients and the test of slope being 1 in a simple linear regression, as well as alternative methods, the limits of agreement [BLA 86, BLA 99], the concordance correlation coefficient [LIN 89], mean squared deviation and total deviation index [LIN 00], and coverage probability approach [LIN 02]. When the response variable is categorical, kappa statistic, Somer's D-statistic and logistic regression are commonly used. When one measure is binary and the other measure is continuous, the methods of the ROC curve and logistic regression approach are often applied. These methods are related to typical concordance correlation between repeated measurements through an underlying linear or nonlinear parametric model. Recently developed concordance probability is a non-parametric approach. The concordance probability is commonly used as a measure of discriminatory power and predictive accuracy of statistical models. We show that the concordance probability also provides a unified measure of agreement for different types of measurement.

In this chapter, we present a review of the statistical models underlying the CAC and the BRC in section 1.2, and the kappa coefficient in section 1.3. In section 1.4, we introduce the concordance probability and describe its relationship with Kendall's τ, Somer's D and area of ROC curve of sensitivity and 1–specificity for different cutoffs. In section 1.5, we review the estimation of concordance probability and present its large sample properties. In section 1.6, we present recent developments on how to use the concordance probability to assess the agreement among different measures. We present the extension of the approach to the right censored data in section 1.7 and conclude with some discussion in section 1.8.

1.2. From reliability to unidimensionality: CAC and curve

1.2.1. *Classical unidimensional models for measurement*

Latent variable models involve a set of observable variables $A = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k\}$ and a latent (unobservable) variable θ of dimension $d \leq k$. In such models, the dimensionality of A is captured by the dimension of θ, the value of d. When $d = 1$, the dimensionality of set A is called unidimensional.

In a health-related quality of life (HrQoL) study, measurements are taken with an instrument: the questionnaire, which consists of questions (or items). In such cases, the $\mathbf{X}_{ij}$ represents the random response of the jth question by the ith subject and the $\mathbf{X}_j$ denotes the random variable generating responses to the jth question.

The parallel model is a classical latent variable model describing the unidimensionality of a set $A = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k\}$ of quantitative observable variables. Let $\mathbf{X}_{ij}$ be the measurement of subject i, given by a variable $\mathbf{X}_j$, $i = 1, \ldots, n, j = 1, \ldots, k$, then:

$$\mathbf{X}_{ij} = \tau_{ij} + \varepsilon_{ij}, \quad [1.1]$$

where τ_{ij} is the unknown true measurement corresponding to the observed measurement $\mathbf{X}_{ij}$ and ε_{ij} a measurement error. The model is called a parallel model if the τ_{ij} can be divided as:

$$\tau_{ij} = \beta_j + \theta_i,$$

where β_j is an unknown fixed parameter (non-random) representing the effect of the jth variable, and θ_i is an unknown random parameter effect of the ith subject.

It is generally assumed that θ_i has zero mean and unknown standard deviation σ_θ. It should be noted that the zero-mean assumption is an arbitrary identifiability constraint with consequence on the interpretation of the parameter: its value must be interpreted comparatively to the mean population value. *In HrQoL setting, θ_i is the true latent HrQoL that the clinician or health scientist wants to measure and analyze*. It is a zero mean individual random part of all observed subject responses $\mathbf{X}_{ij}$, the same whatever the variable $\mathbf{X}_j$ (in practice, a question j of an HrQoL questionnaire). It is also generally assumed that ε_{ij} are independent random errors with zero mean and standard deviation σ corresponding to the additional measurement error. Moreover, the true measure and the error are assumed to be uncorrelated, i.e. $cov(\theta_i, \varepsilon_{ij}) = 0$. This model is known as the parallel model, because the regression lines relating any observed item $\mathbf{X}_j$, $j = 1, \ldots, k$, and the true unique latent measure θ_i are parallel.

Model [1.1] can be obtained in an alternative way through modeling the conditional moments of the observed responses. Specifically, the conditional mean of $\mathbf{X}_{ij}$ can be specified as:

$$E[\mathbf{X}_{ij}|\theta_i; \beta_j] = \beta_j + \theta_i, \quad [1.2]$$

where β_j, $j = 1, \ldots, k$, are fixed effects and θ_i, $i = 1, \ldots, n$, are independent random effects with zero mean and standard deviation σ_θ. The conditional variance of $\mathbf{X}_{ij}$ is specified as:

$$Var[\mathbf{X}_{ij}|\theta_i; \beta_j] = Var(\varepsilon_{ij}) = \sigma^2. \quad [1.3]$$

Assumptions [1.2] and [1.3] are classical in experimental design. The model defines relationships between different kinds of variable: the observed score $\mathbf{X}_{ij}$, the true score τ_{ij} and the measurement error ε_{ij}. It is significant to make some remarks about the assumptions underlying this model. The random part of the true measure given by response by the ith individual does not vary with the question number j as the θ_i does not depend on j, $j = 1, \dots, k$. The model is unidimensional in the sense that the random part of all observed variables (questions $\mathbf{X}_j$) is generated by the common unobserved variable (θ_i). More precisely, let $\mathbf{X}^*_{ij} = \mathbf{X}_{ij} - \beta_j$ be the calibrated version of the response to the jth item by the ith subject, then models [1.2] and [1.3] can be rewritten as:

$$E[X^*_{ij}|\theta_i; \beta_j] = \theta_i, \text{ for } \forall j, \qquad [1.4]$$

along with the same assumptions on β and θ and the conditional variance model [1.3].

When both θ_i and ε_{ij} are normally distributed, then we have the so-called conditional independence property: whatever j and j', two observed items $\mathbf{X}_j$ and $\mathbf{X}_{j'}$ are independent conditional to the latent θ_i.

1.2.2. *Reliability of an instrument: CAC*

A measurement instrument yields values that we call the observed measure. The reliability ρ of an instrument is defined as the ratio of two variances of the true over the observed measure. Under the parallel model, we can show that the reliability of any variable $\mathbf{X}_j$ (as an instrument to measure the true value) is given by:

$$\rho = \frac{\sigma^2_\theta}{\sigma^2_\theta + \sigma^2}. \qquad [1.5]$$

This coefficient is also known as the intra-class coefficient. The reliability coefficient, ρ, can easily be interpreted as a correlation coefficient between the true measure and the observed measure. When the parallel model is assumed, the reliability of the sum of k variables is:

$$\tilde{\rho}_k = \frac{k\rho}{k\rho + (1-\rho)}. \qquad [1.6]$$

This formula is known as the Spearman–Brown formula [BRO 10, SPE 10].

The Spearman–Brown formula shows a simple relationship between $\tilde{\rho}_k$ and k, the number of variables. It is easy to see that $\tilde{\rho}_k$ is an increasing function of k.

The maximum likelihood estimator of $\tilde{\rho}_k$, under the parallel model with normal distribution assumptions, is known as CAC [CRO 51, BLA 97], which is denoted as α:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} S^2_j}{S^2_{tot}}\right), \qquad [1.7]$$

where

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X_j})^2$$

and

$$S_{tot}^2 = \frac{1}{nk-1}\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij} - \overline{X})^2.$$

Under the parallel model, the variance–covariance matrix of the observed items X_j and the latent trait θ is:

$$V_{X,\theta} = \begin{pmatrix} \sigma_\theta^2 + \sigma^2 & \sigma_\theta^2 & \cdots & \cdots\sigma_\theta^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 + \sigma^2 & \sigma_\theta^2 & \cdots\sigma_\theta^2 & \sigma_\theta^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_\theta^2 & \cdots & \sigma_\theta^2 & \sigma_\theta^2 + \sigma^2 & \sigma_\theta^2 \\ \sigma_\theta^2 & \cdots & \cdots & \sigma_\theta^2 & \sigma_\theta^2 \end{pmatrix},$$

and the corresponding correlation matrix of the observed items X_j and the latent trait θ is:

$$R_{X,\theta} = \begin{pmatrix} 1 & \rho & \cdots & \cdots\rho & \sqrt{\rho} \\ \rho & 1 & \rho & \cdots\rho & \sqrt{\rho} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \cdots & \rho & 1 & \sqrt{\rho} \\ \sqrt{\rho} & \cdots & \cdots & \sqrt{\rho} & 1 \end{pmatrix}.$$

The *marginal* covariance V_X and correlation matrix R_X of the k observed variables X_j, under the parallel model, are:

$$V_X = \begin{pmatrix} \sigma_\theta^2 + \sigma^2 & \sigma_\theta^2 & \cdots & \cdots\sigma_\theta^2 \\ \sigma_\theta^2 & \sigma_\theta^2 + \sigma^2 & \sigma_\theta^2 & \cdots\sigma_\theta^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_\theta^2 & \cdots & \sigma_\theta^2 & \sigma_\theta^2 + \sigma^2 \end{pmatrix}$$

and

$$R_X = \begin{pmatrix} 1 & \rho & \cdots & \cdots\rho \\ \rho & 1 & \rho & \cdots\rho \\ \vdots & \vdots & \vdots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}.$$

This structure is known as a *compound symmetry*-type structure. It is easy to show that the reliability of the sum of k items given in [1.7] can be expressed as:

$$\tilde{\rho}_k = \frac{k}{k-1}\left[1 - \frac{trace(V_X)}{J'V_XJ}\right], \qquad [1.8]$$

with J a vector with all components being 1, and

$$\alpha = \frac{k}{k-1}\left[1 - \frac{trace(S_X)}{J'S_XJ}\right], \qquad [1.9]$$

where S_X is the observed variance, empirical estimation of S_X. There is, even in the recent literature, an understanable confusion between Cronbach's alpha as a population parameter (theoretical reliability of the sum of items) or its sample estimate.

In addition, it is easy to show a direct connection between the CAC and the percentage of variance of the first component in principal component analysis (PCA), which is often used to assess unidimensionality. The PCA is mainly based on the analysis of the latent roots of V_X or R_X (or, in practice their sample estimate). The matrix R_X has only two different latent roots, the greater root is $\lambda_1 = (k-1)\rho + 1$, and the other multiple roots are $\lambda_2 = \lambda_3 = \lambda_4 = \cdots = 1 - \rho = \frac{k-\lambda_1}{k-1}$. So, using the Spearman–Brown formula, we can express the reliability of the sum of the k variables as $\tilde{\rho}_k = \frac{k}{k-1}(1 - \frac{1}{\lambda_1})$.

This clearly indicates a monotonic relationship between $\tilde{\rho}_k$, which can be consistently estimated by the CAC, and the first latent root λ_x, which in practice is naturally estimated by the corresponding observed sample correlation matrix and thus the percentage of variance of the first principal component in a PCA. So, CAC can also be considered as a measure of unidimensionality.

Nevertheless such a measure is not very useful, because it is easy to show, using the Spearman–Brown formula [BRO 10, SPE 10], that under the parallel model assumption, the reliability of the total score is an increasing function of the number of variables.

Therefore, *if the parallel model is true*, increasing the number of items will increase the reliability of a questionnaire. Moreover, the coefficient lies between 0 and 1. Zero value indicates a totally unreliable scale, while unit value means that the scale is perfectly reliable. Of course, in practice, these two scenarios never occur.

The CAC is an estimate of the reliability of the raw-score (sum of item responses) of a person *if the model generating those responses is a parallel model.*