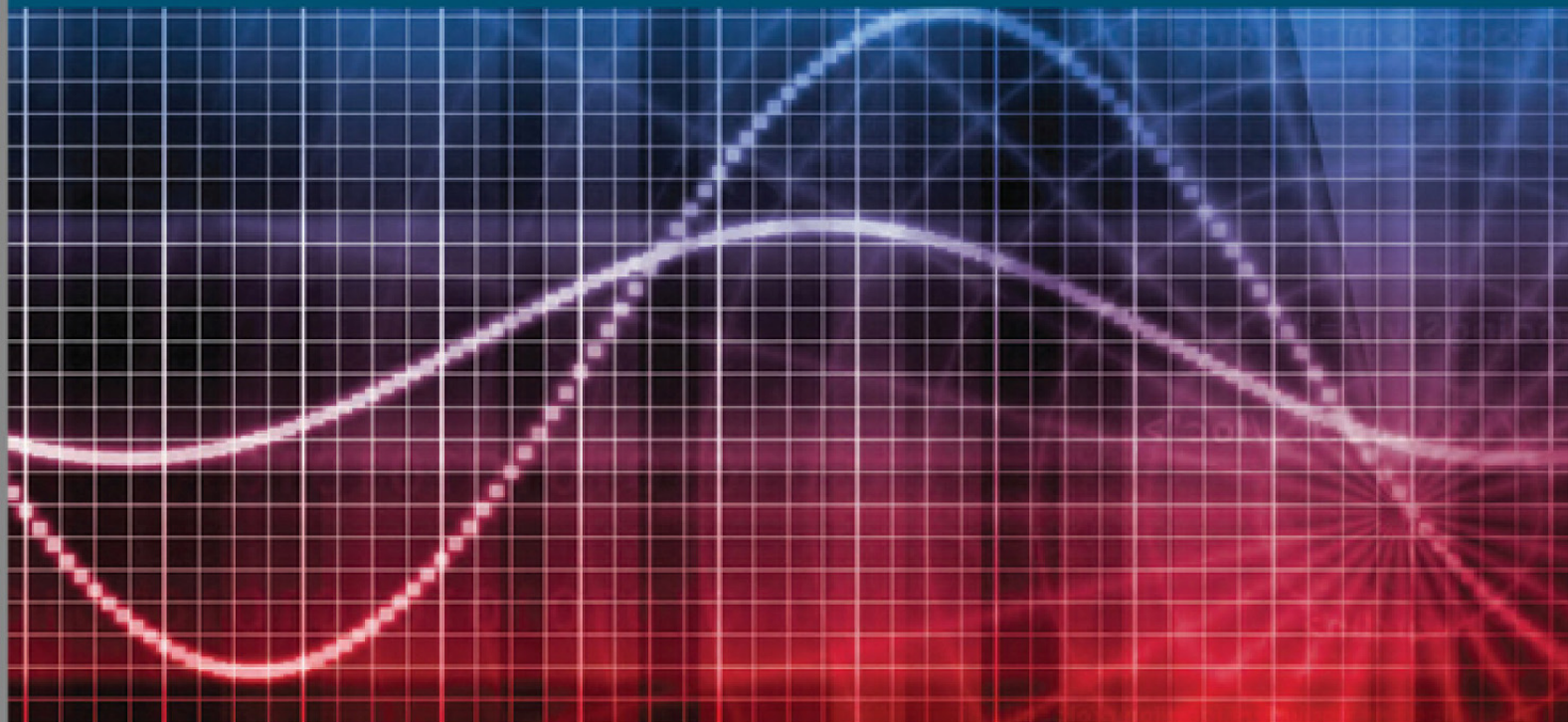


**MATHEMATICS AND STATISTICS SERIES**



# **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.

for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

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

[www.iste.co.uk](http://www.iste.co.uk)

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

[www.wiley.com](http://www.wiley.com)

© ISTE Ltd 2014

The rights of Vincent Couallier, Léo Gerville-Réache, Catherine Huber-Carol, Nikolaos Limnios and Mounir Mesbah to be identified as the authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

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

---



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

## Table of Contents

<b>Preface</b> . . . . .	xv
<b>Biography of Mikhail Stepanovitch Nikouline</b> . . . . . Vincent COUALLIER, Léo GERVILLE-RÉACHE, Catherine HUBER-CAROL, Nikolaos LIMNIOS and Mounir MESBAH	xvii
<b>PART 1. STATISTICAL MODELS AND METHODS</b> . . . . .	1
<b>Chapter 1. Unidimensionality, Agreement and Concordance Probability</b> . . . . .	3
Zhezhen JIN and Mounir MESBAH	
1.1. Introduction . . . . .	3
1.2. From reliability to unidimensionality: CAC and curve . . . . .	4
1.2.1. Classical unidimensional models for measurement . . . . .	4
1.2.2. Reliability of an instrument: CAC . . . . .	6
1.2.3. Unidimensionality of an instrument: BRC . . . . .	9
1.3. Agreement between binary outcomes: the kappa coefficient . . . . .	10
1.3.1. The kappa model . . . . .	10
1.3.2. The kappa coefficient . . . . .	10
1.3.3. Estimation of the kappa coefficient . . . . .	10
1.4. Concordance probability . . . . .	11
1.4.1. Relationship with Kendall's $\tau$ measure . . . . .	12
1.4.2. Relationship with Somer's $D$ measure . . . . .	12
1.4.3. Relationship with ROC curve . . . . .	13
1.5. Estimation and inference . . . . .	14
1.6. Measure of agreement . . . . .	14
1.7. Extension to survival data . . . . .	15
1.7.1. Harrell's $c$ -index . . . . .	15
1.7.2. Measure of discriminatory power . . . . .	16

1.8. Discussion . . . . .	17
1.9. Bibliography . . . . .	18
<b>Chapter 2. A Universal Goodness-of-Fit Test Based on Regression Techniques . . . . .</b>	<b>21</b>
Florence GEORGE and Sneh GULATI	
2.1. Introduction . . . . .	21
2.2. The Brain and Shapiro procedure for the exponential distribution . . . . .	22
2.3. Applications of the Brain and Shapiro test . . . . .	24
2.4. Small sample null distribution of the test statistic for specific distributions . . . . .	25
2.5. Power studies . . . . .	28
2.6. Some real examples . . . . .	28
2.7. Conclusions . . . . .	31
2.8. Acknowledgment . . . . .	32
2.9. Bibliography . . . . .	32
<b>Chapter 3. Entropy-type Goodness-of-Fit Tests for Heavy-Tailed Distributions . . . . .</b>	<b>33</b>
Andreas MAKRIDES, Alex KARAGRIGORIOU and Filia VONTA	
3.1. Introduction . . . . .	33
3.2. The entropy test for heavy-tailed distributions . . . . .	35
3.2.1. Development and asymptotic theory . . . . .	35
3.2.2. Discussion . . . . .	39
3.3. Simulation study . . . . .	40
3.4. Conclusions . . . . .	42
3.5. Bibliography . . . . .	42
<b>Chapter 4. Penalized Likelihood Methodology and Frailty Models . . . . .</b>	<b>45</b>
Emmanouil ANDROULAKIS, Christos KOUKOUVINOS and Filia VONTA	
4.1. Introduction . . . . .	45
4.2. Penalized likelihood in frailty models for clustered data . . . . .	48
4.2.1. Gamma distributed frailty . . . . .	52
4.2.2. Inverse Gaussian distributed frailty . . . . .	52
4.2.3. Uniform distributed frailty . . . . .	54
4.3. Simulation results . . . . .	55
4.4. Concluding remarks . . . . .	57
4.5. Bibliography . . . . .	57



<b>Chapter 5. Interactive Investigation of Statistical Regularities in Testing Composite Hypotheses of Goodness of Fit</b> . . . . .	61
Boris LEMESHKO, Stanislav LEMESHKO and Andrey ROGOZHNIKOV	
5.1. Introduction . . . . .	61
5.2. Distributions of the test statistics in the case of testing composite hypotheses . . . . .	63
5.3. Testing composite hypotheses in “real-time” . . . . .	68
5.4. Conclusions . . . . .	73
5.5. Acknowledgment . . . . .	73
5.6. Bibliography . . . . .	73
<b>Chapter 6. Modeling of Categorical Data</b> . . . . .	77
Henning LÄUTER	
6.1. Introduction . . . . .	77
6.2. Continuous conditional distributions . . . . .	78
6.2.1. Conditional normal distribution . . . . .	78
6.2.1.1. Estimation of parameters . . . . .	78
6.2.2. More general continuous conditional distributions . . . . .	81
6.2.2.1. Conditional distribution . . . . .	82
6.2.2.2. Normal copula . . . . .	83
6.3. Discrete conditional distributions . . . . .	84
6.3.1. Parametric conditional distributions . . . . .	84
6.3.2. Estimation of parameters . . . . .	86
6.4. Goodness of fit . . . . .	86
6.4.1. Distribution of $\hat{X}^2$ . . . . .	87
6.5. Modeling of categorical data . . . . .	88
6.5.1. Contingency tables . . . . .	89
6.5.1.1. General tables . . . . .	89
6.5.1.2. Further examples . . . . .	93
6.6. Bibliography . . . . .	93
<b>Chapter 7. Within the Sample Comparison of Prediction Performance of Models and Submodels: Application to Alzheimer’s Disease</b> . . . . .	95
Catherine HUBER-CAROL, Shulamith T. GROSS and Annick ALPÉROVITCH	
7.1. Introduction . . . . .	95
7.2. Framework . . . . .	96
7.2.1. General description of the data set and the models to be compared . . . . .	96
7.2.2. Definition of the performance prediction criteria: IDI and BRI . . . . .	96
7.3. Estimation of IDI and BRI . . . . .	97
7.3.1. General estimating equations for IDI and BRI . . . . .	98
7.3.2. Estimation of IDI and BRI in the logistic case . . . . .	98
7.3.2.1. Asymptotics of $\widehat{IDI}_{2/1}$ for logistic predictors . . . . .	99

7.3.2.2. Asymptotics of $\widehat{BRI}_{2/1}$ for logistic predictors . . . . .	100
7.4. Simulation studies . . . . .	102
7.4.1. First simulation . . . . .	102
7.4.2. Second simulation: Gu and Pepe's example . . . . .	104
7.5. The three city study of Alzheimer's disease . . . . .	106
7.6. Conclusion . . . . .	108
7.7. Bibliography . . . . .	109
<b>Chapter 8. Durbin–Knott Components and Transformations of the Cramér-von Mises Test</b> . . . . .	111
Gennady MARTYNOV	
8.1. Introduction . . . . .	111
8.2. Weighted Cramér-von Mises statistic . . . . .	111
8.3. Examples of the Cramér-von Mises statistics . . . . .	113
8.3.1. Classical Cramér-von Mises statistic . . . . .	113
8.3.2. Anderson–Darling statistic . . . . .	113
8.3.3. Cramér-von Mises statistic with the power weight function . . . . .	114
8.4. Weighted parametric Cramér-von Mises statistic . . . . .	114
8.4.1. Covariance functions of weighted parametric empirical process . . . . .	114
8.4.2. Eigenvalues and eigenfunctions for weighted parametric Cramér- von Mises statistic . . . . .	116
8.5. Transformations of the Cramér-von Mises statistic . . . . .	117
8.5.1. Preliminary notes . . . . .	117
8.5.2. Replacement of eigenvalues . . . . .	118
8.5.3. Transformed statistics . . . . .	119
8.6. Bibliography . . . . .	122
<b>Chapter 9. Conditional Inference in Parametric Models</b> . . . . .	125
Michel BRONIATOWSKI and Virgile CARON	
9.1. Introduction and context . . . . .	125
9.2. The approximate conditional density of the sample . . . . .	127
9.2.1. Approximation of conditional densities . . . . .	127
9.2.2. The proxy of the conditional density of the sample . . . . .	129
9.2.3. Comments on implementation . . . . .	131
9.3. Sufficient statistics and approximated conditional density . . . . .	131
9.3.1. Keeping sufficiency under the proxy density . . . . .	131
9.3.2. Rao–Blackwellization . . . . .	132
9.4. Exponential models with nuisance parameters . . . . .	135
9.4.1. Conditional inference in exponential families . . . . .	135
9.4.2. Application of conditional sampling to MC tests . . . . .	137
9.4.2.1. Context . . . . .	137
9.4.2.2. Bimodal likelihood: testing the mean of a normal distribution in dimension 2 . . . . .	139

9.4.3. Estimation through conditional likelihood . . . . .	140
9.5. Bibliography . . . . .	142
<b>Chapter 10. On Testing Stochastic Dominance by Exceedance, Precedence and Other Distribution-Free Tests, with Applications . . . . .</b>	<b>145</b>
Paul DEHEUVELS	
10.1. Introduction . . . . .	145
10.2. Results . . . . .	148
10.2.1. The experimental data set . . . . .	148
10.2.2. An application of the Wilcoxon–Mann–Whitney statistics . . . . .	149
10.2.3. One-sided Kolmogorov–Smirnov tests . . . . .	150
10.2.4. Precedence and Exceedance Tests. . . . .	152
10.3. Negative binomial limit laws . . . . .	155
10.4. Conclusion . . . . .	159
10.5. Bibliography . . . . .	159
<b>Chapter 11. Asymptotically Parameter-Free Tests for Ergodic Diffusion Processes . . . . .</b>	<b>161</b>
Yury A. KUTOYANTS and Li ZHOU	
11.1. Introduction . . . . .	161
11.2. Ergodic diffusion process and some limits . . . . .	165
11.3. Shift parameter . . . . .	168
11.4. Shift and scale parameters . . . . .	172
11.5. Bibliography . . . . .	175
<b>Chapter 12. A Comparison of Homogeneity Tests for Different Alternative Hypotheses . . . . .</b>	<b>177</b>
Sergey POSTOVALOV and Petr PHILONENKO	
12.1. Homogeneity tests . . . . .	178
12.1.1. Tests for data without censoring . . . . .	179
12.1.2. Tests for data with censoring . . . . .	180
12.2. Alternative hypotheses . . . . .	184
12.3. Power simulation . . . . .	185
12.3.1. Power of tests without censoring . . . . .	187
12.3.2. Power of tests with censoring . . . . .	189
12.3.2.1. How does the distribution of censoring time affect the power of the test? . . . . .	189
12.3.2.2. How does the censoring rate affect the power of the test? . . . . .	191
12.4. Statistical inference . . . . .	191
12.5. Acknowledgment . . . . .	192
12.6. Bibliography . . . . .	193

<b>Chapter 13. Some Asymptotic Results for Exchangeably Weighted Bootstraps of the Empirical Estimator of a Semi-Markov Kernel with Applications</b>	195
Salim BOUZEBDA and Nikolaos LIMNIOS	
13.1. Introduction	195
13.2. Semi-Markov setting	197
13.3. Main results	201
13.4. Bootstrap for a multidimensional empirical estimator of a continuous-time semi-Markov kernel	205
13.5. Confidence intervals	208
13.6. Bibliography	210
<b>Chapter 14. On Chi-Squared Goodness-of-Fit Test for Normality</b>	213
Mikhail NIKULIN, Léo GERVILLE-RÉACHE and Xuan Quang TRAN	
14.1. Chi-squared test for normality	213
14.2. Simulation study	221
14.3. Bibliography	226
<b>PART 2. STATISTICAL MODELS AND METHODS IN SURVIVAL ANALYSIS</b>	229
<b>Chapter 15. Estimation/Imputation Strategies for Missing Data in Survival Analysis</b>	231
Elodie BRUNEL, Fabienne COMTE and Agathe GUILLOUX	
15.1. Introduction	231
15.2. Model and strategies	233
15.2.1. Model assumptions	233
15.2.2. Strategy involving knowledge of $\zeta$	234
15.2.3. Strategy involving knowledge of $\pi$	235
15.2.4. Estimation of $\zeta$ or $\pi$ : logit or non-parametric regression	236
15.2.5. Computing the hazard estimators	236
15.2.6. Theoretical results	239
15.3. Imputation-based strategy	241
15.4. Numerical comparison	242
15.5. Proofs	244
15.6. Bibliography	251
<b>Chapter 16. Non-Parametric Estimation of Linear Functionals of a Multivariate Distribution Under Multivariate Censoring with Applications</b>	253
Olivier LOPEZ and Philippe SAINT-PIERRE	
16.1. Introduction	253

16.2. Non-parametric estimation of the distribution . . . . .	255
16.3. Asymptotic properties . . . . .	257
16.4. Statistical applications of functionals . . . . .	260
16.4.1. Dependence measures . . . . .	260
16.4.2. Bootstrap . . . . .	261
16.4.3. Linear regression . . . . .	262
16.5. Illustration . . . . .	263
16.6. Conclusion . . . . .	264
16.7. Acknowledgment . . . . .	264
16.8. Bibliography . . . . .	264
<b>Chapter 17. Kernel Estimation of Density from Indirect Observation . . . . .</b>	<b>267</b>
Valentin SOLEV	
17.1. Introduction . . . . .	267
17.1.1. Random partition . . . . .	267
17.1.2. Indirect observation . . . . .	268
17.1.3. Kernel density estimator . . . . .	269
17.2. Density of random vector $\Lambda(X)$ . . . . .	271
17.3. Pseudo-kernel density estimator . . . . .	273
17.3.1. Pointwise density estimation based on indirect data . . . . .	273
17.3.2. Bias of the kernel estimator . . . . .	274
17.3.3. Estimate of variance . . . . .	276
17.4. Bibliography . . . . .	279
<b>Chapter 18. A Comparative Analysis of Some Chi-Square Goodness-of-Fit Tests for Censored Data . . . . .</b>	<b>281</b>
Ekaterina CHIMITOVA and Boris LEMESHKO	
18.1. Introduction . . . . .	281
18.2. Chi-square goodness-of-fit tests for censored data . . . . .	283
18.2.1. NRR $\chi^2$ test . . . . .	283
18.2.2. GPF $\chi^2$ test . . . . .	284
18.3. The choice of grouping intervals . . . . .	285
18.3.1. Equifrequent grouping (EFG) . . . . .	289
18.3.2. Intervals with equal expected numbers of failures (EENFG) . . . . .	289
18.3.3. Optimal grouping (OptG) . . . . .	289
18.4. Empirical power study . . . . .	290
18.5. Conclusions . . . . .	293
18.6. Acknowledgment . . . . .	294
18.7. Bibliography . . . . .	294

<b>Chapter 19. A Non-parametric Test for Comparing Treatments with Missing Data and Dependent Censoring</b> . . . . .	297
Amel MEZAOUER, Kamal BOUKHETALA and Jean-François DUPUY	
19.1. Introduction . . . . .	297
19.2. The proposed test statistic . . . . .	299
19.3. Asymptotic distribution of the proposed test statistic . . . . .	301
19.4. Acknowledgment . . . . .	305
19.5. Appendix . . . . .	306
19.6. Bibliography . . . . .	309
<b>Chapter 20. Group Sequential Tests for Treatment Effect with Covariates Adjustment through Simple Cross-Effect Models</b> . . . . .	311
Isaac Wu HONG-DAR	
20.1. Introduction . . . . .	311
20.2. Notations and models . . . . .	313
20.3. Group sequential test . . . . .	316
20.4. Discussion . . . . .	318
20.5. Acknowledgment . . . . .	318
20.6. Bibliography . . . . .	318
<b>PART 3 . RELIABILITY AND MAINTENANCE</b> . . . . .	321
<b>Chapter 21. Optimal Maintenance in Degradation Processes</b> . . . . .	323
Waltraud KAHLE	
21.1. Introduction . . . . .	323
21.2. The degradation model . . . . .	324
21.3. Optimal replacement after an inspection . . . . .	326
21.4. The simulation of degradation processes . . . . .	327
21.5. Shape of cost functions and optimal $\delta$ and $a$ . . . . .	329
21.6. Incomplete preventive maintenance . . . . .	330
21.7. Bibliography . . . . .	333
<b>Chapter 22. Planning Accelerated Destructive Degradation Tests with Competing Risks</b> . . . . .	335
Ying SHI and William Q. MEEKER	
22.1. Introduction . . . . .	336
22.1.1. Background . . . . .	336
22.1.2. Motivation: adhesive bond C . . . . .	336
22.1.3. Related literature . . . . .	337
22.1.4. Overview . . . . .	338
22.2. Degradation models with competing risks . . . . .	338
22.2.1. Accelerated degradation model for the primary response . . . . .	338

22.2.2. Accelerated degradation model for the competing response . . .	339
22.2.3. Degradation models for adhesive bond C . . . . .	339
22.2.4. Degradation distribution and quantiles . . . . .	340
22.3. Failure-time distribution with competing risks . . . . .	341
22.3.1. Relationship between degradation and failure . . . . .	341
22.3.2. Failure-time distribution and quantiles . . . . .	342
22.4. Test planning with competing risks . . . . .	342
22.4.1. ADDT planning information . . . . .	342
22.4.2. Criterion for ADDT planning with competing risks . . . . .	343
22.5. ADDT plans with competing risks . . . . .	344
22.5.1. Initial optimum ADDT plan with competing risks . . . . .	344
22.5.2. Constrained optimum ADDT plan with competing risks . . . . .	348
22.5.3. General equivalence theorem . . . . .	348
22.5.4. Compromise ADDT plan with competing risks . . . . .	350
22.6. Monte Carlo simulation to evaluate test plans . . . . .	352
22.7. Conclusions and extensions . . . . .	353
22.8. Appendix: technical details . . . . .	354
22.8.1. The Fisher information matrix for ADDT with competing risks . . . . .	354
22.8.2. Large-sample approximate variance of $h_t(\hat{t}_p)$ and $\hat{t}_p$ . . . . .	355
22.9. Bibliography . . . . .	355
<b>Chapter 23. A New Goodness-of-Fit Test for Shape-Scale Families . . . . .</b>	<b>357</b>
Vilijandas BAGDONAVIČIUS	
23.1. Introduction . . . . .	357
23.2. The test statistic . . . . .	358
23.3. The asymptotic distribution of the test statistic . . . . .	359
23.4. The test . . . . .	364
23.5. Weibull distribution . . . . .	364
23.6. Loglogistic distribution . . . . .	365
23.7. Lognormal distribution . . . . .	366
23.8. Bibliography . . . . .	367
<b>Chapter 24. Time-to-Failure of Markov-Modulated Gamma Process with Application to Replacement Policies . . . . .</b>	<b>369</b>
Christian PAROISSIN and Landy RABEHASAINA	
24.1. Introduction . . . . .	369
24.2. Degradation model . . . . .	370
24.2.1. Covariate process . . . . .	370
24.2.2. Degradation process . . . . .	371
24.3. Time-to-failure distribution . . . . .	371
24.3.1. Case of a non-modulated gamma process . . . . .	372
24.3.2. Case of a Markov-modulated gamma process . . . . .	373
24.3.3. Stochastic comparison . . . . .	374

24.4. Replacement policies . . . . .	376
24.4.1. Block replacement policy . . . . .	377
24.4.2. Age replacement policy . . . . .	379
24.5. Conclusion . . . . .	381
24.6. Acknowledgment . . . . .	381
24.7. Bibliography . . . . .	382
<b>Chapter 25. Calculation of the Redundant Structure Reliability for Aging-type Elements . . . . .</b>	<b>383</b>
Alexandr ANTONOV, Alexandr PLYASKIN and Khizri TATAEV	
25.1. Introduction . . . . .	383
25.2. The operation process of the renewal and repaired products . . . . .	384
25.3. The model of the geometric process . . . . .	386
25.4. Task solution . . . . .	387
25.5. Conclusion . . . . .	389
25.6. Bibliography . . . . .	390
<b>Chapter 26. On Engineering Risks of Complex Hierarchical Systems Analysis . . . . .</b>	<b>391</b>
Vladimir RYKOV	
26.1. Introduction . . . . .	391
26.2. Risk definition and measurement . . . . .	392
26.3. Engineering risk . . . . .	393
26.4. Risk characteristics for general model calculation . . . . .	395
26.4.1. Lifelength and appropriate loss size CDF . . . . .	395
26.4.2. Probability of risk event evolution . . . . .	396
26.4.3. Lifelength and loss moments . . . . .	397
26.4.4. Mostly dangerous paths of risk event evolution and sensitivity analysis . . . . .	399
26.5. Risk analysis for short-time risk models . . . . .	400
26.6. Conclusion . . . . .	402
26.7. Bibliography . . . . .	402
<b>List of Authors . . . . .</b>	<b>405</b>
<b>Index . . . . .</b>	<b>409</b>



## 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 H el ene at the end of the same year.

Then they decided to learn French intensively in order to go to Brazzaville, capital of the “R epublique Populaire du Congo”, where they spent three years. There, both were professors of mathematics in the *Coll ege d’Enseignement G en eral* 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, L eo GERVILLE-R EACHE, Catherine HUBER-CAROL, Nikolaos LIMNIOU 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  $\tau$ , 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

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  $\tau$ , 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  $A = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k\}$  and a latent (unobservable) variable  $\theta$  of dimension  $d \leq k$ . In such models, the dimensionality of  $A$  is captured by the dimension of  $\theta$ , 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  $j$ th question by the  $i$ th subject and the  $\mathbf{X}_j$  denotes the random variable generating responses to the  $j$ th question.

The parallel model is a classical latent variable model describing the unidimensionality of a set  $A = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_k\}$  of quantitative observable

variables. Let  $\mathbf{X}_{ij}$  be the measurement of subject  $i$ , given by a variable  $\mathbf{X}_j$ ,  $i = 1, \dots, n, j = 1, \dots, k$ , then:

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

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

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

where  $\beta_j$  is an unknown fixed parameter (non-random) representing the effect of the  $j$ th variable, and  $\theta_i$  is an unknown random parameter effect of the  $i$ th subject.

It is generally assumed that  $\theta_i$  has zero mean and unknown standard deviation  $\sigma_\theta$ . 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,  $\theta_i$  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  $\varepsilon_{ij}$  are independent random errors with zero mean and standard deviation  $\sigma$  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  $\mathbf{X}_j, j = 1, \dots, k$ , and the true unique latent measure  $\theta_i$  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  $\beta_j, j = 1, \dots, k$ , are fixed effects and  $\theta_i, i = 1, \dots, n$ , are independent random effects with zero mean and standard deviation  $\sigma_\theta$ . 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  $\tau_{ij}$  and the measurement error  $\varepsilon_{ij}$ . 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  $i$ th individual does not vary with the question number  $j$  as the  $\theta_i$  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 ( $\theta_i$ ). More precisely, let  $\mathbf{X}_{ij}^* = \mathbf{X}_{ij} - \beta_j$  be the calibrated version of the response to the  $j$ th item by the  $i$ th subject, then models [1.2] and [1.3] can be rewritten as:

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

along with the same assumptions on  $\beta$  and  $\theta$  and the conditional variance model [1.3].

When both  $\theta_i$  and  $\varepsilon_{ij}$  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  $\theta_i$ .

### 1.2.2. Reliability of an instrument: CAC

A measurement instrument yields values that we call the observed measure. The reliability  $\rho$  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:

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

This coefficient is also known as the intra-class coefficient. The reliability coefficient,  $\rho$ , 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:

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

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  $\alpha$ :

$$\alpha = \frac{k}{k-1} \left( 1 - \frac{\sum_{j=1}^k S_j^2}{S_{tot}^2} \right), \quad [1.7]$$

where

$$S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$$

and

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

Under the parallel model, the variance–covariance matrix of the observed items  $X_j$  and the latent trait  $\theta$  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  $\theta$  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:

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

with  $J$  a vector with all components being 1, and

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

where  $S_X$  is the observed variance, empirical estimation of  $V_X$ . There is, even in the recent literature, an understandable 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  $\lambda_2 = \lambda_3 = \lambda_4 = \dots = 1 - \rho = \frac{k-\lambda_1}{k-1}$ . So, using the Spearman–Brown formula, we can express the reliability of the sum of the  $k$  variables as  $\tilde{\rho}_k = \frac{k}{k-1} \left( 1 - \frac{1}{\lambda_1} \right)$ .

This clearly indicates a monotonic relationship between  $\tilde{\rho}_k$ , which can be consistently estimated by the CAC, and the first latent root  $\lambda_x$ , 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*.