Springer Proceedings in Mathematics & Statistics

Volume 136

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Christos P. Kitsos · Teresa A. Oliveira
Alexandros Rigas · Sneh Gulati
Editors

Theory and Practice of Risk Assessment

ICRA 5, Tomar, Portugal, 2013

Springer

Editors
Christos P. Kitsos
Technological Educational Institute of Athens
Athens
Greece

Teresa A. Oliveira
Universidade Aberta
Palácio Ceia, Lisbon
Portugal

Alexandros Rigas
Democritus University
Thrace
Germany

Sneh Gulati
Florida International University
Miami, FL
USA

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-18028-1 ISBN 978-3-319-18029-8 (eBook)
DOI 10.1007/978-3-319-18029-8

Library of Congress Control Number: 2015937949

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Everything humans venture to do has some degree of risk involved in it. Since the degree of risk they face is random, it is only natural that we statisticians feel the need to jump in and try to analyze it. The plethora of papers in this field being published in the field of risk analysis eventually led to the launching of a conference series where statisticians could present their findings.

The International Committee on Risk Analysis initially launched the series as a conference on cancer risk assessment in Athens, Greece on August 22, 2003. The second and the third conferences held in 2007 and 2009, respectively, still retained cancer risk assessment as the main focus; however, after the third conference it was decided to broaden the theme of the conference. This led to the new acronym of the ICRA (**I**nternational **C**onference on **R**isk **A**nalysis) series, so the 4th meeting took place in 2011 at Limassol, Cyprus. Proceedings of these meetings are available.

The fifth conference on risk analysis, ICRA5 was held at Tomar, Portugal, from May 30 to June 1, 2013. The papers presented at this conference covered a number of topics on risk analysis with applications in both biological and industrial fields. This book forms a proceedings volume and includes most of the papers presented at the conference. The book itself is divided into two main parts based on the subject matter covered:

Part I is devoted to *Risk Methods for Bioinformatics* while
Part II focuses on *Risk Methods for Management and Industry.*

The papers in Part I mainly cover topics from Life Sciences and Environmetrics and are divided into subsections based on the primary focus of the papers included. The first subsection in The Bioinformatics section deals with the original theme of the conference: "cancer research" and consists of three chapters covering various aspects of cancer risk analysis. The second subsection consists of two chapters which consider the applications of first time hitting models while the third subsection considers papers on general quantification of risk for diseases. Finally, the last subsection in this first part considers risk analysis as it pertains to the environment.

The second part of the book, “Risk Analysis in Management and Industry,” is also divided into four subsections on the basis of the subject matter covered. The first chapter in this part though deals with sampling strategies and stands on its own. Section 2 considers papers on Industrial Quality Control and consists of two chapters. Next in Section 3, the focus is on Extreme Value Theory with the papers looking at various ways of quantifying extreme quantiles in natural catastrophic events. The last subsection in this part is devoted to general papers in reliability and survival analysis and consists of six chapters.

The total of 30 papers presented in this volume cover a diverse range of topics on risk analysis and we hope our readers find it useful for their research.

Contents

Part I Risk Methods for Bioinformatics

Generalized Information Criteria for the Best Logit Model 3
Christos P. Kitsos and Thomas L. Toulias

Fractal Case Study for Mammary Cancer: Analysis of Interobserver Variability 21
Philipp Hermann, Sarah Piza, Sandra Ruderstorfer, Sabine Spreitzer and Milan Stehlík

On Analytical Methods for Cancer Research 37
Stefan Giebel, Philipp Hermann, Jens-Peter Schenk and Milan Stehlík

Modelling Times Between Events with a Cured Fraction Using a First Hitting Time Regression Model with Individual Random Effects 45
S. Malefaki, P. Economou and C. Caroni

Acceleration, Due to Occupational Exposure, of Time to Onset of a Disease 67
A. Chambaz, D. Choudat and C. Huber-Carol

Transformations of Confidence Intervals for Risk Measures? 79
Karl-Ernst Biebler and Bernd Jäger

Discrete Compound Tests and Dorfman's Methodology in the Presence of Misclassification 85
Rui Santos, João Paulo Martins and Miguel Felgueiras

A Maximum Likelihood Estimator for the Prevalence Rate Using Pooled Sample Tests 99
João Paulo Martins, Rui Santos and Miguel Felgueiras

On Intra-individual Variations in Hair Minerals in Relation to Epidemiological Risk Assessment of Atopic Dermatitis 111
Tomomi Yamada, Todd Saunders, Tsuyoshi Nakamura, Koichiro Sera and Yoshiaki Nose

Assessing Risk Factors for Periodontitis Using Multivariable Regression Analysis 123
J.A. Lobo Pereira, Maria Cristina Ferreira and Teresa A. Oliveira

COPD: On Evaluating the Risk for Functional Decline 133
F. Rodrigues, I. Matias, J. Oliveira, S. Vacas and A. Botelho

Microarray Experiments on Risk Analysis Using R 147
Teresa A. Oliveira, Amílcar Oliveira and Andreia A. Monteiro

Risk Assessment of Complex Evolving Systems Involving Multiple Inputs 159
A.G. Rigas and V.G. Vassiliadis

Monitoring Environmental Risk by a Methodology Based on Control Charts 177
Helton Saulo, Victor Leiva and Fabrizio Ruggeri

Risk Problems Identifying Optimal Pollution Level 199
George E. Halkos and Dimitra C. Kitsou

Part II Risk Methods for Management and Industry

Finite Populations Sampling Strategies and Costs Control 211
Dinis Pestana, Maria Luísa Rocha and Fernando Sequeira

Industrial Production of Gypsum: Quality Control Charts 225
Luís M. Grilo, Helena L. Grilo and Cristiano J. Marques

Risk Analysis with Reference Class Forecasting Adopting Tolerance Regions 235
Vasilios Zarikas and Christos P. Kitsos

Randomly Stopped *k*th Order Statistics 249
Sandra Mendonça, Dinis Pestana and M. Ivette Gomes

The Role of Asymmetric Families of Distributions in Eliminating Risk 267
Fernanda Otília Figueiredo and Maria Ivette Gomes

Parametric and Semi-parametric Approaches to Extreme Rainfall Modelling 279
Isabel Fraga Alves and Pedro Rosário

A Log Probability Weighted Moment Estimator of Extreme Quantiles 293
Frederico Caeiro and Dora Prata Gomes

A Mean-of-Order-*p* Class of Value-at-Risk Estimators 305
M. Ivette Gomes, M. Fátima Brilhante and Dinis Pestana

Adaptive Choice and Resampling Techniques in Extremal Index Estimation 321
Dora Prata Gomes and M. Manuela Neves

Some Estimation Techniques in Reliability and Survival Analysis Based on Record-Breaking Data 333
Inmaculada Barranco-Chamorro and Sneh Gulati

Risk Scoring Models for Trade Credit in Small and Medium Enterprises 349
Manuel Terradez, Renatas Kizys, Angel A. Juan, Ana M. Debon and Bartosz Sawik

Signatures of Systems with Non-exchangeable Lifetimes: Some Implications in the Analysis of Financial Risk 361
Roy Cerqueti and Fabio Spizzichino

Detecting an IO/AO Outlier in a Set of Time Series 377
Vassiliki Karioti

Response Surface Methodology: A Review of Applications to Risk Assessment 385
Teresa A. Oliveira, Conceição Leal and Amílcar Oliveira

FF-Type Multivariate Models in an Enforced Regression Paradigm . . . 399
Jerzy K. Filus and Lidia Z. Filus

Introduction

"Nothing Ventured, Nothing Gained," a well-known proverb implying that to attain something, one has to be willing to take risks. Thus, it is only natural that no professional venture or field can be devoid of risk. As an example, in medical studies, one is concerned with the risk of patient death, the risk of a lack of a cure, the risk of side effects from medications, etc. Engineers are concerned with the risk of structural and mechanical failures; manufacturers are concerned with the risk of producing defective products and so on. Mitigating risk and analyzing it then are integral components of any area, however, risk analysis per se, is a field specific to applied statistics. In an attempt to recognize the role that statistics plays in risk analysis, the International Committee on Risk Analysis decided to launch a conference series to serve as a forum for researchers in this area to get together and discuss their methodologies.

The main focus of the first conference was cancer research, and so the series was actually launched as the International Conference on Cancer Risk Assessment (ICCRA) in Athens, Greece on August 22, 2003. The second conference was held on the island of Santorini, Greece during May 25–27, 2007 (still as ICCRA) while the third and final ICCRA was held at Porto Heli, Greece during July 16–18, 2009. Thereafter it was decided to broaden the theme of the conference and the first ICRA (**I**nternational **C**onference on **R**isk **A**nalysis) labeled ICRA4 took place at Limassol, Cyprus, during May 26–29, 2011. From its inception, support for the series was provided by its main sponsor, the ISI Committee on Risk Analysis. One of the main aims of the committee was to improve and expand the role of statistics in risk analysis. While the committee retained human health, welfare, and survival as its main focus, it decided to more actively pursue risk analysis in other fields such as the environment, ecology, engineering, etc. Toward that end it was formally decided that the committee would engage more actively in conferences with a broader coverage of risk analysis, including identification and quantification of risk (http://www.isi-web.org/sections/44-com/com/126-ra).

The first realization of this goal was manifested in ICRA5 held at Tomar, Portugal, from May 30 to June 1, 2013. The conference drew together a number of scientists working on various aspects of risk analysis with applications in both

biological and industrial fields. This book forms a proceedings volume and includes most of the papers presented at the conference.

While the past ICRA conferences have led to a number of publications, none have been as broad in subject coverage as this book. Examples of statistical analysis using real data are found throughout this book, which we hope will spark interest in the related theoretical results. Despite the fact that research in risk analysis has been burgeoning, the new methodologies have not had the deepest possible penetration among the practitioners of the field. We believe that this is because the relevant articles and papers are scattered in too many journals with different foci. We hope to remedy the situation with the publication of this volume, totally devoted to methods on Risk Analysis.

The book itself is divided into two main parts based on the subject matter covered:

Part I is devoted to *Risk Methods for Bioinformatics* while
Part II focuses on *Risk Methods for Management and Industry.*

The papers in Part I mainly cover topics from Life Sciences and Environmetrics and are divided into subsections based on the primary focus of the papers included. We now briefly describe some of the topics covered in this section:

The first subsection in The Bioinformatics section deals with the original theme of the conference: "cancer research." "Generalized Information Criteria for the Best Logit Model" considers the use of Entropy measures to quantify relative risk and applies it to compute the relative risk of breast cancer for women based on their risk factors which include (but are not limited to): age, use of oral contraceptives, hormone replacement therapy. "Fractal Case Study for Mammary Cancer: Analysis of Interobserver Variability" uses a Fractal Case Study to classify different types of malignancies in cancer tissues. "On Analytical Methods for Cancer Research" uses Statistical Dynamic Shape Analysis to quantify cancer risk. The authors argue that temporal shaping in medicine has to consider the medical relevance for certain time points in the measurement and landmarks to describe the object at these time points. Their analysis is shown to have a distinct advantage in oncology compared to traditional approaches.

The second subsection on Bioinformatics looks at the applications of First Time Hitting Models. In remission studies, one often encounters long-term survivors, or a "cured fraction" of units, which will never experience the event of interest. As a result, the empirical survival function for such studies never tends to zero. First hitting time (FHT) models can be used to account for such phenomena in lifetime models, and an example relevant to treatment of drug users is presented in "Modelling Times Between Events with a Cured Fraction Using a First Hitting Time Regression Model with Individual Random Effects". First time hitting models can also be used to estimate the number of disease-free years lost to occupational exposure and "Acceleration, Due to Occupational Exposure, of Time to Onset of a Disease" uses such a model to estimate the "expected number of disease free years lost due to exposure to asbestos."

The next subsection focuses on general risk quantification for diseases. Relationships between relative risk, odds ratios, and their respective confidence intervals are discussed in "Transformations of Confidence Intervals for Risk Measures?". "Discrete Compound Tests and Dorfman's Methodology in the Presence of Misclassification" presents an overview of the application of compound tests to classify individuals into two groups based on the presence or absence of a disease. In the same spirit, "A Maximum Likelihood Estimator for the Prevalence Rate Using Pooled Sample Tests" presents maximum likelihood methods to determine the prevalence rate of a disease. "On Intra-Individual Variations in Hair Minerals in relation to Epidemiological Risk Assessment of Atopic Dermatitis" discusses risk analysis in a Cohort Study of 842 mother-infant pairs for Atopic Dermatitis in Japan. The chapter looks at the association between hair minerals at one month and the onset of atopic dermatitis (AD) at ten months after birth with the aim of identifying infants with a high risk of getting the disease. "Assessing Risk Factors for Periodontitis Using Multivariable Regression Analysis" presents a deterministic mathematical model to evaluate risk factors for periodontitis (using data from Portugal) and the authors conclude that periodontitis is significantly associated with High Density Lipoproteins (HDL). "COPD: on Evaluating the Risk for Functional Decline" considers patients with Chronic Obstructive Pulmonary Disease (COPD) and uses a longitudinal study to measure their risk of becoming dependent on others for day-to-day activities. The goal is early intervention and assistance in order to reduce their dependence on others. "Microarray Experiments on Risk Analysis Using R" presents several designs to conduct microarray analysis using R, a technique that is being increasingly used to identify individuals at risk of getting a certain disease as well to identify the relevant risk factors. Finally, the last chapter in this subsection, "Risk Assessment of Complex Evolving Systems Involving Multiple Inputs" looks at complex nuerophysical systems with multiple inputs to evaluate whether some of the inputs inhibit the occurrence of new events.

The final subsection in this section deals with risk analysis for environmental sciences. While most of us want to live in a world completely free of pollution, realistically speaking that is an impossible dream. Hence, environmental policies focus on achieving "optimal pollution levels" where the marginal damage cost is equal to the marginal abetment cost. The authors "Monitoring Environmental Risk by a Methodology Based on Control Charts" argue that it is more efficient to focus instead on maximizing the net benefit (difference between abatement costs and damage costs.) They present different methods to evaluate the benefit area, which allows the comparison of different environmental policies. In the last chapter of this section, "Risk Problems Identifying Optimal Pollution Level", the authors propose a method for monitoring environmental risk through the use of control charts when the contaminant concentration follows a Birnbaum-Saunders distribution.

Next we turn to Part II: Risk Analysis in Management and Industry. The success of any industry is heavily dependent on its ability to deliver a product that is consistent and of high quality. Thus risk of failures, breakdowns, losses, inferior quality, etc., must all be identified and mitigated. Some methodologies to do just that are presented in this next section. As in the previous part, this section is also

divided into subsections, although the first chapter in this section is in its own subsection since it deals with sampling strategies.

"Finite Populations Sampling Strategies and Costs Control" presents a brief and compact review of Sampling Techniques. As pointed out by the authors, sophisticated statistical techniques are useless when they use bad data. Thus the authors present a quick overview of sampling strategies to show how to deal with cost control in nonideal circumstances (where the practitioner is unable to sample randomly without replacement).

Thereafter, the section moves on to papers with applications to real-life problems. The first subsection is on Quality Control. Process Control Charts have revolutionized the concept of monitoring quality. However, standard quality control charts are predicated on the assumption of normality, which is often not the case with real data. "Industrial Production of Gypsum: Quality Control Charts" deals with the use of Box-Cox transformations to normalize data in order to construct appropriate control charts. An application to the production of gypsum (marketed only if it meets required specifications) is presented. The benefit of a tolerance region, rather than a confidence region for problems in Industry and Management, is explained and discussed in "Risk Analysis with Reference Class Forecasting Adopting Tolerance Regions".

The previous subsection provides a natural bridge to the next subsection on "Extreme Value Theory." In order to estimate guarantee values and tolerance limits, most manufacturers are concerned with the estimation of extreme quantiles, especially in the context of heavy-tailed distributions. Heavy-tailed distributions are a norm in financial and insurance data. Extreme quantiles in these settings are often called Value at Risk at level q (Var_q) or Probable Maximum Loss (PML). No risk analysis strategy in the business world is complete without an evaluation of Var_q and a number of papers in this section deal with the estimation of the same. Research methodologies presented deal with the enumeration of stable extreme value laws along with a characterization of their domains in "Randomly Stopped kth Order Statistics", the use of a new class of skew-normal distributions to model heavy-tailed distributions in "The Role of Asymmetric Families of Distributions in Eliminating Risk", estimation of extreme rainfall levels using parametric and semi-parametric methods in "Parametric and Semi-Parametric Approaches to Extreme Rainfall Modelling", the use of Pareto Probability Weighted Moments (PPWM) and semi-parametric methods to estimate extreme quantiles in "A Log Probability Weighted Moment Estimator of Extreme Quantiles" and "A Mean-of-order-p Class of Value-at-Risk Estimators" respectively. The last chapter in this subsection, "Adaptive Choice and Resampling Techniques in Extremal Index Estimation" presents the use of resampling techniques to estimate the extreme value index, an important parameter in the estimation of PML.

The remaining papers fall under the general heading of "Applications in Reliability and Survival Analysis" and are summarized below:

In a number of experiments in industrial quality control and reliability, observed data often consist of record-breaking values where only successive maxima or minima are recorded. Such data also routinely arise in fields like Climatology,

Geosciences, and athletics. The authors in "Some Estimation Techniques in Reliability and Survival Analysis Based on Record-Breaking Data", present a review of the results on statistical inference from records that can be used in Reliability and Survival Analysis, including inferential results for heavy-tailed distributions. Commercial credit management is a matter of great importance for most small and medium enterprises (SMEs), since it represents a significant portion of their assets. Commercial lending involves assuming some credit risk due to exposure to default. Thus, the Management of Trade Credit and payment delays are strongly related to the liquidation and bankruptcy of these enterprises. The relationship between Trade Credit Management and the level of risk in SMEs is extensively discussed in "Risk Scoring Models for Trade Credit in Small and Medium Enterprises". The concept of signature is a powerful tool in the analysis of reliability systems and networks. However, most papers on this topic have dealt with a system of i.i.d components. "Signatures of Systems with Non-exchangeable Lifetimes: Some Implications in the Analysis of Financial Risk" considers the expansion of this concept to the non-exchangeable case, which allows applications to systems in different fields, such as Economics, Financial Risk, Environmental Sciences, etc. One of the key factors in quality control and risk quantification is the identification of outliers. In autoregressive time series one encounters two basic types of outliers: additive outliers (AO), affecting only a particular observation, and innovative outliers (IO), which act as an addition to the noise at a point in the entire series. Tests to detect the two types of outliers are presented in "Detecting an IO/AO Outlier in a Set of Time Series". Response Surface Methodology (RSM) is becoming more and more important as a risk assessment tool in this ever-changing world where big data and several dependent variables are the norm. Thus in "Response Surface Methodology: A Review of Applications to Risk Assessment" presents a review of the various aspects on the use of RSM as a risk assessment tool in the environmental, financial and public health fields. The final chapter in our book, "FF-type Multivariate Models in an Enforced Regression Paradigm" considers the use of "Enforced Regression Theory" to describe the relationship between a dependent variable and several independent variables.

In conclusion, the 30 papers included in this volume are diverse in nature, some applied, some theoretical, with a number of them providing the essential bridge between the two. In addition, several review papers included here fulfill the mission of the committee to put forth publications with papers that review various methodologies in risk assessment.

Given its scope and the straightforward nature of the presentation, we believe that this book will help a new generation of statisticians and practitioners to solve complex problems in risk analysis. Therefore, this book can easily serve as a textbook for a special topics course in risk analysis.

All of the papers collected here were reviewed by two referees and by the editors. We would like to extend our heartfelt thanks to all the reviewers who devoted their time to allow us to improve the quality of the submitted papers, and in turn the quality of the volume. At the same time, we express our sincere thanks to

all the authors, not only for their submission of papers, but also for their expeditious revisions and for incorporating the reviewer's suggestions.

Thanks also go out to the ISI Committee on Risk Analysis for sponsoring the conference. We would also like to express our sincere gratitude to all the people who worked on various committees, served as chairpersons, reviewed papers, and of course presented their own work at ICRA5. Without them ICRA5 would have never seen the light of day.

Last but not least, the editors would also like to express their heartfelt thanks and gratitude to SPRINGER for their help and support with this volume, especially, Dr. Eva Hiripi and Udhayakumar Panneerselvam without whose valuable assistance we could never have realized this manuscript.

Christos P. Kitsos
Chair, ISI Committee on Risk Analysis
Technological Educational Institute of Athens

Teresa A. Oliveira
Secretary, ISI Committee on Risk Analysis
Universidade Aberta and Center of Statistics and
Applications of University of Lisbon

Alexandros Rigas
Democritus University of Thrace

Sneh Gulati
Florida International University, Miami, FL