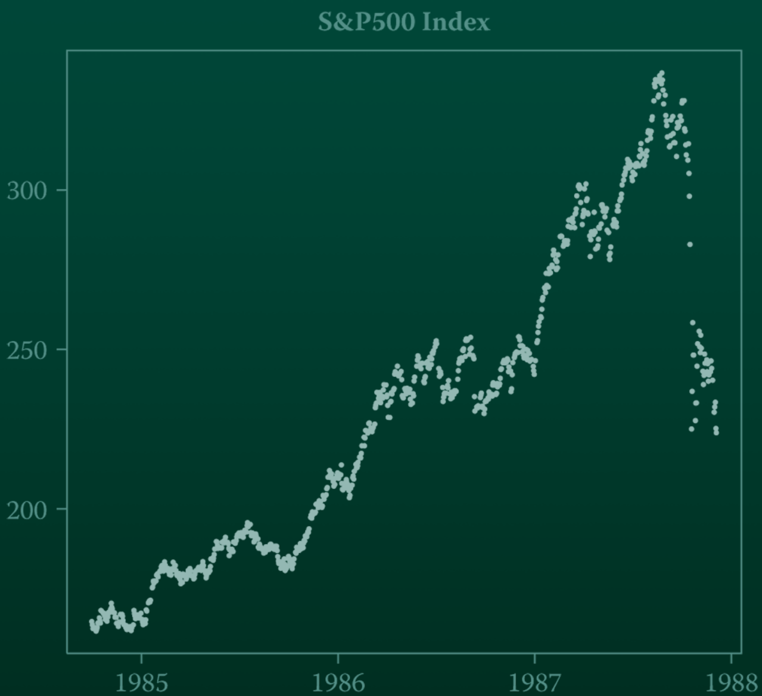
Monographs on Statistics and Applied Probability 122

Extreme Value Methods with Applications to Finance



Serguei Y. Novak

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Extreme Value Methods with Applications to Finance

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

F. Bunea, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong

1 Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2 Queues *D.R. Cox and W.L. Smith* (1961)
3 Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4 The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5 Population Genetics *W.J. Ewens* (1969)
6 Probability, Statistics and Time *M.S. Barlett* (1975)
7 Statistical Inference *S.D. Silvey* (1975)
8 The Analysis of Contingency Tables *B.S. Everitt* (1977)
9 Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10 Stochastic Abundance Models *S. Engen* (1978)
11 Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12 Point Processes *D.R. Cox and V. Isham* (1980)
13 Identification of Outliers *D.M. Hawkins* (1980)
14 Optimal Design *S.D. Silvey* (1980)
15 Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16 Classification *A.D. Gordon* (1981)
17 Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18 Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19 Applications of Queueing Theory, 2nd edition *G.F. Newell* (1982)
20 Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21 Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22 An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23 Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24 Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25 The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26 Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27 Regression Analysis with Applications *G.B. Wetherill* (1986)
28 Sequential Methods in Statistics, 3rd edition *G.B. Wetherill and K.D. Glazebrook* (1986)
29 Tensor Methods in Statistics *P. McCullagh* (1987)
30 Transformation and Weighting in Regression *R.J. Carroll and D. Ruppert* (1988)
31 Asymptotic Techniques for Use in Statistics *O.E. Bandorff-Nielsen and D.R. Cox* (1989)
32 Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)
33 Analysis of Infectious Disease Data *N.G. Becker* (1989)
34 Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)
35 Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)
36 Symmetric Multivariate and Related Distributions *K.T. Fang, S. Kotz and K.W. Ng* (1990)
37 Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)
38 Cyclic and Computer Generated Designs, 2nd edition *J.A. John and E.R. Williams* (1995)
39 Analog Estimation Methods in Econometrics *C.F. Manski* (1988)
40 Subset Selection in Regression *A.J. Miller* (1990)
41 Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)
42 Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)
43 Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)
44 Inspection Errors for Attributes in Quality Control *N.L. Johnson, S. Kotz and X. Wu* (1991)
45 The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)

46 The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)
47 Longitudinal Data with Serial Correlation—A State-Space Approach*R.H. Jones* (1993)
48 Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)
49 Markov Models and Optimization *M.H.A. Davis* (1993)
50 Networks and Chaos—Statistical and Probabilistic Aspects *O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall* (1993)
51 Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)
52 Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)
53 Practical Risk Theory for Actuaries *C.D. Daykin, T. Pentikäinen and M. Pesonen* (1994)
54 Biplots *J.C. Gower and D.J. Hand* (1996)
55 Predictive Inference—An Introduction *S. Geisser* (1993)
56 Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)
57 An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)
58 Nonparametric Regression and Generalized Linear Models *P.J. Green and B.W. Silverman* (1994)
59 Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)
60 Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)
61 Statistics for Long Memory Processes *J. Beran* (1995)
62 Nonlinear Models for Repeated Measurement Data *M. Davidian and D.M. Giltinan* (1995)
63 Measurement Error in Nonlinear Models *R.J. Carroll, D. Rupert and L.A. Stefanski* (1995)
64 Analyzing and Modeling Rank Data *J.J. Marden* (1995)
65 Time Series Models—In Econometrics, Finance and Other Fields *D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen* (1996)
66 Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)
67 Multivariate Dependencies—Models, Analysis and Interpretation *D.R. Cox and N. Wermuth* (1996)
68 Statistical Inference—Based on the Likelihood *A. Azzalini* (1996)
69 Bayes and Empirical Bayes Methods for Data Analysis *B.P. Carlin and T.A Louis* (1996)
70 Hidden Markov and Other Models for Discrete-Valued Time Series *I.L. MacDonald and W. Zucchini* (1997)
71 Statistical Evidence—A Likelihood Paradigm *R. Royall* (1997)
72 Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)
73 Multivariate Models and Dependence Concepts *H. Joe* (1997)
74 Theory of Sample Surveys *M.E. Thompson* (1997)
75 Retrial Queues *G. Falin and J.G.C. Templeton* (1997)
76 Theory of Dispersion Models *B. Jørgensen* (1997)
77 Mixed Poisson Processes *J. Grandell* (1997)
78 Variance Components Estimation—Mixed Models, Methodologies and Applications *P.S.R.S. Rao* (1997)
79 Bayesian Methods for Finite Population Sampling *G. Meeden and M. Ghosh* (1997)
80 Stochastic Geometry—Likelihood and computation *O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout* (1998)
81 Computer-Assisted Analysis of Mixtures and Applications—Meta-Analysis, Disease Mapping and Others *D. Böhning* (1999)
82 Classification, 2nd edition *A.D. Gordon* (1999)
83 Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)
84 Statistical Aspects of BSE and vCJD—Models for Epidemics *C.A. Donnelly and N.M. Ferguson* (1999)
85 Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)
86 The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)
87 Complex Stochastic Systems *O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg* (2001)
88 Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)

89 Algebraic Statistics—Computational Commutative Algebra in Statistics *G. Pistone, E. Riccomagno and H.P. Wynn* (2001)

90 Analysis of Time Series Structure—SSA and Related Techniques *N. Golyandina, V. Nekrutkin and A.A. Zhigljavsky* (2001)

91 Subjective Probability Models for Lifetimes *Fabio Spizzichino* (2001)

92 Empirical Likelihood *Art B. Owen* (2001)

93 Statistics in the 21st Century *Adrian E. Raftery, Martin A. Tanner, and Martin T. Wells* (2001)

94 Accelerated Life Models: Modeling and Statistical Analysis *Vilijandas Bagdonavicius and Mikhail Nikulin* (2001)

95 Subset Selection in Regression, Second Edition *Alan Miller* (2002)

96 Topics in Modelling of Clustered Data *Marc Aerts, Helena Geys, Geert Molenberghs, and Louise M. Ryan* (2002)

97 Components of Variance *D.R. Cox and P.J. Solomon* (2002)

98 Design and Analysis of Cross-Over Trials, 2nd Edition*Byron Jones and Michael G. Kenward* (2003)

99 Extreme Values in Finance, Telecommunications, and the Environment*Bärbel Finkenstädt and Holger Rootzén* (2003)

100 Statistical Inference and Simulation for Spatial Point Processes *Jesper Møller and Rasmus Plenge Waagepetersen* (2004)

101 Hierarchical Modeling and Analysis for Spatial Data *Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand* (2004)

102 Diagnostic Checks in Time Series *Wai Keung Li* (2004)

103 Stereology for Statisticians *Adrian Baddeley and Eva B. Vedel Jensen* (2004)

104 Gaussian Markov Random Fields: Theory and Applications *Håvard Rue and Leonhard Held* (2005)

105 Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition *Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu* (2006)

106 Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood *Youngjo Lee, John A. Nelder, and Yudi Pawitan* (2006)

107 Statistical Methods for Spatio-Temporal Systems *Bärbel Finkenstädt, Leonhard Held, and Valerie Isham* (2007)

108 Nonlinear Time Series: Semiparametric and Nonparametric Methods *Jiti Gao* (2007)

109 Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis *Michael J. Daniels and Joseph W. Hogan* (2008)

110 Hidden Markov Models for Time Series: An Introduction Using R *Walter Zucchini and Iain L. MacDonald* (2009)

111 ROC Curves for Continuous Data *Wojtek J. Krzanowski and David J. Hand* (2009)

112 Antedependence Models for Longitudinal Data *Dale L. Zimmerman and Vicente A. Núñez-Antón* (2009)

113 Mixed Effects Models for Complex Data *Lang Wu* (2010)

114 Intoduction to Time Series Modeling *Genshiro Kitagawa* (2010)

115 Expansions and Asymptotics for Statistics *Christopher G. Small* (2010)

116 Statistical Inference: An Integrated Bayesian/Likelihood Approach *Murray Aitkin* (2010)

117 Circular and Linear Regression: Fitting Circles and Lines by Least Squares *Nikolai Chernov* (2010)

118 Simultaneous Inference in Regression *Wei Liu* (2010)

119 Robust Nonparametric Statistical Methods, Second Edition *Thomas P. Hettmansperger and Joseph W. McKean* (2011)

120 Statistical Inference: The Minimum Distance Approach *Ayanendranath Basu, Hiroyuki Shioya, and Chanseok Park* (2011)

121 Smoothing Splines: Methods and Applications *Yuedong Wang* (2011)

122 Extreme Value Methods with Applications to Finance *Serguei Y. Novak* (2012)

123 Dynamic Prediction in Clinical Survival Analysis *Hans C. van Houwelingen and Hein Putter* (2012)

Monographs on Statistics and Applied Probability 122

Extreme Value Methods with Applications to Finance

Serguei Y. Novak

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20111020

International Standard Book Number-13: 978-1-4398-3575-3 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To my mother,
Novak Ludmila
Ivanovna.

This page intentionally left blank

Contents

This page intentionally left blank

Preface

Extreme value theory (EVT) deals with extreme (rare) events. Such events/variables are often reported as outliers. In some textbooks one can find recommendations to remove outliers (extremes) as they do not fit the model—in other words, to correct reality if it does not fit the picture.

However, any model is only an approximation of reality. It is not data that are wrong by exhibiting outliers, it is the model that does not fit the data. A statistician would like to "let the data speak for itself": we want to extract information about the unknown distribution making as few assumptions as possible.

Extreme value theory is a part of probability theory and statistics that recognizes the importance of outliers (exceedances of a high threshold). The probabilistic part of EVT describes the limiting distribution of sample extremes and numbers of exceedances of high thresholds. In insurance applications one might be interested in the joint limiting distribution of numbers of exceedances of separate levels. For instance, when a hurricane strikes, insurance companies face a cluster of claims. Claim sizes depend on many factors. The need to describe the features of multilevel clustering of extremes led to the development of the theory of processes of exceedances.

A general process of exceedances takes into account locations of extremes as well as their heights. Any question on the limiting distribution of sample extremes can be answered if a limit theorem for a general process of exceedances is established. We describe the limiting distribution of sample extremes, numbers of exceedances, and processes of exceedances in Chapters 1–8.

The statistical part of EVT is concerned with extracting information related to extreme properties of an unknown distribution from a sample of observations. Consider the situation a typical reinsurance company faces. Let $X_1, \ldots, X_n$ be consecutive claims. A reinsurance company pays $X_k - x$ if the kth claim, X_k, exceeds threshold x. Because x is typically high, the probability of a rare event, $\mathbb{P}(X \geq x)$, is low. However, even if an event is rare, its magnitude can be considerable. For instance, Fig. 0.1 presents the empirical distribution function of Danish fire insurance claims for the period Jan. 1980–Dec. 1990. There were 2156 claims in excess of 1 m Danish kroner (DK), 109 claims in excess of 10 m DK, and 7 claims in excess of 50 m DK; the largest claim was over 263 m.

The important practical question is how to estimate the probability $\mathbb{P}(X \geq x)$. The empirical estimator of $\mathbb{P}(X \geq x)$ is obviously inapplicable, as it would base the inference on very few sample elements. However, the question is of vital importance to insurance companies. We present the method of estimating the tail probability $\mathbb{P}(X \geq x)$ in Chapter 9.

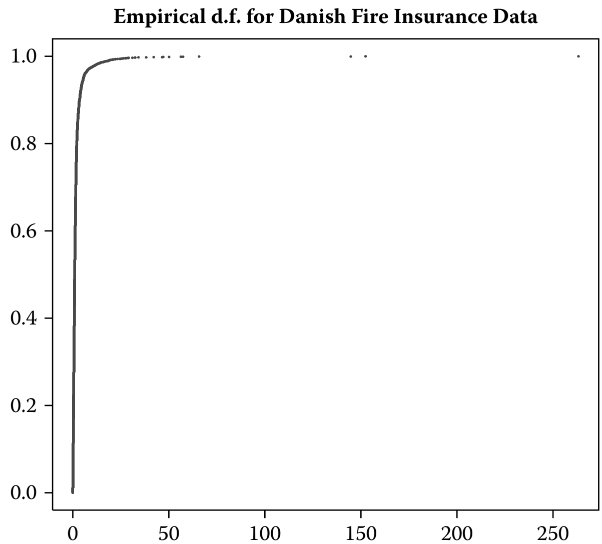


FIGURE 0.1
The empirical distribution function of Danish fire insurance claims.

In finance, an extreme quantile is a popular measure of market risk. A quantile of level 0.05 (5%) or 0.01 (1%) is considered extreme as sample sizes are typically not large and the empirical quantile estimator becomes unreliable.

Regulators require banks to routinely estimate the 1%–quantile of the profit/loss distribution. Using the computed estimate, banks put aside a certain amount of capital to offset market risk. The extreme quantile, known in financial risk management as value-at-risk (VaR), together with the related measure of risk called expected shortfall (ES) or conditional VaR (CVaR), estimate the magnitude of a possible loss.

The 1% level means that a bank aims to offset a possible loss it can face roughly once in 100 days (in the case of daily data). Taking a lower level, one can speak about the magnitude of a worst market crash in decades.

The S&P500 index represents a portfolio of the 500 largest companies in the world. It is a global stock index and is considered a good proxy to the market portfolio.

Figure 0.2 shows the S&P500 index on the eve of the famous "Black Monday" crash in October 1987. On October 19, 1987, the index fell by 20.5%—its worst daily fall from January 1960 until the end of the century. That day alone erased all that the index had gained since March 1986. It took until January 1989 for the index to recover the October 16, 1987 level.

Was it possible to predict the magnitude of the worst market crash in four decades using data available on the eve of Black Monday [222]? The answer is yes.

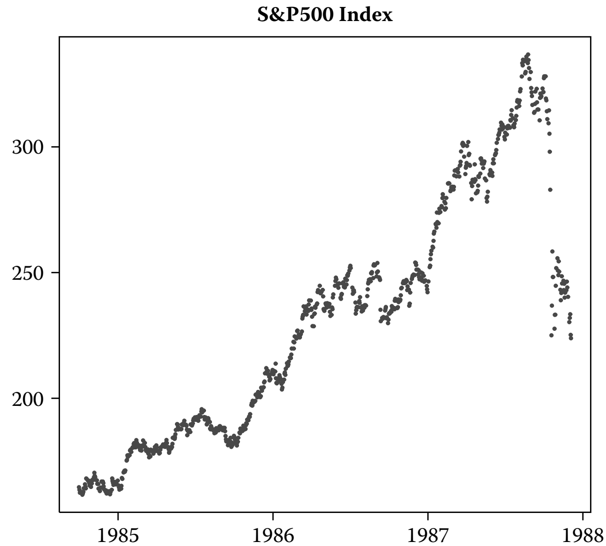


FIGURE 0.2
S&P500 index from October 1984 to December 1987.

We present methods of estimating extreme quantiles, tail probabilities, and measures of market risk in Chapters 9 and 10. The approach is nonparametric (meaning few assumptions on the unknown distribution) and involves an algorithm for choosing a tuning parameter. Concerning the accuracy, the estimate of the magnitude of a worst daily loss of S&P500 in 1960–2000 obtained by the method appears remarkably close to the value of the actual Black Monday fall.

Results in probability and statistics that form a background to Chapters 9 and 10 are collected in Chapters 12 and 13. They are of interest in their own right and can be used in advanced statistical courses. Several miscellaneous and auxiliary results are given in the Appendix.

The book concentrates on the univariate EVT for dependent random variables—the area where the main progress seems to have been achieved during the last two decades. Clustering of extremes was the main phenomenon that fueled the development of the theory that was started by Fisher, Tippett, von Mises, and Gnedenko in the first half of the twentieth century.

The book is intended for PhD and MSc students, data analysts, risk managers, specialists in other branches of probability and statistics who employ certain results from EVT, and all who are interested in EVT and its applications. Parts of the book can be used in lecture courses on extreme value theory, advanced statistical methods, and financial risk management.

This page intentionally left blank

Introduction

Extreme value theory (EVT) has important applications in insurance, finance, hydrology, meteorology, and other fields (cf. [115]). For the case of *independent* random variables (r.v.s) the theory has been fully developed up to the 1970s and is now well presented in a number of textbooks and monographs [115, 132, 202]. These books also cover the situation where r.v.s are dependent but additional restrictions ensure no influence on the asymptotic behavior of extremes.

However, it is now widely accepted that in many applications data are dependent (for instance, this is typical of daily returns of stocks and stock indexes). Dependence causes features, which were not encountered in the classical EVT (e.g., asymptotic clustering of extreme values). This inspired the intensive development of EVT for stationary sequences of random variables.

Many efforts have been made to describe the phenomena of clustering of extremes and develop a rigorous EVT for stationary sequences of r.v.s. However, no book has yet presented a comprehensive theory, as important gaps need to be filled.

This monograph gives a systematic background to a rapidly growing branch of modern probability and statistics: EVT for stationary sequences of random variables.

What You Will Find in This Book

The book is divided into two parts, roughly according to probabilistic and statistical aspects of EVT. Theoretical results are illustrated by examples and applications to particular problems of financial risk management.

Chapter 1 presents basic methods of EVT: Bernstein's "blocks" method, the "runs" approach, and the method of recurrent inequalities. One of those methods is old; the other have evolved during the last two decades.

We investigate the distribution of the Erdös–Rényi maximum of partial sums (MPS) in Chapter 2. MPS is a universal statistic that covers the whole range of statistics from sums to maxima and thus links the limit theory of sums of random variables (LTS) and EVT. Thus, MPS forms a basis of a universal approach with the potential to grow into a general theory combining LTS and EVT.

In Chapter 3 we investigate asymptotics of extreme values in samples of random size. The important particular case is where the sample size is a renewal process. Related problems are those of the length of the longest head run and of the length of the longest match pattern. The problems considered have applications in insurance and statistical analysis of DNA data.

Chapter 4 deals with the number N_n of exceedances of a "high" level. Statistic N_n is the cornerstone of the modern EVT. If data is independent, then N_n has the binomial distribution. It can be well approximated by the Poisson law. Many famous scientists worked on the problem of evaluating the accuracy of Poisson approximation to the binomial distribution. Chapter 4 presents classical as well as new results on the topic.

If data are dependent, then the only possible limiting distribution of the number of exceedances is compound Poisson. Chapter 5 describes the limit theory for N_n. We derive sharp estimates of the accuracy of compound Poisson approximation to $\mathcal{L}(N_n)$.

In insurance/reinsurance applications N_n is the number of claims exceeding a certain level. Over a period of time an insurance company faces a number $N_n(x_1)$ of claims exceeding level x_1, a number $N_n(x_2)$ of claims exceeding level x_2, and so on. Knowing the distribution of the vector $\overline{N}_n = (N_n(x_1), \ldots, N_n(x_m))$ can help deciding on the level of premiums. Chapter 6 is devoted to this topic. We describe the limiting distribution of the vector $\overline{N}_n$ and evaluate the accuracy of compound Poisson approximation to $\mathcal{L}(\overline{N}_n)$.

At the heart of modern EVT lies the notion of the empirical point process of exceedances (EPPE). The key results on the distribution of an EPPE were established by Mori [232], Hsing et al. [167, 169], and the author [257, 262]. A one-dimensional EPPE either counts locations of extremes or their heights. We present results for one-dimensional EPPEs in Chapter 7.

Chapter 8 deals with a general empirical point process of exceedances N_n^* that counts both locations of extremes as well as their heights. We describe the class of possible limiting laws for N_n^* and present necessary and sufficient conditions for the so-called "complete convergence" of N_n^* to a limiting point process. The result can be regarded as an invariance principle for EPPEs. We discuss separately the "central" case where the limiting process is compound Poisson.

The fact that financial/insurance data often exhibit heavy tails is currently the subject of textbooks (see, e.g., [115], p. 404). This is especially common for "frequent" data (e.g., daily log-returns of stock prices). Chapter 9 is devoted to the theory of statistical inference on heavy tails from a sample of dependent data. The main characteristic describing the heavy tail is the so-called "tail index." The chapter deals with the problems of nonparametric estimation of the tail index, extreme quantiles, tail probabilities, and second-order indices.

Evaluating financial risks is a problem of particular importance to financial risk management. Popular measures of risk are Value-at-Risk (VaR) and Expected Shortfall (ES), also known as conditional VaR (CVaR). In statistical terminology VaR is an extreme quantile and ES is a corresponding mean excess function. Chapter 10 presents methods of VaR and ES estimation.

We discuss the notion of the extremal index in relation to the distribution of extremes in Chapter 11.

Chapters 12 and 13 provide a background to the Statistics of Extremes. A number of estimators in the Statistics of Extremes belong to the class of self-normalized sums (SNS) of random variables. For instance, Student's statistic and the ratio estimator of the tail index are members of the SNS family. Self-normalized sums are also needed to construct subasymptotic confidence intervals (confidence intervals that take into account estimates of the accuracy of normal approximation). In Chapter 12 we present results on the asymptotics of SNS and evaluate the accuracy of normal approximation to the distribution of SNS.

Lower bounds to the accuracy of estimation may allow one to decide on the efficiency of a particular estimator as well as to compare different estimators. We present nonparametric lower bounds as well as the classical Fréchet–Rao–Cramér inequality in Chapter 13. The bounds are illustrated on particular estimation problems. The results of this chapter are of interest on their own; they can be used in courses on advanced statistical methods.

Useful auxiliary facts are collected in the Appendix, including the results on sums of dependent random variables. An extensive list of references concludes the monograph.

A number of charts have been created using data from Datastream, Interactive Data, Yahoo!, and the R-project.

The author is grateful to N. Bingham and A. Kukush, who read several chapters and made helpful comments and to I.S. Borisov, who read the brief book proposal and gave a positive reply. I'm very grateful to Richard Jones, my colleague at Middlesex University Business School, for his support. I thank everybody who helped me during my work on this book.

S.Y. Novak
London 2010

This page intentionally left blank

List of Conventions

The operation of multiplication is superior to the division.

A^c	complement to set A
$a_n \lesssim b_n$	$a_n \leq b_n(1 + o(1))$
$a_n \gtrsim b_n$	$a_n \geq b_n(1 + o(1))$
$\mathcal{B}(\cdot)$	Borel σ–field
∂B	boundary of the set B
F_c	$1 - F$
$\mathcal{L}(X)$	distribution of a random variable X
Λ	rate function
$\mathcal{R}(\bar{t})$	the class of sequences (14.72)
$x^{(m)}$	$x(x-1)...(x-m+1)$
sum over ∅	zero
$\Rightarrow$	weak convergence
$\mathbf{B}(p)$	Bernoulli distribution
$\mathbf{B}(n,\, p)$	Binomial distribution
$\mathbf{E}(a)$	Exponential distribution
$\mathbf{\Gamma}(p)$	Geometric distribution
$\mathbf{K}(0;1)$	Cauchy distribution
$\mathbf{\Pi}(\lambda,\, \zeta)$	Compound Poisson distribution
$\mathbf{E}(a)$	Exponential distribution
$\mathcal{N}(\mu;\sigma^2)$	Normal (gaussian) distribution
$\mathbf{\Pi}(\lambda)$	Poisson distribution

This page intentionally left blank

List of Abbreviations

The operation of multiplication is superior to the division.

ACI	asymptotic confidence interval
AMSE	asymptotic mean–squared error
AR	autoregressive model
a.s.	almost surely
CLT	central limit theorem
CP	compound Poisson
CPS	common probability space
CVaR	conditional Value-at-Risk
c.f.	characteristic function
$\mathbb{D}$	variance
Φ	standard normal d.f.
d.f.	distribution function
ES	Expected Shortfall
EVT	Extreme Value Theory
$\mathbb{E}$	mathematical expectation
EMA	exponential moving average
EMH	efficient market hypothesis
EPPE	empirical point process of exceedances
i.i.d.	independent and identically distributed
i.o.	infinitely often
$K_*,\ K^*$	left and right end-points of a distribution
LLN	law of large numbers
LLHR	length of the longest head run
LLMP	length of the longest match pattern
LTS	Limit Theory of Sums of random variables
MA	moving average
MACD	moving average convergence/divergence
MPS	maximum of partial sums
MSE	mean–squared error
NDA	domain of attraction of a normal law
OHLC	Open-High-Low-Close
POT	peak-over-threshold
RE	ratio estimator
RSI	relative strength index
r.v.	random variable
SACI	subasymptotic confidence interval
SLLN	strong law of large numbers
SNS	self-normalized sum

TVD	total variation distance
UOS	upper order statistics
VaR	Value-at-Risk
w.p. 1	with probability one

Author

Dr S.Y. Novak earned his Ph.D. at the Novosibirsk Institute of Mathematics under the supervision of Dr S.A. Utev in 1988. The Novosibirsk group forms a part of Russian tradition in Probability and Statistics that extends its roots to Kolmogorov and Markov.

Dr S.Y. Novak began his teaching carrier at the Novosibirsk Electrotechnical Institute (NETI) and Novosibirsk Institute of Geodesy, held post-doctoral positions at the University of Sussex and Eurandom (Technical University of Eindhoven), and taught at Brunel University of West London, before joining the Middlesex University (London) in 2003. He published over 40 papers, mostly on the topic of Extreme Value Theory, in which he is considered an expert.

This page intentionally left blank

Part I

Distribution of Extremes

This page intentionally left blank

1

Methods of Extreme Value Theory

There are three kinds of lies: lies, damned lies, and statistics.

Mark Twain
on official statistics

CONTENTS

This chapter overviews the methods of extreme value theory (EVT). Section 1.1 presents a number of results on upper-order statistics. Sections 1.2 and 1.3 are devoted to the "blocks" approach and the method of recurrent inequalities.

1.1 Order Statistics

Let $X, X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Rewrite the sample $X_1, \ldots, X_n$ in nonincreasing order:

$$X_{1,n} \geq \cdots \geq X_{n,n}. \tag{1.1}$$

Random variables (1.1) are called *order statistics*.

$$M_n = X_{1,n}$$

is the sample maximum, and $X_{k,n}$ is called the kth *maximum*.

Denote

$$N_n(x) = \sum_{i=1}^{n} \mathbb{1}\{X_i > x\}.$$

The random variable $N_n(x)$ is called the *number of exceedances over the threshold* x. It is easy to see that $(1 \leq m \leq n)$

$$\{X_{m,n} \leq x\} = \{N_n(x) < m\}. \tag{1.2}$$

This entails the representation

$$X_{m,n} = \min\{x : N_n(x) < m\}.$$

From (1.2),

$$\mathbb{P}(X_{m,n} \leq x) = \sum_{k=0}^{m-1} \binom{n}{k} \mathbb{P}^k(X > x)\mathbb{P}^{n-k}(X \leq x).$$

Choosing $x = x(n)$ in such a way that

$$y := n\mathbb{P}(X > x)$$

is bounded away from 0 and ∞, we derive the asymptotic representation

$$\mathbb{P}(X_{m,n} \leq x) \approx e^{-y} \sum_{k=0}^{m-1} y^k/k!$$

The following inequality hints that the tail of $\mathcal{L}(X_{i,n})$ is lighter than that of $\mathcal{L}(X_{j,n})$ as $i > j$: if $X, X_1, X_2, \ldots$ are i.i.d.r.v.s, then

$$\mathbb{P}(X_{m,n} > x) \leq (n\mathbb{P}(X > x))^m/m! \tag{1.3}$$

Denote

$$M_n^+ = \max_{1 \leq i \leq n} |X_i|, \;\; S_n = X_1 + \cdots + X_n.$$

The following proposition compares the tails of M_n^+ and $|S_n|$.

Proposition 1.1 *If r.v.s $\{X_i\}$ are symmetric, then for any $x > 0$, $n \geq 1$,*

$$\mathbb{P}\big(M_n^+ > x\big) \leq 2\mathbb{P}(|S_n| > x). \tag{1.4}$$

By the Khintchin–Kolmogorov strong law of large numbers (SLLN),

$$S_n/n \to const \quad (a.s.)$$

if and only if $\mathbb{E}|X| < \infty$. A similar result holds for M_n^+.

Lemma 1.2 *For any $r > 0$,*

$$M_n^+/n^{1/r} \to 0$$

a.s. if and only if $\mathbb{E}|X|^r < \infty$;

$$M_n^+/n^{1/r} \xrightarrow[p]{} 0$$

if and only if $x^r\mathbb{P}(|X| > x) \to 0$ as $x \to \infty$.

Lemma 1.2 follows from the following fact.

Proposition 1.3 *If $\{x_n\}$ is a nondecreasing sequence of numbers, then*

$$\mathbb{P}(M_n > x_n \; i.o.) = \mathbb{P}(X_n > x_n \; i.o.) = 0 \text{ or } 1$$

depending on whether $\sum_n \mathbb{P}(X > x_n) < \infty$ *or* $\sum_n \mathbb{P}(X > x_n) = \infty$.

If $\mathbb{E}|X| < \infty$, then $\sum_{n\geq 1} \mathbb{P}(|X| > n) < \infty$, and hence

$$\lim_{n\to\infty} M_n^+/n = 0 \quad (a.s.).$$

Denote

$$R_n(t) = \max_{1\leq i\leq n} |X_i|^t / \sum_{i=1}^{n} |X_i|^t. \tag{1.5}$$

Proposition 1.4 *As $n \to \infty$,*

$$R_n(t) \to 0 \; a.s. \;\Leftrightarrow\; \mathbb{E}|X|^t < \infty, \tag{1.6}$$

$$R_n(t) \underset{p}{\to} 0 \;\Leftrightarrow\; \mathbb{E}|X|^t \mathbb{1}\{|X| \leq x\} \text{ is slowly varying}, \tag{1.7}$$

$$R_n(t) \underset{p}{\to} 1 \;\Leftrightarrow\; \mathbb{P}(|X| > x) \text{ is slowly varying}. \tag{1.8}$$

Thus, $M_n^+/|S_n|$ is asymptotically "small" if $\mathbb{E}|X| < \infty$, whereas M_n^+ is comparable to $|S_n|$ if $\mathbb{P}(|X| > x)$ is slowly varying (see also (1.17)).

Assume that $\{X_i\}$ are i.i.d.r.v.s. Then $(X_{1,n}, \ldots, X_{n,n})$ admit the representation

$$(X_{1,n}, \ldots, X_{n,n}) \stackrel{d}{=} \left(F_c^{-1}(T_1/T_{n+1}), \ldots, F_c^{-1}(T_n/T_{n+1})\right), \tag{1.9}$$

where

$$F_c = 1 - F,$$

$T_m = \eta_1 + \cdots + \eta_m$ and $\{\eta_i\}$ are i.i.d.r.v.s with an exponential $\mathbf{E}(1)$ distribution. Applying (1.9) to the uniform $\mathbf{U}(0;1)$ distribution, we get the following representation for the corresponding order statistics $U_{1,n} \geq \cdots \geq U_{n,n}$:

$$(U_{1,n}, \ldots, U_{n,n}) \stackrel{d}{=} (T_n/T_{n+1}, \ldots, T_1/T_{n+1}). \tag{1.10}$$

Example 1.1 If $\mathcal{L}(X) = \mathbf{E}(1)$, then (1.9) entails

$$(X_{i,n} - X_{k+1,n})_{1\leq i\leq k} \stackrel{d}{=} (X_{1,k}, \ldots, X_{k,k}). \tag{1.11}$$

If X has a Pareto distribution, that is,

$$F_c(x) = Cx^{-\alpha} \qquad (x > C^{1/\alpha}),$$

where $\alpha \in (0; \infty)$ and $C > 0$, then

$$(X_{1,n}/X_{k+1,n}, \ldots, X_{k,n}/X_{k+1,n}) \stackrel{d}{=} (X_{1,k}, \ldots, X_{k,k}) \qquad (1 \leq k < n). \tag{1.12}$$

□

Because

$$N_n(X_{k+1,n}) = k,$$

the order statistic $X_{k+1,n}$ is the empirical quantile of level $1 - k/n$. We will call $X_{k+1,n}$ the *empirical upper quantile* of level k/n.

Let $k = k(n)$ depend on n. The following cases have been intensively studied (see, e.g., [315, 352]):

(a) $k/n \to const \in (0; 1)$,
(b) $k \to \infty$, $k/n \to 0$.

Theorem 1.5 is a consequence of (1.9). It shows that the empirical quantile is a proper tool when estimating nonextreme quantiles.

Theorem 1.5 *Suppose that $k/n \to q \in (0; 1)$. If F is continuously differentiable at $x_q = F_c^{-1}(q)$ and $f := F'$, then*

$$(X_{k,n} - x_q) f(x_q) \sqrt{n/q(1-q)} \Rightarrow \mathcal{N}(0; 1). \tag{1.13}$$

According to Theorem 1.5,

$$X_{k,n} = x_q + \frac{\sqrt{q(1-q)}}{f(x_q)\sqrt{n}} \xi_n,$$

where $\xi_n \Rightarrow \mathcal{N}(0; 1)$.

Example 1.2 Let $F_c(x) = 1/x$, $x > 1$, and let $q = 0.01$. Then $x_q = 100$ is the 1%-upper quantile and $X_{k+1,n}$ with $k = [n/100]$ is the empirical upper quantile of level 0.01. Theorem 1.5 states that

$$X_{k,n} = x_q + 1000 \xi_n \sqrt{0.99/n},$$

where $\xi_n \Rightarrow \mathcal{N}(0; 1)$. Obviously, the sample size n must be very large in order to compensate the factor 1000. Thus, the empirical quantile does not appear to be a proper estimate of extreme quantiles. □

We now describe the limiting distribution of $X_{k,n}$ in situation (b) assuming that $\mathcal{L}(X)$ has a heavy right tail:

$$\mathbb{P}(X > x) = L(x) x^{-\alpha}, \tag{1.14}$$

where $\alpha > 0$ and L is a slowly varying function.

Theorem 1.6 *Let $X, X_1, X_2, \ldots$ be i.i.d.r.v.s with the heavy-tailed distribution (1.14). If $k = k(n) \to \infty$, $k/n \to 0$ and F_c is strictly monotone, then*

$$\sqrt{k}\big(X_{k,n}/F_c^{-1}(k/n) - 1\big) \Rightarrow \mathcal{N}(0; \alpha^{-2}). \tag{1.15}$$

Supplements

1. Let $X, X_1, X_2, \ldots$ be i.i.d.r.v.s with a d.f. F. If $\mathbb{E}\max\{X; 0\} \in (0; \infty)$, then for any positive constant b

$$M := \sup_{n \geq 1}\{X_n - (n-1)b\} < \infty \qquad (a.s.). \tag{1.16}$$

If $\lim_{x\to\infty} F_c(x+1)/F_c(x) = 1$, then

$$\mathbb{P}(M > x) \sim \frac{1}{b}\int_x^\infty F_c(y)dy \qquad (x \to \infty).$$

This relation remains valid if $(n-1)b$ in (1.16) is replaced with $\xi_1 + \cdots + \xi_{n-1}$, where $(X, \xi), (X_1, \xi_1), \ldots$ are i.i.d. pairs of r.v.s, $\mathbb{E}\xi = b$ [10, 164, 294, 329].

2. Let $X, X_1, X_2, \ldots$ be i.i.d. nonnegative heavy-tailed r.v.s obeying (1.14). Darling [86] has shown that if $\alpha \in (0; 1)$, then

$$\mathbb{E} S_n/M_n \to 1/(1-\alpha) \qquad (n \to \infty). \tag{1.17}$$

Exercises

1. Prove (1.3).
2. Check (1.9) in the case $n = 1$.
3. Prove (1.11) and (1.12). Derive (1.13).
4. Let $\{U_i\}$ be uniform $\mathbf{U}(0; 1)$ i.i.d.r.v.s. Prove that

$$\max_{1\leq i\leq n} U_i \stackrel{d}{=} \sum_{j=1}^{n} \eta_j \Big/ \sum_{j=1}^{n+1} \eta_j,$$

where $\{\eta_j\}$ are exponential $\mathbf{E}(1)$ i.i.d.r.v.s.

5. Assume conditions of Theorem 1.6. Show

$$\hat{a}_n = (\ln X_{k,n})/\ln(n/k)$$

is a consistent estimator of index a. What can you say about the accuracy of approximation $\hat{a}_n \approx a$?

6. Let $X \in \mathbf{U}(0; 1)$ and $y \equiv y_n = o(1/n)$. Show that

$$\mathbb{P}(X_{1,n} > 1 - y) \sim ny, \ \mathbb{P}(X_{2,n} > 1 - y) \sim (ny)^2/2.$$

7. Assume that $\mathcal{L}(X)$ has a continuous d.f.. Show that

$$\mathbb{P}(X_n = M_n) = \mathbb{P}(X_n \geq M_{n-1}) = 1/n.$$

8. Check that

$$\mathbb{P}(X_n \in dx | M_n = y) = \frac{n-1}{n} \frac{dF(x)}{F(y)} \mathbb{1}\{x < y\} + \frac{1}{n} \delta_y(dx), \tag{1.18}$$

where $\delta_y(\cdot)$ is the unit measure concentrated at point y.

9. Derive from (1.18) that

$$\mathbb{E}\{X_1 | M_n = y\} = (1 - 1/n)\mathbb{E}\{X | X < y\} + y/n.$$

As a consequence, if $\mathcal{L}(X) = \mathbf{U}(0; 1)$, then

$$2\mathbb{E}\{X_1 | M_n\} = (1 + 1/n)M_n.$$

10. Let $F_c(x) = Cx^{-\alpha}$ as $x > C^{1/\alpha}$, where $\alpha \in (0; \infty)$, $C > 0$. Show that Hill's estimator of the tail index is consistent: if $n > k = k(n) \to \infty$, then

$$k \Big/ \sum_{i=1}^{k} \ln(X_{i,n}/X_{k+1,n}) \xrightarrow[p]{} \alpha. \tag{1.19}$$

11. Let $\mathcal{L}(X)$ have a continuous d.f. F. Check that $F(X_{k,n})$ has beta $\mathbf{B}(k, n-k+1)$ distribution.

12. Suppose that $X_i = \max\{\xi_i; \xi_{i+1}\}$, where ξ_i is a sequence of i.i.d.r.v.s $\{\xi_i\}$ with a continuous distribution function. Check that $\mathbb{P}(X_{1,n} = X_{2,n}) = 1 - 2/(n+1)$.

13. Show that

$$\mathbb{E}M_n/S_n \geq 1 - \alpha + o(1) \qquad (n \to \infty) \tag{1.20}$$

in the assumptions of supplement 2.

1.2 "Blocks" and "Runs" Approaches

Bernstein's "blocks" method is probably the most universal tool in EVT. It was originally developed for proving limit theorems for sums of dependent r.v.s.

In this chapter we apply the approach to the sample maximum and the number of exceedances. The asymptotics of empirical point processes in Chapter 5 is also studied using the blocks method.

The idea of the approach is simple: split the sample of size n into blocks of lengths $r = r(n)$, $1 \ll r \ll n$, and subtract subblocks of lengths $l = l(n) \ll r$. Then the "reduced" blocks are almost independent.

The weak point of the method is its poor accuracy of approximation (cf. Remark 1.1 below).

In Theorem 1.7 and Corollaries 1.8 and 1.9, we assume mixing condition $(D\{u_n\})$ (see the Appendix). Recall that $[x]$ denotes the integer part of x and $\{x\}$ is the fractional part of x. Denote

$$u = u_n\,, \;\; p = \mathbb{P}(X > u),$$

and let $\alpha_n(\cdot)$ be the α-mixing coefficient of the sequence $\{\mathbb{1}\{X_i > u\}, i \geq 1\}$.

Theorem 1.7 *If* $1 \leq l < r \leq n$, *then*

$$\left|\mathbb{P}(M_n \leq u) - \mathbb{P}^{n/r}(M_r \leq u)\right| \leq \mathbb{P}(M_{r\{n/r\}} > u) + (\alpha_n(l) + 2lp)n/r + (e[n/r])^{-1}. \tag{1.21}$$

If $1 \leq l \leq n/k \leq n$, *then*

$$\left|\mathbb{P}(M_n \leq u) - \mathbb{P}^k(M_{[n/k]} \leq u)\right| \leq kp + (\alpha_n(l) + 2lp)k. \tag{1.21*}$$

Let $\{u = u_n\}$ be a sequence of numbers such that

$$p = \mathbb{P}(X > u_n) \to 0 \qquad (n \to \infty).$$

According to Theorem 1.7,

$$\left|\mathbb{P}(M_{[n/k]} \leq u) - \mathbb{P}^{1/k}(M_n \leq u)\right| \leq C_k \alpha_n(l) + o(1) \qquad (\forall k, l \in \mathbb{N})$$

as $n \to \infty$. Taking into account mixing condition $(D\{u_n\})$, we obtain

$$\limsup_{n\to\infty} \left|\mathbb{P}(M_{[n/m]} \leq u) - \mathbb{P}^{k/m}(M_{[n/k]} \leq u)\right| = 0 \qquad (\forall k, m \in \mathbb{N}). \tag{1.22}$$

If

$$\limsup_{n\to\infty} n\mathbb{P}(X > u_n) < \infty, \tag{1.23}$$

then there exist $l = l(n)$, $r = r(n)$ such that $1 \ll l \ll r \ll n$ and

$$\lim_{n\to\infty} \left|\mathbb{P}(M_n \leq u) - \mathbb{P}^{n/r}(M_r \leq u)\right| = 0.$$

Let $\{u_n\}$ be a nondecreasing normalizing sequence in a limit theorem for $\mathcal{L}(M_n)$, that is,

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n) = e^{-\lambda} \qquad (\exists \lambda > 0). \tag{1.24}$$

Denote $u_n(t) = u_{[n\lambda/t]}$. If (1.23), (1.24), and mixing condition $\Delta\{u_n\}$ hold, then

$$\lim_{n\to\infty} \mathbb{P}(M_n \leq u_n(t)) = e^{-t} \qquad (\forall t > 0). \tag{1.25}$$

Denote by $\hat{M}_n$ the maximum of n independent copies of X. A well-known theorem by Gnedenko [147] describes the class of possible limit laws for

$b_n(\hat{M}_n - a_n)$ for properly chosen sequences of constants $\{a_n\}$ and $\{b_n\}$. Gnedenko's theorem is valid in the case of weakly dependent r.v.s as well.

Corollary 1.8 *If constants a_n, $b_n > 0$ are chosen so that $b_n(\hat{M}_n - a_n)$ converges weakly to a nondegenerate distribution P, then P belongs to one of the three types of extreme value distributions:*

$$\begin{array}{ll} \textit{Fréchet: } F_{_F}(x) = \exp(-x^{-\alpha}) & (x > 0, \alpha > 0) \\ \textit{Weibull: } F_{_W}(x) = \exp(-|x|^{\alpha}) & (x < 0, \alpha > 0) \\ \textit{Gumbel: } F_{_G}(x) = \exp(-e^{-x}) & (x \in \mathbb{R}) \end{array}$$

Index α is called sometimes the extreme value index.

Corollary 1.9 *If (1.23) holds and*

$$\lim_{k\to\infty} \limsup_{n\to\infty} \sum_{i=1}^{[n/k]} \mathbb{P}(X_{i+1} > u_n | X_1 > u_n) = 0, \qquad (D')$$

then

$$\lim_{n\to\infty} \left| \mathbb{P}(M_n \le u) - \mathbb{P}^n(X > u) \right| = 0.$$

Condition (D') means that a cluster can asymptotically contain only one element. In other words, clustering of extremes is prohibited, and asymptotic behavior of M_n is similar to that of the maximum of n independent copies of X.

The blocks method competes with the "runs" approach initiated by Newell (see also O'Brien [245, 282, 285]). The idea of the runs approach is to consider a new cluster of exceedances starting at a point i if

$$X_i > u,\ X_{i-1} \le u, \ldots, X_{i-r} \le u_n.$$

Intuitively, the runs approach must provide a better accuracy of approximation. Indeed, we count only those blocks that contain at least one extreme value (exceedance over the level u), whereas with the blocks approach we count all blocks of length r (including those without any "extreme" value at all).

Another powerful tool is the *method of generating functions*. The functions

$$g_X(t) = \mathbb{E}\exp(tX),\ \ h_X(t) = \mathbb{E}t^X$$

are called the *moment generating functions*. Properties of g_X and h_X are similar to those of characteristic functions. We call

$$\sum_{k\ge 0} \mathbb{P}(M_n \le k)t^k,\ \sum_{n\ge 1} \mathbb{P}(M_n \le k)t^n$$

the *generating functions* (provided the series converge). In some particular situations they can be found explicitly (e.g., [149, 248]). Further analysis can yield asymptotic expansions in a limit theorem for M_n, as it is done in the case of the length of the longest head run (Chapter 7). The Stein method is presented in Chapters 2 and 10.

Open Problem

1.1. The "runs" method has been worked out for a sample maximum only. The open problem is to apply this method to sample extremes, numbers of exceedances, and processes of exceedances. In order to do this, one would require a renewal theory for dependent r.v.s, which is not well developed yet.

Exercises

14. Prove (1.25).

15. Assuming mixing condition $(D\{u_n\})$, show that (1.24) entails

$$\mathbb{P}(X > u_n) \to 0.$$

16. Assume that the sequence $\{X_i\}$ is φ–mixing. Prove that (1.23) follows from (1.24).

17. Assume $(D\{u_n\})$. Prove that (1.24) and (1.23) yield

$$\liminf_{n\to\infty} n\mathbb{P}(X_n > u_n) > 0. \tag{1.26}$$

1.3 Method of Recurrent Inequalities

Close to the runs approach is the *the method of recurrent inequalities*. It suggests composing and solving recurrent inequalities for $\mathbb{P}(M_n \le x)$ and other quantities of interest. Applied to particular problems of EVT, it yields correct rates of convergence in the corresponding limit theorems.

To demonstrate the idea of the method, assume that

$$p := \mathbb{P}(X > u) > 0$$

and denote

$$b \equiv b(r, u) = \mathbb{P}(B_r),$$

where

$$B_n = \{X_n > u,\, X_{n-1} \le u, \ldots, X_{n-r+1} \le u\}$$

if $r > 1$, $B_n = \{X_n > u\}$ if $r = 1$. Observe that

$$\{M_n \le u\} = \{M_{n-1} \le u\} \setminus \{M_{n-r} \le u,\, B_n\}.$$

Thus,

$$P_n := \mathbb{P}(M_n \le u) = \mathbb{P}(M_{n-1} \le u) - \mathbb{P}(M_{n-r} \le u,\, B_n). \tag{1.27}$$

Events $\{M_{n-r} \le u\}$ and B_n are usually "almost independent." Therefore,

$$P_n \approx P_{n-1} - b\,P_{n-r} \approx (1-b)\,P_{n-1} = \cdots = (1-b)^{n-r}\,P_r.$$

As P_r is typically close to 1,

$$\mathbb{P}(M_n \le u) \approx e^{-nb}. \tag{1.28}$$

The following theorem makes (1.28) more precise in the case of m-dependent r.v.s.

Theorem 1.10 *If the random variables $\{X_i, i \ge 1\}$ are $(m-1)$-dependent, $8mb \le 1$ and $n > 4m$, then*

$$\begin{aligned}(1-b)^{n-4m} - 2m(b+2p) &\le \mathbb{P}(M_n \le u)\\ &\le e^{-(n-4m)b} + (e^{-1} + 4mb)mp/(1-mp).\end{aligned} \tag{1.29}$$

Remark 1.1 Relation (1.29) is established by the method of recurrent inequalities. It implies that the rate of approximation (1.28) is $O(n^{-1} + p)$. Concerning the approximation by the blocks method, it is natural to put $l = m$ in (1.21) in the case of m-dependent r.v.s. Then $r = \sqrt{2mn}$ minimizes the right-hand side of (1.21), and the rate of approximation is $O(n^{-1/2} + n^{1/2}p)$. Thus, the method of recurrent inequalities appears more accurate.

Remark 1.2 Condition $0 < \mathbb{P}(X > u_n) \to 0$ as $n \to \infty$ seems to be natural. If $\mathbb{P}(X > u_n) = 0$, then $\mathbb{P}(M_n \le u_n) = 1$. On the other hand, if $\liminf_{n\to\infty} \mathbb{P}(X > u_n) > 0$ and $\lim_{l\to\infty} \varphi(l) = 0$, where φ is the mixing coefficient, then Lemma 1.14 entails $\lim_{n\to\infty} \mathbb{P}(M_n \le u_n) = 0$.

Comparing (1.28) with (1.21), one can conclude that

$$b(r,u) \approx \mathbb{P}(M_r > u)/r.$$

The following lemma makes this observation more precise.

Lemma 1.11 *If $1 \le i \le r$ and $0 \le l < m$, then*

$$\frac{\mathbb{P}(M_i > u)}{i} \ge b(r,u) \ge \frac{\mathbb{P}(M_{r+m} > u) - \mathbb{P}(M_{r+l} > u)}{m-l}. \tag{1.30}$$

In particular, this inequality with $i = m = r$ and $l = 0$ yields

$$\mathbb{P}(M_r > u)/r \ge b(r,u) \ge (\mathbb{P}(M_{2r} > u) - \mathbb{P}(M_r > u))/r. \tag{1.31}$$

The following theorem describes the asymptotics of the sample maximum in the case of a stationary φ-mixing sequence of r.v.s.

Let $\mu \equiv \mu(r,u) = (1+\sqrt{1-4(r+l)b})/2,\ R_n = 0 \quad \text{if} \quad 4(r+l)b > 1,$

$$\begin{aligned} R_n &\equiv R_n(r,u,l) = \mu^{[\frac{n}{r+l}]}(\mu - (r+l)p) - \varphi(l) \quad \text{if} \quad 4(r+l)b \le 1, \\ Q_n &\equiv Q_n(r,u,l) = (1-b)^{n-2r-l}(1-(2r+l)p) - (r+l)b - \varphi(l), \\ V_n &\equiv V_n(r,u,l) = (1-rb)^{[n/(r+l)]} + \varphi(l). \end{aligned}$$

Quantities R_n, Q_n, V_n approximate e^{-nb}.

Theorem 1.12 *If* $r, l \in \mathbb{N},\ r > 1, r+l \le n$ *and* $n > 3r+2l$, *then*

$$\max\{R_n; Q_n\} \le \mathbb{P}(M_n \le u) \le V_n. \tag{1.32}$$

Theorem 1.12 justifies (1.28) and provides a basis for further results on the distribution of the sample maximum in φ-mixing sequences.

More general than (D') is Watson's [392] condition

$$\lim_{n\to\infty} \mathbb{P}(X_{i+1} > u_n | X_1 > u_n) = 0 \qquad (\forall i \in \mathbb{N}). \tag{1.33}$$

Example 1.3 of a stationary sequence $\{X_i\}$ that obeys (1.33) but not (D'). Let $\tau, \tau_1, \tau_2, \ldots$ and $\{Y_i\}$ be independent sequences of i.i.d.r.v.s, τ takes values in $\mathbb{N}$, Y takes values in $[1;\infty)$, and $\mathbb{P}(\tau = 1) < 1$. We put $T_0 = 0$, $T_j = \sum_{l=1}^{j} \tau_l$, and let $X_i = \sum_{j\ge 1} Y_j \mathbb{1}\{T_{j-1} < i \le T_j\} = Y_{\nu(i)}$, where $\nu(i) = \min\{k : T_k \ge i\}$. $\square$

Theorem 1.13 *Assume that* $\lim_{l\to\infty} \varphi(l) = 0$ *and*

$$0 < \liminf_{n\to\infty} n\mathbb{P}(X > u_n) \le \limsup_{n\to\infty} n\mathbb{P}(X > u_n) < \infty. \tag{1.34}$$

Then

$$\mathbb{P}(M_n \le u_n) - \exp(-n\mathbb{P}(X > u_n)) \to 0 \tag{1.35}$$

if and only if (1.33) holds.

According to Theorem 1.13, if Watson's condition holds, then the limiting distribution of the sample maximum of a stationary φ-mixing sequence of r.v.s is the same as if the sample elements were independent.

Exercise

18. Check that $\mathbb{P}(X_2 > u_n | X_1 > u_n) > 0$ in Example 1.3. Thus, Watson's condition holds, while (D') does not.