Statistical Techniques for Modelling Extreme Value Data and Related Applications

Statistical Techniques for Modelling Extreme Value Data and Related Applications

Bу

Haroon M. Barakat, El-Sayed M. Nigm and Osama M. Khaled

Cambridge Scholars Publishing



Statistical Techniques for Modelling Extreme Value Data and Related Applications

By Haroon M. Barakat, El-Sayed M. Nigm and Osama M. Khaled

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data A catalogue record for this book is available from the British Library

Copyright o 2019 by Haroon M. Barakat, El-Sayed M. Nigm and Osama M. Khaled

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-3207-0 ISBN (13): 978-1-5275-3207-6

Contents

	Pret	face	x
	Note	ations and abbreviations	xii
	List	of illustrations	XV
	List	of tables	xvii
1	Intr	oduction: Some basic and miscellaneous results	1
	1.1	The convergence concept in probability theory	1
		1.1.1 Modes of convergence of RVs	2
		1.1.2 Further limit theorems on weak convergence	4
	1.2	Statistical methods	6
		1.2.1 Maximum likelihood method	6
		1.2.2 Kolmogorov-Smirnov (K-S) test	9
		1.2.3 Genetic algorithms (GA)	11
	1.3	Bootstrap technique	13
2	Asy	mptotic theory of order statistics: A historical ret-	
	\mathbf{rosp}	pective	18
	2.1	Order statistics	18
	2.2	Extreme value theory under linear normalization	20
	2.3	Max-domains of attraction of univariate ℓ -max-stable	
		laws	26
	2.4	Limit theory of intermediate order statistics under	
		linear normalization	30
		2.4.1 Asymptotic theory of intermediate order statistics	30
		2.4.2 Domains of attraction of intermediate limit laws	32
	2.5	Central order statistics—domains of attraction of	
		central limit laws	33
		2.5.1 Asymptotic theory of central order statistics	33
		2.5.2 Domains of attraction of central limit laws	36

Contents

2.6	Asymp	ptotic theory of extremes under nonlinear normal-			
	ization	l	37		
	2.6.1	Characterization of the class of ML-laws and the			
		GMA group	39		
	2.6.2	The class of ML-laws	39		
	2.6.3	The class of max-stable laws (MS-laws)	42		
2.7	Comm	ents on Pancheva's work	43		
2.8	Extrem	ne value theory under power normalization	45		
2.9	Max-d	omains of attraction of univariate p-max-stable			
	laws		47		
2.10	Compa	arison between domains of attraction	49		
2.11	Asymp	ototic central order statistics under nonlinear			
	norma	lization	50		
	2.11.1	The class of weak limits of central order statistics			
		under general normalization	51		
	2.11.2	Asymptotic central order statistics under power			
		normalization	54		
	2.11.3	Examples	55		
	2.11.4	Comparisons between the domains of attraction			
		of weak limits of central order statistics under			
		linear and power normalizing constants	57		
2.12	Asym	ototic intermediate order statistics under nonlin-			
	ear no	rmalization	57		
	2.12.1	The class of weak limits of intermediate order			
		statistics under general normalization	58		
	2.12.2	The domains of attraction of the lower interme-			
0.10		diate power types	62		
2.13	Asym	btotic theory of order statistics under exponential			
0.1.4	norma	lization	64		
2.14	Generalized order statistics and dual generalized order				
	statist	ICS	68 60		
	2.14.1	Distribution theory of the m-gos and m-dgos	69		
	2.14.2	Asymptotic theory of univariate $m-\text{gos}$ and $m-\text{drog}$	71		
	0149	m-ugos	(1 75		
9 1 E	2.14.3 Doctor	ruther minit theorems of gos and agos	61 77		
2.10	Restric	cieu convergence	11		
2.10	reviev	v of extreme value analysis in environmental studies	79		

vi

3	Boo	otstrap	order statistics and calibration of the sub-		
	sam	ple bo	otstrap method	83	
	3.1	Boots	trapping extremes under linear normalization	84	
		3.1.1	Convergence bootstrap distributions when the		
			normalizing constants are known	84	
		3.1.2	Convergence bootstrap distributions when a_n		
			and b_n are unknown	85	
	3.2	Boots	trapping extremes under power normalization	86	
		3.2.1	Convergence bootstrap distributions when the		
			normalizing constants are known	87	
		3.2.2	Convergence bootstrap distributions when the		
			normalizing constants are unknown	88	
	3.3	Verifie	cation of the sub-sample bootstrap method	89	
	3.4	Boots	trapping of order statistics with variable rank in		
		the L-	-model	92	
		3.4.1	Bootstrapping of central order statistics under		
			linear normalization	92	
		3.4.2	Bootstrapping of intermediate order statistics		
			under linear normalization	98	
	3.5	Boots	trapping of order statistics with variable rank in		
		the P-	-model	100	
		3.5.1	Bootstrapping central order statistics under		
			power normalization	100	
		3.5.2	Bootstrapping intermediate order statistics		
			under power normalization	102	
	3.6	Simul	ation study	104	
	3.7	Boots	trapping extreme generalized order statistics	106	
4	Statistical modelling of extreme value data under linear				
	normalization 11				
	4.1	The N	National Environmental Radiation Monitoring		
		Netwo	ork (NERMN) in Egypt	112	
	4.2	Envir	onmental monitoring	113	
	4.3	Chem	ical pollutants	114	
		4.3.1	Particulate matter	114	
		4.3.2	Sulphur dioxide	115	
		4.3.3	Ozone	116	
		4.3.4	Ambient gamma radiation	117	
	4.4	Collec	eted data	118	
	4.5	Data	treatments and simulation study	124	

v	ı	1
_ ^		

		Contents	xiii	
		4.5.1 Mathematical models	124	
	4.6	Data treatments	127	
5	\mathbf{Ext}	reme value modelling under power normalization	139	
	5.1	Generalized extreme value distribution under power		
		normalization	139	
	5.2	Statistical inference using the BM method	141	
	5.3	The GPDP DFs and their related statistical inference	142	
		5.3.1 The derivation of GPDP—The POT stability		
		property	142	
		5.3.2 Estimation of the shape and the scale parameters		
	- .	within the GPDP model	143	
	5.4	Simulation study	145	
	5.5	Parameter estimation for GEVL and GEVP by using	150	
		the GA technique	152	
6	Me	thods of threshold selection	154	
	6.1	Some estimators for the EVI under linear normalization	154	
	6.2	Some methods of threshold selection	159	
		6.2.1 Graphical methods	161	
	6.3	Comparison between $\gamma_{\ell M}$ and $\gamma_{\ell M R}$ via a simulation study		
7	\mathbf{Est}	imations under power normalization for the EVI	167	
•	7.1	Counterparts of Hills estimators under power normal- ization	167	
		7.1.1 Counterparts of HEPs	168	
	7.2	Hill plot under power normalization (HPP)	170	
	7.3	Simulation study	171	
	7.4	Harmonic t-Hill estimator under power normalization	174	
	7.5	Moment and moment-ratio estimators under power		
		normalization	174	
	7.6	Further contemporaneous Hill estimators under power		
		normalization	178	
	7.7	Four HEPs based on GPDP	178	
		7.7.1 Four HEPs that do not have counterparts in the		
		L-model	179	
		7.7.2 Simulation study	180	
	7.8	New Hill plot (NHP)	183	
	7.9	Comparison between estimators under power normal-		
		ization	184	
		7.9.1 The first simulation study	184	

Contents

		7.9.2 The second simulation study	189
	7.10	The weighting between the linear and power models	194
	7.11	Summary and conclusion	200
8	Som	e applications to real data examples	202
	8.1	The first application to real data-related air pollution	202
	8.2	Graphical methods to select the threshold of the given	
		application	208
		8.2.1 HPL and HPP to select a suitable threshold	208
		8.2.2 MEPL and MEPP to select a suitable threshold	211
	8.3	Test for the choice of EVI in the GPDL and GPDP	
		models	215
	8.4	Comparison between graphical methods of threshold	
		selection	217
	8.5	Fitting of the GPDL and GPDP	218
	8.6	Comparison between some estimators of the EVI	219
	8.7	The second application to real data	224
9	Miscellaneous results		
	9.1	Extreme value theory under linear-power normalization	225
		9.1.1 The class of weak limits of lp -model	226
		9.1.2 Statistical inference using BM method in lp -model	227
	9.2	Real data application related to AccuWeather	228
	9.3	Box-Cox transformation to improve the L-model and	
		P-model	230
	9.4	Real data application	231
	9.5	The Kumaraswamy GEVL and GEVP DFs and further	
		generalizations	237
App	endix	A Summary of Hill's estimators in the L-model	
	and	P-model	239
	D		241
	Refe	rences	244
	Auth	lor index	254
	Subi	ect index	- 258

 $_{\rm xiv}$

Preface

Extreme value theory is a progressive branch of statistics dealing with extreme events. The restriction of the statistical analysis to this special field is justified by the fact that the extreme data, or the extreme part of the sample, can be of outstanding importance in studying floods, hurricanes, air pollutants, extreme claim sizes, life spans, etc.

A quick look at the literature reveals that all the known books in the area of extreme value analysis deal with the modelling of extreme value data based on extreme value theory under linear normalization. In this book, we will tackle some modern trends in the modelling of extremes under linear normalization, such as the bootstrap technique. In addition, we consider the problem of the mathematical modelling of extremes under power normalization with the hope that this most recent approach will be more routinely applied in practice. Finally, the present book handles some recent approaches in order to achieve an improved fit of generalized extreme value distribution for block maxima data and of generalized Pareto distribution for peak-overthreshold data, either under linear or power normalization. Among these approaches is the use of Box-Cox transformation, which provides additional flexibility in improving the model fit.

This book is designed as an addition to the series of books about the modelling of extreme value data rather than as a competitor to them. To the best of the author's knowledge, no books now in print cover the modelling of extreme data under power normalization. It is worth mentioning that the advantage of using the power normalization is that the classical linear model (i.e., using extreme value theory under linear normalization) may fail to fit the given extreme data, while the power model (i.e., using extreme value theory under power normalization) succeeds. On the other hand, although the book contains several applications, it meets the needs of readers who are interested in both the theoretical and the practical aspects of extreme value theory. In addition, the prerequisites for reading the book are minimal; readers do not need knowledge of advanced calculus or advanced theory of probability.

The primary readership of this book will be researchers who have a strong mathematical background and are interested in extreme value theory and its applications in modelling extreme value data, including statisticians, and researchers who are interested in environmental and economic issues. In fact, in some cases, the book may be a primary text (for students of departments of statistics in faculties of science and postgraduate students studying ecology) and it may be supplementary or recommended reading for all students or researchers who are interested in environmental studies and economics.

I am indebted to the numerous researchers who have enriched this field, especially in the modelling of extreme data concerning air pollution. Usually, these researchers worked on their own data arising from their particular habitats; consequently, we may find some diversities or even divergences in their results. However, beneath these diversities or even divergences there lies a shared basis of a general theory. Actually, I am pleased to be part of this team. In this book, I am trying with some members of my own research group to present our own experience that has extended over two decades in this field.

Finally, I would like thank my earlier Ph.D student Dr Hafid A. Alaswed for many considerable contributions presented in this book, especially in Chapters 6–8 of this book. I would also like to extend my sincere gratitude to Adam Rummens who encouraged me to write this book.

> The principal author H. M. Barakat June 2018

Notations and abbreviations

AB	Asymptotic bias
AIC	Akaike information criterion
AM	Asymptotic mean squared error
AN	Asymptotic normality
AV	Asymptotic variance
BIC	Bayesian Akaike criterion
Box-Cox-GL	Box-Cox-GEVL
BM	Block maxima
C.V	Coefficient of variation
CVC	Coefficient of variation criterion
CVCL	Coefficient of variation under linear normalization
CVCP	Coefficient of variation under power normalization
DF	Distribution function
DFs	Distribution functions
\mathcal{D}_ℓ	Domain of attraction under linear normalization
\mathcal{D}_p	Domain of attraction under power normalization
dgos	Dual generalized order statistics
EEAA	Egyptian Environmental Affairs Agency
E(X)	Expected value of X
evir	Extreme values in R package
EVT	Extreme value theory
EVI	Extreme value index
GA	Genetic algorithms
GEVL	Generalized extreme value DF under linear normalization
GEVLs	Generalized extreme value DFs under linear normalization
gos	Generalized order statistics
GP	Generalized Pareto distribution
GPDL	Generalized Pareto DF under linear normalization
GEVPs	Generalized extreme value DFs under power normalization
GPDP	Generalized Pareto DF under power normalization
GPDPs	Generalized Pareto DFs under power normalization
GPVLP	Generalized Pareto DF under linear-power normalization
GPVLPs	Generalized Pareto DFs under linear-power normalization
K-S	Kolmogorov-Smirnov (test)
HP	Hill plot
HPL	Hill Plot under linear normalization
HPP	Hill plot under power normalization
HE	Hill estimator
HEs	Hill estimators

HEL	Hill estimator under linear normalization
HELs	Hill estimators under linear normalization
HEP	Hill estimator under power normalization
HEPs	Hill estimators under power normalization
HMEL	Harmonic moment estimator under linear normalization
HMEP	Harmonic moment estimator under power normalization
HMEPs	Harmonic moment estimators under power normalization
iid	Independent and identically distributed
LI	Location-invariant
LAQN	London air quality network
LRT	Likelihood ratio test
MEL	Moment estimators under linear normalization
MEP	Moment estimators under power normalization
MEPL	Mean excesses plot under linear normalization
MEPP	Mean excess plot under power normalization
$m-\mathrm{gos}$	m-generalized order statistics
$m-\mathrm{dgos}$	m-dua generalized order statistics
ML-laws	The class of maximum limit laws
ML	Maximum likelihood
MLE	Maximum likelihood estimate
MLEs	Maximum likelihood estimates
MREL	Moment-ratio estimator under linear normalization
MREP	Moment-ratio estimator under power normalization
MSE	Mean squared error
MSEs	Mean squared errors
MSEL	Mean squared error under linear normalization
MSEP	Mean squared error under power normalization
MS-laws	Class of max-stable laws
NERMN	National Environmental Radiation Monitoring Network
NCNSRC	National Center for Nuclear Safety and Radiation Control
NHP	New Hill plot
NO	Nitric oxide
NO_2	Nitrogen dioxide
PDF	Probability density function, also density function
PDFs	Probability density functions, also density functions
PM	Particulate matter
PM10	PM of diameter less than 10 mm
POT	Peak over threshold
RLP	Return level plot
RV	Random variable

xiv	Notations and abbreviations
RVs	Random variables
SE	Standard error
SP	Stability plot
SO_2	Sulphur dioxide
STD	Standard deviation
SI	Scale invariant
\mathbf{SC}	Strong consistence
TCP	Threshold choice plot
UA	Uniform assumption
$\operatorname{Var}(X)$	Variance of X
WC	Weak consistence
WHO	World health organization
$\overline{F} = 1 - F$	Survival function
\mathcal{R}	Real line
$\mathcal{N}(\mu,\sigma)$	Normal distribution with mean μ and variance σ
$\Phi(.)$	Standard normal distribution function
\xrightarrow{n}	Convergence, as $n \to \infty$
$\xrightarrow{w}{n}$	Weak convergence, as $n \to \infty$
$\xrightarrow{d}{n}$	Convergence in distribution, as $n \to \infty$
$\xrightarrow{p}{n}$	Convergence in probability, as $n \to \infty$
$\xrightarrow[n]{a.s.}$	Convergence almost surly, as $n \to \infty$

Illustrations

4.1	Mobile gas monitoring station used to monitor pollution over	100
4.0	the course of a full year and provided by NCNSRC	120
4.2	Hourly average of particulate matter concentration for 10th of	1.0.0
	Ramadan	120
4.3	Hourly average of particulate matter concentration for 10th of	
	Ramadan	121
4.4	Hourly average sulphur dioxide concentration for 10th of Ra-	
	madan	121
4.5	Hourly average sulphur dioxide concentration for Zagazig	122
4.6	Thirty minutes average ozone concentration for 10th of Ramadan	122
4.7	Fifteen-minute average gamma radiation level for Zagazig	122
4.8	Fifteen-minute average gamma radiation level for Zagazig	123
4.9	SO ₂ in Zagazig	132
4.10	SO_2 in 10th of Ramadan	132
4.11	PM10 in Zagazig	133
4.12	PM10 in 10th of Ramadan	133
4.13	O_3 in 10th of Ramadan after bootstrap	134
4.14	Ambient gamma radiation in Zagazig after bootstrap	134
6.1	The relation between threshold selection and k	161
6.2	The threshold selection by using the MEPL for the River Nidd	
	data. Vertical dashed lines mark these thresholds	162
6.3	The threshold selection by using the MEPP for the River Nidd	
	data. Vertical dashed lines mark these thresholds	163
6.4	The threshold selection by using the HPL. Vertical dashed lines	
	mark these thresholds	164
6.5	The threshold selection by using the TCP function for Nidd data	165
7.1	The left panel is the Hill plot of $\gamma_n^{++\succ}$, with $k = 112$, and the	
	right panel is the Hill plot of $\gamma_n^{++\prec}$, with $k=82$	170
7.2	The threshold selection using the New Hill plot for Nidd data	183
7.3	The Hill plot of $\gamma_{PM}^{+\prec}$, with $\vec{k}=19$	193
	- '1 IVI'	

Illustrations

7.4	The Hill plot of $\gamma_{PM}^{+\succ}$, with k=51	193
7.5	The Hill plot of $\gamma_{PH}^{+\prec}$, with k=94	193
7.6	The Hill plot of $\gamma_{PH}^{+\succ}$, with k=9	194
7.7	Samples generated from GEVL (0.10 $\leq \gamma \leq$ 2.60, $\mu = 7$, $\sigma = 1$): n=100, replicates=1000	199
7.8	Samples generated from GEVP(0.10 $\leq \gamma \leq 2.60$): n=100, repli-	
	cated=1000	200
8.1	Return level plot of two pollutants in the two cities	204
8.2	Four different plots: P-P plot, Q-Q plot, return level plot, and density plot of a daily period (24 hours) of PM10 in 10th of	
	Ramadan	206
8.3	Four different plots: P-P plot, Q-Q plot, return level plot, and density plot of a daily period (24 hours) PM10 in Zagazig	207
8.4	Four different plots: P-P plot, Q-Q plot, return level plot, and	
	density plot of a daily period (24 hours) SO_2 in 10th of Ramadan	207
8.5	Four different plots: P-P plot, Q-Q plot, return level plot, and	
	density plot of a daily period (24 hours) SO_2 in Zagazig	208
8.6	Selection of the threshold of PM10 in 10th of Ramadan. The	
	left, middle, and right panels indicate respectively γ_{ℓ}^{++} , γ_{P}^{++} ,	
	and γ_P^{++}	210
8.7	Selection of the threshold of PM10 in Zagazig. The left, middle,	
	and right panels indicate respectively γ_{ℓ}^{++} , γ_{P}^{++} , and γ_{P}^{++}	210
8.8	Selection of the threshold of SO_2 in 10th of Ramadan. The left,	
	middle, and right panels indicate respectively γ_{ℓ}^{++} , γ_{P}^{++-} , and	
	γ_P^{++}	211
8.9	Selection of the threshold of SO_2 in Zagazig. The left, middle,	
	and right panels indicate respectively γ_{ℓ}^{++} , γ_{P}^{++} , and γ_{P}^{++}	211
8.10	Mean excess plot for selecting the threshold of PM10 in 10th of	
	Ramadan. The left and the right panels indicate respectively the	
	MEPL and the MEPP	212
8.11	Mean excess plot for selecting the threshold of PM10 in Zagazig.	
	The left and the right panels indicate respectively the MEPL and	
	the MEPP	212
8.12	Mean excess plot for selecting the threshold of SO_2 in 10th of	
	Ramadan. The left and the right panels indicate respectively the MEPL and the MEPP	214
8.13	Mean excess plot for selecting the threshold of SO_2 in Zagazig.	
	The left and the right panels indicate respectively the MEPL	
	and the MEPP	214
9.1	Graphical representation of the data set and the fitted distribu-	
	tion $\mathcal{P}_{1;\hat{\gamma}}(x;\hat{c};\hat{a},\tilde{b})$	229

xvi

2.1	Domains of attraction of the most common distributions	29
3.1	Estimated GEVL and GEVP models, for F_1	91
3.2	Estimated GEVL and GEVP models, for F_2	91
3.3	Estimated GEVL and GEVP models, for $F_{3:1}$	91
3.4	Generated data from $\mathcal{N}(\mu, \sigma = 1)$: bootstrap technique for quan-	
	tiles	96
3.5	K-S Test: bootstrap technique for quantiles	97
3.6	Simulation study for $K = 1$	106
3.7	Simulation study for $K = 2$	106
4.1	Zagazig and 10th of Ramadan for GEVL	131
4.2	Zagazig and 10th of Ramadan for GEVL, after bootstrap	131
4.3	K-S test for the data with and without bootstrap	135
4.4	Simulation study for choosing a suitable number of POT (k) — k^{\star}	
	denotes the best value	136
4.5	Simulation study for choosing a suitable number of POT (k) — k^{\star}	
	denotes the best value	136
4.6	Simulation study for chosen m sub-sample bootstrap— m^* de-	
	notes the best value	137
4.7	Simulation study for chosen m sub-sample bootstrap— m^\star de-	
	notes the best value	137
4.8	Zagazig and 10th of Ramadan for GPDL	138
4.9	Zagazig and 10th of Ramadan for GPDL after bootstrap	138
4.10	Zagazig and 10th of Ramadan for GEVL	138
5.1	Estimating the shape parameter γ in the GEVP $(\gamma, 1, 1)$, defined	
	in (2.32) , by using the ML method and the suggested estimate	
	(5.3)—"*" in the superscript of a value means that this value is	
	the best	146
5.2	Estimating the shape parameter γ in the GPDP, defined in (5.4),	
	by using the ML method—"*" in the superscript of a value	
	means that this value is the best	147

5.3	Estimating the shape parameter γ in the GPDP (5.4) by using the suggested estimate (5.11)—"*" in the superscript of a value	
	means that this value is the best	148
5.4	Estimate parameters of GEVP (5.4) by using the ML method	
	for the sub-sample bootstrapping method	148
5.5	Estimating the shape parameter γ in the GEVP(0.4, 1, 1), de-	
	fined in (5.4) by using the suggested estimate (5.3)—" $*$ " in the	140
E C	superscript of a value means that this value is the Dest E_{rest} in the CEVD(0.2.1.1) de	149
0.0	Estimating the snape parameter γ in the GEVP(0.2, 1, 1), defined in (5.4) by using the suggested estimate (5.3) "*" in the	
	superscript of a value means that this value is the best	1/0
5.7	Estimating the shape parameter γ in the GEVP(0, 1, 1) defined	140
0.1	in (5.4), by using the suggested estimate (5.3) —"*" in the su-	
	perscript of a value means that this value is the best	150
5.8	Estimating the shape parameter γ in the GEVP(-0.2, 1, 1), de-	
	fined in (5.4) by using the suggested estimate (5.3) —"*" in the	
	superscript of a value means that this value is the best	150
5.9	Estimating the shape parameter γ in the GEVP(-0.4, 1, 1), de-	
	fined in (5.4) by using the suggested estimate (5.3) —"*" in the	
	superscript of a value means that this value is the best	151
5.10	Simulation study for estimate GEVL and GEVP by using the	150
C 1	GA technique	153
0.1 6.0	Comparison between some desirable properties of HELS for EVI	159
0.2	Simulation output for assessing and comparing the estimators α^+ and α^{++}	166
71	Simulation output for assessing the estimators γ^{++} and γ^{++}	172
7.2	Simulation output for assessing the estimators γ_p^{+-} and γ_p^{+-}	172
7.3	Simulation output for assessing the estimators γ_p^{-+} and γ_p^{-+}	173
74	Simulation output for assessing the estimators $\gamma_p^{}$ and $\gamma_p^{}$	173
7.5	Simulation output for assessing the HEP $\gamma_p^{++} \rightarrow 0$	181
7.6	Simulation output for assessing the HEP γ_p^{+-} , $\gamma > 0$	181
77	Simulation output for assessing the HEP γ_p^{-+} , $\gamma < 0$	182
7.8	Simulation output for assessing the HEP $\gamma^{}_p \sim 0$	182
7.9	Simulation output for assessing and comparing the HEPs γ^{++}	102
1.0	$\gamma_p^{++\succ}, \gamma_p^{++}, \gamma_{PH}^{++\prec}, \text{ and } \gamma_{PH}^{++\succ}$	185
7.10	Simulation output for assessing and comparing the HEPs $\gamma_p^{+-\prec}$,	
	γ_p^{+-} , and γ_p^{+-}	186
7.11	Simulation output for assessing and comparing the HEPs $\gamma_p^{-+\prec}$,	
	γ_p^{-+} , and γ_p^{-+}	187
7.12	Simulation output for assessing and comparing the HEPs γ_p^{\prec} ,	
	$\gamma_p^{}$, and $\gamma_p^{}$	188

xviii

7.13	Simulation output for assessing and comparing the estimators	
	$\gamma_P^{++\prec}, \gamma_{PH}^{++\prec}, \gamma_{PM}^{+\prec}, \text{ and } \gamma_{PMR}^{++\prec}$	191
7.14	Simulation output for assessing and comparing the estimators	
	$\gamma_P^{++\succ}, \gamma_{PH}^{++\succ}, \gamma_{PM}^{+\succ}, \text{and } \gamma_{PMR}^{++\succ}$	192
7.15	Simulation output for assessing the criterion CVC for $\gamma_{\ell M}^+$	196
7.16	Simulation output for assessing the criterion CVC for $\gamma_P^{\mp \pm \succ}$	197
7.17	Simulation output for assessing the criterion CVC for γ_{ℓ}^{++}	198
7.18	Simulation output for assessing the criterion CVC for γ_P^{++}	199
8.1	Statistics summary of the two pollutants in the two cities	203
8.2	ML estimates for the shape parameter of the two pollutants in	
	the two cities	205
8.3	The LRT of the two pollutants, p-value, and decision	205
8.4	Threshold selection by using HPL and HPP	209
8.5	MEPL and MEPP for selecting an appropriate threshold	213
8.6	Gomes and van Monfort test for GPDL and GPDP of PM10 in	
	10th of Ramadan	215
8.7	Gomes and van Monfort test for GPDL and GPDP of PM10 in	
	Zagazig	216
8.8	Gomes and van Monfort test for GPDL and GPDP of SO_2 in	
	10th of Ramadan	216
8.9	Gomes and van Monfort test for GPDL and GPDP of SO_2 in	
	Zagazig	217
8.10	Graphical methods of threshold selection of PM10	217
8.11	Graphical methods of threshold selection of SO_2	218
8.12	Parameter estimates for the GP distribution of PM10, under	
	linear and power normalizing constants	218
8.13	Parameter estimates for the GP distribution of SO ₂ , under linear	
	and power normalizing constants	219
8.14	Different estimates of the EVI for 10th of Ramadan under the	
	linear and power models	220
8.15	Different estimates of the EVI for Zagazig under the linear and	
	power models	221
8.16	Comparison between the linear and power models for pollutants	
	$PM10 and SO_2$	223
8.17	Estimates of the EVI of Danish fire insurance claims under linear	
	and power normalizing constants	224
8.18	Comparison between the linear and power models	224
9.1	Summary statistics	228
9.2	Parameter estimation for maximum temperature	229
9.3	Descriptive statistics for maximum data for air pollution	232
9.4	The MLEs for the GEVL and GPDL models—the application	
	of AIC and BIC	233
9.5	K-S test	233

 $_{\rm xix}$

9.6	The MLEs for Box-Cox-GL model and the application of AIC	
	and BIC	234
9.7	K-S test	234
9.8	The MLEs for GEVP and GPVP models—the application of	
	AIC and BIC	235
9.9	K-S test	235
9.10	The MLEs for the Box-Cox-GP1 model and the application of	
	the AIC and BIC	236
9.11	K-S test	236
A.1	Four HELs	239
A.2	Eight counterparts of HEPs	240
A.3	Four HEPs based on GPDP	241
A.4	Four moment estimators under power normalization	242
A.5	Four moment ratio estimators under power normalization	243

xx

Introduction: Some basic and miscellaneous results

In practice, we usually do not know the true probability models of random phenomena, such as a human behaviour. George Box once said that there is no true model, but there are useful models. Even if there was a true probability model, we would never be able to observe it. Fortunately, in many cases a complicated situation can be replaced by a comparatively simple asymptotic model. The most important example of such cases is when the extremes govern the law of interest (e.g., air pollution, floods, strength of material, etc.). More precisely, the asymptotic theory of extreme order statistics provides approximate probability models that are not true but are definitely useful. Therefore, we must connect what we can observe with these approximate models. The key idea here is that we use a large set of observations (or a set of realizations) to figure out the approximate probability model given the data we have. Clearly, the cornerstone of the approximate probabilities model is the concept of the convergence in probability theory. In Section 1.1, we will discuss different types of convergence in the probability theory and statistics. On the other hand, some important tools of data treatments, such as the Maximum Likelihood Method, Genetic Algorithms (GA), and the Kolmogorov-Smirnov (K-S) test, are discussed in Sections 1.1 and 1.2.

1.1 The convergence concept in probability theory

There are several convergence concepts associated with the limiting behaviour of a sequence of RVs. Convergence in distribution (or weak convergence), convergence in probability, and almost sure convergence are the prominent ones. In the case of the sample mean, these concepts lead us to the classical central limit theorem, weak law of large numbers, and strong law of large numbers, respectively. In this book we will mostly be concerned with weak convergence results for order statistics. In the context of weak convergence, we are interested in identifying the possible non-degenerate limit distributions for appropriately normalized sequences of RVs of interest. These limiting distributions can be of direct use in suggesting inference procedures when the sample size is large. These concepts and some required theorems of a purely analytical nature will be briefly discussed in this section. Throughout what follows the symbol (\xrightarrow{n}) stands for convergence, as $n \to \infty$.

1.1.1 Modes of convergence of RVs

Definition 1.1 (almost sure convergence) We say that a sequence of RVs $X_1, X_2, ...$ converges to a RV X almost surely, written $X_n \xrightarrow[n]{a.s} X$, if

 $\{\omega \in \Omega : X_n(\omega) \xrightarrow{} X(\omega)\}$

is an event whose probability is one, where X_n and X are defined on the same probability space (Ω, \mathcal{F}, P) .

Definition 1.2 (convergence in probability) A sequence of RVs $\{X_n\}$ is said to converge in probability to a RV X, as $n \to \infty$, written $X_n \xrightarrow{p} X$, if for every $\epsilon > 0$ we have $P(|X_n - X| < \epsilon) \xrightarrow{n} 1$, or equivalently $P(|X_n - X| \geq \epsilon) \xrightarrow{n} 0$.

Definition 1.3 (convergence in the *r*th mean) A sequence of RVs $X_1, X_2, ...$ is said to converge in the *r*th mean, or in the norm $||.||_r$, to a RV X, written $X_n \xrightarrow{r}{n} X$, if $r \ge 1$, $E|X_n|^r < \infty$, $\forall n$, and

$$\lim_{n \to \infty} E(|X_n - X|^r) = 0.$$

The most important cases of convergence in rth mean are:

- When X_n converges in rth mean to X, for r = 1, we say that X_n converges in mean to X.
- When X_n converges in rth mean to X, for r = 2, we say that X_n converges in mean square to X.

Convergence in the *r*th mean, for r > 0, implies convergence in probability (by Chebyshev's inequality), while if $r > s \ge 1$, convergence in *r*th mean implies convergence in *s*th mean. Hence, convergence in mean square implies convergence in mean.

Definition 1.4 (convergence in distribution or weak convergence) Assume that X_1, X_2, \dots is a sequence of RVs with corresponding DFs F_1, F_2, \dots and

the RV X has the DF F. We say that the sequence of RVs $\{X_n\}$ converges in distribution to the RV X, as $n \to \infty$, written $X_n \xrightarrow{d} X$ (or the sequence of DFs $\{F_n\}$ converges weakly to the DF F, as $n \to \infty$, written $F_n(x) \xrightarrow{w} F(x)$ if $F_n(x)$ converges pointwise to F(x) at all continuity

Remark Many authors avoid using the notation $X_n \xrightarrow{d} X$, since weak convergence pertains only to the DF of X and not to X itself. However, we only use this notation in this section for the sake of notation uniformity; however, in the sequel we will use the notation $F_n(x) \xrightarrow{w} F(x)$.

points of F, that is $F_n(x) \xrightarrow{} F(x)$ at all points x, where F is continuous.

Remark Unless otherwise stated, we assume that the limiting function F(x)is non-degenerate proper DF, i.e., that there exists a real number x such that 0 < F(x) < 1 and $F(\infty) - F(-\infty) = 1$, in this case, we say that $F_n(x)$ converges properly to F(x) or simply $F_n(x)$ converges weakly to F(x). On the contrary, if $F(\infty) - F(-\infty) < 1$, F(x) will be called improper DF and in this case the aforesaid convergence will be called improper convergence.

Some important relations between the modes of convergence are given in the next theorems.

Theorem 1.5 Assume that $X_1, X_2, ..., X_n$ are RVs on the same probability space (Ω, \mathcal{F}, P) . If so, the following implications hold:

- If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{p} X$.
- If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$.
- If $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{p} X$.

Theorem 1.6 (Continuous Mapping Theorem) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of RVs, $f : \mathcal{R} \to \mathcal{R}$ be a continuous function, and X be an RV.

- If $X_n \xrightarrow[n]{a.s.} X$, then $f(X_n) \xrightarrow[n]{a.s.} f(X)$. If $X_n \xrightarrow[n]{d} X$, then $f(X_n) \xrightarrow[n]{d} f(X)$.
- If $X_n \xrightarrow{p} X$, then $f(X_n) \xrightarrow{p} f(X)$.

The preceding results hold equivalently for a sequence of random vectors and matrices. Also, an important special case here is that X = c, where $c \in$ \mathcal{R} . In this case, we get $f(X_n) \xrightarrow[n]{a.s.} f(c)$, if $X_n \xrightarrow[n]{a.s.} c$. Similarly, if $X_n \xrightarrow[n]{p} c$, then $f(X_n) \xrightarrow{p} f(c)$.

Theorem 1.7 (Slutzky's Theorem) Let $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} C$, where $C \in \mathcal{R}$ is a constant. Then, $Y_n X_n \xrightarrow{d} CX$ and $X_n + Y_n \xrightarrow{d} X + C$.

An important special case of Theorem 1.7 is that if $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} 0$, then $X_n + Y_n \xrightarrow{d} X$. In this case, we say that $Z_n = X_n + Y_n$ and X_n are asymptotically equivalent because $Z_n - X_n \xrightarrow{p} 0$. Clearly, Slutzky's theorem, as well as the convergence concepts, can be readily extended to random vectors and random matrices.

Theorem 1.8 If $F_n \xrightarrow[w]{} F$ and F is continuous, then $\sup_{x} |F_n(x) - F(x)| \xrightarrow[w]{} 0,$

which means that the convergence is uniform with respect to x.

1.1.2 Further limit theorems on weak convergence

The meaning of any limit theorem for a random sequence $\{X_n\}$ is that it gives a sufficiently simple approximation to the DF $F_n(x) = P(X_n < x)$. Namely, let $F_n(G_n(x)) = P(G_n^{-1}(X_n) < x) \xrightarrow{w} P(X < x)$, where $G_n(.)$ is a monotone continuous function (we may take $G_n(x) = a_n x + b_n$) and $G_n^{-1}(.)$ is the inverse of G_n . If the limit F(x) = P(X < x) is continuous, then Theorem 1.8 implies that

$$\epsilon_n = \sup_x |P(G_n^{-1}(X_n) < x) - P(X < x)| = \rho(G_n^{-1}(X_n), X) \longrightarrow 0.$$

Since the metric ρ is invariant with respect to strongly monotone continuous transformations of RVs, we have

$$\rho(X_n, G_n(X)) = \epsilon_n \longrightarrow 0,$$

i.e., we receive a uniform approximation to $P(X_n < x) = F_n(x)$ by means of some universal DF of the RV X (see Pancheva, 1984). Such a viewpoint to the limit theorems deprives the traditionally linear transformation of its exclusiveness. Thus, it makes sense to extend the class of normalizing transformations, $\{G_n(x)\}$, to any strongly monotone continuous transformations for constructing a simplified approximation if only one can prove a suitable limit theorem. Chapter 5 will rely on this idea. The next result gives equivalent characterizations of the weak convergence.

Theorem 1.9 If ψ and $\{\psi_n\}$ are the characteristic functions with the DFs F and $\{F_n\}$, respectively, then the following statements are equivalent: (i) $F_n \xrightarrow{w}{n} F$; (ii) $\psi_n(t) \xrightarrow{}{n} \psi(t)$, for every $t \in \mathcal{R}$; (iii) $\int g(x) dF_n(x) \xrightarrow{}{n} \int g(x) dF(x)$ for every bounded continuous function g.

4

Let F and F_n be the DFs of the RVs X and X_n , respectively (notice that $X_1, X_2, ...$ and X need not to be defined on the same probability space). Let $F_n \xrightarrow{w} F$ (or equivalently $X_n \xrightarrow{d} X$). Then, in this case, the DF F is usually called the asymptotic (or limiting) distribution of the sequence X_n . Clearly, the convergence in distribution depends only on the involved DFs and does not require that the relevant RVs approximate each other. However, the only relationship between the weak convergence and the convergence in probability is given in the following theorem.

Theorem 1.10 If $X_n \xrightarrow[n]{d} C$, where C is a constant, then $X_n \xrightarrow[n]{p} C$.

The following definition and theorem, due to Helly (see Feller, 1979), are basic tools in studying the weak convergence of the sequence of DFs.

Definition 1.11 Let $\{X_n\}$ be a sequence of RVs with corresponding DFs $\{F_n\}$. Then, the sequences $\{X_n\}$ and $\{F_n\}$ are said to be stochastically bound, if for each $\epsilon > 0$, there exists a number c such that

 $P(|X_n| \ge c) < \epsilon$, for all sufficiently large n.

- **Theorem 1.12** (A) Every sequence of DFs $\{F_n\}$ possesses a subsequence $\{F_{n_k}\}$, that converges (properly or improperly) to a limit F (remember that the improper convergence means that the limit is an extended DF, i.e., $F(\infty) F(-\infty) < 1$).
- (B) In order that all such limits be proper it is necessary and sufficient that $\{F_n\}$ be stochastically bounded.
- (C) In order that $F_n \xrightarrow{w} F(x)$, it is necessary and sufficient that the limit of every convergence subsequence equals F.

We will end this section with an important known theorem, which will be needed in the sequel.

Theorem 1.13 (Khinchin's type theorem) Let $F_n(x)$ be a sequence of DFs. Furthermore, let

$$F_n(G_n(x)) \xrightarrow{w} F(x),$$

with $G_n(x) = a_n x + b_n$, $a_n > 0$. Then, with $G_n^*(x) = c_n x + d_n$, $c_n > 0$, we have

$$F_n(G_n^*(x)) \xrightarrow{w} F^*(x), F^* \text{ is a non-degenerate } DF,$$

 $\begin{array}{l} \text{if and only if } G_n^{-1}(G_n^*(x)) = G_n^{-1}oG_n^*(x) \xrightarrow[]{}{} g(x), \, \forall x, \, where \, g(x) = ax + b, \, \frac{c_n}{a_n} \xrightarrow[]{}{} a, \, \frac{d_n - b_n}{a_n} \xrightarrow[]{} b \, and \, F^*(x) = F(g(x)). \end{array}$

Theorem 1.13 leads to the following definition:

Definition 1.14 We say that the DFs G(x) and $G^*(x)$ are of the same type, under linear transformation, if there are real numbers A and B > 0 such that

$$G^*(x) = G(Ax + B).$$

Clearly the relation between G and G^* in Definition 1.14 is symmetrical, reflexive, and transitive. Hence, it gives rise to equivalence classes of DFs. Sometimes we shall indicate a type by one representative of the equivalence classes. These facts convince us that the probability limit theory basically deals with the types of DFs rather than the DFs themselves.

Remark (Why the weak convergence mode?) It is natural to wonder why we use weak convergence in statistical modelling, although it is the weakest mode of convergence. Actually, Barakat and Nigm (1996) have investigated the mixing property of order statistics. The notion of mixing sequences of RVs was first introduced by Rényi (1962, 1970). In the sense of Renyi, a sequence $\{X_n\}$ of RVs is called mixing if for any event E of positive probability, the conditional DF of X_n under the condition E converges weakly to a non-degenerate DF, which does not depend on E. Barakat and Nigm (1996)have shown that any sequence of order statistics (extreme, intermediate, and central), under linear normalization, is mixing. On the other hand, they also showed in the same work that any mixing sequence of RVs $X_1, X_2, ..., X_n$ cannot converge in probability to an RV X_{∞} that has non-degenerate DF. This simply means that any sequence of order statistics, particularly the sequence of extreme order statistics, cannot converge in probability to any RV with non-degenerate DF (except for convergence in probability to a constant) and the only available mode of convergence is the weak convergence.

1.2 Statistical methods

1.2.1 Maximum likelihood method

A general and flexible method of estimation of the unknown parameter θ within a family F is the maximum likelihood method. Each value of $\theta \in \Theta$ defines a model in F that attaches (potentially) different probabilities (or probability densities) to the observed data. The probability of the observed data as a function of θ is called the likelihood function. Plausible values of θ should have a relatively high likelihood. The principle of maximum likelihood estimation is choosing the model with greatest likelihood, among all the models under consideration, i.e., this is the one that assigns highest probability to the observed data. To see this in greater detail, we can refer back to the situation in which we have a data set X whose density is defined by some d-dimensional parametric model with parameter $\theta = (\theta_1, ..., \theta_d)$. Write the density evaluated at X = x in the form

$$f(x;\theta).$$

The likelihood function for θ based on the data X is just $f(x; \theta)$ interpreted as a function of θ . Usually, we work with the log likelihood

$$\ell_X(\theta) = \log[f(x;\theta)].$$

The maximum likelihood estimate (MLE) $\hat{\theta}$ (of the parameter θ) is the value of θ which maximizes $\ell_X(\theta)$. Usually, we assume $\ell_X(\theta)$ is differentiable with a unique interior maximum, so the MLE is given by solving the likelihood equations

$$\frac{\partial \ell_X(\theta)}{\partial \theta_j} = 0, \ j = 1, ..., d.$$

For the maximization of $\ell_X(\theta)$, for a general model indexed by θ , this may be performed using a packaged nonlinear optimization subroutine, of which several excellent versions are available.

Example 1.15 Consider the general extreme value DF under linear normalization (GEVL)

$$G_{\gamma}(x;\mu,\sigma) = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}}\right\}$$
(1.1)

defined on $\{x : 1 + \gamma(x - \mu)/\sigma > 0\}$. In this distribution γ is a shape parameter, μ is a location parameter and σ is a scale parameter. This DF is the foremost pillar of the statistical modelling of extreme value data under linear normalization that will be discussed in detail in Chapter 4. For the GEVL (1.1), the density $g(x; \mu, \sigma, \gamma)$ is obtained by differentiating $G_{\gamma}(x; \mu, \sigma)$ with respect to x. The likelihood function based on observations $x_1, ..., x_k$ is

$$\prod_{i=1}^k g(x_i; \mu, \sigma, \gamma)$$

and so the log likelihood is given by

$$\ell_X(\mu,\sigma,\gamma;x) = -k\log\sigma + \sum_{i=1}^k \left\{ -\left[1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\gamma}} - \left(1 + \frac{1}{\gamma}\right)\log\left[1 + \gamma\left(\frac{x_i - \mu}{\sigma}\right)\right] \right\}, \quad (1.2)$$

provided $\{1 + \gamma(x_i - \mu)/\sigma > 0\}$ for each *i*; otherwise, (1.2) is undefined. The following practical points should be considered for this example:

- 1. Although the maximization is unconstrained, there are some practical constraints. For example, (1.2) requires $\gamma > 0$ as well as $\{1+\gamma(x_i-\mu)/\sigma > 0\}$ for each *i*. It is advisable to test explicitly for such violations and to set $-\ell_X(\theta)$ equal to some very large value if the conditions are indeed violated.
- 2. All Newton-type routines require the user to supply starting values, but the importance of good starting values can be overemphasized. Simple guesses usually suffice, e.g., in (1.2), one might set μ and σ equal to the sample mean and sample standard deviation respectively, with γ equal to some crude guess value such as 0.1. However, it is important to check that the initial conditions are feasible and this can sometimes not be so easy to achieve.
- 3. In cases of doubt about our application, where a true maximum has been found, the algorithm may be re-run from different starting values. If the results are highly sensitive to starting values, this is indicative that the problem may have multiple local maxima, or alternatively that a mistake has been made in programming.

A few further comments are necessary regarding the specific application of numerical MLE to the GEVL family. There is a singularity in the likelihood for $\gamma < 0$, as $\mu \to X_{max} = \max(X_1, ..., X_k)$ in (1.2) and the effect is that $\ell_X(\theta) \to \infty$. However, in the most practical cases, there is a local maximum (of $\ell_X(\theta)$) that is some distance from the singularity and the presence of the singularity does not interfere with the convergence of the nonlinear optimization algorithm to the local maximum. In this case, the correct procedure is to ignore the singularity and use the local maximum. However, it is possible that no local maximum exists and the singularity dominates. In this case, MLE fails and some other method must be sought. However, this very rarely happens with environmental data. Finally, we should say something about the theoretical status of the approximations involved. The asymptotic theory of MLE for the GEVL model is valid provided $\gamma > -0.5$ (cf. Smith, 1985). Cases with $\gamma \leq -0.5$ correspond to an extremely short upper tail and hardly ever occur in environmental applications. A more serious problem is that even when $\gamma > -0.5$, the asymptotic theory may give rather poor results with small sample sizes, see Hosking et al. (1985).

In summary: it is possible that MLEs will fail either numerically or in terms of their asymptotic properties, especially if the sample size is small. The user should be aware of their possible difficulties but should not be deterred from using these extremely powerful and general methods. For more details about this subject, see Prescott and Walden (1980, 1983), Mached (1989), and Smith (1985).

An alternative method for quantifying the uncertainty in the MLE is based on the deviance function, or the likelihood ratio test (LRT) (see Theorems 2.6 and 2.7 in Coles, 2001), which is defined by

$$LRT = -2(\log L_0 - \log L_1), \tag{1.3}$$

where $\log L_0$ and $\log L_1$ are the values of the log-likelihood under the null and alternative hypothesis, respectively. The statistic *LRT* is distributed as χ^2_n , with degrees of freedom equal to the number of restrictions under the null hypothesis. The method of the LRT is summarized as follows:

- 1. Let $L_0(M_0)$ and $L_1(M_1)$ be the maximized values of the log-likelihood for models M_0 and M_1 , respectively.
- 2. Test of the validity of model M_0 relative to M_1 at a suitable chosen level of significance.

Reject M_0 in favour of M_1 if $LRT = -2(\log L_0 - \log L_1) > c_\alpha$, where c_α is the $(1 - \alpha)$ quantile of the χ_n^2 distribution.

1.2.2 Kolmogorov-Smirnov (K-S) test

In statistics, the K-S test is a nonparametric test of the equality of continuous one-dimensional DFs that can be used to compare a sample with a reference DF (one-sample K-S test), or to compare two samples (two-sample K-S test). It is named after Andrey Kolmogorov and Nikolai Smirnov.

The K-S statistic quantifies a distance between the empirical DF of the sample and the DF of the reference distribution, or between the empirical DFs of two samples. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution $\hat{F}(x)$ (in the one-sample case) or that the samples are drawn from the same distribution (in the two-sample case). In each case, the considered distributions under the null hypothesis are continuous DFs, but are otherwise unrestricted.

Let $X_1, X_2, ..., X_n$ be independent and identically random sample distributed under the null-hypothesis H_0 , as F_0 . Therefore, the K-S test statistic D_n is defined by

$$D_n = \sup_x |F_0(x) - F_n(x)|,$$

where $\sup x$ is the supremum of the set of distances and $F_n(x)$ is the empirical DF that increases by $\frac{1}{n}$ at each data value. Namely,

$$F_n = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{[-\infty,x]}(X_i),$$

where $I_{[-\infty,x]}(X_i)$ is the indicator function, which is equal to 1 if $X_i \leq x$ and is equal to 0 otherwise. By the Glivenko-Cantelli theorem, if the sample comes from the DF $F_0(x)$, then the statistic D_n converges to 0 almost surely in the limit when n goes to infinity. Kolmogorov strengthened this result, by effectively providing the rate of this convergence. In practice, the statistic requires a relatively large number of data points to properly reject the null hypothesis.

The K-S statistic has been used for goodness-of-fit testing for continuous populations for decades, although other tests have made slight improvements in terms of power. The K-S test appeal includes the straightforward computation of the test statistic and the distribution-free characteristic of D_n . Its drawback is that the DF of D_n , under the null hypothesis (i.e., the assumption that data was drawn from a population with DF $F_0(x)$), is difficult to determine, leaving one to calculate critical values with various approximation methods. An algorithm for computing the distribution of D_n , for small to moderate values of n, was given by Drew et al. (2000). As the supremum must be achieved at a data value, the computational formula for computing D_n , is $D_n = \max_x(D_n^+, D_n^-)$, where

$$D_n^+ = \sup_x [F_n(x) - F_0(x)] = \max\left[\max_{1 \le i \le n} [\frac{i}{n} - F_0(X_{i:n}], 0]\right],$$
$$D_n^- = \sup_x [F_0(x) - F_n(x)] = \max\left[\max_{1 \le i \le n} [F_0(X_{i:n} - \frac{i-1}{n}], 0]\right]$$

and $X_{1:n}, X_{2;n}, ..., X_{n:n}$ are the order statistics corresponding to the random sample $X_1, X_2, ..., X_n$. The maximum positive difference, D_n , detects the largest vertical deviation between the two DFs, where the fitted DF $F_0(x)$ is below the empirical DF. Likewise, the maximum negative difference detects the largest vertical deviation between the two DFs, where the fitted DF is above the empirical DF. The smallest value of D_n that can be achieved is 1/2, which corresponds to the DF of the fitted DF $F_0(x)$ bisecting all the risers of the steps associated with the empirical DF.

Assume we have the random sample $X_1, X_2, ..., X_n$ and the hypothesistesting situation $H_0: F_X(x) = F_0(x)$, for all x, where $F_0(x)$ is a completely specified continuous DF. The differences between $F_X(x)$ and $F_0(x)$ should be