Handbook of Fitting Statistical Distributions with R

Zaven A. Karian
Edward J. Dudewicz

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

Contents

Preface v

About the Authors ix

Dedication xv

Comments from GLD Pioneers xvii

PART I: Overview 1

1 Fitting Statistical Distributions: An Overview 3
1.1 History and Background 4
1.2 The Organization of the Handbook 10
References for Chapter 1 16

PART II: The Generalized Lambda Distribution 19

2 The Generalized Lambda Family of Distributions 21
2.1 Definition of the Generalized Lambda Distributions 21
2.2 The Parameter Space of the GLD 23
2.3 Shapes of the GLD Density Functions 32
2.4 GLD Random Variate Generation 47
2.5 The Fitting Process 48
Problems for Chapter 2 49
References for Chapter 2 51

3 Fitting Distributions and Data with the GLD via the Method of Moments 53
3.1 The Moments of the GLD Distribution 54
3.2 The (α_3^2, α_4)-Space Covered by the GLD Family 59
3.3 Fitting the GLD through the Method of Moments 64

3.3.1 Fitting through Direct Computation 65
3.3.2 Fitting by the Use of Tables 74
3.3.3 Limitations of the Method of Moments 75
3.4 GLD Approximations of Some Well-Known Distributions 76
3.4.1 The Normal Distribution 81
3.4.2 The Uniform Distribution 82
3.4.3 The Student's t Distribution 83
3.4.4 The Exponential Distribution 85
3.4.5 The Chi-Square Distribution 87
3.4.6 The Gamma Distribution 88
3.4.7 The Weibull Distribution 90
3.4.8 The Lognormal Distribution 91
3.4.9 The Beta Distribution 93
3.4.10 The Inverse Gaussian Distribution 94
3.4.11 The Logistic Distribution 96
3.4.12 The Largest Extreme Value Distribution 98
3.4.13 The Extreme Value Distribution 99
3.4.14 The Double Exponential Distribution 100
3.4.15 The F-Distribution 101
3.4.16 The Pareto Distribution 103
3.5 Examples: GLD Fits of Data, Method of Moments 105
3.5.1 Assessment of Goodness-of-Fit 105
3.5.2 Example: Cadmium in Horse Kidneys 109
3.5.3 Example: Brain (Left Thalamus) MRI Scan Data 110
3.5.4 Example: Human Twin Data for Quantifying Genetic (vs. Environmental) Variance 111
3.5.5 Example: Rainfall Distributions 115
3.6 Moment-Based GLD Fit to Data from a Histogram 116
3.7 The GLD and Design of Experiments 120
Problems for Chapter 3 124
References for Chapter 3 125

4 The Extended GLD System, the EGLD: Fitting by the Method of Moments 129
4.1 The Beta Distribution and its Moments 129
4.2 The Generalized Beta Distribution and its Moments 134
4.3 Estimation of GBD($\beta_1, \beta_2, \beta_3, \beta_4$) Parameters 138
4.4 GBD Approximations of Some Well-Known Distributions 144

4.4.1 The Normal Distribution . . . 145
4.4.2 The Uniform Distribution . . . 146
4.4.3 The Student's t Distribution . . . 147
4.4.4 The Exponential Distribution . . . 148
4.4.5 The Chi-Square Distribution . . . 149
4.4.6 The Gamma Distribution . . . 150
4.4.7 The Weibull Distribution . . . 151
4.4.8 The Lognormal Distribution . . . 153
4.4.9 The Beta Distribution . . . 154
4.4.10 The Inverse Gaussian Distribution . . . 155
4.4.11 The Logistic Distribution . . . 155
4.4.12 The Largest Extreme Value Distribution . . . 155
4.4.13 The Extreme Value Distribution . . . 156
4.4.14 The Double Exponential Distribution . . . 156
4.4.15 The F-Distribution . . . 156
4.4.16 The Pareto Distribution . . . 156
4.5 Examples: GBD Fits of Data, Method of Moments . . . 156
4.5.1 Example: Fitting a GBD to Simulated Data from GBD$(3, 5, 0, -0.5)$ . . . 157
4.5.2 Example: Fitting a GBD to Data Simulated from GBD$(2, 7, 1, 4)$ . . . 158
4.5.3 Example: Cadmium in Horse Kidneys . . . 160
4.5.4 Example: Rainfall Data of Section 3.5.5 . . . 161
4.5.5 Example: Tree Stand Heights and Diameters in Forestry . . . 163
4.6 EGLD Random Variate Generation . . . 167
Problems for Chapter 4 . . . 167
References for Chapter 4 . . . 168

5 A Percentile-Based Approach to Fitting Distributions and Data with the GLD **171**
5.1 The Use of Percentiles . . . 172
5.2 The (ρ_3, ρ_4)-Space of GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ . . . 174
5.3 Estimation of GLD Parameters through a Method of Percentiles . 180
5.4 GLD Approximations of Some Well-Known Distributions . . . 186
5.4.1 The Normal Distribution . . . 186
5.4.2 The Uniform Distribution . . . 188
5.4.3 The Student's t Distribution . . . 188
5.4.4 The Exponential Distribution . . . 190
5.4.5 The Chi-Square Distribution . . . 192
5.4.6 The Gamma Distribution . . . 194
5.4.7 The Weibull Distribution . . . 196

5.4.8 The Lognormal Distribution . . . 197
5.4.9 The Beta Distribution . . . 199
5.4.10 The Inverse Gaussian Distribution . . . 201
5.4.11 The Logistic Distribution . . . 203
5.4.12 The Largest Extreme Value Distribution . . . 203
5.4.13 The Extreme Value Distribution . . . 205
5.4.14 The Double Exponential Distribution . . . 206
5.4.15 The F-Distribution . . . 207
5.4.16 The Pareto Distribution . . . 209
5.4.17 Summary of Distribution Approximations . . . 210
5.5 Comparison of the Moment and Percentile Methods . . . 211
5.6 Examples: GLD Fits of Data via the Method of Percentiles . . . 214
5.6.1 Example: Data from the Cauchy Distribution . . . 215
5.6.2 Data on Radiation in Soil Samples . . . 217
5.6.3 Data on Velocities within Galaxies . . . 218
5.6.4 Rainfall Data of Sections 3.5.5 and 4.5.4 . . . 219
5.7 Percentile-Based GLD Fit of Data from a Histogram . . . 222
Problems for Chapter 5 . . . 224
References for Chapter 5 . . . 225

6 Fitting Distributions and Data with the GLD through L-Moments 227
6.1 L-Moments . . . 227
6.2 The (τ_3, τ_4)-Space of the GLD . . . 229
6.3 Estimation of GLD Parameters through L-Moments . . . 232
6.4 Approximations of Some Well-Known Distributions . . . 239
6.4.1 The Normal Distribution . . . 240
6.4.2 The Uniform Distribution . . . 242
6.4.3 The Student's t Distribution . . . 242
6.4.4 The Exponential Distribution . . . 244
6.4.5 The Chi-Square Distribution . . . 245
6.4.6 The Gamma Distribution . . . 248
6.4.7 The Weibull Distribution . . . 249
6.4.8 The Lognormal Distribution . . . 250
6.4.9 The Beta Distribution . . . 251
6.4.10 The Inverse Gaussian Distribution . . . 253
6.4.11 The Logistic Distribution . . . 253
6.4.12 The Largest Extreme Value Distribution . . . 255

6.4.13 The Extreme Value Distribution . . . 256
6.4.14 The Double Exponential Distribution . . . 256
6.4.15 The F-Distribution . . . 258
6.4.16 The Pareto Distribution . . . 260
6.5 Examples of GLD Fits to Data via L-Moments . . . 261
6.5.1 Example: Cadmium Concentration in Horse Kidneys . . . 261
6.5.2 Example: Brain MRI Scan . . . 263
6.5.3 Example: Human Twin Data . . . 264
6.5.4 Example: Rainfall Distribution . . . 266
6.5.5 Example: Data Simulated from GBD(3, 5, 0, − 0.5) . . . 268
6.5.6 Example: Data Simulated from GBD(2, 7, 1, 4) . . . 269
6.5.7 Example: Tree Stand Heights and Diameters . . . 270
6.5.8 Example: Data from the Cauchy Distribution . . . 273
6.5.9 Example: Radiation in Soil Samples . . . 274
6.5.10 Example: Velocities within Galaxies . . . 275
6.6 Fitting Data Given by a Histogram . . . 275
References for Chapter 6 . . . 277

7 Fitting a Generalized Lambda Distribution Using a Percentile-KS (P-KS) Adequacy Criterion 279
7.1 Introduction . . . 280
7.2 The Generalized Lambda Distributions . . . 281
7.2.1 Definitions . . . 281
7.2.2 Existing Parameter Estimation Methods . . . 282
7.2.3 A New P-KS Method . . . 283
7.3 GLD Modeling of Data Coming from a GLD . . . 287
7.3.1 Results on the Choice of u . . . 287
7.3.2 Influence of the Sample Size . . . 291
7.4 Gaussian Data Approached by a GLD . . . 294
7.4.1 Confidence Intervals . . . 295
7.4.2 Modeling Adequacy. . . . 296
7.5 Comparison with the Method of Moments in Three Specific Cases . . . 299
7.5.1 Gaussian Data . . . 299
7.5.2 Uniform Data . . . 300
7.5.3 Student t Data . . . 300
7.6 Conclusions . . . 301
References for Chapter 7 . . . 303
Appendix for Chapter 7 . . . 305

8 Fitting Mixture Distributions Using a Mixture of Generalized Lambda Distributions with Computer Code 311
8.1 Brief Overview of the Generalized Lambda Distribution . . . 312
8.2 The Problem of Mixture Distributions . . . 314
8.3 Estimation of Parameters of a Mixture of Two GLDs . . . 315
8.4 Graphs of the Mixture Density of Two GLDs . . . 318
8.5 Fitting the Mixture of Two GLDs to Real Data . . . 319
8.5.1 Pearson's Data . . . 319
8.5.2 Cadmium in Horse Kidneys . . . 322
8.5.3 Exchange Rate Data for Japanese Yen . . . 323
8.6 Comparison with Normal Mixtures . . . 327
8.7 Conclusions and Research Problems Regarding the Mixture of Two GLDs . . . 332
References for Chapter 8 . . . 334
Appendix for Chapter 8 . . . 338

9 GLD–2: The Bivariate GLD Distribution 363
9.1 Overview . . . 364
9.2 Plackett's Method of Bivariate d.f. Construction: The GLD–2 . . . 366
9.3 Fitting the GLD–2 to Well-Known Bivariate Distributions . . . 376
9.3.1 The Bivariate Normal (BVN) Distribution . . . 377
9.3.2 Gumbel's Bivariate Exponential Type I (BVE) . . . 382
9.3.3 Bivariate Cauchy (BVC) . . . 383
9.3.4 Kibble's Bivariate Gamma (BVG) . . . 387
9.4 GLD–2 Fits: Distributions with Non-identical Marginals . . . 391
9.4.1 Bivariate Gamma BVG with Non-identical Marginals . . . 391
9.4.2 Bivariate with Normal and Cauchy Marginals . . . 392
9.4.3 Bivariate with Gamma and "Backwards Gamma" Marginals . . . 392
9.5 Fitting GLD–2 to Datasets . . . 396
9.5.1 Algorithm for Fitting the GLD–2 to Data . . . 396
9.5.2 Example: Human Twin Data of Section 3.5.4 . . . 403
9.5.3 Example: The Rainfall Distributions of Section 3.5.5 . . . 404
9.5.4 Example: The Tree Stand Data of Section 4.5.5 . . . 405
9.6 GLD–2 Random Variate Generation . . . 407
9.7 Conclusions and Research Problems Regarding the GLD–2 . . . 409
Problems for Chapter 9 . . . 412
References for Chapter 9 . . . 413

10 Fitting the Generalized Lambda Distribution with Location and Scale-Free Shape Functionals **415**
10.1 Introduction 416
10.1.1 The Generalized Lambda Distribution 416
10.1.2 Shape Functionals 417
10.2 Description of Method 418
10.2.1 Overview 418
10.2.2 Theoretical Values of the Shape Functionals 418
10.2.3 Optimization 420
10.2.4 u, v Selection 420
10.2.5 Location and Scale Parameters 421
10.3 Simulations 422
10.3.1 Effect of Sample Size (RS Parameterization) 422
10.3.2 Effect of Sample Size (FMKL Parameterization) 424
10.3.3 Different Shapes 425
10.3.4 Overall 427
10.4 Example: Particulates 430
10.5 Approximation 431
10.6 Conclusion 431
References for Chapter 10 432

11 Statistical Design of Experiments: A Short Review **433**
11.1 Introduction to DOE 434
11.1.1 Experiments 435
11.1.2 Types of Experiments 436
11.1.3 The Independent Variable (IV) 436
11.1.4 Types of Causal (CV) or Independent (IV) Variables . . . 437
11.1.5 The Dependent Variable (DV) 437
11.1.6 When to Use DOE 438
11.1.7 Factors or Treatments 438
11.1.8 Levels 438
11.1.9 Regression Coefficients 441
11.1.10 Residuals 441
11.1.11 Optimality of Design 441
11.1.12 Optimization 442
11.1.13 Orthogonality 442
11.2 Fundamentals of DOE 446
11.2.1 Basic Principles 446
11.2.2 Practical Considerations 446
11.2.3 Designing 447
11.2.4 Randomization 451
11.2.5 Replication 452

11.2.6 Blocking . . . 452
11.2.7 Degrees of Freedom . . . 453
11.2.8 Example: Full 2^3 Factorial . . . 453
11.2.9 Summary . . . 459
11.2.10 Example: Central Composite Designs (CCD) . . . 459
11.2.11 Taguchi Designs . . . 462
11.2.12 Example . . . 464
11.2.13 Summary of Taguchi Design . . . 468
11.2.14 Latin Hypercube Sampling . . . 469
11.3 Analysis Procedures . . . 471
11.3.1 How Many Runs? . . . 472
11.3.2 Other DOE Patterns and their Usage . . . 480
References for Chapter 11 . . . 482
Appendix for Chapter 11 . . . 486

PART III: Quantile Distribution Methods 501

12 Statistical Modeling Based on Quantile Distribution Functions 503
12.1 Distributions Formulated as Quantile Functions . . . 503
12.2 Describing and Analyzing Distributional Shape . . . 509
12.3 Model Construction . . . 517
12.4 Methods of Fitting Quantile Distributions: An Overview . . . 520
12.5 Minimization Methods of Fitting Quantile Distributions . . . 521
12.5.1 Rankits and Median Rankits . . . 521
12.5.2 Distributional Least Squares (DLS) . . . 524
12.5.3 Distributional Least Absolutes (DLA) . . . 526
12.6 Fitting Parametric Regression Models Based on Quantile Functions . . . 528
12.7 Validation . . . 533
12.8 Conclusion . . . 534
References for Chapter 12 . . . 535

13 Distribution Fitting with the Quantile Function of Response Modeling Methodology (RMM) 537
13.1 The General Approach to Fitting by Response Modeling Methodology (RMM) . . . 538
13.2 The Quantile Function of the RMM Model and its Estimation . . . 542
13.2.1 Derivation of the RMM Quantile Function . . . 542
13.2.2 Estimating RMM Quantile Function . . . 546
References for Chapter 13 . . . 556

14 Fitting GLDs and Mixture of GLDs to Data Using Quantile Matching Method **557**
14.1 Introduction . . . 558
14.2 Methods . . . 559
14.3 Results . . . 561
14.3.1 Performance of Quantile Matching Estimation . . . 578
14.3.2 Quantile Matching Method for Mixture Data . . . 580
14.4 Conclusion . . . 582
References for Chapter 14 . . . 582

15 Fitting GLD to Data Using GLDEX 1.0.4 in R **585**
15.1 Introduction . . . 586
15.2 Installation and Basic GLDEX Functions . . . 589
15.3 Fitting Examples . . . 590
15.4 Fitting Empirical Data . . . 601
15.5 Future Possible Improvements to `GLDEX 1.0.4` . . . 606
15.6 Conclusion . . . 607
References for Chapter 15 . . . 607

PART IV: Other Families of Distributions **609**

16 Fitting Distributions and Data with the Johnson System via the Method of Moments **611**
16.1 Components of the Johnson System . . . 611
16.2 The S_L Component . . . 613
16.3 The S_U Component . . . 618
16.4 The S_B Component . . . 621
16.5 Approximations of Some Well-Known Distributions . . . 624
16.5.1 The Normal Distribution . . . 625
16.5.2 The Uniform Distribution . . . 626
16.5.3 The Student's t Distribution . . . 627
16.5.4 The Exponential Distribution . . . 629
16.5.5 The Chi-Square Distribution . . . 630
16.5.6 The Gamma Distribution . . . 633
16.5.7 The Weibull Distribution . . . 634
16.5.8 The Lognormal Distribution . . . 635
16.5.9 The Beta Distribution . . . 635
16.5.10 The Inverse Gaussian Distribution . . . 638
16.5.11 The Logistic Distribution . . . 639
16.5.12 The Largest Extreme Value Distribution . . . 640
16.5.13 The Extreme Value Distribution . . . 641
16.5.14 The Double Exponential Distribution . . . 641

16.5.15 The F–Distribution 642
16.5.16 The Pareto Distribution 644
16.6 Examples of Johnson System Fits to Data 645
16.6.1 Example: Cadmium Concentration in Horse Kidneys . . . 645
16.6.2 Example: Brain MRI Scan 647
16.6.3 Example: Human Twin Data 648
16.6.4 Example: Rainfall Distribution 650
16.6.5 Example: Data Simulated from GBD(3, 5, 0, -0.5) 652
16.6.6 Example: Data Simulated from GBD(2, 7, 1, 4) 653
16.6.7 Example: Tree Stand Heights and Diameters 654
16.6.8 Example: Data from the Cauchy Distribution 657
16.6.9 Example: Radiation in Soil Samples 658
16.6.10 Example: Velocities within Galaxies 659
16.7 Fitting Data Given by a Histogram 660
References for Chapter 16 663

17 Fitting Distributions and Data with the Kappa Distribution through L-Moments and Percentiles 665
17.1 The Kappa Distribution 666
17.2 Estimation of Kappa Parameters via L-Moments 669
17.3 Estimation of Kappa Parameters via Percentiles 674
17.4 Approximations of Some Well-Known Distributions 679
17.4.1 The Normal Distribution 679
17.4.2 The Uniform Distribution 680
17.4.3 The Student's t Distribution 681
17.4.4 The Exponential Distribution 684
17.4.5 The Chi-Square Distribution 685
17.4.6 The Gamma Distribution 688
17.4.7 The Weibull Distribution 689
17.4.8 The Lognormal Distribution 690
17.4.9 The Beta Distribution 692
17.4.10 The Inverse Gaussian Distribution 694
17.4.11 The Logistic Distribution 695
17.4.12 The Largest Extreme Value Distribution 696
17.4.13 The Extreme Value Distribution 696
17.4.14 The Double Exponential Distribution 696
17.4.15 The F-Distribution 696
17.4.16 The Pareto Distribution 698
17.5 Examples of Kappa Distribution Fits to Data 699
17.5.1 Example: Cadmium Concentration in Horse Kidneys . . . 699
17.5.2 Example: Brain MRI Scan 701
17.5.3 Example: Human Twin Data 702

17.5.4 Example: Rainfall Distribution 705
17.5.5 Example: Data Simulated from GBD(3, 5, 0, −0.5) 708
17.5.6 Example: Data Simulated from GBD(2, 7, 1, 4) 709
17.5.7 Example: Tree Stand Heights and Diameters 711
17.5.8 Data from the Cauchy Distribution 714
17.5.9 Example: Radiation in Soil Samples 714
17.5.10 Velocities within Galaxies 714
17.6 Fitting Data Given by a Histogram 716
References for Chapter 17 718

18 Weighted Distributional L_α Estimates **719**
18.1 Introduction 720
18.1.1 Is the Normal Distribution Normal? 721
18.1.2 Weighted L_α Regression 722
18.2 Probability-Based Partially Adaptive Estimation 724
18.2.1 Not-Necessarily Gaussian Error Distributions 726
18.2.2 Estimation of the Parameters 730
18.2.3 Probability-Based Distributional Regression 732
18.3 Quantile-Based Partially Adaptive Estimation 736
18.3.1 Quantile Models 737
18.3.2 Quantile-Based Distributional Regression 742
18.4 Controlled Random Search 745
18.5 Goodness-of-Fit Assessment 748
18.6 Empirical Examples 752
18.6.1 Mayer's Data 754
18.6.2 Martin Marietta Data 754
18.6.3 Prostate Cancer Data 757
18.6.4 Salinity Data 760
18.6.5 Gaussian Data 760
18.7 Conclusions and Future Research 763
References for Chapter 18 764
Appendix for Chapter 18 771

19 A Multivariate Gamma Distribution for Linearly Related Proportional Outcomes **787**
19.1 Introduction 788
19.2 Definitions 789
19.3 Basic Concepts 793
19.4 The Fatal Shock Model 802
19.5 Example 807
19.6 Conclusion 808
References for Chapter 19 808
Appendix for Chapter 19 811

PART V: The Generalized Bootstrap and Monte Carlo Methods 813

20 The Generalized Bootstrap (GB) and Monte Carlo (MC) Methods 815
20.1 The Generalized Bootstrap (GB) Method 816
20.2 Comparisons of the GB and BM Methods 823
Problems for Chapter 20 823
References for Chapter 20 824

21 The Generalized Bootstrap: A New Fitting Strategy and Simulation Study Showing Advantage over Bootstrap Percentile Methods 827
21.1 Introduction 828
21.2 Algorithms for Three Methods of Distribution Fitting 831
21.2.1 Fitting the GLD through the Method of Moments (MOM) 831
21.2.2 EGLD: Method of Generalized Beta Distribution (MGBD) 831
21.2.3 Method of Percentiles (MOP) 832
21.3 Fitting Strategy and Simulation Study 833
21.4 Specific Examples 837
21.4.1 Sample 1: GBD, Not Covered 838
21.4.2 Sample 5: GBD, Covered 838
21.4.3 Sample 12: MOM, Not Covered 840
21.4.4 Sample 19: MOM, Covered 843
21.4.5 Sample 21: GBD, Covered 843
21.4.6 Sample 28: MOP, Not Covered 846
21.5 Conclusions 852
Acknowledgments 852
References for Chapter 21 852
Appendix for Chapter 21 854

22 Generalized Bootstrap Confidence Intervals for High Quantiles 877
22.1 Introduction 878
22.1.1 High Quantile Estimation 878
22.1.2 The Bootstrap Method 879
22.2 Generalized Lambda Distribution and Generalized Bootstrap 881
22.3 Comparisons of Bootstrap Confidence Intervals for High Quantiles 882
22.3.1 Simulated Distributions 883

22.3.2 Choices of Quantile Levels and Sample Sizes 884
22.3.3 Criteria for Performance Evaluation 886
22.3.4 Simulation Algorithms 887
22.4 Simulation Results and Discussion 888
22.4.1 Performances for the Beta Distributions 890
22.4.2 Performances for the Gamma, Weibull, and Normal Distributions 890
22.5 Conclusions 897
References for Chapter 22 897
Appendix for Chapter 22 901

PART VI: Assessment of the Quality of Fits 915

23 Goodness-of-Fit Criteria Based on Observations Quantized by Hypothetical and Empirical Percentiles 917
23.1 Data and their Statistical Models 918
23.2 Assessment of Goodness-of-Fit 921
23.2.1 Special Distances, Divergences, and Disparities 922
23.2.2 Examples 927
23.3 Criteria of Goodness-of-Fit 931
23.3.1 Disparities, Divergences, and Metric Distances 931
23.3.2 Metricity and Robustness 937
23.4 Disparities Based on Partitions 939
23.4.1 Partitioning by Hypothetical Percentiles 943
23.4.2 Partitioning by Empirical Percentiles 943
23.5 Goodness-of-Fit Statistics Based on Spacings 947
23.5.1 Objectives of the Following Sections 947
23.5.2 Types of Statistics Studied 950
23.5.3 Structural Spacings Statistics 956
23.5.4 Organization of the Following Sections 958
23.6 Asymptotic Properties of Structural Statistics 958
23.6.1 Asymptotic Equivalence 959
23.6.2 Assumptions and Notations 960
23.6.3 Consistency under Hypothesis and Fixed Alternatives . . . 961
23.6.4 Asymptotic Normality under Local Alternatives 962
23.6.5 Asymptotic Normality under Fixed Alternatives 962
23.7 Asymptotic Properties of Power Spacings Statistics 964
23.7.1 Power Spacing Statistics 964
23.7.2 Consistency 968
23.7.3 Asymptotic Normality under Local Alternatives 969
23.7.4 Asymptotic Normality under Fixed Alternatives 971

23.7.5 Discussion 974
23.8 The PODISTAT Program Package 976
23.9 Proofs of Assertions 982
23.9.1 Proofs for Structural Spacings Statistics 982
23.9.2 Proofs for Power Spacings Statistics 986
Acknowledgments 991
References for Chapter 23 991

24 Evidential Support Continuum (ESC): A New Approach to Goodness-of-Fit Assessment, which Addresses Conceptual and Practical Challenges 995
24.1 Introduction 996
24.2 Challenges to Effective G-O-F Testing 997
24.2.1 Conceptual and Theoretical Challenges 997
24.2.2 Practical Application Challenges 1000
24.2.3 Addressing Goodness-of-Fit Challenges: A New Approach . 1001
24.3 Context 1001
24.3.1 Data Description 1001
24.3.2 Distributions Considered 1002
24.4 The Method 1003
24.4.1 Eight Pieces of Evidence 1003
24.4.2 The Evidential Support Continuum (ESC) Method 1003
24.4.3 Quantitative Evaluations 1004
24.4.4 Graphical Evaluations 1008
24.5 Results 1011
24.5.1 The χ^2 Test 1012
24.5.2 The K-S Test 1013
24.5.3 Distribution Support to Spread of Data (Adequacy and Appropriateness 1013
24.5.4 Fit of the Distribution p.d.f. to the Dataset Histogram (Main Body and Tails) 1016
24.5.5 Fit of the Distribution c.d.f. to the Dataset e.d.f. (Main Body and Tails) 1017
24.5.6 Constructing the ESC Diagram 1020
24.6 Conclusions 1021
24.6.1 ESCs Address Conceptual and Theoretical Challenges to Effective G-O-F 1021
24.6.2 ESCs Address Practical Application Challenges to G-O-F 1023
References for Chapter 24 1026
Appendix for Chapter 24 1029

25 Estimation of Sampling Distributions of the Overlapping Coefficient and Other Similarity Measures 1039
25.1 Introduction . . . 1040
25.2 Measures of Similarity . . . 1043
25.2.1 Overlap Coefficient δ . . . 1044
25.2.2 Matusita's Measure ρ . . . 1047
25.2.3 Morisita's Measure λ . . . 1047
25.2.4 MacArthur-Levins' Measure α^* . . . 1049
25.2.5 Common Properties . . . 1049
25.3 Measures of Similarity for Exponential Populations . . . 1050
25.4 Measures of Similarity for Normal Populations . . . 1053
25.4.1 The Equal Means Case . . . 1053
25.4.2 The Equal Variances Case . . . 1057
25.4.3 The General Case . . . 1057
25.5 Sampling Distributions: Exponential Populations Case . . . 1061
25.6 Sampling Distributions: Normal Populations Case . . . 1069
25.7 Conclusions . . . 1083
Acknowledgments . . . 1085
References for Chapter 25 . . . 1085

PART VII: Applications 1091

26 Fitting Statistical Distribution Functions to Small Datasets 1093
26.1 Introduction . . . 1094
26.2 Analysis Techniques . . . 1097
26.3 The Johnson Family of Distributions and the JFit and GAFit Tools . . . 1098
26.4 The Empirical Distribution Function (edf) . . . 1100
26.5 Goodness-of-Fit (GOF) . . . 1101
26.6 Example 1: The SST1 Dataset . . . 1103
26.6.1 Johnson Family Distributions that Minimize KS Scores . . . 1109
26.6.2 The Four-Parameter Generalized Lambda Distribution . . . 1110
26.7 Example 2: The SST2 Dataset . . . 1113
26.8 Example 3: The SST3 Dataset . . . 1118
26.9 Example 4: The SST4 Dataset . . . 1121
26.10 Example 5: The SST5 Dataset . . . 1123
26.11 Example 6: The SST47 Dataset . . . 1127
26.12 Example 7: The SST29 Dataset . . . 1129
26.13 Example 8: Solar Flux (Top 95% Amplitude) at 10.7 MHz . . . 1131

26.14 A Bimodal Fit to Data 1132
26.15 Comparisons 1135
26.16 Model Selection 1136
26.17 Summary 1138
References for Chapter 26 1138
Appendix for Chapter 26 1140

27 Mixed Truncated Random Variable Fitting with the GLD, and Applications in Insurance and Inventory Management 1145
27.1 Introduction 1146
27.2 The Generalized Lambda Distribution and its Partial Moments . . 1148
27.3 Moments of Mixed Type of a Truncated Random Variable 1150
27.4 Applications 1155
27.4.1 Optimum Deductible in Insurance Purchasing 1155
27.4.2 Newsboy Model: Solution under Utility Maximization . . . 1161
27.5 Conclusion 1167
Acknowledgments 1167
References for Chapter 27 1167

28 Distributional Modeling of Pipeline Leakage Repair Costs for a Water Utility Company 1171
28.1 Introduction 1172
28.2 Generalized Lambda Distributions 1172
28.2.1 Basic Theory 1173
28.2.2 Fitting Methods 1173
28.3 Factors Influencing the Costs of Pipe Repairs 1176
28.3.1 Continuous Data: Pipe Length and Age of Pipe 1176
28.3.2 Discrete Data: Pipe Diameter, Pipe Location, and Month 1177
28.3.3 GLD Fits on the Repair Costs in Relation to Pipe Diameter, Pipe Location, and Month 1177
28.3.4 GLD Fits to Repair Costs on Pipelines with Diameter 300 in LOCA and LOCB Regions 1185
28.3.5 Remarks 1185
28.4 Rationally Setting Water Prices to Break Even: A Case of Pipe Repair Costs 1185
28.4.1 Global Pipe Repair Cost Example 1185
28.4.2 Individual Pipe Repair Cost Example 1189
28.4.3 Other Considerations 1200
28.5 Further Considerations in Choosing the GLD for Analysis 1203
28.6 Conclusion 1203
References for Chapter 28 1203

29 Use of the Generalized Lambda Distribution in Materials Science, with Examples in Fatigue Lifetime, Fracture Mechanics, Polycrystalline Calculations, and Pitting Corrosion 1207
29.1 Fatigue Lifetimes . . . 1208
29.1.1 Lifetime Distributions . . . 1209
29.1.2 Crack Initiation and Crack Propagation . . . 1211
29.2 Fracture Mechanics . . . 1217
29.2.1 Fracture of Adhesively Bonded Joints Composed of Pultruted Adherends . . . 1218
29.2.2 Modelling the Nuclear Reactor Pressure Vessel Steel Brittle Fracture . . . 1221
29.3 Extreme Values . . . 1227
29.3.1 Pitting Corrosion . . . 1229
29.3.2 Roughness . . . 1233
29.4 Conclusion . . . 1238
References for Chapter 29 . . . 1238

30 Fitting Statistical Distributions to Data in Hurricane Modeling 1245
30.1 Introduction . . . 1246
30.2 Hurricane Modeling Basics . . . 1247
30.3 Distributional Fitting at Individual Sites . . . 1250
30.4 Data Quality and Preparation Issues . . . 1252
30.5 Predictive Ability and Performance . . . 1255
30.6 Case Study of the New Orleans Levees: Wrath of Mother Nature or Ordinary Extreme Event . . . 1258
30.7 Conclusions . . . 1260
Acknowledgments . . . 1261
References for Chapter 30 . . . 1261

31 A Rainfall-Based Model for Predicting the Regional Incidence of Wheat Seed Infection by *Stagonospora nodorum* in New York 1263
31.1 Introduction . . . 1264
31.2 Materials and Methods . . . 1265
31.2.1 Probability Distribution for Seed Infection Incidence . . . 1265
31.2.2 Relating Mean Seed Infection to Rainfall . . . 1268
31.2.3 Binary Power Law . . . 1270
31.2.4 Model Verification and Validation . . . 1271
31.3 Results . . . 1272
31.3.1 Probability Distribution for Seed Infection Incidence . . . 1272
31.3.2 Relating Mean Seed Infection to Rainfall . . . 1272
31.3.3 Binary Power Law . . . 1272

31.3.4 Model Verification and Validation 1274
31.4 Discussion 1277
References for Chapter 31 1278

32 Reliability Estimation Using Univariate Dimension Reduction and Extended Generalized Lambda Distribution 1281
32.1 Introduction 1282
32.2 Calculation of Statistical Moments 1284
32.3 Estimating the Distribution of Performance Function 1286
32.3.1 Overview of GLD and GBD 1287
32.3.2 Estimation of EGLD Parameters 1289
32.3.3 Probability of Failure Estimation Using EGLD 1291
32.4 The UDR + EGLD Algorithm 1291
32.5 Example Problems 1292
32.5.1 Example 1: A Concave Function 1292
32.5.2 Example 2: A Non-Linear Function 1293
32.5.3 Example 3: Performance Function with Infinite MPP . . . 1294
32.5.4 Example 4: Vehicle Crash 1296
32.5.5 Example 5: I-Beam Design Problem 1297
32.5.6 Example 6: Checking for Formulation Invariance 1300
32.5.7 Effect of Probability Level 1300
32.6 Concluding Remarks 1301
Acknowledgments 1302
References for Chapter 32 1302

33 Statistical Analyses of Environmental Pressure Surrounding Atlantic Tropical Cyclones 1305
33.1 Introduction 1306
33.2 Historical Data Sources 1306
33.3 Distribution of Environmental Pressure 1308
33.4 Environmental Pressure for Specific Storms 1311
33.5 Relationship of Penv to Latitude, Time of Year, Pmin 1313
33.6 Extrapolation to the Full Historical Record 1315
33.7 Wind Field Impacts 1320
33.8 Conclusions 1323
References for Chapter 33 1323

34 Simulating Hail Storms Using Simultaneous Efficient Random Number Generators 1325
34.1 Simulation of a Hail Precipitation System 1326
34.2 Models Used 1331
34.3 Conclusions 1333

34.4 Tests of Simultaneous Generators 1334
34.5 Introduction of Several Generators in a Simulation 1335
34.6 Statistical Analysis 1337
References for Chapter 34 1339
Appendix for Chapter 34 1342

PART VIII: Appendices **1347**

A Programs and their Documentation 1349
A.1 General Computational Issues 1350
A.2 General Functions 1351
A.3 Functions for GLD Computations 1352
A.4 Functions for GBD Fits 1358
A.5 Functions for the Kappa Distribution 1360
A.6 Maple Programs for Johnson System Fits 1364
A.7 The Maple Code for the Bivariate GLD 1365
A.8 Content of the Attached CD 1365
A.9 The **R** Code of the Programs for GLD Fits 1367
A.10 **R** Code of the Programs for Kappa Distribution Fits 1387
A.11 Maple Code for Johnson System Fits 1397
A.12 Maple Code for Bivariate GLD (GLD–2) Fits 1404
B Table B–1 for GLD Fits: Method of Moments 1407
C Table C–1 for GBD Fits: Method of Moments 1429
D Tables D–1 through D–5 for GLD Fits: Method of Percentiles . . 1455
E Tables E–1 through E–5 for GLD Fits: Method of L-Moments . . 1511
F Table F–1 for Kappa Distribution Fits: Method of L-Moments . . 1557
G Table G–1 for Kappa Distribution Fits: Method of Percentiles . . 1589
H Table H–1 for Johnson System Fits in the S_U Region: Method of Moments 1611
I Table I–1 for Johnson System Fits in the S_B Region: Method of Moments 1627
J Table J–1 for p-Values Associated with Kolmogorov-Smirnov Statistics 1659
K Table K–1 Normal Distribution Percentiles 1661

Subject Index **1665**