Past, Present, and Future of Statistical Science

Edited by
Xihong Lin, Christian Genest,
David L. Banks, Geert Molenberghs,
David W. Scott, and Jane-Ling Wang

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

COPSS
Committee of Presidents
of Statistical Societies

Contents

Preface xvii

Contributors xxi

I The history of COPSS 1

1 A brief history of the Committee of Presidents of Statistical Societies (COPSS) 3
Ingram Olkin
1.1 Introduction 3
1.2 COPSS activities in the early years 6
1.3 COPSS activities in recent times 8
1.4 Awards 10

II Reminiscences and personal reflections on career paths 21

2 Reminiscences of the Columbia University Department of Mathematical Statistics in the late 1940s 23
Ingram Olkin
2.1 Introduction: Pre-Columbia 23
2.2 Columbia days 24
2.3 Courses 26

3 A career in statistics 29
Herman Chernoff
3.1 Education 29
3.2 Postdoc at University of Chicago 32
3.3 University of Illinois and Stanford 34
3.4 MIT and Harvard 38

4 "...how wonderful the field of statistics is..." 41
David R. Brillinger
4.1 Introduction 41
4.2 The speech (edited some) 42
4.3 Conclusion 45

5 An unorthodox journey to statistics: Equity issues, remarks on multiplicity **49**

Juliet Popper Shaffer

5.1 Pre-statistical career choices 49
5.2 Becoming a statistician 50
5.3 Introduction to and work in multiplicity 52
5.4 General comments on multiplicity 54

6 Statistics before and after my COPSS Prize **59**

Peter J. Bickel

6.1 Introduction 59
6.2 The foundation of mathematical statistics 59
6.3 My work before 1979 60
6.4 My work after 1979 62
6.5 Some observations 67

7 The accidental biostatistics professor **73**

Donna J. Brogan

7.1 Public school and passion for mathematics 73
7.2 College years and discovery of statistics 74
7.3 Thwarted employment search after college 76
7.4 Graduate school as a fallback option 76
7.5 Master's degree in statistics at Purdue 77
7.6 Thwarted employment search after Master's degree 77
7.7 Graduate school again as a fallback option 77
7.8 Dissertation research and family issues 78
7.9 Job offers — finally! 79
7.10 Four years at UNC-Chapel Hill 79
7.11 Thirty-three years at Emory University 80
7.12 Summing up and acknowledgements 81

8 Developing a passion for statistics **83**

Bruce G. Lindsay

8.1 Introduction 83
8.2 The first statistical seeds 85
8.3 Graduate training 85
8.4 The PhD 88
8.5 Job and postdoc hunting 92
8.6 The postdoc years 92
8.7 Starting on the tenure track 93

9 Reflections on a statistical career and their implications **97**

R. Dennis Cook

9.1 Early years 97
9.2 Statistical diagnostics 100

9.3 Optimal experimental design 104
9.4 Enjoying statistical practice 105
9.5 A lesson learned 106

10 Science mixes it up with statistics **109**
Kathryn Roeder
10.1 Introduction 109
10.2 Collaborators 110
10.3 Some collaborative projects 111
10.4 Conclusions 114

11 Lessons from a twisted career path **117**
Jeffrey S. Rosenthal
11.1 Introduction 117
11.2 Student days 118
11.3 Becoming a researcher 122
11.4 Final thoughts 127

12 Promoting equity **129**
Mary W. Gray
12.1 Introduction 129
12.2 The Elizabeth Scott Award 130
12.3 Insurance 132
12.4 Title IX 134
12.5 Human rights 134
12.6 Underrepresented groups 136

III Perspectives on the field and profession **139**

13 Statistics in service to the nation **141**
Stephen E. Fienberg
13.1 Introduction 141
13.2 The National Halothane Study 143
13.3 The President's Commission and CNSTAT 144
13.4 Census-taking and multiple-systems estimation 145
13.5 Cognitive aspects of survey methodology 146
13.6 Privacy and confidentiality 147
13.7 The accuracy of the polygraph 148
13.8 Take-home messages 149

14 Where are the majors? **153**
Iain M. Johnstone
14.1 The puzzle 153
14.2 The data 153
14.3 Some remarks 154

15 We live in exciting times **157**
Peter G. Hall
15.1 Introduction 157
15.2 Living with change 159
15.3 Living the revolution 161

16 The bright future of applied statistics **171**
Rafael A. Irizarry
16.1 Introduction 171
16.2 Becoming an applied statistician 171
16.3 Genomics and the measurement revolution 172
16.4 The bright future 175

17 The road travelled: From statistician to statistical scientist **177**
Nilanjan Chatterjee
17.1 Introduction 177
17.2 Kin-cohort study: My gateway to genetics 178
17.3 Gene-environment interaction: Bridging genetics and theory of case-control studies 179
17.4 Genome-wide association studies (GWAS): Introduction to big science 181
17.5 The post-GWAS era: What does it all mean? 183
17.6 Conclusion 184

18 A journey into statistical genetics and genomics **189**
Xihong Lin
18.1 The 'omics era 189
18.2 My move into statistical genetics and genomics 191
18.3 A few lessons learned 192
18.4 A few emerging areas in statistical genetics and genomics . . 193
18.5 Training the next generation statistical genetic and genomic scientists in the 'omics era 197
18.6 Concluding remarks 199

19 Reflections on women in statistics in Canada **203**
Mary E. Thompson
19.1 A glimpse of the hidden past 203
19.2 Early historical context 204
19.3 A collection of firsts for women 206
19.4 Awards 209
19.5 Builders 210
19.6 Statistical practice 212
19.7 The current scene 213

20 "The whole women thing" **217**
Nancy M. Reid
20.1 Introduction 217
20.2 "How many women are there in your department?" 218
20.3 "Should I ask for more money?" 220
20.4 "I'm honored" 221
20.5 "I loved that photo" 224
20.6 Conclusion 225

21 Reflections on diversity **229**
Louise M. Ryan
21.1 Introduction 229
21.2 Initiatives for minority students 230
21.3 Impact of the diversity programs 231
21.4 Gender issues 233

IV Reflections on the discipline **235**

22 Why does statistics have two theories? **237**
Donald A.S. Fraser
22.1 Introduction 237
22.2 65 years and what's new 239
22.3 Where do the probabilities come from? 240
22.4 Inference for regular models: Frequency 243
22.5 Inference for regular models: Bootstrap 245
22.6 Inference for regular models: Bayes 246
22.7 The frequency-Bayes contradiction 247
22.8 Discussion 248

23 Conditioning is the issue **253**
James O. Berger
23.1 Introduction 253
23.2 Cox example and a pedagogical example 254
23.3 Likelihood and stopping rule principles 255
23.4 What it means to be a frequentist 257
23.5 Conditional frequentist inference 259
23.6 Final comments 264

24 Statistical inference from a Dempster–Shafer perspective **267**
Arthur P. Dempster
24.1 Introduction 267
24.2 Personal probability 268
24.3 Personal probabilities of "don't know" 269
24.4 The standard DS protocol 271
24.5 Nonparametric inference 275

24.6 Open areas for research 276

25 Nonparametric Bayes **281**
David B. Dunson
25.1 Introduction 281
25.2 A brief history of NP Bayes 284
25.3 Gazing into the future 287

26 How do we choose our default methods? **293**
Andrew Gelman
26.1 Statistics: The science of defaults 293
26.2 Ways of knowing 295
26.3 The pluralist's dilemma 297
26.4 Conclusions 299

27 Serial correlation and Durbin–Watson bounds **303**
T. W. Anderson
27.1 Introduction 303
27.2 Circular serial correlation 304
27.3 Periodic trends 305
27.4 Uniformly most powerful tests 305
27.5 Durbin–Watson 306

28 A non-asymptotic walk in probability and statistics **309**
Pascal Massart
28.1 Introduction 309
28.2 Model selection 310
28.3 Welcome to Talagrand's wonderland 315
28.4 Beyond Talagrand's inequality 318

29 The past's future is now: What will the present's future bring? **323**
Lynne Billard
29.1 Introduction 323
29.2 Symbolic data 324
29.3 Illustrations 325
29.4 Conclusion 331

30 Lessons in biostatistics **335**
Norman E. Breslow
30.1 Introduction 335
30.2 It's the science that counts 336
30.3 Immortal time 338
30.4 Multiplicity 341
30.5 Conclusion 345

31 A vignette of discovery **349**
Nancy Flournoy
31.1 Introduction 349
31.2 CMV infection and clinical pneumonia 350
31.3 Interventions 354
31.4 Conclusions 357

32 Statistics and public health research **359**
Ross L. Prentice
32.1 Introduction 359
32.2 Public health research 361
32.3 Biomarkers and nutritional epidemiology 362
32.4 Preventive intervention development and testing 363
32.5 Clinical trial data analysis methods 365
32.6 Summary and conclusion 365

33 Statistics in a new era for finance and health care **369**
Tze Leung Lai
33.1 Introduction 369
33.2 Comparative effectiveness research clinical studies 370
33.3 Innovative clinical trial designs in translational medicine . . 371
33.4 Credit portfolios and dynamic empirical Bayes in finance . . 373
33.5 Statistics in the new era of finance 375
33.6 Conclusion 376

34 Meta-analyses: Heterogeneity can be a good thing **381**
Nan M. Laird
34.1 Introduction 381
34.2 Early years of random effects for meta-analysis 382
34.3 Random effects and clinical trials 383
34.4 Meta-analysis in genetic epidemiology 385
34.5 Conclusions 387

35 Good health: Statistical challenges in personalizing disease prevention **391**
Alice S. Whittemore
35.1 Introduction 391
35.2 How do we personalize disease risks? 391
35.3 How do we evaluate a personal risk model? 393
35.4 How do we estimate model performance measures? 394
35.5 Can we improve how we use epidemiological data for risk model assessment? 397
35.6 Concluding remarks 401

36 Buried treasures **405**
Michael A. Newton
36.1 Three short stories 405
36.2 Concluding remarks 409

37 Survey sampling: Past controversies, current orthodoxy, and future paradigms **413**
Roderick J.A. Little
37.1 Introduction 413
37.2 Probability or purposive sampling? 415
37.3 Design-based or model-based inference? 416
37.4 A unified framework: Calibrated Bayes 423
37.5 Conclusions 425

38 Environmental informatics: Uncertainty quantification in the environmental sciences **429**
Noel Cressie
38.1 Introduction 429
38.2 Hierarchical statistical modeling 430
38.3 Decision-making in the presence of uncertainty 431
38.4 Smoothing the data 433
38.5 EI for spatio-temporal data 434
38.6 The knowledge pyramid 444
38.7 Conclusions 444

39 A journey with statistical genetics **451**
Elizabeth A. Thompson
39.1 Introduction 451
39.2 The 1970s: Likelihood inference and the EM algorithm . . . 452
39.3 The 1980s: Genetic maps and hidden Markov models 454
39.4 The 1990s: MCMC and complex stochastic systems 455
39.5 The 2000s: Association studies and gene expression 457
39.6 The 2010s: From association to relatedness 458
39.7 To the future 458

40 Targeted learning: From MLE to TMLE **465**
Mark van der Laan
40.1 Introduction 465
40.2 The statistical estimation problem 467
40.3 The curse of dimensionality for the MLE 469
40.4 Super learning 473
40.5 Targeted learning 474
40.6 Some special topics 476
40.7 Concluding remarks 477

41 Statistical model building, machine learning, and the ah-ha moment **481**
Grace Wahba
41.1 Introduction: Manny Parzen and RKHS 481
41.2 Regularization methods, RKHS and sparse models 490
41.3 Remarks on the nature-nurture debate, personalized medicine and scientific literacy 491
41.4 Conclusion 492

42 In praise of sparsity and convexity **497**
Robert J. Tibshirani
42.1 Introduction 497
42.2 Sparsity, convexity and ℓ_1 penalties 498
42.3 An example 500
42.4 The covariance test 500
42.5 Conclusion 503

43 Features of Big Data and sparsest solution in high confidence set **507**
Jianqing Fan
43.1 Introduction 507
43.2 Heterogeneity 508
43.3 Computation 509
43.4 Spurious correlation 510
43.5 Incidental endogeneity 512
43.6 Noise accumulation 515
43.7 Sparsest solution in high confidence set 516
43.8 Conclusion 521

44 Rise of the machines **525**
Larry A. Wasserman
44.1 Introduction 525
44.2 The conference culture 526
44.3 Neglected research areas 527
44.4 Case studies 527
44.5 Computational thinking 533
44.6 The evolving meaning of data 534
44.7 Education and hiring 535
44.8 If you can't beat them, join them 535

45 A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it) **537**
Xiao-Li Meng
45.1 Nobel Prize? Why not COPSS? 537
45.2 Multi-resolution inference 539

45.3 Multi-phase inference 545
45.4 Multi-source inference 551
45.5 The ultimate prize or price 557

V Advice for the next generation 563

46 Inspiration, aspiration, ambition 565
C.F. Jeff Wu
46.1 Searching the source of motivation 565
46.2 Examples of inspiration, aspiration, and ambition 566
46.3 Looking to the future 567

47 Personal reflections on the COPSS Presidents' Award 571
Raymond J. Carroll
47.1 The facts of the award 571
47.2 Persistence 571
47.3 Luck: Have a wonderful Associate Editor 572
47.4 Find brilliant colleagues 572
47.5 Serendipity with data 574
47.6 Get fascinated: Heteroscedasticity 575
47.7 Find smart subject-matter collaborators 575
47.8 After the Presidents' Award 577

48 Publishing without perishing and other career advice 581
Marie Davidian
48.1 Introduction 581
48.2 Achieving balance, and how you never know 582
48.3 Write it, and write it again 586
48.4 Parting thoughts 590

49 Converting rejections into positive stimuli 593
Donald B. Rubin
49.1 My first attempt 594
49.2 I'm learning 594
49.3 My first JASA submission 595
49.4 Get it published! 596
49.5 Find reviewers who understand 597
49.6 Sometimes it's easy, even with errors 598
49.7 It sometimes pays to withdraw the paper! 598
49.8 Conclusion 601

50 The importance of mentors 605
Donald B. Rubin
50.1 My early years 605
50.2 The years at Princeton University 606

50.3 Harvard University — the early years 608
50.4 My years in statistics as a PhD student 609
50.5 The decade at ETS 610
50.6 Interim time in DC at EPA, at the University of Wisconsin, and the University of Chicago 611
50.7 The three decades at Harvard 612
50.8 Conclusions 612

51 Never ask for or give advice, make mistakes, accept mediocrity, enthuse **615**
Terry Speed
51.1 Never ask for or give advice 615
51.2 Make mistakes 616
51.3 Accept mediocrity 617
51.4 Enthuse 618

52 Thirteen rules **621**
Bradley Efron
52.1 Introduction 621
52.2 Thirteen rules for giving a really bad talk 621