## Contents

Preface, xv

Acknowledgments, xix

1 Introduction, 1
   1.1 Systems and their characteristics, 1
      1.1.1 Classes of systems, 1
      1.1.2 System states, 1
      1.1.3 Change of state, 2
      1.1.4 Thermodynamic entropy, 3
      1.1.5 Evolutive connotation of entropy, 5
      1.1.6 Statistical mechanical entropy, 5
   1.2 Informational entropies, 7
      1.2.1 Types of entropies, 8
      1.2.2 Shannon entropy, 9
      1.2.3 Information gain function, 12
      1.2.4 Boltzmann, Gibbs and Shannon entropies, 14
      1.2.5 Negentropy, 15
      1.2.6 Exponential entropy, 16
      1.2.7 Tsallis entropy, 18
      1.2.8 Renyi entropy, 19
   1.3 Entropy, information, and uncertainty, 21
      1.3.1 Information, 22
      1.3.2 Uncertainty and surprise, 24
   1.4 Types of uncertainty, 25
   1.5 Entropy and related concepts, 27
      1.5.1 Information content of data, 27
      1.5.2 Criteria for model selection, 28
      1.5.3 Hypothesis testing, 29
      1.5.4 Risk assessment, 29

Questions, 29

References, 31

Additional References, 32
2 Entropy Theory, 33
  2.1 Formulation of entropy, 33
  2.2 Shannon entropy, 39
  2.3 Connotations of information and entropy, 42
    2.3.1 Amount of information, 42
    2.3.2 Measure of information, 43
    2.3.3 Source of information, 43
    2.3.4 Removal of uncertainty, 44
    2.3.5 Equivocation, 45
    2.3.6 Average amount of information, 45
    2.3.7 Measurement system, 46
    2.3.8 Information and organization, 46
  2.4 Discrete entropy: univariate case and marginal entropy, 46
  2.5 Discrete entropy: bivariate case, 52
    2.5.1 Joint entropy, 53
    2.5.2 Conditional entropy, 53
    2.5.3 Transinformation, 57
  2.6 Dimensionless entropies, 79
  2.7 Bayes theorem, 80
  2.8 Informational correlation coefficient, 88
  2.9 Coefficient of nontransferred information, 90
  2.10 Discrete entropy: multidimensional case, 92
  2.11 Continuous entropy, 93
    2.11.1 Univariate case, 94
    2.11.2 Differential entropy of continuous variables, 97
    2.11.3 Variable transformation and entropy, 99
    2.11.4 Bivariate case, 100
    2.11.5 Multivariate case, 105
  2.12 Stochastic processes and entropy, 105
  2.13 Effect of proportional class interval, 107
  2.14 Effect of the form of probability distribution, 110
  2.15 Data with zero values, 111
  2.16 Effect of measurement units, 113
  2.17 Effect of averaging data, 115
  2.18 Effect of measurement error, 116
  2.19 Entropy in frequency domain, 118
  2.20 Principle of maximum entropy, 118
  2.21 Concentration theorem, 119
  2.22 Principle of minimum cross entropy, 122
  2.23 Relation between entropy and error probability, 123
  2.24 Various interpretations of entropy, 125
    2.24.1 Measure of randomness or disorder, 125
    2.24.2 Measure of unbiasedness or objectivity, 125
    2.24.3 Measure of equality, 125
    2.24.4 Measure of diversity, 126
    2.24.5 Measure of lack of concentration, 126
    2.24.6 Measure of flexibility, 126
2.24.7 Measure of complexity, 126
2.24.8 Measure of departure from uniform distribution, 127
2.24.9 Measure of interdependence, 127
2.24.10 Measure of dependence, 128
2.24.11 Measure of interactivity, 128
2.24.12 Measure of similarity, 129
2.24.13 Measure of redundancy, 129
2.24.14 Measure of organization, 130

2.25 Relation between entropy and variance, 133
2.26 Entropy power, 135
2.27 Relative frequency, 135
2.28 Application of entropy theory, 136
Questions, 136
References, 137
Additional Reading, 139

3 Principle of Maximum Entropy, 142
3.1 Formulation, 142
3.2 POME formalism for discrete variables, 145
3.3 POME formalism for continuous variables, 152
  3.3.1 Entropy maximization using the method of Lagrange multipliers, 152
  3.3.2 Direct method for entropy maximization, 157
3.4 POME formalism for two variables, 158
3.5 Effect of constraints on entropy, 165
3.6 Invariance of total entropy, 167
Questions, 168
References, 170
Additional Reading, 170

4 Derivation of Pome-Based Distributions, 172
4.1 Discrete variable and discrete distributions, 172
  4.1.1 Constraint E[x] and the Maxwell-Boltzmann distribution, 172
  4.1.2 Two constraints and Bose-Einstein distribution, 174
  4.1.3 Two constraints and Fermi-Dirac distribution, 177
  4.1.4 Intermediate statistics distribution, 178
  4.1.5 Constraint: E[N]: Bernoulli distribution for a single trial, 179
  4.1.6 Binomial distribution for repeated trials, 180
  4.1.7 Geometric distribution: repeated trials, 181
  4.1.8 Negative binomial distribution: repeated trials, 183
  4.1.9 Constraint: E[N] = n: Poisson distribution, 183
4.2 Continuous variable and continuous distributions, 185
  4.2.1 Finite interval [a, b], no constraint, and rectangular distribution, 185
  4.2.2 Finite interval [a, b], one constraint and truncated exponential distribution, 186
  4.2.3 Finite interval [0, 1], two constraints E[ln x] and E[ln(1 − x)] and beta
distribution of first kind, 188
  4.2.4 Semi-infinite interval (0, ∞), one constraint E[x] and exponential distribution, 191
  4.2.5 Semi-infinite interval, two constraints E[x] and E[ln x]
and gamma distribution, 192
4.2.6 Semi-infinite interval, two constraints $E[\ln x]$ and $E[\ln(1 + x)]$ and beta distribution of second kind, 194
4.2.7 Infinite interval, two constraints $E[x]$ and $E[x^2]$ and normal distribution, 195
4.2.8 Semi-infinite interval, log-transformation $Y = \ln X$, two constraints $E[y]$ and $E[y^2]$ and log-normal distribution, 197
4.2.9 Infinite and semi-infinite intervals: constraints and distributions, 199

Questions, 203
References, 208
Additional Reading, 208

5 Multivariate Probability Distributions, 213
5.1 Multivariate normal distributions, 213
  5.1.1 One time lag serial dependence, 213
  5.1.2 Two-lag serial dependence, 221
  5.1.3 Multi-lag serial dependence, 229
  5.1.4 No serial dependence: bivariate case, 234
  5.1.5 Cross-correlation and serial dependence: bivariate case, 238
  5.1.6 Multivariate case: no serial dependence, 244
  5.1.7 Multi-lag serial dependence, 245
5.2 Multivariate exponential distributions, 245
  5.2.1 Bivariate exponential distribution, 245
  5.2.2 Trivariate exponential distribution, 254
  5.2.3 Extension to Weibull distribution, 257
5.3 Multivariate distributions using the entropy-copula method, 258
  5.3.1 Families of copula, 259
  5.3.2 Application, 260
5.4 Copula entropy, 265
Questions, 266
References, 267
Additional Reading, 268

6 Principle of Minimum Cross-Entropy, 270
6.1 Concept and formulation of POMCE, 270
6.2 Properties of POMCE, 271
6.3 POMCE formalism for discrete variables, 275
6.4 POMCE formulation for continuous variables, 279
6.5 Relation to POME, 280
6.6 Relation to mutual information, 281
6.7 Relation to variational distance, 281
6.8 Lin’s directed divergence measure, 282
6.9 Upper bounds for cross-entropy, 286
Questions, 287
References, 288
Additional Reading, 289

7 Derivation of POME-Based Distributions, 290
7.1 Discrete variable and mean $E[x]$ as a constraint, 290
  7.1.1 Uniform prior distribution, 291
  7.1.2 Arithmetic prior distribution, 293
7.1.3 Geometric prior distribution, 294
7.1.4 Binomial prior distribution, 295
7.1.5 General prior distribution, 297

7.2 Discrete variable taking on an infinite set of values, 298
7.2.1 Improper prior probability distribution, 298
7.2.2 A priori Poisson probability distribution, 301
7.2.3 A priori negative binomial distribution, 304

7.3 Continuous variable: general formulation, 305
7.3.1 Uniform prior and mean constraint, 307
7.3.2 Exponential prior and mean and mean log constraints, 308

Questions, 308
References, 309

8 Parameter Estimation, 310
8.1 Ordinary entropy-based parameter estimation method, 310
8.1.1 Specification of constraints, 311
8.1.2 Derivation of entropy-based distribution, 311
8.1.3 Construction of zeroth Lagrange multiplier, 311
8.1.4 Determination of Lagrange multipliers, 312
8.1.5 Determination of distribution parameters, 313

8.2 Parameter-space expansion method, 325
8.3 Contrast with method of maximum likelihood estimation (MLE), 329
8.4 Parameter estimation by numerical methods, 331

Questions, 332
References, 333
Additional Reading, 334

9 Spatial Entropy, 335
9.1 Organization of spatial data, 336
9.1.1 Distribution, density, and aggregation, 337
9.2 Spatial entropy statistics, 339
9.2.1 Redundancy, 343
9.2.2 Information gain, 345
9.2.3 Disutility entropy, 352
9.3 One dimensional aggregation, 353
9.4 Another approach to spatial representation, 360
9.5 Two-dimensional aggregation, 363
9.5.1 Probability density function and its resolution, 372
9.5.2 Relation between spatial entropy and spatial disutility, 375
9.6 Entropy maximization for modeling spatial phenomena, 376
9.7 Cluster analysis by entropy maximization, 380
9.8 Spatial visualization and mapping, 384
9.9 Scale and entropy, 386
9.10 Spatial probability distributions, 388
9.11 Scaling: rank size rule and Zipf’s law, 391
9.11.1 Exponential law, 391
9.11.2 Log-normal law, 391
9.11.3 Power law, 392
12.2.2 Minimum cross-entropy-based probability density functions given total expected spectral powers at each frequency, 498

12.2.3 Spectral probability density functions for white noise, 501

12.3 Minimum cross-entropy power spectrum given auto-correlation, 503

12.3.1 No prior power spectrum estimate is given, 504

12.3.2 A prior power spectrum estimate is given, 505

12.3.3 Given spectral powers: $T_k = G_j, G_j = P_k$, 506

12.4 Cross-entropy between input and output of linear filter, 509

12.4.1 Given input signal PDF, 509

12.4.2 Given prior power spectrum, 510

12.5 Comparison, 512

12.6 Towards efficient algorithms, 514

12.7 General method for minimum cross-entropy spectral estimation, 515

References, 515

Additional References, 516

13 Evaluation and Design of Sampling and Measurement Networks, 517

13.1 Design considerations, 517

13.2 Information-related approaches, 518

13.2.1 Information variance, 518

13.2.2 Transfer function variance, 520

13.2.3 Correlation, 521

13.3 Entropy measures, 521

13.3.1 Marginal entropy, joint entropy, conditional entropy and transinformation, 521

13.3.2 Informational correlation coefficient, 523

13.3.3 Isoinformation, 524

13.3.4 Information transfer function, 524

13.3.5 Information distance, 525

13.3.6 Information area, 525

13.3.7 Application to rainfall networks, 525

13.4 Directional information transfer index, 530

13.4.1 Kernel estimation, 531

13.4.2 Application to groundwater quality networks, 533

13.5 Total correlation, 537

13.6 Maximum information minimum redundancy (MIMR), 539

13.6.1 Optimization, 541

13.6.2 Selection procedure, 542

Questions, 553

References, 554

Additional Reading, 556

14 Selection of Variables and Models, 559

14.1 Methods for selection, 559

14.2 Kullback-Leibler (KL) distance, 560

14.3 Variable selection, 560

14.4 Transitivity, 561

14.5 Logit model, 561

14.6 Risk and vulnerability assessment, 574
14.6.1 Hazard assessment, 576
14.6.2 Vulnerability assessment, 577
14.6.3 Risk assessment and ranking, 578

Questions, 578
References, 579
Additional Reading, 580

15 Neural Networks, 581
15.1 Single neuron, 581
15.2 Neural network training, 585
15.3 Principle of maximum information preservation, 588
15.4 A single neuron corrupted by processing noise, 589
15.5 A single neuron corrupted by additive input noise, 592
15.6 Redundancy and diversity, 596
15.7 Decision trees and entropy nets, 598

Questions, 602
References, 603

16 System Complexity, 605
16.1 Ferdinand’s measure of complexity, 605
16.1.1 Specification of constraints, 606
16.1.2 Maximization of entropy, 606
16.1.3 Determination of Lagrange multipliers, 606
16.1.4 Partition function, 607
16.1.5 Analysis of complexity, 610
16.1.6 Maximum entropy, 614
16.1.7 Complexity as a function of N, 616
16.2 Kapur’s complexity analysis, 618
16.3 Cornacchio’s generalized complexity measures, 620
16.3.1 Special case: R = 1, 624
16.3.2 Analysis of complexity: non-unique K-transition points and conditional complexity, 624
16.4 Kapur’s simplification, 627
16.5 Kapur’s measure, 627
16.6 Hypothesis testing, 628
16.7 Other complexity measures, 628

Questions, 631
References, 631
Additional References, 632

Author Index, 633

Subject Index, 639