
Contents

Preface	<i>page</i> xiii
1 Compositional data: some challenging problems	1
1.1 Introduction	1
1.2 Geochemical compositions of rocks	2
1.3 Sediments at different depths	4
1.4 Ternary diagrams	5
1.5 Partial analyses and subcompositions	8
1.6 Supervisory behaviour	12
1.7 Household budget surveys	13
1.8 Steroid metabolite patterns in adults and children	14
1.9 Activity patterns of a statistician	15
1.10 Calibration of white-cell compositions	16
1.11 Fruit evaluation	17
1.12 Firework mixtures	17
1.13 Clam ecology	18
1.14 Bibliographic notes	19
Problems	20
2 The simplex as sample space	24
2.1 Choice of sample space	24
2.2 Compositions and simplexes	25
2.3 Spaces, vectors, matrices	28
2.4 Bases and compositions	30
2.5 Subcompositions	33
2.6 Amalgamations	36
2.7 Partitions	38
2.8 Perturbations	42

2.9	Geometrical representations of compositional data	43
2.10	Bibliographic notes	44
	Problems	45
3	The special difficulties of compositional data analysis	48
3.1	Introduction	48
3.2	High dimensionality	50
3.3	Absence of an interpretable covariance structure	52
3.4	Difficulty of parametric modelling	58
3.5	The mixture variation difficulty	61
3.6	Bibliographic notes	61
	Problems	62
4	Covariance structure	64
4.1	Fundamentals	64
4.2	Specification of the covariance structure	66
4.3	The compositional variation array	68
4.4	Recovery of the compositional variation array from the crude mean vector and covariance matrix	71
4.5	Subcompositional analysis	73
4.6	Matrix specifications of covariance structures	75
4.7	Some important elementary matrices	81
4.8	Relationships between the matrix specifications	82
4.9	Estimated matrices for hongite compositions	83
4.10	Logratios and logcontrasts	83
4.11	Covariance structure of a basis	86
4.12	Commentary	87
4.13	Bibliographic notes	88
	Problems	89
5	Properties of matrix covariance specifications	92
5.1	Logratio notation	92
5.2	Logcontrast variances and covariances	93
5.3	Permutations	93
5.4	Properties of \mathbf{P} and \mathbf{Q}_P matrices	95
5.5	Permutation invariants involving $\boldsymbol{\Sigma}$	95
5.6	Covariance matrix inverses	98
5.7	Subcompositions	100
5.8	Equivalence of characteristics of $\mathbf{T}, \boldsymbol{\Sigma}, \mathbf{I}$	101

5.9	Logratio-uncorrelated compositions	105
5.10	Isotropic covariance structures	107
5.11	Bibliographic notes	110
	Problems	110
6	Logistic normal distributions on the simplex	112
6.1	Introduction	112
6.2	The additive logistic normal class	113
6.3	Density function	115
6.4	Moment properties	116
6.5	Composition of a lognormal basis	117
6.6	Class-preserving properties	118
6.7	Conditional subcompositional properties	122
6.8	Perturbation properties	123
6.9	A central limit theorem	124
6.10	A characterization by logcontrasts	125
6.11	Relationships with the Dirichlet class	126
6.12	Potential for statistical analysis	129
6.13	The multiplicative logistic normal class	130
6.14	Partitioned logistic normal classes	132
6.15	Some notation	135
6.16	Bibliographic notes	136
	Problems	137
7	Logratio analysis of compositions	141
7.1	Introduction	141
7.2	Estimation of μ and Σ	143
7.3	Validation: tests of logistic normality	143
7.4	Hypothesis testing strategy and techniques	148
7.5	Testing hypotheses about μ and Σ	153
7.6	Logratio linear modelling	158
7.7	Testing logratio linear hypotheses	162
7.8	Further aspects of logratio linear modelling	166
7.9	An application of logratio linear modelling	170
7.10	Predictive distributions, atypicality indices and outliers	173
7.11	Statistical discrimination	176
7.12	Conditional compositional modelling	181
7.13	Bibliographic notes	181
	Problems	182

8	Dimension-reducing techniques	184
8.1	Introduction	184
8.2	Crude principal component analysis	186
8.3	Logcontrast principal component analysis	189
8.4	Applications of logcontrast principal component analysis	194
8.5	Subcompositional analysis	196
8.6	Applications of subcompositional analysis	198
8.7	Canonical component analysis	202
8.8	Bibliographic notes	206
	Problems	206
9	Bases and compositions	208
9.1	Fundamentals	208
9.2	Covariance relationships	209
9.3	Principal and canonical component comparisons	214
9.4	Distributional relationships	219
9.5	Compositional invariance	220
9.6	An application to household budget analysis	223
9.7	An application to clinical biochemistry	226
9.8	Reappraisal of an early shape and size analysis	227
9.9	Bibliographic notes	231
	Problems	232
10	Subcompositions and partitions	234
10.1	Introduction	234
10.2	Complete subcompositional independence	234
10.3	Partitions of order 1	241
10.4	Ordered sequences of partitions	247
10.5	Caveat	252
10.6	Partitions of higher order	253
10.7	Bibliographic notes	254
	Problems	254
11	Irregular compositional data	256
11.1	Introduction	256
11.2	Modelling imprecision in compositions	257
11.3	Analysis of sources of imprecision	261
11.4	Imprecision and tests of independence	264
11.5	Rounded or trace zeros	266

11.6	Essential zeros	271
11.7	Missing components	274
11.8	Bibliographic notes	277
	Problems	277
12	Compositions in a covariate role	281
12.1	Introduction	281
12.2	Calibration	282
12.3	A before-and-after treatment problem	283
12.4	Experiments with mixtures	285
12.5	An application to firework mixtures	291
12.6	Classification from compositions	293
12.7	An application to geological classification	297
12.8	Bibliographic notes	302
	Problems	302
13	Further distributions on the simplex	305
13.1	Some generalizations of the Dirichlet class	305
13.2	Some generalizations of the logistic normal classes	307
13.3	Recapitulation	309
13.4	The $\mathcal{A}^d(\alpha, \mathbf{B})$ class	310
13.5	Maximum likelihood estimation	313
13.6	Neutrality and partition independence	316
13.7	Subcompositional independence	318
13.8	A generalized lognormal gamma distribution with compositional invariance	319
13.9	Discussion	321
13.10	Bibliographic notes	323
	Problems	323
14	Miscellaneous problems	324
14.1	Introduction	324
14.2	Multi-way compositions	324
14.3	Multi-stage compositions	327
14.4	Multiple compositions	329
14.5	Kernel density estimation for compositional data	332
14.6	Compositional stochastic processes	335
14.7	Relation to Bayesian statistical analysis	337

14.8	Compositional and directional data	338
	Problems	341
Appendices		
A	Algebraic properties of elementary matrices	342
B	Bibliography	344
C	Computer software for compositional data analysis	353
D	Data sets	354
Author index		
		406
Subject index		
		409