

Contents

Preface	xiii
0 Introduction	1
0.1 What is mathematical modeling?	1
0.2 Purpose of this book	2
0.3 Organization of the book	4
I Describing single variables	7
1 Arithmetic and variables: The lifeblood of modeling	9
1.1 Blood circulation and mathematical modeling	11
1.2 Parameters and variables in models	15
1.2.1 discrete state variables: genetics	17
1.2.2 continuous state variables: concentration	18
1.3 First steps in R programming	19
1.3.1 numbers and arithmetic operations	20
1.3.2 variable assignment	23
1.4 Computational projects	26
2 Functions and their graphs	29
2.1 Dimensions of quantities	30
2.2 Functions and their graphs	33
2.2.1 linear and exponential functions	34
2.2.2 rational and logistic functions	37
2.3 Scripts, functions, and plotting in R	41

2.3.1	writing scripts and calling functions	41
2.3.2	vector variables	42
2.3.3	arithmetic with vector variables	45
2.3.4	plotting graphs	47
2.4	Rates of biochemical reactions	51
2.5	Computational projects	53
3	Describing data sets	57
3.1	Mutations and their rates	57
3.2	Describing data sets	60
3.2.1	central value of a data set	60
3.2.2	spread of a data set	63
3.2.3	graphical representation of data sets	65
3.3	Working with data in R	67
3.4	Computational projects	70
3.4.1	data description	70
3.4.2	plotting data and fitting by eye	71
4	Random variables and distributions	75
4.1	Probability distributions	76
4.1.1	axioms of probability	76
4.1.2	random variables	81
4.1.3	expectation (mean) of random variables	82
4.1.4	variance of random variables	84
4.2	Examples of distributions	86
4.2.1	uniform distribution	86
4.2.2	binomial distribution	87
4.3	Testing for mutants	93
4.4	Random numbers and iteration in R	94
4.4.1	random numbers	94
4.4.2	for loops	97
4.5	Computational projects	100
4.5.1	uniform distribution	100
4.5.2	binomial distribution	101

5	Estimation from a random sample	103
5.1	Law of Large Numbers	104
5.1.1	sample mean	104
5.1.2	sample size and standard error	106
5.2	Central Limit Theorem	107
5.2.1	normal distribution	107
5.2.2	confidence intervals	110
5.3	Relative risk	114
5.4	Sampling in R	116
5.4.1	simulated sampling	116
5.4.2	computing confidence intervals	118
5.5	Computational projects	121
II	Relationship between two variables	123
6	Independence of random variables	125
6.1	Categorical data sets with two variables	126
6.2	Mathematics of independence	127
6.2.1	conditional probability and information	127
6.2.2	independence of events	132
6.2.3	calculation of expected frequencies	133
6.3	Testing for independence	137
6.3.1	hypothesis testing	137
6.3.2	rejecting the null hypothesis	140
6.3.3	the chi-squared statistic	142
6.4	Hypothesis testing in R	145
6.5	Independence in data sets	147
6.5.1	maternal age and Down syndrome	147
6.5.2	stop-and-frisk and race	148
6.6	Computational projects	150
6.6.1	thumb-on-top preference and sex	150
6.6.2	relationship between species and habitat	151
6.6.3	independence testing of simulated data	152

7	Bayes' amazing formula	155
7.1	Prior knowledge	156
7.2	Bayes' formula	157
7.2.1	positive and negative predictive values	159
7.3	Applications of Bayesian thinking	162
7.3.1	when too much testing is bad	162
7.3.2	reliability of scientific studies	164
7.4	Random simulations	167
7.5	Computational project	172
8	Linear regression and correlation	177
8.1	Linear relationship between two variables	178
8.2	Linear least-squares fitting	178
8.2.1	sum of squared errors	178
8.2.2	best-fit slope and intercept	182
8.2.3	correlation and goodness of fit	184
8.3	Linear regression using R	188
8.4	Regression to the mean	191
8.5	Computational projects	193
8.5.1	parental age and new mutations	194
8.5.2	heart rates on two different days	195
9	Nonlinear data fitting	197
9.1	Nonlinear relationships between variables	197
9.2	Fitting using log transforms	199
9.3	Generalized linear fitting in R	201
9.3.1	logarithmic transforms	201
9.3.2	polynomial regression	202
9.4	Allometry and power law scaling	205
9.5	Computational projects	208
III	Chains of random variables	211
10	Markov models with discrete states	213
10.1	Building Markov models	214

10.2	Markov property	216
10.2.1	transition matrices	219
10.2.2	probability of a string of states	220
10.3	Simulation of Markov models	224
10.4	Markov models of medical treatment	226
10.5	Computational projects	229
10.5.1	state strings for a two-state model	229
10.5.2	state strings for a three-state model	230
11	Probability distributions of Markov chains	233
11.1	Distributions evolve over time	234
11.1.1	Markov chains	235
11.1.2	matrix multiplication	237
11.1.3	propagation of probability vectors	241
11.2	Matrix multiplication in R	244
11.3	Mutations in evolution	247
11.4	<i>Computational projects</i>	249
11.4.1	probability vectors of a two-state model	250
11.4.2	probability vectors of a three-state model	251
12	Stationary distributions of Markov chains	253
12.1	The origins of Markov chains: A feud and a poem	254
12.2	Stationary distributions	256
12.2.1	definition of stationary distribution	256
12.2.2	condition for unique stationary distribution	259
12.3	Multiple random simulations in R	264
12.4	Bioinformatics and Markov models	265
12.5	Computational projects	270
12.5.1	multiple two-state model simulations	270
12.5.2	multiple three-state model simulations	272
13	Dynamics of Markov models	275
13.1	Phylogenetic trees	276
13.2	Eigenvalues of Markov models	278
13.2.1	basic linear algebra terminology	278
13.2.2	calculation of eigenvalues on paper	281
13.2.3	calculation of eigenvectors on paper	283
13.2.4	rate of convergence	285

13.3	Matrix diagonalization in R	288
13.4	Molecular evolution	293
13.4.1	Jukes-Cantor model	293
13.4.2	time since divergence	296
13.4.3	calculation of phylogenetic distance	297
13.4.4	divergence of human and chimp genomes	298
13.5	Computational projects	301
13.5.1	eigenvalues of a two-state model	301
13.5.2	eigenvalues of a three-state model	302
13.5.3	analysis of the Jukes-Cantor model	303
IV	Variables that change with time	305
14	Linear difference equations	307
14.1	Discrete-time population models	308
14.1.1	static population	308
14.1.2	exponential population growth	309
14.1.3	example with birth and death	309
14.1.4	dimensions of birth and death rates	311
14.2	Solutions of linear difference models	312
14.2.1	simple linear models	312
14.2.2	models with a constant term	313
14.3	Population growth and decline	314
14.4	Numerical solutions in R	317
14.4.1	functions in R	317
14.4.2	solving difference equations	319
14.5	Computational project	322
15	Linear ordinary differential equations	325
15.1	From discrete time to smooth change	326
15.1.1	bacteria that divide at arbitrary times	326
15.1.2	growth proportional to population size	329
15.1.3	chemical kinetics	330
15.2	Solutions of ordinary differential equations	331
15.2.1	separate-and-integrate method	332
15.2.2	solution of inhomogeneous ODEs	335
15.2.3	Forward Euler method	338

15.3 Numerical solutions of ODEs	341
15.3.1 implementation of Forward Euler	341
15.3.2 error in Forward Euler solutions	343
15.4 Applications of linear ODE models	347
15.4.1 model of pharmacokinetics	347
15.4.2 Cole's membrane potential model	351
15.5 Computational projects	355
15.5.1 error and time step	355
15.5.2 pharmacokinetics model	356
16 Graphical analysis of ordinary differential equations	359
16.1 ODEs with nonlinear terms	360
16.2 Qualitative analysis of ODEs	361
16.2.1 graphical analysis of the defining function	362
16.2.2 fixed points and stability	364
16.3 Modeling infectious disease spread	372
16.4 Computational projects	377
16.4.1 logistic population growth model	378
16.4.2 SIS epidemic model	379
17 Chaos and bifurcations in difference equations	383
17.1 Logistic model in discrete time	384
17.2 Qualitative analysis of difference	385
17.2.1 fixed points or equilibria	385
17.2.2 stability of fixed points	386
17.3 Graphical analysis	389
17.3.1 graphical analysis using R	389
17.3.2 cobweb plots	390
17.4 Discrete-time logistic model and chaos	395
17.5 Computational projects	402
17.5.1 graphical stability analysis	402
17.5.2 investigation of chaotic dynamics	403
Bibliography	405
Index	411