

Contents

Part I The Basics of R

1	Calculating with R	3
1.1	Installing R	3
1.1.1	Demos	3
1.2	Finding help with R	4
1.3	Some interface aids	4
1.4	Arithmetic	5
1.5	Complex numbers	5
1.6	Assigning variables	6
1.6.1	Example: Conversions between units	6
1.7	Standard mathematical functions	7
1.7.1	Rounding	8
1.8	Vectors	8
1.8.1	Operations on vectors	9
1.8.2	Functions that operate on vectors	10
1.8.3	Character vectors	11
1.9	Generating sequences	11
1.9.1	Generating regular sequences	11
1.9.2	Generating sequences of random numbers	12
1.10	Logical vectors	13
1.11	Matrices	14
1.11.1	Arithmetic operations on matrices	15
1.11.2	Matrix multiplication	16
1.11.3	Determinant of a matrix	16
1.11.4	Transpose of a matrix	17
1.11.5	Diagonal matrix	17
1.11.6	Matrix inverse	17
1.11.7	Eigenvalues and eigenvectors	18
1.11.8	Other matrix functions	19
1.12	Other data structures	19

1.12.1	Data frames	19
1.12.2	Factors	19
1.12.3	Lists	20
1.13	Problems	20
2	Plotting with R	23
2.1	Some common plots	23
2.1.1	Data plot	23
2.1.2	Bar plot	25
2.1.3	Function plot	26
2.1.4	Histogram	26
2.1.5	Three-dimensional plot	27
2.2	Customizing plots	28
2.2.1	Different plot characters	28
2.2.2	Plotting data with a line	29
2.2.3	Adding title and axis labels	31
2.2.4	Adding colors	31
2.2.5	Adding straight lines to a plot	32
2.2.6	Adjusting the axes	34
2.2.7	Customizing ticks and axes	34
2.2.8	Setting default graph parameters	35
2.2.9	Adding text to a plot	35
2.2.10	Adding math expressions and arrows	36
2.2.11	Constructing a diagram	37
2.3	Superimposing data series in a plot	39
2.4	Placing two or more plots in a figure	42
2.5	Error bars	44
2.6	Locating and identifying points on a plot	45
2.7	Problems	46
3	Functions and Programming	49
3.1	Built-in functions in R	49
3.1.1	Sorting	50
3.1.2	Splines	51
3.1.3	Sampling	52
3.2	User-defined functions	52
3.2.1	Gaussian function	52
3.2.2	pH titration curves	54
3.3	Programming	55
3.3.1	Conditional execution with <code>if()</code>	56
3.3.2	Looping with <code>for()</code>	56
3.3.3	Looping with <code>while()</code>	57
3.3.4	Looping with <code>repeat</code>	58
3.3.5	Choosing with <code>which()</code>	58
3.4	Numerical analysis with R	58

3.4.1	Finding a zero of a function	59
3.4.2	Finding the roots of a polynomial	60
3.4.3	Solving a system of simultaneous linear equations	61
3.4.4	Solving a system of nonlinear equations	62
3.4.5	Numerical integration of functions	63
3.4.6	Numerical integration of data using <code>splinefun</code>	64
3.4.7	Numerical differentiation	64
3.5	Problems	65
4	Data and Packages	69
4.1	Writing and reading data to files	69
4.1.1	Changing directories	69
4.1.2	Writing data to a file	70
4.1.3	Reading data from a file using <code>scan()</code>	70
4.1.4	Writing and reading tables	71
4.1.5	Reading and writing spreadsheet files	71
4.1.6	Saving the R environment between sessions	71
4.2	Packages	72
4.3	Data frames	73
4.4	Factors	75
4.5	The contributed package <code>seqinr</code>	76
4.6	Problems	81

Part II Simulation of Biological Processes

5	Equilibrium and Steady State Calculations	85
5.1	Equilibrium in reacting mixture	85
5.1.1	Binding of a ligand L to a protein or polymer P	85
5.1.2	Fitting binding data with a Scatchard plot using <code>linear</code> <code>model (lm)</code>	87
5.1.3	Strong and weak binding	89
5.1.4	Cooperative binding	90
5.1.5	Oxygen binding by hemoglobin	92
5.1.6	Hill plot	94
5.1.7	Experimental determination of equilibrium constants	96
5.1.8	Temperature dependence of equilibrium constants: the van't Hoff equation	97
5.2	Single strand–double helix equilibrium in oligonucleotides	98
5.3	Steady-state enzyme kinetics	103
5.3.1	Michaelis-Menten kinetics	103
5.3.2	Lineweaver-Burk fitting	104
5.3.3	Eadie-Hofstee fitting	106
5.3.4	Competitive inhibition	108
5.4	Non-linear least-squares fitting	109
5.5	Problems	110

6	Differential Equations and Reaction Kinetics	113
6.1	Analytically solvable models	113
6.1.1	Exponential growth	113
6.1.2	Exponential decay	114
6.1.3	Chemical Conversion	115
6.1.4	Exponential decay to a constant value	115
6.1.5	Model of limited growth	117
6.1.6	Kinetics of bimolecular reactions	117
6.2	Numerical integration of ODEs	119
6.2.1	Integrating a single ODE by the Euler method	119
6.2.2	Integrating a system of ODEs by the Euler method	121
6.2.3	Integrating a system of ODEs by the improved Euler method	122
6.2.4	Integrating a system of equations using 4th-order Runge-Kutta	124
6.2.5	lsoda, a sophisticated differential equation solver	128
6.2.6	Stability of systems of ODEs	130
6.2.7	Numerical solution of second-order ODEs	133
6.3	Stochastic differential equations	133
6.4	Problems	135
7	Population Dynamics	141
7.1	Models of homogeneous populations of organisms	141
7.1.1	Verhulst-Pearl (logistic) equation	141
7.1.2	Variable carrying capacity and the logistic	143
7.1.3	Lotka-Volterra model of predation	145
7.1.4	Modifications of the Lotka-Volterra model	145
7.1.5	Volterra's model for two-species competition	146
7.2	Models of microbial growth	146
7.2.1	Monod model of microbial growth	147
7.2.2	Microbial growth in batch culture and chemostat	147
7.2.3	Multiple limiting nutrients	148
7.2.4	Competition for limiting nutrients	148
7.2.5	Toxic inhibition of microbial growth	149
7.3	Models of Epidemics	149
7.3.1	The simple SIR model	150
7.3.2	The SIR model with births and deaths	152
7.3.3	SI model of fatal infections	153
7.3.4	SIS model of infection without immunity	153
7.3.5	A model for an epidemic of gonorrhea	154
7.4	Problems	155
8	Diffusion and Transport	159
8.1	Transport by simple diffusion	159
8.1.1	Fick's laws of diffusion	159
8.1.2	Analytical solutions of Fick's second law in one dimension	160

8.1.3	Numerical solutions of Fick's second law in one dimension .	163
8.1.4	Diffusion in spherical and cylindrical geometries	165
8.2	Diffusion in a driving field: electrophoresis	166
8.3	Countercurrent diffusion	166
8.4	Diffusion as a random process: Brownian motion	167
8.5	Compartmental models in physiology and pharmacokinetics	169
8.5.1	Periodic dose administration to a single compartment	170
8.5.2	Liver function—A two-compartment model	170
8.5.3	Multi-compartment model of liver function	171
8.5.4	Oscillations in calcium metabolism	172
8.6	Problems	172
9	Regulation and Control of Metabolism	175
9.1	Successive enzyme reactions	175
9.1.1	One-substrate, one-product reaction	176
9.1.2	Successive one-substrate, one-product reactions	178
9.1.3	Steady-state flux calculation	180
9.2	Metabolic control analysis	181
9.2.1	Flux control coefficients	182
9.2.2	Elasticities	183
9.3	Biochemical systems theory	184
9.3.1	Linear pathway with feedback inhibition	185
9.3.2	Transitions in reaction network behavior	186
9.4	Problems	188
10	Models of Regulation	191
10.1	Regulation of transcription: Feed-forward loops	191
10.2	Regulation of signaling: Bacterial chemotaxis	193
10.2.1	Modeling of chemotaxis as a biased random walk	194
10.2.2	Robust model of chemotaxis	195
10.3	Regulation of development: Morphogenesis	197
10.3.1	Exponential morphogen gradients are <i>not</i> robust	198
10.3.2	Self-enhanced morphogen degradation to form robust gradients	199
10.3.3	Patterning in the dorsal region of <i>Drosophila</i>	200
10.4	Problems	202

Part III Analyzing DNA and Protein Sequences

11	Probability and Population Genetics	207
11.1	Some fundamentals of probability	207
11.1.1	Review of basic probability ideas	207
11.1.2	Conditional probability	208
11.1.3	The law of total probability	208
11.1.4	Bayes' theorem	209
11.2	Stochastic population models	210

11.2.1	Stochastic modeling of population growth	211
11.2.2	Stochastic simulation of radioactive decay	213
11.3	Markov chains	214
11.3.1	Markov chain model of diffusion out of and into a cell	214
11.4	Probability distribution functions	216
11.4.1	R's four basic functions for a statistical distribution	216
11.4.2	Uniform distribution	217
11.4.3	Normal distribution	219
11.4.4	Binomial distribution	222
11.4.5	Poisson distribution	224
11.4.6	Geometric distribution	226
11.4.7	Exponential distribution	227
11.4.8	Other distributions	228
11.5	Population Genetics	228
11.5.1	Hardy-Weinberg Principle	228
11.5.2	Effect of selection	229
11.5.3	Selection for heterozygotes	229
11.5.4	Mutation	230
11.5.5	Selection involving sex-linked recessive genes	230
11.6	Problems	231
12	DNA Sequence Analysis	233
12.1	Getting a sequence from the Web	233
12.1.1	Using Entrez	233
12.1.2	Making the sequence usable by R	234
12.1.3	Converting letters to numbers	234
12.2	Single base sequences and frequencies	237
12.2.1	Counting the number of each type of base	237
12.2.2	Simulating a random sequence of bases	237
12.2.3	Simulating the distribution of the number of A residues in a sequence	238
12.3	Dinucleotide sequences and frequencies	239
12.3.1	Probability of observing dinucleotides	239
12.3.2	Markov chain simulation of dinucleotide frequencies	240
12.4	Simulation of restriction sites	244
12.5	Detecting periodicity in a sequence	246
12.6	Problems	248
Part IV Statistical Analysis in Molecular and Cellular Biology		
13	Statistical Analysis of Data	251
13.1	Summary statistics for a single group of data	251
13.1.1	Summary statistics	251
13.1.2	Histograms	252
13.1.3	Cumulative distribution	253

13.1.4	Test of normal distribution using <code>qqnorm</code>	254
13.1.5	Boxplots	255
13.1.6	Summary statistics for grouped data using <code>tapply()</code>	257
13.2	Statistical comparison of two samples	258
13.2.1	One-sample <i>t</i> test	260
13.2.2	Two-sample <i>t</i> test	262
13.2.3	Two-sample Wilcoxon test	264
13.2.4	Paired <i>t</i> test	264
13.2.5	Statistical power calculations	265
13.3	Analysis of spectral data	267
13.3.1	Fitting to a sum of decaying exponentials	267
13.3.2	Fitting a superposition of Gaussian spectra	269
13.3.3	Processing mass spectrometry data	272
13.4	Problems	276
14	Microarrays	279
14.1	Introduction	279
14.2	Preprocessing overview	280
14.3	The <code>affy</code> package	281
14.4	Importing data not in standard form	284
14.5	The need for preprocessing	284
14.6	Preprocessing steps	286
14.6.1	The Dilution dataset	286
14.6.2	Preliminary inspection	286
14.6.3	MAplot	287
14.6.4	Background correction	288
14.6.5	Normalization	289
14.6.6	PM correction	290
14.6.7	Summarization	290
14.6.8	Combining preprocessing steps with <code>expresso</code>	291
14.7	Using the results of preprocessing	291
14.7.1	<code>ExpressionSet</code>	291
14.7.2	Identifying highly expressed genes	292
14.8	Statistical analysis of differential gene expression	293
14.8.1	Paired samples	293
14.8.2	Unpaired samples	295
14.8.3	Bootstrapping	296
14.8.4	Analysis of variance (anova) for more than two groups	296
14.9	Detecting groups of genes	297
14.9.1	Correlation	298
14.9.2	Distance measures	298
14.9.3	Clustering	300
14.9.4	Principal component analysis	302
14.10	Power analysis	303
14.11	Problems	307

A Basic String Manipulations in R 309

References 313

Index 317



<http://www.springer.com/978-1-4419-0084-5>

Computer Simulation and Data Analysis in Molecular
Biology and Biophysics

An Introduction Using R

Bloomfield, V.

2009, XVI, 321 p., Hardcover

ISBN: 978-1-4419-0084-5