

Index

- AlgLab website 12
- algorithm analysis 2–4, 35, 240
- algorithm design 2, 4–5, 8
- algorithm design hierarchy 7–9, 98
- algorithm engineering 4–5, 7–8
- algorithm paradigms 6, 117
 - brute force 107
 - divide-and-conquer 118
 - dynamic programming 117, 121
 - exhaustive enumeration 117, 131
 - exhaustive search 5, 100–102, 119
 - greedy 121
 - heuristic search 200
 - iterative 121–123
 - recursion-heavy 117–121
 - scanning and sweeping 121
- algorithm tuning 7–9, 98–100, 116–117
 - bin packing case study 100–106
 - cache efficiency 140–141
 - concurrency 143–144
 - controlling subproblem sizes 119–120
 - customizing data structures 123–125
 - essential subgraph case study 106–116
 - filtering 112–113, 119
 - finessing calculations 108–109
 - hybridizing the recursion 120–121
 - input presentation 123–124
 - I/O (external memory) 141–143
 - iterative paradigms 121–122
 - loop abort 110–112
 - memoization 108, 110
 - memory 135–137
 - preprocessing 105–106
 - propagation 104–105
 - pruning 119
 - recursion-heavy paradigms 117–121
 - shrinking cost per stage 120–121
 - See also* tuning process
- algorithmics 2–5
- alias method 172–173
- analysis-friendly data 181–212
 - experimental designs for 183–184
 - See also* simulation shortcuts, variance reduction techniques
- analyzing trends and functions *see* categories of experiments
- antithetic variates. *See* variance reduction techniques
- approximation algorithms 84, 91
- artifacts 23–24
- assessment experiment. *See* categories of experiments
- biased estimator 86–87, 232–233
- bimodal distribution 223
- bin packing, optimal algorithm 100–101
- bin packing problem 85–86, 100–101, 246
- binary search algorithm 140–141
- binary search tree (BST) 4, 175, 177
- bivariate data 236–251
- bootstrapping 234–236
- Box-Cox procedure 244
- boxplots. *See* graphical data analysis
- breakpoint phylogeny algorithm 5, 9
- cache coherence problem 81, 144
- cache-efficient binary search 140–141
- Cachegrind 71–72
- caches
 - hand-built 125, 210
 - structure of 77–80
 - and time measurement 77–78

Cambridge University Press

978-1-107-00173-2 - A Guide to Experimental Algorithmics

Catherine C. McGeoch

Index

[More information](#)

258

Index

- caches (*cont.*)
 - warm vs cold 80, 86
 - See also* cache efficiency under algorithm tuning, code tuning
- categories of data 217–218
- categories of experiments 31
 - analyzing trends and functions 35–37
 - assessment 31–34
 - comparison (horse race) 37–43
- ceiling effects 22, 226–227
- censored data 34, 226–227
- code correctness 20
 - validation 155–156
 - verification 156–158
- code profilers 69–72
- code tuning 7–9, 98–99, 125–126
 - and compiler optimizers 126, 133
 - concurrency 143–144
 - I/O (external memory) 141–143
 - loops 126–130
 - memory 135–137
 - objects and memory structures 134–135
 - procedures 130–134
 - See also* tuning process
- common random numbers. *See* variance reduction techniques
- computer intensive inferential methods 234–236
- concurrency 8–9, 64
 - time measurement 80–82
 - tuning strategies 143–144
- conditional expectation. *See* variance reduction techniques
- confidence intervals 233–235
- control variates. *See* variance reduction techniques
- correlation 236
 - coefficient 237–238
 - graphical analysis of 237–240
 - in overloaded trials 207
 - and variance reduction techniques 190, 193, 194, 202
- covariance 192, 236–238
- CPU times
 - and concurrency 80
 - measurement of 63–65
 - measurement tools 69–73, 82–83
 - and memory access 76–79
 - as performance indicators 51–52, 66–69, 72–74
- cycle counts 65–66, 69
- data analysis 215–217
- data analysis software 254
- descriptive models 240–246
- descriptive statistics 216, 218–221, 229–231
- design of experiments (DOE). *See* experimental design
- design point 26–27
- Dijkstra's algorithm 107–111
- DIMACS Implementation Challenges 153
 - graph coloring 22, 29 40, 43
 - network flows 14
 - shortest paths 5
 - traveling salesman problem 51, 79
- dispersion (statistics) 218–221
- distributions, analysis of 221–228, 229–231
- documentation of experiments 160–161
- doubling experiment 33–35, 57–58
- efficiency
 - of algorithms. *See* algorithm design hierarchy, algorithm tuning
 - of experiments 24
 - of programs. *See* algorithm design hierarchy, code tuning
 - of test programs 152, 158–159
 - See also* simulation shortcuts, variance reduction techniques
- elapsed time. *See* cycle counts, wall clock time
- empirical algorithm analysis 3
- empirical distribution plots. *See* graphical data analysis
- empirical rule 225–226
- environment parameters 26
- essential subgraph (ES) algorithm 106–116
- essential subgraph (ES) problem 106–107
- estimators 86, 188
 - biased 86–87, 93, 232–233
 - of distribution parameters 232–233
 - indirect 196
 - in VRT's 188, 193–197
- exact algorithm 84–85, 101–102, 109
- excess kurtosis statistic 224
- exhaustive enumeration 117–118
- exhaustive search 100–102, 117–118
- expectation of a random variate 85, 187–188, 192
- experimental designs 10–11, 20, 25–45
 - for analysis friendly data 183–184
 - factor reduction 40–45
 - factorial designs 37–40
 - terminology 25–27, 45–46

- experimental process 10–11
- exploratory data analysis 217–218, 244–245
- exploratory experiment. *See* pilot experiment
- exponential distribution 170, 224
- external memory algorithms 141–143
- factors 26
 - choosing 31–33
 - reduction of 43–44
- fat-tailed distribution 222, 225
- first fit algorithm 85–86, 87–90, 246–248
- first fit decreasing algorithm 105
- full factorial designs 37–40
- functions, analysis of. *See* trends and functions
- geometric mean 229
- goals of experimentation 20–25
 - generality 20, 24–25
 - newsworthiness 20, 29
 - reproducibility 20, 22,
 - validity 20–21, 22–24, 45*See also* code correctness, efficiency
- gprof 70–71, 73, 74, 108
- graph classes 27–29, 173–175
- graph coloring problem 17–18
 - See also* DIMACS Implementation Challenges
- graphical data analysis 217, 239–240
 - boxplots 227–228, 231
 - coding 239, 252
 - coplots 252
 - empirical distribution plots 223–227, 228, 231, 235
 - jittering 221–223
 - lowess 243–247, 250–251
 - multidimensional data 252–253
 - scatterplots 238–239
 - segment pairplots 238–239
 - subscales 252
 - trend analysis 250–251
- Greedy graph coloring algorithm 18–19, 27–29
- greedy (paradigm) 117, 121
- Harmonic Number 171, 186, 211
- heuristic algorithms 83–84
- hinges 227–228, 231
- horse race. *See* categories of experiments
- hotlink 125
- hybrid input. *See* input classes
- hybridizing the recursion. *See* algorithm tuning
- hyperthreading 65
- importance sampling. *See* variance reduction techniques
- independence
 - random variate generation 162–163
 - simulation shortcuts 207
 - variance reduction techniques 190–193*See also* nonindependent random variates
- inferential methods 232–234
- input classes
 - generated, algorithm centered 29
 - generated, reality centered 29
 - hybrid 29
 - known optimal solutions 92–93
 - random 28. *See also* random generation
 - real (typical) 29, 98
 - selection of 10, 20–30
 - stress test 27
 - structured random 29
 - testbed 29, 30–31
 - worst case, bad case 27
- input presentation 123–124
- instance parameters 26
- instantiation scale 6–7, 11, 72–75
- interaction effects 38, 44, 113, 128
- interquartile range (IQR) 220, 222–223, 227
- interrupt 64–65, 66, 68, 79
- interval timing. *See* recording under CPU times
- inverse ratio data. *See* ratio test
- iterated greedy (IG) graph coloring algorithm 40
- jittering. *See* graphical data analysis
- joint probability distribution 236, 242
- kurtosis 223–225
- least recently used (LRU) 188, 208, 210
- least squares regression fit 242–247
- levels, choice of 26–27
 - factor reduction 43–44
 - factorial designs 38–40
 - good views 183–184
 - overloading 44, 205–206
 - trend and function analysis 35–37, 249
- linear congruential random number generators 161–162
- linearity of expectation 89–91, 94
- locality
 - and cache efficiency 141
 - in data structure accesses 123, 125
 - generating random variates with 175–177
 - and I/O efficiency 142–143

- locality (*cont.*)
 - in memory references 135–140
 - in procedure scope 134
 - spatial 136, 140
 - temporal 136, 137
 - whack-a-mole principle 140
- location (statistics) 218–220
- logarithmic transformation 228–229, 231, 244–245
- lookup method 171–172, 202
- loop abort 110–112, 116, 122, 124
- loop invariant 156
- loops, tuning. *See* code tuning
- lowess. *See* graphical data analysis
- main effects 38–39
- Markov Chain Monte Carlo (MCMC) algorithm 52–54, 241
- mean
 - distribution (μ) 86, 170, 188
 - geometric 229
 - sample 86, 187–189, 219
 - trimmed 222
 - See also* location
- median
 - estimation of 234–235
 - sample 219–223, 227,
- median selection in quicksort 121–121
- memoization 108–109, 112–114
- memory
 - footprint 78, 138
 - hierarchy 76–80, 135–137
 - I/O (external) 141–143
 - locality 78, 136–137
 - measurement tools 71, 82–83, 134, 138
 - See also* caches, virtual memory
- mems 80
- model validation 240, 243
- modeling tricks (simulation shortcuts) 209–211
- move-to-front algorithm 185
- multi-core processors 64, 80, 143–144
- multivariate data, analysis of 252–253
- mutable objects 135
- narrow vs wide apertures. *See* performance indicators
- negative binomial distribution 170
- nonindependent random variates 175–177
- nonuniform distributions 169–172
- normal distribution 170–171, 224–225
- optimal bin packing algorithm 100–101
- optimal solutions 83–84
 - bounds on 86–88
 - known by construction 92–93
- optimization (compiler). *See* compiler optimizers
- optimization problem 83–84, 118
- order statistics 219
- ordinal data 217
- outliers 221–227, 238–239, 243–245
- overloading 204–209
- paradigms. *See* algorithm paradigms
- parallel speedup 81–82
- parameter estimation 232–233
- parameters (experimental) 25–26
 - algorithm 26, 32, 42–43
 - choosing 30–32
 - environment 26, 32, 43
 - fixed 27
 - instance 26, 28–30
 - noise 27, 44, 192
 - and test programs 158–160
- parameters (of distributions) 232–234
- performance indicators 26
 - choosing 50–52, 72–82, 183–184
 - clocks and timers 63–69
 - code counters 54–63
 - narrow vs wide apertures 74–75
 - profilers 69–72
 - solution quality 83–94
 - timing tools 82–83
- performance metrics 26, 50
- pilot experiment 21–22
- poststratification. *See* variance reduction techniques
- power transformation 229, 245–249
- principle of locality 136
- probability distribution 169
- process (CPU) 6, 63–64
- profilers. *See* performance indicators
- pseudorandom number generators. *See* random number generators
- quicksort 118
 - in algorithm tuning 118–121
 - in MCMC case study 55–59
 - in simulation speedups 210–211
- random generation (of)
 - graphs 173–175

- nonindependent variates 175–177
- nonuniform variates 169–173
- permutations 164–165
- samples 165–169
- uniform variates 161–164
- Random graph coloring algorithm 19
- random inputs. *See* input classes
- random number generators 161–164
- random variate 85–86, 161–162
- ranked data 219
- ratio test 249–251
- real inputs. *See* input classes
- real time. *See* wall clock time
- register spill 134
- replication (of experiments) 24, 155–156
- reporting experimental results 75–76
- resampling 234–236
- residuals 242–247, 253
- resources for experimenters 12–14, 153–154
- response of a performance indicator 182
- roadmap graphs 5, 122–123
- robust statistic 222
- sampling distribution of the mean 233
- scatterplots. *See* graphical data analysis
- segment pairplots. *See* graphical data analysis
- self-organizing sequential search problem 185–187
- simplified iterated greedy (SIG) algorithm 41–45
- simulation shortcuts 204
 - modeling tricks 209–211
 - overloading 204–208
- skewed data sample 221
- solution quality. *See* performance indicators
- spatial locality 136
- spurious result 22–24
- standard deviation
 - confidence intervals for 233–234
 - of a distribution 216, 232
 - empirical rule 225
 - estimation 232–233
 - statistic 220
- startup problem 34
- steady state 34
- stopping rule 34
- stratification. *See* variance reduction techniques
- summary statistics 216, 218–221, 229–231
- system clock, time-of-day clock. *See*
 - performance indicators, clocks and timers
- system load 65
- temporal locality 136
- test apparatus 154
- test environment 152
- test program 6–7, 154–155
 - correctness 155–158
 - documentation 160–161
 - interface 158–160
- text generation problem 52
- timer latency 66
- timers. *See* performance indicators
- timestamp resolution 64
- transformation of data
 - asymptotic analysis 246–251
 - model fitting 244–246
 - skewed data 228–229
- transpose algorithm 185
- trends and functions, analysis of
 - asymptotic 246–251
 - descriptive models 240–246
 - experimental designs for 35–37
- trimmed mean 222, 227
- Tukey's ladder of transformations 244
- tuning. *See* algorithm tuning, code tuning
- tuning process 145–146
- typical inputs. *See* input classes
- uniform distribution 45, 224
- uniform random variate. *See* random generation
- univariate data analysis 218–236
- validating code. *See* code correctness
- validating a model. *See* model validation
- variable (experimental) 45
- variance of a distribution 233
- variance of a sample 219
- variance reduction techniques 184–185
 - antithetic variates 201–202
 - common random numbers 190–193
 - conditional expectation 196–201
 - control variates 193–196
 - importance sampling 203
 - poststratification 202–203
 - stratification 202–203
- virtual address 77
- virtual memory 137
- wall clock time 65
- warm start 80
- whack-a-mole principle 140
- workhorse experiment 21–22
- Zipf's distribution 171