

Subject Index

- accept-reject method, 133
accuracy, 18, 19, 91
 data input, 10
 data issues, 186
 defined, 15
 numeric, *see* numerical accuracy
 sources, 18–19
algorithm, 15–17, 28
 approximation, 29, 31–32
 truncation error, 31
 bias, 3
 defined, 15, 19
 efficiency, 16
 example, bubble sort, 16
 example, find max, 15–16
 example, scanning, 16
 heuristic, *see* heuristic
 implementation dependence, 46
 instance definition, 15
 limitations, 29, 42
 option dependence, 46
 problem definition, 15
 scaling, 98
 wrapper algorithm, 17
aliasing, 173
 constraints, 173
 extrinsic, 173
 intrinsic, 173
alternating conditional sampling, *see*
 Gibbs sampler
American National Election Studies,
 290
American Political Science Review, 189

Aptech Systems, Inc., 67, 268
artificial neural networks (ANN), *see*
 optimization algorithm, artificial
 neural networks (ANN)
auxiliary variables, 136

Bayesian GLS regression, 232
Bayesian Inference, *see* inference,
 Bayesian
Bayesian model averaging, 116
benchmarks, 63
 data input, *see* data input and output,
 benchmarks
 defined, 45
 distribution functions, *see*
 distribution functions, benchmarks
 pseudo-random number generators,
 see pseudo-random number
 generator, benchmarks
Box-Müller method, 270
BUGS, 6

C, 29, 63, 92, 142, 265
C++, 63, 93, 131, 142, 265
central limit theorem, 134, 141
Cephes, 59, 92
Cholesky decomposition, **155**, 156, 158
 algorithm, 156
 Gill–Murray algorithm, 156–157
 Schnabel–Eskow algorithm, 158–159
Cholesky factorization, *see* Cholesky
 decomposition

Boldface page number indicates that the subject is the primary topic of discussion.

Numerical Issues in Statistical Computing for the Social Scientist, by Micah Altman, Jeff Gill,
and Michael P. McDonald
ISBN 0-471-23633-0 Copyright © 2004 John Wiley & Sons, Inc.

- combinatorial optimization, 31
- computational efficiency, 250
- computation error examples, 1–2
- computer, 3
- conditioning, 47–48
- condition number, 47
- confidence interval, 77, 86, 104
 - bootstrapped, 106
 - credible interval, *see* credible interval
 - likelihood ratio, 211, 247, **249**, 251
 - likelihood ratio inference, 104
 - likelihood ratio test (LRT), 105
 - Wald inference, 104–105, **212**, **213**
 - Wald vs. likelihood inference, **215**, 216
- conjugate gradient, *see* optimization algorithm, conjugate gradient
- constrained maximum likelihood, *see* maximum likelihood estimation (MLE), constrained
- Cook's curvature, *see* likelihood curvature
- correlation matrix, standardized, 154
- countable collection, 120
- coupling from the past, 264
- coverage, **216**
- credible interval, 110
- cross-level-inference, *see* ecological inference

- Data Documentation Initiative (DDI), 256
- data input and output, 256
 - benchmarks, 60–63
 - canonical data, 255
 - cyclic redundancy checks, 62–63
 - truncation, 256
- Davidon–Fletcher–Powell method (DFP), *see* optimization algorithm, Davidson–Fletcher–Powell (DFP)
- DDI, *see* Data Documentation Initiative
- derivatives
 - analytic vs. numerical, **201**
 - automatic, 201
 - forward difference method, 32
 - forward vs. central differences, **205**
 - numerical, 205
- DFP, *see* optimization algorithm, Davidson–Fletcher–Powell (DFP)
- distribution functions
 - benchmarks, 58–60, 65
 - cumulative bivariate normal, 180, 181, 194
 - truncated normal, 146
- DSSTAB, 59, 92

- ecological inference, 11
 - choice of optimization algorithm, **187**
 - choice of PRNGS, **187–188**
 - Goodman's regression, **178**
 - method of bounds, **178**, **179**
 - sensitivity to environment, **195–197**
 - sensitivity to implementation, **188–189**, **194–195**
 - sensitivity to perturbation, **185–186**, **191–194**
 - two-stage models, **190**
- ecological inference (EI), 176
 - background, **177–179**
 - definition, **179–180**
 - numerical issues, **180–181**
- EI-R, 196–197
- ELV, 59
- EM algorithm, 114, 142
- empirical cdf, 113
- ergodic theorem, 124
- E-Views, 199
- exact inference, logistic regression, **249**
- Excel, 12–14, 50, 52, 74, 207, 209
 - solver, 206–209, **209**
- EzI, 182, 188, 189, 191, 195, 196
- execution bugs, 189
- simulation variance, 195
- versions, 196

- floating point arithmetic, 23–24, 28
 - approximation, 94
 - bits, 23
 - cancellation, 25
 - compilers, 29
 - example, 25–27

- guard bits, 24
- inaccuracies, 42
- in EI, 180
- intermediate calculation, 28
- machine epsilon, 24
- mantissa, 24
- misconceptions, 28
- normalized number, 24
- overflow, 23
- precision, 24–25
- reproducibility, 28–29
- rounding, 24, 93, 94
- rounding error, 17
- truncation, 24
- underflow, 23, 25
- FORTRAN, xv, 50, 51, 92, 265
- function maximization, *see*
 - optimization algorithm

- Gauss, xv, 7, 9, 66–68, 74, 153, 155, 156, 176, 181, 182, 187, 188, 199, 213, 258
- Gauss–Markov, 79–80
- Gauss–Newton method, *see*
 - optimization algorithm, Gauss–Newton method
- generalized Cholesky, 147, 151, 155–159
 - numerical examples, 158–159
- generalized inverse, 147, **152**, 151–155
 - minimum norm, 153
 - Moore–Penrose, 151
 - normalized, 153
 - numerical examples, 154–155
 - reflexive, 153
 - weak, 153
- generalized linear model, 82
 - interactions, 82
 - link function, 82
- genetic algorithm (GA), *see*
 - optimization algorithm, genetic algorithm
- geographic information systems (GIS), 219, 266
- Gerschgorin bounds, 158
- Gerschgorin circle theorem, 158
- Gibbs sampler, 6, 119, 128, 139, 232
 - full conditional distributions, 129

- GIS, *see* geographic information systems (GIS)
- GLIM, 244, 246, **246**
- global optimality, *see* optimization algorithm, global optimality
- Gnu Multiple Precision (GMP), 93
- Goodman’s regression, *see* ecological inference, Goodman’s regression

- hat matrix, 251
- Hessian, 103, 105, **143**
 - bootstrapping approach, 174–175
 - derivative approach, 174
 - inversion, 10–11, 194–195
 - invertibility, 143
 - nonpositive definite, 149
 - noninvertible, 262–263
 - causes, 149
 - numerical example, 166, 169
 - positive definite, 262
 - processing, numerical example, 163–171
 - respecification, 146, 175
 - singular, 149
- heuristic, 30, 32–40
 - defined, 16
- highest posterior density, 109–113, 148, 263
 - average coverage criterion, 110
 - average length criterion, 110
 - definition, 109
 - highest average coverage criterion, 110
 - simulation methods, 112
 - worst outcome criterion, 111
- HPD, *see* highest posterior density

- ICPSR, *see* Inter-University Consortium for Political and Social Research
- implementation
 - defined, 16, 19
 - error, 21–22, 258
- implementation uncertainty, **197**
- importance ratio, 160
- importance sampling, 147, 160–163
 - algorithm, 160–161

- importance sampling (*continued*)
 - applied to pseudo-variance matrix, 163
 - approximation distribution, 160
 - diagnostics, 163
 - illustration, 162
- importance weight, 161
- independence chain, 125
- inference, 20–21
 - Bayesian, 18, 145, 148
 - logistic regression, 250
 - likelihood model, 18, 145
- influence, 71, 72
- inlier, 140
- instrumental variables, 83
- integer linear programming, 31
- Inter-University Consortium for Political and Social Research (ICPSR), 60, 182, 189, 191, 255, 256, 259
- inverse probability model, 18

- jackknifing, 81
- Java, 63
- JMP, **247**

- KISS, *see* pseudo-random number generator, KISS
- Kolmogorov's zero-one law, 123

- LAPACK, 92
- Leamer bounds, 116, **197**, 261
- likelihood curvature, 77
- likelihood displacement, 77
- likelihood function, 76
 - core region, 90
 - curvature, 77, 85
 - nonconcavity, 85
 - quadratic, 261
 - sensitive, 77
- likelihood inference, *see* inference, likelihood model
- likelihood mean-shift, 77
- likelihood principle, 108
- likelihood ratio, *see* confidence intervals, likelihood ratio
- Limdep, 199, 244, **246**

- linear model
 - Gauss–Markov assumptions, 78
 - generalized, *see* generalized linear model
- logistic regression, 265
 - Bayesian inference, **250**
 - convergence criteria, **246**
 - diagnostics, **247**
 - example, Florida, 168
 - example, Texas, 164
 - likelihood, 168
 - MLE, **238**
 - separation, **240**
 - detection, **245**
 - treatment, **248**
 - software problems, **244**
- logit, *see* logistic regression
- log relative error (LRE), 19, 53, 59
- LogXact, 250
- lunch, no free, 40, 99, 102

- $m - n$ poly- t form, 107
- machine epsilon, 24, 157
- machine precision, 83
- Maple, 93–95, 261
- Markov chain, 119
 - absorbing condition, 122
 - absorbing set, **122**, 139
 - absorbing state, **139–140**, 142
 - auxiliary variables, 135
 - burn-in period, 133
 - chain period, 122
 - closed set, **122**
 - convergence
 - bounds, 264
 - conditions, 126
 - diagnostics, 264
 - convergence of empirical averages, 134
 - convergence parameter, 123
 - convergence to stationarity, 134
 - cycle, 122
 - definition, 121
 - degeneracy, 140
 - detailed balance equation, 127, 135
 - empirical averages, 134
 - equilibrium distribution, 123, 124
 - ergodic, 133, 265
 - geometric convergence, 135

- geometrically ergodic, 129, 134
- Gibbs sampler, 128, 130, 135
 - steps, 129
- hit-and-run algorithm, 127
 - steps, 128
- homogeneous, 129
- invariant distribution, 124
- irreducible set, 121
- limiting distribution, 124
- marginal distribution, 124
- measure conditions, 120
- memory-less property, 121
- Metropolis–Hastings algorithm, 126, 130, 142
 - acceptance ratio, 140
 - actual transaction function, 127
 - candidate generating distribution, 126
 - instrumental distribution, 126
 - jumping distribution, 126
 - proposal distribution, 126
 - transition kernel, 127
- minorization condition, 131
- mixing, 264
- movement, 120
- obtainable set, **122**
- partial sums, 134
- periodicity of generator, 131
- properties
 - aperiodic, **122**
 - communicating, 122
 - ergodic, **124**
 - Harris recurrent, 123, 134
 - Harris *R*-recurrent, 133
 - homogeneous, **122**
 - irreducible, **121**
 - periodic, 122
 - persistent, 122
 - positive recurrent, 123
 - recurrent, **122**, 123
 - transient, **123**
- pseudo-random number generator
 - problems, 137
- ψ -communicating, 121
- random number generation, **129**
- random number generators, 130
 - problems, **130**
- reversibility, 127
- σ -finite kernel, 120
- slice sampler, 135
- stationary distribution, **124**, 133
- symmetry, 127
- transient, 123
- transitioning, 120
- transition kernel, 131
- trapping set, 139
- unbounded and continuous state
 - spaces, 123
- uniform ergodicity, 125
- unobtainable set, 122
- Markov chain Monte Carlo (MCMC), 10, 32, 35, 38, 58, 119, 145, 250, 263
 - background, **118–120**
 - mechanics, **126–129**
 - properties, **121–126**
 - spatial models, **231–234**
 - theory, **120–121**
- Markovian property, 121, 129, 135, 139
- Mathematica, 59, 93–95, 207, 261
- Matlab, 59, 93, 226, 237
- MATLAB, 48
- matrix decomposition
 - diagonalization, 152
 - LDU, 151, 152
- matrix inverse conditions, 151
- maximum likelihood estimation (MLE), 5
 - constrained, 181
 - convergence failure, 243
 - nonexistence, 243
- measurable functions, 120
- measurable space, 120
- measure space, 132
- measurement error, 18
 - generalized linear model, 82
 - linear model, 78–82
 - induced covariance, 79
 - multivariate, 80
 - model effect, 78
 - zero mean, 78
- measures
 - positive, 120
 - signed, 120
- Metropolis–Hastings algorithm, 119, 126, 135, 234, 264
 - acceptance ratio, 127
 - steps, 127
 - symmetry, 127

- Minitab, 244, **246**
 misspecification, 20
 mixture distribution, 113
 mixture model, 113
 MLE, *see* maximum likelihood
 estimation (MLE)
 model specification, 257
 MODULA-2, 138
 Monte Carlo arithmetic, 77–78
 Monte Carlo simulation, **140–141**, 147,
 259
 error term, 141
 expected value, 141
 steps, 141
 Moore–Penrose theorem, 152, 153
 MRC Biostatistics Unit, 137
 multicollinearity, 46, 144, 173
 multiplicative congruential generator,
 138
 MuPad, 93
- National Election Survey, 90, 255
 National Institute of Standards and
 Technology (NIST), 23, 45,
 49–50, 53, 88, 92, 199
 National Science Foundation, 253
 Newton's method, *see* optimization
 algorithm, Newton's method
 Newton–Raphson, 239, 240, 243, 244,
 246, 247, 251
 NIST, *see* National Institute of
 Standards and Technology
 nonlinear estimation
 diagnostics, **210**
 function profiling, **212–213**
 gradient analysis, **211**
 Hessian condition, **212**
 Tobit model, 203
 trace analysis, **211**
 nonlinear least squares, *see*
 optimization algorithm, nonlinear
 least squares
 nonlinear maximum likelihood, 199
 nonlinear regression, *see* optimization
 algorithm, nonlinear least squares
 norm operator, 132
 North Carolina, 6–8
 NP-completeness, 31
- numerical accuracy, 4–6, 19
 backward error, 48
 benchmarks, 48–49
 test results, 64–65
 error detection, 4
 example, rare event counts model, 6,
 9
 examples, 4
 stability, 20
 numerical expected value, 141
- omitted variable bias, 169
 open source software (OSS), *see*
 statistical software, open source
 optimization algorithm, 30, 39–41, 75,
 84, 94, **200–202**, 258–259,
 261–262
 artificial neural networks (ANN),
 101–102
 BFGS, 67, 97–99, 103, 186, 202,
 203, 206, 217
 BHHH, 67, 202, 204
 conjugate gradient, 67, 96
 convergence, 102
 convergence criteria, 98, 103, **205**
 convergence rates, **205**
 Davidon–Fletcher–Powell (DFP), 67,
 97, 151, 202
 downhill simplex method, 98
 example, ecological inference,
 186–187
 Gauss–Newton, 204, 208, 217
 genetic algorithms, 101
 global optimality, 84–92
 global optimality tests
 Starr test, 86, 88–90
 Veall test, 87–90
 global optimum, 39–40, 51, 84
 gradient, 67
 grid search, 87, 109
 implementation, 99
 Levenberg–Marquardt, 204, 217
 line search, 96, 97, 201
 local optima, 84, 103
 multiple modes, 89, 106
 example, 107–109
 Nelder–Mead algorithm, 99
 Newton's method, 96, 98, **201**
 Newton–Raphson, 67, 204

- NL2SOL, 217
- nonlinear, *see* nonlinear estimation
- nonlinear least squares, 199, 261–262
- numerical accuracy, **206–209**
- optimization method, 95–96
- options, 52
- quadratic hill-climbing, 204
- quasi-Newton, 202, 204–206
- simplex algorithm, 99
- simulated annealing, 99–101
- solvers, 52–53, 67
- starting values, **51**, 53, 199, 240
- steepest descent, 67, 96
- stopping rules, 246
- trust region, 204
- trust-region algorithm, 96
- outliers, 71, 72

- parallel analyses, 144
- penalized log likelihood, 251
- penalized maximum likelihood estimation, **250**
- perfect sampling, 264
- PERL, 51, 258
- posterior distribution
 - as objective function, 145
 - plateau, 150
 - ridge, 145
 - saddlepoint, 150
- precedence, 51
- precision, 24
 - defined, 15
- prior, 18
- PRNG, *see* pseudo-random number generator
- probability measure, 121
- pseudo-random number generator, 32–39, 75
 - add with carry, 33–35
 - benchmarks, 54–58, 65
 - birthday-spacings test, 55
 - compound, 36
 - crypto-strength, 132
 - defined, 32–33
 - DIEHARD, 55, 56, 64
 - distribution of draws, 37
 - hardware generators, 39
 - independence, 37
 - inverse CDF method, 36
 - inverse congruential, 35
 - KISS, 132
 - lagged Fibonacci, 33–35
 - linear congruential, 33, 131, 136
 - long-runs test, 55
 - Markov chains, *see* Markov chain, pseudo-random number generator
 - Mersenne twister, 35, 132, 264
 - mother-of-all generator, 132
 - multiple recursive, 33, 132
 - parallel application, 58
 - parallel processor, 37
 - period, 33, **36**, 57, **57**
 - PRNGS, 181
 - KISS, 188
 - RANDMAX, 132
 - RANDU, 33–34, 130, 135, 137
 - RANROT, 132
 - rejection method, 36
 - reproducibility, 37
 - seed, **37**, 57
 - shuffle, 35
 - SPRNG, 56–57, 93
 - TESTU01, 56, 93
- pseudo-variance matrix, 147, 151, 155, 160, 171
- pseudoprime, 30–31
- ψ -irreducibility, 121

- QR factorization, 153
- quasi-Newton, 202, 204

- R, xv, 7, 9, 37, 44, 51, 70, 74, 136, 137, 176, 207, 215, 244, 246, **246**
- R Development Core Team, 51, 294
- R-square, 81
- racial bloc voting, 177–180
- random number generation, 32
 - hardware sources, 38
 - Markov chains, *see* Markov chain, random number generation
 - pseudo, *see* pseudo-random number generator
 - true, 57
- random walk chain, 125

- randomized algorithm, 29–31
 - example, prime number, 30
- recreation, 255
- register overflow, 138
- rejection sampling, 147, 173
- reliability, 3
- replication, 2, 5, 10, **253–254**, 254
 - example, failure to replicate, 254
 - example, scobit, 66
- reproduction, 254–255; *see also* replication
 - independent, 254
 - secondary analysis, 254
 - tasks, 255
- resistance, 71, 72
- ridge regression, 173–174
- robustness, 71, 72
 - Bayesian, 72–73
 - global, 71
 - linear model, 72
 - local, 71
- rodeo, prison, 255
- roll-off, 190

- S-Plus, 7, 9, 37, 74, 153, 155, 199, 208, 213, 251
- sample space, 119
- sampling importance resampling (SIR), *see* importance sampling
- SAS, 7, 9, 42, 44, 74, 199, 213, 244–247, **247**, 249, 251, 296
- Schrage algorithm, 138
- scobit, 65–69, 88–90
- Scythe Library, 132
- sensitivity analysis
 - data perturbation, **182–185**
 - bounded, 185
 - proportional data, 185
 - definition, 73
 - perturbations
 - estimation effect, 76, 77
 - maximizing, 77
 - minimizing, 77
 - perturbation tests, 46, 75–84
 - software, 74
 - tests, 73–91
- SHAZAM, 66–68, 74, 199

- σ -algebra, 120, 121
- signed measure, 132
- simulated annealing (SA), *see* optimization algorithm, simulated annealing
- simulation, 4, 5
 - error, 141
 - variance, 190–191
- smalltalk, 51
- software bugs, 22–23, 41, 46, 257
 - ecological inference, 189
 - examples, 64
- solution, defined, 15
- spatial dependence
 - definition, **219–220**
 - specification using weight matrices, **220–221**
- spatial regression, 266
 - Bayesian heteroskedastic models, **229–231**
 - estimation, **231–236**
 - estimation issues, **222–223**
 - limited dependent variables, 222
 - maximum likelihood estimation, **223–224**
 - numerical accuracy, **226–229**
 - sparse matrix algorithms, **224–225**
 - spatial autoregressive model (SAR), **221–222**
 - spatial Durbin model (SDM), **222**
 - spatial error model (SEM), **222**
 - vectorization, **225–226**
- SPSS, xv, 44, 199, 244, **245**, 246, **247**
- standard deviation, 12–14
- starting values, *see* optimization algorithm, starting values
- Stata, xv, 7, 9, 50, 63, 66–68, 74, 244, **246**, **247**, 258
- state space, 120
- statistical physics, 119, 126
- Statistical Reference Datasets (StRD), 49–51, 70, 199, **199**, 259
 - BoxBOD test, 41, 88, 89, 95
 - Misra1a test, 52, 206–210, 213
 - π -digits test, 50
- statistical software, 44, 60, 65, 74
 - arbitrary precision, 94
 - computer algebra, 94

- external library, 63
- high precision, 91
- multiple precision, 93
- open source, 70, 74
- selecting, 69–70
- speed, 91
- statistics textbooks, **xiv**
- steepest descent, *see* optimization
 - algorithm, steepest descent
- stochastic process, 119, 121
 - H*-valued, 121
 - history, 121
 - random elements, 121
- stochastic simulation, 118
- stopping rules, *see* optimization
 - algorithm, stopping rules
- StRD, *see* Statistical Reference Datasets
- symmetry, 127
- Systat, 244, 246

- Thornburg v. Gingles*, 178
- Tobit, 203
- total variation norm theorem, 133
- Transactions on Mathematical Software*, 92
- transition kernel, 120

- truncated singular normal distribution, 171, 173
- TSP, 207

- U.S. Census, 219, 228

- variance reduction, 141
- VDC, *see* Virtual Data Center
- verification, 254; *see also* replication
 - recommendations, 259–260
 - verification data, 258
- Virtual Data Center (VDC), 256

- Wald inference, **249**
- Wald test, *see* confidence intervals,
 - Wald inference; confidence intervals, likelihood ratio
- warning messages, 7, 14, 246
- Web site, **xv**
- WinBUGS, xv, 6, 112, 131, 135, 137, 142, 250
 - evaluation, **137–139**
 - period of generator, 131
 - random seed, 139
- within-chain versus cross-chain
 - correlation, 265

- Yacas, 93, 95