

Index

- abnormal activity, *see* anomalies
- abort, 48, 393, 405, 457
- absolute deviation, 90, **93**
- absolute value, *see* square vs. absolute value
- abstract descriptive model, 17, 134, 340
- abstraction, 10, 16
- acceptance testing, 46
- accounting, 179
- acf (autocorrelation function), *see* autocorrelation
- ack, 369, 438, 463
- active instrumentation, 30
- adaptivity, 38, 241, 398, 424
- address space, 230, 414
- ADSL, 438
- aest, 195
- agglomerative clustering, 265, 379
- aggregated Whittle estimator, 337
- aggregation, 190, 195, **295**, 316
- aging, 8, 434
- AIC, *see* Akaike's information criterion
- AJAX, 461
- Akaike's information criterion, 19
- aliasing, 32, 328
- α -stable distribution, 170
- amplitude, 327
- analysis of variation, 253
- Anderson-Darling test, 160
- anomalies, 15, **52–63**, 365, 389, 446
- anonymization, 71–72
- anti-persistent process, 308
- AOL search log, 68, 71, 451–456
- application scaling, 3, 480
- AR, *see* autoregressive process
- ARIMA, *see* autoregressive intergrated moving average process
- arithmetic mean, 87
- ARMA, *see* autoregressive moving average process
- arrival process, 7, 64, 279, 284–287, 289, 351–353, 395, 400, 435, 440, 474, 489, *see* daily cycle, interarrival time
- arrivals of new users, 371
- ASCII file, 28, 66, 70
- aspect ratio, 186
- asymmetrical distribution, *see* skewed distribution
- asynchronous request, 376, 431, 461
- attack, 15, 59, 464
- autocorrelation, 44, 239, **275**, 284, 300, 318
- autocovariance, **275**, 300, 303, 325
 - and spectral density, 326, 331
- automatically generated workload, *see* robot
- autoregressive integrated moving average process, 341
- autoregressive moving average process, 341
- autoregressive process, 340
- average, **86–88**, 90
 - trimmed, 88
- average absolute deviation, 93
- b-model, 347
- back button, 444
- backup, 435, 438
- backward operator, 341
- bad data, 34, 47, 164
- bag-of-tasks, 39, 476, 488
- balls and urns, 208
- basic block, 413, 423
- basin of attraction, 171
- batch arrivals, 285
- Bayes information criterion, 19
- bell curve, 114
- bell shape, 85, 114, 427
 - in log scale, 115
- Bellcore Ethernet workload, 436

542 Index

- benchmark, **8–10**, 413, 421–423, 483
 ESP, 485
 HiBench, 471
 MediaBench, 421
 MiBench, 421
 NAS, 485
 PARSEC, 420
 SPEC, *see* SPEC
 SPLASH, 485
 TPC, *see* TPC
- bias model, 347
- BIC, *see* Bayes information criterion
- big data, 469–473
- bimodal distribution, 129, 140, 178, 377, 379
- binary format, 66, 70
- binning, 79, 186
 logarithmic, 80, 194
- BitTorrent, 460
- blogs, 447, 461
- BM, *see* Brownian motion
- book sales, 174
- bootstrap, 160, 161, 190, 239
- bot, *see* robot
- botnet, 464
- bounded Pareto distribution, *see* truncated Pareto distribution
- box plot, 97, 166
- branch prediction, 414, 420
- branching, 31, 413
- Britney Spears, 55, 455
- Brownian bridge, 160
- Brownian motion, 342
- browsing the web, 368, 380, 440, 445
- BSP model, 483
- buddy system, 3
- buffering, 32, 89, 291
- burstiness, 14, 285, **289**, 357
- caching, 34, 35, 179, 415, 430, 445, 458, 473
 disabling the cache, 415
 web pages, 442
- call center, 405
- Cantor set, 289
- capacity planning, 14
- CCDF, *see* survival function
- CDF, *see* cumulative distribution function
- censored data, 47, **147**
- centered data, 96, 275, **296**, 300
- central limit theorem, 114, 116, 169, 190, 211
- central moment, 96
- central tendency, 86–90, 98
- centroid, 263, 264
- characteristic scale, 172, 287
- Chauvenet's criterion, 62, 202
- χ^2 test, 161, 237
- citations, 128, 209
- city sizes, 209
- cleaning of data, 46–63
- clickmap, 443
- clickstream, 451
- client-server, 405, 438, 449, 461, 464
- closed system, 18, 394, 401
- cloud, 14, 462
- clustering, 61, 244, **262–267**, 374
- coast of Britain, 288
- Coda, 397
- coefficient of determination, 253
- coefficient of variation, **91**, 108, 110, 309, 410
- coin toss, 75
- color coding, 248
- comb, 149, 426
- comma-separated values, 67
- comparing distributions, 84, 97, 155–161
- complementary cumulative distribution function,
see survival function
- complex number, 328
- conditional distribution, 258, 261, 270, 398
- conditional expectation, 5, 175, 179, 411
- conditional probability, 103, 227
- confidence intervals, 12, 390
- configuration change, 14, 38, 48
- congestion control, 36, 394, 435, 463
- connection, 438
- conservative workload modeling, 5, 17, 491
- contention, 392, 418, 419
- continuous distribution, 76, 78, 159
- convergence, 86, 165, 167, 204, 228
 in distribution, 169
 in probability, 169
- convolution, 322
- cookies, 445
- correlation, 220, *see* locality, long-range dependence, short-range correlation
- correlation coefficient, **251**, 249–261, 275, *see* rank correlation coefficient, distributional correlation
- correlogram, 276
- covariance, 43, **250**, 275, 298, 301
- covariance matrix, 299, 305, 333
- Coxian distribution, 112, 134
- CPU workload, 8, 41, 234, 413, 421
- crawling the web, 55, 61, 447
- critical request, 432
- cross-correlation, 214, *see* correlation coefficient
- cross-validation, 148, 492
- CSV, *see* comma-separated values
- cumulative distribution function, **81**, 89, 159
 multidimensional, 245
- cumulative histogram, 85
- cumulative process, 295, 298, 306
- curvature, 189
- CV, *see* coefficient of variation
- cycles, 44, *see* daily cycle, periodicity
- daily cycle, 12, 44, 50, **273–280**, 285, 364, 382, 391,
 400, 440, 457, 459, 461, 462, 489
 modeling, 278, 374, 403

- data center, 462
- data cleaning, 46–63
- data errors, 35, 47
- data filtering, 50
- data portability, 64
- data quality, 34–37, 46–50, 52–63
- data sanitization, 71
- data volume, 14, 32, 70, 75
- database, 6, 370, 465
- DDoS, *see* denial-of-service
- dendrogram, 266
- denial-of-service, 15, 464
- density, 79, 83, 116
- dependence, 43, 213, 258, 300, *see* long-range dependence
- dependency, *see* feedback, think time
- descriptive model, 16, 245, 491
- DFT, *see* discrete Fourier transform
- DHCP, 51
- difference operator, 341, 343
- differences, 45, 281, 295
- dilation, 319
- dimension, 288
- directory, 207
- disabled cache, 415
- discrete distribution, 76, 78, *see* modal distribution
- discrete Fourier transform, 327
- discrete wavelet transform, 322, 345
- dispersion, 90–95, 98
- distance, 263
- distributed file system, 430
- distribution
 - α -stable, 170
 - bimodal, 140, 377
 - comb, 149, 426
 - comparing, 84, 97, 155–161
 - continuous, 78
 - description, 75–100
 - discrete, 78
 - empirical, 130, **145–148**, 159, 186
 - Erlang, **109–110**, 136, 155
 - exponential, **102–106**, 135, 172, 309
 - exponential family, 135
 - extreme value, 171
 - gamma, **118–119**, 133, 137
 - generative model, 105, 117, 122, 206–212
 - geometric, 104, 307
 - heavy tail, **164**
 - instances, 122, 410, 427, 439, 441
 - hyper-Erlang, **111**, 137
 - hyper-exponential, **107–109**, 136, 151–154, 270
 - hypo-exponential, 110
 - log-uniform, 128, 379, 475
 - lognormal, **115–117**, 137, 171, 173, 205, 210
 - instances, 115, 206, 427
 - Lomax, 121, 178
 - long tail, 172
 - mixture, 107, 111, 129, **141**, 267
 - modal, 86, 131, **148–151**, 235, 266, 357, 383
 - mode, 58, 154
 - multinormal, 267, 300, 333
 - normal, 85, **114**, 169, 340
 - Pareto, **120–123**, 138, 165, 166, 167, 170, 173, 176, 183, 193, 230
 - instances, 122, 403, 410, 427
 - phase-type, **106–107**, 133, 151
 - Poisson, 103
 - shifted, 101
 - skewed, 84, 85, 88, 93, **98**, 115, 118, 223
 - subexponential, 172
 - tail, 84, 85, 138, 158, 160
 - Weibull, **119–120**, 137, 173
 - Zipf, **123–128**, 191, 223, 460
 - instances, 224, 443, 453, 458
 - Zipf-like, **124**, 242, 244, 458
- distribution function, *see* cumulative distribution function
- distribution of wealth, 122, 184, 209
- distributional correlation, 259, 270, 478
- distributional correlation coefficient, 259
- diurnal cycle, *see* daily cycle
- diversity, **236**, 358, 427, 478, 493
- divisive clustering, 265
- dominance, 166, 174, 202
- double counting, 48
- downtime, 47
- drift, 280
- drunk, 306
- DSL line, 381
- DWT, *see* discrete wavelet transform
- dynamic content, 443, 444, 449
- dynamic linking, 30
- dynamic workload, 7, 270, 428
- Dyninst, 30
- e-commerce, 369, 449
- economic model, 392
- 80/20 rule, 122, 178, *see* joint ratio
- EM algorithm, **142–144**, 154, 378
- email, 15, 50, 55, 463
- empirical CDF, *see* empirical distribution
- empirical distribution, 75, **81**, 130, **145–148**, 159, 186
 - multidimensional, 249
- eMule, 460
- Enron email archive, 50, 441
- enterprise network traffic, 435
- entropy, 223, 349
- epidemic, 447, 462
- equivalent conditions, 8, 394, 486
- ergodicity, 228
- Erlang distribution, **109–110**, 155
 - parameter estimation, 136
- erroneous data, 35, 47
- ESP benchmark, 485
- Ethernet, 148, 435, 436
- Euclidean distance, 263

544 Index

- Euler's formula, 329
- evolution (in biology), 208
- evolution (of workload), 37, 148, 280, 285, 398, 436
- exact self-similarity, 287, 297–299
- expectation, 86
- expectation paradox, 176
- expectation-maximization, *see* EM algorithm
- explained variance, 267
- exploratory data analysis, 17
- exponential cutoff, 205
- exponential distribution, **102–106**, 172, 187
 - generative model, 105
 - maximum likelihood parameter, 135
 - memoryless, 103, 177
 - test, 309
- exponential family, 135, 173
- extrapolation, 146, 202, 364
- extreme value distribution, 171
- eyeball method, 60, 266

- Facebook, 31, 462, 463, 471
- factorial, 118, 344
- failed statistical test, 160
- failure, 8, 48, 105, 177, 407, 449
- fARIMA, *see* fractional ARIMA process
- fast Fourier transform, 329
- fBM, *see* fractional Brownian motion
- feedback, 13, 19, 383, **391–398**, 431, 435, 463, 483
- fGN, *see* fractional Gaussian noise
- Fibonacci numbers, 80
- 50/0 rule, 178
- file access patterns, 429
- file layout, 433
- file sharing, 38, 434, 459
- file sizes, 117, 178, 182, 206, 210, 426–429
- filtering (data), 50
- filtering (signal), 322
- finite dimensional distributions, 298
- finite support, 175
- firewall, 464
- first differences, 45, 281, 295, 341
- first-order distribution, 298
- flash crowd, 389, **446**, 460, 462, 465
- flow, 29, 438
- fluctuations, 12, 44, 241, 281, 289, 292, 309, 363, 364, 372, 390
- flurries, **56–59**, 358, 365, 389, 390
- FMA instruction, 412
- format of log, 24, 28, **66**
- Fourier frequencies, 328
- Fourier transform, 326, 327
- fractal, 230, 287–288
- fractal model, 229
- fractional ARIMA process, 343–345
- fractional Brownian motion, 342
- fractional Gaussian noise, 342
- fragmentation, 4, 36, 149, 426, 433, 486
- France'98 website workload, 35, 183, 280, 446
- frequency, 76, 326

- gaming, 392, 405, 460
- gamma distribution, 118–119
 - parameter estimation, 133, 137
- gamma function, 118, 344
- gang scheduling, 3
- gap in data, 35, 47
- garbage collection, 105
- Gaussian noise, 340
- generalization, 10, 12, 70, 200
- generalized Zipf distribution, 124
- generative model, 17, 245, 340, 362, 464, 491, *see*
 - user-based modeling
 - exponential, 105
 - heavy tail, 206–212
 - hierarchical, 357
 - lognormal, 117, 210
 - Pareto, 122, 209
 - self-similar, 350–355
- geometric distribution, 104, 307
- geometric mean, 87, 94, 115
- geometric standard deviation, 93
- Gini coefficient, 185
- global distribution, 236
- Gnutella, 460
- goodness of fit
 - Anderson-Darling test, 160
 - χ^2 test, 161
 - Kolmogorov-Smirnov test, 159–160
- Google, 31, 462
- Google maps, 461
- GPU workload, 421
- graphics, 421
- graphs
 - aspect ratio, 186
 - box plot, 97
 - CDF, 83
 - correlogram, 276
 - heatmap, 248
 - histogram, 76, 79, 80, 194
 - LLCD, 186
 - logarithmic scale, 77, 85, 100
 - logscale diagram, 324
 - P-P plot, 158
 - pdf, 79
 - periodogram, 331
 - pox plot, 313
 - Q-Q plot, 155
 - scatterplot, 246
 - split scale, 78
 - texture plot, 310
 - variance-time plot, 317
- grid, 410, 462, 488
- guessing, 63–64

- Haar transform, 321, 346
 Haar wavelet, 321
 Hadoop, 471
 harmonic mean, 87
 harmonics, 328
 Harry Potter, 174
 hazard, 147, 175, 177
 header, 28, *see* packet header
 heat map, 248, 350
 heavy tail, 64, 122, **164**, *see* mass-count disparity
 alternatives, 172, 187, 204, 211, 427
 conditional expectation, 5, 176, 411
 instances, 122, 410, 427, 439, 441
 tail index metric, 164, 187
 HGM, *see* hierarchical generative model
 HiBench benchmark, 471
 hidden Markov model, 229, 244, 359, 389, 430
 hierarchical clustering, 265, 379
 hierarchical generative model, 357
 hierarchical modeling, 359–370
 high-pass filter, 322
 high-performance computing, 473
 Hill estimator, 198
 histogram, 76, 79
 binning, 79
 logarithmic bins, 80, 194
 vs. pdf, 80, 116
 weighted, 270
 HMM, *see* hidden Markov model
 horizon, 203
 hotspot, 225, 447, 483
 H-ss, 299
 H-sssi, 299
 HTTP, 28, 35, 43, 183, 219, 287, 368, 435
 HTTP status codes, 28, 48, 50, 442
 Hurst parameter, 294, **298**, 306–308, 356
 estimation, 310–339
 Hurst, Harold Edwin, 292
 hydrology, 292
 hyper-Erlang distribution, 111
 parameter estimation, 137
 hyper-exponential distribution, **107–109**, 205, 270, 402, 410
 approximation of heavy tail, 151–154
 parameter estimation, 136, 154
 hypo-exponential distribution, 110
 I/O modes, 483
 idle time, 410
 iid, *see* independent identically distributed
 immutable file, 459
 increment process, 295, 299, 306
 independent identically distributed, 43
 independent reference model, 224
 infinite moments, 166
 information retrieval, 468
 informational query, 453
 innovations, 295
 input, 6, 413, 420, 422
 instruction count, 413, 422
 instruction mix, 40, 234, 412
 instruction-level parallelism, 234, 421
 instrumentation, 29–32
 insurance, 180
 inter-process locality, 434
 inter-reference gap, 229
 interactive behavior, 392, 402, 421, 462
 interarrival time, 103, 309, 363, **376**, 379, 395, 486
 interference, 32
 Internet protocols, 59, 437
 Internet service provider, 28, 51, 463
 Internet traffic, 434–439
 sanitization, 71
 interpolation, 146
 interquartile range, 94
 interrupt, 367
 intrusion detection, 15
 invariants, 38, 63, 399
 inverse distribution function, 105
 inverse wavelet transform, 323, 345
 I/O behavior, 431
 IP address, 28, 51, 71, 438, 464
 IQR, *see* interquartile range
 IRM, *see* independent reference model
 ISP, *see* Internet service provider
 job mix, 420
 joint distribution, 42, **245**, 298, 300
 joint ratio, 98, **181**, 223, 423
 Joseph effect, 291, 292, 308
 Kaplan-Meier formula, 147
 KaZaA, 460
 Kendall's τ , 256
 Kendall's concordance, 257
 key-value store, 473
 keystroke, 367, 402
 k -means clustering algorithm, 263
 Kolmogorov-Smirnov test, 159–160, 237
 K-S, *see* Kolmogorov-Smirnov test
 k -transform, 192
 ℓ_2 distance, 263
 lack of convergence, 165, 167
 lag, 240, 275, 284
 LAMP architecture, 449
 LAN traffic, 289, 291, 435
 LANL CM-5 workload, 56–58, 86, 232, 383
 last, 47, 51
 lastcomm, 36, 411
 law of large numbers, 86, 167
 learning, 38, 398
 least squares, 252
 lifetime, 411
 likelihood, 19, 134, 197, 333
 limiting distribution of Markov chain, 228

546 Index

- linear process, 341
- linear regression, 187, 252–253
- LLCD, *see* log-log complementary distribution
- load, 255, 279, 285, 363, 392, 393, 394, 486
- load balancing, 4, 179, 410
- load manipulation, 363, 372, 390, 485–487
- local Whittle estimator, 337
- locality, 34, 63, **215–245**, 415, 445, 465
 - Denning's definition, 215
- locality of sampling, 231–245, 350, 357, 358, 383
 - spatial, 428
- location parameter, **101**, 120, 121
- log, 10, 272, 390
 - format, 24, 28, **66**
- log histogram, 309
- log-likelihood, 134, 197, 333
- log-log complementary distribution plot, 100, 122, 126, 170, **186–189**
- log-uniform distribution, 128, 379, 475
- logarithmic scale, 77, 80, 85, 100, 115, 126, 158
- logarithmic transformation, 84, 115, 128, 262, 282
- lognormal distribution, **115–117**, 171, 188
 - as alternative to heavy tail, 173, 205, 210, 427
 - generative model, 117
 - instances, 115, 206, 427
 - parameter estimation, 137
- logscale diagram, 324
- Lomax distribution, 121, 178, *see* shifted Pareto distribution
- long tail, 172, 174, 188
 - in Internet economics, 174
- long-range dependence, **300**, 318
- Lorenz curve, 184
- low-pass filter, 322
- LRD, *see* long-range dependence
- LRU stack, 219
- lunch, 274, 375, 403

- MA, *see* moving average process
- MAIG, *see* maximum allowable inter-car gap
- malicious traffic, 441, 464
- Mandelbrot, Benoit, 288, 292
- Manhattan distance, 264
- manipulating the workload, 11, 18, 363, 390, 485
- MAP, *see* Markovian arrival process
- MapReduce, 36, 470
- marginal distribution, 246
- Markov chain, 104, 106, **227–228**, 284, 423
- Markov reference model, 229
- Markov-modulated Poisson process, 285
- Markovian arrival process, 284
- Markovian model, 151, 227–229, 359, 388, 430, 450
- mass distribution, 178
- mass-count disparity, 98, **178–185**, 223, 225, 426, 484
 - $N_{1/2}$ metric, 181
 - $W_{1/2}$ metric, 181
 - joint ratio metric, 181
 - median-median distance metric, 182
- mass-count disparity plot, 178
- matching moments, *see* method of moments
- maximum allowable inter-car gap, 376
- maximum likelihood, **134–136**, 333
 - exponential, 135, 137
 - Pareto, 197
- maximum transmission unit, 36, 148, 436
- mean, 86, *see* average
- MediaBench benchmark, 421
- median, **88–90**, 139
- median-median distance, 99, **182**, 447
- memory access, 215, 229, 365, 415
- memory allocation, 234
- memoryless, 64, **103**, 177, 272
- message-passing interface, 482, 485
- metadata, 65
- Metcalfe's law, 128
- method of moments, **133–134**, 138–140
 - Erlang, 137
 - gamma, 133
 - hyper-exponential, 136
 - lognormal, 137
- M/G/ ∞ queue, 350
- MiBench benchmark, 421
- mice and elephants, 178
- micro model, 418, 431
- microbenchmark, 9
- migration, *see* process migration
- MinneSPEC, 422
- misconfiguration, 15, 51
- mixed data, *see* scrambled data
- mixture of distributions, 107, 111, 129, **141**, 267, 270, 427
- mixture of Gaussians, 150, 267
- MLE, *see* maximum likelihood
- MMPP, *see* Markov-modulated Poisson process
- mobile device, 406
- modal distribution, 83, 86, 131, **148–151**, 235, 266, 267, 357, 383, 408, 426, 478
- mode, 58, 88, 154, 357
- modeling, 10, 390
 - abstract, 17, 134, 340
 - conservative, 5, 17, 491
 - descriptive, 16, 130, 145
 - generative, 17, 117, 206–212, *see* generative model
 - parsimonious, 19, 491
 - what to model, 45
- modeling abnormal situations, 59
- modeling at the source, 340, 394, 464
- moldable parallel job, 480
- moment, **95**, 133, 138–139, 166, *see* method of moments
- monitoring, 29
- monkey typing, 124

- monofractal, 356
- monotonicity, 82, 154, 256
- mother wavelet, 319
- mouse click, 402
- move-to-front, 219
- moving average, 273, 341, 342
- moving average process, 340
- MPI, *see* message-passing interface
- MTU, *see* maximum transmission unit
- multiclass workload, 50, 147, 262, 270, 389
- multidimensional empirical distribution, 249
- multifractal, 356
- multifractal wavelet model, 346
- multimedia, 420, 459
- multinormal distribution, 267, 300, 333
- multiplicative process, 210
- multiplicative standard deviation, 93, 115
- multiply-add, 412
- ×/, 94
- multiprogramming, 26, 418
- multiresolution analysis, 319
- multivariate normal distribution, *see* multinormal distribution

- $N_{1/2}$ metric, 181, 223
- Napster, 59, 437, 460
- NAS parallel benchmark, 485
- NASA Ames iPSC/860 workload, 24–26, 52
- NAT, *see* network address translation
- navigation burst, 368, 440
- navigational query, 453
- nearly completely decomposable, 229
- negative exponential distribution, 102
- negative feedback, 392, 463
- network address translation, 51
- network file system, 435
- network intrusion detection system, 464
- network weather service, 15
- NIDS, *see* network intrusion detection system
- nighttime, 26, 279, 475
- Nile River, 292
- 90/10 rule, 178, *see* joint ratio
- Noah effect, 165, 291
- noise (in data), 47
- noise (in stochastic model), 172, 340
- nonhomogeneous Poisson process, 285, 440
- nonstationarity, 12, 272, 279, 285, 298, 318, 341, 450
- norm, 263
- normal distribution, 85, 93, **114**, 169, 259, 333, 340
- normalization, 84, 91, 121, 124, 172, 223, 238, 251, 315
- number of samples, 75
- numerical derivative, 189

- octave, 320
- OLTP, 465
- on-off process, 353, 446

- 1/ f noise, 172, 326
- online, 241, 393, 456
- online control, 14
- open system, 18, 394, 401
- order of operations, 361, 430, 450–451
- order statistics, 96, 139, 198
- ordered set, 216
- outliers, 35, 47, 52, **62–63**, 88, 93, 138, 148, 202, 255
- overfitting, 12, 19, 70, 148
- overhead, 4

- packet header, 29, 148, 438
- packet sizes, 62, 148, 270, 436
- packet train, 369, 376
- paging, 415
- PAPI, 31
- paradox, 176, 272
- parallel file system, 434, 483
- parallel I/O, 483
- parallel jobs, 2, 3, 473–482, 487
- Parallel Workloads Archive, 68
- parallelism profile, 481
- parameter estimation, 131–140
 - Erlang, 136
 - exponential, 135, 136
 - gamma, 137
 - hyper-Erlang, 137
 - hyper-exponential, 136, 154
 - lognormal, 137
 - Pareto, 138, 193–200
 - Weibull, 137
- Pareto distribution, **120–123**, 165, 166, 167, 170, 173, 176, 183, 189, 230
 - generative model, 122
 - instances, 122, 403, 410, 427
 - parameter estimation, 138, 193–200
 - shifted, **121**, 178
 - truncated, **121**, 199, 202
- Pareto principle, 122, 178, *see* joint ratio
- Pareto, Vilfredo, 122
- PARSEC benchmark, 420
- parsimony, 11, 19, 244, 491
- partitional clustering, 265
- passive instrumentation, 29
- PatchWrx, 30
- patience, 405
- pdf, *see* probability density function
- Pearson's correlation coefficient, 251
- peer-to-peer, 459
- Pentium, 31
- percent point function, *see* inverse distribution function
- percentile, 89, 95, **96**, 155
- performance counters, 31
- performance metric, 394, 397
- periodicity, 44, 275–280, *see* daily cycle
 - detection, 275, 330
- permutation, 218, 233

548 Index

- persistent process, 308
- personality, 374
- PH, *see* phase-type distribution
- phase, 327
- phase transition, 227, 229, 417, 419
- phase-type distribution, **106–107**, 133, 151
- physical model, *see* generative model
- pink noise, 172
- Poisson process, 64, **102**, 104, 210, 284, 291, 293, 371, 440
 - test, 309
- polling asynchronous operation, 431
- popularity, 64, 125, 127, 174, 191, 217, 219, 223, 224, 350, 443, 457, 458, 460, 461
- port, 438
- positive feedback, 208, 463
- power law, **164–172**, 187, 194, 300, 326
- power spectrum, 330
- powers of two, 4, 39, 148, 233, 426, 475, 481, 486
- pox plot, 312
- P-P plot, 158, 184
- PQRS model, 350
- predictability, 14, 34, 217, 240, 241, 253, 414, 424
- preferential attachment, 206–210
- prime time, 26, 50, 279, 364
- privacy, 14, 28, 71–72, 441, 460
- probability density function, 78, **82**, 141
 - multidimensional, 245
- probability mass function, 78
- process lifetime, 411
- process migration, 4, 104, 179, 410
- process runtime, 5, 179, 182, 410
- processor allocation, 3
- product moment correlation coefficient, 251
- protocol, 438, *see* TCP
 - usage, 59, 437, 460
- proxy, 28, 35, 445
- P2P, *see* peer-to-peer
- pwd, 26, 52

- QoS, *see* quality of service
- Q-Q plot, **155–158**, 184, 195
- quality of service, 15, 435
- quantization error, 36, 67, 78
- quartile, 94
- queries (database), 465
- queries (search engine), 369, 451–456

- random graph, 208
- random mixing, 218, 233
- random number generator, 12, 106, 167, 168, 203
- random text, 124
- random variable, 74
- random variate, 74
- random variate generation, **105**
 - empirical, 145
 - exponential, 106
 - gamma, 119
 - lognormal, 117
 - normal, 115
 - Pareto, 123, 203
 - phase-type, 107
 - Weibull, 120
 - Zipf, 128
- random walk, 229, 261, 306, 342
- rank correlation coefficient, 255
- rank-size plot, 123, 126, 191, 193
- ranking, 125, 174
- rare event, 164, 167, 180
- redirection, 48
- reference stream, 216, 415
- regression, 252–253
- regularity, 39, 217, 241, 358, 382, 424
- relevance, 34, 399
- repetition, 239, 242–245, 358, 383, 442
- representative input, 6, 413, 421
- representative slice, 39, 419
- representativeness, 8, 19, **34–42**, 58, 76, 88, 138, 140, 389
- request-response, *see* client-server
- resampling, 390, *see* bootstrap, user resampling
- rescaled range, 311–316
- residence time, 371
- resolution, 36, 66, 67, 68, 78, 79, 85, 237, 249, 279, 295
- response time, 3, 392, 394, 398
- reuse distance, 219, 419
- reuse set, 418
- rich get richer, 122, 207
- right tail, 85, 98, 174, *see* tail
- rigid parallel job, 475, 485
- RMS, *see* root mean square
- roaming, 406
- robot, 29, **52–56**, 61, 365, 375, 389, 390, 447, 449, 451
- root mean square, 307
- round-trip time, 405, 439, 463
- R/S, *see* rescaled range
- running average, 167, 170
- runtime, 5, 179, 182, 363, 410
- runtime estimates, 21, 407–409, 478

- sample moments, 95
- sampling, 8, 32–34, 70, 160, 178, 248, 272
- sanitizing data, 71
- satisfaction, 367, 394
- saturation, 4, 363, 393
- scale free, *see* scale invariance
- scale invariance, 171, 207, 230
- scale parameter, **101**, 102, 118, 121, 151
- scaling, 118, 171, 262, 297, 356, 467
- scatterplot, 233, **246**
- scheduling, 2, 6, 213, 442
- scientific citations, 128, 209
- scrambled data, **218**, 233
- SDSC Paragon workload, 53, 235, 236, 290, 474

- search behavior, *see* web search
 search engine, 444, 451
 seasonal cycle, 273
 second moment, 96, 139
 second-order distribution, 298
 second-order self-similarity, 303–306, 342
 second-order stationarity, 43, 303
 seed, 12, 167, 231
 self-regulation, 393
 self-similarity, 64, **287–291**, 364, *see* long-range dependence
 due to heavy-tailed data, 291
 exact, 287, 297–299
 Hurst parameter, 298, 306, 310
 modeling, 342–355
 nomenclature, 294, 307
 second order, 303–306, 342
 semi interquartile range, 95
 sensitivity, 64
 to outliers, 138
 to tail, 138, 158, 160
 sequential access, 417, 484
 serial correlation, 285, 310
 server farm, 442
 service-level agreement, 14, 15, 435
 session, 47, 51, 374–384, 391, 400, 439
 sessionlet, 450
 set-associative cache, 34
 SETI@home, 410
 shape of distribution, 98, 112, 140, *see* tail
 shape parameter, **101**, 118, 120, 124
 shaped traffic, 464
 shared memory, 485
 shift (in time), 42, 275
 shifted distribution, *see* location parameter
 shifted Pareto distribution, **121**, 178
 short tail, 175
 short-range correlation, 237, 241
 short-range dependence, 214, 301, 315, 338, 343
 shortest job first, 3, 409
 shortest remaining processing time, 442
 Sierpinski triangle, 288
 sigmoid, 114
 sign function, 90, 261
 SIMD extensions, 413, 420
 Simon, Herbert Alexander, 208
 simple LRU stack model, 225
 simulation, 10, 12, 39, 106, 107, 168, 218, 222, 366
 single link clustering algorithm, 265, 379
 SIQR, *see* semi interquartile range
 SJF, *see* shortest job first
 skewed distribution, 64, 84, 85, 88, 93, **98**, 115, 118, 163, 223, 262, 468
 skewness, 98
 Skype, 459
 SLA, *see* service-level agreement
 slashdot effect, 447
 slice distribution, 236
 slowdown, 398
 slowly varying function, 301
 SLRUM, *see* simple LRU stack model
 snapshot, 8, 270, 428
 social network, 462, 463
 software agent, *see* robot
 source model, 340, 394, 464, *see* generative model
 spam, 441, 465
 spatial locality, 216, 218, 222
 spatial locality of sampling, 428
 spatial regularity, 217
 Spearman's rank correlation coefficient, 256
 SPEC, 8
 SPEC CPU, 8, 220, 413, 421, 424
 SPECweb, 446
 spectral analysis, 356
 spectral density, 326, 330
 speedup, 480
 spike, 446
 SPLASH benchmark, 485
 split scale, 78, 380
 SQL queries, 370, 465
 square vs. absolute value, 90, 92, 238, 255, 306
 SRPT, *see* shortest remaining processing time
 stable distribution, 114, 170, 190
 stack distance, 219
 stack model, 219, 225, 245
 standard deviation, **91**
 standard workload format, 68
 static web pages, 443
 static workload, 7
 stationarity, 12, **42–44**, 272, 298, 300, 303
 statistical graphics, *see* graphs
 statistics, 75, 160
 steady state, 42, 168, 179, 292, *see* stationarity
 stochastic dominance, 259
 stochastic process, 74, 227
 dependence, 342–345
 self-similar, 297
 stationary, 42
 stock market, 180
 stream algorithms, 457
 stretched exponential distribution, *see* Weibull distribution
 strided access, 217, 417, 485
 subexponential distribution, 172
 sufficient statistics, 136
 sum squares, 253
 supply and demand, 392
 support, 62, 127, 175
 surfing, *see* browsing the web
 SURGE, 200, 445
 survival function, **82**, 126, 147, 152, 164, 172, 186
 Swing, 440
 symmetrical distribution, 85
 SYN flood, 464

550 Index

- synchronization, 482
 synthetic application, 9
- tail, 84, 85, 98, 138, 158, 160, 195, *see* heavy tail,
 long tail
 tail index, 98, 121, 164, 170, 187
 estimation, 194–200
 taken rate, 414
 TCP congestion control, 36, **394**, 435, **463**
 TCP connection, 368, 405, 438
 telnet, 368
 temporal locality, 216, 217, 218–222, 458
 texture plot, 310
 think time, 367, **376**, 383, 395–398, 402
 throttling, 393, 463, 483
 throughput, 394
 time series, 275, 285
 time to failure, 105
 time window, 222, 416
 time zone, 68, 403
 time-zone offset, 28, 68
 timestamp, 28, 68
 tolerance, 396
 TPC, 9, 466
 TPC-C, 9, 366, 466
 TPC-DS, 467
 TPC-E, 466
 TPC-H, 466
 TPC-VMS, 467
 TPC-W, 449, 466
- trace, 10, 415
 trace-driven simulation, 10, 432
 trading, 180
 traffic in network, 434–439
 traffic spike, 446
 transactional query, 453
 transformation
 logarithmic, 115, 128, 262
 transient, 44, 168, 179, 418
 transition matrix, 227, 229
 transition rate, 414
 translation, 319
 transmission window, 463
 TREC, 468
 trend, 45, 280–281, 285
 trimmed mean, 88
 truncated Pareto distribution, **121**, 169, 199, 202,
 203
 Tstat, 29
 2-norm, 263
 typing, 124, 367, 402
- UCML, *see* user community modeling language
 unbalanced loading, 180
 Unix file sizes, 178, 182, 206, 210
 Unix process runtime, 5, 179, 182, 410
 Unix time, 68
 upload, 438, 461
- URL, 443
 urns and balls, 208
 use scenario, *see* order of operations
 user adaptation, 38, 398
 user at terminal, 367, 402
 user behavior graph, 244, 388
 user community modeling language, 450
 user equivalent, 366, 446
 user mobility, 406
 user personality, 374
 user population, 370–374, 461
 user resampling, 339, 390
 user satisfaction, 367, 394
 user sessions, 47, 51, 374–384, 400, 448
 user-based modeling, 370–398, 399–406
 motivation, 362–366
 user-generated content, 461
 utilization, 4, 26, 363, 410, *see* load
- validation, 190, 239, 267, 338, 492, *see*
 cross-validation, visual validation
 value of network, 128
 variability, 12, 15, 20, 33, 59, 154, 204, 232, 291,
 293, 312, 324, 375, 389
 variance, **90**, 236, 300, 316
 Winsorized, 93
 variance-time plot, 317
 verification, 492
 video-on-demand, 457
 virus, 15, 441
 visual validation, 154, 186, 266
 voice over IP, 439, 459
 VoIP, *see* voice over IP
 volume of data, 14, 32, 70, 75
- $W_{1/2}$ metric, 181, 223
 waiting for asynchronous operation, 431
 waiting time paradoxes, 272
 WAN traffic, 435
 wavelets, **319**, 319–324, 356
 weak stationarity, 43
 wealth, 122, 184, 209
 web archive, 64
 web browsing, 368, 380, 440, 445
 web crawl, 55, 61, 447
 web search, 71, 369, 380, 451–457
 web server, 28, 34, 48, 441
 log format, 28
 weekend, 26, 279, 391, 404, 461
 Weibull distribution, **119–120**, 188
 as alternative to heavy tail, 173
 parameter estimation, 137
 weighted distribution, 178, 270, 383
 Welchia worm, 15, 59, 437
 white noise, 172
 Whittle estimator, 332–337
 Wi-Fi, 406
 WIDE backbone, 15, 59, 437

- Wikipedia, 381, 461
- window size, 416
- Winsorized variance, 93
- word frequencies, 123, 191, 209
- workflow, 392
- working set, 222, 416
- workload evolution, 37, 280, 285, 398, 436
- workload examples
 - AOL search, 71, 451–456
 - Bellcore Ethernet, 436
 - Enron email, 50
 - Facebook MapReduce, 471
 - France'98 website, 35, 183, 280, 446
 - LANL CM-5, 56–58, 86, 232, 383
 - NASA Ames iPSC/860, 24–26, 52
 - SDSC Paragon, 53, 235, 236, 290, 474
 - WIDE backbone, 15, 59, 437
- workload representativeness, 34–42, 58
- world wide web, 207, 441
- worm, 15, 59, 464
- WWW, *see* world wide web
- XML, 67, 70
- Yahoo, 31
- YouTube, 458, 463
- Yule, Udney, 208
- Z score, 262
- Zipf distribution, **123–128**, 191, 223, 460
 - instances, 224, 443, 453, 458
 - population limited, 191, 462
- Zipf's law, 125, 208
- Zipf, George Kingsley, 123
- Zipf-like distribution, **124**, 242, 244, 458