

Index

- Γ , 600
 $\Phi(z)$, 516
 χ^2 , 40, 325, 460, 461
 λ , 119
 σ_N and σ_{N-1} , 322
:=, 602
?, 421
2s, 156
- acceptance rate, 367, 369, 371, 382, 385, 396
acceptance ratio method, 381
accumulator, 254, 572, 584
activation, 473
activation function, 473
activity, 473
activity rule, 472, 473
adaptive direction sampling, 395
adaptive models, 101
adaptive rejection sampling, 372
address, 201, 470
Aiyer, Sree, 520
Alberto, 56
alchemists, 74
algorithm
 covariant, 444
 EM, 434
 exact sampling, 415
 expectation–maximization, 434
 function minimization, 475
 genetic, 397, 398
 Hamiltonian Monte Carlo, 389, 498
 independent component analysis, 445
 Langevin Monte Carlo, 498
 leapfrog, 391
 max–product, 341
 perfect simulation, 415
 sum–product, 336
 Viterbi, 342
Alice, 199
Allias paradox, 456
alphabetical ordering, 194
America, 356
American, 238, 260
amino acid, 201, 204, 279, 364
anagram, 200
Angel, J. R. P., 531
annealed importance sampling, 381
annealing, 381, 394, 399
 deterministic, 520
 importance sampling, 381
antiferromagnetic, 402
- ape, 269
approximation
 by Gaussian, 2, 303, 343, 352, 498
 Laplace, 343, 549
 of complex distribution, 185, 284, 366, 424, 435
 of density evolution, 569
 saddle-point, 343
 Stirling, 1
 variational, 424
arabic, 127
architecture, 472, 531
arithmetic coding, 101, **110**, 111
 decoder, 118
 software, 121
 uses beyond compression, 118, 250, 255
arithmetic progression, 346
arms race, 278
associative memory, 470, 507, 509
assumptions, 26
astronomy, 553
asymptotic equipartition, 80, 386
 why it is a misleading term, 83
Atlantic, 173
AutoClass, 308
automatic relevance determination, 546
automobile data reception, 596
average, 26, *see* expectation
AWGN, 177
- background rate, 309
backpropagation, 475, 477, 530, 537
backward pass, 244
bad, *see* error-correcting code, bad
Balakrishnan, Sree, 520
balance, 66
Baldwin effect, 279
ban (unit), 264
Banburismus, 265
Banbury, 265
bandwidth, 178
bar-code, 262, 401
base transitions, 375
base-pairing, 280
basis dependence, 308, 344
bat, 213, 214
battleships, 71
Bayes’ theorem, 6, 24, 25, 27, 28, 48–50, 53, 148, 326, 346, 349, 448, 495, 524
Bayes, Rev. Thomas, 51
- Bayesian, 26
Bayesian belief networks, 295
Bayesian inference, 459
BCH codes, 13
BCJR, 580
BCJR algorithm, 332
Belarusian, 238
belief, 57
belief propagation, 332, 559, *see* sum–product algorithm
Benford’s law, 448
bent coin, 51
Berlekamp, Elwyn, 172, 213
Bernoulli distribution, 117
Berrou, C., 186
bet, 200, 209, 457
beta distribution, 318
Bethe free energy, 436
Bhattacharyya parameter, 215
bias, 347, 508
 in neural net, 473
 in statistics, 308, 309, 323
biased, 323
biexponential distribution, 88, 315, 450
bifurcation, 89, 293
binary entropy function, 2, 15
binary erasure channel, **148**, 151
binary images, 401
binary representations, 132
binary symmetric channel, 4, 148, **148**, 149, 151, 211, 215, 229
binding DNA, 201
binomial distribution, 1, 313
bipartite graph, 19
birthday, 156, 157, 160, 198, 200
bit (unit), 264
bits back, 104, 108, 355
bivariate Gaussian, 390
black, 357
Bletchley Park, 265
Blind Watchmaker, 269, 398
block code, 9, *see* source code or error-correcting code
block-sorting, 121
blow up, 308
blur, 551
Bob, 199
Boltzmann entropy, 85
bombes, 265
book ISBN, 235
Bottou, Leon, 121
bound, 85
bounded-distance decoder, 207, 212

- bounding chain, 421
- box, 345, 353
- boyish matters, 58
- brain, 470
- Braunstein, A., 342
- Bridge, 126
- British, 260
- broadcast channel, 237, 239, 596
- Brody, Carlos, 246
- Brownian motion, 280, 318, 537
- BSC, *see* channel, binary symmetric
- budget, 94, 96
- Buffon's needle, 38
- BUGS, 373, 433
- burglar alarm and earthquake, 295
- Burrows–Wheeler transform, 121
- burst errors, 185, 186
- bus-stop paradox, 39, 46, 107

- cable labelling, 175
- calculator, 322
- camera, 551
- canonical, 88
- capacity, 14, **146**, **150**, 151, 183, 486
 - Hopfield network, 516
 - neural network, 485
 - neuron, 485
 - symmetry argument, 151
- car data reception, 596
- card, 233
- casting out nines, 198
- Cauchy distribution, 85, 88, 315, 364
- caution, *see* sermon
 - equipartition, 83
 - Gaussian distribution, 314
 - importance sampling, 364, 384
 - sampling theory, 64
- cave, 214
- caveat, *see* caution
- cellphone, *see* mobile phone
- cellular automaton, 130
- centre of gravity, 35
- chain rule, 530
- channel
 - AWGN, 177
 - binary erasure, **148**, 151
 - binary symmetric, 4, 146, 148, **148**, 149, 151, 206, 211, 215, 229
 - broadcast, 237, 239, 596
 - bursty, 185, 559
 - capacity, 14, 146, 150
 - connection with physics, 257
 - complex, 184, 559
 - constrained, 248, 255, 256
 - continuous, 178
 - discrete memoryless, 147
 - erasure, 219, 591
 - extended, 153
 - fading, 186
 - Gaussian, 155, 177, 186
 - input ensemble, 150
 - multiple access, 237
 - multiterminal, 239
 - noiseless, 248
 - noisy, 3, 146
 - noisy typewriter, **148**, 152
 - symmetric, 171
 - two-dimensional, 262
 - unknown noise level, 238
 - variable symbol durations, 256
 - with correlated sources, 236
 - with memory, 559
 - Z channel, **148**, 149, 150, 172
- cheat, 200
- Chebyshev inequality, 81, 85
- checkerboard, 406
- Chernoff bound, 85
- chess board, 522
- chi-squared, 27, 40, 325, 460
- Cholesky decomposition, 554
- chromatic aberration, 554
- cinema, 187
- circle, 318
- classical statistics, 64
 - criticisms, 32, 50, 459
- classifier, 534
- Claude Shannon, 3
- Clockville, 39
- clustering, 286, **286**, 305
- coalescence, 415
- cocked hat, 309
- code, *see* error-correcting code, source code (for data compression), symbol code, arithmetic coding, linear code, random code or hash code
 - dual, *see* error-correcting code, dual
 - for constrained channel, 249
 - variable-length, 249, 255
- code-equivalent, 578
- codebreakers, 265
- codeword, *see* source code, symbol code, or error-correcting code
- coding theory, 4, 205, 215
- coin, 38, 63
- coincidence, 267, 345, 353
- collision, 200
- coloured noise, 179
- combination, 2, 492, 600
- communication, v, 3, 16, 138, 146, 156, 162, 167, 178, 182, 186, 192, 205, 210, 215, 396, 558, 564, 598
 - broadcast, 237
 - of correlated information, 236
 - over noiseless channels, 248
 - perspective on learning, 485, 514
- competitive learning, 287
- complexity, 533, 550
- complexity control, 291, 348, 349, 351
- compress, 119
- compression, *see* source code
 - lossless, 74
 - lossy, 74, 286, 287
 - of already-compressed files, 74
 - of *any* file, 74
- computer, 372
- concatenation, 185, 214, 220
 - error-correcting codes, 16, 21, 184, 185
 - in compression, 92
 - in Markov chains, 375
- concave \curvearrowright , 35
- conditional entropy, 138, 146
- cones, 556
- confidence interval, 459, 466
- confidence level, 466
- confused gameshow host, 57
- conjugate gradient, 481
- conjugate prior, 321
- conjuror, 233
- connection between
 - channel capacity and physics, 257
 - error correcting code and latent variable model, 439
 - pattern recognition and error-correction, 483
 - supervised and unsupervised learning, 517
 - vector quantization and error-correction, 287
- connection matrix, 253
- constrained channel, 248, 257, 260, 401
- constraint satisfaction, 518
- content-addressable memory, 192, 193, 471, 507
- continuous channel, 178
- control treatment, 460
- conventions, *see* notation
 - error function, 156
 - logarithms, 2
 - matrices, 147
 - vectors, 147
- convex hull, 102
- convex \smile , 35
- convexity, 372
- convolution, 570
- convolutional code, 184, 186
- Conway, John H., 86, 522
- Copernicus, 348
- correlated sources, 237
- cost function, 180
- cost of males, 277
- counting argument, 21, 222
- coupling from the past, 415, 416, *see* Monte Carlo methods, exact sampling
- covariance, 442
- covariance function, 537
- covariance matrix, 176
- covariant algorithm, 444
- Cover, Thomas, 458
- Cox axioms, 26
- crib, 265, 268
- critical path, 246
- cross-validation, 355
- crossover, 398
- crossword, 260
- cryptanalysis, 580
- cryptography, 200
 - digital signatures, 199
 - tamper detection, 199

- cumulative probability function, 156
- cycles in graphs, 242
- cyclic, 19
- Dasher, 119
- data compression, 73, *see* source code
- data entry, 118
- data modelling, *see* modelling
- data set
 - little 'n' large, 290
- Davey, Matthew C., 571
- death penalty, 356, 357
- deciban (unit), 264
- decibel, 186
- decision theory, 348, **453**
- decoder, 4, 146, 152
 - bitwise, 220, 326
 - bounded-distance, 207
 - codeword, 220, 326
 - probability of error, 221
- degree, 570
- degree sequence, *see* profile
- degrees of belief, 26
- degrees of freedom, 324, 461
- d'ej'a vu, 121
- delay line, 577
- Delbrück, Max, 448
- deletions, 187
- delta function, 602
- density evolution, 568, 569, 594
- density modelling, 286, 305
- depth of lake, 361
- design theory, 209
- detailed balance, 393
- detection of forgery, 199
- deterministic annealing, 520
- dictionary, 72, 119
- difference-set cyclic code, 571
- differentiator, 254
- diffusion, 318
- digital cinema, 187
- digital fountain, 592
- digital signature, 199, 200
- digital video broadcast, 595
- dimer, 204
- directory, 193
- Dirichlet distribution, 318
- Dirichlet model, 117
- discriminant function, 179
- discriminative training, 554
- disease, 25, 460
- disk drive, 3, 188, 215, 248, 255
- distance, 205
 - D_{KL} , 34
 - bad, 207, 214
 - distance distribution, 206
 - entropy distance, 140
 - Gilbert–Varshamov, 212, 221
 - good, 207
 - Hamming, 206
 - isn't everything, 215
 - of code, 206, 214, 220
 - good/bad, 207
 - of code, and error probability, 221
 - of concatenated code, 214
 - of product code, 214
 - relative entropy, 34
 - very bad, 207
- distribution
 - beta, 318
 - biexponential, 315
 - binomial, 313
 - Cauchy, 88, 314
 - Dirichlet, 318
 - exponential, 313, 315
 - gamma, 315
 - Gaussian, 314
 - sample from, 314
 - inverse-cosh, 315
 - log-normal, 317
 - Luria–Delbrück, 448
 - normal, 314
 - over periodic variables, 317
 - Poisson, 175, 313, 317
 - Student-*t*, 314
 - useful, 313
 - Von Mises, 317
- divergence, 34
- DjVu, 121
- DNA, 3, 55, 201, 204, 257, 423
 - replication, 280
- do the right thing, 453
- dodecahedron code, 20, 206, 207
- dongle, 560
- doors, on game show, 57
- Dr. Bloggs, 464
- draw straws, 233
- DSC, *see* difference-set cyclic code
- dual, 216
- dumb Metropolis, 396, 498
- E_b/N_0 , 177, 178, 223
- earthquake and burglar alarm, 295
- earthquake, during game show, 57
- Ebert, Todd, 222
- edge, 251
- eigenvalue, 411
- Elias, 135
- Elias, Peter, 111, 135
- EM algorithm, 285, 434
- email, 201
- empty string, 119
- encoder, 4
- energy, 293, 403, 603
- English, 72, 110, 260
- Enigma, 265, 268
- ensemble, **67**
 - extended, 76
- ensemble learning, 431
- entropic distribution, 320
- entropic prior, 553
- entropy, 67, 603
 - Boltzmann, 85
 - conditional, 138
 - Gibbs, 85
 - joint, 138
 - marginal, 139
 - mutual information, 139
 - of continuous variable, 180
 - relative, 34
- entropy distance, 140
- epicycles, 348
- equipartition, 80
- erasure channel, 219, 591
- erasure-correction, 188, 190, 220
- erf, 156
- ergodic, 120, 375
- error bars, 303, 503
- error correction, *see* error-correcting code
 - in DNA replication, 280
 - in protein synthesis, 280
- error detection, 198, 199, 203
- error floor, 583
- error function, 156, 475, 492, 516, 531, 601
- error probability
 - block, 152
 - in compression, 74
- error-correcting code, 188, 203
 - bad, 183, 207
 - block code, 9, **151**, 183
 - concatenated, 184–186, 214
 - convolutional, 184
 - cyclic, 19
 - decoding, 184
 - density evolution, 568
 - difference-set cyclic, 571
 - distance, *see* distance of code
 - dodecahedron, 20, 206, 207
 - dual, 216, 218
 - erasure channel, 591
 - Gallager, *see* error-correcting codes, low-density parity-check
 - Golay, 209
 - good, 183, 184, 207, 214, 218
 - Hamming, 19, 214
 - in DNA replication, 280
 - in protein synthesis, 280
 - interleaving, 186
 - linear, **9**, 171, 183, 184, 229
 - low-density generator-matrix, 218, 592
 - low-density parity-check, 20, 187, 218, 559, 598
 - fast encoding, 571
 - profile, 571
 - LT code, 592
 - maximum distance separable, 220
 - nonlinear, 187
 - P_3 , 218
 - parity-check code, 220
 - pentagonful, 221
 - perfect, 208, 211, 212
 - practical, 183, 187
 - product code, 184, 214
 - quantum, 574
 - random, 184
 - random linear, 211, 212
 - rate, 152, 229
 - rateless, 592
 - rectangular, 184
 - Reed–Solomon code, 573, 591
 - repetition, 183
 - simple parity, 218

- sparse graph, 558
 - density evolution, 568
- syndrome decoding, 373
- variable rate, 238, 592
- very bad, 207
- very good, 183
- weight enumerator, 206
- with varying level of protection, 239
- error-reject curves, 535
- errors, *see* channel
- estimate, 461
- estimator, 48, 309, 322, 448
- eugenics, 273
- euro, 63
- evidence, 29, 53, 300, 324, 349, 533
 - typical behaviour of, 54, 60
- evolution, 269, 279
 - as learning, 278
 - Baldwin effect, 279
 - colour vision, 556
 - of the genetic code, 279
- evolutionary computing, 396, 397
- exact sampling, 415, *see* Monte Carlo methods
- exchange rate, 603
- exchangeability, 263
- exclusive or, 592
- EXIT chart, 569
- expectation, 27, 35, 37
- expectation propagation, 342
- expectation-maximization algorithm, 434, *see* EM algorithm
- experimental design, 465
- experimental skill, 311
- explaining away, 295, 297
- exploit, 455
- explore, 455
- exponential distribution, 45, 315
 - on integers, 313
- exponential-family, 310
- expurgation, 167, 171
- extended channel, 153, 159
- extended code, **92**
- extended ensemble, 76
- extra bit, **98**, 101
- extreme value, 448
- eye movements, 556
- factor analysis, 439, 446
- factor graph, 336–338, 436, 558, 559, 582, 585
- factorial, 2
- fading channel, 186
- feedback, 508
- female, 277
- ferromagnetic, 402
- Feynman, Richard, 424
- Fibonacci, 253
- field, 607
- file storage, 188
- finger, 119
- finite field theory, *see* Galois field
- Finland, 209
- fitness, 269, 279
- fixed point, 510
- Florida, 357
- fluctuation analysis, 448
- fluctuations, 403, 406
- focus, 531
- football pools, 209
- forensic, 423
- forgery, 199, 200
- forward pass, 244
- forward probability, 27
- forward-backward algorithm, 328, 332
- Fotherington-Thomas, 241
- Fourier transform, 88, 219, 341, 546, 570
- fovea, 556
- free energy, 257, 409, 411, 412
 - minimization, 425
 - variational, 425
- frequency, 26
- frequentist, 322, *see* sampling theory
- Frey, Brendan J., 355
- Frobenius-Perron theorem, 412
- frustration, 408
- full probabilistic model, 156
- function minimization, 475
- functions, 246
- gain, 509
- Galileo code, 186
- Gallager code, 559, *see* error-correcting codes, low-density parity-check
- Gallager, Robert, 170, 172, 187
- Galois field, 185, 607
- game, *see* puzzle
 - Bridge, 126
 - guess that tune, 204
 - guessing, 110
 - life, 522
 - sixty-three, 70
 - submarine, 71
 - three doors, 57, 60, 456
 - twenty questions, 70
- game show, 57, 456
- game-playing, 453
- gamma distribution, 315, 321
- gamma function, 600
- ganglion cells, 493
- Gaussian channel, 155, **177**
- Gaussian distribution, 2, **176**, 314, 323, 400, 551
 - N -dimensional, 124
 - approximation, 503
 - parameters, 321
 - sample from, 314
- Gaussian processes, 537
 - variational Gaussian process classifier, 549
- generalization, 485
- generalized parity-check matrix, 583
- generating function, 88
- generative model, 27, 156
- generator matrix, 9, 183
- genes, 201
- genetic algorithm, 397, 398
- genetic code, 279
- genome, 201, 280
- geometric progression, 258
- George, E.I., 395
- geostatistics, 538, 550
- $GF(q)$, 570
- Gibbs entropy, 85
- Gibbs sampling, **372**, 393, 420, *see* Monte Carlo methods
- Gibbs' inequality, 34, 37, 44
- Gilbert-Varshamov conjecture, 212
- Gilbert-Varshamov distance, 212, 221
- Gilbert-Varshamov rate, 212
- Gilks, W.R., 395
- girlie stuff, 58
- Glauber dynamics, 372, *see* Monte Carlo methods, Gibbs sampling
- Glavieux, A., 186
- Golay code, 209
- golden ratio, 253
- good, *see* error-correcting code, good
- Good, Jack, 265
- gradient descent, 478, 481, 500, 531
 - natural, 445
- graduated non-convexity, 520
- Graham, Ronald L., 175
- grain size, 180
- graph, 251
 - factor graph, 336
 - of code, 20, 558
- graphs and cycles, 242
- graphs and error-correcting codes, 19
- guessing game, 110, 111, 115
- gzip, 119
- Haldane, 278
- Hamilton, 278
- Hamiltonian Monte Carlo, **389**, 399, 498, **498**, 499
- Hamming code, 8, 9, 12, 13, 17–19, 183, 184, 190, 208, 209, 214, 219
 - graph, 19
- Hamming distance, 206
- handwritten digits, 156
- hard drive, 595
- hash code, 193, 231
- hash function, 195, 200, 228
 - linear, 231
 - one-way, 200
- hat puzzle, 222
- heat bath, 603, *see* Monte Carlo methods, Gibbs sampling
- heat capacity, 403, 406
- Hebb, Donald, 507
- Hebbian learning, 507
- Hertz, 178
- Hessian, 503
- hidden Markov models, 439
- hidden neurons, 527
- hierarchical clustering, 286
- hierarchical model, 381, 550
- high dimensions, life in, 37
- hint for computing mutual information, 149
- Hinton, Geoffrey E., 355, 431, 434, 524
- hitchhiker, 280

- homogenous, 546
- Hooke, Robert, 200
- Hopfield network, 285, **507**, 508, 519
 - capacity, 516
- Hopfield, John J., 246, 280, 519
- hot-spot, 275
- Huffman code, 91, **99**, 103
 - 'optimality', 99, 101
 - disadvantages, 100, 115
 - general alphabet, 104, 107
- human, 269
- human-machine interfaces, 119, 127
- hybrid Monte Carlo, 389, *see*
 - Hamiltonian Monte Carlo
- hydrogen bond, 280
- hyperparameter, 64, 320, 321, 381, 481
- hypersphere, 42
- hypothesis testing, *see* model
 - comparison, sampling theory
- i.i.d., 80
- ICA, *see* independent component analysis
- ICF (intrinsic correlation function), 553
- identical twin, 111
- identity matrix, 602
- ignorance, 448
- image, 551
- image analysis, 345, 353
- image compression, 74, 286
- image models, 401
- image processing, 246
- image reconstruction, 553
- implicit assumptions, 186
- implicit probabilities, **97**, 98, 102
- importance sampling, **363**, 364, 381,
 - see* Monte Carlo methods
 - weakness of, 384
- improper, 316, 318, 321, 322, 344
- improper prior, 355
- in-car navigation, 596
- independence, 138
- independent component analysis, 315, **439**, 445
- indicator function, 602
- inequality, 35, 81
- inference, 27, 531
 - and learning, 495
- inference problems
 - bent coin, 51
- information, 66
- information content, 72, 73, 91, 351
 - how to measure, 67
 - Shannon, 67
- information maximization, 445
- information retrieval, 193
- information theory, 4
- inner code, 184
- Inquisition, 348
- insertions, 187
- instantaneous, 92
- integral image, 246
- interleaving, 184, 186, 581
- internet, 188, 591
- intersection, 66, 222
- intrinsic correlation function, 551, 553
- invariance, 447
- invariant distribution, 374
- inverse probability, 27
- inverse-arithmetic-coder, 118
- inverse-cosh distribution, 315
- inverse-gamma distribution, 316
- inversion of hash function, 199
- investment portfolio, 457
- irregular, 570
- ISBN, 235
- Ising model, 130, 285, 401, 402
- iterative probabilistic decoding, 559
- Jaakkola, Tommi S., 435, 549
- Jeffreys prior, 318
- Jensen's inequality, 35, 44
- Jet Propulsion Laboratory, 186
- joint ensemble, 138
- joint entropy, 138
- joint typicality, 162
- joint typicality theorem, 163
- Jordan, Michael I., 435, 549
- journal publication policy, 465
- judge, 55
- juggling, 15
- junction tree algorithm, 342
- jury, 26, 55
- K-means clustering, **287**, **305**
 - derivation, 305
 - soft, 291
- kaboom, 308, 435
- Kalman filter, 537
- kernel, 550
- key points
 - how much data needed, 53
 - how to solve probability problems, 61
- keyboard, 119
- Kikuchi free energy, 436
- KL distance, 34
- Knowlton-Graham partitions, 175
- Knuth, Donald, xii
- Kolmogorov, Andrei Nikolaevich, 550
- Kraft inequality, **94**, 523
- Kraft, L.G., 95
- kriging, 538
- Kullback-Leibler divergence, 34, *see*
 - relative entropy
- Lagrange multiplier, 174
- lake, 361
- Langevin method, 500
- Langevin process, 537
- language model, 119
- Laplace approximation, *see* Laplace's method
- Laplace model, 117
- Laplace prior, 318
- Laplace's method, 343, 356, 498, 503, 539, 549
- Laplace's rule, 52
- latent variable, 439
- latent variable model, 285
 - compression, 355
- law of large numbers, 36
- lawyer, 55, 58, 61
- Le Cun, Yann, 121
- leaf, 338
- leapfrog algorithm, 391
- learning, 473
 - as communication, 485
 - as inference, 494, 495
 - Hebbian, 507
 - in evolution, 278
- learning algorithms, 470
 - backpropagation, 530
 - Boltzmann machine, 524
 - classification, 477
 - competitive learning, 287
 - Hopfield network, 507
 - K-means clustering, **288**, **291**, 305
 - multilayer perceptron, 530
 - single neuron, 477
- learning rule, 472
- Lempel-Ziv coding, 110, 119-122
 - criticisms, 128
- life, 522
- life in high dimensions, 37, 124
- likelihood, 6, 28, 49, 152, 326, 531, 560
 - contrasted with probability, 28
 - subjectivity, 30
- likelihood equivalence, 449
- likelihood principle, 32, 61, 466
- limit cycle, 510
- linear block code, 9
- linear feedback shift register, 184
- linear regression, 344, 529
- Litsyn, Simon, 574
- little 'n' large data set, 290
- log-normal, 317
- logarithms, 2
- logit, 318
- long thin strip, 411
- loopy, 342, 558
- loopy belief propagation, 436
- loopy message-passing, 340
- lossy compression, 168, 286, 287
- low-density generator-matrix code, 207, 592
- low-density parity-check code, 558, **559**
 - staircase, 571
- LT code, 592
- Luby, Michael G., 570, 592
- Luria, Salvador, 448
- Lyapunov function, 289, 293, **510**, 522, 523
- Mézard, Marc, 342
- macho, 321
- MacKay, David, 187, 498
- magician, 233
- magnetic recording, 595
- majority vote, 5
- male, 277
- Mandelbrot, Benoit, 262
- MAP, *see* maximum a posteriori
- MAP decoding, 327
- mapping, 92

- marginal entropy, 139
- marginal likelihood, 29, 300, 324, *see* evidence
- marginalization, 29, 297, 321
- Markov chain, 141, 168
- Markov chain Monte Carlo, *see* Monte Carlo methods
- Markov model, 111, *see* Markov chain and hidden Markov model
- marriage, 456
- matrix, 411
- matrix identities, 440
- max-product, 341
- maxent, 310, *see* maximum entropy
- maximum distance separable, 219, *see* error-correcting code, maximum distance separable
- maximum entropy, 310, 553
- maximum likelihood, 6, 302, 349
- maximum likelihood decoder, 152
- maximum *a posteriori* decoder, 327
- maximum *a posteriori*, 6, 309, 327, 540
- MCMC (Markov chain Monte Carlo), *see* Monte Carlo methods
- McMillan, B., 95
- MD5, 200
- MDL, *see* minimum description length
- MDS, 220, *see* error-correcting code, maximum distance separable
- mean, 1
- mean field theory, 424, 427
- melody, 201, 203
- memory, 470
 - address-based, 470
 - associative, 470, 507
 - content-addressable, 192, 471
- MemSys, 553
- message passing, 187, 241, 248, 285, 326, 409, 593
 - BCJR, 332
 - belief propagation, 332
 - forward-backward, 332
 - in graphs with cycles, 340
 - loopy, 340
 - sum-product algorithm, 338
 - Viterbi, 331
- metacode, 104, 108
- metric, 514
- Metropolis method, 498, *see* Monte Carlo methods
- micro-saccades, 556
- microsoftus, 460
- microwave oven, 127
- min-sum algorithm, 245, 327, 331, 580, 583
- mine (hole in ground), 453
- minimax, 457
- minimization, 475
- minimum description length, 354, **354**
- minimum distance, 206, 214, *see* distance of code
- Minka, Thomas, 342
- mirror, 531
- Mitzenmacher, Michael, 570
- mixing coefficients, 300, 314
- mixture
 - in Markov chains, 375
- mixture distribution, 375
- mixture modelling, 284, 286, 305, 439
- mixture of Gaussians, 314
- MLP, *see* multilayer perceptron
- MML, *see* minimum description length
- mobile phone, 186
- model
 - latent variable, 439
- model comparison, 198, 348, 349, 351
 - typical behaviour of evidence, 54, 60
- modelling, 287
 - density modelling, 286, 305
- models of images, 526
- moderation, 29, 500, *see* marginalization
- molecules, 201
- Molesworth, 241
- momentum, 389, 481
- Monte Carlo methods, 359, 500
 - acceptance rate, 396
 - acceptance ratio method, 381
 - annealed importance sampling, 381
 - coalescence, 415
 - dependence on dimensionality, 360
 - exact sampling, 415
 - for visualization, 553
 - Gibbs sampling, **372**, 393, 420
 - Hamiltonian Monte Carlo, 389, 498
 - hybrid Monte Carlo, *see* Hamiltonian Monte Carlo
 - importance sampling, **363**, 381
 - weakness of, 384
 - information communication in, 396
 - Langevin method, 500
 - Markov chain Monte Carlo, **367**, 368
 - Metropolis method
 - dumb Metropolis, 396, 498
 - Metropolis-Hastings, **367**
 - multi-state, 394, 397, 400
 - overrelaxation
 - ordered, 393
 - perfect simulation, 415
 - random walk suppression, 372
 - random-walk Metropolis, 390
 - rejection sampling, **366**
 - adaptive, 372
 - reversible jump, 381
 - simulated annealing, 381, 394
 - thermodynamic integration, 381
 - umbrella sampling, 381
- Monty Hall problem, 57
- Morse, 256
- motorcycle, 110
- movie, 553
- multilayer perceptron, 531, 537
- multiple access channel, 237
- multiterminal networks, 239
- multivariate Gaussian, 176
- Munro-Robbins theorem, 443
- murder, 26, 58, 61, 356
- music, 201, 203
- mutation rate, 448
- mutual information, 139, 146, 150, 151
 - how to compute, 149
- myth, 349
 - compression, 74
- nat (unit), 264, 603
- natural gradient, 445
- natural selection, 269
- navigation, 596
- Neal, Radford, 111, 121, 187, 376, 381, 393, 394, 399, 421, 422, 431, 434, 498
- needle, Buffon's, 38
- network, 531
- neural network, 470, 472
 - capacity, 485
 - learning as communication, 485
 - learning as inference, 494
- neuron, 473
 - capacity, 485
- Newton algorithm, 443
- Newton, Isaac, 200, 554
- Newton-Raphson, 305
- nines, 198
- noise, 3, *see* channel
 - coloured, 179
 - spectral density, 177
 - white, 179
- noisy channel, *see* channel Gaussian, 177
- noisy typewriter, **148**, 152, 154
- noisy-channel coding theorem, **15**, 152, 162, 171, 229
 - poor man's version, 216
- noisy-or, 296
- non-confusable inputs, 152
- noninformative, 321
- nonlinear, 537
- nonlinear code, 187
- nonparametric data modelling, 540
- nonrecursive, 577
- noodle, Buffon's, 38
- normal, 314, *see* Gaussian
- normal graph, 219, 586
- normalizing constant, *see* partition function
- not-sum, 337
- notation, 600, *see* conventions
 - absolute value, 33, 601
 - conventions of this book, 147
 - expectation, 37
 - intervals, 90
 - logarithms, 2
 - matrices, 147
 - set size, 33, 601
 - vectors, 147
- NP-complete, **184**, 327, 519

- nucleotide, 201, 204
- nuisance parameters, 321
- numerology, 208

- objective function, 475
- Occam factor, 324, 347, **350**, 352, 354
- Occam's razor, **345**
- octal, **577**
- octave, 480
- Ode to Joy, 203
- Oliver, 56
- one-way hash function, 200
- optic nerve, 493
- optimal decoder, 152
- optimal input distribution, **150**, 162
- optimal linear filter, 551
- optimal stopping, 456
- optimization, 169, 394, 431, 481, 507, 518, 533
 - gradient descent, 478
 - Newton algorithm, 443
 - of model complexity, 533
- ordered overrelaxation, 393
- orthodox statistics, 322, *see* sampling theory
- outer code, 184
- overfitting, 308, 324, 531
- overrelaxation, 392
 - ordered, 393

- p*-value, 64, 459, 464
- packet, 188
- paradox, 107
 - Alias, 456
 - bus-stop, 39
 - heat capacity and fluctuations, 403
- paramagnetic, *see* Ising model
- paranormal, 233
- parasite, 278
- parent, 561
- parity, 9
- parity-check bits, 9, 203
- parity-check constraints, 20
- parity-check matrix, 12, 229
 - generalized, 583
- parity-check nodes, 19, 219, 569, 570, 585
- parse, 119, 450
- parthenogenesis, 273
- partial order, 420
- partial partition functions, 409
- particle filter, 398
- partition, 174
- partition function, 403, 409, 411, 424, 425, 603, 605
 - analogy with lake, 362
- partitioned inverse, 545
- Pasco, Richard, 111
- pattern recognition, 156, 179, 201
- pentagonful code, 21, 221
- perfect code, 208, 210, 211, 219, 591
- perfect simulation, 415, *see* Monte Carlo methods, exact sampling
- periodic variable, 317

- permutation, 19, 268
- phase transition, 363, 603
- philosophy, 26, 119, 386
- phone, 596
- phone directory, 193
- phone number, 58, 129
- photon counter, 309, 344, 450
- physics, 85
- pigeon-hole, 575
- pigeon-hole principle, 86
- pitchfork bifurcation, 293
- plaintext, 265
- plankton, 361
- point estimate, 434
- point spread function, 551
- pointer, 119
- Poisson distribution, 2, 175, 309, 313, 344
- Poisson process, 39, 46, 450
- Poissonville, 39, 315
- polymer, 257
- poor man's coding theorem, 216
- porridge, 280
- positive definite, 541
- positivity, 553
- posterior probability, 6, 152
- power cost, 180
- power law, 586
- practical, 183, *see* error-correcting code, practical
- precision, **176**, 181, 314, 322, 385
- prediction, 29, 52
- predictive distribution, 111
- prefix code, **92**, 95
- prior, 6, 310, 531
 - assigning, 310
 - improper, 355
 - subjectivity, 30
- prior equivalence, 449
- priority of bits in a message, 239
- prize, on game show, 57
- probabilistic model, 111, 120
- probabilistic movie, 553
- probability, 26, 38
 - Bayesian, 50
 - contrasted with likelihood, 28
- probability distributions, 313, *see* distribution
- probability of block error, 152
- probability propagation, *see* sum-product algorithm
- product code, 184, 214
- profile, of random graph, 570
- pronunciation, 34
- proper, 541
- proposal density, **366**, 367
- Propp, Jim G., 415, 420
- prosecutor's fallacy, 25
- prospecting, 453
- protein, 201, 204
 - regulatory, 201, 204
- protein synthesis, 280
- protocol, 591
- pseudoinverse, 552
- Punch, 450

- puncturing, 222, 582
- puzzle, *see* game
 - cable labelling, 173
 - chessboard, 522
 - fidelity of DNA replication, 280
 - hat, 222, 223
 - life, 522
 - magic trick, 233, 234
 - poisoned glass, 103
 - southeast, 522
 - transatlantic cable, 173
 - weighing 12 balls, 68

- quantum error-correction, 574
- queue, 456
- QWERTY, 119

- R₃, *see* repetition code
- race, 356
- radial basis function, 537, 538
- RAID, 188, 190, 219
- random, 359
- random code, 156, 161, 164, 165, 184, 192, 195, 214, 567
 - for compression, 231
- random variable, 26, 465
- random walk, 369
 - suppression, 372
- random-coding exponent, 171
- random-walk Metropolis method, 390
- rant, *see* sermon
 - confidence level, 467
 - p*-value, 465
- Raptor codes, 596
- rate, **152**
- rate-distortion theory, 167
- reading aloud, 531
- receiver operating characteristic, 535
- recognition, 204
- record breaking, 448
- rectangular code, 184
- reducible, 375
- redundancy, 4, 33
 - in channel code, 146
- redundant array of independent disks, 188, 190
- redundant constraints in code, 20
- Reed-Solomon code, 185, 186, 573, 591
- regression, 344, 538
- regret, 457
- regular, 559
- regularization, 531, 552
- regularization constant, 481
- Reinforcement learning, 455
- rejection, 366, 368, 535
- rejection sampling, **366**, *see* Monte Carlo methods
- relative entropy, 34, 98, 102, 142, 424, 431, 437, 477
- reliability function, 171
- repeat-accumulate code
 - connection to low-density parity-check code, 589
- repetition code, 5, 13, 15, 16, 46, 183
- responsibility, 291

- retransmission, 591
- reverse, 110
- reversible jump, 381
- Richardson, Thomas J., 572, 597
- Rissanen, Jorma, 111
- Roberts, Gareth O., 395
- ROC, 535
- roman, 127
- rule of thumb, 382
- runlength, 256
- runlength-limited channel, 249

- saccades, 556
- saddle-point approximation, 343
- sample, 314
 - from Gaussian, 314
- sampler density, 364
- sampling distribution, 461
- sampling theory, 38, 322
 - criticisms, 32
- sandwiching method, 421
- satellite, 596
- satellite communications, 186
- scaling, 203
- Schönberg, 203
- Schottky anomaly, 406
- secret, 200
- secretary problem, 456
- security, 199, 201
- seek time, 595
- Sejnowski, Terry J., 524
- self-delimiting, 132
- self-dual, 218
- self-orthogonal, 218
- self-punctuating, 92
- sequence, 346
- sequential decoding, 583
- sequential probability ratio test, 466
- sermon, *see* caution
 - classical statistics, 64
 - confidence level, 467
 - gradient descent, 443
 - importance sampling, 384
 - interleaving, 189
 - MAP method, 285
 - maximum entropy, 310
 - maximum likelihood, 308
 - maximum *a posteriori* method, 308
 - most probable is atypical, 285
 - p*-value, 465
 - sampling theory, 64
 - sphere-packing, 209, 212
 - stopping rule, 465
 - turbo codes, 583
 - unbiased estimator, 309
 - worst-case-ism, 207
- set, 66
- Shannon, *see* noisy-channel coding
 - theorem, source coding
 - theorem, information content
- shannon (unit), 265
- Shannon information content, 67, 91, 115
- Shannon, Claude, 14, 15, 152, 164, 212, 215, 262
- Shevelev, Vladimir, 574
- shifter ensemble, 526
- Shokrollahi, M. Amin, 570
- shortening, 222
- Siegel, Paul, 262
- sigmoid, 475, 529
- signal to noise ratio, 177
- significance, 465
- significance level, 51, 64, 459
- simplex, 173, 318
- Simpson's paradox, 357
- Simpson, O.J., *see* wife-beaters
- simulated annealing, 381, 394, *see* annealing
- Skilling, John, 394
- Slepian-Wolf, *see* correlated sources
- slice sampling, 376, *see* Monte Carlo methods
 - multi-dimensional, 380
- soft K-means clustering, 291
- softmax, 318
- softmax, softmin, 291, 341
- software, xi
 - arithmetic coding, 121
 - BUGS, 373
 - Dasher, 119
 - free, xii
 - Gaussian processes, 536
 - hash function, 200
 - VIBES, 433
- solar system, 348
- soliton distribution, 594
- sound, 187
- source code, 73, 75
 - algorithms, 119, 121
 - block code, 76
 - block-sorting compression, 121
 - Burrows-Wheeler transform, 121
 - for complex sources, 355
 - for constrained channel, 249
 - for integers, 132
 - Huffman, *see* Huffman code
 - implicit probabilities, 102
 - optimal lengths, 97, 102
 - prefix code, 95
 - software, *see* software
 - stream codes, 110–130
 - supermarket, 96, 104, 112
 - symbol code, 91
 - optimal, 91
 - uniquely decodeable, 94
 - variable symbol durations, 125, 256
- source coding theorem, 78, 91, 229, 231
- southeast puzzle, 522
- span, 333
- space graph
 - profile, 571
- sparse graph code, 558
 - density evolution, 568
- sparsifiers, 255
- species, 269
- speculation about vision, 556
- spell, 201
- sphere packing, 205
- sphere-packing exponent, 172
- Spielman, Daniel A., 570
- spin system, 402
- spines, 526
- spline, 540
- spring, 293
- spy, 466
- square, 38
- staircase, 571, 589
- stalactite, 214
- standard deviation, 322
- stars, 309
- state diagram, 251
- statistic, 460
 - sufficient, 302
- statistical physics, 257, 403
- statistical test, 51, 460
- steepest descents, 443
- stereoscopic vision, 526
- stiffness, 291
- Stirling's approximation, 1, 8
- stochastic, 474
- stochastic dynamics, *see* Hamiltonian
 - Monte Carlo
- stochastic gradient, 478
- stop-when-it's-done, 563, 585
- stopping rule, 465
- straws, drawing, 233
- stream codes, 110–130
- student, 125
- Student-*t* distribution, 314, 325
- subjective probability, 26, 30
- submarine, 71
- subscriber, 595
- subset, 66
- substring, 119
- sufficient statistics, 302
- sum rule, 39, 46
- sum-product algorithm, 187, 245, 328, 336, **338**, 409, 436, 558, 559, 574, 580
- summary, 337
- summary state, 421
- summation convention, 440
- super-channel, 184
- supermarket for codewords, 96, 104, 112
- support vector, 550
- surprise value, 264
- survey propagation, 342
- suspicious coincidences, 353
- symbol code, **91**
 - budget, 94, **96**
 - codeword, **92**
 - disadvantages, 100
 - optimal, 91
 - self-delimiting, 132
 - supermarket, 112
- symmetric channel, 171
- symmetry argument, 151
- synchronization, 249
- synchronization errors, 187

- syndrome, 10, 11, 20
- syndrome decoding, 11, 216, 229, 373
- system, 4
- systematic, 577

- t*-distribution, *see* Student-*t*
- tail, 85, 314, 315, 442, 448, 505, 586
- tamper detection, 199
- Tank, David W., 519
- Tanner product code, 573
- Tanner, Michael, 571
- Tanzanite, 453
- tap, 577
- telephone, 125
- telephone directory, 193
- telephone number, 58, 129
- telescope, 531
- temperature, 394, 603
- termination, 581
- terminology, 600
 - Monte Carlo methods, 374
- test
 - fluctuation, 448
 - statistical, 51, 460
- text entry, 118, 119
- thermal distribution, 88
- thermodynamic integration, 381
- thermodynamics, 406, 603
- Thiele, T.N., 550
- thin shell, 37, 125
- Thitimajshima, P., 186
- three cards, 142
- three doors, 57
- threshold, 569
- tiling, 422
- time-division, 237
- timing, 187
- training data, 531
- transatlantic, 173
- transfer matrix method, 409
- transition, 251
- transition probability matrix, 147, 609
- translation-invariant, 411
- travelling salesman problem, 246, 519
- tree, 242, 338, 345, 353
- trellis, 251
 - termination, 581
- trellis section, 251, 257
- triangle, 309
- truth function, 211, 602
- tube, 257
- turbo code, 186, 558
- turbo product code, 573
- Turing, Alan, 265
- twenty questions, 70, 103
- twin, 111
- twos, 156
- typical set, **80**, 154, 365
 - for compression, 80
 - for noisy channel, 154
- typical set decoder, 165, 230
- typicality, **78**, 80, 162
 - behaviour of evidence, 54, 60

- umbrella sampling, 381
- unbiased estimator, 309, 323, 451
- uncompression, 231
- union, 66
- union bound, 216
- uniquely decodeable, **93**, 94
- units, 264
- universal, 110, 120, 121, 135
- universality, 402
- Urbanke, Rüdiger, 572, 597
- urn, 31
- user interfaces, 118
- utility, 453

- vaccination, 460
- Vapnik–Chervonenkis dimension, 491
- variable-length code, 249, 255
- variable-rate error-correcting codes, 238, 592
- variance, 1, 27, 88, 323
- variance-covariance matrix, 176
- variational Bayes, 431
- variational free energy, 424, 425
 - minimization, 425
- variational methods, 424, 435, 498, 510
 - typical properties, 437
 - variational Gaussian process classifier, 549
- VC dimension, 491
- vector quantization, 286, 292

- very good, *see* error-correcting code, very good
- VGC (variational Gaussian process classifier), 549
- VIBES, 433
- Virtakallio, Juhani, 209
- Viterbi algorithm, 245, 331, 342, 580
- volume, 42
- Von Mises distribution, 317

- Wainwright, Martin, 342
- waiting for a bus, 39, 46
- warning, *see* caution
- Watson–Crick base pairing, 280
- weighing babies, 164
- weighing problem, 66, 68
- weight
 - importance sampling, 364
 - in neural net, 473
 - of binary vector, 20
- weight decay, 481, 531
- weight enumerator, 206, 211, 214, 216
 - typical, 574
- weight space, 475, 476, 489
- Wenglish, 72, 260
- what number comes next?, 346
- white, 357
- Wiberg, Niclas, 187
- Wiener process, 537
- Wiener, Norbert, 550
- wife-beater, 58, 61
- Wilson, David B., 415, 420
- window, 309
- Winfrey, Erik, 522
- Wolf, Jack, 262
- word-English, 260
- world record, 448
- worst-case-ism, 207, 213
- writing, 118

- Yedidia, Jonathan, 342

- Z channel, **148**, 149–151, 155
- Zecchina, R., 342
- Zipf plot, 319
- Zipf’s law, 40
- Zipf, George K., 262