
Index

- $\binom{n}{m}$, binomial coefficient, 26
 μ , mean, 17
 σ , standard deviation, 18
- Akaike information criterion, 282
 analysis of variance, 122
 ANOVA, 122
- Bayes, 188–257, 421–427
 conjugate priors, 198
 evidence, 191
 hyperparameters, 191
 likelihood, 189
 MAP, 191
 prior, 190
 rule, 7, 188
 uninformative prior, 205–211
 updating, 197
- Bayesian belief networks, 246
 Bayesian information criterion, 282
 Bayesian optimisation, 356
 belief networks, 246
 Benford's law, 207, 255, 426
 Bennett's inequality, 158, 176, 414
 Bernoulli distribution, 446
 Bernoulli trial, 60
 Bernstein's inequality, 158, 176, 417
 Berry–Esseen bound, 66
 beta distribution, 34, 198, 447
 beta function, 34, 43
 bias-variance dilemma, 22, 394
 binomial distribution, 26, 41, 61–66, 397, 446
 Boltzmann distribution, 177, 290, 420, 429
 Boole's inequality, 165
 Box–Muller method, 54
 branching process, 295–297, 307, 431
- categorical distribution, 251
 Cauchy distribution, 35, 107–109, 408, 447
 Cauchy–Schwarz inequality, 142, 176, 414
 central limit theorem, 30, 66, 81–89, 408
- central moment, 64
 channel capacity, 262
 Chapman–Kolmogorov equation, 356
 characteristic function, 63, 107–109
 Chebyshev's inequality, 149
 chemical reactions, 385–388
 master equation, 387
 Chernoff bound, 157–159
 chi-squared distribution, 103, 411, 447
 clustering, 282
 compressed sensing, 282
 concentration theorem, 152
 conjugate priors, 198–205
 constrained optimisation, 182–186
 convergence, 133
 almost sure, 135
 in distribution, 134
 in moments, 135
 in probability, 134
 correlation, *see* Pearson correlation
 covariance, 18
 covariance matrix, 36
 Cramér–Rao bound, 287
 cumulant, 18, 64–66
 generating function, 64, 72, 176, 295, 401–402, 419
 normal distribution, 78
 cumulative distribution function, 144
 binomial distribution, 405
 normal distribution, 90–92, 103
 Poisson distribution, 73, 403
 curve fitting, 278
- de Finetti's representation theorem, 20
 defective
 matrix, 348
 degrees of freedom, 128
 detailed balance, 318
 deviate, 46, 47
 Cauchy, 58, 399
 exponential, 52
 linear-congruential, 50

- multinomial, 56
- multivariate normal, 56, 99
- normal, 53, 55, 346, 436
 - rejection sampling, 54
 - transformation method, 51
- diffusion process, 359–380
- digamma function, 43
- Dirac delta function, 103–107
- Dirichlet distribution, 37, 255, 421, 448
- disordered systems, 304–306
- distribution
 - continuous, 30–35
 - beta, 34, 447
 - Cauchy, 35, 58, 107–109, 399, 408, 447
 - chi-squared, 32, 103, 411, 447
 - exponential, 32, 52, 447
 - gamma, 31, 41, 146, 204, 398, 405, 447
 - log-normal, 85, 144, 146, 366, 389, 442, 447
 - normal, 30, 53, 55, 58, 75–204, 290, 399, 429, 447
 - Student's *T*, 126, 130, 255, 426, 447
 - uniform, 447
 - Weibull, 33, 398, 447
 - discrete, 26–27, 60–71
 - Bernoulli, 60, 446
 - binomial, 26, 41, 61–66, 397, 446
 - geometric, 58, 72, 400, 401, 446
 - hypergeometric, 27, 41, 397–398, 446
 - negative binomial, 72, 402, 446
 - Poisson, 28, 73, 144, 202, 403, 446
 - long tailed, 17, 35
 - multivariate, 35–37
 - Boltzmann, 420
 - categorical, 68, 251
 - Dirichlet, 37, 255, 421, 448
 - multinomial, 35, 68–69, 72, 202, 255, 406, 421, 448
 - normal, 37, 96, 103, 204–448
 - Wishart, 204, 448
- dual problem, 185
- earth-mover's distance, 170
- eigenvalue, 347
- eigenvector, 347
- EM algorithm, 229
- entropy, 258–290, 427–430
 - Fisher's information, 285–289
 - joint, 260
 - Kolmogorov's complexity, 276–285
 - Shannon's entropy, 259–276
- ergodic
 - stochastic matrix, 316
- error bars, 113
- error function, 90
- error in the mean, 113
- estimator, 19
 - consistent, 61
 - efficient, 287
 - maximum likelihood, 60
 - unbiased, 61, 111
- events, 3
- expectation, 15
- expectation-maximisation, 229
- exponential distribution, 32, 52, 72, 403, 447
- exponential family, 38, 205, 289, 413
- extreme value distributions, 93–95
- F*-test, 121
- filtering, 330–343
 - Kalman, 332–337, 347, 439
 - particle, 338–343, 347, 439
- first-passage time, 345, 433
- Fisher's *F*-test, 121
- Fisher's information, 285–289
- Fisher, R. A., 285, 376
- Fokker–Planck, 370
- functional, 290–292
- gamma distribution, 41, 146, 202, 398, 405, 447
- gamma function, 31, 41–43
- Gaussian distribution, *see* normal distribution
- Gaussian process, 351–356, 389, 441
- generating function, 62
 - characteristic function, 63, 107–109
 - cumulant, 64, 72, 295, 401–402, 419
 - moment, 62
- generative adversarial networks, 174
- generative model, 251
- geometric distribution, 58, 72, 400, 401, 446
- geometric mean, 164
- Gillespie algorithm, 387, 390, 443
- graph generation models, 297
- graphical models, 245–254
- Heaviside function, 105
- hidden Markov model (HMM), 235
- hierarchical model, 214–215
- Hinton, G. E., 243
- Hoeffding's inequality, 154–157
- hybrid Monte Carlo, 329
- hypergeometric distribution, 27, 41, 397–398, 446
- hyperparameter, 204, 214
- iid, 46
- improper distribution, 206
- improper priors, 211

- independence, 8
 - conditional, 10
 - pairwise, 10
- inequalities, 142–182
 - AM–GM, 165
 - Bennett's, 158, 176, 414
 - Bernstein's, 158, 176, 417
 - Cauchy–Schwarz, 142, 176, 414
 - Chebyshev's, 149
 - Chernoff bound, 157–159
 - Cramér–Rao bound, 287
 - Gibbs, 167
 - Hoeffding's, 154–157
 - Jensen's, 159, 176, 177, 418
 - Markov's, 143, 144, 150
 - Shepp's, 150
 - union bound, 165
- inference, 19
- information geometry, 289
- information theory, 259
- Ising model, 299–302, 346, 420, 436
- Itô
 - calculus, 363
 - isometry, 363
 - lemma, 364
- Jacobian, 14
- Jaynes, E. T., 188, 206, 209, 256, 259
- Jeffreys, H., 207
- Jensen's inequality, 159, 176, 177, 418
- Kálmán, R. E., 332
- Kalman filter, 332–337, 347, 439
- Kelly's strategy, 89
- KL divergence, *see* Kullback–Leibler divergence
- Kolmogorov complexity, 276–285
- Kolmogorov, A. N., 4, 23, 276
- Kramer–Moyal's expansion, 372–374
- Kullback–Leibler divergence, 166–170, 177, 221, 421
- kurtosis, 79
- Lagrange multipliers, 182–185
- Langevin equation, 361
- large deviation, 152
- latent Dirichlet allocation, 249–253
- latent variable models, 227–242
- law of large numbers, 138–139
- LeCun, Y., 243
- Legendre transform, 153
- location parameter, 206
- log-normal distribution, 15, 85, 144, 146, 305, 366, 389, 442, 447
- MAP solution, 215
- marginalisation, 21
 - multivariate normal, 100
- Markov chain, 308–347, 432–441
 - absorbing state, 315
 - ergodic, 316
 - higher-order, 309
 - rapid mixing, 327
- Markov Chain Monte Carlo, 317–330, 346, 436–438
- Markov chain Monte Carlo
 - detailed balance, 318
 - heat bath, 319
 - hybrid Monte Carlo, 329
 - Metropolis, 319
- Markov process, 356–359
- Markov random fields, 246
- Markov's inequality, 143, 144, 150
- martingale, 139–142
- master equation, 357, 372–374, 387
- maximum entropy, 267
- maximum likelihood, 60, 191
- Maxwell's demon, 275
- MCMC, *see* Markov Chain Monte Carlo
- mean, 17
- mean field approximation, 222
- measure theory, 23–24
- median, 17, 176, 418
- minimum description length, 277–285
- mode, 17
- model selection, 211–214, 277
- moment, 62, 63, 78
 - central, 64
- moment generating function, 62
- Monte Carlo, 45–57
 - integration, 47
 - MCMC
 - detailed balance, 318
 - heat bath, 319
 - hybrid Monte Carlo, 329
 - Metropolis, 319
 - Metropolis–Hastings, 324
 - rejection method, 54
 - transformation method, 51
- multinomial distribution, 35, 68–69, 73, 202, 255, 406, 421, 448
- multivariate normal distribution, 37, 96–103, 204–205, 409, 448
- negative binomial distribution, 72, 402, 446
- negatively correlated random variables, 179
- networks, 297
- non-informative priors, *see* uninformative prior
- normal distribution, 30, 58, 75–102, 144, 201–204, 290, 399, 429, 447
- null hypothesis, 117

Index

457

- optimal stopping theorem, 140
 Ornstein–Uhlenbeck process, 360, 372
 outlier, 17

 particle filter, 338–343, 347, 439
 partition function, 268, 300
 Pearson’s correlation, 19, 176, 414
 percolation, 297–299, 306
 Perron–Frobenius theorem, 316
 phase transition, 298
 point process, 380–388
 Poisson distribution, 28, 67, 73, 144, 202, 403, 446
 Poisson process, 380–385
 Poissonisation, 69
 positive definite matrix, 97
 prior
 improper, 211
 uninformative, 205–211
 probability
 conditional probability, 6
 cumulative distribution function, 12
 density function, 13
 joint probability, 6
 mass function, 12
 process
 diffusion, 359–380
 Gaussian, 351–356, 389, 441
 Markov, 356–359
 Ornstein–Uhlenbeck, 360, 372
 point, 380–388
 Poisson, 380–385
 Wiener, 360
 properties
 stochastic matrix, 313
 pseudo-counts, 255, 424, 425

 quasi-random numbers, 49
 queues, 384–385

 Rademacher variables, 60
 random deviate, 46
 random variable, 11
 random walk, 294–295, 307, 359, 430
 recommender system, 216
 reflecting boundary conditions, 377
 regression, 122, 278
 regulariser, 282
 robust statistic, 17

 scale parameter, 207
 Schur complement, 101–102
 second law of thermodynamics, 272–275
 second-order statistics, 18
 self-averaging, 305

 self-organised criticality, 302–304
 Shannon’s entropy, 259–276
 Shannon, C. E., 259
 simplex
 discrete, 36
 unit, 36
 simulated annealing, 323
 skewness, 79
 standard deviation, 18
 stationarity, 201
 stationary
 Markov chain, 311
 statistical significance, 116
 statistics, 17
 central moment, 64
 cumulant, 64–66
 mean, 17
 moment, 62
 standard deviation, 18
 variance, 17
 Stirling’s approximation, 42
 stochastic matrix, 310
 ergodic, 316
 stock price, 365, 389
 Stratonovich’s calculus, 363
 Student’s T distribution, 130, 255, 426, 447
 sub-Gaussian random variables, 154
 sufficient statistic, 40, 61, 124

 t-test, 117–121, 126, 411
 tail bound, 152–159
 Taylor expansion, 177–179
 topic model, 249

 unbiased estimator, 61
 uniform distribution, 447
 uninformative prior, 205–211, 255–256, 426–427
 union bound, 165
 universality, 302

 variance, 17
 variate, 46
 variational approximation, 220
 variational auto-encoder, 240–242, 283
 variational free energy, 221

 Wasserstein distance, 170–175
 Weibull distribution, 33, 398, 447
 Wiener
 measure, 360
 process, 360
 Wishart distribution, 204, 448
 Woodbury matrix identity, 101, 334